# A Geometric Approach to Variance Analysis in System Identification

Håkan Hjalmarsson, *Member, IEEE*, and Jonas Mårtensson

***Abstract*—This paper addresses the problem of quantifying the model error ("variance-error") in estimates of dynamic systems. It is shown that, under very general conditions, the asymptotic (in data length) covariance of an estimated system property (represented by a smooth function of estimated system parameters) can be interpreted in terms of an orthogonal projection of a certain function, associated with the property of interest, onto a subspace determined by the model structure and experimental conditions. The presented geometric approach simplifies structural analysis of the model variance and this is illustrated by analyzing the influence of inputs and sensors on the model accuracy.**

***Index Terms*—Asymptotic covariance, model accuracy, stochastic systems, system identification.**

## I. Introduction

QUANTIFICATION of the model error is a core issue in system identification and much research effort has been devoted to this issue; the following list of contributions is by no means complete: [1]–[21]. In the stochastic setting [22] that we will consider in this contribution, unknown system parameters $\theta = [\theta_1, \ldots, \theta_n] \in \mathbb{R}^{1\times n}$ are estimated using a data set consisting of measured inputs and outputs resulting in the parameter estimate $\hat{\theta}_N \in \mathbb{R}^{1\times n}$. We assume that a true parameter value exists, denoted by $\theta^o$, and we assume that the (normalized) model error $\sqrt{N}(\hat{\theta}_N - \theta^o)$ becomes normally distributed as the sample size $N$ of the data set grows to infinity

$$\sqrt{N}\left(\hat{\theta}_N - \theta^o\right) \in \mathrm{As}\mathcal{N}(0, \mathrm{AsCov}\hat{\theta}_N). \tag{1}$$

The asymptotic covariance matrix $\mathrm{AsCov}\hat{\theta}_N$ of the limit distribution is a measure of the model accuracy. This is reinforced by that, under mild conditions [22]

$$\lim_{N\to\infty} N \cdot \mathbf{E}\left[(\hat{\theta}_N - \mathbf{E}\hat{\theta}_N)^T(\hat{\theta}_N - \mathbf{E}\hat{\theta}_N)\right] = \mathrm{AsCov}\hat{\theta}_N.$$

The asymptotic covariance matrix was early used for optimal experiment design [23]–[26] and Ljung's [27] model-order asymptotic variance expression led to experiment designs for high-order models [28]–[31]. Optimal experiment design problems have received a renewed interest in recent years [32]–[37].

### A. The Importance of Asymptotic Covariance Theory

As a motivating example we would like to mention Zhu, [38], [39], who has developed an experiment design procedure, entirely based on the asymptotic covariance theory, that has significantly reduced the time and cost related to modeling in several industrial plants. We quote [39]:

> Model identification plays a crucial role in MPC technology and it is also the most time consuming and difficult task in MPC projects and maintenance. [...]
>
> Considerable benefits are obtained using the new identification technology: 1) reduction of identification test time and model building time by over 70%; 2) higher model quality for control; [...]

One specific example is given in the case study [40], which deals with the re-tuning of an MPC controller at the Hovensa refinery in Virgin Islands, United States, which is one of the largest refineries in the world, with a crude oil processing capacity of 495.000 barrels per day. The MPC application at hand had 34 inputs and 90 outputs, and the last time the model was tuned it took one month to gather the data and one month to identify the model. With the new method, the total modeling time was reduced to just five days; a reduction of over 90%.

This example more than well justifies the use of the asymptotic covariance theory for assessing the model quality, and we will now return to the study of the asymptotic covariance matrix.

### B. The Asymptotic Covariance Matrix

Often the asymptotic covariance matrix can be written as

$$\mathrm{AsCov}\hat{\theta}_N = \langle \Psi, \Psi \rangle^{-1} \tag{2}$$

where $\Psi : \mathbb{C} \to \mathbb{C}^{n\times m}$, for some integer $m > 0$ depending on the model structure, and whose elements belong to $\mathcal{L}_2$ and where $\langle \Psi, \Psi \rangle$ denotes the integral $1/2\pi \int_{-\pi}^{\pi} \Psi(e^{j\omega})\Psi^*(e^{j\omega})\mathrm{d}\omega$ (superscript $*$ denotes complex conjugate transpose). The relation (2) will be our standing assumption throughout this paper.

*1) Example 1:* Consider a system described by the linear regression

$$y_t = \theta^o \varphi_t^T + e_t \tag{3}$$

where $\theta^o \in \mathbb{R}^{1\times n}$, where $\varphi_t \in \mathbb{R}^{1\times n}$ are elements of a stationary stochastic process with bounded moments and $\mathbf{E}\left[\varphi_t^T \varphi_t\right] > 0$, and where $\{e_t\}$ is white noise with bounded

Manuscript received March 25, 2009; revised December 11, 2009, June 01, 2010, and August 04, 2010; accepted August 04, 2010. Date of publication September 13, 2010; date of current version May 11, 2011. This work was supported in part by the Swedish Research Council under Contract 621-2007-6271 and Contract 621-2009-4017. Recommended by T. Zhou.

H. Hjalmarsson and J. Mårtensson are with the ACCESS Linnaeus Center, KTH School of Electrical Engineering, KTH—Royal Institute of Technology, S-100 44 Stockholm, Sweden (e-mail: hakan.hjalmarsson@ee.kth.se; jonas.martensson@ee.kth.se).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2010.2076213

moments, variance $\lambda_o$ and independent of $\varphi_t$. The least-squares estimate of $\theta^o$ has asymptotic covariance matrix

$$\mathrm{AsCov}\hat{\theta}_N = \lambda_o \left[\mathbf{E}\left[\varphi_t^T \varphi_t\right]\right]^{-1} \tag{4}$$

see [22]. If we take $\Psi$ to be a Cholesky factor of $1/\lambda_o \mathbf{E}\left[\varphi_t^T \varphi_t\right]$, (2) holds for this $\Psi$. ■

When the regressor vector $\varphi_t$ in Example 1 has some structure, structure can be provided to $\Psi$ as well.

*2) Example 2:* The $n$th order finite impulse response (FIR) model

$$y_t = G(q,\theta)u_t + e_t = \sum_{k=1}^{n} \theta_k u_{t-k} + e_t \tag{5}$$

with $\{e_t\}$ being white noise with bounded moments and variance $\lambda_o$, corresponds to (3) with $\varphi_t = [u_{t-1}, \ldots, u_{t-n}] = \Gamma_n^T(\mathrm{q})u_t$ where $\Gamma_n(\mathrm{q}) = [\, q^{-1} \quad \ldots \quad q^{-n}\,]^T$. Here $q^{-k}$ is the delay operator: $q^{-k}u_t = u_{t-k}$. Now, if $\{u_t\}$ is stationary with spectrum $\Phi_u$, it follows from the multivariable counterpart to [22, Theorem 2.2] that the spectrum for $\{\varphi_t\}$ is given by

$$\Gamma_n(e^{j\omega})\Phi_u(\omega)\Gamma_n^*(e^{j\omega})$$

and, from the definition of a spectrum as the Fourier transform of the correlation sequence, it thus follows that:

$$\begin{aligned}\mathbf{E}\left[\varphi_t^T \varphi_t\right] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_n(e^{j\omega})\Phi_u(\omega)\Gamma_n^*(e^{j\omega})d\omega \\ &= \langle \Gamma_n R_u, \Gamma_n R_u\rangle \end{aligned} \tag{6}$$

where $R_u$ denotes a stable minimum phase spectral factor of the spectrum $\Phi_u$ (so that $R_u R_u^* = \Phi_u$). Thus, for an FIR model structure with the true system in the model set and with a stationary input sequence, (2) holds with

$$\Psi = \frac{1}{\sqrt{\lambda_o}}\Gamma_n R_u. \tag{7}$$

■

Example 2 can be generalized to prediction error models where the true system belongs to the model structure [22]. The regressor $\varphi_t$ is then replaced by the prediction error gradient.

The primary interest is often not the model parameters $\theta$ themselves but some "system theoretic" quantity such as the frequency response, impulse response coefficients, poles and zeros, some norm of the system, or the performance of a closed loop system where the controller is designed using the identified model. We will let such a quantity be represented by a differentiable function $J : \mathbb{R}^{1\times n} \to \mathbb{C}^{1\times p}$. Given an estimate $\hat{\theta}_N$ of the true parameter $\theta^o$, which we assume belongs to the set of model parameters, a natural estimate of $J(\theta^o)$ is $J(\hat{\theta}_N)$. This estimator is motivated by that if $\hat{\theta}_N$ is asymptotically efficient, i.e., it is consistent and its asymptotic covariance matrix equals the Cramér-Rao lower bound, then $J(\hat{\theta}_N)$ is also asymptotically efficient. This follows from the *invariance principle*, see [41].

It follows from (1) (and some additional mild conditions, see [22]) that

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

where, using Gauss' approximation formula and (2), it can be shown that

$$\mathrm{AsCov}J(\hat{\theta}_N) = \Lambda^T \langle \Psi, \Psi\rangle^{-1}\overline{\Lambda}$$

where $\Lambda$ is the derivative $\Lambda := J'(\theta^o) \in \mathbb{C}^{n\times p}$. We shall be slightly more general and allow cases where $\langle \Psi, \Psi\rangle$ is singular. Then the parameter estimate $\hat{\theta}_N$ is non-unique but there may be other properties $J(\hat{\theta}_N)$ that are identifiable from the data. In that case, provided that $\Lambda$ belongs to the column space of $\langle \Psi, \Psi\rangle$, the correct variance is given by

$$\mathrm{AsCov}J(\hat{\theta}_N) = \Lambda^T \langle \Psi, \Psi\rangle^{\dagger}\overline{\Lambda}. \tag{8}$$

We refer to [42], [43] for details. In the following, we will assume that $\Lambda$ belongs to the column space of $\langle \Psi, \Psi\rangle$ and consider (8) as a *definition* of $\mathrm{AsCov}J(\hat{\theta}_N)$.

From knowledge of the model structure, the true system and experimental conditions, the asymptotic covariance of $J(\hat{\theta}_N)$ can be computed, at least, numerically from (8). However, it is not straightforward to obtain structural insights into how quantities such as model structure, model order and experimental conditions influence the asymptotic covariance matrix from (8). Understanding the nature of this expression for different problem settings has been subject of rather intense research. An example of a well known structural result is that the variance of the model parameters increases with the model order [44], [45]. Another, more recent, result in the same vein has been to determine when estimating a transfer function in a multiple input single output system may help improve the quality of estimates of other transfer functions in the model [46]. Also related to our work is the very interesting paper [47] where conditions for the asymptotic covariance matrix to be non-singular are studied.

Most of what is known regarding the structure of (8) pertains to models of linear time-invariant systems. An exception is [48] where the asymptotic variance for the linear part of a Hammerstein model is characterized asymptotically as the model order tends to infinity and when the linear part has a fixed denominator.

Recently, there have been several interesting contributions to non-asymptotic assessment of confidence regions [3], [49]–[51]. In [52] it is discussed under which conditions the asymptotic theory is valid.

*3) Contributions and Outline:* This contribution has its origin in [53], where exact expressions for the asymptotic variance of frequency function estimates for LTI models were derived using the theory of reproducing kernels. It was in this work that the importance of the subspace generated by the rows of $\Psi$ was recognized. Our work can be seen as an extension of the work in [53] to general system properties (represented by the function $J$) and a more general class of model structures

using new techniques which deepens the geometric interpretation of (8) initiated in [53]. Related to our work is [54], where a geometric interpretation is given to the Cramér-Rao lower bound for the special case where the estimate of the parameter vector $\theta$ is based on an observation of a normal distributed variable $y$ with mean $x(\theta)$ and known covariance. An approach similar to [54] is taken in [55]. A preliminary version of this paper has appeared as [56].

More precisely, the contributions of this paper are:

i) Section II: *A geometric interpretation of the asymptotic variance (8).* The results referred to above are by no means obvious by studying the initial expression (8), instead many of the results are proved using rather intricate matrix manipulations. A new geometric interpretation is presented which facilitates the understanding of how different quantities such as model structure, model order, input spectra and the quantity of interest, influence the asymptotic accuracy.

ii) Section III: *General structural results*. Here we provide some general results concerning what happens with (8) when the structure of $\Psi$ changes.

iii) Section IV: *Structural results for LTI-systems*. In this section we apply the results of Section III to some structural problems in estimation of linear time invariant (LTI) systems. First we illustrate the method by examining how system complexity impacts the model accuracy. The particular example of non-minimum phase zero identification is studied and both matrix algebra and the geometric approach are used so that the reader may compare these two methods. Secondly, we examine the impact different inputs have whereby we extend results in [46]. Finally, we examine the impact different number of sensors have on the estimation accuracy.

### NOTATION

We will consider vector valued complex functions as row vectors and the inner product of two such functions $f, g : \mathbb{C} \to \mathbb{C}^{1\times m}$ is defined as $\langle f, g\rangle = 1/2\pi \int_{-\pi}^{\pi} f(e^{j\omega})g^*(e^{j\omega})\mathrm{d}\omega$ where $g^*$ denotes the complex conjugate transpose of $g$. When $f$ and $g$ are matrix-valued functions, we will still use the notation $\langle f, g\rangle$ to denote $1/2\pi \int_{-\pi}^{\pi} f(e^{j\omega})g^*(e^{j\omega})\mathrm{d}\omega$ whenever the dimensions of $f$ and $g$ are compatible. The elements of $\langle f, g\rangle$ will consist of the inner products between the functions corresponding to rows of $f$ and rows of $g$. In particular, $\langle f, f\rangle$ is the Gramian matrix of the rows of $f$. The $\mathcal{L}_2$-norm of $f : \mathbb{C} \to \mathbb{C}^{n\times m}$ is given by $\|f\| = \sqrt{\mathrm{Tr}\langle f, f\rangle}$. The space $\mathcal{L}_2^{n\times m}$ consists of all functions $f : \mathbb{C} \to \mathbb{C}^{n\times m}$ such that $\|f\| < \infty$ and when $n = 1$, the notation is simplified to $\mathcal{L}_2^m$. For $f : \mathbb{C} \to \mathbb{C}^{n\times m}$, $f_i : \mathbb{C} \to \mathbb{C}^{1\times m}$ denotes the $i$th row of $f$. When $f$ is partitioned into blocks, with $ij$th block $f_{i,j}$, $f_{\star j}$ denotes the function $[\, f_{1,j}, f_{2,j}, \dots]^T$. Similarly, $f_{i\star}$ denotes $[\, f_{i,1}, f_{i,2}, \dots]$. If $\Psi \in \mathcal{L}_2^{n\times m}$ for some positive integers $n$ and $m$, then $\mathcal{S}_\Psi$ denotes the subspace of $\mathcal{L}_2^m$ generated by the rows of $\Psi$.

For a differentiable function $f : \mathbb{R}^{1\times n} \to \mathbb{C}^{1\times p}$, $f'(\bar{x})$ is a $n \times p$ matrix with $\partial f_j(x)/\partial x_i\big|_{x=\bar{x}}$ as $ij$th entry, the partial derivative $\partial f(\bar{x})/\partial x_i$ is defined analogously.

With $\mathcal{X}, \mathcal{Y}$ being closed subspaces of a Hilbert space, we use $\mathcal{X} + \mathcal{Y}$ to denote the subspace $\{x + y : x \in \mathcal{X}, y \in \mathcal{Y}\}$, $\mathcal{Z} = \mathcal{X} \oplus \mathcal{Y}$ to denote that $\mathcal{Z}$ is the direct sum of $\mathcal{X}$ and $\mathcal{Y}$, i.e., that $\mathcal{Y}$ is the orthogonal complement of $\mathcal{X}$ in $\mathcal{Z}$. We use $\mathcal{X}^\perp$ to denote the orthogonal complement of $\mathcal{X}$ with respect to the entire Hilbert space. Furthermore, we define $\mathcal{Y} \ominus \mathcal{X}$ to be the orthogonal complement of $\mathcal{X}$ in $\mathcal{X} + \mathcal{Y}$, i.e., $\mathcal{X} + \mathcal{Y} = \mathcal{X} \oplus (\mathcal{Y} \ominus \mathcal{X})$.

By $x \sim \mathcal{N}(m, P)$ we mean that $x$ is a normal distributed random vector with mean $m$ and covariance matrix $P$. The Moore-Penrose pseudo-inverse of a matrix $A$ is denoted $A^\dagger$.

## II. GENERAL RESULTS

The purpose of this paper is to provide a geometric interpretation of (8). After recalling some standard results in Section II-A, we will in Section II-B show how an isometric isomorphism, between the stochastic model error $\hat{\theta}_N - \theta^o$ and a deterministic object related to $\Psi$, leads to the paper's main technical result, Theorem II.5. This result will serve as foundation for the sequel. In Sections II-C–II-E follow some results that can be used as building blocks when analyzing (8).

Technically, the results concern orthogonal projections in the Hilbert space $\mathcal{L}_2$ and several results are well known; the main contribution of this section is the linking of these results to the asymptotic covariance expression (8).

### A. Technical Preliminaries

*Lemma II.1:* Let $\Psi \in \mathcal{L}_2^{n\times m}$ for some finite positive integers $n$ and $m$. Then the number of linearly independent rows of $\Psi$, or, equivalently, the dimension of $\mathcal{S}_\Psi$, the subspace of $\mathcal{L}_2^m$ generated by the rows of $\Psi$, is given by the rank of $\langle \Psi, \Psi\rangle$.

*Proof:* Straightforward. ∎

Next, we have the following standard result from Hilbert space theory.

*Lemma II.2:* Let $f \in \mathcal{L}_2^m$ and let $S$ be a closed subspace of $\mathcal{L}_2^m$ with orthonormal basis $\{\mathcal{B}_k\}_{k=1}^n$, $n \in \{1, 2, \dots, \infty\}$. Then

$$\mathbf{P}_S\{f\} := \sum_{k=1}^{n} \langle f, \mathcal{B}_k\rangle \mathcal{B}_k$$

is said to be the orthogonal projection of $f$ on $S$ and is the unique solution to

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

*Proof:* See, e.g., Lemma 6.2.1, Theorem 6.4.2 and Theorem 6.4.4 in [57]. ∎

*Remark:* Notice that with $S$ and $\{\mathcal{B}_k\}_{k=1}^n$ as in Lemma II.2 but with $f \in \mathcal{L}_2^{p\times m}$ for some positive integer $p$

$$\sum_{k=1}^{n} \langle f, \mathcal{B}_k\rangle \mathcal{B}_k = [\mathbf{P}_S\{f_1\}^T, \dots, \mathbf{P}_S\{f_p\}^T]^T$$

so we write also in this case $\mathbf{P}_S\{f\} = \sum_{k=1}^n \langle f, \mathcal{B}_k\rangle \mathcal{B}_k$.

*Lemma II.3:* Let $\gamma \in \mathcal{L}_2^{p\times m}$ and $\Psi \in \mathcal{L}_2^{n\times m}$. Then the orthogonal projection of the rows of $\gamma$ on $\mathcal{S}_\Psi$ (the subspace of $\mathcal{L}_2^m$ spanned by the rows of $\Psi$) is given by

$$\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\} = \langle \gamma, \Psi\rangle \langle \Psi, \Psi\rangle^\dagger \Psi. \tag{9}$$

Furthermore

$$\langle \gamma, \Psi\rangle\langle \Psi, \Psi\rangle^{\dagger}\langle \Psi, \gamma\rangle = \langle \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}\rangle. \tag{10}$$

Finally, it holds that

$$\langle \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}\rangle = \sum_{k=1}^{r}\langle \gamma, \mathcal{B}_k\rangle\langle \mathcal{B}_k, \gamma\rangle \tag{11}$$

where $\{\mathcal{B}_k\}_{k=1}^{r}$, for some $r \leq n$, is any orthonormal basis for $\mathcal{S}_\Psi$.

*Proof:* This is a standard result and we only point out that by taking the inner product of the right-hand side of (9) with itself, (10) readily follows. Furthermore, the equality in (11) follows using Lemma II.2 when evaluating the left hand side. ■

*Lemma II.4:* With $\Psi$ and $\gamma$ as in Lemma II.3, the rows of $\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}$ span a subspace of $\mathcal{S}_\Psi$ with dimension equal to the rank of $\langle \gamma, \Psi\rangle$

$$\dim \mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}} = \mathrm{Rank}\langle \gamma, \Psi\rangle. \tag{12}$$

In particular, $\dim \mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}} = 0$ if and only if $\gamma \perp \mathcal{S}_\Psi$.

*Proof:* By writing $\Psi = T\Gamma$ where $T \in \mathbb{C}^{n\times r}(r \leq n)$ has full (column) rank and with $\Gamma$ being orthonormal, i.e., $\langle \Gamma, \Gamma\rangle = I_{r\times r}$, the result follows straightforwardly from Lemma II.1 and Lemma II.3. ■

### B. Basic Result

Estimation is closely linked to (orthogonal) projection, cf. optimal filtering and parameter estimation. Consider for example the least-squares problem in Example 1. With $Y = [\,y_1 \;\; \dots \;\; y_N\,]$ and $\Phi = [\,\varphi_1^T \;\; \dots \;\; \varphi_N^T\,]$, the problem becomes that of finding the best predictor $\hat{Y}(\theta) = \theta\Phi$ of $Y$, which effectively means that the predictor should be the orthogonal projection of $Y$ onto the subspace spanned by the rows of $\Phi$. The least-squares estimate is given by

$$\hat{\theta}_N = Y\Phi^T[\Phi\Phi^T]^{-1} = \theta^o + E\Phi^T[\Phi\Phi^T]^{-1} \tag{13}$$

where $E = [\,e_1 \;\; \dots \;\; e_N\,]$, and if we introduce the inner product $\langle x, z\rangle = xz^T$ in the linear vector space of real valued row vectors of dimension $N$, we have from (13) that [1]

$$\hat{Y}(\hat{\theta}_N) = \hat{\theta}_N\Phi = \langle Y, \Phi\rangle\langle \Phi, \Phi\rangle^{-1}\Phi \tag{14}$$

which is exactly the right-hand side of (9) so that $\hat{Y}(\hat{\theta}_N) = \mathbf{P}_{\mathcal{S}_\Phi}\{Y\}$, i.e., the optimal predictor is indeed the orthogonal projection of $Y$ onto the linear subspace spanned by the rows of $\Phi$. One may wonder whether the parameter estimate $\hat{\theta}_N$ itself can be related to some projection in a meaningful way. As it turns out, this can be done for the estimation error $\hat{\theta}_N - \theta^o$. For simplicity of argument we will assume that $\Phi$ is deterministic and that $\langle \Phi, \Phi\rangle > 0$. Under these assumptions and those made on the noise $\{e_k\}$ in Example 1, the elements of $\hat{\theta}_N - \theta^o$ are random variables with zero mean and finite second order moments so we can consider them as elements of the Hilbert space with inner product[2] $\langle x, y\rangle_{\mathbf{E}} = \mathbf{E}\,[x^T y\,]$. Notice that

$$\mathbf{E}\left[(\hat{\theta}_N - \theta^o)^T(\hat{\theta}_N - \theta^o)\right] = \langle \hat{\theta}_N - \theta^o, \hat{\theta}_N - \theta^o\rangle_{\mathbf{E}} \tag{15}$$

i.e., the covariance matrix consists of all inner products between the elements of $\hat{\theta}_N - \theta^o$. Furthermore, from (13) it follows that this covariance matrix can be expressed as

$$\mathbf{E}\left[(\hat{\theta}_N - \theta^o)^T(\hat{\theta}_N - \theta^o)\right] = \lambda_o\langle \Phi, \Phi\rangle^{-1}. \tag{16}$$

We will now argue that we can identify, through an isometric isomorphism, $\hat{\theta}_N - \theta^o$ with an element, which itself is a projection, in another Hilbert space which consists of deterministic elements. To this end notice first that the projection (14) gives

$$\langle \mathbf{P}_{\mathcal{S}_\Phi}\{Y\}, \mathbf{P}_{\mathcal{S}_\Phi}\{Y\}\rangle = \langle Y, \Phi\rangle\langle \Phi, \Phi\rangle^{-1}\langle \Phi, Y\rangle.$$

This calculation suggests that if we instead of projecting $Y$, project the rows of some $\gamma \in \mathbb{R}^{n\times N}$ with the property that $\langle \gamma, \Phi\rangle = \sqrt{\lambda_o}I_{n\times n}$, we will obtain a matrix[3] $\mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}$ whose rows are the projection of the rows of $\gamma$ onto $\mathcal{S}_\Phi$ and for which it holds that

$$\langle \mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}\rangle = \lambda_o\langle \Phi, \Phi\rangle^{-1}. \tag{17}$$

Combining (15)–(17) we have shown that

$$\langle \hat{\theta}_N - \theta^o, \hat{\theta}_N - \theta^o\rangle_{\mathbf{E}} = \langle \mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}\rangle \tag{18}$$

which means that the elements of $\hat{\theta}_N - \theta^o$ (as seen as elements of the Hilbert space consisting of zero mean random variables with finite variance and inner product $\langle x, y\rangle_{\mathbf{E}} = \mathbf{E}[x^T y]$) can be related to the rows of $\mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}$ (which are seen as elements of the Hilbert space of $\mathbb{R}^{1\times N}$ vectors with inner product $\langle x, y\rangle = xy^T$) through an isometric isomorphism. This means that the second order statistical properties of $\hat{\theta}_N - \theta^o$ can be deduced from $\mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}$. In particular (15) and (18) give

$$\mathbf{E}\left[(\hat{\theta}_N - \theta^o)^T(\hat{\theta}_N - \theta^o)\right] = \langle \mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}\rangle.$$

Furthermore, if we are interested in the variance of a linear combination $J(\hat{\theta}_N) = \hat{\theta}_N\Lambda$ of the parameter estimates, we obtain

$$\begin{aligned}\mathbf{E}\left[|J(\hat{\theta}_N) - J(\theta^o)|^2\right] &= \Lambda^T\mathbf{E}\left[(\hat{\theta}_N - \theta^o)^T(\hat{\theta}_N - \theta^o)\right]\Lambda \\ &= \|\mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}\|^2\end{aligned}$$

for any $\gamma$ for which $\langle \Phi, \gamma\rangle = \sqrt{\lambda_o}\Lambda$.

This leads us to our main technical result which generalizes the idea above to asymptotic covariance matrices.

*Theorem II.5:* Suppose that $J : \mathbb{R}^{1\times n} \to \mathbb{C}^{1\times p}$ is differentiable and let the asymptotic covariance matrix $\mathrm{AsCov}J(\hat{\theta}_N)$ be

[1]Notice that for $X \in \mathbb{R}^{r_1\times N}$ and $Z \in \mathbb{R}^{r_2\times N}$, $\langle X, Z\rangle = XZ^T$ is a matrix consisting of the inner products between the rows of $X$ and the rows of $Z$.

[2]The definition includes a transpose of the first argument so that when $x$ and $y$ are random row vectors with elements of this Hilbert space, then $\langle x, y\rangle_{\mathbf{E}} = \mathbf{E}\,[x^T y\,]$ is a matrix with the inner products between all elements of $x$ and $y$.

[3]Notice that the notation is such that the same formula $\mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\} = \langle \gamma, \Phi\rangle\langle \Phi, \Phi\rangle^{-1}\Phi$ as before gives as result a matrix whose rows are the rows of $\gamma$ projected on $\mathcal{S}_\Phi$.

defined by (8) where $\Psi \in \mathcal{L}_2^{n\times m}$. Suppose that $\gamma \in \mathcal{L}_2^{p\times m}$ is such that

$$\Lambda = \langle \Psi, \gamma \rangle \tag{19}$$

then

$$\mathrm{AsCov} J(\hat{\theta}_N) = \langle \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}\rangle^T \tag{20}$$

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

In particular, when $J$ is scalar

$$\mathrm{AsVar} J(\hat{\theta}_N) = \|\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}\|^2. \tag{21}$$

*Proof:* Follows directly from (8) and Lemma II.3. ∎

The relations (19)–(21) provide a geometric interpretation of the asymptotic covariance matrix. Expression (21) shows that the asymptotic variance of $J(\hat{\theta}_N)$ is given by the squared norm of the projection of $\gamma$ onto the subspace spanned by the rows of $\Psi$ in (8). Notice that $\Psi$ typically depends on the model structure, operating conditions (such as input and noise excitations and feedback configuration) but is independent of the quantity of interest $J$. The link between $\Psi$ and $J$ is given by $\gamma$ through (19). The geometric interpretation together with these two observations can provide useful insights, especially when comparing asymptotic covariances for two different scenarios. For example when $J$ is changed, the function which is projected will change, whereas the subspace onto which the projection takes place will remain the same. Furthermore there are several situations when structural changes in the identification problem result in a change of the subspace whereas the function to be projected can be kept the same, see Section II-D for further discussion. Qualitative information such as which structure has higher asymptotic covariance can then sometimes be obtained by studying the corresponding subspaces without ever computing the function $\gamma$ to be projected. The remaining sections of the paper explore this possibility. For example, in Section IV-A we encounter a case where a whole range of model structures can be compared. However, before embarking on this track we provide some basic results that will be useful in the sequel.

### C. Bounds

It may be cumbersome to compute the projection in (20). Lower and upper bounds can then be computed using the following lemma.

*Lemma II.6:* Let $\mathcal{X}$ and $\mathcal{Y}$ be two closed subspaces of $\mathcal{L}_2^m$ such that $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{L}_2^m$ and let $\gamma \in \mathcal{L}_2^{p\times m}$.

It holds that

$$\langle \mathbf{P}_{\mathcal{Y}}\{\gamma\}, \mathbf{P}_{\mathcal{Y}}\{\gamma\}\rangle - \langle \mathbf{P}_{\mathcal{X}}\{\gamma\}, \mathbf{P}_{\mathcal{X}}\{\gamma\}\rangle = \langle \mathbf{P}_{\mathcal{Y}\ominus\mathcal{X}}\{\gamma\}, \mathbf{P}_{\mathcal{Y}\ominus\mathcal{X}}\{\gamma\}\rangle. \tag{22}$$

*Proof:* By definition it follows that $\mathcal{X}+\mathcal{Y} = \mathcal{X}\oplus(\mathcal{Y}\ominus\mathcal{X})$ and since $\mathcal{X} \subseteq \mathcal{Y}$ we have $\mathcal{X}+\mathcal{Y} = \mathcal{Y}$. Thus, $\mathbf{P}_{\mathcal{Y}}\{\gamma\} = \mathbf{P}_{\mathcal{X}}\{\gamma\} + \mathbf{P}_{\mathcal{Y}\ominus\mathcal{X}}\{\gamma\}$ and by taking the inner product of each side of the equation with itself, the result (22) follows. ∎

Upper bounds for (20) are obtained by taking $\mathcal{X} = \mathcal{S}_\Psi$ and $\mathcal{Y}$ such that the projection $\mathbf{P}_{\mathcal{S}_\mathcal{Y}}\{\gamma\}$ is easy to compute, for example $\mathcal{Y} = \mathcal{L}_2^m$ is one possibility. Lower bounds can be obtained by taking $\mathcal{Y} = \mathcal{S}_\Psi$ and by choosing $\mathcal{X}$ suitably.

The next lemma is useful for characterizing $\mathcal{Y}\ominus\mathcal{X}$ when $\mathcal{Y}$ and $\mathcal{X}$ are the spans of the rows of different $\mathcal{L}_2^m$-functions.

*Lemma II.7:* Let $\Psi \in \mathcal{L}_2^{n_1\times m}$, $\Phi \in \mathcal{L}_2^{n_2\times m}$ and $\gamma \in \mathcal{L}_2^m$ for some positive integers $n_1$, $n_2$ and $m$. Then

$$\mathcal{S}_\Phi \ominus \mathcal{S}_\Psi = \mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Psi^\perp}\{\Phi\}} \tag{23}$$

and

$$\mathbf{P}_{\mathcal{S}_\Phi\ominus\mathcal{S}_\Psi}\{\gamma\} = \mathbf{P}_{\mathcal{S}_\Psi^\perp}\{\gamma_o\} \tag{24}$$

where $\gamma_o := \mathbf{P}_{\mathcal{S}_\Phi}\{\gamma\}$.

*Proof:* $\mathcal{S}_\Phi \ominus \mathcal{S}_\Psi$ is the set of $\mathcal{L}_2^m$-functions in $\mathcal{S}_\Psi + \mathcal{S}_\Phi$ that are orthogonal to $\mathcal{S}_\Psi$. This must thus be a subset of $\mathcal{S}_\Phi = \mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Psi}\{\Phi\}} \oplus \mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Psi^\perp}\{\Phi\}}$. But $\mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Psi}\{\Phi\}} \subseteq \mathcal{S}_\Psi$ and $\mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Psi^\perp}\{\Phi\}} \perp \mathcal{S}_\Psi$ and (23) follows.

Since $\mathcal{S}_\Phi \ominus \mathcal{S}_\Psi \subseteq \mathcal{S}_\Phi$

$$\mathbf{P}_{\mathcal{S}_\Phi\ominus\mathcal{S}_\Psi}\{\gamma\} = \mathbf{P}_{\mathcal{S}_\Phi\ominus\mathcal{S}_\Psi}\{\gamma_o\}. \tag{25}$$

Now we make the orthogonal decomposition of $\gamma_o$ into $\gamma_o = \mathbf{P}_{\mathcal{S}_\Psi^\perp}\{\gamma_o\} + \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma_o\}$ where the first term belongs to $\mathcal{S}_\Phi \ominus \mathcal{S}_\Psi$ and where the second term is orthogonal to $\mathcal{S}_\Phi \ominus \mathcal{S}_\Psi$. By combining these observations with (25) the result (24) follows. ∎

### D. Existence of Suitable Functions $\gamma$

The applicability of Theorem II.5 for rewriting the asymptotic covariance matrix as (20) hinges on the existence of an $\mathcal{L}_2$-function $\gamma$ such that (19) holds. The next lemma shows that there is a whole family of such functions.

*Lemma II.8:* Let $\Psi \in \mathcal{L}_2^{n\times m}$ and let $\Lambda \in \mathbb{C}^{n\times p}$ be such that its columns are in the column space of $\langle \Psi, \Psi\rangle$.

Then all solutions $\gamma \in \mathcal{L}_2^{p\times m}$ to the equation $\Lambda = \langle \Psi, \gamma\rangle$ are given by $\gamma = \Lambda^* \langle \Psi, \Psi\rangle^\dagger \Psi + s^\perp$, where $s^\perp$ is any $\mathcal{L}_2^{p\times m}$-function orthogonal to $\mathcal{S}_\Psi$.

*Proof:* Straightforward. ∎

The possibility to include an arbitrary term $s^\perp$ (as long as it is orthogonal to $\mathcal{S}_\Psi$) in $\gamma$ opens up the possibility of using the same $\gamma$ for different model structures. When this is possible, it will only be the subspaces onto which the projections take place that differ. This can facilitate comparisons significantly. For example, when two model structures are nested in the sense that one subspace is a subset of the other, then Lemma II.6 immediately gives that the asymptotic covariance for the structure corresponding to the larger subspace will be no less than the other one. We refer to Section III-B for more details on this. Another advantage of this degree of freedom in the choice of $\gamma$ is that it may be used so that an upper bound (obtained using Lemma II.6) becomes tractable to compute. We will not pursue this idea here (due to space limitations), instead we refer to [58], [59] where this idea is used to obtain model structure independent upper bounds when the underlying system is linear time invariant.

### E. Explicit Expression for the Projection

In some cases it is easy to compute the projection $\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}$ explicitly.

*Lemma II.9:* Let $J$, $\Psi$ and $\mathcal{S}_\Psi$ be as in Theorem II.5 and suppose that $\Lambda = \Psi(z_o)L$ for some $z_o \in \mathbb{C}$ and $L \in \mathbb{C}^{m\times p}$. Let $\{\mathcal{B}_k\}_{k=1}^r$, $r \leq n$, be an orthonormal basis for $\mathcal{S}_\Psi$. Then (20) can be expressed as

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

*Proof:* Let $\Gamma = [\mathcal{B}_1^T, \ldots, \mathcal{B}_r^T]^T$. Then there exists a full (column) rank $T \in \mathbb{C}^{n\times r}$ such that $\Psi = T\Gamma$. Inserting this in $\Lambda^*\left[\langle\Psi,\Psi\rangle\right]^\dagger\Lambda$ and some straightforward algebra gives the result. ∎

## III. Structural Results

In this section we will consider structural properties of the asymptotic covariance (8). We will derive expressions for how (8) changes when $\Psi$ (and $\Lambda$) changes.

### A. Introduction

We will use some simple examples to illustrate how structural changes affect the estimation accuracy. The purpose is to illuminate the basic principles rather than the final expressions for the asymptotic variances (which in these simple examples can be derived in simpler ways).

Consider again the simple FIR model (5) in Example 2, where we for simplicity assume the noise variance $\lambda_o = 1$. One example of a structural property of the asymptotic covariance is how the variance of a particular parameter (or set of parameters) changes when the system order changes. Let us first assume that the true system order is $n = 1$, i.e.,

$$y_t = \theta_1 u_{t-1} + e_t.$$

Following (7) and taking

$$\Psi = \Psi_1 := \left[\, z^{-1}R_u(z) \,\right] \tag{26}$$

we have from (4) and (6) that the asymptotic covariance of the estimate of $\theta_1$ is given by $P_1 := \langle\Psi_1,\Psi_1\rangle^{-1} = 1/r_0$ where $r_k = \mathbf{E}[u_t u_{t-k}]$. If now the system instead is of order $n = 2$, $\Psi$ changes to

$$\Psi = \Psi_2(z) := \begin{bmatrix} z^{-1}R_u(z) \\ z^{-2}R_u(z) \end{bmatrix} \tag{27}$$

with associated asymptotic covariance

$$P_2 := \langle\Psi_2,\Psi_2\rangle^{-1} = \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix}^{-1} = \frac{1}{r_0^2 - r_1^2}\begin{bmatrix} r_0 & -r_1 \\ -r_1 & r_0 \end{bmatrix}.$$

Since $r_0/(r_0^2 - r_1^2) = 1/(r_0 - r_1^2/r_0) \geq 1/r_0$ we see that the variance of the $\theta_1$-estimate increases when two parameters have to be estimated (unless $r_1 = 0$ in which the variances become equal). Using Schur-complements, this well known fact can be generalized to higher dimensions, see, e.g., Example 9.1 in [60].

Now, if we take a look at what happens structurally when the number of estimated parameters changes we see that the only change is that rows are added to $\Psi$ (compare (26) to (27)). We conclude that when rows are added to $\Psi$ there is an information decrease regarding the original parameters.

Now we will examine how the information can be increased. One obvious method is to increase the excitation. Suppose that a second input $w_t$ (independent of $u_t$) can be used to excite the system, leading to the model

$$y_t = \theta_1(u_{t-1} + w_{t-1}) + e_t.$$

The asymptotic variance is now determined by

$$\Psi(z) = \Psi_3(z) := \left[\, z^{-1}R_u(z) \quad z^{-1}R_w(z) \,\right] \tag{28}$$

resulting in $P_3 :=$

$$\frac{1}{\langle\Psi_3,\Psi_3\rangle} = \frac{1}{\langle R_u, R_u\rangle + \langle R_w, R_w\rangle} = \frac{1}{r_0 + s_0} \leq P_1 \tag{29}$$

where $s_k = \mathbf{E}[w_t w_{t-k}]$. Thus, we have shown the obvious fact that the second input will increase the information in the data and thus improve the estimate.

Another way to increase the information contents is to use more sensors. Suppose that we can use also measurements from

$$z_t = \theta_1 x_{t-1} + v_t \tag{30}$$

where $v_t$ is also zero mean white noise with unit variance (and uncorrelated with $e_t$) and where $x_t$ is a measured input, e.g., $x_t = u_t$. Now the asymptotic variance of the least-squares estimate is again determined by (28) (with $R_w$ exchanged for $R_x$), so that the variance is given by (29). From the last two simple examples we conclude that there is an information increase when columns are added to $\Psi$.

When both rows and columns are added there will be a joint effect of information decrease and increase and one objective of this section is to provide tools for analyzing this situation. Another objective will be to derive general conditions for when the asymptotic variance remains constant to structural changes. This is instrumental when understanding the limitations that need to be considered when complex systems are identified as well as how such systems should be identified. This is also important in order to understand when adding new actuators and sensors to a system is beneficial from a system identification point of view.

### B. A Comparison Theorem

The following theorem is a technical result for comparison of asymptotic covariances for two different structures. We will illustrate its use in ensuing subsections. In Section IV, these results will be used for structural analysis when identifying linear systems.

*Theorem III.1:* Let $\Psi_1 \in \mathcal{L}_2^{n_1\times m}$, $\Psi_2 \in \mathcal{L}_2^{n_2\times m}$ and $\gamma \in \mathcal{L}_2^{p\times m}$ for some positive integers $n_1$, $m$, $n_2$, and $p$. Define

$$X := \langle \mathbf{P}_{\mathcal{S}_{\Psi_1}}\{\gamma\}, \mathbf{P}_{\mathcal{S}_{\Psi_1}}\{\gamma\}\rangle - \langle \mathbf{P}_{\mathcal{S}_{\Psi_2}}\{\gamma\}, \mathbf{P}_{\mathcal{S}_{\Psi_2}}\{\gamma\}\rangle \tag{31}$$

and

$$\Delta := \mathbf{P}_{\mathcal{S}_{\Psi_2}^{\perp}}\{\Psi_1\}. \tag{32}$$

Suppose $\gamma \perp \mathcal{S}_{\Psi_2} \ominus \mathcal{S}_{\Psi_1}$.

Then

$$X = \langle \mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\}\rangle \tag{33}$$

and

$$\mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\} = \mathbf{P}_{\mathcal{S}_{\Psi_1}}\{\gamma\} - \mathbf{P}_{\mathcal{S}_{\Psi_2}}\{\gamma\}. \tag{34}$$

Furthermore,

$$\mathrm{Rank}X = \dim \mathcal{S}_{\mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\}} = \mathrm{Rank}\langle \Delta, \gamma\rangle. \tag{35}$$

*Proof:* The result (33) is an almost immediate consequence of Lemma II.6. We have $\mathcal{S}_\Delta \oplus \mathcal{S}_{\Psi_2} = \mathcal{S}_{\Psi_1} + \mathcal{S}_{\Psi_2} = \mathcal{S}_{\Psi_1} \oplus (\mathcal{S}_{\Psi_2} \ominus \mathcal{S}_{\Psi_1})$, but since $\gamma \perp \mathcal{S}_{\Psi_2} \ominus \mathcal{S}_{\Psi_1}$ we get

$$\mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\} + \mathbf{P}_{\mathcal{S}_{\Psi_2}}\{\gamma\} = \mathbf{P}_{\mathcal{S}_{\Psi_1}}\{\gamma\} \tag{36}$$

which gives (34) upon observing that $\mathcal{S}_\Delta = \mathcal{S}_{\Psi_1} \ominus \mathcal{S}_{\Psi_2}$ (this follows from Lemma II.7). Taking the inner product of (36) with itself gives (33) due to that $\Delta \perp \Psi_2$. Finally, (35) follows from Lemma II.1 and Lemma II.4. ∎

Theorem III.1 explores the possibility of using the same function $\gamma$ to be projected for two different structures that was discussed below Theorem II.5 and in Section II-D. Suppose that we are to estimate some quantity $J$ of a system and that we are given two identification settings (including the experimental conditions, model structure, etc) $\mathcal{I}_i$ which correspond to $\Psi = \Psi_i$, $i = 1, 2$, respectively, in the asymptotic covariance expression (8). When one can find $\gamma \perp \mathcal{S}_{\Psi_2} \ominus \mathcal{S}_{\Psi_1}$ such that $\langle \Psi_i, \gamma\rangle$, $i = 1, 2$ are the derivatives of $J$, the quantity of interest, corresponding to $\mathcal{I}_i$, $i = 1, 2$ then, since $X \geq 0$, Theorem III.1 shows that the asymptotic covariance for $\mathcal{I}_2$ is no larger than that for $\mathcal{I}_1$.

In the proof of Theorem III.1 it was noticed that $\mathcal{S}_\Delta = \mathcal{S}_{\Psi_1} \ominus \mathcal{S}_{\Psi_2}$. Thus, (33) and (34) show that how much $\mathcal{S}_{\Psi_2}$ differs from $\mathcal{S}_{\Psi_1}$ is important for how big the difference $X$ in asymptotic covariance will be. For example, when $\mathcal{S}_{\Psi_1} \subseteq \mathcal{S}_{\Psi_2}$ then $\mathcal{S}_{\Psi_1} \ominus \mathcal{S}_{\Psi_2} = 0$ and $X = 0$ regardless of $\gamma$. Furthermore, (35) gives that the rank of $X$ is upper bounded by $\dim \mathcal{S}_{\Psi_1} \ominus \mathcal{S}_{\Psi_2}$.

One way to explicitly compute the right-hand side of (33) is to express $\Delta$ as $\Delta = \Psi_1 - \langle \Psi_1, \Psi_2\rangle\langle \Psi_2, \Psi_2\rangle^{\dagger}\Psi_2$ using Lemma II.3 and then to use (10) in the same lemma.

### C. Structure Extension

In Section III-A we saw that an information decrease occurs when rows are added to $\Psi$ and an information increase occurs when columns are added. For given $\Psi_{ij} \in \mathcal{L}_2^{n_i \times m_j}$, $i, j = 1, 2$, we are therefore interested in comparing the asymptotic covariance when $\Psi = \Psi_{11}$ and when

$$\Psi = \Psi^e := \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}. \tag{37}$$

In order to work in (row) spaces of the same dimension, we extend the single blocks of $\Psi^e$ with zeros into

$$\begin{aligned} \Psi_{i1e} &:= \begin{bmatrix} \Psi_{i1} & 0_{n_i \times m_2} \end{bmatrix}, \quad i = 1, 2 \\ \Psi_{i2e} &:= \begin{bmatrix} 0_{n_i \times m_1} & \Psi_{i2} \end{bmatrix}, \quad i = 1, 2. \end{aligned} \tag{38}$$

Notice that $\langle \Psi_{11}, \Psi_{11}\rangle = \langle \Psi_{11e}, \Psi_{11e}\rangle$ so whether we use $\Psi_{11}$ or $\Psi_{11e}$ is immaterial when we are computing asymptotic covariances. We will thus compare

$$\langle \mathbf{P}_{\Psi_{11e}}\{\gamma\}, \mathbf{P}_{\Psi_{11e}}\{\gamma\}\rangle \quad \text{with} \quad \langle \mathbf{P}_{\Psi^e}\{\gamma\}, \mathbf{P}_{\Psi^e}\{\gamma\}\rangle \tag{39}$$

where we will comment on the function $\gamma \in \mathcal{L}_2^{p \times (m_1 + m_2)}$ in Section III-D. By choosing either $\Psi_1 = \Psi^e$ and $\Psi_2 = \Psi_{11e}$ or $\Psi_1 = \Psi_{11e}$ and $\Psi_2 = \Psi^e$ in Theorem III.1 we obtain results for when there will be an information decrease or an information increase, respectively, when $\Psi$ is augmented from $\Psi_{11}$ to $\Psi^e$. By using the structure that $\Psi^e$ is an extension of $\Psi_{11}$ we can provide more detailed expressions for the projections that appear in Theorem III.1. This we do next.

*1) Information Decrease:* We start with the case where $\Psi_1 = \Psi^e$ and $\Psi_2 = \Psi_{11e}$. Since $X$ in Theorem III.1 is positive (semi-)definite, this corresponds to the case where augmenting $\Psi$ from $\Psi_{11}$ to $\Psi^e$ will result in a net decrease of information.

*Theorem III.2:* This is an application of Theorem III.1 for $\Psi_1 = \Psi^e$ and $\Psi_2 = \Psi_{11e}$ according to (37)–(38). Let $\gamma \in \mathcal{L}_2^{p \times (m_1 + m_2)}$ and let $\Psi_{ij} \in \mathcal{L}_2^{n_i \times m_j}$, $i, j = 1, 2$ for some integers $p > 0, n_1 > 0, m_1 > 0, n_2 \geq 0$ and $m_2 \geq 0$. Suppose that $\gamma \perp \mathcal{S}_{\Psi_{11e}} \ominus \mathcal{S}_{\Psi^e}$. Define $\Psi_0 := \begin{bmatrix} \mathbf{P}_{\mathcal{S}_{\Psi_{11}}^{\perp}}\{\Psi_{21}\} & \mathbf{P}_{\mathcal{S}_{\Psi_{12}}^{\perp}}\{\Psi_{22}\} \end{bmatrix}$ and let $\Delta$ be as in (32).

Then

$$\mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\} = \mathbf{P}_{\mathcal{S}_{\Psi_{12e}}}\{\gamma\} + \mathbf{P}_{\mathcal{S}_{\Psi_0}}]\{\gamma\} \tag{40}$$

and

$$\langle \Delta, \gamma\rangle = \langle \Psi^e, \gamma\rangle - \begin{bmatrix} I \\ Y \end{bmatrix} \langle \Psi_{11e}, \gamma\rangle \tag{41}$$

where $Y := \langle \Psi_{21}, \Psi_{11}\rangle\langle \Psi_{11}, \Psi_{11}\rangle^{\dagger}$.

Furthermore,

$$\dim \mathcal{S}_\Delta = \dim \mathcal{S}_{\Psi_{12e}} + \dim \mathcal{S}_{\Psi_0} \tag{42}$$

and, finally, $X$ defined in (31) satisfies

$$\mathrm{Rank}X \leq \mathrm{Rank}\langle \Psi_{12e}, \gamma\rangle + \mathrm{Rank}\langle \Psi_0, \gamma\rangle. \tag{43}$$

*Proof:* First (40) and (42) follow from the orthogonal decomposition $\mathcal{S}_\Delta \oplus \mathcal{S}_{\Psi_{11e}} = \mathcal{S}_{\Psi^e} + \mathcal{S}_{\Psi_{11e}} = S_{\Psi_{11e}} \oplus \mathcal{S}_{\Psi_{12e}} \oplus \mathcal{S}_{\Psi_0}$. Furthermore, it follows from the decomposition above that $\mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\} = \mathbf{P}_{\mathcal{S}_{\Psi_{12e}} \oplus \mathcal{S}_{\Psi_0}}\{\gamma\}$, which, combined with Lemma II.4, gives (43). Finally, (41) follows from

$$\begin{aligned} \Delta &= \mathbf{P}_{\mathcal{S}_{\Psi_{11e}}^{\perp}}\{\Psi^e\} = \Psi^e - \begin{bmatrix} \mathbf{P}_{\mathcal{S}_{\Psi_{11e}}}\{\Psi_{1\star}\} \\ \mathbf{P}_{\mathcal{S}_{\Psi_{11e}}}\{\Psi_{2\star}\} \end{bmatrix} \\ &= \Psi^e - \begin{bmatrix} \Psi_{11e} \\ \mathbf{P}_{\mathcal{S}_{\Psi_{11e}}}\{\Psi_{21e}\} \end{bmatrix} = \Psi^e - \begin{bmatrix} \Psi_{11e} \\ Y\Psi_{11e} \end{bmatrix} \end{aligned}$$

where the last equality is due to Lemma II.3. ∎

One case of particular interest is when $\Psi_{12} = \emptyset$ and $\Psi_{22} = \emptyset$, i.e., when only rows are added to $\Psi_{11}$ when $\Psi^e$ is formed. We

will return to this setup in Section IV-A; here we only notice that (41) simplifies to

$$\langle \Delta, \gamma \rangle = \begin{bmatrix} 0 \\ \langle \mathbf{P}_{\mathcal{S}^{\perp}_{\Psi_{11}}}\{\Psi_{21}\}, \gamma \rangle \end{bmatrix}$$

so that (35) and (43) give the condition $\mathrm{Rank}\ \langle \mathbf{P}_{\mathcal{S}^{\perp}_{\Psi_{11}}}\{\Psi_{21}\}, \gamma \rangle = 0$ for the difference $X$ in (31) to have zero rank.

*2) Information Increase:* Taking $\Psi_1 = \Psi_{11e}$ and $\Psi_2 = \Psi^e$ in Theorem III.1 corresponds to the case where there is an information increase when the structure is extended. From Theorem III.1 it follows that in this case it is the properties of the space $\mathcal{S}_{\Psi_{11e}} \ominus \mathcal{S}_{\Psi^e}$ that will be important. To this end, for any $\mathcal{L}_2$ function $f$ let $\widetilde{f}$ denote the projection of $f$ on $\mathcal{S}^{\perp}_{\Psi^e_{2\star}}$, e.g., $\widetilde{\Psi}_{12e} := \mathbf{P}_{\mathcal{S}^{\perp}_{\Psi^e_{2\star}}}\{\Psi_{12e}\}$.

*Theorem III.3:* This is an application of Theorem III.1 for $\Psi_1 = \Psi_{11e}$ and $\Psi_2 = \Psi^e$ according to (37)–(38). Let $\gamma \in \mathcal{L}_2^{p\times(m_1+m_2)}$ and let $\Psi_{ij} \in \mathcal{L}_2^{n_i\times m_j}, i,j = 1,2$ for some integers $p > 0$, $n_1 > 0$, $m_1 > 0$, $n_2 \geq 0$ and $m_2 \geq 0$. Suppose that $\gamma \perp \mathcal{S}_{\Psi^e} \ominus \mathcal{S}_{\Psi_{11e}}$ and let $\Delta$ be defined by (32). Suppose in addition that $\Psi_{21} \perp \Psi_{11}$. Then

$$\langle \Psi^e, \gamma \rangle = \begin{bmatrix} \langle \Psi_{11e}, \gamma \rangle \\ 0 \end{bmatrix}. \tag{44}$$

Furthermore

$$\mathbf{P}_{\mathcal{S}_\Delta}\{\gamma\} = \mathbf{P}_{\mathcal{S}_{\Psi_{11e}}}\{\gamma\} - \mathbf{P}_{\mathcal{S}_{\widetilde{\Psi}^e_{1\star}}}\{\gamma\} \tag{45}$$

and

$$\langle \Delta, \gamma \rangle = (I - Z)\langle \Psi_{11e}, \gamma \rangle \tag{46}$$

where $Z = \langle \Psi_{11}, \Psi_{11} \rangle \left[ \langle \Psi_{11}, \Psi_{11} \rangle + \langle \widetilde{\Psi}_{12e}, \widetilde{\Psi}_{12e} \rangle \right]^{\dagger}$.

Furthermore,

$$\dim \mathcal{S}_\Delta = \dim \mathcal{S}_{\Psi_{11}} + \dim \mathcal{S}_{\widetilde{\Psi}_{12e}} - \dim \mathcal{S}_{\widetilde{\Psi}^e_{1\star}} \tag{47}$$

and, finally $X$ defined in (31) satisfies

$$\mathrm{Rank} X \leq \min(\mathrm{Rank}(I - Z), \mathrm{Rank}\langle \Psi_{11e}, \gamma \rangle). \tag{48}$$

*Proof:* From the condition $\gamma \perp \mathcal{S}_{\Psi^e} \ominus \mathcal{S}_{\Psi_{11e}}$ it follows that $\langle \Psi^e, \gamma \rangle = \langle \mathbf{P}_{\mathcal{S}_{\Psi_{11e}}}\{\Psi^e\}, \gamma \rangle$ and from the orthogonality between $\Psi_{11}$ and $\Psi_{21}$ we get $\mathbf{P}_{\mathcal{S}_{\Psi_{11e}}}\{\Psi^e\} = \begin{bmatrix} \Psi_{11e} \\ 0 \end{bmatrix}$ which gives (44). The result (45) follows from (34) if we make the decomposition $\mathcal{S}_{\Psi^e} = \mathcal{S}_{\widetilde{\Psi}^e_{1\star}} \oplus \mathcal{S}_{\Psi^e_{2\star}}$ and use the orthogonality between $\gamma$ and $\Psi^e_{2\star}$. From the same decomposition and the fact that $\Psi_{11e} \perp \Psi^e_{2\star}$ we have

$$\begin{aligned} \Delta &= \mathbf{P}_{\mathcal{S}^{\perp}_{\Psi^e}}\{\Psi_{11e}\} = \Psi_{11e} - \mathbf{P}_{\mathcal{S}_{\widetilde{\Psi}^e_{1\star}}}\{\Psi_{11e}\} \\ &= \Psi_{11e} - \langle \Psi_{11e}, \widetilde{\Psi}^e_{1\star} \rangle \langle \widetilde{\Psi}^e_{1\star}, \widetilde{\Psi}^e_{1\star} \rangle^{\dagger} \widetilde{\Psi}^e_{1\star}. \end{aligned} \tag{49}$$

Now we observe that $\Psi^e_{1\star} = \Psi_{11e} + \Psi_{12e}$ so that $\widetilde{\Psi}^e_{1\star} = \widetilde{\Psi}_{11e} + \widetilde{\Psi}_{12e} = \Psi_{11e} + \widetilde{\Psi}_{12e}$. Furthermore, both $\gamma \perp \widetilde{\Psi}_{12e}$ and $\Psi_{11e} \perp \widetilde{\Psi}_{12e}$ since they are orthogonal to both $\Psi_{12e}$ and $\Psi^e_{2\star}$. Inserting this in (49) gives (46) and then (48) follows from a standard rank inequality $\mathrm{Rank}\,(AB) \leq \min(\mathrm{Rank}\,(A), \mathrm{Rank}\,(B))$.

Finally, (47) follows from the following decompositions:

$$\begin{aligned} \mathcal{S}_\Delta \oplus \mathcal{S}_{\Psi^e} &= \mathcal{S}_{\Psi_{11e}} + \mathcal{S}_{\Psi^e} = \mathcal{S}_{\Psi_{11e}} \oplus \mathcal{S}_{\widetilde{\Psi}_{12e}} \oplus \mathcal{S}_{\Psi^e_{2\star}} \\ \mathcal{S}_{\Psi^e} &= \mathcal{S}_{\widetilde{\Psi}^e_{1\star}} \oplus \mathcal{S}_{\Psi^e_{2\star}} \end{aligned}$$

which gives that $\mathcal{S}_\Delta \oplus \mathcal{S}_{\widetilde{\Psi}^e_{1\star}} = \mathcal{S}_{\Psi_{11e}} \oplus \mathcal{S}_{\widetilde{\Psi}_{12e}}$. ∎

The expressions (44)–(48) can be used in Theorem III.1 to quantify the decrease in the asymptotic variance when $\Psi$ is extended from $\Psi_{11}$ to $\Psi^e$. When $\langle \Psi_{11}, \Psi_{11} \rangle > 0$, a simple interpretation of the rank result (47) in Theorem III.3 can be given. Then in (47) and (48) we get

$$\mathrm{Rank}(I - Z) = \dim \mathcal{S}_\Delta = \dim \mathcal{S}_{\widetilde{\Psi}_{12e}}.$$

However, $\mathcal{S}_{\Psi_{12e}} + \mathcal{S}_{\Psi^e_{2\star}} = \mathcal{S}_{\Psi^e_{2\star}} \oplus \mathcal{S}_{\widetilde{\Psi}_{12e}}$ so

$$\dim \mathcal{S}_{\widetilde{\Psi}_{12e}} = \dim(\mathcal{S}_{\Psi_{12e}} + \mathcal{S}_{\Psi^e_{2\star}}) - \dim \mathcal{S}_{\Psi^e_{2\star}}$$

and we arrive at the interpretation that $\dim \mathcal{S}_{\Psi_{11e}} \ominus \mathcal{S}_{\Psi^e}$ equals the degrees of freedom offered by adding $\begin{bmatrix} 0 & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}$ to $\Psi_{11}$ reduced by the degrees of freedom $(\dim \mathcal{S}_{\Psi^e_{2\star}})$ that have to be used for the new rows that have been added. In particular, $\mathcal{S}_\Delta = 0 \Leftrightarrow S_{\Psi_{12e}} \subseteq S_{\Psi^e_{2\star}}$. Furthermore, since $\dim \mathcal{S}_{\Psi^e_{2\star}} \leq n_2$ and $\dim \mathcal{S}_{\Psi_{12e}} \leq n_1$ (recall that $n_1$ and $n_2$ are the number of rows of $\Psi_{12e}$ and $\Psi^e_{2\star}$, respectively), it follows that $\dim \mathcal{S}_\Delta = \dim \mathcal{S}_{\Psi_{11}} = n_1$ when $\dim(\mathcal{S}_{\Psi_{12e}} + \mathcal{S}_{\Psi^e_{2\star}}) = n_1 + n_2$. We summarize this.

*Lemma III.4:* Let the notation and assumptions in Theorem III.3 be in force as well as the condition $\langle \Psi_{11}, \Psi_{11} \rangle > 0$.

Then

i) $\mathcal{S}_\Delta = 0 \Leftrightarrow S_{\Psi_{12e}} \subseteq S_{\Psi^e_{2\star}}$

ii) $\dim \mathcal{S}_\Delta = \dim \mathcal{S}_{\Psi_{11}} = n_1$ when $\dim(\mathcal{S}_{\Psi_{12e}} + \mathcal{S}_{\Psi^e_{2\star}}) = n_1 + n_2$.

Notice that i) in Lemma III.4 is a simple condition for when there will be no information increase no matter what $\gamma$ is.

### *D. How to Use the Results to Compare Asymptotic Covariance Matrices*

The intended use of the results above is to compare covariances given by $\Lambda_i^T \langle \Psi_i, \Psi_i \rangle^{\dagger} \Lambda_i, i = 1,2$. Of course, a crucial aspect for the usefulness of Theorem III.1 is how easy it is to establish the existence of a $\gamma$ so that $\Lambda_i = \langle \Psi_i, \gamma \rangle$ is satisfied for pre-specified $\{\Lambda_1, \Psi_1\}$ and $\{\Lambda_2, \Psi_2\}$. Following the discussion around the expression (8) in the Introduction, we require that $\Lambda_1$ and $\Lambda_2$ belong to the column spaces of $\langle \Psi_1, \Psi_1 \rangle$ and $\langle \Psi_2, \Psi_2 \rangle$, respectively, which also ensures the existence of functions $\gamma_1$ and $\gamma_2$ such that $\Lambda_1 = \langle \Psi_1, \gamma_1 \rangle$ and $\Lambda_2 = \langle \Psi_2, \gamma_2 \rangle$. Lemma II.8 gives the general solutions

$$\gamma_i = \Lambda_i^* \langle \Psi_i, \Psi_i \rangle^{\dagger} \Psi_i + s_i^{\perp}, \quad s_i^{\perp} \perp \mathcal{S}_{\Psi_i}, \quad i = 1,2. \tag{50}$$

When applying Theorem III.1 it is required that $\gamma_1 = \gamma_2 = \gamma$ and also that $\gamma \perp \mathcal{S}_{\Psi_2} \ominus \mathcal{S}_{\Psi_1}$ and that limits the possible

combinations of $\Lambda_1$ and $\Lambda_2$. The latter condition is equivalent to

$$\langle \Psi_2 - \mathbf{P}_{\mathcal{S}_{\Psi_1}}\{\Psi_2\}, \gamma\rangle = 0 \Leftrightarrow \langle \Psi_2, \gamma\rangle = \langle \Psi_2, \Psi_1\rangle\langle \Psi_1, \Psi_1\rangle^{\dagger}\langle \Psi_1, \gamma\rangle$$

which implies

$$\Lambda_2 = \langle \Psi_2, \Psi_1\rangle\langle \Psi_1, \Psi_1\rangle^{\dagger}\Lambda_1 \tag{51}$$

and hence while $\Lambda_1$ can be chosen freely, $\Lambda_2$ is directly determined by $\Lambda_1$. The existence of a solution to (50) for $\gamma_1 = \gamma_2 = \gamma$ follows from (51) since we then have

$$\begin{bmatrix}\Lambda_1\\ \Lambda_2\end{bmatrix} = \begin{bmatrix}\langle \Psi_1, \Psi_1\rangle & \langle \Psi_1, \Psi_2\rangle\\ \langle \Psi_2, \Psi_1\rangle & \langle \Psi_2, \Psi_2\rangle\end{bmatrix}\begin{bmatrix}\langle \Psi_1, \Psi_1\rangle^{\dagger}\Lambda_1\\ 0\end{bmatrix}$$

and the general solution to the joint problem $\Lambda_i = \langle \Psi_i, \gamma\rangle, i = 1, 2$ is given by

$$\gamma = \Lambda_1\langle \Psi_1, \Psi_1\rangle^{\dagger}\Psi_1 + s^{\perp}, \quad s^{\perp} \perp \mathcal{S}_{\Psi_1} + \mathcal{S}_{\Psi_2}. \tag{52}$$

That $\gamma$ cannot be found so that $\Lambda_i = \langle \Psi_i, \gamma\rangle$ is satisfied for arbitrary $\Lambda_1$ and $\Lambda_2$ may seem as a severe limitation. However, it is quite natural since for given $\Psi_1$ and $\Psi_2$

$$\Lambda_1^T\langle \Psi_1, \Psi_1\rangle^{\dagger}\overline{\Lambda}_1 - \Lambda_2^T\langle \Psi_2, \Psi_2\rangle^{\dagger}\overline{\Lambda}_2 \geq 0$$

does not hold for arbitrary $\Lambda_1$ and $\Lambda_2$.

One special case which sometimes is studied is when $\Psi_2 = \alpha\Psi_1$ for some $\alpha \in \mathbb{C}^{n_2\times n_1}$, in which case it is required that $\Lambda_2 = \langle \Psi_2, \gamma\rangle = \alpha\langle \Psi_1, \gamma\rangle = \alpha\Lambda_1$. Condition (51) is then automatically fulfilled

$$\begin{aligned}\langle \Psi_2, \Psi_1\rangle\langle \Psi_1, \Psi_1\rangle^{\dagger}\Lambda_1 &= \alpha\langle \Psi_1, \Psi_1\rangle\langle \Psi_1, \Psi_1\rangle^{\dagger}\Lambda_1\\ &= \alpha\Lambda_1 = \Lambda_2.\end{aligned}$$

An example of this is in Theorem III.2 in the case when $\Psi_1 = \begin{bmatrix}\Psi_{11}\\ \Psi_{21}\end{bmatrix}$, $\Psi_2 = \Psi_{21}$, $\Lambda_1 = \begin{bmatrix}\Lambda_{11}\\ \Lambda_{21}\end{bmatrix}$ and $\Lambda_2 = \Lambda_{11}$ i.e., when $\alpha = [0 \quad I]$. Thus, when only rows are added to $\Psi_{11}$, Theorem III.1 and Theorem III.2 can be used to compare the covariances when $\Lambda_1$ is an extension of $\Lambda_2$.

Another example of how the relations between $\Lambda_1$ and $\Lambda_2$ are restricted is in Theorem III.3 where we have the condition (44) which implies $\Lambda_2 = \begin{bmatrix}\Lambda_1\\ 0\end{bmatrix} = \begin{bmatrix}I\\ 0\end{bmatrix}\Lambda_1$.

## IV. Structural Results for LTI Systems

We will now use the results in the preceding sections to analyze some structural issues that arise when identifying LTI systems.

### A. Influence of Model Complexity

For simplicity we will assume that the true system is of FIR type, i.e., the system is described by (5), with noise variance $\lambda_o = 1$, and that it is operated in open loop. We recall that the asymptotic covariance matrix of the parameter estimates is given by (2) with $\Psi$ given by (7).

Consider the problem of estimating a real-valued non-minimum phase (NMP) zero located at $z = z_o$. We are interested in understanding how the asymptotic variance of the corresponding zero estimate, obtained from an FIR model, depends on the model complexity, as represented by the order $n$ which here is assumed to be larger than or equal to the system order $n_o$, (meaning that the true system always is in the model set). It is well known that the asymptotic variances of individual parameter estimates do not decrease when the model order is increased, compare with the discussion in Section III-A. The same actually holds for the variance of a zero estimate, but this fact is not immediately evident, (at least not to us), since a zero depends on all parameters in a rather complicated way. The usual way to show this is by matrix algebra using Schur complements, a routine that will be explained in Section IV-A-I.

We will also ask the question whether there may be experimental conditions such that the asymptotic variance becomes independent of the model order $n$. This issue is of independent interest since the system order is typically not known beforehand. We will analyze these two questions using both matrix algebra and the geometric approach presented in this paper.

It is straightforward to show that the asymptotic variance of the estimate of a zero at $z = z_o$ is given by (8) with $\Lambda = c\Gamma_n(z_o)$ (recall that $\Gamma_n$ is defined in Example 2) where the constant $c$, which will be of no concern for us, depends on the true system but not on the model order $n$, see [61], [62] for details. The asymptotic variance of the estimate of $z_o$ is thus given by

$$\mathrm{AsVar}\hat{z}_o = c^2\Gamma_n^*(z_o)\langle \Gamma_n R_u, \Gamma_n R_u\rangle^{-1}\Gamma_n(z_o) \tag{53}$$

where $R_u$ is a stable minimum-phase spectral factor of the input. Introducing $\Gamma_{n,m}(\mathrm{q}) = [\mathrm{q}^{-(n+1)} \quad \dots \quad \mathrm{q}^{-m}]^T$, the asymptotic variances for model orders $n$ and $m(> n)$ are given by

$$\Gamma_n^*(z_o)\langle \Gamma_n R_u, \Gamma_n R_u\rangle^{-1}\Gamma_n(z_o) = \Gamma_n^*(z_o)A^{-1}\Gamma_n(z_o) \tag{54}$$

and

$$\begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix}^*\left\langle \begin{bmatrix}\Gamma_n\\ \Gamma_{n,m}\end{bmatrix}R_u, \begin{bmatrix}\Gamma_n\\ \Gamma_{n,m}\end{bmatrix}R_u\right\rangle^{-1}\begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix} = \begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix}^*\begin{bmatrix}A & B\\ B^T & C\end{bmatrix}^{-1}\begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix} \tag{55}$$

respectively, with obvious definitions of the matrices $A$, $B$ and $C$.

*1) Analysis Using Matrix Algebra:* Using a standard expression for the inverse of a block matrix (63), we can write

$$\begin{bmatrix}A & B\\ B^T & C\end{bmatrix}^{-1} = \begin{bmatrix}A^{-1} & 0\\ 0 & 0\end{bmatrix} + \begin{bmatrix}A^{-1}BS_A^{-1}B^TA^{-1} & -A^{-1}BS_A^{-1}\\ -S_A^{-1}B^TA^{-1} & S_A^{-1}\end{bmatrix} \tag{56}$$

where $S_A = C - B^TA^{-1}B$ is a Schur complement. However, the block matrix in the last term of (56) can be factorized. This

gives the following expression for the variance when the model order is $m$

$$\Gamma_n^*(z_o)A^{-1}\Gamma_n(z_o)+\begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix}^*\begin{bmatrix}A^{-1}B\\ -I\end{bmatrix}\times S_A^{-1}\begin{bmatrix}A^{-1}B\\ -I\end{bmatrix}^T\begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix}. \quad (57)$$

Since the first term in (57) corresponds to the asymptotic variance when the model order is $n$ and the second term is non-negative definite, we have shown that the variance is non-decreasing with the model order.

We now turn to the question of whether the variance can be independent of the model order $n$. For this to hold, the second term in (57) has to be identically zero, and this has to hold regardless of $m$. This implies the condition

$$\begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix}^*\begin{bmatrix}A^{-1}B\\ -I\end{bmatrix}=0 \quad (58)$$

which can be expressed as

$$\Gamma_n^*(z_o)R_n^{-1}R_{n,m}=\Gamma_{n,m}^*(z_o) \quad (59)$$

where $R_n$ and $R_{n,m}$ are Toeplitz matrices

$$R_n=\begin{bmatrix}r_0 & \dots & r_{n-1}\\ \vdots & \ddots & \vdots\\ r_{n-1} & \dots & r_0\end{bmatrix},R_{n,m}=\begin{bmatrix}r_n & \dots & r_{m-1}\\ \vdots & \ddots & \vdots\\ r_1 & \dots & r_{m-n}\end{bmatrix}$$

where $r_k:=\mathbf{E}[u_tu_{t-k}]$. Thus, if the second order properties of the input are such that (59) is satisfied for any $m(>n)$ then the asymptotic variance is independent of the model order $n$.

We now outline how one can determine if there exist $r_0,\dots,r_{m-1}$ such that (59) is satisfied. First we determine all possible $F\in\mathbb{R}^{n\times(m-n)}$ such that

$$\Gamma_n^*(z)F=\Gamma_{n,m}^*(z) \quad (60)$$

Note that (60) defines $m-n$ constraints on the $n(m-n)$ parameters in $F$. The set of allowable parameters of $F$:s can then be parametrized by an $n(m-n)-(m-n)$ dimensional subspace of $\mathbb{R}^{n(m-n)}$. Using this parametrization and solving the linear system of equations

$$R_nF=R_{n,m} \quad (61)$$

for all $F:s$ in this subspace with respect to $r_0,\dots,r_{m-1}$ will produce all possible (if any) solutions $r_0,\dots,r_{m-1}$ to (59). For a fixed $F$, the relation (61) defines $n(m-n)$ constraints on the $m$ parameters of the matrices $R_n$ and $R_{n,m}$.

Compounding factors when solving this problem are that only those solutions which correspond to *bona fide* autocorrelation sequences are allowed and, furthermore, that we are looking for solutions which are valid regardless of the particular value of $m$. All in all, this gives rise to a rather complex problem and even though the solution can be obtained we will not make the complete derivation. Instead we will analyze the problem using the geometric tools that have been presented.

*2) Geometric Analysis:* Lemma II.8 gives that there exists $\gamma$ such that

$$\langle\Gamma_mR_u,\gamma\rangle=\begin{bmatrix}\Gamma_n(z_o)\\ \Gamma_{n,m}(z_o)\end{bmatrix}. \quad (62)$$

But with $\Psi_1=\Gamma_mR_u$ and $\Psi_2=\Gamma_nR_u$, then (33) in Theorem III.1 gives that the asymptotic variance for the zero estimate is non-decreasing with model order and that the difference between (55) and (54) is given by

$$\|\mathbf{P}_{\mathcal{S}_{\Psi_1}\ominus\mathcal{S}_{\Psi_2}}\{\gamma\}\|^2. \quad (63)$$

Notice that we did not have to compute $\gamma$ explicitly in order to arrive at this result and that the result holds for all $n$ and $m$, $m>n$.

We now turn to the question of whether the variance can be independent of the model order $n$. We notice that if $\gamma$ can be taken to belong to $\mathcal{S}_{\Psi_2}$ then the projection of $\gamma$ on $\mathcal{S}_{\Psi_1}\ominus\mathcal{S}_{\Psi_2}$ will be zero. For example we can examine if $\gamma(z)=R_u(z)z^{-1}$ (the first element of $\Psi_2$, which clearly belongs to $\mathcal{S}_{\Psi_2}$) is possible.

The conditions (62) that $\gamma$ has to satisfy can be expressed as

$$\begin{aligned}z_o^{-l}&=\langle z^{-l}R_u(z),\gamma(z)\rangle\\ &=\langle z^{-l}R_u(z),z^{-1}R_u(z)\rangle\\ &=\langle R_u(z^{-1})R_u(z),z^{l-1}\rangle\\ &=\langle\Phi_u(z),z^{l-1}\rangle,\quad l=1,\dots m.\end{aligned}$$

Notice that these are constraints on the $m$ first coefficients in the series expansion of the input spectrum. For example, the input spectrum

$$\begin{aligned}\Phi_u(z)&=z_o^{-1}\sum_{l=-\infty}^{\infty}z_o^{-|l|}z^{-l}\\ &=z_o^{-1}\frac{1-z_o^{-2}}{(1-z_o^{-1}z^{-1})(1-z_o^{-1}z)}\end{aligned} \quad (64)$$

satisfies (62). In fact, this spectrum satisfies (62) regardless of the order $m$.

Thus, using the geometric approach, we have with little effort shown that the asymptotic variance of an estimate of a NMP zero does not depend on the model complexity if the input spectrum has a single pole at the zero itself (and at its mirror image in the unit circle). The intuition is that this input spectrum provides an infinite signal to noise ratio *at the zero in question*, since $\Phi_u(z_o)=\infty$, which makes the estimate independent of the system complexity. Compare this with the well known fact that a sinusoid with frequency $\omega$ results in the asymptotic variance of the frequency function estimate at $\omega$ being independent of the system complexity, see, e.g., [8]. The analysis presented here can be generalized and we refer the reader to [64]–[66] for details. The result is illustrated with a simulation example.

*3) Example 3:* In this example the identification of the zero $z_o=1.67$ of the system $G_o(z)=1-z_oz^{-1}$ is considered. The model is $G(z)=b_o+b_1z^{-1}+\dots+b_nz^{-n}$ where $n=1$, 2, 3, 10. White noise with variance 0.01 is added to the output and the input is a unit variance white noise sequence, filtered by $F(z)=k(f)/(1-fz^{-1})$. The pole $f$ of the filter is varied between

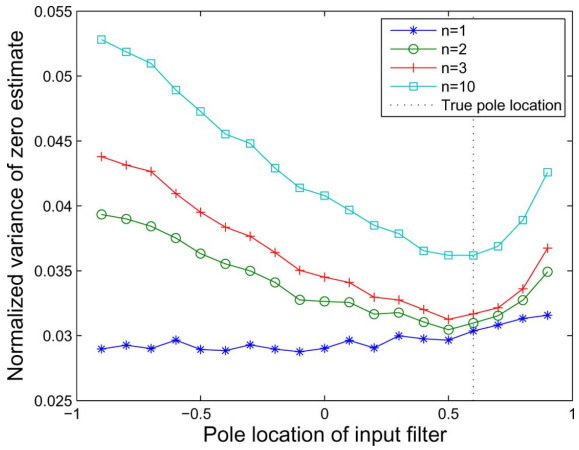


Fig. 1. Normalized sample variance from the numerical simulations of Example 3.

$-0.9$ and 0.9 in steps of 0.1 and the value corresponding to (64) is $f = 0.6$. Sequences of length $N = 100$ of the two white inputs are drawn from a Gaussian distribution and the output is simulated for the different pole locations $f$. For each data set the four models with different $n$ are estimated. The simulation is repeated 25.000 times with different noise realizations. In Fig. 1 the sample variance of the estimated zero, normalized by $N$, is plotted against the pole location $f$ of the input filter. The gain $k(f)$ of the input filter is adjusted so that the asymptotic variance of the zero, see (53), is equal for all $f$ when $n = 1$. When $n$ increases to 10, the normalized sample variance increases with as much as 80% (for $f = -0.9$), but when $f = 0.6$ the increase is only 20%. The asymptotic theory says that there should be no increase at all for $f = 0.6$. This is not true when the sample size is small (as in this example), but note that the minimum variance is achieved for $f = 0.6$. This example shows that the asymptotic results also are useful when the sample size is small.

### B. Influence of Inputs

Next we will consider the problem of when adding inputs may increase the information. The following result shows that there are some limitations.

*Theorem IV.1:* Let

$$\Psi_{11} = \begin{bmatrix} \Phi_{11} \\ \Phi_{21} \end{bmatrix}, \quad \Psi_{12} = \begin{bmatrix} 0_{n_1 \times m_2} \\ \Phi_{22} \end{bmatrix}, \quad \Psi^e = [\,\Psi_{11} \quad \Psi_{12}\,]$$

where $\Phi_{ij} \in \mathcal{L}_2^{n_i \times m_j}, i, j = 1, 2$. Suppose that $\langle \Psi_{11}, \Psi_{11} \rangle > 0$ and $\langle \Phi_{22}, \Phi_{22} \rangle > 0$. Let $\Lambda = [\,I_{n_1 \times n_1} \quad 0_{n_1 \times n_2}\,]^T$. Then

$$\Lambda^T \langle \Psi_{11}, \Psi_{11} \rangle^{-1} \Lambda - \Lambda^T \langle \Psi^e, \Psi^e \rangle^{-1} \Lambda \geq 0 \tag{65}$$

and the rank of the left-hand side is given by

$$\dim \mathcal{S}_{\mathbf{P}_{\mathcal{S}}[\Phi_{11} \quad 0_{n_1 \times m_2}]\{[\Phi_{21} \quad \Phi_{22}]\}} = \mathrm{Rank} \langle \Phi_{21}, \Phi_{11} \rangle. \tag{66}$$

*Proof:* See Appendix. ∎

We illustrate the use of Theorem IV.1 with an example.

*1) Example 4:* Consider the linear regression model

$$y_t = \theta_1 u_{t-1} + \theta_2 (u_{t-2} + w_{t-1}) + e_t \tag{67}$$

where $u$ is stationary. The question now is if the use of a second input $w$, which we assume also to be stationary and uncorrelated with $u$, will reduce the asymptotic variance of the estimate of $\theta_1$.

Denoting by $R_u$ and $R_w$, the stable minimum phase spectral factors of the spectra of $u$ and $w$, we have that the asymptotic covariance matrix of the estimate of $\theta = [\,\theta_1 \quad \theta_2\,]$ is given by (2) with

$$\Psi(z) = \Psi_{11}(z) := \begin{bmatrix} R_u(z) z^{-1} \\ R_u(z) z^{-2} \end{bmatrix}$$

when only $u$ is used, and by

$$\Psi(z) = \Psi^e(z) := \begin{bmatrix} R_u(z) z^{-1} & 0 \\ R_u(z) z^{-2} & R_w(z) z^{-1} \end{bmatrix}$$

when also $w$ is used. Identifying $\Phi_{11}(z) = R_u(z) z^{-1}$ and $\Phi_{21}(z) = R_u(z) z^{-2}$, Theorem IV.1 gives that the asymptotic variance of the estimate of $\theta_1$ will decrease when $w$ is used only if

$$0 \neq \langle R_u(z) z^{-1}, R_u(z) z^{-2} \rangle = \langle \Phi_u(z), z^{-1} \rangle = \mathbf{E}[u_t u_{t-1}]. \tag{68}$$

For example, if $u$ is white noise, there will be no improvement.

For the readers' benefit we provide the expression for $\gamma$ here. The comparison that is made is between $\|\mathbf{P}_{\mathcal{S}_{\Psi_{11e}}}\{\gamma\}\|^2$ and $\|\mathbf{P}_{\mathcal{S}_{\Psi^e}}\{\gamma\}\|^2$ for a $\gamma$ such that $\Lambda = [1\ 0]^T = \langle \Psi_{11e}, \gamma \rangle = \langle \Psi^e, \gamma \rangle$. From (52) we get $\gamma = \Lambda \langle \Psi_{11}, \Psi_{11} \rangle^{-1} \Psi_{11e} + s^\perp$ where $s^\perp$ is orthogonal to $\mathcal{S}_{\Psi_{11e}} + \mathcal{S}_{\Psi^e}$. If we let $r_k = \langle |R_u(z)|^2, z^{-k} \rangle$ we can express $\gamma$ as

$$\gamma = \frac{(r_0 z^{-1} - r_1 z^{-2}) R_u(z)}{r_0^2 - r_1^2} + s^\perp.$$

In [46] it is analyzed, for linear time-invariant MISO systems, when a particular input improves the accuracy of a certain parameter estimate. The rank result above gives more precise information regarding the possible improvements than can be obtained as compared with Theorem 4 in [46]. In particular, the example above illustrates that not only the connectedness of the transfer functions (see [46] for a definition) are important for when an added input will improve the estimate of a parameter not directly connected to it, also the input properties are important.

### C. Influence of Sensors

We will now turn to a problem that in a sense is dual to the problem studied in Section IV-B. Consider a system $G^o$ that can be modelled by an output error model

$$y_t = G(q, \theta) u_t + e_t \tag{69}$$

where $\theta \in \mathbb{R}^{1 \times n_\theta}$ and where $\{e_t\}$ is white noise with $e_t \sim \mathcal{N}(0, \lambda_e)$. The issue is now whether the accuracy of the estimated parameters $\hat{\theta}_N$ can be improved by adding a second

sensor to the system. The new sensor will give measurements according to

$$z_t = M^o(q)G^o(q)u_t + w_t \tag{70}$$

where $\{w_t\}$ is white noise and independent of $\{e_t\}$ with $w_t \sim \mathcal{N}(0, \lambda_w)$, and where $M^o$ represents unknown sensor dynamics that also have to be estimated. The sensor dynamics is modeled by

$$M(q, \beta) \tag{71}$$

where $\beta \in \mathbb{R}^{1\times n_\beta}$. As the signal $\{z_t\}$ contains information about $G$ it may seem obvious that a second sensor would improve the accuracy but let us analyze the problem using the tools we have developed in this paper.

We first study the case where only $y$ is used. Then the output error model (69) will have asymptotic parameter covariance given by (2) with

$$\Psi = \Psi_{11} := \frac{1}{\sqrt{\lambda_e}} R_u G'(\theta^o) \tag{72}$$

where $R_u$ is the spectral factor of the input.

When a second sensor is added and $\beta$ is estimated together with $\theta$, the maximum likelihood criterion is given by

$$\det \left\{ \sum_{t=1}^{N} \begin{bmatrix} y_t - G(\theta)u_t \\ z_t - M(\beta)G(\theta)u_t \end{bmatrix} \begin{bmatrix} y_t - G(\theta)u_t \\ z_t - M(\beta)G(\theta)u_t \end{bmatrix}^T \right\} \tag{73}$$

and it follows, see [67], that the asymptotic parameter covariance is given by (2) with

$$\Psi = \Psi^e := \begin{bmatrix} \frac{R_u}{\sqrt{\lambda_e}} G'(\theta^o) & \frac{R_u}{\sqrt{\lambda_w}} M^o G'(\theta^o) \\ 0_{n_\beta \times 1} & \frac{R_u}{\sqrt{\lambda_w}} G^o M'(\beta^o) \end{bmatrix}. \tag{74}$$

It is easy to verify that the condition (51) holds for $\Lambda_2 = [\Lambda_1^T \ 0]^T$, $\Psi_1 = \Psi_{11e} := [\Psi_{11} \ 0]$ with $\Psi_{11}$ as in (72) and $\Psi_2 = \Psi^e$ with $\Psi^e$ as in (74). This ensures that there exists a $\gamma \perp \mathcal{S}_{\Psi^e} \ominus \mathcal{S}_{\Psi_{11e}}$ such that $\Lambda_1 = \langle \Psi_{11e}, \gamma \rangle$ and $\Lambda_2 = \langle \Psi^e, \gamma \rangle$. The expression for $\gamma$ is given by (52)

$$\gamma = \langle \Psi_{11}, \Psi_{11} \rangle^{-1} \Psi_{11e} + s^\perp, \quad s^\perp \perp \mathcal{S}_{\Psi_{11e}} + \mathcal{S}_{\Psi^e}.$$

Now taking $\Psi_1 = \Psi_{11e}$ and $\Psi_2 = \Psi^e$ in Theorem III.1 (as in Section III-C-II) and then referring to i) in Lemma III.4 now gives the following result.

*Theorem IV.2:* Consider the system

$$y_t = G^o(q)u_t + e_t$$

where $\{e_t\}$ is white noise with $e_t \sim \mathcal{N}(0, \lambda_e)$.

Let the system be modelled by the output error model (69) which we assume to be globally identifiable. Assume also that the input signal is persistently exciting of order at least equal to the number of estimated parameters.

Adding the sensor (70), where $\{w_t\}$ is white noise with $w_t \sim \mathcal{N}(0, \lambda_w)$ with variance $\lambda_w > 0$ and independent of $\{e_t\}$ and where $M^o$ is unknown but can be modelled by (71), and estimating all unknown parameter using maximum likelihood estimation does not improve the asymptotic covariance for the estimate of $\theta$ as compared to only using measurements of $y$ if the elements of $M^o G'(\theta^o)$ are spanned by the elements of $G^o M'(\beta^o)$.

*Remark IV.3:* With some additional effort, the result of Theorem IV.2 can be extended to the singular case when $\lambda_w = 0$.

The condition in Theorem IV.2 holds, for example, when $M(\beta)$ has the same parametrization as $G(\theta)$ and $M^o$ is proportional to $G^o$. The perhaps surprising result from Theorem IV.2 is thus that in this situation the extra sensor does not provide any new information about $G^o$. Perhaps even more surprising is that this holds regardless of the noise variance for the second sensor, i.e., even very high quality measurements from the extra sensor are useless (as measured by the asymptotic covariance matrix) for estimating $G^o$ if also $y$ is used.

When the condition in Theorem IV.2 does not hold, there may still be estimates of certain properties of $G^o$ that does not improve when a second sensor is added. Drawing on the results above, variance analysis for cascade systems is presented in [68].

## V. Conclusions

We have in Theorem II.5 presented a geometric interpretation of the asymptotic covariance expression

$$\mathrm{AsCov} J(\hat{\theta}_N) = \Lambda^T [\langle \Psi, \Psi \rangle]^{-1} \overline{\Lambda}.$$

While all quantitative results concerning this expression, e.g., those presented in this paper, can equally well be obtained using matrix algebra such as Schur complements and the matrix inversion lemma, we believe that the geometrical interpretation can be useful for unveiling structural relationships. In Section IV-A this was illustrated by showing that a specific choice of input spectrum makes the asymptotic variance of a NMP zero estimate independent of the system complexity. For comparison purposes, we also presented a derivation based entirely on matrix algebra. It is up to the reader to decide which approach suits him/her. Furthermore, in Section IV-B we re-visited the problem of determining when new inputs can decrease the asymptotic covariance. Again using geometric arguments, we were able to show that the problem is more intricate than previously believed in [46]. We have also analyzed the dual problem of when adding sensors may help decrease the asymptotic covariance matrix (Section IV-C) and it was shown that there are situations in which high quality measurements may be useless. We believe that this problem and its generalizations will turn out to be important in distributed sensor networks. Our insights have already paved the way for results concerning cascade systems [68].

We point to two properties that we believe are of key importance in the geometric approach to variance analysis:

i) For qualitative structural analysis, it is often not necessary to compute $\gamma$ such that $\Lambda = \langle \Psi, \gamma \rangle$, it is sufficient to know that such a $\gamma$ exists. Notice that the only occasion in entire Section IV when we needed $\gamma$ was in Section IV-A

and then only when we wanted to explicitly determine the input spectrum.

ii) The freedom in the choice of $\gamma$ represented by $s^\perp$ in Lemma II.8. For some problems, where asymptotic covariance matrices for two (or more) identification setups (or structures) are to be compared, this allows the same $\gamma$ (which still may not have to be computed explicitly) to be used. Hence, it will only be the subspaces on which $\gamma$ is projected that differ. This will simplify the analysis. All results in Section III are based on this. For example, (62) shows that the same $\gamma$ can be used for analyzing the variance of a NMP zero up to some finite, but arbitrary high, system order.

Another use of the freedom represented by $s^\perp$ is that it can be used together with Lemma II.6 to obtain tractable explicit upper bounds. We have not explored this possibility in this contribution but refer to [58], [59] where model structure independent upper bounds are derived for a number of different estimated system properties when the underlying system is linear time invariant.

Our work has its origin in [53], where the importance of the subspace spanned by the prediction error gradient was recognized. Perhaps the most important contribution is the analysis technique as such. Hopefully the methodology will be of use in future research. It has already been used in [58] for deriving variance expressions for estimates obtained from single input single output causal finite dimensional LTI systems. We see as interesting applications to LTI multi-input and multi-output systems and non-linear systems, for which at present the available results on the structural properties of the asymptotic covariance matrix are limited. Experiment design is another area where the approach has potential, e.g., the result in Section IV-A has been generalized in [64]–[66].

## Appendix
## Proof of Theorem IV.1

It is easily verified that $\langle \Psi^e, \Psi_{11e}\rangle\langle \Psi_{11e}, \Psi_{11e}\rangle^{-1}\Lambda = \Lambda$ for any $\Lambda$ and Theorem III.3 shows that there exists a $\gamma$ such that $\Lambda_{11e} = \langle \Psi_{11e}, \gamma\rangle = \langle \Psi^e, \gamma\rangle$ where $\Lambda_{11e} := \begin{bmatrix} I_{n_1\times n_1} & 0_{n_1\times n_2}\end{bmatrix}^T$. Theorem III.1 gives (65) is positive and that the rank of the left-hand side of (65) is given by the rank of $\langle \gamma, \mathbf{P}_{\mathcal{S}^\perp_{\Psi^e}}\{\Psi_{11e}\}\rangle$ which in turn by (46) in Theorem III.3 equals the rank of $\langle \gamma, \mathbf{P}_{\mathcal{S}_{\Psi^e}}\{\Psi_{12e}\}\rangle$ where $\Psi_{12e} = \begin{bmatrix} 0_{n_1\times m_1} & \Psi_{12}\end{bmatrix}$.

Introduce $\Phi = \begin{bmatrix} \Phi_{11} & 0_{n_1\times m_2} \\ \Phi_{21} & \Phi_{22}\end{bmatrix}$, $\Phi_{21e} = \begin{bmatrix} \Phi_{21} & 0_{n_2\times m_2}\end{bmatrix}$ and $\Phi_{22e} = \begin{bmatrix} 0_{n_2\times m_1} & \Phi_{22}\end{bmatrix}$ and the notation $\widetilde{f} = \mathbf{P}_{\mathcal{S}^\perp_{\Phi_{1\star}}}\{f\}$. Next, observe that $W := \langle \widetilde{\Phi}_{2\star}, \widetilde{\Phi}_{2\star}\rangle > 0$ since $\widetilde{\Phi}_{2\star} = \widetilde{\Phi}_{21e} + \Phi_{22e}$ with these two terms being orthogonal and where, by assumption, $\langle \Phi_{22e}, \Phi_{22e}\rangle = \langle \Phi_{22}, \Phi_{22}\rangle > 0$. Decomposing $\mathcal{S}_{\Psi^e} = \mathcal{S}_{\Phi_{1\star}} \oplus \mathcal{S}_{\widetilde{\Phi}_{2\star}}$ and observing that $\Psi_{12e} \perp \Phi_{1\star}$ gives that

$$\begin{aligned}\mathbf{P}_{S_{\Psi^e}}\{\Psi_{12e}\} &= \mathbf{P}_{\mathcal{S}_{\widetilde{\Phi}^e_{2\star}}}\{\Psi_{12e}\} \\ &= \langle \Psi_{12e}, \widetilde{\Phi}_{2\star}\rangle W^{-1}\widetilde{\Phi}_{2\star} \\ &= \begin{bmatrix} 0_{n_1\times n_1} \\ \langle \Phi_{22}, \Phi_{22}\rangle\end{bmatrix} W^{-1}\widetilde{\Phi}_{2\star}\end{aligned}$$

where we have used the same decomposition of $\widetilde{\Phi}_{2\star}$ as above. Now, since $\widetilde{\Phi}_{2\star} = \Phi_{2\star} - \mathbf{P}_{\mathcal{S}_{\Phi_{1\star}}}\{\Phi_{2\star}\}$ and $\gamma \perp \Phi_{2\star}$ by assumption, we obtain

$$\langle \mathbf{P}_{S_{\Psi^e}}\{\Psi_{12e}\}, \gamma\rangle = -\begin{bmatrix} 0_{n_1\times n_2} \\ \langle \Phi_{22}, \Phi_{22}\rangle\end{bmatrix} W^{-1}\langle \mathbf{P}_{\mathcal{S}_{\Phi_{1\star}}}\{\Phi_{2\star}\}, \gamma\rangle.$$

Since $\langle \Phi_{22}, \Phi_{22}\rangle > 0$ and $\langle \Phi_{11}, \Phi_{11}\rangle^{-1} > 0$ by assumption, the rank of this expression equals the rank of $\langle \mathbf{P}_{\mathcal{S}_{\Phi_{1\star}}}\{\Phi_{2\star}\}, \gamma\rangle$ but since $\langle \Phi_{1\star}, \gamma\rangle = I_{n_1\times n_1}$ by assumption the rank of this quantity equals the rank of $\langle \mathbf{P}_{\mathcal{S}_{\Phi_{1\star}}}\{\Phi_{2\star}\}, \mathbf{P}_{\mathcal{S}_{\Phi_{1\star}}}\{\Phi_{2\star}\}\rangle$ which according to Lemmas II.1 and II.4 equals the right-most expression in (66).

## Acknowledgment

The authors would like to thank the six anonymous referees for many constructive comments. In particular the interpretation at the beginning of Section II-B is due to one of the reviewers. The first author wishes to thank Prof. Brett Ninness, colleague and friend, for a very rewarding collaboration throughout the last decade, it has been instrumental for arriving at the results in this paper.

## References

[1] J. Antoni and J. Schoukens, "A comprehensive study of the bias and variance of frequency-response-function measurements: Optimal window selection and overlapping strategies," *Automatica*, vol. 43, no. 10, pp. 1723–1736, 2007.

[2] X. Bombois, B. Anderson, and M. Gevers, "Mapping parametric confidence ellipsoids to Nyquist plane for linearly parametrized transfer functions," in *Model Identification and Adaptative Control*, G. Goodwin, Ed. New York: Springer Verlag, 2000, pp. 53–71.

[3] M. Campi and E. Weyer, "Finite sample properties of system identification methods," *IEEE Trans. Autom. Contr.*, vol. 47, no. 8, pp. 1329–1334, Aug. 2002.

[4] S. Douma and P. V. den Hof, "Relations between uncertainty structures in identification for robust control," *Automatica*, vol. 41, no. 3, pp. 439–457, 2005.

[5] M. Gevers, L. Ljung, and P. Van den Hof, "Asymptotic variance expressions for closed-loop identification," *Automatica*, vol. 37, no. 5, pp. 781–786, 2001.

[6] R. Hakvoort and P. Van den Hof, "Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors," *IEEE Trans. Autom. Contr.*, vol. 42, no. 11, pp. 1516–1528, Nov. 1997.

[7] R. Hildebrand and M. Gevers, "Quantification of the variance of estimated transfer functions in the presence of undermodelling," in *Proc. 13th IFAC Symp. System Identification*, Rotterdam, The Netherlands, 2003, pp. 1850–1855.

[8] H. Hjalmarsson, "From experiment design to closed loop control," *Automatica*, vol. 41, no. 3, pp. 393–438, Mar. 2005.

[9] L. Ljung, "Perspectives on system identification," in *Proc. 17th IFAC World Congress*, Seoul, Korea, Jul. 2008, pp. 7172–7184.

[10] L. Ljung and L. Guo, "The role of model validation for assessing the size of the unmodeled dynamics," *IEEE Trans. Autom. Cotntr.*, vol. 42, no. 9, pp. 1230–1239, 1997.

[11] M. Milanese and C. Novara, "Model quality in nonlinear SM identification," in *Proc. 42nd IEEE Conf. Decision Control*, Maui, HI, Dec. 2003, pp. 6021–6026.

[12] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica*, vol. 27, no. 6, pp. 997–1009, 1991.

[13] M. Milanese and A. Vicino, "$H_\infty$ set membership identification: A survey," *Automatica*, vol. 41, no. 12, pp. 2019–2032, 2005.

[14] B. Ninness, "On the CRLB for combined model and model order estimation of stationary stochastic processes," *IEEE Signal Process. Lett.*, vol. 11, pp. 293–296, Feb. 2004.

[15] B. Ninness and G. Goodwin, "Estimation of model quality," *Automatica*, vol. 31, no. 12, pp. 32–74, 1995.

[16] B. Ninness and S. Henriksen, "System identification via a computational Bayesian approach," in *Proc. 13th IFAC Symp. System Identification*, Rotterdam, The Netherlands, 2003, pp. 785–790.

[17] B. Ninness and H. Hjalmarsson, "On the frequency domain accuracy of closed-loop estimates," *Automatica*, vol. 41, no. 7, pp. 1109–1122, 2005.

[18] R. Pintelon, P. Guillaume, and J. Schoukens, "Uncertainty calculation in (operational) modal analysis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2359–2373, 2007.

[19] R. Pintelon, J. Schoukens, and Y. Rolain, "Uncertainty of transfer function modelling using prior estimated noise models," *Automatica*, vol. 39, no. 10, pp. 1721–1733, 2003.

[20] W. Reinelt, A. Garulli, and L. Ljung, "Comparing different approaches to model error modeling in robust identification," *Automatica*, vol. 38, no. 5, pp. 787–803, May 2002.

[21] B. Wahlberg and L. Ljung, "Hard frequency-domain model error bounds from least-squares like identification techniques," *IEEE Trans. Autom. Contr.*, vol. 37, no. 7, pp. 900–912, Jul. 1992.

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

[23] V. Fedorov, *Theory of Optimal Experiments*, ser. Probability and Mathematical Statistics. New York: Academic, 1972, vol. 12.

[24] G. Goodwin and R. Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, ser. Mathematics in Science and Engineering. New York: Academic, 1977, vol. 136.

[25] R. Mehra, "Optimal input signals for parameter estimation in dynamic systems—Survey and new results," *IEEE Trans. Autom. Contr.*, vol. 19, no. 12, pp. 753–768, Dec. 1974.

[26] M. Zarrop, *Optimal Experiment Design for Dynamic System Identification*. Berlin, Germany: Springer Verlag, 1979, Lecture Notes in Control and Information Sciences. Sci. 21.

[27] L. Ljung, "Asymptotic variance expressions for identified black-box transfer function models," *IEEE Trans. Autom. Contr.*, vol. 30, no. 9, pp. 834–844, Sep. 1985.

[28] U. Forssell and L. Ljung, "Some results on optimal experiment design," *Automatica*, vol. 36, no. 5, pp. 749–756, May 2000.

[29] M. Gevers and L. Ljung, "Optimal experiment designs with respect to the intended model application," *Automatica*, vol. 22, pp. 543–554, 1986.

[30] H. Hjalmarsson, M. Gevers, and F. D. Bruyne, "For model-based control design, closed-loop identification gives better performance," *Automatica*, vol. 32, no. 12, pp. 1659–1673, 1996.

[31] Z. Yuan and L. Ljung, "Unprejudiced optimal open loop input design for identification of transfer functions," *Automatica*, vol. 21, no. 6, pp. 697–708, 1985.

[32] G. Belforte and P. Gay, "Optimal experiment design for regression polynomial models identification," *Int. J. Contr.*, vol. 75, no. 15, pp. 1178–1189, Oct. 2002.

[33] G. Franceschini and S. Macchietto, "Model-based design of experiments for parameter precision: State of the art," *Chem. Eng. Sci.*, vol. 63, no. 19, pp. 4846–4872, 2008.

[34] C. Jauberthie, L. Denis-Vidal, P. Coton, and G. Joly-Blanchard, "An optimal input design procedure," *Automatica*, vol. 42, no. 5, pp. 881–884, 2006.

[35] L. Pronzato, "Adaptive optimization and D-optimum experimental design," *Ann. Statist.*, vol. 28, no. 6, pp. 1743–1761, 2000.

[36] L. Pronzato, "Optimal experimental design and some related control problems," *Automatica*, vol. 44, no. 2, pp. 303–325, 2008.

[37] 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.

[38] Y. Zhu, *Multivariable System Identification for Process Control*. Oxford, U.K.: Elsevier Science, 2001.

[39] Y. Zhu, "System identification for process control: recent experience and outlook," *Int. J. Modelling, Identification Contr.*, vol. 6, no. 2, pp. 89–103, 2009.

[40] P. Celaya, R. Tkatch, Y. Zhu, and R. Patwardhan, "Closed-loop identification at the Hovensa refinery," presented at the NPRA Plant Autom. Decision Support Conf., San Antonio, TX, 2004.

[41] P. Zehna, "Invariance of maximum likelihood estimators," *Ann. Math. Statist.*, vol. 37, no. 755, pp. 744–744, 1966.

[42] H. Hjalmarsson, "System identification of complex and structured systems," *Eur. J. Control/Plenary Address Eur. Contr. Conf.*, vol. 15, no. 3–4, pp. 275–310, May–Aug. 2009.

[43] P. Stoica and T. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, pp. 87–90, Jan. 2001.

[44] G. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*. Oakland, CA: Holden-Day, 1976.

[45] I. Gustavsson, L. Ljung, and T. Söderström, "Identification of processes in closed-loop—Identifiability and accuracy aspects," *Automatica*, vol. 13, no. 1, pp. 59–75, 1977.

[46] M. Gevers, L. Miskovic, 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.

[47] M. Gevers, A. Bazanella, X. Bombois, and L. Miskovic, "Identification and the information matrix: How to get just sufficiently rich?," *IEEE Trans. Autom. Contr.*, vol. 54, no. 12, pp. 2828–2840, Dec. 2009.

[48] B. Ninness and S. Gibson, "Quantifying the accuracy of Hammerstein model estimation," *Automatica*, vol. 38, no. 12, pp. 2037–2051, 2002.

[49] M. Campi, S. Ooi, and E. Weyer, "Non-asymptotic quality assessment of generalised FIR models with periodic inputs," *Automatica*, vol. 40, no. 12, pp. 2029–2041, 2004.

[50] M. Campi and E. Weyer, "Guaranteed non-asymptotic confidence regions in system identification," *Automatica*, vol. 41, no. 10, pp. 1751–1764, 2005.

[51] S. Douma and P. Van den Hof, "Probabilistic model uncertainty bounding: An approach with finite-time perspectives," in *Proc. 4th IFAC Symp. System Identification*, Newcastle, Australia, Mar. 27–29, 2006, pp. 1021–1026.

[52] S. Garatti, M. Campi, and S. Bittanti, "Assessing the quality of identified models through the asymptotic theory—When is the result reliable?," *Automatica*, vol. 40, no. 8, pp. 1319–1332, 2004.

[53] B. Ninness and H. Hjalmarsson, "Variance error quantifications that are exact for finite model order," *IEEE Trans. Autom. Contr.*, vol. 49, pp. 1275–1291, 2004.

[54] L. Scharf and L. McWorter, "Geometry of the Cramer-Rao bound," *Signal Process.*, vol. 31, no. 3, pp. 301–311, 1993.

[55] E. Dilaveroglu, "Simple general expression for Cramér-Rao bound in presence of nuisance parameters," *Electron. Lett.*, vol. 38, no. 25, pp. 1750–1751, 2002.

[56] H. Hjalmarsson and J. Mårtensson, "A geometric approach to variance analysis in system identification: Theory and nonlinear systems," in *Proc. 46th IEEE Conf. Decision Contr.*, New Orleans, LA, Dec. 2007, pp. 5092–5097.

[57] A. Friedman, *Foundations of Modern Analysis*. New York: Dover, 1982.

[58] J. Mårtensson and H. Hjalmarsson, "Variance analysis in SISO linear systems identification," *IEEE Trans. Autom. Contr.*, 2011, provisionally accepted for publication.

[59] J. Mårtensson and H. Hjalmarsson, "A geometric approach to variance analysis in system identification: Linear time-invariant systems," in *46th IEEE Conf. Decision Contr.*, Dec. 2007, pp. 4269–4274.

[60] H. Hjalmarsson, "Aspects on incomplete modeling in system identification," Ph.D. dissertation, Linköping University, Sweden, Sweden, 1993.

[61] K. Lindqvist and H. Hjalmarsson, "Identification of performance limitations in control," in *Proc. Eur. Contr. Conf.*, Porto, Portugal, 2001, pp. 1446–1451.

[62] J. Mårtensson and H. Hjalmarsson, "Variance error quantification for identified poles and zeros," *Automatica*, vol. 45, no. 11, pp. 2512–2525, Nov. 2009.

[63] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[64] H. Hjalmarsson, J. Mårtensson, and B. Wahlberg, "On some robustness issues in input design," in *Proc. 14th IFAC Symp. Syst. Identification*, Mar. 2006, pp. 511–516.

[65] J. Mårtensson and H. Hjalmarsson, "How to make bias and variance errors insensitive to system and model complexity in identification," *IEEE Trans. Autom. Contr.*, vol. 56, no. 1, pp. 100–112, 2011.

[66] J. Mårtensson, Geometric Analysis of Stochastic Model Errors in System Identification KTH, Stockholm, Sweden, 2007 [Online]. Available: http://www.ee.kth.se/~jonas1/thesis.pdf

[67] T. Söderström and P. Stoica, *System Identification*. : Prentice Hall International series in systems and control engineering, 1989.

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

**Håkan Hjalmarsson** (M'98) was born in 1962. He received the M.Sc. degree in electrical engineering, the Licentiate degree, the Ph.D. degree in automatic control from Linköping University, Sweden, in 1988, 1990, and 1993, respectively.

He has held visiting research positions at California Institute of Technology, Louvain University and at the University of Newcastle, Australia. His research interests include system identification, signal processing, control and estimation in communication networks and automated tuning of controllers.

Dr. Hjalmarsson has served as an Associate Editor for Automatica (1996–2001), and IEEE TRANS. AUTOM. CONTR. (2005–2007) and been Guest Editor for *European Journal of Control and Control Engineering Practice*. He is Professor at the School of Electrical Engineering, KTH, Stockholm, Sweden. He is Chair of the IFAC Technical Committee on Modeling, Identification and Signal Processing. In 2001 he received the KTH award for outstanding contribution to undergraduate education.

**Jonas Mårtensson** was born in 1976. He received the M.Sc. degree in vehicle engineering and the Ph.D. degree in automatic control from KTH Royal Institute of Technology, Stockholm, Sweden, in 2002 and 2007, respectively.

Since 2007 he is a Postdoctoral research associate at the School of Electrical Engineering at KTH. His research interests include system identification, experiment design and modeling and control of vehicular systems, with a present focus on internal combustion engines.