On the variance analysis of identified linear MIMO models

Conference Paper · December 2015

DOI: 10.1109/CDC.2015.7402414

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

On the Variance Analysis of identified Linear MIMO Models

Niklas Everitt, Giulio Bottegal, Cristian R. Rojas and Håkan Hjalmarsson

***Abstract*— We study the accuracy of identified linear time-invariant multi-input multi-output (MIMO) systems. Under a stochastic framework, we quantify the effect of the spatial correlation and choice of model structure on the covariance matrix of the transfer function estimates. In particular, it is shown how the variance of a transfer function estimate depends on signal properties and model orders of other modules composing the MIMO system.**

***Index Terms*— System identification, Asymptotic variance, Linear MIMO models, Least-squares.**

I. Introduction

Modern control systems for industrial plants are based on mathematical models of the plant dynamics. These control systems need to handle several decision variables (input signals), having access, through sensing devices, to a possibly large number of measured variables (output signals). The whole structure can be modeled as a multi-input multi-output (MIMO) system and its identification is a crucial task. In particular, assessing the quality of the identified model by quantifying model uncertainties is an important aspect that must be taken into account when designing model-based and robust control schemes [1].

We focus on quantifying these model uncertainties in the identification of MIMO systems, where each system module is expressed as the linear combination of (known) basis functions and where the linear coefficients are unknown. Adopting a stochastic framework, we assume that measurements are corrupted by additive noise, with a given probabilistic description.

We assume the system is in the model set and study the accuracy of the identified model in terms of the parameter error covariance matrix [2]. Our aim is to understand how the experimental conditions and model structure influence this covariance matrix. To this end, we simplify our analysis by assuming that the input signals are temporally white, but may be spatially correlated. We make a similar assumption for the output noise. Under these conditions, we derive an insightful expression for the parameter error covariance matrix. In particular, we characterize the behavior of this covariance matrix in terms of the input and noise correlation matrices and the model orders of the modules composing the MIMO system. Our results show that the combination of suitable input and noise correlation structures and proper model orders of the modules, may significantly improve the accuracy of the estimated model for a specific module. This has important implications in experiment design for MIMO systems [3] and fault-detection and diagnosis [4].

The authors are with the Automatic Control Lab and ACCESS Linnaeus Center, School of Electrical Engineering, KTH–Royal Institute of Technology, SE-100 44 Stockholm, Sweden. (e-mail: {neveritt, bottegal, crro, hjalmars}@kth.se). This work was supported in part by the European Research Council under the advanced grant LEARN, contract 267381, and in part by the Swedish Research Council under contract 621-2009-4017.

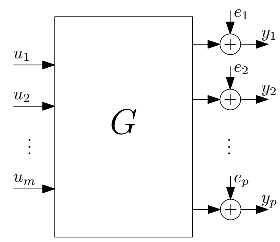

Fig. 1. *Block scheme of the linear MIMO system.*

In the literature, there are readily available formulas quantifying the model error of identified MIMO systems. However they provide little insight in what affects the model error. There are other expressions that try to give this insight, however, they are typically valid only asymptotically (in both the number of samples and model orders of the modules). One classic result, valid for large data length N and large model order n, is given by the following expression [6], [7]

$$\operatorname{Cov}\Big[\operatorname{vec}\hat{G}\Big] \approx \frac{n}{N}\Phi_u^{-1}\otimes\Phi_v\,,$$

where Φ_u is the input spectrum and Φ_v the noise spectrum. This expression was extended in [8] to a general set of orthonormal basis functions. , denoted by $\{\mathcal{B}_k(z)\}$:

$$\operatorname{Cov}\Big[\operatorname{vec}\hat{G}\Big] \approx \frac{1}{N}\sum_{k=1}^{n_{\mathcal{B}}}|\mathcal{B}_k(z)|^2\Phi_u^{-1}\otimes\Phi_v.$$

Note that both expressions are valid only for large model orders. Our results, instead, are exact for finite model orders (but still asymptotic in data length). Furthermore, we hereby generalize variance results for multi-input single-output (MISO) systems [9] and single-input multi-output (SIMO) systems [10]. The significance of our results will be clarified through simulation examples.

In Section II we give the problem formulation. Some technical lemmas are formulated in Section III. In Section IV we give the developed formulas of the model accuracy. In Section V we illustrate the developed results on a 2-by-2 system. The paper ends with some conclusions in Section VI.

II. Problem Statement

We consider linear time-invariant dynamic systems with m inputs and p outputs (see Fig. 1). Let us introduce the vector notation

$$y(t) := \begin{bmatrix} y_1(t) & y_2(t) & \dots & y_p(t)\end{bmatrix}^T$$

and the input $u(t)$ and noise $e(t)$ are defined in the same manner. The model is described as follows:

$$y(t) = G(q)u(t) + e(t)$$

where

$$G(q) = \begin{bmatrix} G_{11}(q) & G_{12}(q) & \dots & G_{1m}(q) \\ G_{21}(q) & G_{22}(q) & \dots & G_{2m}(q) \\ \vdots & \vdots & \ddots & \vdots \\ G_{p1}(q) & G_{p2}(q) & \dots & G_{pm}(q) \end{bmatrix}.$$

Here q denotes the forward shift operator, i.e., $qu(t) = u(t+1)$ and the $G_{ij}(q)$ are causal stable rational transfer functions. The modules G_{ij} are modeled as

$$G_{ij}(q, \theta_{ij}) = \Gamma_{ij}(q)\theta_{ij}, \quad \theta_{ij} \in \mathbb{R}^{n_{ij}}, \tag{1}$$

where $\Gamma_{ij}(q) = [\mathcal{B}_1(q), \dots, \mathcal{B}_{n_{ij}}(q)]$, $i = 1, \dots, p$ and $j = 1, \dots, m$, for some orthonormal basis functions $\{\mathcal{B}_k(q)\}_{k=1}^{n_{\mathcal{B}}}$. The basis functions can be tailored according to a priori knowledge on the system dynamics; for example, Laguerre basis functions can be tailored to the time constant of the system [11]. Examples of more general basis functions can be found in [12], [13]. 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 $\langle f, g\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\omega})g^*(e^{i\omega})\,\mathrm{d}\omega$. The noise sequence $\{e(t)\}$ is zero mean and temporally white, but may be correlated in the spatial domain:

$$\mathrm{E}\left[e(t)\right] = 0, \quad \mathrm{E}\left[e(t)e(s)^T\right] = \delta_{t-s}\Lambda, \tag{2}$$

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

$$\Lambda = \Lambda_{CH}\Lambda_{CH}^T, \tag{3}$$

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

$$\Lambda_{CH} = \begin{bmatrix} \rho_{11} & 0 & \dots & 0 \\ \rho_{21} & \rho_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & \rho_{pp} \end{bmatrix}, \tag{4}$$

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

$$\Lambda^{-1} = \Lambda_{CH}^{-T}\Lambda_{CH}^{-1}. \tag{5}$$

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

$$\mathrm{E}\left[u(t)\right] = 0, \quad \mathrm{E}\left[u(t)u(s)^T\right] = \delta_{t-s}\Sigma, \tag{6}$$

for some positive definite matrix covariance matrix Σ. We express Σ in terms of its *upper* triangular Cholesky factorization[1]

$$\Sigma = \Sigma_{CH}\Sigma_{CH}^T. \tag{7}$$

[1]The standard Cholesky factorization is defined with a lower triangular factor but for reasons that will be clear in following sections, we need the upper triangular counterpart.

Also Σ_{CH}^{-1} is upper triangular and we denote its elements by $\{\gamma_{ij}\}$ according to

$$\Sigma_{CH}^{-1} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1m} \\ 0 & \gamma_{22} & \dots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \gamma_{mm} \end{bmatrix}. \tag{8}$$

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

Assumption 2.1: The input $\{u(t)\}$ is zero mean temporally white noise with finite moments of all orders, and (6) holds with $\Sigma > 0$. The noise $\{e(t)\}$ is zero mean and temporally white, i.e, (2) holds with $\Lambda > 0$. It is assumed that $\mathrm{E}\left[|e(t)|^{4+\rho}\right] < \infty$ for some $\rho > 0$. The data are 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 (1) for some parameters $\theta_{ij}^o \in \mathbb{R}^{n_{ij}}$, $i = 1, \dots, p$, $j = 1, \dots, m$. The orthonormal basis functions $\{\mathcal{B}_k(q)\}$ are assumed stable.

A. Weighted least-squares estimate

By introducing

$$\theta = \left[\theta_{11}^T, \theta_{21}^T, \dots, \theta_{p1}^T, \dots, \theta_{1m}^T, \theta_{2m}^T, \dots, \theta_{pm}^T\right]^T \in \mathbb{R}^n,$$

$n := \sum_{i=1}^{p}\sum_{j=1}^{m} n_{ij}$, the transfer function matrix

$$\underset{(n\times pm)}{\tilde{\Psi}(q)} = \mathrm{diag}\{\Gamma_{11}^T, \Gamma_{21}^T, \dots, \Gamma_{pm}^T\}$$

and the notation

$$\underset{(pm\times 1)}{\mathrm{vec}\, G(q)} = \begin{bmatrix} G_{.1}(q) & G_{.2}(q) & \dots & G_{.m}(q) \end{bmatrix}^T,$$

we can write $\mathrm{vec}\, G(q) = \tilde{\Psi}(q)^T\theta$ and describe the model (1) as a linear regression model

$$y(t) = \varphi^T(t)\theta + e(t), \tag{9}$$

where, using Theorem T2.13 in [14],

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

An unbiased and consistent 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., [2], [15]). 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 $\hat{\theta}$ [16]. However, in certain applications, not knowing Λ will increase the covariance of $\hat{\theta}$. Under Assumption 2.1, the noise sequence is zero mean, hence $\hat{\theta}_N$ is unbiased. It also follows that the asymptotic covariance matrix of the parameter estimates is given by

$$\mathrm{AsCov}\,\hat{\theta}_N = \mathrm{E}\left[\varphi(t)\Lambda^{-1}\varphi^T(t)\right]^{-1}, \tag{10}$$

where the expectation is over the input sequence. By repeated use of the mixed product rule of Kronecker products (T2.4 in [14]) (10) can be expressed as

$$\mathrm{AsCov}\,\hat{\theta}_N = \mathrm{E}\left[\tilde{\Psi}(q)\left\{u(t)u^T(t) \otimes \Lambda^{-1}\right\}\tilde{\Psi}(q)^T\right]^{-1}. \tag{11}$$

Here 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[2]

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

In the problem we consider, using Parseval's formula, (3) and (5), the asymptotic covariance matrix, (11), can be written as[3]

$$\text{AsCov } \hat{\theta}_N = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \Psi(e^{j\omega})\Psi^*(e^{j\omega})\,\mathrm{d}\omega\right]^{-1} = \langle \Psi, \Psi\rangle^{-1}, \quad (12)$$

where $\Psi(q) = \tilde{\Psi}(q)(\Sigma_{CH} \otimes \Lambda_{CH}^{-T})$. Note that $\Psi(q)$ is block upper triangular since $\tilde{\Psi}(q)$ is block diagonal, Λ_{CH}^{-T} is upper triangular and Σ_{CH} is upper triangular[4]. The rest of this contribution tries to analyze how model structure, input and noise properties affect (12).

III. Geometric variance expression

The following lemma is included for completeness.

Lemma 3.1: (Lemma II.9 in [17]) Let $J : \mathbb{R}^n \to \mathbb{C}^{1\times q}$ be differentiable with respect to θ, and $\Psi \in \mathcal{L}_2^{n\times m}$; let $\mathcal{S}_\Psi$ be the subspace of $\mathcal{L}_2^{1\times m}$ spanned by the rows of Ψ and $\{\mathcal{B}_k^{\mathcal{S}}\}_{k=1}^{r}, r \le n$ be an orthonormal basis for $\mathcal{S}_\Psi$. Suppose that $J'(\theta^o) \in \mathbb{C}^{n\times q}$ is the gradient of J with respect to θ and $J'(\theta^o) = \Psi(z_o)L$ for some $z_0 \in \mathbb{C}$ and $L \in \mathbb{C}^{m\times q}$. Then

$$\text{AsCov } J(\hat{\theta}_N) = L^* \sum_{k=1}^{r} \mathcal{B}_k^{\mathcal{S}}(z_o)^* \mathcal{B}_k^{\mathcal{S}}(z_o)\, L. \quad (13)$$

To make use of the aforementioned lemma, we notice that the asymptotic variance is given by (12) with $\Psi(q) = \tilde{\Psi}(q)(\Sigma_{CH} \otimes \Lambda_{CH}^{-T})$. Note that

$$\frac{\partial \operatorname{vec} G(q)}{\partial \theta} = \Psi(\Sigma_{CH}^{-1} \otimes \Lambda_{CH}^{T}).$$

Introduce $L := (\Sigma_{CH}^{-1} \otimes \Lambda_{CH}^{T})$. Then, using Lemma 3.1,

$$\text{AsCov}\Big[\operatorname{vec} \hat{G}(e^{j\omega_0})\Big] = L^T \sum_{k=1}^{n} \mathcal{B}_k^{\mathcal{S}}(e^{j\omega_0})^* \mathcal{B}_k^{\mathcal{S}}(e^{j\omega_0}) L. \quad (14)$$

To analyze (14), we will consider a few special cases. To this end, we need to characterize the basis functions $\{\mathcal{B}_k^{\mathcal{S}}\}$.

A. Basis functions

We need the following assumption on the model structure:

Assumption 3.1: For each row i, if G_{ij} contains $\mathcal{B}_k$, then $G_{i(j+1)}$ contains $\mathcal{B}_k$.

Assumption 3.1 states that the modules that do not contain $\mathcal{B}_k$ are located in the upper left corner of G, which is somewhat restrictive since this may not be achieved by renaming the modules in all cases. The rows of Ψ that contain $\mathcal{B}_k$ are given by

$$\mathcal{B}_k Q_k (\Sigma_{CH} \otimes \Lambda_{CH}^{-T}). \quad (15)$$

We will also need the following definition:

Definition 3.1: For each basis function $\mathcal{B}_k$, let

$$Q_k := \operatorname{diag}(q_1, \ldots, q_{mp}),$$

where

$$q_i := \begin{cases} 1 & \text{if entry i of vec G(q) does contain } \mathcal{B}_\text{k} \\ 0 & \text{if entry i of vec G(q) does not contain } \mathcal{B}_\text{k} \end{cases}$$

Lemma 3.2: Assume Assumption 3.1 holds and let $\mathcal{S}_\Psi$ be the rowspace of Ψ. Let $\{\mathcal{B}_l^{\mathcal{S}}(q)\}_{l=1}^{n}$ be such that for every basis function $\mathcal{B}_k$ and every row i of $\operatorname{vec} G$ that has the basis function $\mathcal{B}_k$, there is a basis function of the form

$$\mathcal{B}_l^{\mathcal{S}}(q) := \begin{bmatrix} 0 & \ldots & 0 & \mathcal{B}_k(q) & 0 & \ldots & 0 \end{bmatrix},$$

i.e., the entry in column i of $\mathcal{B}_l^{\mathcal{S}}(q)$ is $\mathcal{B}_k$. Then $\{\mathcal{B}_l^{\mathcal{S}}(q)\}_{l=1}^{n}$ is a set of basis functions for $\mathcal{S}_\Psi$.

Proof: Available upon request to the authors. ∎

Example 3.1: To illustrate Lemma 3.2, we consider a system with 2 inputs and 2 outputs. When all modules contain $\mathcal{B}_k$, Assumption 3.1 is satisfied, and (15) gives

$$\mathcal{B}_k \begin{bmatrix} \gamma_{11}\rho_{11} & \gamma_{11}\rho_{21} & \gamma_{12}\rho_{11} & \gamma_{12}\rho_{21} \\ 0 & \gamma_{11}\rho_{22} & 0 & \gamma_{12}\rho_{22} \\ 0 & 0 & \gamma_{22}\rho_{11} & \gamma_{22}\rho_{21} \\ 0 & 0 & 0 & \gamma_{22}\rho_{22} \end{bmatrix},$$

and since the matrix is full rank, a set of basis functions can be given by Lemma 3.2. We now look at which non-full parametrizations comply with Assumption 3.1. Assumption 3.1 tells us that the modules in G that contain $\mathcal{B}_k$ should be located in the lower right corner. Naturally, if only G_{22} contains $\mathcal{B}_k$, then (15) is given by

$$\mathcal{B}_k \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_{22}\rho_{22} \end{bmatrix},$$

and the only basis function is $\mathcal{B}_1^{\mathcal{S}}(q) = [0, 0, 0, \mathcal{B}_k(q)]$. If Assumption 3.1 still is to hold, we are allowed to add $\mathcal{B}_k$ to either G_{12} or G_{21}. In the first case, (15) is given by

$$\mathcal{B}_k \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{22}\rho_{11} & \gamma_{22}\rho_{21} \\ 0 & 0 & 0 & \gamma_{22}\rho_{22} \end{bmatrix},$$

and the second basis function is $\mathcal{B}_2^{\mathcal{S}}(q) = [0, 0, \mathcal{B}_k(q), 0]$. If instead $\mathcal{B}_k$ is added to G_{21}, (15) is given by

$$\mathcal{B}_k \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \gamma_{11}\rho_{22} & 0 & \gamma_{12}\rho_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma_{22}\rho_{22} \end{bmatrix},$$

and the second basis function is $\mathcal{B}_2^{\mathcal{S}}(q) = [0, \mathcal{B}_k(q), 0, 0]$. To illustrate what happens when Assumption 3.1 is not satisfied,

[2]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, nevertheless, all results in this paper are symmetric.

[3]Non-singularity of $\langle \Psi, \Psi\rangle$ usually requires parameter identifiability and persistence of excitation [2], both of which are satisfied under Assumption 2.1.

[4]The reason for introducing the upper triangular factorization Σ_{CH} is to obtain this block upper triangular structure.

consider the case when G_{12} does not contain $\mathcal{B}_k$ and the rest of the modules do. Then, (15) is given by

$$\mathcal{B}_k \begin{bmatrix} \gamma_{11}\rho_{11} & \gamma_{11}\rho_{21} & \gamma_{12}\rho_{11} & \gamma_{12}\rho_{21} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma_{22}\rho_{11} & \gamma_{22}\rho_{21} \\ 0 & 0 & 0 & \gamma_{22}\rho_{22} \end{bmatrix},$$

and the set of basis functions are not on the required form:

$$\mathcal{B}_1^{\mathcal{S}}(q) = \begin{bmatrix} 0 & 0 & 0 & \mathcal{B}_k, \end{bmatrix}, \quad \mathcal{B}_2^{\mathcal{S}}(q) = \begin{bmatrix} 0 & 0 & \mathcal{B}_k & 0 \end{bmatrix},$$
$$\mathcal{B}_3^{\mathcal{S}}(q) = \mathcal{B}_k(\rho_{11}^2 + \rho_{21}^2)^{-1/2} \begin{bmatrix} \rho_{11} & \rho_{21} & 0 & 0 \end{bmatrix}.$$

IV. Main results

The configuration of basis function prescribed by Lemma 3.2 allows us to derive a neat expression for the asymptotic covariance matrix in (14). which in turn will be exemplified for a few special structures.

Theorem 4.1: Let Assumptions 2.1 and 3.1 hold. Then

$$\operatorname{AsCov}\Big[\operatorname{vec}\hat{G}(e^{j\omega_0})\Big] = \sum_{k=1}^{n_{\mathcal{B}}} |\mathcal{B}_k(e^{j\omega_0})|^2 L^T Q_k L. \tag{16}$$

where Q_k is given by Definition 3.1 and

$$L := \Sigma_{CH}^{-1} \otimes \Lambda_{CH}^T.$$

Proof: Follows from (14) using Lemma 3.2. ■

Remark 4.1: The formula (16) is a decomposition of the variance of $\operatorname{vec}\hat{G}(q)$ as the sum of the contributions of each module basis function $\mathcal{B}_k$. The weighting factor L^TQ_kL is determined by which other modules also contain $\mathcal{B}_k$. In fact, it is possible to associate L^TQ_kL as the covariance of the parameters related to the basis function $\mathcal{B}_k$ similar to the results found in [10]. This has been left out due to space limitations.

Remark 4.2: If Assumption 3.1 does not hold, (16) still holds, but, in this case Q_k in (16) must be replaced by P^TQ_kP, where P is an orthogonal matrix such that the rowspace of $\mathcal{B}_kQ_kP$ is the same as the rowspace of $\mathcal{B}_kQ_kL$.

We will now consider a few special cases where we can give some insights into the expression (14). We will analyze the weighting factor

$$L^TQ_kL \tag{17}$$

that depends on the other modules that contain $\mathcal{B}_k$.

A. Full parametrization

We first consider the case when all modules G_{ij} contain the same set of basis functions, i.e., the model orders are identical: $n_{11} = n_{21} = \ldots = n_{pm}$. Naturally, in this case Assumption 3.1 holds and Theorem 4.1 gives that (16) holds with $Q_k = I_{mp}$ which inserted in in (17) gives

$$L^TQ_kL = \Sigma^{-1} \otimes \Lambda. \tag{18}$$

Hence, we have the following corollary:

Corollary 4.1: Let Assumption 2.1 hold. Let all modules have the same model order. Then

$$\operatorname{AsCov}\Big[\operatorname{vec}\hat{G}\Big] = \sum_{k=1}^{n_{\mathcal{B}}} |\mathcal{B}_k|^2 (\Sigma^{-1} \otimes \Lambda). \tag{19}$$

Remark 4.3: This is a generalization of (1) derived in [8], to finite model orders, when there is no temporal correlation in neither input nor noise.

B. Unused input

We consider next the case when we do not estimate parameters related to basis function $\mathcal{B}_k$ for modules with input $\{u_i\}_{i=1}^{\tau_k-1}$, i.e., modules $G_{11}, G_{21}, \ldots, G_{p(\tau_k-1)}$ do not contain $\mathcal{B}_k$. Also in this case Assumption 3.1 holds and Theorem 4.1 gives that (16) holds with

$$Q_k = \begin{bmatrix} 0_{p(\tau_k-1)} & 0 \\ 0 & I_{m-\tau_k-1} \end{bmatrix} \otimes I_p, \tag{20}$$

where $0_{p(\tau_k-1)}$ is a matrix of zeros of dimension $p(\tau_k-1) \times p(\tau_k-1)$, and $I_{m-\tau_k-1}$ is an identity matrix of dimension $(m-\tau_k-1) \times (m-\tau_k-1)$. Inserting (20) in (17) and using Lemma 1.2 gives

$$L^TQ_kL = \begin{bmatrix} 0_{p(\tau_k-1)} & 0 \\ 0 & \Sigma_{\tau_k:m}^{-1} \end{bmatrix} \otimes \Lambda, \tag{21}$$

where $\Sigma_{\tau_k:m}^{-1}$ is the inverse of the covariance matrix for the inputs $[u_\tau(t), \ldots, u_m(t)]^T$. Inserting (21) in Theorem 4.1 leads to the following corollary:

Corollary 4.2: Let Assumption 2.1 hold. Assume that for each basis function $\mathcal{B}_k$, there is a τ_k such that no module with input $\{u_k\}_{k=1}^{\tau_k-1}$ contains $\mathcal{B}_k$. Then

$$\operatorname{AsCov}\Big[\operatorname{vec}\hat{G}\Big] = \sum_{k=1}^{n_{\mathcal{B}}} |\mathcal{B}_k|^2 \begin{bmatrix} 0_{p(\tau_k-1)} & 0 \\ 0 & \Sigma_{\tau_k:m}^{-1} \end{bmatrix} \otimes \Lambda. \tag{22}$$

Remark 4.4: If there is only one output, $p = 1$, we recover the MISO case of Theorem 4 of [9], i.e., the main diagonal of (22) corresponds to the module variance results of Theorem 4 of [9].

Remark 4.5: Depending on the input correlation structure, the contribution from a basis function to the variance of a parameter may be greatly reduced, given that the basis function is not present in modules with inputs strongly correlated to the input related to the parameter. Define $\sigma_{i|l:o}^2$ as the variance of $u_i(t)$ conditioned on $[u_l(t), \ldots, u_o(t)]^T$ (see [9]). From (22), Lemma 1.1 in the Appendix and Lemma 1.2, it can be seen that the variance of a module depends on the inverse of the variance of the input to that module, conditioned on the inputs of modules that contain the basis function. To give an example, consider the lower right element of $\Sigma_{\tau_k:m}^{-1}$, which is given by $1/\sigma_{m|\tau_k:m-1}^2$. If there is information about $u_m(t)$ in $\{u_1, \ldots, u_{\tau_{k-1}}\}$ (not found in $\{u_{\tau_k}, \ldots, u_{m-1}\}$), then $\sigma_{m|\tau_k:m-1}^2 > \sigma_{m|1:m-1}^2$ and the variance of $\hat{G}_{mp}$ will be lower.

C. Unused output

We now consider the case when we do not estimate a parameter for basis function $\mathcal{B}_k$ for all modules that affect output $\{y_i(t)\}_{i=1}^{\tau_k-1}$, i.e., modules $G_{11}, G_{12}, \ldots, G_{(\tau_k-1)m}$ do not contain $\mathcal{B}_k$. Also in this case Assumption 3.1 holds and Theorem 4.1 gives that (16) holds with

$$Q_k = I_m \otimes \begin{bmatrix} 0_{p(\tau_k-1)} & 0 \\ 0 & I_{p-\tau_k-1} \end{bmatrix}, \tag{23}$$

where $0_{p(\tau_k-1)}$ is a matrix of zeros of dimension $p(\tau_k-1) \times p(\tau_k-1)$, and $I_{p-\tau_k-1}$ is an identity matrix of dimension

$(p-\tau_k-1)\times(p-\tau_k-1)$. Inserting (23) in (17) and using Lemma 5 of [10] we obtain

$$L^T Q_k L = \Sigma^{-1} \otimes \begin{bmatrix} 0_{p(\tau_k-1)} & 0 \\ 0 & \Lambda_{\tau_k:m|\tau_k-1} \end{bmatrix}, \qquad (24)$$

where $\Lambda_{\tau_k:m|\tau_k-1}$ is the covariance of $[e_\tau(t),\ldots,e_m(t)]^T$ given the other noise sources $[e_1(t),\ldots,e_{\tau-1}(t)]^T$. In this case, it follows from Theorem 4.1 that the following corollary holds:

Corollary 4.3: Let Assumption 2.1 hold. Assume that for each basis function $\mathcal{B}_k$, there is a τ_k such that no module that affect output $\{y_k(t)\}_{k=1}^{\tau_k-1}$ contains $\mathcal{B}_k$. Then

$$\operatorname{AsCov}\Big[\operatorname{vec}\hat{G}\Big] = \sum_{k=1}^{n_{\mathcal{B}}} |\mathcal{B}_k|^2 \Sigma^{-1} \otimes \begin{bmatrix} 0_{p(\tau_k-1)} & 0 \\ 0 & \Lambda_{\tau_k:m|\tau_k-1} \end{bmatrix}. \qquad (25)$$

Remark 4.6: Result (25) is a generalization of the SIMO result of Theorem 7 in [10] in the sense that if there is only one input, $m=1$, we recover Theorem 7 in [10].

Remark 4.7: Depending on the noise correlation structure, the parameter variance contribution from a basis function may be greatly reduced, given that the basis function is not present in modules with outputs affected by noise strongly correlated to the noise affecting the output related to the parameter. Define λ_i as the covariance of $e_i(t)$ and $\lambda_{i|\tau_k-1}$ as the covariance of $e_i(t)$ conditioned on the noise sources $[e_1(t),\ldots,e_{\tau_k-1}(t)]^T$. Consider the lower right element of $\Lambda_{\tau_k:m|\tau_k-1}$, which is given by $\lambda_{m|\tau_k-1}$. If there is information about $e_m(t)$ in $\{e_1,\ldots,e_{\tau_{k-1}}\}$, then $\lambda_{m|\tau_k-1} < \lambda_m$ and the variance of $\hat{G}_{mp}$ will be lower.

V. Numerical examples

A. First Experiment

In this section we illustrate the developed results on a system with 2 inputs and 2 outputs. To better illustrate the results, the parameter variances will be studied instead of AsCov $\operatorname{vec}\hat{G}$. However, the results have a direct link to the parameter variance, cf. Remark 4.1. The system is given by

$$G(q) = \begin{bmatrix} q^{-1} & 2q^{-1}+q^{-2} \\ q^{-1}+4q^{-2} & q^{-1}+2q^{-2} \end{bmatrix}. \qquad (26)$$

The model is

$$G(q,\theta) = \begin{bmatrix} \theta_{11}^1 q^{-1} & \theta_{12}^1 q^{-1} + \theta_{12}^2 q^{-2} \\ \theta_{21}^1 q^{-1} + \theta_{21}^2 q^{-2} & \theta_{22}^1 q^{-1} + \theta_{22}^2 q^{-2} \end{bmatrix}, \qquad (27)$$

where all parameters have been given a superscript that denotes which basis function they correspond to. One parameter, related to the basis function q^{-2}, is not estimated for G_{11}, and this will lead to a decrease in variance for the other parameters related to the basis function q^{-2} compared to those related to the basis function q^{-1}. We will investigate how the amount of correlation in the input determines the variance reduction. The Cholesky factors of the noise and input are given by

$$\Lambda_{CH} = \begin{bmatrix} 1 & 0 \\ 0.8 & (1-0.8^2)^{1/2} \end{bmatrix}, \quad \Sigma_{CH} = \begin{bmatrix} (1-\beta^2)^{1/2} & \beta \\ 0 & 1 \end{bmatrix},$$

respectively. Thus $\sigma_1^2 = \sigma_2^2 = 1$, while the parameter β determines the amount of correlation between the inputs, i.e., $\operatorname{E}[u_1(t)u_2(t)] = \beta$. For each β, we generate $MC = 2000$ Monte Carlo experiments, where in each of them we collect $N = 2000$ input-output samples. At the i-th Monte Carlo run, we generate new trajectories for the input and the noise. The sample asymptotic covariance matrix, for each β, is computed as the sample covariance matrix multiplied by the number of samples N.

$$\operatorname{AsCov}\hat{\theta} = \frac{N}{MC}\sum_{i=1}^{MC}(\hat{\theta}^i - \theta^o)(\hat{\theta}^i - \theta^o)^T,$$

It is seen in Figure 2 that the variance is reduced for all parameters related to the basis function q^{-2} compared to those related to q^{-1}, as long as there is correlation between the inputs (the variance would have been the same if also G_{11} had a parameter related to q^{-2}). The variance of $\hat{\theta}_{12}^2$ is independent of the input correlation and behaves exactly as the parameters of Corollary 4.2, even though the Corollary is not applicable since θ_{21}^2 is still estimated. This indicates that it should be possible to strengthen Corollary 4.2. The whole input excitation of $q^{-2}u_2(t)$ is used for this estimate i.e., AsVar $\hat{\theta}_{12}^2 = \lambda_1/\sigma_2^2 = 1$, since $y_1(t)$ does not depend on $q^{-2}u_1(t)$. The variance of $\hat{\theta}_{21}^2$ is strictly smaller than $\hat{\theta}_{21}^1$, regardless of the input correlation. The variance of this parameter behaves exactly as in Corollary 4.3, even though the assumptions of Corollary 4.3 are not met (θ_{12}^2 is still estimated). This indicates that it should be possible to strengthen Corollary 4.3. The measurement $y_1(t)$ can be used to estimate the noise affecting $y_2(t)$ without corrupting the estimate $\hat{\theta}_{21}^2$, i.e., AsVar $\hat{\theta}_{21}^2 = \lambda_{2|1}/\sigma_{1|2}^2$, since $y_1(t)$ does not depend on $q^{-2}u_1(t)$. The estimate $\hat{\theta}_{22}^2$ also benefits, as long as there is correlation between the inputs.

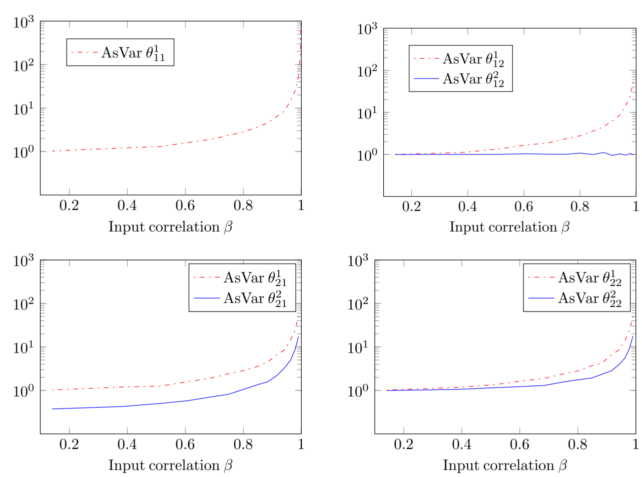

Fig. 2. *Variance of the parameters of G_{11}, G_{12}, G_{21} and G_{22} respectively. The parameters related to the basis function q^{-2}, θ_{12}^2, θ_{21}^2 and θ_{22}^2, are estimated with reduced variance compared to those related to the q^{-1}. The variance of θ_{12}^2 does not depend on the input correlation.*

B. Second Experiment

Let again the system be given by (26) and the model by (27). In this experiment we vary the noise correlation instead of the input correlation. Let the Cholesky factors of the noise and input be given by

$$\Lambda_{CH} = \begin{bmatrix} 1 & 0 \\ \beta & (1-\beta^2)^{1/2} \end{bmatrix}, \quad \Sigma_{CH} = \begin{bmatrix} 0.8 & (1-0.8^2)^{1/2} \\ 0 & 1 \end{bmatrix},$$

respectively. It is seen in Figure 3 that the variance is reduced in all parameters related to the basis function q^{-2} compared to those related to q^{-1} as long as there is correlation between the outputs. The variance of $\hat{\theta}_{12}^2$ behaves as the MISO case, i.e., $\mathrm{AsVar}\,\hat{\theta}_{12}^2 = \lambda_1/\sigma_2^2 = 1$ and $\mathrm{AsVar}\,\hat{\theta}_{12}^2$ is independent of the correlation between the outputs. The variance of $\hat{\theta}_{21}^2$ and $\hat{\theta}_{22}^2$ are lower than $\hat{\theta}_{21}^1$ and $\hat{\theta}_{22}^1$ respectively for nonzero correlation, and the variances are the same when there is zero correlation. When β goes to one, the non-estimable part of $e_2(t)$ conditioned on $e_1(t)$ goes to zero and we achieve perfect estimation of $\hat{\theta}_{21}^2$.

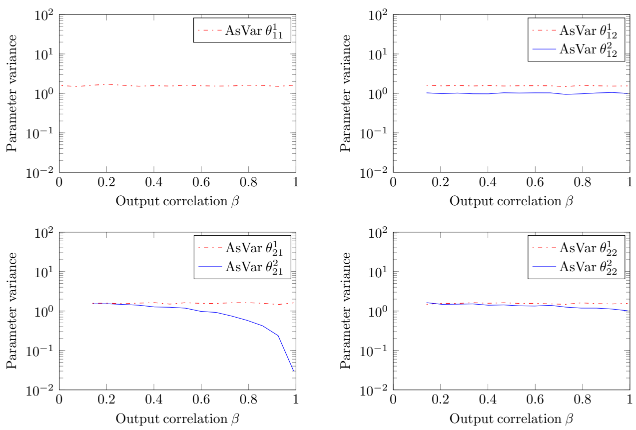

Fig. 3. *Variance of the parameters of G_{11}, G_{12}, G_{21} and G_{22} respectively. The parameters related to the basis function q^{-2}, θ_{12}^2, θ_{21}^2 and θ_{22}^2, are estimated with reduced variance compared to those related to the q^{-1}. The variance of θ_{12}^2 does not depend on the noise correlation.*

VI. Conclusions

In this paper, we have extended existing variance formulas for MIMO systems to finite model orders, assuming temporally uncorrelated input and noise. We have shown that model structure strongly influences the effect spatial correlation in inputs and noise have on the variance of module transfer function estimates. We have highlighted connections with recently developed results for SIMO and MISO models. The interplay of model structure, input correlation and noise correlation has been exemplified by numerical simulations. We believe there is ample space for further research. For instance, including temporal correlation on input and noise still presents a challenging problem.

References

[1] M. Barenthin and H. Hjalmarsson, "Identification and control: Joint input design and state feedback with ellipsoidal parametric uncertainty via LMIs," *Automatica*, vol. 44, no. 2, pp. 543 – 551, 2008.

[2] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall, 1999.

[3] H. Hjalmarsson, "System identification of complex and structured systems," *European Journal of Control*, vol. 15, no. 3-4, pp. 275–310, Aug. 2009.

[4] R. Isermann, "Model–based fault–detection and diagnosis – status and applications," *Annual Reviews in Control*, vol. 29, no. 1, pp. 71 – 85, 2005.

[5] C. R. Rojas, J. S. Welsh, G. C. Goodwin, and A. Feuer, "Robust optimal experiment design for system identification," *Automatica*, vol. 43, no. 6, pp. 993–1008, 2007.

[6] Y. Zhu, *Multivariable system identification for process control*. Elsevier, 2001.

[7] Z. D. Yuan and L. Ljung, "Black-box identification of multivariable transfer functions–asymptotic properties and optimal input design," *International Journal of Control*, vol. 40, no. 2, pp. 233–256, 1984.

[8] B. Ninness and J. C. Gómez, "Asymptotic analysis of MIMO system estimates by the use of orthonormal bases," in *Proceedings of the 13th IFAC World Congress*, 1996.

[9] P. Ramazi, H. Hjalmarsson, and J. Mårtensson, "Variance analysis of identified linear MISO models having spatially correlated inputs, with application to parallel Hammerstein models," *Automatica*, vol. 50, no. 6, pp. 1675 – 1683, 2014.

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

[11] B. Wahlberg, "System identification using Laguerre models," *IEEE Transactions on Automatic Control*, vol. 36, pp. 551–562, 1991.

[12] P. S. Heuberger, P. M. Van den Hof, and O. H. Bosgra, "A generalized orthonormal basis for linear dynamical systems," *Automatic Control, IEEE Transactions on*, vol. 40, no. 3, pp. 451–465, 1995.

[13] B. Ninness and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," *IEEE Transactions on Automatic Control*, vol. 42, pp. 515–521, April 1997.

[14] J. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. 25, no. 9, pp. 772–781, Sep 1978.

[15] T. Söderström and P. Stoica, *System identification*. Englewood Cliffs: Prentice-Hall, 1989.

[16] D. R. Cox and N. Reid, "Parameter orthogonality and approximate conditional inference," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 49, no. 1, pp. pp. 1–39, 1987.

[17] H. Hjalmarsson and J. Mårtensson, "A geometric approach to variance analysis in system identification," *IEEE Transactions on Automatic Control*, vol. 56, no. 5, pp. 983 –997, may 2011.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

Appendix

A. Useful lemmas

Lemma 1.1: Let $u(t) \in \mathbb{R}^m$ have zero mean and covariance matrix $\Sigma > 0$. Let Σ_{CH} be the upper triangular Cholesky factor of Σ so that (7) holds, and let the elements of its inverse be denoted as in (8). Then, for $k \leq i-1$

$$\sigma_{i|k:i-1}^2 = \frac{1}{\sum_{j=k}^{i} \gamma_{ji}^2}.$$

Proof: Available upon request to the authors ■

Lemma 1.2: Let $u(t) \in \mathbb{R}^m$ have zero mean and covariance matrix $\Sigma > 0$. Let Σ_{CH} be the upper triangular Cholesky factor of Σ so that (7) holds. Let Σ_{CH}^{-1} be partitioned according $[u_1(t), \ldots, u_{i-1}(t)]^T$ and $[u_i(t), \ldots, u_m(t)]^T$ as:

$$\Sigma_{CH}^{-1} = \begin{bmatrix} [\Sigma_{CH}^{-1}]_{11} & [\Sigma_{CH}^{-1}]_{12} \\ 0 & [\Sigma_{CH}^{-1}]_{22} \end{bmatrix}.$$

Then

$$\Sigma_{i:m}^{-1} = [\Sigma_{CH}^{-1}]_{22}^T [\Sigma_{CH}^{-1}]_{22}$$

Proof: Available upon request to the authors ■