Identification of dynamic networks operating in the presence of algebraic loops

To appear in Proc. 55th IEEE Conf. Decision and Control, 12-14 December 2016, Las Vegas, AZ.

Identification of dynamic networks operating in the presence of algebraic loops

Harm Weerts, Paul M.J. Van den Hof and Arne Dankers

***Abstract*— When identifying all modules in a dynamic network it is natural to treat all node variables in a symmetric way, i.e. not having pre-assigned roles of 'inputs' and 'outputs'. In a prediction error setting this implies that every node signal is predicted on the basis of all other nodes. A usual restriction in direct and joint-io methods for dynamic network and closed-loop identification is the need for a delay to be present in every loop (absence of algebraic loops). It is shown that the classical one-step-ahead predictor that incorporates direct feedthrough terms in models can not be used in a dynamic network setting. It has to be replaced by a network predictor, for which consistency results are shown when applied in a direct identification method. The result is a one-stage direct/joint-io method that can handle the presence of algebraic loops. It is illustrated that the identified models have improved variance properties over instrumental variable estimation methods.**

I. INTRODUCTION

In most prediction error identification schemes the roles of 'input' and 'output' are pre-assigned to the data. Typically outputs are predicted by models (predictors) while inputs are not. The inputs typically represent external excitation variables that are not influenced by the system. On the other hand in closed-loop systems and dynamic networks the variables that appear as input to some dynamic module can be generated as output of another dynamic module. A clear distinction between inputs and outputs can no longer be made, and therefore these variables will be called *nodes*.

In this paper we want to detect the topology of a network, i.e. identify the network structure and all the dynamic modules. It seems appropriate to consider all node variables in a symmetric way, i.e. all node variables will be predicted on the basis of all other node variables. For a classical closed-loop system this means that the dynamics of both plant and controller are identified, and the outputs of both plant and controller are predicted. A closed-loop identification method that follows this symmetric approach is the Joint-IO method [1], [2], [3], in which the plant/module dynamics are identified in a two-stage procedure.

Joint-IO methods (and other closed-loop identification methods like the direct method) typically have the requirement that the closed-loop does not operate in the presence of an algebraic loop [1], i.e. $\lim_{z\to\infty} G(z)C(z) = 0$. The conditions are also studied (and slightly relaxed) in [4].

There are several situations in which it is restrictive to discard algebraic loops. This is e.g. the case when considering continuous time systems approximated by discrete time models. Secondly the particular structure of a dynamic network can originate from a structured physical system, where physical variables interact with each other, without the direct presence of a computer-controlled (digital) operation. Direct couplings between physical variables is then a natural situation to take into account, and an assumption on absence of algebraic loops is too restrictive.

It is our goal to formulate an identification method that results in consistent estimates of systems operating in the presence of an algebraic loop. Some methods that can handle the presence of algebraic loops are the IV [5], [6] and Two-Stage methods [7]. For dynamic network versions of these methods see [8], [9]. In both these methods the internal signals are 'projected' onto an external signal which is not affected by noise. However the price is that any excitation coming from noise is not used, and therefore the estimator does not have minimum variance properties. External excitation must be available for these methods to work, and this excitation must be of 'sufficient' power and order of persistence of excitation.

The classical predictor that deals with direct feedthrough terms in models appears to be unsuitable for application in situations where algebraic loops are present. Therefore we need to define a new predictor. This new predictor avoids the algebraic loop problem by making explicit use of external excitation, but only for estimation of the direct feedthrough terms of the modules. It is shown that the direct identification scheme with our new predictor leads to consistent estimates of the network even in the presence of an algebraic loop, and actually can be interpreted as a generalized single-stage joint-io method.

First the general dynamic network will be defined (Section II). Then the problems with the traditional predictor are investigated (Section III-A), and the new predictor is formulated (Section III-B). An identification method is specified based on the parametrized new predictor (Section IV), and consistency of this method is proven (Section V). A discussion on properties of the new method is provided (Section VI), after which results are illustrated by simulations (Section VII), and conclusions are drawn (Section VIII).

II. SYSTEM DEFINITION

A general dynamic network setup [8] will be used to formulate the results. The dynamic network consists of L *internal variables* or *nodes* $w_1(t), \cdots, w_L(t)$. We do not distinguish between 'inputs' and 'outputs', all nodes are

Harm Weerts and Paul M.J. Van den Hof are with the Dept. of Electrical Engineering, Eindhoven University of Technology, The Netherlands h.h.m.weerts, p.m.j.vandenhof@tue.nl

Arne Dankers is with the Electrical and Computer Engineering Dept. at the University of Calgary, Canada, adankers@hifieng.com

treated symmetrically. The network is defined by the equation

$$w(t) = G(q)w(t) + R(q)r(t) + H(q)e(t), \tag{1}$$

where

- nodes $w(t) \in \mathbb{R}^L$;
- q is the forward shift operator: $qw(t) = w(t+1)$;
- $r(t) \in \mathbb{R}^K$, are K quasi-stationary *external variables* that can directly be manipulated by the user;
- $G \in \mathbb{R}^{L\times L}(z)$, $R \in \mathbb{R}^{L\times K}(z)$, are proper rational transfer function matrices;
- $H(q)e(t)$ is the *process noise* affecting the nodes, it is modeled as a stationary stochastic process with rational spectral density;
- $e(t) \in \mathbb{R}^L$, a stationary white noise process with diagonal covariance matrix $\Gamma > 0$.

The diagonal of $G(q)$ is 0, i.e. nodes are not directly connected to themselves. $H \in \mathbb{R}^{L\times L}(z)$ is monic, stable and stably invertible.

In this paper we are modeling the relations between measured variables only. When actually the network contains some unmeasured (hidden) nodes, then we are identifying the so-called immersed network [9], [10]. Some methods to identify the hidden nodes are available in literature, e.g. [11].

A transfer function G_{ij} has a *direct term* or *feedthrough* denoted

$$G_{ij}^{\infty} := \lim_{z\to\infty} G_{ij}(z).$$

As a shorthand notation we use $G^{\infty} := \lim_{z\to\infty} G(z)$ to indicate the direct terms of the whole network. When a transfer function G_{ij} has a delay then $G_{ij}^{\infty} = 0$. It is possible that there are algebraic loops in the network. A dynamic network operates in the presence of an algebraic loop if there exists a sequence $n_1, \cdots, n_k$ such that

$$G_{n_1 n_2}^{\infty} G_{n_2 n_3}^{\infty} \cdots G_{n_k n_1}^{\infty} \neq 0.$$

Even though algebraic loops are present, we do require that the network is well-posed. For a discussion on well-posedness see [9]. Well-posedness implies that $(I - G(q))$ and $(I - G^{\infty})$ have a proper inverse.

As an illustrative example system named $\mathcal{S}_1$ we consider the symmetric closed-loop defined in Figure 1. This system is a network consisting of the variables $w(t) = \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}$, $r(t) = \begin{bmatrix} r_1(t) \\ r_2(t) \end{bmatrix}$, $e(t) = \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix}$, and the filters

$$G(q) = \begin{bmatrix} 0 & G_{12}(q) \\ G_{21}(q) & 0 \end{bmatrix},$$
$$H(q) = \begin{bmatrix} H_1(q) & 0 \\ 0 & H_2(q) \end{bmatrix},$$
$$R(q) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

This closed-loop system is similar to the classical loop used in the Joint-IO method [1], however now additional external excitations r_1, r_2 are present. Well-posedness for this loop implies that the direct terms in the loop satisfy

$$G_{12}^{\infty} G_{21}^{\infty} \neq 1. \tag{2}$$

This example will be used as a base-case for developing our approach to the general problem.

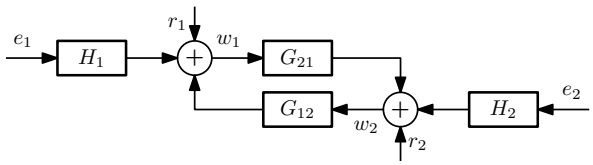

Fig. 1. Symmetrical closed-loop system $\mathcal{S}_1$.

III. PREDICTORS

A. Traditional predictor

When there is a direct feedthrough term present in $G_{21}(q)$ the traditional one-step-ahead predictor for w_2 ([12]) is

$$\breve{w}_2(t|t-1) := \mathbb{E}\{w_2(t) \mid w_2^{t-1}, w_1^t\} \tag{3}$$

where w_i^t refers to the past of w_i up until time moment t. In an open-loop situation ($G_{12} = 0$) or in absence of algebraic loops, this predictor is given by

$$\breve{w}_2(t|t-1) = H_2^{-1}(q)G_{21}(q)w_1(t) + \Big(1 - H_2^{-1}(q)\Big)w_2(t). \tag{4}$$

Note that through the direct feedthrough term G_{21}^{∞}, the predictor will be dependent/conditioned on $w_1(t)$. If we apply this predictor in a situation where $w_1(t)$ is correlated with $w_2(t)$ through a feedthrough term G_{12}^{∞}, the predictor (3) is not given by (4) anymore, and does not have a simple explicit expression in terms of just module dynamics G_{12} and G_{21}.

We can conclude that the traditional predictor for w_2 works fine in open-loop ($G_{12} = 0$) and closed-loop with delay ($G_{12}^{\infty} G_{21}^{\infty} = 0$). When a loop does not contain delay, then the expression beloning to (3) becomes more complicated than (4) and less attractive to use.

B. Network predictor

A new predictor will be defined to replace the traditional predictor. This new predictor has an expression which can be used for identification purposes in a direct method even when there are algebraic loops in the network. The predictor will be multivariable and fully symmetric to take the nature of the dynamic network into account.

Definition 1 (Network predictor): The network predictor is defined as the conditional expectation

$$\hat{w}(t|t-1) := \mathbb{E}\left\{w(t) \middle| w^{t-1}, r^t\right\}. \qquad \square \tag{5}$$

The predictor is conditioned with respect to the delayed values w^{t-1}. This seems strange for a network which contains transfer functions without delay since it seems like we discard useful information that is present in $w(t)$. But by leaving out $w(t)$ we are sure that $e(t)$ is not part of the predictor. Additional conditioning onto r^t ensures that we do use information on the feedthrough terms of the modules.

Proposition 1: The network predictor (5) is given by (omitting arguments q, t)

$$\begin{aligned} \hat{w}(t|t-1) = &\Big(I - (I - G^{\infty})^{-1} H^{-1}(I - G)\Big)w + \\ &+ (I - G^{\infty})^{-1} H^{-1} R r. \qquad \square \end{aligned} \tag{6}$$

The proof of the proposition is included in appendix A.

Predictor expression $\hat{w}$ has some interesting properties. Note that through the factor $(I-G^{\infty})^{-1}$, the filter $(I-G^{\infty})^{-1}H^{-1}(I-G)$ becomes monic, which implies that the predictor filter of w is strictly proper. This conforms to the conditioning on w^{t-1} that is used in the predictor definition (5).

The innovation related to the network predictor $\hat{w}$ is

$$\hat{e}(t) := w(t) - \hat{w}(t|t-1), \tag{7}$$

which is equal to a scaled version of the driving noise process

$$\hat{e}(t) = (I-G^{\infty})^{-1}e(t). \tag{8}$$

Due to the scaling of $e(t)$ the innovation is correlated over the channels, but it is still a white noise.

IV. IDENTIFICATION SETUP

An identification setup will be introduced which has similar structure as the direct method for multivariable systems [5]. The difference with the direct method is the definition of the predictor. First the network predictor $\hat{w}$ will be parameterized, and then an identification criterion is defined.

Definition 2: [13] A network model set is defined as a set of parametrized transfer functions:

$$\mathscr{M} := \{G(q,\theta),\, H(q,\theta),\, R(q,\theta),\, \theta \in \Theta\}$$

with $G \in \mathbb{R}^{L\times L}(z)$, with zeros on the diagonal and all entries proper, $H \in \mathbb{R}^{L\times L}(z)$, monic, diagonal, stable and stably invertible, $R \in \mathbb{R}^{L\times K}(z)$, all entries proper. □

Particular models will be indicated with $M(\theta) := \{G(q,\theta),H(q,\theta),R(q,\theta)\}$. The direct terms of the model are part of the dynamics of $G(q,\theta)$ and are $G^{\infty}(\theta) := \lim_{z\to\infty} G(z,\theta)$.

With the use of the parameterized model set the parameterized predictor is defined as

$$\hat{w}(t|t-1;\theta) := W_w(q,\theta)w(t) + W_r(q,\theta)r(t) \tag{9}$$

with

$$\begin{aligned} W_w(q,\theta) &= I-(I-G^{\infty}(\theta))^{-1}H^{-1}(q,\theta)(I-G(q,\theta)) \\ W_r(q,\theta) &= (I-G^{\infty}(\theta))^{-1}H^{-1}(q,\theta)R(q,\theta). \end{aligned} \tag{10}$$

The prediction error is defined as $\hat{\varepsilon}(t,\theta) := w(t) - \hat{w}(t|t-1;\theta)$ such that we obtain

$$\hat{\varepsilon}(t,\theta) = (I-G^{\infty}(\theta))^{-1}\varepsilon(t,\theta), \tag{11}$$

with

$$\varepsilon(t,\theta) = H^{-1}(q,\theta)\Big((I-G(q,\theta))w(t) - R(q,\theta)r(t)\Big). \tag{12}$$

In [13] ε was used as the prediction error. With this new predictor we use $\hat{\varepsilon}$, being a scaled version of ε.

As identification criterion a weighted least squares criterion will be applied:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta), \tag{13a}$$

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} \hat{\varepsilon}^T(t,\theta)\Lambda^{-1}\hat{\varepsilon}(t,\theta), \tag{13b}$$

where the matrix $\Lambda > 0$ is chosen by the user. If the 'true system' is in the model set and indicated by θ_0, then for $\theta = \theta_0$ the prediction error is a white noise. More precisely, for $\theta = \theta_0$ the prediction error is the innovation

$$\hat{\varepsilon}(t,\theta_0) = \hat{e}(t) = (I-G_{\infty})^{-1}e(t). \tag{14}$$

In order to reduce the variance of the estimator an appropriate choice for Λ is the covariance matrix of $\hat{e}(t)$ being given by $(I-G_{\infty})^{-1}\Gamma(I-G_{\infty})^{-T}$. This will be further commented upon in section VI.

V. CONSISTENCY

In this section we will show that the estimator $\mathscr{M}(\hat{\theta}_N)$ produces consistent estimates.

One of the conditions for consistency is that the data must contain sufficient information on the system. We define *informative data* in line with [12].

Definition 3 (Informative data): A quasi-stationary data sequence is called informative with respect to the model set $\mathscr{M}$ if for any two $\theta_1, \theta_2 \in \Theta$

$$\bar{\mathbb{E}}\left(\hat{\varepsilon}(t,\theta_1) - \hat{\varepsilon}(t,\theta_2)\right)^T \Lambda^{-1}\left(\hat{\varepsilon}(t,\theta_1) - \hat{\varepsilon}(t,\theta_2)\right) = 0 \tag{15}$$

implies that $\hat{W}_w(e^{i\omega},\theta_1) = \hat{W}_w(e^{i\omega},\theta_2)$ and $\hat{W}_r(e^{i\omega},\theta_1) = \hat{W}_r(e^{i\omega},\theta_2)$ for almost all ω. □

In classical direct and joint-io methods the presence of external signals r is not strictly necessary for arriving at informative data. In our new setup the presence of signals r is necessary to be informative in view of the presence of direct feedthrough terms in the models.

The next property that we need is that the identification criterion has a unique solution. To this end we use a property, denoted as global network identifiability, to reflect the unique relationship between predictor filters and model structure.

Definition 4: [13] A network model set is globally network identifiable if for all $\theta_1, \theta_2 \in \Theta$ the following implication holds

$$\left.\begin{aligned} W_w(q,\theta_1) &= W_w(q,\theta_2) \\ W_r(q,\theta_1) &= W_r(q,\theta_2) \end{aligned}\right\} \Rightarrow M(\theta_1) = M(\theta_2). \quad \square \tag{16}$$

This is a definition on dynamics level, not on parameter level. A precise investigation of network identifiability is outside the scope of this paper, for a thorough discussion see [13], [14], [15]. A sufficient condition for global network identifiability is that $R(q,\theta) = I$

The final step in this section is to prove consistency of the estimator $M(\hat{\theta}_N)$.

Theorem 1: $M(\hat{\theta}_N)$ is a consistent estimate if all of the following conditions are satisfied:

1) The network system is in the model set, i.e. $\exists\, \theta_0 \in \Theta$ such that $G(q,\theta_0) = G(q)$, $H(q,\theta_0) = H(q)$, and $R(q,\theta_0) = R(q)$;
2) The data is informative with respect to $\mathscr{M}$;
3) The model set $\mathscr{M}$ is globally network identifiable. □

The proof of the theorem is included in appendix B

VI. DISCUSSION

We will now further interpret and comment on the results that have been derived.

The identification setup that we have chosen is basically a direct identification method that apparently can estimate a dynamic network / closed-loop system while algebraic loops are present. The basic step that we have made in this respect is to exclude direct feedthrough terms in the predictor models from the node variables w, but we include the direct feedthrough terms in the predictor models from the external signals r. As a result, when algebraic loops are present, the presence of an external excitation signal is necessary. However different from the alternative projection methods (IV, two-stage), we keep on using the full signals w as predictor inputs, rather than projecting them onto external signals first. This has two consequences:

- Firstly, the requirements on the persistence of excitation properties of the external signals will be limited, as the r signals only serve to identify the direct feedthrough terms;
- Secondly, the variance of the estimated models will be driven by the signal power of the w signals, rather than by their projections onto r, thus substantially improving the variance of the estimate.

Since external excitation is used for identification of the direct terms, one could wonder whether it would be necessary to have external excitation signals on all node variables when only a few loops are algebraic. This question can be answered by the conditions under which the model set is network identifiable. For an in-depth reasoning on network identifiability see [13], [15]. In short we can say that for the symmetric closed-loop system $\mathcal{S}_1$ we need r_j to be present when $G_{ij}^{\infty}(\theta) \neq 0$. This implies that two external excitations are necessary when the closed-loop contains an algebraic loop.

One could wonder whether it would be necessary in dynamic network identification to have external excitation signals on all node variables, and to apply the presented network predictor to all node variables. However, without addressing this problem in detail now, it seems feasible to use the traditional predictor for predicting those node variables that are a priori known to have no algebraic loops, while applying the new network predictor for the variables that can be part of an algebraic loop.

Concerning the asymptotic variance of the estimate, it can be stated that minimum variance is achieved when the covariance of the innovations process is used as weighting Λ in the identification criterion. According to [12] the resulting asymptotic (minimum) variance is equal to the asymptotic variance of the maximum likelihood estimator under Gaussian assumptions, resulting in the criterion

$$V_N(\theta) = \frac{1}{N} \det \sum_{t=1}^{N} \hat{\varepsilon}(t,\theta)\hat{\varepsilon}^T(t,\theta). \tag{17}$$

The method introduced in this paper can be linked to the classical results for MIMO identification [12], as the network predictor (5) is actually the same as the MIMO predictor defined in [12]. Network identification can be interpreted as MIMO identification with a structured (grey box) model.

VII. SIMULATION EXAMPLE

The direct identification method with network predictor $\hat{w}$ will be validated by numerical simulations. A comparison to the extended instrumental variable (EIV) method [6] is made since it is one of the methods that can deal with algebraic loops. $\mathcal{S}_1$ is used to generate data, and has the following dynamical components:

$$\begin{aligned} G_{12}(q) &= 0.3 + 0.7q^{-1} + 0.3q^{-2}, \\ G_{21}(q) &= 0.15 + 0.9q^{-1} - 0.5q^{-2}, \\ H_1(q) &= 1, \\ H_2(q) &= 1, \\ R(q) &= I. \end{aligned}$$

Low order FIR filters are used to keep numerical computation relatively easy.

In total two sets of experiments are performed. Each set of experiments is performed on 1000 random realizations of data. The external excitation is known but generated as normally distributed white noise $r_i = \mathcal{N}(0, \sigma_r^2)$, and randomized in each experiment. The first set of experiments is performed with the power of the noise ($e_i = \mathcal{N}(0, \sigma_e^2)$), and the power of the external excitation equal, $\sigma_e^2 = \sigma_r^2 = 1$. In the second set of experiments the external excitation has less power to illustrate the benefit of the additional excitation coming from the noise, $\sigma_e^2 = 1, \sigma_r^2 = 0.01$. For all experiments 1000 data samples are drawn, and initial conditions are 0.

For the identification methods the following setup is used. The prediction error $\hat{\varepsilon}(t,\theta)$ is considered with the criterion defined in (13), and the model set is chosen to contain the network system. As weighting matrix the covariance of the innovation will be used

$$\Lambda = \mathbb{E}\hat{e}_0(t)\hat{e}_0^T(t).$$

Here we use the true covariance matrix to obtain the optimal weighting for the least squares criterion. In case one does not know the true covariance matrix, then a determinant minimization criterion (17) can be used, leading to the same asymptotic variance [12]. The cost function (13) is minimized in Matlab by the function `fmincon()`.

Evaluation of estimates is performed in two ways. First the sample mean of the estimated parameters over the 1000 experiments is calculated. For the first set of experiments we obtain the sample means shown in Figure 2. This figure shows that on average both methods produce unbiased estimates of the true parameters. In order to indicate the variance the Best Fit Ratio (BFR) of all these estimated systems is calculated on a validation data set. The Best Fit Ratio is calculated in the following way

$$BFR = \frac{1}{2}\sum_{i=1}^{2}\left(1 - \frac{\|w_i(t) - \hat{w}_i(t|t-1,\hat{\theta}_N)\|_2}{\|w_i(t) - \text{mean}(w_i(t))\|_2}\right). \tag{18}$$

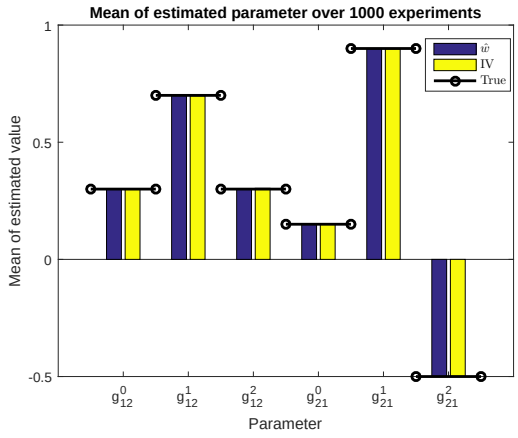

Fig. 2. Sample mean for each parameter over 1000 experiments with $\sigma_r^2 = \sigma_e^2$. The black horizontal lines indicate the parameter value of the true system, the purple (left) bar indicates the mean value of estimated parameters using the network predictor, the yellow (right) bar indicates the mean value of estimated parameters using the EIV method.

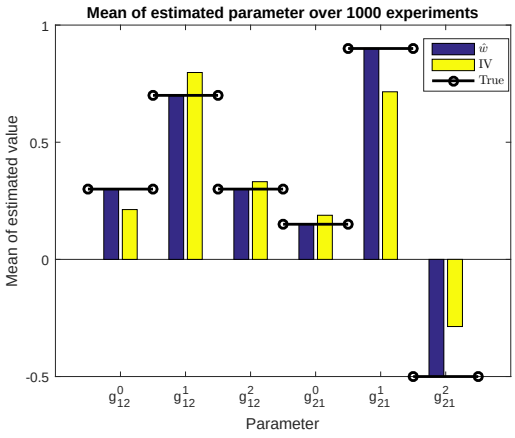

Fig. 4. Sample mean for each parameter over 1000 experiments with $\sigma_r^2 = 0.01\sigma_e^2$. The black horizontal lines indicate the parameter value of the true system, the purple (left) bar indicates the mean value of estimated parameters using the network predictor, the yellow (right) bar indicates the mean value of estimated parameters using the EIV method.

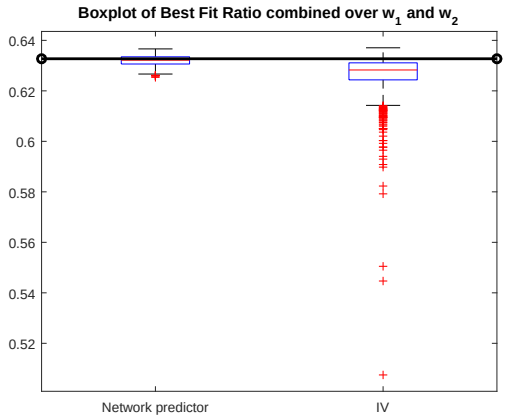

Fig. 3. BFR over 1000 experiments with $\sigma_r^2 = \sigma_e^2$. The horizontal black line is the BFR obtained by the true system on the validation data, the left box indicates the BFR for the estimated network predictors on a validation data set, the right box indicates the BFR for the EIV method on the validation data set.

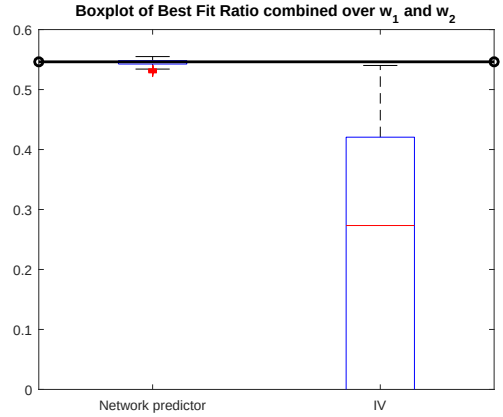

Fig. 5. BFR over 1000 experiments with $\sigma_r^2 = 0.01\sigma_e^2$. The horizontal black line is the BFR obtained by the true system on the validation data, the left box indicates the BFR for the estimated network predictors on a validation data set, the right box indicates the BFR for the EIV method on the validation data set

Best Fit Ratios depend on the noise level of the true system, therefore the BFR of the true system is plotted as a comparison. Best Fit Ratios for the first set of experiments are plotted in Figure 3. In this figure we can see that the network predictor obtains a better fit ratio when compared to the IV when compared over 1000 experiments. The variance of the network predictor is lower than that of the EIV as expected by the amount of excitation used by the estimator.

In the second set of experiments the power of the external excitation is lowered by a factor 100 ($\sigma_r^2 = 0.01$). Again the sample means of the parameter estimates are plotted, see Figure 4. In this figure it can be seen that on average the IV estimate over 1000 experiments does not produce the true parameter values. It is an unbiased estimate, but the external excitation did not have enough power to provide a decent estimation. In contrast the network predictor produces on average still the true parameter values with this low external excitation. A similar mechanic is seen in the BFR plot in Figure 5. The BFR of the network predictor has small variance, all experiments result in a model that is rougly as good as the true system. The BFR of the IV shows that there is a large variance on the estimate, and that many of the estimated models are much worse than the true system. In the case that external excitation is of low power, the IV is significantly outperformed by the network predictor.

VIII. CONCLUSIONS

An identification method has been presented that can consistently identify all modules in a dynamic network while allowing the presence of algebraic loops. The presented method combines a fully symmetric treatment of node variables, as present in classical joint-io methods, with a one-stage algorithm as present in direct identification methods. External excitation signals are necessary in case algebraic

loops are present. The presented method makes maximum use of module excitation through noisy signals in the loops, thereby outperforming projection methods (IV, two-stage) in terms of variance. Numerical aspects of the algorithm, in particular for large scale networks, will need to be considered in future work.

APPENDIX

A. Proof of Proposition 1

In the proof, arguments q and t will be omitted where possible. Starting with (1) add a multiplication with identity after H, and subtract w from both sides of the equation:

$$0 = -(I-G)w + Rr + H(I-G^{\infty})(I-G^{\infty})^{-1}e. \tag{19}$$

Pre-multiplying the equation with $(I-G^{\infty})^{-1}H^{-1}$ and adding w to both sides of the equation delivers

$$\begin{aligned} w = &\left\{I-(I-G^{\infty})^{-1}H^{-1}(I-G)\right\}w + \\ &+(I-G^{\infty})^{-1}H^{-1}Rr + (I-G^{\infty})^{-1}e. \end{aligned} \tag{20}$$

Since the first filter on the right hand side is strictly proper, the second filter is proper, and e is white noise, it follows directly, by applying the definition of the network predictor, that the predictor is given by the first two terms on the right hand side.

B. Proof of Theorem 1

The proof is divided into 3 parts. The first part is the convergence of $V_N(\theta)$ to $\bar{V}(\theta) := \bar{\mathbb{E}}\,\hat{\varepsilon}^T(t,\theta)\Lambda^{-1}\hat{\varepsilon}(t,\theta)$ for $N \to \infty$. This convergence can be shown by applying the convergence proof found in [12].

As second part of the proof it is shown that the true system minimizes the quadratic function $\bar{V}(\theta)$. Rewrite the prediction error in terms of its driving variables with the use of $w = (I-G)^{-1}(He+Rr)$ (omitting arguments q, t)

$$\begin{aligned} &\hat{\varepsilon}(\theta) = \big(I-G_{\infty}(\theta)\big)^{-1}H^{-1}(\theta)\big(I-G(\theta)\big)(I-G)^{-1}H^{-1}e + \cdot \\ &\cdot + \big(I-G_{\infty}(\theta)\big)^{-1}H^{-1}(\theta)\big\{\big(I-G(\theta)\big)(I-G)^{-1}R(q) - R(q,\theta)\big\}r. \end{aligned} \tag{21}$$

The above equation contains a mix of parameterized and non-parameterized transfer functions. In the above equation the e terms can be split into a delayed and non-delayed part

$$\begin{aligned} &\Big\{\big(I-G_{\infty}(\theta)\big)^{-1}H^{-1}(\theta)\big(I-G(\theta)\big)(I-G)^{-1}H^{-1} + \\ &-\big(I-G_{\infty}\big)^{-1}\Big\}e(t) + \big(I-G_{\infty}\big)^{-1}e(t), \end{aligned} \tag{22}$$

where the first e term is delayed, and the second non-delayed. The two terms are uncorrelated since $e(t)$ is a white noise. In the quadratic function $\bar{V}(\theta)$ any cross-term between the non-delayed e term and r or the delayed e term is zero, due to uncorrelatedness. The choice of parameter has no effect on the non-delayed and non-parameterized e term, and it has no cross-terms in the expectation of the quadratic expression. The choice $\theta = \theta_0$ results in the first term in (22) and the second (r-dependent) term in (21) to be equal to 0. This minimizes $\bar{V}(\theta)$ and the prediction error is then equal to the innovation $\varepsilon(t,\theta_0) = \hat{e}(t)$. In the last step we consider any model θ_1 which realizes the same criterion $\bar{V}(\theta)$ as θ_0,

$$\bar{V}(\theta_0) = \bar{V}(\theta_1). \tag{23}$$

It can be shown that

$$\begin{aligned} \bar{V}(\theta_1) - \bar{V}(\theta_0) = &\bar{\mathbb{E}}\,(\hat{\varepsilon}_1 - \hat{\varepsilon}_0)^T\Lambda^{-1}(\hat{\varepsilon}_1 - \hat{\varepsilon}_0) + \\ &+2\bar{\mathbb{E}}(\hat{\varepsilon}_1 - \hat{\varepsilon}_0)^T\Lambda^{-1}\hat{\varepsilon}_0 \end{aligned}$$

where $\hat{\varepsilon}_i := \hat{\varepsilon}(t,\theta_i)$.
Analysing the second term we can use the fact that $\hat{\varepsilon}_0 = e_0(t)$ being a white noise process, while $\hat{\varepsilon}_1 - \hat{\varepsilon}_0$ can be shown to be dependent on data up to $t-1$ only. Therefore this latter term will be uncorrelated with $e(t)$, and the second term in the above equation will be 0, so that

$$\bar{V}(\theta_1) - \bar{V}(\theta_0) = \bar{\mathbb{E}}\,(\hat{\varepsilon}_1 - \hat{\varepsilon}_0)^T\Lambda^{-1}(\hat{\varepsilon}_1 - \hat{\varepsilon}_0). \tag{24}$$

With the condition on informativity of data in Definition 3 it now follows that $\bar{V}(\theta_0) - \bar{V}(\theta_1) = 0$ implies that the corresponding predictor filters must be equal. Then with the use of network identifiability (Definition 4) we know that this must also imply that the models are equal

$$M(\theta_0) = M(\theta_1), \tag{25}$$

which concludes the proof.

REFERENCES

[1] P. E. Caines and C. W. Chan, "Feedback between stationary stochastic processes," *IEEE Trans. Automatic Control*, vol. 20, no. 4, pp. 498–508, 1975.

[2] T. S. Ng, G. C. Goodwin, and B. D. O. Anderson, "Identifiability of mimo linear dynamic systems operating in closed loop," *Automatica*, vol. 13, no. 5, pp. 477–485, 1977.

[3] B. D. O. Anderson and M. R. Gevers, "Identifiability of linear stochastic systems operating under linear feedback," *Automatica*, vol. 18, no. 2, pp. 195–213, 1982.

[4] P. M. J. Van den Hof, D. K. de Vries, and P. Schoen, "Delay structure conditions for identifiability of closed loop systems," *Automatica*, vol. 28, no. 5, pp. 1047–1050, 1992.

[5] T. Söderström and P. Stoica, *System identification*. Prentice-Hall, Inc., 1988.

[6] M. Gilson and P. M. J. Van den Hof, "Instrumental variable methods for closed-loop system identification," *Automatica*, vol. 41, no. 2, pp. 241–249, 2005.

[7] P. M. J. Van den Hof and R. J. P. Schrama, "An indirect method for transfer function estimation from closed loop data," *Automatica*, vol. 29, no. 6, pp. 1523–1527, 1993.

[8] P. M. J. Van den Hof, A. G. 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*, vol. 49, no. 10, pp. 2994–3006, 2013.

[9] A. G. Dankers, "System identification in dynamic networks," PhD dissertation, Delft University of Technology, 2014.

[10] A. G. Dankers, P. M. J. Van den Hof, P. S. C. Heuberger, and X. Bombois, "Identification of dynamic models in complex networks with predictior error methods - predictor input selection," *IEEE Trans. Automatic Control*, vol. 61, no. 4, 2016, to appear.

[11] D. Materassi and M. Salapaka, "Identification of network components in presence of unobserved nodes," in *Proc. 2015 IEEE 54th Conf. Decision and Control, Osaka, Japan*, 2015, pp. 1563–1568.

[12] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[13] H. H. M. Weerts, A. G. Dankers, and P. M. J. Van den Hof, "Identifiability in dynamic network identification," in *Prepr. 17th IFAC Symposium on System Identification*, 2015, pp. 1409–1414.

[14] M. Gevers and A. S. Bazanella, "Identification in dynamic networks: identifiability and experiment design," in *Proc. 2015 IEEE 54th Conf. Decision and Control, Osaka, Japan*, 2015, pp. 4006–4011.

[15] H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers, "Identifiability of dynamic networks with noisy and noise-free nodes," `ArXiv:1609.00864[cs.SY]`, 2016.