

# Variance Results for Parallel Cascade Serial Systems <sup>\*</sup>

Niklas Everitt <sup>\*</sup> Cristian R. Rojas <sup>\*</sup> Håkan Hjalmarsson <sup>\*</sup>

<sup>\*</sup> Department of Automatic Control and ACCESS Linnaeus Centre,  
School of Electrical Engineering,  
KTH–Royal Institute of Technology, SE-100 44 Stockholm, Sweden.  
E-mail: {neveritt, crro, hjalmars}@kth.se

---

**Abstract:** Modelling dynamic networks is important in different fields of science. At present, little is known about how different inputs and sensors contribute to the statistical properties concerning an estimate of a specific dynamic system in a network. We consider two forms of parallel serial structures, one multiple-input-multiple-output structure and one single-input-multiple-output structure. The quality of the estimated models is analysed by means of the asymptotic covariance matrix, with respect to input signal characteristics, noise characteristics, sensor locations and previous knowledge about the remaining systems in the network. It is shown that an additive property applies to the information matrix for the considered structures. The impact of input signal selection, sensor locations and incorporation of previous knowledge is illustrated by simple examples.

---

## 1. INTRODUCTION

Considerable research effort has been devoted to control of dynamic networks. The modelling problem is less understood. Some contributions to structured systems can be found in [Dayal and MacGregor, 1997], [Massioni and Verhaegen, 2009], [Van den Hof et al., 2013], and [Wahlberg et al., 2009]. The paper [Gevers et al., 2006] provides an analysis of which parameter estimates are improved by different inputs. The aim of this paper is to quantify the improvement in two specific structures. The first structure has its origin in boiler control, and has been considered in the paper [Hägg et al., 2011]; key results concern the case when there is common dynamics in the subsystems. The structure also appears, as a special case, when estimating a subsystem in a dynamic network, using the two-stage method [Van den Hof et al., 2013]. The second structure can be an example of a sensor network of spatially distributed sensors where the sensor dynamics are not completely known. Given the structure of the network and a subsystem of interest, our aim is to quantify how different inputs and sensors improve the estimate of that particular subsystem, in terms of the asymptotic statistical properties of the estimator. The main contribution of this paper is to provide an upper bound on the variance of an estimate of a subsystem. Additionally the variance reduction is characterized as a projection onto a certain row space.

The outline of this paper is as follows. In Section 2 we state the problem formulation, Section 3 gives some technical preliminaries. Section 4 contains results for the first structure, both for an example of  $m = 2$  inputs and the general case. Similarly, in Section 5, results for the second

structure for an example of  $m = 2$  additional sensors and the general case can be found. Results are exemplified in low-order FIR examples in Section 6, and Section 7 concludes the paper.

### Notation

We will treat vector valued complex functions as row vectors, and the inner product of two such functions  $f(z), g(z) : \mathbb{C} \rightarrow \mathbb{C}^{1 \times m}$  is defined as

$$\langle f, g \rangle \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\omega}) g^*(e^{i\omega}) d\omega \quad (1)$$

where  $g^*$  denotes the complex conjugate transpose of  $g$ . Furthermore  $\bar{f}$  denotes the complex conjugate of  $f$ . In case  $f, g$  are matrix valued functions we keep the same notation whenever the matrix dimensions are compatible. We denote by  $\|f\| = \sqrt{\text{Tr} \langle f, f \rangle}$  the  $\mathcal{L}_2$ -norm of  $f : \mathbb{C} \rightarrow \mathbb{C}^{n \times m}$ . We call two functions  $f, g$  orthogonal if  $\langle f, g \rangle = 0$ ; if  $f, g$  are matrix valued, they are considered orthogonal if every entry of  $\langle f, g \rangle$  is zero. A set of functions  $\{\mathcal{B}_k\}_{k=1}^n$  is said to be orthonormal if they are mutually orthogonal with unit  $\mathcal{L}_2$ -norm. If  $\Psi \in \mathcal{L}_2^{n \times m}$ , we denote by  $\mathcal{S}_\Psi \subset \mathcal{L}_2^m$  the subspace spanned by the rows of  $\Psi$ . For two subspaces  $\mathcal{X}$  and  $\mathcal{Y}$ , such that  $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{L}_2^m$ ,  $\mathcal{X}^\perp(\mathcal{Y})$  denotes the set of functions in  $\mathcal{Y}$  that are orthogonal to  $\mathcal{X}$ . We denote the orthogonal projection of  $f$  onto the space  $\mathcal{S}_\Psi$  by  $P_{\mathcal{S}_\Psi}[f]$ , i.e.,  $P_{\mathcal{S}_\Psi}[f]$  is the unique solution to

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

A sequence of subspaces  $\{X^n\}$ ,  $X^n \subseteq \mathcal{L}_2^m$  is said to converge to  $Y \subseteq \mathcal{L}_2^m$  if for any  $f \in \mathcal{L}_2^m$

$$\lim_{n \rightarrow \infty} \|P_{X^n}[f] - P_Y[f]\| = 0. \quad (2)$$

We denote this by  $\lim_{n \rightarrow \infty} X^n = Y$  or simply  $X^n \rightarrow Y$  as  $n \rightarrow \infty$ . The asymptotic covariance matrix of a stochastic sequence  $\{f_N\}_{N=1}^\infty$ ,  $f_N \in \mathbb{C}^q$  is defined as

---

<sup>\*</sup> This work was partially supported by the Swedish Research Council under contract 621-2009-4017, and by the European Research Council under the advanced grant LEARN, contract 267381.

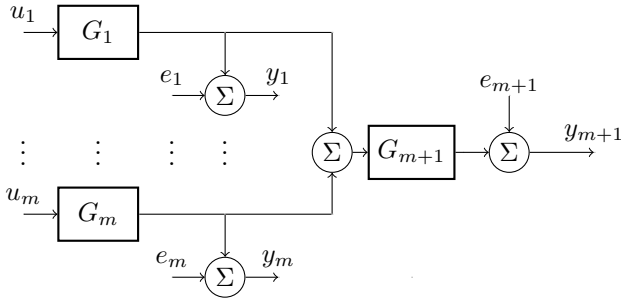


Fig. 1. Structure 1: Parallel serial structure.

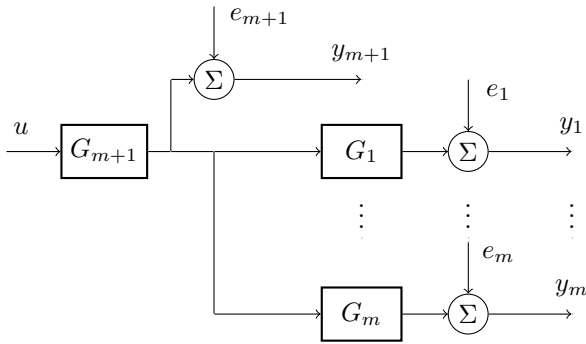


Fig. 2. Structure 2: Multi-sensor structure.

$$\text{AsCov } f_N \triangleq \lim_{N \rightarrow \infty} N \cdot \mathbf{E} \left[ (f_N - \mathbf{E}f_N)^T (f_N - \mathbf{E}f_N) \right].$$

For a differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{C}^q$ ,  $f'(\bar{x})$  is a  $n \times q$  matrix with  $\left. \frac{df_j(x)}{dx_i} \right|_{x=\bar{x}}$  as  $(i, j)$ th entry. For a row vector  $X$ , we will denote by  $\text{diag}\{X\}$  the matrix with the elements of  $X$  in the main diagonal and where all other elements equal zero. The  $\text{vec}\{X\}$  operator transforms a matrix into a vector by stacking the columns of  $X$  on top of each other [Seber, 2008].  $\otimes$  denotes the Kronecker product [Seber, 2008]. We use the notation  $A^\dagger$  for the Moore-Penrose pseudo inverse of  $A$ . Function arguments will, for clarity and lack of space, often be omitted. However, they should be clear from the context.

## 2. PROBLEM FORMULATION

We consider two types of networks of linear dynamic systems, where we wish to identify subsystem  $G_{m+1}$ . The first is the parallel serial (cascade) structure considered in [Hägg et al., 2011], see Figure 1.

**Structure 1:**

$$y_{m+1}(t) = \sum_{k=1}^m G_{m+1}(q)G_k(q)u_k(t) + e_{m+1}(t), \quad (3a)$$

$$y_k(t) = G_k(q)u_k(t) + e_k(t), \quad k = 1, \dots, m, \quad (3b)$$

where  $q$  denotes the forward shift operator, i.e.,  $q^{-1}u(t) = u(t-1)$  using normalized sampling time. The second is the multi-sensor structure of Figure 2. The structure is modelled as

**Structure 2:**

$$y_k(t) = G_k(q)G_{m+1}(q)u(t) + e_k(t), \quad k = 1, \dots, m, \quad (4a)$$

$$y_{m+1}(t) = G_{m+1}(q)u(t) + e_{m+1}(t). \quad (4b)$$

We assume that the additive zero mean white noise sequences  $\{e_i(t)\}$  are mutually independent, and independent of the input  $u(t)$ , with variances  $\lambda_i, i = 1, \dots, m+1$ . The input is assumed to be a realization of a weakly stationary stochastic process with spectrum  $\Phi_u$ . The models of the subsystems are independently parametrized with  $\theta = [\theta_1, \dots, \theta_{m+1}]$ , where  $\theta_i \in \mathbb{R}^{d_i}, d_i \in \mathbb{R}$  for all  $i = 1, \dots, m+1$ . We assume the model structure is uniformly stable (see [Ljung, 1999]), the true system is in the model set and we denote the true parameters by  $\theta^o$ , that is,

$$G_k(q) = G_k(q, \theta_k^o), \quad k = 1, \dots, m. \quad (5)$$

We assume that the parameter vector  $\theta$  is estimated from a data set of measured inputs and outputs of sample size  $N$  using the prediction error method, and we denote the estimate by  $\hat{\theta}_N$ . Under mild regularity conditions (see [Ljung, 1999] for details), as  $N$  goes to infinity, the parameter error  $\sqrt{N}(\hat{\theta}_N - \theta^o)$  converges in distribution to a normal distribution with zero mean and covariance matrix  $P$ , which we conveniently denote by

$$\sqrt{N}(\hat{\theta}_N - \theta^o) \in \text{AsN}(0, P). \quad (6)$$

Here  $P$  is the asymptotic covariance matrix of the parameter estimates, which we assume for the moment can be written as

$$P_\theta = \text{AsCov } \hat{\theta}_N = \langle \Psi, \Psi \rangle^\dagger, \quad (7)$$

where  $\Psi : \mathbb{C} \rightarrow \mathbb{C}^{n \times m}$ , for some integer  $m > 0$ , which in our case corresponds to the number of subsystems. All the elements of  $\Psi$  are assumed to belong to  $\mathcal{L}_2^1$ . Let  $J : \mathbb{R}^{1 \times n} \rightarrow \mathbb{C}^{1 \times q}$  be a differentiable function of  $\theta$ . From (6), it follows that

$$\sqrt{N}(J(\hat{\theta}_N) - J(\theta^o)) \in \text{AsN}(0, \text{AsCov } J(\hat{\theta}_N)). \quad (8)$$

Using the Gauss' approximation formula (or delta method) [Ljung, 1999] and (6) it can be shown that

$$\text{AsCov } J(\hat{\theta}_N) = \Lambda^T [\langle \Psi, \Psi \rangle]^\dagger \Lambda, \quad (9)$$

where  $\Lambda$  is the derivative  $\Lambda \triangleq J'(\theta^o) \in \mathbb{C}^{n \times q}$ . We will use the formulation of the asymptotic covariance matrix given in [Agüero et al., 2012] (Theorem 4) which, under our assumptions<sup>2</sup> and adapted to our notation, is

$$P_\theta^{-1} = \left\langle \left[ \frac{\partial \mathcal{L}}{\partial \theta^T} \right]^H W_\chi, \left[ \frac{\partial \mathcal{L}}{\partial \theta^T} \right]^H \right\rangle, \quad (10)$$

where

$$\mathcal{L} = \text{vec}\{G\}, \quad W_\chi = \Phi_u^T \otimes \Phi_e^{-1}, \quad (11)$$

and  $G$  is the transfer function matrix between all inputs and all outputs.

## 3. TECHNICAL PRELIMINARIES

Here we recall some technical preliminaries that reformulate the Schur complement into orthogonal projections.

*Lemma 3.1.* Let  $f \in \mathcal{L}_2^l$  and let  $\mathcal{S}_n^m$  be a (closed) subspace of  $\mathcal{L}_2^m$  with orthonormal basis  $\{\mathcal{B}_k\}_{k=1}^n$ . Then

$$P_{\mathcal{S}_n^m}[f] \triangleq \sum_{k=1}^n \langle f, \mathcal{B}_k \rangle \mathcal{B}_k \quad (12)$$

is the orthogonal projection of  $f$  onto  $\mathcal{S}_n^m$ .

<sup>1</sup> This is the standard situation when the true parameter vector corresponds to a stable predictor in the prediction error method, see [Ljung, 1999].

<sup>2</sup> The main simplification of the general formula of [Agüero et al., 2012] comes from knowing the noise models and the noise variances.

**Proof.** See, e.g., [Friedman, 1970].  $\square$

*Lemma 3.2.* (Lemma II.3 in [Hjalmarsson and Mårtensson, 2011]) Let  $\gamma \in \mathcal{L}_2^{q \times m}$ ,  $\Psi \in \mathcal{L}_2^{n \times m}$ . Then the orthogonal projection of the rows of  $\gamma$  onto  $\mathcal{S}_\Psi$  (the subspace to  $\mathcal{L}_2^m$  spanned by the rows of  $\Psi$ ) is given by

$$P_{\mathcal{S}_\Psi}[\gamma] = \langle \gamma, \Psi \rangle [\langle \Psi, \Psi \rangle]^\dagger \Psi. \quad (13)$$

Furthermore

$$\langle \gamma, \Psi \rangle [\langle \Psi, \Psi \rangle]^\dagger \langle \Psi, \gamma \rangle = \langle P_{\mathcal{S}_\Psi}[\gamma], P_{\mathcal{S}_\Psi}[\gamma] \rangle. \quad (14)$$

Finally it holds that

$$\langle P_{\mathcal{S}_\Psi}[\gamma], P_{\mathcal{S}_\Psi}[\gamma] \rangle = \sum_{k=1}^r \langle \gamma, \mathcal{B}_k \rangle \langle \mathcal{B}_k, \gamma \rangle \quad (15)$$

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

*Lemma 3.3.* Assume that the asymptotic covariance matrix for a vector  $\theta = [\theta_1 \ \theta_2]$ ,  $\theta_1 \in \mathbb{R}^{n_1}$ ,  $\theta_2 \in \mathbb{R}^{n_2}$ , can be written as

$$P_\theta^{-1} = \langle \Psi, \Psi \rangle = \begin{bmatrix} \langle \Psi_1, \Psi_1 \rangle & \langle \Psi_1, \Psi_2 \rangle \\ \langle \Psi_2, \Psi_1 \rangle & \langle \Psi_2, \Psi_2 \rangle \end{bmatrix} \quad (16)$$

where  $\Psi = [\Psi_1^T \ \Psi_2^T]^T$ ,  $\Psi_1 \in \mathcal{L}_2^{n_1 \times m}$ ,  $\Psi_2 \in \mathcal{L}_2^{n_2 \times m}$ , and  $\Psi_1(e^{j\omega})\Psi_1^T(e^{-j\omega})$  is positive semidefinite and has rank  $p$ . Then the asymptotic covariance matrix for  $\theta_2$  is

$$P_{\theta_2} = [\langle \Psi_2, \Psi_2 \rangle - \langle P_{\mathcal{S}_{R_1}}[\gamma], P_{\mathcal{S}_{R_1}}[\gamma] \rangle]^\dagger \quad (17)$$

with  $\gamma = \Psi_2 \Psi_1^H [R_1^\dagger]^H$ , where  $R_1$  is a spectral factor of  $\Psi_1(e^{j\omega})\Psi_1^T(e^{-j\omega})$ , that is,  $\Psi_1(e^{j\omega})\Psi_1^T(e^{-j\omega}) = R_1(e^{j\omega})R_1^T(e^{-j\omega})$  such that the function  $R_1(z)$  is analytic in the unit disc and has rank  $p$  for all  $z$  in this domain.

**Proof.** The spectral factor  $R_1$  exists under the given assumptions, see Theorem 10.1 in [Rozanov, 1967]. Rewriting

$$\langle \Psi_2, \Psi_1 \rangle = \langle \Psi_2 \Psi_1^H [R_1^\dagger]^H, R_1 \rangle, \quad (18)$$

and applying the standard inverse of a partitioned matrix [Horn and Johnson, 1990] and Lemma 3.2 proves the Lemma.  $\square$

In the next lemma, we let number of estimated parameters in the  $m$  first subsystem grow large to make the projection trivial to calculate.

*Lemma 3.4.* Let  $\Psi$  be defined as in Lemma 3.3 and assume that  $\Psi_1 = \Gamma_1 \tilde{\Psi}_1$  and  $\Psi_2 = \Gamma_2 \tilde{\Psi}_2$  for some  $\Gamma_1 \in \mathcal{L}_2^{n_1 \times m_1}$ ,  $\Gamma_2 \in \mathcal{L}_2^{n_2 \times m_2}$ ,  $\tilde{\Psi}_1 \in \mathcal{L}_2^{m_1 \times m}$  and  $\tilde{\Psi}_2 \in \mathcal{L}_2^{m_2 \times m}$ , and that  $\text{rank} \left\{ \tilde{\Psi}_1 \tilde{\Psi}_1^H \right\} = m_1$ . If  $\mathcal{S}_{\Gamma_1} = \mathcal{L}_2^{m_1}$ , then

$$P_{\theta_2} = \left[ \left\langle \Gamma_2 (\tilde{\Psi}_2 \tilde{\Psi}_2^H - \gamma \gamma^H), \Gamma_2 \right\rangle \right]^{-1}, \quad (19)$$

with  $\gamma = \tilde{\Psi}_2 \tilde{\Psi}_1^H [R_1^{-1}]^H$ , where  $R_1$  is a spectral factor of  $\Psi_1(e^{j\omega})\Psi_1^T(e^{-j\omega})$ , analytic in the unit disc with rank  $m_1$  for all  $z$  in this domain.

**Proof.**  $R_1$  is an invertible mapping, hence  $\mathcal{S}_{\Gamma_1 R_1} = \mathcal{S}_{\Gamma_1} = \mathcal{L}_2^{m_1}$ , which implies  $P_{\mathcal{S}_{\Gamma_1 R_1}}[\Gamma_2 \gamma] = P_{\mathcal{L}_2^{m_1}}[\Gamma_2 \gamma] = \Gamma_2 \gamma$  in Lemma 3.3.  $\square$

#### 4. STRUCTURE 1: PARALLEL SERIAL STRUCTURE

In this section we study the parallel cascade structure described by Equation 3, visualized in Figure 1. Before

giving the general theorem, it is instructive to consider the case  $m = 2$ . When  $m = 2$ ,

$$P_\theta^{-1} = \langle \Gamma A, \Gamma \rangle,$$

where

$$A = \begin{bmatrix} \tilde{\Psi}_1 \tilde{\Psi}_1^H & \tilde{\Psi}_1 \tilde{\Psi}_2^H \\ \tilde{\Psi}_2 \tilde{\Psi}_1^H & \tilde{\Psi}_2 \tilde{\Psi}_2^H \end{bmatrix}$$

$$\tilde{\Psi}_2 \tilde{\Psi}_1^H = \begin{bmatrix} \frac{\Phi_{u_1} \overline{G_1 G_3}}{\lambda_3} & \frac{\Phi_{u_2} \overline{G_2 G_3}}{\lambda_3} \end{bmatrix}$$

$$\tilde{\Psi}_1 \tilde{\Psi}_1^H = \text{diag} \left\{ \frac{\Phi_{u_1}}{\lambda_1} + \frac{\Phi_{u_1} |G_3|^2}{\lambda_3}, \frac{\Phi_{u_2}}{\lambda_2} + \frac{\Phi_{u_2} |G_3|^2}{\lambda_3} \right\}$$

$$\tilde{\Psi}_2 \tilde{\Psi}_2^H = \frac{\Phi_{u_1} |G_1|^2}{\lambda_3} + \frac{\Phi_{u_2} |G_2|^2}{\lambda_3},$$

$$\Gamma_1 = \text{diag} \left\{ \overline{G_1'}, \overline{G_2'} \right\}, \quad \Gamma_2 = \overline{G_3'}.$$

Notice that  $\text{rank} \left\{ \tilde{\Psi}_1 \tilde{\Psi}_1^H \right\} = 2$  when  $\Phi_{u_1}, \Phi_{u_2} > 0$ . We let the number of estimated parameters in  $\theta_1$  and  $\theta_2$  grow large. Naturally, if  $\mathcal{S}_{G_1'}, \mathcal{S}_{G_2'} \rightarrow \mathcal{L}_2$ , then,  $\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^2$ . From Lemma 3.4 we have that

$$\lim_{\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^2} P_{\theta_3} = \left[ \left\langle \overline{G_3'} (M_{u_1} + M_{u_2}), \overline{G_3'} \right\rangle \right]^{-1} \quad (20)$$

where

$$M_{u_k} = \frac{\Phi_{u_k} |G_k|^2}{\lambda_3 + |G_3|^2 \lambda_k}, \quad k = 1, 2. \quad (21)$$

*Remark 1.* Notice that if we only use input  $u_1$ , we obtain the covariance matrix

$$P_{\theta_3} = \left[ \left\langle \overline{G_3'} M_{u_1}, \overline{G_3'} \right\rangle \right]^{-1} \quad (22)$$

and similar results hold if we only use  $u_2$ . The information is additive, which may come as no surprise considering that the same kind of relations hold for optimal combination of uncorrelated estimators cf. [Kailath et al., 2000].

For  $m$  input signals we have the following theorem.

*Theorem 4.1.* Consider structure 1 with  $m$  inputs, and define

$$\Gamma_1 = \text{diag} \left\{ \overline{G_1'}, \dots, \overline{G_m'} \right\}, \quad \Gamma_2 = \overline{G_{m+1}'}$$

If  $\mathcal{S}_{G_1'}, \mathcal{S}_{G_2'}, \dots, \mathcal{S}_{G_m'} \rightarrow \mathcal{L}_2$ ,  $\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m$ . Then

$$\lim_{\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m} P_{\theta_{m+1}} = \left[ \left\langle \overline{G_{m+1}'} \sum_{k=1}^m M_{u_k}, \overline{G_{m+1}'} \right\rangle \right]^{-1} \quad (23)$$

where

$$M_{u_k} = \frac{\Phi_{u_k} |G_k|^2}{\lambda_{m+1} + |G_{m+1}|^2 \lambda_k}, \quad k = 1, \dots, m. \quad (24)$$

**Proof.** The proof is provided in Appendix A

#### 5. STRUCTURE 2: MULTI SENSOR STRUCTURE

In this section we study the multi sensor structure described by Equation 4, visualized in Figure 2. For this structure, we also first consider the case  $m = 2$ , before providing the general formulation. For  $m = 2$  additional sensors

$$P_\theta^{-1} = \langle \Gamma A, \Gamma \rangle$$

where

$$A = \begin{bmatrix} \tilde{\Psi}_1 \tilde{\Psi}_1^H & \tilde{\Psi}_1 \tilde{\Psi}_2^H \\ \tilde{\Psi}_2 \tilde{\Psi}_1^H & \tilde{\Psi}_2 \tilde{\Psi}_2^H \end{bmatrix}$$

$$\tilde{\Psi}_2 \tilde{\Psi}_1^H = \begin{bmatrix} \frac{\Phi_u \overline{G_1 G_3}}{\lambda_1} & \frac{\Phi_u \overline{G_2 G_3}}{\lambda_2} \end{bmatrix}$$

$$\tilde{\Psi}_1 \tilde{\Psi}_1^H = \text{diag} \left\{ \frac{\Phi_u |G_3|^2}{\lambda_1}, \frac{\Phi_u |G_3|^2}{\lambda_2} \right\}$$

$$\tilde{\Psi}_2 \tilde{\Psi}_2^H = \frac{\Phi_u |G_1|^2}{\lambda_1} + \frac{\Phi_u |G_2|^2}{\lambda_2} + \frac{\Phi_u}{\lambda_3}$$

$$\Gamma_1 = \text{diag} \left\{ \overline{G'_1}, \overline{G'_2} \right\}, \quad \Gamma_2 = \overline{G'_3}.$$

Notice that  $\text{rank} \left\{ \tilde{\Psi}_1 \tilde{\Psi}_1^H \right\} = 2$  when  $\Phi_u |G_3|^2 > 0$ . That is, we assume that

$$\Phi_u(\omega) > 0, \quad |G_3(e^{i\omega})|^2 > 0, \quad \omega \in [-\pi, \pi]. \quad (25)$$

We let the number of estimated parameters in  $\theta_1$  and  $\theta_2$  grow large. Naturally, if  $\mathcal{S}_{G'_1}, \mathcal{S}_{G'_2} \rightarrow \mathcal{L}_2$ , then,  $\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^2$ . From Lemma 3.4 we obtain

$$\lim_{\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^2} P_{\theta_3} = \left[ \left\langle \Gamma_2 (\tilde{\Psi}_{22} \tilde{\Psi}_{22}^H - \gamma \gamma^H), \Gamma_2 \right\rangle \right]^{-1} \quad (26)$$

$$\gamma \gamma^H = \frac{\Phi_u |G_1|^2}{\lambda_1} + \frac{\Phi_u |G_2|^2}{\lambda_2}. \quad (27)$$

We see that

$$\lim_{\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^2} P_{\theta_3} = \left[ \left\langle \overline{G'_3} \Phi_u \lambda_3^{-1}, \overline{G'_3} \right\rangle \right]^{-1} \quad (28)$$

Notice that this is the same covariance matrix as if we would only use output  $y_3$ . This also holds for any number of outputs satisfying the constraints of structure 2.

*Theorem 5.1.* Consider structure 2 with  $m$  additional sensors, and assume that

$$\Phi_u(\omega) > 0, \quad |G_{m+1}(e^{i\omega})|^2 > 0, \quad \omega \in [-\pi, \pi]. \quad (29)$$

Define

$$\Gamma_1 = \text{diag} \left\{ \overline{G'_1}, \dots, \overline{G'_m} \right\}, \quad \Gamma_2 = \overline{G'_{m+1}}.$$

If  $\mathcal{S}_{G'_1}, \mathcal{S}_{G'_2}, \dots, \mathcal{S}_{G'_m} \rightarrow \mathcal{L}_2$ ,  $\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m$ , then

$$\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m$$

and

$$\lim_{\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m} P_{\theta_{m+1}} = \left[ \left\langle \overline{G'_{m+1}} \Phi_u \lambda_{m+1}^{-1}, \overline{G'_{m+1}} \right\rangle \right]^{-1}. \quad (30)$$

**Proof.** The proof is constructive and provided in Appendix B

*Remark 2.* The gain in information becomes arbitrary small when the number of estimated parameters grow large. This might (wrongly) lead us to the conclusion that if we do not know the sensor dynamics completely we should not bother with the additional sensors. However, if we can restrict the dimension of the space  $\mathcal{S}_{\Gamma_1}$  (knowing some parts of the dynamics for example) we still gain information. In fact, we can quantify the gain in information as a projection.

*Theorem 5.2.* Consider the same assumptions as in Theorem 5.1. Then

$$P_{\theta_{m+1}} = \left[ \left\langle \Gamma_2 S, \Gamma_2 \right\rangle - \left\langle P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}} [\Gamma_2 \gamma], P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}} [\Gamma_2 \gamma] \right\rangle \right]^{-1} \quad (31)$$

$$= \left[ \left\langle \Gamma_2 \Phi_u \lambda_{m+1}^{-1}, \Gamma_2 \right\rangle + M \right]^{-1} \quad (32)$$

where the gain in information  $M$  is given by

$$M = \left\langle P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}^\perp} [\Gamma_2 \gamma], P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}^\perp} [\Gamma_2 \gamma] \right\rangle, \quad (33)$$

**Proof.** The proof is provided in Appendix C

*Remark 3.* The information gain is a continuum where another extreme is knowing the  $m$  additional sensors exactly, which corresponds to  $\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}^\perp = \mathcal{L}_2^m$ . In that case,  $M$  is given by

$$M = \left\langle \Gamma_2 \gamma \gamma^H, \Gamma_2 \right\rangle = \left\langle \Gamma_2 \sum_{k=1}^m \Phi_u |G_k|^2 \lambda_k^{-1}, \Gamma_2 \right\rangle. \quad (34)$$

## 6. FIR EXAMPLES

We verify the correctness of the presented results on Monte-Carlo simulations of FIR systems. In all examples  $N = 1000$  measurements are used and the sample variances of the frequency function estimate of 500 noise realisations is computed using (9). The noise source and input are assumed mutually independent zero mean Gaussian white noise with unit variance. When input  $r_2$  is not considered it is put to zero. The estimates are computed as the minimizer of<sup>3</sup>

$$f(\theta) = \frac{N}{2} \ln \det \left\{ \sum_{t=1}^N \epsilon(t)^T \Lambda \epsilon(t) \right\} \quad (35)$$

where  $\epsilon(t) = y(t) - \hat{y}(t)$ ,  $\Lambda = \text{diag} \{ \lambda_1^{-1}, \dots, \lambda_{m+1}^{-1} \}$ . We consider examples with  $m = 2$  inputs and 3 FIR systems, all with true order  $p = 3$ , i.e.

$$G_1 = G_2 = 1 + 0.5q^{-1} + 0.25q^{-2},$$

$$G_3 = 1 + 0.2q^{-1} + 0.04q^{-2}.$$

The systems  $G_1, G_2$  are estimated with 30 parameters and the system of interest in all examples,  $G_3$ , is estimated with 3 parameters:

$$\hat{G}_i = \sum_{k=0}^{29} \hat{g}_{i,k} q^{-k} \quad i = 1, 2, \quad \hat{G}_3 = \sum_{k=0}^2 \hat{g}_{3,k} q^{-k}.$$

For structure 1, the parallel serial structure, the sample covariance of the transfer function estimates shows strong similarity to the asymptotic (both in samples and parameters) theoretic expression, see Figure 3. In the case  $m = 1$ , then  $r_2 = 0$  and  $G_3$  is estimated. Knowing the first impulse response coefficient of  $G_1$  and  $G_2$  gives only a minor reduction in variance of  $\hat{G}_3$  as seen in Figure 4.

For structure 2, the multi sensor structure, the variance of the transfer function estimate  $\hat{G}_3$  does not improve by using also  $y_2$ , and is the same as what would be achieved by only using  $y_3$ , cf. Figure 5. When we know the first coefficient of  $G_1$  and  $G_2$ , the estimate of the first impulse response coefficient  $g_{3,1}$  is improved which results in a lower variance for the estimated transfer function  $\hat{G}_3$ , cf. Figure 6. In contrast to Structure 1, knowing some parameters in  $G_1$  and  $G_2$  makes all the difference, cf. Figure 4 and Figure 6.

## 7. CONCLUSION

We have examined the variance of the estimate of one systems in the network, when little assumptions were made on the remaining systems in the network. We derived asymptotic variance expressions for two types of structured dynamic systems. The information from using additional inputs was shown to be additive. Previous

<sup>3</sup> The prediction error method is efficient in the Gaussian case

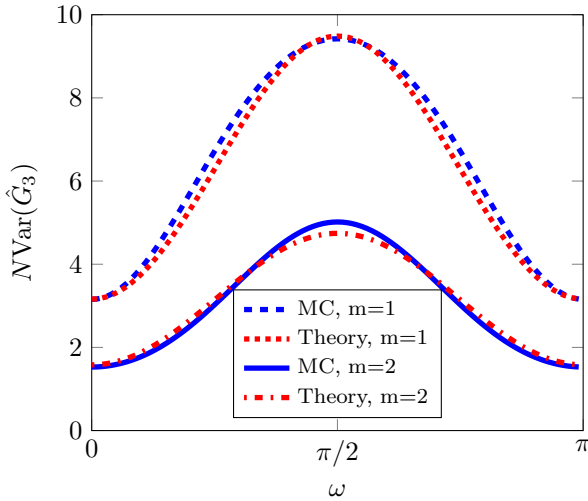


Fig. 3. Structure 1: Comparison of Monte-Carlo simulations (MC) and the asymptotic theory for  $m = 1, 2$ .

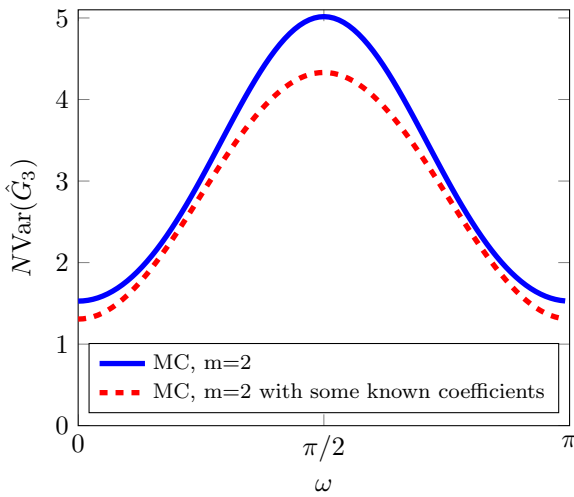


Fig. 4. Structure 1: Comparison of Monte-Carlo simulations (MC) for  $m = 2$  when the first impulse response coefficient of  $G_1$  and  $G_2$  are known.

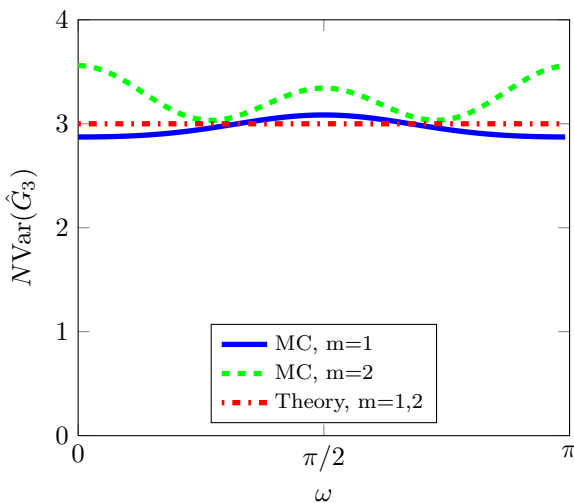


Fig. 5. Structure 2: Comparison of Monte-Carlo simulations (MC) and the asymptotic theory for  $m = 1, 2$ .

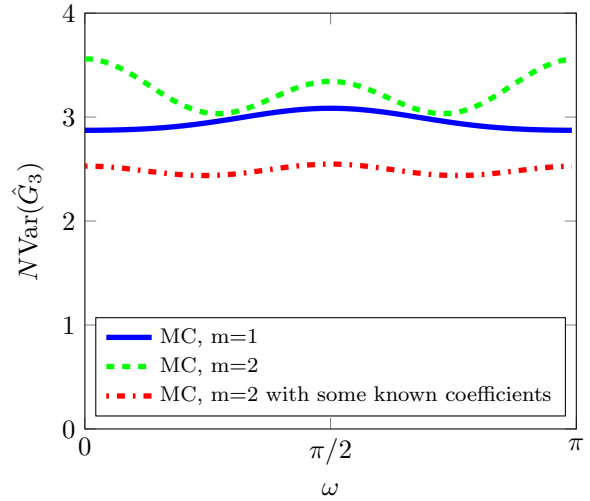


Fig. 6. Structure 2: Comparison of Monte-Carlo simulations (MC) for  $m = 2$  when the first impulse response coefficient of  $G_1$  and  $G_2$  are known.

knowledge about the additional sensors in the multi sensor structure is imperative for a variance reduction, in fact, without prior knowledge there is no asymptotic variance reduction.

#### REFERENCES

- J. C. Agüero, C. R. Rojas, H Hjalmarsson, and G. C. Goodwin. Accuracy of linear multiple-input multiple-output (MIMO) models obtained by maximum likelihood estimation. *Automatica*, 48(4):632–637, 2012.
- B. S. Dayal and J. F. MacGregor. Multi-output process identification. *Journal of Process Control*, 7(4):269–282, 1997.
- A. Friedman. *Foundations of Modern Analysis*. Dover, 1970.
- M. Gevers, L. Mišković, D. Bonvin, and A. Karimi. Identification of multi-input systems: variance analysis and input design issues. *Automatica*, 42(4):559 – 572, 2006.
- P. Hägg, B. Wahlberg, and H. Sandberg. On identification of parallel cascade serial systems. In *Proceedings of the 18th IFAC World Congress*, 2011.
- H. Hjalmarsson and J. Mårtensson. A geometric approach to variance analysis in system identification. *Automatic Control, IEEE Transactions on*, 56(5):983–997, may 2011.
- R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.
- T. Kailath, A. H. Sayed, and B. Hassibi. *Linear estimation*. Prentice Hall, 2000.
- L. Ljung. *System Identification: Theory for the User*. Prentice Hall, 2 edition, 1999.
- P. Massioni and M. Verhaegen. Subspace identification of distributed decomposable systems. In *48th IEEE Conference on Decision and Control*, 2009.
- Yu. A. Rozanov. *Stationary Random Processes*. Holden-Day, 1967.
- G. A. F. Seber. *A Matrix Handbook for Statisticians*. Wiley, 2008.
- P. M.J. Van den Hof, Arne Dankers, P. S. C. Heuberger, and X Bombois. Identification of dynamic models in complex networks with prediction error methods - Basic

methods for consistent module estimates. *Automatica*, 49(10):2994–3006, 2013.

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

#### Appendix A. PROOF OF THEOREM 4.1

We define  $V_k \in \mathcal{L}^{(m+1) \times m}$  as

$$[V_k]_{ij} = \begin{cases} 1, & i = j = k, \\ 0, & \text{otherwise,} \end{cases}$$

and  $Q_k \in \mathcal{L}^{(m+1) \times 1}$  as

$$[Q_k]_i = \begin{cases} \overline{G_{m+1}}, & i = k, \\ \overline{G_k}, & i = m + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$W_{\chi_k} = \Phi_{u_k} \cdot \text{diag} \{ \lambda_1^{-1}, \dots, \lambda_{m+1}^{-1} \}.$$

We can write

$$\left[ \frac{\mathcal{L}_\pi}{\partial \theta^T} \right]^H = \Gamma L, \quad \Gamma = \text{diag} \{ \overline{G'_1}, \dots, \overline{G'_{m+1}} \}$$

and  $\Gamma$  should be interpreted as block wise diagonal, where

$$L = [V_1 \ Q_1 \ \dots \ V_m \ Q_m].$$

Then,

$$P_\theta^{-1} = \langle \Gamma L W_\chi, \Gamma L \rangle = \langle \Gamma A, \Gamma \rangle$$

where

$$A = \sum_{k=1}^m [V_k \ Q_k] W_{\chi_k} \begin{bmatrix} V_k \\ Q_k \end{bmatrix}.$$

It is readily verified that

$$A = \begin{bmatrix} D & T \\ T^H & S \end{bmatrix}$$

where

$$D_{ij} = \begin{cases} \Phi_{u_i} (\lambda_i^{-1} + |G_{m+1}|^2 \lambda_{m+1}^{-1}), & i = j, \\ 0, & \text{otherwise,} \end{cases}$$

$T \in \mathcal{L}^{m+1 \times 1}$  is given by

$$[T]_i = \Phi_{u_i} \overline{G_i} G_{m+1} \lambda_{m+1}^{-1}$$

and

$$S = \sum_{k=1}^m \Phi_{u_k} |G_k|^2 \lambda_{m+1}^{-1}.$$

We identify

$$\tilde{\Psi}_2 \tilde{\Psi}_1^H = T^H, \quad \tilde{\Psi}_1 \tilde{\Psi}_1^H = D, \quad \tilde{\Psi}_2 \tilde{\Psi}_2^H = S$$

and

$$\Gamma_1 = \text{diag} \{ \overline{G'_1}, \dots, \overline{G'_m} \}, \quad \Gamma_2 = \overline{G'_{m+1}}.$$

Notice that  $\text{rank} \{ D \} = m$  when  $\Phi_{u_1}, \dots, \Phi_{u_m} > 0$ . We let the number of estimated parameters in  $\theta_1, \dots, \theta_m$  grow large. Naturally, if  $\mathcal{S}_{G'_1}, \mathcal{S}_{G'_2}, \dots, \mathcal{S}_{G'_m} \rightarrow \mathcal{L}_2$ ,  $\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m$ . From Lemma 3.4 we have that

$$\lim_{\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m} P_{\theta_{m+1}} = [\langle \Gamma_2 (S - \gamma \gamma^H), \Gamma_2 \rangle]^{-1} \quad (\text{A.1})$$

$$\gamma \gamma^H = \sum_{k=1}^m \frac{\Phi_{u_k} |G_k G_{m+1}|^2 \lambda_{m+1}^{-2}}{\lambda_k^{-1} + |G_{m+1}|^2 \lambda_{m+1}^{-1}}, \quad (\text{A.2})$$

and the theorem follows after some simplifications.

#### Appendix B. PROOF OF THEOREM 5.1

We can write

$$\left[ \frac{\mathcal{L}_\pi}{\partial \theta^T} \right]^H = \Gamma L, \quad \Gamma = \text{diag} \{ \overline{G'_1}, \dots, \overline{G'_{m+1}} \}$$

where  $\Gamma$  should be interpreted as block wise diagonal, and

$$L = \begin{bmatrix} \overline{G_{m+1}} & & & \\ & \ddots & & \\ & & \overline{G_{m+1}} & \\ \overline{G_1} & \dots & G_m & 1 \end{bmatrix}$$

$$W_\chi = \Phi_u \otimes \text{diag} \{ \lambda_1^{-1}, \dots, \lambda_{m+1}^{-1} \}.$$

Then,

$$P_\theta^{-1} = \langle \Gamma L W_\chi, \Gamma L \rangle = \langle \Gamma A, \Gamma \rangle, \quad (\text{B.1})$$

where

$$A = \begin{bmatrix} D & T \\ T^H & S \end{bmatrix}$$

with

$$D_{ij} = \begin{cases} |G_{m+1}|^2 \lambda_i^{-1}, & i = j, \\ 0, & \text{otherwise,} \end{cases}$$

$T \in \mathcal{L}^{m+1 \times 1}$  is given by

$$[T]_i = \Phi_u \overline{G_{m+1}} G_i \lambda_i^{-1}$$

and

$$S = \Phi_u \lambda_{m+1}^{-1} + \sum_{k=1}^m \Phi_u |G_k|^2 \lambda_k^{-1}.$$

The next step is to apply Lemma 3.4. We identify

$$\tilde{\Psi}_2 \tilde{\Psi}_1^H = T^H, \quad \tilde{\Psi}_1 \tilde{\Psi}_1^H = D, \quad \tilde{\Psi}_2 \tilde{\Psi}_2^H = S$$

and

$$\Gamma_1 = \text{diag} \{ \overline{G'_1}, \dots, \overline{G'_m} \}, \quad \Gamma_2 = \overline{G'_{m+1}}.$$

To ensure that  $\text{rank} \{ \tilde{\Psi}_1 \tilde{\Psi}_1^H \} = m$ , we again assume that

$$\Phi_u(\omega) > 0, \quad |G_{m+1}(e^{i\omega})|^2 > 0, \quad \omega \in [-\pi, \pi]. \quad (\text{B.2})$$

We let the number of estimated parameters in  $\theta_1, \dots, \theta_m$  grow large. Naturally, if  $\mathcal{S}_{G'_1}, \dots, \mathcal{S}_{G'_m} \rightarrow \mathcal{L}_2$ , then,  $\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m$ . From Lemma 3.4 we have that

$$\lim_{\mathcal{S}_{\Gamma_1} \rightarrow \mathcal{L}_2^m} P_{\theta_{m+1}} = [\langle \Gamma_2 (S - \gamma \gamma^H), \Gamma_2 \rangle]^{-1} \quad (\text{B.3})$$

$$\gamma \gamma^H = \sum_{k=1}^m \Phi_u |G_k|^2 \lambda_k^{-1}, \quad (\text{B.4})$$

and the theorem follows after some simplifications.

#### Appendix C. PROOF OF THEOREM 5.2

We notice that

$$\Gamma_2 (S - \Phi_u \lambda_{m+1}^{-1}) \Gamma_2^H = \Gamma_2 \gamma (\Gamma_2 \gamma)^H \quad (\text{C.1})$$

$$= \langle P_{\mathcal{L}_2^m} [\Gamma_2 \gamma], P_{\mathcal{L}_2^m} [\Gamma_2 \gamma] \rangle. \quad (\text{C.2})$$

Applying Lemma 3.3 to Equation (B.1) we see that

$$P_{\theta_{m+1}} = \left[ \langle \Gamma_2 S, \Gamma_2 \rangle - \left\langle P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}} [\Gamma_2 \gamma], P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}} [\Gamma_2 \gamma] \right\rangle \right]^{-1} \quad (\text{C.3})$$

$$= [\langle \Gamma_2 \Phi_u \lambda_{m+1}^{-1}, \Gamma_2 \rangle + M]^{-1} \quad (\text{C.4})$$

where the gain in information  $M$  is given by

$$M = \left\langle P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}^\perp} [\Gamma_2 \gamma], P_{\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}^\perp} [\Gamma_2 \gamma] \right\rangle, \quad (\text{C.5})$$

where  $\mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}^\perp = \mathcal{S}_{\Gamma_1 \tilde{\Psi}_1}^\perp (\mathcal{L}_2^m)$ .