52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

A Geometric Approach to Variance Analysis of Cascaded Systems

Niklas Everitt, Cristian R. Rojas and Håkan Hjalmarsson

***Abstract*— Modeling complex and interconnected systems is a key issue in system identification. When estimating individual subsystems of a network of interconnected system, it is of interest to know the improvement of model-accuracy in using different sensors and actuators. In this paper, using a geometric approach, we quantify the accuracy improvement from additional sensors when estimating the first of a set of subsystems connected in a cascade structure. We present results on how the zeros of the first subsystem affect the accuracy of the corresponding model. Additionally we shed some light on how structural properties and experimental conditions determine the accuracy. The results are particularized to FIR systems, for which the results are illustrated by numerical simulations. A surprising special case occurs when the first subsystem contains a zero on the unit circle; as the model orders grows large, the variance of the frequency function estimate, evaluated at the corresponding frequency of the unit-circle zero, is shown to be the same as if the other subsystems were completely known.**

***Index Terms*— Asymptotic covariance, cascade systems, system identification.**

I. Introduction

Many dynamical systems are, or can be modeled as, composed of subsystems interconnected in a network structure. When estimating models of structured systems, including structural information may improve the accuracy of the models and ease their interpretability. In this paper we will regard the structure as known a priori. In a structured system, there might be a large set of sensors and actuators available, and choosing which of them to use and their location is an important issue. Therefore, it is interesting to understand how each sensor and actuator affect model accuracy (model identifiability is also of interest, but is outside of the scope of this paper, see e.g. [1]). Some contributions in this vein includes multi-input systems considered in [2], and cascaded systems studied in [3].

Recently, a geometric approach developed in [4], [5] has shed some light into how structural properties (e.g. model structure and model order), and experimental conditions (e.g. input spectrum and noise variances) determine the asymptotic covariance matrix. This type of analysis concern orthogonal projections in the Hilbert space $\mathcal{L}_2$ and has already given valuable insights into cascaded and multi-sensor systems [5]. The key results are conditions on the subsystems, under which adding more sensors do not improve the accuracy. The remaining problem that we will address is quantifying the improvement in accuracy of using more sensors.

Given a cascaded setup and experimental conditions, there are expressions for the resulting model's accuracy, measured as the asymptotic covariance of the estimated parameters. However, these expressions do not provide insight into how structural properties and experimental conditions affect the covariance. Our aim is to use the geometric approach to give such insights for the cascaded structure. In that sense, our work is an extension of the work presented in [4]. The main contribution of this paper is to present some results on how the zeros of the first subsystem affect the information gain of using multiple sensors in the cascaded structure, when estimating that first subsystem. We try to quantify the contribution of each sensor to a model's accuracy, with focus on a frequency range dependent on the zeros of the transfer function of interest, and the number of model parameters. This information might be useful in determining if we should use additional sensors. There is a trade-off to be made since additional sensors means more data have to be collected, and more parameters have to be identified. Hence, if the gain in accuracy is concentrated mainly in regions of little interest, the additional effort may not be worthwhile.

After presenting the problem in Section II we recall some technical preliminaries and variance expressions in Section III. Section IV contains the main results, which covers the calculation of one basis function and its implication on the frequency response estimate. Our theoretical findings are illustrated and confirmed by simulations on finite impulse response (FIR) models in Section V. The paper is finished with conclusions and some final remarks in Section VI.

Notation

We will treat vector valued complex functions as row vectors, and the inner product of two such functions $f(z), g(z) : \mathbb{C} \to \mathbb{C}^{1\times m}$ is defined as $\langle f, g\rangle \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\omega})g^*(e^{i\omega})\,\mathrm{d}\omega$ where g^* denotes the complex conjugate transpose of g. Furthermore $\overline{f}$ denotes the complex conjugate of f. In case f, g are matrix valued functions we keep the same notation whenever the matrix dimensions are compatible. We denote by $\|f\|$ the $\mathcal{L}_2$-norm of $f : \mathbb{C} \to \mathbb{C}^{n\times m}$ and it is given by $\sqrt{\mathrm{Tr}\,\langle f, f\rangle}$. We call two functions f, g orthogonal if $\langle f, g\rangle = 0$; if f, g are matrix valued, they are considered orthogonal if every entry of the resulting matrix is zero. A set of functions $\{\mathcal{B}_k\}_{k=1}^n$ is said to be orthonormal if they are mutually orthogonal with unit $\mathcal{L}_2$-norm. If $\Psi \in \mathcal{L}_2^{n\times m}$, we denote by $\mathcal{S}_\Psi \subset \mathcal{L}_2^m$ the subspace spanned by the rows of Ψ. We denote the orthogonal projection of f onto the space

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.

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, crro, hjalmars}@kth.se.)

$\mathcal{S}_\Psi$ by $\mathrm{P}_{\mathcal{S}_\Psi}[f]$, i.e., $\mathrm{P}_{\mathcal{S}_\Psi}[f]$ is the unique solution to

$$\min_{g\in\mathcal{S}_\Psi}\|g-f\|.$$

For a differentiable function $f:\mathbb{R}^n\to\mathbb{C}^q$, $f'(\bar{x})$ is a $n\times q$ matrix with $\frac{\mathrm{d}f_j(x)}{\mathrm{d}x_i}\Big|_{x=\bar{x}}$ as (i,j)th entry. For a row vector X, we will denote by $\mathrm{diag}(X)$ the matrix with the elements of X in the main diagonal and where all other elements equals zero. We use the notation $A^\dagger$ for the Moore-Penrose pseudo inverse of A. Function arguments will, for clarity and lack of space, often be omitted. However, they should be clear from the context.

II. Problem Statement

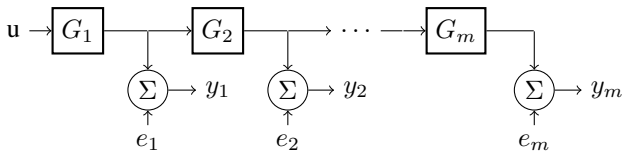

Fig. 1: Cascaded Structure

Consider the single input single output systems connected in cascade as depicted in Fig 1, which are described by the following set of equations:

$$y_k(t)=\prod_{i=1}^{k}G_i(q)u(t)+e_k(t),\qquad k=1,\dots,m,\quad(1)$$

where q denotes the forward shift operator, *i.e.*, $q^{-1}u(t)=u(t-1)$ using normalized sampling time. We assume that the additive noise sequences $\{e_i(t)\}$ are mutually independent zero mean white noise sequences independent of the input $u(t)$ with variances $\lambda_i, i=1,\dots,m$.

The input is assumed to be a realization of a weakly stationary stochastic process with spectrum Φ_u. The models of the subsystems are independently parameterized with $\theta=\big[\theta_1,\dots,\theta_m\big]$, where $\theta_i\in\mathbb{R}^{d_i}$, $d_i\in\mathbb{R}$ for all $i=1,\dots,m$. We assume the true system is in the model set and denote the true parameters by θ^o, that is,

$$G_k(q)=G_k(q,\theta_k^o),\qquad k=1,\dots,m.\quad(2)$$

We are interested in estimating the first subsystem G_1. We assume that the parameter vector θ is estimated from a data set of measured inputs and outputs of sample size N and we denote the estimate by $\hat{\theta}_N$. Under mild regularity conditions (see [6] for details), as N goes to infinity, the parameter error $\sqrt{N}(\hat{\theta}_N-\theta^o)$ converges in distribution to the normal distribution with zero mean and covariance matrix P, which we conveniently denote by

$$\sqrt{N}(\hat{\theta}_N-\theta^o)\in\mathrm{As}\mathcal{N}(0,P).\quad(3)$$

Here P is the asymptotic covariance matrix of the parameter estimates, in the sense that the asymptotic covariance matrix of a stochastic sequence $\{f_N\}_{N=1}^{\infty}, f_N\in\mathbb{C}^q$ is defined as

$$\mathrm{AsCov}\,f_N\triangleq\lim_{N\to\infty}N\cdot\mathbf{E}\left[(f_N-\mathbf{E}f_N)^T\overline{(f_N-\mathbf{E}f_N)}\right].\quad(4)$$

Assume for the moment that the asymptotic covariance matrix can be written as

$$P_N=\mathrm{AsCov}\,\hat{\theta}_N=\langle\Psi,\Psi\rangle^\dagger\quad(5)$$

where $\Psi:\mathbb{C}\to\mathbb{C}^{n\times m}$, for some integer $m>0$ that in our case corresponds to the number of subsystems. All the elements of Ψ are assumed to belong to $\mathcal{L}_2$[1]

Let $J:\mathbb{R}^{1\times n}\to\mathbb{C}^{1\times q}$ be a differentiable function of θ. From (3), it follows that

$$\sqrt{N}(J(\hat{\theta}_N)-J(\theta^o))\in\mathrm{As}\mathcal{N}(0,\mathrm{AsCov}\,J(\hat{\theta}_N)).\quad(6)$$

Using Gauss' approximation formula (or the delta rule) [6] and (3) it can be shown that

$$\mathrm{AsCov}\,J(\hat{\theta}_N)=\Lambda^T[\langle\Psi,\Psi\rangle]^\dagger\overline{\Lambda},\quad(7)$$

where Λ is the derivative $\Lambda\triangleq J'(\theta^o)\in\mathbb{C}^{n\times q}$.

III. Technical Preliminaries and Variance Expressions

In this section we recall some technical preliminaries.

Theorem 3.1: (Theorem II.5 in [4]) Suppose that $J:\mathbb{R}^{1\times n}\to\mathbb{C}^{1\times q}$ is differentiable and let the asymptotic covariance matrix $\mathrm{AsCov}\,J(\hat{\theta}_N)$ be defined by (7) where $\Psi\in\mathcal{L}_2^{n\times m}$. Suppose that $\gamma\in\mathcal{L}_2^{q\times m}$ is such that

$$\Lambda=\langle\Psi,\gamma\rangle,\quad(8)$$

then

$$\mathrm{AsCov}\,J(\hat{\theta}_N)=\langle\mathrm{P}_{\mathcal{S}_\Psi}[\gamma],\mathrm{P}_{\mathcal{S}_\Psi}[\gamma]\rangle^T,\quad(9)$$

where $\mathcal{S}_\Psi$ is the subspace of $\mathcal{L}_2^m$ spanned by the rows of Ψ.

In some cases it is possible to calculate the projection explicitly, like in our case. The following corollary to Theorem 3.1 will be instrumental in calculating the projection.

Lemma 3.1: (Lemma II.9 in [4]) Let J, Ψ, and $\mathcal{S}_\Psi$ be as in Theorem 3.1 and suppose that $\Lambda=\Psi(z_o)L$ for some $z_0\in\mathbb{C}$ and $L\in\mathbb{C}^{m\times q}$. Let $\{\mathcal{B}_k\}_{k=1}^{l}, l\le n$, be an orthonormal basis for $\mathcal{S}_\Psi$. Then

$$\mathrm{AsCov}\,J(\hat{\theta}_N)=L^T\sum_{k=1}^{l}\mathcal{B}_k^T(z_o)\overline{\mathcal{B}_k(z_o)}\,\overline{L}.\quad(10)$$

Let F_u be a stable minimum-phase spectral factor of Φ_u. Then, for the cascaded system, we can express Ψ in (5) as

$$\Psi=\begin{bmatrix}G_1' & G_2G_1' & \cdots & G_m\cdots G_2G_1'\\ 0 & G_2'G_1 & \ddots & G_m\cdots G_3G_2'G_1\\ \vdots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & G_m'G_{m-1}\cdots G_1\end{bmatrix}F_uW^{-1},\quad(11)$$

where

$$W=\mathrm{diag}([\lambda_1^{1/2},\dots,\lambda_m^{1/2}])\quad(12)$$

with G_j' being the partial derivative of G_j with respect to θ_j. Define T as the row vector containing all frequency responses between u and $\begin{bmatrix}y_1 & \cdots & y_m\end{bmatrix}$, that is

$$T\triangleq\begin{bmatrix}G_1 & G_2G_1 & \cdots & G_m\cdots G_1\end{bmatrix},\quad(13)$$

[1]This is the standard situation when the true parameter vector corresponds to a stable predictor in the prediction error method, see [6].

then

$$\begin{aligned}\Lambda &= T' \\ &= \begin{bmatrix} G_1' & G_2G_1' & \cdots & G_m\cdots G_2G_1' \\ 0 & G_2'G_1 & \ddots & G_m\cdots G_3G_2'G_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & G_m'G_{m-1}\cdots G_1 \end{bmatrix} \\ &= \Psi W F_u^{-1}.\end{aligned}$$

This allows us to apply Lemma 3.1 to express the asymptotic covariance of T as

$$\text{AsCov}\,\hat{T}(e^{j\omega}) = \Phi_u^{-1}(e^{j\omega})W^T \sum_{k=1}^{l} \mathcal{B}_k^T(e^{j\omega})\overline{\mathcal{B}_k(e^{j\omega})}\,\overline{W}. \quad (14)$$

We only need to consider the top left corner of (14), because we are only interested in $\hat{G}_1$. Without loss of generality, the basis functions can be expressed as

$$\mathcal{B}_k = \begin{bmatrix} c_k^1\mathcal{B}_k^1 & \dots & c_k^m\mathcal{B}_k^m \end{bmatrix}, \quad (15)$$

where $c_k^j \in \mathbb{R}^+$, $\|\mathcal{B}_k^j\| = 1$ if $\mathcal{B}_k^j \neq 0$ for all $k \leq l, j \leq m$. For each basis function, the coefficients $\{c_k^j\}_{j=1}^m$ satisfy $\|\begin{bmatrix} c_k^1 & \dots & c_k^m \end{bmatrix}\|_2^2 = 1$, which ensures that $\|\mathcal{B}_k\| = 1$. If $\mathcal{B}_k^j = 0$ then c_k^j is defined to be zero. With this notation, we obtain

$$\text{AsVar}\,\hat{G}_1(e^{j\omega}) = \frac{\lambda_1}{\Phi_u(e^{j\omega})} \sum_{k=1}^{l} |c_k^1\mathcal{B}_k^1(e^{j\omega})|^2. \quad (16)$$

In the next section we will provide a specific basis that will be used to obtain insights of (16).

IV. Calculating Basis Functions and Variance Results

In this section we present a theorem that let us determine one basis function, $\mathcal{B}_1$, of a set $\{\mathcal{B}_k\}_{k=1}^l$ that spans Ψ and satisfies

$$\langle \mathcal{B}_1^i, \mathcal{B}_j^i \rangle = 0, \qquad i = 1, \dots, l \quad j = 2, \dots, l, \quad (17)$$

allowing us to compute the variance in (16) at a specific frequency point.

Theorem 4.1: Let G_1 have a zero in $\xi \in \mathbb{C}$ and none of G_1-G_m have a pole in ξ, i.e. there is no pole-zero cancelation of the zero ξ in Ψ. Define $\{\mathcal{A}_k^i\}_{k=1}^{l_i}$ as a basis for the space spanned by the rows of the ith column in Ψ for $i = 1, \dots, m$. Then we can choose a basis $\{\mathcal{B}_k\}_{k=1}^l$ for Ψ with

$$\mathcal{B}_k = \begin{bmatrix} c_k^1\mathcal{B}_k^1 & \dots & c_k^m\mathcal{B}_k^m \end{bmatrix}, \quad k = 1, \dots, l, \quad (18)$$

such that

$$\langle \mathcal{B}_1^i, \mathcal{B}_j^i \rangle = 0, \qquad i = 1, \dots, m \quad j = 2, \dots, l, \quad (19)$$

where

$$\begin{aligned} c_1^1 = b_1\lambda_1^{-1/2}\Big(& b_1^2\lambda_1^{-1} + b_2^2\lambda_2^{-1}|G_2(\xi)|^2 + \\ & + \cdots + b_l^2\lambda_m^{-1}|G_m(\xi)\cdots G_2(\xi)|^2 \Big)^{-1/2} \end{aligned} \quad (20a)$$

$$\mathcal{B}_1^i = b_i \sum_{k=1}^{l_i} \overline{\mathcal{A}_k^i(\xi)}\mathcal{A}_k^i, \quad i = 1, \dots, m, \quad (20b)$$

and $\{b_1^i\}_{i=1}^m$ are normalization constants such that $\|\mathcal{B}_1^i\| = 1$.

Proof: See Appendix B. ∎

Notice that (19) does not follow from the orthogonality of the basis function $\{\mathcal{B}_k\}$ as the next example illustrates.

Example 4.1: Consider

$$\Psi(z) = \begin{bmatrix} z^{-1} & z^{-2} \\ z^{-1} & -z^{-2} \end{bmatrix},$$

formed from $\{\mathcal{A}_k^1\}_{k=1}^1 = z^{-1}$ and $\{\mathcal{A}_k^2\}_{k=1}^1 = z^{-2}$. Recall that $\{z^{-k}\}_{k=0}^\infty$ forms a (complete) orthonormal set [7]. Then one choice of basis functions is

$$\mathcal{B}_1(z) = 2^{-1/2}\begin{bmatrix} z^{-1} & z^{-2} \end{bmatrix}, \quad \mathcal{B}_2(z) = 2^{-1/2}\begin{bmatrix} z^{-1} & -z^{-2} \end{bmatrix}$$

and $\langle \mathcal{B}_1, \mathcal{B}_2 \rangle = 0$, even though $\langle \mathcal{B}_1^1, \mathcal{B}_2^1 \rangle = 1$.

The work invested in calculating a specific basis function will now be put to use. The covariance of the transfer function estimate will, at the frequency of the zero, be strongly linked to that specific basis function as the next theorem shows.

Theorem 4.2: Assume that G_1 has a zero at $\xi = re^{j\omega_\xi}$, where $r \in \mathbb{R}^+$ and that the assumptions in Theorem 4.1 hold. Then

$$\begin{aligned} \text{AsVar}\,\hat{G}_1(e^{j\omega_\xi}) \leq & \frac{\lambda_1}{\Phi_u(e^{j\omega_\xi})}\|f_1\|^2\Big(1 - \\ & - (1 - (c_1^1)^2)\langle \mathcal{B}_1, f_1/\|f_1\| \rangle^2\Big), \end{aligned} \quad (21)$$

where

$$f_1 = \sum_{k=1}^{l_1} \overline{\mathcal{A}_k^1(e^{j\omega_\xi})}\mathcal{A}_k^1. \quad (22)$$

If $r = 1$, then (21) simplifies to

$$\text{AsVar}\,\hat{G}_1(e^{j\omega_\xi}) = \frac{\lambda_1}{\Phi_u(e^{j\omega_\xi})}(c_1^1)^2 b_1^{-2}, \quad (23)$$

or equivalently

$$\text{AsVar}\,\hat{G}_1(e^{j\omega_\xi}) = \frac{\lambda_1}{\Phi_u(e^{j\omega_\xi})}Z^{-1}, \quad (24)$$

where

$$\begin{aligned} Z = & \frac{b_1^2}{\lambda_1} + \frac{b_2^2}{\lambda_2}|G_2(e^{j\omega_\xi})|^2 + \\ & + \cdots + \frac{b_m^2}{\lambda_m}|G_m(e^{j\omega_\xi})\cdots G_2(e^{j\omega_\xi})|^2. \end{aligned} \quad (25)$$

Proof: Using Lemma A.2 we can rewrite (16) as

$$\text{AsVar}\,\hat{G}_1(re^{j\omega_\xi}) = \frac{\lambda_1}{\Phi_u}\sum_{k=1}^{l}(c_k^1)^2|\langle \mathcal{B}_k^1, f_1 \rangle|^2. \quad (26)$$

To arrive at (21), we notice that we can bound the contribution from the other basis functions through $c_k \leq 1$ and $\sum_{k=1}^r \langle \mathcal{B}_k, f_1/\|f_1\| \rangle^2 = 1$. If $r = 1$, then $f_1 = b_1^{-1}\mathcal{B}_1$, and the orthogonality of $\{\mathcal{B}_k^1\}$ gives

$$\text{AsVar}\,\hat{G}_1(e^{j\omega_\xi}) = \frac{\lambda_1}{\Phi_u}(c_1^1)^2 b_1^{-2}. \quad (27)$$

∎

The expressions (21-23) give insight into how much the variance of $G(e^{j\omega_\xi})$ would increase if we remove the sensor

that measures y_2, or decrease if we add another sensor. Similar observations can be made when $r \neq 1$ regarding the upper bound in (21). If we take a closer look at c_1^1, we notice that it resembles the signal to noise ratio of the sensors. That means, a higher ratio between the gain from the output of G_1 to a given sensor and the noise variance, leads to better accuracy. The constants $\{b_j\}_{j=1}^m$ provide a scaling that in some sense depends on the "size" of the corresponding spaces, so that sensors far away from G_1 contribute less to the accuracy of the model of G_1. Note that if one subsystem is zero at $e^{j\omega_\xi}$, i.e., $G_j(e^{j\omega_\xi}) = 0$, then any sensor downstream of G_j will not reduce the variance of $G_1(e^{j\omega_\xi})$ regardless of their signal-to-noise ratio, which should come as no surprise. This generalizes to the following observation: if $G_j(e^{j\omega_\xi})$ is small, then the benefit from sensors downstream of G_j is limited. These results are related to those found in [3], where, in the case of three subsystems, knowing y_2 and y_3 will not improve the quality of the estimate of θ_1, if $G_2(q,\theta_2^o)G_1'(q,\theta_1^o) = G_2'(q,\theta_2^o)G_1(q,\theta_1^o)$. Our results have a weaker condition and provide a weaker result, that is, if one zero is common between G_1 and G_2, then y_2 and y_3 will not considerably improve the quality of the estimate around the corresponding frequency.

Corollary 4.1: Let the assumptions of Theorem 4.2 hold. Then

$$\begin{aligned}\text{AsVar } \hat{G}_1(e^{j\omega}) \leq & \frac{\lambda_1}{\Phi_u(e^{j\omega})}\|f_1\|^2\Big(1- \\ & -(1-(c_1^1)^2)\langle \mathcal{B}_1, f_1/\|f_1\|\rangle^2\Big),\end{aligned} \tag{28}$$

where

$$f_1 = \sum_{k=1}^{l_1} \overline{\mathcal{A}_k^1(e^{j\omega})}\mathcal{A}_k^1(z). \tag{29}$$

Proof: Follows from the proof of Theorem 4.2, by changing ω_ξ to ω in the definition of f_1. ∎

Theorem 4.2 and Corollary 4.1 only consider one basis function determined by one zero of G_1. If we know additional zeros, Theorem 4.2 provides different sets of basis functions for the different zeros and corresponding lower bounds for the variance at the corresponding frequencies. However using Gram-Smith orthogonalization we can determine orthonormal basis functions of one set and refine the lower bounds in Theorem 4.2 and Corollary 4.1 accordingly.

V. FIR Case

In this section, we consider FIR systems, all with true order p, that is, subsystem j can be expressed for some $\{g_{j,k}\}_{k=1}^p \in \mathbb{R}^p$ as

$$G_j(q) = \sum_{k=1}^{p} g_{j,k}q^{-k}.$$

The first system G_1 has a zero at $\xi = e^{j\omega_\xi}$, $\omega_\xi \in [0, 2\pi]$, $\Phi_u = 1$, and we estimate each transfer function with n parameters. We only give the expressions for two subsystems to ease notation, however, the generalization to more sensors is straightforward. We assume that we can take $\{\mathcal{A}_k^1\} = \{z^{-k}\}_{k=l}^{n+(l-1)p}$ as a basis for l:th column of Ψ, then $b_k^2 = n + (l-1)p - (l-1)$ (which does not hold if for example $G_1 = G_2$). Equation (24) becomes

$$\text{AsVar } \hat{G}_1(e^{j\omega_\xi}) = \frac{n}{\lambda_1^{-1} + \frac{n}{n+p-1}\lambda_2^{-1}|G_2(e^{j\omega_\xi})|^2}. \tag{30}$$

We would expect that when we estimate both G_1 and G_2 with many parameters, the benefit of the second sensor would diminish. However, this is not the case when G_1 has a zero on the unit circle:

$$\lim_{n\to\infty}\lim_{N\to\infty}\frac{N}{n}\text{Var } \hat{G}_1(e^{j\omega_\xi}) = \frac{1}{\frac{1}{\lambda_1} + \frac{1}{\lambda_2}|G_2(e^{j\omega_\xi})|^2}.$$

This is the same asymptotic variance as if the subsystem G_2 was completely known (the noise signals are uncorrelated, so the optimal variance is the inverse of the sum of the information in each signal). To us, this is a surprising result. However, it fits with the discussion on optimal input design of structured systems [8], where the input signal should be designed to hide unimportant system properties. Here, since $G_1(e^{j\omega_\xi}) = 0$, the output signal y_2 does not contain any information about $G_2(e^{j\omega_\xi})$.

TABLE I: Comparison of the asymptotic variance of $\hat{G}_1(e^{j\pi})$ with Monte-Carlo simulations.

$n^{-1}\text{AsVar } \hat{G}_1(e^{j\pi})$	Equation (23)	Directly (7)	Monte-Carlo
$n = 2$	0.6000	0.6000	0.5913
$n = 10$	0.5238	0.5238	0.5413

In the following example and Example 5.2, (30), and its generalization, are verified by MATLAB simulations.

Example 5.1: We consider a setup according to (1), with two FIR transfer functions: $G_1 = z^{-1} + z^{-2}$ and $G_2 = z^{-1}$, where $\lambda_1 = \lambda_2 = 1$. G_1 thus have a zero at $z = e^{j\pi}$. The input signal is white noise with variance $\sigma_u^2 = 1$. The variance at $\omega_\xi = \pi$ is estimated in 3 ways: calculated from (24), calculated directly from (7) and calculated as the variance from 1,000 Monte-Carlo simulations. When estimating the given system with FIR models of order n, (24) reduces to

$$\text{AsVar } \hat{G}_1(e^{j\omega_\xi}) = \frac{1}{\frac{1}{n} + \frac{1}{n+1}}. \tag{31}$$

Each estimate in the Monte-Carlo simulations is the minimizer of the cost function

$$\sum_{i=1}^{N}(y_1(i) - G_1(q)u(i))^2 + (y_2(i) - G_1(q)G_2(q)u(i))^2, \tag{32}$$

with $N = 10^4$ data points. The minimization is solved with Matlab's built in function *lsqnonlin*, initialized at the true parameter values of the systems.

The Monte-Carlo simulations come close to what is predicted by Equation (24) (cf. Table I). The reduction in variance, compared to only using y_1, is centered around the frequency of the zero (see Figures 2 and 3). Note that, if we only use y_1 then $(N/n)Var\hat{G}_1(e^{j\omega_\xi}) = 1$ for all frequencies. The intuition is that the input to the second system is zero

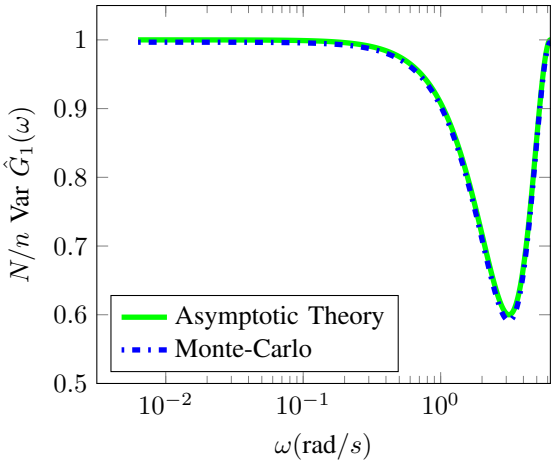

Fig. 2: Asymptotic variance of $\hat{G}_1$, weighted by number of estimated parameters $n = 2$, and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The variance is reduced around the frequency where G_1 is zero, when also the output y_2 from G_2 is also used.

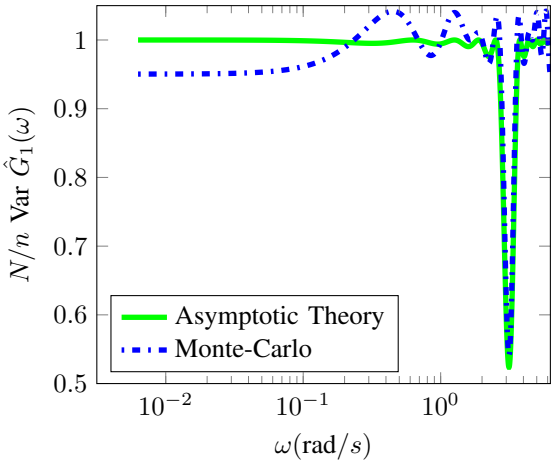

Fig. 3: Asymptotic variance of $\hat{G}_1$, weighted by number of estimated parameters $n = 10$, and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The reduction in variance is more focused around the frequency where G_1 is zero compared to Figure 2.

at the frequency $e^{j\omega_\xi}$ and therefore, with a slight abuse of notation, $y_2(e^{j\omega_\xi})$ is a signal uncorrelated with the input $u(e^{j\omega_\xi})$. Thus, $y_2(e^{j\omega_\xi})$ tells us that either $G_1(e^{j\omega_\xi}) = 0$, or $G_2(e^{j\omega_\xi}) = 0$ (or both). Furthermore, the transfer functions in our model set are all continuous, which implies that if G_2 is large around $e^{j\omega_\xi}$, it is unlikely to drop to zero at $e^{j\omega_\xi}$. Thereby, y_2 gives us information that $G_1(e^{j\omega_\xi})$ is small. A related aspect not presented in the figures, is that the estimate of $G_2(e^{j\omega_\xi})$ is poor because of the poor signal-to-noise ratio, as expected. When the number of estimated parameters increases, the reduction in variance becomes more focused around ω_ξ. This is due to the increased order of the basis functions (cf. Figure 2 and 3).

Example 5.2: Consider the case when two additional subsystems are appended to the setup of Example 5.1, still satisfying Equation (1), with $G_3 = z^{-2}$, $G_4 = (z^{-1} - z^{-2})/\sqrt{2}$ and $\lambda_3 = \lambda_4 = 1$. Simulations are performed in the same way as in Example 5.1. Again, the Monte-Carlo simulations come close to what is predicted by Equation (24) (cf. Table II). Notice the similarities between the sets of Figures 2 and 3 compared to Figures 4 and 5, and the scaling effect that comes from the additional sensors. The fit is significantly better in Figure 2 because the ratio between number of samples and estimated parameters is larger.

TABLE II: Comparison of the asymptotic variance of $\hat{G}_1(e^{j\pi})$ with Monte-Carlo simulations. The outputs y_1-y_4 are used here.

$n^{-1}\text{AsVar}\ \hat{G}_1(e^{j\pi})$	Equation (23)	Directly (7)	Monte-Carlo
$n = 2$	0.3371	0.3309	0.3305
$n = 10$	0.2336	0.2314	0.2561

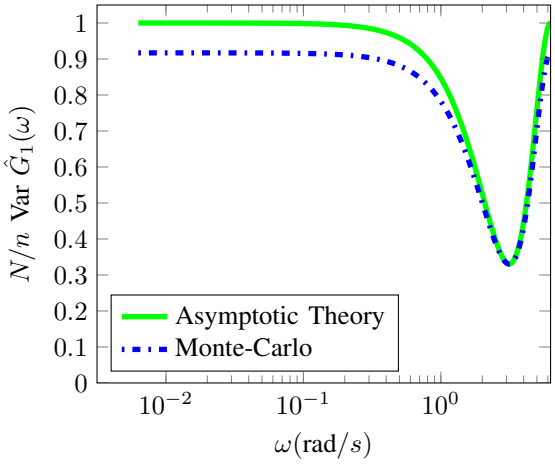

Fig. 4: Asymptotic variance of $\hat{G}_1$, weighted by number of estimated parameters $n = 2$, and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The variance is reduced more than in Figure 2, when also y_3 and y_4 are used.

VI. Conclusions

In this paper we have analyzed the asymptotic variance of the first of a set of subsystem connected in a cascade structure. The main contribution is results on how the zeros of the first subsystem affect the accuracy of the corresponding model. The surprising special case is when the first subsystem contains a zero on the unit circle; as the model orders grow large, the variance at the corresponding frequency of the unit-circle zero is shown to be the same as if the other subsystems were completely known. Additionally, we have derived quantifications of the increase in accuracy for a broader frequency range, the extent of which depends on the number of estimated parameters and model structure.

References

[1] P. M. Van den Hof, A. G. Dankers, P. S. Heuberger, and X. Bombois, "Identification in dynamic networks with known interconnection topology," in *51st IEEE Conference on Decision and Control*, 2012.

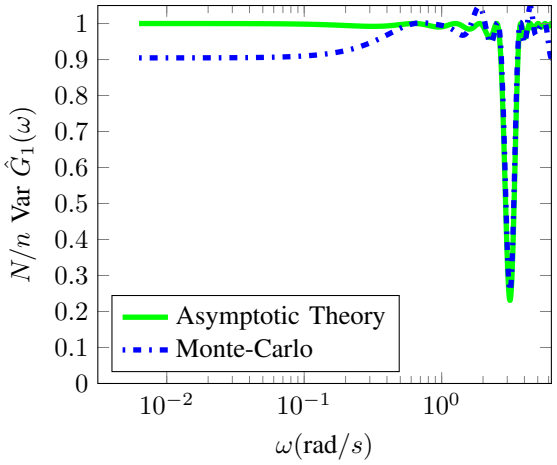

Fig. 5: Asymptotic variance of $\hat{G}_1$, weighted by number of estimated parameters $n = 10$, and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The variance is reduced more than in Figure 3, when also y_3 and y_4 are used.

[2] M. Gevers, L. Mišković, D. Bonvin, and A. Karimi, "Identification of multi-input systems: variance analysis and input design issues," *Automatica*, vol. 42, no. 4, pp. 559 – 572, 2006.

[3] B. Wahlberg, H. Hjalmarsson, and J. Mårtensson, "Variance results for identification of cascade systems," *Automatica*, vol. 45, no. 6, pp. 1443 – 1448, 2009.

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

[5] J. Mårtensson, "Geometric analysis of stochastic model errors in system identification," Ph.D. Thesis, KTH, Automatic Control, oct 2007.

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

[7] A. Friedman, *Foundations of Modern Analysis*, ser. Dover Books on Mathematics Series. Dover, 1970.

[8] H. Hjalmarsson, "System Identification of Complex and Structured Systems," *European Journal of Control*, vol. 15, no. 3-4, pp. 275–310, Aug. 2009.

[9] B. Ninness and H. Hjalmarsson, "Variance Error Quantifications That are Exact for Finite Model-Order," *IEEE Transactions on Automatic Control*, vol. 49, no. 8, pp. 1275–1291, 2004.

Appendix A
Preliminary Lemmas

Lemma A.1: Let $f \in \mathcal{L}_2^l$ and let $\mathcal{S}_n^m$ be a (closed) subspace of $\mathcal{L}_2^m$ with orthonormal basis $\{\mathcal{B}_k\}_{k=1}^n$. Then

$$\mathrm{P}_{\mathcal{S}_n^m}[f] \triangleq \sum_{k=1}^{n} \langle f, \mathcal{B}_k \rangle \mathcal{B}_k \tag{A.1}$$

is the orthogonal projection of f on $\mathcal{S}_n^m$, i.e., it is the unique solution to

$$\min_{g \in \mathcal{S}_n^m} \|g - f\|.$$

Proof: See, e.g., [7]. ∎

Lemma A.2: Let $\mathcal{X}$ be a finite-dimensional subspace of $\mathcal{L}_2^m$, $\{\mathcal{A}_k\}_{k=1}^{\infty}$ be an orthonormal basis for $\mathcal{L}_2^m$ and $\{\mathcal{A}_k\}_{k=1}^{r}$ a basis for $\mathcal{X}$. If the function $f : \mathbb{C} \to \mathbb{C}^m$ given by

$$f(z) = \sum_{k=1}^{r} \overline{\mathcal{A}_k(\xi)} \mathcal{A}_k(z), \tag{A.2}$$

where $\xi \in \mathbb{C}$, then for any $G \in \mathcal{X}$

$$\langle G, f \rangle = G(\xi). \tag{A.3}$$

Proof: We can express G as

$$G = \sum_{j=1}^{r} G_j \mathcal{A}_j, \tag{A.4}$$

for some scalars $\{G_j\}$. From the orthogonality of the basis functions we have that

$$\langle G, f \rangle = \langle \sum_{j=1}^{r} G_j \mathcal{A}_j, \sum_{k=1}^{r} \overline{\mathcal{A}_k(\xi)} \mathcal{A}_k \rangle \tag{A.5}$$

$$= \sum_{k=1}^{r} G_k \mathcal{A}_k(\xi) = G(\xi). \tag{A.6}$$

∎

Remark A.1: The function f is strongly related to the so-called "reproducing kernel" for the space $\mathcal{X}$ (see [9] and the references therein).

Corollary A.1: Let $G \in \mathcal{X}$ and f be as in Lemma A.3. Then

$$\langle G, f \rangle = 0, \tag{A.7}$$

if and only if

$$G(\xi) = 0. \tag{A.8}$$

Proof: Follows directly from Lemma A.2. ∎

Appendix B
Proof of Theorem 4.1

For ease of notation we will assume $\Phi_u(e^{j\omega}) = 1$ for $\omega \in [0, 2\pi]$ and $\lambda_1 = \cdots = \lambda_m = 1$; it is straightforward to adjust the derivations for the general setting. It is not obvious that $\mathcal{B}_1$ lies in the space spanned by the rows of Ψ; however, this will be clear from how we construct $\mathcal{B}_1$. To construct $\mathcal{B}_1$, we project Ψ onto the space $\mathcal{X}$, defined as the span of the m functions that have $\mathcal{B}_1^k$ in the kth column and zeros in the other columns, e.g. the first one being $\begin{bmatrix}\mathcal{B}_1^1 & 0 & \cdots\end{bmatrix}$. Here, $\mathcal{B}_1^k$ is defined as in (20b). Applying Lemma A.1, we obtain

$$\mathrm{P}_{\mathcal{X}}[\Psi] = \begin{bmatrix} \langle G_1', \mathcal{B}_1^1 \rangle \mathcal{B}_1^1 & \cdots & \langle G_m \cdots G_1', \mathcal{B}_1^m \rangle \mathcal{B}_1^m \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix}, \tag{B.1}$$

where all rows are zero except the first because of Corollary A.1. Furthermore, we have used the fact that $\mathcal{B}_1^k, k = 1, \ldots, l$ have unit norm. Applying Lemma A.2 to (B.1) we obtain

$$\mathrm{P}_{\mathcal{X}}[\Psi] = \begin{bmatrix} G_1'(\xi) \\ 0 \\ \vdots \end{bmatrix} \begin{bmatrix} b_1 \mathcal{B}_1^1 & \cdots & b_m G_m(\xi) \cdots G_2(\xi) \mathcal{B}_m^2 \end{bmatrix}.$$

Normalize the vector $\begin{bmatrix} b_1 \mathcal{B}_1^1 & \cdots & b_m G_m(\xi) \cdots G_2(\xi) \mathcal{B}_1^m \end{bmatrix}$ to form $\mathcal{B}_1$. Note that $\mathrm{P}_{\mathcal{X}}[\Psi] = \begin{bmatrix} C & 0 \end{bmatrix}^T \mathcal{B}_1 = \mathrm{P}_{\mathcal{B}_1}[\Psi]$ for some constant vector C, that is, $\dim(\mathrm{P}_{\mathcal{X}}[\Psi]) = 1$ and $\mathcal{B}_1$ is a basis vector for this space. Consequently:

$$\Psi - \mathrm{P}_{\mathcal{B}_1}[\Psi] = \mathrm{P}_{\mathcal{X}^\perp}[\Psi] \perp \mathcal{X}, \tag{B.2}$$

which is equivalent to (19).