On the Effect of Noise Correlation in Parameter Identification of SIMO Systems

Article · December 2015

DOI: 10.1016/j.ifacol.2015.12.148

4 authors:

Niklas Everitt
KTH Royal Institute of Technology

Giulio Bottegal
Technische Universiteit Eindhoven

Cristian Ricardo Rojas
KTH Royal Institute of Technology

Håkan Hjalmarsson
KTH Royal Institute of Technology

Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 48-28 (2015) 326–331

On the Effect of Noise Correlation in Parameter Identification of SIMO Systems ⋆

Niklas Everitt ∗, **Giulio Bottegal** ∗, **Cristian R. Rojas** ∗, **Håkan Hjalmarsson** ∗

∗ *Department of Automatic Control and ACCESS Linnaeus Centre, School of Electrical Engineering, KTH–Royal Institute of Technology, Stockholm, Sweden.*
E-mail: {neveritt, bottegal, crro, hjalmars}@kth.se

Abstract: The accuracy of identified linear time-invariant single-input multi-output (SIMO) models can be improved when the disturbances affecting the output measurements are spatially correlated. Given a linear parametrization of the modules composing the SIMO structure, we show that the correlation structure of the noise sources and the model structure of the other modules determine the variance of a parameter estimate. In particular we show that increasing the model order only increases the variance of other modules up to a point. We precisely characterize the variance error of the parameter estimates for finite model orders. We quantify the effect of noise correlation structure, model structure and signal spectra.

1. INTRODUCTION

Dynamic networks are gaining popularity within the system identification community, see e.g., Van den Hof et al. [2013], Dankers et al. [2013a,b, 2014], Ali et al. [2011], Materassi et al. [2011], Torres et al. [2014], Haber and Verhaegen [2014], Gunes et al. [2014], Chiuso and Pillonetto [2012]. A dynamic network is composed of a set of nodes and edges, where nodes represent signals and edges represent transfer functions. In this framework, the user has the freedom to choose which signals to include in the estimation. However, questions such as how to chose these signals, and how large the potential is for variance reduction, have not been extensively studied. For a few specific network structures, some results are available see e.g., Gevers et al. [2006], Hägg et al. [2011], Wahlberg et al. [2009], Everitt et al. [2014], Ramazi et al. [2014].

To understand the potential of including additional signals in the identification process, we will focus on a special case of dynamic networks, namely, single-input multi-output (SIMO) orthonormal basis function models. We will investigate if, and by how much, an added sensor improve the accuracy of an estimate of a certain target transfer function. In particular, we will focus on the role the correlation structure between noise sources play in the variance of a module estimate. The assumption of orthonormal basis function models is not restrictive since a model consisting of non-orthonormal basis functions can be transformed to this form by a linear transformation.

In this paper, we assume that the dynamics of the true system can be accurately described by the model structure, i.e., the true system lies within the set of models considered, and thus the bias (systematic) error is zero. Then, the model error mainly consists of the variance error, which is caused by disturbances and noise when the model is estimated using a finite number of input-output samples. The variance error will be quantified in terms of the noise covariance matrix, input spectrum and model structure. The paper also extends some results reported in Bottegal and Hjalmarsson [2014] to more general model structures.

As we shall see throughout the paper, there are some interesting aspects related to model parametrization and noise correlation structure. To give a flavor of these aspects, we will consider the following two output example:

$$y_1(t) = \theta_{1,1}u(t-1) + e_1(t),$$
$$y_2(t) = \theta_{2,2}u(t-2) + e_2(t),$$

where the input $u(t)$ is white noise and e_k $(k = 1, 2)$ is measurement noise. We consider two different types of measurement noise (uncorrelated with the input). In the first case, the noise is perfectly correlated. For simplicity, let us assume that $e_1(t) = e_2(t)$. As a second case, assume that $e_1(t)$ and $e_2(t)$ are independent. For the first case, it turns out that we can perfectly recover the parameters $\theta_{1,1}$ and $\theta_{2,2}$. However, in the second case, the accuracy of the estimate of $\theta_{1,1}$ is not improved by also using the measurement $y_2(t)$. The reason is that, in the first case, we can construct the noise free equation

$$y_1(t) - y_2(t) = \theta_{1,1}u(t-1) - \theta_{2,2}u(t-2)$$

and we can perfectly recover $\theta_{1,1}$ and $\theta_{2,2}$. However, in the second case, neither $y_2(t)$ nor $e_2(t)$ contain information about $e_1(t)$. In this paper we will generalize these observations to a wider class of SIMO models.

The paper has the following organization: in Section 2 we define the SIMO model structure under study and provide an expression for the covariance matrix of the parameter estimates. Section 3 contains the main results, namely, novel variance expression for LTI SIMO orthonormal basis function models. In Section 4, numerical experiments are presented that illustrate the application of the derived results. A final discussion ends the paper in Section 5.

⋆ This work was partially supported by the Swedish Research Council under contract 621-2009-4017, and by the European Research Council under the advanced grant LEARN, contract 267381.

10.1016/j.ifacol.2015.12.148

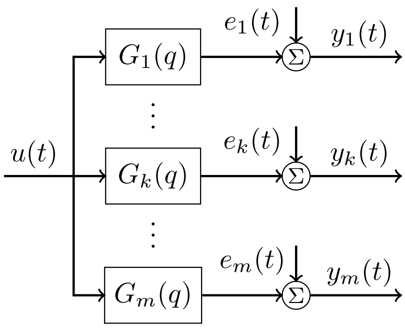

Fig. 1. *Block scheme of the linear SIMO system.*

2. SIMO MODELLING AND IDENTIFICATION

We consider linear time-invariant dynamic systems with one input and m outputs (see Fig. 1). The model is described as follows:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} = \begin{bmatrix} G_1(q) \\ G_2(q) \\ \vdots \\ G_m(q) \end{bmatrix} u(t) + \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_m(t) \end{bmatrix}, \qquad (1)$$

where q denotes the forward shift operator, i.e., $qu(t) = u(t+1)$ and the $G_i(q)$ are causal stable rational transfer functions. The G_i are modeled as

$$G_i(q, \theta_i) = \Gamma_i(q)\theta_i, \quad \theta_i \in \mathbb{R}^{n_i}; \quad i = 1, \ldots, m, \qquad (2)$$

where, without loss of generality, $n_1 \leq \ldots \leq n_m$ and $\Gamma_i(q) = [\mathcal{B}_1(q), \ldots, \mathcal{B}_{n_i}(q)]$, for some orthonormal basis functions $\{\mathcal{B}_k(q)\}_{k=1}^{n_m}$. Here, orthogonality is defined with respect to the scalar product defined for complex functions $f(z), g(z) : \mathbb{C} \to \mathbb{C}^{1\times m}$ as $\langle f, g\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\omega})g^*(e^{i\omega})\, \mathrm{d}\omega$. Let us introduce the vector notation

$$y(t) := \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix}, \quad e(t) := \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_m(t) \end{bmatrix}.$$

The noise sequence $\{e(t)\}$ is zero mean and temporally white, but may be correlated in the spatial domain:

$$\mathrm{E}\left[e(t)\right] = 0$$
$$\mathrm{E}\left[e(t)e(s)^T\right] = \delta_{t-s}\Lambda \qquad (3)$$

for some positive definite matrix covariance matrix Λ, and where $\mathrm{E}\left[\cdot\right]$ is the expectation operator. We express Λ in terms of its Cholesky factorization

$$\Lambda = \Lambda_{CH}\Lambda_{CH}^T, \qquad (4)$$

where $\Lambda_{CH} \in \mathbb{R}^{m\times m}$ is lower triangular, i.e.,

$$\Lambda_{CH} = \begin{bmatrix} \gamma_{11} & 0 & \ldots & 0 \\ \gamma_{21} & \gamma_{22} & \ldots & 0 \\ \vdots & \ldots & \ddots & 0 \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm} \end{bmatrix} \qquad (5)$$

for some $\{\gamma_{ij}\}$. Also notice that since $\Lambda > 0$,

$$\Lambda^{-1} = \Lambda_{CH}^{-T}\Lambda_{CH}^{-1}. \qquad (6)$$

We summarize the assumptions on input, noise and model as follows:

Assumption 1. The input $\{u(t)\}$ is zero mean stationary white noise with finite moments of all orders, and variance $\sigma^2 > 0$. The noise $\{e(t)\}$ is zero mean and temporally white, i.e, (3) holds with $\Lambda > 0$. It is assumed that $\mathrm{E}\left[|e(t)|^{4+\rho}\right] < \infty$ for some $\rho > 0$. The data is generated in open loop, that is, the input $\{u(t)\}$ is independent of the noise $\{e(t)\}$. The true input-output behavior of the data generating system can be captured by our model structure, i.e., the true system can be described by (1) and (2) for some parameters $\theta_i^o \in \mathbb{R}^{n_i}$, $i = 1, \ldots, m$, where $n_1 \leq \ldots \leq n_m$. The orthonormal basis functions $\{\mathcal{B}_k(q)\}$ are assumed stable. □

2.1 Weighted least-squares estimate

By introducing $\theta = \left[\theta_1^T, \ldots, \theta_m^T\right]^T \in \mathbb{R}^n$, $n := \sum_{i=1}^m n_i$ and the $n \times m$ transfer function matrix

$$\tilde{\Psi}(q) := \begin{bmatrix} \Gamma_1^T & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Gamma_m^T \end{bmatrix},$$

we can write the model (1) as a linear regression model

$$y(t) = \varphi^T(t)\theta + e(t), \qquad (7)$$

where

$$\varphi^T(t) = \tilde{\Psi}(q)^T u(t).$$

A consistent and unbiased estimate of the parameter vector θ can be obtained from weighted least-squares, with weighting matrix Λ^{-1} giving the linear unbiased estimator with lowest variance (see, e.g., Ljung [1999], Söderström and Stoica [1989]). The covariance Λ is assumed known, however, this assumption is not restrictive since Λ can be estimated from data and replacing Λ by a consistent estimate does not affect the asymptotic covariance of $\hat{\theta}$ [Cox and Reid 1987]. The estimate of θ is given by

$$\hat{\theta}_N = \left(\sum_{t=1}^N \varphi(t)\Lambda^{-1}\varphi^T(t)\right)^{-1} \sum_{t=1}^N \varphi(t)\Lambda^{-1}y(t). \qquad (8)$$

Inserting (7) in (8) gives

$$\hat{\theta}_N = \theta + \left(\sum_{t=1}^N \varphi(t)\Lambda^{-1}\varphi^T(t)\right)^{-1} \sum_{t=1}^N \varphi(t)\Lambda^{-1}e(t).$$

Under Assumption 1, the noise sequence is zero mean, hence $\hat{\theta}_N$ is unbiased. It can be noted that this is the same estimate as the one obtained by the prediction error method and, if the noise is Gaussian, by the maximum likelihood method [Ljung 1999]. It also follows that the asymptotic covariance matrix of the parameter estimates is given by

$$\mathrm{AsCov}\,\hat{\theta}_N = \left(\mathrm{E}\left[\varphi(t)\Lambda^{-1}\varphi^T(t)\right]\right)^{-1}. \qquad (9)$$

Here $\mathrm{AsCov}\,\hat{\theta}_N$ is the asymptotic covariance matrix of the parameter estimates, in the sense that the asymptotic covariance matrix of a stochastic sequence $\{f_N\}_{N=1}^{\infty}$, $f_N \in \mathbb{C}^{1\times q}$ is defined as [1]

$$\mathrm{AsCov}\, f_N := \lim_{N\to\infty} N \cdot \mathrm{E}\left[(f_N - \mathrm{E}\left[f_N\right])^*(f_N - \mathrm{E}\left[f_N\right])\right].$$

In our problem, using Parseval's formula and (6), the asymptotic covariance matrix, (9), can be written as [2]

[1] This definition is slightly non-standard in that the second term is usually conjugated. For the standard definition, in general, all results have to be transposed, however, all results in this paper are symmetric.

[2] Non-singularity of $\langle\Psi, \Psi\rangle$ usually requires parameter identifiability and persistence of excitation [Ljung 1999].

$$\begin{aligned}\text{AsCov}\,\hat{\theta}_N &= \left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi(e^{j\omega})\Psi^*(e^{j\omega})\,\mathrm{d}\omega\right]^{-1}\\ &= \langle\Psi,\Psi\rangle^{-1},\end{aligned} \tag{10}$$

where

$$\Psi(q) = \frac{1}{\sigma}\tilde{\Psi}(q)\Lambda_{CH}^{-T}. \tag{11}$$

Note that $\Psi(q)$ is block upper triangular since $\tilde{\Psi}(q)$ is block diagonal and Λ_{CH}^{-T} is upper triangular.

3. MAIN RESULTS

In this section, we present expressions for the variance error of an estimated frequency response function and we also present expressions for the covariance between transfer function estimates. In the derived expressions, it is clear how the noise correlation structure, model orders and input variance affect the variance error. Before presenting our results for SIMO models, we need to introduce some concepts which will be used to understand those results.

3.1 Non-estimable part of the noise

Strong noise correlation may reduce the variance of an estimated module, as demonstrated by the introductory example. In fact, the variance error will depend on the non-estimable part of the noise, i.e., the part of the noise that cannot be linearly estimated from other noise sources. To be more specific, define the signal vector $e_{j\backslash i}(t)$ to include the noise sources from module 1 to module j, with the one from module i excluded, i.e.,

$$e_{j\backslash i}(t) := \begin{cases} [e_1(t),\ldots,e_j(t)]^T & j<i,\\ [e_1(t),\ldots,e_{i-1}(t)]^T & j=i,\\ [e_1(t),\ldots,e_{i-1}(t),e_{i+1}(t),\ldots,e_j(t)]^T & j>i.\end{cases}$$

Now, the linear minimum variance estimate of e_i given $e_{j\backslash i}(t)$, is given by

$$\hat{e}_{i|j}(t) := \varrho_{ij}^T e_{j\backslash i}(t), \tag{12}$$

where the vector ϱ_{ij} in (12) is given by

$$\varrho_{ij} = \left[\text{Cov}\, e_{j\backslash i}(t)\right]^{-1}\mathrm{E}\left[e_{j\backslash i}(t)e_i(t)\right].$$

Introduce the notation

$$\lambda_{i|j} := \text{Var}\,[e_i(t) - \hat{e}_{i|j}(t)], \tag{13}$$

with the convention that $\lambda_{i|0} := \lambda_i$. We call

$$e_i(t) - \hat{e}_{i|j}(t)$$

the non-estimable part of $e_i(t)$ given $e_{j\backslash i}(t)$.

Definition 2. When $\hat{e}_{i|j}(t)$ does not depend on $e_k(t)$, where $1 \le k \le j$, $k \ne i$, we say that $e_i(t)$ is orthogonal to $e_k(t)$ conditionally to $e_{j\backslash i}(t)$.

The variance of the non-estimable part of the noise is closely related to the Cholesky factor of the covariance matrix Λ. We have the following lemma.

Lemma 3. Let $e(t) \in \mathbb{R}^m$ have zero mean and covariance matrix $\Lambda > 0$. Let Λ_{CH} be the lower triangular Cholesky factor of Λ, i.e., Λ_{CH} satisfies (4), with $\{\gamma_{ik}\}$ as its entries as defined by (5). Then for $j < i$,

$$\lambda_{i|j} = \sum_{k=j+1}^{i}\gamma_{ik}^2.$$

Furthermore, $\gamma_{ij} = 0$ is equivalent to that $e_i(t)$ *is orthogonal to* $e_j(t)$ *conditionally to* $e_{j\backslash i}(t)$.

Proof. See Everitt et al. [2015]. ■

As a small example of why this formulation is useful, consider the covariance matrix below, where there is correlation between any pair $(e_i(t), e_j(t))$:

$$\Lambda = \begin{bmatrix}1 & 0.6 & 0.9\\ 0.6 & 1 & 0.54\\ 0.9 & 0.54 & 1\end{bmatrix} = \underbrace{\begin{bmatrix}1 & 0 & 0\\ 0.6 & 0.8 & 0\\ 0.9 & 0 & 0.44\end{bmatrix}}_{\Lambda_{CH}}\begin{bmatrix}1 & 0.6 & 0.9\\ 0 & 0.8 & 0\\ 0 & 0 & 0.44\end{bmatrix}.$$

From the Cholesky factorization above we see that, since γ_{32} is zero, Lemma 3 gives that $e_3(t)$ is orthogonal to $e_2(t)$ conditionally to $e_{2\backslash 3}(t)$, i.e., there is no information about $e_3(t)$ in $e_2(t)$ if we already know $e_1(t)$. This is not apparent from Λ where every entry is non-zero. If we know $e_1(t)$ a considerable part of $e_2(t)$ and $e_3(t)$ can be estimated. Without knowing $e_1(t)$, $\lambda_1 = \lambda_2 = \lambda_3 = 1$, while if we know $e_1(t)$, $\lambda_{2|1} = 0.64$ and $\lambda_{3|1} = 0.19$.

Similar to the above, for $i \le m$, we also define

$$e_{i:m}(t) := [e_i(t)\ \ldots\ e_m(t)]^T,$$

and for $j < i$ we define $\hat{e}_{i:m|j}(t)$ as the linear minimum variance estimate of $e_{i:m}(t)$ based on the signals in $e_{j\backslash i}(t)$, i.e.,

$$\hat{e}_{i:m|j}(t) := \left[\hat{e}_{i|j}(t)\ \ldots\ \hat{e}_{m|j}(t)\right]^T.$$

Furthermore, we define

$$\Lambda_{i:m|j} := \text{Cov}\,[e_{i:m}(t) - \hat{e}_{i:m|j}(t)].$$

3.2 Variance results for SIMO models

We are now ready to present the results on the variance error of the estimated frequency response function and for the covariance between transfer function estimates. To this end, collect all m transfer functions into

$$G := [G_1\ G_2\ \ldots\ G_m].$$

For convenience, we will simplify notation according to the following definition:

Definition 4. The asymptotic covariance of $\hat{G}(e^{j\omega_0}) := G(e^{j\omega_0}, \hat{\theta}^N)$ for the fixed frequency ω_0 is denoted by

$$\text{AsCov}\,\hat{G}.$$

In particular, the variance of $\hat{G}_i(e^{j\omega_0}) := G_i(e^{j\omega_0}, \hat{\theta}_i^N)$ for the fixed frequency ω_0 is denoted by

$$\text{AsVar}\,\hat{G}_i.$$

Define χ_k as the index of the first system model that contains the basis function $\mathcal{B}_k(e^{j\omega_0})$. Notice that, because of the ordering of the modules, $\chi_k - 1$ is the number of system models that do not contain the k-th basis function.

Let the entries of θ be arranged as follows:

$$\begin{aligned}\bar{\theta} = [&\theta_{1,1}\ \ldots\ \theta_{m,1}\ \theta_{1,2}\ \ldots\ \theta_{m,2}\ \ldots\ \theta_{1,n_1}\ \ldots\\ &\ldots\ \theta_{m,n_1}\ \theta_{2,n_1+1}\ \ldots\ \theta_{m,n_1+1}\ \ldots\ \theta_{m,n_m}]^T.\end{aligned} \tag{14}$$

and the corresponding weighted least-squares estimate be denoted by $\hat{\bar{\theta}}$.

Theorem 5. Let Assumption 1 hold. Suppose that the parameters $\theta_i \in \mathbb{R}^{n_i}$, $i = 1,\ldots,m$, are estimated using weighted least-squares (8). Then, the covariance of $\hat{\bar{\theta}}$ is

$$\begin{aligned}\text{AsCov}\,\hat{\bar{\theta}} = \frac{1}{\sigma^2}\,\text{diag}(&\Lambda_{1:m}, \Lambda_{\chi_2:m|\chi_2-1}, \ldots,\\ &\ldots, \Lambda_{\chi_{n_m}:m|\chi_{n_m}-1}).\end{aligned} \tag{15}$$

In particular, the covariance of the parameters related to the k-th basis function is given by

$$\text{AsCov}\,\hat{\bar{\theta}}_k = \frac{1}{\sigma^2}\Lambda_{\chi_k:m|\chi_k-1}, \tag{16}$$

where

$$\hat{\bar{\theta}}_k = \left[\hat{\theta}_{\chi_k,k}\ \ldots\ \hat{\theta}_{m,k}\right]^T,$$

and where, for $\chi_k \le i \le m$,

$$\text{AsVar}\,\hat{\theta}_{i,k} = \frac{\lambda_{i|\chi_k-1}}{\sigma^2}. \tag{17}$$

Proof. See Everitt et al. [2015]. ■

3.3 Interpretation of Theorem 5

It is clear that from Theorem 5 and Remark 7 that the covariance of the transfer function estimates in $\hat{G}$, is decoupled in terms of the basis functions $\mathcal{B}_k$. In this section, we will try to give some intuition why this decoupling appears. It turns out that the orthogonal basis functions introduces a decomposition of the output signals into uncorrelated components and decouples the problem. As an example, consider the system described by:

$$\begin{aligned} y_1(t) &= \theta_{1,1}\mathcal{B}_1(q)u(t) + e_1(t),\\ y_2(t) &= \theta_{2,1}\mathcal{B}_1(q)u(t) + e_2(t),\\ y_3(t) &= \theta_{3,1}\mathcal{B}_1(q)u(t) + \theta_{3,2}\mathcal{B}_2(q)u(t) + e_3(t). \end{aligned} \tag{18}$$

Suppose that we are interested in estimating $\theta_{3,2}$. For this parameter, (17) becomes

$$\text{AsVar}\,\hat{\theta}_{3,2} = \frac{\lambda_{3|2}}{\sigma^2} \tag{19}$$

To understand the mechanisms behind this expression, let $u_1(t) = \mathcal{B}_1(q)u(t)$, and $u_2(t) = \mathcal{B}_2(q)u(t)$ so that the system can be visualized as in Figure 2, i.e., we can consider u_1 and u_2 as separate inputs.

First we observe that it is only y_3 that contains information about $\theta_{3,2}$, and the term $\theta_{3,1}u_1$ contributing to y_3 is a nuisance from the perspective of estimating $\theta_{3,2}$. This term vanishes when $u_1 = 0$ and we will not be able to achieve better accuracy than the optimal estimate of $\theta_{3,2}$ for this idealized case. So let us study this setting first. Straightforward application of the least-squares method, using u_2 and y_3, gives an estimate of $\theta_{3,2}$ with variance λ/σ^2, which is larger than (19) when e_3 depends on e_1 and e_2. However, in this idealized case, $y_1 = e_1$ and $y_2 = e_2$, and these signals can thus be used to estimate e_3. This estimate can then be subtracted from y_3 before the least-squares method is applied. The remaining noise in y_3 will have variance $\lambda_{3|2}$, if e_3 is optimally estimated (see (12)–(13)), and hence the least-squares estimate will now have variance $\lambda_{3|2}/\sigma^2$, i.e., the same as (19).

To understand why it is possible to achieve the same accuracy as this idealized case when u_1 is non-zero, we need to observe that our new inputs $u_1(t)$ and $u_2(t)$ are orthogonal (uncorrelated) [3]. Returning to the case when only the output y_3 is used for estimating $\theta_{3,2}$, this implies that we pay no price for including the term $\theta_{3,1}u_1$ in our model, and then estimating $\theta_{3,1}$ and $\theta_{3,2}$ jointly, i.e., the variance of $\hat{\theta}_{3,2}$ will still be λ/σ^2 [4]. The question now is if we can use y_1 and y_2 as before to estimate e_3. Perhaps surprisingly, we can use the same estimate as when u_1 was zero. The reader may object that this estimate will now, in addition to the previous optimal estimate of e_3, contain a term which is a multiple of u_1. However, due to the orthogonality between u_1 and u_2, this term will only affect the estimate of $\theta_{3,1}$ (which we anyway were not interested in, in this example), and the accuracy of the estimate of $\theta_{3,2}$ will be $\lambda_{3|2}/\sigma^2$, i.e. (19). Figure 3 illustrates the setting with $\tilde{y}_3$ denoting y_3 subtracted by the optimal estimate of e_3. In the figure, the new parameter $\tilde{\theta}_{3,1}$ reflects that the relation between u_1 and $\tilde{y}_3$ is different from $\theta_{3,1}$ as discussed above. A key insight from this discussion is that for the estimate of a parameter in the path from input i to output j, it is only outputs that are not affected by input i that can be used to estimate the noise in output j; when this particular parameter is estimated, using outputs influenced by input i will introduce a bias, since the noise estimate will then contain a term that is not orthogonal to this input. In (17), this manifests itself in that the numerator is $\lambda_{i|\chi_k-1}$, only the $\chi_k - 1$ first systems do not contain u_i.

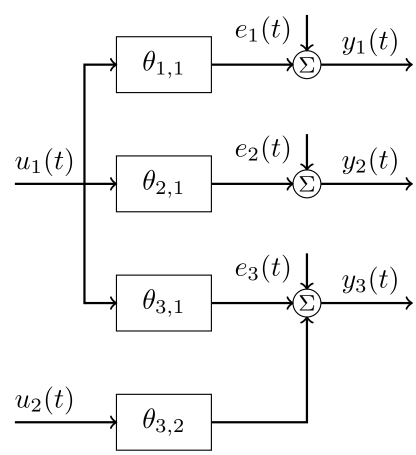

Fig. 2. The SIMO system of Remark 3.3, described by (18).

3.4 Variance of the transfer function estimates

We now turn our attention to the variance of the individual transfer function estimates.

Corollary 6. Let the same assumptions as in Theorem 5 hold. Then, for any frequency ω_0, it holds that

$$\text{AsCov}\,\hat{G} = \sum_{k=1}^{n_m} \begin{bmatrix} \mathbf{0}_{\chi_k-1} & \mathbf{0} \\ \mathbf{0} & \text{AsCov}\,\hat{\bar{\theta}}_k \end{bmatrix} |\mathcal{B}_k(e^{j\omega_0})|^2, \tag{20}$$

where $\text{AsCov}\,\hat{\bar{\theta}}_k$ is given by (16) and $\mathbf{0}_{\chi_k-1}$ is a $(\chi_k - 1)\times(\chi_k - 1)$ matrix with all entries equal to zero. For $\chi_k = 1$, $\mathbf{0}_{\chi_k-1}$ is an empty matrix. In (20), $\mathbf{0}$ denotes zero matrices of dimensions compatible to the diagonal blocks. It also holds that

$$\text{AsVar}\,\hat{G}_i = \sum_{k=1}^{n_i} |\mathcal{B}_k(e^{j\omega_0})|^2\, \text{AsVar}\,\hat{\theta}_{i,k}, \tag{21}$$

where

$$\text{AsVar}\,\hat{\theta}_{i,k} = \frac{\lambda_{i|\chi_k-1}}{\sigma^2}, \tag{22}$$

[3] This since $u(t)$ is white and $\mathcal{B}_1$ and $\mathcal{B}_2$ are orthonormal.

[4] With u_1 and u_2 correlated, the variance will be higher, see Ramazi et al. [2014], Everitt et al. [2015] for a further discussion of this topic.

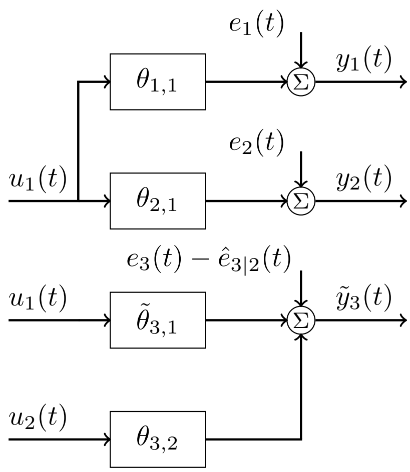

Fig. 3. The SIMO system of Remark 3.3, described by (18), but with $\tilde{y}_3$ denoting y_3 subtracted by the optimal estimate of e_3 and $\tilde{\theta}_{3,1}$ reflects that the relation between u_1 and $\tilde{y}_3$ is different from $\theta_{3,1}$.

and $\lambda_{i|j}$ is defined in (13).

Proof. See Everitt et al. [2015]. ■

Remark 7. The covariance of $\hat{\bar{\theta}}_k$, which contains the parameters related to the k-th basis function, is determined by which other models share the basis function $\mathcal{B}_k$. The asymptotic covariance of $\hat{G}$ can be understood as a sum of the contributions from each of the n_m basis functions. The covariance contribution from a basis function $\mathcal{B}_k$ is weighted by $|\mathcal{B}_k(e^{j\omega_0})|^2$ and only affects the covariance between systems that contain that basis function, as visualized in Figure 4.

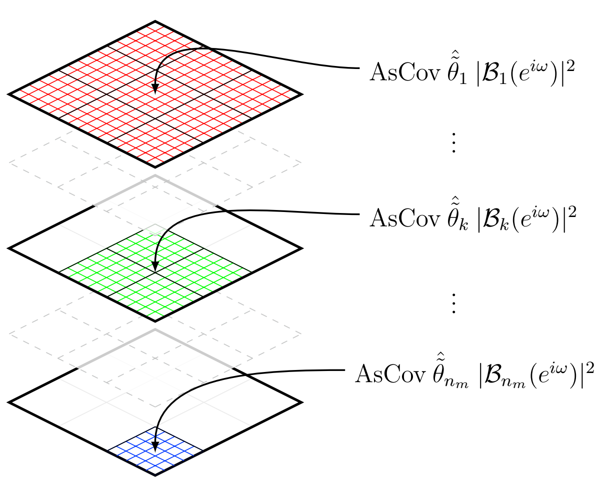

Fig. 4. A graphical representation of AsCov $\hat{G}$ where each term of the sum in (20) is represented by a layer. A basis function only affects the covariance between modules that also contain that basis function. Thus, the first basis function affects the complete covariance matrix while the last basis function n_m only affects modules $\chi_{n_m}, \ldots, m$.

From Corollary 6, we can tell when increasing the model order of G_j will increase the asymptotic variance of $\hat{G}_i$.

Corollary 8. Under the same conditions as in Theorem 5, if we increase the number of estimated parameters of G_j from n_j to $n_j + 1$, the asymptotic variance of G_i will increase, if and only if all the following conditions hold:

(1) $n_j < n_i$,
(2) $e_i(t)$ is not orthogonal to $e_j(t)$ conditionally to $e_{j\backslash i}(t)$,
(3) $|\mathcal{B}_{n_{j+1}}(e^{j\omega_0})|^2 \neq 0$.

Proof. See Everitt et al. [2015]. ■

Remark 9. Corollary 8 explicitly tells when an increased in the model order of G_j from n_j to n_j+1 will increase the variance of G_i. If $n_j \geq n_i$, there will be no increase in the variance of G_i, no matter how many additional parameters we introduce to the model G_j because G_i do not contain any of the basis functions introduced. Naturally, if $e_i(t)$ is orthogonal to $e_j(t)$ conditionally to $e_{j\backslash i}(t)$, $\hat{e}_{i|j}(t)$ does not depend on $e_j(t)$ and there is no increase in variance of $\hat{G}_i$. Similarly, if the basis function $\mathcal{B}_{n_j+1}(e^{j\omega_0})$ is zero, while the variance of the corresponding parameter $\hat{\theta}_{i,n_j+1}$ increases, the variance of $\hat{G}_i$ remain the same.

4. NUMERICAL EXAMPLES

In this section, the effect of Corollary 8 is illustrated in Figure 5, where the following systems are identified using $N = 500$ input-output measurements:

$$G_i = \tilde{\Gamma}_i\theta_i, \qquad \tilde{\Gamma}_i(q) = F(q)^{-1}\Gamma_i(q),$$
$$\Gamma_i(q) = [\mathcal{B}_1(q), \ldots, \mathcal{B}_{n_i}(q)], \qquad \mathcal{B}_k(q) = q^{-k}, \tag{23}$$

for $i = 1, 2, 3$ with

$$F(q) = \frac{1}{1 - 0.8q^{-1}}, \qquad \theta_1^0 = [1\ 0.5\ 0.7]^T,$$
$$\theta_2^0 = [1\ -1\ 2]^T, \qquad \theta_3^0 = [1\ 1\ 2\ 1\ 1]^T.$$

The input $u(t)$ is drawn from a Gaussian distribution with variance $\sigma^2 = 1$, filtered by $F(q)$. The measurement noise is normally distributed with covariance matrix $\Lambda = \Lambda_{CH}\Lambda_{CH}^T$, where

$$\Lambda_{CH} = \begin{bmatrix} 1 & 0 & 0 \\ 0.6 & 0.8 & 0 \\ 0.7 & 0.7 & 0.1 \end{bmatrix},$$

thus $\lambda_1 = \lambda_2 = \lambda_3 = 1$. The sample variance is computed using

$$\text{Cov } \hat{\theta}_s = \frac{1}{MC}\sum_{k=1}^{MC} |G_3(e^{j\omega_0}, \theta_3^0) - G_3(e^{j\omega_0}, \hat{\theta}_3)|^2,$$

where $MC = 2000$ is the number of realizations of the input and noise. The same realizations of the input and noise are used for all model orders.

The variance of $G_3(e^{j\omega}, \hat{\theta}_3)$ increases with increasing n_i, $i = 1, 2$, but only up to the point where $n_i = n_3 = 5$. After that, any increase in n_1 or n_2 does not increase the variance of $G_3(e^{j\omega}, \hat{\theta}_3)$, as can be seen in Figure 5. The behavior can be explained by Corollary 8: when $n_3 \geq n_1, n_2$, G_3 is the last module, having the highest number of parameters, and any increase in n_1, n_2 increases the variance of G_3. When for example $n_1 \geq n_3$, the modules should be reordered so that G_3 comes before G_1. In this case, when n_1 increases the first condition of Corollary 8 does not hold and hence the variance of $G_3(e^{j\omega}, \hat{\theta}_3)$ does not increase further.

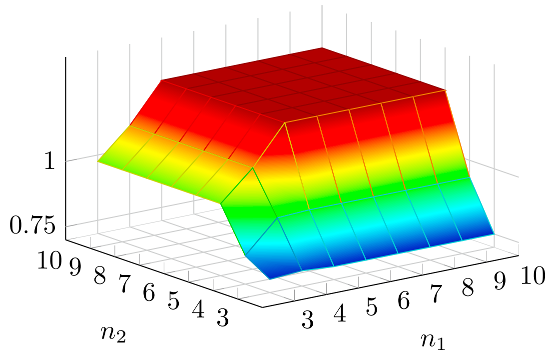

Fig. 5. Sample variance of $G_3(e^{j\omega}, \hat{\theta}_3)$ as a function of the number of estimated parameters of G_1 and G_2.

5. CONCLUSIONS

In this paper, we have investigated how the accuracy of an identified linear SIMO model depends on the correlation structure of the noise and model structure and model order. We have quantified the asymptotic covariance of the frequency response function estimate and the model parameters, for a linear-in-the-parameter model structure, in the case of temporally white, but possibly spatially correlated additive noise. Modules estimated with less parameters lead to a reduction of the variance of other modules since parts of the noise can be linearly estimated from measurement of the first modules. It is shown that the order of the different modules and the noise correlation affect the variance of one module. In particular, the variance of the module of interest levels off when the number of estimated parameters in another module reaches the number of estimated parameters of the module of interest. We have illustrated this aspect with numerical simulations.

REFERENCES

Ali, M., Ali, A., Chughtai, S.S., and Werner, H. (2011). Consistent identification of spatially interconnected systems. In *Proceedings of the 2011 American Control Conference*, 3583–3588.

Bottegal, G. and Hjalmarsson, H. (2014). On the variance of identified SIMO systems with spatially correlated output noise. In *Proceedings of the 53nd IEEE Conference on Decision and Control*.

Chiuso, A. and Pillonetto, G. (2012). A Bayesian approach to sparse dynamic network identification. *Automatica*, 48(8), 1553 – 1565.

Cox, D.R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(1), pp. 1–39.

Dankers, A., Van den Hof, P.M.J., Bombois, X., and Heuberger, P.S.C. (2013a). Predictor input selection for two stage identification in dynamic networks. In *Proceeding of the 2013 European Control Conference*, 1422–1427.

Dankers, A., Van den Hof, P.M.J., Bombois, X., and Heuberger, P.S.C. (2014). Errors-in-variables identification in dynamic networks. In *Proceedings of the 19th IFAC World Congress*.

Dankers, A., Van den Hof, P.M.J., and Heuberger, P.S.C. (2013b). Predictor input selection for direct identification in dynamic networks. In *Proceeding of the 52nd IEEE Conference on Decision and Control*, 4541–4546.

Everitt, N., Bottegal, G., Rojas, C.R., and Hjalmarsson, H. (2015). Variance analysis of linear SIMO models with spatially correlated noise. *Submitted to Automatica arXiv preprint arXiv:1501.03077*.

Everitt, N., Rojas, C.R., and Hjalmarsson, H. (2014). Variance results for parallel cascade serial systems. In *Proceedings of the 18th IFAC World Congress*.

Gevers, M., Mišković, L., Bonvin, D., and Karimi, A. (2006). Identification of multi-input systems: variance analysis and input design issues. *Automatica*, 42(4), 559 – 572.

Gunes, B., Dankers, A., and Van den Hof, P.M.J. (2014). A variance reduction technique for identification in dynamic networks. In *Proceedings of the 19th IFAC World Congress*.

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

Hägg, P., Wahlberg, B., and Sandberg, H. (2011). On identification of parallel cascade serial systems. In *Proceedings of the 18th IFAC World Congress*.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall, 2nd edition.

Materassi, D., Salapaka, M.V., and Giarre, L. (2011). Relations between structure and estimators in networks of dynamical systems. In *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, 162–167.

Ramazi, P., Hjalmarsson, H., and Mårtensson, J. (2014). Variance analysis of identified linear MISO models having spatially correlated inputs, with application to parallel Hammerstein models. *Automatica*, 50(6), 1675 – 1683.

Söderström, T. and Stoica, P. (1989). *System identification*. Prentice-Hall, Englewood Cliffs.

Torres, P., van Wingerden, J.W., and Verhaegen, M. (2014). Output-error identification of large scale 1d-spatially varying interconnected systems. *IEEE Transactions on Automatic Control*, 60(1), 130–142.

Van den Hof, P.M.J., Dankers, A., 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 Mårtensson, J. (2009). Variance results for identification of cascade systems. *Automatica*, 45(6), 1443–1448.