Predictor Input Selection for Two Stage Identification in Dynamic Networks

Arne Dankers, Paul M.J. Van den Hof, Xavier Bombois and Peter S.C. Heuberger

Abstract—Recently, the Two-Stage method has been proposed as a tool to obtain consistent estimates of modules embedded in dynamic networks [1], [2]. However, for this method the variables that are included in the predictor model are currently not considered as a user choice. In this paper it is shown that there is considerable freedom as to which variables can be included in the predictor model as inputs, and still obtain consistent estimates of the module of interest. Conditions that the choice of predictor inputs must satisfy are presented. The conditions could be used to find the smallest number of predictor inputs for instance. Algorithms are presented for checking the conditions and obtaining the estimates.

I. INTRODUCTION

Obtaining models of complex dynamic networks from data is an increasingly important area of research. In many fields of science and engineering such as power systems, biological systems, flexible mechanical structures, economic systems, it is becoming possible to collect data at various locations, or of different variables that have dynamic interrelations (i.e. form an interconnected network). However, in some applications, although it is possible to take measurements at many different locations in the network, it may be expensive to do so. Therefore, there is a motivation to use the minimum number of required measurement locations in order to identify a particular module embedded in a network. Secondly, it may be unsafe, or practically infeasible to measure some variables in the network. Therefore, it would be preferable if it is not necessary to measure these variables in order to obtain estimates of the dynamics of interest.

The question addressed in this paper is: given a dynamic network with known interconnection structure, which variables must be included as inputs in the predictor model so that it is possible to obtain consistent estimates of a particular module of interest embedded in the network? Conditions are presented that the set of predictor inputs must satisfy. In this paper the conditions are derived for the Two-Stage Prediction-Error Method as described in [1], [2].

This problem can also be interpreted as determining which variables should be measured (or where should sensors be placed) in order to obtain consistent estimates of a particular module in the network. Using this interpretation, conditions on the sensor placement scheme are presented such that the Two-Stage Method results in consistent estimates of the module of interest.

If sensors are expensive then these conditions can be used to find the smallest number of required sensors. If obtaining data at particular locations in the network is difficult, then the conditions can be used to obtain a predictor input set which avoids using the particular variables.

There is a growing interest in dynamic network identification, ([3], [4], [5] and references therein). If the interconnection structure is not known, then all variables must be used in the predictor input (no choices based on the interconnection structure can be made). However, in the case where the interconnection structure is known, it becomes possible to choose the set of predictor inputs which are optimal in some sense. The results in this paper build on the results for the Two-Stage Prediction-Error Method found in [2], [1].

In Section II required background material is presented, Section III contains the main result and Section IV contains algorithms for a practical implementation of the result.

II. BACKGROUND

In this section dynamic networks are formally defined, then the Prediction-Error framework and Two-Stage Method are presented, and finally some graph theory is presented.

A. Dynamic Networks and Problem Setup

A dynamic network consists of $L$ internal (or node) variables $\{w_1, \ldots, w_L\}$ that are dynamically interrelated:

$$w_j(t) = \sum_{k \in N_j} G_{jk}^0(q) w_k(t) + r_j(t) + v_j(t)$$

with $G_{jk}^0, k \in N_j$ proper transfer functions, $q^{-1}$ the delay operator ($qu(t) = u(t - 1)$), and

- $N_j$ is the set of indices of node variables with direct causal connections to $w_j$, i.e. $k \in N_j$ if $G_{jk}^0 \neq 0$;
- $v_j$ is an unmeasured variable that is a realization of a stationary stochastic process with rational spectral density: $v_j = H_j^0(q) e_j$ where $e_j$ is white noise, and $H_j^0$ is a monic, stable, minimum phase filter;
- $r_j$ is a external excitation term that is known to the user;

It may be that the disturbance term and/or the external excitation term are not present at some nodes.

All the node variables can be written in one equation as:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G_{12}^0 & \cdots & \cdots & G_{1L}^0 \\ G_{21}^0 & 0 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \ddots & \cdots & \cdots \\ G_{L1}^0 & \cdots & \cdots & 0 & \cdots \\ G_{L2}^0 & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_L \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_L \end{bmatrix}$$
Using an obvious notation this can be expressed as

\[ w = G^0 w + r + H^0 e = (I - G^0)^{-1}(r + v) \]  

(2)

where, if an external variable or disturbance term is absent at node \( i \) then the \( i \)th entry of \( r \) or \( v \) respectively is 0.

A path from \( w_k \rightarrow w_j \) will be understood to mean that there are transfer functions such that \( G_{jn_1} G_{n_1 n_2} \cdots G_{n_k i} \) is nonzero. A loop is a path from \( w_j \rightarrow w_j \).

Since (2) is assumed to describe some physical phenomenon, the equations should be realizable. To characterize the suitability of the equations in describing a physical system, the property of well-posedness is used [6]. The dynamic networks considered are assumed to satisfy the following general conditions.

**Assumption 1:**

(a) The network is well-posed in the sense that all principal minors of \((I - G^0(\infty))\) are non-zero.\(^1\)

(b) \((I - G^0)^{-1}\) is stable.

(c) The noise process \( v \) has a positive semi-definite spectral density, \( \Phi_v(\omega) \geq 0 \), not necessarily diagonal

### B. Prediction Error Identification

The Prediction-Error framework is an identification framework that is based on the one-step-ahead predictor model. See [7] for a detailed description and analysis.

Let \( w_j \) denote the variable which is to be predicted. Let \( w_k, k \in D_j \) and \( r_k, k \in R^p \) denote the predictor inputs (the set of node variables and external variables that will be used to predict \( w_j \)). The one-step-ahead predictor for \( w_j \) is: [7]

\[ \hat{w}_j(t|t-1, \theta) = \sum_{k \in D_j} G_{jk}(q, \theta)w_k(t) + \sum_{k \in R^p} F_{jk}(q, \theta)r_k(t). \]  

(3)

Note that no noise model has been included in the predictor (in the next section it will be shown that noise models are not required for the Two-Stage Method). The prediction error is:

\[ \varepsilon_j(t, \theta) = w_j(t) - \hat{w}_j(t|t-1, \theta) = w_j(t) - \sum_{k \in D_j} G_{jk}(q, \theta)w_k(t) - \sum_{k \in R^p} F_{jk}(q, \theta)r_k(t) \]  

(4)

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

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

(5)

Under standard (weak) assumptions the estimated parameter vector \( \hat{\theta}_N \) converges in the number of data \( N \) as \( N \rightarrow \infty \):

\[ \hat{\theta}_N \rightarrow \theta^* \text{ with probability 1 as } N \rightarrow \infty, \]

where

\[ \theta^* = \arg \min_{\theta} \mathbb{E}[\varepsilon_j^2(t, \theta)] \quad \text{and} \quad \mathbb{E} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^{N-1} \mathbb{E}, \]

\(^1\)This condition is adopted from [6] and imposes weak restrictions on allowable feed-through terms in the network but still allows for the occurrence of algebraic loops.

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 main result of this paper is to derive rules which can help in choosing appropriate sets \( D_j \) and \( R^p \) such that it is possible to obtain a consistent estimate of a particular module \( G^0_{ji} \) embedded in the network using the Two-Stage Method.

### C. Two Stage Method

One problem in the identification in networks is that the noise of the predicted variable \( w_j \) may be correlated to the predictor inputs \( w_k, k \in D_j \) (if there are loops). Without special attention consistent estimates will not be obtained if the noise is correlated to the predictor inputs [8], [9].

The Two-Stage Method is one way to deal with this problem [8], [9], [1], [2]. The method relies on two facts:

1. Any internal variable \( w_k \) can be written as:

\[ w_k = \sum_{m \in R_k} F_{km} r_m + \sum_{m \in V_k} H_{km} v_m = \sum_{m \in R_k} w_m^{(r_m)} + \sum_{m \in V_k} w_m^{(v_m)} \]

where \( R_k \) and \( V_k \) denote the sets of indices of all external and noise variables respectively that have a path to \( w_k \).

A new notation is introduced in the second equality. The term \( w_m^{(r_m)} \) is referred to as the projection of \( w_k \) onto \( r_m \).

2. The transfer function from \( w_m^{(r_m)} \) to \( w_j \) is the same as the transfer function from \( w_k \) to \( w_j \):

\[ w_j = \sum_{k \in D_j} G_{jk}(q, \theta^*) w_k^{(r_m)} + \sum_{k \in R^p} G_{jk}(q, \theta^*) w_k^{(v_m)} + v_j + r_j \]

where \( w_k^{(r_m)} = w_k - w_k^{(r_m)} \).

In the first stage, an estimate of \( w_k^{(r_m)} \) is obtained. This is an open-loop problem since \( r_m \) is assumed to be uncorrelated to all noise sources and all other external variables.

In the second stage, an estimate of \( G_{jk}(q, \theta^*) \) is obtained by estimating the dynamics from \( w_k^{(r_m)} \) to \( w_j \). By the second fact, this transfer is the desired transfer.

The generalized procedure is summarized in the following algorithm adapted from [1], [2]. It has been slightly modified to emphasize that (in this paper) which inputs to include in the predictor model is a choice that the user must make.

**Algorithm 1: Two Stage Method**

1. Choose a set of external excitation signals \( \{ r_m \}, m \in R^s \) to project onto.
2. Choose a set of internal and external variables \( w_k, k \in D_j \), and \( r_k, k \in R^p \) to include as inputs to the predictor.
3. Determine \( w_k^{(R^s_j)} \) for \( k \in D_j \), where \( w_k^{(R^s_j)} = w_k \) is the projection of \( w_k \) onto all \( r_m, m \in R^s \).
4. Construct the predictor

\[ \hat{w}_j(t|t-1, \theta) = \sum_{k \in D_j} G_{jk}(\theta) w_k^{(R^s_j)} + \sum_{k \in R^p} F_{jk}(\theta)r_k. \]

(6)

5. Obtain estimates of \( G_{jk}(q, \theta^*) \) by minimizing the sum of squared prediction errors (5).

Recall the objective is to obtain consistent estimates of a particular module, \( G_{ji}^0 \). In [1] the reasoning is as follows. Choose a set of external variables to project onto that have
paths to $w_j$. Choose $D_j = N_j$ and $R^p_j = \{ j \}$ if $r_j$ is present, otherwise choose $R^p_j = \emptyset$. If a particular $w_k$, $k \in N_j$ is not correlated to any $r_m$, $m \in R^p_j$, remove $k$ from $D_j$. In [1] it is shown that for these choices of $R^p_j$, $R^p_j$ and $D_j$ consistent estimates of $G^j_i$ are possible using Algorithm 1.

This particular choice of predictor inputs works, but it is not the only possible choice. In this paper it will be shown that there is considerable flexibility in the choice of predictor inputs. The chosen set $D_j$ need not even be a subset of $N_j$.

D. Some Results From Graph Theory

Some useful concepts from graph theory are briefly presented. A graph $G$ is made up of nodes interconnected by edges. The set of nodes of $G$ is denoted $V(G)$.

Definition 1 (A-B path): Given a directed graph $G$ and sets of nodes $A$ and $B$. Denote the nodes in the graph $x_k$. A path $P = x_0x_1 \cdots x_k$, where the $x_i$ are all distinct, is an A-B path if $V(P) \cap A = \{x_0\}$, and $V(P) \cap B = \{x_k\}$. [10].

Definition 2 (A-B Separating Set): Consider a directed graph $G$. Given $A, B \subseteq V(G)$, a set $X \subseteq V(G)$ is an A-B separating set if the removal of the nodes in $X$ results in a graph with no $A-B$ paths. [10].

III. PREDICTOR INPUT SELECTION

In this section conditions that the set of predictor inputs must satisfy are presented such that it is possible to obtain a consistent estimate of $G^j_i$. This enables the user to choose a set of signals from a given data set such that the conditions are satisfied. Or, equivalently, it enables the user to place sensors in a network in order to collect the required data.

Before stating the main result, two guiding principles behind the main proposition will be illustrated using a few simple examples. The guiding principles for this paper are

1. choose $D_j$ so that the expression of $w_j$ in terms of $w_k$, $k \in D_j$ is such that the transfer between $w_j$ and $w_k$ is $G^j_i$ (since this will be the dynamics that will be estimated by the predictor) and
2. choose $R^p_j$ such that the unmodeled component of $w_j$ is uncorrelated to the predictor inputs $w_k, k \in D_j$ (i.e. the predictor inputs are uncorrelated to the “noise”).

The following two examples illustrate the first point.

Example 1: Consider the dynamic network described by:

$$
\begin{bmatrix}
  w_1 \\
  w_2 \\
  w_3 \\
  w_4 \\
  w_5
\end{bmatrix} = \begin{bmatrix}
  0 & G^0_{22} & 0 & 0 & 0 \\
  G^0_{42} & 0 & G^0_{23} & 0 & G^0_{25} \\
  G^0_{43} & 0 & 0 & 0 & 0 \\
  G^0_{41} & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & G^0_{54} & 0
\end{bmatrix} \begin{bmatrix}
  w_1 \\
  w_2 \\
  w_3 \\
  w_4 \\
  w_5
\end{bmatrix} + \begin{bmatrix}
  v_1 + r_1 \\
  v_2 \\
  v_3 \\
  v_4 \\
  v_5 + r_5
\end{bmatrix}
$$

Suppose the objective is to obtain an estimate of $G^0_{21}$. Expressing $w_2$ in terms of only $w_1$ plus a disturbance:

$$
w_2 = (G^0_{21} + G^0_{23}G^0_{31} + G^0_{25}G^0_{54}G^0_{41})w_1 + G^0_{25}v_3 + G^0_{25}(v_5 + r_5) + G^0_{25}v_4 + v_2 \quad (7)
$$

where the relation between $w_1$ and $w_2$ is not $G^0_{21}$ as desired. If a predictor were constructed to predict $w_2$ with only $w_1$ as the input, then from (7), an estimate of the transfer $(G^0_{21} + G^0_{23}G^0_{31} + G^0_{25}G^0_{54}G^0_{41})$ would be obtained. For this example, apparently it is not sufficient to simply estimate the dynamics between $w_1$ and $w_2$. Expressing $w_2$ in terms of both $w_1$, $w_5$, and a disturbance results in:

$$
w_2 = (G^0_{21} + G^0_{23}G^0_{31})w_1 + G^0_{25}v_3 + G^0_{25}v_5 + v_2
$$

where again, the relationship between $w_1$ and $w_2$ is not $G^0_{21}$ as desired. For this network, the only way to ensure that the dynamics between $w_1$ and $w_2$ are equal to $G^0_{21}$ is to consider the (proper) mapping from \{w_1, w_3, w_4, v, r\} → \{w_2\} or \{w_1, w_3, w_5, v, r\} → \{w_2\} \quad \Box

It will be shown that one internal variable $w_k$ from every independent path $w_i \rightarrow w_j$ must be included as a variable in the proper mapping to $w_j$ to ensure that the dynamical relationship between $w_i$ and $w_j$ is $G^0_{ij}$, as desired.

Example 2: Consider the network described by:

$$
\begin{bmatrix}
  w_1 \\
  w_2 \\
  w_3
\end{bmatrix} = \begin{bmatrix}
  0 & G^0_{22} & 0 \\
  G^0_{21} & 0 & G^0_{23} \\
  0 & G^0_{32} & 0
\end{bmatrix} \begin{bmatrix}
  w_1 \\
  w_2 \\
  w_3
\end{bmatrix} + \begin{bmatrix}
  v_1 + r_1 \\
  v_2 \\
  v_3
\end{bmatrix}
$$

Suppose the objective is to obtain an estimate of $G^0_{21}$. Expressing $w_2$ in terms of $w_1$ plus a disturbance results in:

$$
w_2 = \frac{1}{1 - G^0_{21}G^0_{32}} (G^0_{21}w_1 + G^0_{23}v_3 + v_2).
$$

The relation between $w_1$ and $w_2$ is not $G^0_{21}$ as desired. The reason the factor $\frac{1}{1 - G^0_{21}G^0_{32}}$ appears is due to the loop from $w_2 \rightarrow w_2$. In the proper mapping from \{w_1, w_3, v, r\} → \{w_2\} the relation between $w_1$ and $w_2$ is $G^0_{21}$ as desired. \Box

It will be shown that one internal variable $w_k$ from every independent loop $w_j \rightarrow w_i$ must be included as a variable in the proper mapping to $w_j$ to ensure that the dynamical relationship between $w_i$ and $w_j$ is $G^0_{ij}$, as desired.

The second guiding principle deals with the relation between the unmodeled component of $w_j$ and the chosen predictor inputs $w_k, k \in D_j$. Again it is first illustrated by example and then proved in the main proposition.

Example 3: Recall the network of Example 1. Again, suppose that the objective is to obtain an estimate of $G^0_{21}$. Choose to project the predictor inputs onto both $r_1$ and $r_5$ ($R^p_2 = \{1, 5\}$). Choose to express $w_2$ in terms of $w_1$, $w_3$, and $w_4$ ($D_2 = \{1, 3, 4\}$):

$$
w_2 = G^0_{21}w_1 + G^0_{23}w_3 + G^0_{25}G^0_{54}w_4 + G^0_{25}v_5 + G^0_{25}v_7 + v_2.
$$

The unmodeled term (or noise) of $w_2$ is $G^0_{25}v_7 + G^0_{25}v_5 + v_2$ which is a function of $r_5$. Note also that the predictor inputs are also a function of $r_5$ since the projection of $w_1$, $w_3$ and $w_4$ onto $r_5$ is non-zero. The point is that the noise term is correlated to the predictor inputs. Therefore, biased estimates would result using the choices $D_2 = \{1, 3, 4\}$, $R^p_2 = \{1, 5\}$ and $R^p_2 = \emptyset$.

However, choosing $R^p_2 = \{5\}$ would mean that the transfer between $r_5$ and $w_2$ is modeled, and consequently, the unmodeled component of $w_2$ is only a function of $v_5$ and $v_2$ which are uncorrelated to the predictor inputs. This is the idea behind the second guiding principle. \Box

These ideas are now be formalized and proved.
Proposition 1: Consider a dynamic network as defined in Section II-A that satisfies Assumption 1. Let \( \{r_m\}, m \in R^s_j \) be the external input(s) onto which will be projected. Let \( \{w_k\}, k \in D_j \) and \( \{r_k\}, k \in R^s_j \) be the sets of internal and external variables respectively that are included as inputs to the predictor (6). Consistent estimates of \( G^0_{ji} \), are obtained using Algorithm 1 if the following conditions hold:

(a) The external excitation signals \( r_m \), \( m \in R^s_j \) are uncorrelated to all disturbance signals.
(b) The set \( D_j \) satisfies the following conditions:
   1. \( i \in D_j \), \( j \notin D_j \),
   2. every path from \( w_i \rightarrow w_j \) excluding the path \( G^0_{ji} \) passes through a node \( w_k \), \( k \in D_j \),
   3. every path from \( w_j \rightarrow w_j \) passes through a node \( w_k \), \( k \in D_j \).
(c) The set \( R^p_j \), satisfies the following condition:
   1. \( k \in R^p_j \) if there is a path from \( r_m \rightarrow w_j \), \( m \in R^s_j \) that does not pass through a node \( w_k \), \( k \in D_j \).
   2. \( R^p_j \) is positive definite for \( w \in [-\pi, \pi] \).
   3. The parameterization is chosen flexible enough such that there exists a parameter \( \theta^0 \) which can exactly describe the dynamics \( w_k \rightarrow w_j \), \( k \in D_j \) and \( r_k \rightarrow w_j \), \( k \in R^p_j \).

Before proceeding to the proof of the Proposition, consider an example which highlights the usefulness of the result.

Example 4: Consider the dynamic network in Fig. 1. Suppose the objective is to obtain consistent estimates of \( G_{ji}^0 \) (marked in green) and \( r_1 \), is chosen as the external excitation signal to project onto \( (R^s_j = \{1\}) \). From Proposition 1 the set of predictor inputs should satisfy Condition (b). Firstly, there are two paths from \( w_2 \) to \( w_3 \) (excluding \( G^0_{32} \)):

\[
\begin{align*}
  &w_2 \rightarrow w_1 \rightarrow w_4 \rightarrow w_5 \rightarrow w_3 \\
  &w_2 \rightarrow w_1 \rightarrow w_4 \rightarrow w_6 \rightarrow w_3
\end{align*}
\]

Therefore one internal variable from each of the two paths must be included as an input in the predictor. There are no paths from \( w_3 \rightarrow w_3 \) so Condition (b3) is automatically satisfied. Consequently, choosing \( D_3 = \{1, 2\} \) or \( \{2, 4\} \) satisfies Condition (b) (denoted in blue in Fig.1).

There are two paths from \( r_1 \rightarrow w_4 \), both of which pass through \( D_3 \), so choosing \( R^s_j = \emptyset \) satisfies Condition (c).

If sensors are to be placed in the network in order to obtain consistent estimates of \( G^0_{ji} \) then the fewest number of required sensors is 3, and they should be placed so that they measure the predicted variable \( w_3 \) (which is required in Step 5 of Algorithm 1 to calculate the prediction error) and the predictor inputs \( w_2 \) and either \( w_1 \) or \( w_4 \).

By the reasoning in [1] \( w_2 \), \( w_3 \) and \( w_4 \) must be included as inputs in the predictor (6) i.e. \( D_3 = N_3 \). Clearly, the results of Proposition 1 allow for a strictly smaller set of predictor inputs for this example. Moreover, note the set of chosen predictor inputs \( \{w_2, w_3\} \) is not a subset of \( N_3 \).

Note that in order for Condition (d) to hold, there must be a path from at least one \( r_m \), \( m \in R^s_j \) to \( w_j \). If not the \( w_i^{(R^s_j)} = 0 \) and the power spectral density of Condition (d) will not be positive definite.

Another interesting point about Proposition 1 is that \( R^p_j \subseteq D_j \) (i.e. the set of external inputs that must be included in the predictor is a subset of the set of external inputs that the node variables \( w_k \), \( k \in D_j \) are being projected onto). This makes sense in light of Guiding Principle 2: the predictor inputs are functions of \( r_m \), \( m \in R^s_j \), and so the only unmodeled terms of \( w_j \) that could cause a problem (i.e. a correlation to the predictor inputs) are those that are functions \( r_k \), \( k \in R^p_j \).

The following lemma is used to prove Proposition 1.

Lemma 1: Consider a dynamic network with transfer matrix \( G_0 \) that satisfies all conditions of Assumption 1. Let \( G^0_{ab} \), be the \((a, b)\)th entry of \((I - G_0)^{-1}\). If there is no path from \( b \) to \( a \) then \( G^0_{ab} = 0 \).

The result follows directly from Mason’s rule [11]. Next Proposition 1 is proved.

Proof: The proof will proceed as follows:

1. Show that \( w_j \) can be expressed in terms of \( w_k \), \( k \in D_j \):

\[
\begin{align*}
  w_j = & \sum_{k \in D_j} G^0_{jk} w_k + \sum_{k \in R^s_j} \tilde{F}^0_{jk} r_k + \sum_{k \in \mathcal{R}_j \setminus R^s_j} F^0_{jk} r_k + \sum_{k \in \mathcal{V}_j} \bar{H}^0_{jk} v_k
\end{align*}
\]

where the dynamics between \( w_i \rightarrow w_j \) are \( G^0_{ji} \), and \( \mathcal{V}_j \) and \( \mathcal{R}_j \) denote the sets of indices of all noise sources and external variables respectively that have a path to \( w_j \).

2. Show that at the global minimum of \( V_j(\theta) \) it must be that \( G^0_{ji}(q, \theta) \) is not equal to \( G^0_{ji}(q) \).

Step 1. With an abuse of notation, let \( D_j \) denote the set of indices \( k \), such that \( k \notin \{j, D_j\} \). Let \( w_{k_1} \) denote the vector \([w_{k_1}, w_{k_2}, \ldots]\) \( k_1 \in D_j \), and let \( G^0_{ji} \) denote the vector \([G^0_{j1}, G^0_{j2}, \ldots]\) \( k_1 \in D_j \). Let \( v_0 \) and \( r_0 \) denote vectors of size \( \text{card}(D_j) \) where the \( i \)th entry is replaced by a 0 if \( v_i \) or \( r_i \) respectively are not present in the data generating system. Using this notation, the network equations (2) are

\[
\begin{bmatrix}
  w_j \\
  w_{v_0} \\
  w_{r_0}
\end{bmatrix} = \begin{bmatrix}
  0 & G_{j0} & G_{j0} \\
  G_{jv} & G_{jv} & G_{jv} \\
  G_{jr} & G_{jr} & G_{jr}
\end{bmatrix} \begin{bmatrix}
  w_j \\
  v_0 \\
  r_0
\end{bmatrix} + \begin{bmatrix}
  v_j \\
  v_{v_0} \\
  v_{r_0}
\end{bmatrix} + \begin{bmatrix}
  r_j \\
  r_{v_0} \\
  r_{r_0}
\end{bmatrix}.
\]
The variables $w_D^\perp$ can be eliminated from the equations:

$$
\begin{bmatrix}
  w_j
  w_D
\end{bmatrix} = \begin{bmatrix}
  G_{1j} & G_{2j} \\
  G_{1D} & G_{2D}
\end{bmatrix} \begin{bmatrix}
  I - G_{0D} \end{bmatrix}^{-1} \begin{bmatrix}
  G_{1D} & G_{2D}
\end{bmatrix} \begin{bmatrix}
  w_j
  w_D
\end{bmatrix} + \begin{bmatrix}
  G_{1D} & G_{2D}
\end{bmatrix} \begin{bmatrix}
  I - G_{0D} \end{bmatrix}^{-1} (v_\phi + r_\phi) + \begin{bmatrix}
  v_j + r_j
  v_D + r_D
\end{bmatrix}.
$$

(9)

Note that by Assumption 1a the inverse $(I - G_{0D})^{-1}$ exists, and that all transfers in the last line are proper.2

The following reasoning will show how Condition (b) ensures that $G_{ij}^0 = 0$ and $G_{ij}^0 = G_{ij}^0$. Consider the term $G_{ij}^0$:

$$
G_{ij}^0 = G_{ij}^0 (I - G_{0D})^{-1} G_{ij}^0 = \sum_{k_1, k_2 \in D} G_{j k_1}^0 c_{k_1 k_2}^0 G_{k_2 j}^0
$$

(10)

where $c_{k_1 k_2}^0$ is the $(k_1, k_2)$ entry of $(I - G_{0D})^{-1}$. By Lemma 1 if there is no path from $w_{k_2} \to w_{k_1}$, then the transfer $G_{j k_1}^0 c_{k_1 k_2}^0 G_{k_2 j}^0$ is zero. By Condition (b) there is no path from $w_j \to w_j$ that does not pass through at least one node in $D_j$. Therefore, by (10) $\tilde{G}_{ij}^0 = 0$.

Secondly, by (9)

$$
\tilde{G}_{ji} = G_{j1}^0 + G_{jk} (I - G_{0D})^{-1} G_{ij} = G_{j1}^0 + \sum_{k_1, k_2 \in D} G_{j k_1}^0 c_{k_1 k_2}^0 G_{ij}^0.
$$

Again, by Lemma 1 and Condition (b) the second term of the above equation is zero. Consequently, the dynamics between $w_j \to w_j$ are $G_{j1}^0$.

Finally, in order to show that (8) holds, it must be shown that $w_j$ is only a function of $r_k, k \in R^+_j$ and $k \in R^+_{j} \backslash R^+_{j}$. From (9) the transfer between $r_k \to w_j, k \in D_j$ is

$$
\tilde{G}_{jk} = [(G_{jD} (I - G_{0D})^{-1})]_k = \sum_{k \in D} G_{j k} \tilde{G}_{kj}
$$

(11)

where the $[\cdot]_k$ denotes the $k$th entry of the vector in square brackets. By Lemma 1, if there is no path from $r_k, k \in D_j$ to $w_j$, that only passes through nodes in $D_j$, then (11) is zero. If there is such a path, then by Condition (c) $m$ is in $R^+_{j}$ if $m$ is in $R^+_{j}$ and otherwise $m$ is in $R^+_{j} \backslash R^+_{j}$.

From (9) and by combining the last three results, it follows that $w_j$ can be expressed as in (8).

**Step 2.** In order to determine if consistent estimates of $G_{ij}^0$ will result by minimizing (5) it must be determined if the minimizer of

$$
\tilde{V}_j(\theta) = \lim_{N \to \infty} E[V_N(\theta)] = \mathbb{E}[e_j^2(t, \theta)]
$$

is equal to $\tilde{G}_{ij}^0$. Substitute the predictor (6) into (12):

$$
\tilde{V}_j(\theta) = \mathbb{E}\left[\left(\sum_{k \in \tilde{D}_j} G_{jk}(\theta) w_{k}^{(\perp j)} + \sum_{k \in R^+_{j}} F_{jk}(\theta) r_k - s_j\right)^2\right]
$$

$$
= \mathbb{E}\left[\left(\sum_{k \in \tilde{D}_j} \Delta G_{jk}(\theta) w_{k}^{(\perp j)} + \sum_{k \in R^+_{j}} \Delta F_{jk}(\theta) r_k + s_j\right)^2\right]
$$

(12)

where $\Delta G_{jk}(\theta) = G_{jk}^0(q) - G_{jk}(q, \theta)$, and $\Delta F_{jk}(\theta) = F_{jk}(q, \theta) - F_{jk}(q)$. For $s_j = \sum_{k \in \tilde{D}_j} \tilde{G}_{jk} w_k^{(\perp j)} + \sum_{k \in \tilde{D}_j} \tilde{F}_{jk} r_k + \sum_{k \in \tilde{R}^+_{j}} \tilde{H}_{jk} v_k$

where $w_k^{(\perp j)} = w_k - w_k^{(\perp j)}$ and $\tilde{R}^+_{j}$ is the set of indices of all noise terms $w_k$ that are present in the expression of $w_j$ for the particular choice of $D_j$. All terms that make up $s_k$ are uncorrelated to $r_m, m \in \tilde{R}^+_{j}$, which results in the following simplification:

$\tilde{V}_j(\theta) = \mathbb{E}\left[\left(\sum_{k \in \tilde{D}_j} \Delta G_{jk}(\theta) w_k^{(\perp j)} + \sum_{k \in \tilde{R}^+_{j}} \Delta F_{jk}(\theta) r_k\right)^2\right] + \mathbb{E}[|s_j|^2]$

Let

$$
\Delta X^+(\epsilon, \omega) = \left|\Delta G_{j1}(\epsilon, \omega) \cdots \Delta G_{jn}(\epsilon, \omega)\right|
$$

then, by Parseval’s theorem, $\tilde{V}_j(\theta)$ is equal to

$$
\frac{1}{2\pi} \int_{-\pi}^{\pi} \Delta X^+(\epsilon, \omega) f_{\tilde{D}_j}(\omega) \Delta X(\epsilon, \omega) + |s_j(\epsilon, \omega)|^2 d\omega.
$$

By Condition (d) $f_{\tilde{D}_j}(\omega)$ is positive definite for all $\omega \in [\pi, \pi]$, therefore both terms in the sum are positive for all $\omega$. Consequently, $\tilde{V}_j(\theta)$ will be a minimum when $\Delta X(\theta) = 0$.

By Condition (e) $\Delta X(\theta) = 0$ when $G_{jk}(q, \theta) = G_{jk}^0(q, D_j)$ and $F_{jk}(q, \theta) = G_{jk}^0(q, D_j)$. Consequently, it follows that $G_{j1}(q, \theta) = G_{j1}^0(q)$.

**IV. ALGORITHMS**

Two algorithms will be presented. The first one provides a way checking Condition (b), or of finding a set $D_j$ that satisfies Condition (b). The second one is an efficient parameterization to obtain the estimates.

Condition (b) can be reformulated using the notion of separating sets defined in Definition 2. Let the node $w_j$ be split into two nodes: $w_j^+$ to which all paths coming into $w_j$ are connected and $w_j^-$ to which all paths leaving $w_j$ are connected. The node $w_j^+$ is connected to $w_j^-$ with the path $G_{j1}^+$ = 1. The set $D_j$ satisfies the following conditions:

1. $D_j \backslash \{i\}$ is a $\{w_i\} \{w_j\}$ separating set for the network with path $G_{ij}^+$ removed,
2. $D_j$ is a $\{w_j^+\} \{w_j^-\}$ separating set for the network with $G_{ij}^+$ removed.

These two conditions can be formulated as a single condition. Let $w_j^+$ and $w_j^-$ be defined similar to $w_j^+$ and $w_j^-$. Then the set $D_j$ is a $\{w_j^+\} \{w_j^-\}$ separating set for the network with edge $G_{ij}^+$ removed.

The advantage of reformulating the conditions in terms of separating sets is that there exist tools from graph theory to check if a given set is a separating set or to find (the smallest possible) separating sets [10], [12]. The following example will illustrate the use of separating sets.

Example 5: Consider the network shown in Fig. 2. Suppose that the objective is to obtain consistent estimates of $G_{21}(\text{denoted in green})$. Both $w_1$ and $w_2$ have been split into two nodes as described above. Choose to project onto all external variables that are present (3, 4, 5, 8).
By the above reasoning using separating sets, Conditions (b2)-(b3) are satisfied for the given network if \( \mathcal{D}_2 \) is a \( \{ w_1, w_2 \} \{ w_3 \} \) separating set for the network with \( \mathcal{G}_0^{(1)} \),\( \mathcal{G}_1^{(1)} \),\( \mathcal{G}_2^{(1)} \),\( \mathcal{G}_3^{(1)} \),\( \mathcal{G}_4^{(1)} \),\( \mathcal{G}_5^{(1)} \),\( \mathcal{G}_6^{(1)} \),\( \mathcal{G}_7^{(1)} \),\( \mathcal{G}_8^{(1)} \),\( \mathcal{G}_9^{(1)} \),\( \mathcal{G}_{10}^{(1)} \),\( \mathcal{G}_{11}^{(1)} \),\( \mathcal{G}_{12}^{(1)} \),\( \mathcal{G}_{13}^{(1)} \) removed. These variables are marked in red in the figure.

There are many possible choices of \( \mathcal{D}_2 \), but the smallest choice is \( \{ w_1, w_6, w_3 \} \) denoted in blue. Again, this is fewer predictor inputs than required in [1].

The second algorithm that is presented provides an efficient method to obtain the estimates. A main theme of this paper is that it is often necessary to include more than just \( w_i \) as an input to the predictor model. However, the dynamics that are estimated in addition to \( \mathcal{G}_0^{(1)} \), are not really of interest. A complicating factor is that the parameterization of all transfers \( \mathcal{G}_{jk}(\theta) \), \( k \in \mathcal{D}_2 \) must be chosen such that the data generating system can be represented by a parameter vector \( \theta_0 \) in the model set (Condition (e) of Proposition 1).

Often (particularly in control applications) it is desirable to parameterize \( \mathcal{G}_0^{(1)} \),\( \mathcal{G}_1^{(1)} \) as

\[
G_{ji}(q, \theta) = \frac{b_{0j}^i q^{-a_{0j}} + \cdots + b_{nj}^i q^{-a_{nj}}}{1 + a_{1j}^i q^{-1} + \cdots + a_{nj}^i q^{-a_{nj}}},
\]

where \( n_a \) and \( n_b \) must be chosen a priori. However, each transfer \( \hat{G}_{jk}(q, \theta), k \in \mathcal{D}_2 \setminus \{ i \} \) could be parameterized as higher order FIR models or using basis functions:

\[
G_{jk}(q, \theta) = \sum_{\ell=0}^n c_{jk}^{\ell} F_{\ell}(q),
\]

where \( F_{\ell}(q) \) are basis functions. Using these types of models does not require much a priori knowledge, and since the models of these transfer functions are not the primary objective of the identification it does not matter that the models are not in the form (13).

An advantage of this parameterization is that the parameters \( c_{jk}^{\ell} \) appear quadratically in the objective function (5) which means a separable least squares algorithm [13] can be used to minimize the objective function.

Remark 1: As a (rough) rule of thumb, the fewer the parameters to estimate, the lower the variance of the parameters [7] as long as the unmodeled disturbance term on \( w_j \) stays the same by removing the parameters. By this reasoning, using fewer predictor inputs may result in estimated parameters with a lower variance.

Consider the network in Example 4. Choose \( \mathcal{D}_3 = \mathcal{N}_3 \). Suppose that 100 tap FIR models are used to parameterize \( \mathcal{G}_0^{(1)} \) and \( \mathcal{G}_0^{(2)} \). Then \( 200 + n_a + n_b + 1 \) parameters need to be estimated.

Next choose \( \mathcal{D}_3 = \{ w_4, w_2 \} \) as suggested in Example 4, then only \( 100 + n_a + n_b + 1 \) parameters need to be estimated. The disturbance term (denoted \( s_3 \) in the proof of the proposition) affecting \( w_8 \) has not changed suggesting that the estimates in the second case will have lower variance.

V. CONCLUSION

Conditions on the predictor inputs have been presented such that it is possible to obtain consistent estimates of a module embedded in a dynamic network using the Two Stage method of identification. This enables the user to design the least expensive sensor placement scheme for instance. Moreover, large networks do not pose a problem since there exist efficient algorithms to find separating sets.

VI. ACKNOWLEDGMENT

Thanks to Yasin Gunes for noticing Mason’s rules are an easy proof of Lemma 1.

REFERENCES


