Conditions for handling confounding variables in dynamic networks ⋆

Arne Dankers * **Paul M. J. Van den Hof** ** **Donatello Materassi** *** **Harm H. M. Weerts** **

* *Dept. of Electrical and Computer Engineering, University of Calgary, Canada (email: adankers@hifieng.com).*
** *Dept. of Electrical Engineering, Eindhoven University of Technology, The Netherlands (email: h.h.m.weerts@tue.nl, p.m.j.vandenhof@tue.nl).*
*** *Dept. of Electrical and Computer Engineering, University of Tennessee, Knoxville, USA (email: dmateras@utk.edu).*

Abstract: In this paper we focus on consistently identifying a transfer function (module) embedded in a dynamic network. When identifying a module embedded in a dynamic network, a critical choice is which variables to include as predictor inputs. In the system identification literature sufficient conditions have been derived. One condition is that there should be no *confounding variables*. We show that this condition can be relaxed.

1. INTRODUCTION

Dynamic networks are pervasive in engineering domains such as power systems, pipelines, and distributed control systems. These systems cannot be designed, operated or maintained without models of the system. Data is becoming easier to collect from these systems, and so system identification is well poised to play a critical role in the advancement of these technologies. Developing tools for identifying transfer functions embedded in dynamic networks is an active research field [Gonçalves and Warnick, 2008, Materassi and Innocenti, 2010, Haber and Verhaegen, 2013]. Conditions have been proposed to ensure that consistent estimates of the transfer function(s) of interest [Dankers et al., 2016, Materassi and Salapaka, 2015].

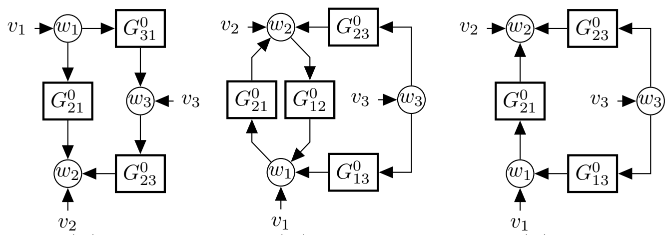

Fig. 1. (a) Example of a network, circles denote internal variables, and boxes denote transfer functions. (b), (c) Networks analyzed in Examples 1 and 3.

When deciding how to identify a particular module embedded in a dynamic network, a critical choice is which variables to include as predictor inputs. Suppose that the objective is to consistently identify a transfer function G^0_{ji} embedded in a dynamic network. Simply choosing w_i as the only input and w_j as the variable to be predicted (i.e. the output) will generally not lead to consistent estimates of G^0_{ji}. Consider the example network shown in Fig. 1a to illustrate this statement. For this network, there are two parallel paths from $w_1 \to w_2$. Suppose that only w_1 and w_2 are known. It can be proved that when using w_2 as the output, and w_1 as the input, an estimate of $G^0_{21} + G^0_{13}G^0_{31}$ is obtained. In other words, instead of identifying G^0_{21}, an estimate of a sum of the parallel paths is obtained. On the other hand, it can also be shown that if both w_1 and w_3 are used as predictor inputs, then consistent estimates of both G^0_{21} and G^0_{23} are obtained. The main point is that including w_3 as a predictor input enabled the possibility of consistent estimates of G^0_{21}.

As illustrated, an important question is: what variables must be included as predictor inputs to ensure consistent estimates of a module embedded in a dynamic network? This question has been addressed in both Dankers et al. [2016] and Materassi and Salapaka [2015]. The paper of Materassi and Salapaka [2015] is based on extending notions developed in the field of probabilistic inference [Pearl, 2000], where typically the networks that are dealt with are *directed acyclic graphs* (networks with no loops). The paper of Dankers et al. [2016] on the other hand is based on extending closed-loop identification methods. The conditions for predictor input selection presented in Materassi and Salapaka [2015] are less restrictive when considering networks that are directed acyclic graphs, and the conditions of Dankers et al. [2016] are less restrictive when considering causal networks with loops. One of the key differences between the two papers is how *confounding variables* are handled. A confounding variable is an unmeasured variable that directly affects both the output and the predictor inputs. They are well studied in the statistics literature (see Pearl [2009] for instance). The question we address in this paper is: can the conditions of Dankers et al. [2016] be relaxed when handling confounding variables?

⋆ The work of A. Dankers is supported by Mitacs of Canada. The work of P. Van den Hof and H. Weerts is supported by the European Research Council (ERC), Advanced Research Grant SYSDYNET, under the European Union's Horizon 2020 research and innovation programme (grant agreement No 694504).

2. BACKGROUND

In this section we present dynamic networks, confounding variables, and the *Direct Method* for identifying a module embedded in a dynamic network.

2.1 Dynamic Networks

We use the framework presented in Van den Hof et al. [2013], where a dynamic network is built up of L nodes, related to L scalar *internal variables* w_j, $j = 1, \dots, L$. Each internal variable is such that it can be written as:

$$w_j(t) = \sum_{k \in \mathcal{N}_j} G^0_{jk}(q) w_k(t) + r_j(t) + v_j(t) \tag{1}$$

where $G^0_{jk}(q)$, $k \in \mathcal{N}_j$ is a proper transfer function, q^{-1} is the delay operator (i.e. $q^{-1}u(t) = u(t-1)$) and,

- $\mathcal{N}_j$ is the set of indices of internal variables with direct connections to w_j, i.e. $i \in \mathcal{N}_j$ iff $G^0_{ji} \neq 0$;
- v_j is an unmeasured *process disturbance* variable modelled as a stationary stochastic process with rational spectral density, i.e. $v_j = H^0_j(q) e_j$ where e_j is a white noise process, and H^0_j is a monic, stable, minimum phase transfer function;
- r_j is an *external variable* that is known and can be manipulated by the user.

If a disturbance and/or external variable are not present at a node, the corresponding v_k or r_k term is set to zero. The entire network is defined by:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G^0_{12} & \cdots & G^0_{1L} \\ G^0_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & G^0_{L-1\ L} \\ G^0_{L1} & \cdots & G^0_{L\ L-1} & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_L \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_L \end{bmatrix},$$

Using an obvious notation this results in:

$$w = G^0 w + r + v \tag{2}$$

where, w, r and v are vectors. For the remainder of the paper we assume there are no external variables present in the network. This simplifies the notation and reasoning.

A *path* from $w_i \to w_j$ means there are transfer functions such that $G_{jn_1} G_{n_1 n_2} \cdots G_{n_k i}$ is non-zero. A *loop* is a path from $w_j \to w_j$. Let w_k, $k \in \mathcal{D}$ and w_z, $z \in \mathcal{Z} = \{1, \dots, L\} \setminus \mathcal{D}$ be a partition of internal variables. A *path from* $w_i \to w_j$ *that passes only through nodes in* $\mathcal{Z}$ is defined as a path for which $G_{jz_1} G_{z_1 z_2} \cdots G_{z_n i}$ are non zero and all z_k are in $\mathcal{Z}$. A direct path between w_i and w_j (i.e. $G_{ji} \neq 0$ is also considered a path that passes only through nodes in $\mathcal{Z}$. A path from w_i to w_j is *blocked* by w_k if the path passes through w_k.

The following assumption will hold throughout this paper.

Assumption 1. (General Assumptions).

(a) The network is *well posed.*
(b) The process noise variables are mutually uncorrelated, i.e. the power spectral density of v is diagonal.
(c) The process noise is full rank, i.e. the power spectral density is full rank for almost all $\omega \in [-\pi, \pi)$.
(d) $(I - G^0)^{-1}$ is stable.

The focus of this paper is how to handle confounding variables when estimating a particular transfer function embedded in a network. In the probabilistic inference literature, a confounding variable is defined as an unmeasured variable that causally affects both w_i and w_j [Pearl, 2009]. Here, we formulate the same concept in our framework.

Definition 1. (Confounding Variables). Consider a dynamic network as defined by (2). Partition the nodes into the following three sets: a set of variables w_k, $k \in \mathcal{A}$, a variable w_j, $j \notin \mathcal{A}$ and the set $\mathcal{Z} = \{1, \dots, L\} \setminus (\mathcal{A} \cup \{j\})$. Given a partition $\mathcal{A}$, $\{j\}$ and $\mathcal{Z}$, construct the set $\mathcal{C}(j, \mathcal{A})$ such that for every $c \in \mathcal{C}(j, \mathcal{A})$, v_c satisfies the following conditions:

(a) There is a path from v_c to w_j that passes only through nodes in $\mathcal{Z}$.
(b) There is a path from v_c to any w_k, $k \in \mathcal{A}$ that passes only through nodes in $\mathcal{Z}$.

A variable v_c, $c \in \mathcal{C}(j, \mathcal{A})$ is a *confounding variable.* □

Remark 2. Whether or not a variable v_k is a confounding variable depends on the choice of j and $\mathcal{A}$. To emphasize this dependence we use the notation v_c, $c \in \mathcal{C}(j, \mathcal{A})$. □

Example 1. Consider the network in Fig. 1b. Let $j = 2$, and $\mathcal{A} = \{1\}$. By Definition 1, v_3 is a confounding variable, i.e. $\mathcal{C}(2, \{1\}) = \{3\}$. Note that v_2 is not a confounding variable. Even though there are paths from v_2 to both w_2 and w_1, the path from $v_2 \to w_1$ passes through $\{j\} \cup \mathcal{A} = \{1, 2\}$, and so the conditions for a confounding variable are not satisfied. □

Why are confounding variables important? If they are not accounted for properly, they can induce a correlation in the noise terms of w_j and w_k, $k \in \mathcal{D}$, even if v_j and v_k, $k \in \mathcal{D}$ are all mutually uncorrelated. This is illustrated in the following example.

Example 2. Consider the same setup as in Example 1. The network equations in this case are:

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0 & G^0_{12} & G^0_{13} \\ G^0_{21} & 0 & G^0_{23} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} + \begin{bmatrix} H^0_1 & 0 & 0 \\ 0 & H^0_2 & 0 \\ 0 & 0 & H^0_3 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

where e_1, e_2, and e_3 are mutually uncorrelated. Both w_1 and w_2 can be expressed in terms of only w_1 and w_2 (eliminate w_3 from the expressions) as:

$$w_1 = G^0_{12} w_2 + \breve{v}_1 \quad \text{and} \quad w_2 = G^0_{21} w_1 + \breve{v}_2.$$

where $\breve{v}_1 = v_1 + G^0_{13} v_3$ and $\breve{v}_2 = v_2 + G^0_{23} v_3$. The key observation is that $\breve{v}_1$ and $\breve{v}_2$ are correlated since they are both functions of the confounder v_3. □

Remark 3. Consider a network with uncorrelated process noise variables, i.e. each process noise variable v_k only has a path to w_k. For given sets j and $\mathcal{A}$, the variables v_k, $k \in \mathcal{A} \cup \{j\}$ cannot be confounding variables since every path from these v_k's to any other internal variable w_n, $n \neq k$ necessarily passes through a node w_k, $k \in \mathcal{A} \cup \{j\}$. Therefore, for a network with uncorrelated process noise variables, we always have $\mathcal{C}(j, \mathcal{A}) \cap \big(\mathcal{A} \cup \{j\}\big) = \emptyset$.

When using a direct prediction error identification method, a confounding variable that affects the output to be predicted and the predictor inputs can result in biased estimates. Before illustrating this point, we first present the direct prediction error identification method.

2.2 Prediction Error Identification

In the following text we show how to obtain an estimate of G^0_{ji} embedded in a dynamic network using the *prediction-error framework* [Ljung, 1999, Van den Hof et al., 2013].

Let w_j denote the variable which is to be predicted. Let w_k, $k \in \mathcal{D}_j$ denote the *predictor inputs* (the set of internal variables that will be used to predict w_j). The one-step-ahead predictor for w_j is [Ljung, 1999]:

$$\hat{w}_j(t|t-1,\theta) = \sum_{k\in\mathcal{D}_j} H_j^{-1}(q,\theta)G_{jk}(q,\theta)w_k(t) + \big(1-H_j^{-1}(q,\theta)\big)w_j(t), \quad (3)$$

where $H_j(q,\theta)$ is the noise mode which is parameterized as a monic, stable and minimum phase transfer function. The prediction error is:

$$\begin{aligned}\varepsilon_j(t,\theta) &= w_j(t) - \hat{w}_j(t|t-1,\theta)\\ &= H_j(q,\theta)^{-1}\big(w_j(t) - \sum_{k\in\mathcal{D}_j} G_{jk}(q,\theta)w_k(t)\big). \quad (4)\end{aligned}$$

The unknown parameters, θ, are estimated by minimizing the sum of squared (prediction) errors (SSE):

$$V_j(\theta) = \frac{1}{N}\sum_{t=0}^{N-1}\varepsilon_j^2(t,\theta). \quad (5)$$

where N is the data length. Under standard (weak) assumptions the estimated parameter vector $\hat{\theta}_N$ converges in the number of data N as $\hat{\theta}_N \to \theta^*$ with probability 1 as $N \to \infty$ [Ljung, 1999] where $\theta^* = \arg\min_\theta \bar{\mathbb{E}}[\varepsilon_j^2(t,\theta)]$ and $\bar{\mathbb{E}} := \lim_{N\to\infty}\frac{1}{N}\sum_{t=0}^{N-1}\mathbb{E}$, and $\mathbb{E}$ is the expected value operator. If $G_{jk}(q,\theta^*) = G^0_{jk}$ the module transfer is said to be estimated *consistently*.

The identification method presented in (3) - (5) is referred to as the *Direct Method* [Ljung, 1999, Van den Hof et al., 2013]. The following example illustrates why confounding variables are difficult to handle using the Direct Method.

Example 3. Consider the network shown in Fig. 1c. Suppose the objective is to obtain an estimate of G^0_{21}. Select w_2 as the output to be predicted, and select w_1 as the predictor input. By Definition 1, $\mathcal{C}(2,\{1\}) = \{3\}$, i.e. v_3 is a confounding variable in this case. From (4), the prediction error for this situation is

$$\varepsilon_2(\theta) = H_2(\theta)^{-1}\big(w_2 - G_{21}(\theta)w_1\big).$$

After some algebra, it follows that

$$\varepsilon_2(\theta) = H_2(\theta)^{-1}\Big(\big(G^0_{21} - G_{21}(\theta)\big)v_1 + v_2 + \Big(G^0_{23} + G^0_{21}G^0_{13} - G^0_{13}G_{21}(\theta)\Big)v_3\Big). \quad (6)$$

Because v_1, v_2 and v_3 are uncorrelated, ε_2 consists of three independent terms. For argument's sake, suppose that $v_3 = 0$. Then minimizing the power of ε is achieved by $G_{21}(\theta) = G^0_{21}$ (setting the v_1 term to 0), and $H_2(\theta) = H^0_2$. Now suppose that $v_1 = 0$, and $v_3 \neq 0$. Then minimizing the power of ε_2 is acheived by $G_{21}(\theta) = G^0_{21} + G^0_{23}/G^0_{13}$ (setting the v_3 term to 0) and $H_2(\theta) = H^0_2$. By continuity, when all three noise terms are present the estimate lies between G^0_{21} and $G^0_{21} + G^0_{23}/G^0_{13}$.[1] Thus, the estimate of G^0_{21} is biased due to the confounding variable v_3. □

The tool that we employ to handle confounding variables is to use additional variables in the data generating system as predictor inputs. The availability of additional variables to use as predictor inputs is a unique advantage of dealing with dynamic networks (as opposed to open or closed-loop systems). The key question is: which additional variables need to be included in the predictor to ensure consistent estimates of the module of interest? In Dankers et al. [2016] it is shown that consistent estimates of G^0_{ji} are possible if a set of predictor inputs is selected that has the following properties: (1) all parallel paths from w_i to w_j are blocked by a predictor input, (2) all loops through w_j are blocked by a predictor input, and (3) there are no confounding variables. The first two conditions are handled using a set denoted $\mathcal{A}_j$ and the third property is handled using a set denoted $\mathcal{B}_\mathcal{A}$. In this paper we focus on the third condition. In particular, we show that the third condition can be relaxed. In the following text we summarize the results of Dankers et al. [2016].

Property 1. Consider a set of internal variables w_k, $k \in \mathcal{A}_j$. Let $\mathcal{A}_j$ satisfy the conditions:

(a) $i \in \mathcal{A}_j$, $j \notin \mathcal{A}_j$,
(b) Every path from w_i to w_j except G^0_{ji} is blocked by a node w_k, $k \in \mathcal{A}_j$.
(c) Every loop through w_j is blocked by a node w_k, $k \in \mathcal{A}_j$. □

In Dankers et al. [2016] confounding variables are handled by selecting the internal variable associated with the confounding variable as a predictor input. This is formalized the in following property.

Property 2. Suppose that w_j and a set internal variables w_k, $k \in \mathcal{A}_j$ have been selected. Consider an additional set of internal variables, $\mathcal{B}_\mathcal{A}$. Let $\mathcal{B}_\mathcal{A} = \mathcal{C}(j,\mathcal{A}_j)$, where $\mathcal{C}(j,\mathcal{A}_j)$ satisfies the conditions of Definition 1. □

If a set $\mathcal{B}_\mathcal{A}$ has Property 2, then it is the set of indeces of all confounding variables associated with $\mathcal{A}_j$ and j.

Algorithm 1. (Direct Method for Estimating G_{ji}).

(1) Select w_j as the output variable to be predicted.
(2) Select sets $\mathcal{A}_j$ and $\mathcal{B}_\mathcal{A}$. The set of predictor inputs is w_k, $k \in \mathcal{D}_j = \mathcal{A}_j \cup \mathcal{B}_\mathcal{A}$.
(3) Construct the predictor (3) with w_k, $k \in \mathcal{D}_j$ as the predictor inputs.
(4) Obtain estimates $G_{jk}(q,\theta)$, $k \in \mathcal{D}_j$ and $H_j(q,\theta)$, by minimizing the sum of squared errors (5).

Proposition 4. Consider a dynamic network as defined in (2) that satisfies Assumption 1. Consistent estimates of G^0_{ji} are obtained using Algorithm 1 if:

(a) There is a delay in every loop w_j to w_j,
(b) The set $\mathcal{A}_j$ has Property 1,
(c) The set $\mathcal{B}_\mathcal{A}$ has Property 2,
(d) The power spectral density of $[w_j\ w_{k_1}\ \cdots\ w_{k_n}]^T$, $k_* \in \mathcal{D}_j$ is positive definite for a sufficiently large number of frequencies $\omega_k \in (-\pi,\pi]$.
(e) The parameterization is chosen flexible enough.
(f) $G_{jk}(\theta)$ is parameterized with a delay if all paths from w_k to w_j have a delay. □

[1] It can be shown that $G_{21}(\theta)$ that minimizes (5) when all three noise sources are present is $G^0_{21} + \breve{H}_{21}\breve{H}_{11}^{-1}$, where $\breve{H}_{11}$ and $\breve{H}_{21}$ are elements of the spectral factor of the noise process $[\breve{v}_1\ \breve{v}_2]^T$ as defined in Example 2.

For a proof and further discussion of Proposition 4 see Proposition 7 in Dankers et al. [2016]. In the remainder of this paper we will present some examples that show that Property 2 can be made less restrictive. Through examples, we aim to present a reasoning as to why it is possible to relax this property. At the end of the paper we present a new, less restrictive set of conditions to handle confounding variables.

3. EXAMPLES TO ILLUSTRATE HOW CONFOUNDERS CAN BE HANDLED

In this section we present three examples, each illustrating a different aspect of the proposed conditions. In the first example, we illustrate the first key element of the new conditions: if either the paths from a confounding variable to w_j or to all w_k, $k \in \mathcal{A}_j$ can be blocked then consistent estimates of G_{ji} are possible.

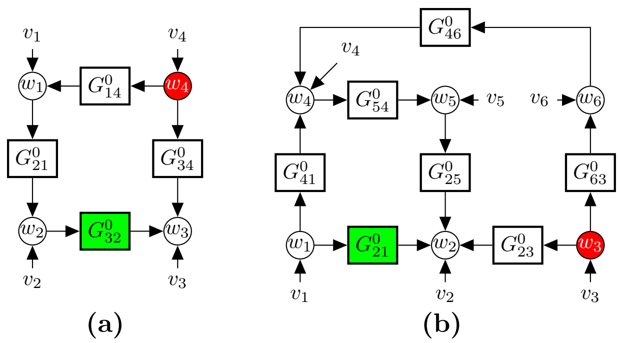

Fig. 2. (a) Networks analyzed in Examples 4 and 5.

Example 4. Consider the network shown in Fig. 2a. The objective is to consistently estimate G^0_{32} without using w_4 as a predictor input (because it is unmeasurable for instance). Select $\mathcal{A}_3 = \{2\}$. In this case $\mathcal{Z} = \{1, 2, 3, 4\} \setminus \{\{j\} \cup \mathcal{A}_j\} = \{1, 4\}$. This set $\mathcal{A}_3$ has Property 1. However, note that there are paths from $v_4 \to w_2$ and $v_4 \to w_3$ that pass only through nodes in $\mathcal{Z}$. Thus, by Definition 1, $\mathcal{C}(3, \{2\}) = \{4\}$, i.e v_4 is a confounding variable in this case. If v_1 were to be zero, then the network considered in this example is the same as that of Example 3. Thus, by the same reasoning, it follows that G^0_{32} cannot be consistently estimated using w_2 as a predictor input and w_3 as the output.

The main problem is that the predictor input w_2 is correlated to the confounding variable v_4. In the following text we show that (due to the topology of the network) it is possible to *partition* w_2 into two components, one is uncorrelated and the other is uncorrelated to the confounding variable. The component of w_2 that is uncorrelated to the confounding variable is then used as a predictor input to identify G^0_{32}. In the second step a consistent estimate of G^0_{32} is obtained because the effect of the confounding variable has been removed from the predictor input.

First, we show how to partition w_2 to obtain a component that is not correlated to v_4. The expression for w_2 is:

$$w_2 = G^0_{21} w_1 + v_2.$$

Thus, if G^0_{21} can be (consistently) estimated then we can obtain an estimate of v_2:

$$w_2^{(\perp w_1)} = w_2 - G_{21}(\hat{\theta}) w_1 = \hat{v}_2.$$

In this situation, v_2 is a component of w_2 that is uncorrelated to w_4. Using w_1 and w_2, the transfer function G^0_{21} can be consistently estimated ($\mathcal{A}_2 = \{1\}$ and $\mathcal{B}_{\mathcal{A}} = \emptyset$ have Properties 1 and 2 respectively, and so by Proposition 4 consistent estimates of G^0_{21} are possible using only w_1 as a predictor input). Now estimate the dynamics from $w_2^{(\perp w_1)}$ to w_3. From (4), using w_3 as the "output", and the estimate of $w_2^{(\perp w_1)}$ as the predictor input, and $H_3(\theta) = 1$, the prediction error is:

$$\varepsilon_3(\theta) = w_3 - G_{32}(\theta) w_2^{(\perp w_1)} = \big(G^0_{32} - G_{32}(\theta)\big) v_2 + p, \quad (7)$$

where in the second equality $p = (G^0_{32} G^0_{21} G^0_{14} + G^0_{34}) v_4 + G^0_{32} G^0_{21} v_1 + v_3$. Because v_1, v_3, and v_4 are uncorrelated to v_2, it follows that p is uncorrelated to $w_2^{(\perp w_1)}$. Consequently, from open loop identification theory, it follows that it is possible to consistently identify G^0_{32} by minimizing the prediction error (7) [Ljung, 1999]. □

In Example 4, we have shown that it is possible to consistently estimate G_{32} using w_1 and w_2, even though $\mathcal{B}_{\mathcal{A}} = \{1\}$ does not have Property 2. The key feature of $\mathcal{B}_{\mathcal{A}} = \{1\}$ is that all paths from the confounder v_4 to w_2, are blocked by w_1. We will show in the main result of the paper that if the paths from the confounding variables to either w_j or w_k, $k \in \mathcal{A}_j$ can be blocked by nodes w_n, $n \in \mathcal{B}_{\mathcal{A}}$, then consistent estimates of G_{ji} are possible.

In the next example we present a second feature of the proposed conditions: the variables selected to block the paths of the confounding variables should not block any paths from $w_i \to w_k$, $k \in \mathcal{A}_j$.

Example 5. Consider the network shown in Fig. 2b. The objective is to consistently estimate G^0_{21} without using w_3. Thus, $j = 2$. Choose $\mathcal{A}_2 = \{1, 5\}$. This set of predictor inputs has Property 1. In this case, by Definition 1 $\mathcal{C}(2, \{1, 5\}) = \{3\}$, i.e. v_3 is a confounding variable (there are paths from $v_3 \to w_2$ and $v_3 \to w_5$ that pass only through nodes in $\mathcal{Z} = \{1, \ldots 6\} \setminus \{\{j\} \cup \mathcal{A}_2\} = \{3, 4, 6\}$).

Using a similar methodology as in Example 4, we want to eliminate the component of the predictor input w_5 that is correlated to the confounder v_3. In Example 4 it was shown that this can be achieved using a variable that blocks the paths from v_3 to the predictor input w_5. The variable w_4 has this property. Now, we follow the same reasoning as in Example 4. From Fig. 2b the equation for w_5 is:

$$w_5 = G^0_{54} w_4 + v_5 \quad (8)$$

From (8) we see that the component of w_5 due to v_5 can be reconstructed if an estimate of G^0_{54} is available, and that v_5 is uncorrelated to the confounder, as desired. By Proposition 4 G^0_{54} can be consistently estimated using w_5 as the output, and w_4 as the predictor input ($\mathcal{A}_5 = \{4\}$ and $\mathcal{B}_{\mathcal{A}} = \emptyset$ have properties 1 and 2 respectively). Thus the component of w_5 due to v_5 is equal to:

$$\hat{w}_5^{(\perp w_4)} = w_5 - G_{54}(\hat{\gamma}) w_4 = v_5.$$

Now we use the (estimated) version of w_5 that is uncorrelated to the counfounder as a predictor input in the primary identification problem in order to obtain an estimate of G^0_{21}. The equation for the prediction error ε_2 using w_1 and $\hat{w}_5^{(\perp w_4)}$ as predictor inputs is:

$$\varepsilon_2(\theta) = w_2 - G_{21}(\theta)w_1 - G_{25}(\theta)w_5^{(\perp w_4)}$$
$$= (G_{21}^0 + G_{25}^0 G_{54}^0 G_{41}^0 - G_{21}(\theta))w_1 + (G_{25} - G_{25}(\theta))w_5^{(\perp w_4)} + p,$$
where $p = v_2 + G_{25}G_{54}v_4 + G_{25}G_{54}G_{46}w_6$. Because p is uncorrelated to the predictor inputs $w_5^{(\perp w_4)}$ and w_1, it follows from open loop system identification theory that consistent estimates of $G_{21}^0 + G_{25}^0 G_{54}^0 G_{41}^0$ and G_{25}^0 will be obtained, not G_{21}^0 as was desired. The reason is that $w_5^{(\perp w_4)} = v_5$ does not block the effect of the parallel path from $w_1 \to w_2$ that passes though w_4 and w_5. The component of w_5 that was blocking the parallel path was removed by removing the component of w_5 due to w_4. The component of w_5 blocking the parallel path is needed, otherwise an estimate of the sum of parallel paths is obtained. Indeed, this is shown in the above prediction error, where an estimate of $G_{21}^0 + G_{25}^0 G_{54}^0 G_{41}^0$ is obtained.

In the following text we show that if an internal variable is selected as an additional predictor input (to handle the effects of the confounder) that both blocks the path from the confounder, v_3 to w_k, $k \in \mathcal{A}_2 = \{1\}$, but it does not block any paths from w_1 to w_2 then a consistent estimate of G_{21}^0 is possible. All effects of the confounder v_3 can be removed from w_5 by removing the component due to w_6 (because there is only one path from w_3 to w_5, and it passes through w_6). The expression for w_5 in terms of w_1 and w_6 is:
$$w_5 = G_{54}^0 G_{41}^0 w_1 + G_{54}^0 G_{46}^0 w_6 + G_{54}^0 v_4 + v_5.$$
Thus we need to obtain an estimate of $G_{54}^0 G_{46}^0$. This is achieved as follows. Choose w_5 as an "output", and w_6 as a predictor input ($\mathcal{A}_5 = \{6\}$ has Property 1 and $\mathcal{B}_\mathcal{A} = \emptyset$ has Property 2 for this choice of $\mathcal{A}_5$). Consequently, by Proposition 4 $G_{54}^0 G_{46}^0$ can be consistently estimated. Denote the estimated transfer function $G_{56}(\hat{\gamma})$. Now the effect of w_3 can be removed from w_5 as follows:
$$\hat{w}_5^{(\perp w_6)} = w_5 - G_{56}(\hat{\gamma})w_6$$
$$= G_{54}^0 G_{41}^0 w_1 + G_{54}^0 v_4 + v_5.$$
The key features of $\hat{w}_5^{(\perp w_6)}$ is that it is still a function of w_1 (still blocks the path from w_1 to w_2) and it is uncorrelated to the confounder w_3. Now construct a predictor of w_2 using w_1 and $\hat{w}_5^{(\perp w_6)}$ as predictor inputs:
$$\varepsilon_2(\theta) = w_2 - G_{21}(\theta)w_1 - G_{25}(\theta)w_5^{(\perp w_6)} + G_{25}^0 w_5^{(w_6)} + G_{23}^0 w_3 + v_2$$
$$= (G_{21}^0 - G_{21}(\theta))w_1 + (G_{25}^0 - G_{25}(\theta))w_5^{(\perp w_6)} + p, \quad (9)$$
where $p = G_{25}^0 w_5^{(w_6)} + G_{23}^0 w_3 + v_2$. In this case the predictor inputs are functions of v_1, v_4, v_5, and the term p is a function of v_6 and v_3. Therefore, p is uncorrelated to both predictor inputs. Again, from open loop identification theory it follows that it is possible to consistently estimate G_{32}^0 by minimizing the prediction error (9). □

The main point of this example is to show that when choosing variables to block a path from a confounder to w_j or a variable w_k, $k \in \mathcal{A}_j$, the selected variable must not block a path from $w_i \to w_k$, $k \in \mathcal{A}_j$.

The last feature of the new conditions is that it can happen that after selecting additional variables to include as predictor inputs, there is a "new" confounding variable, v_k, that has paths to the newly selected predictor inputs, and either w_j or any w_k, $k \in \mathcal{A}_j$. Therefore, sometimes more than one variable may be required to partition the predictor inputs such that the effect of the confounding variable is neutralized. This is illustrated in the following example.

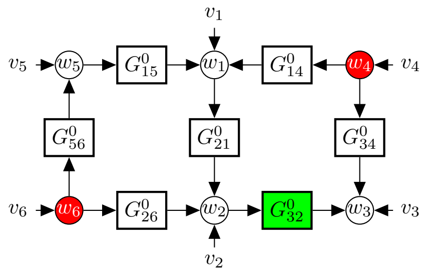

Fig. 3. Network analyzed in Example 6.

Example 6. Consider the network shown in Fig. 3. This network is the network of Example 4 plus two additional internal variables w_5 and w_6. The objective is to consistently estimate G_{32}^0 without using w_4 or w_6. The set $\mathcal{A}_3 = \{2\}$ has Property 1. Recall from Example 4 that we partitioned w_2 into a component uncorrelated to the confounder v_4 using w_1. This was achieved was by estimating G_{21}^0 (consistently). In the network of Figure 3 we are not able to consistently estimate G_{21}^0 using w_2 as an output, and w_1 as a predictor input (as was done in Example 4) because there is a confounder v_6 (by Definition 1 we have $\mathcal{C}(2, \{1\}) = \{6\}$). However, G_{21}^0 can be consistently estimated, if the effect of w_6 is first removed from w_1.

In the following text we show that by using w_2, w_1 and w_5 and solving 3 identification problems, it is possible to obtain a consistent estimate of G_{32}^0.

By Proposition 4, G_{15}^0 can be consistently estimated with w_1 as the output, since $\mathcal{A}_1 = \{5\}$ and $\mathcal{B}_\mathcal{A} = \emptyset$ have Properties 1 and 2 respectively. The expression for w_1 is
$$w_1 = G_{15}^0 w_5 + G_{14}^0 w_4 + v_1.$$
Using the estimated transfer function $G_{15}(\hat{\gamma})$, $w_1^{(\perp w_5)}$ can be constructed as: $w_1^{(\perp w_5)} = w_1 - G_{15}(\hat{\theta})w_5 = G_{14}^0 w_4 + v_1$. It can be verified that $w_1^{(\perp w_5)}$ is only a function of v_1 and v_4, and so it is uncorrelated to v_6.

Now we return to the secondary identification problem where the objective is to consistently estimate G_{21}^0. Choose w_2 as the "output" and $w_1^{(\perp w_5)}$ as the predictor input (recall that $\mathcal{C}(2, \{1\}) = \{6\}$). However, G_{21}^0 can be consistently estimated because the effect of the confounding variable v_6 has been removed from $w_1^{(\perp w_5)}$. Now, using the estimate $G_{21}(\hat{\gamma})$, $w_2^{(\perp w_1)}$ can be constructed as $w_2^{(\perp w_4)} = w_2 - G_{21}(\hat{\gamma})w_1 = G_{26}^0 w_6 + v_2$.

Finally, we return to the primary identification problem with the objective to identify G_{32}^0. Choose w_3 as the output, and $w_2^{(\perp w_1)}$ as the predictor input (recall that $\mathcal{C}(3, \{2\}) = \{4\}$). However, G_{32}^0 can be consistently identified because the effect of the confounding variable v_4 has been removed from $w_2^{(\perp w_1)}$. □

The recursive reasoning as presented in this example can be captured by the following definition.

Definition 5. (Sequence of linked confounders). Consider a network as defined in (2) with a node w_j, and the set of nodes w_k, $k \in \mathcal{A}_j$, $j \notin \mathcal{A}_j$. Consider an additional set of nodes w_n, $n \in \mathcal{L}$, where $\mathcal{L} \cap \{\mathcal{A}_j \cup \{j\}\} = \emptyset$. Let $\mathcal{Z}$ be the set of nodes that are not in $\{j\}$, $\mathcal{A}_j$ or $\mathcal{L}$, (i.e. $\mathcal{L} = \{1, \ldots, L\} \setminus \{\{j\} \cup \mathcal{A}_j \cup \mathcal{L}\}$). The set $\mathcal{L}$ induces a *sequence of linked confounders* v_{z_1}, v_{z_2}, ..., v_{z_n} between w_j and w_k, $k \in \mathcal{A}_j$ if there exists a set of nodes such that

$$\begin{aligned} &v_{z_1} \to w_j, \quad v_{z_1} \to w_{\ell_1}, \\ &v_{z_2} \to w_{\ell_1}, \quad v_{z_2} \to w_{\ell_2}, \\ &v_{z_3} \to w_{\ell_2}, \quad v_{z_3} \to w_{\ell_3}, \\ &\qquad \vdots \\ &v_{z_n} \to w_{\ell_n}, \quad v_{z_n} \to w_k, \quad k \in \mathcal{A}_j \end{aligned}$$

where all $z_k \in \mathcal{Z}$, all $\ell_k \in \mathcal{L}$ and each path passes only through nodes in $\mathcal{Z}$. □

Example 7. Consider the network shown in Fig. 3. Choose $j = 3$, $\mathcal{A}_3 = \{2\}$, and $\mathcal{L} = \{1\}$. In this case $\mathcal{Z} = \{4, 5, 6\}$. The set $\mathcal{L}$ induces the sequence of linked confouders v_4 and v_5 between w_3 and w_2 because there exist paths

$$\begin{aligned} &v_4 \to w_3, \quad v_4 \to w_1, \\ &v_6 \to w_1, \quad v_6 \to w_2, \end{aligned}$$

where all paths pass only through nodes in $\mathcal{Z}$. □

The third feature of the new conditions we propose in this paper is that the set $\mathcal{B}_{\mathcal{A}}$ should not induce a sequence of confounders from any w_k, $k \in \mathcal{A}_j$ to w_j. In the next section we combine all elements for handling confounding variables as highlighted in Examples 4, 5, and 6 into one general property that $\mathcal{B}_{\mathcal{A}}$ should have.

4. GENERALIZATION OF PROPERTY 2

In the previous examples we have illustrated the three features of our proposed conditions to handle confounding variables. Namely that after selecting a set $\mathcal{A}_j$ that has Property 1, additional internal variables must be selected that (a) block the paths from any confounders to either w_j or all w_k, $k \in \mathcal{A}_j$ (b) do not block any paths from $w_i \to w_j$, or any loops through w_j and (c) do not induce a sequence of linked confounders between any w_k, $k \in \mathcal{A}_j$ and w_j. In the examples we used a multi-step approach to demonstrate these core ideas. However, the final algorithm that we propose is not a two step algorithm. We simply choose sets $\mathcal{A}_j$ and $\mathcal{B}_{\mathcal{A}}$ and use all the selected variables w_k, $k \in \mathcal{D}_j = \mathcal{A}_j \cup \mathcal{B}_{\mathcal{A}}$ as the predictor inputs (i.e. we use Algorithm 1). As long as $\mathcal{A}_j$ and $\mathcal{B}_{\mathcal{A}}$ satisfy certain conditions, then consistent estimates of G^0_{ji} will be obtained using Algorithm 1. The proposed conditions that $\mathcal{B}_{\mathcal{A}}$ must satisfy are presented in the following property.

Property 3. Consider a dynamic network as defined in (2), a node w_j and a set of nodes w_k, $k \in \mathcal{A}_j$, $j \notin \mathcal{A}_j$. Let $\mathcal{C}(j, \mathcal{A}_j)$ denote the set of confounding variables as defined in Definition 1. For each v_c, $c \in \mathcal{C}(j, \mathcal{A}_j)$, let $\mathcal{B}_{\mathcal{A}}$ satisfy the following conditions:

(a) For each $c \in \mathcal{V}_{\mathcal{A}}$, one of the following conditions is satisfied:
   (i) all paths from $v_c \to w_j$ are blocked by a node w_n, $n \in \mathcal{B}_{\mathcal{A}}$, or
   (ii) all paths from $v_c \to w_k$, $k \in \mathcal{A}_j$ are blocked by a node w_n, $n \in \mathcal{B}_{\mathcal{A}}$.

(b) $\mathcal{B}_{\mathcal{A}}$ does not induce a sequence of linked confounders between any node w_k $k \in \mathcal{A}_j$ and w_j.
(c) No nodes w_n, $n \in \mathcal{B}_{\mathcal{A}}$ block any paths from w_i to w_k, $k \in \mathcal{A}_j$.
(d) No nodes w_n, $n \in \mathcal{B}_{\mathcal{A}}$ block any loops through w_j. □

The following proposition is a generalization of Proposition 4 that allows for more general conditions when handling confounding variables.

Proposition 6. Consider a network as defined in (2) that satisfies Assumption 1. Consistent estimates of G^0_{ji} are obtained using Algorithm 1 if Condtions (a)-(b), (d)-(f) of Proposition 4 are satisified, and $\mathcal{B}_{\mathcal{A}}$ has Property 3.

The proof is in Appendix A

5. CONCLUSION AND FUTURE WORK

In this paper we have made the conditions for handling confounding variables more flexible. There are likely many connections between the work presented in this paper and other works in the literature. For instance, Properties 1 and Properties 3 likely define a notion similar to the concept of d-separation [Pearl, 2000]. Secondly, the conditions presented in this paper are likely closely related to the conditions for network identifiability as defined in Weerts et al. [2015]. Finally, in this paper we did not investigate the use of external variables for handling confounders, but there is likely a significant advantage in doing so.

REFERENCES

A. Dankers, P. M. J. Van den Hof, P. S. C. Heuberger, and X. Bombois. Identification of dynamic models in complex networks with prediction error methods - predictor input selection. *IEEE Transactions on Automatic Control*, 61(4):937–952, 2016.

J. Gonçalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of LTI networks. *IEEE Transactions on Automatic Control*, 53 (7):1670–1674, August 2008.

A. Haber and M. Verhaegen. Moving horizon estimation for large-scale interconnected systems. *IEEE Transactions on Automatic Control*, 58(11):2834–2847, Nov 2013.

L. Ljung. *System Identification. Theory for the User.* Prentice Hall, 2 edition, 1999.

D. Materassi and G. Innocenti. Topological identification in networks of dynamical systems. *IEEE Transactions on Automatic Control*, 55(8):1860–1871, 2010.

D. Materassi and M.V. Salapaka. Identification of network components in presence of unobserved nodes. In *Proceedings of 54th IEEE Conference on Decision and Control (CDC)*, pages 1563–1568, Osaka, Japan, 2015.

J. Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 40, West 20th Street, New York, 2000.

J. Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009.

P. M. J. Van den Hof, A. Dankers, P. S. C. Heuberger, and X. Bombois. Identification of dynamic models in complex networks with prediction error methods - basic methods for consistent module estimates. *Automatica*, 49:2994–3006, October 2013.

H. H. M. Weerts, A. G. Dankers, and P. M. J. Van den Hof. Identifiability in dynamic network identification. In *Proceedings of the 17th IFAC Symposium on System Identification*, pages 1409–1414, Beijing, P.R. China, 2015.

D. C. Youla. On the factorization of rational matrices. *IRE Transactions on Information Theory*, 7:172–189, July 1961.

Appendix A. PROOF OF PROPOSITION 6

Before presenting the proof of Proposition 6, we present a proposition and a lemma from Dankers et al. [2016] that will be used in the proof.

Proposition 7. Consider a dynamic network as defined in (2) that satisfies Assumption 1. Consider a set of internal variables w_k, $k \in \mathcal{D}_j \cup \{j\}$. There exists a network

$$\begin{bmatrix} w_j(t)\\ w_{\mathcal{D}}(t)\end{bmatrix} = \breve{G}^0(q,\mathcal{D}_j)\begin{bmatrix} w_j(t)\\ w_{\mathcal{D}(t)}\end{bmatrix} + \breve{F}^0(q,\mathcal{D}_j)\begin{bmatrix} v_j(t)\\ v_{\mathcal{D}}(t)\\ v_{\mathcal{Z}}(t)\end{bmatrix}. \qquad \text{(A.1)}$$

where $\breve{G}^0$ and $\breve{F}^0$ are unique transfer matrices given by (arguments q and $\mathcal{D}_j$ have been dropped for clarity):

$$\begin{bmatrix} 0 & \breve{G}^0_{j\mathcal{D}}\\ \breve{G}^0_{\mathcal{D}j} & \breve{G}^0_{\mathcal{D}\mathcal{D}}\end{bmatrix} = \begin{bmatrix} 1-\tilde{G}_{jj} & \\ & I-\mathrm{diag}(\tilde{G}^0_{\mathcal{D}\mathcal{D}})\end{bmatrix}^{-1} \begin{bmatrix} 0 & \tilde{G}^0_{j\mathcal{D}}\\ \tilde{G}^0_{\mathcal{D}j} & \tilde{G}^0_{\mathcal{D}\mathcal{D}}-\mathrm{diag}(\tilde{G}^0_{\mathcal{D}\mathcal{D}})\end{bmatrix}$$

$$\begin{bmatrix} \breve{F}^0_{jj} & 0 & \breve{F}^0_{j\mathcal{Z}}\\ 0 & \breve{F}^0_{\mathcal{D}\mathcal{D}} & \breve{F}^0_{\mathcal{D}\mathcal{Z}}\end{bmatrix} = \begin{bmatrix} 1-\tilde{G}_{jj} & \\ & I-\mathrm{diag}(\tilde{G}^0_{\mathcal{D}\mathcal{D}})\end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & \tilde{F}^0_{j\mathcal{Z}}\\ 0 & I & \tilde{F}^0_{\mathcal{D}\mathcal{Z}}\end{bmatrix}$$

where

$$\begin{bmatrix} \tilde{G}_{jj} & \tilde{G}_{j\mathcal{D}}\\ \tilde{G}_{\mathcal{D}j} & \tilde{G}_{\mathcal{D}\mathcal{D}}\end{bmatrix} = \begin{bmatrix} 0 & G^0_{j\mathcal{D}}\\ G^0_{\mathcal{D}j} & G^0_{\mathcal{D}\mathcal{D}}\end{bmatrix} + \begin{bmatrix} G^0_{j\mathcal{Z}}\\ G^0_{\mathcal{D}\mathcal{Z}}\end{bmatrix}(I-G^0_{\mathcal{Z}\mathcal{Z}})^{-1}\begin{bmatrix} G^0_{\mathcal{Z}j} & G^0_{\mathcal{Z}\mathcal{D}}\end{bmatrix},$$

$$\begin{bmatrix} \tilde{F}_{j\mathcal{Z}}\\ \tilde{F}_{\mathcal{D}\mathcal{Z}}\end{bmatrix} = \begin{bmatrix} G^0_{j\mathcal{Z}}\\ G^0_{\mathcal{D}\mathcal{Z}}\end{bmatrix}(I-G^0_{\mathcal{Z}\mathcal{Z}})^{-1}.$$

□

In Dankers et al. [2016] the network defined by (A.1) is referred to as the *immersed network*.

The following lemma relates the transfer functions $\breve{F}_{mn}$ in the immersed network to a property of the original network. For a proof of the lemma see Dankers et al. [2016]

Lemma 8. Consider a dynamic network as defined in (2) that satisfies Assumption 1. If all paths from v_n to w_m, $n \in \mathcal{Z}$, $m \in \mathcal{D}_j \cup \{j\}$, pass through at least one node w_ℓ, $\ell \in \mathcal{D}_j \cup \{j\}$ (in the original network), then $\breve{F}^0_{mn} = 0$ in the immersed network. □

The following lemma presents a characteristic of the noise model in the immersed network if $\mathcal{B}_{\mathcal{A}}$ has Property 3.

Lemma 9. Consider a network defined by (2) that satisfies Assumption 1 where the variables have been partitioned into sets w_j, w_k, $k \in \mathcal{A}_j$, w_n, $n \in \mathcal{B}_{\mathcal{A}}$, and w_z, $z \in \mathcal{Z}$, where $\mathcal{Z} = \{1,\dots,L\} \setminus \{\mathcal{A}_j \cup \mathcal{B}_{\mathcal{A}} \cup \{j\}\}$. Partition $\mathcal{B}_{\mathcal{A}}$ into two sets, $\mathcal{S}$ and $\mathcal{T}$ according to the following criterion: let w_s, $s \in \mathcal{S}$ be the set of all variables in $\mathcal{B}_{\mathcal{A}}$ for which there are paths $w_z \to w_j$ and $w_z \to w_s$, $z \in \mathcal{Z}$, that pass only through nodes in $\mathcal{Z}$. Let $\mathcal{T}$ be the remaining variables in $\mathcal{B}_{\mathcal{A}}$, i.e. $\mathcal{T} = \mathcal{B}_{\mathcal{A}} \setminus \mathcal{S}$. If $\mathcal{B}_{\mathcal{A}}$ has Property 3 then the noise model of the immersed network will have the structure

$$\Phi_{\breve{v}} = \begin{bmatrix} \Phi_{\breve{v}_j} & \Phi_{\breve{v}_j\breve{v}_{\mathcal{S}}} & 0 & 0\\ \Phi_{\breve{v}_{\mathcal{S}}\breve{v}_j} & \Phi_{\breve{v}_{\mathcal{S}}} & 0 & 0\\ 0 & 0 & \Phi_{\breve{v}_{\mathcal{T}}} & \Phi_{\breve{v}_{\mathcal{T}}\breve{v}_{\mathcal{A}}}\\ 0 & 0 & \Phi_{\breve{v}_{\mathcal{A}}\breve{v}_{\mathcal{T}}} & \Phi_{\breve{v}_{\mathcal{A}}}\end{bmatrix}. \qquad \text{(A.2)}$$

□

Proof: From Proposition 7 the noise model of the immersed network can be expressed as:

$$\begin{bmatrix} \breve{v}_j\\ \breve{v}_{\mathcal{S}}\\ \breve{v}_{\mathcal{T}}\\ \breve{v}_{\mathcal{A}}\end{bmatrix} = \begin{bmatrix} \breve{F}_{jj} & 0 & 0 & 0 & \breve{F}_{j\mathcal{Z}}\\ 0 & \breve{F}_{\mathcal{S}\mathcal{S}} & 0 & 0 & \breve{F}_{\mathcal{S}\mathcal{Z}}\\ 0 & 0 & \breve{F}_{\mathcal{T}\mathcal{T}} & 0 & \breve{F}_{\mathcal{T}\mathcal{Z}}\\ 0 & 0 & 0 & \breve{F}_{\mathcal{A}\mathcal{A}} & \breve{F}_{\mathcal{A}\mathcal{Z}}\end{bmatrix} \begin{bmatrix} v_j\\ v_{\mathcal{S}}\\ v_{\mathcal{T}}\\ v_{\mathcal{A}}\\ v_{\mathcal{Z}}\end{bmatrix}.$$

The power spectral density of $\breve{v}$ is equal to $\breve{H}\breve{H}^*$, where $\breve{H}$ is the matrix defined above. Thus, the expression for $\Phi_{\breve{v}}$ is

$$\begin{bmatrix} |\breve{F}_{jj}|^2+|\breve{F}_{j\mathcal{Z}}|^2 & \breve{F}_{j\mathcal{Z}}\breve{F}^*_{\mathcal{S}\mathcal{Z}} & \breve{F}_{j\mathcal{Z}}\breve{F}^*_{\mathcal{T}\mathcal{Z}} & \breve{F}_{j\mathcal{Z}}\breve{F}^*_{\mathcal{A}\mathcal{Z}}\\ \breve{F}_{\mathcal{S}\mathcal{Z}}\breve{F}^*_{j\mathcal{Z}} & |\breve{F}_{\mathcal{S}\mathcal{S}}|^2+|\breve{F}_{\mathcal{S}\mathcal{Z}}|^2 & \breve{F}_{\mathcal{S}\mathcal{Z}}\breve{F}^*_{\mathcal{T}\mathcal{Z}} & \breve{F}_{\mathcal{S}\mathcal{Z}}\breve{F}^*_{\mathcal{A}\mathcal{Z}}\\ \breve{F}_{\mathcal{T}\mathcal{Z}}\breve{F}^*_{j\mathcal{Z}} & \breve{F}_{\mathcal{T}\mathcal{Z}}\breve{F}^*_{\mathcal{S}\mathcal{Z}} & |\breve{F}_{\mathcal{T}\mathcal{T}}|^2+|\breve{F}_{\mathcal{T}\mathcal{Z}}|^2 & \breve{F}_{\mathcal{T}\mathcal{Z}}\breve{F}^*_{\mathcal{A}\mathcal{Z}}\\ \breve{F}_{\mathcal{A}\mathcal{Z}}\breve{F}^*_{j\mathcal{Z}} & \breve{F}_{\mathcal{A}\mathcal{Z}}\breve{F}^*_{\mathcal{S}\mathcal{Z}} & \breve{F}_{\mathcal{A}\mathcal{Z}}\breve{F}^*_{\mathcal{T}\mathcal{Z}} & |\breve{F}_{\mathcal{A}\mathcal{A}}|^2+|\breve{F}_{\mathcal{A}\mathcal{Z}}|^2\end{bmatrix}$$

In the following text, we will show that the off-block diagonal terms in $\Phi_{\breve{v}}$ are 0 resulting in the required form. First consider the term $\breve{F}_{j\mathcal{Z}}\breve{F}^*_{\mathcal{A}\mathcal{Z}}$. By Condition (a) of Property 3, there cannot be a variable present in $\mathcal{Z}$ that has a path to both w_j and w_k, $k \in \mathcal{A}_j$ that passes only through nodes in $\mathcal{Z}$. Therefore, by Lemma 8 it must be that either $\breve{F}_{j\mathcal{Z}}$ or $\breve{F}_{k\mathcal{Z}}$ equal zero for every node $k \in \mathcal{A}_j$. It follows that $\breve{F}_{j\mathcal{Z}}\breve{F}^*_{k\mathcal{Z}} = 0$ for all $k \in \mathcal{A}_j$.

Next consider the term $\breve{F}_{j\mathcal{Z}}\breve{F}^*_{\mathcal{S}\mathcal{Z}}$. By construction, there is no node w_z, $z \in \mathcal{Z}$ such that there is are paths $w_z \to w_j$ and $w_z \to w_t$, $t \in \mathcal{T}$. If there were such a path, t would be in $\mathcal{S}$. Thus, by Lemma 8, it must be that $\breve{F}_{j\mathcal{Z}}\breve{F}^*_{\mathcal{S}\mathcal{Z}} = 0$.

Next consider the term $\breve{F}_{\mathcal{S}\mathcal{Z}}\breve{F}^*_{\mathcal{A}\mathcal{Z}}$. Recall that for every node w_s, $s \in \mathcal{S}$ there exists a node w_z, $z \in \mathcal{Z}$ with paths $w_z \to w_s$ and $w_z \to w_j$ that pass only through nodes in $\mathcal{Z}$. Suppose that there would exist a variable w_c that has paths $w_c \to w_s$, $s \in \mathcal{S}$ and $w_c \to w_k$, $k \in \mathcal{A}_j$. If such a w_c existed, there would be a sequence of linked confounders from $w_k \to w_j$. By Condition (b) of Property 3, this cannot be. It follows that $\breve{F}_{\mathcal{S}\mathcal{Z}}\breve{F}^*_{\mathcal{A}\mathcal{Z}} = 0$.

Finally, consider the term $\breve{F}_{\mathcal{S}\mathcal{Z}}\breve{F}^*_{\mathcal{T}\mathcal{Z}}$. By construction of $\mathcal{T}$ there is a sequence of linked confounders from each node w_t, $t \in \mathcal{T}$ to a node w_k in $k \in \mathcal{A}_j$. Thus, by reasoning similar as above, it follows that it must be that $\breve{F}_{\mathcal{S}\mathcal{Z}}\breve{F}^*_{\mathcal{T}\mathcal{Z}} = 0$. □

The following is the proof of Proposition 6.

Proof: Let $\mathcal{B}_{\mathcal{A}}$ be partitioned into $\mathcal{S}$ and $\mathcal{T}$, as defined in the statement of Lemma 9. From Proposition 7 the equations for the immersed network using variables w_k, $k \in \mathcal{D}_j$, with $\mathcal{D}_j = \mathcal{A}_j \cup \mathcal{T} \cup \mathcal{S}$ is

$$\begin{bmatrix} w_j\\ w_{\mathcal{S}}\\ w_{\mathcal{T}}\\ w_{\mathcal{A}}\end{bmatrix} = \begin{bmatrix} 0 & \breve{G}_{j\mathcal{S}} & \breve{G}_{j\mathcal{T}} & \breve{G}_{j\mathcal{A}}\\ \breve{G}_{\mathcal{S}j} & \breve{G}_{\mathcal{S}\mathcal{S}} & \breve{G}_{\mathcal{S}\mathcal{T}} & \breve{G}_{\mathcal{S}\mathcal{A}}\\ \breve{G}_{\mathcal{T}j} & \breve{G}_{\mathcal{T}\mathcal{S}} & \breve{G}_{\mathcal{T}\mathcal{T}} & \breve{G}_{\mathcal{T}\mathcal{A}}\\ \breve{G}_{\mathcal{A}j} & \breve{G}_{\mathcal{A}\mathcal{S}} & \breve{G}_{\mathcal{A}\mathcal{T}} & \breve{G}_{\mathcal{A}\mathcal{A}}\end{bmatrix} \begin{bmatrix} w_j\\ w_{\mathcal{S}}\\ w_{\mathcal{T}}\\ w_{\mathcal{A}}\end{bmatrix} + \begin{bmatrix} \breve{v}_j\\ \breve{v}_{\mathcal{S}}\\ \breve{v}_{\mathcal{T}}\\ \breve{v}_{\mathcal{A}}\end{bmatrix} \qquad \text{(A.3)}$$

From Lemma 9 the power spectral density of $\breve{v}$ has the form (A.2). From the Spectral Factorization Theorem of

Youla [1961] we have $\Phi_{\breve v} = \breve H \breve H^*$ where $\breve H$ is a unique, monic, stable, minimum phase spectral factor of $\Phi_{\breve v}$. The upper and lower blocks of $\Phi_{\breve v}$ can be factored as:

$$\Phi_{\breve v}=\begin{bmatrix} F_{11}\Lambda_{11}F_{11}^* & 0 \\ 0 & F_{22}\Lambda_{22}F_{22}^*\end{bmatrix} = \begin{bmatrix}F_{11} & 0\\ 0 & F_{22}\end{bmatrix}\begin{bmatrix}\Lambda_{11}&0\\0&\Lambda_{22}\end{bmatrix}\begin{bmatrix}F_{11}&0\\0&F_{22}\end{bmatrix}^*.$$

In the last equality the left spectral factor is the noise model of the immersed network. Thus the equations for the immersed network (A.3) can be expressed as:

$$\begin{bmatrix}w_j\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal A}\end{bmatrix}=\begin{bmatrix}0 & \breve G_{j\mathcal S} & \breve G_{j\mathcal T} & \breve G_{j\mathcal A}\\ \breve G_{\mathcal S j} & \breve G_{\mathcal S\mathcal S} & \breve G_{\mathcal S\mathcal T} & \breve G_{\mathcal S\mathcal A}\\ \breve G_{\mathcal T j} & \breve G_{\mathcal T\mathcal S} & \breve G_{\mathcal T\mathcal T} & \breve G_{\mathcal T\mathcal A}\\ \breve G_{\mathcal A j} & \breve G_{\mathcal A\mathcal S} & \breve G_{\mathcal A\mathcal T} & \breve G_{\mathcal A\mathcal A}\end{bmatrix}\begin{bmatrix}w_j\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal D}\end{bmatrix}+\begin{bmatrix}\breve H_j & \breve H_{j\mathcal S} & 0 & 0\\ \breve H_{\mathcal S j} & \breve H_{\mathcal S} & 0 & 0\\ 0 & 0 & \breve H_{\mathcal T} & \breve H_{\mathcal T\mathcal A}\\ 0 & 0 & \breve H_{\mathcal A\mathcal T} & \breve H_{\mathcal A}\end{bmatrix}\begin{bmatrix}\breve e_j\\ \breve e_{\mathcal S}\\ \breve e_{\mathcal T}\\ \breve e_{\mathcal A}\end{bmatrix}.$$

In this expression for the network, the noise $\breve v_j$ is correlated to the noise $\breve v_{\mathcal S}$. In the following text we aim to derive an expression for the network, where the noise $\breve v_j$ is uncorrelated to all noise sources. Premultiply both sides of the network equations by

$$\begin{bmatrix}1 & -\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1} & 0 & 0\\ 0 & I & 0 & 0\\ 0 & 0 & I & 0\\ 0 & 0 & 0 & I\end{bmatrix}.$$

The resulting network equations are:

$$\begin{bmatrix}w_j-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}w_{\mathcal S}\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal D}\end{bmatrix}=\begin{bmatrix}\breve G'_{jj} & \breve G'_{j\mathcal S} & \breve G'_{j\mathcal T} & \breve G'_{j\mathcal A}\\ \breve G_{\mathcal S j} & \breve G_{\mathcal S\mathcal S} & \breve G_{\mathcal S\mathcal T} & \breve G_{\mathcal S\mathcal A}\\ \breve G_{\mathcal T j} & \breve G_{\mathcal T\mathcal S} & \breve G_{\mathcal T\mathcal T} & \breve G_{\mathcal T\mathcal A}\\ \breve G_{\mathcal A j} & \breve G_{\mathcal A\mathcal S} & \breve G_{\mathcal A\mathcal T} & \breve G_{\mathcal A\mathcal A}\end{bmatrix}\begin{bmatrix}w_j\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal A}\end{bmatrix}$$
$$+\begin{bmatrix}\breve H_j-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve H_{\mathcal S j} & 0 & 0 & 0\\ \breve H_{\mathcal S j} & \breve H_{\mathcal S} & 0 & 0\\ 0 & 0 & \breve H_{\mathcal T} & \breve H_{\mathcal T\mathcal A}\\ 0 & 0 & \breve H_{\mathcal A\mathcal T} & \breve H_{\mathcal A}\end{bmatrix}\begin{bmatrix}\breve e_j\\ \breve e_{\mathcal S}\\ \breve e_{\mathcal T}\\ \breve e_{\mathcal A}\end{bmatrix}.$$

where

$$\begin{aligned}\breve G'_{jj} &= -\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S j}\\ \breve G'_{j\mathcal S} &= \breve G_{j\mathcal S}-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S\mathcal S}\\ \breve G'_{j\mathcal T} &= \breve G_{j\mathcal T}-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S\mathcal T}\\ \breve G'_{j\mathcal A} &= \breve G_{j\mathcal A}-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S\mathcal A}\end{aligned}$$

After rearanging the terms in the top row, and removing the $\breve G'_{jj}$ term, the resulting network equation is:

$$\begin{bmatrix}w_j\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal A}\end{bmatrix}=\begin{bmatrix}0 & \breve G''_{j\mathcal S} & \breve G''_{j\mathcal T} & \breve G''_{j\mathcal A}\\ \breve G_{\mathcal S j} & \breve G_{\mathcal S\mathcal S} & \breve G_{\mathcal S\mathcal T} & \breve G_{\mathcal S\mathcal A}\\ \breve G_{\mathcal T j} & \breve G_{\mathcal T\mathcal S} & \breve G_{\mathcal T\mathcal T} & \breve G_{\mathcal T\mathcal A}\\ \breve G_{\mathcal A j} & \breve G_{\mathcal A\mathcal S} & \breve G_{\mathcal A\mathcal T} & \breve G_{\mathcal A\mathcal A}\end{bmatrix}\begin{bmatrix}w_j\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal A}\end{bmatrix}+\begin{bmatrix}\breve H'_j & 0 & 0 & 0\\ \breve H_{\mathcal S j} & \breve H_{\mathcal S} & 0 & 0\\ 0 & 0 & \breve H_{\mathcal T} & \breve H_{\mathcal T\mathcal A}\\ 0 & 0 & \breve H_{\mathcal A\mathcal T} & \breve H_{\mathcal A}\end{bmatrix}\begin{bmatrix}\breve e_j\\ \breve e_{\mathcal S}\\ \breve e_{\mathcal T}\\ \breve e_{\mathcal D}\end{bmatrix}.$$

where

$$\breve G''_{j\mathcal S} = \frac{1}{1+\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S j}}\left(\breve G_{j\mathcal S}-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S\mathcal S}+\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\right) \tag{A.4}$$

$$\breve G''_{j\mathcal T} = \frac{1}{1+\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S j}}\left(\breve G_{j\mathcal T}-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S\mathcal T}\right) \tag{A.5}$$

$$\breve G''_{j\mathcal A} = \frac{1}{1+\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S j}}\left(\breve G_{j\mathcal A}-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve G_{\mathcal S\mathcal A}\right) \tag{A.6}$$

$$\breve H'_j = \breve H_j-\breve H_{j\mathcal S}\breve H_{\mathcal S}^{-1}\breve H_{\mathcal S j} \tag{A.7}$$

The same matrix operations can be repeated to remove the term $\breve H_{\mathcal S j}$ from the network equations, resulting in:

$$\begin{bmatrix}w_j\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal A}\end{bmatrix}=\begin{bmatrix}0 & \breve G''_{j\mathcal S} & \breve G''_{j\mathcal T} & \breve G''_{j\mathcal A}\\ \breve G''_{\mathcal S j} & \breve G''_{\mathcal S\mathcal S} & \breve G''_{\mathcal S\mathcal T} & \breve G''_{\mathcal S\mathcal A}\\ \breve G_{\mathcal T j} & \breve G_{\mathcal T\mathcal S} & \breve G_{\mathcal T\mathcal T} & \breve G_{\mathcal T\mathcal A}\\ \breve G_{\mathcal A j} & \breve G_{\mathcal A\mathcal S} & \breve G_{\mathcal A\mathcal T} & \breve G_{\mathcal A\mathcal A}\end{bmatrix}\begin{bmatrix}w_j\\ w_{\mathcal S}\\ w_{\mathcal T}\\ w_{\mathcal A}\end{bmatrix}+\begin{bmatrix}\breve H'_j & 0 & 0 & 0\\ 0 & \breve H_{\mathcal S} & 0 & 0\\ 0 & 0 & \breve H_{\mathcal T} & \breve H_{\mathcal T\mathcal A}\\ 0 & 0 & \breve H_{\mathcal A\mathcal T} & \breve H_{\mathcal A}\end{bmatrix}\begin{bmatrix}\breve e_j\\ \breve e_{\mathcal S}\\ \breve e_{\mathcal T}\\ \breve e_{\mathcal A}\end{bmatrix}.$$

For the purpose of this proof, we are not interested in the specific expressions for $\breve G''_{j\mathcal S}$, $\breve G''_{\mathcal S\mathcal S}$, $\breve G''_{\mathcal S\mathcal T}$, $\breve G''_{\mathcal S\mathcal A}$. The important feature of the above network equations is that the output noise $\breve v'_j = \breve H'_j \breve e_j$ is uncorrelated to all other noise variables in the network. Consequently, by Proposition 4 the transfer functions $[\breve G''_{j\mathcal S}\ \breve G''_{j\mathcal T}\ \breve G''_{j\mathcal A}]$ can be consistently estimated.

The last item that we need to prove is that under the proposed conditions, $\breve G''_{ji} = G_{ji}$. The expression for $\breve G''_{ji}$ is shown in (A.6) (i is always a member of $\mathcal A_j$ by Condition (a) of Property 1). The proof will conclude by showing that $\breve G_{\mathcal S j} = 0$ and $\breve G_{\mathcal S i} = 0$ for all $s \in \mathcal S$.

Start with $\breve G_{\mathcal S j}$. By Condition (d) of Property 3 there is no loop through w_j that passes through a node in $\mathcal B_{\mathcal A}$. There is a path from each w_s, $s \in \mathcal S$ to w_j. Thus, if $\breve G_{\mathcal S j}$ were to be non-zero, then there would be a loop through w_j that passes through a node in $\mathcal B_{\mathcal A}$. This is a contradiction, and so we conclude that $\breve G_{\mathcal S j} = 0$ for all $s \in \mathcal S$. The same reasoning can be applied to conclude that $\breve G_{\mathcal S i} = 0$. This concludes the proof. □