This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

Sufficient and Necessary Graphical Conditions for MISO Identification in Networks with Observational Data

Sina Jahandari, Donatello Materassi

***Abstract*—This article addresses the problem of consistently identifying a single transfer function in a network of dynamic systems using only observational data. It is assumed that the topology is partially known, the forcing inputs are not measured, and that only a subset of the nodes outputs is accessible. The developed technique is applicable to scenarios encompassing confounding variables and feedback loops, which are complicating factors potentially introducing bias in the estimate of the transfer function. The results are based on the prediction of the output node using the input node along with a set of additional auxiliary variables which are selected only from the observed nodes. Similar prediction error methods provide only sufficient conditions for the appropriate choice of auxiliary variables and assume a priori information about the location of strictly causal operators in the network. In this article, such an a priori knowledge is not required. A most remarkable feature of our approach is that the conditions for the selection of the auxiliary variables are purely graphical. Furthermore, within single-output prediction methods such conditions are proven to be necessary and sufficient to consistently identify all networks with a given topology. A fundamental consequence of this characterization is to enable the search of a set of auxiliary variables minimizing a suitable cost function for single-output prediction error identification. In the article, we suggest possible approaches to tackle such optimal identification problems.**

I. Introduction

Designing a suitable input to be injected into a system in order to identify its dynamics is a common strategy in identification theory [1]. Apart from injecting known inputs [2], networked systems offer other options of active intervention to facilitate the identification process. For example, standard approaches involve removing certain edges and/or knocking out specific nodes [3], [4].

However, in several applications the network cannot be actively manipulated and data are merely observational. Namely, the network measurements are usually acquired while the system is responding to excitations that are not necessarily known [5], [6].

A wealth of methodologies has also been developed to deal with the problem of identifying a network of dynamic systems from observational data. These methods rely on different a priori assumptions and have different identification goals. On one hand, some techniques have as primary goal the recovery of the unknown network graph [7]–[9], while the quantitative identification of the network dynamics (i.e. transfer functions) might only be a complementary outcome [10]. On the other hand, some techniques assume that the underlying graph is at least partially known and the very objective becomes to identify the transfer functions describing the dynamics coupling the nodes [11]–[16].

The results of this article fall into the category of techniques aiming at identifying an individual transfer function in a network where the graph topology is partially known. A defining feature of our techniques is that they rely on conditions that are purely graphical and are inspired by the theory of probabilistic graphical models of random variables. The main advantage of graphical model techniques is that they tend to be particularly suited to deal with confounding variables. However, graphical models are typically defined on directed acyclic graphs. Thus, they might not be considered an adequate model to describe scenarios involving feedback loops which are instead central in the theory of automatic control. Conversely, the problem of determining a transfer function involved in a feedback loop within a network is an active topic of research in identification theory. In [14] classical closed-loop prediction error techniques such as direct, two-stage, and joint-input-output methods are extended to be applicable in local network identification settings. Improvements upon the same general ideas were presented in [17] using tools from graph theory, in [18] incorporating Bayesian/kernel methods and in [19] to deal with sensor noise. Furthermore, the possibility of parametric identification strategies based on instrumental variable methods was explored in [20]. As a recent development some closed loop identification techniques [21]–[24] have incorporated graphical conditions to effectively deal with confounding variables and offering the opportunity to create connections with the theory of graphical models.

A typical limitation of prediction error methods is that some information about the locations of strictly causal transfer functions needs to be available. This knowledge is typically formalized by requiring that there is no algebraic loop for any value of the parameters in the full network parameterization [17].

In this respect, our first contribution can be interpreted as an attempt to extend certain graphical model tools to deal with the problem of closed loop identification combining the best of the two worlds: an effective way to take into account the unknown locations of strictly causal transfer functions while obtaining an unbiased closed loop identification in presence

Sina Jahandari is with the Electrical and Computer Engineering department, University of Minnesota, Twin Cities, USA jahan046@umn.edu

Donatello Materassi is with the Electrical and Computer Engineering department, University of Minnesota, Twin Cities, USA mater013@umn.edu

of confounding variables. In particular, our identification procedure falls in the class of prediction error methods, while the selection of auxiliary variables borrows elements from the theory of graphical models.

The result is a technique which guarantees the consistent identification of a transfer function in a partially observed network by selecting auxiliary predictors using only graphical conditions. Unlike other works our technique is capable of detecting the location of strictly causal transfer functions directly from data. Remarkably, the derived graphical conditions are also proven to be necessary providing a complete characterization of the sets of auxiliary variables that lead to a consistent identification in single-output error prediction methods. This characterization is the basis for the search for a set of auxiliary variables minimizing a suitable cost function for the identification.

The article is organized as follows. Section II reviews preliminary definitions and concepts of dynamic networks, their graphical representations and introduces the concept of pointing separation. In Section III, sufficient graphical conditions for consistent identification in causal networks are presented. In Section IV, two data-driven tests are proposed to detect strictly causal transfer functions and presence of feedthroughs in dynamic networks. Section V shows that the condition proposed in this paper are necessary and casts a formal optimal identification framework for dynamic networks. Section VI provides a numerical example quantitatively verifying the results of the paper. Concluding remarks are given in Section VII.

NOMENCLATURE

$\hat{x}_j(t)$	Estimate of $x_j(t)$
$\perp\!\!\!\perp$	Independence
$\mathrm{an}_G(j)$	The set of ancestors of node j in graph G
$\mathrm{ch}_G(j)$	The set of children of node j in graph G
$\mathrm{de}_G(j)$	The set of descendents of node j in graph G
$\mathrm{pa}_G(j)$	The set of parents of node j in graph G
D^+	Set of nodes used in prediction up to time t
D^-	Set of nodes used in prediction up to time $t-1$
G	The graphical representation of network $\mathscr{G}$
$G^{\not\,t}$	Graph of instantaneous propagationss
$I_A(t)$	Natural filtration generated by x_A up to time t
O	Set of measurable nodes
$W_{ji}(z)$	The component of the Wiener filter corresponding to $x_i(t)$ when estimating $x_j(t)$
$x_j(t)$	The output of node j at time t
Z	Set of auxiliary predictors

II. PRELIMINARIES

In this section, we introduce the class of models that is going to be the object of our investigation along with some preliminary concepts and notions from the area of graphical models.

Similar network models have been considered and investigated in [12], [24]–[26].

Definition 1. *A network $\mathscr{G}$ is a pair $(H(z), n)$ where $H(z)$ is a proper rational discrete-time $v \times v$ transfer matrix and n is a vector of v mutually independent stochastic processes with rational power spectral density. The output signals of the network are defined by the relation*

$$x_j(t) = n_j(t) + \sum_{i \in V} H_{ji}(z) x_i(t), \qquad \text{for } j = 1, \ldots, v \tag{1}$$

Using a vector notation and defining $V = \{1, \ldots, v\}$ we can represent the model in a more compact way as

$$x_V(t) = n_V(t) + H(z) x_V(t) \tag{2}$$

The dynamics of the matrix $H(z)$ in model (2) indicates how the process x_i directly affects the process x_j. If $H_{ji}(z) = 0$ there is no direct effect of x_i on x_j (even though x_i could still affect x_j indirectly through other processes). For this reason models described by (2) lend themselves to be represented via graphs. We assume that the reader is already familiar with basic notions of graph theory [27] and in this section we just introduce our notation and nomenclature. For a directed graph G, defined by the pair (V, E) where $V = \{1, 2, \cdots, v\}$ is the set of nodes and $E \subseteq V \times V$ is the set of edges, we denote an edge $(i, j) \in E$ as $i \to j$ or $j \leftarrow i$ and say that the edge is oriented from i to j. We also say that two distinct edges $i \to j$ and $k \to \ell$ are adjacent if they share at least a node, namely $\{i, j\} \cap \{k, \ell\} \neq \emptyset$. In a directed graph, a path between i and j is a sequence of distinct edges such that the first edge contains i, the last edge contains j and each two consecutive edges in the sequence are adjacent. A path can be suggestively denoted by using the arrow symbols ($\to$ and $\leftarrow$) to separate the nodes involved in the path while at the same time representing the orientation of the edges. For example, in the graph of Figure 1,

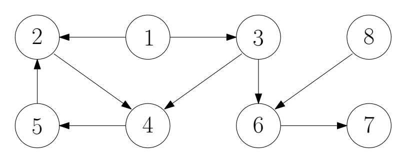


Fig. 1. Representation of a directed graph.

there are four paths between nodes 3 and 5 which can be denoted as $\{3 \to 4 \to 5\}, \{3 \to 4 \leftarrow 2 \leftarrow 5\}, \{3 \leftarrow 1 \to 2 \to 4 \to 5\}, \{3 \leftarrow 1 \to 2 \leftarrow 5\}$. Furthermore, if the edges have all the same orientation (as in $\{1 \to 3 \to 4 \to 5\}$) the path is called a dipath or a chain. For a graph G, we also recall the following relations among its nodes

- node j is a child of node i if the edge $i \to j$ is present in the graph. We also say that i is a parent of j. We denote the set containing all children of node j by $\mathrm{ch}_G(j) = \{v \in V | j \to v \in E\}$ and the set containing all parents of node i by $\mathrm{pa}_G(i) = \{v \in V | v \to i \in E\}$. Moreover for a set $A \subseteq V$ we define $\mathrm{ch}_G(A) = \cup_{j \in A} \mathrm{ch}_G(j)$ and $\mathrm{pa}_G(A) = \cup_{i \in A} \mathrm{pa}_G(i)$.
- node j is a descendant of node i if $j = i$ or if there is a dipath from i to j. Equivalently, we say that i is an ancestor of j. We denote the set containing all descendants of node i as $\mathrm{de}_G(i)$ and the set containing all ancestors of node i as $\mathrm{an}_G(i)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

For the description of model (1) we are going to use a special instance of multi-typed graphs [28] which are an extension of standard directed graphs.

Definition 2. *A multi-arrowed graph is a triple $G=(V,E_1,E_2)$ where E_1, the set of single-headed edges, and E_2, the set of double-headed edges, are disjoint subsets of $V\times V$.*

We represent a multi-arrowed graph in the same way we represent a standard graph but we draw a single-headed edge to represent $i\to j\in E_1$ and a double-headed edge to represent $i\to j\in E_2$. Note that multi-arrowed graphs generalize directed graphs and all the graphical notions extend naturally to multi-arrowed graphs, as well, by simply considering $E_1\cup E_2$ as a set of standard edges. For example, vertex i is a parent of j whether there is a single-headed or double-headed edge from i to j. If $i\to j\in E_1$ we say that i is a single-headed parent of j, while if $i\to j\in E_2$ we say that i is a double-headed parent of j.

Definition 3. *We say that the multi-arrowed graph $G=(V,E_1,E_2)$ is recursive if in every directed loop there is at least one double-headed edge.*

We can use multi-arrowed graphs to describe the sparsity pattern of $H(z)$ in model (1) along with some information about the strict causality of the entries in $H(z)$.

Definition 4. *Let $\mathscr{G}=(H,n)$ be a network with output processes x_V, where $V:=\{1,\dots,v\}$, and let E_1 and E_2 be two disjoint subsets of $V\times V$ such that*

(a) $i\to j\notin E_1\cup E_2$ *implies* $H_{ji}(z)=0$
(b) $i\to j\notin E_1$ *implies* $H_{ji}(z)$ *is strictly causal.*

We say that the multi-arrowed graph $G=(V,E_1,E_2)$ is a graphical representation of the network. Furthermore, if the implications (a) and (b) hold also in the opposite direction, we say that $G=(V,E_1,E_2)$ is a perfect graphical representation of the network.

In other words, the absence of the edge $i\to j$ in a graphical representation implies that $H_{ji}(z)=0$ while the presence of a double-headed edge $i\to j$ implies that $H_{ji}(z)$ is strictly causal (potentially zero). Thus, a network can have different graphical representations each providing different degrees of information on its dynamics, as the following example illustrates.

Example 1. *Consider a dynamic network $\mathscr{G}=(H(z),n)$ with four nodes governed by the following equations.*

$$\begin{aligned}x_1(t)&=n_1(t)+H_{12}(z)x_2(t)+H_{14}(z)x_4(t)\\x_2(t)&=n_2(t)\\x_3(t)&=n_3(t)+H_{31}(z)x_1(t)\\x_4(t)&=n_4(t)+H_{43}(z)x_3(t).\end{aligned}$$

The nonzero entries of $H(z)$ in this network are

$$H_{12}(z)=\frac{z}{z+\frac{1}{2}},\qquad H_{14}(z)=\frac{1}{z+\frac{1}{2}},$$
$$H_{31}(z)=\frac{1}{z^2},\qquad H_{43}(z)=\frac{1}{2}.$$

Figure 2 (a) shows the perfect graphical representation G^p of the network $\mathscr{G}$. The information that transfer functions $H_{31}(z)$ and $H_{14}(z)$ are strictly causal and $H_{23}(z)=0$ is available in G^p. Figure 2 (b) shows a recursive graphical representation G of the network $\mathscr{G}$. Unlike G^p, the information that transfer function $H_{14}(z)$ is strictly causal or that $H_{23}(z)=0$ is not available from G.

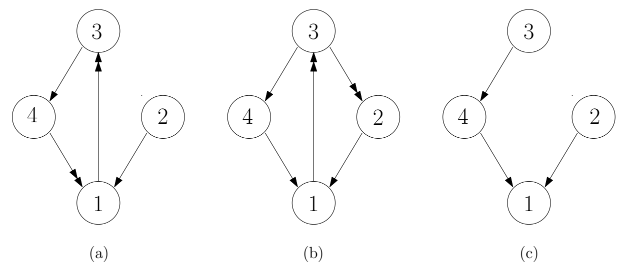


Fig. 2. (a) The perfect graphical representation G^p of the dynamic network $\mathscr{G}$ discussed in Example 1; (b) A recursive graphical representation G of $\mathscr{G}$; (c) The graph of instantaneous propagations $G^{\not{t}}$ associated with G.

Observe that a graphical representation provides partial information about the network's topology. Indeed, given a graphical representation of a network it is always possible to obtain another less informative graphical representation by introducing additional single-headed or double-headed edges and/or replacing a double-headed edge with a single-headed one.

From its definition if a network has recursive graphical representation then it has no algebraic loops. The following definition of graph of instantaneous propagations is an important tool to deal with the presence of direct feed-throughs.

Definition 5. *Consider a multi-headed graph $G=(V,E_1,E_2)$. Its associated graph of instantaneous propagations, denoted as $G^{\not{t}}$, is the standard directed graph (V,E_1) obtained from G by removing the double-headed edges.*

For example, Figure 2 (c) shows the graph of instantaneous propagations $G^{\not{t}}$ associated with graph G depicted in Figure 2 (b). It is an immediate consequence of the definition that if G is recursive, $G^{\not{t}}$ is a directed acyclic graph. We refer to $G^{\not{t}}$ as the graph of instantaneous propagations because, if G is a graphical representation, new information entering a node k at time t can potentially propagate to all nodes in $\mathrm{de}_{G^{\not{t}}}(k)$ at the same time t with no delay.

Throughout the paper, we will sometimes refer to nodes, edges, paths and chains of a network even though, formally, we should refer to them as nodes, edges, paths and chains of its graphical representation or its perfect directed graph.

As shown in [24], there is a strong relationship between signal estimators and graphical representations in a network. Such a relationship will play a central role in the development of our results. For this reason we recall some fundamental notions from estimation theory and introduce our notation.

Definition 6. *Given a probability space, for a set of stochastic processes x_A where $A\subseteq V$, we denote the natural filtration generated by the processes x_A up to time t as $I_A(t)$.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

In this article we typically consider the estimate $\hat{x}_j(t)$ of $x_j(t)$ using the information of processes x_{D^+} up to time t and the information of processes x_{D^-} up to time $t-1$. Using the notation introduced in Definition 6 the least square estimate $\hat{x}_j(t)$ can be written as

$$\hat{x}_j(t) = \mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)). \tag{3}$$

In the linear Gaussian case this estimation problem can be solved via Wiener filters, reducing (3) to

$$\hat{x}_j = \sum_{k \in D^+} W_{jk}(z)x_k + \sum_{k \in D^-} W_{jk}(z)x_k \tag{4}$$

where $W_{jk}(z)$ for $k \in D^+$ are causal transfer functions and for $k \in D^-$ are strictly causal transfer functions. So long as the power spectral density matrix of the signals x_j, x_{D^+} and x_{D^-} is the same, the expressions of the Wiener filter components are the same when considering a least square estimation even in the linear non-Gaussian case. In the following, we assume for simplicity that all the processes are jointly Gaussian even though the same results can be easily shown to hold in the linear non-Gaussian case, as well.

The sparsity properties of the Wiener filters $W_{jk}(z)$ are connected to the notion of conditional independence.

Definition 7. *We say that $x_j(t)$ and the information of x_i up to time t are conditionally independent given $I_{D^+}(t)$ and $I_{D^-}(t-1)$ if*

$$\mathbb{E}(x_j(t) \mid I_{\{i\}\cup D^+}(t), I_{D^-}(t-1)) = \mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)). \tag{5}$$

We denote this by $x_j(t) \perp\!\!\!\perp I_i(t) \mid I_{D^+}(t), I_{D^-}(t-1)$. Similarly, if $\mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{\{i\}\cup D^-}(t-1)) = \mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1))$ we say that $x_j(t)$ and the information of x_i up to time $t-1$ are conditionally independent which we denote by $x_j(t) \perp\!\!\!\perp I_i(t-1) \mid I_{D^+}(t), I_{D^-}(t-1)$.

In the linear Gaussian case, using a Wiener filter formulation, the estimate $\hat{x}_j(t)$ of $x_j(t)$ from the processes x_{D^+} up to time t, the processes x_{D^-} up to time $t-1$ and the process x_i can be expressed as

$$\hat{x}_j = W_{ji}(z)x_i + \sum_{k \in D^+} W_{jk}(z)x_k + \sum_{k \in D^-} W_{jk}(z)x_k \tag{6}$$

where $W_{jk}(z)$ are causal for $k \in D^+$, strictly causal for $k \in D^-$ and $W_{ji}(z)$ is causal if the information of x_i is used up to time t and strictly causal if the information of x_i is used up to time $t-1$. The relation of conditional independence between $x_j(t)$ with $I_i(t)$ (or analogously $I_i(t-1)$) translates into having $W_{ji}(z) = 0$ in Equation (6).

In the theory of graphical models the internal nodes of a path are classified as forks, colliders or chain links.

Definition 8. *Given a path π in a graph G we say that a node j is*

- *a fork, when there exist two consecutive edges in the path of the form $i \leftarrow j$ and $j \rightarrow k$*
- *a collider (or an inverted fork), when there exist two consecutive edges in the path of the form $i \rightarrow j$ and $j \leftarrow k$*
- *a chain link, when there exist two consecutive edges in the path of the form $i \rightarrow j$ and $j \rightarrow k$*

Specifically, the notion of colliders allows one to define if a path π is blocked by a set Z.

Definition 9. *In a directed graph G, a path π between nodes i and j is blocked by a set of nodes Z if*

- *there is at least a non-collider on π that belongs to Z; or*
- *there is at least a collider c on π such that $de_G(c) \cap Z = \emptyset$.*

Otherwise, we say that the path π is activated by Z.

In the theory of graphical models, a fundamental concept defined over the nodes of a directed graph is d-separation [29].

Definition 10. *In a directed graph $G = (V, E)$ let A, B, and C be disjoint subsets of V. A and B are d-separated by C if for all nodes $a \in A$ and $b \in B$, all paths between a and b are blocked by C. If A and B are not d-separated by C in G, we say that they are d-connected by C in G.*

Example 2. *In the directed graph depicted in Fig. 1 $A = \{1\}$ and $B = \{8\}$ are d-separated by $C = \emptyset$ because 6 is a collider on a path from 1 to 8. For the same reason, $A = \{1\}$ and $B = \{8\}$ become d-connected if we choose $C = \{6\}$ and also if we choose $C = \{7\}$ or $C = \{6,7\}$. Again, $A = \{2\}$ and $B = \{6\}$ are d-connected by $C = \emptyset$ because of the path $2 \leftarrow 1 \rightarrow 3 \rightarrow 6$. If we consider $C = \{1\}$ to "block" such a path, $A = \{2\}$ and $B = \{6\}$ are still d-connected because of the other path $2 \leftarrow 5 \leftarrow 4 \leftarrow 3 \rightarrow 6$. If we now consider $C = \{1,4\}$ to "block" this other path, $A = \{2\}$ and $B = \{6\}$ are still d-connected because now 4 is a collider in C on the path $2 \rightarrow 4 \leftarrow 3 \rightarrow 6$. By choosing $C = \{1,3,4\}$, we make $A = \{2\}$ and $B = \{6\}$ d-separated. Alternatively, $C = \{3\}$ would have been a smaller set making $A = \{2\}$ and $B = \{6\}$ d-separated.*

In [24], some criteria for consistent identification are derived using the notion of d-separation. This article obtains more powerful criteria by using a weaker notion of separation that involves only a subset of the paths between the nodes i and j.

Definition 11. *A path π between nodes i and j is called j-pointing if the last edge in the path π is of the form $k \rightarrow j$ for some node k. If all the j-pointing paths between nodes i and j, with the exception of the path constituted by only the edge $i \rightarrow j$, are blocked by a set of nodes Z, we say that i and j are j-pointing separated.*

Note that a j-pointing path between i and j might or might not be i-pointing.

III. Consistent Identification in Networks

Consider a simple two-node system with x_i as input, x_j as output, and n_j as additive output error independent of x_i, namely

$$x_j(t) = n_j(t) + H_{ji}(z)x_i(t),$$

for some causal transfer function $H_{ji}(z)$. The block diagram of this system is depicted in Figure 3 (a), while a graphical representation of the system is shown in Figure 3 (b). Following [30], a possible technique to identify $H_{ji}(z)$ is to compute a linear least square prediction $\hat{x}_j(t)$ for $x_j(t)$ by using the past

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

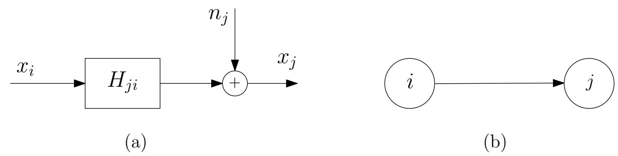


Fig. 3. (a) Block diagram of a two-node network; (b) Corresponding graphical representation.

information of x_j and the past and present information of x_i. Namely, we compute

$$\hat{x}_j(t) = \mathbb{E}[x_j(t) \mid I_i(t), I_j(t-1)] = W_{jj}(z)x_j(t) + W_{ji}(z)x_i(t),$$

where $W_{jj}(z)$ is strictly causal and $W_{ji}(z)$ is causal. After computing $\hat{x}_j(t)$, we calculate the quantity

$$\hat{H}_{ji}(z) = \frac{W_{ji}(z)}{1 - W_{jj}(z)}$$

which can be proven to be a consistent estimate for $H_{ji}(z)$. In other words, in this simple two-node system, the transfer function $H_{ji}(z)$ can be consistently identified via the following procedure by setting $D^- = \{j\}$ and $D^+ = \{i\}$.

Procedure 1 Identification via prediction error

Given: Sets of nodes D^-, D^+ and $i, j \in D^- \cup D^+$
Output: $\hat{H}_{ji}(z)$

$$\mathbb{E}(x_j(t) \mid I_{D^-}(t-1),\ I_{D^+}(t)) = \sum_{k \in D^- \cup D^+} W_{jk}(z)x_k(t)$$

$\hat{H}_{ji}(z) = \frac{W_{ji}(z)}{1 - W_{jj}(z)}$

We stress that Procedure 1 is not an algorithm but more precisely a meta-algorithm since the computation of $\hat{x}_j(t)$ could be obtained using a variety of methods either parametric or non-parametric [30]. In Procedure 1, the estimate of $x_j(t)$ is obtained using the information from the past for the variables in D^- and information from the past and present for the variables in D^+. In other words, the transfer functions $W_{jk}(z)$ are strictly causal if $k \in D^-$ and causal if $k \in D^+$.

However, when dealing with more complex networks, applying Procedure 1 with $D^- = \{j\}$ and $D^+ = \{i\}$, leads, in general, to an estimate $\hat{H}_{ji}(z)$ for $H_{ji}(z)$ which is not consistent because of the presence of feedback loops or because other variables in the network might act as confounders between i and j.

In several recent results, it has been shown that, by appropriately introducing additional measured variables to the sets of predictors D^- and D^+, Procedure 1 (or substantially equivalent tools such as the methods in [20]) can still achieve a consistent estimate of $H_{ji}(z)$.

This idea has been explored in the extension of closed-loop identification techniques to network identification [17], [23] and by applying graphical model tools [24]. The drawback of using graphical model techniques is that they tend to be limited to acyclic networks. In particular, the related results in [24] are not as powerful when the target node j is involved in a directed feedback loop. The existing closed-loop identification techniques, on the other hand, can successfully deal with loops, but require information about the presence of direct feed-throughs. However, all these methodologies basically try to solve or are applicable to the following problem [2], [17], [19], [23], [24].

Problem 1. *Consider a network $\mathcal{G} = (H(z), n)$ with a known graphical representation $G = (V, E_1, E_2)$. Suppose that the forcing inputs n are unknown and that a subset $O \subseteq V$ of the node outputs is observable with $i, j \in O$. Find sets of predictors D^- and D^+ with $\{i, j\} \subseteq D^- \cup D^+ \subseteq O$ such that Procedure 1 guarantees a consistent identification of the transfer function $H_{ji}(z)$.*

A unifying feature of most of these approaches is to exploit additional measurements (apart from i and j) as auxiliary predictors. More specifically, apart from the nodes i and j, these methods require an extra set of nodes Z to be observable and a way of partitioning $Z \cup \{i, j\}$ into the sets $D^- \cup D^+$ for Procedure 1 to consistently estimate $H_{ji}(z)$.

A first contribution of this article is a solution to Problem 1 that can be interpreted as an attempt to combine graphical model and closed loop identification methods in order to effectively deal with confounding variables and feedback loops. Namely, we provide conditions, of purely graphical nature, to determine the set Z of auxiliary predictors along with a way to partition $Z \cup \{i, j\}$ into the sets D^- and D^+ in order to guarantee that Procedure 1 obtains a consistent identification.

Furthermore, we prove that such graphical conditions on Z are also necessary for consistency given the known graphical representation G. Having sufficient and necessary conditions for the set of auxiliary predictors enables the search for an optimal Z that provides a consistent identification while at the same time minimizing an assigned cost function to select the auxiliary variables.

A standard assumption when dealing with the problem of identifying a module in a dynamic network is the absence of algebraic loops. Furthermore, most identification methods also need to include among their assumptions some a priori information about the location of the strictly causal transfer functions in each loop [2], [14], [18]–[20]. In this article we still keep the assumption that the network has no algebraic loops, but, as an important distinction from other methods, we also reduce the need of a priori information about the locations of strictly causal transfer functions. Indeed, as we later show in Section IV, we provide some methods to infer the locations of strictly causal transfer functions directly from data. In the derivation of the result for this section, we temporarily assume that such information is obtained and available in the form of a recursive graphical representation.

A. Sufficient conditions for consistent identification

As a first observation, if all the parents of the target node j are available, then it is possible to consistently identify the transfer function $H_{ji}(z)$ when some knowledge about the delays of the transfer functions $H_{jk}(z)$, $k \in \mathrm{pa}_G(j)$ is available from a recursive graphical representation G of the network.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

Proposition III.1. *Consider a network $\mathscr{G}$ with no algebraic loops. Let $G = (V, E_1, E_2)$ be a recursive graphical representation of $\mathscr{G}$ and P^+ and P^- be respectively the sets of single-headed parents and double-headed parents of the node j in G.*

Then

$$\mathbb{E}(x_j(t)|I_{\{j\}\cup P^-}(t-1), I_{P^+}(t)) = W_{jj}(z)x_j(t) + \sum_{k\in P^-\cup P^+} [1 - W_{jj}(z)]H_{jk}(z)x_k(t). \quad (7)$$

Proof. See the appendix. □

A consequence of Proposition III.1 is the following corollary.

Corollary III.1.1. *Under the assumptions of Proposition III.1, let the power spectral density matrix of x_i, x_j and $x_{P^+\cup P^-}$ be non-singular. The application of Procedure 1 with $D^- = P^- \cup \{j\}$ and $D^+ = P^+$ leads to a consistent estimate of $H_{ji}(z)$.*

Proof. See the appendix. □

Corollary III.1.1 is quite intuitive and can be interpreted in terms of the results in [17]. We use Corollary III.1.1 as a starting point for the derivation of our results and also to illustrate how simple manipulations on a graphical representation can lead to different selections of the sets D^+ and D^- in Procedure 1, as done in the following example.

Example 3. *Consider a network with a graphical representation G depicted in Figure 4 (a). The objective is the identification of the transfer function $H_{21}(z)$. Since G is recursive and all the parents of node 2 are measured, we can apply Corollary III.1.1. Node 1 is a single-headed parent of node 2, and node 4 is a double-headed parent of node 2. Therefore, the application of Corollary III.1.1 for $j = 2$ and $i = 1$ in the graphical representation G, leads to the choice of $D^- = \{4, 2\}$ and $D^+ = \{1\}$ in Procedure 1 which gives a consistent estimate of $H_{21}(z)$. However, if the graph G of Figure 4 (a) is a graphical representation of the network under study, so must be the graph G' depicted in Figure 4(b). The difference between the two graphical representations is that the information that $H_{24}(z)$ is strictly causal is available in G, but is not available in G'. Since G' is also recursive, we can still apply Corollary III.1.1 to it. When applied to G', Corollary III.1.1 leads to a different choice for the sets D^- and D^+ (namely $D^- = \{2\}$ and $D^+ = \{1, 4\}$) which also provides a consistent estimate of $H_{21}(z)$ via Procedure 1. Furthermore, if the graph G of Figure 4 (a) is a graphical representation of the network under study, so must be the graph G'' depicted in Figure 4(c). There are a few differences between G and G''. It can be seen from G that $H_{32}(z)$ is strictly causal, and $H_{26}(z) = H_{63}(z) = 0$. This information is not available in G''. Since G'' is also recursive, we can still apply Corollary III.1.1 to it. When applied to G'', Corollary III.1.1 leads to yet a different choice for the sets D^- and D^+ (namely $D^- = \{2, 4\}$ and $D^+ = \{1, 6\}$) which also provides a consistent estimate of $H_{21}(z)$ via Procedure 1.*

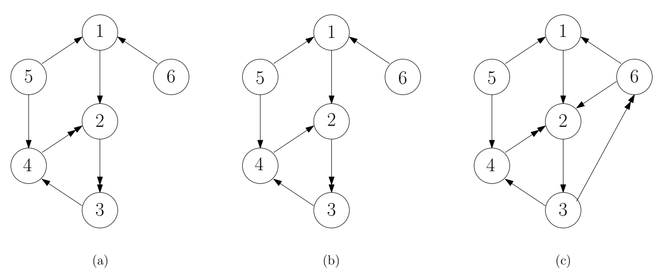


Fig. 4. The networks discussed in Example 3.

This example shows that if a recursive graphical representation G is available, we can still apply Corollary III.1.1 to a less informative graphical representation which can be obtained by introducing additional edges in G or by replacing double-headed edges in G with single-headed edges. The only requirement for the application of Corollary III.1.1 is that such a less informative graphical representation has still to be recursive.

Observe that in Proposition III.1 we have $D^- \cup D^+ = \text{pa}_G(j) \cup \{j\}$. Hence, Proposition III.1 substantially states that Procedure 1 can consistently identify the transfer function $H_{ji}(z)$, but, in order to do so, it requires the observation of all parents of j in a given graphical representation. In some scenarios, assuming that all parents of j are being observed might be overly restrictive, since missing information from some parents of the target node j does not necessarily hinder the consistent identification of $H_{ji}(z)$. Indeed, information from other observed nodes can be exploited to compensate the missing information from the unmeasured parents, so that a consistent identification of $H_{ji}(z)$ can still be achieved. The first main contribution of this article is the following result providing a criterion to appropriately select the sets D^- and D^+ in Procedure 1 to guarantee an unbiased estimate of $H_{ji}(z)$ under significantly more general conditions than Corollary III.1.1.

Theorem III.2. *Consider a network $\mathscr{G} = (H(z), n)$ with recursive graphical representation $G = (V, E_1, E_2)$. Let $Z \cap \{i, j\} = \emptyset$ be a set such that*

(i) Z is j-pointing separating the nodes i and j in G; and
(ii) $Z \cup \{i\}$ blocks all j-pointing paths from j to itself G.

Let $G^{\sharp}$ be the graph of instantaneous propagations associated to G and let D^- and D^+ be the following two disjoint sets partitioning $Z \cup \{i, j\}$

- $D^+ := an_{G^{\sharp}}(j) \cap (Z \cup \{i\})$
- $D^- := (Z \cup \{i, j\}) \setminus D^+$

The application of Procedure 1 with D^- and D^+ leads to a consistent estimate of $H_{ji}(z)$ when the power spectral density matrix of (x_i, x_j, x_Z) is non-singular.

Proof. See the appendix. □

Theorem III.2 presents a systematic procedure for selecting two sets of predictors D^- and D^+ to identify a specific transfer function $H_{ji}(z)$ via Procedure 1. Observe that the fact that the graphical representation is recursive allows one to determine D^+ and D^- in a unique way. The expressions for D^+ and D^- in Theorem III.2 state that we always have $j \in D^-$ and for any $k \in \{i\} \cup Z$ if there is no delay from k to j, given the information by the recursive graphical representation, we have $k \in D^+$ and D^- contains the remaining variables. It is possible to apply Theorem III.2 also in presence of feedback loops and/or confounding variables affecting the nodes i and j. The following example illustrates how Theorem III.2 can successfully deal with unobserved confounding variables.

Example 4. *Consider the network with a recursive graphical representation shown in Figure 5. Suppose the objective is to identify the transfer function $H_{21}(z)$. Assume that nodes 3 and 4 which are depicted with a dashed line are not measured $3,4 \notin O = \{1,2,5,6,7\}$. Note that Proposition III.1 could not be applied because node 3 is a parent of the target node 2 but is not observable. Instead, we could search for a set $Z \subseteq O$ that satisfies conditions (i) and (ii) of Theorem III.2. The 2-pointing path $\{1 \leftarrow 5 \leftarrow 3 \rightarrow 2\}$ needs to be blocked and since $3 \notin O$ we need to have $5 \in Z$. However, when node 5 is observed, it becomes an activated collider in the path $\{1 \leftarrow 4 \rightarrow 7 \rightarrow 5 \leftarrow 3 \rightarrow 2\}$. To block $\{1 \leftarrow 4 \rightarrow 7 \rightarrow 5 \leftarrow 3 \rightarrow 2\}$ we need to have $7 \in Z$. Since all the 2-pointing path from 2 to itself are also blocked by $\{5,7\} \cup \{1\}$, we obtain that $Z = \{5,7\}$ is a subset of $O = \{1,2,5,6,7\}$ that satisfies conditions (i) and (ii) of Theorem III.2. Thus, applying Procedure 1 with $D^- = \{2\}$ and $D^+ = \{1,5,7\}$ leads to a consistent estimate of $H_{21}(z)$.*

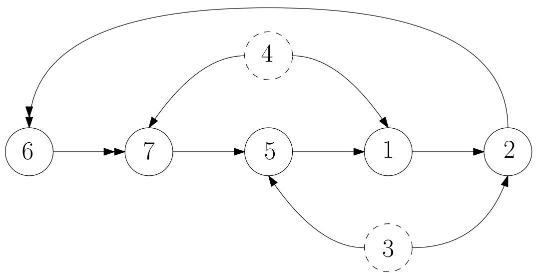


Fig. 5. The graphical representation of the network discussed in Example 4. The nodes 3 and 4 act as confounders.

Observe that, by setting $Z = \mathrm{pa}_G(j)$, Theorem III.2 becomes equivalent to Corollary III.1.1. One main advantage of Theorem III.2 is that it can successfully deal with confounding variables in a way similar to the formulation of the Single Door Criterion for dynamic systems [24], [31]. However, contrary to the Single Door Criterion, Theorem III.2 can be easily applied to networks where the node j is involved in feedback loops. Furthermore, the graphical conditions on the nodes i and j for the application of Single Door Criterion are stronger than the graphical condition required for Theorem III.2. Indeed, Single Door Criterion needs all the paths between i and j to be blocked by a set Z that does not contain descendants of j. Conversely, Theorem III.2 only needs Z to block the j-pointing ones and Z also can contain descendants of j.

Note that the choice of predictors is not unique since multiple sets Z might satisfy the conditions of Theorem III.2. It might also happen that none of the predictors sets Z satisfying the conditions of Theorem III.2 are contained in the set of observable nodes O, namely, $Z \nsubseteq O$. In addition to this, for a fixed Z a further degree of flexibility can be obtained as follows: so long as another recursive graphical representation can be obtained from the graphical representation G of the network, different choices of the sets D^- and D^+ are also possible. The following example computes all the possible sets Z satisfying conditions (i) and (ii) of Theorem III.2.

Example 5. *Consider a network with a graphical representation G depicted in Figure 6. The objective is the identification of the transfer function $H_{21}(z)$. Node 4 is not observable. Node 3 should be measured since it is the only choice to block the 2-loop $\{2 \rightarrow 3 \rightarrow 2\}$ and the 2-pointing path $1 \rightarrow 3 \rightarrow 2$. Now, since $3 \in Z$ node 3 will act as an activated collider on the 2-pointing path $\pi_1 = \{1 \rightarrow 3 \leftarrow 5 \leftarrow 4 \rightarrow 2\}$. Since 4 is not measured, the only choice for blocking π_1 is to have $5 \in Z$. Since $5 \in Z$ node 5 will act as an activated collider on the 2-pointing paths $\pi_2 = \{1 \leftarrow 7 \rightarrow 6 \rightarrow 5 \leftarrow 4 \rightarrow 2\}$ and $\pi_3 = \{1 \leftarrow 8 \rightarrow 6 \rightarrow 5 \leftarrow 4 \rightarrow 2\}$. To block π_2 and π_3 we need to measure either node 6 or nodes 7 and 8 together. Table 5, lists all choices for Z that satisfy conditions (i) and (ii) of Theorem III.2 and their corresponding D^- and D^+. In Section V we will also show that the choices listed in Table I are the only possible choices for Z guaranteeing a consistent identification of $H_{21}(z)$ using Procedure 1 for all networks with graphical representation G.*

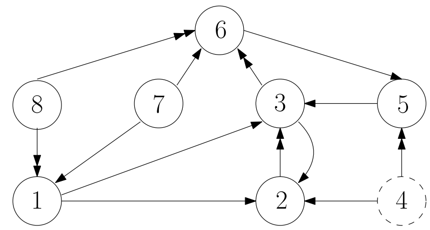


Fig. 6. The graphical representation of the network discussed in Example 5.

TABLE I
ALL POSSIBLE PREDICTOR SETS TO CONSISTENTLY IDENTIFIY $H_{21}(z)$ IN EXAMPLE 5 FOR THE GIVEN GRAPHICAL REPRESENTATION

	Z	D^-	D^+
1	$\{3,5,6\}$	$\{2\}$	$\{1,3,5,6\}$
2	$\{3,5,7,8\}$	$\{2,8\}$	$\{1,3,5,7\}$
3	$\{3,5,6,7\}$	$\{2\}$	$\{1,3,5,6,7\}$
4	$\{3,5,6,8\}$	$\{2,8\}$	$\{1,3,5,6\}$
5	$\{3,5,6,7,8\}$	$\{2,8\}$	$\{1,3,5,6,7\}$

IV. DETECTING DELAYS IN A DYNAMIC NETWORK

While deriving the results of the previous section, it was assumed that some partial knowledge about causality or strict causality of the transfer functions of the network was available. This knowledge was embedded in the fact that a recursive graphical representation of the network had to be a priori available.

If such a recursive graphical representation is not available, the application of Procedure 1 to the sets D^- and D^+ given by Theorem III.2 leads, in the general case, to an inconsistent estimate of $H_{ji}(z)$. Other network identification methods require

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

analogous information about the locations of strictly causal transfer functions. This knowledge is typically formalized by requiring that there is no algebraic loop for any value of the parameters in the full network parameterization [17], which substantially implies the knowledge of a recursive graphical representation.

However, in some cases a recursive graphical representation might not be a priori available. In this section, we provide sufficient criteria to obtain a recursive graphical representation in a network with no algebraic loops, from a known graphical representation which is not necessarily recursive. These results have the main advantage of widening not only the applicability of Theorem III.2 but also the applicability of other network identification methods such as the ones described in [17], [21].

This first result gives a sufficient criterion to determine if a transfer function is strictly causal directly from observational data.

Theorem IV.1. *Consider a network with no algebraic loops and with (non-necessarily recursive) graphical representation* $G=(V,E_1,E_2)$ *and* $i\in pa_G(j)$. *Let* $Z\cap\{i,j\}=\emptyset$ *be a set that j-pointing separates nodes i and j in G. Let* A,Z^-,Z^+ *be a partition of Z such that*

- $Z^- := \{\ell\in Z : \ell\notin an_{G^\sharp}(j)\}$
- $Z^+ := \{k\in Z : k\notin de_{G^\sharp}(j)\}\setminus Z^-$
- $A := Z\setminus(Z^-\cup Z^+)$.

If

$$\lim_{z\to\infty} W_{ji}(z)=0 \tag{8}$$

in

$$\mathbb{E}(x_j(t)\mid I_{\{j\}\cup Z^-\cup A^-}(t-1), I_{Z^+\cup A^+\cup\{i\}}(t)) = \sum_{r\in Z^-\cup Z^+\cup A^-\cup A^+\cup\{i,j\}} W_{jr}(z)x_r(t). \tag{9}$$

for all possible combinations of disjoint A^- *and* A^+ *with* $A^-\cup A^+=A$, *then the transfer function* $H_{ji}(z)$ *is strictly causal.*

Proof. See the appendix. □

The following example revisits Example 4 showing that the consistent identification can be achieved even without knowing which transfer functions are strictly causal.

Example 6. *Consider a network with a perfect graphical representation as given in Figure 5. Suppose though that the information about the strict causality of the transfer functions is not available. Hence, what is known about the network is given by the graphical representation of Figure 7. Suppose the objective is to identify the transfer function* $H_{21}(z)$. *Assume that nodes 3 and 4 which are depicted with a dashed line are not measured* $3,4\notin O=\{1,2,5,6,7\}$. *Similar to Example 4* $Z=\{5,7\}$ *is a subset of* $O=\{1,2,5,6,7\}$ *that satisfies conditions (i) and (ii) of Theorem III.2. However, since the available graphical representation is not recursive, the sets* D^+ *and* D^- *cannot be determined. We will show, however, a recursive graphical representation can be obtained using the results of this section. Applying Theorem IV.1 on the transfer function* $H_{76}(z)$ *we get that the set* $\{2\}$ *7-pointing separates nodes 6 and 7 and* $Z^-=Z^+=\emptyset$. *To consider all possible combinations of* A^+ *and* A^-, *we need to consider two cases. In the first case, we have* $A^+=\{2\}$ *and* $A^-=\{\emptyset\}$. *It turns out that in*

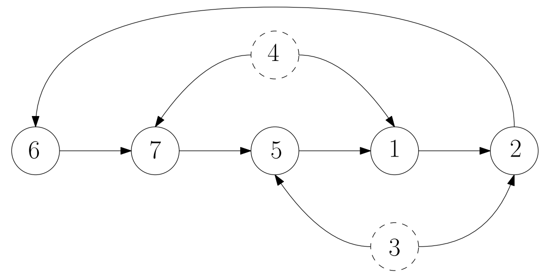


Fig. 7. The non-recursive graphical representation of the network discussed in Example 6. The nodes 3 and 4 are not measured and act as confounders.

$$\mathbb{E}(x_7(t)\mid I_7(t-1), I_{2,6}(t)) = \sum_{r\in\{2,6,7\}} W_{7r}(z)x_r(t) \tag{10}$$

the transfer function $W_{76}(z)$ *is strictly causal. In the second case, we have* $A^+=\{\emptyset\}$ *and* $A^-=\{2\}$. *Similarly, it turns out that in*

$$\mathbb{E}(x_7(t)\mid I_{2,7}(t-1), I_6(t)) = \sum_{r\in\{2,6,7\}} W_{7r}(z)x_r(t) \tag{11}$$

the transfer function $W_{76}(z)$ *is strictly causal. Thus, according to Theorem IV.1 we can conclude that the transfer function* $H_{76}(z)$ *is strictly causal. This information enables us to obtain a recursive graphical representation for the network, determine the sets* $D^+=\{1,5,7\}$ *and* $D^-=\{2\}$ *and consistently estimate the transfer function* $H_{21}(z)$ *using Procedure 1.*

This second result instead gives a sufficient criterion to determine if a transfer function has a nonzero feedthrough.

Theorem IV.2. *Consider a network with no algebraic loops and with (non-necessarily recursive) graphical representation* $G=(V,E_1,E_2)$ *and* $i\in pa_G(j)$. *Let* $Z\cap\{i,j\}=\emptyset$ *be a set that*

(i) j-pointing separates i and j in G
(ii) i-pointing separates i and j in G

Let Z^-,Z^+,A *be a partition of Z such that*

- $Z^- := \{\ell\in Z : \ell\notin an_{G^\sharp}(j)\}$
- $Z^+ := \{k\in Z : k\notin de_{G^\sharp}(j)\}\setminus Z^-$
- $A = Z\setminus(Z^-\cup Z^+)$.

If

$$\lim_{z\to\infty} W_{ji}(z)\neq 0 \tag{12}$$

in

$$\mathbb{E}(x_j(t)\mid I_{\{j\}\cup Z^-\cup A^-}(t-1), I_{Z^+\cup A^+\cup\{i\}}(t)) = \sum_{r\in Z^-\cup Z^+\cup A^-\cup A^+\cup\{i,j\}} W_{jr}(z)x_r(t) \tag{13}$$

and

$$\lim_{z\to\infty} W_{ij}(z)\neq 0 \tag{14}$$

in

$$\mathbb{E}(x_i(t)\mid I_{\{i\}\cup Z^-\cup A^-}(t-1), I_{Z^+\cup A^+\cup\{j\}}(t)) = \sum_{\ell\in Z^-\cup Z^+\cup A^-\cup A^+\cup\{i,j\}} W_{ir}(z)x_\ell(t) \tag{15}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

for all possible combinations of disjoint A^- and A^+ with $A^- \cup A^+ = A$, then either the transfer function $H_{ji}(z)$ is not strictly causal or the transfer function $H_{ij}(z)$ is not strictly causal.

Proof. See the appendix. □

Theorems IV.1 and IV.2 provide sufficient conditions to determine if a transfer function in the network is strictly causal or not, respectively. These conditions are only sufficient. Hence, there could be situations where their application would be inconclusive.

However, in several scenarios the information obtained from these two theorems, might be enough to determine a recursive graphical representation from a non-recursive one. The following example illustrates a situation when this occurs.

Example 7. *Consider a network with an unknown recursive graphical representation $\overline{G}$ depicted in Figure 8 (a). Assume, instead, that the less informative non-recursive graphical representation G of Figure 8 (b) is available, even though the network is known not to have any algebraic loops. As can be seen in $\overline{G}$, transfer functions $H_{43}(z)$, $H_{32}(z)$ and $H_{76}(z)$ are strictly causal. This information, however, is not available from G. The objective is identification of transfer function $H_{21}(z)$ given the topology and outputs of nodes $O = \{1,2,3,4,6,7\} \subset V = \{1,2,3,4,5,6,7\}$. Node 5 is not measured. The set $\{4,6\}$ satisfies graphical conditions (i) and (ii) of Theorem III.2. However, since there is no information available about the locations of strictly causal transfer functions, standard techniques cannot be applied to identify $H_{12}(z)$. For instance, in order to apply Procedure 1 to identify $H_{12}(z)$ we need to know a recursive graphical representation to determine D^- and D^+. Theorems IV.1 and IV.2 instead could be effectively applied in this case. If we consider the set $\{2\}$ we notice that such a set 4-pointing separates nodes 3 and 4. Hence, we can apply Theorem IV.1 on transfer function $H_{43}(z)$. Since $Z^- = Z^+ = \emptyset$, to consider all possible combinations of A^+ and A^-, we need to consider two cases. In the first case, we have $A^+ = \{2\}$ and $A^- = \{\emptyset\}$. In the second case, we have $A^+ = \{\emptyset\}$ and $A^- = \{2\}$. Taking similar steps as in Example 6, it turns out that the transfer function $H_{43}(z)$ is strictly causal. On the other hand, if we consider the set $\{7\}$ we notice that such a set 2-pointing and 6-pointing separates nodes 2 and 6. Hence, we can apply Theorem IV.2 on transfer function $H_{62}(z)$. Since $Z^- = Z^+ = \emptyset$, to consider all possible combinations of A^+ and A^-, again we need to consider two cases. In the first case, we have $A^+ = \{7\}$ and $A^- = \{\emptyset\}$. It turns out that in*

$$\mathbb{E}(x_6(t) \mid I_6(t-1), I_{2,7}(t) = \sum_{r\in\{2,6,7\}} W_{7r}(z)x_r(t) \tag{16}$$

the transfer function $W_{62}(z)$ is not strictly causal and in

$$\mathbb{E}(x_2(t) \mid I_2(t-1), I_{6,7}(t) = \sum_{r\in\{2,6,7\}} W_{7r}(z)x_r(t) \tag{17}$$

the transfer function $W_{26}(z)$ is not strictly causal. In the second case, we have $A^+ = \{\emptyset\}$ and $A^- = \{7\}$. It turns out that in

$$\mathbb{E}(x_6(t) \mid I_{6,7}(t-1), I_2(t) = \sum_{r\in\{2,6,7\}} W_{7r}(z)x_r(t) \tag{18}$$

the transfer function $W_{62}(z)$ is not strictly causal and in

$$\mathbb{E}(x_2(t) \mid I_{2,7}(t-1), I_6(t) = \sum_{r\in\{2,6,7\}} W_{7r}(z)x_r(t) \tag{19}$$

the transfer function $W_{26}(z)$ is not strictly causal. Thus, it follows Theorem IV.2 that the transfer function $H_{62}(z)$ is not strictly causal.

Since the network has no algebraic loops, we can conclude that either the transfer function $H_{76}(z)$ or $H_{27}(z)$ or both are strictly causal. As a consequence, the graph in Figure 8 (c) or the graph in Figure 8 (d) is a graphical representation of network and they are both recursive. Therefore, Theorem III.2 can be applied to either graph leading to the same choice of $D^+ = \{1,4\}$ and $D^- = \{2,6\}$ for the consistent identification of $H_{21}(z)$. Note that Theorem IV.2 can also be applied on the transfer function $H_{27}(z)$ revealing that it is not strictly causal. As a consequence, since the network has no algebraic loops, it can be inferred that the transfer function $H_{76}(z)$ is strictly causal.

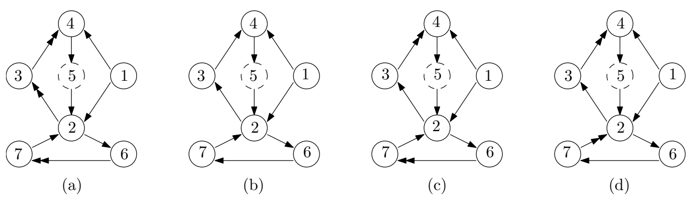


Fig. 8. (a) The unknown recursive graphical representation of a network discussed in Example 7; (b) The given non-recursive graphical representation of the network discussed in Example 7; The application of Theorems IV.1 and IV.2 allows one to conclude that either (c) or (d) is a valid graphical representation of the network. The application of Theorem III.2 with $Z = \{4,6\}$ leads to the same sets $D^- = \{2,6\}$ and $D^+ = \{1,4\}$ for the identification of $H_{21}(z)$ via Procedure 1.

V. Necessity of the Graphical Conditions for the Selection of Auxiliary Variables

Theorem III.2 provides sufficient conditions on how to select the set of auxiliary variables Z in order to consistently identify the transfer function $H_{ji}(z)$, namely, Z has to j-pointing separate i and j and block all the j-pointing paths from j to itself. In this section we show that these conditions are also necessary for the successful application of Procedure 1 when the only information about the network is given by a graphical representation.

Theorem V.1. *For any recursive graph $G = (V, E_1, E_2)$, if the graphical conditions (i) and (ii) of Theorem III.2 are not met by a set Z, there exists a network $\mathcal{G} = (H(z), n)$ with graphical representation G such that the estimate $\hat{H}_{ji}(z)$ of Procedure 1 will be inconsistent for all the possible choices of sets D^+ and D^- such that $D^+ \cup D^- = Z \cup \{i, j\}$ and $j \in D^-$.*

Proof. Suppose the first condition of Theorem III.2 is not met. That is, there is at least a j-pointing path between nodes i and j that is not blocked by Z and such a path is not the edge from i to j. Let π be the activated path with the least number of colliders. If π is collider-free, none of the nodes on π is a member of the separating set Z. Choose $H_{ji}(z)$ and all the other transfer functions outside of π as zero. Let the transfer

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

function from the parent of j on π to j be equal to z^{-m} where m is the length of π. Let all the other transfer functions on π be equal to z^{-1}. This reduces the perfect graphical representation of the network to only path π as shown in Figure 9. Let the noise processes on all the chain links and j be zero, and let the noise process on the fork, if there is one, and i be white with nonzero power spectral density. Since the transfer functions are all strictly causal the graph G is a graphical representation of this network. Notice that all the nodes in Z are independent of i and j. Also notice that the past of x_i is correlated with $x_j(t)$, and $x_i(t)$ and $x_j(t)$ are independent of each other. Thus, for every possible choice of D^+ and D^- we have

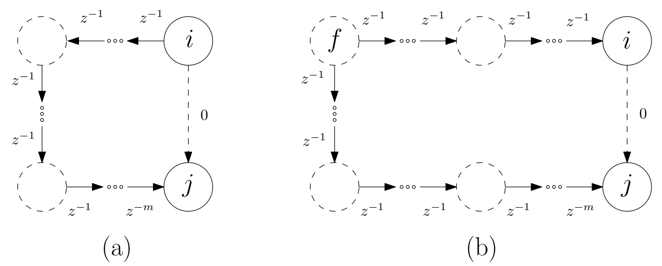


Fig. 9. Scenario one in the proof of Theorem V.1.

$$\mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)) = \mathbb{E}(x_j(t) \mid I_{\{i,j\}}(t-1)) \neq 0$$

that is a biased estimate of $H_{ji}(z)$.

In the second scenario there is at least one collider on π. Since the path is activated by Z, each collider on π must be either in Z or have at least one of its descendants in Z according to graph G. Furthermore, none of the non-colliders on π is in Z. Let every transfer function on the dipath from each collider to its descendants be equal to one. Let the noise process on each collider have variance one and let the noise on its other descendants be equal to zero. Since π has the minimal number of colliders, distinct colliders need to have disjoint descendants. Thus, measuring each descendant of each collider is the same as measuring the collider itself. Hence, we can assume without any loss of generality, that Z is a set of cooliders on π.

In the second scenario, suppose there exist r colliders $c_h \in V$, $h = 1,2,3,\cdots,r$ on π. Choose the transfer function entering j to be z^{-2}, all the transfer functions entering all the colliders to be z^{-1}, and all other transfer functions on π to be one. Let all the noise processes on all the chain links and j be zero, and the noise processes on the forks and colliders be white with variance one. If π is not i-pointing (Figure 10 (a)), let the noise on i be one. Otherwise, if π is i-pointing (Figure 10 (b)), let the the noise on i be white with variance zero.

The power spectral density associated with random processes $x_i, x_{c_1}, x_{c_2}, \cdots, x_{c_r}, x_j$ is given by the $(r+2)\times(r+2)$ matrix

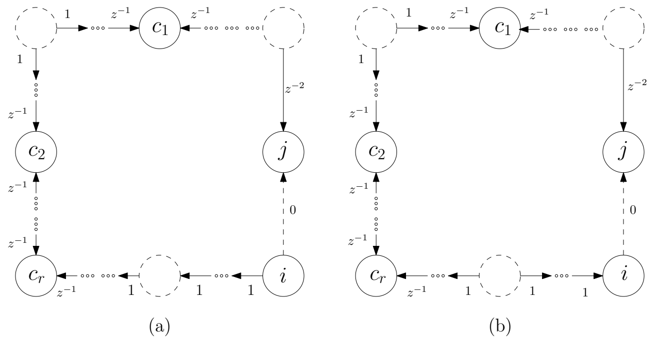


Fig. 10. Generic configurations of scenario two in the proof of Theorem V.1.

$$\Sigma = \begin{bmatrix} 1 & z & 0 & 0 & 0 & \cdots & 0 \\ z^{-1} & 3 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 3 & 1 & 0 & \cdots & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & 0 & \cdots & 1 & 3 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 & 3 & z \\ 0 & 0 & \cdots & 0 & 0 & z^{-1} & 1 \end{bmatrix}. \tag{20}$$

From Σ and the inverse of tridiagonal matrices formula [32] it is possible to compute the non-causal Wiener filter estimating x_j from x_i, $x_{c_1}, x_{c_2}, \cdots, x_{c_r}$. In particular, the component of the Wiener filter associated with x_i and x_{c_h} for $h = 1,2.\cdots,r$ are given by

$$W_{ji}(z) = \frac{z^{-2}}{\theta_{r+1}} \qquad W_{jc_h}(z) = \frac{z^{-1}\theta_h}{\theta_{r+1}} \tag{21}$$

where $\theta_k = 3\theta_{k-1} - \theta_{n-2}$ with $\theta_0 = 1$ and $\theta_1 = 1$ for $k = 2,...,r+1$. Since the non-causal Wiener filter is a strictly causal transfer function, it matches the Wiener-Hopf filter. Also, the strict causality of the Wiener filter implies that the expression of $W_{ji}(z)$ does not change for all choices of D^+ and D^- such that $D^+ \cup D^- = Z \cup \{i,j\}$ proving the estimate of $H_{ji}(z)$ via Procedure 1 is biased.

Now suppose the second condition of Theorem III.2 is not met. That is, there exists a j-pointing path between j and itself which is not blocked. There are two cases. Either the unblocked path is directed or not. First, we consider the case where there exists a directed feedback loop from the target node j to itself. Assume a network where all the transfer functions that are not involved in such a loop are zero. Instead, let all the transfer functions on the loop be $\frac{\alpha}{z}$ with $0 < |\alpha| < 1$. Let $W_{jj}(z)$ be the product of all the transfer functions on this directed feedback loop. Then the estimate of $H_{ji}(z)$ for all choices of D^- and D^+ will be given by

$$\hat{H}_{ji}(z) = \frac{H_{ji}(z)}{1 - W_{jj}(z)} \tag{22}$$

which is biased. Now we consider the case where there is a j-pointing path ℓ between j and itself which is not directed. That is, there is at least one collider on ℓ. Since ℓ is activated by Z, each collider on ℓ must be either in Z or have at least one of its descendants in Z according to graph G. Furthermore, none of the non-colliders on ℓ is in Z. Similar to above, we can assume without any loss of generality, that Z is a set of colliders on ℓ. Suppose there exist r colliders $c_h \in V$, $h = 1,2,3,\cdots,r$ on ℓ. Choose $H_{ji}(z)$ to be z^{-1} and all the other transfer functions outside of ℓ as zero. Choose the transfer function entering j from its parent j to be z^{-2}, all the transfer functions entering all the colliders on ℓ to be z^{-1}, and all other transfer functions on ℓ to be one. Let all the noise processes on all the chain links and j be zero, and the noise processes on the forks, colliders, and i be white with variance one. We have that $D^+ = \emptyset$ and $D^- = \{i, j, Z\}$

$$\begin{aligned}
&\mathbb{E}(x_j(t) \mid I_{\{i,j\}\cup Z}(t-1)) = \\
&\quad \mathbb{E}(H_{ji}(z)x_i(t) + H_{jp}(z)x_p(t) + n_j(t) \mid I_{\{i,j\}\cup Z}(t-1)) = \\
&\qquad \mathbb{E}(z^{-1}x_i(t) + z^{-2}x_p(t) + n_j(t) \mid I_{\{i,j\}\cup Z}(t-1)) = \\
&\qquad z^{-1}x_i(t) + \mathbb{E}(z^{-2}x_p(t) + n_j(t) \mid I_{\{i,j\}\cup Z}(t-1)) = \\
&\qquad z^{-1}x_i(t) + \mathbb{E}(z^{-2}x_p(t) + n_j(t) \mid I_{\{j\}\cup Z}(t-1)) = \\
&\qquad\qquad z^{-1}x_i(t) + \sum_{k\in\{j\}\cup Z} W_{jk}(z)x_k(z) \qquad (23)
\end{aligned}$$

Computing the Wiener filters $W_{jk}(z)$, $k \in \{j\} \cup Z$, using the power spectral density matrix like above, it turns out that $W_{jj}(z)$ is nonzero. Since $W_{ji}(z) = H_{ji}(z) = z^{-1}$, the estimate of $H_{ji}(z)$ via Procedure 1 is $\hat{H}_{ji}(z) = \frac{W_{ji}(z)}{1-W_{jj}(z)} = \frac{H_{ji}(z)}{1-W_{jj}(z)}$ which is biased.

□

Since the choice of the set Z is not unique, in some applications we might be interested in finding an optimal predictors set according to some cost function. The sufficiency of the conditions of Theorem III.2, along with their necessity as proven in Theorem V.1, are the basis to enable the search for an optimal predictors set which guarantees a consistent identification.

VI. Numerical Verification

The purpose of this section is to explore the identification performance of our variable selection method in the case of finite data and also provide a numerical illustration of the consistency properties proven in the theoretical sections.

Consider a network $\mathcal{G}$ with a recursive graphical representation G shown in Figure 11. Suppose the objective is the

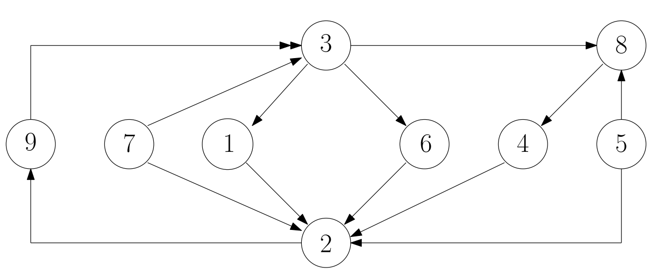


Fig. 11. The graphical representation of the benchmark network discussed in Section VI.

identification of the transfer function $H_{21}(z)$ using Procedure 1. To verify the consistency property of the identification when choosing a set of predictors satisfying the graphical conditions of Theorem III.2, we numerically simulated the network $\mathcal{G}$ and obtained time-series data. Using such generated time-series data, we considered different sets of predictors and computed the bias and the variance of the estimated transfer function.

We consider a parameterization $H(z,\theta)$ of the network and we denote as θ_{21} the subset of parameters associated with the transfer function $H_{21}(z)$ and as $\hat{\theta}_{21}$ the estimated parameters. We chose two predictor sets $Z_1 = \{4,5,6,7\}$ and $Z_2 = \{3,9\}$ satisfying conditions of Theorem III.2 and proceeded to the identification using time series of different lengths. For each set of predictors and for each time series length, we simulated the network $\mathcal{G}$ and used a linear regression technique to obtain the estimate $\hat{\theta}_{21}$. We repeated this procedure 1000 times in order to estimate $\mathbb{E}(\hat{\theta}_{21})$ and the covariance matrix of $\hat{\theta}_{21}$.

In Figure 12 we have reported the results of our Monte Carlo simulations for the set Z_1. On the horizontal axis we have the different time series lengths. The red squares represent the estimates of $\mathbb{E}\|\theta_{21} - \hat{\theta}_{21}\|_1$. We observe that for longer time series this quantity goes to zero numerically verifying that the bias of the estimated $\hat{\theta}_{21}$ asymptotically vanishes. The blue candle sticks define an interval the semi-amplitude of which is the square root of the trace of our estimate of the covariance matrix of $\hat{\theta}_{21}$. Since the amplitudes of these intervals go to zero for longer time series we have numerically verified the consistency property of our estimate.

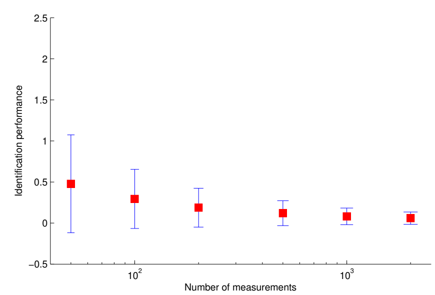


Fig. 12. Identification performance of the predictor set $Z_1 = \{4,5,6,7\}$ for different number of measurements.

We ran a similar set of simulations for Z_2 and the results are reported in Figure 13.

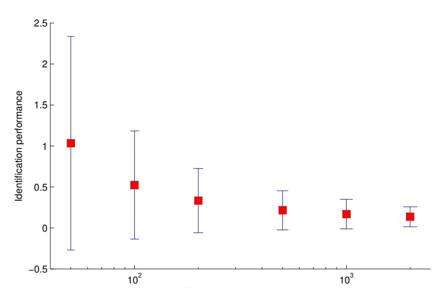


Fig. 13. Identification performance of the predictor set $Z_2 = \{3,9\}$ for different number of measurements.

Notice that even though both sets Z_1 and Z_2 guarantee consistent identification, in the case of finite data they provide different performance in terms of bias and variance of estimated parameters.

VII. Conclusion

The article introduced graphical conditions on the set of auxiliary variables in order to consistently identify a certain transfer function in a partially observed causal dynamic network via a single-output prediction error algorithm using only observational data. The results extend previous techniques borrowing elements from the theory of Structural Equation Models, Graphical Models and System Identification. One main advantage is that a consistent identification can be obtained for a network with no algebraic loops even when the class of parameterized models is allowed to contain algebraic loops. This is achieved by devising specific tests to detect strictly causal transfer functions. Most importantly, the graphical conditions on the set of auxiliary variables are proven to be sufficient and necessary. This characterization allows one to formulate identification problems while at the same time optimizing a cost function to take into account the potential cost of the observations.

References

[1] G. C. Goodwin and R. L. Payne, *Dynamic system identification: experiment design and data analysis*. Academic press, 1977.

[2] M. Gevers, A. S. Bazanella, and G. V. da Silva, "A practical method for the consistent identification of a module in a dynamical network," *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 862–867, 2018.

[3] M. Nabi-Abdolyousefi and M. Mesbahi, "Network identification via node knockout," *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 3214–3219, 2012.

[4] S. Shahrampour and V. M. Preciado, "Topology identification of directed dynamical networks via power spectral analysis," *IEEE Transactions on Automatic Control*, vol. 60, no. 8, pp. 2260–2265, 2014.

[5] C. Alexander, *Market models: A guide to financial data analysis*. John Wiley & Sons, 2001.

[6] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences*, vol. 95, no. 25, pp. 14 863–14 868, 1998.

[7] D. Materassi and M. V. Salapaka, "On the problem of reconstructing an unknown topology via locality properties of the wiener filter," *IEEE transactions on automatic control*, vol. 57, no. 7, pp. 1765–1777, 2012.

[8] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Exact topology identification of large-scale interconnected dynamical systems from compressive observations," in *Proceedings of the 2011 American Control Conference*. IEEE, 2011, pp. 649–656.

[9] S. Jahandari and D. Materassi, "Identification of dynamical strictly causal networks," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4739–4744.

[10] D. Hayden, Y. Yuan, and J. Gonçalves, "Network identifiability from intrinsic noise," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3717–3728, 2016.

[11] A. Haber and M. Verhaegen, "Subspace identification of large-scale interconnected systems," *IEEE Transactions on Automatic Control*, vol. 59, no. 10, pp. 2754–2759, 2014.

[12] J. Gonçalves and S. Warnick, "Necessary and sufficient conditions for dynamical structure reconstruction of lti networks," *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1670–1674, 2008.

[13] Y. Yuan, G.-B. Stan, S. Warnick, and J. Goncalves, "Robust dynamical network structure reconstruction," *Automatica*, vol. 47, no. 6, pp. 1230–1235, 2011.

[14] P. M. Van den Hof, A. Dankers, P. S. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with prediction error methods—basic methods for consistent module estimates," *Automatica*, vol. 49, no. 10, pp. 2994–3006, 2013.

[15] A. S. Bazanella, M. Gevers, J. M. Hendrickx, and A. Parraga, "Identifiability of dynamical networks: which nodes need be measured?" in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 5870–5875.

[16] H. H. Weerts, P. M. Van den Hof, and A. G. Dankers, "Identifiability of linear dynamic networks," *Automatica*, vol. 89, pp. 247–258, 2018.

[17] A. Dankers, P. M. Van den Hof, X. Bombois, and P. S. Heuberger, "Identification of dynamic models in complex networks with prediction error methods: Predictor input selection," *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 937–952, 2015.

[18] K. R. Ramaswamy, G. Bottegal, and P. M. Van den Hof, "Local module identification in dynamic networks using regularized kernel-based methods," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4713–4718.

[19] N. Everitt, G. Bottegal, and H. Hjalmarsson, "An empirical bayes approach to identification of modules in dynamic networks," *Automatica*, vol. 91, pp. 144–151, 2018.

[20] A. Dankers, P. M. Van den Hof, X. Bombois, and P. S. Heuberger, "Errors-in-variables identification in dynamic networks—consistency results for an instrumental variable approach," *Automatica*, vol. 62, pp. 39–50, 2015.

[21] K. R. Ramaswamy and P. M. Vandenhof, "A local direct method for module identification in dynamic networks with correlated noise," *IEEE Transactions on Automatic Control*, 2020.

[22] P. M. Van den Hof, K. R. Ramaswamy, A. G. Dankers, and G. Bottegal, "Local module identification in dynamic networks with correlated noise: the full input case," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 5494–5499.

[23] A. Dankers, P. M. Van den Hof, D. Materassi, and H. H. Weerts, "Conditions for handling confounding variables in dynamic networks," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3983–3988, 2017.

[24] D. Materassi and M. V. Salapaka, "Signal selection for estimation and identification in networks of dynamic systems: a graphical model approach," *IEEE Transactions on Automatic Control*, 2019.

[25] A. G. Dankers, P. M. Van den Hof, P. S. Heuberger, and X. Bombois, "Dynamic network structure identification with prediction error methods-basic examples," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 876–881, 2012.

[26] C. E. Shannon, *The theory and design of linear differential equation machines*, 1942.

[27] R. Diestel, "Graph theory 3rd ed," *Graduate texts in mathematics*, vol. 173, 2005.

[28] T. Shafie, "A multigraph approach to social network analysis." *Journal of Social Structure*, vol. 16, 2015.

[29] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[30] L. Ljung, "System identification," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–19, 1999.

[31] J. Pearl, *Causality: models, reasoning and inference*. Springer, 2000, vol. 29.

[32] R. A. Usmani, "Inversion of a tridiagonal jacobi matrix," *Linear Algebra and its Applications*, vol. 212, no. 213, pp. 413–414, 1994.

Appendix

A. Proof of Proposition III.1 and corollary III.1.1

Proof of Proposition III.1: Consider an estimate of $x_j(t)$ based on the information $I_{\{j\}\cup P^-}(t-1), I_{P^+}(t)$.

$$\mathbb{E}(x_j(t) \mid I_{\{j\}\cup P^-}(t-1), I_{P^+}(t)) = \sum_{k\in\{j\}\cup P^-} H_{jk}(z)x_k(t) + \sum_{k\in P^+} H_{jk}(z)x_k(t) + \mathbb{E}(n_j(t) \mid I_{\{j\}\cup P^-}(t-1), I_{P^+}(t)). \quad (24)$$

Define

$$\varepsilon_j(t) = n_j(t) - \hat{n}_j(t) \quad (25)$$

where $\hat{n}_j(t) = \mathbb{E}(n_j(t)|I_{\{n_j\}}(t-1)) = W_{jj}(z)n_j(t)$ with $W_{jj}(z)$ being strictly causal. Observe that $\mathbb{E}(n_j(t) \mid I_{\{n_j\}}(t-1))$ is $(I_{\{j\}\cup P^-}(t-1), I_{P^+}(t))$-measurable

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

since $n_j(t) = y_j(t) - \sum_{k\in P^-} H_{jk}(z)x_k(t) + \sum_{k\in P^+} H_{jk}(z)x_k(t)$. Since $\varepsilon_j(t) \perp\!\!\!\perp I_{\{j\}\cup P^-}(t-1), I_{P^+}(t)$, we get

$$\begin{aligned}
&\mathbb{E}(x_j(t))|I_{\{j\}\cup P^-}(t-1), I_{P^+}(t)) = \\
&\sum_{k\in P^-} H_{jk}(z)x_k(t) + \sum_{l\in P^+} H_{jk}(z)x_k(t) + \hat{n}_j(t) = \\
&\sum_{k\in P^-} H_{jk}(z)x_k(t) + \sum_{k\in P^+} H_{jk}(z)x_k(t) + \\
&W_{jj}(z)[x_j(t) - \sum_{k\in P^-} H_{jk}(z)x_k(t) + \sum_{k\in P^+} H_{jk}(z)x_k(t)] = \\
&W_{jj}(z)x_j(t) + \sum_{k\in P^-} [1-W_{jj}(z)]H_{jk}(z)x_k(t) \\
&\qquad + \sum_{k\in P^+} [1-W_{jj}(z)]H_{jk}(z)x_k(t). \quad (26)
\end{aligned}$$

Proof of Corollary III.1.1: Since, the minor of the power spectral density matrix corresponding to $\mathrm{pa}_G(j)$ is non-singular, the Wiener filter components estimating x_j from x_i and $x_{P^+\cup P^-}$ is unique. Hence, we have

$$W_{ji}(z) = [1-W_{jj}(z)]H_{ji}(z). \quad (27)$$

B. Proof of Theorem III.2

To prove Theorem III.2, we first need to provide a few lemmas.

Lemma VII.1. *Consider a network $\mathcal{G} = (H(z), n)$ with graphical representation $G = (V, E_1, E_2)$ and output processes x_V described by (2). Suppose $j \in V$ is a node in the network such that $ch_{G^{\not t}}(j) = \emptyset$. Define a network $\bar{\mathcal{G}} = (\bar{H}, \bar{n}_V)$ with output processes $\bar{x}_V$, where*

$$\begin{aligned}
\bar{H}_{jj}(z) &= H_{jj}(z) & \\
\bar{H}_{ab}(z) &= H_{ab}(z) & \text{for } a,b \neq j\\
\bar{H}_{aj}(z) &= zH_{aj}(z) & \text{for } a \neq j\\
\bar{H}_{ja}(z) &= z^{-1}H_{ja}(z) & \text{for } a \neq j\\
\bar{n}_a(t) &= n_a(t) & \text{for } a \neq j\\
\bar{n}_a(t) &= n_a(t-1) & \text{for } a = j.
\end{aligned}$$

Then, in $\bar{\mathcal{G}}$ we have that

$$\begin{aligned}
\bar{x}_a(t) &= x_a(t) & \text{for } a \neq j\\
\bar{x}_a(t) &= x_a(t-1) & \text{for } a = j.
\end{aligned}$$

Furthermore, $\bar{\mathcal{G}}$ has a graphical representation given by $\bar{G} = (V, \bar{E}_1, \bar{E}_2)$, where

$$\bar{E}_1 = (E_1 \cup_{c\in ch_G(j)} \{j \to c\}) \setminus \{\cup_{p\in pa_G(j)} p \to j\} \quad (28)$$
$$\bar{E}_2 = (E_2 \cup_{p\in pa_G(j)} \{p \to j\}) \setminus \{\cup_{c\in ch_G(j)} j \to c\} \quad (29)$$

and the relation $de_{\bar{G}^{\not t}}(k) = de_{G^{\not t}}(k) \setminus \{j\}$ holds for every node $k \neq j$.

Proof. Define the square matrix $M^j(z) = [m_{ab}(z)]$, $a,b,j \in V$ such that all off-diagonal entries $m_{ab}(z) = 0$ for $a \neq b$, $m_{ab}(z) = 1$ for $a = b \neq j$, and $m_{jj}(z) = z^{-1}$. Therefore, we have that $\bar{H}(z) = M^j(z)H(z)M^{j^{-1}}(z)$ and $\bar{n} = M^j(z)n$. Thus, the output processes of $\bar{\mathcal{G}}$ could be calculated in terms of output processes of $\mathcal{G}$ as follows.

$$\begin{aligned}
\bar{x}_V = \bar{H}(z)\bar{x}_V + \bar{n} = M^j(z)H(z)M^{j^{-1}}(z)\bar{x}_V + M^j n = \\
M^j(z)H(z)x_V + M^j(z)n = M^j(z)(H(z)x_V + n) = M^j(z)x_V,
\end{aligned}$$

which verifies that $\bar{x}_a(t) = x_a(t)$ for $a \neq j$ and $\bar{x}_j(t) = x_j(t-1)$. Moreover, since all the transfer functions $\bar{H}_{jb}(z) = z^{-1}H_{jb}(z)$, for $b \neq j$, are strictly causal, all the edges $p \to j$ for $p \in \mathrm{pa}_{\bar{G}}(j)$ can be double-headed. Therefore, $\bar{G}$ is a graphical representation of $\bar{\mathcal{G}}$. Finally, since all the edges $p \to j$ for $p \in \mathrm{pa}_{\bar{G}}(j)$ can be double-headed, and $\mathrm{ch}_{G^{\not t}}(j) = \emptyset = \mathrm{de}_{G^{\not t}}(j) \setminus j$, we have that $\mathrm{de}_{\bar{G}^{\not t}}(k) = \mathrm{de}_{G^{\not t}}(k) \setminus \{j\}$ for any node $k \neq j$. □

Lemma VII.2. *Given a recursive graph G, for any node j there exists at least a node $d \in de_G(j)$ such that $ch_{G^{\not t}}(d) = \emptyset$.*

Proof. The result is an immediate consequence of the fact that for any recursive graph G, the graph of instantaneous propagations $G^{\not t}$ is a directed acyclic graph. □

The following lemma provides a connection between the standard notion of d-separation and the notion of pointing separation adopted in this article.

Lemma VII.3. *Let $i,j \in V$, $Z \subset V$, and $\{i,j\} \cap Z = \emptyset$ in a network with graphical representation $G = (V,E)$. If all the j-pointing paths between nodes i and j are blocked by Z, then i and $pa(j) \setminus \{Z \cup \{j\}\}$ are d-separated by $Z \cup \{j\}$ in G.*

Proof. By contradiction suppose $w \in \mathrm{pa}(j) \setminus \{Z \cup \{j\}\}$ such that there is a connected path $\tilde{\pi}$ between nodes i and w not blocked by $Z \cup \{j\}$. We can have two cases. Either j is in $\tilde{\pi}$ or not. Suppose j is not in $\tilde{\pi}$ and $\tilde{\pi}$ is not blocked by $Z \cup \{j\}$. If j is not a descendent of colliders in $\tilde{\pi}$, then $\pi = (\tilde{\pi}, w \to j)$ is a j-pointing path connecting nodes i and j not blocked by Z, which is a contradiction. If j is a descendent of some colliders in $\tilde{\pi}$, let c be the closest such collider to i. Then, $\pi = (i \cdots \to c \to \cdots \to j)$ is a j-pointing path connecting nodes i and j not blocked by Z, which is a contradiction. Now, suppose j is in $\tilde{\pi}$ and $\tilde{\pi}$ is not blocked by $Z \cup \{j\}$. Then, $\tilde{\pi}$ is either of the form $\tilde{\pi} = (\hat{\pi} \to j \cdots w)$ or $\tilde{\pi} = (\hat{\pi} \leftarrow j \cdots w)$. If $\tilde{\pi} = (\hat{\pi} \to j \cdots w)$, then $\pi = \hat{\pi} \to j$ is a j-pointing path connecting nodes i and j not blocked by Z, which is a contradiction. If $\tilde{\pi} = (\hat{\pi} \leftarrow j \cdots w)$, then π is blocked by $Z \cup \{j\}$, which is a contradiction. □

We are now ready to prove Theorem III.2.

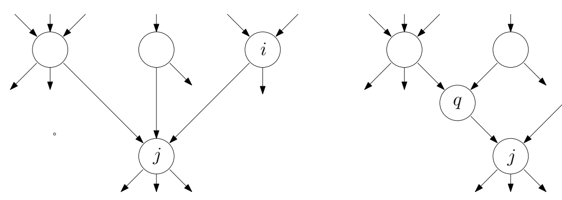
Proof of Theorem III.2: Let $G' = (V, (E_1 \cup E_2) \setminus \{i \to j\})$ be the standard directed graph associated to G after removing the edge $i \to j$. Define $E = E_1 \cup E_2$. Also define a new process $x_q(t) = x_j(t) - H_{ji}(z)x_i(t)$. We are going to define a new network $\mathcal{G}'' = (H'', n'')$ with all the variables of the original network and the additional variable x_q. Let

$$\begin{aligned}
&H''_{jq}(z) = 1,\\
&H''_{qr}(z) = H_{jr}(z) \text{ for } r \in \mathrm{pa}_G(j) \setminus i,\\
&H''_{k\ell}(z) = H_{k\ell}(z) \text{ in all other cases}
\end{aligned}$$

and $n''_q = n_j$, $n''_j = 0$ and $n''_k = n_k$ in all other cases. From G we can obtain a graphical representation for $\mathscr{G}''$ given by $G'' = (V'', E''_1, E''_2)$ in the following way. Let $K_1 := \{k | k \neq i \text{ and } k \to j \in E_1\}$ be the set of single-headed parents of j in G that are not node i and $K_2 := \{k | k \neq i \text{ and } k \to j \in E_2\}$ be the set of double-headed parents of j in G that are not node i. Then

$$\begin{aligned}
V'' &:= V \cup \{q\} \\
E''_1 &:= E_1 \cup \{q \to j\} \cup \{k \to q | k \in K_1\} \setminus \{k \to j | k \in K_1\} \\
E''_2 &:= E_2 \cup \{k \to q | k \in K_2\} \setminus \{k \to j | k \in K_2\}.
\end{aligned}$$

Namely, in $G'' = (V'', E''_1, E''_2)$, the additional node q is placed in between j and its original parents in G that are not the node i, see Figure 14. Also, notice that since G is recursive,



(a) (b)

Fig. 14. The new variable q in G'' is introduced in between node j and all its parents that are not the node i.

G'' is trivially recursive, as well. Define $P = \mathrm{pa}_{G''}(q) \setminus Z$. Decompose P into P^- and P^+ where P^- contains all double headed parents of q that are not in Z and $P^+ = P \setminus P^-$ contains all the single-headed parents of q that are not in Z. Observe that $j \in D^-$, hence D^- is never empty. Conversely, the node i belongs to either D^- or D^+. First, we consider the case $i \in D^+$. Let S_0 be the set containing all descendants of the nodes in D^- in $G''^{\not{z}}$, $S_0 := \mathrm{de}_{G''^{\not{z}}}(D^-)$. Since G'' is recursive and D^- is not empty, by Lemma VII.2 there exists a node $w_0 \in S_0$ such that $\mathrm{ch}_{G''^{\not{z}}}(w_0) = \emptyset$. Apply Lemma VII.1 on w_0 obtaining a new network with recursive graphical representation G_1. Define $S_1 := \mathrm{de}_{G_1^{\not{z}}}(D^-)$. From Lemma VII.1 it follows that $S_1 = S_0 \setminus \{w_0\}$. Again, by Lemma VII.2 there exists a node $w_1 \in S_1$ such that $\mathrm{ch}_{G_1^{\not{z}}}(w_1) = \emptyset$. Apply Lemma VII.1 on w_1, represent the resulting network with G_2 and let $S_2 := \mathrm{de}_{G_2^{\not{z}}}(D^-)$. Again, from Lemma VII.1 it follows that $S_2 = S_1 \setminus \{w_1\}$. Repeat the procedure N times, till $S_N = \emptyset$, for $N \in \mathbb{Z}$, the number of elements in S_0. Let (H_N, n_N) be the resulting network with recursive graphical representation G_N. Build a new graphical representation $\underline{G}$ for (H_N, n_N) by adding single-headed edges from nodes in Z to q. Observe that in $\underline{G}$, the parents of q are now given by $P^- \cup P^+ \cup Z$. Let $Z^- := S_0 \cap Z$ and $Z^+ := Z \setminus Z^-$. Since S_0 contains all the elements in D^- we have that $D^- = Z^- \cup \{j\}$. Consequently, $D^+ = Z^+ \cup \{i\}$. Also, it follows from Lemma VII.1 that in (H_N, n_N) the output processes of nodes k in Z^- are now $x_k(t-1)$ while the output processes of nodes i (since $i \in D^+$) and ℓ in Z^+ remain unchanged, $x_\ell(t)$. Applying Proposition III.1 on q in $\underline{G}$, which is recursive, gives

$$x_q(t) \perp\!\!\!\perp I_i(t) \mid I_{P^- \cup Z^- \cup \{q\}}(t-1), I_{P^+ \cup Z^+}(t). \quad (30)$$

Apply Lemma VII.1 one more time on q in G_N to get $\bar{G}$. Every q-pointing path $\bar{\pi}_1 = \{i \cdots k \to q\}$ between i and q in $\bar{G}$ that does not pass through j could be mapped to a j-pointing path $\pi_1 = \{i \cdots k \to j\}$ between i and j in G'. Thus, from assumption (i) we have that Z also blocks all paths $\bar{\pi}_1$ in $\bar{G}$. Every q-pointing path $\bar{\pi}_2 = \{i \cdots j \to \hat{\pi}_3 \to q\}$, where $\hat{\pi}_3$ is a node or a path that might or might not be directed, between i and q in $\bar{G}$ that passes through j corresponds to a j-pointing path $\pi_2 = \{j \to \hat{\pi}_3 \to j\}$ between j and itself in G. Thus, from assumption (ii) we have that Z also blocks all paths $\bar{\pi}_2$ in $\bar{G}$. Therefore, Z blocks all the q-pointing paths between i and q in $\bar{G}$. By Lemma VII.3, we have that $\mathrm{pa}_{\bar{G}}(q) \setminus Z$ are d-separated from i by $Z \cup \{q\}$. Theorem 24 in [24] states that if two sets of nodes A and B are d-separated by C in a graphical representation of a network we have that $I_A(t) \perp\!\!\!\perp I_B(t) \mid I_C(t)$. Therefore, applying Theorem 24 in [24] to $\bar{G}$ we get

$$I_{P^-}(t-1), I_{P^+}(t) \perp\!\!\!\perp I_i(t) \quad \mid \quad I_{\{q\} \cup Z^-}(t-1), I_{Z^+}(t). \quad (31)$$

By Contraction property of conditional independence [31, Chapter 1, Page 11], combining (30) and (31), we obtain

$$I_{P^-}(t-1), I_{P^+ \cup \{q\}}(t) \perp\!\!\!\perp I_i(t) \mid I_{Z^- \cup \{q\}}(t-1), I_{Z^+}(t). \quad (32)$$

By Decomposition property of conditional independence [31, Chapter 1, Page 11], this yields

$$x_q(t) \perp\!\!\!\perp I_i(t) \mid I_{Z^- \cup \{q\}}(t-1), I_{Z^+}(t). \quad (33)$$

Therefore, when estimating $x_q(t)$ from $I_{\{i\} \cup Z^+}(t), I_{\{q\} \cup Z^-}(t-1)$, the transfer function corresponding x_i will be zero:

$$\begin{aligned}
&\mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{q\} \cup Z^-}(t-1)) = \\
&\qquad \mathbb{E}(x_q(t) \mid I_{Z^+}(t), I_{\{q\} \cup Z^-}(t-1)) = \\
&\qquad F_{qq}(z) x_q(t) + \sum_{k \in Z^-} F_{qk}(z) x_k(t) + \sum_{k \in Z^+} F_{qk}(z) x_k(t), \quad (34)
\end{aligned}$$

where $F_{qq}(z)$ and $F_{qk}(z)$, $k \in Z^-$ are strictly causal transfer functions, and $F_{qk}(z)$, $k \in Z^+$ are causal transfer functions. Since $H''_{jq}(z) = 1$ is causally invertible, it follows that the filtration induced by random processes x_i and x_{Z^+} till time t and x_{Z^-} and x_q till time $t-1$ is equal to the filtration induced by random processes x_i and x_{Z^+} till time t and x_{Z^-} and x_j till time $t-1$. Since $x_j(t) = x_q(t) + H_{ji}(z) x_i(t)$, when $x_j(t)$ is projected on $I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)$, we get

$$\begin{aligned}
&\mathbb{E}(x_j(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)) = \\
&\quad \mathbb{E}(x_q(t) + H_{ji}(z) x_i(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)) = \\
&\quad H_{ji}(z) x_i(t) + \mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)) = \\
&\quad H_{ji}(z) x_i(t) + \mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{q\} \cup Z^-}(t-1)) = \\
&\qquad H_{ji}(z) x_i(t) + F_{qq}(z) x_q(t) + \\
&\qquad \sum_{k \in Z^-} F_{qk}(z) x_k(t) + \sum_{k \in Z^+} F_{qk}(z) x_k(t) = \\
&\qquad H_{ji}(z) x_i(t) + F_{qq}(z)(x_j(t) - H_{ji}(z) x_i(t)) + \\
&\qquad \sum_{k \in Z^-} F_{qk}(z) x_k(t) + \sum_{k \in Z^+} F_{qk}(z) x_k(t) = \\
&\qquad F_{qq}(z) x_j(t) + [H_{ji}(z) - F_{qq}(z) H_{ji}(z)] x_i(t) \\
&\qquad\quad + \sum_{k \in Z^-} F_{qk}(z) x_k(t) + \sum_{k \in Z^+} F_{qk}(z) x_k(t). \quad (35)
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

Procedure 1 in step 3 computes

$$\mathbb{E}(x_j(t) \mid I_{D^-}(t-1),\ I_{D^+}(t)) = \sum_{k \in D^- \cup D^+} W_{jk}(z) x_k(t). \quad (36)$$

Since the power spectral density matrix associated with (x_i, x_j, x_Z) is non-singular, comparing the two expressions for $\mathbb{E}(x_j(t) \mid I_{\{i\}\cup Z^+}(t), I_{\{j\}\cup Z^-}(t-1))$ we can conclude $W_{jk}(z) = F_{qk}(z)$ for all $k \in Z^- \cup Z^+$, $W_{jj}(z) = F_{qq}(z)$, and

$$W_{ji}(z) = H_{ji}(z) - F_{qq}(z) H_{ji}(z) = H_{ji}(z) - W_{jj}(z) H_{ji}(z). \quad (37)$$

This verifies the assertion for the case where $i \in D^+$. Analogous steps apply to prove the case where $i \in D^-$.

C. Proofs of Theorems IV.1 and IV.2

To prove Theorems IV.1 and IV.2, we first introduce a Lemma.

Lemma VII.4. *Consider a network with recursive graphical representation $G = (V, E_1, E_2)$ and $i \in pa_G(j)$. Let $Z \cap \{i, j\} = \emptyset$ be a set that j-pointing separates nodes i and j in G. Let $G^{\not{t}}$ be the graph of instantaneous propagations associated to G and let D^- and D^+ be the following two disjoint sets partitioning $Z \cup \{i, j\}$*

- $D^+ := an_{G^{\not{t}}}(j) \cap (Z \cup \{i\})$
- $D^- := (Z \cup \{i, j\}) \setminus D^+$

Let

$$\mathbb{E}(x_j(t) \mid I_{D^-}(t-1), I_{D^+}(t)) = \sum_{r \in D^- \cup D^+} W_{jr}(z) x_r(t). \quad (38)$$

Then, $H_{ji}(z)$ is strictly causal if and only if $W_{ji}(z)$ is strictly causal.

Proof. Let $G' = (V, (E_1 \cup E_2) \setminus \{i \to j\})$ be the standard directed graph associated to G after removing the edge $i \to j$. Define $E = E_1 \cup E_2$. Also define a new processes $x_q(t) = x_j(t) - H_{ji}(z) x_i(t)$ and $x_w(t) = x_i(t)$. We are going to define a new network $\mathscr{G}'' = (H'', n'')$ with all the variables of the original network and the additional variables x_q and x_w. Let

$$\begin{aligned} &H''_{jq}(z) = 1, \\ &H''_{jw}(z) = H_{ji}(z), \\ &H''_{qr}(z) = H_{jr}(z) \text{ for } r \in \mathrm{pa}_G(j) \setminus i, \\ &H''_{wi}(z) = H_{ji}(z) \\ &H''_{k\ell}(z) = H_{k\ell}(z) \text{ in all other cases} \end{aligned}$$

and $n''_q = n_j$, $n''_w = 0$, $n''_j = 0$, and $n''_k = n_k$ in all other cases. From G we can obtain a graphical representation for $\mathscr{G}''$ given by $G'' = (V'', E''_1, E''_2)$ in the following way. Let $K_1 := \{k | k \neq i \text{ and } k \to j \in E_1\}$ be the set of single-headed parents of j in G that are not node i and $K_2 := \{k | k \neq i \text{ and } k \to j \in E_2\}$ be the set of double-headed parents of j in G that are not node i. Then

$$\begin{aligned} V'' &:= V \cup \{q, w\} \\ E''_1 &:= E_1 \cup \{q \to j, w \to j, i \to w\} \cup \{k \to q | k \in K_1\} \\ &\quad \setminus \{k \to j | k \in K_1\} \\ E''_2 &:= E_2 \cup \{k \to q | k \in K_2\} \setminus \{k \to j | k \in K_2\}. \end{aligned}$$

Namely, in $G'' = (V'', E''_1, E''_2)$, the additional node q is placed in between j and its original parents in G that are not the node i, and node w is placed in between nodes i and j. Also, notice that since G is recursive, G'' is trivially recursive, as well. Observe that $i \in D^+$ and $j \in D^-$, hence D^- is never empty. Let S_0 be the set containing all descendants of the nodes in D^- in $G''^{\not{t}}$, $S_0 := \mathrm{de}_{G''^{\not{t}}}(D^-)$. Since G'' is recursive and D^- is not empty, by Lemma VII.2 there exists a node $w_0 \in S_0$ such that $\mathrm{ch}_{G''^{\not{t}}}(w_0) = \emptyset$. Apply Lemma VII.1 on w_0 obtaining a new network with recursive graphical representation G_1. Define $S_1 := \mathrm{de}_{G_1^{\not{t}}}(D^-)$. From Lemma VII.1 it follows that $S_1 = S_0 \setminus \{w_0\}$. Again, by Lemma VII.2 there exists a node $w_1 \in S_1$ such that $\mathrm{ch}_{G_1^{\not{t}}}(w_1) = \emptyset$. Apply Lemma VII.1 on w_1, represent the resulting network with G_2 and let $S_2 := \mathrm{de}_{G_2^{\not{t}}}(D^-)$. Again, from Lemma VII.1 it follows that $S_2 = S_1 \setminus \{w_1\}$. Repeat the procedure N times, till $S_N = \emptyset$, for $N \in \mathbb{Z}$, the number of elements in S_0. Let (H_N, n_N) be the resulting network with recursive graphical representation G_N. Let $Z^- := S_0 \cap Z$ and $Z^+ := Z \setminus Z^-$. Since S_0 contains all the elements in D^- we have that $D^- = Z^- \cup \{j\}$. Consequently, $D^+ = Z^+ \cup \{i\}$. Also, it follows from Lemma VII.1 that in (H_N, n_N) the output processes of nodes k in Z^- are now $x_k(t-1)$ while the output processes of nodes i (since $i \in D^+$) and ℓ in Z^+ remain unchanged, $x_\ell(t)$. Apply Lemma VII.1 one more time on w in G_N to get $\bar{G}$. By Lemma VII.3, we have that node q is d-separated from node i given $\{w, j\} \cup Z^- \cup Z^+$ in $\bar{G}$. Note that the output process of node w in $\bar{G}$ is $x_i(t-1)$. Therefore, applying Theorem 24 in [24] yields

$$x_q(t) \perp\!\!\!\perp x_i(t) \mid I_{\{i,j\}\cup Z^-}(t-1), I_{Z^+}(t). \quad (39)$$

Thus, we have that

$$\begin{aligned} &\mathbb{E}(x_q(t) \mid I_{\{i\}\cup Z^+}(t),\ I_{\{j\}\cup Z^-}(t-1)) = \\ &\qquad \mathbb{E}(x_q(t) \mid I_{Z^+}(t),\ I_{\{i,j\}\cup Z^-}(t-1)). \quad (40) \end{aligned}$$

Therefore, we can write

$$\begin{aligned} &\mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)) = \\ &\mathbb{E}(x_j(t) \mid I_{\{i\}\cup Z^+}(t), I_{\{j\}\cup Z^-}(t-1)) = \\ &\mathbb{E}(x_q(t) + H_{ji}(z) x_i(t) \mid I_{\{i\}\cup Z^+}(t), I_{\{j\}\cup Z^-}(t-1)) = \\ &H_{ji}(z) x_i(t) + \mathbb{E}(x_q(t) \mid I_{\{i\}\cup Z^+}(t), I_{\{j\}\cup Z^-}(t-1)) = \\ &H_{ji}(z) x_i(t) + \mathbb{E}(x_q(t) \mid I_{Z^+}(t),\ I_{\{i,j\}\cup Z^-}(t-1)) = \\ &H_{ji}(z) x_i(t) + \sum_{r \in \{i,j\}\cup Z^- \cup Z^+} F_{qr}(z) x_r(t) = \\ &H_{ji}(z) x_i(t) + F_{qi}(z) x_i(t) + F_{qj}(z) x_j(t) + \sum_{r \in Z^- \cup Z^+} F_{qr}(z) x_r(t) = \\ &[H_{ji}(z) + F_{qi}(z)] x_i(t) + F_{qj}(z) x_j(t) + \sum_{r \in Z^- \cup Z^+} F_{qr}(z) x_r(t) \quad (41) \end{aligned}$$

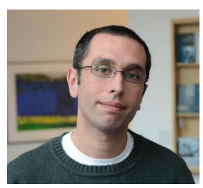
where $F_{qi}(z)$, $F_{qj}(z)$, and $F_{qr}(z)$, $r \in Z^-$ are strictly causal transfer functions. Since the power spectral density matrix associated with (x_i, x_j, x_Z) is non-singular, comparing the two expressions for $\mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1))$ we can conclude $W_{jk}(z) = F_{qk}(z)$ for all $k \in Z^- \cup Z^+$, $W_{jj}(z) = F_{qj}(z)$ and $W_{ji}(z) = H_{ji}(z) + F_{qi}(z)$. Since $F_{qi}(z)$ is strictly causal, $W_{ji}(z) = H_{ji}(z) + F_{qi}(z)$ is strictly causal if and only if $H_{ji}(z)$ is strictly causal. Also, $W_{ji}(z)$ is not strictly causal if and only if $H_{ji}(z)$ is not strictly causal. □

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2021.3130885, IEEE Transactions on Automatic Control

Proof of Theorem IV.1: Let $G^p = (V, E_1^p, E_2^p)$ be the perfect graphical representation of the network. Since the network has no algebraic loops, G^p is recursive. Build a new graphical representation $\overline{G} = (V, \overline{E}_1, \overline{E}_2)$ of the network by adding single-headed edges from all nodes $k \in Z^+$ to j in G^p. That is, $\overline{E}_2 = E_2^p$ and $\overline{E}_1 = E_1^p \cup_{k \in Z^+} \{k \to j\}$. This implies that $Z^+ \subseteq \mathrm{an}_{\overline{G}^{\not\leftrightarrow}}(j)$. Note that $\overline{G}$ is recursive because for all edges $k \to j$ that we added to E_1^p to obtain $\overline{E}_1$, we have that $k \notin \mathrm{de}_{\overline{G}^{\not\leftrightarrow}}(j)$. Assume, by contradiction that $H_{ji}(z)$ is not strictly causal. Since $Z^- := \{\ell \in Z : \ell \notin \mathrm{an}_{G^{\not\leftrightarrow}}(j)\}$, we have that $Z^- \cap \mathrm{an}_{G^{p\not\leftrightarrow}}(j) = \emptyset$. Since Z^- does not contain any ancestor of Z^+ in $G^{\not\leftrightarrow}$, it also follows that $Z^- \cap \mathrm{an}_{\overline{G}^{\not\leftrightarrow}}(j) = \emptyset$. Hence, applying Lemma VII.4 on $\overline{G}$, we get that $Z^- \subset D^-$. On the other hand, since $Z^+ \subseteq \mathrm{an}_{\overline{G}^{\not\leftrightarrow}}(j)$, we have that $Z^+ \subset D^+$. Since i is a parent of j in $\overline{G}^{\not\leftrightarrow}$, which is a recursive graph, there is one choice of A_1 and A_2 where $D^- = Z^- \cup A_1 \cup \{j\}$ and $D^+ = Z^+ \cup A_2 \cup \{i\}$ meeting the conditions of Lemma VII.4 on $\overline{G}$. For those A_1 and A_2 Lemma VII.4 gives that $H_{ji}(z)$ is strictly causal which is a contradiction.

Proof of Theorem IV.2: Without any loss of generality, assume that $\{j \to i\} \in E_1$, otherwise we can redefine E_1 and E_2 respectively as $E_1 \cup \{j \to i\}$, and $E_2 \setminus \{j \to i\}$, since this would still give us a (non-necessarily recursive) graphical representation of the network where the set Z still satisfies the Theorem's assumption. Let $G^p = (V, E_1^p, E_2^p)$ be the perfect graphical representation of the network. Since the network has no algebraic loops, G^p is recursive. Since G^p is recursive it holds that (i) every dipath from j to i has at least a double headed edge or (ii) every dipath from i to j has at least a double headed edge. Consider first case (i). Build a graphical representation $\overline{G} = (V, \overline{E}_1, \overline{E}_2)$ of the network by adding the single headed edge $i \to j$ and single-headed edges from all nodes $k \in Z^+$ to j in G^p. That is, $\overline{E}_2 = E_2^p$ and $\overline{E}_1 = E_1^p \cup \{i \to j\} \cup_{k \in Z^+} \{k \to j\}$. This implies that $Z^+ \cup \{i\} \subseteq \mathrm{an}_{\overline{G}^{\not\leftrightarrow}}(j)$. Since Z^- does not contain any ancestor of Z^+ in $G^{\not\leftrightarrow}$, it also follows that $Z^- \cap \mathrm{an}_{\overline{G}^{\not\leftrightarrow}}(j) = \emptyset$. Observe also that because of (i) and the definition of Z^+, $\overline{G}$ is recursive. Hence, by applying Lemma VII.4 on $\overline{G}$, we get that there exist disjoint A_1 and A_2 such that $D^- = A_1 \cup Z^- \cup \{j\}$ and $D^+ = A_2 \cup Z^+ \cup \{i\}$ giving a non-strictly causal estimate of the transfer function $H_{ji}(z)$. Since for all choices of A_1 and A_2 the transfer function estimate that would result from Equation (12) and Equation (13) has a non-zero feedthrough component, $H_{ji}(z)$ needs to be non-strictly causal under scenario (i). If instead scenario (ii) holds, we build a graphical representation $\overline{G} = (V, \overline{E}_1, \overline{E}_2)$ of the network by adding the single headed edge $j \to i$ and single-headed edges from all nodes $k \in Z^+$ to i in G^p. By repeating steps similar to scenario (i) with reversed roles for the nodes i and j, we would find that, in scenario (ii), because of Equation (14) and Equation (15) the transfer function $H_{ij}(z)$ needs to be non-strictly causal. Now, we do not know if scenario (i) or scenario (ii) holds, thus, we can only conclude that either $H_{ji}(z)$ is strictly causal or $H_{ij}(z)$ is strictly causal.

Sina Jahandari received the B.S. degree in electrical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2012, and the M.S. degree in mechatronics engineering from the University of Tehran, Tehran, in 2015. He is currently pursuing the Ph.D. degree in electrical engineering with the University of Minnesota, Twin Cities, MN, USA. His current research interests include stochastic processes and system identification.



Donatello Materassi holds a Laurea in "Ingegneria Informatica" and a "Dottorato di Ricerca" in Electrical Engineering/Nonlinear Dynamics and Complex Systems from Università degli Studi di Firenze, Italy. He has been a research associate at University of Minnesota (Twin Cities) from 2008 till 2011. He has been concurrently both a post-doctoral researcher at Laboratory for Information and Decision Systems (LIDS) at the Massachusetts Institute of Technology and a lecturer at Harvard University till 2014. Since 2014 he has been an assistant professor at University of Tennessee in Knoxville and since 2021 an associate professor at University of Minnesota. He received the NSF CAREER award in 2015. His research interests include nonlinear dynamics, system identification and classical control theory with applications to atomic force microscopy, single molecule force spectroscopy, biophysics, statistical mechanics and quantitative finance.