Preprints of the 20th World Congress
The International Federation of Automatic Control
Toulouse, France, July 9-14, 2017

Identification of dynamic networks with rank-reduced process noise ⋆

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

* *Control Systems Group, Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands (email: h.h.m.weerts@tue.nl, p.m.j.vandenhof@tue.nl)*
** *Department of Electrical Engineering, University of Calgary, Canada (email: adankers@hifieng.com)*

Abstract: In dynamic network identification usually the assumption is made that there is a full rank process noise affecting the network. For large scale networks with many variables this assumption is not realistic as the noise could be generated by a limited number of sources. We extend prediction error identification methods by allowing rank-reduced process noise in the network. The developed method is based on a modification of the typical predictor expression and an appropriate modification of the identification criterion. It is shown that this method leads to consistent estimates, and we provide a method to reduce the variance of the estimates, which is confirmed by simulations.

Keywords: System identification, dynamic networks, rank-reduced noise

1. INTRODUCTION

Systems which are identified from data are becoming more complex, large scale, and can have an interconnected structure. Examples of these systems can be found in electrical networks, systems biology, and distributed process control. A way of modeling interconnected systems that has been gaining popularity in recent years is to treat the dynamical system as a dynamic network, which includes the structure in the dynamic model. In dynamic networks the traditional input-output structure is replaced by nodes which act as both inputs and outputs of dynamic modules. Identification methods for these networks can be divided into 3 categories: Identification of a particular module in the network [Van den Hof et al., 2013, Gevers and Bazanella, 2015, Materassi and Salapaka, 2015, Dankers et al., 2015, Linder and Enqvist, 2017, Dankers et al., 2016], identification of the full network dynamics [Haber and Verhaegen, 2014, Weerts et al., 2016a], and identification of the interconnection structure of the network [Hayden et al., 2016, Zorzi and Chiuso, 2015].

In the current paper we aim at consistently identifying all dynamic modules in a network. A typical assumption in the dynamic network identification setup is that there are as many independent noise sources as nodes in the network, i.e. the network is excited by a full rank noise process. However, especially when the network consists of many nodes, this assumption might be unrealistic. Rank-reduced noise refers to the situation that there are fewer independent white noise sources than node disturbance signals, and is generally not treated in prediction error identification. Rank-reduced noise does appear in dynamic factor analysis [Deistler et al., 2015] for autoregressive models. In previous work we have allowed some nodes in the network to be completely noise-free [Weerts et al., 2016b,c], which is a special case of rank-reduced process noise. We have also studied the situation of general rank-reduced noise for an open-loop SIMO system [Van den Hof et al., 2017]. It has been shown that appropriate treatment of the rank-reduced noise can lead to significant variance reduction. In [Everitt et al., 2015] an example was shown where rank-reduced noise leads to a noise-free expression from which parameters can be identified variance-free, but the topic was not further elaborated in that paper.

In the current paper we are extending the results in [Van den Hof et al., 2017] from an open-loop SIMO case to dynamic networks of any structure. When modeling a rank-reduced noise as a filtered white noise, restricting the filter to be square and monic does not automatically lead to a unique model. Consequently the predictor expression and model set have to be chosen appropriately. Using an appropriate parameterization for the prediction error we can show that a standard identification criterion leads to consistent estimates. Achieving minimum variance properties of the estimate is typically done by choosing an optimal weight in the criterion. However due to the rank-reduced situation constructing an optimal weight is not straightforward. Based on a maximum likelihood reasoning we define a constrained criterion that reduces variance, or even results in a variance-free estimate. Then we show that this constrained criterion can be relaxed to a weighted least squares criterion with a particular weight.

The paper proceeds with definitions of a dynamic network and rank-reduced noise (Section 2). Then predictors,

⋆ This work has received funding from the European Research Council (ERC), Advanced Research Grant SYSDYNET, under the European Union's Horizon 2020 research and innovation programme (grant agreement No 694504).

model set and prediction errors are formulated (Section 3). Identification criteria are defined and consistency of the estimate is shown (Section 4). Variance-free estimates due to the constraint are discussed (Section 5). The results are illustrated by simulations (Section 6).

2. DYNAMIC NETWORK DEFINITION

Following the basic setup of [Van den Hof et al., 2013], a dynamic network is defined by L scalar *internal variables* or *nodes* w_j, $j = 1, \ldots, L$, and K *external variables* r_k, $k = 1, \cdots K$. Each internal variable is described as:

$$w_j(t) = \sum_{\substack{l=1 \\ l \neq j}}^{L} G^0_{jl}(q) w_l(t) + \sum_{k=1}^{K} R^0_{jk}(q) r_k(t) + v_j(t) \quad (1)$$

where q^{-1} is the delay operator, i.e. $q^{-1}w_j(t) = w_j(t-1)$;

- G^0_{jl} is a stricly proper rational transfer function referred to as a *module* in the network;
- R^0_{jk} are proper rational transfer functions, being the (j,k) elements of R^0;
- r_k are *external variables* that can directly be manipulated by the user.

The full network is constructed by combining all nodes

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

denoted by the matrix equation

$$w = G^0 w + R^0 r + v. \quad (2)$$

The *process noise* is denoted v_j, where the vector process $v = [v_1 \cdots v_L]^T$ is modeled as a stationary stochastic process with rational spectral density, such that there exists a white noise process $e := [e_1 \cdots e_L]^T$ with covariance matrix $\Lambda^0 \geq 0$ of rank p, such that

$$v(t) = H^0_s(q) e(t)$$

with $H^0_s \in \mathbb{R}^{L \times L}(z)$ a proper rational transfer function matrix which is monic, stable, and stably invertible. The process noise is called *rank-reduced* or *singular* if $p < L$, and a node j is called *noise-free* if $v_j(t) = 0$ for all t.

We assume that the nodes are ordered such that the first p nodes are affected by a full rank process noise, i.e. v can be modeled as $v = \left[\begin{smallmatrix} v_a \\ v_b \end{smallmatrix}\right]$ with v_a a p dimensional noise process that is full rank, which is modeled as

$$\begin{bmatrix} v_a \\ v_b \end{bmatrix} = H^0 e_a = \begin{bmatrix} H^0_a \\ H^0_b \end{bmatrix} e_a \quad (3)$$

with H^0 having all poles and zeros inside the unit circle, and H^0_a monic, stable and stably invertible. Define $\lim_{z\to\infty} H^0_b =: \Gamma^0$. Using this noise model the network is written as

$$\begin{aligned} \begin{bmatrix} w_a(t) \\ w_b(t) \end{bmatrix} = & \underbrace{\begin{bmatrix} G^0_{aa}(q) & G^0_{ab}(q) \\ G^0_{ba}(q) & G^0_{bb}(q) \end{bmatrix}}_{G^0(q)} \begin{bmatrix} w_a(t) \\ w_b(t) \end{bmatrix} + \\ & + \underbrace{\begin{bmatrix} R^0_a(q) \\ R^0_b(q) \end{bmatrix}}_{R^0(q)} r(t) + \begin{bmatrix} H^0_a(q) \\ H^0_b(q) \end{bmatrix} e_a(t). \end{aligned} \quad (4)$$

For a dynamic network as defined above, the resulting identification problem then becomes to identify the topology and/or the network dynamics (G^0, R^0, H^0) on the basis of measured node variables $\{w_j, j = 1, \cdots L\}$ and external variables $\{r_k, k = 1, \cdots K\}$. The topology is typically defined as the interconnection structure, specified by the set of modules which are unequal to 0.

3. PREDICTOR AND PREDICTION ERROR

As a first step the one-step-ahead predictor that will be used as a basis in our identification setup is defined.

Definition 1. The one-step-ahead predictor for node signals $w(t)$ is defined as the conditional expectation

$$\hat{w}(t|t-1) := \bar{\mathbb{E}}\left\{ w(t) \mid w^{t-1},\ r^t \right\} \quad (5)$$

where $w^{t-1} := \{w(0), \cdots, w(t-1)\}$, $r^t := \{r(0), \cdots, r(t)\}$.

□

All nodes are conditioned onto the same variables, such that no distinction is made between inputs and outputs and hence the nodes are treated symmetrically. The predictor can be represented by different expressions due to the rank-reduced noise [Weerts et al., 2016b,c].

The noise process v can be modeled in different ways that each have different properties. Three noise models that are used in this paper are

$$\begin{bmatrix} v_a \\ v_b \end{bmatrix} = \underbrace{\begin{bmatrix} H^0_a \\ H^0_b \end{bmatrix}}_{H^0} e_a = \underbrace{\begin{bmatrix} H^0_a & 0 \\ \tilde{H}^0_b & I \end{bmatrix}}_{H^0_s} \begin{bmatrix} e_a \\ e_b \end{bmatrix} = \underbrace{\begin{bmatrix} H^0_a & 0 \\ 0 & I + \tilde{H}^0_b \Gamma^{0\dagger} \end{bmatrix}}_{H^0_d} \begin{bmatrix} e_a \\ e_b \end{bmatrix} \quad (6)$$

where a white noise process e_b has been added, such that H^0_s can be chosen monic, stable, and stably invertible with $\tilde{H}^0_b := H^0_b - \Gamma^0$. Alternatively H^0_d allows for a (block) diagonal model which can be useful in implementation, see Section 6 for an example. The model $H^0_d e$ only exists whenever Γ^0 has a left inverse $\Gamma^{0\dagger}$. In (6) we find that e_a and e_b are strongly related through $H^0_b e_a = \tilde{H}^0_b e_a + e_b$, such that when we subtract $\tilde{H}^0_b e_a$ from both sides we have

$$\Gamma^0 e_a = e_b. \quad (7)$$

The covariance matrix of $e = \left[\begin{smallmatrix} e_a \\ e_b \end{smallmatrix}\right]$ is the singular matrix

$$\Lambda^0 = \begin{bmatrix} \Lambda_a & \Lambda_a \Gamma^{0T} \\ \Gamma^0 \Lambda_a & \Gamma^0 \Lambda_a \Gamma^{0T} \end{bmatrix} = \begin{bmatrix} I \\ \Gamma^0 \end{bmatrix} \Lambda_a \begin{bmatrix} I & \Gamma^{0T} \end{bmatrix}. \quad (8)$$

Modeling v as $H^0 e_a$ has the property that it is driven by a full rank white noise of dimension p, but also that H^0 is not square and does not have a unique inverse. Models H^0_s and H^0_d are monic and invertible, but are driven by a rank-reduced white noise of dimension L. For noise model $v = H^0_s e$ the expression for the one-step-ahead predictor can be given as:

Proposition 1. The one-step-ahead predictor is given by

$$\hat{w}(t|t-1) = w - (H^0_s(q))^{-1} \left((I - G^0(q)) w - R^0(q) r \right). \quad (9)$$

Proof: Can be found by following the predictor derivation steps in [Weerts et al., 2015]. □

Other predictor filters can be found by modeling v using different expressions as in (6). Now that a predictor expression has been obtained it can be parameterized using a set of models.

Definition 2. (network model set). A network model set for a network of L nodes, and K external signals, is defined as a set of parametrized transfer functions:

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

with

$$G(z,\theta) = \begin{bmatrix} G_{aa}(z,\theta) & G_{ab}(z,\theta) \\ G_{ba}(z,\theta) & G_{bb}(z,\theta) \end{bmatrix} \in \mathbb{R}^{L\times L}(z),$$

$$H(z,\theta) = \begin{bmatrix} H_a(z,\theta) \\ H_b(z,\theta) \end{bmatrix} \in \mathbb{R}^{L\times p}(z),$$

$$R(z,\theta) = \begin{bmatrix} R_a(z,\theta) \\ R_b(z,\theta) \end{bmatrix} \in \mathbb{R}^{L\times K}(z)$$

while

- $G_{aa}(z,\theta)$, $G_{bb}(z,\theta)$ have zeros on the diagonal, $G(z,\theta)$ is strictly proper, and $(I-G(z,\theta))^{-1}$ proper and stable;
- $H_a(z,\theta) \in \mathbb{R}^{p\times p}$ monic, $H_b(z,\theta) \in \mathbb{R}^{(L-p)\times p}$ proper;
- $H_s(z,\theta)$ is defined with the same structure as H_s^0, and H_a and H_b are such that $H_s(z,\theta)$ is monic, proper, stable and stably invertible;
- $R(z,\theta)$ is proper.

Additionally $\Gamma(\theta) := \lim_{z\to\infty} H_b(z,\theta)$, and $\tilde{H}_b(z,\theta) := H_b(z,\theta) - \Gamma(\theta)$. One particular model in $\mathcal{M}$ is denoted by $M(\theta) := [G(q,\theta), H(q,\theta), R(q,\theta)]$. □

Requiring $(I - G(z,\theta))^{-1}$ to be proper and stable is induced by requiring all mappings from external and noise signals to node signals to be proper and stable. Predictor (9) will be parameterized using the model set to create the parameterized predictor

$$\begin{aligned} &\hat{w}(t|t-1,\theta) = w(t) + \\ &-(H_s(q,\theta))^{-1}\{(I-G(q,\theta))w(t) - R(q,\theta)r(t)\}. \end{aligned} \tag{10}$$

By choosing this particular form we guarantee that the parameterized transfer functions G_{ba}, G_{bb}, and R_b appear in the parameterized predictor, as this is not necessarily the case for other predictor expressions [Weerts et al., 2016b]). The definition of the prediction error is

$$\varepsilon(t,\theta) = \begin{bmatrix} \varepsilon_a(t,\theta) \\ \varepsilon_b(t,\theta) \end{bmatrix} := \begin{bmatrix} w_a(t) \\ w_b(t) \end{bmatrix} - \begin{bmatrix} \hat{w}_a(t|t-1,\theta) \\ \hat{w}_b(t|t-1,\theta) \end{bmatrix}. \tag{11}$$

It must be noted that the prediction error does not take the direct feedthrough of $H_b(\theta)$ into account, i.e. $\Gamma(\theta)$ is not a part of the prediction error defined above.

4. IDENTIFICATION CRITERION

An estimate of the network is obtained by application of an identification criterion. Typically the asymptotic prediction error identification criterion for multivariable estimators is [Ljung, 1999]

$$\theta^* = \arg\min_\theta \bar{\mathbb{E}}\, \varepsilon^T(t,\theta)\, Q\, \varepsilon(t,\theta) \tag{12}$$

with Q strictly positive definite. One of the conditions for consistent estimates is network identifiability, which ensures that we can distinguish between different network topologies and dynamics. For more details on network identifiability and conditions under which a model set is network identifiable see [Weerts et al., 2015, 2016c]. The following definition is in line with those publications.

Definition 3. Model set $\mathcal{M}$ is globally network identifiable at $M(\theta_0)$ if for all $M(\theta_1) \in \mathcal{M}$ the following implication holds

$$T(\theta_1) = T(\theta_0) \Rightarrow M(\theta_1) = M(\theta_0), \tag{13}$$

where $T(\theta) := (I - G(\theta))^{-1}[H(\theta)\ R(\theta)]$. □

A sufficient condition for network identifiability is when every node has an independent excitation source coming from either noise or external excitation. The consistency result is then as follows.

Proposition 2. Consider a model set $\mathcal{M}$ and data generated by (2) with $M^0(q) := (G^0, H^0, R^0)$. Let $\theta^\star$ be the solution of the criterion (12), Then $G(q,\theta^\star) = G^0(q)$, $H_a(q,\theta^\star) = H_a^0(q)$, $\tilde{H}_b(q,\theta^\star) = \tilde{H}_b^0(q)$, and $R(q,\theta^\star) = R^0(q)$ provided that:

(1) The data generating system is in the model set, i.e. $\exists\theta_0 \in \Theta$ such that $M(q,\theta_0) = M^0(q)$, and
(2) external excitation r is persistently exciting of sufficiently high order, and
(3) $\mathcal{M}$ is globally network identifiable at θ_0.

Proof: Collected in the appendix. □

Note that the prediction errors are not dependent on $\Gamma(\theta)$ and therefore no estimate is made of $\Gamma(\theta)$. This is the reason that $\tilde{H}_b$ is estimated consistently, rather than H_b. Whenever we want to determine Γ^0 in (7) then it can be estimated using the obtained model by

$$\Gamma^\star = \left(\bar{\mathbb{E}}\, \varepsilon_b(\theta^*)\varepsilon_a^T(\theta^*)\right)\left(\bar{\mathbb{E}}\, \varepsilon_a(\theta^*)\varepsilon_a^T(\theta^*)\right)^{-1}. \tag{14}$$

Since $\theta^\star$ is a consistent estimate we have $\varepsilon(\theta^\star) = e$ and

$$\Gamma^\star = \left(\bar{\mathbb{E}}\, e_b e_a^T\right)\left(\bar{\mathbb{E}}\, e_a e_a^T\right)^{-1} = \Gamma^0\Lambda_a\Lambda_a^{-1} = \Gamma^0. \tag{15}$$

Consistency of the estimate has been shown under fairly general conditions. We can conclude that for a network with rank-reduced noise a standard weighted least squares criterion can be used for identification, in combination with a modified noise model. Typically, in order to achieve minimum variance, weight Q in the criterion must be the inverse of the covariance matrix of e, but this is impossible since this covariance matrix is singular. The criterion (12) is based on ε_a and ε_b, without taking into account that the innovation process satisfies the relation (7). Based on a maximum likelihood reasoning we would like to include (7) in the criterion similar to the approach in [Van den Hof et al., 2017]. Therefore a constrained identification criterion is defined which enforces $\Gamma\varepsilon_a = \varepsilon_b$ by including it as a constraint. Let $Q_a > 0$, then the criterion is

$$\begin{aligned} \theta^\star = &\arg\min_\theta \bar{\mathbb{E}}\, \varepsilon_a^T(t,\theta)\, Q_a\, \varepsilon_a(t,\theta) \\ &\text{under constraint: } \Gamma(\theta)\varepsilon_a(t,\theta) = \varepsilon_b(t,\theta) \quad \forall t. \end{aligned} \tag{16}$$

Note that $\Gamma(\theta)$ is parameterized and will be estimated from data. Consistency (now including Γ^0) is shown under the same conditions as for the unconstrained criterion.

Proposition 3. Consider data generated by (2) with $(G^0, H^0, R^0) = M_0(q)$, and let $\theta^\star$ be the solution of the criterion (16). Then $M(q,\theta^\star) = M_0(q)$ provided that:

(1) The datagenerating system is in the model set, i.e. $\exists\theta_0 \in \Theta$ such that $M(q,\theta_0) = M_0(q)$, and
(2) external excitation r is persistently exciting of sufficiently high order,
(3) $\mathcal{M}$ is globally network identifiable at θ_0.

Proof: Collected in the appendix. □

The constraint naturally represents the presence of dependencies in the innovation process. In comparison with the unconstrained criterion (12) we can expect that the variance of the estimate will be reduced.

When considering implementation of the constrained criterion in practice then we must keep in mind that there is the possibility that the constraint is not feasible. In case the true network is not in the model set, or when initial conditions are not taken into account, it is possible that no model in the model set satisfies the constraint. An always feasible relaxed version of the criterion can be constructed by adding the constraint (in a quadratic version) as a weighted penalty term to the optimization problem

$$\theta^\star = \arg\min_\theta \bar{\mathbb{E}}\left\{\varepsilon_a^T(\theta)Q_a\varepsilon_a(\theta) + \lambda Z(\theta)\right\}, \qquad (17)$$

with $Z(\theta) = (\Gamma(\theta)\varepsilon_a(t,\theta) - \varepsilon_b(t,\theta))^T(\Gamma(\theta)\varepsilon_a(t,\theta) - \varepsilon_b(t,\theta))$ which can also be written as criterion (12) with a parameterized weighting matrix

$$Q(\theta) = \begin{bmatrix} Q_a + \lambda\Gamma^T(\theta)\Gamma(\theta) & -\lambda\Gamma^T(\theta) \\ -\lambda\Gamma(\theta) & \lambda I \end{bmatrix}. \qquad (18)$$

For $\lambda \to \infty$ the unconstrained criterion (17) is equivalent to the constrained criterion. In cases where the constraint (from (16)) is not feasible, $\lambda \to \infty$ will likely not be the optimal choice for λ. An analysis of appropriately choosing λ in such a situation is deferred to future work.

5. VARIANCE-FREE ESTIMATION

Improving the variance of the estimate is the main reason for choosing the constrained criterion over the unconstrained criterion. In [Everitt et al., 2015] the following basic example has been shown which allows variance-free estimation of parameters. Consider this system without dynamics

$$y_1(t) = \theta_1^0 u_1(t) + e(t), \quad y_2(t) = \theta_2^0 u_2(t) + e(t). \qquad (19)$$

By using the equality relation of the noise in both equations, we can substitute $e = y_2 - \theta_2^0 u_2$ into the y_1 equation to obtain $y_1(t) - \theta_1^0 u_1(t) = y_2(t) - \theta_2^0 u_2(t)$. This equation can be parameterized with θ_1 and θ_2 such that

$$\theta_1^0 u_1(t) + e(t) - \theta_1 u_1(t) = \theta_2^0 u_2(t) + e(t) - \theta_2 u_2(t). \quad (20)$$

There is no contribution of $e(t)$ and what is left is a variance-free estimation problem.

In the constrained criterion (16), the parameter values are determined through the constraint $\Gamma\varepsilon_a = \varepsilon_b$. Another way to write the constraint is as a filter driven by a full rank process

$$[\Gamma(\theta)\ -I]\, X(\theta) \begin{bmatrix} e_a(t) \\ r(t) \end{bmatrix} = 0 \quad \forall t \qquad (21)$$

for some filter $X(\theta)$, which implies that (provided r is persistently exciting of sufficiently high order)

$$[\Gamma(\theta)\ -I]\, X(\theta) = 0. \qquad (22)$$

Under particular conditions only one model in $\mathcal{M}$ satisfies (22), leading to a variance-free estimate.

As an example for variance-free estimation we use the network $\mathcal{S}_1$ (Figure 1)

$$\begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} = \begin{bmatrix} 0 & G_{12}^0(q) \\ G_{21}^0(q) & 0 \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} + \begin{bmatrix} r_1(t) \\ r_2(t) \end{bmatrix} + v(t) \quad (23)$$

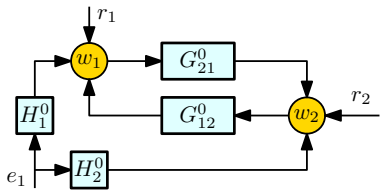

Fig. 1. Dynamic network $\mathcal{S}_1$ with a rank-reduced noise.

with $v(t) = \begin{bmatrix} H_1^0(q) \\ H_2^0(q) \end{bmatrix} e_1$ where $H_2^0(q) = 1 + \tilde{H}_2^0(q)$ such that $\Gamma^0 = 1$. For this network w_a could be either w_1 or w_2, but we choose $w_a = w_1$.

Proposition 4. Consider a 2 node network ($L = 2$) with rank 1 noise ($p = 1$) and 2 external excitations ($R = I$). Let $\mathcal{M}$ be a model set with $\mathcal{S} \in \mathcal{M}$ and $R(\theta) = I$. Then the constraint

$$\Gamma(\theta)\varepsilon_a(t,\theta) = \varepsilon_b(t,\theta)\ \forall t \quad \Rightarrow \quad G(q,\theta) = G^0(q). \qquad (24)$$

Proof: Enclosed in the appendix. □

A variance-free estimate is obtained since the implication holds for every realization of data. The proposition shows that in case of network $\mathcal{S}_1$ only a constraint, and no minimization criterion is needed to obtain a consistent variance-free estimate of G. Weight $Q(\theta)$ from (18) can in that case be approximated with

$$\lim_{\lambda\to\infty} Q(\theta) \propto Q_\infty(\theta) = \begin{bmatrix} \Gamma^T(\theta)\Gamma(\theta) & -\Gamma^T(\theta) \\ -\Gamma(\theta) & I \end{bmatrix} \qquad (25)$$

which is a singular matrix. It must be noted that Proposition 4 can not be generalized to any network. Only in special cases can a variance-free estimate be obtained.

6. SIMULATIONS

In order to illustrate the theoretical results obtained in the previous sections a simulation example is included. Network $\mathcal{S}_1$ will be used as data generating system. The noise of the network can be modeled with diagonal form

$$v(t) = \begin{bmatrix} H_1(q) & 0 \\ 0 & H_2(q) \end{bmatrix} \begin{bmatrix} e_1 \\ e_1 \end{bmatrix}. \qquad (26)$$

A diagonal noise model is favorable over a non-diagonal noise model for optimization. For simulation the following systems will be used: $G_{12}^0(q) = 0.7q^{-1} + 0.3q^{-2}$, $G_{21}^0(q) = 0.9q^{-1} + -0.5q^{-2}$, $H_1^0(q) = \frac{1}{1+0.3q^{-1}}$, and $H_2^0(q) = \frac{1}{1-0.4q^{-1}}$. r_1,r_2 and e_1 are realizations of normally distributed white noise, with $\sigma_e^2 = 1$, $\sigma_{r_i}^2 = 0.01$.

A network model set will be used with G of FIR structure $G_{12}(q,\theta) = b_1^{12}q^{-1} + b_2^{12}q^{-1}$, $G_{21}(q,\theta) = b_1^{21}q^{-1} + b_2^{21}q^{-1}$, H of the form (26) where the inverse has FIR structure $H_1^{-1}(q,\theta) = 1 + d_1^1 q^{-1}$, $H_2^{-1}(q,\theta) = 1 + d_1^2 q^{-1}$ such that $\Gamma(\theta) = 1$, and $R(q,\theta) = I$. In total 6 parameters must be estimated. $\mathcal{M}$ satisfies the conditions of Proposition 4, so the constrained criterion can make a variance-free estimation of G. Identification criterion (12) will be used with weights $Q_I = I$, $Q_\lambda = \begin{bmatrix} 1+\lambda & -\lambda \\ -\lambda & \lambda \end{bmatrix}$ for $\lambda = 10, 10^3$, and $Q_\infty = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. A bilinear optimization problem results from the used model/criterion. The criterion is optimized using Matlab's function `fmincon()` and initialized with parameters equal to 0.

100 different realizations of 1000 datapoints are used for 100 identifications. The resulting parameter estimates are plotted in Figure 2 as boxplots. The criterion with $Q = I$

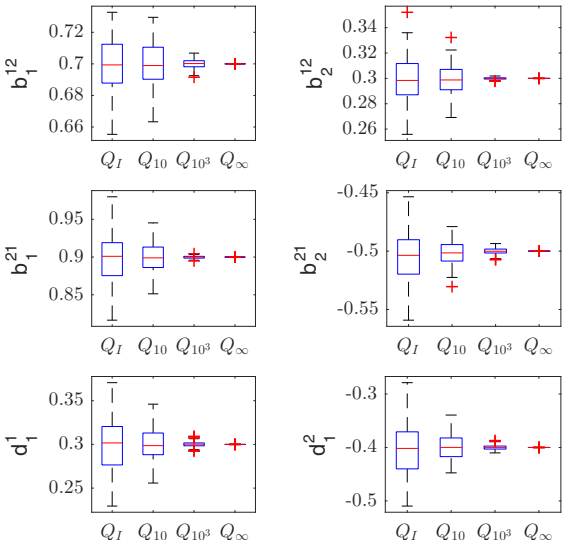

Fig. 2. Resulting parameter estimates over 100 realizations of data for different weights.

as weight shows that treating noises as being completely independent leads to parameter estimates close to the true value. Using Q_λ leads to smaller variance than Q_I for increasing λ. But most remarkable is the weight Q_∞ which leads to parameter estimates which deviate from the true value by an error with order of magnitude 10^{-5}, which is the error tolerance of the solver. So the weight Q_∞ leads to unbiased variance-free estimates for this example system.

7. CONCLUSIONS

In this paper we have extended prediction error identification in dynamic networks to handle rank-reduced noise. It has been shown that a typical weighted least squares criterion leads to consistent estimates when the noise model is chosen appropriately. A constrained criterion that explicitly accounts for the rank-reduced nature of the noise will also lead to consistent estimates. Additionally the constrained criterion reduces variance, and in particular in certain situations can result in a variance-free estimate.

Appendix A. PROOF OF PROPOSITION 2

The proof consists of 2 steps. First it will be shown that θ_0 is a minimum of the criterion, after which it will be shown that $M(q, \theta_0)$ is the only minimum.

The multivariable prediction error is

$$\varepsilon(\theta) = H_s^{-1}(\theta)\big(I - G(\theta)\big)w - H_s^{-1}(\theta)R(\theta)r, \tag{A.1}$$

which can be expanded as a function of just e and r

$$\begin{aligned}\varepsilon(\theta) =& H_s^{-1}(\theta)\big(I - G(\theta)\big)(I - G^0)^{-1}\big(H_s^0 e + R^0 r\big) + \\ &- H_s^{-1}(\theta)R(\theta)r.\end{aligned} \tag{A.2}$$

The innovation e can be separated to create 3 terms on the right hand side that are mutually uncorrelated

$$\begin{aligned}\varepsilon(\theta) =& \Big(H_s^{-1}(\theta)\big(I - G(\theta)\big)(I - G^0)^{-1}H_s^0 - I\Big)e + \\ &+ e + H_s^{-1}(\theta)\Big(\big(I - G(\theta)\big)(I - G^0)^{-1}R^0 - R(\theta)\Big)r.\end{aligned} \tag{A.3}$$

The first term has a strictly proper filter, hence it is uncorrelated to the second term and the 3 terms are uncorrelated. Since the 3 terms are uncorrelated the power of each term can be minimized individually. For the parameter θ_0 the first and last terms are 0, and the other term does not contain parameters, hence the minimum is reached.

In the second step it is shown that any θ_1 which reaches the minimum of the cost function must result in the same dynamic model as θ_0. The cost of the criterion for a particular parameter θ will be denoted with $V(\theta)$, so the minimum cost satisfies $V(\theta_0) = V(\theta_1)$. It can be shown that [Ljung, 1999]

$$V(\theta_0) - V(\theta_1) = \bar{\mathbb{E}}\Big(\varepsilon(\theta_0) - \varepsilon(\theta_1)\Big)^T Q\Big(\varepsilon(\theta_0) - \varepsilon(\theta_1)\Big) = 0 \tag{A.4}$$

where ε can be written as a function of e_a and r (A.2)

$$\varepsilon(\theta) = H_s^{-1}(\theta)\Big(\big(I - G(\theta)\big)w - R(\theta)r\Big) \tag{A.5}$$

with $w = (I - G^0)^{-1}\begin{bmatrix}H^0 & R^0\end{bmatrix}\begin{bmatrix}e_a \\ r\end{bmatrix}$. Define shorthand notation $\varepsilon(\theta) = \alpha(\theta)\begin{bmatrix}e_a \\ r\end{bmatrix}$ with

$$\begin{aligned}\alpha(\theta) =& H_s^{-1}(\theta)(I - G(\theta))(I - G^0)^{-1}\begin{bmatrix}H^0 & R^0\end{bmatrix} + \\ &- H_s^{-1}(\theta_1)\begin{bmatrix}0 & R(\theta)\end{bmatrix}.\end{aligned} \tag{A.6}$$

Then (A.4) can be written as

$$\bar{\mathbb{E}}\Big(\big(\alpha(\theta_0) - \alpha(\theta_1)\big)\begin{bmatrix}e_a \\ r\end{bmatrix}\Big)^T Q\Big(\big(\alpha(\theta_0) - \alpha(\theta_1)\big)\begin{bmatrix}e_a \\ r\end{bmatrix}\Big) = 0 \tag{A.7}$$

which implies that $\alpha(\theta_0) = \alpha(\theta_1)$ since $\begin{bmatrix}e_a \\ r\end{bmatrix}$ is a full rank process (provided that r is persistently exciting of sufficiently high order) and Q is strictly positive definite. The expression for $\alpha(\theta_0)$ can be simplified since θ_0 is equal to the real network

$$\alpha(\theta_0) = H_s^{-1}(\theta_0)\begin{bmatrix}H^0 & 0\end{bmatrix} = \begin{bmatrix}I & 0 \\ \Gamma^0 & 0\end{bmatrix} \tag{A.8}$$

such that the equation for α reduces to

$$\begin{bmatrix}I & 0 \\ \Gamma^0 & 0\end{bmatrix} = \alpha(\theta_1). \tag{A.9}$$

When the above equation is pre-multiplied with $(I - G(\theta_1))^{-1}H_s(\theta_1)$ then we obtain, when using (A.6):

$$\begin{aligned}&(I - G(\theta_1))^{-1}\begin{bmatrix}H_a(\theta_1) & 0 \\ \tilde{H}_b(\theta) + \Gamma^0 & 0\end{bmatrix} = \\ &(I - G^0)^{-1}\begin{bmatrix}H^0 & R^0\end{bmatrix} - (I - G(\theta_1))^{-1}\begin{bmatrix}0 & R(\theta_1)\end{bmatrix}.\end{aligned} \tag{A.10}$$

Note that the above equation contains Γ^0 at the left hand side. Then we can write $T(\theta_1) = T(\theta_0)$ with

$$T(\theta) = (I - G(\theta))^{-1}\begin{bmatrix}H_a(\theta) & R_a(\theta) \\ \tilde{H}_b(\theta) + \Gamma^0 & R_b(\theta)\end{bmatrix}. \tag{A.11}$$

The model set which corresponds to the above situation with fixed Γ^0 is defined as

$$\tilde{\mathcal{M}} := \big\{M \in \mathcal{M} \;\big|\; \Gamma(\theta) = \Gamma^0\big\} \subseteq \mathcal{M}.$$

$\tilde{\mathcal{M}}$ is globally network identifiable at θ_0 when $\mathcal{M}$ is globally network identifiable at θ_0 since it is a subset of $\mathcal{M}$. Since $\tilde{\mathcal{M}}$ is globally network identifiable at θ_0 we then have

$$T(q, \theta_0) = T(q, \theta_1) \Rightarrow \begin{cases} G(q, \theta^\star) &= G^0(q) \\ H_a(q, \theta^\star) &= H_a^0(q) \\ \tilde{H}_b(q, \theta^\star) &= \tilde{H}_b^0(q) \\ R(q, \theta^\star) &= R^0(q). \end{cases} \tag{A.12}$$

Appendix B. PROOF OF PROPOSITION 3

Using the proof of Proposition 2 it is easy to show that θ_0 is a minimum of the cost function. Moreover θ_0 satisfies the constraint, hence it is a solution of the criterion. Next we show that it is a unique solution. Any model θ_1 in the minimum of the cost function has $\varepsilon_a(t,\theta_1) = \varepsilon_a(t,\theta_0) = e_a(t)$, or expanded

$$e_a = H_a^{-1}(\theta_1)\left\{[I - G_{aa}(\theta_1)\ -G_{ab}(\theta_1)]\begin{bmatrix} w_a \\ w_b\end{bmatrix} - R_a(\theta_1)r\right\}. \tag{B.1}$$

Then the constraint can be written as $\Gamma(\theta_1)e_a(t) = \varepsilon_b(t,\theta_1)$, or expanded

$$\Gamma(\theta_1)e_a = -\tilde{H}_b(\theta_1)e_a + [-G_{ba}(\theta_1)\ I - G_{bb}(\theta_1)]\begin{bmatrix} w_a \\ w_b\end{bmatrix} - R_b(\theta_1)r. \tag{B.2}$$

By bringing $\tilde{H}_b e_a$ to the left-hand side, (B.1) and (B.2) can be stacked into one matrix

$$\begin{bmatrix} H_a(\theta) \\ H_b(\theta)\end{bmatrix} e_a = (I - G(\theta))w - R(\theta)r. \tag{B.3}$$

w can be replaced by $w = T(\theta_0)\left[\begin{smallmatrix} e_a \\ r\end{smallmatrix}\right]$. Then with some algebra we obtain

$$T(\theta_1)\left[\begin{smallmatrix} e_a \\ r\end{smallmatrix}\right] = T(\theta_0)\left[\begin{smallmatrix} e_a \\ r\end{smallmatrix}\right]. \tag{B.4}$$

Since the equation is driven by a full rank process provided that r is persistently exciting of sufficiently high order we obtain $T(\theta_1) = T(\theta_0)$. and a consistent estimate under the condition that $\mathcal{M}$ is globally network identifiable at θ_0.

Appendix C. PROOF OF PROPOSITION 4

For the considered model the constraint can be written as

$$0 = [\Gamma(\theta)\ -I]\, H_s^{-1}(\theta)\big((I - G(\theta))w - Ir\big). \tag{C.1}$$

In this equation w can be replaced by $w = (I - G^0)^{-1}(H^0 e + Ir)$, such that we have two equations (provided r is persistently exciting of sufficiently high order)

$$0 = [\Gamma(\theta)\ -I]\, H_s^{-1}(\theta)(I - G(\theta))(I - G^0)^{-1}H^0, \tag{C.2}$$

$$0 = [\Gamma(\theta)\ -I]\, H_s^{-1}(\theta)\Big((I - G(\theta))(I - G^0)^{-1} - I\Big). \tag{C.3}$$

When (C.3) is post-multiplied with H^0 and (C.2) is subtracted then $0 = [\Gamma(\theta)\ -I]\, H_s^{-1}(\theta)H^0$ is obtained, which leads to

$$H_b(\theta)H_a^{-1}(\theta) = H_b^0(H_a^0)^{-1}. \tag{C.4}$$

Taking only the feedthrough terms

$$\lim_{z\to\infty} H_b(z,\theta)H_a^{-1}(z,\theta) = \lim_{z\to\infty} H_b^0(z)(H_a^0)^{-1}(z) \tag{C.5}$$

shows due to H_a being monic that $\Gamma(\theta) = \Gamma^0$. Then when (C.3) is postmultiplied with $(I - G^0)$ we obtain

$$0 = \big[H_b(\theta)H_a^{-1}(\theta)\ -I\big]\,(G^0 - G(\theta)). \tag{C.6}$$

Since we have a 2 node network this implies

$$0 = G_{21}^0 - G_{21}(\theta),$$
$$0 = H_b(\theta)H_a^{-1}(\theta)(G_{12}^0 - G_{12}(\theta)).$$

For this network $\Gamma(\theta) = \Gamma^0 = 1$, so $H_b(\theta)H_a^{-1}(\theta)$ is some nonzero transfer function, hence we have $G^0 = G(\theta)$.

REFERENCES

A. G. Dankers, P. M. J. Van den Hof, X. Bombois, and P. S. C. Heuberger. Errors-in-variables identification in dynamic networksconsistency results for an instrumental variable approach. *Automatica*, 62:39–50, 2015.

A. G. 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 Trans. Automatic Control*, 61(4):937–952, 2016.

M. Deistler, W. Scherrer, and B. D. O. Anderson. The structure of generalized linear dynamic factor models. In J. Beran, Y. Feng, and H. Hebbel, editors, *Empirical Economic and Financial Research*. Springer, 2015.

N. Everitt, G. Bottegal, C. R. Rojas, and H. Hjalmarsson. On the effect of noise correlation in parameter identification of SIMO systems. In *Proc. 17th IFAC Symposium on System Identification*. IFAC, 2015. (IFAC-PapersOnLine, 48(28):326-331).

M. Gevers and A. S. Bazanella. Identification in dynamic networks: identifiability and experiment design issues. In *Proc. 54th IEEE Conf. Decision and Control (CDC)*, pages 4005–4010. IEEE, 2015.

A. Haber and M. Verhaegen. Subspace identification of large-scale interconnected systems. *IEEE Trans. Automatic Control*, 59(10):2754–2759, 2014.

D. Hayden, Y. H. Chang, J. Gonçalves, and C. J. Tomlin. Sparse network identifiability via compressed sensing. *Automatica*, 68:9–17, 2016.

J. Linder and M. Enqvist. Identification of systems with unknown inputs using indirect input measurements. *Intern. J. Control*, 90(4):729–745, 2017.

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

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

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*, 49(10):2994–3006, 2013.

P. M. J. Van den Hof, H. H. M. Weerts, and A.G. Dankers. Prediction error identification with rank-reduced output noise. In *Proc. 2017 American Control Conference (ACC)*, 2017. (to appear).

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, Beijing, China*, pages 1409–1414, 2015. (IFAC-PapersOnLine, 48(28):1409-1414).

H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Identification of dynamic networks operating in the presence of algebraic loops. In *Proc. 55nd IEEE Conf. Decision and Control (CDC)*, pages 4606–4611, 2016a.

H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers. Identifiability of dynamic networks with part of the nodes noise-free. In *Proc. 12th IFAC Workshop ALCOSP*, 2016b. (IFAC-PapersOnLine, 49(13):19-24).

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 preprint arXiv:1609.00864*, 2016c.

M. Zorzi and A. Chiuso. A bayesian approach to sparse plus low rank network identification. In *Proc. 54nd IEEE Conf. Decision and Control (CDC)*, pages 7386–7391, 2015.