
# Identification of Modules in Acyclic Dynamic Networks: A Geometric Analysis of Stochastic Model Errors

NIKLAS EVERITT

Licentiate Thesis
Stockholm, Sweden 2015

KTH Royal Institute of Technology
School of Electrical Engineering
Department of Automatic Control
SE-100 44 Stockholm
SWEDEN

TRITA-EE 2015:005
ISSN 1653-5146
ISBN 978-91-7595-432-5

Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framläggs till offentlig granskning för avläggande av teknologie licentiatexamen i elektro- och systemteknik onsdagen den 11 februari 2015 klockan 10.15 i sal E3, Kungliga Tekniska högskolan, Osquarsbacke 14, Stockholm.



Tryck: Universitetsservice US AB

**Abstract**

Systems in engineering are becoming ever more complex and interconnected, due to advancing technology with cheaper sensors and increased connectivity. For example in process industry, sensors that monitor the operation of the plant can be connected through wireless connections and used for monitoring and control. In this thesis, we study the problem of identifying one module, i.e., one transfer function from one internal variable to another, in the dynamic network. We investigate how accurate models will be obtained using different gathered measurements. Model errors are assumed to originate from random disturbances that affect the dynamic network. The variance of the model errors are analyzed using the classical assumption that a large amount of data is available. By using a geometric approach, the (co-)variance of the model errors can be analyzed in a way that brings forward how input signal properties, noise variance, noise correlation structure and model structure affect the asymptotic model errors. Several different network structures are analyzed in order to investigate how different signals can reduce the asymptotic model errors in dynamic networks.

For SISO systems we develop reparametrization formulas for the asymptotic variance of the model errors of functions of the estimated system parameters. In particular, we demonstrate that one can use the experimental conditions to make the asymptotic variance independent of model order and model structure in some cases. These expressions are used to derive simple model structure independent upper bounds of the asymptotic covariance for commonly estimated quantities such as system zeros and impulse response coefficients.

The variance of the first of a set of estimated modules connected in a cascade structure is analyzed. The main contribution is the characterization of the variance of the frequency function estimate of a module with a zero close to the unit circle. It is shown that a variance reduction of the first estimated module is possible, compared to only using the first measurement in the estimation. The variance reduction is concentrated around the frequency of the unit-circle-zero.

For a parallel cascade structure and a multi sensor structure, upper bounds on the asymptotic covariance of the parameter estimates are derived when the model order of the system of interest was fixed, while the model order of every other module is large.

The effect of the noise correlation structure is examined for single input multiple output (SIMO) systems. For the case of temporally white, but possibly spatially correlated additive noise, we develop a formula for the asymptotic covariance of the frequency response function estimates and a formula for the asymptotic covariance of the model parameters. It is shown that when parts of of the noise can be linearly estimated from measurements of other blocks with less estimated parameters, the variance decreases. The effect of the input spectrum is shown to have a less significant effect than expected. We determine the optimal correlation structure for the noise, for the case when one block has one parameter less than the other blocks.

# Acknowledgements

Many people contributed to the completion of this thesis and I would like to mention a few.

First of all I would like to thank my supervisor Håkan Hjalmarsson for taking me on as a PhD student and guiding me along the way.

I would like to thank my co-supervisor Cristian Rojas for his endless knowledge, patience and insight.

I would like to thank my collaborators Jonas Mårtensson and Giulio Bottegal.

Many thanks to Afrooz Ebadat, Håkan Terelius and Patricio Valenzuela for the help with proofreading.

I would like to thank all past and present colleagues at the department of Automatic for all fruitful discussions, fun times and making the department a wonderful place to work.

I would like to thank the administrators Anneli, Hanna, Karin, Kristina and Gerd for making the department run smoothly.

A special thanks to my family and friends for your support and providing me with a meaningful life outside of research.

# Contents

# Notation

| | |
|---|---|
| $A^*$ | Complex conjugate transpose of $A$ |
| $A^{\mathrm{T}}$ | Transpose of $A$ |
| $A^\dagger$ | Moore-Penrose pseudoinvers of $A$ |
| $\mathrm{AsCov}\, x$ | Asymptotic covariance matrix of $x$ |
| $\mathrm{As}\mathcal{N}(x, P)$ | Asymptotic normal distribution with mean $x$ and covariance matrix $P$ |
| $\mathrm{AsVar}\, x$ | Asymptotic variance of $x$ |
| $\delta(t)$ | Dirac delta function |
| $\det A$ | Determinant of A |
| $\mathrm{diag}\,\{v\}$ | Matrix with the entries of the vector $v$ in the main diagonal |
| $\dim X$ | Dimension of $X$ |
| $\overline{A}$ | Conjugate of $A$ |
| $\mathbf{P}_{\mathcal{S}_\mathcal{X}}\{Y\}$ | Projection of $Y$ onto the space $\mathcal{S}_\mathcal{X}$ |
| $\mathrm{rank}\,\{A\}$ | Rank of A |
| $\mathrm{Tr}\, A$ | Trace of $A$ |
| $f'(\theta)$ | Gradient of $f$ with respect to $\theta$ |
| $\mathbb{C}^n$ | Set of $n$-dimensional column vectors with complex entries |
| $\mathbb{C}^{n\times q}$ | Set of $n \times q$ dimensional matrices with complex entries |
| $\mathbb{C}$ | Set of complex numbers |
| $\mathbb{R}^n$ | Set of $n$-dimensional column vectors with real entries |
| $\mathbb{R}_{\geq 0}$ | Set of non-negative real numbers, $\{x \in \mathbb{R} : x \geq 0\}$ |
| $\mathbb{R}$ | Set of real numbers |
| $\mathbb{Z}_{\geq 0}$ | Set of non-negative integers |
| $\mathbb{Z}$ | Set of integers |
| $:=$ | Definition |
| $\hat{\theta}$ | Estimated parameter vector |
| $\hat{y}(t\|t-1,\theta)$ | Mean square optimal one-step ahead predictor |
| $\langle f, g\rangle$ | Inner product between $f$ and $g$: $\langle f, g\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\omega})g^*(e^{i\omega})\,\mathrm{d}\omega$ |
| $\|f\|$ | The $\mathcal{L}_2$-norm of $f$: $\|f\| := \sqrt{\mathrm{Tr}\,\langle f, f\rangle}$ |
| $\theta^o$ | True parameter vector |
| $\theta$ | Model parameter vector |
| $q$ | Time shift operator, $qu(t) = u(t+1)$ |

| | |
|---|---|
| $\mathcal{H}_2$ | The vector space of all functions with finite $\mathcal{L}_2$-norm that are analytic on the unit disc |
| $\mathcal{L}_2$ | The vector space of all functions with finite $\mathcal{L}_2$-norm |
| $\text{Span}\,\{\mathcal{X}\}$ | Linear span of the rows of the rows of $\mathcal{X}$ |
| $\mathcal{S}_\mathcal{X}$ | Linear span of the rows of the rows of $\mathcal{X}$, i.e., $\mathcal{S}_\mathcal{X} = \text{Span}\,\{\mathcal{X}\}$ |
| $\mathcal{X} + \mathcal{Y}$ | $\{x + y : x \in \mathcal{X}, y \in \mathcal{Y}\}$ |
| $\oplus$ | $\mathcal{Z} = \mathcal{X} \oplus \mathcal{Y}$ is the direct sum of $\mathcal{X}$ and $\mathcal{Y}$, i.e., $\mathcal{Y}$ is the orthogonal complement of $\mathcal{X}$ in $\mathcal{Z}$ |
| $\ominus$ | $\mathcal{Y} \ominus \mathcal{X}$ is the orthogonal complement of $\mathcal{X}$ in $\mathcal{X} + \mathcal{Y}$, i.e., $\mathcal{X} + \mathcal{Y} = \mathcal{X} \oplus (\mathcal{Y} \ominus \mathcal{X})$ |

# Abbreviations

| | |
|---|---|
| AR | autoregressive |
| ARMA | autoregressive moving average |
| ARMAX | autoregressive moving average with exogenous input |
| ARX | autoregressive with exogenous input |
| BCLIV | basic closed-loop instrumental variable |
| BL | band limited |
| CRLB | Cramér-Rao lower bound |
| FIR | finite impulse response |
| IO | input output |
| IV | instrumental variable |
| LTI | linear time invariant |
| MIMO | multiple input multiple output |
| MISO | multiple input single output |
| OE | output error |
| PEM | prediction error method |
| SIMO | single input multiple output |
| SISO | single input single output |
| ZOH | zero order hold |

# Chapter 1

# Introduction

Systems in engineering are becoming ever more complex and interconnected, due to advancing technology with cheaper sensors and increased connectivity. For example in process industry, sensors that monitor the operation of the plant can be connected through wireless connections and used for monitoring and control. Other examples are to be found in various engineering disciplines, e.g., power systems, telecommunication systems, process manufacturing and distributed control systems. Accurate models are of course essential to analyze and control these systems. Furthermore, these networks of modules are not necessarily static in structure, as new users, actuators and sensors might join to the network after commissioning. When a module is added to the network, there is possibly some unknown dynamics in the network that needs to be identified. However, we should not have to identify the whole network all over. We thus need to incorporate the network structure in our models so that we do not have to throw away all we know about a system when only a minor part has changed. This is not the only benefit from taking a network perspective, we can also retain some of the physical interpretation in the model of the system and improve the accuracy of the models. In this chapter we will provide a few high level examples of dynamic networks, give a literature review concerning identification in dynamic networks and outline the contribution of this thesis.

## 1.1 Motivating examples of dynamic networks

In this section, as a motivation, we provide a few examples where it is natural to model the system as a dynamic network.

### Water distribution system

The first example of a dynamic network is taken from water supply systems. The water supply system is a network of components that together provide water to consumers (which may be residential, industrial, commercial or governmental insti-

tutions). We consider pressure control of the pipe network, which transfers water from the set of supply nodes, e.g., water purification facilities or water storage facilities, to the consumers. The goal of the pressure control is to have enough pressure in each node of the network to supply sufficient amount of water to consumers in the network. At the same time, water is a scarce resource in many parts of the world, and too high pressure increases the risk of pipes breaking and increases the amount of leaking water. Furthermore, considerable amount of energy can be saved with efficient pressure control. Thus, proper management of the pressure in the network is needed for safe and effective operation. The system can be modeled as a set of pipes, storage tanks, consumer nodes and supply nodes. The pipes transport the water between different nodes in the network. We can describe the dynamics of a tank in the network as

$$\begin{aligned} \dot{V}_i &= \sum_{j \in \mathcal{N}_i} q_{ij} \\ q_{ij} &= \left( \frac{p_i - p_j}{R_{ij}} \right)^{1/a}, \end{aligned}$$

where $V$ is the volume in the tank located in node $i$, $q_{ij}$ is the flow in the pipe between nodes $i$ and $j$, $p_i$ is the pressure at node $i$, $\mathcal{N}_i$ is the set of neighbors of node $i$, $a$ is the flow exponent and $R_{ij}$ is the resistance coefficient. If the pressure can be controlled in a node $i$, we consider the pressure $p_i$ a control input, otherwise it is modeled as determined by the pressure at neighboring nodes.

### Flotation plant

The second example of a dynamic network is concentration of ore, using froth flotation. Froth flotation is a process that is used to separate hydrophobic from hydrophilic materials and is common in processing industries. In the flotation plant, the objective is to separate the valuable mineral from ore, while minimizing the amount of undesired minerals in the extracted concentrate. At the same time, as much of the mineral as possible should be harvested. Thus, the residual tailings should mainly be composed of finely ground waste rock. In a flotation cell, froth flotation is done by adding certain chemical reagents to render the desired mineral hydrophobic, so air bubbles then lift the mineral. The resulting froth layer is then skimmed to produce the concentrate. Normally a flotation process consists of several flotation cells connected in cascade together with cyclones, mills, and mixing tanks as seen in the schematics of a typical plant in Figure 1.1. The plant can be described as a network composed of interconnected systems with simple dynamics.

#### Steam pressure control

Industrial power plants are often equipped with several parallel boilers with controlled and close to constant loads. Together they produce steam at high pressure

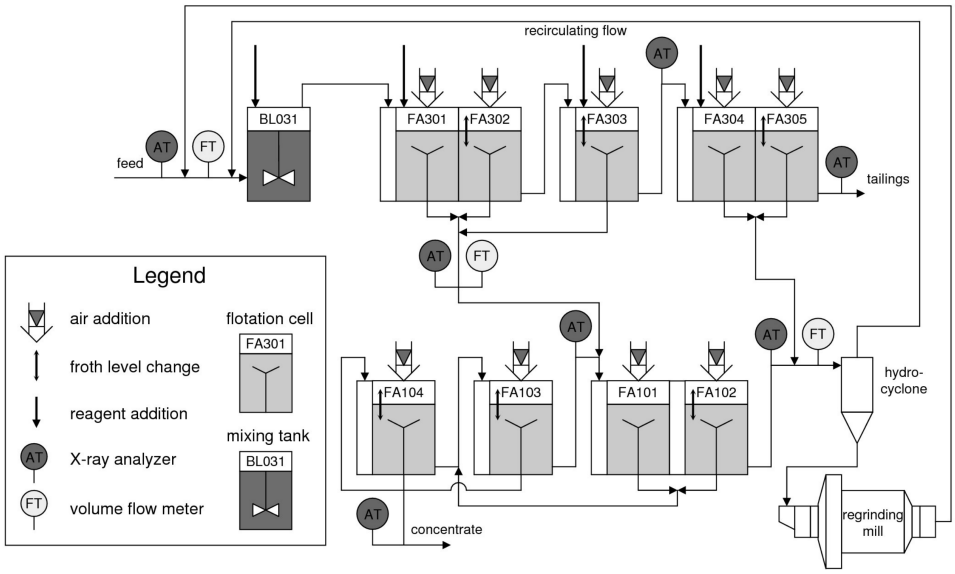


Figure 1.1: Schematic of a floatation plant with several cascades of flotation cells. (Image courtesy of ABB).

that will be used in the plant. However, steam is consumed at intermediate and low pressure in one intermediate pressure (IP) header and one low pressure (LP) header respectively, with rapid and large variations in consumption (Majanne, 2005). In the middle a set of back pressure turbines feed the IP header and LP header appropriate amounts of steam respectively. In order to control the pressure in the IP header and LP header, it is necessary to accurately control the pressure in the high pressure (HP) header as well. The boilers should, for energy effectiveness, operate at close to constant loads. Therefore, an accumulator tank is used to handle the rapid fluctuations in load. When the demand for steam is low, the accumulator tank may store steam, as long as its internal pressure is less than the pressure in the IP header. When demand is high, the steam tank may discharge steam into the LP header as long as the steam tanks internal pressure is higher than the pressure in the LP header. The plant can be modeled as in Figure 1.2, where neither the demand nor the feedback structure of the control system has been modeled.

## 1.2 Identification of dynamic networks

The main point of the previous section was to show that there are many examples of processes which we may model as an interconnection of simpler systems in a dynamic network. We will call such a simple system a module. From the system

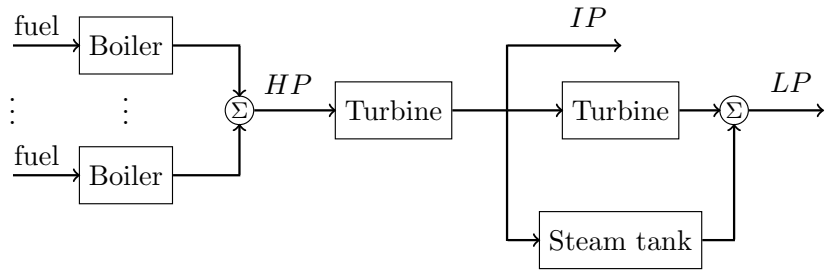


Figure 1.2: Steam pressure control model.

identification literature, there seem to be two separate tasks related to the network: identify the structure of the network and identify the modules that make up the network. The first task is referred to as *topology detection* and is only briefly mentioned in this chapter. In the second line of research the structure of the interconnections are assumed to be known, and this is the setting for this work.

### Topology detection

Topology detection is very much an active research topic with a rich literature. The topic is intrinsically linked with the notion of causality, since a transfer function in the network structure determines that there exists a causal link between the variable on the input side and the variable at the output side. The topic dates back to at least Granger (1969); Wiener (1956), with early contributions by Caines and Chan (1975); Granger (1980); and Anderson and Gevers (1982). Some more recent contributions that consider dynamic networks are given in Bottegal and Picci (2014); Hayden *et al.* (2014a); Marques *et al.* (2013); Materassi and Salapaka (2012); Materassi *et al.* (2011); Sanandaji *et al.* (2012); and Hayden *et al.* (2014b)

### Identification in dynamic networks

Identification in dynamic networks is fundamentally different from unstructured multiple input multiple output (MIMO) identification for several reasons. Firstly, we may only wish to model a module or a subset of modules within the dynamic network. Using methods tailored for this task imposes less restrictions on the excitation and modeling and we gain flexibility in which signals we need to measure, see e.g., Van den Hof *et al.* (2013). Secondly, as discussed in Hägg *et al.* (2011), a considerable variance reduction is possible if we incorporate previous knowledge about the modules. We may also have known feedback structures in the dynamic network, i.e., some of the modules may be known. Lastly, by explicitly taking the network structure into account, we preserve the physical interpretation of the modules. However, if the sampling is not done fast enough the physical structure

might be lost in the modeling. This is because we measure the signals in the network only at a certain rate and assume that the signals change much more slowly, so that they can be approximated to be constant between measurements. This point is further discussed in Section 2.1.

### Can we apply MIMO methods?

In the prediction error method (PEM) it is possible to incorporate previous structural knowledge. However, PEM requires solving a non-convex optimization problem and thus there is a risk of getting stuck in a local minimum. For other classical methods such as the subspace method it is not straightforward how to impose a certain structure. One result, considering OE models and autoregressive moving average with exogenous input (ARMAX) models, is found in Lyzell *et al.* (2009). Another approach is to identify a full MIMO model and then use $\mathcal{H}_\infty$ model reduction techniques to make the reduced order model conform to the interconnection structure (Sturk *et al.*, 2011, 2012; Wahlberg and Sandberg, 2008). In Sandberg *et al.* (2014), it was shown that for cascaded systems the approximation error could be bounded by the weighted Hankel singular values. This low order model can then be used as initialization for PEM. However, for large networks the computational complexity grows prohibitively large, and other methods are needed.

### Identification of the whole network

For very large networks there is a line of research that tries to identify the whole network when the interconnection structure is known. To do so, assumptions are often made on the network, either that it is composed of identical modules or that modules are from spatially distributed systems where the modules are only connected to a few neighbors. This structure is motivated by practical applications were partial differential equations have been discretized, for example heat conduction or flexible structures. Since the objective is to identify large networks, the focus is on algorithms that scale well with the number of modules, which renders traditional methods out of the question since they scale badly with the number of modules. A key feature in the considered problem formulation is that the sensor noise is local in nature and does not propagate in the network, i.e., there is in general neither feedback nor unmeasured disturbances present. The assumption that the modules are identical is made in Massioni and Verhaegen (2008, 2009). In Ali *et al.* (2011a,b, 2009, 2011c), more complex noise structures, parameter varying modules, as well as the closed loop case are considered. Using a special type of matrices, called Sequentially Semi-separable matrices, a linear scaling in time complexity is achieved (Torres *et al.*, 2014; van Wingerden and Torres, 2012). Any matrix can be written in this form; however, if the order of the Sequentially Semi-separable matrices grows large, the nice scaling in the number of modules is lost. Under the assumption that the systems only interact with spatially close neighbors, Haber

and Verhaegen (2012, 2014) propose a distributed algorithm to efficiently estimate all modules.

### Identification of a module

The main interest in this work is to identify a module or a set of modules in the network. Recently, this topic has gained popularity, see e.g., Chiuso and Pillonetto (2012); Dankers *et al.* (2013a, 2014b, 2013b); Gunes *et al.* (2014); Van den Hof *et al.* (2013). Also in this case the interconnection structure is assumed known. To estimate a transfer function in the network, a large number of methods have been proposed. Some have been shown to give consistent estimates (Dankers *et al.*, 2013a,b), provided that a certain subset of signals is included in the identification process. In these methods, the user has the freedom to include additional signals. However, little is known about how these signals should be chosen, and how large the potential is for variance reduction. In Van den Hof *et al.* (2013), under the assumption that there is no measurement noise, the direct method and joint input-output method (Ljung, 1999) are generalized and conditions are given under which the methods give consistent estimates. The conditions are that the noise signals are mutually uncorrelated and uncorrelated with the reference signals. The spectral density of the measured internal variables should be positive definite (persistence of excitation) and both the model set for the modules and noise models have to be flexible enough to capture the true system. The two-stage method and the instrumental variable (IV) method (Ljung, 1999) are also generalized in Van den Hof *et al.* (2013). However, in contrast, these methods rely more on the external reference signals. The persistence of excitation condition in this case concern the spectral density of the measured internal variables projected onto the reference signals. The condition is that the spectral density of the projected internal variables needs to be positive definite. The benefit is that it is not necessary to include noise models. Some flexibility in how to choose the set of signals to use in the predictor is introduced for the direct method in Dankers *et al.* (2013b) and for the two-stage method in Dankers *et al.* (2013a). Conditions are given on the set of signals in order to achieve consistency. Sensor noise is added to the framework in Dankers *et al.* (2014b) and three generalizations of the basic closed-loop instrumental variable (BCLIV) method of Gilson and Van den Hof (2005) are presented. However, an assumption is made that sensor noise does not propagate to the other internal variables in the network, i.e., the noise term enters after the signal is fed back. Hence there is an assumption that excludes physical systems were feedback is used on measured signals. The analysis of the accuracy of the presented methods is far less developed than the consistency analysis. Before discussing what has been done in is this regard, we will pause for a moment and consider how one can analyze the accuracy of a model. We will come back to this issue towards the end of this section.

## 1.3 Model accuracy

Depending on which method we use, we have to include some measurements and inputs to achieve consistent models. However, consistency is not the whole story. In order for us to trust a model that our identification algorithm gives us, we need some kind of guarantee that the model lies sufficiently close to the true module. To analyze the accuracy, mainly two approaches are possible, namely, either the error sources are regarded as stochastic or they are regarded as deterministic. Considering the errors as stochastic leads to confidence regions of the model error (see, Goodwin and Payne (1977); Ljung (1999); Söderström and Stoica (1989)). In the deterministic case, hard bounds on the model error can be given in the frequency domain or in the time domain (see e.g., Milanese and Novara (2011); Milanese and Vicino (1991); Ninness and Goodwin (1995); Wahlberg and Ljung (1992)).

### Confidence ellipsoids

This thesis takes the classical stochastic approach where we describe the accuracy of the estimated parameters $\hat{\theta}$ as confidence ellipsoids around the true system parameters $\theta^o$. This is motivated by that, under reasonable assumptions (see Ljung (1999) for details), as the number of measurements $N$ grows large, the random variable $\sqrt{N}(\hat{\theta}-\theta^\circ)$ converges in distribution to a Gaussian random variable with zero mean and covariance matrix $P$. For finite data, the covariance matrix is approximated as

$$\text{Cov}\,(\hat{\theta}-\theta^\circ) \approx \frac{1}{N}P. \tag{1.1}$$

In some cases, we may not be interested in the model parameters themselves, but in some system theoretic property $J$, e.g., the frequency response function or the system poles. Assuming sufficient smoothness of $J$ (with respect to the parameters $\theta$) and bounded moments of sufficiently high order of the noise, it follows that

$$\sqrt{N}(J(\hat{\theta}_N) - J(\theta^o)) \in \text{As}\mathcal{N}(0, \text{AsCov}\; J(\hat{\theta}_N)), \tag{1.2}$$

where, using Gauss' approximation formula (also known as the delta method (Casella and Berger, 2002)) (Ljung, 1999), it can be shown that $\text{AsCov}\,J(\hat{\theta}_N)$ in (1.2) is given by

$$\text{AsCov}\; J(\hat{\theta}_N) = J'(\theta^\circ)P\overline{J}'(\theta^\circ). \tag{1.3}$$

### Motivation of asymptotic results

The asymptotic variance is based on an assumption that the amount of input-output data available tends to infinity. The assumption is that for large enough $N$, (1.2) provides a reasonable approximation. In Garatti *et al.* (2004), some answers on when this is a reasonable assumption are given. Non-ellipsoidal confidence regions are considered in Bombois *et al.* (2009), which seem to give better approximations

for small $N$, but are still based on a large number of samples. An interesting approach for non-asymptotic confidence regions has been developed in Campi and Weyer (2005, 2010); Csáji *et al.* (2012a,b); Kolumbán *et al.* (2015). From this approach, promising methods based on hypothesis testing are emerging, which provide non-asymptotic confidence regions under mild assumptions on the noise distribution (Csáji *et al.*, 2012a,b; Kolumbán *et al.*, 2015). In Csáji *et al.* (2012a,b), the noise terms are assumed independent and symmetrically distributed around zero, and in Kolumbán *et al.* (2015) it is shown that the symmetry requirement may be replaced by exchangeability. The non-asymptotic case is also considered in (Douma, 2006; Hjalmarsson and Ninness, 2006).

### Parameter accuracy

The covariance matrix $P$ in (1.3) has a long history of study. Early on, it was realized that some scalar measures of the covariance matrix $P$, e.g., the determinant of $P$ and weighted trace $\mathrm{Tr}\,(WP)$, grow with the model order, see for example Box and Jenkins (1976); Gustavsson *et al.* (1977). Some more recent results can be found in (Aguero and Goodwin, 2006; Agüero and Goodwin, 2007; Bombois *et al.*, 2005; Forssell and Ljung, 1999). In these contributions it is analyzed under which settings open loop or closed loop identification is optimal, e.g., it is shown that under input constraints open loop identification is preferred (Aguero and Goodwin, 2006), while under output power constraints, typically closed loop is better (Aguero and Goodwin, 2006).

### Frequency function estimate

Assuming that $\mathcal{S} \in \mathcal{M}$, the classical open loop variance approximation

$$\mathrm{AsCov}\, G(\mathrm{e}^{\mathrm{j}\omega}, \hat{\theta}_N) \approx \frac{n}{N} \frac{\Phi_v(\omega)}{\Phi_u(\omega)} \tag{1.4}$$

was derived in Ljung (1985). The expression tells us that the variance of the estimated frequency response function $\hat{G}$ evaluated at the frequency $\omega$, depends on the noise spectrum to signal spectrum ratio at that frequency. The variance increases linearly with the model order $n$. The result (1.4) is only valid when both $n$ and $N$ go to infinity. For finite model order, the expression can be quite misleading. Refinements of (1.4) can be found in Hildebrand and Gevers (2004); Hjalmarsson and Ninness (2006); Ninness and Hjalmarsson (2004); Wahlberg *et al.* (2012); Xie and Ljung (2001, 2004). Variance expressions that are exact for finite model order are derived for some model structures in Hjalmarsson and Ninness (2006); Ninness and Hjalmarsson (2004); Xie and Ljung (2001, 2004) and expressions for spectral estimates of AR Models are presented in Xie and Ljung (2004). For closed loop identification, variance expressions that are exact for finite model order are derived in Ninness and Hjalmarsson (2005a,b). It is also known that the variance of the

frequency response function satisfies a waterbed effect, similar to the Bode integral (Rojas *et al.*, 2009).

### Geometric approach

The geometric approach, the main analysis tool used in this thesis, was developed in Hjalmarsson and Mårtensson (2011), where the asymptotic variance of a smooth function of the model parameters, is expressed as an orthogonal projection onto the subspace spanned by the predictor error gradient (see also Mårtensson (2007)). The importance of this subspace, and the geometric properties of prediction error estimates were first recognized in a series of papers (Ninness and Hjalmarsson, 2004, 2005a,b). The geometric analysis has since been applied to analyze a series of settings: In Mårtensson and Hjalmarsson (2009) the variance of identified poles and zeros are quantified and it is shown that non minimum phase zeroes and unstable poles can be estimated with finite variance, even when the model order tends to infinity; in Hjalmarsson *et al.* (2011) the difference in variance between the error-in-variables setting compared to when the input noise is zero is studied; when the minimum variance controller is the optimal experiment for closed loop identification is addressed in Mårtensson *et al.* (2011); some results on optimal and robust experiment design are presented in Mårtensson and Hjalmarsson (2011). The geometric approach has also been applied to MISO systems (Ramazi *et al.*, 2014).

### MIMO results

There are far less results that consider the variance of MIMO system estimates. The paper Agüero *et al.* (2012) shows that the variance of the parameter estimates satisfies a waterbed effect. For fixed denominator models, it is not possible to simultaneously minimize both the bias error and the variance error at a particular frequency (Ninness and Gómez, 1996). In Bazanella *et al.* (2010), it is established that it is not necessary to excite all inputs in closed loop control of MIMO systems, provided the controller is sufficiently complex and the noise sufficiently exciting. It is however preferable to excite all inputs at the same time in MIMO identification (Mišković *et al.*, 2008).

### Accuracy in identification of a module

There are but a few results that try to quantify the variance of different methods when a module in a network is estimated. In general, the available results are restricted to special cases of networks, in order to be able to say something meaningful. Cascaded modules are such a special case considered in Wahlberg *et al.* (2009). The main result is that measurements downstream do not improve the variance of the estimate of the first module, if the modules are identical and have the same parametrization. Cascaded modules are also considered in Chapter 4. In Hägg *et al.* (2011), a generalization of cascade modules is considered. Here, the effect of

sensor placement, input spectrum, and common dynamics is considered. The same structure is considered in Chapter 5, where high order models are used for some of the modules. In a MISO system setting, the paper Gevers *et al.* (2006) studies which parameters are identified with decreased variance when an input signal is added (the added input is considered to be zero to start with). For MISO systems, it is shown in Ramazi *et al.* (2014) that the estimation accuracy decreases when inputs are correlated, and it is shown how this effect depends on the correlation structure and the model structure. The above contributions all consider a direct approach. A technique to reduce the variance of a two stage method is presented in Gunes *et al.* (2014). The two step method first tries to obtain estimates of the inputs to the module of interest, and in a second step the module of interest is estimated (Van den Hof *et al.*, 2013). The main idea in Gunes *et al.* (2014), is to simultaneously minimize the prediction error of the two steps in the two step method. This leads to a variance reduction compared to the two stage method, however, it is not clear how large the reduction will be.

### Problem formulation

As seen from the literature review, rapid progress is made when it comes to developing new methods for identification of modules in dynamic networks. However, a thorough analysis of the accuracy of the proposed methods is often lacking. Similar to the expression (1.4), the contribution of different system properties (e.g., model structures, model order), interconnection properties (feedback connections, cascades, parallel branches) and signal properties (noise covariance, input spectra, noise spectra), should be as clearly distinguishable as possible. Ideally, given a dynamic network and a set of signals to include in the estimation, we would like to provide a variance quantification that gives insight into how system properties, interconnection properties and signal properties influence the asymptotic variance. With this information, we could understand several important questions regarding identification in dynamic networks, such as, what would be the effect on the variance of an estimated module if:

- another output of the network is measured,
- a disturbance is measured and included in the predictor,
- the model order of a module is increased,
- the power spectrum of a reference signal is changed.

To show that interesting phenomena appear in the dynamic network setting, let us introduce the following one-input-two-output example.

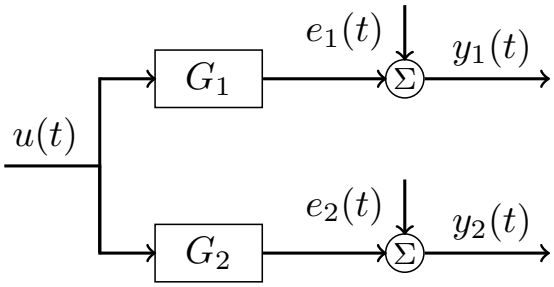


Figure 1.3: Two parallel modules with the same input.

**Example 1.3.1.** *Consider the model visualized in Figure 1.3, where the system dynamics is captured in the following equations:*

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

*where the input $u(t)$ is white noise and we consider two different types of noise (uncorrelated with the input). In the first case, the noise is perfectly correlated. Let us for simplicity assume that $e_1(t) = e_2(t)$. For the second case, $e_1(t)$ and $e_2(t)$ are independent. It turns out that in the first case we can perfectly recover $\theta_{1,1}$ and $\theta_{2,2}$, while, in the second case we do not improve the accuracy of the estimate of $\theta_{1,1}$ by also using the measurement $y_2(t)$. The reason for this difference is that, in the first case, we can construct the noise free equation*

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

*and we can perfectly recover $\theta_{1,1}$ and $\theta_{2,2}$, while in the second case the noise signals do not cancel. We will hint that also the model structure plays an important role in determining the benefit of the second sensor. To this end, we consider a third case, where again $e_1(t) = e_2(t)$. But this time, the model structure is slightly different:*

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

*In this case, we can construct the noise free equation*

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

*The fundamental difference is that now only the difference $(\theta_{1,1} - \theta_{2,1})$ can be perfectly recovered, but not the parameters $\theta_{1,1}$ and $\theta_{2,1}$ themselves, but they can be identified from $y_1(t)$ and $y_2(t)$ separately. A similar consideration is made in Ljung et al. (2011), where SIMO cascade systems are considered.*

## 1.4 Contribution and outline

This section gives an outline of the chapters contained in the thesis and the corresponding publications. The results developed in this thesis rely on the geometric approach, and the results reported in Chapter 3 owe a lot to the work of my co-authors.

### Chapter 2 – Background

A background is given on system identification and prediction error identification. The chapter also contains properties of Hilbert spaces, orthogonal functions and some results from the geometric approach to variance analysis.

### Chapter 3 – SISO models

In this chapter, expressions for the asymptotic (co)variance of system properties are derived for causal single input single output linear time invariant systems. It can be considered the minimal example of a dynamic network and will serve as a springboard for subsequent chapters, which consider other network structures. A connection is established to results on frequency response function estimation. Variance expressions and bounds are provided for common system properties such as impulse response coefficients and non-minimum phase zeros. As an illustration of the insights the expressions provide, they are used to derive conditions on the input spectrum which makes the asymptotic variance of non-minimum phase zero estimates independent of the model order and model structure. This chapter also serves as a review of the variance analysis provided by the geometric approach for variance analysis. Indeed, most of the results in this chapter can be found in Mårtensson (2007). The chapter is based on the publication:

J. Mårtensson, N. Everitt, and H. Hjalmarsson. 2015. Variance analysis in SISO linear systems identification. *Automatica*. Submitted

### Chapter 4 – Cascade models

Cascaded modules are considered in this chapter. We quantify the accuracy improvement from additional sensors when estimating the first of a set of modules connected in a cascade structure. We present results on how the zeros of the first module affect the accuracy of the corresponding model. The results are illustrated on finite impulse response (FIR) systems. The chapter is based on the publication:

N. Everitt, C.R. Rojas, and H. Hjalmarsson. 2013. A geometric approach to variance analysis of cascaded systems. In *Proceedings of the 52nd IEEE Conference on Decision and Control*

### Chapter 5 – Generalized parallel cascade models

Two types of generalized cascaded modules are considered in this chapter. The first structure may represent a system where several actuators with unknown dynamics are used to excite a system. Upper and lower bounds are provided for the variance of the estimated plant dynamics. The second structure may represent a sensor network where additional sensors are used to increase the accuracy of the estimated plant dynamics. Again, upper and lower bounds are provided for the variance of the estimated plant dynamics. The chapter is based on the publication:

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

### Chapter 6 – SIMO models with spatially correlated noise

In this chapter the effect of the noise correlation structure is examined in SIMO models. It is shown how the estimation accuracy depends on the correlation structure of the noise, model structure and model order. A formula for the asymptotic covariance of the frequency response function estimates and the model parameters is developed for the case of temporally white, but possibly spatially correlated additive noise. It is shown that when parts of the noise can be linearly estimated from measurements of other blocks with less estimated parameters, the variance decreases. The expressions reveal how the order of the different blocks and the correlation of the noise affects the variance of one block. In particular, it is shown that the variance of the block of interest levels off when the number of estimated parameters in another block reaches the number of estimated parameters of the block of interest. We show that the effect of the input spectrum is less significant effect than expected. The optimal correlation structure for the noise is determined for the case when one block has one parameter less than the other blocks. The chapter is based on the publication:

N. Everitt, G. Bottegal, C.R. Rojas, and H. Hjalmarsson. 2015. Variance analysis of linear SIMO models with spatially correlated noise. *Automatica*. Submitted

### Chapter 7 – Conclusions

In the final chapter, we draw some conclusions and outline directions for future work.

### Author's Contributions

The contributions of the thesis are principally the results of the author's own work, in collaboration with the respective coauthors. The order of the authors reflect their contributions in the mentioned papers. As mentioned above, most of the results in the single input single output (SISO) chapter can be found in Mårtensson (2007).

# Chapter 2

# Background

This thesis concerns the accuracy of models identified from system identification experiments. The objective of this chapter is to provide a theoretical background for the results presented in this thesis. The system identification method considered is prediction error identification. The geometric analysis that is instrumental in this thesis is based on Hilbert space theory and a brief background is provided in the second part of this chapter.

## 2.1 System identification

System identification concerns building mathematical models from observed data from the system. The mathematical model should provide a good approximation of the behavior of the system relevant to the intended use of the model, e.g. simulation, prediction or control. We consider linear time invariant (LTI) dynamic systems with $m$ inputs collected in a column vector $u_c(t)$ and $p$ outputs collected in a column vector $y_c(t)$ (the subscript $c$ denotes continuous time, which we will drop when we move to discrete time). The output of the system can be described by the relation

$$y_c(t) = \int_{-\infty}^{\infty} g(t-\tau)u_c(\tau)\mathrm{d}\tau$$

where $g(t)$ is the impulse response. We collect measurements of the output of the system at equidistant samples with sample time $T_s$. We thus have samples $y(t) = y(kT_s)$, $k = 1, \ldots,$ according to

$$y(t) = y(kT_s) = \int_{-\infty}^{\infty} g(kT_s - \tau)u(\tau)\mathrm{d}\tau. \tag{2.1}$$

We assume that we also have access to the control signal at these samples, i.e., $u_k = u(kT_s)$. The inter sample behavior of the input signal is assumed to satisfy a zero order hold (ZOH) assumption. Under this assumption the signal $u$ is assumed to

be constant in between sample instances, i.e.

$$u(t) = u(kT_s), \qquad (k-1)T_s \leq t \leq kT_s. \tag{2.2}$$

For convenience, we let $t$ enumerate the sampling instances. Under the ZOH assumption, the output samples $y(t)$ given by (2.1) can be written as

$$y(t) = \sum_{l=1}^{\infty} g_l u(l)(t-l), \qquad t = 0, 1, \ldots \tag{2.3}$$

where

$$g_l = \int_{(l-1)T_s}^{lT_S} g(\tau)\mathrm{d}\tau.$$

Introducing the time shift operator $q$, defined by

$$qy(t) = y(t+1),$$

(2.3) can be written as

$$y(t) = G_o(q)u(t),$$

where $G_o(q)$ is the transfer function $G_o(q) = \sum_{l=1}^{\infty} g_l q^{-l}$. However, in practice, we cannot measure the output of the system exactly. There are always measurement noise and disturbances acting on the system. These are modeled as a zero mean white noise signal $e(t)$ with variance $\Lambda$, filtered through an inversely stable filter $H_o(q)$, so that our basic description of a (discrete) linear system is

$$y(t) = G_o(q)u(t) + H_o(q)e(t). \tag{2.4}$$

### Prediction error identification

We model the system (2.4) as

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t), \tag{2.5}$$

where $G(q,\theta)$ and $H(q,\theta)$ are rational functions which are parametrized with the parameter vector $\theta \in \mathbb{R}^n$. Thus, (2.5) describes a set of models. For a given $\theta$, (2.5) can be used to predict the future output of the system given past samples of the output and input. The mean square optimal one-step ahead predictor is the conditional expectation, denoted by $\hat{y}(t|t-1,\theta)$ and is given by

$$\hat{y}(t|t-1,\theta) = H^{-1}(q,\theta)G(q,\theta)u(t) + \left(I - H^{-1}(q,\theta)\right)y(t). \tag{2.6}$$

The prediction error is

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t|t-1,\theta). \tag{2.7}$$

The objective is to find the model within the set of models that most accurately describes the system. In PEM, the most accurate model is the one that minimizes a cost function $V_N(\theta)$ based on the prediction errors of the observed inputs and outputs. The cost function is a sum of some scalar norm $l$ of the prediction error, i.e.

$$V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} l\left(\varepsilon(t,\theta)\right). \tag{2.8}$$

In this work, a quadratic norm will be used with the noise variance as weighting matrix $\Lambda^{-1}$, i.e.

$$l\left(\varepsilon(t,\theta)\right) = \frac{1}{2}\varepsilon^{\mathrm{T}}(t,\theta)\Lambda^{-1}\varepsilon(t,\theta). \tag{2.9}$$

This norm uses the (usually) unknown noise (co)-variance. However, this covariance can be estimated from the data. Since we are interested in the asymptotic properties of the estimates $\hat{\theta}$, it is interesting to note that minimizing the cost function

$$\det\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta)\varepsilon(t,\theta)^{\mathrm{T}}\right] \tag{2.10}$$

gives the same asymptotic covariance matrix as minimizing (2.8) with the norm (2.9) based on the true noise covariance.

### Statistical properties of the estimates

We assume that the model is in the model set, i.e., there exists a $\theta^o$ such that $G_o(q) = G(q,\theta^o)$ and $H_o(q) = H(q,\theta^o)$. 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 the normal distribution with zero mean and covariance matrix $\operatorname{AsCov}\hat{\theta}_N$, which we conveniently denote by

$$\sqrt{N}(\hat{\theta}_N - \theta^o) \in \operatorname{As}\mathcal{N}(0, \operatorname{AsCov}\hat{\theta}_N). \tag{2.11}$$

where

$$\begin{aligned}\operatorname{AsCov}\hat{\theta}_N &:= \left[\mathrm{E}\left[\psi(t,\theta^o)\Lambda^{-1}\psi^{\mathrm{T}}(t,\theta^o)\right]\right]^{-1},\\ \psi(t,\theta^o) &:= \frac{\mathrm{d}}{\mathrm{d}\theta}\varepsilon(t,\theta)\Big|_{\theta=\theta^o}.\end{aligned} \tag{2.12}$$

In following chapters we will express (2.12) as

$$\operatorname{AsCov}\hat{\theta}_N = \langle \Psi, \Psi\rangle^{-1}, \tag{2.13}$$

where $\Psi$ will depend on the problem at hand. Often, we are not interested in the covariance of the parameters themselves, but in some "system theoretic" quantity.

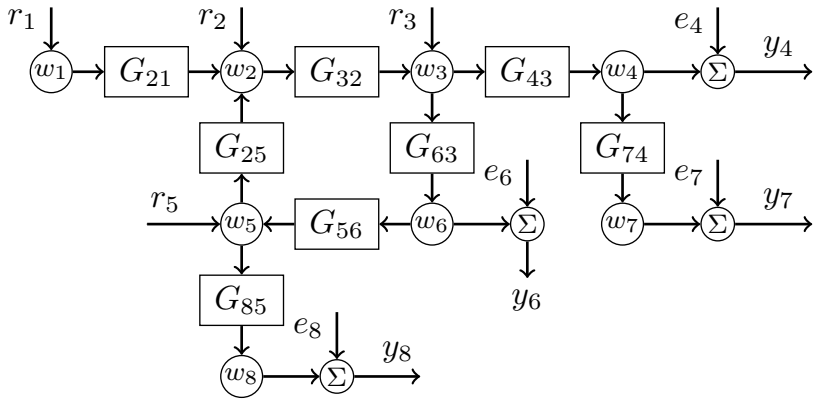


Figure 2.1: An example of a dynamic network. The internal variables $\{w_k\}$ are described by the dynamics (2.16), where $\{r_k\}$ is the set of reference signals. The set of measurements $\{y_k\}$ are described by (2.18), where $\{e_k\}$ is the measurement noise.

Let $J : \mathbb{R}^n \to \mathbb{C}^{1\times q}$ be a differentiable function of $\theta$ such that $J(\theta^o)$ is the quantity of interest. Assuming sufficient smoothness of $J$ (with respect to $\theta$), bounded noise moments of sufficiently high order, and using (2.11), it follows that

$$\sqrt{N}(J(\hat{\theta}_N) - J(\theta^o)) \in \text{As}\mathcal{N}(0, \text{AsCov}\, J(\hat{\theta}_N)). \tag{2.14}$$

where, using Gauss' approximation formula (or the delta method (Casella and Berger, 2002)) (Ljung, 1999) and (2.11), it can be shown that[1]

$$\begin{aligned}\text{AsCov}\, J(\hat{\theta}_N) &:= \lim_{N\to\infty} N \cdot \text{E}\left[(J(\hat{\theta}_N - J(\theta^o))^*(J(\hat{\theta}_N - J(\theta^o))\right] \\ &= J'(\theta^o)^* \left\langle \Psi, \Psi \right\rangle^{-1} J'(\theta^o),\end{aligned} \tag{2.15}$$

where $J'(\theta^o) \in \mathbb{C}^{n\times q}$ is the gradient of $J$ with respect to $\theta$.

### Dynamic networks

In this thesis we will work with networks of dynamic systems. We will spend some time in this section to formalize the different quantities in the network and the network structures we consider. Consider as an example the dynamic network given in Figure 2.1. In the network we have a set of internal variables $\{w_k\}$ which encode the states of the network. Their dynamics can be described by

$$w = Gw + r. \tag{2.16}$$

[1] This definition is slightly non-standard in that the second term is usually conjugated. For the standard definition, all results in the thesis have to be transposed.

The example network of Figure 2.1 is described by

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ G_{21} & 0 & 0 & 0 & G_{25} & 0 & 0 & 0 \\ 0 & G_{32} & 0 & 0 & 0 & G_{36} & 0 & 0 \\ 0 & 0 & G_{43} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & G_{56} & 0 & 0 \\ 0 & 0 & G_{63} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & G_{74} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & G_{85} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_8 \end{bmatrix} + \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ 0 \\ r_5 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We could also consider adding process noise $v$ so that the system would be described by

$$w = Gw + r + v. \tag{2.17}$$

However, this case is not considered in this thesis. The internal variables are measured with additive white noise $e$, i.e.,

$$y = w + e = (I - G)^{-1}r + e, \tag{2.18}$$

where the inverse $(I-G)^{-1}$ is assumed to exist. In the example only $\{w_4, w_6, w_7, w_8\}$ are measured so that

$$\begin{bmatrix} y_4 \\ y_6 \\ y_7 \\ y_8 \end{bmatrix} = \begin{bmatrix} w_4 \\ w_6 \\ w_7 \\ w_8 \end{bmatrix} + \begin{bmatrix} e_4 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix} = G_{cl} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_5 \end{bmatrix} + \begin{bmatrix} e_4 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix},$$

for some $G_{cl}$ that depends on the dynamics of the modules in $G$. In the subsequent chapters we will consider examples without feedback loops where $G_{cl}$ is easily parameterized by the individual modules in $G$. The input $\{r(t)\}$ is a zero mean process with finite moments of all orders and power spectrum $\Phi_r(\omega)$. The noise $\{e(t)\}$ is a zero mean temporally white noise process, but may be correlated in the spatial domain:

$$\begin{aligned} \mathrm{E}\left[e(t)\right] &= 0 \\ \mathrm{E}\left[e(t)e(s)^{\mathrm{T}}\right] &= \delta_{t-s}\Lambda, \end{aligned} \tag{2.19}$$

where $\Sigma > 0$ is a positive definite matrix.

### Zero order hold assumption

As discussed above, any modeling effort should take into account the intended use of the model. The assumption in this work is that we have access to one or several discrete control signals and samples of the outputs. In general, we would like to

be able to model networks of continuous time systems. Each of the modules are modeled under the ZOH assumption. The inter sample behavior of signals in the network will however not (in general) satisfy this assumption. For this to hold at least approximately, all systems need to have low pass behavior and we need to sample at a high rate. Thus, in general, discrete time models of the modules will not be good approximations of their continuous time counterparts. The models will still be able to predict the sampled input output behavior of the modules, but the accuracy of a particular model is intrinsically linked to the other modules in the network. Even the network structure may be different from the continuous time network (Dankers *et al.*, 2014a). If the network changes, the model will change too. As discussed in Pintelon and Schoukens (2012, Chapter 13), if the objective is to accurately model the module it would be preferred to take another route using a band limited (BL) assumption on the signal and continuous time models. If the goal of the identification is continuous time models, the IV based method of Dankers *et al.* (2014a) can be used. A signal $u(t)$ with power spectrum $\Phi(\omega)$ is considered BL if $\Phi(\omega) = 0$ for all $\omega > \omega_{\max}$ for some $\omega_{\max}$. If the physical interpretation of the parameters is not the main concern, then the sensor could be equipped with anti aliasing filters (to realize a BL setup), which would violate the ZOH assumption. However, discrete time models would accurately capture the dynamics of the modules, even though the dynamics of the anti-aliasing filters would be part of the models. From a control perspective, this is not that bad. If we are only interested in accurately modeling the behavior of the system up to some frequency, and we sample fast, the error will be small. For a further discussion on these issue we refer to Pintelon and Schoukens (2012, Chapter 13).

## 2.2 Hilbert space fundamentals

Much of the results derived in this thesis has its foundation on Hilbert space theory. In this section we introduce some notation and review a few fundamental results from this theory.

### Inner product

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

$$\langle f, g\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\omega})g^*(e^{i\omega})\,\mathrm{d}\omega \tag{2.20}$$

where $g^*$ denotes the complex conjugate transpose of $g$. In case $f, g$ are matrix valued functions we keep the same notation whenever the matrix dimensions are compatible.

### The spaces $\mathcal{L}_2$ and $\mathcal{H}_2$

We denote by $\|f\|$ the $\mathcal{L}_2$-norm of $f : \mathbb{C} \to \mathbb{C}^{n\times m}$ and it is given by

$$\|f\|=\sqrt{\text{Tr}\,\langle f, f\rangle} \tag{2.21}$$

where Tr denotes the trace operator. The vector space that consists of all functions with finite $\mathcal{L}_2$-norm is denoted $\mathcal{L}_2^{n\times m}$. We call a function $f$ an $\mathcal{L}_2$-function if $f \in \mathcal{L}_2$. The subspace of all $\mathcal{L}_2$-functions that are analytic on the unit disc is denoted by $\mathcal{H}_2$.

### Orthonormal functions

We call two functions $f, g$ orthogonal if $\langle f, g\rangle = 0$; if $f, g$ are matrix valued, they are considered orthogonal if every entry of the resulting matrix is zero. A set of functions $\{\mathcal{B}_k\}_{k=1}^n$ is said to be orthonormal if they are mutually orthogonal with unit $\mathcal{L}_2$-norm. Sometimes we will introduce a positive definite weighting $W > 0$, such that

$$\langle f, g\rangle_W = \langle fW, g\rangle$$

and

$$\|f\|_W = \sqrt{\text{Tr}\ \langle fW, f\rangle}.$$

### Subspaces

For an $\mathcal{L}_2$-function $\Psi \in \mathcal{L}_2^{n\times m}$, we denote by $\mathcal{S}_\Psi \subset \mathcal{L}_2^m$ the $r$-dimensional subspace spanned by the rows of $\Psi$, $r \leq n$. An orthonormal basis of $\mathcal{S}_\Psi$ consists of a set of $r$, orthonormal functions that span $\mathcal{S}_\Psi$, $r$ also correspond to the dimension of the subspace $\mathcal{S}_\Psi$. Given an $\mathcal{L}_2$-function $\Psi \in \mathcal{L}_2^{n\times m}$, it is straightforward to construct an orthonormal basis of $\mathcal{S}_\Psi$ as a linear combination of the rows of $\Psi$, e.g., by using the Gram-Schmidt method.

### Takenaka-Malmquist functions

In some cases it is possible to derive explicit expressions for the basis functions $\mathcal{B}_k$. A well known case (Ninness and Gustafsson, 1997) is when

$$\text{Span}\left\{\frac{\mathit{\Gamma}_n}{A(q)}\right\} = \text{Span}\left\{\frac{q^{-1}}{A(q)}, \frac{q^{-2}}{A(q)}, \ldots, \frac{q^{-n}}{A(q)}\right\} \tag{2.22}$$

where $A(q) = \prod_{k=1}^{n_a}(1-\xi_k q^{-1})$, $|\xi_k| < 1$ for some set of specified poles $\{\xi_1, \ldots, \xi_{n_a}\}$ and where $n \geq n_a$. Then, it holds (Ninness and Gustafsson, 1997) that

$$\text{Span}\left\{\frac{\mathit{\Gamma}_n}{A(q)}\right\} = \text{Span}\,\{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$$

where $\{\mathcal{B}_k\}$ are the Takenaka-Malmquist functions given by

$$\begin{aligned}\mathcal{B}_k(q) &:= \frac{\sqrt{1-|\xi_k|^2}}{q-\xi_k}\,\phi_{k-1}(q),\ k=1,\ldots,n\\ \phi_k(q) &:= \prod_{i=1}^{k}\frac{1-\overline{\xi_i}q}{q-\xi_i},\ \ \phi_0(q):=1\end{aligned} \tag{2.23}$$

and with $\xi_k = 0$ for $k = n_a + 1, \ldots, n$. In Ninness and Hjalmarsson (2004) it is shown that the structure (2.22) holds for common model structures such as Output-Error and Box-Jenkins provided the input spectrum has no zeros and sufficiently many numerator coefficients are estimated. Notice that the system zeros do not affect the basis functions above.

### Orthogonal projections

We denote the orthogonal projection of $f$ onto the space $\mathcal{S}_\Psi$ by $\mathbf{P}_{\mathcal{S}_\Psi}\{f\}$, i.e., $\mathbf{P}_{\mathcal{S}_\Psi}\{f\}$ is the unique solution to

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

Given an orthonormal basis $\{\mathcal{B}_k\}_{k=1}^r$ of $\mathcal{S}_\Psi$, the projection is readily calculated as

$$\mathbf{P}_{\mathcal{S}_\Psi}\{f\} = \sum_{k=1}^{r}\langle f,\mathcal{B}_k\rangle\,\mathcal{B}_k. \tag{2.24}$$

It can alternatively be expressed as

$$\mathbf{P}_{\mathcal{S}_\Psi}\{f\} = \langle f,\Psi\rangle\,\langle\Psi,\Psi\rangle\,\Psi. \tag{2.25}$$

## 2.3 Geometric tools for variance analysis

Many results in this thesis are based on the following results from Hjalmarsson and Mårtensson (2011), restated here for completeness.

**Lemma 2.3.1.** *(Lemma II.3 in Hjalmarsson and Mårtensson (2011)) Suppose that $J : \mathbb{R}^n \to \mathbb{C}^{1\times q}$ is differentiable and let the asymptotic covariance matrix $\mathrm{AsCov}\,J(\hat{\theta}_N)$ be defined by (2.15) where $\Psi \in \mathcal{L}_2^{n\times m}$. Suppose that $\gamma \in \mathcal{L}_2^{q\times m}$ is such that*

$$J'(\theta^o) = \langle\Psi,\gamma\rangle\,, \tag{2.26}$$

*then*

$$\begin{aligned}\mathrm{AsCov}\,J(\hat{\theta}_N) &= \langle\gamma,\Psi\rangle\,\langle\Psi,\Psi\rangle^{-1}\,\langle\Psi,\gamma\rangle\\ &= \langle\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\},\mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}\rangle\,.\end{aligned} \tag{2.27}$$

*Finally it holds that*

$$\text{AsCov } J(\hat{\theta}_N) = \sum_{k=1}^{r} \langle \gamma, \mathcal{B}_k \rangle \langle \mathcal{B}_k, \gamma \rangle , \tag{2.28}$$

*where $\mathcal{S}_\Psi$ is the subspace of $\mathcal{L}_2^m$ spanned by the rows of $\Psi$, and $\{\mathcal{B}_k\}_{k=1}^r$ is any orthonormal basis for this space.*

*Proof.* Taking the inner product of (2.25) with itself, (2.27) follows. Similarly, (2.28) follows from taking the inner product of (2.24) with itself. □

Since there are many functions $\gamma$ for which (2.26) holds, there is a large degree of freedom in the choice of $\gamma$. In Hjalmarsson and Mårtensson (2011, Lemma II.8) it is shown that all solutions $\gamma \in \mathcal{L}_2^{p\times m}$ to the equation $J'(\theta^o) = \langle \Psi, \gamma \rangle$ are given by

$$\gamma = J'(\theta^o)^* \langle \Psi, \Psi \rangle^\dagger \Psi + s^\perp, \tag{2.29}$$

where $s^\perp$ is any $\mathcal{L}_2^{p\times m}$-function orthogonal to $\mathcal{S}_\Psi$. We will explore this degree of freedom in the next lemma, where a re-parametrization of $J(\theta)$ is used to find an expression for a $\gamma$ that fulfills the condition (2.29).

**Lemma 2.3.2.** *(Lemma II.9 in Hjalmarsson and Mårtensson (2011)) Let $J$, $\Psi$, and $\mathcal{S}_\Psi$ be as in Lemma 2.3.1 and suppose that $J'(\theta^o) = \Psi(z_o)L$ for some $z_0 \in \mathbb{C}$ and $L \in \mathbb{C}^{m\times q}$. Let $\{\mathcal{B}_k\}_{k=1}^r, r \leq n$, be an orthonormal basis for $\mathcal{S}_\Psi$. Then*

$$\text{AsCov } J(\hat{\theta}_N) = L^* \sum_{k=1}^{r} \mathcal{B}_k^*(z_o)\mathcal{B}_k(z_o)\, L. \tag{2.30}$$

*Proof.* Let $\Gamma = [\mathcal{B}_1^{\mathrm{T}}, \dots, \mathcal{B}_r^{\mathrm{T}}]^{\mathrm{T}}$. Then there exists a full (column) rank $T \in \mathbb{C}^{n\times r}$ such that $\Psi = T\Gamma$. Inserting this in (2.27) and some straightforward algebra gives the result. □

The next lemma will be useful to derive upper bounds for (2.15).

**Lemma 2.3.3.** *(Lemma II.6 in Hjalmarsson and Mårtensson (2011)) 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 $\gamma \in \mathcal{L}_2^{q\times m}$. It holds that*

$$\langle \mathbf{P}_\mathcal{Y}\{\gamma\}, \mathbf{P}_\mathcal{Y}\{\gamma\} \rangle - \langle \mathbf{P}_\mathcal{X}\{\gamma\}, \mathbf{P}_\mathcal{X}\{\gamma\} \rangle = \langle \mathbf{P}_{\mathcal{Y}\ominus\mathcal{X}}\{\gamma\}, \mathbf{P}_{\mathcal{Y}\ominus\mathcal{X}}\{\gamma\} \rangle . \tag{2.31}$$

*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 (2.31) follows. □

Upper bounds for (2.27) can now be constructed by taking $\mathcal{X} = \Psi$ and $\mathcal{Y} = \mathcal{L}_2^m$ for example. It is immediate from (2.31) that

$$\langle \mathbf{P}_\mathcal{Y}\{\gamma\}, \mathbf{P}_\mathcal{Y}\{\gamma\} \rangle - \langle \mathbf{P}_\mathcal{X}\{\gamma\}, \mathbf{P}_\mathcal{X}\{\gamma\} \rangle \geq 0.$$

Lemma 2.3.3 will prove useful when deriving upper bounds for dynamics networks in Chapter 5.

# Chapter 3

# SISO models

In this chapter we consider SISO LTI systems. This chapter serves as a review of the variance analysis provided by the geometric approach for variance analysis for the case of SISO LTI systems. Indeed, most of the results in this chapter can be found in Mårtensson (2007). Expressions for the asymptotic (co)variance of system properties are derived. These expressions delineate the impacts of model structure, model order, true system dynamics, and experimental conditions. A connection to results on frequency function estimation is established. Also, simple model structure independent upper bounds are established. Explicit variance expressions and bounds are provided for common system properties such as impulse response coefficients and non-minimum phase zeros. As an illustration of the insights the expressions provide, they are used to derive conditions on the input spectrum which makes the asymptotic variance of non-minimum phase zero estimates independent of the model order and model structure.

More precisely, the contributions of this chapter are:

i) Section 3.1: *Re-parametrization formulae.* We provide formulae for re-expressing (2.15) when the quantity of interest is parameterized in other parameters than those used in the system identification. These expressions are useful when comparing different model parametrizations in terms of the asymptotic variance they yield for the estimate of a specific system property.

ii) Sections 3.1:*A general characterization of* (2.15) *for SISO LTI systems.* Here we provide general formulas, and bounds, for (2.15), valid for different experimental conditions and model parametrizations.

iii) Section 3.3: *Expressions for the asymptotic variance for some properties of SISO LTI systems.* We provide expressions and upper bounds for the asymptotic variance of estimated frequency functions, impulse response coefficients, $\mathcal{L}_2$-gains and NMP-zeros.

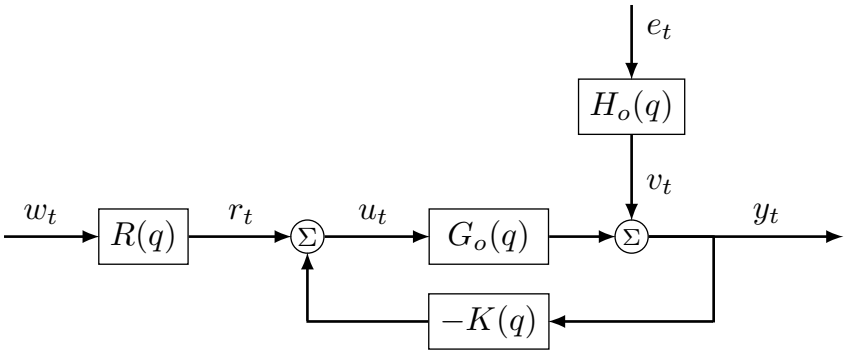


Figure 3.1: Block diagram of SISO LTI system with output feedback

iv) *Model structure independent upper bounds for* (2.15). At present there are surprisingly few rules of thumbs available regarding model quality in system identification; the expression (1.4) for the variance of the frequency function estimate and some similar variance expressions for pole/zero estimates, are sole exceptions. Thus determining suitable experiment length and excitation in order to achieve a certain accuracy of, for example, an impulse response coefficient or an estimate of the $\mathcal{L}_2$ gain of the system, requires extensive calculations based on (2.15). A spin-off of our new expression for (2.15) is that it is easy to provide simple model structure independent upper bounds for (2.15). We hope this to be of value to practitioners.

## 3.1 Problem formulation

In this section we present the system and model assumptions, and provide the reparametrization formulae that can be used to characterize the variance of specific system properties.

### System and model assumptions

Throughout the chapter we will assume that the true system is given by a causal finite dimensional SISO LTI system $G_o$(q) as depicted in Figure 3.1, where $u_t$ and $y_t$ represent the measured input and output, respectively, and where $e_t$ and $w_t$ are zero mean white noise sequences with variances $\lambda_o$ and 1, respectively. The causal finite dimensional LTI filter $R$ represents a stable minimum phase spectral factor of the reference signal $r_t$, and $H_o$ is an inversely stable finite dimensional LTI filter that is normalized to be monic, i.e., $\lim_{z\to\infty} H_o(z) = 1$. The system $G_o$ includes at least one unit time delay, so that the feedback loop is well defined, and we also assume the entire system to be internally stabilized by the causal finite dimensional

LTI controller $K$. Furthermore, we will assume that neither $G_o$ nor $K$ have poles on the unit circle. The system is said to be operating in open loop when $K = 0$. Next, we introduce a quite general family of model structures that will be covered.

The system is modeled by

$$y_t = T(\mathrm{q}, \theta)\chi_t, \qquad \chi_t = [u_t,\ e_t]^{\mathrm{T}}, \tag{3.1}$$

where $T(\mathrm{q}, \theta) = [G(\mathrm{q}, \theta), H(\mathrm{q}, \theta)]$ is a causal finite dimensional LTI model of the system and the noise dynamics, parameterized by the vector $\theta \in \mathbb{R}^n$. The noise model may also be independently parameterized by a separate vector $\eta$, in which case we write $H(\mathrm{q}, \eta)$. This distinction is only used when it has important implications, and for the general treatment we can consider the noise model $H(\mathrm{q}, \theta)$.

The model parametrization is such that the *true* system is in the model set, that is, there is a, not necessarily unique, parameter $\theta^o$ such that

$$G_o(\mathrm{q}) = G(\mathrm{q}, \theta^o), \quad H_o(\mathrm{q}) = H(\mathrm{q}, \theta^o).$$

The model $T(z, \theta)$ is continuously differentiable with respect to $\theta$ in a neighborhood of $\theta^o$. The type of model described above includes all standard black-box model structures such as ARMAX, output error and Box-Jenkins.

Now, introduce the spectral factor of the signal-to-noise ratio

$$R_{\text{SNR}}(z) = R_\chi(z)R_v^{-1}(z),$$

where $R_v = \sqrt{\lambda_o}H_o$ is a minimum phase spectral factor of the noise spectrum $\Phi_v$ and where $R_\chi$ is a stable spectral factor of the spectrum $\Phi_\chi$ of $\chi$, i.e.

$$R_\chi \coloneqq \begin{bmatrix} S_oR & -KS_oH_o \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\lambda_o} \end{bmatrix}, \tag{3.2}$$

where $S_o(\mathrm{q}) = 1/(1 + K(\mathrm{q})G_o(\mathrm{q}))$ is the closed loop sensitivity function. It is straightforward to show that the predictor gradient, normalized by $\sqrt{\lambda_o}$, is given by

$$\Psi(z) = T'(z, \theta^o)R_{\text{SNR}}(z), \tag{3.3}$$

where $T'(z, \theta) = \left[\frac{\partial G(z,\theta)}{\partial \theta} \quad \frac{\partial H(z,\theta)}{\partial \theta}\right]$.

We will assume that the model parametrization is such that $\Psi$ is stable. The stability assumption on the closed loop system and the assumptions on $G_o$ and $K$ imply that $R_{\text{SNR}}(z)$ and its inverse are real rational functions without poles on the unit circle and hence are $\mathcal{L}_2^{2\times 2}$ functions, as well as bounded on the unit circle.

Our main assumption is that prediction error identification results in an asymptotic covariance AsCov $J(\hat{\theta}_N)$ of the quantity of interest $J$ given by (2.15). We refer to Ljung (1999) for exact conditions and to Hjalmarsson and Mårtensson (2011, Section I) for a discussion of the case when $\theta^o$ is non-unique and $\langle \Psi, \Psi \rangle$ singular.

## 3.2 Asymptotic covariance of LTI system properties

In this section we will derive an expression for the asymptotic covariance (2.15) of the estimate $J(\hat{\theta}_N)$ of an arbitrary differentiable quantity $J : \mathbb{R}^n \to \mathbb{C}^{1\times p}$ when $\Psi$ in (2.15) is given by (3.3). While this can be done on a case by case basis for different model structures using Lemma 2.3.1, we will instead use (a generalization of) impulse response coefficients as an intermediate parametrization in order to obtain an expression that is valid regardless of the model structure.

Take $\{\mathcal{G}_k(z)\}_{k=1}^{\infty}$ and $\{\mathcal{H}_k(z)\}_{k=1}^{\infty}$ to be two sequences of orthonormal $\mathcal{L}_2$-functions and for $k = 1, 2, \ldots$ define the orthonormal functions

$$\mathcal{T}_{2k-1}(z) = [\mathcal{G}_k(z) \;\; 0], \;\; \mathcal{T}_{2k}(z) = [0 \;\; \mathcal{H}_k(z)]. \tag{3.4}$$

With $\tau = [\tau_1 \;\; \tau_2 \;\; \cdots]$, any transfer function $T = [G \;\; H]$ satisfying the assumptions in Section 3.1 can be represented by

$$T(z) = [G(z) \;\; H(z)] = \sum_{k=1}^{\infty} \tau_k \, \mathcal{T}_k(z) \tag{3.5}$$

on the unit circle for suitable choices of $\{\mathcal{G}_k(z)\}_{k=1}^{\infty}$ and $\{\mathcal{H}_k(z)\}_{k=1}^{\infty}$. We will assume that the sum on the right hand side of (3.5) has a region of convergence that includes the unit circle. For an asymptotically stable system $G(z)$ we can for example use the impulse response representation $G(z) = \sum_{k=1}^{\infty} g_k z^{-k}$ (cf. the single-sided z-transform). If $\sum_{k=1}^{\infty} |g_k| < \infty$, the sum converges uniformly to $G(z)$ on the unit circle. A more general representation is via the Takenaka-Malmquist functions, see (2.23) in Section 2.2, for which the convergence is uniform on the unit circle if $\sum_{k=1}^{\infty}(1 - |\xi_k|) = \infty$ for the set of pre-specified poles $\{\xi_k\}$. It is worth noticing that also unstable $G(z)$ can be represented by (3.5) on the unit circle, for example by a Laurent series expansion (cf. the double-sided z-transform). Since $G(z)$ is analytic on an annulus around the unit circle we can always find a series $\sum_{k=-\infty}^{\infty} a_k z^{-k}$ that converges uniformly to $G(z)$ on the unit circle. Note that also the Takenaka-Malmquist basis functions, which form a basis for $\mathcal{H}_2$, can be extended to a basis for $\mathcal{L}_2$ by appending a function $\mathcal{A}_k(z) \coloneqq \frac{1}{z}\mathcal{B}_k(1/z)$ for each basis function $\mathcal{B}_k(z)$. For general orthonormal $\mathcal{L}_2$-sequences $\{\mathcal{G}_k(z)\}_{k=1}^{\infty}$ and $\{\mathcal{H}_k(z)\}_{k=1}^{\infty}$, it is only required that the sum in (3.5) converges in $\mathcal{L}_2$ mean, since the expression is always used within the inner product brackets $\langle \cdot, \cdot \rangle$. This type of convergence will be assumed whenever the representation (3.5) is used.

The original model (3.1), which is parameterized by the vector $\theta$, can also be expressed through the parametrization (3.5):

$$T(z, \theta) = \sum_{k=1}^{\infty} \tau_k(\theta) \mathcal{T}_k(z) \tag{3.6}$$

or

$$G(z, \theta) = \sum_{k=1}^{\infty} g_k(\theta) \mathcal{G}_k(z), \quad H(z, \theta) = \sum_{k=1}^{\infty} h_k(\theta) \mathcal{H}_k(z),$$

where $g_k = \tau_{2k-1},\ h_k = \tau_{2k}$. We will denote by $\tau^o$ the model parameters corresponding to $\theta^o$, i.e., $\tau^o = \tau(\theta^o)$.

We will first establish some properties of the maps $\tau_k : \mathbb{R}^n \to \mathbb{C}$, $k = 1, \ldots$.

**Lemma 3.2.1.** *Under the assumptions in Section 3.1, $\tau_k(\theta)$, $k = 1, \ldots$ are differentiable at $\theta^o$ and*

$$T'(z, \theta^o) = \sum_{k=1}^{\infty} \tau_k'(\theta^o)\mathcal{T}_k(z) \in \mathcal{L}_2^{n\times 2}. \tag{3.7}$$

*Proof.* By assumption, the elements of $T(z, \theta)$ are finite dimensional real rational functions with no poles on the unit circle, i.e., they can be written as $B_i(z, \theta)/A_i(z, \theta)$, $i = 1, 2$ for some polynomials $B_i$ and $A_i$ with real coefficients where $A_i(z, \theta^o)$, $i = 1, 2$, does not have any roots on the unit circle. Thus $T(z, \theta^o)$ belongs to $\mathcal{L}_2^2$ and hence $\tau_k$ can be expressed through the inverse transformation

$$\tau_k(\theta^o) = \langle T(z, \theta^o), \mathcal{T}_k(z)\rangle. \tag{3.8}$$

By assumption $T(z, \theta)$ is continuously differentiable with respect to $\theta$ in a neighborhood of $\theta^o$ and hence the right hand side of (3.8) is differentiable at $\theta^o$ under the integral sign (Rudin, 1976, Theorem 9.42). Thus

$$\tau_k'(\theta^o) = \langle T'(z, \theta^o), \mathcal{T}_k(z)\rangle,\ k = 1, \ldots. \tag{3.9}$$

Now the elements of $T'(z, \theta^o)$ are given by

$$\frac{B_i'(z, \theta^o)}{A_i(z, \theta^o)} - \frac{B_i(z, \theta^o)A_i'(z, \theta^o)}{A_i^2(z, \theta^o)},\ i = 1, 2$$

and $T'(z, \theta^o) \in \mathcal{L}_2^{n\times 2}$ since by assumption $A_i(z, \theta^o)$, $i = 1, 2$ does not have any roots on the unit circle. Therefore, (3.7) follows from the inverse transformation of (3.9). We remark that $\{\mathcal{T}_k\}$ does not have to be complete. It is only required that $T(z)$ is representable in this orthonormal system.

□

**Theorem 3.2.1.** *Suppose that $J_\tau(\tau^o) \in \mathbb{C}^p$ is estimated by $J(\hat{\theta}_N) = J_\tau(\tau(\hat{\theta}_N))$. Assume that*

*i) The system and model assumptions in Section 3.1 hold.*

*ii) The partial derivatives of $J_\tau$ with respect to $\tau_k$, $k = 1, \ldots, n_\tau$ exist at $\theta^o$ and satisfy*

$$\nabla J_\tau(z) \coloneqq \sum_{k=1}^{n_\tau} \left(\frac{\partial J_\tau(\tau^o)}{\partial \tau_k}\right)^* \mathcal{T}_k(z) \in \mathcal{L}_2^{p\times m}. \tag{3.10}$$

*iii) The following chain rule applies:*

$$J'(\theta^o) = \sum_{k=1}^{n_\tau} \tau_k'(\theta^o) \frac{\partial J_\tau(\tau(\theta^o))}{\partial \tau_k} \ . \tag{3.11}$$

*Then*

$$\text{AsCov } J(\hat{\theta}_N) = \langle \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\}, \mathbf{P}_{\mathcal{S}_\Psi}\{\gamma\} \rangle \tag{3.12}$$

*holds with*

$$\gamma = \nabla J_\tau R_{\text{SNR}}^{-*} \tag{3.13}$$

*When $J$ is scalar,* (3.12) *becomes*

$$\text{AsVar } J(\hat{\theta}_N) = \left\| \mathbf{P}_{\mathcal{S}_\Psi}\left\{ \nabla J_\tau \ R_{\text{SNR}}^{-*} \right\} \right\|^2 . \tag{3.14}$$

*Proof.* All that has to be proven is that (2.26) holds with $\gamma$ as in (3.13), i.e., $\langle \Psi(z), \gamma(z) \rangle = J'(\theta^o)$. First notice that, from 3.3,

$$\langle \Psi, \nabla J_\tau R^{-*} \rangle = \langle \Psi R^{-1}, \nabla J_\tau \rangle = \langle T'(z, \theta^o), \nabla J_\tau \rangle.$$

Due to Lemma 3.2.1, (3.10) and the orthonormality of $\{\mathcal{T}_k\}$ it follows that

$$\langle \Psi, \nabla J_\tau R^{-*} \rangle = \sum_{k=1}^{n_\tau} \tau_k'(\theta^o) \ \frac{\partial J_\tau(\tau(\theta^o))}{\partial \tau_k},$$

which, according to assumption (3.11), equals $J'(\theta^o)$. □

Some simplifications of (3.14) are given in the following corollary. For those results we introduce the notation $\nabla J_\tau(z) =: \left[ \nabla J_\tau^g(z) \quad \nabla J_\tau^h(z) \right]$.

**Corollary 3.2.1** (Simplifications of (3.14))**.** *When the property $J$ does not depend on the noise model $H$, the asymptotic variance of $J(\hat{\theta}_N)$ is given by*

$$\text{AsVar } J(\hat{\theta}_N) = \left\| \mathbf{P}_{\mathcal{S}_\Psi} \left\{ \nabla J_\tau^g \left[ \frac{\sqrt{\lambda_0} H_o^*}{S_o^* R^*} \quad 0 \right] \right\} \right\|^2 , \tag{3.15}$$

*where, as before, $\mathcal{S}_\Psi$ denotes the row span of $\Psi$. When $K = 0$ the asymptotic variance of $J(\hat{\theta}_N)$ is given by*

$$\text{AsVar } J(\hat{\theta}_N) = \left\| \mathbf{P}_{\mathcal{S}_\Psi} \left\{ \left[ \nabla J_\tau^g \frac{\sqrt{\lambda_0} H_o^*}{R^*} \quad \nabla J_\tau^h H_o^* \right] \right\} \right\|^2 , \tag{3.16}$$

*where in this case $\mathcal{S}_\Psi$ is the row span of*

$$\left[ \frac{G'(z, \theta^o) R(z)}{\sqrt{\lambda_o} H_o(z)}, \ \frac{H'(z, \theta^o)}{H_o(z)} \right].$$

*When $K = 0$, and the model $G(z, \theta)$ with the noise model $H(z, \eta)$ are independently parameterized, the asymptotic variance of $J(\hat{\theta}_N, \hat{\eta}_N)$ is given by*

$$\text{AsVar } J(\hat{\theta}_N, \hat{\eta}_N) = \left\| \mathbf{P}_{\mathcal{S}_{\Psi_G}} \left\{ \nabla J_\tau^g \frac{\sqrt{\lambda_0} H_o^*}{R^*} \right\} \right\|^2 + \left\| \mathbf{P}_{\mathcal{S}_{\Psi_H}} \left\{ \nabla J_\tau^h H_o^* \right\} \right\|^2 , \tag{3.17}$$

*where $\mathcal{S}_{\Psi_G}$ is the row span of $\Psi_G = R(z) G'(z, \theta^o)/H_o(z)$ and $\mathcal{S}_{\Psi_H}$ is the row span of $\Psi_H = H'(z, \eta_o)/H_o(z)$.*

### Upper bounds

One advantage with the new expression (3.12) for the asymptotic covariance (2.15) is that it is easy to provide simple, model structure independent, bounds for (2.15). These bounds are obtained by replacing the projection onto $\mathcal{S}_\Psi$ with projections onto the spaces $\mathcal{H}_2^2$ or $\mathcal{L}_2^2$. Below, when we are projecting onto $\mathcal{L}_2^2$, i.e., when the projection is removed, the bounds derived are typically conservative even as the model order increases since $\mathcal{S}_\Psi \subseteq \mathcal{H}_2^m$ regardless of the model order and model structure, while the function that is projected, $\nabla J_\tau \, R_{\text{SNR}}^{-*}$, typically has a term that belongs to the complement of $\mathcal{H}_2^m$.

**Theorem 3.2.2.** *Let the conditions of Theorem 3.2.1 be fulfilled. An upper bound of the asymptotic covariance of $J(\hat{\theta}_N)$ is then given by*

$$\text{AsCov } J(\hat{\theta}_N) \leq \langle \nabla J_\tau \Phi_v \Phi_\chi^{-1}, \nabla J_\tau \rangle. \tag{3.18}$$

*When $J$ is scalar we get*

$$\text{AsVar } J(\hat{\theta}_N) \leq \|\nabla J_\tau\|^2_{\Phi_v \Phi_\chi^{-1}}. \tag{3.19}$$

*Proof.* By removing the projection in (3.12) of Theorem 3.2.1 we get an upper bound, cf. Lemma II.6 in Hjalmarsson and Mårtensson (2011).

□

Some simplifications of (3.19) are given next.

**Corollary 3.2.2** (Simplifications of (3.19))**.** *When the property $J$ does not depend on the noise model $H$, the upper bound* (3.19) *is given by*

$$\text{AsVar } J(\hat{\theta}_N) \leq \|\nabla J_\tau^g\|^2_{\Phi_v / \Phi_u^r}, \tag{3.20}$$

*where $\Phi_u^r$ is the part of the input spectrum that is due to $r_t$.*

*When the system is identified in open loop, the upper bound* (3.19) *is given by*

$$\text{AsVar } J(\hat{\theta}_N) \leq \|\nabla J_\tau^g\|^2_{\Phi_v / \Phi_u} + \left\|\nabla J_\tau^h\right\|^2_{\Phi_v / \lambda_0}. \tag{3.21}$$

We remark that the bounds in Theorem 3.2.2 typically (but not always) depend on the true underlying system through $\nabla J_\tau$ and $R_{\text{SNR}}^{-1}(z)$.

Notice that the upper bounds above are valid for *any* model structure, which also means that they apply to any model order.

It is obvious that the inverse of the signal to noise ratio, i.e., $\Phi_v \Phi_\chi^{-1}$, plays an important role for the variance. For clarity we write out the expression (3.19) in integral form

$$\text{AsVar } J(\hat{\theta}_N) \leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \nabla J_\tau(\mathrm{e}^{\mathrm{j}\omega}) \frac{\Phi_v(\mathrm{e}^{\mathrm{j}\omega})}{\Phi_\chi(\mathrm{e}^{\mathrm{j}\omega})} \nabla J_\tau^*(\mathrm{e}^{\mathrm{j}\omega}) \, \mathrm{d}\omega.$$

The function $\nabla J_\tau$ is weighted by the inverse of the signal to noise ratio. Thus, $\Phi_v \Phi_\chi^{-1}$ must be small in the directions where $\nabla J_\tau$ is large in order to produce an accurate estimate of $J$.

### Curse of Complexity

In this section we will compare the results based on Theorem 3.2.1 with the asymptotic covariance expressions implied by the results in Ljung (1985) that are asymptotic in model order.

Using the orthonormality of $\{\mathcal{T}_k\}$ in the representation (3.5) gives that $\tau_l = \langle T, \mathcal{T}_l \rangle$ and hence (with $\tilde{\tau}_k \coloneqq \tau_k(\hat{\theta}_N) - \tau_k^o$, $\widetilde{T}(z) \coloneqq T(z, \hat{\theta}_N) - T_o(z)$ and with $m$ being the model order)

$$\begin{aligned}\lim_{m\to\infty} \frac{1}{m} \mathrm{E}\,[\tilde{\tau}_k^* \tilde{\tau}_l] &= \lim_{m\to\infty} \frac{1}{m} \mathrm{E}\left[\left\langle \mathcal{T}_k(z), \widetilde{T}(z)\right\rangle \left\langle \widetilde{T}(\zeta), \mathcal{T}_l(\zeta)\right\rangle\right] \\ &= \lim_{m\to\infty} \frac{1}{m} \mathrm{E}\left[\left\langle \left\langle \mathcal{T}_k(z), \widetilde{T}^*(\zeta)\widetilde{T}(z)\right\rangle, \mathcal{T}_l(\zeta)\right\rangle\right] \\ &= \left\langle \left\langle \mathcal{T}_k(z), \lim_{m\to\infty} \frac{1}{m} \mathrm{E}\left[\widetilde{T}^*(\zeta)\widetilde{T}(z)\right]\right\rangle, \mathcal{T}_l(\zeta)\right\rangle, \end{aligned} \tag{3.22}$$

assuming that the limit operation and the integration commute. If we now use the asymptotic result

$$\lim_{m\to\infty} \frac{1}{m} \mathrm{AsCov}\, T(\mathrm{e}^{\mathrm{j}\omega}, \hat{\theta}_N) = \Phi_v(\mathrm{e}^{\mathrm{j}\omega}) \Phi_\chi^{-1}(\mathrm{e}^{\mathrm{j}\omega}) \tag{3.23}$$

derived in Ljung (1985) and another result from Ljung (1985), namely that frequency function estimates at different frequencies become uncorrelated as the model order $m \to \infty$, (3.22) collapses to zero, which in turn suggests that for any $J$ of the type described in Theorem 3.2.1

$$\lim_{m\to\infty} \frac{1}{m} \mathrm{AsCov}\, J(\hat{\theta}_N) = 0. \tag{3.24}$$

We have thus obtained, by direct application of the model order asymptotic results in Ljung (1985) that the asymptotic variance is of the order $o(m)$. Note that $\mathrm{AsCov}\, T(\mathrm{e}^{\mathrm{j}\omega}, \hat{\theta}_N)$ grows unbounded with $m$ as in (3.23), which seems to contradict (3.24), but $T(\mathrm{e}^{\mathrm{j}\omega})$ does not belong to the allowed class of functions $J$ in Theorem 3.2.1. This situation will be addressed later in Section 3.3.

The result (3.24) is considerably weaker than the upper bound (3.18) derived in this chapter. Thus we have shown that, when the conditions of Theorem 3.2.1 hold, the upper bounds derived in this chapter are significantly more accurate expressions for the asymptotic covariance than the asymptotic covariance expressions implied by the results in Ljung (1985).

In order to go from the results in Ljung (1985) and arrive at something similar to (3.18) in Theorem 3.2.2 we must remove the scaling factor $1/m$ and interpret (3.23) as

$$\lim_{m\to\infty} \mathrm{E}\left[\widetilde{T}^*(\xi)\widetilde{T}(z)\right] = \Phi_v(z)\Phi_\chi^{-1}(z)\delta(z-\xi), \tag{3.25}$$

with $\delta(x)$ being the Dirac delta. Note that Ljung (1985) does *not* give support for this interpretation. Removing the factor $1/m$ in (3.22) and then using (3.25) yields

$$\begin{aligned}\lim_{m\to\infty} \mathrm{E}\,[\tilde{\tau}_k^* \tilde{\tau}_l] &= \left\langle \left\langle \mathcal{T}_k(z), \Phi_v(z)\Phi_\chi^{-1}(z)\delta(z-\xi) \right\rangle, \mathcal{T}_l(\zeta) \right\rangle \\ &= \left\langle \mathcal{T}_k, \mathcal{T}_l \Phi_v \Phi_\chi^{-1} \right\rangle\end{aligned}$$

and with $J$, $J_\tau$ and $\nabla J_\tau$ as in Theorem 3.2.1 we get

$$\lim_{m\to\infty} \mathrm{AsCov}\, J(\hat{\theta}_N) = \left\langle \nabla J_\tau \Phi_v \Phi_\xi^{-1}, \nabla J_\tau \right\rangle,$$

which is the same as the upper bound (3.18).

The fact that the scaling factor $1/m$ is not present is especially important as it shows that certain properties, even of highly complex systems, are not subject to what is known as the "curse of complexity", i.e., there are system properties that can be accurately identified using full order models also when the system is highly complex. In Section 3.3 we will see some examples of such properties. For more details on this important topic we refer the reader to Hjalmarsson (2005); Mårtensson and Hjalmarsson (2009); Rojas *et al.* (2008, 2010).

### Interpretation of Theorem 3.2.1

The result in Theorem 3.2.1 is basically applicable whenever the predictor gradient is given by (3.3) and thus very general. The expression (3.12) is an exact representation of the asymptotic variance (2.15) which is valid for a wide range of LTI model structures, including commonly used structures such as ARMAX, output-error and Box-Jenkins, and it can be used for both open loop and closed loop identification. Furthermore it expresses the variance of any property of the estimated model, provided this property can be expressed as a differentiable function of the (impulse response) coefficients $\tau_k$ satisfying the conditions in the theorem.

Recall the covariance expression (3.12)

$$\mathrm{AsCov}\, J(\hat{\theta}_N) = \left\langle \mathbf{P}_{\mathcal{S}_\Psi}\left\{\nabla J_\tau R_{\mathrm{SNR}}^{-*}\right\}, \mathbf{P}_{\mathcal{S}_\Psi}\left\{\nabla J_\tau R_{\mathrm{SNR}}^{-*}\right\} \right\rangle$$

and the expression (3.3) for the prediction error gradient

$$\Psi(z) = T'(z,\theta^o) R_{\mathrm{SNR}}(z).$$

There is a certain decoupling between the property of interest $J$, the experimental conditions $\Phi_v(\mathrm{e}^{\mathrm{j}\omega})\Phi_\chi^{-\mathrm{T}}(\mathrm{e}^{\mathrm{j}\omega})$ and model structure $T(z,\theta)$ in the expression (3.12).

The property of interest enters the expression only through the function $\nabla J_\tau$ which describes the sensitivity of the property $J$ to changes in the transfer function $T$. One could interpret $\nabla J_\tau$ as a Fréchet derivative (Luenberger, 1969) by viewing $J$ as a functional of the system transfer function, i.e., $J = f(T(\cdot))$, or, if we include the $\theta$-dependence $J(\theta) = f(T(\cdot,\theta))$. For small changes $\delta\theta$ we have $\delta J(\theta) = [\delta\theta]^{\mathrm{T}} J'(\theta)$

and $\delta T(z) = [\delta\theta]^{\mathrm{T}} T'(z,\theta)$. Similarly we have, by the Riesz representation theorem (Kreyszig, 1978), that $\delta f = \langle \delta T, y\rangle$ for some $y \in \mathcal{L}_2^{p\times 2}$ where $y$ should be interpreted as the "functional derivative" $y = \dfrac{\delta f}{\delta T}$. Now consider

$$\delta J = \delta f \circ \delta T = [\delta\theta]^{\mathrm{T}}\langle T', y\rangle = [\delta\theta]^{\mathrm{T}}\langle \Psi, y R_{\mathrm{SNR}}^{-*}\rangle$$

and recall that by definition

$$\delta J = [\delta\theta]^{\mathrm{T}} J'(\theta) = [\delta\theta]^{\mathrm{T}}\langle \Psi, \gamma\rangle = [\delta\theta]^{\mathrm{T}}\langle \Psi, \nabla J_\tau R_{\mathrm{SNR}}^{-*}\rangle.$$

By comparing these two expressions it can be seen that $\nabla J_\tau = y = \dfrac{\delta f}{\delta T}$. Thus, when $J$ is viewed as a functional on $T$, $\nabla J_\tau$ has the interpretation of a Fréchet derivative of $J$ with respect to $T$ .

$\nabla J_\tau$ is weighted by $R_{\mathrm{SNR}}^{-*}(z^{-*})$ which is a spectral factor of the ratio $\Phi_v(z)\Phi_\chi^{-1}(z)$. This ratio is known from the expression (3.23) and can be interpreted as the frequency-wise noise to signal ratio.

The space $\mathcal{S}_\Psi$ is the span of the rows of

$$\Psi(z) = T'(z,\theta^o)\; R_{\mathrm{SNR}}^{-1}(z) = T'(z,\theta^o) R_\chi(z) R_v^{-1}(z).$$

The structure of this space is thus to a large extent determined by the model structure (through $T'$). However, the true system also determines $T'(z,\theta^o)$ (through $\theta^o$) and together with the experimental conditions also acts as translation through the factor $R_\chi(z)R_v^{-1}(z)$. Furthermore, the projection only depends on the span of $\Psi$, i.e., the subspace $\mathcal{S}_\Psi$. From these two observations it follows that all model structures whose predictor gradients span the same space will have exactly the same asymptotic covariance. For example, order $n$ Laguerre models (Wahlberg, 1991) with poles in $\xi$ will have the same asymptotic variance as fixed denominator models of order $n$ with a pole of multiple $n$ at $\xi$. It also follows that scaling the model structure, i.e., replacing $T(z,\theta)$ with $\tilde{T}(z,\theta) = \alpha T(z,\theta)$ will not change the asymptotic variance, again since both the function to be projected and the subspace do not change. On the other hand if the experimental conditions are changed so that the signal to noise ratio $R_{\mathrm{SNR}}$ is scaled by a factor $\beta$, then since $\mathcal{S}_\Psi$ will remain the same (even though $\Psi$ is re-scaled), the asymptotic variance is scaled by $1/\beta^2$.

We also remark that Theorem 3.2.1 illustrates the flexibility offered by (2.26). The function $\nabla J_\tau(z)\; R_{\mathrm{SNR}}^{-*}(z^{-*})$ is a function in the set (2.29) of functions $\gamma$ that can be used in Lemma 2.3.1 such that $\langle \Psi, \gamma\rangle$ is the sensitivity of the quantity of interest with respect to the model parameters. However, this function is chosen with care so that it can be used regardless of the model structure (which determines $T'(z,\theta^o)$). It is due to this that the decoupling between the function of interest and the model structure, discussed above, is obtained. This also opens up the possibility to derive the upper bounds found in Theorem 3.2.2 and Corollary 3.2.2 for the asymptotic variance that are model structure independent. This is one of the features offered by the geometric approach employed in this thesis. For further discussion on the geometric approach we refer to Hjalmarsson and Mårtensson (2011).

### Explicit variance expressions

Recall from Lemma 2.3.2 that if $J'(\theta^o) = \Psi(z_o)L$ for some $z_o \in \mathbb{C}$ and $L \in \mathbb{C}^{m\times p}$ we get that (2.15) can be written as

$$\text{AsCov } J(\hat{\theta}_N) = L^* \sum_{k=1}^{r} \mathcal{B}_k^*(z_o)\mathcal{B}_k(z_o)L. \tag{3.26}$$

The next result is an adaptation of (3.26) to the re-parametrization in Lemma 3.2.1.

**Lemma 3.2.2.** *Let the assumptions in Theorem 3.2.1 hold and let $\{\mathcal{B}_k\}_{k=1}^r$, $r \le n$, be an orthonormal basis for $\mathcal{S}_\Psi$. Assume also that*

$$\frac{\partial J_\tau(\tau(\theta^o))}{\partial \tau_k} = \mathcal{T}_k(z_o)\alpha \tag{3.27}$$

*for some $\alpha \in \mathbb{C}^{m\times p}$ and $z_o \in \mathbb{C}$, and let $\gamma$ be defined by* (3.13).

*Then* (3.26) *holds with*

$$L = R_{\text{SNR}}^{-1}(z_o)\alpha. \tag{3.28}$$

*Proof.* We need to prove that

$$J'(\theta^o) = \langle \Psi, \gamma \rangle = \Psi(z_o)R_{\text{SNR}}^{-1}(z_o)\alpha. \tag{3.29}$$

With $\gamma$ as in (3.13) with (3.27) it holds that

$$\begin{aligned}\langle \Psi, \gamma \rangle &= \sum_{l=1}^{n_\tau} \langle \Psi, \mathcal{T}_l R_{\text{SNR}}^{-*} \rangle \ \mathcal{T}_l(z_o)\ \alpha \\ &= \sum_{l=1}^{n_\tau} \langle \Psi R_{\text{SNR}}^{-1}, \mathcal{T}_l \rangle \ \mathcal{T}_l(z_o)\ \alpha = \mathbf{P}_{\mathcal{Y}}\{\Psi R_{\text{SNR}}^{-1}\}(z_o)\ \alpha,\end{aligned}$$

where $\mathcal{Y}$ is the space spanned by $\{\mathcal{T}_l\}_{l=1}^{n_\tau}$. However, due to (3.3) and (3.7), $\Psi R_{\text{SNR}}^{-1} \in \mathcal{Y}$, so the projection can be removed giving (3.29). □

Notice that under the conditions in Lemma 3.2.2, the condition that $\nabla J_\tau R_{\text{SNR}}^{-*}(z^{-*}) \in \mathcal{L}_2^{p\times m}$ in (3.10) can be written as

$$\alpha^* \sum_{k=1}^{n_\tau} \mathcal{T}_k^*(z_o)\mathcal{T}_k(z)\ R_{\text{SNR}}^{-*}(z^{-*}) \in \mathcal{L}_2^{p\times m}.$$

This may restrict the set of points $z_o \in \mathbb{C}$ for which Lemma 3.2.2 is applicable when $n_\tau = +\infty$.

### Upper bounds

Here we will describe a case when a simple bound for (3.26) can be found by a projection onto the subspace $\mathcal{H}_2^2 \subset \mathcal{L}_2^2$. This gives a a tighter bound than Theorem 3.2.2 where the projection was made onto $\mathcal{L}_2^2$. Let the asymptotic variance be given by (3.26) for a $z_o$ such that $|z_o| > 1$. An upper bound of (3.26) is obtained by exchanging the basis functions $\{\mathcal{B}_k\}$ for basis functions $\{\widetilde{\mathcal{B}}_k\}$ for the larger subspace $\mathcal{H}_2^2$. This is an upper bound since we project onto a larger space in (3.12) and the inner product is always positive definite. Since all elements of both $G'(z,\theta^o)$ and $H'(z,\theta^o)$ have at least one time delay, all elements of $\Psi$ will also have at least one time delay, and therefore we will exclude constant functions. One such orthonormal basis is given by $\{\widetilde{\mathcal{B}}_k(z)\}_{k=1}^{\infty}$ where $\widetilde{\mathcal{B}}_k(z) = [z^{-(k+1)/2}\ 0]$ when $k$ is odd, and $\widetilde{\mathcal{B}}_k(z) = [0\ z^{-k/2}]$ when $k$ is even. For $|z_o| > 1$ we then get

$$\sum_{k=1}^{\infty} \overline{\widetilde{\mathcal{B}}_k}^*(z_o)\widetilde{\mathcal{B}}_k(z_o) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \sum_{k=1}^{\infty} |z_o|^{-2k} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{|z_o|^2 - 1},$$

which when inserted in (3.26) gives the upper bound

$$\text{AsCov } J(\hat{\theta}_N) \leq \frac{1}{|z_o|^2 - 1}\ \alpha^*\ \Phi_v(z_o)\Phi_\chi^{-1}(z_o)\ \alpha. \tag{3.30}$$

In some cases, we can find a similar bound that holds with equality, as we will see in the following section.

### Model structure independent example

Now, we turn to a less obvious insight that does not seem to be generally known. Suppose that the explicit expression (3.26) holds for some $z_o$ strictly outside the unit circle and that $\Psi$ contains a pole at $z_o^{-1}$. Suppose further that the orthonormal basis used in (3.26) is of the form (2.23). If we then order the poles in $\Psi$ such that $\xi_1 = z_o^{-1}$, we obtain from (2.23) that $\mathcal{B}_1(z_o) = \sqrt{1 - |z_o|^{-2}}/(z_o - z_o^{-1})$ and $\mathcal{B}_k(z_o) = 0$, $k = 2, \ldots, n$, resulting in that

$$\text{AsCov } J(\hat{\theta}_N) = \frac{1 - |z_o|^{-2}}{|z_o - z_o^{-1}|^2}\ \alpha^*\ \Phi_v(z_o)\Phi_\chi^{-1}(z_o)\ \alpha. \tag{3.31}$$

This expression is remarkable in that it is independent of the model structure and model order. Now recall that $\Psi(z) = T'(z,\theta^o)R_{\text{SNR}}(z_o)$. Thus in the cases when (3.26) holds and when the experimental conditions can be chosen such that $R_{\text{SNR}}(z)$ has a pole at $z_o^{-1}$, this choice makes the asymptotic covariance the same for different model structures and arbitrary model order. This insight is important in order to come to terms with the so called "curse of dimensionality" discussed in Section 3.2.2. We will illustrate this idea in Section 3.3 where the objective is to identify NMP-zeros. The geometric approach has been used in Hjalmarsson *et al.* (2006); Mårtensson and Hjalmarsson (2009) to generalize this result as well as to show that certain optimality properties also hold from an experiment design perspective.

### Parametrization of explicit variance expressions

In this section we wish to point out that the parametrization does not affect the explicit variance expressions of Section 3.2.2. Let the assumptions in Section 3.1 be in force. Assume that the condition (3.27) of Lemma 3.2.2 holds. This implies via the chain rule (3.11) and Lemma 3.2.1 that

$$J'(\theta^o) = \sum_{k=1}^{\infty} \tau_k'(\theta^o)\mathcal{T}_k(z_o)\alpha = T'(z_o, \theta^o)\alpha. \tag{3.32}$$

The conditions (3.27) and (3.32) are closely related. They can be formulated as

$$\frac{\partial J_\tau}{\partial \tau_k} = \frac{\partial T(z_o)}{\partial \tau_k}\alpha \quad \text{and} \quad \frac{\partial J}{\partial \theta_k} = \frac{\partial T(z_o)}{\partial \theta_k}\alpha$$

respectively, and we see that the only difference lies in the parametrization and that it does not matter if the system and the property $J$ are described by the model parameters $\theta$ or by the coefficients $\tau$ of the rational orthonormal basis representation.

## 3.3 Analysis of some LTI system properties

In this section we apply the results from Sections 3.1 and 3.2.2 to some specific examples of the function $J(\theta)$. The main purpose is to show how the geometric approach can be used in the analysis, but each of these results are also of independent interest.

### Frequency response

We will first look at the covariance of the frequency response estimate, i.e., $J(\theta) = T(e^{j\omega_o}, \theta)$ for a fixed frequency $\omega_o$ when $T_o$ is stable (so that the frequency response is well defined). Then we get

$$J'(\theta^o) = T'(e^{j\omega_o}, \theta^o) = \Psi(e^{j\omega_o})R_{\text{SNR}}^{-1}(e^{j\omega_o}),$$

where $\Psi$ is given by (3.3). Now (3.26) can be applied to get the covariance expression

$$\text{AsCov}\, T(e^{j\omega_o}, \hat{\theta}_N) = R_{\text{SNR}}^{-*}(e^{j\omega_o}) \sum_{k=1}^{n} \mathcal{B}_k^*(e^{j\omega_o})\mathcal{B}_k(e^{j\omega_o})\; R_{\text{SNR}}^{-1}(e^{j\omega_o}), \tag{3.33}$$

where $\{\mathcal{B}_k\}_{k=1}^n$ is any orthonormal basis for the space $\mathcal{S}_\Psi$.

It is instructive to also consider the formulation in Theorem 3.2.1 for expressing $\text{AsCov}\, T(e^{j\omega_o}, \hat{\theta}_N)$. We will use basis functions $\mathcal{T}_k$ (3.4) defined by $\mathcal{G}_k(z) = \mathcal{H}_k(z) = z^{-(k-1)}$ so that

$$T(z, \theta) = \sum_{k=1}^{\infty} \tau_k(\theta)\mathcal{T}_k(z)$$

is parameterized in terms of the impulse responses of $G(z,\theta)$ and $H(z,\theta)$. For this problem $J_\tau(\tau(\theta)) = \sum_{k=1}^{\infty} \tau_k(\theta)\mathcal{T}_k(\mathrm{e}^{\mathrm{j}\omega_o})$ and hence $\partial J_\tau(\tau)/\partial\tau_k = \mathcal{T}_k(\mathrm{e}^{\mathrm{j}\omega_o})$ which implies that $\nabla J_\tau$, defined in (3.10), is not an $\mathcal{L}_2$-function. Thus we instead look at $J(\theta) = T(z_o,\theta)$, $z_o = r\mathrm{e}^{\mathrm{j}\omega_o}$, $r > 1$ and later we let $r \to 1$. The function $\nabla J_\tau$ is now given by

$$\nabla J_\tau(z) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \sum_{k=1}^{\infty} \bar{z}_o^{-k} z^{-k} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \frac{\bar{z}_o^{-1} z^{-1}}{1 - \bar{z}_o^{-1} z^{-1}},$$

which is a function in $\mathcal{L}_2$ so that Assumption ii) in Theorem 3.2.1 holds. Furthermore, Assumption iii) in the same lemma follows directly from Lemma 3.2.1 with $z = \mathrm{e}^{\mathrm{j}\omega_o}$ for this problem. Thus Theorem 3.2.1 applies under the assumptions in Section 3.1.

In order to calculate the projection $\mathbf{P}_{\mathcal{S}_\Psi}\{\nabla J_\tau R_{\text{SNR}}^{-*}\}$ in Theorem 3.2.1, let $\{\mathcal{B}_k\}_{k=1}^n$ be an orthonormal basis for $\mathcal{S}_\Psi$, then

$$\begin{aligned} \langle \nabla J_\tau(z) R_{\text{SNR}}^{-*}(z^{-*}), \mathcal{B}_k(z) \rangle &= \frac{1}{2\pi j} \oint_{|z|=1} \frac{\bar{z}_o^{-1} z^{-1}}{1 - \bar{z}_o^{-1} z^{-1}} R_{\text{SNR}}^{-*}(z^{-*}) \mathcal{B}_k^*(z^{-*}) \frac{\mathrm{d}z}{z} \\ &= R_{\text{SNR}}^{-*}(z_o)\mathcal{B}_k^*(z_o). \end{aligned} \tag{3.34}$$

In the second equality we use that $z^{-1} R_{\text{SNR}}^{-*}(z^{-*})\mathcal{B}_k^*(z^{-*})$ has all poles outside the unit circle (since all $\mathcal{B}_k$ contain at least one unit time delay which cancels the factor $z^{-1}$) and residue calculus, see e.g., Wunsch (1993), gives the result (3.34). The projection $\mathbf{P}_{\mathcal{S}_\Psi}\{\nabla J_\tau R_{\text{SNR}}^{-*}\}$ can now be computed as

$$\mathbf{P}_{\mathcal{S}_\Psi}\{\nabla J_\tau R_{\text{SNR}}^{-*}\} = \sum_{k=1}^{n} \langle \nabla J_\tau R_{\text{SNR}}^{-*}, \mathcal{B}_k \rangle \mathcal{B}_k = R_{\text{SNR}}^{-*}(z_o) \sum_{k=1}^{n} \mathcal{B}_k^*(z_o)\mathcal{B}_k.$$

With $z_o = r\mathrm{e}^{\mathrm{j}\omega_o}$ and $r \to 1$, we get the asymptotic covariance

$$\begin{aligned} \text{AsCov}\, T(\mathrm{e}^{\mathrm{j}\omega_o}, \hat{\theta}_N) &= R_{\text{SNR}}^{-*}(\mathrm{e}^{\mathrm{j}\omega_o}) \sum_{k=1}^{n}\sum_{p=1}^{n} \mathcal{B}_k^*(\mathrm{e}^{\mathrm{j}\omega_o}) \langle \mathcal{B}_k, \mathcal{B}_p \rangle \mathcal{B}_p(\mathrm{e}^{\mathrm{j}\omega_o})\ R_{\text{SNR}}^{-1}(\mathrm{e}^{\mathrm{j}\omega_o}) \\ &= R_{\text{SNR}}^{-*}(\mathrm{e}^{\mathrm{j}\omega_o}) \sum_{k=1}^{n} \mathcal{B}_k^*(\mathrm{e}^{\mathrm{j}\omega_o}) \mathcal{B}_k(\mathrm{e}^{\mathrm{j}\omega_o})\ R_{\text{SNR}}^{-1}(\mathrm{e}^{\mathrm{j}\omega_o}), \end{aligned} \tag{3.35}$$

which, of course, is the same as (3.33).

The covariance expression (3.33) was first established in Ninness and Hjalmarsson (2004) by employing the theory of reproducing kernels, see also Ninness and Hjalmarsson (2005b). Above we have shown that the results in Ninness and Hjalmarsson (2004) can be given an alternative system theoretic interpretation as resulting from a projection of the weighted $z$-transform of the sensitivity of the system frequency function with respect to the impulse response on a subspace determined by the model structure, the true system and the experimental conditions. The weighting function depends on the noise to signal ratio during the experiment (which

in turn depends on the experimental conditions and the true system). This work can also be seen as an extension of the work in Ninness and Hjalmarsson (2004) regarding variance analysis in frequency function estimation to general quantities $J$.

### Impulse response

In this example we look at the asymptotic variance of the coefficients $\tau_k$ of the estimated model $T(z,\theta) = \sum_{k=1}^{\infty} \tau_k(\theta)\mathcal{T}_k(z)$, cf. (3.6), but we assume that only the first $n_\tau$ coefficients are of interest and we let

$$J_\tau(\tau) = \tau = \begin{bmatrix} \tau_1 & \cdots & \tau_{n_\tau} \end{bmatrix}.$$

Now $\frac{\mathrm{d}J_\tau}{\mathrm{d}\tau} = I$ (the identity matrix) and hence

$$\nabla J_\tau(z) = \begin{bmatrix} \mathcal{T}_1(z) \\ \vdots \\ \mathcal{T}_{n_\tau}(z) \end{bmatrix}.$$

It is straightforward to verify the chain rule, Assumption ii) in Theorem 3.2.1. Thus the asymptotic covariance can be expressed as

$$\begin{aligned} \text{AsCov } \tau(\hat{\theta}_N) &= \langle \mathbf{P}_{\mathcal{S}_\Psi}\{\nabla J_\tau R_{\text{SNR}}^{-*}\}, \mathbf{P}_{\mathcal{S}_\Psi}\{\nabla J_\tau R_{\text{SNR}}^{-*}\}\rangle \\ &\leq \langle \nabla J_\tau \Phi_v \Phi_\chi^{-1}, \nabla J_\tau \rangle, \end{aligned} \tag{3.36}$$

where the inequality comes from Theorem 3.2.2. If we consider the impulse response coefficients $g_k$ and $h_k$ corresponding to $\mathcal{G}_k(z) = \mathcal{H}_k(z) = z^{-k}$ in (3.4) we get for the $2 \times 2$ diagonal blocks of (3.36) that

$$\text{AsCov } \begin{bmatrix} g_k(\hat{\theta}_N) & h_k(\hat{\theta}_N) \end{bmatrix} \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_v(\mathrm{e}^{\mathrm{j}\omega})\Phi_\chi^{-1}(\mathrm{e}^{\mathrm{j}\omega})\,\mathrm{d}\omega.$$

### $\mathcal{L}_2$-norm

Now we consider the asymptotic variance of the $\mathcal{L}_2$-norm of the estimated model $G(z,\theta) = \sum_{k=1}^{\infty} g_k(\theta)\mathcal{G}_k(z)$, cf. (3.5). The $\mathcal{L}_2$-norm is given by

$$\|G(\cdot,\theta)\| = \sqrt{\langle G(\cdot,\theta), G(\cdot,\theta)\rangle} = \sqrt{\sum_{k=1}^{\infty} g_k^2(\theta)}$$

and the function $\nabla J_\tau^g(z)$ is

$$\nabla J_\tau^g(z) = \sum_{k=1}^{\infty} \frac{g_k(\theta^o)\mathcal{G}_k(z)}{\sqrt{\sum_{l=1}^{\infty} g_l^2(\theta^o)}} = \frac{G_o(z)}{\|G_o\|} \in \mathcal{L}_2.$$

It is straightforward to verify the chain rule, Assumption ii) in Theorem 3.2.1. Thus we can use Corollary 3.2.1 to express the asymptotic variance as

$$\text{AsVar } \|G(\cdot,\hat{\theta}_N)\| = \frac{\left\|\mathbf{P}_{\mathcal{S}_\Psi}\left\{\begin{bmatrix}\frac{\sqrt{\lambda_o}G_oH_o^*}{S_o^*R^*} & 0\end{bmatrix}\right\}\right\|^2}{\|G_o\|^2}. \tag{3.37}$$

The projection in (3.37) may be cumbersome to calculate, but we can use Corollary 3.2.2 to get an upper bound of the asymptotic variance:

$$\text{AsVar } \|G(\cdot,\hat{\theta}_N)\| \leq \frac{\|G_o\|^2_{\Phi_v/\Phi_u^r}}{\|G_o\|^2}.$$

The bound can also be written in the form

$$\text{AsVar } \|G(\cdot,\hat{\theta}_N)\| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi_v(\omega)}{\Phi_u^r(\omega)}w_G(\omega)\,\mathrm{d}\omega,$$

where $w_G(\omega) = |G_o(\mathrm{e}^{\mathrm{j}\omega})|^2/\|G_o\|^2$ is a weighting function with $\frac{1}{2\pi}\int_{-\pi}^{\pi} w_G(\omega)\,\mathrm{d}\omega = 1$. The weighting function $w_G(\omega)$ gives more weight to frequencies where the gain is large.

### Non-minimum phase zeros

Next we consider estimation of NMP-zeros of a stable system $G_o$. The zeros of the system are defined as the solutions $z$ to the equation $G(z,\theta) = \sum_{k=1}^{\infty} g_k(\theta)z^{-k} = 0$ and we assume that the zero of interest, $z_o$, is non-minimum phase, i.e., $|z_o| > 1$. The quantity of interest is thus $J(\hat{\theta}_N) = z_o(\hat{\theta}_N)$. Theorem 3.2.2 is here, and since $J$ is independent of the noise model $H$ we can use the simplified version (3.20) in Corollary 3.2.1. Similar to Mårtensson and Hjalmarsson (2009) we obtain

$$\frac{\partial J_\tau^g(\tau^o)}{\partial g_k} = -\frac{z_o}{\widetilde{G}_o(z_o)}z_o^{-k},$$

where $\widetilde{G}_o(z) = G_o(z)/(1 - z_o z^{-1})$, which gives, for $|z| > 1/|z_o|$, that

$$\begin{aligned}\nabla J_\tau^g(z) &= -\frac{\bar{z}_o}{\overline{\widetilde{G}_o(z_o)}}\sum_{k=1}^{\infty}\bar{z}_o^{-k}z^{-k}\\ &= -\frac{\bar{z}_o}{\overline{\widetilde{G}_o(z_o)}}\frac{\bar{z}_o^{-1}z^{-1}}{(1-\bar{z}_o^{-1}z^{-1})},\end{aligned}$$

which is in $\mathcal{L}_2$. It is straightforward to verify that the chain rule (Assumption iii) in Theorem 3.2.1) applies: Suppose, for simplicity, that the numerator polynomial, $B(\mathrm{q},\theta)$, in $G(\mathrm{q},\theta) = B(\mathrm{q},\theta)/A(\mathrm{q},\theta)$ is independently parameterized of $A(\mathrm{q},\theta)$.

Lemma 3.2.1 gives that $g_k'(\theta^o) = \langle G'(z,\theta^o), z^{-k}\rangle$ so that

$$\begin{aligned}\sum_{k=1}^{\infty} g_k'(\theta^o)\,\frac{\partial J_\tau^g(\tau^o)}{\partial g_k} &= -\frac{z_o}{\widetilde{G}_o(z_o)}\,\sum_{k=1}^{\infty}\langle G'(z,\theta^o), z^{-k}\rangle\, z_o^{-k}\\ &= -\frac{z_o}{\widetilde{G}_o(z_o)}\,G'(z_o,\theta^o)\\ &= -\frac{z_o}{\widetilde{B}_o(z_o)}\,\begin{bmatrix}B'(z_o(\theta^o),\theta^o)\\ 0\end{bmatrix},\end{aligned}$$

where $\widetilde{B}_o(z) = B(z,\theta^o)/(1-z_o z^{-1})$. The last expression then equals $J'(\theta^o)$ (Mårtensson and Hjalmarsson, 2009).

We have thus verified the conditions in Theorem 3.2.1 and the asymptotic variance can be calculated in the same way as in (3.34)-(3.35), or alternatively (3.26) could be applied, and, without giving all details, we get

$$\text{AsVar}\; z_o(\hat{\theta}_N) = \frac{\lambda_0|z_o|^2}{|\widetilde{G}_o(z_o)|^2}\,\frac{|H_o(z_o)|^2}{|R(z_o)|^2}\,\sum_{k=1}^{n}|\mathcal{B}_k^1(z_o)|^2, \tag{3.38}$$

where $\mathcal{B}_k \coloneqq [\mathcal{B}_k^1, \mathcal{B}_k^2]$ and $\{\mathcal{B}_k\}_{k=1}^n$ is an orthonormal basis for $\mathcal{S}_\Psi$. The expression (3.38) is used in Mårtensson and Hjalmarsson (2009) where also explicit expressions and bounds for $\sum_{k=1}^n|\mathcal{B}_k^1(z_o)|^2$ are derived for certain model structures. When an orthonormal basis of the type given in Section 2.2 can be used, we see from (3.38) that the asymptotic variance for a NMP-zero will be large if there is another NMP-zero nearby. This follows from that the factor $|\widetilde{G}_o(z_o)|^2$ in the denominator will be small in this case and that the basis functions are independent of the system zeros.

Bounds on the asymptotic variance for NMP zeros can be derived using (3.30):

$$\text{AsVar}\; z_o(\hat{\theta}_N) \le \frac{\lambda_0}{(1-|z_o|^{-2})|\widetilde{G}_o(z_o)|^2}\,\frac{|H_o(z_o)|^2}{|S_o(z_o)R(z_o)|^2}, \tag{3.39}$$

or using Theorem 3.2.2 (see also Corollary 3.2.2.):

$$\begin{aligned}\text{AsVar}\; z_o(\hat{\theta}_N) &\le \frac{1}{|\widetilde{G}_o(z_o)|^2}\left\|\frac{1}{(1-\bar{z}_o^{-1}z^{-1})}\right\|^2_{\Phi_v/\Phi_u^r}\\ &= \frac{\lambda_0}{(1-|z_o|^{-2})|\widetilde{G}_o(z_o)|^2}\left\|\frac{H_o}{S_oR}\right\|^2_{w_Z},\end{aligned} \tag{3.40}$$

where $w_Z(\omega) = (1-|z_o|^{-2})/|1-z_o^{-1}\mathrm{e}^{\mathrm{j}\omega}|^2$ can be seen as a weighting function with $\frac{1}{2\pi}\int_{-\pi}^{\pi} w_Z(\omega)\mathrm{d}\omega = 1$. The weighting function $w_Z(\omega)$ focuses on frequencies around $\omega = \arg(z_o)$. Sufficient accuracy of an NMP-zero estimate is thus guaranteed if $|H_o/(S_oR)|$ is small in this frequency range.

The bound (3.40) is always larger than the bound (3.39) which is a tight bound in the sense that equality holds (in many cases) when the model order $n$ goes to

infinity, see Mårtensson and Hjalmarsson (2009), which provides a quite complete asymptotic variance analysis of both zero and pole estimates.

Before closing this section, we illustrate the idea outlined at the end of Section 3.1, i.e., that the input can be used to make the asymptotic variance the same for different model structures and arbitrary model orders. For simplicity we will assume that the NMP zero is real and that the system is operating in open loop so that the prediction error gradient (3.3) is given by

$$\Psi(z) = T'(z, \theta^o) \begin{bmatrix} \frac{R(z)}{\sqrt{\lambda_o}\, H_o(z)} & 0 \\ 0 & \frac{1}{H_o(z)} \end{bmatrix}.$$

Assume first that an output error model is used and that the spectral factor of the input is chosen as $R(z) = 1/(z - z_o^{-1})$. When $n_b \geq n_f + 1$, as discussed in Section 2.2, it then follows that (2.23) is an orthonormal basis with $\xi_1 = z_o^{-1}$ and the other poles being the poles of the true system dynamics are counted twice. Similar to (3.31), (3.38) then collapses to

$$\text{AsVar } z_o(\hat{\theta}_N) = \frac{\lambda_0 |z_o|^2}{|\widetilde{G}_o(z_o)|^2} \frac{|H_o(z_o)|^2}{|R(z_o)|^2} \sum_{k=1}^{n} |\mathcal{B}_k^1(z_o)|^2 = \frac{\lambda_0 \left(|z_o|^2 - 1\right)}{|\widetilde{G}_o(z_o)|^2}.$$

Due to the presence of $R(z)$ in $\Psi$ it can easily be shown that exactly the same result is obtained for Box-Jenkins models (under the same order condition on $n_b$ and $n_f$). Thus an input with spectral factor $R(z) = 1/(z - z_o^{-1})$ ensures that the asymptotic variance of an estimate of an NMP-zero at $z_o$ becomes independent both of the model and system order and also the model structure. One may argue that this choice of input is infeasible since it depends on the to be estimated zero. However, this insight is of independent value; it has been shown that using an estimate of the NMP-zero in $R$ instead, an adaptive scheme will achieve the optimal asymptotic variance, under quite mild conditions, see Rojas *et al.* (2011).

## 3.4 Conclusions

The main results in this chapter are the formulae (3.12) and reparametrization of (3.26) which expresses the asymptotic covariance as defined by (2.15). We have shown that these geometric expressions provide insights into how model structure, model order, true system dynamics, and experimental conditions affect the asymptotic covariance. In particular we demonstrated that one can use the experimental conditions to make the asymptotic variance independent of model order and model structure in some cases.

We have also used these expressions to derive novel model structure independent upper bounds of the asymptotic covariance, in particular for a number of commonly estimated quantities such as system zeros and gains and impulse response coefficients. We have shown that these bounds are significantly less conservative as

compared to the variance expressions that result from using the (asymptotic in model order) variance formulae for frequency function estimation in Ljung (1985).

Our work has its foundation in Ninness and Hjalmarsson (2004), where the significance of the subspace spanned by the prediction error gradient was acknowledged and we have shown that the results in Ninness and Hjalmarsson (2004) are recovered from the results in this chapter.

# Chapter 4

# Cascade models

We make the interconnection structure more complicated compared to the previous chapter by considering cascaded modules. Cascaded modules are very common in applications; consider for example the flotation plant example in Section 1.1. Our concern in this chapter is the accuracy of the model of the first module, and we try to give some insight into how much the accuracy is improved by additional measurements. It turns out that the zero locations of the first module play an important role and we analyze how the zeros of the first module affect the accuracy. We quantify the contribution of each sensor to the accuracy, with focus on a frequency range dependent on the zeros of the transfer function of interest. This information might be useful in determining if we should use additional sensors. There is a trade-off to be made since additional sensors means more data have to be collected, and more parameters have to be identified. Hence, if the gain in accuracy is concentrated mainly in regions of little interest, the additional effort may not be worthwhile. For FIR systems, the results are illustrated by numerical simulations. A surprising special case occurs when the first module contains a zero on the unit circle; as the model orders of all modules grow large at the same rate, the variance of the frequency function estimate, evaluated at the corresponding frequency of the unit-circle zero, is shown to be the same as the case where the other modules are completely known.

## 4.1 Problem formulation

Consider the SISO LTI systems connected in cascade as depicted in Figure 4.1, the output of which can be described by the following set of equations:

$$y_k(t)=\prod_{i=1}^{k} G_i(q)u(t) + e_k(t), \qquad k=1,\ldots,m. \tag{4.1}$$

Define $y(t) = [y_1(t), \ldots, y_m]^{\mathrm{T}}$, $e(t) = [e_1(t), \ldots, e_m]^{\mathrm{T}}$ and $T$ as the row vector containing all frequency responses between $u$ and $y(t)^{\mathrm{T}}$, i.e.,

$$T \coloneqq \begin{bmatrix} G_1 & G_2 G_1 & \ldots & G_m \cdots G_1 \end{bmatrix}.$$

Then we can express the relationship between the input, noise and outputs as follows:

$$y(t) = T^{\mathrm{T}} u(t) + e(t).$$

We assume that the additive noise sequences, $\{e_i(t)\}$ are mutually independent zero mean white noise sequences, independent of the input $u(t)$, with variances $\lambda_i, i = 1, \ldots, m$, i.e.,

$$\operatorname{Cov} e(t) = \Lambda,$$

where $\Lambda = \operatorname{diag}\{[\lambda_1, \ldots, \lambda_m]\}$. The models of the modules are independently parameterized, module $G_i$ is parameterized with $\theta_i \in \mathbb{R}^{d_i}$, $n_i \in \mathbb{R}$ for all $i = 1, \ldots, m$, and let $\theta = \left[\theta_1^{\mathrm{T}}, \ldots, \theta_m^{\mathrm{T}}\right]^{\mathrm{T}}$ and $n = \sum_{i=1}^{m} n_i$. We assume the true system is in the model set and denote the true parameters by $\theta^o$, i.e.,

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

We are interested in estimating the first module $G_1$. The predictor of $y(t)$ based on $\theta$ is given by

$$\hat{y}(t) = T(\theta)^{\mathrm{T}} u(t).$$

The predictor error gradient defined in (2.12), of dimension $n \times m$, is then given by

$$\psi(t) = -T'(\theta) u(t),$$

where $T'(\theta) \in \mathbb{C}^{n \times m}$ is the gradient of $T$ with respect to $\theta$ given by

$$T' = \begin{bmatrix} G_1' & G_2 G_1' & \cdots & G_m \cdots G_2 G_1' \\ 0 & G_2' G_1 & \ddots & G_m \cdots G_3 G_2' G_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & G_m' G_{m-1} \cdots G_1 \end{bmatrix},$$

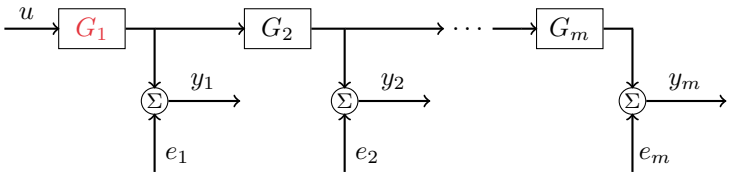


Figure 4.1: Cascaded structure.

where $G_i'$ is the gradient of $G_i$ with respect to $\theta$, for $i = 1, \ldots, m$. From (2.12) and Parseval's relation we can express the asymptotic covariance of the parameter estimate as

$$\begin{aligned} \operatorname{AsCov} \hat{\theta}_N &= \langle \Psi, \Psi \rangle^{-1} \\ \Psi &= T'(\theta^o) R \Lambda^{-1/2} \in \mathcal{L}_2^{n \times m}, \end{aligned}$$

where $R$ is the spectral factor of the input power spectrum $\Phi_u(\omega)$. From this equation it is clear that we can rewrite

$$T'(\theta^o) = \Psi \Lambda^{1/2} R^{-1},$$

and hence, with $L = \Lambda^{1/2} R^{-1}$, we may use Lemma 2.3.2 to express the asymptotic covariance matrix for the estimate of the transfer functions from the input to the outputs as

$$\operatorname{AsCov} \hat{T}(e^{j\omega}) = \frac{1}{\Phi_u(\omega)} \Lambda^{1/2} \sum_{k=1}^{l} \mathcal{B}_k^*(e^{j\omega}) \mathcal{B}_k(e^{j\omega}) \Lambda^{1/2}. \tag{4.2}$$

We are only interested in $\hat{G}_1$, so we only need to consider the top left corner of (4.2). To this end, without loss of generality, the basis functions can be expressed according to the following definition:

**Definition 4.1.1.** *The basis functions $\{\mathcal{B}_k\}$ are expressed*

$$\mathcal{B}_k =: \begin{bmatrix} c_k^1 \mathcal{B}_k^1 & \ldots & c_k^m \mathcal{B}_k^m \end{bmatrix}, \tag{4.3}$$

*where $c_k^j \in \mathbb{R}_{\geq 0}$. If $\mathcal{B}_k^j \neq 0$, then $\mathcal{B}_k^j$ has unit $\mathcal{L}_2$-norm for all $j$ and $k$, i.e., $\|\mathcal{B}_k^j\| = 1$ for all $k \leq l, j \leq m$, where $\|\cdot\|$ is defined in (2.21). For each basis function $\mathcal{B}_k$, the coefficients $\{c_k^j\}_{j=1}^m$ satisfy $\| \begin{bmatrix} c_k^1 & \ldots & c_k^m \end{bmatrix} \|_2^2 = 1$, which ensures that $\|\mathcal{B}_k\| = 1$. If $\mathcal{B}_k^j = 0$ then $c_k^j$ is defined to be zero.*

With this notation, we obtain

$$\operatorname{AsVar} \hat{G}_1(e^{j\omega}) = \frac{\lambda_1}{\Phi_u(\omega)} (c_k^1)^2 \sum_{k=1}^{l} |\mathcal{B}_k^1(e^{j\omega})|^2. \tag{4.4}$$

In the next section we will provide a specific basis function $\mathcal{B}_1$ that is used to obtain insights into (4.4), insight into how much $\operatorname{AsVar} \hat{G}_1(e^{j\omega})$ is reduced by additional measurements. The basis function will be determined by a zero of $G_1(e^{j\omega})$ and we will analyze $\operatorname{AsVar} \hat{G}_1(e^{j\omega})$ with focus on a frequency range dependent on this zero.

## 4.2 Basis functions and variance results

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

$$\left\langle \mathcal{B}_1^i, \mathcal{B}_j^i \right\rangle = 0, \qquad i = 1, \ldots, l \quad j = 2, \ldots, m. \tag{4.5}$$

Not only is $\mathcal{B}_1$ orthogonal to all other basis functions $\mathcal{B}_k$, $k = 2, \ldots, l$, but also, every column of $\mathcal{B}_1$ is orthogonal the the corresponding column in all other basis functions $\mathcal{B}_k$, $k = 2, \ldots, l$. This allows us to compute the variance expression (4.4) at a specific frequency point.

**Lemma 4.2.1.** *Let $G_1$ have a zero in $\xi \in \mathbb{C}$ and none of $\{G_k\}_{k=1}^{m}$ have a pole in $\xi$, i.e., there is no pole-zero cancellation of the zero $\xi$ in $\Psi$. Let $\{\mathcal{A}_k^i\}_{k=1}^{l_i}$ be a basis (not necessarily orthonormal) for the space spanned by the rows of the $i$-th column of $\Psi$ for $i = 1, \ldots, m$. Then we can choose an orthonormal basis $\{\mathcal{B}_k\}_{k=1}^{l}$ for $\Psi$ such that (cf. (4.3))*

$$\left\langle \mathcal{B}_1^i, \mathcal{B}_j^i \right\rangle = 0, \qquad i = 1, \ldots, m \quad j = 2, \ldots, l, \tag{4.6}$$

*where*

$$c_1^1 = \frac{b_1}{\lambda_1^{1/2}} \left( \frac{b_1^2}{\lambda_1^{1/2}} + \frac{b_2}{\lambda_2^{1/2}} |G_2(\xi)|^2 + \cdots + \frac{b_m^2}{\lambda_m^{1/2}} |G_m(\xi) \cdots G_2(\xi)|^2 \right)^{-1/2}, \tag{4.6a}$$

$$\mathcal{B}_1^i = b_i \sum_{k=1}^{l_i} \overline{\mathcal{A}_k^i(\xi)} \mathcal{A}_k^i, \quad i = 1, \ldots, m, \tag{4.6b}$$

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

$$b_i^2 = \left( \sum_{k=1}^{l_i} |\mathcal{A}_k^i(\xi)|^2 \right)^{-1},$$

*and $b_i^2 = l_i^{-1}$ when $\mathcal{A}_k^i$ is an orthonormal basis, e.g., the basis $\{z^{-k}\}$.*

*Proof.* See Appendix 4.B. □

Notice that (4.6) does not follow from the orthogonality of the basis functions $\{\mathcal{B}_k\}$ as the next example illustrates.

**Example 4.2.1.** *Consider*

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

*formed from* $\{\mathcal{A}_k^1\}_{k=1}^1 = z^{-1}$ *and* $\{\mathcal{A}_k^2\}_{k=1}^1 = z^{-2}$. *Recall that* $\{z^{-k}\}_{k=0}^\infty$ *forms a (complete) orthonormal set (Friedman, 1970). Then one choice of basis functions is*

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

*and* $\langle \mathcal{B}_1, \mathcal{B}_2\rangle = 0$, *even though* $\left\langle \mathcal{B}_1^1, \mathcal{B}_2^1\right\rangle = \left\langle z^{-1}, z^{-1}\right\rangle = 1$.

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

**Theorem 4.2.1.** *Assume that* $G_1$ *has a zero at* $\xi = re^{j\omega_\xi}$, *where* $r \in \mathbb{R}_{\geq 0}$ *and that the assumptions in Lemma 4.2.1 hold. Then*

$$\operatorname{AsVar} \hat{G}_1(e^{j\omega_\xi}) \leq \frac{\lambda_1}{\Phi_u(\omega_\xi)}\|f_1\|^2\left(1 - (1-(c_1^1)^2)\left\langle \mathcal{B}_1^1, f_1/\|f_1\|\right\rangle^2\right), \tag{4.7}$$

*where* $c_1^1$ *and* $\mathcal{B}_1^1$ *are chosen according to Lemma 4.6 and where*

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

*If* $r = 1$, *i.e., the zero is on the unit circle, then*

$$\operatorname{AsVar} \hat{G}_1(e^{j\omega_\xi}) = \frac{\lambda_1}{\Phi_u(\omega_\xi)}(c_1^1)^2 b_1^{-2}, \tag{4.8}$$

*or equivalently*

$$\operatorname{AsVar} \hat{G}_1(e^{j\omega_\xi}) = Z^{-1}, \tag{4.9}$$

*where*

$$Z = b_1^2\frac{\Phi_u(\omega_\xi)}{\lambda_1} + b_2^2\frac{\Phi_u(\omega_\xi)}{\lambda_2}\left|G_2(e^{j\omega_\xi})\right|^2 + \cdots + b_m^2\frac{\Phi_u(\omega_\xi)}{\lambda_m}\left|G_m(e^{j\omega_\xi})\cdots G_2(e^{j\omega_\xi})\right|^2.$$

*Proof.* Using Lemma 4.A.1 we can rewrite (4.4) as

$$\operatorname{AsVar} \hat{G}_1(re^{j\omega_\xi}) = \frac{\lambda_1}{\Phi_u(\omega_\xi)}\sum_{k=1}^{l}(c_k^1)^2\left|\left\langle \mathcal{B}_k^1, f_1\right\rangle\right|^2. \tag{4.10}$$

To arrive at (4.7) from (4.10), we notice that we can bound the contribution from the other basis functions through $c_k \leq 1$ and $\sum_{k=1}^l \left|\left\langle f_1, \mathcal{B}_k^1\right\rangle\right|^2 = \|f_1\|^2$, where the second equality is due to Lemma 4.A.1 in Appendix 4.A. We can form the upper

bound

$$\begin{aligned}\sum_{k=1}^{l}(c_k^1)^2\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1\rangle\right|^2 &= (c_1^1)^2\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1,\rangle\right|^2+\sum_{k=2}^{l}(c_k^1)^2\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1\rangle\right|^2\\ &\le (c_1^1)^2\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1,\rangle\right|^2+\sum_{k=2}^{l}\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1\rangle\right|^2\\ &\le (c_1^1)^2\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1,\rangle\right|^2+(1-\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1,\rangle\right|^2)\\ &= 1-\left(1-(c_1^1)^2\right)\left|\langle f_1/\|f_1\|,\mathcal{B}_k^1,\rangle\right|^2.\end{aligned}$$

Using this bound in (4.10) gives (4.7). If $r = 1$, i.e., when the zero is on the unit circle, then $f_1 = b_1^{-1}\mathcal{B}_1^1$, i.e., it is a scaled version of the first basis function. The orthogonality of $\{\mathcal{B}_k^1\}$ gives that (4.10) simplifies to

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

By inserting the value of $c_1^1$ provided by Lemma 4.6, (4.8) and (4.9) follows. □

**Remark 4.2.1.** *If $G_1$ has a unit circle zero in $\xi = re^{j\omega_\xi}$, then* (4.9) *and* (4.9) *in Theorem 4.2.1 are exact expression for the asymptotic variance of the transfer function estimate* $\text{Var}\,\hat{G}_1(e^{j\omega})$ *at the frequency $\omega = \omega_\xi$. In* (4.9)*, $Z$ is a weighted sum of terms, sensor $k$ corresponds to a term*

$$b_k^2\frac{\Phi_{SNR}^k(\omega)}{|G_1(\mathrm{e}^{\mathrm{j}\omega})|^2},$$

*where $\Phi_{SNR}^k(\omega)$ is the signal to noise ratio at sensor $k$, i.e., for sensor $k$ we have that*

$$\Phi_{SNR}^k(\omega) = \frac{\Phi_u(\omega)|G_1(\mathrm{e}^{\mathrm{j}\omega})\cdots G_k(\mathrm{e}^{\mathrm{j}\omega})|^2}{\lambda_k}$$

*This means that a higher ratio between the gain from the output of $G_1$ to a given sensor and the noise variance leads to better accuracy. The constants $\{b_j\}_{j=1}^m$, i.e., the weights for the different signal to noise ratios in $Z$, provide a scaling that in some sense depends on the "size" of the corresponding spaces. This weighting gives that sensors far away from $G_1$ (more to the right in Figure 4.1) contribute less to the accuracy of the model of $G_1$. If we only use the first measurement, we get*

$$\text{AsVar}\,\hat{G}_1(e^{j\omega_\xi})) = \frac{\lambda_k}{\Phi_u(\omega)}\frac{1}{b_k^2}$$

*For fixed denominator models, e.g., FIR models, this is exactly the high model order approximation 1.4, so this constant reflects the contribution to the model error from the model complexity. The expressions* (4.9) *and* (4.9)*) give insight into how much the variance of $\hat{G}_1(e^{j\omega_\xi})$ would increase if we for example remove one sensor since this corresponds to removing the corresponding term in $Z$.*

**Remark 4.2.2.** *Similar observations can be made when $r \neq 1$ regarding the upper bound in* (4.7)*. If we take a closer look at $c_1^1$, we see that a similar weighted sum of the signal to noise ratio $\Phi_{SNR}^k(\omega)$ of the sensors appears, almost the same as in $Z$. This means that a higher ratio between the gain from the output of $G_1$ to a given sensor and the noise variance leads to better accuracy. Again, $\{b_j\}_{j=1}^m$, i.e., the weights for the different signal to noise ratios in $c_1^1$, provide a scaling that in some sense depends on the "size" of the corresponding spaces. This weighting gives that sensors far away from $G_1$ (more to the right in Figure 4.1) contribute less to the accuracy of the model of $G_1$.*

**Remark 4.2.3.** *Note that if one module is zero at $e^{j\omega_\xi}$, i.e., $G_j(e^{j\omega_\xi}) = 0$, then sensors downstream of $G_j$ will not reduce the variance of $G_1(e^{j\omega_\xi})$ regardless of their signal-to-noise ratio. This should come as no surprise since any measurement downstream of $G_j$ will be zero, regardless of $G_1(e^{j\omega_\xi})$. This generalizes to the following observation: if $G_j(e^{j\omega_\xi})$ is small, then the benefit to the accuracy of $\hat{G}_1(e^{j\omega_\xi})$ from sensors downstream of $G_j$ is limited. These results are related to those found in Wahlberg et al. (2009), where, in the case of three modules, knowing $y_2$ and $y_3$ will not improve the quality of the estimate of $\theta_1$, if $G_2(q,\theta_2^o)G_1'(q,\theta_1^o) = G_2'(q,\theta_2^o)G_1(q,\theta_1^o)$. Our results have a weaker condition and provide a weaker result, that is, if one zero is common between $G_1$ and $G_2$, then $y_2$ and $y_3$ will not considerably improve the quality of the estimate around the corresponding frequency.*

The previous result is valid for $\hat{G}_1(e^{j\omega_\xi})$, i.e., at the same frequency as the zero. We will in the following corollary extend (4.7) to hold for all frequencies.

**Corollary 4.2.1.** *Let the assumptions of Theorem 4.2.1 hold, where $G_1$ has a zero at $\xi = re^{j\omega_\xi}$. Then*

$$\text{AsVar}\,\hat{G}_1(e^{j\omega}) \leq \frac{\lambda_1}{\Phi_u(e^{j\omega})}\|f_1\|^2\left(1-(1-(c_1^1)^2)\left\langle \mathcal{B}_1^1, f_1/\|f_1\|\right\rangle^2\right), \tag{4.11}$$

*where*

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

*and where $c_1^1$ and $\mathcal{B}_1^1$ are chosen according to Lemma 4.6.*

*Proof.* Follows from the proof of Theorem 4.2.1, by changing $\omega_\xi$ to $\omega$ in the definition of $f_1$, but keeping $c_1^1$ and $\mathcal{B}_1^1$ according to Lemma 4.6. □

**Remark 4.2.4.** *if $\omega \approx \omega_\xi$, we can expect $\mathcal{B}_1^1$ and $f_1$ to be similar and that* (4.11) *will provide a rather tight upper bound. However, if $f \perp \mathcal{B}_1$, i.e., $\langle f, \mathcal{B}_1\rangle = 0$, then* (4.11) *gives the upper bound*

$$\text{AsVar}\,\hat{G}_1(e^{j\omega}) \leq \frac{\lambda_1}{\Phi_u(e^{j\omega})}\|f_1\|^2,$$

*which is the same as if we would only use the first measurement to estimate* $\hat{G}_1(e^{j\omega_\xi})$.

**Remark 4.2.5.** *Theorem 4.2.1 and Corollary 4.2.1 only consider one basis function determined by one zero of* $G_1$*. If we know additional zeros, Theorem 4.2.1 provides different sets of basis functions for the different zeros and corresponding lower bounds for the variance at the corresponding frequencies. However, using Gram-Smith orthogonalization we can determine orthonormal basis functions of one set and refine the lower bounds in Theorem 4.2.1 and Corollary 4.2.1 accordingly.*

## 4.3 Several branches

In this section we generalize the previous result to an arbitrary number of branches. We only consider the case of one additional branch, even though the generalization

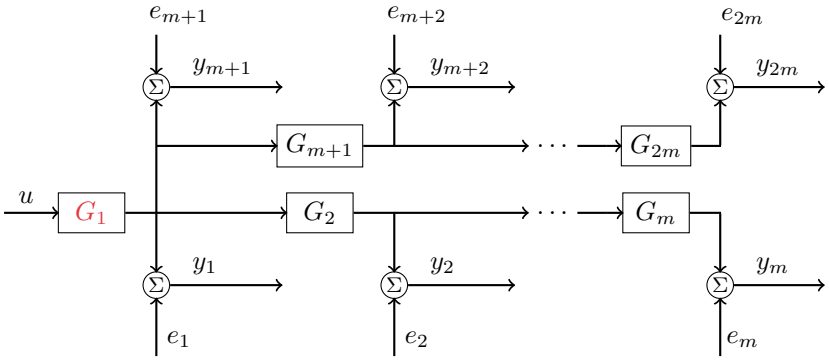


Figure 4.2: Cascaded structure with two branches.

to a tree structure is straightforward. We assume that we have an extra branch of $m$ systems, which are connected in cascade, with the output of $G_1$ as input to the first additional system as depicted in Figure 4.2. The additional module $G_i$ is parameterized with $\theta_i \in \mathbb{R}^{n_i}$, $n_i \in \mathbb{Z}_{\geq 0}$ for all $i = 1, \ldots, 2m$. The parameter vector is extended to $\theta_e = \left[\theta^{\mathrm{T}}, \theta_{m+1}^{\mathrm{T}}, \ldots, \theta_{2m}^{\mathrm{T}}\right]^{\mathrm{T}}$ and let $n_e = \sum_{i=1}^{2m} n_i$. The measured extra outputs can be described by the following set of equations:

$$y_k(t) = \left(\prod_{i=1}^{k} G_{m+i}(q)\right) G_1(q)u(t) + e_k(t), \qquad k = m+1, \ldots, 2m. \tag{4.12}$$

The noise satisfies

$$\operatorname{Cov} e(t) = \Lambda_e,$$

where $\Lambda_e = \operatorname{diag}\{\lambda_1, \ldots, \lambda_{2m}\}$. Before we continue, we partition $T$ and $\Psi$ according to

$$T = \begin{bmatrix} G_1 & T_1 \end{bmatrix}.$$

We append $T$ as follows:

$$T_e \coloneqq \begin{bmatrix} G_1 & T_1 & T_2 \end{bmatrix}, \quad T_2 = \begin{bmatrix} G_{m+1}G_1 & G_{m+2}G_{m+1}G_1 & \cdots & G_{2m}\cdots G_{m+1}G_1 \end{bmatrix}.$$

The predictor error gradient is then of dimension $n_e \times 2m$, and is given by

$$\psi = -T_e'(\theta)u(t),$$

where $T'(\theta) \in \mathbb{C}^{n_e \times 2m}$ is the gradient of $T_e$ with respect to $\theta$, and is given by

$$T_e' = \begin{bmatrix} G_1' & G_1'T_1 & G_1'T_2 \\ 0 & T_1'G_1 & 0 \\ 0 & 0 & T_2'G_1 \end{bmatrix}. \tag{4.13}$$

The variance of the estimated parameters can also this time be expressed as

$$\begin{aligned} \operatorname{AsCov} \hat{\theta}_e &= \langle \Psi_e, \Psi_e \rangle^{-1} \\ \Psi_e &= T_e'(\theta^o) R \Lambda_e^{-1/2} \in \mathcal{L}_2^{n_e \times 2m}. \end{aligned}$$

It is straightforward to extend the previous results to this case. The only thing that is different is the constant $c_1^1$ in Lemma 4.2.1, Theorem 4.2.1 and corollary 4.2.1. The constant $c_1^1$ now becomes:

$$\begin{aligned} c_1^1 = b_1 \lambda_1^{-1/2} \big( & b_1^2 \lambda_1^{-1} + b_2^2 \lambda_2^{-1} |G_2(\xi)|^2 + \cdots + b_l^2 \lambda_m^{-1} |G_m(\xi) \cdots G_2(\xi)|^2 \\ & b_{m+1}^2 \lambda_{m+1}^{-1} |G_{m+1}(\xi)|^2 + \cdots + b_{2m}^2 \lambda_{2m}^{-1} |G_{2m}(\xi) \cdots G_{m+1}(\xi)|^2 \big)^{-1/2} \end{aligned} \tag{4.14}$$

and subsequently $Z$ in (4.9) is changed accordingly. We get

$$\begin{aligned} Z = & b_1^2 \frac{\Phi_u(\omega_\xi)}{\lambda_1} \\ & + b_2^2 \frac{\Phi_u(\omega_\xi)}{\lambda_2} \left| G_2(e^{j\omega_\xi}) \right|^2 + \ldots + b_m^2 \frac{\Phi_u(\omega_\xi)}{\lambda_m} \left| G_m(e^{j\omega_\xi}) \cdots G_2(e^{j\omega_\xi}) \right|^2 \\ & + b_{m+1}^2 \frac{\Phi_u(\omega_\xi)}{\lambda_{m+1}} \left| G_{m+1}(e^{j\omega_\xi}) \right|^2 + \ldots + b_{2m}^2 \frac{\Phi_u(\omega_\xi)}{\lambda_{2m}} \left| G_{2m}(e^{j\omega_\xi}) \cdots G_{m+1}(e^{j\omega_\xi}) \right|^2. \end{aligned} \tag{4.15}$$

The remarks concerning Theorem 4.2.1 and corollary 4.2.1 also applies in this case.

**Remark 4.3.1.** *We see that a sensor $k$ adds a term*

$$b_k^2 \frac{\Phi_{SNR}^k(\omega)}{|G_1(\mathrm{e}^{\mathrm{j}\omega})|^2},$$

*where $\Phi_{SNR}^k(\omega)$ is the signal to noise ratio at sensor $k$. We see that also in this case, the weighting $b_k^2$ gives that sensors far away from $G_1$ (more to the right in Figure 4.2) contribute less to the accuracy of $\hat{G}_1$. Also note that adding one branch, i.e., adding modules $G_{m+1}, \ldots, G_{2m}$, do not reduce (nor increase) the weights $b_k$ for $2 \leq k \leq m$; the contribution of the first branch stays the same.*

## 4.4 Numerical FIR examples

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

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

The first system $G_1$ has a zero at $\xi = e^{j\omega_\xi}$, $\omega_\xi \in [0, 2\pi]$, $\Phi_u \equiv 1$, and we estimate each transfer function with $n_1$ parameters. We only give the expressions for two modules to ease notation, but, the generalization to more sensors is straightforward. We assume that we can take $\{\mathcal{A}_k^l\} = \{z^{-k}\}_{k=l}^{n_1+(l-1)p}$ as a basis for $l$-th column of $\Psi$. This assumption is a rank condition, based on that the $l$-th column will consist of the functions $\{z^{-k}\}_{k=l}^{n_1+(l-1)p}$. Thus, we exclude degenerate cases, e.g., when $G_1 = G_2$. In the case we consider, $b_l^{-2} = n_1 + (l-1)p - (l-1)$ and (4.9) becomes

$$\text{AsVar } \hat{G}_1(e^{j\omega_\xi}) = \frac{n_1}{\lambda_1^{-1} + \frac{n_1}{n_1+p-1}\lambda_2^{-1}|G_2(e^{j\omega_\xi})|^2}. \tag{4.16}$$

We would expect that when we estimate both $G_1$ and $G_2$ with many parameters, the benefit of the second sensor would diminish. However, this is not the case when $G_1$ has a zero on the unit circle, as then, using (4.9) in Theorem 4.2.1,

$$\lim_{n_1\to\infty} \lim_{N\to\infty} \frac{N}{n} \text{Var } \hat{G}_1(e^{j\omega_\xi}) = \frac{1}{\frac{1}{\lambda_1} + \frac{1}{\lambda_2}|G_2(e^{j\omega_\xi})|^2}.$$

This is the same asymptotic variance as if the module $G_2$ was completely known (the noise signals are uncorrelated, so the optimal variance is the inverse of the sum of the information in each signal). To us, this is a surprising result. However, it fits with the discussion on optimal input design of structured systems, where the input signal should be designed to hide unimportant system properties (see e.g., Hjalmarsson (2009)). Here, since $G_1(e^{j\omega_\xi}) = 0$, the output signal $y_2$ does not contain any information about $G_2(e^{j\omega_\xi})$ in the limit when the number of estimated parameters tends to infinity. However, what we measure in $y_2$ is 0 + noise, but the zero is the quantity$G_1(e^j\omega_\xi)$ we try to identify, and hence $y_2$ can be used entirely to estimate this quantity.

In the following example and Example 4.4.2, (4.16) and its generalization are verified by MATLAB simulations.

**Example 4.4.1.** *We consider a setup according to* (4.1)*, with two FIR transfer functions:* $G_1 = z^{-1} + z^{-2}$ *and* $G_2 = z^{-1}$*, where* $\lambda_1 = \lambda_2 = 1$*.* $G_1$ *thus has a zero at* $z = e^{j\pi}$*. The input signal is white noise with variance* $\sigma_u^2 = 1$*. The variance at* $\omega_\xi = \pi$ *is estimated in 3 ways: from* (4.9)*, directly from* (2.12) *and as the sample variance from 1,000 Monte-Carlo simulations. When estimating the given system with FIR models of order* $n_1$*,* (4.9) *reduces to*

$$\text{AsVar } \hat{G}_1(e^{j\omega_\xi}) = \frac{1}{\frac{1}{n_1} + \frac{1}{n_1+1}}.$$

Table 4.1: Comparison of the asymptotic variance of $\hat{G}_1(e^{j\pi})$ with Monte-Carlo simulations.

| $n_1^{-1}\text{AsVar}\,\hat{G}_1(e^{j\pi})$ | (4.8) | Directly from (2.12) | Monte-Carlo |
|---|---|---|---|
| $n_1 = 2$ | 0.6000 | 0.6000 | 0.5913 |
| $n_1 = 10$ | 0.5238 | 0.5238 | 0.5413 |

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

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

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

*The Monte-Carlo simulations come close to what is predicted by* (4.9) *(cf. Table 4.1). The reduction in variance, compared to only using $y_1$, is centered around the frequency of the zero (see Figures 4.3 and 4.4). Note that if we only use $y_1$ then $(N/n_1)\text{Var}\hat{G}_1(e^{j\omega_\xi}) = 1$ for all frequencies. The intuition is that the input to the second system is zero at the frequency $e^{j\omega_\xi}$ and therefore, with a slight abuse of notation, $y_2(e^{j\omega_\xi})$ is a signal uncorrelated with the input $u(e^{j\omega_\xi})$. Thus, $y_2(e^{j\omega_\xi})$ tells us that either $G_1(e^{j\omega_\xi}) = 0$, or $G_2(e^{j\omega_\xi}) = 0$ (or both). Furthermore, the transfer functions in our model set are all continuous, which implies that if $G_2$ is large around $e^{j\omega_\xi}$, it is unlikely to drop to zero at $e^{j\omega_\xi}$. Thereby, $y_2$ gives us information that $G_1(e^{j\omega_\xi})$ is small. A related aspect, not presented in the figures, is that the estimate of $G_2(e^{j\omega_\xi})$ is poor because of the poor signal-to-noise ratio, as expected. When the number of estimated parameters increases, the reduction in variance becomes more focused around $\omega_\xi$. This is due to the increased order of the basis functions (cf. Figures 4.3 and 4.4).*

**Example 4.4.2.** *Consider the case when two additional modules are appended to the setup of Example 4.4.1, still satisfying* (4.1)*, with $G_3 = z^{-2}$, $G_4 = (z^{-1} - z^{-2})/\sqrt{2}$ and $\lambda_3 = \lambda_4 = 1$. Simulations are performed in the same way as in Example 4.4.1. Again, the Monte-Carlo simulations come close to what is predicted by* (4.9) *(cf. Table 4.2). Notice the similarities between the sets of Figures 4.3 and 4.4 compared to Figures 4.5 and 4.6; we get a variance reduction that is focused around the frequency of the unit circle zero, but, when we use more sensor the reduction is larger, and for both sets, as more parameters are estimated, the variance reduction becomes more focused.*

The fit is significantly better in Figure 4.3 compared to Figures 4.4–4.6 because the ratio between number of samples and estimated parameters is larger. The difference between the Monte-Carlo simulations and the theory is smooth because only a relatively small number of parameters of $\hat{G}_1$ is estimated compared to the number of frequencies.

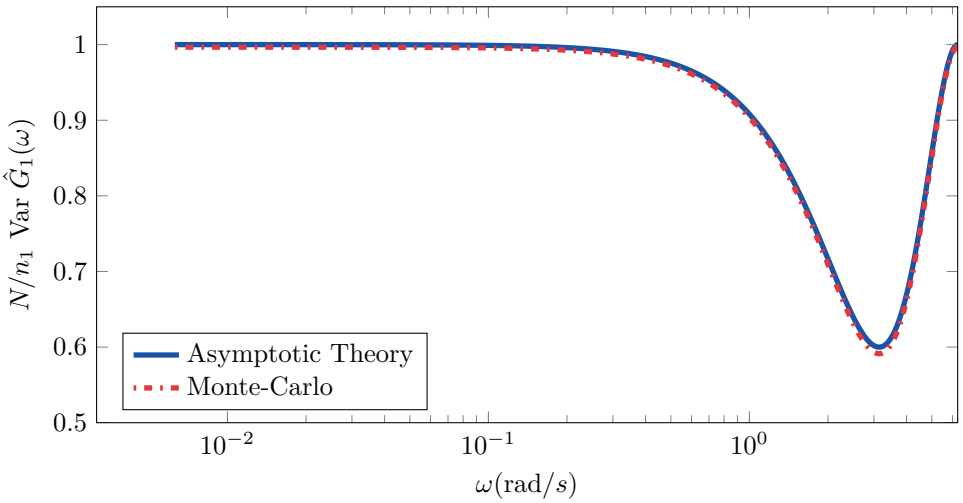


Figure 4.3: Asymptotic variance of $\hat{G}_1$ in Example 4.4.1, weighted by number of estimated parameters of $\hat{G}_1$ ($n_1 = 2$), and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The variance is reduced around the frequency where $G_1$ is zero, when the output $y_2$ from $G_2$ is also used.

Table 4.2: Comparison of the asymptotic variance of $\hat{G}_1(e^{j\pi})$ with Monte-Carlo simulations. The outputs $y_1$-$y_4$ are used here.

| $n_1^{-1}$AsVar $\hat{G}_1(e^{j\pi})$ | (4.8) | Directly from (2.12) | Monte-Carlo |
|---|---|---|---|
| $n_1 = 2$ | 0.3371 | 0.3309 | 0.3305 |
| $n_1 = 10$ | 0.2336 | 0.2314 | 0.2561 |

## 4.5 Conclusions

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

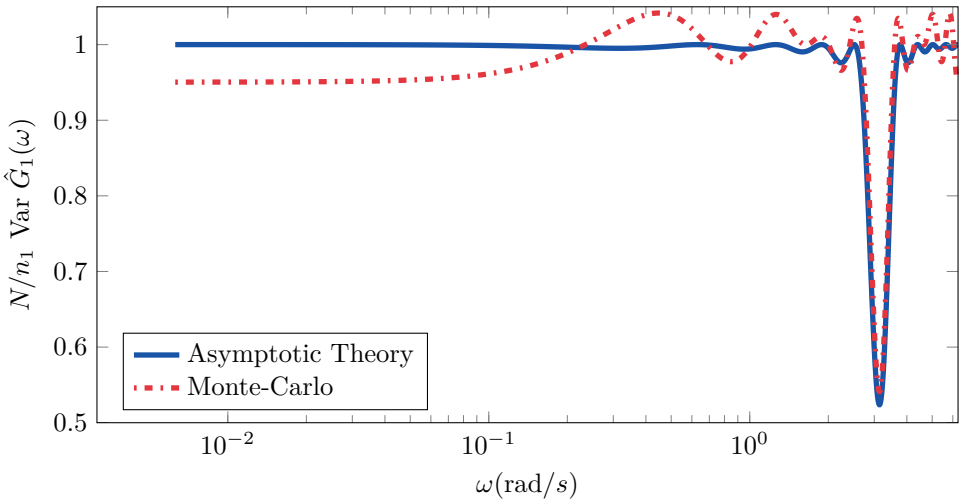


Figure 4.4: Asymptotic variance of $\hat{G}_1$ in Example 4.4.1, weighted by number of estimated parameters of $\hat{G}_1$ ($n_1 = 10$), and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The reduction in variance is more focused around the frequency where $G_1$ is zero compared to Figure 4.3.

## 4.A Preliminary lemmas

**Lemma 4.A.1.** *Let $\mathcal{X}$ be a finite-dimensional subspace of $\mathcal{L}_2^m$, $\{\mathcal{A}_k\}_{k=1}^{\infty}$ be an orthonormal basis for $\mathcal{L}_2^m$ and $\{\mathcal{A}_k\}_{k=1}^{r}$ a basis for $\mathcal{X}$. If the function $f : \mathbb{C} \to \mathbb{C}^m$ is given by*

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

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

$$\langle G, f \rangle = G(\xi). \tag{4.17}$$

*Proof.* We can express $G$ as

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

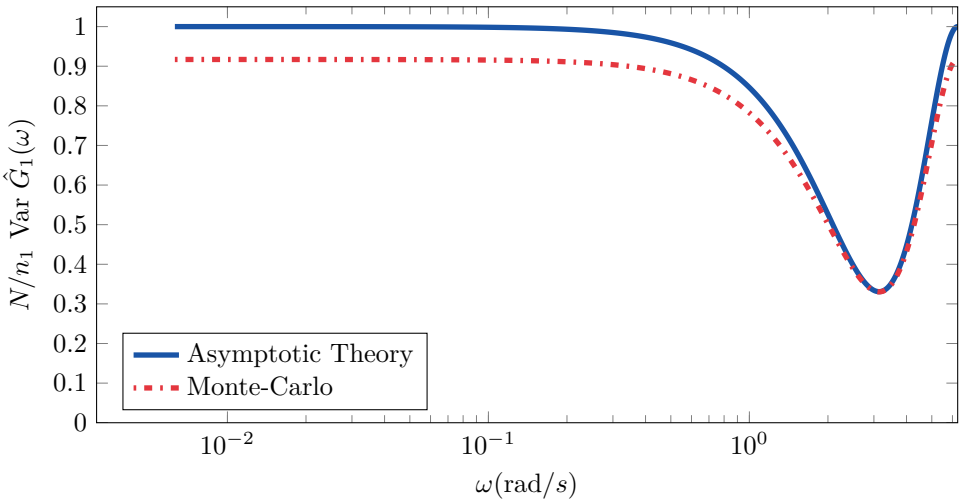


Figure 4.5: Asymptotic variance of $\hat{G}_1$ in Example 4.4.2, weighted by number of estimated parameters of $\hat{G}_1$ ($n_1 = 2$), and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The variance is reduced more than in Figure 4.3, when also $y_3$ and $y_4$ are used.

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

$$\begin{aligned}\langle G, f\rangle &= \langle \sum_{j=1}^{r} G_j \mathcal{A}_j, \sum_{k=1}^{r} \overline{\mathcal{A}_k(\xi)} \mathcal{A}_k \rangle \\ &= \sum_{k=1}^{r} G_k \mathcal{A}_k(\xi) = G(\xi).\end{aligned}$$

□

**Remark 4.A.1.** *The function $f$ is strongly related to the so-called "reproducing kernel" for the space $\mathcal{X}$ (see Ninness and Hjalmarsson (2004) and the references therein).*

**Corollary 4.A.1.** *Let $G \in \mathcal{X}$ and $f$ be as in Lemma 4.17. Then*

$$\langle G, f\rangle = 0 \tag{4.18}$$

*if and only if*

$$G(\xi) = 0.$$

*Proof.* Follows directly from Lemma 4.A.1. □

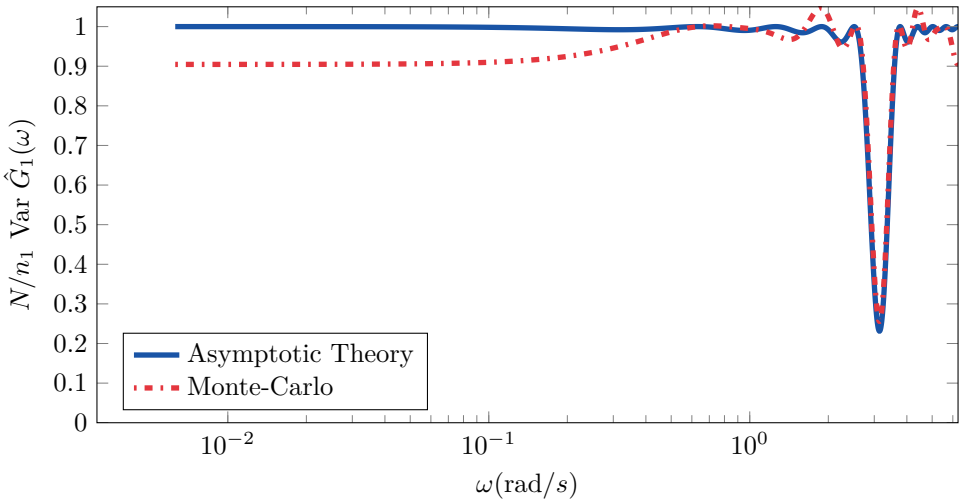


Figure 4.6: Asymptotic variance of $\hat{G}_1$ in Example 4.4.2, weighted by number of estimated parameters of $\hat{G}_1$ ($n_1 = 10$), and number of data points $N = 10^4$, at frequencies $[0, 2\pi]$. The variance is reduced more than in Figure 4.4, when also $y_3$ and $y_4$ are used.

## 4.B Proof of Lemma 4.2.1

For ease of notation we will assume that $\Phi_u(e^{j\omega}) = 1$ for $\omega \in [0, 2\pi]$ and $\lambda_1 = \cdots = \lambda_m = 1$; it is straightforward to adjust the derivations for the general setting. It is not obvious that $\mathcal{B}_1 = \begin{bmatrix} c_k^1 \mathcal{B}_k^1 & \dots & c_k^m \mathcal{B}_k^m \end{bmatrix}$, as defined by (4.3), lies in the space spanned by the rows of $\Psi$; however, this will be clear from how we construct $\mathcal{B}_1$. To construct $\mathcal{B}_1$, we project $\Psi$ onto the space $\mathcal{X}$, defined as the span of the $m$ functions $\{\alpha_i\}$ that have $\mathcal{B}_1^i$ in the $i$th column and zeros in the other columns, e.g., the first one being $\alpha_1 = \begin{bmatrix} \mathcal{B}_1^1 & 0 & \cdots \end{bmatrix}$. Here, $\mathcal{B}_1^k$ is defined as in (4.6b). The reason for this construction is that we can write (4.6) as

$$\langle \alpha_i, \mathcal{B}_j \rangle = 0, \qquad \text{for } i = 1, \dots, m \qquad \text{for } j = 2, \dots, m, \tag{4.19}$$

since $\langle \alpha_i, \mathcal{B}_j \rangle = 0$ if and only if $\langle \mathcal{B}_1^i, \mathcal{B}_j^i \rangle = 0$. Furthermore, the basis functions for $\Psi$, $\{\mathcal{B}_k\}_{k=1}^l$, satisfy (4.19) if and only if they are orthogonal to $\mathcal{X}$ for $k \geq 2$. Applying the projection of $\Psi$ onto the space $\mathcal{X}$ according to (2.24), we obtain

$$\mathbf{P}_{\mathcal{X}} \Psi = \begin{bmatrix} \langle G_1', \mathcal{B}_1^1 \rangle \mathcal{B}_1^1 & \langle G_2 G_1', \mathcal{B}_1^2 \rangle \mathcal{B}_1^2 & \cdots & \langle G_m \cdots G_1', \mathcal{B}_1^m \rangle \mathcal{B}_1^m \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}, \tag{4.20}$$

where all rows are zero except the first because of Corollary 4.A.1. Furthermore, we have used the fact that $\mathcal{B}_1^k, k = 1, \ldots, l$ have unit norm. Applying Lemma 4.A.1 to (4.20) we obtain

$$\mathbf{P}_{\mathcal{X}}\Psi = \begin{bmatrix} G_1'(\xi) \\ 0 \\ \vdots \end{bmatrix} \begin{bmatrix} b_1\mathcal{B}_1^1 & b_2G_2(\xi)\mathcal{B}_1^2 & \cdots & b_mG_m(\xi)\cdots G_2(\xi)\mathcal{B}_1^m \end{bmatrix}.$$

Normalize the vector $\begin{bmatrix} b_1\mathcal{B}_1^1 & b_2G_2(\xi)\mathcal{B}_1^2 & \cdots & b_mG_m(\xi)\cdots G_2(\xi)\mathcal{B}_1^m \end{bmatrix}$ to form $\mathcal{B}_1$. Note that $\mathbf{P}_{\mathcal{X}}\Psi = \begin{bmatrix} C & 0 \end{bmatrix}^{\mathrm{T}} \mathcal{B}_1 = \mathbf{P}_{\mathcal{B}_1}\Psi$ for some constant vector $C$, i.e., rank $\{\mathbf{P}_{\mathcal{X}}\Psi\} = 1$ and $\mathcal{B}_1$ is a basis vector for the space spanned by the rows of $\mathbf{P}_{\mathcal{B}_1}\Psi$. Consequently, the part of $\Psi$ that is orthogonal to $\mathcal{B}_1$ is also orthogonal to $\mathcal{X}$,

$$\Psi - \mathbf{P}_{\mathcal{B}_1}\Psi = \mathbf{P}_{\mathcal{X}^\perp}\Psi \perp \mathcal{X},$$

where $\mathcal{X}^\perp$ is the orthogonal complement of $\mathcal{X}$ in $\mathcal{L}_2$. However, for $2 \leq k \leq l$, $\mathcal{B}_k$ lies in the space spanned by the rows of $\Psi - \mathbf{P}_{\mathcal{B}_1}\Psi$, since this corresponds to he part of $\Psi$ that is orthogonal to $\mathcal{B}_1$ ($\{\mathcal{B}_k\}_{k=1}^l$ is an orthonormal basis). As noted above, (4.6) is equivalent to (4.19), which only holds if $\mathcal{B}_k$, for $2 \leq k \leq l$, is orthogonal to $\mathcal{X}$. But, we know that $\mathcal{B}_k$, for $2 \leq k \leq l$, lies in $\Psi - \mathbf{P}_{\mathcal{B}_1}\Psi$ which we showed was orthogonal to $\mathcal{X}$ and the theorem follows.

# Chapter 5

# Generalized parallel cascade models

In this chapter we consider two forms of parallel serial structures, one multiple-input-multiple-output structure and one single-input-multiple-output structure. The first structure is an abstraction of the structure found in the motivating example on steam pressure control. This structure has also been considered in Hägg *et al.* (2011), whose key results concern the case when there is common dynamics among the modules. The second structure can be an example of a network of spatially distributed sensors where the sensor dynamics are not completely known and therefore need to be estimated if these sensors are to be used. The structure considered corresponds to a star topology with the module of interest in the center (Yang *et al.*, 2007). In the previous chapter, the model order of all the modules was the same. In this chapter, the model order of the system of interest will be fixed, while the model order of every other module will be allowed to increase. By letting the model order grow large we provide upper bounds on the variance of the parameter estimates. A lower bound can in turn be found by assuming the other modules to be known.

## 5.1 Problem formulation

We consider two types of networks of linear dynamic systems depicted in Figure 5.1 and Figure 5.2. For both networks, we wish to identify a model for module $G$. The first network is the parallel serial (cascade) structure considered in Hägg *et al.* (2011), see Figure 5.1. For this network, the dynamics are given as
**Structure 1**:

$$y(t) = \sum_{k=1}^{m} G(q)G_k(q)u_k(t) + e(t), \tag{5.1a}$$

$$y_k(t) = G_k(q)u_k(t) + e_k(t), \quad k = 1, \dots, m, \tag{5.1b}$$

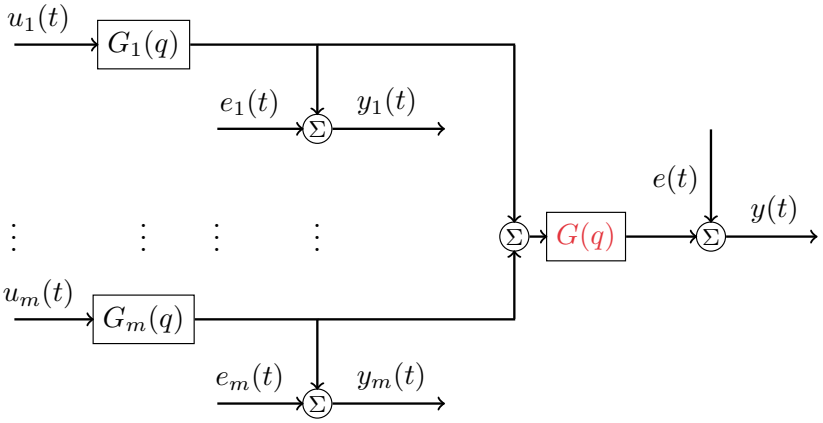


Figure 5.1: Structure 1: Parallel serial structure.

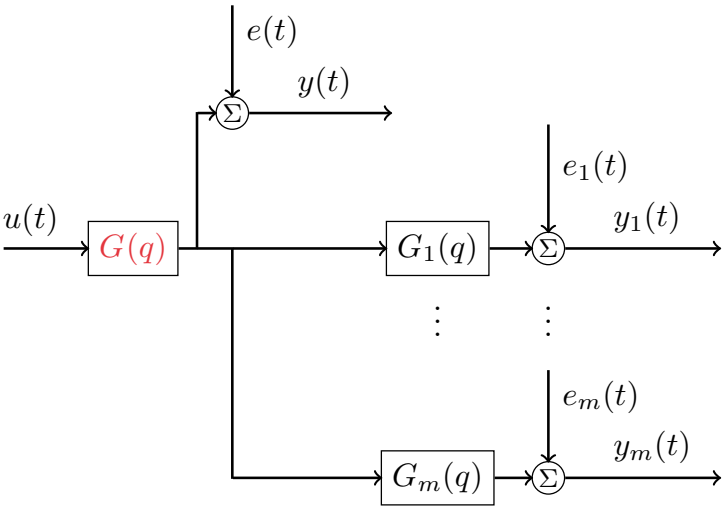


Figure 5.2: Structure 2: Multi-sensor structure.

where each input signal $u_k(t)$ is feed to the actuator module $G_k(q)$, the output of which is measured in $y_k(t)$ with additive noise $e_k(t)$. The sum of the outputs of all actuators is fed to the module $G(q)$, which is the one we are interested in. The output of $G(q)$ is measured as $y(t)$ with additive noise $e(t)$. For the second network structure, depicted in Figure 5.2, the dynamics are given as
**Structure 2**:

$$y_k(t) = G_k(q)G(q)u(t) + e_k(t), \quad k = 1, \dots, m, \tag{5.2a}$$
$$y(t) = G(q)u(t) + e(t), \tag{5.2b}$$

where there is only one input $u(t)$, which is fed to the module of interest $G(q)$ and measured as $y(t)$ with additive noise $e(t)$. The output of $G(q)$ is also fed to additional sensor modules $G_k(q)$ whose outputs are measured with additive noise. We assume that the additive zero mean stationary sequences $\{e_i(t)\}$ are mutually independent, and independent of the input signals $u_k(t), k = 1, \dots, m$ (input signal $u(t)$), with spectra $\Phi_{e_i}(\omega), i = 1, \dots, m$, the same holds for $e(t)$ with spectrum $\Phi_e(\omega)$. We will in the derivations, for improved readability, assume that the noise sequences are white noise with with variances $\lambda_i, i = 1, \dots, m$, the same holds for $e(t)$ with variance $\lambda$. The derivations for the general case follows analogously to the white noise case. The input is assumed to be a realization of a weakly stationary stochastic process with spectrum $\Phi_u(\omega)$. The inputs $u_k(t), k = 1, \dots, m$ (input $u(t)$) are assumed to be a realization of a weakly stationary stochastic process with spectra $\Phi_{u_k}(\omega), k = 1, \dots, m$ (spectrum $\Phi_u(\omega)$). The models of the modules are independently parametrized, i.e., $G_k(q) = G_k(q, \theta_k)$, $k = 1, \dots, m$ and $G(q) = G(q, \eta)$. We stack the parameters together in $\theta = \left[\theta_1^{\mathrm{T}}, \dots, \theta_m^{\mathrm{T}}, \eta^{\mathrm{T}}\right]^{\mathrm{T}}$, where $\theta_i \in \mathbb{R}^{d_i}$, $d_i \in \mathbb{Z}_{\geq 0}$ for all $i = 1, \dots, m+1$, $\eta \in \mathbb{R}^d$.

## 5.2 Structure 1: parallel serial structure

In this section, we study the parallel cascade structure described by (5.1),and visualized in Figure 5.1. The main idea is to express the covariance of $\eta$, the parameters of the module of interest, as a sum of two parts, where the second part is a projection onto a space whose size depends on the model structure of the other modules. From this expression it will be possible to determine upper and lower bounds on the covariance of $\eta$. If we assume that the other modules are known exactly the projection becomes zero and we get a lower bound on the covariance. On the other hand, if we assume that the space is large, the projection can be removed and we get an upper bound.

The one step ahead predictor of $y(t)$ based on $\theta$ is given by

$$\hat{y}(t) = T(\theta)^{\mathrm{T}} u(t),$$

where

$$T = \begin{bmatrix} G_1(q,\theta_1) & 0 & 0 & G(q,\eta)G_1(q,\theta_1) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & G_m(q,\theta_m) & G(q,\eta)G_m(q,\theta_m) \end{bmatrix}.$$

The predictor error gradient is then given by

$$\psi = -\begin{bmatrix} G_1'(q,\theta_1)u_1(t) & 0 & 0 & G_1'(q,\theta_1)G(q,\eta)u_1(t) \\ 0 & \ddots & 0 & \vdots \\ 0 & 0 & G_m'(q,\theta_m)u_m(t) & G_m'(q,\theta_m)G(q,\eta)u_m(t) \\ 0 & 0 & 0 & G'(q,\eta)\left(\sum_{k=1}^{m} G_k(q,\theta_k)u_k(t)\right) \end{bmatrix}.$$

Utilizing the assumption that the input signals are independent, (2.12) and Parseval's relation, we can express the asymptotic covariance of the parameter estimate as

$$\operatorname{AsCov}\hat{\theta} = \langle \Gamma A, \Gamma \rangle^{-1}, \tag{5.3}$$

where

$$\Gamma = \operatorname{diag}\left\{G_1'(q,\theta_1^o), \ldots, G_1'(q,\theta_m^o), G'(q,\eta^o)\right\},$$

and

$$A = \begin{bmatrix} D & B^* \\ B & S \end{bmatrix},$$

with $D \in \mathcal{L}^{m\times m}$ with elements

$$D_{ij} = \begin{cases} \Phi_{u_i}(\lambda_i^{-1} + |G|^2\lambda^{-1}), & i = j, \\ 0, & \text{otherwise}, \end{cases} \tag{5.4}$$

$B \in \mathcal{L}^{1\times m}$ has elements

$$B_i = \Phi_{u_i}\overline{G_i}G\lambda^{-1}, \quad i = 1, \ldots, m,$$

and $S$ is the scalar function

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

We now have an expression for the asymptotic covariance for $\hat{\eta}$ in the lower right corner of (5.3). As our next step, we will apply Lemma 5.A.1 to reformulate the covariance of $\hat{\eta}$ in (5.3) as the sum of two parts where the second part is a projection onto a space that depends on the model structure of the other modules. Then we will use Lemma 2.3.3 to derive lower and upper bounds. We are interested in $G$,

so we need make the split to have all quantities related to $G_1, \ldots, G_m$ in the first part and the ones related to $G$ in the second part. Thus, we make the split

$$\Gamma_1 = \operatorname{diag}\{G_1', \ldots, G_m'\}, \quad \Gamma_2 = G',$$

and

$$\Psi_2\Psi_1^* = G'B\Gamma_1^*, \quad \Psi_1\Psi_1^* = \Gamma_1 D\Gamma_1^*, \quad \Psi_2\Psi_2^* = G'SG'^*.$$

We assume that $\operatorname{rank}\{D\} = m$, which holds, e.g., when $\Phi_{u_1}(\omega), \ldots, \Phi_{u_m}(\omega) > 0$ and furthermore, we assume that also $\Gamma_1 D\Gamma_1^*$ has full rank. From Lemma 5.A.1 we have that the asymptotic covariance matrix of $\hat{\eta}$ is

$$\operatorname{AsCov}\hat{\eta} = \left[\langle G'S, G'\rangle - \left\langle \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\}, \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\}\right\rangle\right]^{-1} \tag{5.5}$$

with $\gamma = G'B\Gamma_1^*[R_1^\dagger]^*$, where $R_1$ is a spectral factor of $\Gamma_1 D\Gamma_1^*$, i.e.,

$$\Gamma_1(e^{j\omega}, \theta^o)D(e^{j\omega}, \theta^o)\Gamma_1^*(e^{j\omega}, \theta^o) = R_1(e^{j\omega})R_1^{\mathrm{T}}(e^{-j\omega}),$$

such that the function $R_1(z)$ is analytic in the unit disc and has full rank on the unit circle. From (5.5) it is still not clear how different quantities affect AsCov $\hat{\eta}$. To allow for an interpretation of (5.5), we will consider two extreme cases of the model order of the modules $\{G_k\}_{k=1}^m$. We first consider the modules $\{G_k\}_{k=1}^m$ to be known. Then no basis function of $\{G_k\}_{k=1}^m$ have to be estimated and $\left\langle \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\}, \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\}\right\rangle = 0$, since the projection is made onto the empty set $\mathcal{S}_{R_1}$. This corresponds to a lower bound on AsCov $\hat{\eta}$ since the projection is made onto a smaller subspace, cf. Lemma 2.3.3. The bound is given by

$$\operatorname{AsCov}\hat{\eta} \geq \left[\left\langle G', \sum_{k=1}^m M_{u_k}^L, G'\right\rangle\right]^{-1}, \tag{5.6}$$

where

$$M_{u_k}^L = \frac{\Phi_{u_k}|G_k|^2}{\lambda}, \quad k = 1, \ldots, m.$$

The second case is when the model order grows large for the models associated with the modules $\{G_k\}_{k=1}^m$. Formally, if the rows of $G_k'$ span $\mathcal{L}_2$ for all $k$, we have that $\mathcal{S}_{G_k'} = \mathcal{L}_2$ for all $k$. This implies that $\mathcal{S}_{\Gamma_1} = \mathcal{L}_2^m$. This is a conservative upper bound since $\mathcal{S}_{G_k'} \subseteq \mathcal{H}_2 \subset \mathcal{L}_2$. In this case, the projection in Lemma 5.A.2 can be removed (since the projection is made onto $\mathcal{L}_2^m$ and $\tilde{\gamma} \in \mathcal{L}_2^m$), which gives that

$$\operatorname{AsCov}\hat{\eta} < \left[\langle G'(S - \tilde{\gamma}\tilde{\gamma}^*), G'\rangle\right]^{-1},$$

where

$$\tilde{\gamma}\tilde{\gamma}^* = BD^{-1}B^* = \sum_{k=1}^m \frac{\Phi_{u_k}|G_kG|^2\lambda^{-2}}{\lambda_k^{-1} + |G|^2\lambda^{-1}}. \tag{5.7}$$

Simplifying the above expression leads to

$$\text{AsCov}\,\hat{\eta} < \left[\left\langle G' \sum_{k=1}^{m} M_{u_k}^U, G' \right\rangle\right]^{-1}, \tag{5.8}$$

where,

$$M_{u_k}^U = \frac{\Phi_{u_k}|G_k|^2}{\lambda + |G|^2\lambda_k}, \quad k = 1, \ldots, m. \tag{5.9}$$

We summarize the results in the following theorem for non white noise spectra.

**Theorem 5.2.1.** *Let the system dynamics be described* (5.1)*. We assume that the additive zero mean stationary sequences* $\{e_i(t)\}$ *are mutually independent, and independent of the input signals* $u_k(t), k = 1, \ldots, m$*, with spectra* $\Phi_{e_i}(\omega), i = 1, \ldots, m$*; the same holds for* $e(t)$ *with spectrum* $\Phi_e(\omega)$*. The inputs* $u_k(t), k = 1, \ldots, m$ *are assumed to be a realization of a weakly stationary stochastic process with spectra* $\Phi_{u_k}(\omega), k = 1, \ldots, m$*. The models of the modules are independently parametrized, i.e.,* $G_k(q) = G_k(q, \theta_k)$*,* $k = 1, \ldots, m$ *and* $G(q) = G(q, \eta)$*. We assume that* $D$ *in* (5.4) *is full rank (*$\text{rank}\,\{D\} = m$*). Then*

$$\left[\left\langle G' \sum_{k=1}^{m} M_{u_k}^L, G' \right\rangle\right]^{-1} \leq \text{AsCov}\,\hat{\eta} < \left[\left\langle G' \sum_{k=1}^{m} M_{u_k}^U, G' \right\rangle\right]^{-1}, \tag{5.10}$$

*where*

$$M_{u_k}^L = \frac{\Phi_{u_k}|G_k|^2}{\Phi_e}, \quad k = 1, \ldots, m, \tag{5.11a}$$

$$M_{u_k}^U = \frac{\Phi_{u_k}|G_k|^2}{\Phi_e + |G|^2\Phi_{e_k}}, \quad k = 1, \ldots, m. \tag{5.11b}$$

**Remark 5.2.1.** *Comparing* (5.11a) *and* (5.11b) *in* (5.10) *with open loop SISO identification, we see that* $\sum_{k=1}^{m} M_{u_k}$ *plays the same role as the signal to noise ratio. For the corresponding open loop SISO case, we have*

$$\text{AsCov}\,\hat{\eta} = \left[\left\langle G' \frac{\Phi_u}{\Phi_e}, G' \right\rangle\right]^{-1},$$

*The lower bound directly corresponds to this case with* $\Phi_u = \sum_{k=1}^{m} \Phi_{u_k}|G_k|^2$*. When we know the dynamics of the actuators, the measurements of* $\{y_k(t)\}_{k=1}^m$ *are redundant. On the other hand, when the actuator dynamics* $\{G_k(q)\}_{k=1}^m$ *are estimated with high order models, the spectrum of the noise also plays a role. In* (5.9)*, the power spectrum of the input signal fed to the system of interest* $G(q)$*, is the same as in* (5.7)*. However, the denominator is given by the spectrum of the noise affecting* $y(t)$*,* $\Phi_e$*, and by* $\Phi_{e_k}$*, the spectrum of the noise affecting the input measurement* $y_k$*, weighted by the module of interest* $|G|^2$*. Since* $\Phi_{e_k}$*, the noise spectrum at the input*

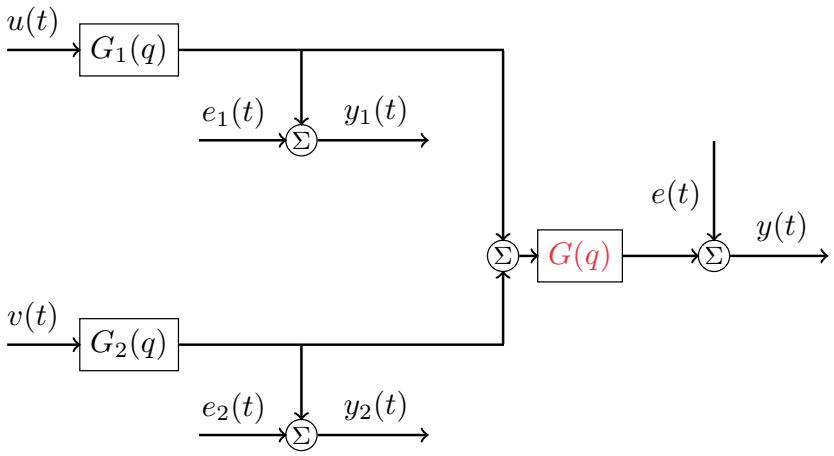


Figure 5.3: Network considered in Example 5.2.1.

*measurements, is scaled by the system of interest $G$, high noise variance in the input measurements hurts us the most at frequencies where the gain of the system of interest is high.*

The next example illustrates the benefit of knowing a disturbance.

**Example 5.2.1.** *We consider a system with one input $u(t)$ and a measurable disturbance $v(t)$ with spectrum $\Phi_v(\omega)$. The noise $e(t)$ affecting the output of the system of interest $y(t)$ is white noise with variance $\lambda$. The system is shown in Figure 5.3. We will compare the case when the disturbance $v(t)$ is not measured with the case when it is measured and used in the predictor ($v(t)$ known). We assume that we estimate many parameters in $G_1$ and $G_2$, and thus we can consider the upper bound given by* (5.11b) *in* (5.10) *In the first case, the noise affecting $y(t)$ is given by $e(t) + G(q)G_2(q)v(t)$, which has spectrum $\lambda + |G(q)G_2(q)|^2\Phi_v(\omega)$. Hence, the upper bound in* (5.11b) *is given by*

$$\text{AsCov}\,\hat{\eta} \leq \left[\left\langle G' \frac{\Phi_u |G_1|^2}{\lambda + |GG_2|^2\Phi_v + |G|^2\lambda_1}, G' \right\rangle\right]^{-1}, \tag{5.12}$$

*when the disturbance $v(t)$ is not measured. The disturbance thus increases the variance of $\hat{\eta}$, since it increases the denominator in* (5.12)*. If we measure the disturbance, the upper bound on the covariance is instead given by* (5.11b) *as*

$$\text{AsCov}\,\hat{\eta} \leq \left[\left\langle G' \left( \frac{\Phi_u |G_1|^2}{\lambda + |G|^2\lambda_1} + \frac{\Phi_v |G_2|^2}{\lambda + |G|^2\lambda_2} \right), G' \right\rangle\right]^{-1}, \tag{5.13}$$

*The variance is decreased in* (5.13) *compared with* (5.12) *for two reasons: (i) we reduce the noise affecting $y(t)$, and (ii) we provide a new excitation. From the*

*reduced noise (i), the first term inside the parenthesis in* (5.13) *is larger than the corresponding term in* (5.12)*. The second term inside the parenthesis in* (5.13) *corresponds to the new excitation (ii).*

## 5.3 Structure 2: multi-sensor structure

In this section we study the multi-sensor structure described by (5.2),and visualized in Figure 5.2. The main idea is to express the covariance of $\eta$, the parameters of the module of interest, as a sum of two parts, where the second part is a projection onto a space whose size depends on the model structure of the other modules. From this expression it will be possible to determine upper and lower bounds on the covariance of $\eta$. If we assume that the other modules are known exactly the projection becomes zero and we get a lower bound on the covariance. If, on the other hand, we assume the space is large the projection can be removed and we get an upper bound. The minimum variance one step ahead predictor of $y(t)$ based on $\theta$ is given by

$$\hat{y}(t) = T(\theta)^{\mathrm{T}} u(t), \tag{5.14}$$

where

$$T = \begin{bmatrix} G_1(q,\theta_1)G(q,\eta) & \dots & G_m(q,\theta_m)G(q,\eta) & G(q,\eta) \end{bmatrix}.$$

The predictor error gradient is then given by

$$\psi(t) = - \begin{bmatrix} G_1'(q,\theta_1)G(q,\eta) & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & G_m'(q,\theta_m)G(q,\eta) & 0 \\ G'(q,\eta)G_1(q,\theta_1) & \dots & G'(q,\eta)G_m(q,\theta_1) & G'(q,\eta) \end{bmatrix} u(t).$$

Using (2.12) and Parseval's relation, we can express the asymptotic covariance of the parameter estimate as

$$\operatorname{AsCov} \hat{\theta} = \langle \Gamma A, \Gamma \rangle^{-1}, \tag{5.15}$$

$$\Gamma = \operatorname{diag} \{ G_1'(q,\theta_1^o), \dots, G_1'(q,\theta_m^o), G'(q,\eta^o) \},$$

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

$$A = \begin{bmatrix} D & B^* \\ B & S \end{bmatrix},$$

with $D \in \mathcal{L}^{m\times m}$ with elements

$$D = \operatorname{diag} \left\{ |G|^2\lambda_1^{-1}, |G|^2\lambda_2^{-1}, \dots, |G|^2\lambda_m^{-1} \right\}. \tag{5.16}$$

$B \in \mathcal{L}^{1\times m+1}$ has elements

$$B_i = \Phi_u \overline{G} G_i \lambda_i^{-1}$$

and $S$ is the scalar function

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

We now have an expression for the asymptotic covariance for $\hat{\eta}$ in the lower right corner of (5.15). As our next step, we will apply Lemma 5.A.1 to reformulate the covariance of $\hat{\eta}$ in (5.15) as the sum of two parts where the second part is a projection onto a space that depends on the model structure of the other modules. Then we will use Lemma 2.3.3 to derive lower and upper bounds. We are interested in $G$, so we need make the split to have all quantities related to $G_1, \ldots, G_m$ in the first part and the ones related to $G$ in the second part. Thus, we make the split

$$\Psi_2 \Psi_1^* = G' B \Gamma_1^*, \quad \Psi_1 \Psi_1^* = \Gamma_1 D \Gamma_1^*, \quad \Psi_2 \Psi_2^* = G' S G'^*,$$

and

$$\Gamma_1 = \operatorname{diag}\{G_1', \ldots, G_m'\}, \quad \Gamma_2 = G'.$$

We assume that $\tilde{\Psi}_1 \tilde{\Psi}_1^*$ has full rank (rank $\left\{\tilde{\Psi}_1 \tilde{\Psi}_1^*\right\} = m$), which essentially is an identifiability condition. This holds, e.g., , when

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

From Lemma 5.A.1 we have that the asymptotic covariance matrix for $\hat{\eta}$ is given by (5.5) with $\gamma = G' T \Gamma_1^* [R_1^\dagger]^*$, where $R_1$ is a minimum phase spectral factor of $\Gamma_1 D \Gamma_1^*$, i.e., $\Gamma_1(e^{j\omega}, \theta^o) D(e^{j\omega}, \theta^o) \Gamma_1^*(e^{j\omega}, \theta^o) = R_1(e^{j\omega}) R_1^{\mathrm{T}}(e^{-j\omega})$ such that the function $R_1(z)$ is analytic in the unit disc and has full rank on the unit circle. As in Section 5.2, we consider two extreme cases of the model order of the estimated models of the modules $\{G_k(q)\}_{k=1}^m$ to give an interpretation of AsCov $\hat{\eta}$. The first is to consider the modules $\{G_k(q)\}_{k=1}^m$ to be known, so no basis functions of $\{G_k(q)\}_{k=1}^m$ have to be estimated and $\left\langle \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\}, \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\} \right\rangle = 0$ (the space $\mathcal{S}_{R_1}$ is empty and hence the projection onto this space is zero). This case corresponds to a lower bound on AsCov $\hat{\eta}$. The bound is given by

$$\text{AsCov } \hat{\eta} \geq \left[ \left\langle G' \left( \frac{\Phi_u}{\lambda} + \sum_{k=1}^{m} M_{u_k}^L \right), G' \right\rangle \right]^{-1}, \tag{5.17}$$

where,

$$M_{u_k}^L = \frac{\Phi_u |G_k|^2}{\lambda_k}, \quad k = 1, \ldots, m.$$

The second case is when the model order grows large. Formally, we assume the rows of $G'_k$ span $\mathcal{L}_2$ for all $k$, i.e., we assume that $\mathcal{S}_{G'_k} = \mathcal{L}_2$ for all $k$, so then $\mathcal{S}_{\Gamma_1} \to \mathcal{L}_2^m$. The upper bound will be conservative since the space $\mathcal{S}_{G'_k}$ is a subset of $\mathcal{H}_2$, i.e., $\mathcal{S}_{G'_k} \subseteq \mathcal{H}_2 \subset \mathcal{L}_2$. Thus we make the projection onto a larger space and $\tilde{\gamma}$ has a component that is not in $\mathcal{H}_2$. In this case, Lemma 5.A.2 gives that

$$\text{AsCov}\,\hat{\eta} < \left[\langle G'(S - \tilde{\gamma}\tilde{\gamma}^*), G'\rangle\right]^{-1},$$

where

$$\tilde{\gamma}\tilde{\gamma}^* = BD^{-1}B^* = \sum_{k=1}^{m} \frac{\Phi_u |G_k|^2}{\lambda_k}.$$

Simplifying the above expression leads to

$$\text{AsCov}\,\hat{\eta} < \left[\left\langle G' \frac{\Phi_u}{\lambda}, G' \right\rangle\right]^{-1}. \tag{5.18}$$

We summarize the results in the following theorem for non white noise spectra.

**Theorem 5.3.1.** *Let the system dynamics be described* (5.2)*. We assume that the additive zero mean stationary sequences* $\{e_i(t)\}$ *are mutually independent, and independent of the input signal* $u(t)$*, with spectra* $\Phi_{e_i}(\omega), i = 1, \ldots, m$*; the same holds for* $e(t)$ *with spectrum* $\Phi_e(\omega)$*. The input is assumed to be a realization of a weakly stationary stochastic process with spectrum* $\Phi_u(\omega)$*. The models of the modules are independently parametrized, i.e.,* $G_k(q) = G_k(q, \theta_k)$*,* $k = 1, \ldots, m$ *and* $G(q) = G(q, \eta)$*. We assume that* $D$ *in* (5.16) *is full rank (*$\text{rank}\,\{D\} = m$*). Then*

$$\left[\left\langle G' \left(\frac{\Phi_u}{\Phi_e} + \sum_{k=1}^{m} M_{u_k}^L\right), G' \right\rangle\right]^{-1} \leq \text{AsCov}\,\hat{\eta} < \left[\left\langle G' \frac{\Phi_u}{\Phi_e}, G' \right\rangle\right]^{-1}, \tag{5.19}$$

*where*

$$M_{u_k}^L = \frac{\Phi_u |G_k|^2}{\Phi_{e_k}}, \quad k = 1, \ldots, m. \tag{5.20}$$

**Remark 5.3.1.** *Comparing* (5.19) *with open loop SISO identification of* $G$*, we see that the upper bound coincides with the expression for the corresponding SISO case:*

$$\text{AsCov}\,\hat{\eta} = \left[\left\langle G' \frac{\Phi_u}{\lambda}, G' \right\rangle\right]^{-1}.$$

*The lower bound* (5.17) *tells us that if high model orders are used for the models* $\{\hat{G}_k(q)\}_{k=1}^m$*, the benefit of including measurements* $\{y_k(t)\}_{k=1}^m$ *in the predictor* (5.14) *is limited. If we do not have to estimate the dynamics of the sensor modules*

*$\{G_k\}_{k=1}^m$, each sensor gives a contribution $M_{u_k} = \Phi_u|G_k|^2/\Phi_{e_k}$. We see that a sensor $k$ adds a term*

$$\frac{\Phi_{SNR}^k(\omega)}{|G(e^{j\omega})|^2},$$

*where $\Phi_{SNR}^k(\omega)$ is the signal to noise ratio at sensor $k$. This is the same quantity that determines the variance of the first estimated module of the cascaded systems in Chapter 4, cf. Remark 4.2.1 and Remark 4.3.1 in Chapter 4.*

## 5.4 Numerical FIR examples

In this section, we verify the correctness of the presented results for both parallel cascade systems (Structure 1) and multi-sensor systems (Structure 2) by 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 realizations is computed using (1.3). The noise source and input are assumed mutually independent zero mean Gaussian white noise with unit variance. When the input $r_2$ is not considered it is set to zero. The estimates are computed as the minimizer of

$$V_N(\theta) = \frac{1}{2N}\left\{\sum_{t=1}^{N} \varepsilon(t)^{\mathrm{T}} \Lambda^{-1} \varepsilon(t)\right\},$$

where $\varepsilon(t) = y(t) - \hat{y}(t)$. We consider examples with $m = 1$ and $m = 2$ input signals ($m = 1$ and $m = 2$ additional sensors) and 3 FIR systems, all with true order $p = 3$, i.e.

$$G_1 = G_2 = 1 + 0.5q^{-1} + 0.25q^{-2},$$
$$G = 1 + 0.2q^{-1} + 0.04q^{-2}.$$

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

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

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 5.4. In the case of one input signal ($m = 1$), then $r_2 = 0$ and $G$ is estimated. Knowing the first impulse response coefficient of $G_1$ and $G_2$ gives a reduction in variance of $\hat{G}$ as seen in Figure 5.5.

For Structure 2, the multi-sensor structure, the variance of the transfer function estimate $\hat{G}$ does not improve by using also using $y_2$, and is the same as what would be achieved by only using $y$, cf. Figure 5.6. When we know the first coefficient of $G_1$

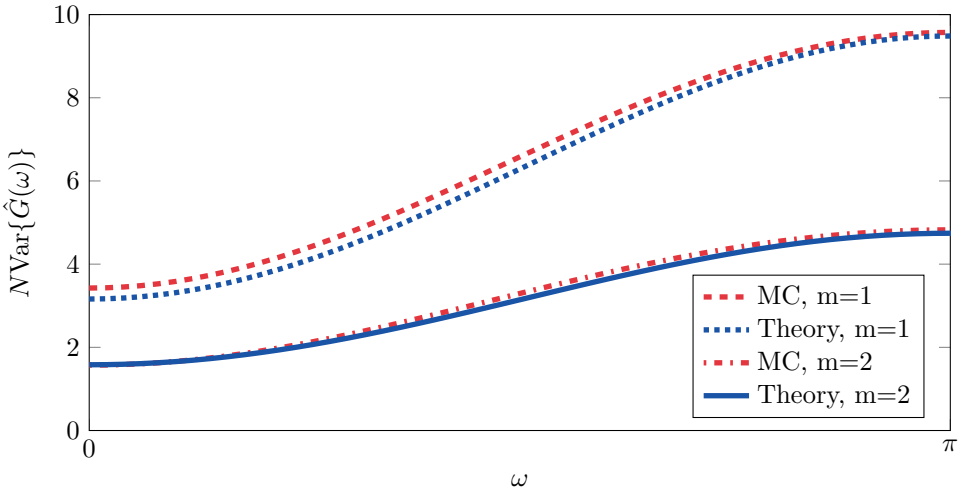


Figure 5.4: Structure 1: Comparison of the variance of $\hat{G}(\omega)$ for Monte-Carlo simulations (MC) and the asymptotic theory for $m = 1, 2$.

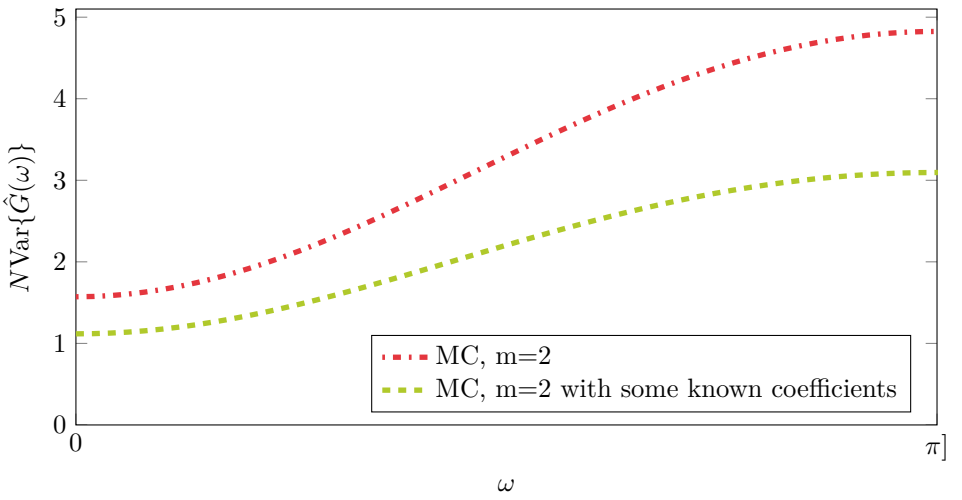


Figure 5.5: Structure 1: Comparison of the variance of $\hat{G}(\omega)$ for Monte-Carlo simulations (MC) and the asymptotic theory for $m = 2$, when the first impulse response coefficients of $G_1$ and $G_2$ are known.

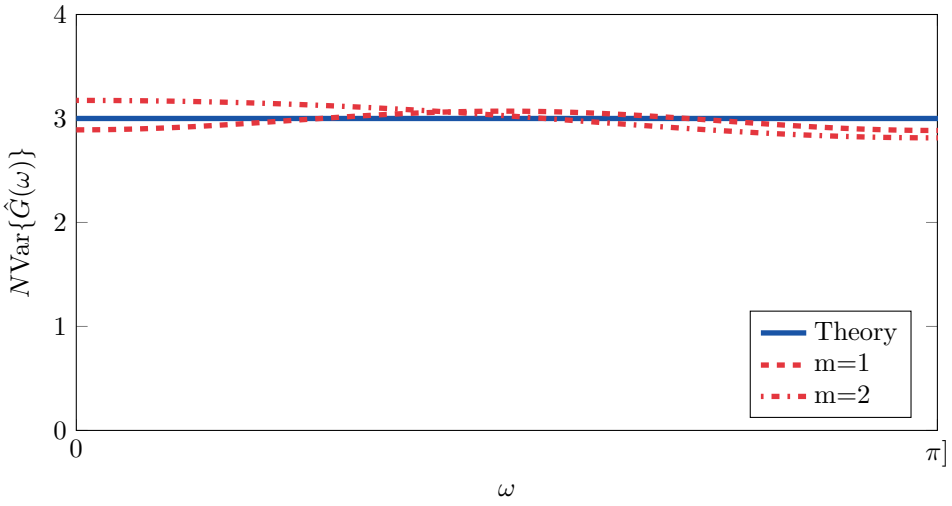


Figure 5.6: Structure 2: Comparison of the variance of $\hat{G}(\omega)$ for Monte-Carlo simulations (MC) and the asymptotic theory for $m = 1, 2$.

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 $\hat{G}$, cf. Figure 5.7. Knowing some parameters in $G_1$ and $G_2$ makes all the difference.

## 5.5 Conclusion

We have examined the covariance of the parameter estimates of one module in a parallel cascade structure and a multi-sensor structure. Conservative upper bounds on the asymptotic covariance of the parameter estimates were derived when little assumptions were made on the remaining systems in the network. We derived asymptotic variance expressions for two types of structured dynamic systems: parallel cascade systems (Structure 1) and multi-sensor systems (Structure 2). For the parallel cascade structure, when the actuator dynamics $\{G_k\}_{k=1}^m$ are completely unknown, the variance reduction from knowing one input (compared to neglecting it and considering it as noise), was shown to be two-folded. One part corresponded to increased excitation and the other to noise reduction. Previous 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.

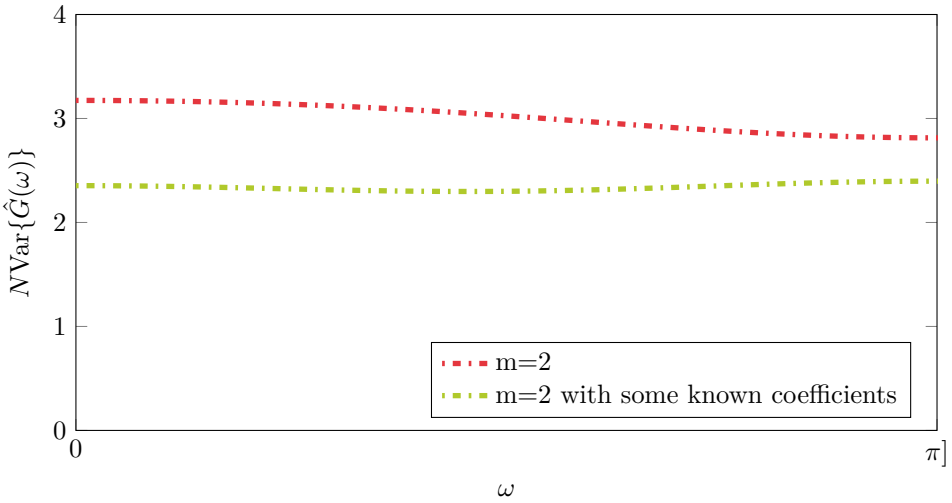


Figure 5.7: Structure 2: Comparison of the variance of $\hat{G}(\omega)$ for Monte-Carlo simulations (MC) and the asymptotic theory for $m = 2$, when the first impulse response coefficients of $G_1$ and $G_2$ are known.

## 5.A Technical preliminaries

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

**Lemma 5.A.1.** *Assume that the asymptotic covariance matrix for a vector* $\theta = \begin{bmatrix}\theta_1^{\mathrm{T}} & \theta_2^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}}$, $\theta_1 \in \mathbb{R}^{n_1}$, $\theta_2 \in \mathbb{R}^{n_2}$, *can be written as*

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

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

$$P_{\theta_2} = \left[\langle \Psi_2, \Psi_2 \rangle - \left\langle \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\}, \mathbf{P}_{\mathcal{S}_{R_1}}\{\gamma\} \right\rangle\right]^\dagger$$

*with* $\gamma = \Psi_2 \Psi_1^*[R_1^\dagger]^*$, *where* $R_1$ *is a spectral factor of* $\Psi_1(e^{j\omega})\Psi_1^{\mathrm{T}}(e^{-j\omega})$, *i.e.,* $\Psi_1(e^{j\omega})\Psi_1^{\mathrm{T}}(e^{-j\omega}) = R_1(e^{j\omega})R_1^{\mathrm{T}}(e^{-j\omega})$ *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

$$\langle \Psi_2, \Psi_1 \rangle = \left\langle \Psi_2 \Psi_1^*[R_1^\dagger]^*, R_1 \right\rangle,$$

and applying the standard inverse of a partitioned matrix (Horn and Johnson, 1990) and Lemma 2.3.1 proves Lemma 5.A.1. □

In the next lemma, we let the number of estimated parameters in the $m$ first modules grow large to make the projection in Lemma 5.A.1 trivial to calculate.

**Lemma 5.A.2.** *Let $\Psi$ be defined as in Lemma 5.A.1 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* $\operatorname{rank}\left\{\tilde{\Psi}_1\tilde{\Psi}_1^*\right\} = m_1$. *If $\mathcal{S}_{\Gamma_1} = \mathcal{L}_2^{m_1}$, then*

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

*with $\tilde{\gamma} = \tilde{\Psi}_2\tilde{\Psi}_1^*[R_1^{-1}]^*$, where $R_1$ is a spectral factor of $\Psi_1(e^{j\omega})\Psi_1^{\mathrm{T}}(e^{-j\omega})$, 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 that $\mathbf{P}_{\mathcal{S}_{\Gamma_1 R_1}}\Gamma_2\tilde{\gamma} = \mathbf{P}_{\mathcal{L}_2^{m_1}}\Gamma_2\tilde{\gamma} = \Gamma_2\tilde{\gamma}$ in Lemma 5.A.1. □

# Chapter 6

# SIMO models with spatially correlated noise

In this chapter we consider SIMO models. SIMO models have a simple multi-output structure, which can have correlation between noise signals. The aim is to investigate how correlation between noise sources can be utilized for variance reduction. SIMO models are interesting in themselves. They find applications in various disciplines, such as signal processing and speech enhancement (Benesty *et al.*, 2005), (Doclo and Moonen, 2002), communications (Bertaux *et al.*, 1999), (Schmidt, 1986), (Trudnowski *et al.*, 1998), biomedical sciences (McCombie *et al.*, 2005) and structural engineering (Ulusoy *et al.*, 2011). Some of these applications are concerned with spatio-temporal models, in the sense that the measured output can be strictly related to the location at which the sensor is placed (Stoica *et al.*, 1994), (Viberg *et al.*, 1997), (Viberg and Ottersten, 1991). In these cases, it is reasonable to expect that measurements collected at locations close to each other are affected by disturbances of the same nature. In other words, noise on the outputs can be correlated; understanding how this noise correlation affects the accuracy of the estimated model is a key issue in data-driven modeling of SIMO systems. We characterize the covariance between the estimated parameters of the estimated modules as well as the covariance between the estimated module transfer functions. The results of this chapter are related to the ones found in Ramazi *et al.* (2014), where multiple input single output (MISO) models are considered, which instead can have correlated inputs. In the MISO case, correlation is detrimental for the accuracy. For white inputs, it was recently shown in Ramazi *et al.* (2014) that an increment in the model order of one transfer function leads to an increment in the variance of another estimated transfer function only up to a point, after which the variance levels off. For MISO models, the concept of connectedness (Gevers *et al.*, 2006) gives conditions on when one input can help reduce the variance of an identified transfer function. These results were later refined in Mårtensson (2007). Note that variance expressions for the special case of SIMO cascade structures are

found in Wahlberg *et al.* (2009) and in Chapter 4.

As a motivation, let us recall Example 1.3.1 of Chapter 1. Consider the model:

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

where the input $u(t)$ is white noise and $e_k, k = 1, 2$ is measurement noise. We consider two different types of measurement noise (uncorrelated with the input). In the first case, the noise is perfectly correlated; let us for simplicity assume that $e_1(t) = e_2(t)$. For the second case, $e_1(t)$ and $e_2(t)$ are independent. It turns out that in the first case we can perfectly recover the parameters $\theta_{11}$ and $\theta_{22}$, while, in the second case we do not improve the accuracy of the estimate of $\theta_{11}$ by also using the measurement $y_2(t)$. The reason for this difference is that, in the first case, we can construct the noise free equation

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

and we can perfectly recover $\theta_{1,1}$ and $\theta_{2,2}$, while in the second case neither $y_2(t)$ nor $e_2(t)$ contain information about $e_1(t)$.

Also the model structure plays an important role for the benefit of the second sensor. To this end, we consider a third case, where again $e_1(t) = e_2(t)$. This time, the model structure is slightly different:

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

In this case, we can construct the noise free equation

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

The fundamental difference is that now only the difference $(\theta_{1,1} - \theta_{2,1})$ can be recovered exactly, but not the parameters $\theta_{1,1}$ and $\theta_{2,1}$ themselves. They can be identified from $y_1(t)$ and $y_2(t)$ separately, as long as $y_1$ and $y_2$ are measured. A similar consideration is made in Ljung *et al.* (2011), where SIMO cascade systems are considered.

We will generalize these observations in the following contributions:

- Novel expressions are provided for the variance-error of an estimated frequency response function of a SIMO linear model in orthonormal basis form, or equivalently, in fixed denominator form (Ninness *et al.*, 1999). The expression reveals how the noise correlation structure, model orders and input variance affect the variance-error of the estimated frequency response function.

- For a non-white input spectrum, we show where in the frequency spectrum the benefit of the correlation structure is focused.

- When one module is identified using less parameters, we derive the noise correlation structure under which the mentioned model parametrization gives the lowest total variance.

The chapter is organized as follows: in Section 6.1 we define the SIMO model structure under study and provide an expression for the covariance matrix of the parameter estimates. Section 6.2 contains the main results, namely a novel variance expression for LTI SIMO orthonormal basis function models. The connection with MISO models is explored in Section 6.3. In Section 6.4, the main results are applied to a non–white input spectrum. In section 6.5 we derive the correlation structure that gives the minimum total variance, when one block has less parameters than the other blocks. Numerical experiments illustrating the application of the derived results are presented in Section 6.6. A final discussion ends the chapter in Section 6.7.

## 6.1 Problem formulation

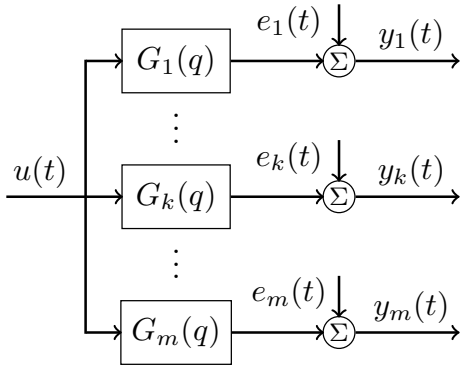


Figure 6.1: *Block scheme of the linear SIMO system.*

We consider linear time-invariant dynamic systems with one input and $m$ outputs (see Fig. 6.1). The model is described as follows:

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix} = \begin{bmatrix} G_1(q) \\ G_2(q) \\ \vdots \\ G_m(q) \end{bmatrix} u(t) + \begin{bmatrix} e_1(t) \\ e_2(t) \\ \vdots \\ e_m(t) \end{bmatrix}, \tag{6.1}$$

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

$$G_i(q, \theta_i) = \Gamma_i(q)\theta_i, \quad \theta_i \in \mathbb{R}^{n_i}; \tag{6.2}$$

where $n_1 \leq \ldots \leq n_m$ and $\Gamma_i(q) = \left[\mathcal{B}_1(q), \ldots, \mathcal{B}_{n_i}(q)\right], i = 1, \ldots, m$, for some orthonormal basis functions $\{\mathcal{B}_k(q)\}_{k=1}^{n_m}$. Let us introduce the vector notation

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

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

$$\begin{aligned} \mathrm{E}\left[e(t)\right] &= 0 \\ \mathrm{E}\left[e(t)e(s)^{\mathrm{T}}\right] &= \delta_{t-s}\Lambda \end{aligned} \tag{6.3}$$

for some positive definite matrix covariance matrix $\Lambda$. We express $\Lambda$ in terms of its Cholesky factorization

$$\Lambda = \Lambda_{CH}\Lambda_{CH}^{\mathrm{T}}, \tag{6.4}$$

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

$$\Lambda_{CH} = \begin{bmatrix} \gamma_{11} & 0 & \ldots & 0 \\ \gamma_{21} & \gamma_{22} & \ldots & 0 \\ \vdots & \ldots & \ddots & 0 \\ \gamma_{m1} & \gamma_{m2} & \ldots & \gamma_{mm} \end{bmatrix} \tag{6.5}$$

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

$$\Lambda^{-1} = \Lambda_{CH}^{-T}\Lambda_{CH}^{-1}. \tag{6.6}$$

**Remark 6.1.1.** *Generalization of the model structure:*

- *The assumption that the modules have the same orthogonal parametrization in* (6.2) *is less restrictive than it might appear at first glance, and made for clarity and ease of presentation. A model consisting of non-orthonormal basis functions can be transformed into this format by a linear transformation, which can be computed by the Gram-Schmidt procedure (Trefethen and Bau, 1997). Notice that fixed denominator models, with the same denominator in all transfer functions, also fit this framework (Ninness et al., 1999). However, it is essential for our analysis that the modules share poles.*

- *It is not necessary to restrict the input to be white noise. A colored input introduces a weighting by its spectrum $\Phi_u$, which means that the basis functions need to be orthogonal with respect to the inner product $\langle f, g\rangle_{\Phi_u} := \langle f\Phi_u, g\rangle$. If $\Phi_u(z) = \sigma^2 R(z)R^*(z)$, where $R(z)$ is a monic stable minimum phase spectral factor, the transformation $\tilde{\Gamma}_i(q) = \sigma^{-1}R(q)^{-1}\Gamma_i(q)$ is a procedure that gives a set of orthogonal basis functions in the weighted space. If we use this parametrization, all the main results carry over naturally. However, in general, the new parametrization does not contain the same set of models as the original parametrization. Another way is to use the Gram-Schmidt method, which maintains the same model set and the main results are still valid. If we would like to keep the original parametrization, we may, in some cases, proceed as in Section 6.4 where the input signal is generated by an AR-spectrum.*

- *Note that the assumption that $n_1 \leq \ldots \leq n_m$ is not restrictive as it only represents an ordering of the modules in the system.*

### Weighted least-squares estimate

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

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

we can write the model (6.1) as a linear regression model

$$y(t) = \varphi^{\mathrm{T}}(t)\theta + e(t), \tag{6.7}$$

where

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

An unbiased and consistent estimate of the parameter vector $\theta$ can be obtained from weighted least-squares, with optimal weighting matrix $\Lambda^{-1}$ (see, e.g., Ljung (1999); Söderström and Stoica (1989)). The estimate of $\theta$ is given by

$$\hat{\theta}_N = \left(\sum_{t=1}^N \varphi(t)\Lambda^{-1}\varphi^{\mathrm{T}}(t)\right)^{-1} \sum_{t=1}^N \varphi(t)\Lambda^{-1}y(t). \tag{6.8}$$

Inserting (6.7) in (6.8) gives

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

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

$$\operatorname{AsCov} \hat{\theta}_N = \left(\operatorname{E}\left[\varphi(t)\Lambda^{-1}\varphi^{\mathrm{T}}(t)\right]\right)^{-1}. \tag{6.9}$$

In the problem we consider, using Parseval's formula and (6.6), the asymptotic covariance matrix, (6.9), can be written as[1]

$$\begin{aligned}\operatorname{AsCov} \hat{\theta}_N &= \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \Psi(e^{j\omega})\Psi^*(e^{j\omega})\,\mathrm{d}\omega\right]^{-1} \\ &= \langle \Psi, \Psi \rangle^{-1},\end{aligned} \tag{6.10}$$

where

$$\Psi(q) = \frac{1}{\sigma}\tilde{\Psi}(q)\Lambda_{CH}^{-T}. \tag{6.11}$$

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

### The introductory example

With formal assumptions in place, we now consider the introductory example in greater detail. Consider the model

$$y_1(t) = \theta_{1,1}q^{-1}u(t) + e_1(t), \tag{6.12}$$
$$y_2(t) = \theta_{2,1}q^{-1}u(t) + \theta_{2,2}q^{-2}u(t) + e_2(t) \tag{6.13}$$

which uses the delays $q^{-1}$ and $q^{-2}$ as orthonormal basis functions. With $\theta = [\theta_{1,1}\,\theta_{2,1}\,\theta_{2,2}]^{\mathrm{T}}$; the corresponding regression matrix is

$$\varphi(t)^{\mathrm{T}} = \begin{bmatrix} u(t-1) & 0 & 0 \\ 0 & u(t-1) & u(t-2) \end{bmatrix}.$$

The noise vector is generated by

$$\begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix} = Lw(t) = \begin{bmatrix} 1 & 0 \\ \sqrt{1-\beta^2} & \beta \end{bmatrix} \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix}, \tag{6.14}$$

where $w_1(t)$ and $w_2(t)$ are uncorrelated white processes with unit variance. The parameter $\beta \in [0,\, 1]$ tunes the correlation between $e_1(t)$ and $e_2(t)$. When $\beta = 0$, the

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

two processes are perfectly correlated (i.e., identical); conversely, when $\beta = 1$, they are completely uncorrelated. Note that, for every $\beta \in [0, 1]$, one has $\mathrm{E}\left[e_1(t)^2\right] = \mathrm{E}\left[e_2(t)^2\right] = 1$. In fact, the covariance matrix of $e(t)$ becomes

$$\Lambda = LL^{\mathrm{T}} = \begin{bmatrix} 1 & \sqrt{1-\beta^2} \\ \sqrt{1-\beta^2} & 1 \end{bmatrix}.$$

Then, when computing (6.10) in this specific case gives

$$\mathrm{AsCov}\,\hat{\theta}_N = \frac{1}{\sigma^2}\begin{bmatrix} 1 & \sqrt{1-\beta^2} & 0 \\ \sqrt{1-\beta^2} & 1 & 0 \\ 0 & 0 & \beta^2 \end{bmatrix}. \tag{6.15}$$

We note that:

$$\mathrm{AsCov}\,\begin{bmatrix}\hat{\theta}_{1,1} & \hat{\theta}_{2,1}\end{bmatrix}^{\mathrm{T}} = \frac{1}{\sigma^2}\Lambda,$$
$$\mathrm{AsCov}\,\hat{\theta}_{2,2} = \frac{1}{\sigma^2}\beta^2.$$

The above expressions reveals two interesting facts:

1. The (scalar) variances of $\hat{\theta}_{1,1}$ and $\hat{\theta}_{2,1}$, namely the estimates of parameters of the two modules related to the same time lag, are not affected by possible correlation of the noise processes, i.e., they are independent of the value of $\beta$. However, note that the cross correlation between $\hat{\theta}_{1,1}$ and $\hat{\theta}_{2,1}$ in (6.15):

$$\begin{aligned}\mathrm{Var}\,(\hat{\theta}_{1,1} - \sqrt{1-\beta^2}\hat{\theta}_{2,1}) &= \begin{bmatrix} 1 \\ -\sqrt{1-\beta^2}\end{bmatrix}^{\mathrm{T}} \frac{1}{\sigma^2}\Lambda \begin{bmatrix} 1 \\ -\sqrt{1-\beta^2}\end{bmatrix} \\ &= \frac{1}{\sigma^2}\beta^2. \end{aligned} \tag{6.16}$$

   This cross correlation will induce a cross correlation in the transfer function estimates as well.

2. As seen in (6.15), the variance of $\hat{\theta}_{1,2}$ strongly depends on $\beta$. In particular, when $\beta$ tends to 0, one is ideally able to estimate $\hat{\theta}_{1,2}$ perfectly. Note that in the limit case $\beta = 0$ one has $e_1(t) = e_2(t)$, so that (6.1) can be rearranged to obtain the noise-free equation

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

   which shows that both $\theta_{1,2}$ and the difference $\theta_{1,1} - \theta_{2,1}$ can be estimated perfectly. This can of course also be seen from (6.15), cf. (6.16).

The example shows that correlated measurements can be favorable for estimating for estimating certain parameters, but not necessarily for all. The main focus of this chapter is to generalize these observations to arbitrary basis functions, number of systems and number of estimated parameters. Additionally, the results are used to derive the optimal correlation structure of the noise. But first, we need some technical preliminaries.

### Non-estimable part of the noise

As seen in the example of Section 6.1, strong noise correlation may be helpful in the estimation. In fact, the variance error will depend on the non-estimable part of the noise, i.e., the part that cannot be linearly estimated from other noise sources. To be more specific, define the signal vector $e_{j\setminus i}(t)$ to include the noise sources from module 1 to module $j$, with the one from module $i$ excluded, i.e.,

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

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

$$\hat{e}_{i|j}(t) := \varrho_{ij}^{\mathrm{T}} e_{j\setminus i}(t). \tag{6.17}$$

Introduce the notation

$$\lambda_{i|j} := \mathrm{Var}\,[e_i(t) - \hat{e}_{i|j}(t)], \tag{6.18}$$

with the convention that $\lambda_{i|0} := \lambda_i$. The vector $\varrho_{ij}$ in (6.17) is given by

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

We call

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

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

**Definition 6.1.1.** *When $\hat{e}_{i|j}(t)$ does not depend on $e_k(t)$, where $1 \le k \le j$, $k \neq i$, we say that $e_i(t)$ is orthogonal to $e_k(t)$ conditionally to $e_{j\setminus i}(t)$.*

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

**Lemma 6.1.1.** *Let $e(t) \in \mathbb{R}^m$ have zero mean and covariance matrix $\Lambda > 0$. Let $\Lambda_{CH}$ be the lower triangular Cholesky factor of $\Lambda$, i.e., $\Lambda_{CH}$ satisfies* (6.4)*, with $\{\gamma_{ik}\}$ as its entries as defined by* (6.5)*. Then for $j < i$,*

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

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

*Proof.* See Appendix 6.A. □

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

$$e_{i:m}(t) := \begin{bmatrix} e_i(t) & \dots & e_m(t) \end{bmatrix}^{\mathrm{T}},$$

and for $j < i$ we define $\hat{e}_{i:m|j}(t)$ as the linear minimum variance estimate of $e_{i:m}(t)$ based on the other signals $e_{j\setminus i-1}(t)$, and

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

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

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

From the Cholesky factorization above we see that, since $\gamma_{32}$ is zero, Lemma 6.1.1 gives that $e_3(t)$ is orthogonal to $e_2(t)$ given $e_{2\setminus 3}(t)$, i.e., there is no information about $e_3(t)$ in $e_2(t)$ if we already know $e_1(t)$. This is not apparent from $\Lambda$ where every entry is non-zero. If we know $e_1(t)$ a considerable part of $e_2(t)$ and $e_3(t)$ can be estimated. Without knowing $e_1(t)$, $\lambda_1 = \lambda_2 = \lambda_3 = 1$, while if we know $e_1(t)$, $\lambda_{2|1} = 0.64$ and $\lambda_{3|1} = 0.19$.

## 6.2 Covariance of frequency response estimates

In this section, we present novel expressions for the variance-error of an estimated frequency response function. The expression reveals how the noise correlation structure, model orders and input variance affect the variance-error of the estimated frequency response function. We will analyze the effect of the correlation structure of the noise on the transfer function estimates. To this end, collect all $m$ transfer functions into

$$G := \begin{bmatrix} G_1 & G_2 & \dots & G_m \end{bmatrix}.$$

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

**Definition 6.2.1.** *The asymptotic covariance of $\hat{G}(e^{j\omega_0}) := G(e^{j\omega_0}, \hat{\theta}^N)$ for the fixed frequency $\omega_0$ is denoted by*

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

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

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

*Define $\chi_k$ as the index of the first system that contains the basis function $\mathcal{B}_k(e^{j\omega_0})$. Notice that $\chi_k - 1$ is the number of systems that do not contain the basis function.*

**Theorem 6.2.1.** *Suppose that the parameters $\theta_i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, m$, are estimated using weighted least-squares* (6.8)*. Let the entries of $\theta$ be arranged as follows:*

$$\bar{\theta} = [\theta_{1,1} \ \ldots \ \theta_{m,1} \ \theta_{1,2} \ \ldots \ \theta_{m,2} \ \ldots \ \theta_{1,n_1} \ \ldots \\ \ldots \ \theta_{m,n_1} \ \theta_{2,n_1+1} \ \ldots \ \theta_{m,n_1+1} \ \ldots \ \theta_{m,n_m}]^{\mathrm{T}}. \tag{6.19}$$

*and the corresponding weighted least-squares estimate be denoted by $\hat{\bar{\theta}}$. Then, the covariance of $\hat{\bar{\theta}}$ is*

$$\operatorname{AsCov} \hat{\bar{\theta}} = \frac{1}{\sigma^2} \operatorname{diag} \left\{ \Lambda_{1:m}, \Lambda_{\chi_2:m|\chi_2-1}, \ldots, \Lambda_{\chi_{n_m}:m|\chi_{n_m}-1} \right\}. \tag{6.20}$$

*In particular, the covariance of the parameters related to basis function number $k$ is given by*

$$\operatorname{AsCov} \hat{\bar{\theta}}_k = \frac{1}{\sigma^2} \Lambda_{\chi_k:m|\chi_k-1}, \tag{6.21}$$

*where*

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

*and where, for $\chi_k \leq i \leq m$,*

$$\operatorname{AsVar} \hat{\theta}_{i,k} = \frac{\lambda_{i|\chi_k-1}}{\sigma^2}. \tag{6.22}$$

*It also holds that*

$$\operatorname{AsCov} \hat{G} = \sum_{k=1}^{n_m} \begin{bmatrix} \mathbf{0}_{\chi_k-1} & \mathbf{0} \\ \mathbf{0} & \operatorname{AsCov} \hat{\bar{\theta}}_k \end{bmatrix} |\mathcal{B}_k(e^{j\omega_0})|^2; \tag{6.23}$$

*where* $\operatorname{AsCov} \hat{\bar{\theta}}_k$ *is given by* (6.21) *and $\mathbf{0}_{\chi_k-1}$ is a $\chi_k - 1 \times \chi_k - 1$ matrix with all entries equal to zero. For $\chi_k = 1$, $\mathbf{0}_{\chi_k-1}$ is an empty matrix. In* (6.23)*, $\mathbf{0}$ denotes zero matrices of dimensions compatible to the diagonal blocks.*

*Proof.* See Appendix 6.B □

**Remark 6.2.1.** *The covariance of $\hat{\bar{\theta}}_k$, which contain the parameters related to basis function $k$, is determined by which other models share the basis function $\mathcal{B}_k$; cf.* (6.16) *of the introductory example. The asymptotic covariance of $\hat{G}$ can be understood as a sum of the contributions from each of the $n_m$ basis functions. The covariance contribution from a basis function $\mathcal{B}_k$ is weighted by $|\mathcal{B}_k(e^{j\omega_0})|^2$ and only affects the covariance between systems that contain that basis function, as visualized in Figure 6.2.*

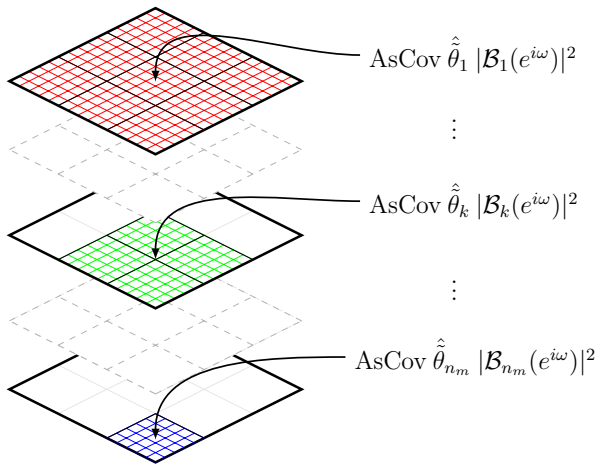


Figure 6.2: A Graphical representation of AsCov $\hat{G}$ where each term of the sum in (6.23) is represented by a layer. A basis function only affect the covariance between modules that also contain that basis function. Thus, the first basis function affects the complete covariance matrix while the last basis function $n_m$ only affects modules $\chi_{n_m}, \ldots, m$.

**Remark 6.2.2.** *The orthogonal basis functions correspond to a decomposition of the output signals into orthogonal components and the problem in a sense becomes decoupled. As an example, consider the system described by*

$$\begin{aligned} y_1(t) &= \theta_{1,1}\mathcal{B}_1(q)u(t) + e_1(t), \\ y_2(t) &= \theta_{2,1}\mathcal{B}_1(q)u(t) + e_2(t), \\ y_3(t) &= \theta_{3,1}\mathcal{B}_1(q)u(t) + \theta_{3,2}\mathcal{B}_2(q)u(t) + e_3(t), \end{aligned} \tag{6.24}$$

*Suppose that we are interested in estimating $\theta_{3,2}$. For this parameter,* (6.22) *becomes*

$$\text{AsVar}\,\hat{\theta}_{3,2} = \frac{\lambda_{3|2}}{\sigma^2} \tag{6.25}$$

*To understand the mechanisms behind this expression, let $u_1(t) = \mathcal{B}_1(q)u(t)$, and $u_2(t) = \mathcal{B}_2(q)u(t)$ so that the system can be visualized as in Figure 6.3, i.e., we can consider $u_1$ and $u_2$ as separate inputs.*

*First we observe that it is only $y_3$ that contains information about $\theta_{3,2}$, and the term $\theta_{3,1}u_1$ contributing to $y_3$ is a nuisance from the perspective of estimating $\theta_{3,2}$. This term vanishes when $u_1 = 0$ and we will not be able to achieve better accuracy than the optimal estimate of $\theta_{3,2}$ for this idealized case. So let us study this setting first. Straightforward application of the least-squares method, using $u_2$ and $y_3$, gives an estimate of $\theta_{3,2}$ with variance $\lambda/\sigma^2$, which is larger than* (6.25) *when $e_3$ depends*

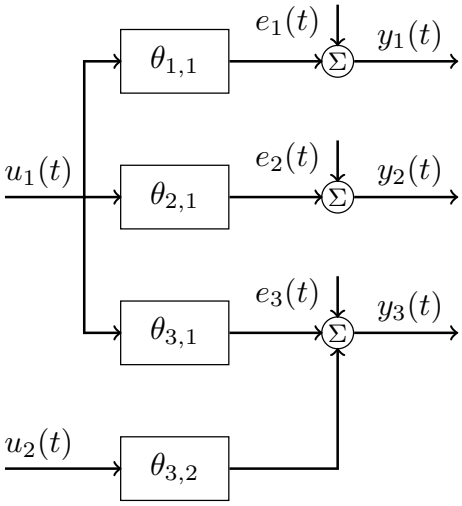


Figure 6.3: The SIMO system of Remark 6.2.2, described by (6.24).

*on $e_1$ and $e_2$. However, in this idealized case, $y_1 = e_1$ and $y_2 = e_2$, and these signals can thus be used to estimate $e_3$ and this estimate can then be subtracted from $y_3$ before the least-squares method is applied. The remaining noise in $y_3$ will have variance $\lambda_{3|2}$ if $e_3$ is optimally estimated (see* (6.17)–(6.18)*), and hence the least-squares estimate will now have variance $\lambda_{3|2}/\sigma^2$, i.e., the same as* (6.25)*.*

*To understand why it is possible to achieve the same accuracy as this idealized case when $u_1$ is non-zero, we need to observe that our new inputs $u_1(t)$ and $u_2(t)$ are orthogonal (uncorrelated)*[2]*. Returning to the case when only the output $y_3$ is used for estimating $\theta_{3,2}$, this implies that we pay no price for including the term $\theta_{3,1}u_1$ in our model, and then estimating $\theta_{3,1}$ and $\theta_{3,2}$ jointly, i.e., the variance of $\hat{\theta}_{3,2}$ will still be $\lambda/\sigma^2$*[3]*. The question now is if we can use $y_1$ and $y_2$ as before to estimate $e_3$? Perhaps surprising, we can use the same estimate as when $u_1$ was zero. The reader may object that this estimate will now, in addition to the previous optimal estimate of $e_3$, contain a term which is a multiple of $u_1$. However, due to the orthogonality between $u_1$ and $u_2$, this term will only affect the estimate of $\theta_{3,1}$ (which we anyway were not interested in, in this example), and the accuracy of the estimate of $\theta_{3,2}$ will be $\lambda_{3|2}/\sigma^2$, i.e.* (6.25)*. Figure 6.4 illustrates the setting with $\tilde{y}_3$ denoting $y_3$ subtracted by the optimal estimate of $e_3$. A key insight from this discussion is that, for the estimate of a parameter in the path from input $i$ to output*

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

[3]With $u_1$ and $u_2$ correlated, the variance will be higher, see Section 6.3 for a further discussion of this topic.

*$j$, it is only outputs that are not affected by input $i$ that can be used to estimate the noise in output $j$ when this particular parameter is estimated, using outputs influenced by input $i$ will introduce a bias since the noise estimate will then contain a term that is not orthogonal to this input.*

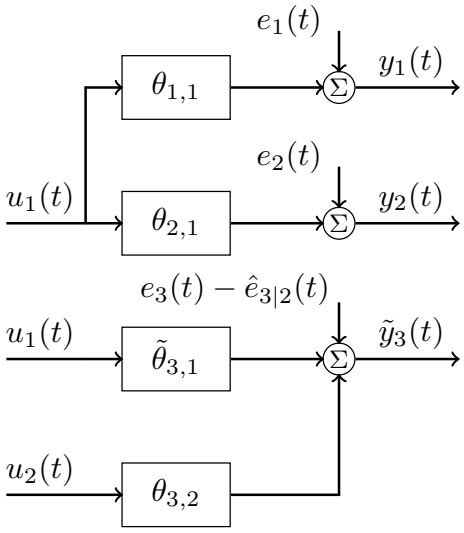


Figure 6.4: The SIMO system of Remark 6.2.2, described by (6.24), but with $\tilde{y}_3$ denoting $y_3$ subtracted by the optimal estimate of $e_3$.

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

**Corollary 6.2.1.** *Let the same assumptions as in Theorem 6.2.1 hold. Then, for any frequency $\omega_0$, it holds that*

$$\text{AsVar } \hat{G}_i = \sum_{k=1}^{n_i} |\mathcal{B}_k(e^{j\omega_0})|^2 \text{AsVar } \hat{\theta}_{i,k}, \tag{6.26}$$

*where*

$$\text{AsVar } \hat{\theta}_{i,k} = \frac{\lambda_{i|\chi_k - 1}}{\sigma^2}, \tag{6.27}$$

*and $\lambda_{i|j}$ is defined in* (6.18).

*Proof.* Follows from Theorem 6.2.1, since (6.26) is a diagonal element of (6.23). □

**Remark 6.2.3.** *Similar to what was noted in Remark 6.2.1 the variance of $\hat{G}_i$ in* (6.26) *is determined as a weighted sum of the estimated parameters of $\hat{G}_i$, where the variance of each estimated parameter $\hat{\theta}_{i,k}$ is weighted by $|\mathcal{B}_k(e^{j\omega_0})|^2$.*

From Corollary 6.2.1, we can tell when increasing the model order of $G_j$ will increase the asymptotic variance of $\hat{G}_i$.

**Corollary 6.2.2.** *Under the same conditions as in Theorem 6.2.1, if we increase the number of estimated parameters of $G_j$ from $n_j$ to $n_j+1$, the asymptotic variance of $G_i$ will increase, if and only if all the following conditions hold:*

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

*Proof.* See Appendix 6.C. □

**Remark 6.2.4.** *Corollary 6.2.2 explicitly tells when an increase in the model order of $G_j$ from $n_j$ to $n_j + 1$ will increase the variance of $G_i$. Notice that if $n_j \geq n_i$ then there will be no increase in the variance of $\hat{G}_i$, no matter how many additional parameters we introduce to the model $G_j$. Naturally, if $e_i(t)$ is orthogonal to $e_j(t)$ conditioned on $e_{j\setminus i}(t)$, $\hat{e}_{i|j}(t)$ does not depend on $e_j(t)$ and there is no increase in the variance of $\hat{G}_i$, cf. Remark 6.2.2.*

### A graphical representation of Corollary 6.2.2

Following the notation in Bayesian Networks (Koski and Noble, 2012), Conditions 1) and 2) in Corollary 6.2.2 can be interpreted graphically. Each module is represented by a vertex in a weighted directed acyclic graph $\mathcal{G}$. Let the vertices be ordered by model order, i.e., let the first vertex correspond to $\hat{G}_1$. Under the additional assumption that module $i$ is the first module with order $n_i$, let there be an edge, denoted by $j \to i$, from vertex $j$ to $i$, $j < i$, if $e_i(t)$ is not orthogonal to $e_j(t)$ conditioned on $e_{j\setminus i}(t)$. Notice that this is equivalent to $\gamma_{ij} \neq 0$. Let the weight of the edge be $\gamma_{ij}$ and define the parents of vertex $i$ to be all nodes with a link to vertex $i$, i.e., $pa_{\mathcal{G}}(i) := \{j : j \to i\}$. Then, (6.27), together with Lemma 6.1.1, shows that only outputs corresponding to parents of node $i$ affect the asymptotic variance. Indeed, a vertex without parents has variance

$$\text{AsVar } \hat{G}_i = \frac{\lambda_i}{\sigma^2} \sum_{k=1}^{n_i} |\mathcal{B}_k(e^{j\omega_0})|^2, \tag{6.28}$$

which corresponds to (6.26) with

$$\lambda_{i|0} = \ldots = \lambda_{i|i-1} = \lambda_i.$$

Thus, AsVar $\hat{G}_i$ is independent of the model order of the other modules.

As an example, consider four systems with the lower Cholesky factor of the covariance of $e(t)$ given by:

$$\begin{aligned}\Lambda &= \begin{bmatrix} 1 & 0.4 & 0.2 & 0 \\ 0.4 & 1.16 & 0.08 & 0.3 \\ 0.2 & 0.08 & 1.04 & 0 \\ 0 & 0.3 & 0 & 1.09 \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.4 & 1 & 0 & 0 \\ 0.2 & 0 & 1 & 0 \\ 0 & 0.3 & 0 & 1 \end{bmatrix}}_{\Lambda_{CH}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.4 & 1 & 0 & 0 \\ 0.2 & 0 & 1 & 0 \\ 0 & 0.3 & 0 & 1 \end{bmatrix}^{\mathrm{T}}\end{aligned} \tag{6.29}$$

If the model orders are distinct, the corresponding graph is given in Figure 6.5, where one can see that AsVar $\hat{G}_4$ depends on $y_2$ (and on $y_4$ of course), but depends neither on $y_3$ nor $y_1$ since $\gamma_{43} = \gamma_{41} = 0$, AsVar $\hat{G}_3$ depends on $y_1$, but not on $y_2$ since $\gamma_{32} = 0$ and AsVar $\hat{G}_2$ depends on $y_1$, while the variance of $\hat{G}_1$ is given by (6.28). If $n_2 = n_4$, the first condition of Corollary 6.2.2 is not satisfied and we have to cut the edge between $\hat{G}_4$ and $\hat{G}_2$. Similarly, if $n_1 = n_2$, we have to cut the edge between $\hat{G}_2$ and $\hat{G}_1$, and if additionally $n_1 = n_2 = n_3$, we have to cut the edge between $\hat{G}_3$ and $\hat{G}_1$.

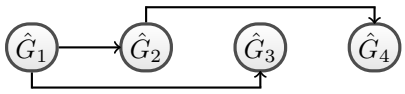


Figure 6.5: Graphical representation of Conditions 1) and 2) in Corollary 6.2.2 for the Cholesky factor given in (6.29).

## 6.3 Connection between MISO and SIMO

There is a strong connection between the results presented here and those regarding MISO systems presented in Ramazi *et al.* (2014). We briefly restate the problem formulation and some results from Ramazi *et al.* (2014) to show the connection. The MISO data generating system is in some sense the dual of the SIMO case. With $m$ spatially correlated inputs and one output, a MISO system is described by

$$y(t) = \begin{bmatrix} G_1(q) & G_2(q) & \dots & G_m(q) \end{bmatrix} \begin{bmatrix} u_1(q) \\ u_2(q) \\ \vdots \\ u_m(q) \end{bmatrix} + e(t).$$

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

$$\begin{aligned} \mathrm{E}\left[u(t)\right] &= 0 \\ \mathrm{E}\left[u(t)u(s)^{\mathrm{T}}\right] &= \delta_{t-s}\Sigma_u, \end{aligned}$$

for some positive definite matrix covariance matrix $\Sigma_u = \Sigma_{CH}\Sigma_{CH}^{\mathrm{T}}$, where $\Sigma_{CH}$ is the lower triangular Cholesky factor of $\Sigma$. The noise $e(t)$ is zero mean and has variance $\lambda$. The asymptotic covariance of the estimated parameters can be expressed using (6.10) with

$$\Psi = \Psi^{\mathrm{MISO}} := \tilde{\Psi}\Sigma_{CH}. \tag{6.30}$$

We make the convention that $\sum_{k=k_1}^{k_2} x_k = 0$ whenever $k_1 > k_2$.

**Theorem 6.3.1** (Theorem 4 in Ramazi *et al.* (2014))**.** *With $n_1 \geq n_2 \geq \ldots, \geq n_m$, for any frequency $\omega_0$ it holds that*

$$\mathrm{AsVar}\ \hat{G}_i = \sum_{j=i}^{m} \frac{\lambda}{\sigma^2_{i|j}} \sum_{k=n_{j+1}+1}^{n_j} |\mathcal{B}_k(e^{j\omega_0})|^2, \tag{6.31}$$

*where $n_{m+1} := 0$ and $\sigma^2_{i|j}$ is the variance of the non-estimable part of $u_i(t)$ given $u_{j\setminus i}(t)$.*

**Corollary 6.3.1** (Corollary 6 in Ramazi *et al.* (2014))**.** *With $n_1 \geq n_2 \geq \ldots, \geq n_m$. Suppose that the order of block $j$ is increased from $n_j$ to $n_{j+1}$. Then there is an increase in the asymptotic variance of $\hat{G}_i$ if and only if all the following conditions hold:*

1. *$n_j < n_i$,*
2. *$u_i(t)$ is not orthogonal to $u_j(t)$ conditioned on $u_{j\setminus i}(t)$,*
3. *$|\mathcal{B}_{n_{j+1}}(e^{j\omega_0})|^2 \neq 0$.*

**Remark 6.3.1.** *The similarities between Corollary 6.2.1 and Theorem 6.3.1, and between Corollary 6.2.2 and Corollary 6.3.1 are striking. In both cases it is the non-estimable part of the input and noise, respectively, along with the estimated basis functions that are the key determinants for the resulting accuracy. Just as in Corollary 6.2.1, Theorem 6.3.1 can be expressed with respect to the basis functions:*

$$\mathrm{AsVar}\ \hat{G}_i = \sum_{k=1}^{n_i} \mathrm{AsVar}\ \hat{\theta}_{i,k}\, |\mathcal{B}_k(e^{j\omega_0})|^2. \tag{6.32}$$

*However, now*

$$\mathrm{AsVar}\ \hat{\theta}_{i,k} = \frac{\lambda}{\sigma^2_{i|\chi_k}} \tag{6.33}$$

*where* $\sigma^2_{i|_l}$ *is determined by the correlation structure of the inputs* $u_i(t)$ *to the systems* $G_i(q,\theta_i)$ *that* do *share basis function* $\mathcal{B}_k(q)$ *(*$i = 1,\dots,\chi_k$*). Note that in the SIMO case we had*

$$\operatorname{AsVar}\hat{\theta}_{i,k} = \frac{\lambda_{i|\chi_k}}{\sigma^2}$$

*where* $\lambda_{i|\chi_k}$ *is determined by the correlation structure of the noise sources* $e_i(t)$ *affecting systems* $G_i(q,\theta_i)$ *that* do not *share basis function* $\mathcal{B}_k(q)$ *(*$i = 1,\dots,\chi_k$*). Note that* (6.31) *found in Ramazi et al. (2014) does not draw the connection to the variance of the parameters. This is made explicit in the alternate expressions* (6.33) *and* (6.32)*.*

The correlation between parameters related to the same basis functions is not explored in Ramazi *et al.* (2014). In fact, it is possible to follow the same line of reasoning leading to Theorem 6.2.1 and arrive at the counter-part for MISO systems. Let the first $\chi_k$ systems contain basis function $k$, so

$$\operatorname{AsVar}\hat{\bar{\theta}}_k^{MISO} = \lambda\, \Sigma^{-1}_{1:\chi_k}$$

where $\Sigma_{1:\chi_k}$ denotes the covariance matrix of the first $\chi_k$ inputs. Hence

$$\operatorname{AsCov}\hat{G} = \lambda \sum_{k=1}^{n_1} \begin{bmatrix} \Sigma^{-1}_{1:\chi_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{m-\chi_k} \end{bmatrix} |\mathcal{B}_k(e^{j\omega_0})|^2,$$

and

$$\operatorname{AsCov}\hat{\bar{\theta}}^{MISO} = \lambda \operatorname{diag}\left\{ \Sigma^{-1}_{1:\chi_1}, \Sigma^{-1}_{1:\chi_2}, \dots, \Sigma^{-1}_{1:\chi_{n_m}} \right\}. \tag{6.34}$$

Note that, while the correlation between the noise sources is beneficial, the correlation in the input is detrimental for the estimation accuracy. Intuitively, if we use the same input to parallel systems, and only observe the sum of the outputs, there is no way to determine the contribution from the individual systems. On the other hand, as observed in the example in Section 6.1, if the noise is correlated, we can construct equations with reduced noise and improve the accuracy of our estimates.

This difference may also be understood from the structure of $\Psi$, which through (6.10) determines the variance properties of any estimate. Consider a single SISO system $G_1$ as the basic case. For the SIMO structure considered in this chapter, as noted before, $\Psi^{\text{SIMO}}$ of (6.11) is block upper triangular with $m$ columns (the number of outputs), while $\Psi^{\text{MISO}}$ is block lower triangular with as many columns as inputs. $\Psi^{\text{MISO}}$ is block lower triangular since $\tilde{\Psi}$ is block diagonal and $\Sigma_{CH}$ is lower triangular in (6.30). Adding an output $y_j$ to the SIMO structure corresponds to extending $\Psi^{\text{SIMO}}$ with one column (and $n_j$ rows):

$$\Psi_e^{\text{SIMO}} = \begin{bmatrix} \Psi^{\text{SIMO}} & \star \\ 0 & \star \end{bmatrix}, \tag{6.35}$$

where the zero comes from that $\Psi_e^{\text{SIMO}}$ also is block upper triangular. Since $\Psi^{\text{MISO}}$ is block lower triangular, adding an input $u_j$ to the MISO structure extends $\Psi^{\text{MISO}}$ with $n_j$ rows (and one column):

$$\Psi_e^{\text{MISO}} = \begin{bmatrix} \Psi^{\text{MISO}} & 0 \\ \star & \star \end{bmatrix}, \tag{6.36}$$

where $\star$ denotes the added column and added row respectively. Addition of columns to $\Psi$ decreases the variance of $G_1$, while addition of rows increases the variance. First, a short motivation of this will be given. Second, we will discuss the implication for the variance analysis.

Addition of one more column to $\Psi$ in (6.35) decreases the variance of $G_1$. With $\Psi = \begin{bmatrix} \psi_1 & \dots & \psi_m \end{bmatrix}$, $\langle \Psi, \Psi \rangle = \sum_k^m \langle \psi_k, \psi_k \rangle$, where $\langle \psi_k, \psi_k \rangle \geq 0$ for every $k$. The variance of the parameter estimate $\hat{\theta}_N$ decreases with the addition of a column, since

$$\langle \Psi_e, \Psi_e \rangle^{-1} \leq \left[ \langle \psi_{m+1}, \psi_{m+1} \rangle + \langle \Psi, \Psi \rangle \right]^{-1}.$$

On the other hand, addition of rows leads to an increase in variance of $\hat{G}_1$, e.g., consider (2.30) in Lemma 2.3.2,

$$\text{AsCov } G_1(\hat{\theta}_N) = L^{\text{T}} \sum_{k=1}^{r} \mathcal{B}_k^{\mathcal{S}}(z_o)^* \mathcal{B}_k^{\mathcal{S}}(z_o)\, L,$$

where $L = \Sigma_{CH}^{-T} \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}^{\text{T}}$ for any number of inputs, and $\{\mathcal{B}_k^{\mathcal{S}}\}_{k=1}^{r}$ is a basis for the linear span of the rows of $\Psi^{\text{MISO}}$. As can be seen from (6.36), the first rows of $\Psi_e^{\text{MISO}}$ are the same as for $\Psi^{\text{MISO}}$ and the first $r$ basis functions can therefore be taken the same (with a zero in the last column). To accommodate for the extra rows, $n_e$ extra basis functions $\{\mathcal{B}_k^{\mathcal{S}}\}_{k=r+1}^{r+n_e}$ are needed. Thus, $\{\mathcal{B}_k^{\mathcal{S}}\}_{k=1}^{r+n_e}$ is a basis for the linear span of $\Psi_e^{\text{MISO}}$. We see that the variance of $G_1(\hat{\theta}_N^e)$ is larger than $G_1(\hat{\theta}_N)$ since

$$\text{AsCov } G_1(\hat{\theta}_N^e) = \text{AsCov } G_1(\hat{\theta}_N) + L^{\text{T}} \sum_{k=r+1}^{r+n_e} \mathcal{B}_k^{\mathcal{S}}(z_o)^* \mathcal{B}_k^{\mathcal{S}}(z_o)\, L,$$

and $L^{\text{T}} \mathcal{B}_k^{\mathcal{S}}(z_o)^* \mathcal{B}_k^{\mathcal{S}}(z_o) L \geq 0$ is positive semidefinite for every $k$.

Every additional input of the MISO system corresponds to addition of rows to $\Psi$. The increase is strictly positive provided that the explicit conditions in Corollary 6.3.1 hold.

Every additional output of the SIMO system corresponds to the addition of one more column to $\Psi$. However, the benefit of the additional columns is reduced by the additional rows arising from the additional parameters that need to be estimated, cf. Corollary 6.2.2 and the preceding discussion. When the number of additional parameters has reached $n_1$ or if $e_1(t)$ is orthogonal to $e_j(t)$ conditioned on $u_{j\setminus 1}(t)$ the benefit vanishes completely. The following examples clarify this point.

**Example 6.3.1.** *We consider the same example as in Section 6.1 for three cases of model orders of the second model, $n_2 = 0, 1, 2$. These cases correspond to $\Psi$ given by the first $n_2 + 1$ rows of*

$$\Psi_e^{SIMO}(q) = \begin{bmatrix} q^{-1} & -\sqrt{1-\beta^2}/\beta q^{-1} \\ 0 & 1/\beta q^{-1} \\ 0 & 1/\beta q^{-2} \end{bmatrix},$$

*respectively. When only $y_1$ is used ($\Psi = q^{-1}$) :*

$$\operatorname{AsVar} \hat{\theta}_{1,1} = \langle \Psi, \Psi \rangle^{-1} = 1.$$

*When $n_2 = 0$, the second measurement gives a benefit determined by how strong the correlation is between the two noise sources:*

$$\operatorname{AsVar} \hat{\theta}_{1,1} = \langle \Psi, \Psi \rangle^{-1} = (1 + (1-\beta^2)/\beta^2)^{-1} = \beta^2.$$

*However, already if we have to estimate one parameter in $G_2$ the benefit vanishes completely, i.e., for $n_2 = 1$:*

$$\operatorname{AsCov} \hat{\theta} = \langle \Psi, \Psi \rangle^{-1} = \begin{bmatrix} 1 & \sqrt{1-\beta^2} \\ \sqrt{1-\beta^2} & 1 \end{bmatrix}.$$

*The third case, $n_2 = 2$, corresponds to the example in Section 6.1, which shows that the first measurement $y_1$ improves the estimate of $\theta_{2,2}$ (compared to only estimating $\hat{G}_2$ using $y_2$):*

$$\operatorname{AsVar} \hat{\theta}_{2,2} = \lambda_{2|1} = \beta^2.$$

**Example 6.3.2.** *We consider the corresponding MISO system with unit variance noise $e(t)$ and $u(t)$ instead having the same spectral factor*

$$\Sigma_{CH} = \begin{bmatrix} 1 & 0 \\ \sqrt{1-\beta^2} & \beta \end{bmatrix}.$$

*for $\beta \in (0, 1]$. We now study the impact of the second input $u_2(t)$ when*

$$G_1(q) = \theta_{1,1} q^{-1} + \theta_{1,2} q^{-2}, \quad G_2(q) = \sum_{k=1}^{n_2} \theta_{2,k} q^{-k}$$

*and $n_2 = 0, 1$. These two cases correspond to $\Psi$ given by the first $n_2 + 2$ rows of*

$$\Psi_e^{MISO}(q) = \begin{bmatrix} q^{-1} & 0 \\ q^{-2} & 0 \\ \sqrt{1-\beta^2} q^{-1} & \beta q^{-1} \end{bmatrix},$$

*respectively. When only $u_1$ is used or $G_2$ is known ($\Psi = \begin{bmatrix} q^{-1} & q^{-2} \end{bmatrix}^{\mathrm{T}}$):*

$$\operatorname{AsCov} \hat{\theta}_1 = \langle \Psi, \Psi \rangle^{-1} = I.$$

*When $n_2 = 1$, the variance of $\hat{\theta}_{1,1}$ is increased depending on the correlation between the two inputs:*

$$\text{AsVar}\,\hat{\theta} = \frac{1}{\beta^2}\begin{bmatrix} 1 & 0 & -\sqrt{1-\beta^2} \\ 0 & \beta^2 & 0 \\ -\sqrt{1-\beta^2} & 0 & 1 \end{bmatrix}.$$

*Also notice that the asymptotic covariance of $\begin{bmatrix}\hat{\theta}_{1,1} & \hat{\theta}_{2,1}\end{bmatrix}^{\mathrm{T}}$ is given by $\Sigma^{-1}$, the inverse of the covariance matrix of $u(t)$ and that* AsVar $\hat{\theta}_{2,2} = \sigma_1^{-1}$. *As $\beta$ goes to zero the variance of $\begin{bmatrix}\hat{\theta}_{1,1} & \hat{\theta}_{2,1}\end{bmatrix}^{\mathrm{T}}$ increases and at $\beta = 0$ the two inputs are perfectly correlated and we loose identifiability.*

## 6.4 Effect of input spectrum

In this section we will see how a non white input spectrum changes the results of Section 6.2. Using white noise filtered through an AR-filter as input, we may use the developed results to show where in the frequency range the benefit of the correlation structure is focused. An alternative approach, when a non-white input spectrum is used, is to change the basis functions as discussed in Remark 6.1.1. However, the effect of the input filter is internalized in the basis functions and it is hard to distinguish the effect of the input filter. We will instead use FIR basis functions for the SIMO system, which are not orthogonal with respect to the inner product induced by the input spectrum, cf. Remark 6.1.1. We let the input filter be given by

$$u(t) = \frac{1}{A(q)}w(t) \tag{6.37}$$

where $w(t)$ is a white noise sequence with variance $\sigma_w^2$ and the order $n_a$ of $A$ is less than the order of $G_1$, i.e., $n_a \leq n_1$. In this case, following the derivations of Theorem 6.2.1, it can be shown that

$$\text{AsVar}\,\hat{G}_i = \sum_{k=1}^{n_i} \text{AsVar}\,\hat{\theta}_{i,k}|\mathcal{B}_k(e^{j\omega_0})|^2 \tag{6.38}$$

where

$$\text{AsVar}\,\hat{\theta}_{i,k} = \frac{\lambda_{i|\chi_k - 1}}{\Phi_u(\omega_0)}$$

and the basis functions $\mathcal{B}_k$ have changed due to the input filter. The solutions boil down to finding explicit basis functions $\mathcal{B}_k$ for the case when

$$\text{Span}\left\{\frac{\Gamma_n}{A(q)}\right\} = \text{Span}\left\{\frac{q^{-1}}{A(q)}, \frac{q^{-2}}{A(q)}, \ldots, \frac{q^{-n}}{A(q)}\right\}$$

where $A(q) = \prod_{k=1}^{n_a}(1-\xi_k q^{-1})$, $|\xi_k| < 1$ for some set of specified poles $\{\xi_1, \ldots, \xi_{n_a}\}$ and where $n \geq n_a$. As discussed in Section 2.2, it then holds that

$$\text{Span}\left\{\frac{\mathcal{I}_n}{A(q)}\right\} = \text{Span}\{\mathcal{B}_1, \ldots, \mathcal{B}_n\}$$

where $\{\mathcal{B}_k\}$ are the Takenaka-Malmquist functions given by

$$\begin{aligned}\mathcal{B}_k(q) &:= \frac{\sqrt{1-|\xi_k|^2}}{q-\xi_k}\,\phi_{k-1}(q),\ k = 1, \ldots, n \\ \phi_k(q) &:= \prod_{i=1}^{k}\frac{1-\overline{\xi_i}q}{q-\xi_i},\ \ \phi_0(q) := 1\end{aligned} \tag{6.39}$$

and with $\xi_k = 0$ for $k = n_a + 1, \ldots, n$. We summarize the result in the following theorem:

**Theorem 6.4.1.** *Let the same assumptions as in Theorem 6.2.1 hold. Additionally the input $u(t)$ is generated by an AR-filter as in* (6.37)*. Then for any frequency $\omega_0$ it holds that*

$$\text{AsVar}\,\hat{G}_i = \frac{1}{\Phi_u(\omega_0)}\left(\lambda_i\sum_{k=1}^{n_a}\frac{1-|\xi_k|^2}{|e^{j\omega_0}-\xi_k|^2} + \lambda_i(n_1-n_a) + \sum_{j=2}^{i}\lambda_{i|j-1}(n_j - n_{j-1})\right) \tag{6.40}$$

*Proof.* The proof follows from (6.38) with the basis functions given by the Takenaka-Malmquist functions and using that $\phi_{k-1}(q)$ is all-pass, and for $k > n_a$, also $\mathcal{B}_k(q)$ is all-pass. This means that $|\mathcal{B}_k(\mathrm{e}^{\mathrm{j}\omega})|^2 = 1$ for all $k > n_a$. □

**Remark 6.4.1.** *The second sum in* (6.40) *is where the benefit from the correlation structure of the noise at the other sensors enters through $\lambda_{i|j-1}$. This contribution is weighted by $1/\Phi_u(\omega)$. The benefit thus enters mainly where $\Phi_u(\omega)$ is small. The first sum gives a variance contribution that is less focused around frequencies close to the poles ($|e^{j\omega_0} - \xi_k|^2$ is canceled by $\Phi_u(\omega)$). This contribution is not reduced by correlation between noise sources. Shaping the input thus gives a weaker effect on the asymptotic variance than what is suggested by the asymptotic in model order result* (1.4)*, and what would be the case if there would be no correlation between noise sources (replacing $\lambda_{i|j-1}$ by $\lambda_i$ in* (6.40)*).*

For the example in Section 6.1 for filtered input with $n_1 = 2$ , $n_2 = 3$ and $n_a = 1$, (6.40) simplifies to

$$\text{AsVar}\,\hat{G}_2 = \frac{\lambda_2}{\sigma^2} + \frac{\lambda_2}{\Phi_u(\omega_0)} + \frac{\lambda_{2|1}}{\Phi_u(\omega_0)}. \tag{6.41}$$

In Figure 6.6 the variance of $\hat{G}_2$ for an input filtered with $A(q) = 1 - 0.8q^{-1}$ is presented. The filter is of low-pass type and thus gives high input power at low

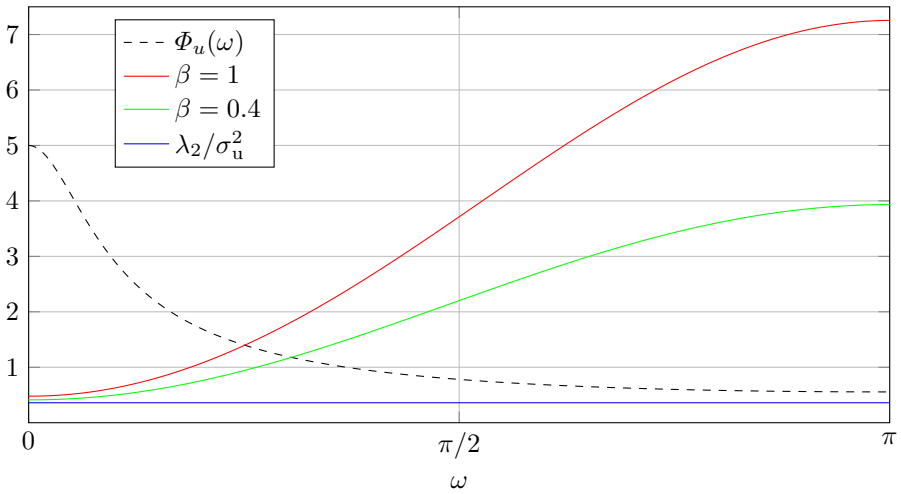


Figure 6.6: Asymptotic variance of $\hat{G}_3(e^{j\omega}, \hat{\theta}_2)$ for $\beta = 1$ and $\beta = 0.4$. Also shown is $\lambda_2/\sigma^2$, the first term of AsVar $\hat{G}_3(e^{j\omega}, \hat{\theta}_2)$ in (6.41).

frequencies which results in low variance at those frequencies. Correlation between noise sources decreases the variance mainly where $\Phi_u(\omega)$ is small, i.e., at higher frequencies.

## 6.5 Optimal correlation structure

In this section we characterize the optimal correlation structure, in order to minimize the total variance. In the previous section, we have seen that not estimating a basis function $k$ leads to a reduction in variance of the other parameters related to basis function $k$. In this section, we will find the optimal correlation structure when the parametrization of the system is such that $n_1 + 1 = n_2 = \ldots = n_m$, i.e., the first module has one parameter less than the others. Let $\tilde{\theta} := [\theta_{2,n_2}, \theta_{3,n_2}, \ldots, \theta_{m,n_2}]^{\mathrm{T}}$, i.e., the sub vector of parameters related to basis function $\mathcal{B}_{n_2}(q)$. Assume the SIMO structure given by (6.1) and let the input be white noise with unit variance. Recalling Theorem 6.2.1, the covariance matrix of $\hat{\theta}$ is given by $\Lambda_{2:m|1}$. In particular, the variance of the entries of $\tilde{\theta}$ is

$$\text{AsVar}\, \hat{\theta}_{k,n_2} = \frac{\lambda_{k|1}}{\sigma^2}, \quad k = 2, \ldots, m. \tag{6.42}$$

As before, $\lambda_{k|1}$ is the non-estimable part of $e_k(t)$ given $e_1(t)$.

Recalling Theorem 6.2.1, the covariance matrix of $\tilde{\theta}$ is given by $\frac{1}{\sigma^2}\Lambda_{2:m|2}$. The

total variance of $\hat{\tilde{\theta}}$ is defined as

$$\operatorname{tvar} \tilde{\theta} := \sum_{i=2}^{m} \operatorname{AsVar} \hat{\theta}_{i,n_2} = \operatorname{Tr} \frac{1}{\sigma^2} \Lambda_{2:m|1} \,.$$

We are interested in understanding how the correlation of $e_1(t)$ with $e_2(t)$, ..., $e_m(t)$, i.e., $\Lambda_{12}$ ..., $\Lambda_{1m}$, should be in order to obtain the minimum value of the above total variance. This problem can be expressed as follows:

$$\begin{aligned} \underset{\Lambda_{12} \ \ldots, \Lambda_{1m}}{\text{minimize}} \quad & \operatorname{Tr} \Lambda_{2:m|1} \\ \text{s.t.} \quad & \Lambda \geq 0, \end{aligned} \tag{6.43}$$

where the constraint on $\Lambda$ implies that not all choices of the entries $\Lambda_{1i}$ are allowed. Directly characterizing the noise structure using this formulation of the problem appears to be hard. Therefore, it turns out convenient to introduce an upper triangular Cholesky factorization of $\Lambda$, namely define $B$ *upper triangular* such that $\Lambda = BB^{\mathrm{T}}$. Note that

1. the rows of $B$, $b_i^{\mathrm{T}}$, $i = 1, \ldots, m$, are such that $\|b_i\|^2 = \lambda_i$;
2. $\mathbf{E}\{e_1 e_i\} = \Lambda_{1i} = b_1^{\mathrm{T}} b_i$;
3. there always exists an orthonormal matrix $Q$ such that $B = \Lambda_{CH} Q$, where $\Lambda_{CH}$ is the previously defined lower triangular Cholesky factor.

**Lemma 6.5.1.** *Let*

$$B = \begin{bmatrix} \eta & p^{\mathrm{T}} \\ 0 & M \end{bmatrix}, \; M \in \mathbb{R}^{m-1 \times m-1}, \; p \in \mathbb{R}^{m-1 \times 1}, \; n \in \mathbb{R} \,;$$

*then*

$$\Lambda_{2:m|1} = M(I - \frac{1}{\lambda_1}) p p^{\mathrm{T}} M^{\mathrm{T}} \,. \tag{6.44}$$

*Proof.* See Appendix 6.D. □

Using the previous lemma we reformulate (6.43); keeping $M$ fixed and letting $p$ vary, we have

$$\begin{aligned} \underset{b_1}{\text{maximize}} \quad & \operatorname{Tr} \frac{1}{\lambda_1} M p p^{\mathrm{T}} M^{\mathrm{T}} \\ \text{s.t.} \quad & \|b_1\|^2 = \lambda_1 \,, \end{aligned} \tag{6.45}$$

with $b_1^{\mathrm{T}} = [\eta \quad p^{\mathrm{T}}]$. Note that the constraint $\Lambda \geq 0$ is automatically satisfied. Let us define $v_1, \ldots, v_{m-1}$ as the right-singular vectors of $M$, namely the columns of the matrix $V$ in the singular value decomposition (SVD) $M = USV^{\mathrm{T}}$. The following result provides the structure of $B$ that solves (6.43).

**Theorem 6.5.1.** *Let the input be white noise. Let $n_1 + 1 = n_2 = \ldots = n_m$. Then* (6.43) *is solved by an upper triangular Cholesky factor $B$ such that its first row is*

$$b_1^{*T} = \lambda_1 [0 \quad v_1^{\mathrm{T}}] \,. \tag{6.46}$$

*Proof.* Observe the following facts:

1. Tr $\frac{1}{\lambda_1} M p p^{\mathrm{T}} M^{\mathrm{T}} = \frac{1}{\lambda_1} p^{\mathrm{T}} M^{\mathrm{T}} M p = \frac{1}{\lambda_1} \|Mp\|^2$;
2. since $b_1^{\mathrm{T}} = [\eta \quad p^{\mathrm{T}}]$, it is clear that a candidate solution to (6.45) is of the form $b_1^{*T} = [0 \quad p^{*T}]$.

It follows that Problem (6.45) can be written as

$$\begin{aligned} \underset{b_1}{\text{maximize}} \quad & \|Mp\|^2 \\ \text{s.t.} \quad & \|p\|^2 = \lambda_1 \,, \end{aligned} \tag{6.47}$$

whose solution is known to be (a rescaling of) the first right singular vector of $M$, namely $v_1$. Hence $b_1^{*T} = \lambda_1 [0 \quad v_1^{\mathrm{T}}]$. □

**Remark 6.5.1.** *As has been pointed out in Section 6.1, $\Lambda$ is required to be positive definite. Thus, the ideal solution provided by Theorem 6.5.1 is not applicable in practice, where one should expect that $\eta$, the first entry of $b_1$, is always nonzero. In this case, the result of Theorem 6.5.1 can be easily adapted, leading to* $b_1^{*T} = \left[\eta \quad \sqrt{1 - \frac{\eta^2}{\lambda_1}} v_1^{\mathrm{T}}\right]$.

## 6.6 Numerical examples

In this section, we illustrate the results derived in the previous sections through three sets of numerical Monte Carlo simulations.

### Effect of model order

Corollary 6.2.2 is illustrated in Figure 6.7, where the following systems are identified using $N = 500$ input-output measurements:

$$\begin{aligned} G_i &= \tilde{\Gamma}_i \theta_i, \quad i = 1, 2, 3 & \tilde{\Gamma}_i(q) &= F(q)^{-1} \Gamma_i(q), \\ \Gamma_i(q) &= \left[\mathcal{B}_1(q), \ldots, \mathcal{B}_{n_i}(q)\right], & \mathcal{B}_k(q) &= q^{-k}, \end{aligned} \tag{6.48}$$

with

$$\begin{aligned} F(q) &= \frac{1}{1 - 0.8q^{-1}}, & \theta_1^0 &= \begin{bmatrix} 1 & 0.5 & 0.7 \end{bmatrix}^{\mathrm{T}}, \\ \theta_2^0 &= \begin{bmatrix} 1 & -1 & 2 \end{bmatrix}^{\mathrm{T}}, & \theta_3^0 &= \begin{bmatrix} 1 & 1 & 2 & 1 & 1 \end{bmatrix}^{\mathrm{T}}. \end{aligned}$$

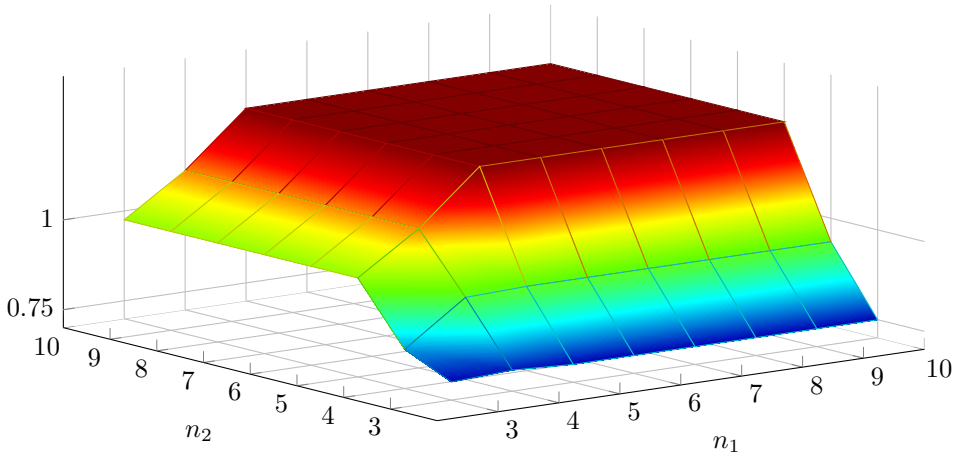


Figure 6.7: Sample variance of $G_3(e^{j\omega}, \hat{\theta}_3)$ as a function of the number of estimated parameters of $G_1$ and $G_2$.

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

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

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

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

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

The variance of $G_3(e^{j\omega}, \hat{\theta}_3)$ increases with increasing $n_i$, $i = 1, 2$, but only up to the point where $n_i = n_3 = 5$. After that, any increase in $n_1$ or $n_2$ does not increase the variance of $G_3(e^{j\omega}, \hat{\theta}_3)$, as can be seen in Figure 6.7. The behavior can be explained by Corollary 6.2.2: when $n_3 \geq n_1, n_2$, $G_3$ is the last block, having the highest number of parameters, and any increase in $n_1, n_2$ increases the variance of $G_3$. When for example $n_1 \geq n_3$, the blocks should be reordered so that $G_3$ comes before $G_1$. In this case, when $n_1$ increases the first condition of Corollary 6.2.2 does not hold and hence the variance of $G_3(e^{j\omega}, \hat{\theta}_3)$ does not increase further.

### Effect of removing one parameter

In the second set of numerical experiments, we simulate the same type of system as in Section 6.1. We let $\beta$ vary in the interval $[0.001, 1]$. For each $\beta$, we generate $MC = 2000$ Monte Carlo experiments, where in each of them we collect $N = 500$ input-output samples. At the $i$-th Monte Carlo run, we generate new trajectories for the input and the noise and we compute $\hat{\theta}_i$ as in (6.8). The sample covariance matrix, for each $\beta$, is computed as

$$\operatorname{Cov} \hat{\theta}_{\mathrm{s}} = \frac{1}{MC} \sum_{i=1}^{MC} (\hat{\theta}_i - \theta)(\hat{\theta}_i - \theta)^{\mathrm{T}}.$$

Figure 6.8 shows that the variance of $\hat{\theta}_{21}$ is always close to one, no matter what the value of $\beta$ is. It also shows that the variance of the estimate $\hat{\theta}_{22}$ behaves as $\beta^2$. In particular, when $\beta$ approaches 0 (i.e., almost perfectly correlated noise processes), the variance of such estimate tends to 0. All of the observations are as predicted by Corollary 6.2.1 and the example in Section 6.1.

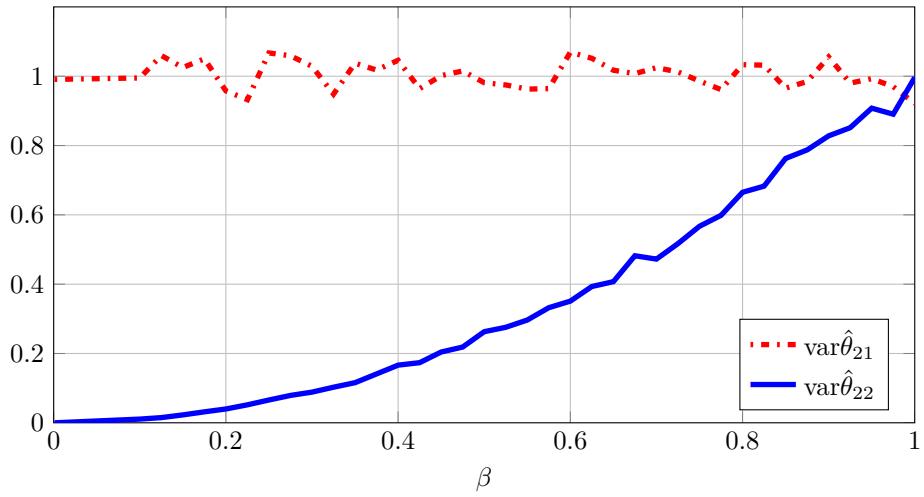


Figure 6.8: Sample variance of the parameters of the first module (as functions of $\beta$) when the second module has one parameter.

### Optimal correlation structure

A third simulation experiment is performed in order to illustrate Theorem 6.5.1. We consider a system with $m = 3$ outputs; the modules are

$$G_1(q) = 0.3q^{-1}, \qquad G_2(q) = 0.8q^{-1} - 0.4q^{-2},$$
$$G_3(q) = 0.1q^{-1} + 0.2q^{-2},$$

so that

$$\theta = [\theta_{11}\,\theta_{21}\,\theta_{31}\,\theta_{22}\,\theta_{32}]^{\mathrm{T}} = [0.3\;\; 0.8\;\; 0.1\;\; -0.4\;\; 0.2]^{\mathrm{T}}\,.$$

The noise process is generated by the following upper triangular Cholesky factor:

$$R = \begin{bmatrix} \varepsilon & \sqrt{1-\varepsilon^2}\cos\alpha & \sqrt{1-\varepsilon^2}\sin\alpha \\ 0 & 0.8 & 0.6 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \varepsilon & p^{\mathrm{T}} \\ 0 & M \end{bmatrix},$$

where $\varepsilon = 0.1$ and $\alpha \in [0,\, \pi]$ is a parameter tuning the correlation of $e_1(t)$ with $e_2(t)$ and $e_3(t)$. The purpose of this experiment is to show that, when $\alpha$ is such that $p = [\sqrt{1-\varepsilon^2}\cos\alpha \quad \sqrt{1-\varepsilon^2}\sin\alpha]^{\mathrm{T}}$ is aligned with the first right-singular vector $v_1$ of $M$, then the total variance of the estimate of the sub-vector $\tilde{\theta} = [\theta_{12}\,\theta_{22}]^{\mathrm{T}}$ is minimized. In the case under analysis, $v_1 = [0.447\;\; 0.894]^{\mathrm{T}} = [\cos\alpha_0\;\; \sin\alpha_0]^{\mathrm{T}}$, with $\alpha_0 = 1.11$. We let $\alpha$ take values in $[0,\, \pi]$ and for each $\alpha$ we generate $MC = 2000$ Monte Carlo runs of $N = 500$ input-output samples each. We compute the sample total variance of $\tilde{\theta}$ as

$$\text{tvar}\;\hat{\tilde{\theta}}_{\mathrm{S}} = \frac{1}{MC}\sum_{i=1}^{MC}\Big((\hat{\theta}_{22,i} - \theta_{22})^2 + (\hat{\theta}_{32,i} - \theta_{32})^2\Big),$$

where $\hat{\theta}_{22,i}$ and $\hat{\theta}_{32,i}$ are the estimates obtained at the $i$-th Monte Carlo run.

The results of the experiment are reported in Figure 6.9. It can be seen that the minimum total variance of the estimate of $\hat{\theta}$ is attained for values close to $\alpha_0$ (approximations are due to low resolution of the grid of values of $\alpha$). An interesting observation regards the value of $\alpha$ for which the total variance is maximized: this happens when $\alpha = 2.68$, which yields the second right-singular vector of $M$, namely $v_2 = [-0.894\;\; 0.447]^{\mathrm{T}}$.

## 6.7 Conclusions

In this chapter, we have examined how the estimation accuracy of a linear SIMO model depends on the correlation structure of the noise, model structure and model order. A formula for the asymptotic covariance of the frequency response function estimate and the model parameters has been developed for the case of temporally white, but possibly spatially correlated additive noise. It has been shown that when parts of the noise can be linearly estimated from measurements of other blocks with less estimated parameters, the variance decreases. The expressions reveal how the order of the different blocks and the correlation of the noise affect the variance of one block. In particular, it has been shown that the variance of the block of interest levels off when the number of estimated parameters in another block reaches the number of estimated parameters of the block of interest. The optimal correlation structure for the noise was determined for the case when one block has one parameter less than the other blocks.

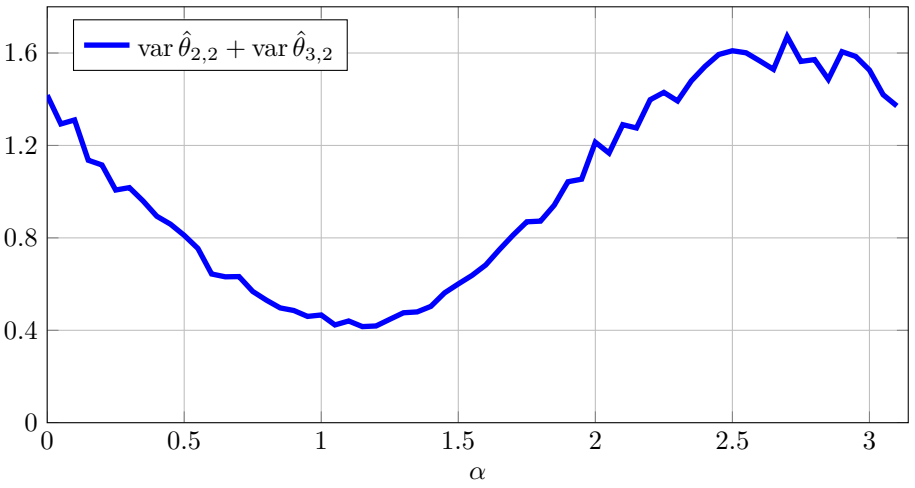


Figure 6.9: Total sample variance of the parameter vector $\tilde{\theta}$ as function of $\alpha$.

## 6.A Proof of Lemma 6.1.2

Let $v(t) = \Lambda_{CH}^{-1} e(t)$ for some real valued random variable $e(t)$ ($\Lambda_{CH}^{-1}$ exists and is unique for $\Lambda > 0$ (Horn and Johnson, 1990)). Then Cov $v(t) = I$. Similarly $e(t) = \Lambda_{CH} v(t)$. The set $\{v_1(t), \ldots, v_j(t)\}$ is a function of $e_1(t), \ldots, e_j(t)$ only and vice versa, for all $1 \leq j \leq m$. Thus, if $e_1(t), \ldots, e_j(t)$ are known, then also $\{v_1(t), \ldots, v_j(t)\}$ are known, but nothing is known about $\{v_{j+1}(t), \ldots, v_m(t)\}$. Thus, for $j < i$ the best linear estimator of $e_i(t)$ given $e_1(t), \ldots, e_j(t)$, is

$$\hat{e}_{i|j}(t) = \sum_{k=1}^{j} \gamma_{ik} v_k(t), \tag{6.49}$$

and

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

has variance

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

For the last part of the lemma, we realize that the dependence of $\hat{e}_{i|j}(t)$ on $e_j(t)$ in Equation (6.49) is given by $\gamma_{ij}/\gamma_{jj}$ (since $v_1(t), \ldots, v_{j-1}(t)$ do not depend on $e_j(t)$). Hence $\hat{e}_{i|j}(t)$ depends on $e_j(t)$ if and only if $\gamma_{ij} \neq 0$.

## 6.B Proof of Theorem 6.2.1

Before giving the proof of Theorem 6.2.1 we need the following auxiliary lemma.

**Lemma 6.B.1.** *Let $\Lambda > 0$ and real and its Cholesky factor $\Lambda_{CH}$ be partitioned according to $e_{1:\chi_k-1}$ and $e_{\chi_k:m}$,*

$$\Lambda = \begin{bmatrix} \Lambda_1 & \Lambda_{12} \\ \Lambda_{21} & \Lambda_2 \end{bmatrix}, \qquad \Lambda_{CH} = \begin{bmatrix} (\Lambda_{CH})_1 & 0 \\ (\Lambda_{CH})_{21} & (\Lambda_{CH})_2 \end{bmatrix}.$$

*Then*

$$\Lambda_{\chi_k:m|\chi_k-1} = (\Lambda_{CH})_2 (\Lambda_{CH})_2^{\mathrm{T}}.$$

*Proof.* By the derivations of Lemma 6.A, for some $v(t)$ with $\operatorname{Cov} v(t) = I$, $e(t) = \Lambda_{CH} v(t)$ and $v_{1:\chi-1}(t)$ are known since $e_{1:\chi-1}(t)$ are known. Furthermore $\hat{e}_{\chi_k:m|\chi_k-1}(t) = (\Lambda_{CH})_{21} v_{1:\chi-1}(t)$, which implies $e_{\chi_k:m|\chi_k-1}(t) - \hat{e}_{\chi_k:m|\chi_k-1}(t) = (\Lambda_{CH})_2 v_{\chi_k:m}(t)$ and the results follows since $\operatorname{Cov} v_{\chi_k:m}(t) = I$. □

The asymptotic variance is given by (6.10) with

$$\Psi(q) = \tilde{\Psi}(q) \Lambda_{CH}^{-T}.$$

Let $n = n_1 + \cdots + n_m$. From the upper triangular structure of $\Lambda_{CH}^{-T}$ and $n_1 \leq n_2 \leq \ldots \leq n_m$, an orthonormal basis for $\mathcal{S}_\Psi$, the subspace spanned by the rows of $\Psi$, is given by

$$\begin{aligned}
\mathcal{B}_k^{\mathcal{S}}(e^{j\omega}) &:= \begin{bmatrix} \mathcal{B}_k & 0 & \ldots & 0 \end{bmatrix}, && k{=}1, \ldots, n_1 \\
\mathcal{B}_k^{\mathcal{S}}(e^{j\omega}) &:= \begin{bmatrix} 0 & \mathcal{B}_{k-n_1} & 0 & \ldots \end{bmatrix}, && k{=}n_1+1, \ldots, n_2 \\
&\vdots \\
\mathcal{B}_k^{\mathcal{S}}(e^{j\omega}) &:= \begin{bmatrix} \ldots & 0 & \mathcal{B}_{k-n+n_m} \end{bmatrix}, && k{=}n-n_m{+}1, \ldots, n.
\end{aligned} \tag{6.50}$$

First note that

$$\frac{\partial G}{\partial \theta} = \Psi \Lambda_{CH}^{\mathrm{T}}.$$

Then, using Theorem 2.3.2,

$$\sigma^2 \operatorname{AsVar} \hat{G} = \Lambda_{CH} \sum_{k=1}^{n} \mathcal{B}_k^{\mathcal{S}}(e^{j\omega_0})^* \mathcal{B}_k^{\mathcal{S}}(e^{j\omega_0}) \Lambda_{CH}^{\mathrm{T}}.$$

Sorting the sum with respect to the basis functions $\mathcal{B}_k(e^{j\omega_0})$, we get

$$\sigma^2 \operatorname{AsCov} \hat{G}_i = \Lambda_{CH} \sum_{k=1}^{n_m} |\mathcal{B}_k(e^{j\omega_0})|^2 \begin{bmatrix} 0_{\chi_k-1} & 0 \\ 0 & I \end{bmatrix} \Lambda_{CH}^{\mathrm{T}}.$$

Using Lemma 6.B.1

$$\text{AsCov}\,\hat{G} = \frac{1}{\sigma^2}\sum_{k=1}^{n_m}\begin{bmatrix}0_{\chi_k-1} & 0\\ 0 & \Lambda_{\chi_k:m|\chi_k-1}\end{bmatrix}|\mathcal{B}_k(e^{j\omega_0})|^2.$$

Thus (6.23) follows. We now show the first part of the theorem, that

$$\text{AsCov}\,\hat{\tilde{\theta}} = \frac{1}{\sigma^2}\text{diag}\left\{\Lambda_{1:m}, \Lambda_{\chi_2:m|\chi_2-1}, \ldots, \Lambda_{\chi_{n_m}:m|\chi_{n_m}-1}\right\}.$$

The covariance of $\hat{G}$ can be expressed as

$$\text{AsCov}\,\hat{G} = T\text{AsCov}\,\hat{\tilde{\theta}}\,T^* \tag{6.51}$$

where

$$T = \begin{bmatrix}\mathcal{B}_1 I(1) & \mathcal{B}_2 I(2)\ldots & \mathcal{B}_{n_m} I(n_m)\end{bmatrix},$$
$$I(k) = \begin{bmatrix}\mathbf{0}\\ I_{m-\chi_k+1}\end{bmatrix} \in \mathbb{R}^{m\times(m-\chi_k+1)}.$$

However, (6.51) equals (6.23) for all $\omega$ and the theorem follows.

## 6.C Proof of Corollary 6.2.2

To prove Corollary 6.2.2 we will use (6.26). First, we make the assumption that $j$ is the last module that has $n_j$ parameters. This assumption is made for convenience since reordering all modules with $n_j$ estimated parameters does not change (6.26). First of all, we see that if

$$n_j \geq n_i,$$

then (6.26) does not increase when $n_j$ increases. If instead

$$n_j < n_i,$$

the increase in variance is given by

$$\gamma_{ij}^2|\mathcal{B}_{n_j+1}(e^{j\omega_0})|^2,$$

which is non-zero iff $\gamma_{ij} \neq 0$ and $|\mathcal{B}_{n_j+1}(e^{j\omega_0})|^2 \neq 0$. From Lemma 6.1.1 the theorem follows.

## 6.D Proof of Lemma 6.5.1

The inverse of $B$ is

$$B^{-1} = \begin{bmatrix}\eta^{-1} & -q^{\mathrm{T}}\\ 0 & M^{-1}\end{bmatrix}, \quad q^{\mathrm{T}} := \eta^{-1}p^{\mathrm{T}}M^{-1},$$

so that

$$\Lambda^{-1} = B^{-T}B^{-1} = \begin{bmatrix} \eta^{-2} & -q^{\mathrm{T}}\eta^{-1} \\ -q\eta^{-1} & M^{-T}M^{-1} + qq^{\mathrm{T}} \end{bmatrix}.$$

Hence, using the Sherman–Morrison formula (Bartlett, 1951)

$$\begin{aligned} \Lambda_{2:m|1} &= \left(M^{-T}M^{-1} + qq^{\mathrm{T}}\right)^{-1} \\ &= MM^{\mathrm{T}} - \frac{1}{k}MM^{\mathrm{T}}qq^{\mathrm{T}}MM^{\mathrm{T}} \\ &= MM^{\mathrm{T}} - \frac{1}{k}\frac{Mpp^{\mathrm{T}}M^{\mathrm{T}}}{\eta^2}, \end{aligned}$$

where

$$\begin{aligned} k &= 1 + q^{\mathrm{T}}MM^{\mathrm{T}}q = 1 + \frac{p^{\mathrm{T}}p}{\eta^2} \\ &= \frac{\eta^2 + p^{\mathrm{T}}p}{\eta^2} = \frac{\|b_1\|_2^2}{\eta^2} \\ &= \frac{\lambda_1}{\eta^2}, \end{aligned}$$

so (6.44) follows.

# Chapter 7

# Conclusions

In this thesis we have analyzed the variance of the parameter estimates, frequency response functions and zeros of estimated modules in certain dynamic networks. The underlying identification method considered in the thesis is PEM. The results of this thesis are in a sense an extension of some of the results of Mårtensson (2007) to dynamic networks. The chapter on SISO linear systems, Chapter 3, served as a springboard to the results derived for dynamic networks. For a dynamic network we have analyzed three different techniques which can reduce the variance of an estimated module in the network. The techniques are additional input signals for increased excitation (see the parallel cascade structure of Chapter 5), additional measurements for increased accuracy (the additional measurements carry information about the module of interest, cf. Chapter 4 and the multiple sensor structure of Chapter 4), and additional measurements for noise cancellation (the additional measurements contain no information about the module of interest but information about the noise, cf. Chapter 6). The contribution of each chapter is listed below, after which we provide an outlook on possible future work.

## 7.1 Summary

### Chapter 3: SISO models

The main results in this chapter are the reparametrization formulas for the asymptotic covariance of functions of the estimated system parameters. In particular we demonstrated that one can use the experimental conditions to make the asymptotic variance independent of model order and model structure in some cases. We have also used these expressions to derive novel model structure independent upper bounds of the asymptotic covariance, in particular for a number of commonly estimated quantities such as system zeros, gains and impulse response coefficients. We have shown that these bounds are significantly less conservative as compared to the variance expressions that result from using the (asymptotic in model order) variance formulae for frequency function estimation in Ljung (1985).

### Chapter 4: Cascade models

The variance of the first of a set of modules connected in a cascade structure has been analyzed. The main contribution was the characterization of the variance of a zero close to the unit circle. It was shown that a variance reduction was possible and concentrated around the unit-circle-zero. The variance reduction capability of a sensor depends on the signal to noise ratio at the sensor and a distance metric. It was also proven that, for a unit-circle-zero and large model order, the variance at the corresponding frequency of the unit-circle zero is the same as if the other modules were completely known. Simulations on FIR systems illustrate the variance reduction for a unit-circle-zero, in the case of two and four modules, respectively, for different model orders.

### Chapter 5: Generalized parallel cascade models

We have examined the covariance of the parameter estimates of one module in a parallel cascade structure and a multi sensor structure. Upper bounds on the asymptotic covariance of the parameter estimates were derived when little assumptions were made on the remaining systems in the network, i.e., in contrast to the previous chapter, the model order of the system of interest was fixed, while the model order of every other module was large. For a measured disturbance, the covariance reduction potential was shown to be two folded. Previous knowledge about the additional sensors in the multi sensor structure is imperative for a variance reduction; in fact, without prior knowledge it was shown that there is no covariance reduction.

### Chapter 6: SIMO models

The effect of the noise correlation structure was examined in SIMO models. It was shown how the estimation accuracy depends on the correlation structure of the noise, model structure and model order. A formula for the asymptotic covariance of the frequency response function estimates and the model parameters has been developed for the case of temporally white, but possibly spatially correlated additive noise. It has been shown that when parts of the noise can be linearly estimated from measurements of other blocks with less estimated parameters, the variance decreases. The expressions reveal how the order of the different blocks and the correlation of the noise affects the variance of one block. In particular, it has been shown that the variance of the block of interest levels off when the number of estimated parameters in another block reaches the number of estimated parameters of the block of interest. The effect of the input spectrum was shown to have a less significant effect than expected. The optimal correlation structure for the noise was determined for the case when one block has one parameter less than the other blocks.

## 7.2 Future work

For some basic dynamic network structures, we have analyzed three different techniques which can reduce the variance of an estimated module in the network. Namely, the techniques are additional input signals for increased excitation, additional measurements for increased accuracy, and additional measurements for noise cancellation. There are some straightforward generalizations, for example, to have a temporally non-white noise spectra in Chapter 4 and Chapter 6. The main contribution of the thesis is the characterization of the variance reduction effect additional measurements and input signals can have in a dynamic network. However, the results are, in their present form, not very general. One challenge ahead is to extend these results to a broader class of dynamic networks, such as networks that may include feedback loops, and where process noise also may be present.

### Upper bounds

As seen in Chapter 3, the geometric approach is readily used to develop model independent upper bounds. With more complex network structures, it will be harder to derive covariance expressions that can give insights in the effect of, for example, input excitation and noise correlations. Therefore, proceeding as in Chapter 3, and deriving upper bounds might be one path forward towards obtaining meaningful insights when the network structure becomes more complex. A slight variation of this approach was the upper bounds derived for high order models in Chapter 5, which also may be a path forward. However, as seen in Chapter 5 for the multi sensor structure, assuming high model order for the modules of less interest led to the variance reduction potential shrinking to zero. It was also seen for SIMO models in Chapter 6 that the variance reduction was strongly linked to the model structures and model order of the modules.

### Under modeling

In this thesis and prediction error identification in general, we assume the existence of a true system and a true model order. This is hardly ever true in practical applications. From a pragmatic perspective, we want our identified model to be able to describe the true system accurate enough with some application of the model in mind. As we have framed our estimation problem, we are mainly interested in a subset of modules, maybe even only one module. From Van den Hof *et al.* (2013) it is known that to consistently estimate a module it is necessary to have model structures that are flexible enough. However, if we are willing to trade some bias to reduce the variance of the module estimates we could perhaps achieve a lower mean squared error. The methods applied in this work might provide some insight into how this trade-off should be done.

### Noise assumptions and feedback

As concluded in Chapter 6, noise correlation may have a significant effect on the covariance of estimates in the dynamic network. In a dynamic network, there are several scenarios that can justify correlation between sensor noise and process noise. Consider that we have a controller in the network that measures some internal variable in the network with sensor noise. Through the controller, this noise is fed back into the network and affects other internal variables. This effect can not be captured by (2.18), i.e., measurement noise is not allowed to affect internal variables. In order to capture this effect, we have to introduce correlation between process noise and measurement noise. However, for consistency in the methods mentioned in the introduction that take sensor noise into account, one assumption is precisely that process noise and measurement noise are uncorrelated. Feedback still provides us with some open questions regarding consistency. From an accuracy perspective, feedback is the component that is most strikingly missing in this thesis. All structures considered are open loop structures, with the exception being SISO systems in Chapter 3.

### Application to available methods

An interesting direction of research is to apply results of this thesis to the methods already available for networks with feedback loops. Perhaps, we could explain and compare the methods' performance and provide further guidance in how to choose measurements to include in the predictor. For example, the rough principle for predictor input selection in the indirect method of Dankers *et al.* (2013a) is to minimize the number of inputs in the predictor. The rationale is that this minimizes the number of estimated parameters. However, as is known for MISO models discussed in Chapter 6, both the number of estimated parameter of the other modules and the correlation structure between the input signals are important. It would be interesting to see if this also holds for input selection for the indirect method, and if it holds for input selection for the direct method as well.

# Bibliography


J.C. Aguero and G.C. Goodwin. 2006. On the optimality of open and closed loop experiments in system identification. In *Proceedings of 45th IEEE Conference on Decision and Control*, pages 163–168.

J.C. Agüero and G.C. Goodwin. 2007. Choosing between open-and closed-loop experiments in linear system identification. *IEEE Transactions on Automatic Control*, 52(8):1475–1480.

J.C. Agüero, C.R. Rojas, H Hjalmarsson, and G.C. Goodwin. 2012. Accuracy of linear multiple-input multiple-output (MIMO) models obtained by maximum likelihood estimation. *Automatica*, 48(4):632–637.

M. Ali, A. Ali, H. Abbas, and H. Werner. 2011a. Identification of Box-Jenkins models for parameter-varying spatially interconnected systems. In *Proceedings of the 2011 American Control Conference*, pages 145–150.

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

M. Ali, S.S. Chughtai, and H. Werner. 2009. Identification of spatially interconnected systems. In *Proceedings of the 48th IEEE Conference on Decision and Control, held jointly with the 28th Chinese Control Conference*, pages 7163–7168.

M. Ali, A.P. Popov, H. Werner, and H. Abbas. 2011c. Identification of distributed systems with identical subsystems. In *Proceedings of the 18th IFAC World Congress*, pages 5633–5638.

B.D.O. Anderson and M. Gevers. 1982. Identifiability of linear stochastic systems operating under linear feedback. *Automatica*, 18(2):195–213.

M.S. Bartlett. 1951. An inverse matrix adjustment arising in discriminant analysis. *The Annals of Mathematical Statistics*, 22(1):107–111.

A.S. Bazanella, M. Gevers, and L. Mišković. 2010. Closed-loop identification of MIMO systems: A new look at identifiability and experiment design. *European Journal of Control*, 16(3):228–239.

J. Benesty, S. Makino, and J. Chen. 2005. *Speech enhancement*. Springer.

N. Bertaux, P. Larzabal, C. Adnet, and E. Chaumette. 1999. A parameterized maximum likelihood method for multipaths channels estimation. In *Proceedings of the 2nd IEEE Workshop on Signal Processing Advances in Wireless Communications*, pages 391–394.

X. Bombois, A.J. den Dekker, M. Barenthin, and P.M.J. Van den Hof. 2009. Effect of model structure and signal-to-noise ratio on finite-time uncertainty bounding in prediction error identification. In *Proceedings of the 48th IEEE Conference on Decision and Control, held jointly with the 28th Chinese Control Conference*, pages 494–499.

X. Bombois, M. Gevers, and G. Scorletti. 2005. Open-loop versus closed-loop identification of Box-Jenkins models: a new variance analysis. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pages 3117–3122.

G. Bottegal and G. Picci. 2014. Analysis and identification of complex stochastic systems admitting a flocking structure. In *Proceedings of the 19th IFAC World Congress*.

G.E.P. Box and G.M. Jenkins. 1976. *Time series analysis: forecasting and control*. Holden-Day series in time series analysis and digital processing. Holden-Day.

P.E. Caines and C. Chan. 1975. Feedback between stationary stochastic processes. *IEEE Transactions on Automatic Control*, 20(4):498–508.

M.C. Campi and E. Weyer. 2005. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751–1764.

M.C. Campi and E. Weyer. 2010. Non-asymptotic confidence sets for the parameters of linear transfer functions. *IEEE Transactions on Automatic Control*, 55(12): 2708–2720.

G. Casella and R.L. Berger. 2002. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning.

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

B.C. Csáji, M.C. Campi, and E. Weyer. 2012a. Non-asymptotic confidence regions for the least-squares estimate. In *Proceeding of the 16th IFAC Symposium on System Identification*, pages 227–232.

B.C. Csáji, M.C. Campi, and E. Weyer. 2012b. Sign-Perturbed Sums (SPS): A method for constructing exact finite-sample confidence regions for general linear systems. In *Proceeding of the 51st IEEE Conference on Decision and Control*, pages 7321–7326.

A. Dankers, P.M.J. Van den Hof, and X. Bombois. 2014a. An instrumental variable method for continuous-time identification in dynamic networks. In *Proceeding of the 53rd IEEE Conference on Decision and Control*.

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

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

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

S. Doclo and M. Moonen. 2002. GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Transactions on Signal Processing*, 50 (9):2230–2244.

S. G. Douma. 2006. Probabilistic uncertainty bounding in output error models with unmodelled dynamics. In *Proceedings of the 2006 American Control Conference*, pages 1677–1682.

N. Everitt, G. Bottegal, C.R. Rojas, and H. Hjalmarsson. 2015. Variance analysis of linear SIMO models with spatially correlated noise. *Automatica*. Submitted.

N. Everitt, C.R. Rojas, and H. Hjalmarsson. 2013. A geometric approach to variance analysis of cascaded systems. In *Proceedings of the 52nd IEEE Conference on Decision and Control*.

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

U. Forssell and L. Ljung. 1999. Closed-loop identification revisited. *Automatica*, 35:1215–1241.

A. Friedman. 1970. *Foundations of Modern Analysis*. Dover.

S. Garatti, M.C. Campi, and S. Bittanti. 2004. Assessing the quality of identified models through the asymptotic theory - when is the result reliable? *Automatica*, 40(8):1319–1332.

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

M. Gilson and P.M.J. Van den Hof. 2005. Instrumental variable methods for closed-loop system identification. *Automatica*, 41(2):241 – 249.

G.C. Goodwin and R.L. Payne. 1977. *Dynamic System Identification: Experiment Design and Data Analysis*. Developmental Psychology Series. Academic Press.

C.W.J. Granger. 1969. Investigating causal relations by econometric models and cross–spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

C.W.J. Granger. 1980. Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and control*, 2:329–352.

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

I. Gustavsson, L. Ljung, and T. Söderström. 1977. Identification of processes in closed loop - identifiability and accuracy aspects. *Automatica*, 13(1):59 – 75.

A. Haber and M. Verhaegen. 2012. Identification of spatially distributed discrete-time state-space models. In *Proceeding of the 16th IFAC Symposium on System Identification*, pages 410–415.

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

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

D. Hayden, Y. Yuan, and J. Gonçalves. 2014a. Network reconstruction from intrinsic noise: Minimum-phase systems. In *Proceedings of the 2014 American Control Conference*, pages 4391–4396.

D. Hayden, Y. Yuan, and J. Goncalves. 2014b. Network reconstruction from intrinsic noise: Non-minimum-phase systems. In *Proceedings of the 19th IFAC World Congress*.

R. Hildebrand and M. Gevers. 2004. Quantification of the variance of estimated transfer functions in the presence of undermodeling. *IEEE Transactions on Automatic Control*, 49(8):1345–1350.

H. Hjalmarsson. 2005. From experiment design to closed loop control. *Automatica*, 41:393–438.

H. Hjalmarsson. 2009. System identification of complex and structured systems. *European Journal of Control*, 15(3-4):275–310.

H. Hjalmarsson and J. Mårtensson. 2011. A geometric approach to variance analysis in system identification. *IEEE Transactions on Automatic Control*, 56(5):983 – 997.

H. Hjalmarsson, J. Mårtensson, C.R. Rojas, and T. Söderström. 2011. On the accuracy in errors–in–variables identification compared to prediction–error identification. *Automatica*, 47(12):2704 – 2712.

H. Hjalmarsson, J. Mårtensson, and B. Wahlberg. 2006. On some robustness issues in input design. In *Proceedings of the 14th IFAC Symposium on System Identification*, pages 511–516.

H. Hjalmarsson and B. Ninness. 2006. Least-squares estimation of a class of frequency functions: A finite sample variance expression. *Automatica*, 42:589–600.

R.A. Horn and C.R. Johnson. 1990. *Matrix Analysis*. Cambridge University Press.

S. Kolumbán, I. Vajk, and J. Schoukens. 2015. Perturbed datasets methods for hypothesis testing and structure of corresponding confidence sets. *Automatica*, 51(0):326 – 331.

T.J.T. Koski and J. Noble. 2012. A review of Bayesian networks and structure learning. *Mathematica Applicanda*, 40(1):51–103.

E. Kreyszig. 1978. *Introductory Functional Analysis With Applications*. John Wiley & Sons, Inc, New York.

L. Ljung. 1985. Asymptotic variance expressions for identified black-box transfer function models. *IEEE Transactions on Automatic Control*, 30(9):834 – 844.

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

L. Ljung, H. Hjalmarsson, and H. Ohlsson. 2011. Four encounters with system identification. *European Journal of Control*, 17(5):449–471.

D.G. Luenberger. 1969. *Optimization by vector space methods*. John Wiley & Sons, Inc, New York.

C. Lyzell, M. Enqvist, and L. Ljung. 2009. Handling certain structure information in subspace identification. In *Proceedings of the 15th IFAC Symposium on System Identification*, pages 90–95.

Y. Majanne. 2005. Model predictive pressure control of steam networks. *Control Engineering Practice*, 13(12):1499–1505.

V.M. Marques, C.J. Munaro, and S.L. Shah. 2013. Data-based causality detection from a system identification perspective. In *Proceedings of the 2013 European Control Conference*, pages 2453–2458.

J. Mårtensson. 2007. *Geometric analysis of stochastic model errors in system identification*. Ph.D. thesis, KTH, Automatic Control.

J. Mårtensson, N. Everitt, and H. Hjalmarsson. 2015. Variance analysis in SISO linear systems identification. *Automatica*. Submitted.

J. Mårtensson and H. Hjalmarsson. 2009. Variance-error quantification for identified poles and zeros. *Automatica*, 45(11):2512 – 2525.

J. Mårtensson and H. Hjalmarsson. 2011. How to make bias and variance errors insensitive to system and model complexity in identification. *IEEE Transactions on Automatic Control*, 56(1):100–112.

J. Mårtensson, C. R. Rojas, and H. Hjalmarsson. 2011. Conditions when minimum variance control is the optimal experiment for identifying a minimum variance controller. *Automatica*, 47(3):578 – 583.

P. Massioni and M. Verhaegen. 2008. Subspace identification of circulant systems. *Automatica*, 44(11).

P. Massioni and M. Verhaegen. 2009. Subspace identification of distributed, decomposable systems. In *Proceedings of the 48th IEEE Conference on Decision and Control, held jointly with the 28th Chinese Control Conference*.

D. Materassi and M.V. Salapaka. 2012. On the problem of reconstructing an unknown topology via locality properties of the wiener filter. *IEEE Transactions on Automatic Control*, 57(7):1765–1777.

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

D.B. McCombie, A.T. Reisner, and H.H. Asada. 2005. Laguerre-model blind system identification: Cardiovascular dynamics estimated from multiple peripheral circulatory signals. *IEEE Transactions on Biomedical Engineering*, 52(11):1889–1901.

M. Milanese and C. Novara. 2011. Unified Set Membership theory for identification, prediction and filtering of nonlinear systems. *Automatica*, 47(10):2141–2151.

M. Milanese and A. Vicino. 1991. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica*, 27(6):997–1009.

L. Mišković, A Karimi, D. Bonvin, and M. Gevers. 2008. Closed-loop identification of multivariable systems: With or without excitation of all references? *Automatica*, 44(8):2048–2056.

B. Ninness and J. C. Gómez. 1996. Asymptotic analysis of MIMO system estimates by the use of orthonormal bases. In *Proceedings of the 13th IFAC World Congress*.

B Ninness and G.C. Goodwin. 1995. Estimation of model quality. *Automatica*, 31 (12):1771–1797.

B. Ninness and F. Gustafsson. 1997. A unifying construction of orthonormal bases for system identification. *IEEE Transactions on Automatic Control*, 42:515–521.

B. Ninness and H. Hjalmarsson. 2004. Variance error quantifications that are exact for finite model order. *IEEE Transactions on Automatic Control*, 49(8):1275–1291.

B. Ninness and H. Hjalmarsson. 2005a. Analysis of the variability of joint input–output estimation methods. *Automatica*, 41(7):1123 – 1132.

B. Ninness and H. Hjalmarsson. 2005b. On the frequency domain accuracy of closed-loop estimates. *Automatica*, 41:1109–1122.

B. Ninness, H. Hjalmarsson, and F. Gustafsson. 1999. The fundamental role of general orthonormal bases in system identification. *IEEE Transactions on Automatic Control*, 44(7):1384–1406.

R. Pintelon and J. Schoukens. 2012. *System identification: a frequency domain approach*, 2nd edition. John Wiley & Sons.

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

C.R. Rojas, M. Barenthin, J.S. Welsh, and H. Hjalmarsson. 2008. The cost of complexity in identification of FIR systems. In *Proceedings of the 17th IFAC World Congress*, pages 11451–11456, Seoul, South Korea.

C.R. Rojas, M. Barenthin Barenthin, J.S. Welsh, and H. Hjalmarsson. 2010. The cost of complexity in system identification: Frequency function estimation of finite impulse response systems. *IEEE Transactions on Automatic Control*, 55 (10):2298–2309.

C.R. Rojas, H. Hjalmarsson, L. Gerencsér, and J. Mårtensson. 2011. An adaptive method for consistent estimation of real-valued non-minimum phase zeros in stable LTI systems. *Automatica*, 47(7):1388 – 1398.

C.R. Rojas, J.S. Welsh, and J.C. Aguero. 2009. Fundamental limitations on the variance of estimated parametric models. *IEEE Transactions on Automatic Control*, 54(5):1077–1081.

Y.A. Rozanov. 1967. *Stationary Random Processes*. Holden-Day.

W. Rudin. 1976. *Principles of Mathematical Analysis*. McGraw-Hill, London.

B. M. Sanandaji, T. Vincent, and M. Wakin. 2012. A review of sufficient conditions for structure identification in interconnected systems. In *Proceedings of the 16th IFAC Symposium on System Identification*, pages 1623–1628.

H. Sandberg, P. Hägg, and B. Wahlberg. 2014. Approximative model reconstruction of cascade systems. *Systems & Control Letters*, 69(0):90 – 97.

R.O. Schmidt. 1986. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation*, 34(3):276–280.

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

P. Stoica, M. Viberg, and B. Ottersten. 1994. Instrumental variable approach to array processing in spatially correlated noise fields. *IEEE Transactions on Signal Processing*, 42(1):121–133.

C. Sturk, H. Sandberg, P. Trnka, V. Havlena, and J. Rehor. 2011. Structured model order reduction of boiler-header models. In *Proceedings of the 18th IFAC World Congress*.

C. Sturk, L. Vanfretti, F. Milano, and H. Sandberg. 2012. Structured model reduction of power systems. In *Proceedings of the 2012 American Control Conference*, pages 2276–2282.

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

L.N. Trefethen and D. Bau. 1997. *Numerical Linear Algebra*. Society for Industrial and Applied Mathematics.

D.J. Trudnowski, J.M. Johnson, and J.F. Hauer. 1998. SIMO system identification from measured ringdowns. In *Proceedings of the 1998 American Control Conference*, volume 5, pages 2968–2972.

H.S. Ulusoy, M.Q. Feng, and P.J. Fanning. 2011. System identification of a building from multiple seismic records. *Earthquake Engineering & Structural Dynamics*, 40(6):661–674.

P. M. J. Van den Hof, Arne Dankers, P. S. C. Heuberger, and X Bombois. 2013. Identification of dynamic models in complex networks with prediction error methods - basic methods for consistent module estimates. *Automatica*, 49(10):2994–3006.

J.W. van Wingerden and P. Torres. 2012. Identification of spatially interconnected systems using a sequentially semi-separable gradient method. In *Proceeding of the 16th IFAC Symposium on System Identification*, pages 173–178.

M. Viberg and B. Ottersten. 1991. Sensor array processing based on subspace fitting. *IEEE Transactions on Signal Processing*, 39(5):1110–1121.

M. Viberg, P. Stoica, and B. Ottersten. 1997. Maximum likelihood array processing in spatially correlated noise fields using parameterized signals. *IEEE Transactions on Signal Processing*, 45(4):996–1004.

B. Wahlberg. 1991. System identification using Laguerre models. *IEEE Transactions on Automatic Control*, 36:551–562.

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

B. Wahlberg, H. Hjalmarsson, and P. Stoica. 2012. On the performance of optimal input signals for frequency response estimation. *IEEE Transactions on Automatic Control*, 57(3):766–771.

B. Wahlberg and L. Ljung. 1992. Hard frequency-domain model error bounds from least-squares like identification techniques. *IEEE Transactions on Automatic Control*, 37(7):900–912.

B. Wahlberg and H. Sandberg. 2008. Cascade structural model approximation of identified state space models. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 4198–4203.

N. Wiener. 1956. The theory of prediction. In E.F. Beckenbachn, editor, *Modern Mathematics for Engineer*. McGraw Hill Book Company, New York.

D.A. Wunsch. 1993. *Complex variables with applications*, second edition. Addison Wesley Publishing Company.

L.L. Xie and L. Ljung. 2001. Asymptotic variance expressions for estimated frequency functions. *IEEE Transactions on Automatic Control*, 46(12):1887–1899.

L.L. Xie and L. Ljung. 2004. Variance expressions for spectra estimated using auto-regressions. *Journal of econometrics*, 118(1):247–256.

Y. Yang, F. Lambert, and D. Divan. 2007. A survey on technologies for implementing sensor networks for power delivery systems. In *Proceedings of the 2007 IEEE Power Engineering Society General Meeting*, pages 1–8.