2017 American Control Conference
Sheraton Seattle Hotel
May 24–26, 2017, Seattle, USA

Prediction error identification with rank-reduced output noise

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

Abstract— **In data-driven modelling in dynamic networks, it is commonly assumed that all measured node variables in the network are noise-disturbed and that the network (vector) noise process is full rank. However when the scale of the network increases, this full rank assumption may not be considered as realistic, as noises on different node signals can be strongly correlated. In this paper it is analyzed how a prediction error method can deal with a noise disturbance whose dimension is strictly larger than the number of white noise signals than is required to generate it (rank-reduced noise). Based on maximum likelihood considerations, an appropriate prediction error identification criterion will be derived and consistency will be shown, while variance results will be demonstrated in a simulation example.**

I. INTRODUCTION

It is becoming more common to use dynamic networks as a modelling tool to represent interconnected systems of increasing complexity. Some examples of these interconnected systems can be found in smart grids, social networks and systems biology. Classical system identification literature has focused mostly on open-loop or closed-loop controlled systems, but over the last couple of years attention for data-driven modelling in a more structured setting, as e.g. in dynamic networks, has received increasing attention. Among different non-parametric and parametric approaches [1], [2], [3], [4], [5], a framework for the extension of prediction error approaches to the case of dynamic networks has been presented in [6]. In most network identification papers the assumption is made that all measured nodes in the network are disturbed by a stationary stochastic noise process, being of full rank, and being uncorrelated over the different nodes. The full rank property means that the spectral density of the vector process has full rank, implying that each disturbance signal is originating from an independent white noise source. If the size of a dynamic network increases, the assumption of having a full rank noise process becomes more and more unrealistic. Different node signals in the network are likely to experience noise disturbances that are highly correlated with and possibly dependent on other node signals in its direct neighbourhood. One could think e.g. of a network of temperature measurements in a spatial area, where unmeasured external effects (e.g. wind) affect all measured nodes in a strongly related way, or waves that affect different locations of a ship. In a different modeling setting, dynamic factor models (see e.g. [7]) have been used to deal with rank-reduced noises, most dominantly applied in economic data analysis. In a prediction error modelling setting the situation of rank reduced noises has -to the authors' best knowledge- not yet been considered. A first step in addressing this situation has recently been made in [8], [9], where network identifiability has been studied for the situation that some of the network nodes are noise-free, leading to a rank-reduced noise process.

In the current paper we develop the theory for the handling of a general rank-reduced (singular) noise process in a standard open-loop identification problem. For simplicity we will treat the situation of a one-input two-output system with an output noise process that is driven by a single scalar white noise.

First the system set-up will be described in Section II, after which the appropriate predictor filters will be analyzed in Section III. In section IV a new identification criterion will be derived and its consistency properties will be shown. Through a simulation example in Section V experimental results of the identification method will be illustrated, including an illustration of the improved variance results of the method, compared to an alternative multiple SISO approach to the problem.

II. SYSTEM SETUP

We consider a one-input two-output system, as sketched in Figure 1, described by

$$y(t) = G(q)u(t) + H(q)e(t) \tag{1}$$

with $G \in \mathbb{R}^{2\times 1}(q)$ a linear, time-invariant system, $H \in \mathbb{R}^{2\times 1}(q)$, a rational, stable and minimum-phase transfer function with full column rank [7], e a scalar white noise process with variance σ_e^2, and q the forward shift operator. The system is operating in open-loop, implying that u and

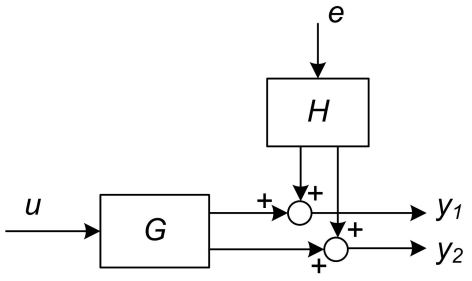

Fig. 1. One input - two output system with singular output noise

Paul Van den Hof and Harm Weerts are with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands {p.m.j.vandenhof, h.h.m.weerts}@tue.nl
Arne Dankers is with the University of Calgary a.g.dankers@hifieng.com

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).

e are uncorrelated. The output noise $v(t) = H(q)e(t)$ has spectral density $\Phi_v(z) = H(z)\sigma_e^2 H^T(z^{-1})$ which has rank 1 over the field of rational functions. We assume that we have ordered the output signals y_1, y_2 in such a way that -by appropriately choosing σ_e^2- H is normalized according to

$$\lim_{z\to\infty} H(z) := H^\infty = \begin{bmatrix} 1 \\ \eta \end{bmatrix}, \quad \eta \in \mathbb{R}. \tag{2}$$

Note that e.g. in the situation when one of the outputs is noise-free, such a reordering is necessary to make y_1 the noisy variable, which is required for (2). When defining H^+ as a proper and stable left inverse of H, satisfying

$$H(z)^+ H(z) = I, \tag{3}$$

then the white noise process e can be explicitly written as

$$e(t) = H^+(q)[y(t) - G(q)u(t)]. \tag{4}$$

By defining the one-step-ahead predictor ([10]) as $\hat{y}(t|t-1) := \bar{\mathbb{E}}\{y(t) \mid y^{t-1}, u^t\}$ with $y^{t-1} := \{y(0), y(1), \cdots y(t-1)\}$ and u^t defined accordingly, we can then construct the expression for the innovation process.

Proposition 1: For the considered system (1) the innovation process is determined by

$$y(t) - \hat{y}(t|t-1) = H^\infty e(t) = \tilde{H}(q)[y(t) - G(q)u(t)] \tag{5}$$

with

$$\tilde{H}(q) := \begin{bmatrix} I - [H(q) - H^\infty]H^+(q) \end{bmatrix} \quad \in \mathbb{R}^{2\times 2}(q). \tag{6}$$

Proof: Starting from (1) it follows that

$$y(t) = G(q)u(t) + [H(q) - H^\infty]e(t) + H^\infty e(t). \tag{7}$$

When subsituting the expression (4) in the second term on the righthand side, it follows that

$$\begin{aligned} y(t) &= G(q)u(t) + [H(q) - H^\infty]H^+(q)[y(t) - G(q)u(t)] \\ &\quad + H^\infty e(t). \end{aligned} \tag{8}$$

Taking conditional expectation then shows that the first two terms on the righthand side form the predictor:

$$\hat{y}(t|t-1) = G(q)u(t) + [H(q) - H^\infty]H^+(q)[y(t) - G(q)u(t)] \tag{9}$$

while the innovation process is given by

$$H^\infty e(t) = \begin{bmatrix} I - [H(q) - H^\infty]H^+(q) \end{bmatrix} [y(t) - G(q)u(t)].$$

■

Note that the left inverse $H^+(z)$ is not uniquely determined. There are multiple filters $H^+(z)$ that satisfy (3). As a result, the filter expressions for the predictor (9) are not unique, while the predicted outputs are unique by definition. Therefore the predictor filter expressions are not suitable for parametrizing models in an identification setting. For this purpose we first have to rework the predictor filters into a unique form.

III. TOWARDS UNIQUE PREDICTOR FILTERS

For purpose of identification we would like to represent the system through (unique) prediction filters that can be parametrized with unknown parameters. The prediction properties of the system can be represented by (5), which actually constitute two equations; one equation that generates the predictor on the basis of y and u, and one that represents a constraint on the innovation process, since according to (2), $\begin{bmatrix} \eta & -1 \end{bmatrix} H^\infty e(t) = 0$. As a result the prediction properties of the system are characterized by the following set of equations that result from combining (9), (6) with (5), (2):

$$\begin{bmatrix} \hat{y}(t|t-1) \\ 0 \end{bmatrix} = \begin{bmatrix} I - \tilde{H} & \tilde{H}G \\ [\eta \quad -1]\tilde{H} & -[\eta \quad -1]\tilde{H}G \end{bmatrix} \begin{bmatrix} y(t) \\ u(t) \end{bmatrix}. \tag{10}$$

In order to further analyze the role of $\tilde{H}$ in this equation, the freedom in choosing the left inverse $H^+(z)$ needs to be specified.

Lemma 1: Let H be denoted as $H = \begin{bmatrix} H_a \\ H_b \end{bmatrix}$. Then

$$H^+ = \begin{bmatrix} (1 - BH_b)H_a^{-1} & B \end{bmatrix} \tag{11}$$

with B any scalar rational transfer function. For a proper and stable H^+, B will need to be proper and stable too.

Proof: Denote $H^+ = \begin{bmatrix} A & B \end{bmatrix}$. Then (3) shows that $AH_a + BH_b = I$, and with H_a being invertible it follows that $A = (1 - BH_b)H_a^{-1}$. which directly leads to the result. ■

Substituting (11) into the expression for $\tilde{H}$ (6) shows that

$$\begin{aligned} \tilde{H} &= I - \begin{bmatrix} H_a - 1 \\ H_b - \eta \end{bmatrix} \begin{bmatrix} (1 - BH_b)H_a^{-1} & B \end{bmatrix} \\ &= \begin{bmatrix} 1 - (H_a - 1)(1 - BH_b)H_a^{-1} & 0 - (H_a - 1)B \\ 0 - (H_b - \eta)(1 - BH_b)H_a^{-1} & 1 - (H_b - \eta)B \end{bmatrix} \\ &= \begin{bmatrix} BH_b - (BH_b - 1)H_a^{-1} & (1 - H_a)B \\ (H_b - \eta)(BH_b - 1)H_a^{-1} & 1 - (H_b - \eta)B \end{bmatrix}. \end{aligned} \tag{12}$$

Expressions (10) and (12) reveal that the predictor filters that generate $\hat{y}(t|t-1)$ are non-unique. On the one hand this is induced by the 0-equality in (10), which shows that the two equations in (10) e.g. can be added together without changing the left hand side. On the other hand, H^+ and therefore also $\tilde{H}$ are dependent on a free rational function B. The underlying reason for this non-uniqueness of the predictor filters is the fact that the vector signal $[y(t)\ u(t)]^T$ that serves as input to the filters, is not a full rank process. By manipulating the expression (10) we can construct an equivalent form that only uses y_1 and u as input, as a result of which the predictor filters become unique. This is reflected in the following Theorem.

Theorem 1: A unique expression for the predictor filters is given by

$$\begin{bmatrix} \hat{y}_1(t|t-1) \\ \hat{y}_2(t|t-1) \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} y_1(t) \\ u(t) \end{bmatrix} \tag{13}$$

with

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} 1 - H_a^{-1} & [H_a^{-1} \;\; 0]G \\ H_a^{-1}(H_b - \eta) & [-H_a^{-1}(H_b - \eta) \;\; 1]G \end{bmatrix} \tag{14}$$

Proof: Enclosed in appendix. ∎

By removing output signal y_2, with a noise contribution that is fully dependent on the noise on y_1, the two one-step-ahead output predictions can be constructed with predictor filters that are uniquely determined by the system properties H_a, H_b, G. As a result, the expression (14) for P now lends itself well to construct a parametrized predictor that can be used in identification.

IV. An identification criterion based on maximum likelihood considerations

For identification purposes we need to formulate an identification criterion through which we can (consistently) identify a parametric model from measurement data. In the prediction error identification framework the identification criterion is developed on the basis of the innovation process of the considered system, which in the situation of the considered process is given by (5):

$$H^{\infty}e(t) = y(t) - \hat{y}(t|t-1). \tag{15}$$

It is clear that $H^{\infty}e(t)$ is a vector of two random variables that are actually the same (except for a scaling). The vector process $H^{\infty}e(t)$ has a singular covariance matrix, and -when e.g. assuming that $e(t)$ has a Gaussian distribution- the vector process does not have a joint probability density function (pdf). However $e(t)$ does have a pdf and this term can be isolated from the vector equation (15) as follows:

$$e(t) = y_1(t) - \hat{y}_1(t|t-1) \tag{16}$$

$$\eta e(t) = y_2(t) - \hat{y}_2(t|t-1). \tag{17}$$

In terms of identification, we are parametrizing the two predictors $\hat{y}_1(t|t-1;\theta)$ and $\hat{y}_2(t|t-1;\theta)$ with parameter vector θ on the basis of parametrized versions of the predictor filter $P(q,\theta)$ (14), where θ is composed of the parametrized elements of the transfer functions $G(q,\theta)$ and $H(q,\theta)$ with $H_a(q,\theta)$ monic. A parametrized pair $(G(q,\theta), H(q,\theta))$ will be denoted by $M(q,\theta)$.

Following maximum likelihood considerations, it is obvious that the identification criterion should not be to minimize the power of the innovation process on each of the two components (16)-(17) independently. Since there is only one degree of freedom in the vector innovation process, the natural choice of identification criterion is to minimize the power of one component of the innovation, under a constraint on the second component:

Constrained identification criterion:

$$\begin{cases} \min\limits_{\theta\in\Theta} \bar{\mathbb{E}}\varepsilon_1(t,\theta)^2 \\ \text{under the constraint } \varepsilon_2(t,\theta) = \eta\varepsilon_1(t,\theta) \;\forall t \end{cases} \tag{18}$$

with $\bar{\mathbb{E}}x^2(t) := \lim_{N\to\infty} \frac{1}{N}\sum_{t=0}^{N-1} \mathbb{E}x^2(t)$, and

$$\begin{bmatrix} \varepsilon_1(t,\theta) \\ \varepsilon_2(t,\theta) \end{bmatrix} := y(t) - \hat{y}(t|t-1;\theta) = y(t) - P(q,\theta)\begin{bmatrix} y_1(t) \\ u(t) \end{bmatrix}.$$

The reduced-rank property of the noise process induces a constraint in the identification criterion. This is a direct generalization of the zero-noise situation that has been addressed in [8], [9]. Note that the parametrized predictor filters do not incorporate parameter η, since η only appears in $P(q,\theta)$ through the term $H_b - \eta$ which is actually independent of η[1]. However occurence of η in the constraint equation allows the consistent identification of η too, as is shown in the following result.

Proposition 2: Consider input-output data that has been generated by system (1) with $(G,H) = M_0(q)$, and let θ^* be in the solution set of the identification criterion (18). Then $M(q,\theta^*) = M_0(q)$ 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) input u is persistently exciting of a sufficiently high order.

Proof: Enclosed in appendix. ∎

With slight abuse of the notion, we refer to the above result as a consistency property[2]. Consistency of the estimator is actually not very much surprising since it is also possible to obtain consistent estimates by separately estimating two independent SISO models on the basis of u, y_1 and u, y_2 respectively. Nevertheless it is attractive that consistency can be formally proved for the multi-output model and the constrained identification criterion.

While the constrained identification criterion can be applied to situations where the data generating system is in the model set, application beyond that situation can easily lead to lack of feasibility of the constrained optimization problem. Therefore we apply a relaxation to the constrained criterion, by adding the constraint as a quadratic term to the original cost function

Identification criterion after constraint relaxation:

$$\min_{\theta\in\Theta} \bar{\mathbb{E}}\left[\varepsilon_1(t,\theta)^2 + \lambda(\varepsilon_2(t,\theta) - \eta\varepsilon_1(t,\theta))^2\right] \tag{19}$$

with $\lambda \in \mathbb{R}$ a tuning parameter.

The attractive feature of this relaxed criterion is that it has the same consistency properties as the original criterion.

Proposition 3: The consistency result of Proposition 2 holds true if we replace the constrained identification criterion (18) by the unconstrained criterion (19).

Proof: The result follows directly from the fact that the minimum of $\bar{\mathbb{E}}(\varepsilon_2(t,\theta) - \eta\varepsilon_1(t,\theta))^2$ is equal to 0, which is indeed achieved when the system is in the model set. This does not constrain the minimum of $\bar{\mathbb{E}}\varepsilon_1(t,\theta)^2$ which is equal to σ_e^2. ∎

One of the nice properties of the criterion with constraint relaxation, is that it allows an asymptotic analysis of the parameter variance, according to the multivariate theory as presented in [11]. This will be presented elsewhere.

[1] Note that the constant (algebraic) term of H_b is equal to η.

[2] Since we have formulated the identification criterion as an infinite-data expected value criterion, we do not consider the convergence property of the related finite-time criterion, see [10].

V. Simulations

This section is dedicated to simulating the proposed method and comparing it to performing two separate SISO estimations. For simulations it turns out to be easier to first rewrite the constraint into a form that is simpler to implement.

Proposition 4: The constraint $\eta\varepsilon_1(t,\theta) = \varepsilon_2(t,\theta)$ can equivalently be written as

$$H_b(q,\theta)(y_1 - G_1(q,\theta)u) = H_a(q,\theta)(y_2 - G_2(q,\theta)u), \quad (20)$$

or, provided that $\eta \neq 0$:

$$H_a^{-1}(q,\theta)(y_1 - G_1(q,\theta)u) = H_b^{-1}(q,\theta)(y_2 - G_2(q,\theta)u). \quad (21)$$

Proof: Included in the appendix. ∎

This form of the constraint is useful for the current system setup, but will generally not be possible for systems of larger dimensions because for example H_b will not be invertible.

In the optimization ε_1 will be optimized subject to the constraint defined in (21). To this end we use the system in Figure 1 with

$$G_1(q) = 0.3 + 0.7q^{-1} + 0.3q^{-2},$$
$$G_2(q) = 0.15 + 0.9q^{-1} - 0.5q^{-2},$$
$$H_a(q) = \frac{1}{1+0.3q^{-1}+0.4q^{-2}}, \quad H_b(q) = \frac{1}{2-0.4q^{-1}+0.2q^{-2}},$$

such that $\eta = 0.5$. The plant has a FIR structure while the inverse of the noise filters are also of the FIR type. In this way a parametrization can be used where both the plant and noise inverse are parameterized by FIR models of order 3, i.e.

$$\begin{aligned} G_i(q,\theta) &= b_{0,i} + b_{1,i}q^{-1} + b_{2,i}q^{-2}, \qquad i = \{1,2\} \\ H_j(q,\theta)^{-1} &= d_{0,j} + d_{1,j}q^{-1} + d_{2,j}q^{-2}, \qquad j = \{a,b\}. \end{aligned}$$

Prediction error $\varepsilon_1(t,\theta)$ and the constraint (21) are then bilinear in the parameters. The input and process noise are both generated as a realization of two independent white noise processes.

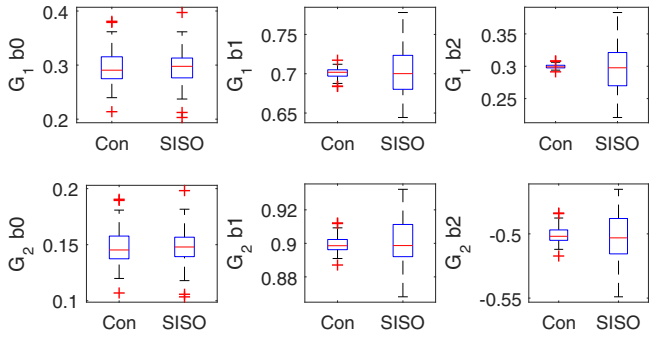

Fig. 2. Results of parameter estimates of G over 100 experiments, for parameters b_0, b_1, b_2 (left, middle, right), and models G_1, G_2 (top, bottom). In all 6 plots the left box is the result of the constrained criterion method (Con), and the right box is the result of the SISO identifications (SISO).

Results of applying the constrained criterion are compared with an identification where the SIMO problem is decomposed into two unconstrained SISO problems. In the SISO problems similar model parametrizations are used, i.e. an FIR models of order 3 for the plant models, and an FIR of order 3 for the inverse of the noise models. This again results in parametrizations that are bilinear in the parameters.

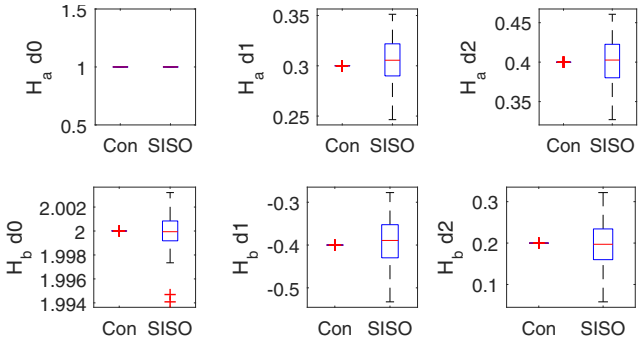

Fig. 3. Results of parameter estimates of H over 100 experiments, for parameters d_0, d_1, d_2 (left, middle, right), and models H_a, H_b (top, bottom). In all 6 plots the left box is the result of the constrained method (Con), and the right box is the result of the SISO identifications (SISO).

In total 100 Monte-Carlo experiments have been performed where each time a new realization of the input and noise is generated. The results of estimation using the constrained SIMO criterion are compared to the separate SISO identifications in Figures 2 and 3. It can be observed that the estimates of G_1 and G_2 are consistent for both methods, and that the variance of the parameters is smaller in the case of the constrained criterion. For the estimates of H_a both methods are consistent, and the constrained method has very small variance on the parameter estimates.

Since in the SISO estimation, the noise models are monic by definition, the estimation results of H_b for the SISO estimation, are scaled by $\sigma_{\varepsilon_1}/\sigma_{\varepsilon_2}$ to make the results comparable to the H_b-estimate from the constrained method. Both methods deliver a consistent H_b and again the constrained method has smaller variance when compared to separating the identification problem in two SISO problems. This is due to the fact that in the constrained case, we appropriately take account of the relation that exists between the noise processes on the two different outputs. These result are also in line with the variance results of [12].

Next we have implemented the criterion (19) with the constraint relaxation, and we have applied it to the same example, with different values of the tuning parameter λ: λ $1, 10, 100$. For $\lambda \to \infty$ the criterion converges to the result of the constrained criterion. Simulation results for 100 Monte-Carlo experiments are shown in Figure 4. It can be observed that the variance of the relaxation criterion is somehow in between the case of the constrained criterion and of the SISO identifications. For increasing values of the tuning parameter λ, the variance decreases.

VI. Conclusions

Motivated by applications in large scale dynamic networks, the prediction error identification method has been extended to be able to deal with output disturbances that are generated

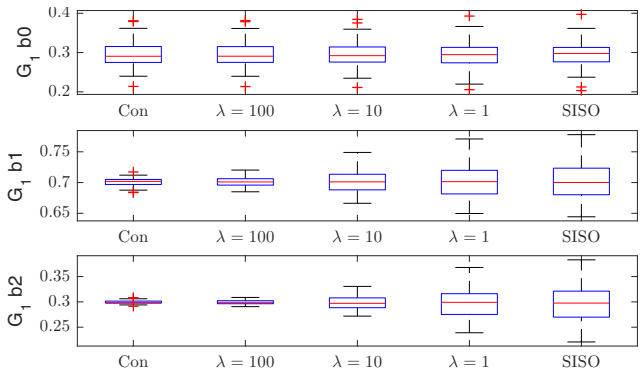

Fig. 4. Results of parameter estimates of G_1 over 100 experiments for the relaxed criterion for $\lambda = 100, 10, 1$ (3 middle columns), compared to the constrained criterion (left column) and the separate SISO identifications (right column) for parameters b_0, b_1, b_2 (top, middle, bottom) of G_1.

by a reduced-rank noise process. To this end the one-step-ahead predictor is written in a particular form in order to arrive at unique predictor filters, and a constrained identification criterion based on maximum likelihood considerations has been analyzed. Through constraint relaxation an alternative criterion can also be formulated that shares consistency properties with the constrained criterion. Consistency and variance properties of the two methods have been illustrated and compared with the separate identification of SISO models, showing improved variance properties. Extensions of this work to a more complex dynamic network situation are addressed in [13].

APPENDIX

Corollary 1: On the basis of the expression (12) for $\tilde{H}$, and denoting $\tilde{H} = \begin{bmatrix} \tilde{H}_{11} & \tilde{H}_{12} \\ \tilde{H}_{21} & \tilde{H}_{22} \end{bmatrix}$, the following expressions hold true:

a. $X := [\eta\tilde{H}_{12} - \tilde{H}_{22}]^{-1}[\eta\tilde{H}_{11} - \tilde{H}_{21}] = -H_b H_a^{-1}$
b. $-\tilde{H}_{11} + 1 + \tilde{H}_{12}X = 1 - H_a^{-1}$
c. $-\tilde{H}_{21} + [\tilde{H}_{22} - 1]X = H_a^{-1}[H_b - \eta]$

Proof:

Part a. When using the expressions for $\tilde{H}$ it follows that

$$\begin{aligned}
&[\eta\tilde{H}_{12} - \tilde{H}_{22}]^{-1}[\eta\tilde{H}_{11} - \tilde{H}_{21}] = \\
&\quad [\eta(1-H_a)B - [1-(H_b-\eta)B]]^{-1} \cdot \\
&\quad \left[\eta H_a^{-1}(1-BH_b) + \eta BH_b - [H_b-\eta][BH_b-1]H_a^{-1}\right] \\
&= [B[H_b - \eta H_a] - 1]^{-1}\left[H_b(\eta B + H_a^{-1} - H_a^{-1}H_b B\right] \\
&= [B[H_b - \eta H_a] - 1]^{-1} H_b H_a^{-1}[1 - B[H_b - \eta H_a]] = \\
&= -H_b H_a^{-1}.
\end{aligned}$$

Part b.

$$\begin{aligned}
&-\tilde{H}_{11} + 1 + \tilde{H}_{12}X = \\
&= -BH_b + (BH_b - 1)H_a^{-1} + 1 - (1-H_a)BH_b H_a^{-1} = \\
&= 1 - H_a^{-1}
\end{aligned}$$

Part c.

$$\begin{aligned}
&-\tilde{H}_{21} + [\tilde{H}_{22} - 1]X = -(H_b - \eta)(BH_b - 1)H_a^{-1} + \\
&\quad + [1 - [1 - (H_b - \eta)B]]\, H_b H_a^{-1} = \\
&= H_a^{-1}\left[-BH_b^2 - H_b + \eta BH_b - \eta + H_b + (H_b - \eta)BH_b\right] \\
&= H_a^{-1}[H_b - \eta].
\end{aligned}$$

■

Proof of Theorem 1

On the basis of (10), we can write

$$\begin{bmatrix} \hat{y}_1(t|t-1) \\ \hat{y}_2(t|t-1) \\ 0 \end{bmatrix} = A \cdot \begin{bmatrix} y_1(t) \\ y_2(t) \\ u(t) \end{bmatrix}$$

where A can be written in the form of its scalar entries, as $A =$

$$\begin{bmatrix} -\tilde{H}_{11}+1 & -\tilde{H}_{12} & [\tilde{H}_{11} \;\; \tilde{H}_{12}]G \\ -\tilde{H}_{21} & -\tilde{H}_{22}+1 & [\tilde{H}_{21} \;\; \tilde{H}_{22}]G \\ \eta\tilde{H}_{11}-\tilde{H}_{21} & \eta\tilde{H}_{12}-\tilde{H}_{22} & [\tilde{H}_{21}-\eta\tilde{H}_{11} \;\; \tilde{H}_{22}-\eta\tilde{H}_{12}]G \end{bmatrix}.$$

We can now manipulate with the equations to remove the dependence on $y_2(t)$ as "input" to the above equations. Premultiplying the set of equations with

$$\begin{bmatrix} 1 & 0 & \tilde{H}_{12}\left[\eta\tilde{H}_{12} - \tilde{H}_{22}\right]^{-1} \end{bmatrix}$$

delivers:

$$\begin{aligned}
\hat{y}_1 = 1 - \tilde{H}_{11} + \tilde{H}_{12}&\left[\eta\tilde{H}_{12} - \tilde{H}_{22}\right]^{-1}[\eta\tilde{H}_{11} - \tilde{H}_{21}]y_1(t) \\
&+ \Big\{[\tilde{H}_{11} \;\; \tilde{H}_{12}]G + \tilde{H}_{12}\left[\eta\tilde{H}_{12} - \tilde{H}_{22}\right]^{-1} \cdot \\
&[\tilde{H}_{21} - \eta\tilde{H}_{11} \;\; \tilde{H}_{22} - \eta\tilde{H}_{12}]G\Big\}\, u(t). \qquad (22)
\end{aligned}$$

Employing part (b.) of Corollary 1 then shows the correct expression for P_{11}, while P_{12} can simply be verified to have the structure $P_{12} = [1 - P_{11} \;\; 0]G$.

For obtaining $\hat{y}_2(t|t-1)$ we premultiply the original set of equations with

$$\begin{bmatrix} 0 & 1 & [\tilde{H}_{22} - 1]\left[\eta\tilde{H}_{12} - \tilde{H}_{22}\right]^{-1} \end{bmatrix}$$

and obtain:

$$\begin{aligned}
\hat{y}_2 = -\tilde{H}_{21} + [\tilde{H}_{22} - 1]&\left[\eta\tilde{H}_{12} - \tilde{H}_{22}\right]^{-1}[\eta\tilde{H}_{11} - \tilde{H}_{21}]y_1(t) \\
&+ \Big\{[\tilde{H}_{21} \;\; \tilde{H}_{22}]G + [\tilde{H}_{22} - 1]\left[\eta\tilde{H}_{12} - \tilde{H}_{22}\right]^{-1} \cdot \\
&[\tilde{H}_{21} - \eta\tilde{H}_{11} \;\; \tilde{H}_{22} - \eta\tilde{H}_{12}]G\Big\}\, u(t). \qquad (23)
\end{aligned}$$

Employing part (c.) of Corollary 1 then shows the correct expression for P_{21}, while P_{22} can simply be verified to have the structure $P_{22} = [-P_{21} \;\; 1]G$. □

Proof of Proposition 2:

The first part of the proof is to show that the true system minimizes the quadratic function $\bar{V}(\theta) = \bar{\mathbb{E}}\varepsilon_1^2(t,\theta)$ and satisfies the constraint. For this we rewrite the prediction

error in terms of its driving variables u and e by filling (1) in into the prediction error[3]

$$\varepsilon_1(t,\theta) = H_a^{-1}(\theta)H_a e + H_a^{-1}(\theta)(G_1 - G_1(\theta))u.$$

In ε_1 we can split the e term in a delayed and a non-delayed part

$$\varepsilon_1(t,\theta) = (H_a^{-1}(\theta)H_a - 1)e + H_a^{-1}(\theta)(G_1 - G_1(\theta))u + e$$

such that the white noise term e is uncorrelated to the other terms on the right hand side. It follows directly that $\theta = \theta_0$ leads to the minimum value of the cost function $\bar{V}(\theta)$ being equal to σ_e^2. Similarly, because $\varepsilon_1(t,\theta_0) = e(t)$ it can simply be verified that the constraint equation $\varepsilon_2(t,\theta) = \eta\varepsilon_1(t,\theta)$ is satisfied for $\theta = \theta_0$.

In the second step we will show that any θ_1 that reaches the same cost function value as θ_0, i.e. $\bar{V}(\theta_1) = \bar{V}(\theta_0)$, while satisfying the constraint, must have the same dynamics as θ_0.

So consider any model θ_1 that satisfies $\bar{V}(\theta_1) = \bar{V}(\theta_0)$, and also satisfies the constraint $\varepsilon_2(t,\theta_1) = \eta(\theta_1)\varepsilon(t,\theta_1)$. It can be shown that

$$\begin{aligned}\bar{V}(\theta_1) - \bar{V}(\theta_0) = \bar{\mathbb{E}}\left(\varepsilon_1(t,\theta_1) - \varepsilon_1(t,\theta_0)\right)^2 + \\ +2\bar{\mathbb{E}}\left(\varepsilon_1(t,\theta_1) - \varepsilon_1(t,\theta_0)\right)\varepsilon_1(t,\theta_0).\end{aligned} \quad (24)$$

Analysing the second term we can use the fact that $\varepsilon_1(t,\theta_0) = e(t)$ being a white noise process, while $\varepsilon_1(t,\theta_1) - \varepsilon_1(t,\theta_0)$ 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 (24) will be 0, so that

$$\bar{V}(\theta_1) - \bar{V}(\theta_0) = \bar{\mathbb{E}}\left(\varepsilon_1(t,\theta_1) - \varepsilon_1(t,\theta_0)\right)^2 = 0 \quad (25)$$

which can equivalently be formulated as

$$\bar{V}(\theta_1) - \bar{V}(\theta_0) = \bar{\mathbb{E}}\left\{\begin{bmatrix}\tilde{P}_{11} & \tilde{P}_{12}\end{bmatrix}\begin{bmatrix}y_1(t)\\ u(t)\end{bmatrix}\right\}^2 = 0 \quad (26)$$

with $\tilde{P}_{ij} := P_{ij}(\theta_1) - P_{ij}(\theta_0)$ for $i,j = 1,2$. Since u is persistently exciting of a sufficiently high order, the process $\begin{bmatrix}y_1(t)\\ u(t)\end{bmatrix}$ is full rank which then implies that $\tilde{P}_{11} = \tilde{P}_{12} = 0$, which in turn implies that $H_a(\theta_1) = H_a(\theta_0)$ and $G_1(\theta_1) = G_1(\theta_0)$. For the constraint equation we will take a different path. We know that the constraint is satisfied for θ_0, so we can add it to the constraint of θ_1 as in

$$\eta(\theta_1)\varepsilon_1(\theta_1) - \varepsilon_2(\theta_1) - \eta(\theta_0)\varepsilon_1(\theta_0) + \varepsilon_2(\theta_0) = 0. \quad (27)$$

Next we define

$$\tilde{Q}_{1j} := \eta(\theta_1)P_{1j}(\theta_1) - \eta(\theta_0)P_{1j}(\theta_0) \quad (28)$$

for $j = 1,2$ such that we can write (27) as

$$\begin{bmatrix}\eta(\theta_1) - \eta(\theta_0) - \tilde{Q}_{11} + \tilde{P}_{21} & -\tilde{Q}_{12} + \tilde{P}_{22}\end{bmatrix}\begin{bmatrix}y_1\\ u\end{bmatrix} = 0.$$

[3]We will use the shorthand notation $H_a = H_a(\theta_0)$, $G_1 = G_1(\theta_0)$, $G_2 = G_2(\theta_0)$.

Since this constraint holds for all t and (y_1, u) is persistently exciting, this implies that the filters should be equal to 0, i.e.

$$\begin{aligned}0 &= \eta(\theta_1) - \eta(\theta_0) - \tilde{Q}_{11} + \tilde{P}_{21},\\ 0 &= -\tilde{Q}_{12} + \tilde{P}_{22}.\end{aligned}$$

When we combine the above two equations with the knowledge that G_1 and H_a are unique then we obtain that $H_b(\theta_1) = H_b(\theta_0)$ and $G_2(\theta_1) = G_2(\theta_0)$ which concludes the proof. □

Proof of Proposition 4: With the predictor form of Theorem 1 the constraint $\eta\varepsilon_1 = \varepsilon_2$ can be written as

$$\eta H_a^{-1}(y_1 - G_1 u) = y_2 - H_a^{-1}H_b^d y_1 + H_a^{-1}H_b^d G_1 u - G_2 u$$

where $H_b^d := H_b - \eta$. Multiplication with H_a, and moving all y_1 and G_1 terms to the left hand side delivers (20). Then when $\eta \neq 0$, H_b^{-1} is proper and we can multiply with H_a^{-1} and H_b^{-1} to obtain (21). □

REFERENCES

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

[2] D. Materassi and G. Innocenti, "Topological identification in networks of dynamical systems," *IEEE Trans. Automatic Control*, vol. 55, no. 8, pp. 1860–1871, 2010.

[3] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, "Exact topology identification of large-scale interconnected dynamical systems from compressive observations," in *Proc. American Control Conference (ACC)*, San Francisco, CA, USA, 2011, pp. 649–656.

[4] A. Chiuso and G. Pillonetto, "A Bayesian approach to sparse dynamic network identification," *Automatica*, vol. 48, no. 8, pp. 1553–1565, 2012.

[5] P. Torres, J. van Wingerden, and M. Verhaegen, "Hierarchical PO-MOESP subspace identification for directed acyclic graphs," *Intern. J. Control*, vol. 88, no. 1, pp. 123–137, 2015.

[6] 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.

[7] M. Deistler, B. D. O. Anderson, A. Filler, C. Zinner, and W. Chen, "Generalized linear dynamic factor models: An approach via singular autoregressions," *European J. Control*, vol. 16, no. 3, pp. 211 – 224, 2010.

[8] 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 Intern. Workshop on Adaptation and Learning in Control and Signal Processing (ALCOSP 2016)*, Eindhoven, The Netherlands, 2016.

[9] ——, "Identifiability of linear dynamic netwoks," 2017, arXiv:1609.00864[cs.SY].

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

[11] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, UK: Prentice-Hall International, 1989.

[12] N. Everitt, G. Bottegal, C. R. Rojas, and H. Hjalmarrsson, "On the effect of noise correlation in parameter identification of SIMO systems," in *Proc. 17th IFAC Symposium on System Identification (SYSID2015)*, 2015, pp. 326–331.

[13] H. H. M. Weerts, P. M. J. Van den Hof, and A. G. Dankers, "Identification of dynamic networks with rank-reduced process noise," in *Proc. 20th IFAC World Congress*, Toulouse, France, 2017, to appear.