Proceedings of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

A variance reduction technique for identification in dynamic networks

Bilal Gunes * **Arne Dankers** * **Paul M.J. Van den Hof** **

* *Delft Center for Systems and Control, Delft University of Technology, The Netherlands (e-mail:{bgunes@student., a.g.dankers@}tudelft.nl).*
** *Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands (e-mail: p.m.j.vandenhof@tue.nl).*

Abstract: With advancing technology, systems are becoming increasingly interconnected and form more complex networks. Additionally, more measurements are available from systems due to cheaper sensors. Hence there is a need for identification methods specifically designed for networks. For dynamic networks with known interconnection structures, several methods have been proposed for obtaining consistent estimates. We suppose that the internal variables in the network are measured with noise, but that there are external reference signals present in the network that are known exactly. A method that is able to deal with this situation is the two stage method, which solves several open loop identification problems sequentially. In this paper it is shown that solving the problems simultaneously leads to estimates with lower variance.

Keywords: System identification, dynamic networks, linear systems, variance

1. INTRODUCTION

A network is a set of subsystems (or modules) embedded according to an interconnection structure (Van den Hof et al. (2013)). Dynamic networks are becoming increasingly complex in engineering. In addition, the ability to take measurements using sensors is also increasing. Thus the identification in networks problem is becoming increasingly important. It is advantageous to address these identification problems as explicit *network* identification problems, because networks exhibit phenomena which do not appear in classic open and closed-loop systems.

The quality of an estimate can be assessed by determining if it is consistent and what its variance is. Consistent models with low variance are in demand. The variance of the parameters of the estimate define the confidence regions. The prediction error method provides tools to assess them.

Consistent estimates of a module embedded in a network can be obtained using several methods presented in the literature in Van den Hof et al. (2013). However, when the internal variables are measured with sensor noise, only the two-stage method (Van den Hof et al. (2013)) still results in consistent estimates. To use this method, there must be external variables present, which are known exactly. Examples of such signals could be reference signals in a control loop. In the current literature limited analysis is performed on the variance of network prediction error method estimates. What are the variance expressions? And can the variance of the two-stage method be reduced in a smart way?

⋆ The work of Arne Dankers is supported in part by the National Science and Engineering Research Council (NSERC) of Canada.

In Wahlberg et al. (2009) a method is proposed to reduce the variance of estimates in a cascade. It is shown that their alternative objective leads to a reduction of the variance of the estimate of the upstream transfer function in the cascade. The question arises how this variance reduction technique can be exploited for identification of a target module embedded in a dynamic network.

In this paper it is shown that the results of Wahlberg et al. (2009) for cascade systems can be extended to directed acyclic graphs. Additionally, it is shown that not only is the upstream module estimated with lower variance, but so is the downstream module. Furthermore, it is shown that the two stage method effectively transforms a network into a directed acyclic graph. Consequently, we combine these three results to obtain estimates of modules embedded in a network with lower variance than the two stage method. The mechanism by which the new method achieves lower variance is by simultaneously minimizing the set of prediction errors that are sequentially minimized in the two-stage method.

The background material is presented in section 2. The proposed method and its consistency properties are presented in Section 3. Its variance is treated in Section 4, and compared with the two-stage method in Sections 5, 6. Simulation results are in Section 7. and compared with the two-stage method in Sections 5, 6.

2. BACKGROUND

2.1 Dynamic networks

A dynamic network is built up of L elements related to L measured scalar *internal variables* $w_j, j = 1, \dots L$. Every internal variable in this network can be written as:

$$w_j(t) = \sum_{j \in \mathcal{N}_j} G_{jk}(q) w_k(t) + r_j(t) \tag{1}$$

- With G_{jk} a proper rational transfer function;
- with $\mathcal{N}_j$ the set of indices of internal variables with direct causal connection to w_j. $k \in \mathcal{N}_j$ iff $G_{jk} \neq 0$.
- with r_j an external variable that can possibly be manipulated by the user;
- with q^{-1} the delay operator (i.e. $q^{-1}u(t) = u(t-1)$)

The internal variables can be expressed as:

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \begin{bmatrix} 0 & G_{12} & \dots & G_{1L} \\ G_{21} & 0 & \ddots & G_{2L} \\ \vdots & \ddots & \ddots & \vdots \\ G_{L1} & G_{L2} & \dots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_L \end{bmatrix} \tag{2a}$$

$$= G(q)w + r(t) \tag{2b}$$

$$= (I - G(q))^{-1} r(t) \tag{2c}$$

Where it is assumed that $(I-G)^{-1}$ exists. Define $S = (I-G)^{-1}$. Some r_i may not be present: define $\mathcal{R}$ as the set of indices of present r_i. Any measurement can be expressed as:

$$\tilde{w}_k = w_k + s_k \tag{3}$$

Where s_k is a sensor error. It is a stationary stochastic process with power spectral density $\Phi_{s_k}(\omega) = \lambda_k$ (i.e. white noise). The proposed dynamic network has sensor noise, and no process noise. Every real sensor has some noise. This noise is presumed to originate from the internal workings of the individual sensors. Then every sensor will have a different error. Because networks have a large amount of measurements, it is important to deal with this explicitly. More complex noise frameworks will be investigated in future work. The following assumptions are made on dynamic networks:

Assumption 2.1.

- The network is well-posed in the sense that all principal minors of $(I - G(\infty))^{-1}$ are non-zero.
- $(I-G)^{-1}$ is stable
- The measurement noise sources are independent white noise sources $\Phi_n = \Lambda = diag(\lambda_1 \dots \lambda_L)$
- The external excitation signals are uncorrelated, i.e.: $R_{r_1 r_2}(\tau) = 0 \ \forall \tau$.

Define $w_{\mathcal{N}} = [w_{k_1} \dots w_{k_n}]^T$ and $G_{j\mathcal{N}} = [G_{jk_1} \dots G_{jk_n}]^T$, where $\{k_1 \dots k_n\} = \mathcal{N}_j$. Any measurement $\tilde{w}$ can then be written in either global or local form respectively as:

$$\tilde{w}_k(t) = \sum_{l \in \mathcal{R}} S_{kl}(q) r_l(t) + s_k(t) \tag{4a}$$

$$\tilde{w}_j(t) = \sum_{k \in \mathcal{N}_j} G_{jk}(q) w_k(t) + r_j(t) + s_j(t) \tag{4b}$$

Where S_{kl} is the (k,l)th element of $(I-G)^{-1}$.

2.2 The Prediction Error method

The prediction error method (Ljung (1999)) predicts the output with a one-step-ahead predictor. The prediction error is minimized to attain a model. The one-step ahead predictor:

$$\hat{w}_k(t|t-1) = \sum_{l \in \mathcal{R}} S_{kl}(q,\theta) r_l(t) \tag{5}$$

And the corresponding prediction error is:

$$\varepsilon_k(t) = \tilde{w}_k(t) - \hat{w}_k(t|t-1;\theta) \tag{6}$$

The unknown parameters are then estimated through a prediction error criterion based on a cost function V_N:

$$\hat{\theta}_N = arg\min_{\theta} V_N(\theta), \tag{7a}$$

$$V_N(\theta) = \sum_{t=0}^{N-1} \varepsilon^2(t,\theta) \tag{7b}$$

Where $V_N(\theta)$ is the sum of squared prediction errors.

Definition 2.1. An estimate $G_{jk}(q,\theta_N)$ is consistent if

$$G_{jk}(q,\hat{\theta}_N) \to G_{jk}(q), \text{ w.p. 1 as } N \to \infty$$

The variance of the estimate of the parameter vector θ in (7) is characterized by the following proposition:

Proposition 2.1. Suppose the Assumption 2.1 holds. Assume also the data set is informative enough. Then the covariance matrix of θ denoted P_θ is (Ljung (1999)):

$$P_\theta = M^{-1}, \tag{9a}$$

$$M = \bar{E}\psi(t,\theta_0)\Lambda^2(\psi(t,\theta_0))^T, \tag{9b}$$

$$\psi(t,\theta_0) = \left.\frac{\partial \varepsilon(t,\theta)}{\partial \theta}\right|_{\theta=\theta^0}, \tag{9c}$$

where $\bar{E}$ is the mean over time and ensemble (Ljung (1999)) and $\psi(t,\theta_0)$ represents the gradient of the prediction error evaluated at θ_0, and Λ is a diagonal matrix with the noise powers. M Represents the information matrix.

2.3 The two-stage method

The two-stage method (Van den Hof et al. (2013)) attempts to obtain a consistent estimate of a target module in a dynamic network. Its (second stage) predictor inputs are (asymptotically) noise-free estimates of internal variables. The two-stage method performs consecutive minimization of prediction errors.

In the first stage, the goal is to reconstruct the internal variables w_k (4). For this purpose, estimates of $S_{kl}, k \in \mathcal{N}_j, l \in \mathcal{R}$ are estimated by minimizing the quadratic cost function (7) of the prediction error:

$$\varepsilon_k(t,\alpha) = \tilde{w}_k(t) - \sum_{l \in \mathcal{R}} S_{kl}(q,\alpha) r_l(t) \tag{10}$$

Where α is a parameter vector. Let $S_{kl}(q,\hat{\alpha})$ denote the estimate of S_{kl}. The estimate of w_k is then:

$$\hat{w}_k(\hat{\alpha}) = \sum_{l \in \mathcal{R}} S_{kl}(q,\hat{\alpha}) r_l(t) \tag{11}$$

In the second stage, the estimates of the noise-free internal variables are used to identify the target module G_{ji} in an open-loop problem. Estimates are obtained by minimizing the quadratic cost function (7) of the prediction error:

$$\varepsilon_j(t,\hat{\alpha},\beta) = \tilde{w}_j(t) - r_j(t) - \sum_{k\in\mathcal{N}_j} G_{jk}(q,\beta)\hat{w}_k(\hat{\alpha}) \quad (12)$$

The two-stage method is defined as follows.

Algorithm 2.1. The Two Stage method

(1) Obtain estimates $\hat{w}_k$ of w_k for each $k \in \mathcal{N}$ using (10) and (7)
(2) Using $\hat{w}(\alpha)$ obtain estimates of the target module G_{ji} using (12) and (7)

Proposition 2.2. Consider a dynamic network as defined in section 2.1. Algorithm 2.1 provides consistent estimates of G_{ji} if the following conditions hold:

- The power spectral densities of $[\hat{w}_{k_1}\cdots\hat{w}_{k_n}], k_* \in \mathcal{N}_j$; $[r_{l_1}\cdots r_{l_n}], l_* \in \mathcal{R}$ are positive definite for $\omega \in [-\pi,\pi]$
- The parametrization is chosen flexible enough such that there exists a parameter vector θ^0 such that $G_{jk}(\beta^0) = G_{jk}$ and $S_{kl}(\alpha^0) = S_{kl}$

Notice that the two-stage method tackles the network identification problem as two sequential open-loop problems. The parametrization of the two-stage method can be represented as a directed acyclic graph. This is shown in figure 1 and the following equations:

$$w_{\mathcal{N}}(t) = S_{\mathcal{N}\mathcal{R}}(q)r(t) \quad (13a)$$
$$w_j(t) - r_j(t) = G_{j\mathcal{N}}(q)w_{\mathcal{N}}(t) \quad (13b)$$

Where $w_{\mathcal{N}} = [w_{k_1}\cdots w_{k_n}]', k_* \in \mathcal{N}_j$ is a vector.

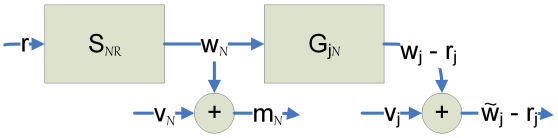

Fig. 1. The parametrization of the two-stage method can be represented as a directed acyclic graph (13).

2.4 Variance reduction technique for cascade systems

In Wahlberg et al. (2009), a similar framework is investigated for a cascade systems. Using our notation:

$$\tilde{w}_2(t) = S_{21}(q)r_1(t) + s_2(t) \quad (14a)$$
$$\tilde{w}_3(t) = G_{32}(q)w_2(t) + s_3(t) \quad (14b)$$

It is possible to consistently estimate S_{21} using only $\tilde{w}_2$, but Wahlberg et al. (2009) show that the variance of the estimate of S_{21} can be reduced by minimizing:

$$V_N(\theta) = \frac{1}{N}\sum_{t=0}^{N-1} \frac{\varepsilon_2(t,\theta)^2}{\lambda_1} + \frac{\varepsilon_3(t,\theta)^2}{\lambda_2}\ , \text{where:} \quad (15a)$$
$$\varepsilon_2(t,\theta) = \tilde{w}_2(t) - S_{21}(q,\alpha)r_1(t) \quad (15b)$$
$$\varepsilon_3(t,\theta) = \tilde{w}_3(t) - G_{32}(q,\beta)S_{21}(q,\alpha)r_1(t) \quad (15c)$$

Notice how the modified cost function utilizes an extra measurement. The information of S_{kl} in $\tilde{w}_3$ is being exploited. In our paper, no extra measurements are used. And the focus is on the target module G_{ji} embedded in a dynamic network rather than the direct open-loop identification of S. The question arises how this variance reduction technique can be modified for identification of the target module G embedded in a dynamic network.

3. SIMULTANEOUS MINIMIZATION OF PREDICTION ERRORS

In this section it is shown that the modified cost function presented in section 2.4 can be applied to dynamic networks. We use it to link the prediction errors of the two-stage method together. Instead of sequentially minimizing the prediction errors in the two-stage method, they can be simultaneously minimized. To apply this reasoning, the results of Wahlberg et al. (2009) must be extended in two ways. It will be shown that the downstream module is also estimated with lower variance. And it will be shown that the results of Wahlberg et al. (2009) hold for directed acyclic graphs as well. It has already been presented that the two-stage method parametrization transforms a dynamic network into a directed acyclic graph expression.

Consequently, we combine these results, and can obtain estimates of modules embedded in a dynamic *network* (including loops) with lower variance than the two stage method. It uses the same parametrization as the two-stage method, but uses a similar cost function as Wahlberg et al. (2009). The method can be presented as following:

Algorithm 3.1.

(1) Construct the prediction errors:

$$\varepsilon_{\mathcal{N}}(t,\alpha) = [\tilde{w}_{\mathcal{N}}(t) - \sum_{p=1}^{n_r} S_{\mathcal{N}p}(q,\alpha_p)r_p(t)]^T \quad (16)$$

$$\varepsilon_j(t,\theta) = \tilde{w}_j(t) - r_j(t) - G_{j\mathcal{N}}(q,\beta)\sum_{p=1}^{n_r} S_{\mathcal{N}p}(q,\alpha_p)r_p(t)$$

Where α is partitioned into $[\alpha_1\ldots\alpha_{n_r}]^T$, such that $S_{\mathcal{N}p}(q,\alpha_p)$. Note that these are the same prediction errors as (10) and (12) for the two-stage method. And that both $\varepsilon_{\mathcal{N}}(t,\alpha)$ and $\varepsilon_j(t,\alpha,\beta)$ are a function of $S_{\mathcal{N}p}(q,\alpha_p)$

(2) Obtain estimates of $G_{jk}(q)$ by minimizing:

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}[\frac{\varepsilon_j(t,\alpha,\beta)^2}{\lambda_j} + \sum_{k\in\mathcal{N}_j}\frac{\varepsilon_k(t,\alpha)^2}{\lambda_k}] \quad (17)$$

The simultaneous cost function and common parametrization in S_{kl} links the prediction errors of the two stages, such that $\tilde{w}_j(t)$ does get utilized in the estimation of S_{kl}. This results in a variance reduction of the estimate of S_{kl} by extending the reasoning of Wahlberg to multiple inputs. In this paper, it is shown that the variance of the estimate of the target module G_{jk} reduces as well.

Proposition 3.1. Under the conditions of Proposition 2.2, Algorithm 3.1 results in consistent estimates of G^0_{jk}, S^0_{kl}. The proof is presented in Gunes (2013).

Note that minimizing (17) is a non-convex optimization problem. However, the optimizer can always be initialized by using initial estimates and model orders attained by performing the two-stage method on the dataset first, because the consistency conditions (Proposition 3.1) match.

4. VARIANCE EXPRESSIONS FOR SIMULTANEOUS MINIMIZATION OF PREDICTION ERRORS

In this section, the variance expressions for simultaneous minimization of prediction errors will be presented. Consider a dynamic network as defined in Section 2.1 that satisfies Assumption 2.1 and the conditions of 3.1. In order to derive the covariance matrix of θ an expression of the prediction error gradient is required. The system equations have been presented in (4), and the prediction errors are presented in (16). Define $\varepsilon = [\varepsilon_{\mathcal{N}} \quad \varepsilon_j]$ as a vector of the prediction errors. The prediction error gradient is:

$$\psi(t,\alpha,\beta) = -\begin{bmatrix} \dfrac{\partial \varepsilon_{\mathcal{N}}(t,\alpha)}{\partial \alpha} & \dfrac{\partial \varepsilon_j(t,\alpha,\beta)}{\partial \alpha} \\ 0 & \dfrac{\partial \varepsilon_j(t,\alpha,\beta)}{\partial \beta} \end{bmatrix} \tag{18}$$

Define α_p as the parameter vector associated to the column vector $S_{\mathcal{N}p}$, and r_p. Define $S'_{\mathcal{N}p} = \frac{\partial S_{\mathcal{N}p}(q,\alpha_p)}{\partial \alpha_p}\big|_{\theta=\theta^0}$. Notice $G_{j\mathcal{N}}$ is a row vector. Define $G'_{j\mathcal{N}} = [\frac{\partial G^T_{j\mathcal{N}}(q,\beta)}{\partial \beta}]^T\big|_{\theta=\theta^0}$ for ease of notation. The prediction error gradient blocks of (18) can then be written as:

$$-\frac{\partial \varepsilon_{\mathcal{N}}(t,\alpha)}{\partial \alpha_p} = [S'_{\mathcal{N}p}(q) r_p(t)]^T \tag{19a}$$

$$-\frac{\partial \varepsilon_j(t,\alpha,\beta)}{\partial \alpha_p} = [G_{j\mathcal{N}}(q) S'_{\mathcal{N}p}(q) r_p(t)]^T \tag{19b}$$

$$-\frac{\partial \varepsilon_j(t,\alpha,\beta)}{\partial \beta} = G'_{j\mathcal{N}}(q) \sum_{p=1}^{n_r} S_{\mathcal{N}p}(q) r_p(t) \tag{19c}$$

Then M as defined in (9), can be partitioned according to α and β as:

$$M = E[\psi(t,\theta^0)\Lambda\psi(t,\theta^0)^H] \tag{20a}$$

$$= \begin{bmatrix} A+F & C^H \\ C & D \end{bmatrix}, \text{ where:} \tag{20b}$$

$$A_{pq} = E[(S'_{\mathcal{N}p} r_p)^T \Lambda_{\mathcal{N}}^{-1} S'_{\mathcal{N}q} r_q]$$
$$F_{pq} = E[(G_{j\mathcal{N}} S'_{\mathcal{N}p} r_p)^T \lambda_j^{-1} G_{j\mathcal{N}} S'_{\mathcal{N}q} r_q]$$
$$C = [C_1 \dots C_{n_r}] \text{ where, for } p = 1, \dots, n_r:$$
$$C_p = E[G'_{j\mathcal{N}}(q) S_{\mathcal{N}p}(q) r_p(t) \lambda_j^{-1} G_{j\mathcal{N}}(q) S'_{\mathcal{N}p}(q) r_p(t)]$$
$$D = \sum_{p=1}^{n_r} E\big[G'_{j\mathcal{N}}(q) S_{\mathcal{N}p}(q) r_p(t) \lambda_j^{-1} r_p(t) S^T_{\mathcal{N}p}(q) G'^T_{j\mathcal{N}}(q)\big]$$

Note that simplifications have been performed using the non-correlation between the external excitation signals. I.e., the off-diagonal blocks of A and F are zero and D is symmetric. Recall $P = M^{-1}$ (9). Using Schur's complement, the following proposition can be set up:

Proposition 4.1. Consider a dynamic network as defined in section 2.1. Suppose that the conditions of Proposition 3.1 are met. The variance expressions of $\theta = [\alpha\beta]$ using Algorithm 3.1 is:

$$P^{si} = \begin{bmatrix} P^{si}_{\alpha} & -P^{si}_{\alpha} C^T D^{-1} \\ -D^{-1} C P^{si}_{\alpha} & D^{-1} + D^{-1} C P^{si}_{\alpha} C^T D^{-1} \end{bmatrix}$$
$$P^{si}_{\alpha} = [A + F - C^T D^{-1} C]^{-1}$$

Where the superscript ^{si} indicates the use of the method 3.1. And where the top-left block is the covariance matrix of α and the bottom-right block is the covariance matrix of β: P^{si}_{β}. The other variables are defined in (20). Note that to evaluate these expressions in practice, the noise powers λ_j and $\lambda_k, k \in \mathcal{N}_j$ need to be known.

5. VARIANCE EXPRESSIONS FOR THE TWO-STAGE METHOD

In this section, the variance expressions for the two-stage method for the current framework will be presented. Unlike Forssell and Ljung (1999) and Gevers et al. (2001), the variance expressions are independent of individual realizations of the first stage estimate due to the fact only sensor noise is present (and not process noise).

Proposition 5.1. Consider a dynamic network as defined in section 2.1. Suppose that the conditions of Proposition 2.2 are met. The variance expressions for α and β obtained using the two-stage method presented in Algorithm 2.1 are:

$$P^{2S}_{\alpha} = A^{-1}$$
$$P^{2S}_{\beta} = D^{-1} + \frac{1}{\lambda_j^2} D^{-1} Q_2 D^{-1}$$

Where A and D are defined in (20), and where Q_2 is:

$$Q_2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} G'_{j\mathcal{N}} \sum_{p=1}^{n_r} S_{\mathcal{N}p} \bar{\Phi}_{d+} \Phi_{r_p} S^H_{\mathcal{N}p} G'^H_{j\mathcal{N}} \, d\omega$$
$$\bar{\Phi}_{d+}(\alpha) = G_{j\mathcal{N}} [\sum_{q=1}^{n_r} S'_{\mathcal{N}q} P^{2S}_{\alpha_q} S'^H_{\mathcal{N}q} \Phi_{r_q}] G^H_{j\mathcal{N}}$$

The proof is presented in appendix B. Again, to evaluate these expressions in practice, the noise powers need to be known.

6. VARIANCE COMPARISON

The variance expressions for the two-stage method (Algorithm 2.1) and the simultaneous minimization of prediction errors (Algorithm 3.1) are briefly presented in table 1. The goal is to achieve variance reduction of the target module estimate. Does the simultaneous minimization of prediction errors indeed result in target module estimates with lower variance than the two-stage does?

Define $Z = F - C^T D^{-1} C$. It is proven in Appendix A that $Z \geq 0$. Hence $A^{-1} \geq [A+Z]^{-1}$ and the simultaneous minimization of prediction errors results in a lower and better P_{α}. Z also appears in P_{β}, hinting to an improvement of the variance of the target module estimate. In the following proposition it is shown that P_{β} is also less.

Proposition 6.1. Consider the variance results summarized in table 1. The variance of estimates (including the target module) of simultaneous minimization of prediction

Table 1. Variance results summary (5.1, 4.1)

Method	Two-stage method	Simultaneous minimization
P_α	A^{-1}	$[A+Z]^{-1}$
P_β	$D^{-1}+D^{-1}Q_2D^{-1}$	$D^{-1}+D^{-1}C[A+Z]^{-1}C^TD^{-1}$

errors is equal to or smaller than the two-stage method ones. The proof is presented in Appendix C.

7. SIMULATION RESULTS

In this section, Monte Carlo simulation results are presented to illustrate the results from the previous sections. The results are based on 200 Monte Carlo simulations of the dynamic network presented in Fig. 2. The noise is white with powers 0.03, 0.0001 and 0.03. The external excitation signals are white and unit power. The data size is 250. The results are presented in Fig. 3.

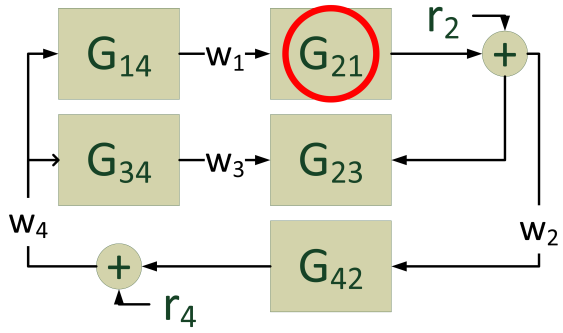

Fig. 2. Example dynamic network with target module G_{21}.

It appears that both methods provide consistent estimates, but that simultaneous minimization of prediction errors results in estimates with lower variance.

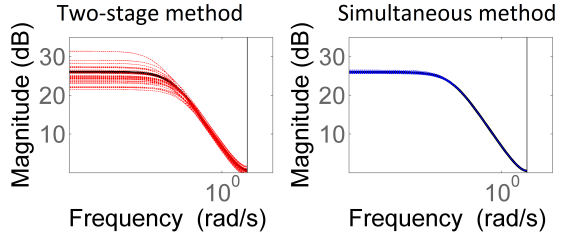

Fig. 3. Magnitude plots of the estimates of G_{21}. The thick black line represents the true system.

8. CONCLUSION AND FUTURE WORK

In this paper, a novel method for network prediction error identification has been presented. Simultaneous minimization of prediction errors is a combination of the two-stage method (Van den Hof et al. (2013)) and simultaneous minimization of prediction errors (Wahlberg et al. (2009)). It has the same consistency properties as the two-stage method, but considerably lower variance in the presence of measurement noise. The mechanism by which the new method achieves lower variance is by simultaneously minimizing the set of prediction errors that are sequentially minimized in the two-stage method. In future work, the proposed method will be extended to deal with more general cases of available measurements. This has been done in the literature for the two-stage method in Dankers et al. (2013). Other future work will be extending to cases with both measurement and process noise.

REFERENCES

Dankers, A.G., van den Hof, P.M.J., Heuberger, P.S.C., and Bombois, X. (2013). Identification of dynamic models in complex networks with prediction error methods - predictor input selection. *Submitted to IEEE Transactions on Automatic Control*.

Forssell, U. and Ljung, L. (1999). Closed-loop identification revisited. *Automatica*, 35(7), 1215 – 1241.

Gevers, M., Ljung, L., and van den Hof, P.M.J. (2001). Asymptotic variance expressions for closed-loop identification. *Automatica*, 37(5), 781 – 786.

Gunes, B. (2013). *A novel network prediction error identification method*. Master's thesis, Delft University of Technology.

Ljung, L. (1999). *System identification (2nd ed.): theory for the user*. Prentice Hall PTR, Upper Saddle River, NJ, USA.

Van den Hof, P.M.J., Dankers, A.G., Heuberger, P.S.C., and Bombois, X. (2013). Identification of dynamic models in complex networks with prediction error methods: basic methods for consistent module estimates. *Automatica*, 49(10), 2994 – 3006.

Wahlberg, B., Hjalmarsson, H., and Martensson, J. (2009). Variance results for identification of cascade systems. *Automatica*, 45(6), 1443 – 1448.

Willems, J. (2008). Modeling interconnected systems. In *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, 421–424.

Appendix A. PROOF FOR $Z \succeq 0$

Recall $Z = F - C^TD^{-1}C$. $Z \succeq 0$ Is equivalent to:

$$\begin{bmatrix} F & C^T \\ C & D \end{bmatrix} = \begin{bmatrix} E[y_1y_1^T] & E[y_1y_2^T] \\ E[y_2y_1^T] & E[y_2y_2^T] \end{bmatrix} \succeq 0 \tag{A.1a}$$

$$y_1(t) = [G_{j\mathcal{N}}S'_{\mathcal{N}1}r_1 \dots G_{j\mathcal{N}}S'_{\mathcal{N}n_r}r_{n_r}]^T\Lambda_2^{-0.5} \tag{A.1b}$$

$$y_2(t) = G'_{j\mathcal{N}}\sum_{p=1}^{n_r} S_{\mathcal{N}p}(q)r_p(t)\Lambda_2^{-0.5} \tag{A.1c}$$

Which is the power spectral density of $[y_1(t)\ y_2(t)]$ and hence by definition positive semi-definite.

Appendix B. PROOF OF PROPOSITION 5.1

Consider the two-stage method presented in Algorithm 2.1. In the first stage, the parameter vector α is estimated using the prediction error (10). Using proposition 2.1, the covariance matrix of α can be derived to be A^{-1} ((20)).

In the second stage, $G_{j\mathcal{N}}(q)$ is estimated as an open-loop problem. Because its input is not available, the estimated input $\hat{w}_\mathcal{N}$ (based on $\hat{\alpha}$) is used instead. Rewriting the system equation in $\hat{w}_\mathcal{N}$ results into:

$$m_j(t) = G_{j\mathcal{N}}(q)\hat{w}_{\mathcal{N}}(t) + r_j(t) + d(t) \tag{B.1a}$$
$$\hat{w}_{\mathcal{N}}(t) = \sum_{p=1}^{n_r} S_{\mathcal{N}p}(q,\hat{\alpha}_p)r_p(t) \tag{B.1b}$$
$$d(t) = n_j(t) + d_+(t) \tag{B.1c}$$
$$d_+(t) = G_{j\mathcal{N}}(q)\sum_{p=1}^{n_r}\big(S_{\mathcal{N}p}(q) - S_{\mathcal{N}p}(q,\hat{\alpha}_p)\big)r_p(t) \tag{B.1d}$$

Where d is the predictor noise term. The prediction error is (12). The covariance matrix is Forssell and Ljung (1999):

$$P_\beta = R^{-1}Q_1R^{-1} + R^{-1}Q_2R^{-1} \tag{B.2a}$$
$$R = \frac{1}{2\pi}\int_{-\pi}^{\pi} G'_{j\mathcal{N}}\Phi_{\hat{w}_\mathcal{N}}G'^H_{j\mathcal{N}}\,d\omega \tag{B.2b}$$
$$Q = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_d G'_{j\mathcal{N}}\Phi_{\hat{w}_\mathcal{N}}G'^H_{j\mathcal{N}}\,d\omega \tag{B.2c}$$

Notice that $n_j(t)$ and $d_+(t)$ are uncorrelated because in the framework we consider, $n_\mathcal{N}$ is uncorrelated with n_j. Only the noise source $n_\mathcal{N}$ affects $S_{\mathcal{N}p}(q,\hat{\alpha}_p)$ and d_+. As a result, $\Phi_d = \Phi_{n_j} + \Phi_{d+}$. Use this to split Q into $Q_1 + Q_2$:

$$Q_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{n_j} G'_{j\mathcal{N}}\Phi_{\hat{w}_\mathcal{N}}G'^H_{j\mathcal{N}}\,d\omega \tag{B.3a}$$
$$Q_2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{d+} G'_{j\mathcal{N}}\Phi_{\hat{w}_\mathcal{N}}G'^H_{j\mathcal{N}}\,d\omega \tag{B.3b}$$

First, obtain an expression for Φ_{d+}. From (B.1), Φ_{d+} can be expressed as:

$$\Phi_{d+}(\alpha) = G_{j\mathcal{N}}\Delta S_{\mathcal{N}\mathcal{R}}(\alpha)\Phi_r\Delta S^H_{\mathcal{N}\mathcal{R}}(\alpha)G^H_{j\mathcal{N}}$$

Where $\Delta S_{\mathcal{N}\mathcal{R}}(q,\alpha) = S^0_{\mathcal{N}\mathcal{R}}(q) - S_{\mathcal{N}\mathcal{R}}(q,\alpha)$. The external excitation signals are uncorrelated, such that:

$$\Phi_{d+}(\alpha) = G_{j\mathcal{N}}[\sum_{p=1}^{n_r}\Delta S_{\mathcal{N}p}(\alpha)\Delta S^H_{\mathcal{N}p}(\alpha)\Phi_{r_p}(\omega)]G^H_{j\mathcal{N}}$$

Where $\Phi_{r_p}(\omega)$ is a scalar. For every element of this sum, a first order Taylor approximation can be performed:

$$S_{\mathcal{N}p}(e^{j\omega},\alpha_p) \approx S_{\mathcal{N}p}(e^{j\omega},\alpha_p^0) + S'_{\mathcal{N}p}(e^{j\omega},\alpha_p^0)(\alpha_p - \alpha_p^0)$$
$$\Delta S_{\mathcal{N}p}(e^{j\omega},\alpha_p) \approx -S'_{\mathcal{N}p}(e^{j\omega},\alpha_p^0)(\alpha_p - \alpha_p^0)$$

The approximation holds for α_p close to α_p^0. From (B.4), using this approximation, it follows that:

$$\Delta S_{\mathcal{N}p}(e^{j\omega},\alpha_p)\Delta S^H_{\mathcal{N}p}(e^{j\omega},\alpha_p) = \tag{B.6}$$
$$S'_{\mathcal{N}p}(e^{j\omega})(\alpha_p - \alpha_p^0)(\alpha_p - \alpha_p^0)^T S'^H_{\mathcal{N}p}(e^{j\omega}) \tag{B.7}$$

$\Phi_{d+}(e^{j\omega},\alpha)$ Depends on the realization of α. By averaging over the ensemble of α, a useful measure $\bar{\Phi}_{d+}(e^{j\omega},\alpha)$ can be attained. Notice the covariance of α is defined as:

$$P_\alpha^{2S} = E[(\alpha_p - \alpha_p^0)(\alpha_p - \alpha_p^0)^T] \tag{B.8}$$

Substituting this expression into (B.6) leads to:

$$\bar{\Phi}_{d+}(\alpha) = G_{j\mathcal{N}}[\sum_{p=1}^{n_r} S'_{\mathcal{N}p}P^{2S}_{\alpha_p}S'^H_{\mathcal{N}p}]G^H_{j\mathcal{N}}\Phi_{r_p}(\omega) \tag{B.9}$$

In similar fashion, replace $\Phi_{\hat{w}_\mathcal{N}}$ by $\Phi_{w_\mathcal{N}}$ (13). Also notice that $R^{-1}QR^{-1} = R^{-1}\lambda_j$, because $\Phi_{n_j} = \lambda_j$. Use $R = \lambda_j D$ (20). This leads to the expressions in Proposition 5.1.

Appendix C. PROOF OF PROPOSITION 6.1

Consider Proposition 6.1 as the lemma. Appendix A proofs $Z \succeq 0$, such that $A + Z \succeq A$ and $[A+Z]^{-1} \preceq A^{-1}$. This concludes the proof for P_α. For P_β, the proof reduces to:

$$\frac{1}{\lambda_j^2}Q_2 \succeq CP^{si}_\alpha C^H \tag{C.1}$$

First, the proof will be presented for the case of $Z = 0$, and afterwards a generalization to $Z \succeq 0$ will be presented. For $Z = 0$, $P^{si}_\alpha = [A+0]^{-1} = P^{2S}_\alpha$, P_α. A and P_α are block-diagonal. Then $CP^{si}_\alpha C^H = \sum_{p=1}^{n_r} C_pP^{si}_{\alpha_p}C_p^H$. Rewrite Q_2:

$$Q_2 = \left[\sum_{p=1}^{n_r}\frac{1}{2\pi}\int_{-\pi}^{\pi} G'_{j\mathcal{N}}S_{\mathcal{N}p}G_{j\mathcal{N}}S'_{\mathcal{N}p}P^{2S}_{\alpha_p}S'^H_{\mathcal{N}p}G^H_{j\mathcal{N}} S^T_{\mathcal{N}p}G'^H_{j\mathcal{N}}\Phi^2_{r_p}\,d\omega\right] + \left[\sum_{p=1}^{n_r}\sum_{q=1}^{n_r,q\neq p}\frac{1}{2\pi}\int_{-\pi}^{\pi} G'_{j\mathcal{N}}S_{\mathcal{N}p} G_{j\mathcal{N}}S'_{\mathcal{N}q}P^{2S}_{\alpha_q}S'^H_{\mathcal{N}q}G^H_{j\mathcal{N}}S^T_{\mathcal{N}p}G'^H_{j\mathcal{N}}\Phi_{r_p}\Phi_{r_q}\,d\omega\right]$$

Where Q_2 is split into two terms. Denote the second term V. V Has the form $KP^{2S}_{\alpha_p}K^T\Phi_{r_p}\Phi_{r_q}$. Since Φ_{r_p}, Φ_{r_q} are scalar magnitudes $\succeq 0$ and $P^{2S}_{\alpha_p} \succeq 0$: $V \succeq 0$. Next, define $A_p = G'_{j\mathcal{N}}S_{\mathcal{N}p}G_{j\mathcal{N}}S'_{\mathcal{N}p}\Phi_{r_p}F$, where F, $P_{\alpha_p} = FF^T$. Using this notation, (C.1) can be rewritten as:

$$\frac{1}{\lambda_j^2}\sum_{p=1}^{n_r}\frac{1}{2\pi}\int_{-\pi}^{\pi} A_pA_p^H d\omega + V \succeq \tag{C.3a}$$
$$\sum_{p=1}^{n_r}\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} A_p\frac{1}{\lambda_j}d\omega\right]\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} A_p^H\frac{1}{\lambda_j}d\omega\right] \tag{C.3b}$$

Consider the stronger lemma without the summer and V:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} A_pA_p^H d\omega \succeq \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} A_p d\omega\right]\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} A_p^H d\omega\right]$$

Suppose $\mathcal{F}[a_p(t)] = A_p(e^{j\omega})$, where $\mathcal{F}$ denotes the Fourier transformation. Then:

$$a_p(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A_p(e^{j\omega})e^{j\omega t}d\omega$$
$$a_p(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A_p(e^{j\omega})d\omega$$
$$\frac{1}{2\pi}\int_{-\pi}^{\pi} A_p(e^{j\omega})A_p(e^{j\omega})^H d\omega = \sum_{t=-\infty}^{\infty} a_p(t)a_p^T(t) \succeq 0\ \forall t$$

The lemma (C.4) can then be rewritten as:

$$\sum_{t=-\infty}^{-1} a_p(t)a_p^T(t) + a_p(0)a_p^T(0) + \sum_{t=1}^{\infty} a_p(t)a_p^T(t) \succeq a_p(0)a_p^T(0)$$

Which holds. This concludes the proof for $Z = 0$. For the remaining cases $Z \succ 0$ (A), the two-stage method variance expressions 5.1 are unaffected, but the simultaneous estimation of prediction errors variance expressions 4.1 are strictly decreasing. The proof thus extends to any Z.