KTH Electrical Engineering

Using Structural Information in System Identification

PER HÄGG

Licentiate Thesis
Stockholm, Sweden 2012

KTH School of Electrical Engineering
Automatic Control Lab
SE-100 44 Stockholm
SWEDEN

TRITA-EE 2012:012
ISSN 1653-5146
ISBN 978-91-7501-299-5

Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie licentiatexamen i reglerteknik torsdagen den 26 April 2012 klockan 15.15 i Q2 Kungl Tekniska högskolan, Osquldasväg 10, Stockholm.

© Per Hägg, April 2012

Tryck: Universitetsservice US AB

Abstract

Recent advances in small and cheap communication and sensing have opened up for large scale systems with intricate interconnections and interactions. These applications pose new challenges for analysis and control design. To keep up with the increasing demand on performance and efficiency, accurate models of these systems are needed. Often some prior knowledge of the system, such as system structure, is available. Prior knowledge should whenever possible be used in system identification to improve the model estimate.

This thesis addresses the problem of using prior information about the overall structure of the system in system identification. Two special structures are considered, the cascade and the parallel serial structure. The motivation for looking at these structures are two folded; they are common in industrial applications and they can be used to build up almost all interesting feedforward interconnected systems. The effect of sensor placement, input signals and common dynamics of the subsystems on the quality of the estimated models for these two structures is considered.

In many control applications it is vital that the model has a physical interpretation. Hence, it is important that the system identification method retains the physical interpretation of the identified model. However, it has proven hard to incorporate prior knowledge of structure in subspace methods. This thesis presents two methods for identifying systems with known structures using subspace methods. The first method utilizes that the state-space matrices of a system on cascade form have a certain structure. The idea is to find a transform that takes the identified system back to this form. The second method uses the known structure of the extended observability matrix. The state-space matrices for the subsystems can then be found by solving linear least squares problems. However, the method is only applicable if the second subsystem has order one. But this is a common case in practice. The two methods are applied to a two tank lab process with promising results.

Nonparametric estimates of the frequency response function of systems are used in most engineering fields. The second contribution of this thesis is a new method for estimating the frequency response. The method uses the known structure of the transient or leakage error. The feasibility of the method is tested in simulations. For the two cases considered, one with a large amount of random systems, the second with a resonant system, the method shows good performance compared to current state of the art methods.

Acknowledgment

Lite text... fds fds fds ghd hgf

Contents

Chapter 1

Introduction

We all use models, simplifications of the reality, in our day to day life, even if we are not always aware of it. Consider the simple task of catching a ball. You use a model, based on previous experience, to predict the trajectory and to hopefully catch the ball. The meteorologist uses measurements and models to predict the weather the coming days. Central banks forecast how a change in the interest rate will affect the economy, based on models of the world economy, and so on.

A natural question is then how does one find a model of a system? In engineering applications, models are often expressed in terms of mathematical and statistical relationships or as a set of rules describing the modeled process. One way to find these relations is to use first principles modeling, that is to use knowledge of the system to derive a model, for example from the basic laws of physics. Another way is to use empirical data from the system under consideration. The task is to establish a model explaining the observed data as good as possible.

The two approaches can be combined. The underlying structure of the system is often known but all internal relations are not and have to be estimated from data. Contrary to a model based only on empirical data, the combined approach retains the physical interpretation of the model, something that is vital in many applications. Hence prior knowledge that improves the quality of the model should always be used in the identification.

The development of advanced communications systems and the increasing demand on performance and efficiency have led to a growth in large-scale interconnected engineering systems. Systems are nowadays viewed as parts of a greater system instead of individual entities. Often the structure of the system, *i.e.,* the interconnections, is known but not the dynamics of the individual subsystem. Deriving good models of these systems are essential from a performance perspective. For the models to be useful it is also desirable to retain the interconnected structure.

The main goal of this thesis is to study methods for incorporating prior knowledge of the system in the model identification step.

1.1 Motivating examples

To motivate the work in this thesis, two examples are given. The first example illustrates the need for structured system identification, that is identification methods that retain the physical interpretation of the model. The second example highlights the usefulness of accurate nonparametric frequency response estimation methods.

Example 1.1 (Identification of Steam Boiler System)
Consider the power plant illustrated in Figure 1.1. Steam is generated by three parallel boilers, were combustion of fuel heats water. The steam from the boilers is collected in a steam header and lead through a turbine. The turbine is connected to a generator, generating electrical power to the grid. A block diagram of the steam boiler system is shown 1.2. The system has a certain structure that will be denoted parallel serial structure.

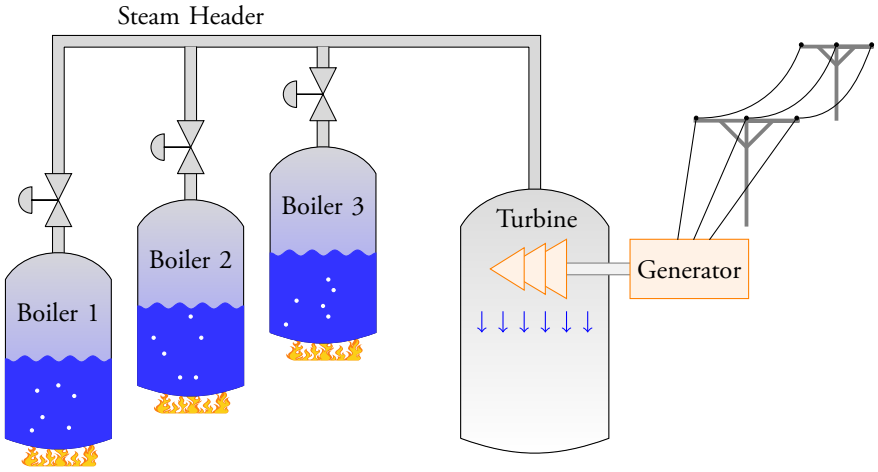


Figure 1.1 Sketch of a power plant with parallel steam boilers.

To be able to adjust for changes in power demand, and hence steam demand in the turbine, the boilers can be disconnected from the header by closing valves. The aim is to identify a model from the inputs to the system, the fuel rate to the combustion

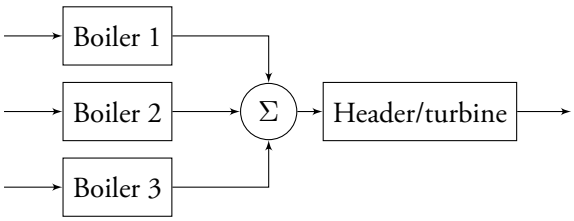


Figure 1.2 Block diagram of power plant with parallel steam boilers.

chamber, to the output, the generated power. The intended use of the model will be for controller design.

The system can be in many different configurations depending on which boilers are turned on and which are turned off. Performing identification experiments for each of the possible on/off configurations is not practically feasible. If, on the other hand, the system is identified with a method that retains the physical structure of the system it is easy to implement the different configurations in the model by removing or adding the corresponding parallel subsystem. Hence system identification methods that preserve the underlying physical structure are important in this case.

The example shows the need for system identification algorithms that take advantage of the structure. Methods to identify structured systems will be discussed in Chapter 4. The boiler system has a parallel serial structure. The fundamental limitations of the quality of the estimates when identifying systems with this specific structure will be studied in Chapter 3.

Example 1.2 (Nonparametric System Identification)
A simple time-variant discrete system is given by

$$y(t) = b(t)u(t) + v(t)$$

where $u(t)$ is the input signal, $v(t)$ is zero mean measurement noise and the parameter $b(t)$ is given by

$$b(t) = \begin{cases} 0.7 & 0 \leq t \leq 35 \\ -0.4 & 36 \leq t \leq 85 \\ 0.4 & t \geq 86. \end{cases}$$

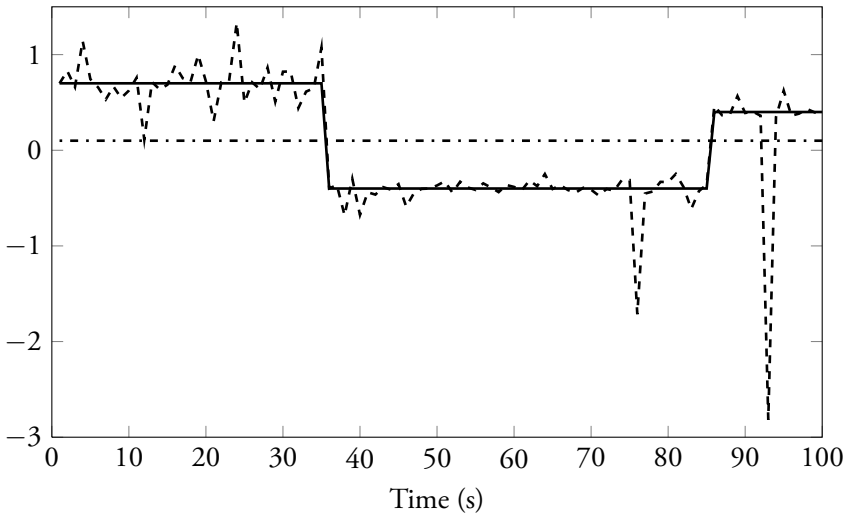


Figure 1.3 Estimate of $b(t)$ with parametric method (· - - -) and with nonparametric method (- - -). True value shown as (——).

Assume that the structure of the true system is known. If it is not known that $b(t)$ is time varying a reasonable model of the system is given by

$$\hat{y}(t) = bu(t) + v(t).$$

Figfig:nonparam1 shows the estimate, $\hat{b}$ of b, in a least squares sense, from 100 input and output measurements. Since the model does not capture the behavior of the true system, the estimate of the parameter b is inadequate. A simple nonparametric estimate, *i.e.,* an estimate for each time t, of the parameter $b(t)$ is for example

$$\hat{b}(t) = \frac{y(t)}{u(t)}.$$

The nonparametric estimate is also shown in Figure 1.3. The estimate is noisy (there is no smoothing of the noise) and there are large peaks or dips (where $|u(t)|$ is small and there is noise present). The nonparametric estimate can be improved by adding some smoothing. A smoothed nonparametric estimate is shown in Figure 1.4. From the estimates it is possible to get an insight into the problem; it looks like the term b is piecewise constant with three different regions.

The information from the nonparametric estimate can be used to design a better parametric model. In Figure 1.4 a parametric estimate with three parameters, one for

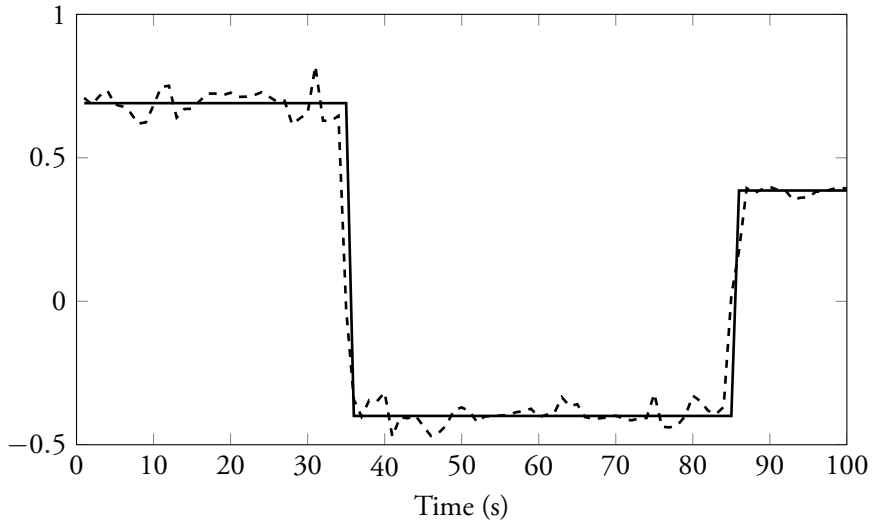


Figure 1.4 Estimate of $b(t)$ with parametric method (——) and with improved nonparametric method (- - -).

each time interval of $b(t)$, is presented. The estimate is close to the true value of the parameter.

Example 1.2 shows the usefulness of nonparametric methods. A good nonparametric method should not be restricted by model assumptions that could reduce the flexibility to accurately reproduce the true system. The nonparametric model can hence be used to gain insight in the system properties and help in choosing a model structure used in a more accurate parametric method.

In this thesis the nonparametric methods will be used in a slightly different setting; they will be used to estimate the frequency response function of a linear dynamic system. This introduces some additional questions and problems.

1.2 Problem formulation

The overall problem studied in this thesis is how to include prior structural knowledge of the system in system identification. Two different types of prior information are studied. The first part of this thesis involves a system consisting of multiple subsystems where the interconnections, but not their respective dynamics, are known. Some

existing methods for identification of structured system are considered and two novel methods based on subspace identification are proposed. Further, the fundamental limitations when identifying structured systems are considered. For example, how does the choice of input signal, sensor placement and dynamics of the subsystems affect the achievable performance?

The second part of this thesis considers nonparametric frequency response estimation. Here, the a priori information is the structure of the leakage or transient error, introduced by the finite data length. This structural information is utilized in a novel method. The method is compared to current state of the art methods and pros and cons are discussed.

1.3 Related Work

In this section, research related to the major topics of this thesis are discussed.

Identification of Structured systems

The first part of this thesis is focused on identification of structured systems. Here a structured system means a system consisting of multiple interconnected subsystems. From a system identification point of view, these systems give rise to two interesting problems.

The first problem is to find the structure of the system from measurement data. Early work on this problem can be found in for example Granger (1969) and Caines (1976), were the problem of finding the causal relation between two signals and whether there is feedback present between the two signals is addressed. Recent advances in wireless communications and electronics have led to an increased interest in networked control systems. Some recent work on identification of networked systems can be found in Materassi and Innocenti (2010) and Sanandaji et al. (2011).

The second problem is to identify the dynamics of the subsystems when the interconnections and the structure of the system is known. If only the input-output relations are of interest this problem can be seen as identification of a system with multiple inputs and multiple outputs something that many of the classical system identification methods can handle. However, if one wants to preserve the underlying structure then the classical methods are in many cases not directly applicable. How does one incorporate the prior knowledge of the structure into system identification? It is possible to apply a prediction error method or a maximum likelihood method to a constrained model structure. On the other hand this often involves solving a non-convex optimization problem. One way to identify these systems were proposed by

Wahlberg and Sandberg (2008). The idea is to in a first step identify a MIMO model of the system, then find high-order but structured model approximation of the subsystems using standard H_∞ model matching techniques. The final step is to use structured balanced model reduction to find low order approximations of the subsystems taking the known structure of the system into account. This method has been further refined in Sturk et al. (2011).

Subspace Identification with Prior Information

Subspace identification methods have proven useful in many applications. The user-friendliness and their ability to handle multiple-input multiple-output (MIMO) systems are parts of their success. However, it has proven hard to incorporate prior knowledge of structure in subspace methods.

Different approaches to incorporate prior information in subspace identification have been proposed in the literature. Maciejowski (1995); Van Gestel et al. (2001); Lacy and Bernstein (2003) consider the problem of guaranteeing that the identified model, estimated with a subspace method, is stable when the true system is known to be stable. Okada and Sugie (1996) proposed a method that uses the known location of the poles. Other system properties such as known static gain, smoothness of the step response and the time constant have been incorporated by Trnka and Havlena (2009) where a Bayesian approach is used and the prior information is used to shape the covariance matrices. Alenany et al. (2011) on the other hand incorporates the known system properties by expressing them as equality constraints in the impulse response coefficients, leading to a constrained least squares problem.

Identifying the subsystems in a structural system using a subspace based method can be seen as a problem of imposing structure in the identified model. One example of this can be found in Lyzell et al. (2009) where subspace identification of OE and ARMAX models are considered. The idea is to find a suitable transformation, taking the identified model to a form corresponding to a OE or ARMAX model. More related to the problem considered in this thesis is the work by Massioni and Verhaegen (2008). They consider subspace identification of a class of large scale systems called "circulant systems". Circulant systems have a special property that allows them to be decomposed into simpler subsystems through a state transformation, a property that can be exploited in the subspace identification method.

Nonparametric Frequency Response Estimation

Nonparametric models of the Frequency Response Function (FRF) are commonly used in almost all fields of engineering ranging from acoustics to advanced processes. They

can also be used as a first step to get insight into the identification problem at hand, even if the final goal is a parametric model. Hence it is important that the nonparametric model captures the important properties of the system and it should be unprejudiced to the system under consideration. It is also important that the nonparametric method should limit the required user-interactions and user choices; else a parametric method could equally well be used.

A major problem when estimating the FRF is the leakage, or transient, error. The leakage error is introduced by the finite measurement time. The classical way to overcome this is to use windowing methods (Stoica and Moses, 2005). The recent introduction of the Local Polynomial Method (LPM) (Schoukens et al., 2006) has led to a renewed interest in the area. LPM is a method for estimating the frequency response function utilizing that the leakage errors are smooth functions of the frequency, allowing them to be approximated well by a low order Taylor series expansion. The Taylor coefficients are then estimated, together with the Frequency response, at one frequency at a time. LPM has been demonstrated to provide superior accuracy on a number of problems compared to traditional methods. The method has been further refined in Gevers et al. (2011) where the relationship between the estimated coefficients at neighboring frequencies is taken into account, making the method "more global". Instead of using a local polynomial approximation, McKelvey and Guérin (2012) proposed the use of a local rational model at each frequency.

Since the frequency response function is the Fourier transform of the impulse response, methods for estimating the impulse response are closely related to FRF estimation. In this area, some new, rather intriguing, approaches have been proposed by Pillonetto and Nicolao (2010) and the follow up paper Pillonetto et al. (2011). The impulse response is modeled as a realization of a Gaussian process. The priors are chosen to reflect information on smoothness and on BIBO-stability and are parameterized by a few hyper-parameters. The hyper-parameters are found via an empirical Bayes method. The relation of this method to more standard regularization has been studied in Chen et al. (2012). The new methods require very little user-interaction and yield good results in many cases compared to standard parametric and nonparametric techniques.

1.4 Contributing Papers

This thesis is based on the following publications:

- Chapter 3

  P. Hägg, B. Wahlberg, and H. Sandberg. On identification of parallel cascade

serial systems. In *Proceedings of the 18th World Congress*, 9978–9983 (2011b)

- Chapter 4

  P. Hägg, B. Wahlberg, and H. Sandberg. On subspace identification of cascade structured systems. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2843 –2848 (2010)

- Chapter 5

  P. Hägg, H. Hjalmarsson, and B. Wahlberg. A least squares approach to direct frequency response estimation. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 2160 –2165 (2011a)

  P. Hägg and H. Hjalmarsson. Non-parametric frequency function estimation using transient impulse response modelling. In *16th IFAC Symposium on System Identification* (2012). To appear

- Other publications not directly related or not considered in this thesis:

  M. Gevers, P. Hägg, H. Hjalmarsson, R. Pintelon, and J. Schoukens. The transient impulse response modeling method and the local polynomial method for nonparametric system identification. In *16th IFAC Symposium on System Identification* (2012). To appear

  C. A. Larsson and P. Hägg. Generation of excitation signals with prescribed autocorrelation for input and output constrained system. In *Decision and Control, 2012. CDC 2012. 51st IEEE Conference on* (2012). Submitted

1.5 Outline

The outline of the thesis is as follows. In Chapter 2 a brief introduction to system identification is given and some background material needed for understanding the following chapters are presented. Chapter 3 considers the fundamental limitations of the quality of the estimates when a Prediction Error Method is applied to a structured system. Two novel subspace based methods for estimating systems with cascade structures are presented in Chapter 4 and they are evaluated on a lab two tank process in Chapter 5. In Chapter 6 a new method for nonparametric frequency

response estimation is presented. Finally, Chapter 7 summarizes the thesis and some future directions for research are given.

Chapter 2

Background

The methods and results presented in this thesis are based on previous work in the field on system identification. In this chapter the theoretical background needed in the coming chapters is presented. Furthermore some related work is discussed in more details and references to relevant topics are given. First a short introduction to system identification is given.

2.1 System Identification

System identification concerns the problem of construction and validation of mathematical models of dynamical systems from experimental data. The objective is to create a model that captures the properties relevant to the intended use of the model. Numerous methods have been proposed, some more general, some more focused on solving a specific problem. In this thesis both parametric and nonparametric methods will be considered. The identification can be performed both in the time domain (Ljung, 1999; Söderström and Stoica, 2001) and in the frequency domain (Pintelon and Schoukens, 2001). Often the final result is independent of the domain. The difference is the calculations needed to arrive to the solution. Preferably the choice of domain should consider where the problem becomes easier to solve. However, the chosen domain, and also the used method, is often influenced by tradition within the field of application.

In this thesis, the first part, considering identification of systems with structure, is done in time domain (due to the Swedish tradition) while the second part dealing with frequency response estimation for the most part is done in the frequency domain.

2.2 Prediction Error Identification

Consider the stable linear system

$$y(t) = G_0(q)u(t) + v(t) \tag{2.1}$$

where $G_0(q)$ is the true system, $u(t)$ the input signal and $v(t)$ captures the additive disturbance acting on the system and can be written as

$$v(t) = H_0(q)e(t)$$

where $H_0(q)$ is a stable and inversely stable, monic linear filter and $e(t)$ is zero mean white noise with variance λ.

From N data-points of the input and the output of the system (2.1) $\{u(t), y(t), t = 1, \ldots, N\}$, the aim is to find a model that describe the system as good as possible. The model that is to be fitted is parameterized by a d-dimensional parameter vector $\theta \in \mathbb{R}^d$, *i.e.,*

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t). \tag{2.2}$$

Normally $G(q, \theta)$ and $H(q, \theta)$ are represented by rational functions and the parameters are the numerator and denominator coefficients. Depending on the relations between the numerators and denominators of the two rational function, different model structures are achieved, for example FIR, ARMAX and Box-Jenkins. In Ljung (1999) these, and other common, model structures are explained in detail.

The optimal one step ahead predictor for the model (2.2) is given by

$$\hat{y}(t|t-1, \theta) = H^{-1}(q, \theta)G(q, \theta)u(t) + \left[1 - H^{-1}(q, \theta)\right] y(t)$$

with the resulting prediction error

$$\epsilon\,(t, \theta) = y(t) - \hat{y}(t|t-1, \theta) = H^{-1}(q, \theta)\,(y(t) - G(q, \theta)u(t))\,. \tag{2.3}$$

The objective of a prediction-error method (PEM) is to find a θ from data such that the prediction error (2.3) is as small as possible for $t = 1, \ldots, N$. One measure of the size of the prediction error is

$$V_N\,(\theta) = \frac{1}{N}\sum_{t=1}^{N} l\,(\epsilon(t, \theta))$$

where $l(\cdot)$ is a scalar valued function. In this thesis the standard quadratic norm

$$l(\epsilon) = \frac{1}{2}\,\epsilon^2,$$

will be used throughout. This corresponds to a Maximum Likelihood (ML) estimate when the noise is Gaussian. The PEM parameter estimate is then given by

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta). \tag{2.4}$$

Statistical Properties of the Estimates

Assume that true system is in the model set, *i.e.,* that there exist some θ_0 such that $G(e^{j\omega}, \theta_0) = G_0(e^{j\omega})$ and $H(e^{j\omega}, \theta_0) = H_0(e^{j\omega})$. Then under some mild assumptions the parameter estimate from 2.4) has the asymptotical distribution

$$\sqrt{N}\left(\theta_0 - \hat{\theta}_N\right) \in \text{As}\, \mathcal{N}(0, P_\theta)$$

where the covariance matrix is given by

$$P_\theta = \lambda \left[\bar{E}\left(\Psi(t,\theta_0)\Psi^T(t,\theta_0)\right)\right]^{-1},$$
$$\Psi(t,\theta_0) = -\frac{\mathrm{d}}{\mathrm{d}\,\theta}\hat{y}(t,\theta)\Big|_{\theta=\theta_0}.$$

More general results and further details can be found in Ljung (1999).

2.3 Subspace Identification

The family of subspace identification methods is used to estimate a state-space realization of a system. The popularity of subspace based identification is mainly due to its two main advantages. Firstly its simplicity for the user, given input-output data it produces estimates of the state-space matrices and also gives a suitable model order without the need for highly technical decisions by the user. Secondly, it is no difference in the algorithm whether the data is from a SISO or a MIMO system.

The different steps in the subspace identification methods will be summarized based on the treatment in Ljung (1999). A broader overview of subspace identification can be found in for example Van Overschee and De Moor (1996).

The aim is to estimate the matrices **A**, **B**, **C** and **D** in a state-space model

$$x(t+1) = \mathbf{A}x(t) + \mathbf{B}u(t) + w(t)$$
$$y(t) = \mathbf{C}x(t) + \mathbf{D}u(t) + v(t)$$

from input output data $\{u(t), y(t), t = 1, \ldots, N\}$. The system has m inputs and p outputs. The input signal $u(t) \in \mathbb{R}^m$ is a m-dimensional vector while the output signal $y(t) \in \mathbb{R}^p$ is a p-dimensional vector.

Finding The Extended Observability Matrix

The first step is to find an estimate of the extended observability matrix $\mathbf{O}_r$. The output of the system at time $t+k$ can be expressed as

$$\begin{aligned} y(t+k) = {} & \mathbf{C}\mathbf{A}^k x(t) + \mathbf{C}\,\mathbf{A}^{k-1}\mathbf{B}u(t) + \mathbf{C}\mathbf{A}^{k-2}\mathbf{B}u(t+1) + \dots \\ & + \mathbf{C}\mathbf{B}u(t+k-1) + \mathbf{D}u(t+k) \\ & + \mathbf{C}\mathbf{A}^{k-1}w(t) + \mathbf{C}\mathbf{A}^{k-2}w(t+1) + \dots \\ & + \mathbf{C}w(t+k-1) + v(t+k). \end{aligned} \tag{2.5}$$

Stacking (2.5) from time t to $t+r-1$ gives

$$Y_r(t) = \mathbf{O}_r\, x(t) + \mathbf{S}_r U_r(t) + V(t) \tag{2.6}$$

with

$$Y_r(t) = \begin{bmatrix} y(t) \\ y(t+1) \\ \vdots \\ y(t+r-1) \end{bmatrix}, \quad U_r(t) = \begin{bmatrix} u(t) \\ u(t+1) \\ \vdots \\ u(t+r-1) \end{bmatrix}$$

and

$$\mathbf{S}_r = \begin{bmatrix} \mathbf{D} & 0 & \cdots & 0 & 0 \\ \mathbf{C}\mathbf{B} & \mathbf{D} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{C}\mathbf{A}^{r-2}\mathbf{B} & \mathbf{C}\mathbf{A}^{r-3}\mathbf{B} & \cdots & \mathbf{C}\mathbf{B} & \mathbf{D} \end{bmatrix}, \quad \mathbf{O}_r = \begin{bmatrix} \mathbf{C} \\ \mathbf{C}\mathbf{A} \\ \vdots \\ \mathbf{C}\mathbf{A}^{r-1} \end{bmatrix}.$$

Equation (2.6) collected for $\{u(t), y(t), t = 1, \dots, N\}$ can be rewritten on matrix form as

$$\mathbf{Y} = \mathbf{O}_r\mathbf{X} + \mathbf{S}_r\mathbf{U} + \mathbf{V} \tag{2.7}$$

where

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} Y_r(1) & Y_r(2) & \cdots & Y_r(N-r+1) \end{bmatrix}, \\ \mathbf{X} &= \begin{bmatrix} x(1) & x(2) & \cdots & x(N-r+1) \end{bmatrix}, \\ \mathbf{U} &= \begin{bmatrix} U_r(1) & U_r(2) & \cdots & U_r(N-r+1) \end{bmatrix}, \\ \mathbf{V} &= \begin{bmatrix} V_r(1) & V_r(2) & \cdots & V_r(N-r+1) \end{bmatrix}. \end{aligned}$$

The objective is to estimate the extended observability matrix $\mathbf{O}_r$ from given data $\mathbf{U}$ and $\mathbf{Y}$. This is done by eliminating the other terms in (2.7). To eliminate $\mathbf{S}_r\mathbf{U}$, the

expression (2.7) is post multiplied by a projection matrix $\Pi_U^{\perp} = \mathbf{I} - \mathbf{U}^T(\mathbf{U}\mathbf{U}^T)^{-1}\mathbf{U}$ giving

$$\mathbf{Y}\Pi_U^{\perp} = \mathbf{O}_r\mathbf{X}\Pi_U^{\perp} + \mathbf{V}\Pi_U^{\perp}. \tag{2.8}$$

Observing that the noise $v(t)$ is uncorrelated with $u(\tau)$ and $y(\tau)$ for $t > \tau$ the noise term at time t can be correlated away by multiplying with

$$\phi_s(t) = \begin{bmatrix} y(t-1) \\ \vdots \\ y(t-s_1) \\ u(t-1) \\ \vdots \\ u(t-s_2) \end{bmatrix}. \tag{2.9}$$

Post-multiplying (2.8) by the matrix $\Phi^T = \begin{bmatrix} \phi_s(1) & \cdots & \phi_s(N) \end{bmatrix}$, gives

$$\underbrace{\frac{1}{N}\mathbf{Y}\,\Pi_U^{\perp}\Phi^T}_{\mathbf{G}} = \mathbf{O}_r \underbrace{\frac{1}{N}\mathbf{X}\Pi_U^{\perp}\Phi^T}_{\tilde{\mathbf{T}}_N} + \underbrace{\frac{1}{N}\mathbf{V}\Pi_U^{\perp}\Phi^T}_{\mathbf{V}_N} \xrightarrow{N\to\infty} \mathbf{O}_r\tilde{\mathbf{T}}.$$

The parameters s_1 and s_2 are design parameters and chosen by the user.

An estimate of the transformed extended observability matrix $\mathbf{O}_r\tilde{\mathbf{T}}$ is hence $\mathbf{G}$. Furthermore the estimate only depends on the input and output data and not on the unknown states of the system.

Remark 2.1 A state-space realization is not unique with respect to the input-output relation. Multiplying the extended observability matrix O_r from the right with any invertible matrix will only change the basis of representation of the system, not the system itself.

Order Selection

From the noisy estimate of the extended observability matrix $\mathbf{G} = \mathbf{O}_r\tilde{\mathbf{T}} + \mathbf{V}_N$ the order of the system can be estimated. The rank of the extended observability matrix $\mathbf{O}_r$ equals the order of the true system. Therefore, in the noise free case, and when $\tilde{\mathbf{T}}$ has full rank, the order of the system can be calculated from the rank of $\mathbf{G}$. But when $\mathbf{V}_N$ is nonzero the rank of $\mathbf{G}$ will in general not be n. Instead a SVD of $\mathbf{G}$ is performed and the order is chosen by inspecting the number of singular values that are larger than some small predetermined threshold. The extended observability matrix is hence approximated by

$$\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^T \approx \mathbf{U}_1\mathbf{S}_1\mathbf{V}_1^T \tag{2.10}$$

where the last approximation is done by keeping the most significant singular values in $\mathbf{S}_1$ and setting the rest to zero.

Estimating A and C

From the reduced order extended observability matrix it is possible to find an estimate of $\mathbf{A}$ and $\mathbf{C}$. The matrix can be written as

$$\mathbf{O}_r = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{r-1} \end{bmatrix}$$

and an estimate $\hat{\mathbf{C}}$ of $\mathbf{C}$ can be taken as the first block row of $\mathbf{O}_r$. The estimate $\hat{\mathbf{A}}$ can then be found by the shift property of $\mathbf{O}_r$ and should be solved in a least squares sense, *i.e.,*

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \|\mathbf{O}_r(p+1:pr) - \mathbf{O}_r(1:p(r-1))\|_2^2$$

where $\mathbf{O}_r(p:q)$ denotes row p to row q of $\mathbf{O}_r$.

Estimation of B and D

Finally the matrices $\mathbf{B}$ and $\mathbf{D}$ and the initial state, x_0 of the system can be solved from the linear regression problem

$$\arg\min_{\mathbf{B},\mathbf{D},x_0} \frac{1}{N}\sum_{t=1}^{N} \|y(t) - \hat{\mathbf{C}}(qI-\hat{\mathbf{A}})^{-1}\mathbf{B}u(t) - \mathbf{D}u(t) - \hat{\mathbf{C}}(qI-\hat{\mathbf{A}})^{-1}x_0\delta(t)\|^2.$$

Implementation Issues

Different methods use different choices of the design variables. The typical choices are

- The number of past inputs, s_1 and the number of past outputs, s_2 in the instrument matrix (2.9). Typically these are set to $s_1 = s_2 = s$ to reduce the problem to one design parameter. Model validation techniques could be used to find a good value of s.
- The maximal prediction horizon r is set to $r = s$ in many algorithms.

- The matrix $\mathbf{G}$ is usually weighted before the SVD step (2.10), *i.e.,* $\hat{\mathbf{G}} = \mathbf{W}_1\mathbf{G}\mathbf{W}_2 = \mathbf{U}\mathbf{S}\mathbf{V}^T$. The choice of the weighting matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ is one of the most important for the accuracy of the method. Different algorithms use different weights. See Van Overschee and De Moor (1996) for an overview of existing methods.

2.4 Nonparametric Frequency Response Estimation

Consider the asymptotically stable linear time-invariant system

$$\begin{aligned} x(t+1) =& \mathbf{A}x(t) + \mathbf{B}u(t) \\ y(t) =& \mathbf{C}x(t) + v(t) \end{aligned} \tag{2.11}$$

where $u(t)$ is the input signal, $y(t)$ is the output signal of the system and $v(t)$ is measurement noise. The Frequency Response Function (FRF) of system (2.11) is defined as

$$G_0(e^{j\omega}) = \sum_{t=0}^{\infty} g_t^0 e^{-j\omega t} = \sum_{t=1}^{\infty} \mathbf{C}\mathbf{A}^t\mathbf{B}e^{-j\omega t}. \tag{2.12}$$

where $g_t^0 = \mathbf{C}\mathbf{A}^t\mathbf{B}$ is the impulse response of the system.

The objective is to find a nonparametric estimate of the FRF, $G(e^{j\omega_k})$ from input-output data $\{u(t), y(t), t = 1, \ldots, N\}$, for all frequencies on the frequency grid given by

$$\omega_k = \frac{2\pi}{N}k, \quad k = 0, \ldots, N-1.$$

If N samples of the input and output from system (2.11) are available then applying the N-point Discrete Fourier Transform (DFT)

$$Z_N(\omega_k) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} z(t)e^{-j\omega_k t}$$

to the input and output data gives the relationship

$$\begin{aligned} Y_N(\omega_k) =& G_0(e^{j\omega_k})U_N(\omega_k) + \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y_t^{\text{tra}} e^{j\omega_k t} + V_N(\omega_k) \\ =& G_0(e^{j\omega_k})U_N(\omega_k) + T_N(e^{j\omega_k}) + V_N(\omega_k) \end{aligned} \tag{2.13}$$

where $y_t^{\text{tra}} = \mathbf{C}\mathbf{A}^t(x_0 - x_0^{\text{per}})$, x_0 is the initial condition of system (2.11) and x_0^{per} is the initial state that would lead to a periodic output. The term $T_N(e^{j\omega_k})$ is the transient or leakage and is due to initial and end effects induced by the finite measurement time. See Pintelon and Schoukens (2001) and McKelvey (2000) for details.

In Chapter 6 the relation (2.13) will be used to develop a new method, the Transient and Impulse Modeling Method, TRIMM, but it can also be used to understand existing methods.

Classical Methods

Many different approaches and methods for estimating the Frequency Response Function have been proposed in the literature over the years. In this section some classical methods still widely in use will be presented.

Empirical Transfer Function Estimate

The simplest method to estimate the frequency response function is to divide the DFT of the output sequence and the DFT of the input sequence. This is called the Empirical Transfer Function Estimate or ETFE. Using relation (2.13) the resulting estimate is given by

$$\hat{G}^{\text{ETFE}}(e^{j\omega_k}) = \frac{Y_N(\omega_k)}{U_N(\omega_k)} = G_0(e^{j\omega_k}) + \frac{T_N(e^{j\omega_k})}{U_N(\omega_k)} + \frac{V_N(\omega_k)}{U_N(\omega_k)}. \tag{2.14}$$

Equation (2.14) reveals the problem with this method. Even if there is no noise ($V_N(\omega_k) = 0$) an error is introduced by the leakage or transient term $T_N(\omega_k)$. Furthermore since there is no smoothing or averaging the estimate will be sensitive to noise. The quality of this estimate is well studied, see for example Ljung (1999) or Pintelon and Schoukens (2001) and can be separated into two cases; the first when the input signal is periodic and the second case when the input signal is a realization of a stochastic process.

The main results when the input signal $u(t)$ is periodic and N is a multiple of the period are:

- The estimate $\hat{G}^{\text{ETFE}}(e^{j\omega_k})$ is only defined for a fixed number of frequencies.
- The ETFE is unbiased and its variance decays as $1/N$.

The main results when the input signal $u(t)$ is a realization of a stochastic process are:

- The ETFE is an asymptotically, in N, unbiased estimate of $G_0(e^{j\omega_k})$.
- The variance of the ETFE is given by the noise-to-signal ratio at the frequency. The variance does not decrease as N increases since there is no smoothing.
- The estimation error can be arbitrary large since for random excitation, the amplitude $|U(\omega_k)|$ can be arbitrary small, *cf.,* (2.14).
- The estimates at different frequencies are asymptotically uncorrelated.

Spectral Analysis

A common approach to counteract the problem with the possible unbounded variance of the ETFE estimate is to average the signals before the division. This method was proposed by Bartlett (1948). The N data is divided into p bins, and for each bin the discrete Fourier transform is calculated. The Bartlett estimate is then

$$\hat{G}^{\mathrm{B}}(e^{j\omega_k}) = \frac{\hat{\Phi}_{uy}}{\hat{\Phi}_u}$$

where

$$\hat{\Phi}_{uy} = \frac{1}{p}\sum_{i=1}^{p} Y^{(i)}(\omega_k)U^{(i)}(\omega_k)^*$$
$$\hat{\Phi}_{u} = \frac{1}{p}\sum_{i=1}^{p} U^{(i)}(\omega_k)U^{(i)}(\omega_k)^*$$

are estimates of the spectrum and cross spectrum. The properties of this estimate are well documented, see for example Broersen (1995) or Pintelon and Schoukens (2001). Ljung (1999) observed that this estimate is the same as taking the weighted average of the ETFE calculated over the p bins, *i.e.,*

$$\hat{G}^{\mathrm{B}}(e^{j\omega_k}) = \frac{\sum_{i=1}^{p} \hat{G}^{\mathrm{ETFE}}_{(i)}(e^{j\omega_k})|U^{(i)}(\omega_k)|^2}{\sum_{i=1}^{p}|U^{(i)}(\omega_k)|^2}.$$

Other choices of weights, instead of $|U^{(i)}(\omega_k)|^2$, are possible in the sum. An study of how different weights affects the estimates can be found in Heath (2007).

With Bartlett's method the variance of the estimate is reduced by a factor of $1/p$ but this comes at a cost of lower resolution. The resolution is reduced by a factor p since the data is split up into p bins. The choice of the number of bins is hence a trade-off between resolution and variance of the estimate.

An extension of the Barlett method is the Welch method (Welch, 1967). The difference is that the data bins in Welch method are allowed to overlap. This gives slightly different properties of the estimate.

The averaging methods are design to reduce the variance of the estimates, still there is the problem with the leakage. To reduce this effect, windowing methods, such as the Blackman-Tukey method was introduced (Blackman and Tukey, 1958). The idea is to use a time window to reduce the initial and end effects caused by the finite data length which gives rise to the leakage error. The Blackman-Tukey estimate of the frequency response function is

$$\hat{G}^{\mathrm{BT}}(e^{j\omega_k}) = \frac{\hat{\Phi}_{uy}}{\hat{\Phi}_u}$$

where now

$$\hat{\Phi}_{uy} = \sum_{\tau=-M}^{M} \hat{R}_{yu}(\tau) w_M(\tau) e^{-j\omega_k t},$$
$$\hat{\Phi}_{u} = \sum_{\tau=-M}^{M} \hat{R}_{u}(\tau) w_M(\tau) e^{-j\omega_k t}.$$

The terms $\hat{R}_u$ and $\hat{R}_{yu}$ are the covariance of $u(t)$ and the cross-covariance between $u(t)$ and $y(t)$, respectively. $w_M(\tau)$ is the window function of width M. The properties of the estimate are directly related to the choice of the window function.

- The length of the window, M, is a tradeoff between resolution and variance of the estimate.
- The shape of the window is a tradeoff between smearing the estimate and the effect of the leakage error.

Numerous windows, trading off the previous mentioned errors in different ways, have been proposed in the literature. An overview of the different windows and their effects can be found in most standard books on spectral analysis, see for example Bendat and Piersol (1980); Stoica and Moses (2005).

Local Polynomial Method

Recently the Local Polynomial Method (LPM) were presented as an alternative to the classical frequency smoothing methods (Schoukens et al., 2006). It uses the assumption that the frequency response and the leakage term, *cf.,* (2.13), both are smooth function of the frequency. The idea is to approximate these by Taylor series expansion and to simultaneously estimate the coefficients of the expansion together with the frequency response at one frequency.

The frequency response function and the leakage term can be approximated at the frequency ω_{k+r} with the following Taylor series

$$\begin{aligned} G_0(e^{j\omega_{k+r}}) =& G_0(e^{j\omega_k}) + \sum_{s=1}^{R} g_s(k) r^s + \mathcal{O}\left(\left(\frac{r}{N}\right)^{R+1}\right) \\ T_N(e^{j\omega_{k+r}}) =& T_N(e^{j\omega_k}) + \sum_{s=1}^{R} t_s(k) r^s + \frac{1}{\sqrt{N}}\mathcal{O}\left(\left(\frac{r}{N}\right)^{R+1}\right). \end{aligned}$$

Applying the Taylor approximation in (2.13), neglecting the higher order terms, gives the approximation

$$\begin{aligned} Y_N(\omega_{k+r}) =& G_0(e^{j\omega_k}) U_N(\omega_{k+r}) + \left(\sum_{s=1}^{R} g_s(k) r^s\right) U_N(\omega_{k+r}) \\ &+ T_N(e^{j\omega_k}) + \sum_{s=1}^{R} t_s(k) r^s + V_N(\omega_{k+r}). \end{aligned} \tag{2.15}$$

To estimate the frequency response function, $G_0(e^{j\omega_k})$, from (2.15) the $2n$ neighboring frequencies to ω_k are used, *i.e.,* by setting $r = 0, \pm 1, \ldots, \pm n$.

Collecting the parameters $G_0(e^{j\omega_k})$, $T_N(e^{j\omega_k})$, the Taylor series coefficients $g_s(k)$ and $t_s(k)$ in a vector θ and their respective coefficients in a matrix $\mathbf{K}_n$ for $r = 0, \pm 1, \ldots, \pm n$, (2.15) can be written on vector form as

$$Y_n = \mathbf{K}_n \theta + V_n \tag{2.16}$$

where Y_n and V_n are vectors with $Y(k+r)$ and $V(k+r)$, respectively, stacked up for $r = 0, \pm 1, \ldots, \pm n$.

The estimate of the FRF, $G_0(e^{j\omega_k})$ and the extra parameters, are calculated by solving (2.16) in a least squares sense. This is then repeated for all frequencies $k = 0, \ldots, N-1$ giving N local least squares problem to solve.

The design parameters for the LPM are the number of neighboring frequencies, n, and the order of the Taylor series expansions R. To get more equations than unknowns in the least squares problem the number of neighboring frequencies has to satisfy $n \geq R + 1$. The optimal choice of these parameters depends on the true system. However, Schoukens et al. (2009) argues that a reasonable choice is $R = 2$ and $n = 3$ in many cases. In this thesis these recommended values will be used in the comparison with the proposed method.

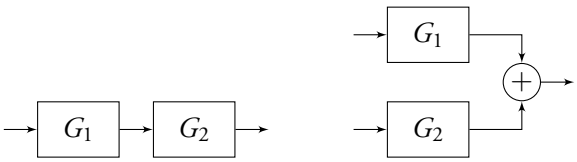
The local polynomial method has been analyzed in a series of papers (Pintelon et al., 2010a,b). The method has been demonstrated to provide superior accuracy on a number of problems compared to traditional smoothing algorithms.

Chapter 3

Structured Systems

Recent advances in small and cheap communication and sensing have opened up for large scale systems with intricate interconnections and interactions. Examples of interconnected systems can be found in many different technological areas; power distribution systems, communication networks, formations control of unmanned aerial vehicles and in the manufacturing industry to name only a few. These applications pose new challenges for analysis and control design and hence accurate models are needed.

The systems that will be studied in this thesis are feedforward systems built up by the basic building blocks shown in Figure 3.1. Two special cases of structured systems will be considered; the cascade system and the parallel cascade serial system. The motivation for this choice is firstly that they both are common in industrial applications as will be shown. Secondly the insights and results obtained from these structures can be used to understand more general and larger feed forward structured systems since many systems can be constructed from these basic building blocks.



(a) Cascade interconnection (b) Parallel interconnection

Figure 3.1 The basic building blocks of structured systems considered in this thesis.

In system identification dealing with structured systems introduces two main problems:

1. if the structure and the interconnections of the subsystems are unknown a priori the task is to identify the unknown structure of the system from input and output measurements,

2. if the overall structure of the system is known but the dynamics of the individual subsystems are unknown, the problem is to identify the subsystems from input-output measurements and to analyze the properties of the estimate.

This thesis will focus on the second problem when the overall structure is assumed known. The objective in this chapter is to analyze the fundamental limitations of the quality of the identified model when using the Prediction Error Method for structured systems.

3.1 Cascade Systems

Cascade systems are common in industrial applications such as process control or control of servo mechanics. A system with cascade, or serial, structure is illustrated in Figure 3.2. In process control the second output y_2 is often the quantity one would like to control, for example temperature or level while the first output is some intermediate variable such as flow or pressure. In servo applications the first output is usually some rate while the second output is a position. These systems are often controlled by a cascade controller (Åström and Hägglund, 2006).

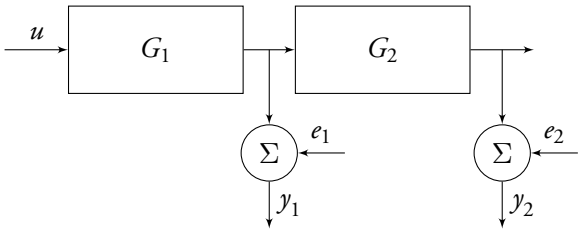


Figure 3.2 Cascade system.

The input-output relations for a cascade structured system are

$$\begin{aligned} y_1(t) &= G_1(q)\,u(t) + e_1(t) \\ y_2(t) &= G_1(q)\,G_2(q)\,u(t) + e_2(t)\,. \end{aligned} \tag{3.1}$$

The input signal is denoted $u(t)$ and the output signals, $y_1(t)$ and $y_2(t)$, respectively. The signals $e_1(t)$ and $e_2(t)$ are measurement noise processes.

Fundamental limitations

In a series of papers (Wahlberg et al., 2009a,b) the fundamental limitation of the quality of the estimates is studied when PEM is used on a cascade structured systems. The main result is that for cascade structured systems with two identical subsystems there is no improvement in the estimation of the first subsystem obtained from the second measurement. Here the results will be summarized.

The model structure used in the identification is

$$\begin{aligned} y_1(t) &= G_1(q,\theta_1)\,u(t) + e_1(t) \\ y_2(t) &= G_1(q,\theta_1)\,G_2(q,\theta_2)\,u(t) + e_2(t) \end{aligned} \tag{3.2}$$

where $e_1(t)$ and $e_2(t)$ are assumed to be zero mean white Gaussian measurement noise processes with known variances λ_1 and λ_2, respectively. Furthermore it is assumed that the true underlying system belongs to the model set (3.2) with the true parameters θ_1^0 and θ_2^0.

Given a data set $\{u(t), y_1(t), y_2(t), t = 1, \ldots, N\}$ the PEM estimates of the model parameters are given by, see Section 2.2,

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \arg\min_{\theta_1,\theta_2} \left(\frac{1}{N}\sum_{t=1}^{N} \frac{[y_1(t) - G_1(q,\theta_1)u(t)]^2}{\lambda_1} + \frac{1}{N}\sum_{t=1}^{N} \frac{[y_2(t) - G_2(q,\theta_2)G_1(q,\theta_1)u(t)]^2}{\lambda_2} \right).$$

If the two systems are identical and the same structure is used for both transfer functions, *i.e.,*

$$G_1^0 = G_2^0$$

and

$$G_2(q,\theta_2^0)G_1'(q,\theta_1^0) = G_2'(q,\theta_2^0)G_1(q,\theta_1^0)$$

where $G'(q,\theta) = \frac{\partial G}{\partial \theta}$, then the asymptotic, in the number of data, covariance is given by

$$\begin{aligned} \operatorname{Cov}\hat{\theta}_1 &\sim \mathbf{A}^{-1} \\ \operatorname{Cov}\hat{\theta}_2 &\sim \mathbf{A}^{-1} + \mathbf{D}^{-1} \end{aligned} \tag{3.3}$$

where

$$\mathbf{A} = \frac{N}{\lambda_1} E\left\{[G_1'(q,\theta_1)u(t)][G_1'(q,\theta_1)u(t)]^T\right\},$$
$$\mathbf{D} = \frac{N}{\lambda_2} E\left\{[G_2'(q,\theta_2)G_1(q,\theta_1)u(t)][G_2'(q,\theta_2)G_1(q,\theta_1)u(t)]^T\right\}.$$

The notation $\sim$ is used to stress the asymptotic relation. If the true subsystems are identical the conclusions form (3.3) are

- Since the matrix $\mathbf{A}^{-1}$ is the asymptotic covariance matrix of $\hat{\theta}_1$ when only $y_1(t)$ is measured, the quality of the estimate $\hat{\theta}_1$ is not improved by also measuring $y_2(t)$.
- The matrix $\mathbf{D}^{-1}$ is positive definite, hence $\mathbf{A}^{-1} + \mathbf{D}^{-1} \succ \mathbf{A}^{-1}$ and the quality of the estimate $\hat{\theta}_2$ will always be worse than the quality of $\hat{\theta}_1$.

If it is known beforehand that the two subsystems are equal, this information should be incorporated in the model structure. If this is done it can be shown that the asymptotic variance of the estimate becomes

$$\operatorname{Cov}\hat{\theta}_1 \sim (\mathbf{A} + 4\mathbf{D})^{-1}$$

see Wahlberg et al. (2009a) for details. Since $\mathbf{A}$ and $\mathbf{D}$ are positive definite matrices the variance could potentially be much smaller in this case.

3.2 Parallel Serial Systems

Now a more general structure will be considered, the parallel serial structure shown in Figure 3.3. This class of systems generalizes cascaded systems. If the number of parallel systems is $p = 1$ it boils down to the cascade structure. The motivation for studying this structure was partly a collaboration with Honeywell Prague Laboratory on boiler control as shown in Example 1.1, see *e.g.,* Baramov et al. (2007). In this case the parallel systems are a set of boilers feeding steam into a common header, G_{p+1}, which then, for example, feeds the steam to a turbine. But this structure can be found in many other industrial applications, typical examples are parallel working pumps, turbines, chemical reactors and so on.

Several important questions have to be answered and user choices have to be made when applying system identification to this special structure.

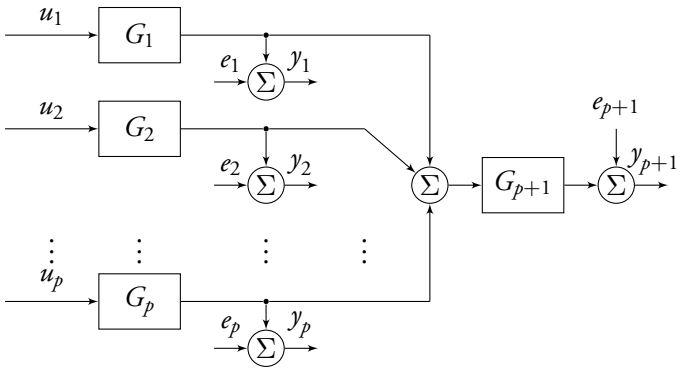


Figure 3.3 A parallel serial structured system.

- Contrary to the cascade structured system, the parallel serial system has multiple inputs. This raises new questions. Which input signals should be applied and how should the total power be distributed among the p inputs?

- Do the dynamics of the subsystems create any fundamental limitation for the identification result as it does for cascade systems?

- If only a subset of the $p + 1$ output signals can be measured and with different sensor qualities, where should the sensors be located for best result?

3.3 Fundamental limitations

To answer these questions, variance results for identification of parallel serial structured systems will be derived and the implications for input and sensor design will be discussed. Mainly the case of two parallel subsystems together with a common serial sub-system will be analyzed. To get some insight into the problem, first, simple FIR systems are considered. Finally some of the results are extended to larger structures with general linear dynamics.

Variance Analysis: FIR Example

Consider the special case with three subsystems in Figure 3.4. Here the model structure

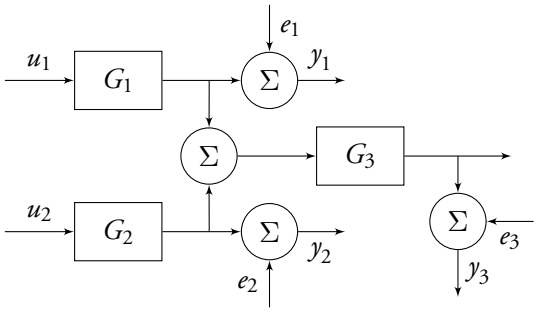


Figure 3.4 Three subsystem parallel serial system.

can be written as

$$\begin{aligned} y_1(t) &= G_1(q,\theta_1)u_1(t) + e_1(t), \\ y_2(t) &= G_2(q,\theta_2)u_2(t) + e_2(t), \\ y_3(t) &= G_3(q,\theta_3)\left(G_1(q,\theta_1)u_1(t) + G_2(q,\theta_2)u_2(t)\right) + e_3(t). \end{aligned} \tag{3.4}$$

To illustrate the analysis the subsystems are assumed to be first order FIR transfer functions, *i.e.,*

$$\begin{aligned} G_1(q,\theta_1) &= 1 + b_1 q^{-1}, \quad \theta_1 = b_1, \\ G_2(q,\theta_2) &= 1 + b_2 q^{-1}, \quad \theta_2 = b_2, \\ G_3(q,\theta_3) &= 1 + b_3 q^{-1}, \quad \theta_3 = b_3. \end{aligned} \tag{3.5}$$

The objective is to identify the FIR parameters b_1, b_2 and b_3 from measurement of $\{u_1(t), u_2(t), y_1(t), y_2(t), y_3(t), t = 1, \ldots, N\}$. Again it is assumed that the true system can be described by the model structure (3.4) with the true FIR parameters b_1^0, b_2^0 and b_3^0, respectively. Furthermore, it is assumed that $e_1(t)$, $e_2(t)$ and $e_3(t)$ are independent white measurement noise processes with given, known, variances λ_1, λ_2 and λ_3, respectively.

To start, the input signals $u_1(t)$ and $u_2(t)$ are chosen as independent white noise processes with variances 1 as this is a common choice in practice. Given the data set $\{u_1(t), u_2(t), y_1(t), y_2(t), y_3(t), t = 1, \ldots, N\}$ the PEM estimate of the model

parameters is given by, see Section 2.2,

$$\begin{aligned}\begin{pmatrix}\hat{b}_1\\ \hat{b}_2\\ \hat{b}_3\end{pmatrix} &= \arg\min_{b_1,b_2,b_3}\left[\frac{1}{N}\sum_{t=1}^{N}\frac{\left[y_1(t)-u_1(t)-b_1u_1(t-1)\right]^2}{\lambda_1}\right.\\ &+\frac{1}{N}\sum_{t=1}^{N}\frac{\left[y_2(t)-u_2(t)-b_2u_2(t-1)\right]^2}{\lambda_2}\\ &+\frac{1}{N}\sum_{t=1}^{N}\frac{1}{\lambda_3}\left[y_3(t)-u_1(t)-u_2(t)-(b_1+b_3)u_1(t-1)-\right.\\ &\left.\left.(b_2+b_3)u_2(t-1)-b_1b_3u_1(t-2)-b_2b_3u_2(t-2)\right]^2\right].\end{aligned} \tag{3.6}$$

The asymptotic (in N) covariance matrix of the parameter estimates, which in this case corresponds to the Cramér-Rao lower bound, is given by

$$\mathbf{P}_\theta = \frac{1}{N}\left[E\left\{\Psi(t)\Psi^T(t)\right\}\right]^{-1},$$

where

$$\Psi(t) = \begin{bmatrix} \frac{u_1(t-1)}{\sqrt{\lambda_1}} & 0 & \frac{u_1(t-1)+b_3^0u_1(t-2)}{\sqrt{\lambda_3}} \\ 0 & \frac{u_2(t-1)}{\sqrt{\lambda_2}} & \frac{u_2(t-1)+b_3^0u_2(t-2)}{\sqrt{\lambda_3}} \\ 0 & 0 & \frac{u_1(t-1)+u_2(t-1)+b_1^0u_1(t-2)+b_2^0u_2(t-2))}{\sqrt{\lambda_3}} \end{bmatrix}$$

is the normalized predictor gradient. Due to the inverse, the expressions for the parameter variances, that is the diagonal elements of $\mathbf{P}_\theta$ are lengthy and difficult to access, instead some special cases will be studied. Again the notation $\sim$ is used to stress that the relations are asymptotic in the number of data.

Preliminaries

If only the outputs from the parallel subsystems $y_1(t)$ and $y_2(t)$ are used the asymptotic variances of the estimates of b_1 and b_2 equal

$$\begin{aligned}\operatorname{Var}\hat{b}_1 &\sim \frac{\lambda_1}{N},\\ \operatorname{Var}\hat{b}_2 &\sim \frac{\lambda_2}{N}.\end{aligned} \tag{3.7}$$

This corresponds to setting $\lambda_3 = \infty$ in the general expression for $\Psi(t)$ and $\mathbf{P}_\theta$. Since no information about the last subsystem is available, it is not possible to estimate b_3 in this case.

In case the quality of the two sensors $y_1(t)$ and $y_2(t)$ are much worse than for the last output $y_3(t)$, *i.e.,* letting $\lambda_1 \to \infty$ and $\lambda_2 \to \infty$, the asymptotic variance will be

$$\operatorname{Var} \hat{b}_1 \sim \frac{\lambda_3}{N} \frac{(2 + (b_1^0)^2 + (b_2^0)^2) - \frac{(1+b_2^0 b_3^0)^2}{1+(b_3^0)^2}}{(b_3^0 - b_1^0)^2 + (b_3^0 - b_2^0)^2}, \tag{3.8a}$$

$$\operatorname{Var} \hat{b}_3 \sim \frac{\lambda_3}{N} \frac{1 + (b_3^0)^2}{(b_3^0 - b_1^0)^2 + (b_3^0 - b_2^0)^2}, \tag{3.8b}$$

which will be large if $b_1^0 \approx b_2^0 \approx b_3^0$. Due to symmetry the expression for the variance of $\hat{b}_2$ is the same as the expression (3.8a) with b_1^0 exchanged for b_2^0 and vice versa. From the expressions (3.8) it is seen that it is possible to estimate both subsystems accurately, measuring only the output from the last subsystem, as long as the subsystems dynamics are sufficiently different and the quality of the measurement is good enough (*i.e.,* λ_3 is small enough). It seems that common dynamics of the subsystems could affect the results negatively. Let us look at this a bit closer.

Subsystems with common dynamics

As shown in Section 3.1, for cascade structured systems with two identical transfer functions, not taking this into account in the estimation, nothing is gained in the estimation of the first subsystem by also measuring the second output. For this parallel serial system the variance of the parameter estimate when the three subsystems are identical, *i.e.,* $b_2^0 = b_3^0 = b_1^0$ becomes

$$\operatorname{Var} \hat{b}_1 \sim \frac{1}{N} \left(\lambda_1 - \frac{\lambda_1^2 (1 + (b_1^0)^2)}{2\lambda_3 + (\lambda_1 + \lambda_2)(1 + (b_1^0)^2)} \right) \leq \frac{\lambda_1}{N}. \tag{3.9}$$

Contrary to cascade systems, the estimates of b_1 and b_2 are improved by also measuring $y_3(t)$, *cf.,* (3.7).

However, if again the subsystems are equal, but this time $y_2(t)$ is not measured

($\lambda_2 = \infty$) then

$$\begin{aligned}
\operatorname{Var} \hat{b}_1 &\sim \frac{\lambda_1}{N}, \\
\operatorname{Var} \hat{b}_2 &\sim \frac{1}{N}\left(\lambda_1 + \frac{2\lambda_3}{1+(b_1^0)^2}\right), \\
\operatorname{Var} \hat{b}_3 &\sim \frac{1}{N}\left(\lambda_1 + \frac{\lambda_3}{1+(b_1^0)^2}\right).
\end{aligned} \tag{3.10}$$

The expressions for $\hat{b}_1$ and $\hat{b}_3$ are exactly the expressions obtained by Wahlberg et al. (2009a) for cascade structured systems. Hence when the subsystems are equal and $y_2(t)$ is not measured, nothing is gained in the estimation of b_1 by also measuring y_3 since the expression is the same as if b_1 would be estimated using only y_1, see (3.7).

An interesting observation is that the variance for the parameter $\hat{b}_2$ belonging to the unmeasured subsystem always is larger ($\lambda_3 > 0$) than the variance for the two other parameter estimates. It is tempting to think that this always is true, that if only one of the outputs from the parallel systems can be measured, you should measure the output corresponding to the subsystem that are of most interest. However this observation is not true for general dynamic of the subsystems. It is quite easy to find cases such that $\operatorname{Var} \hat{b}_1 > \operatorname{Var} \hat{b}_2$ when $y_2(t)$ is not measured. If $y_2(t)$ is not measured then

$$\begin{aligned}
&\lambda_1\lambda_3 b_3^0(b_1^0 - b_2^0)(2 + b_1^0 b_3^0 + b_2^0 b_3^0) - \lambda_3^2(2 + (b_1^0)^2 + (b_2^0)^2) > 0 \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Leftrightarrow \\
&\qquad\qquad\qquad\qquad\qquad\qquad \operatorname{Var} \hat{b}_1 - \operatorname{Var} \hat{b}_2 > 0.
\end{aligned}$$

If $\lambda_1 = \lambda_3 = 1$ and $b_1^0 = 3$, $b_2^0 = -0.5$, $b_3^0 = 2$ then

$$\operatorname{Var} \hat{b}_1 - \operatorname{Var} \hat{b}_2 \approx \frac{0.40}{N}.$$

Hence, if only one of the two output signals from the first subsystems can be measured, it is not always best to follow the intuition and measure the output from the system corresponding to the parameter one wants to identify.

If instead only two subsystems are equal, $b_3^0 = b_1^0$, $b_2^0 \neq b_1^0$ and $y_2(t)$ is not measured ($\lambda_2 = \infty$) then

$$\operatorname{Var} \hat{b}_1 \sim \frac{1}{N}\left[\lambda_1 - \frac{\lambda_1^2}{\lambda_1 + \lambda_3 \frac{(b_1^0 - b_2^0)^2 + (b_1^0)^2 + 1}{((b_1^0)^2+1)(b_1^0 - b_2^0)}}\right],$$

which always is smaller than the expression (3.10) obtained when all three subsystems are equal.

If it is known in advance that $G_3 = G_2 = G_1$, one should incorporate this information to further constrain the model. In this case this means that only one parameter, b_1, should be estimated. The asymptotic variance, when y_2 is not measured, is

$$\operatorname{Var} \hat{b}_1 \sim \frac{1}{N} \frac{\lambda_1}{1 + \frac{8\lambda_1(1+(b_1^0)^2)}{\lambda_3}}.$$

which could be considerably lower than (3.7).

Effects of input signals

Up to now the input signals have been assumed to be independent of each other. The question is: can the estimation be improved by using correlated input signals? For the case when the same input signal is used as inputs to both subsystems, *i.e.,* $u_1(t) = u_2(t)$ and that $y_2(t)$ is not measured ($\lambda_2 = \infty$), the asymptotic variance for $\hat{b}_1$ equals

$$\operatorname{Var} \hat{b}_1 \sim \frac{\lambda_1}{N}.$$

Hence, nothing is gained by measuring $y_3(t)$ if the objective is to only estimate b_1.

However, if the aim of the identification is to identify b_3, this input signal could be good in some cases. Consider the case when the second input is not measured, *i.e.,* $\lambda_2 = \infty$ and the true system is given by $\lambda_1 = \lambda_3 = b_1^0 = 1$, $b_2^0 = 2$ and $b_3^0 = 3$. The variance when the input signals are independent white noise sequences is given by

$$\operatorname{Var} \hat{b}_3 \sim \frac{1}{N} \frac{110}{71} \approx \frac{1.55}{N},$$

and for the case $u_1(t) = u_2(t)$

$$\operatorname{Var} \hat{b}_3 \sim \frac{1}{N} \frac{10}{9} \approx \frac{1.11}{N}.$$

Here the variance is smaller when using the same input signal to both subsystems. Consider instead $\lambda_1 = \lambda_3 = b_3^0 = 1$, $b_2^0 = 2$ and $b_1^0 = 3$ then for the independent input signal case the variance is

$$\operatorname{Var} \hat{b}_3 \sim \frac{1}{N} \frac{6}{31} \approx \frac{0.19}{N},$$

and when the same input signal is applied to both subsystems the variance is given by

$$\operatorname{Var} \hat{b}_3 \sim \frac{1}{N}\frac{2}{9} \approx \frac{0.22}{N}.$$

These simple examples above show the importance of choosing input signals depending on what one wants to identify. More about how the input signal affects the quality of the estimate for systems with multiple inputs but only one output can be found in Gevers et al. (2006).

3.4 General number of subsystems

For the simple FIR systems it was shown that if the systems have the same dynamics and only one of the outputs from the parallel subsystems were measured, this influences the estimation performance. Here these results will be extended to a system with arbitrary number of parallel subsystems and not necessary FIR-type systems.

Consider now the general parallel serial structure in Figure 3.3. The corresponding transfer functions are

$$\begin{aligned} y_1(t) &= G_1(q)u_1(t) + e_1(t), \\ &\vdots \\ y_p(t) &= G_p(q)u_p(t) + e_p(t), \\ y_{p+1}(t) &= G_{p+1}(q)\sum_{i=1}^{p} G_i(q)u_i(t) + e_{p+1}(t), \end{aligned}$$

The system has p inputs and $p+1$ outputs. All transfer functions $G_k(q)$, $k = 1, \ldots, (p+1)$, are assumed stable.

Given a data set $\{u_1(t), \ldots, u_p(t), y_1(t), \ldots, y_{p+1}(t), t = 1, \ldots, N\}$ the PEM estimate of the model parameters,

$$\begin{aligned} \theta_1 &= \left[\theta_{1,1}, \theta_{1,2}, \ldots, \theta_{1,n_1}\right]^T, \\ &\vdots \\ \theta_{p+1} &= \left[\theta_{p+1,1}, \theta_{p+1,2}, \ldots, \theta_{p+1,n_{p+1}}\right]^T, \end{aligned}$$

where n_i denotes the number of parameters in θ_i, is given by

$$\begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_{p+1} \end{pmatrix} = \arg \min_{\theta_1,\ldots,\theta_{p+1}} \left[\frac{1}{N} \sum_{t=1}^{N} \frac{\left[y_1(t) - G_1(q,\theta_1) u_1(t) \right]^2}{\lambda_1} + \right.$$
$$+ \ldots$$
$$+ \frac{1}{N} \sum_{t=1}^{N} \frac{\left[y_p(t) - G_p(q,\theta_p) u_p(t) \right]^2}{\lambda_p}$$
$$\left. + \frac{1}{N} \sum_{t=1}^{N} \frac{\left[y_{p+1}(t) - G_{p+1}(q,\theta_{p+1}) \left(\sum_{i=1}^{p} G_i(q,\theta_i) u_i(t) \right) \right]^2}{\lambda_{p+1}} \right].$$

Define the $(p+1) \times (p+1)$ block matrix

$$\Psi(t) = \begin{bmatrix} \frac{G_1'(q,\theta_1) u_1(t)}{\sqrt{\lambda_1}} & \cdots & \mathbf{0} & \frac{G_{p+1}(q,\theta_3) G_1'(q,\theta_1) u_1(t)}{\sqrt{\lambda_{p+1}}} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \frac{G_p'(q,\theta_1) u_p(t)}{\sqrt{\lambda_p}} & \frac{G_{p+1}(q,\theta_3) G_p'(q,\theta_2) u_p(t)}{\sqrt{\lambda_{p+1}}} \\ \mathbf{0} & \cdots & \mathbf{0} & \frac{G_{p+1}'(q,\theta_{p+1})}{\sqrt{\lambda_{p+1}}} \sum_{i=1}^{p} G_i(q,\theta_i) u_i(t) \end{bmatrix}$$

where prime denotes differentiation with respect to the respective parameter vector. The asymptotic covariance matrix of the parameter estimates is then given by

$$\mathrm{Cov} \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_{p+1} \end{pmatrix} \sim \mathbf{P}_\theta$$

where $\mathbf{P}_\theta = \frac{1}{N} \left[E\{\Psi(t)\Psi(t)^T\} \right]^{-1}$.

Now consider the case when all subsystems are identical and have the same structure. Then

$$G_i(q,\theta_i^0) G_j'(q,\theta_j^0) = G_i'(q,\theta_i^0) G_j(q,\theta_j^0)$$

for $i,j = 1,\ldots,(p+1)$. This implies that the number of parameters in all subsystems is equal. If all $u_i(t)$ and $u_j(t)$ are independent for $i \neq j$ then

$$\begin{aligned}\mathbf{P}_\theta^{-1} &= NE\{\Psi(t)\Psi(t)^T\} \\ &= \begin{pmatrix} \mathbf{A}_1+\mathbf{B}_1 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{B}_1 \\ \mathbf{0} & \mathbf{A}_2+\mathbf{B}_2 & & \mathbf{0} & \mathbf{B}_2 \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{A}_p+\mathbf{B}_p & \mathbf{B}_p \\ \mathbf{B}_1 & \mathbf{B}_2 & \ldots & \mathbf{B}_p & \sum_{i=1}^{p}\mathbf{B}_i \end{pmatrix},\end{aligned} \tag{3.11}$$

where the matrices $\mathbf{A}_i$ and $\mathbf{B}_i$ are defined as

$$\begin{aligned}\mathbf{A}_i &= \frac{N}{\lambda_i} E\{[G_i'(q,\theta_i)u_i(t)][G_i'(q,\theta_i)u_i(t)]^T\}, \\ \mathbf{B}_i &= \frac{N}{\lambda_{p+1}} E\{[G_{p+1}(q,\theta_{p+1})G_i'(q,\theta_i)u_i(t)][G_{p+1}(q,\theta_{p+1})G_i'(q,\theta_i)u_i(t)]^T\}.\end{aligned}$$

If only $y_1(t)$ and $y_{p+1}(t)$ are measured, *i.e.,* $\lambda_i = \infty$ for $i = 2,\ldots,p$ then

$$\lim_{\lambda_i \to \infty} \mathbf{A}_i = \mathbf{0} \quad \forall i = 2,\ldots,p$$

and the matrix (3.11) simplifies to

$$\mathbf{P}_\theta^{-1} = \begin{pmatrix} \mathbf{A}_1+\mathbf{B}_1 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{B}_1 \\ \mathbf{0} & \mathbf{B}_2 & & \mathbf{0} & \mathbf{B}_2 \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{B}_p & \mathbf{B}_p \\ \mathbf{B}_1 & \mathbf{B}_2 & \ldots & \mathbf{B}_p & \sum_{i=1}^{p}\mathbf{B}_i \end{pmatrix}. \tag{3.12}$$

To find the parameter covariance matrix, (3.12) needs to be inverted. The matrix can be block diagonalized by the following transformation matrix

$$\mathbf{T} = \begin{pmatrix} \mathbf{I} & \mathbf{I} & \ldots & \mathbf{I} & -\mathbf{I} \\ \mathbf{0} & \mathbf{I} & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{I} \end{pmatrix},$$

where $\mathbf{I}$ denotes an identity matrix of suitable size. Applying the transformation gives the block diagonal matrix

$$\bar{\mathbf{P}}_\theta^{-1} = \mathbf{T}\mathbf{P}_\theta^{-1}\mathbf{T}^T = \begin{pmatrix} \mathbf{A}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2 & & \mathbf{0} & \mathbf{B}_2 \\ \vdots & & \ddots & & \vdots \\ \mathbf{0} & \mathbf{0} & & \mathbf{B}_p & \mathbf{B}_p \\ \mathbf{0} & \mathbf{B}_2 & \dots & \mathbf{B}_p & \sum_{i=1}^{p} \mathbf{B}_i \end{pmatrix}.$$

By direct matrix calculations it can be verified that the inverse of the lower right block is

$$\begin{pmatrix} \mathbf{B}_2 & & \mathbf{0} & \mathbf{B}_2 \\ & \ddots & & \vdots \\ \mathbf{0} & & \mathbf{B}_p & \mathbf{B}_p \\ \mathbf{B}_2 & \dots & \mathbf{B}_p & \sum_{i=1}^{p} \mathbf{B}_i \end{pmatrix}^{-1} =$$

$$\begin{pmatrix} \mathbf{B}_1^{-1} + \mathbf{B}_2^{-1} & \mathbf{B}_1^{-1} & \cdots & \mathbf{B}_1^{-1} & -\mathbf{B}_1^{-1} \\ \mathbf{B}_1^{-1} & \mathbf{B}_1^{-1} + \mathbf{B}_3^{-1} & \ddots & \vdots & -\mathbf{B}_1^{-1} \\ \vdots & \ddots & \ddots & & \vdots \\ \mathbf{B}_1^{-1} & \mathbf{B}_1^{-1} & & \mathbf{B}_1^{-1} + \mathbf{B}_p^{-1} & -\mathbf{B}_1^{-1} \\ -\mathbf{B}_1^{-1} & -\mathbf{B}_1^{-1} & \cdots & -\mathbf{B}_1^{-1} & \mathbf{B}_1^{-1} \end{pmatrix}.$$

Now $\mathbf{P}_\theta = \mathbf{T}^T\bar{\mathbf{P}}_\theta\mathbf{T}$, which gives

$$\mathbf{P}_\theta = \begin{pmatrix} \mathbf{A}_1^{-1} & \mathbf{A}_1^{-1} & \cdots & & \mathbf{A}_1^{-1} & -\mathbf{A}_1^{-1} \\ \mathbf{A}_1^{-1} & \mathbf{C} + \mathbf{B}_2^{-1} & \mathbf{C} & \cdots & \mathbf{C} & -\mathbf{C} \\ \vdots & \mathbf{C} & \mathbf{C} + \mathbf{B}_3^{-1} & \ddots & \vdots & \vdots \\ & \vdots & \ddots & \ddots & \mathbf{C} & -\mathbf{C} \\ \mathbf{A}_1^{-1} & \mathbf{C} & \cdots & \mathbf{C} & \mathbf{C} + \mathbf{B}_p^{-1} & -\mathbf{C} \\ -\mathbf{A}_1^{-1} & -\mathbf{C} & \cdots & -\mathbf{C} & -\mathbf{C} & \mathbf{C} \end{pmatrix}.$$

where $\mathbf{C} = \mathbf{A}_1^{-1} + \mathbf{B}_1^{-1}$. The results are summarized in the following proposition.

Proposition 3.1

Assume that the true subsystems are all identical and that identical submodel structures

are used. Furthermore, assume that only the signals $y_1(t)$ and $y_{p+1}(t)$ are measured. Denoting

$$\mathbf{A}_i = \frac{N}{\lambda_i} E\{[G_i'(q,\theta_i)u_i(t)][G_i'(q,\theta_i)u_i(t)]^T\},$$
$$\mathbf{B}_i = \frac{N}{\lambda_{p+1}} E\{[G_{p+1}(q,\theta_{p+1})G_i'(q,\theta_i)u_i(t)][G_{p+1}(q,\theta_{p+1})G_i'(q,\theta_i)u_i(t)]^T\},$$

the asymptotic covariance matrices of the parameter estimates equal

$$\begin{aligned} \operatorname{Cov}\hat{\theta}_1 &\sim \mathbf{A}_1^{-1}, \\ \operatorname{Cov}\hat{\theta}_i &\sim \mathbf{A}_1^{-1} + \mathbf{B}_1^{-1} + \mathbf{B}_i^{-1}, \quad i = 2,\dots,p \\ \operatorname{Cov}\hat{\theta}_{p+1} &\sim \mathbf{A}_1^{-1} + \mathbf{B}_1^{-1}. \end{aligned}$$

Remark 3.1 Due to symmetry the result in Proposition 3.1 is also valid if y_1 is not measured but instead one of the signals from the other parallel subsystems is measured; just change the index 1 to the index of the measured subsystem in the result.

The key results when all the subsystems are identical and only $y_1(t)$ and $y_{p+1}(t)$ are measured are:

- Since $\mathbf{A}_1^{-1}$ is the asymptotic covariance matrix of $\hat{\theta}_1$ when only $y_1(t)$ is measured, the quality of the estimate of θ_1 is not improved by also measuring $y_{p+1}(t)$.
- The asymptotic variance for $\hat{\theta}_{p+1}$ is always worse than for $\hat{\theta}_1$ since $\mathbf{B}_1^{-1}$ is a positive definite matrix.
- The asymptotic variances for $\hat{\theta}_i$, $i = 2,\dots,p$ are always worse than both the one for $\hat{\theta}_1$ and the one for $\hat{\theta}_{p+1}$ since $\mathbf{B}_i^{-1}$ are positive definite matrices.
- The estimates of $\hat{\theta}_i$, $i = ,\dots,p+1$ all contains terms of $\mathbf{A}_1^{-1}$. Hence if the quality of the measurement of $y_1(t)$ is bad, *i.e.,* λ_1 is large, the estimates of $\hat{\theta}_i$, $i = 2,\dots,p$ and $\hat{\theta}_{p+1}$ are inaccurate irrespective of how good our measurement of $y_{p+1}(t)$ is and irrespective of how much input power that is used in $u_i(t)$ for $i = 2,\dots,p$.
- If it is known beforehand that all the systems are identical, this information should be incorporated in the model structure. If this information is used in the PEM parameter estimate, the asymptotic covariance matrix $\mathbf{P}_\theta$ can be calculated using

$$\Psi(t) = \left[\frac{G_1'(q,\theta_1)u_1(t)}{\lambda_1}, \quad \dots \quad , \frac{G_1'(q,\theta_1)u_p(t)}{\lambda_p}, \frac{2G_1'(q,\theta_1)G_1(q,\theta_1)\sum_{i=1}^{p} u_i(t)}{\lambda_{p+1}}\right]$$

in $P_\theta = \frac{1}{N} E\{\Psi(t)\Psi(t)^T\}^{-1}$. This gives

$$\operatorname{Cov} \hat{\theta}_1 \sim \left(\mathbf{A}_1 + 4 \sum_{i=1}^{p} \mathbf{B}_i \right)^{-1}$$

which is smaller than A_1^{-1} which is the asymptotic covariance matrix if the structural information is not used. Hence it is possible to improve the estimates if the structural information is used during the identification.

- The asymptotic covariances $\mathbf{A}_i$ and $\mathbf{B}_i$ are directly proportional to the power in the input signal $u_i(t)$. Assume that there is some constraint on the total power that can be used to excite the system, and the aim is to estimate θ_1 or θ_{p+1}, all input power should be used in $u_1(t)$.

  On the other hand, if the goal is to estimate θ_i for $i = 2, \ldots, p$ the input signal should be divided between $u_1(t)$ and $u_i(t)$. How the power should be distributed is decided by the relations between $\mathbf{A}_1$, $\mathbf{B}_1$ and $\mathbf{B}_i$. If independent white noise signals are used one could for example formulate the following optimization problem to find the signal powers σ_1^2 and σ_i^2 for $u_1(t)$ and $u_i(t)$ respectively

$$\begin{aligned} \min_{\sigma_1^2, \sigma_i^2} \quad & \operatorname{tr} \left(\frac{\mathbf{A}_1^{-1}}{\sigma_1^2} + \frac{\mathbf{B}_1^{-1}}{\sigma_1^2} + \frac{\mathbf{B}_i^{-1}}{\sigma_i^2} \right) \\ \text{s. t.} \quad & \sigma_1^2 + \sigma_i^2 = S_{tot} \end{aligned}$$

  where S_{tot} is the maximum input power and $\mathbf{A}_1$, $\mathbf{B}_1$ and $\mathbf{B}_i$ are calculated with white noise inputs with unit variance.

3.5 Summary

Structured systems are common in many industrial applications. Therefore, it is interesting to study the fundamental limitations when identifying these kinds of systems. The previous work on cascade systems were extended to a more general class of systems, the parallel serial structure. The motivation for this extension was mainly industrial collaboration on identification of a steam boiler system, see Example 1.1. Secondly, most system consisting of feedforward interconnected subsystems can be build up by this structure and the cascaded system. Hence the presented theory should cover most interesting cases.

To gain some insight into the problem simple FIR-subsystems were considered and the effect of sensor placement, input signals and common dynamics of the subsystems

on the asymptotic properties was presented. When the subsystems have the same dynamics, and only one of the outputs from the parallel subsystems is measured, the quality of the estimate of the parameters belonging to the measured subsystem is not improved by also measuring the output from the final subsystem. This result was also proven for subsystems with common, general dynamics.

However, if it is known beforehand that the subsystems have common dynamics this should be utilized in the identification. Potentially, this can improve the quality of the estimates considerably.

Chapter 4

Identification Methods for Structured Systems

STRUCTURED systems, such as the cascade system and the parallel serial system, can be seen as special cases of Multiple-Input Multiple-Output (MIMO) systems. Most of the classical identification methods were first design to deal with Single-Input Single-Output (SISO) systems but many of them have with time been be generalized to include MIMO systems. So why not directly use the classical methods to identify structured systems? In many applications it is favorable, or even vital, that the structure of the identified model relates to the physical process and to retain as much as possible of the physical interpretation. This is important if the model will be used for controller design, for example tuning of Kalman filter parameters or cascade controllers. Another motivation can be seen from the steam boiler example in Section 1.1. The parallel boilers can be turned on or off according to steam demand. Hence the system can be in many different configurations depending on which boilers are on and which are turned off. Plants with up to ten boilers are not uncommon and performing identification experiments for each of the possible on/off configurations is not practically feasible. If, on the other hand, the system is identified with a method that retains the physical structure of the system, it is easy to implement the different configurations in the model by removing or adding the corresponding parallel subsystem.

It has however proven hard to keep the structural information in many of the classical identification methods. In this chapter two methods for identification of cascade structured systems based on subspace identification are presented. But first some existing methods are reviewed.

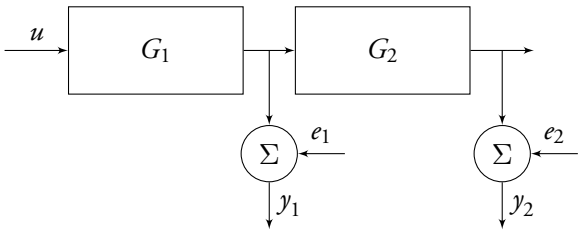


Figure 4.1 Cascade system.

4.1 Related Methods

In this chapter identification of systems with cascade structure, as shown in Figure 4.1, will be studied. The objective is to identify the subsystems $G_1(q)$ and $G_2(q)$ from input and output data. Any single-input multiple-output method could be used, but it is often not straightforward to impose the cascade model structure. The methods presented here are easy to extend to more general structures such as the parallel serial structure.

In Chapter 3 a Prediction Error Method was used to identify the two subsystems and the statistical properties of the estimates were derived. This works good for the simple FIR model structure but because of the product $G_1(q)\,G_2(q)$, other linear model structures, such as OE or ARX, are not directly applicable. To find the parameter estimate with a Prediction Error based method or with a Maximum Likelihood method a non-convex optimization problem has to be solved. It can therefore be difficult to guarantee that during the optimization, the global optimum will be found. A way to solve this is to try to find good initial estimates, hopefully lying in the region of attraction of the global optimum. It can however be shown that structured PEM and structured ML are statistically optimal methods for this problem if one is able to find the global optimum.

Another more direct approach to identify the subsystems would be to first identify $G_1(q)$ from data $\{u(t), y_1(t)\}$ and then in a second step identify $G_2(q)$ from data $\{\hat{u}_2(t), y_2(t)\}$, where $\hat{u}_2$ is an estimate of the input to the second subsystem $G_2(q)$. If the model estimate $\hat{G}_1(q)$ is good, one could use $\hat{u}_2 = \hat{G}_1(q)\,u(t)$. If the noise variance is low for the first measurement noise one could use $\hat{u}_2(t) = y_1(t)$.

The identification of the second subsystem can also be seen as a Errors-in-variables (EIV) problem. An EIV model has measurement noises acting on both the input and the output of the system. This is exactly the case for cascade systems, the measurement $y_1(t)$ can be seen as a noisy measurement of the true input to the second subsystem. A good

overview of error-in-variables identification can be found in Söderström (2007) and the references therein. However, using this approach, the information that the input to the second subsystem is generated by the known input, $u(t)$, filtered through the first subsystem, is not taken into account.

4.2 Subspace Identification of Cascade Systems

If the two subsystems of a cascade system have the state-space representations

$$G_1 : \begin{cases} x_1(t+1) &= \mathbf{A}_1 x_1(t) + \mathbf{B}_1 u_1(t) \\ y_1(t) &= \mathbf{C}_1 x_1(t) + e_1(t) \end{cases}$$
$$G_2 : \begin{cases} x_2(t+1) &= \mathbf{A}_2 x_2(t) + \mathbf{B}_2 u_2(t) \\ y_2(t) &= \mathbf{C}_2 x_2(t) + e_2(t) \end{cases}$$

where $x_1(t) \in \mathbb{R}^{n_1}$, $x_2(t) \in \mathbb{R}^{n_2}$ and,

$$G_1(q) = \mathbf{C}_1(q\mathbf{I} - \mathbf{A}_1)^{-1}\mathbf{B}_1, \quad G_2(q) = \mathbf{C}_2(q\mathbf{I} - \mathbf{A}_2)^{-1}\mathbf{B}_2,$$

one natural realization of a cascade system on state-space form is

$$\begin{aligned} \begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} &= \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{B}_2\mathbf{C}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{0} \end{bmatrix} u(t) \\ \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} &= \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} e_1(t) \\ e_2(t) \end{bmatrix}. \end{aligned} \tag{4.1}$$

This special structure of the state-space matrices, where the states $x_1(t)$ correspond to the first subsystem and the states $x_2(t)$ correspond to the second subsystem, will be called a *realization in cascade form*. Note that this implies that the rank of the lower left part of the matrix, corresponding to $\mathbf{B}_2\mathbf{C}_1$, should be one.

Just applying a standard subspace method to a system on cascade form would return an estimate in the form

$$\begin{aligned} x(t+1) &= \mathbf{A}x(t) + \mathbf{B}u(t) \\ y_1(t) &= \mathbf{C}_1 x(t) \\ y_2(t) &= \mathbf{C}_2 x(t). \end{aligned} \tag{4.2}$$

A state-space realization is not unique with respect to the input-output relation, *i.e.,* the system is only identified up to an unknown similarity transform. In general this similarity transform will mix the states from the first subsystem with the states from the second subsystem. The first and second subsystems could hence not be directly separated from the identified model (4.2).

Example 4.1
Consider a system on cascade form where the state-space matrices of the two subsystems are given by

$$\mathbf{A}_1 = \begin{bmatrix} -1 & -0.5 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{C}_1 = \begin{bmatrix} 0 & 1 \end{bmatrix},$$
$$\mathbf{A}_2 = \begin{bmatrix} -0.9 \end{bmatrix}, \quad \mathbf{B}_2 = \begin{bmatrix} 1 \end{bmatrix}, \quad \mathbf{C}_2 = \begin{bmatrix} 2 \end{bmatrix}.$$

The system is excited with a pseudo random white Gaussian input signal with variance 1 and $N = 200$ samples are collected. A model is identified using *n4sid* from the MATLAB System Identification Toolbox (Ljung, 2011) and is given by

$$\hat{\mathbf{A}} = \begin{bmatrix} -0.24 & 0.53 & 0.018 \\ -0.57 & -0.75 & -0.0090 \\ -0.37 & 0.17 & 0.89 \end{bmatrix}, \quad \hat{\mathbf{B}} = \begin{bmatrix} 0.063 \\ 0.024 \\ -0.026 \end{bmatrix},$$
$$\hat{\mathbf{C}} = \begin{bmatrix} 7.27 & -18.72 & 0.11 \\ -14.38 & 15.92 & -19.90 \end{bmatrix}.$$

Even if the dynamics are identified correctly (eigenvalues $\lambda_{1,2} = -0.5 \pm 0.5j, \lambda_3 = 0.9$) it is not possible to directly separate the two subsystems.

If the similarity transform somehow was known, the system could be transformed back to cascade form (4.1) and the state-space matrices for G_1 and G_2 could directly be recovered. In the first proposed, indirect, method a transform that brings the system back to cascade form is found and from this the state-space matrices for the two subsystems are recovered.

Extended Observability Matrix for Cascade Systems

As was seen in Section 2.3 the first step in the subspace identification algorithm is to form an estimate of the extended observability matrix from input-output data. The estimated extended observability matrix, $\tilde{\mathbf{O}}_r$, has the form

$$\tilde{\mathbf{O}}_r = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{r-1} \end{bmatrix} \tilde{\mathbf{T}} + \tilde{\mathbf{E}}_N$$

where $\tilde{\mathbf{T}}$ is an unknown transformation of full rank and $\tilde{\mathbf{E}}_N$ is an unknown matrix due to noise, see Section 2.3 for details.

For the system on cascade form (4.1) the extended observability matrix, disregarding the noise contribution, becomes

$$\tilde{\mathbf{O}}_r\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{r-1} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{T}}_{11} & \tilde{\mathbf{T}}_{12} \\ \tilde{\mathbf{T}}_{21} & \tilde{\mathbf{T}}_{22} \end{bmatrix} =$$

$$\begin{bmatrix} \mathbf{C}_1\tilde{\mathbf{T}}_{11} & \mathbf{C}_1\tilde{\mathbf{T}}_{12} \\ \mathbf{C}_2\tilde{\mathbf{T}}_{21} & \mathbf{C}_2\tilde{\mathbf{T}}_{22} \\ \mathbf{C}_1\mathbf{A}_1\tilde{\mathbf{T}}_{11} & \mathbf{C}_1\mathbf{A}_1\tilde{\mathbf{T}}_{12} \\ \mathbf{C}_2\mathbf{B}_2\mathbf{C}_1\tilde{\mathbf{T}}_{11} + \mathbf{C}_2\mathbf{A}_2\tilde{\mathbf{T}}_{21} & \mathbf{C}_2\mathbf{B}_2\mathbf{C}_1\tilde{\mathbf{T}}_{12} + \mathbf{C}_2\mathbf{A}_2\tilde{\mathbf{T}}_{22} \\ \mathbf{C}_1\mathbf{A}_1^2\tilde{\mathbf{T}}_{11} & \mathbf{C}_1\mathbf{A}_1^2\tilde{\mathbf{T}}_{12} \\ \mathbf{C}_2(\mathbf{B}_2\mathbf{C}_1\mathbf{A}_1+\mathbf{A}_2\mathbf{B}_2\mathbf{C}_1)\tilde{\mathbf{T}}_{11}+\mathbf{C}_2\mathbf{A}_2^2\tilde{\mathbf{T}}_{21} & \mathbf{C}_2(\mathbf{B}_2\mathbf{C}_1\mathbf{A}_1+\mathbf{A}_2\mathbf{B}_2\mathbf{C}_1)\tilde{\mathbf{T}}_{12}+\mathbf{C}_2\mathbf{A}_2^2\tilde{\mathbf{T}}_{22} \\ \vdots & \vdots \end{bmatrix}. \tag{4.3}$$

This fundamental structure of the extended observability matrix will be used to derive the second, direct, method to identify the subsystems.

4.3 Indirect Method

The main idea of this method is to find a similarity transform for the identified system such that the transformed system is in cascade form (4.1). This basic idea was proposed in Wahlberg et al. (2008) where the transformation matrix is parameterized giving a set of equations that has to be solved in order to get the system to cascade form. Due to uncertainties this problem does not in general have an exact solution. In Wahlberg et al. (2008) no method of actually solve this set of equations is presented. Since there is no exact solution to this problem, the proposed indirect method finds a similarity transform that takes the system to cascade form while minimizing the mean square error between the estimated output and the measured output.

Initial Estimate

The first step is to apply a standard subspace identification method to data collected from the first subsystem, $\{u(t), y_1(t), t = 1, \ldots, N\}$, giving an estimate $\hat{G}_1$ of G_1. The state-space matrices of the identified $\hat{G}_1$ are denoted $(\hat{\mathbf{A}}_1, \hat{\mathbf{B}}_1, \hat{\mathbf{C}}_1)$ and the chosen order

of the system is denoted as n_1. Next the system from the input to the second output is identified, *i.e.,* identify a model from $\{u(t), y_2(t), t = 1, \ldots, N\}$. This gives a model, $\hat{G}_3$ of order n_3, as an estimate of the product G_2G_1. The state-space matrices for this system will be denoted $(\hat{\mathbf{A}}_3, \hat{\mathbf{B}}_3, \hat{\mathbf{C}}_3)$. The estimated order of the second subsystem can be calculated as $n_2 = n_3 - n_1$.

Example 4.2

Consider again the same setup as in Example 4.1. The identified model of G_1 from $\{u(t), y_1(t), t = 1, \ldots, 200\}$ is given by

$$\hat{\mathbf{A}}_1 = \begin{bmatrix} -0.25 & -0.51 \\ 0.61 & -0.75 \end{bmatrix}, \hat{\mathbf{B}}_1 = \begin{bmatrix} -0.063 \\ 0.021 \end{bmatrix}, \hat{\mathbf{C}}_1 = \begin{bmatrix} -6.32 & -19.24 \end{bmatrix}$$

with order $n_1 = 2$. The estimate of the product G_2G_1 from $\{u(t), y_2(t), t = 1, \ldots, 200\}$, is given by the state-space matrices

$$\hat{\mathbf{A}}_3 = \begin{bmatrix} -0.24 & -0.53 & -0.018 \\ 0.57 & -0.75 & 0.0090 \\ -0.37 & -0.17 & 0.89 \end{bmatrix}, \quad \hat{\mathbf{B}}_3 = \begin{bmatrix} -0.063 \\ 0.024 \\ 0.026 \end{bmatrix},$$
$$\hat{\mathbf{C}}_3 = \begin{bmatrix} 14.38 & 15.92 & 19.90 \end{bmatrix}$$

with order $n_3 = 3$. The estimated order of the second subsystem is hence $n_2 = 3-2 = 1$.

Finding a Transform

To find a transformation, $\mathbf{T}$, that takes the system back to cascade form while minimizing the mean square error, the following optimization problem need to be solved:

$$\begin{aligned} \min_{\mathbf{T}, \tilde{\mathbf{B}}_1} \quad & \frac{1}{\lambda_1} \sum_{t=1}^{N} \left(\hat{\mathbf{C}}_1 \left(q\mathbf{I} - \hat{\mathbf{A}}_1 \right)^{-1} \hat{\mathbf{B}}_1 u(t) - y_1(t) \right)^2 + \\ & \frac{1}{\lambda_2} \sum_{t=1}^{N} \left(\hat{\mathbf{C}}_3 \mathbf{T} \left(q\mathbf{I} - \mathbf{T}^{-1} \hat{\mathbf{A}}_3 \mathbf{T} \right)^{-1} \begin{pmatrix} \tilde{\mathbf{B}}_1 \\ \mathbf{0} \end{pmatrix} u(t) - y_2(t) \right)^2 \\ \text{s.t} \quad & \hat{\mathbf{C}}_3 \mathbf{T} \text{ and } \mathbf{T}^{-1} \hat{\mathbf{A}}_3 \mathbf{T} \text{ are in cascade form.} \end{aligned} \tag{4.4}$$

The factors λ_1 and λ_2 are the measurement noise variances or if these are unknown, some user defined weighting.

The constraints are such that the upper left corner of $\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}$ has approximately the same dynamics as the identified system $\hat{G}_1$ and that the upper right corner should contain zeros. Furthermore the lower left corner of $\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}$ corresponding to $\mathbf{B}_2\mathbf{C}_1$ should have rank 1. The optimization problem (4.4) is non-convex and is in general hard to solve.

Structural Constraints

The first constraint, that $\hat{\mathbf{C}}_3 T$ should be on cascade form, means that the first n_1 elements of $\hat{\mathbf{C}}_3 T$ should be equal to zero. This corresponds to that the second output should only depend on the states of the second subsystem. Thus, the first n_1 columns of $\mathbf{T}$ must be in the kernel space of $\hat{\mathbf{C}}_3$.

The second set of constraints is that $\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}$ should be on cascade form. This implies that the upper left corner should have similar dynamics as the first identified subsystem $\hat{G}_1$ and that the upper right corner should be the zero matrix.

If $\hat{\mathbf{A}}_3$ is Schur factorized (Horn and Johnson, 1990)

$$\hat{\mathbf{A}}_3 = \mathbf{U}\tilde{\mathbf{A}}_3\mathbf{U}^* \tag{4.5}$$

then $\tilde{\mathbf{A}}_3$ is similar to $\hat{\mathbf{A}}_3$ and the upper right corner equals the zero matrix. If the Schur factorization is performed such that the eigenvalues closest to the eigenvalues of the first subsystem are located in the upper left corner of $\tilde{\mathbf{A}}_3$, the last n_2 columns of $\mathbf{T}$ should be chosen as the last columns of $\mathbf{U}$. This is summarized in the following proposition.

Proposition 4.1 *If the first n_1 columns in $\mathbf{T}$ spans the kernel space of $\mathbf{C}_3$ and the last n_2 columns span the space corresponding to the n_2 columns of $\mathbf{U}$, where $\mathbf{U}$ is given from the Schur factorization of $\mathbf{A}_3 = \mathbf{U}\tilde{\mathbf{A}}_3\mathbf{U}^*$ such that the dynamics from the first system is in the upper left corner of the block-triangular matrix $\tilde{\mathbf{A}}_3$. The system transformed with $\mathbf{T}$ will be on cascade form, save for the rank constraint on the lower left corner of $\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}$.*

Using Proposition 4.1, the optimization problem (4.4) could be somewhat simplified. Instead of optimizing over all parameters in the transformation matrix $\mathbf{T}$, it can be performed over linear combinations of the vectors spanning the kernel space of $\hat{\mathbf{C}}_3$ and over scaled versions of the appropriate columns of $\mathbf{U}$ from the Schur factorization. Denoting the order of the kernel space of $\hat{\mathbf{C}}_3$ by $n_k = n_2 + n_1 - 1$ gives that the transformation matrix can be written as

$$\mathbf{T} = \begin{bmatrix} \sum_{i=1}^{n_k} k_i c_i & \cdots & \sum_{i=1}^{n_k} k_{i+n_k(n_1-1)} c_i & k_{n_k n_1+1} u_1 & \cdots & k_{n_k n_1+n_2} u_{n_2} \end{bmatrix} \tag{4.6}$$

where u_i is the i:th column of U from the Schur factorization (4.5) and $c_i \in \mathrm{Ker}(\hat{\mathbf{C}}_3)$. The optimization is performed over k_i, $i = 1 \ldots (n_2 + n_1 n_k)$. Compared to the

fully parameterized transformation $\mathbf{T}$, with $(n_1 + n_2)^2$ parameters, the number of optimization variables is reduced.

Rank Constraint

Using the $\mathbf{T}$ defined in (4.6) it is guaranteed that the system is in cascade form, with the exception of the rank constraint; the lower left corner of $\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}$ should have rank one. Instead of explicitly demanding this part of the matrix to have rank one, some heuristics will be used.

The rank one constraint stems from the fact that this part of the matrix corresponds to the product $\mathbf{B}_2\mathbf{C}_1$. Denoting the estimate of the system matrix of the first subsystem from $\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}$ as $\bar{\mathbf{A}}_1$, ideally the eigenvalues of $\hat{\mathbf{A}}_1$ should be the same as the eigenvalues $\bar{\mathbf{A}}_1$. If this is the case a similarity transform between $\hat{\mathbf{A}}_1$ and $\bar{\mathbf{A}}_1$ can directly be found. However, in general this is not the case; often the eigenvalues of the two matrices differ slightly. One solution to this is to transform both $\hat{\mathbf{A}}_1$ and $\bar{\mathbf{A}}_1$ to upper triangular form, then replace the eigenvalues in the diagonal of $\bar{\mathbf{A}}_1$ by the corresponding in $\hat{\mathbf{A}}_1$. This way the two matrices are similar and a similarity transform, $\Upsilon(\mathbf{T})$, could be calculated. An estimate of $\mathbf{B}_2$ is then given by

$$\hat{\mathbf{B}}_2 = \hat{\mathbf{A}}_{21}(\hat{\mathbf{C}}_1\Upsilon(\mathbf{T}))^{\dagger}$$

where $(\dagger)$ denotes the Moore-Penrose pseudoinverse and $\hat{\mathbf{A}}_{21}$ is the lower left corner of $\hat{\mathbf{A}}_3$. The following optimization problem can now be formulated

$$\begin{aligned}\min_{\mathbf{T},\tilde{\mathbf{B}}_1} \quad & \frac{1}{\lambda_1}\sum_{t=1}^{N}\left(\hat{\mathbf{C}}_1\left(q\mathbf{I}-\hat{\mathbf{A}}_1\mathbf{T}\right)^{-1}\tilde{\mathbf{B}}_1 u(t) - y_1(t)\right)^2 + \\ & \frac{1}{\lambda_2}\sum_{t=1}^{N}\left(\hat{\mathbf{C}}_2\left(q\mathbf{I}-\hat{\mathbf{A}}_2\right)^{-1}\hat{\mathbf{B}}_2\hat{\mathbf{C}}_1\left(q\mathbf{I}-\hat{\mathbf{A}}_1\mathbf{T}\right)^{-1}\tilde{\mathbf{B}}_1 u(t) - y_2(t)\right)^2\end{aligned} \tag{4.7}$$

where

$$\begin{aligned}\hat{\mathbf{C}}_2 &= \left(\hat{\mathbf{C}}_3\mathbf{T}\right)_{1:n_2} \\ \hat{\mathbf{B}}_2 &= \hat{\mathbf{A}}_{12}(\hat{\mathbf{C}}_1\Upsilon(\mathbf{T}))^{\dagger} \\ \hat{\mathbf{A}}_2 &= \left(\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}\right)_{1:n_2,1:n_2}.\end{aligned}$$

The optimization is performed over columns of $\mathbf{T}$ defined in (4.6) and $\tilde{\mathbf{B}}_1$. This heuristics solves the problem with the rank condition in many cases.

Remark 4.1 The optimization problem is still non-convex and could suffer from local minima points. It is however possible to verify if a solution is good or not. The difference between the mean square error of the signals from the initially estimated unstructured system and the system with the cascade structure imposed should be small.

Calculating the estimates

When the optimization problem (4.7) has been solved, the transformation matrix is used to transform the identified system to cascade form. From this the estimates of the state-space matrices for the second subsystem $\hat{\mathbf{A}}_2$ and $\hat{\mathbf{C}}_2$ can easily be recovered, *cf.,* (4.1). The matrix estimate $\hat{\mathbf{B}}_2$ could be calculated from the matrix product $\mathbf{B}_2\mathbf{C}_1$.

Example 4.3
Solving the optimization problem (4.7) to the identified state-space models from Example 4.2 gives

$$\mathbf{T} = \begin{bmatrix} 0.0097 & -1.91 & -10.08 \\ -7.09 \cdot 10^{-5} & 1.78 & 6.92 \\ -0.63 & -0.030 & 1.75 \end{bmatrix}.$$

Transforming the system back to cascade form gives

$$\left(\mathbf{T}^{-1}\hat{\mathbf{A}}_3\mathbf{T}\right) = \begin{bmatrix} -5.98 & -25.59 & 0 \\ 1.18 & 4.98 & 0 \\ 2.98 & 8.56 & 0.90 \end{bmatrix}, \quad \hat{\mathbf{C}}_3\mathbf{T} = \begin{bmatrix} -12.38 & 0.27 & 0 \end{bmatrix}$$

From this the state-space matrices for the second subsystem can be recovered. The resulting error between the true system and the estimated system is shown in Figure 4.2.

4.4 Direct Method

The second method uses the fact that the structure of the extended observability matrix is known for cascade systems, see (4.3). As will be shown, this method is only applicable to systems where the second subsystem has order one. However, in practice, this is a common case. The second subsystem is often a sensor with simple dynamics or an integrator, for example in a servo application where the first measurement is the velocity and the second output is the position.

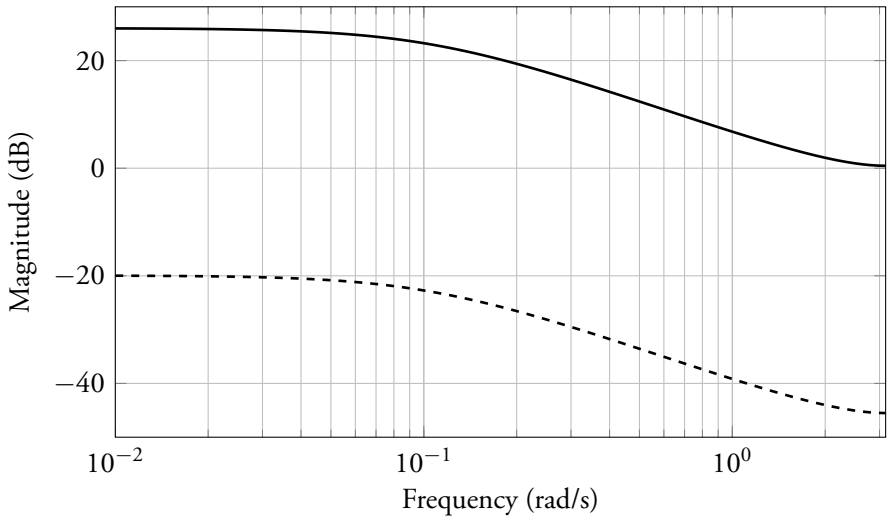


Figure 4.2 True system G_2 (——) and estimation error for indirect method (- - -) for Example 4.3.

Estimating the Extended Observability Matrix

The first step of the algorithm is to estimate the extended observability matrix for the Single-Input Multiple-Output (SIMO) system from u to y_1 and y_2 according to the description in Section 2.3. The extended observability matrix will have the structure (4.3) and will be denoted

$$\mathbf{O}_r\tilde{\mathbf{T}} = \begin{bmatrix} \xi_{1,1} & \xi_{1,2} \\ \eta_{1,1} & \eta_{1,2} \\ \vdots & \vdots \\ \xi_{r,1} & \xi_{r,2} \\ \eta_{r,1} & \eta_{r,2} \end{bmatrix}$$

where $\xi_{i,1}$ has size $(1 \times n_1)$ and $\eta_{i,2}$ has size $(1 \times n_2)$ for $i = 1, \ldots, r$.

Estimating First Subsystem

From (4.3) it can be seen that for the first subsystem the state-space matrices can be solved with least squares just as in the standard subspace formulation. An estimate of C_1 is

$$\hat{\mathbf{C}}_1 = \mathbf{C}_1\tilde{\mathbf{T}}_{11} = \xi_{1,1}.$$

Noting that every other row of (4.3) is just the previous row multiplied with $\tilde{\mathbf{T}}^{-1}\mathbf{A}_1\tilde{\mathbf{T}}$, an estimate of $\mathbf{A}_1$ can be calculated from the linear least squares problem

$$\hat{\mathbf{A}}_1 = \arg\min_{\mathbf{A}_1} \sum_{i=1}^{r-1} \left\| \xi_{i+1,1} - \xi_{i,1}\mathbf{A}_1 \right\|_2^2 .$$

Estimating Second Subsystem

For the second subsystem the state-space matrix C_2 can again easily be estimated by

$$\hat{\mathbf{C}}_2 = \mathbf{C}_2\tilde{\mathbf{T}}_{22} = \eta_{1,2}$$

It is not obvious how to form a least squares problem for estimating the system matrix of the second subsystem $\mathbf{A}_2$. However, under the assumption that the second subsystem has a order of one, this is actually possible. Denoting the product $\chi = \mathbf{C}_2\mathbf{B}_2$ gives the following least squares problem in $\hat{\mathbf{A}}_2$ and χ.

$$\arg\min_{\mathbf{A}_2,\chi} \sum_{i=1}^{r-1} \left\| \eta_{i+1,2} - \eta_{i,2}\mathbf{A}_2 - \chi\xi_{i,2} \right\|_2^2$$

When $\chi = \mathbf{B}_2\mathbf{C}_2$ has been found, $\hat{\mathbf{B}}_2$ can be calculated

$$\hat{\mathbf{B}}_2 = \frac{\chi}{\hat{\mathbf{C}}_2}.$$

Finally $\hat{\mathbf{B}}_1$ can be calculated as in the standard subspace formulation as a linear regression problem, see Section 2.3.

Example 4.4

From the extended observability matrix, used to estimate the state-space matrices in Example 4.1, the following estimates are obtained with the direct method

$$\hat{\mathbf{C}}_2 = -0.65, \quad \hat{\mathbf{A}}_2 = 0.9, \quad \chi = 2.$$

and $\hat{\mathbf{B}}_2$ can be calculated as

$$\hat{\mathbf{B}}_2 = \frac{\chi}{\hat{\mathbf{C}}_2} = -3.09.$$

The resulting error between the true system and the estimated system is shown in Figure 4.3.

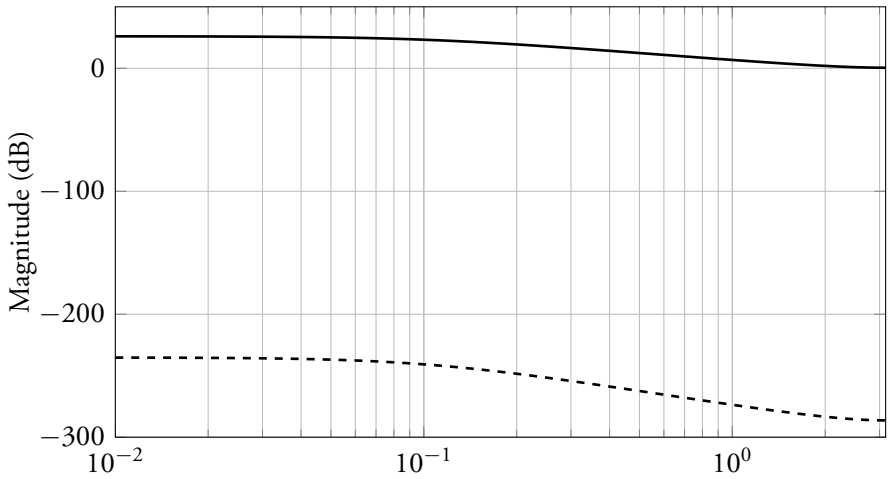


Figure 4.3 True system G_2 (——) and estimation error for the direct method (- - -) for Example 4.4.

4.5 Summary

In this chapter two new subspace based methods for estimating cascade structured systems were presented. Standard subspace methods can be applied to structured systems directly but it is not obvious how to retain the structure in the identification. Keeping the physical interpretation of the system can be important in some applications, for example in control design.

The first presented, indirect, method used that a system on cascade form could be written as a special state-space realization. The idea was then to find a transform that takes the system back to this cascade form while minimizing the error between the measurements and the predicted output from the model. It was shown that this leads to a non-convex optimization problem and it might be hard to find the global minimizer.

Taking a step back, a subspace method tries firstly to find an estimate of the extended observability matrix. For cascade system it was shown that this matrix has a special form that can be utilized in the identification. In this case, the state-space matrices could be estimated by solving simpler least squares problem. However this method was shown to only work, in its current state, when the second system can be approximated well by a first order system.

With the drawbacks of the two methods they might not be practically usable. However they show interesting ideas how structure could be implemented into

subspace methods, something that should be further studied.

Chapter 5

Applications

In this chapter the two subspace based methods presented in Chapter 4 are applied to a lab system, the double tank process. The methods performances are compared and some issues related to identification of a continuous time cascade structured systems are discussed.

5.1 The Double Tank Process

The Double Tank Process from *Quanser Inc* (Quanser, 2010) is an educational process mostly used for teaching of the basics in automatic control. It consists of two equivalent cylindrical water tanks, one situated above the other. A schematic of the process is shown in Figure 5.1. A pump driven by a DC-motor pumps water from a water basin into the upper tank. Water then flows through small orifices located in the bottom of each tank from the upper tank into the lower tank and from the lower tank into the water basin. The water levels in the two tanks are measured by pressure sensors located in the bottom of each tank. Since the pressure is proportional, with known proportionality constant, to the water level the conversion to level is straightforward. The input to the system is the voltage applied to the DC-motor and the outputs are the measured tank levels in the upper and lower tank, respectively.

First a physical model of the plant will be derived. This model will be used to validate the performance of the methods and also to gain some insight into the problem.

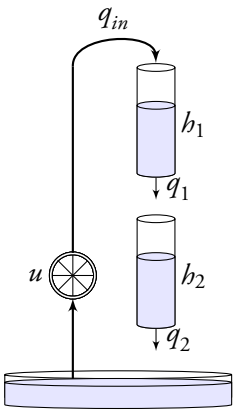


Figure 5.1 A schematic of the double tank process.

Nonlinear Model

From Bernoulli's law the following relation between the outflow from a tank i, q_i, and the water level in the tank, $h_i(t)$, can be derived

$$q_i(t) = a\sqrt{2gh_i(t)}$$

where g is the gravitational constant and a is the outflow orifice area of the tank. This result is credited to Torricelli.

The change in the volume of water in a tank is the difference in the inflow and the outflow, *i.e.,*

$$\frac{\mathrm{d}\,V(t)}{\mathrm{d}\,t} = q_{in}(t) - q_{out}(t).$$

Assuming that the DC-motor's dynamics is much faster than the dynamics of the water tank the relation between the applied input voltage and the water flow produced by the pump, q_{in}, can be described by

$$q_{in}(t) = k_m u(t)$$

where k_m is a constant.

Table 5.1 The nominal parameter values for the water tank process.

Parameter	Nominal value	Description
a_i	0.18 cm^2	Tank outflow orifice area.
A	16.5 cm^2	Tank cross section area.
k_m	4.6 $cm^3/s/V$	Voltage to flow proportionality constant.
g	9.8 m/s^2	Standard gravity constant.

Utilizing that the water level is given by $h_i(t) = V_i(t)A$, where A is the cross section area of the tanks, the following differential equations for the water level can be derived

$$\frac{\mathrm{d}\,h_1(t)}{\mathrm{d}\,t} = \frac{q_{in}(t) - q_1(t)}{A} = \frac{k_m}{A}u(t) - \frac{a_1}{A}\sqrt{2gh_1(t)}$$
$$\frac{\mathrm{d}\,h_2(t)}{\mathrm{d}\,t} = \frac{q_1(t) - q_2(t)}{A} = \frac{a_1}{A}\sqrt{2gh_1(t)} - \frac{a_2}{A}\sqrt{2gh_2(t)}.$$

The values of the parameters are taken from the data sheet for the tanks in Quanser (2010). However due to wear and tear and aging of the process these values cannot be expected to be the exact values of the process. The parameters and their nominal value are given in Table 5.1.

The methods that are to be applied are inherently for linear and time discrete systems. So the next step is to linearize and discretize the model.

Linearized Model

The linearization of the system around the linearization point h_1^0, h_2^0, u^0 can be written on state-space form as

$$\begin{aligned}\dot{h}(t) &= \mathbf{A}_c h(t) + \mathbf{B}_c h(t) = \begin{bmatrix} -\frac{1}{T_1} & 0 \\ \frac{1}{T_1} & -\frac{1}{T_2} \end{bmatrix} h(t) + \begin{bmatrix} \frac{k_m}{A} \\ 0 \end{bmatrix} u(t) \\ y(t) &= \mathbf{C}_c h(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} h(t)\end{aligned} \tag{5.1}$$

where the time constants are

$$T_i = \frac{A}{a_i}\sqrt{\frac{2h_i^0}{g}}.$$

Using a zero-order-hold approximation with sampling time T_s

$$\mathbf{A}_d = e^{\mathbf{A}_c T_s}, \quad \mathbf{B}_d = \left(\int_{\tau=0}^{T_s} e^{\mathbf{A}_c \tau} \, \mathrm{d}\tau \right) \mathbf{B}_c, \quad \mathbf{C}_d = \mathbf{C}_c,$$

the state-space matrices for the discretized system becomes

$$\begin{aligned} \mathbf{A}_d &= \begin{bmatrix} e^{-\frac{T_s}{T_1}} & 0 \\ \frac{e^{-\frac{T_s}{T_1}} - e^{-\frac{T_s}{T_2}}}{T_1 - T_2} T_2 & e^{-\frac{T_s}{T_2}} \end{bmatrix} \\ \mathbf{B}_d &= \frac{k_m}{A} \begin{bmatrix} T_1 \left(1 - e^{-\frac{T_s}{T_1}} \right) \\ \frac{T_2}{T_1 - T_2} \left(T_1 \left(1 - e^{-\frac{T_s}{T_1}} \right) - T_2 \left(1 - e^{-\frac{T_s}{T_2}} \right) \right) \end{bmatrix} \\ \mathbf{C}_d &= \mathbf{I}. \end{aligned} \tag{5.2}$$

The Problem with Continuous Time Cascade Systems

While the continuous time state-space model of the double tank system (5.1) is on cascade form, the discretized model (5.2) is not due to the non-zero element in $\mathbf{B}_d$. The problem is the zero-order hold assumption of the inputs. Since the water tank is a continuous system, the output from the first tank will not be constant during one sampling period. As this output is the input to the second tank in the cascade model, this will not satisfy the zero-order hold assumption and an error will be introduced. Hence from the discretized model (5.2) it is not possible to get a model of the second subsystem.

One way to overcome this is to directly discretize the continuous time model of the second water tank, *i.e.,* to discretize

$$\begin{aligned} \dot{h}_2(t) &= \left[-\tfrac{1}{T_2} \right] h_2(t) + \left[\tfrac{1}{T_1} \right] h_1(t) \\ y_2(t) &= h_2(t). \end{aligned}$$

With this approach it is assumed that $h_1(t)$ satisfy the zero-order hold condition. The validity of this approximation depends on how close the output signal from the first tank is to a constant during a sample interval.

Another approach is to divide the discretized transfer function from u to y_2 by the discretized transfer function from u to y_1.

The two different ways of handling the problem are compared. In Figure 5.2 the error between the frequency response of the continuous model $G_2(j\omega)G_1(j\omega)$ and

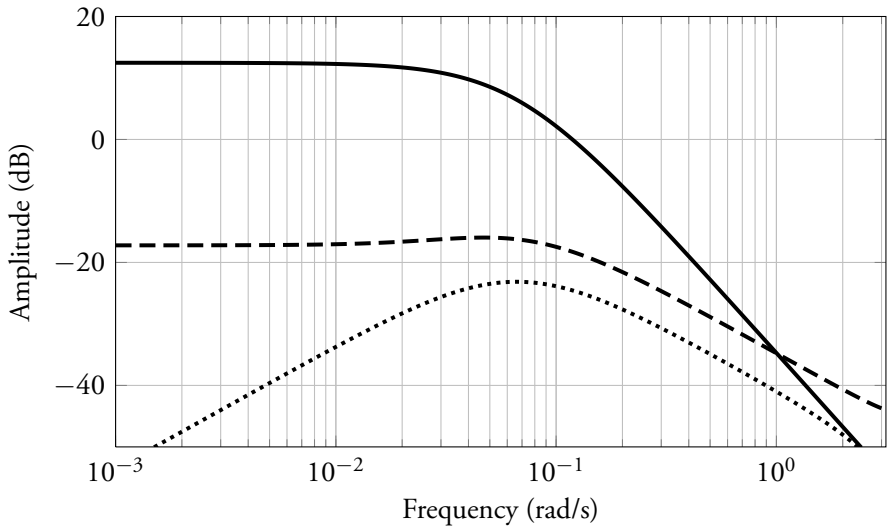


Figure 5.2 Continuous time system G_2G_1 (——), error for directly discretized G_2 times directly discretized G_1 (– –) and error for discretized transfer function from u to y_2 divided by the discretized transfer function from u to y_1 (·····).

the frequency response of the discretized model $G_2^d(j\omega)G_1^d(j\omega)$ for the two different approaches, are shown. The parameter values used are the nominal from Table 5.1 and the sampling time is set to $T_s = 1$s. In this case the second approximation has smaller error for all frequency and will be the one used. This problem when identifying structured continuous systems with discrete time models is well known. One way to get around this is to use continuous time identification methods. A similar problem appears in errors in variables identification, see for example Söderström (2007).

5.2 Experimental Setup

The experimental setup consists of the two tank process as described earlier. The process is powered by a power module, amplifying the measurements and the voltage to the DC-motor. The signals are then sampled by a National Instrument DAQ-card connected to a PC running MATLAB. From MATLAB the user can then apply voltage to the DC-motor and process the sampled measurements from the process.

Data Collection

For the identification of the true tank system a white Gaussian noise process is used as input. Since the water tank system is nonlinear the amplitude of the input signal has to be chosen carefully. To high inputs may drive the system outside the linear range of the system. On the other hand, the input has to be large enough to excite the system and give output signals large enough compare to the noise. By trial and error a variance of the input of Var $u_t = 1$ seems to be a good tradeoff. The sampling time is chosen as $T_s = 1$ s and 400 samples of the input and outputs are collected, 200 for identification and 200 for validation.

5.3 Identification

The two methods presented in Chapter 4 are applied to the data and their results are compared.

Indirect Subspace Based Method

For the direct method an estimate of the first subsystem, from u to y_1 and an estimate of the product of the two subsystems, from u to y_2, are needed. The corresponding state-space matrices, $\hat{\mathbf{A}}_1$, $\hat{\mathbf{B}}_1$, $\hat{\mathbf{C}}_1$, $\hat{\mathbf{A}}_3$ and $\hat{\mathbf{C}}_3$ are identified with *n4sid*, a standard subspace algorithm, see Van Overschee and De Moor (1996) for details. The orders of the identified models are chosen by studying the singular values of their respective extended observability matrices. The singular values are shown in Figure 5.3.

From this the order of G_1 is chosen as $n_1 = 1$ and the order for G_3 is chosen as $n_3 = 2$. The estimate of the model order for the second subsystem is hence $n_2 = n_3 - n_1 = 1$.

A similarity transform is found by solving the optimization problem (4.7) numerically. Since the orders of the subsystems are both one, the rank condition is automatically fulfilled. When the transformation matrix $\mathbf{T}$ and $\tilde{\mathbf{B}}_1$ are found then it is straightforward to find $\hat{\mathbf{A}}_2$, $\hat{\mathbf{B}}_2$ and $\hat{\mathbf{C}}_2$.

Direct Subspace Based Method

Again *n4sid* is used, but now on the SIMO system to get an estimate of the extended observability matrix. The singular values of the extended observability matrix suggest that the order should be chosen as 2. From the results of the indirect method it is known that the first subsystem can be well approximated by a first order system. The

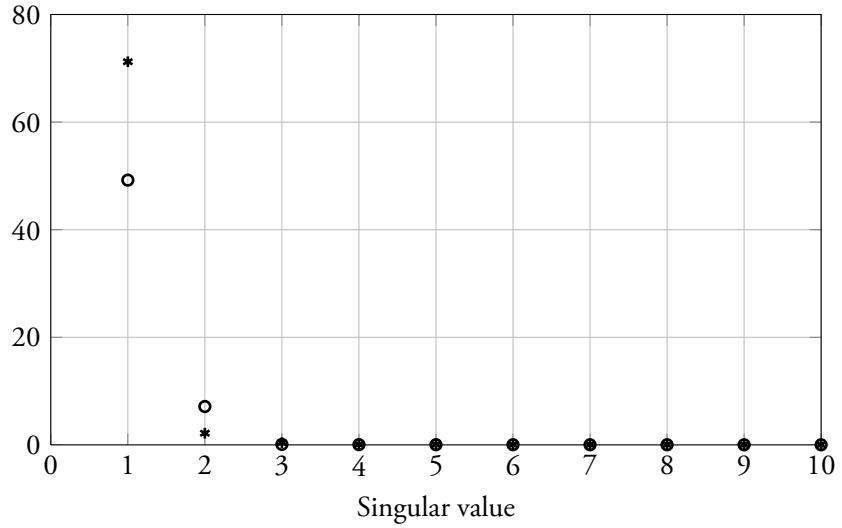


Figure 5.3 The singular values for the extended observability matrices for G_1 (*) and G_3 (o).

direct method is now applied to the extended observability matrix and the state-space matrices for the two subsystems are calculated.

5.4 Results

The resulting estimates for the first subsystem are shown in Figure 5.4. It is seen that for the first subsystem the two proposed methods gives similar result. This is expected since both methods are basically estimating this subsystem in the same way. They also have about the same dynamics as the physical model and only a very small difference in gain. The difference in gain could originate from wear and tear in the real process as well as from the discretization and linearization in the physical model. The main dynamics are however captured.

The Bode magnitude plot for the second system for the two different methods is shown in Figure 5.5. For the second subsystem the methods gives slightly different results. The dynamics are about the same for the two proposed methods, i.e. the eigenvalues of the system matrix are almost identical. As the two tanks are equal it is expected that the dynamics of the first and second subsystem should be equal as can be seen in the physical model. This is the case in both of the proposed methods. The small difference in gain that can be seen between the two proposed methods is due to

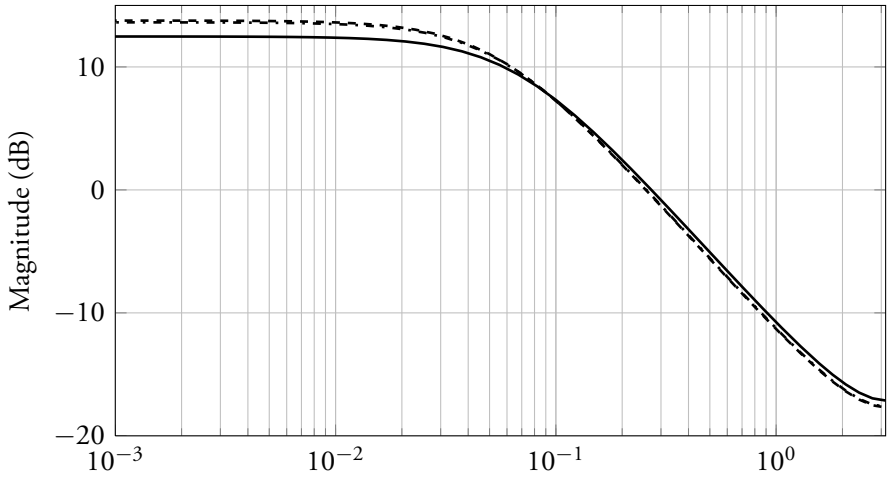


Figure 5.4 Bode magnitude plot of the first subsystem G_1 for the proposed indirect and direct method (overlapping $\cdot - \cdot -$) and the physical model (——).

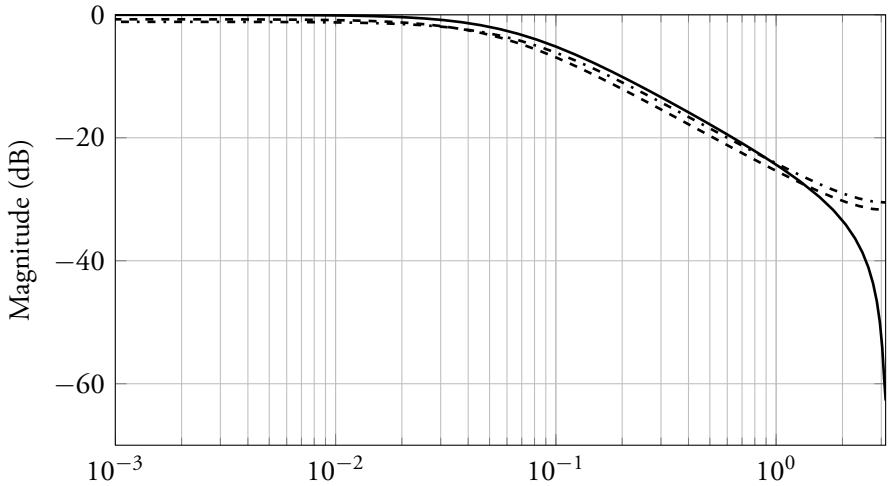


Figure 5.5 Bode magnitude plot of the second subsystem G_2 for the proposed indirect(- - -), the direct method ($\cdot - \cdot -$) and the physical model (——).

the different way the two methods calculates the state-space matrices. It is impossible to say which one is better from this since the true system is unknown.

To evaluate the performance instead the prediction error for the methods are calculated for the 200 validation data points. The prediction error is given by

$$\begin{aligned} V\left(\hat{G}_1, \hat{G}_2\right) = &\frac{1}{N}\sum_{t=1}^{N}\left(y_1(t) - \hat{G}_1(q)\,u(t)\right)^2 \\ &+ \frac{1}{N}\sum_{t=1}^{N}\left(y_2(t) - \hat{G}_1(q)\,\hat{G}_2(q)\,u(t)\right)^2 \end{aligned} \tag{5.3}$$

where $\hat{G}_1$ and $\hat{G}_2$ are the estimated models of the two subsystems. The resulting prediction errors are about the same for the two methods, 0.40 for the indirect method and 0.42 for the direct method. The norm of the measured output is 7.83 so the prediction error is relatively small. The execution time is much smaller for the direct method since the optimization, *i.e.,* the solving of the least squares problem instead of a non-convex problem, is much more computational efficient.

5.5 Summary

The two methods presented in Chapter 4 were applied to a double tank system. To be able to evaluate the methods and to gain some insight into the problem first a physical model of the system was derived. The discretization of the derived model revealed an interesting problem. While the continuous time model was in cascade form the discretized model was not. This is a well-known problem in for example errors in variable identification and is due to the violation of the zero-order hold condition for the input to the second subsystem. A few ways to handle this was also presented.

The methods were successfully applied to data from the double tank process and gave about the same performance.

Chapter 6

Nonparametric Frequency Response Estimation

The previous chapters used parametric identification algorithms, such as the prediction error method, to estimate the frequency response function, FRF, $G(e^{j\omega})$. One inherent challenge with these methods is that the choice of model structure and model order has a great impact on the quality of the estimate. One way to get around this problem is to in a first step estimate a nonparametric model of the frequency response function. With this approach no assumptions that could reduce the flexibility to accurately reproduce the true FRF are imposed. The estimates can then be used to obtain insight into various system properties, such as system order and noise characteristics, in order to proceed in a second step, with more accurate parametric methods.

Frequency response functions have in their own, of course, many applications and are used intensively in many engineering fields. Examples can be found in audio applications, power systems and vibration analysis.

In this chapter a novel method, TRIMM, is presented for estimating the FRF. The method was first introduced in Hägg et al. (2011a).

6.1 Transient Impulse Response Modeling Method

Inspired by the local polynomial method a new method to estimate the FRF was introduced by Hägg et al. (2011a). The method was named TRIMM, the Transient Impulse Response Modelling Method (Hägg and Hjalmarsson, 2012).

The Method

The method will make use of the relation (2.13). To facilitate readability these result are repeated. For an asymptotically stable system $G_0(e^{j\omega})$ the following relation holds

$$Y_N(\omega_k) = G_0(e^{j\omega_k})U_N(\omega_k) + T_N(e^{j\omega_k}) + V_N(\omega_k). \tag{6.1}$$

where the transient, or leakage error, was defined as

$$T_N(e^{j\omega_k}) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y_t^{\text{tra}} e^{-j\omega_k t}$$

and

$$y_t^{\text{tra}} = \mathbf{C}\mathbf{A}^t(x_0 - x_0^{\text{per}}).$$

The objective is to estimate the frequency response function $G_0(e^{j\omega_k})$ for the whole frequency range,

$$\omega_k = \frac{2\pi}{N}k, \quad k = 0, \ldots, N-1$$

from the DFT of the input sequence, $U_N(\omega_k)$ and the DFT of the output sequence $Y_N(\omega_k)$.

Basic Idea

The relationship between the DFT of the input signal and the output signal is given by (6.1). If the leakage or transient term, $T_N(e^{j\omega_k})$, somehow would be known, this could be used to greatly improve the quality of the estimate. The basic idea of TRIMM is to, in addition to the FRF, also try to estimate the transient terms y_t^{tra}, $t = 0, \ldots, N-1$.

However, only N complex equations are available but there are $3N$ unknowns in total, N complex parameters $G_0(e^{j\omega_k})$ and N real parameters y_t^{tra}. Hence both the frequency response and the transient terms cannot be identified simultaneously. The next step is to increase the number of equations.

Adding More Equations

Using the definition of the frequency response function (2.12), the following relation between the frequency response at two different frequencies ω_k and ω_{k+l} is derived

$$\begin{aligned} G_0(e^{j\omega_{k+l}}) - G_0(e^{j\omega_k}) &= \sum_{t=0}^{\infty} g_t^0 e^{-j\omega_{k+l}t} - \sum_{t=0}^{\infty} g_t^0 e^{-j\omega_k t} \\ &= \sum_{t=0}^{\infty} g_t^0 \left(e^{-j\omega_{k+l}t} - e^{-j\omega_k t}\right) \\ &\triangleq \sum_{t=1}^{\infty} g_t^0 \, \varphi_t \left(\omega_{k+l}, \omega_k\right), \end{aligned} \tag{6.2}$$

where $\varphi_t \left(\omega_{k+l}, \omega_k\right) \triangleq e^{-j\omega_{k+l}t} - e^{-j\omega_k t}$.

Inserting (6.2) into (6.1) gives

$$\begin{aligned} Y_N(\omega_{k+l}) = G_0(e^{j\omega_k}) U_N(\omega_{k+l}) + \sum_{t=1}^{\infty} g_t^0 \, \varphi_t(\omega_{k+l}, \omega_k) U_N(\omega_{k+l}) \\ + \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} y_t^{\text{tra}} e^{-j\omega_{k+l}t} + V_N(\omega_{k+l}). \end{aligned} \tag{6.3}$$

To increase the number of equations equation (6.3) can be used. For each frequency k, the $2L$ neighbors corresponding to $l = -L, \ldots, L$ in (6.3). This increases the number of equations from N to $(2L+1)N$.

Remark 6.1 The use of the neighboring frequencies is not necessary; any frequency pair can be used to increase the number of equations.

The increase of the number of equations comes with a cost; a new term has been added

$$\sum_{t=1}^{\infty} g_t^0 \, \varphi_t(\omega_{k+l}, \omega_k) \, U_N(\omega_{k+l}) \tag{6.4}$$

which introduces additional errors. To counteract this, the impulse response coefficients of the system, $\{g_t^0\}_{t=1}^{\infty}$ in the sum of (6.4), are also included in the estimation problem. However, this gives infinitely many parameters from finite data. The final step is hence to reduce the number of parameters to estimate.

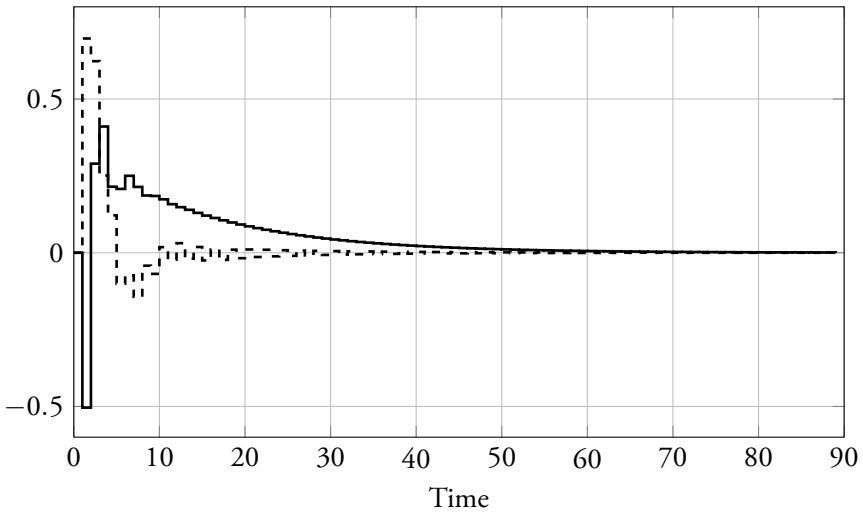


Figure 6.1 The impulse responses of two different asymptotically stable systems.

Reducing the Number of Parameters

One of the assumptions of the true system is that it is asymptotically stable. It is well known that impulse response coefficients- for asymptotically stable systems- decay to zero as time tends to infinity, *i.e.,*

$$g_t^0 = \mathbf{C}\mathbf{A}^t\mathbf{B} \to 0, \quad \text{as } t \to \infty,$$

since the eigenvalues of $\mathbf{A}$ have a magnitude less than one. Two typical impulse responses for asymptotically stable systems can be seen in Figure 6.1. The same property can be noted for the transients

$$y_t^{\text{tra}} = \mathbf{C}\mathbf{A}^t(x_0 - x_0^{\text{per}}) \to 0, \quad \text{as } t \to \infty.$$

Hence it is natural to reduce the number of estimated parameters by truncating the sums corresponding to these coefficients,

$$\sum_{t=0}^{N-1} y_t^{\text{tra}} e^{-j\omega_{k+l}t} \approx \sum_{t=0}^{n_1-1} y_t^{\text{tra}} e^{-j\omega_{k+l}t}$$

$$\sum_{t=1}^{\infty} g_t^0 \, \varphi_t(\omega_{k+l}, \omega_k) \approx \sum_{t=1}^{n_2} g_t^0 \, \varphi_t(\omega_{k+l}, \omega_k)$$

and only estimate n_1 parameters of the transient and n_2 parameters of the impulse response.

With these approximations the regressor model can be written as

$$\begin{aligned}\hat{Y}_N(\omega_{k+l}) = G_0(e^{j\omega_k})U_N(\omega_{k+l}) + \frac{1}{\sqrt{N}}\sum_{t=0}^{n_1-1} y_t^{\text{tra}} e^{-j\omega_{k+l}t} \\ + \sum_{t=1}^{n_2} g_t\,\varphi_t(\omega_{k+l},\omega_k)U_N(\omega_{k+l}) + V_N(\omega_{k+l}).\end{aligned} \tag{6.5}$$

Least Squares Problem

Collecting the extra parameters in a vector

$$\theta = \begin{bmatrix} y_0^{\text{tra}} & \cdots & y_{n_1-1}^{\text{tra}} & g_1 & \cdots & g_{n_2}\end{bmatrix}^T$$

the estimate of the FRF and the additional variables is then calculated from a global linear least squares problem over all frequencies

$$\arg \min_{\{G(e^{j\omega_k})\}_{k=0}^{N-1},\theta} \sum_{k=0}^{N-1}\sum_{l=-L}^{L} |Y_N(\omega_{k+l}) - \hat{Y}_N(\omega_{k+l})|^2 \tag{6.6}$$

where $\hat{Y}_N(\omega_{k+l})$ is the regressor model (6.5).

Remark 6.2 It is possible to make this method local by estimating a new θ for each frequency ω_k.

Comparison with local polynomial method

The presented method was inspired by the local polynomial method. Hence the methods have some similarities, but also some differences. Here the main differences between the two methods are highlighted.

- The local polynomial method uses the fact that the transient term is smooth and approximates it with a Taylor series expansion while TRIMM explores the structure (6.1) of the transient term when making the approximation.
- The same "surplus" variables $\{y_k^{\text{tra}}\}_{k=0}^{n_1-1}$ and $\{g_k\}_{k=1}^{n_2}$ are used for all frequencies. This leads to a global least squares problem. The local polynomial method uses different parameters for all frequencies and can hence be solved locally. This leads to a smaller least squares problem at each frequency.

- The Local Polynomial Method makes use of the smoothness of the frequency response when introducing more equations. The relation between the FRF for two frequencies is approximated with a Taylor series. In TRIMM this relation is modeled in terms of the impulse response of the system, derived from the definition of the frequency response.

Efficient Computation

The least squares problem (6.6) has $(2L+1)N$ equations and $N+n_1+n_2$ unknowns. Even for modest N, n_1 and n_2 the direct solution becomes computational prohibitive. The problem has, however, a special structure that can be explored.

The regressor model (6.5) can be written as

$$Y = \Phi_1 G + \Phi_2 \theta$$

where

$$Y = \begin{bmatrix} Y_0 \\ \vdots \\ Y_{N-1} \end{bmatrix}, \quad Y_k = \begin{bmatrix} Y_N(\omega_{k-L}) \\ \vdots \\ Y_N(\omega_{k+L}) \end{bmatrix}, \quad G = \begin{bmatrix} G_0(e^{j\omega_0}) \\ G_0(e^{j\omega_1}) \\ \vdots \\ G_0(e^{j\omega_{N-1}}) \end{bmatrix}, \quad \theta = \begin{bmatrix} y_0^{\text{tra}} \\ \vdots \\ y_{n_1-1}^{\text{tra}} \\ g_1 \\ \vdots \\ g_{n_2} \end{bmatrix}$$

and Φ_1, Φ_2 are regressor matrices defined in Appendix A.1. This is a *block angular least squares problem*, see for example Cox (1990). Here orthogonal transformations will be used to decompose the problem into two sub problems; the first to estimate the extra parameters, θ, and then in a second step estimate the FRF.

The orthogonal decomposition of the regressor matrix Φ_2 onto the space spanned by Φ_1 can be written as

$$\Phi_2 = \Phi_2^{\parallel} + \Phi_2^{\perp}$$

where $\Phi_2^{\perp}$ is orthogonal to both Φ_1 and $\Phi_2^{\parallel}$, *i.e.,*

$$\begin{aligned}\Phi_2^{\perp} &= (\mathbf{I} - \Phi_1(\Phi_1^*\Phi_1)^{-1}\Phi_1^*)\Phi_2 \\ &= \mathrm{Diag}\{(\mathbf{I} - U_k(U_k^*U_k)^{-1}U_k^*)\}\Phi_2 \\ &= \begin{bmatrix} (\mathbf{I} - U_0(U_0^*U_0)^{-1}U_0^*)\Phi_2^0 \\ \vdots \\ (\mathbf{I} - U_{N-1}(U_{N-1}^*U_{N-1})^{-1}U_{N-1}^*)\Phi_2^{N-1} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \Phi_{2,0}^{\perp} \\ \vdots \\ \Phi_{2,N-1}^{\perp} \end{bmatrix}.\end{aligned} \tag{6.7}$$

This gives

$$Y = \Phi_1 G + \Phi_2^{\parallel}\theta + \Phi_2^{\perp}\theta.$$

Multiplying both sides with $(\Phi_2^{\perp})^*$ yields

$$\begin{aligned}(\Phi_2^{\perp})^*Y &= \underbrace{(\Phi_2^{\perp})^*\Phi_1}_{=0} G + \underbrace{(\Phi_2^{\perp})^*\Phi_2^{\parallel}}_{=0}\theta + (\Phi_2^{\perp})^*\Phi_2^{\perp}\theta \\ &= (\Phi_2^{\perp})^*\Phi_2^{\perp}\theta.\end{aligned} \tag{6.8}$$

Now a least squares estimate $\hat{\theta}$ of θ can be calculated from (6.8) as

$$\hat{\theta} = \left((\Phi_2^{\perp})^*\Phi_2^{\perp}\right)^{-1}(\Phi_2^{\perp})^*Y. \tag{6.9}$$

Finally the least squares estimate $\hat{G}$ of G can be computed as

$$\begin{aligned}\hat{G} = (\Phi_1^*\Phi_1)^{-1}\Phi_1^*\Delta Y &= \mathrm{Diag}\{(U_k^*U_k)^{-1}U_k^*\}\Delta Y \\ &= \begin{bmatrix} (U_k^*U_k)^{-1}U_k^*\Delta Y_0 \\ \vdots \\ (U_{N-1}^*U_{N-1})^{-1}U_{N-1}^*\Delta Y_{N-1} \end{bmatrix}\end{aligned} \tag{6.10}$$

where

$$\Delta Y = Y - \Phi_2\hat{\theta}. \tag{6.11}$$

To summarize, the estimate $\hat{G}$ can be calculated as

i) Compute the projections $\Phi_2^{\perp}$ from (6.7).

ii) Calculate an estimate of θ using (6.9). This corresponds to solving a system of $n_1 + n_2$ equations in $n_1 + n_2$ unknowns.

iii) Compute ΔY defined in (6.11).

iv) Compute $\hat{G}$ from (6.10).

6.2 Asymptotic Distribution of Estimates

In this section the asymptotic properties of the estimated parameters from TRIMM will be studied. In particular two extreme cases will be considered; the first case is when no neighboring frequencies are used in the estimation, *i.e.,* $L = 0$. The second case is when, for each frequency, all other frequencies available are used as neighbors, *i.e.,* $L = N/2$.

The case $L = 0$

The design parameter L governs how many neighboring frequencies that is to be included in the estimation. If $L = 0$ there are $N + n_1$ unknowns to estimate but only N equations. Hence it is only possible to estimate N unknowns. In this case, by setting $n_1 = 0$, TRIMM is equivalent to the Empirical Transfer-Function Estimate ETFE, *i.e.,*

$$\hat{G}(e^{j\omega_k}) = \hat{G}^{\text{ETFE}}(e^{j\omega_k}) = \frac{Y_N(\omega_k)}{U_N(\omega_k)}.$$

Hence for this case the conclusions from Section 2.4 hold. The main observations were

- The ETFE is an asymptotically, in N, unbiased estimate of $G_0(e^{j\omega_k})$.
- The variance of the ETFE is given by the noise-to-signal ratio at the frequency. The variance does not decrease as N increases as there is no smoothing unless the input is periodic in which case the number of frequencies remains fixed as the sample size grows.
- The estimates at different frequencies are asymptotically uncorrelated.

The case $L = N/2$

Now let us look at the other extreme case, when $L = N/2$. This means that the estimation of the frequency response at one frequency involves all other frequencies. To simplify the analysis the method will be modified slightly. This is done to obtain

N equations instead of $N+1$ equations as this will simplify the expressions. The neighboring frequencies used will now be

$$\omega_{k+l} = \frac{2\pi}{N}(k+l), \quad l = -\frac{N}{2}+1,\ldots,\frac{N}{2}. \tag{6.12}$$

With this small modification the following theorem can be used to simplify the analysis.

Theorem 6.1 (Equivalence of global and local problem) *If the frequencies given by (6.12) are used, the solution to the linear least squares problem (6.6) is equivalent to the solutions from the N local linear least squares problems*

$$\arg \min_{G(e^{j\omega_k}),\theta_1^k,\theta_2^k} \|Y^k - \hat{Y}^k\|_2^2, \quad k = 0,\ldots,N-1$$

where

$$\hat{Y}^k = \Phi_1^k G(e^{j\omega_k}) + \Phi_2^k\theta_1^k + \Phi_3^k\theta_2^k.$$

The expressions for $\hat{Y}^k$, Φ_1^k, Φ_2^k, Φ_3^k, θ_1^k and θ_2^k are given in Appendix A.2. Furthermore the estimate $\begin{bmatrix}\hat{\theta}_1^k & \hat{\theta}_2^k\end{bmatrix}^T$ will be equal for all $k = 0,\ldots,N-1$.

Proof. See Appendix A.2. □

Remark 6.3 Since the solution $\begin{bmatrix}\hat{\theta}_1^k & \hat{\theta}_2^k\end{bmatrix}^T$ is equal for all local problems this only has to be calculated once, something that can be used to further improve the efficiency of the method in this special case.

Time equivalent problem

Taking the inverse discrete Fourier transform (IDFT) of the local regressor model from Theorem 6.1

$$\hat{Y}^k = \Phi_1^k G(e^{j\omega_k}) + \Phi_2^k\theta_1^k + \Phi_3^k\theta_2^k$$

gives the time domain predictor model

$$\hat{\mathbf{y}} = \phi_1 G(e^{j\omega_k}) + \phi_2\theta_1 + \phi_3^k\theta_2$$

where

$$\hat{\mathbf{y}} = \begin{bmatrix}\hat{y}(0)\\ \hat{y}(1)\\ \vdots\\ \hat{y}(N-1)\end{bmatrix}, \quad \phi_1 = \begin{bmatrix}u(0)\\ u(1)\\ \vdots\\ u(N-1)\end{bmatrix}, \quad \phi_2 = \begin{bmatrix}\mathbf{I}_{n_2}\\ 0\end{bmatrix}$$

$$\phi_3^k = \begin{bmatrix} u(N-1) - u(0)e^{-j\frac{2\pi}{N}k} & \cdots & u(N-n_2) - u(0)e^{-j\frac{2\pi}{N}n_2k} \\ u(0) - u(1)e^{-j\frac{2\pi}{N}k} & \cdots & u(N-n_2+1) - u(1)e^{-j\frac{2\pi}{N}n_2k} \\ \vdots & & \vdots \\ u(N-2) - u(N-1)e^{-j\frac{2\pi}{N}k} & \cdots & u(N-n_2-1) - u(N-1)e^{-j\frac{2\pi}{N}n_2k} \end{bmatrix}.$$

From this the time equivalent method of TRIMM when $L = N/2$ can be derived. For time t the model equation can be written as

$$y(t) = G(e^{j\omega_k})u(t) + y_t^{\text{tra}} + \sum_{p=1}^{n_2} g_p \left(u(t-p) - u(t)e^{-j\frac{2\pi}{N}kp} \right) + v(t).$$

The time equivalent local least squares problem can then be written as

$$\arg \min_{G(e^{j\omega_k}),\theta_1^k,\theta_2^k} \sum_{t=0}^{N-1} |y(t) - \hat{y}(t)|^2 \tag{6.13}$$

where $y(t)$ are measurements of the output from the system under consideration.

For the time equivalent method it is now possible to analyze the asymptotic properties of the estimates using well-known results from system identification, see for example Ljung (1999).

Variance

First the variance of the estimates when the true system can be described by the model used by TRIMM is considered. The following theorem of the variance of the estimates can be formulated.

Theorem 6.2 (Variance of the parameter estimates) *Let the observed output $y(t)$ be given by*

$$y(t) = G_0(q)u(t) + y_t^{\text{tra}} + e(t).$$

Assume that $u(t)$ and $e(t)$ are both independent zero mean white noise processes with variances λ_u and λ_e, respectively. The transient is denoted y_t^{tra}. Furthermore assume that the true system, $G_0(q)$ can be described by a FIR-filter with impulse response coefficients g_t and $g_t = 0$ for $t > n_2$.

Then the asymptotic distribution of the estimates with (6.13) for the frequency response $G(e^{j\omega})$ and the impulse response coefficients, g_t, $t = 1, \ldots, n_2$ at a fix frequency, not

changing with N, is given by

$$\sqrt{N}\left(\begin{bmatrix}\hat{G}(e^{j\omega})\\ \hat{\theta}_2^k\end{bmatrix} - \begin{bmatrix}G_0(e^{j\omega})\\ \theta_{20}^k\end{bmatrix}\right) \in \text{As}\,\mathcal{N}(0, \mathbf{P}_\theta)$$

with

$$\mathbf{P}_\theta = \frac{\lambda_e}{\lambda_u}\begin{bmatrix}(n_2+1) & \Delta^*\\ \Delta & \mathbf{I}\end{bmatrix}$$

and

$$\Delta = \begin{bmatrix}e^{-j\omega} & e^{-j2\omega} & \cdots & e^{-jn_2\omega}\end{bmatrix}^T.$$

Proof. See Appendix A.3. □

From Theorem 6.2 the following observations can be made:

- The impulse response coefficients will always have the same variance, λ_e/λ_u, regardless of how many extra parameters that are estimated.
- The variance of $G(e^{j\omega_k})$ is proportional to the number estimated impulse response parameters. Furthermore the variance is equal for all frequencies.
- If a FIR-model with n_2 parameters is directly estimated the asymptotical variance of the FRF would be $n_2\lambda_e/\lambda_u$. The extra "1" for TRIMM is the price for simultaneously estimating the FIR-parameters and the frequency response function.

Example 6.1 (Variance of estimates)
Consider the FIR system:

$$y(t) = 0.7u(t-1) + 0.9u(t-2) - 0.5u(t-3) + e(t) \tag{6.14}$$

where $e(t)$ is white zero mean Gaussian noise with variance $\lambda_e = 1$. The objective is to estimate the frequency response using TRIMM. The design parameters for the method are set to $n_2 = 3$ and $L = N/2$, hence the true system can be captured by the model structure used in TRIMM.

The system (6.14) is simulated with the input signal as a white Gaussian noise processes with zero mean and unit variance. The number of samples is $N = 1000$ and

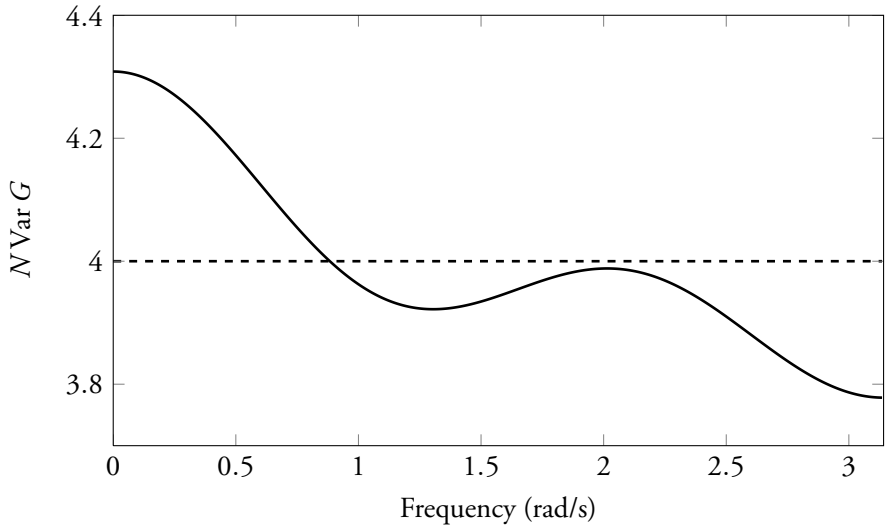


Figure 6.2 The theoretical variance (- - -) and the simulated variance (——) for Example 6.1.

1000 Monte-Carlos simulations are performed. From Theorem 6.2 it is expected that the asymptotic variance of the estimate should be

$$\text{Var } G(e^{j\omega}) = \frac{\lambda_e}{\lambda_u}\frac{1}{N}(n_2+1) = \frac{4}{N}.$$

The variance of the estimates from the simulated data is then calculated at each frequency and shown in Figure 6.2. It is seen that the theoretical values and the simulated values are close. Furthermore the variances of the estimates of the impulse response coefficients, not shown, are close to their theoretical values.

Bias

Next the bias introduced by undermodeling is studied. The results are summarized in the following theorem.

Theorem 6.3 (Bias of the estimates) *Let the observed output $y(t)$ be given by*

$$y(t) = \sum_{k=1}^{n_2} g_k^0 u(t-k) + \sum_{k=n_2+1}^{\infty} g_k^0 u(t-k) + e(t)$$

where $u(t)$ and $e(t)$ are both independent zero mean white noise processes with variances λ_u and λ_e, respectively. Estimating the parameters θ^k using (6.13) gives the following asymptotic, in N, bias

$$\begin{bmatrix} \hat{G}(e^{j\omega}) \\ \hat{g}_1 \\ \vdots \\ \hat{g}_{n_2} \end{bmatrix} - \begin{bmatrix} G(e^{j\omega}) \\ g_1^0 \\ \vdots \\ g_{n_2}^0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \sum_{t=n_2+1}^{\infty} g_t^0 e^{-j\omega t}. \tag{6.15}$$

Proof. See Appendix A.4. □

From Theorem 6.3 the following observations can be made:

- The estimates of the impulse response coefficients are unbiased, just as if a FIR model would be directly estimated.
- The estimate of $G(e^{j\omega_k})$ is biased, and the bias is equal to the fourier transform of the unmodeled impulse response coefficients. The bias is in this case frequency dependent.

Example 6.2 (Bias of estimates)
Consider the true system

$$y(t) = 0.7u(t-1) + 0.9u(t-2) - 0.5u(t-3) + 0.1u(t-4).$$

The input to the system is white gaussian noise with unit variance. The design parameters for TRIMM is chosen as $n_2 = 2$, *i.e.,* there is undermodeling. The theoretical bias is given by (6.15), and can be calculated to

$$\hat{\theta} - \theta_0 = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \sum_{t=n_2+1}^{\infty} g_t^0 e^{-j\omega t} = \begin{bmatrix} 0.5e^{-j3\omega} - 0.1e^{-j4\omega} \\ 0 \\ 0 \end{bmatrix}.$$

Monte-Carlo simulations with $N = 1000$ are performed and the bias is calculated. The absolute value of the error between the theoretical bias and the mean of the simulated bias is shown in Figure 6.3. The bias of the estimates of the impulse response coefficients are close to zero.

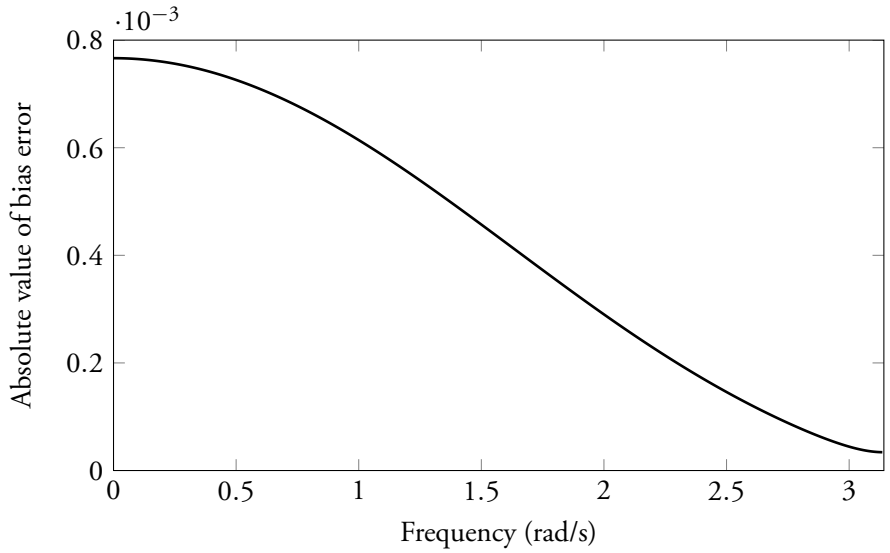


Figure 6.3 The absolute value of the error between the theoretical bias and the simulated bias for Example 6.2.

Discussion

As shown in Theorem 6.2, the asymptotic variance is $\frac{\lambda_e}{\lambda_u}(n_2+1)$, when $L = N/2$ is used in TRIMM. However, if one instead would directly estimating a FIR-model with n_2 parameters the asymptotical variance would be $\frac{\lambda_e}{\lambda_u}n_2$. The extra "1" for TRIMM is the price for simultaneously estimating the FIR-parameters and the frequency response function. From this one might say that one should always estimate a FIR-model and then calculate the FRF as the DFT of the estimated FIR-parameters. While this is true for this extreme case, this observation is not true for general values of L.

For $L = 0$ it was showed that the method boils down to the well-known Empirical Transfer Function Estimate. The ETFE is asymptotically unbiased but the variance does not decrease as the number of data N tends to infinity as there is no smoothing. For $L = N/2$ on the other hand, it was showed that the variance decreases as N increases, while the asymptotic bias is constant, not decreasing with increasing N. So the design parameter L seems to be a smoothing factor and the choice of L is a tradeoff between bias and variance. For small values of L the smoothing is low. This gives a reduced bias but at a cost of higher variance and vice versa for large L.

Table 6.1 Distribution of the randomized system parameters.

Parameter	Distribution
Order of system model G	Uniform$[1, 20]$
Order of noise model H	Uniform$[1, 20]$
Samples, N	Uniform$[50, 600]$
Noise variance, Var $e(t)$	Uniform$[0, 1.5]$
Initial state x_0	Normal$(0, I)$

6.3 Simulations

To verify the feasibility of the proposed method, TRIMM, simulations are performed. The aim is to compare it to some of the existing algorithms.

Random Systems

First the performance of TRIMM is compared to the Local Polynomial Method, when applied to a large amount of randomized linear systems. The input-output relation for the simulated data is

$$y(t) = G_0(q)u(t) + H_0(q)e(t)$$

where $y(t)$ is the output signal, $u(t)$ the input signal and $e(t)$ is white Gaussian noise. In each simulation both the system model, $G_0(q)$, and the noise model $H_0(q)$ are randomized with the MATLAB command *drss* (Matlab, 2011). The systems are then normalized to unit H_2 norm. Furthermore the order of the system, the order of the noise model, the length of the sampled data sequence, N, and the noise variance are independently randomized according to Table 6.1. Four of the random systems are shown in Figure 6.4.

Remark 6.4 The MATLAB command *drss* basically places poles and zeros randomly. Hence zero-poles cancelations or poles close to zeros are possible and the system might be approximated well by a lower order system. The randomized system order here is the number of randomized poles by *drss* and not the true order of the system. The randomized system has often a lower order, or can be approximated well by a lower order system.

The input is chosen as white Gaussian noise with unit variance and new realizations of the noise and the input are generated in each simulation. The used settings are $n_1 = n_2 = 20$ and $L = 10$ for the proposed method and the recommended values (Schoukens et al., 2009) $R = 2$ and $n = 3$ for the Local Polynomial Method. A total

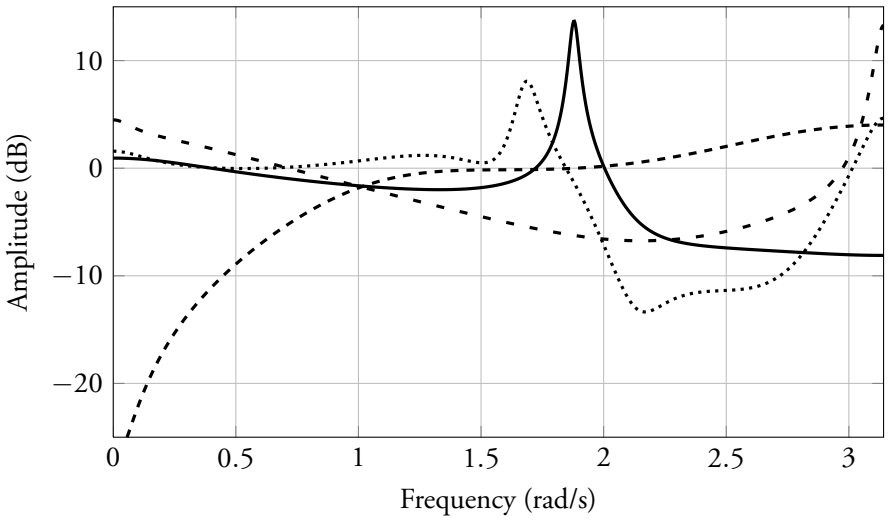


Figure 6.4 Four realizations of the random systems.

of 4000 simulations after each simulation and after each simulation the ratio between the mean square errors

$$MSE = \frac{1}{N}\sum_{k=0}^{N-1}\left|G_0(e^{j\frac{2\pi k}{N}}) - \hat{G}(k)\right|^2,$$

of the proposed method and the local polynomial method is calculated. A histogram of the ratios is shown in Figure 6.5. Values above 1 means that the local polynomial method has a lower MSE and a value below 1 means that TRIMM has a lower MSE. It is clearly seen that the proposed method in general gives lower MSE with an average improvement of about a factor 9. In 95% of the simulations the proposed method performed better.

The computational complexity of TRIMM is much larger than for LPM, the average time per iteration was 4.8 s compared to 0.06 s.

Resonant System

Next a single resonant system is considered. The performance will be compared to the local polynomial method and to the Blackman-Tukey spectral analysis method.

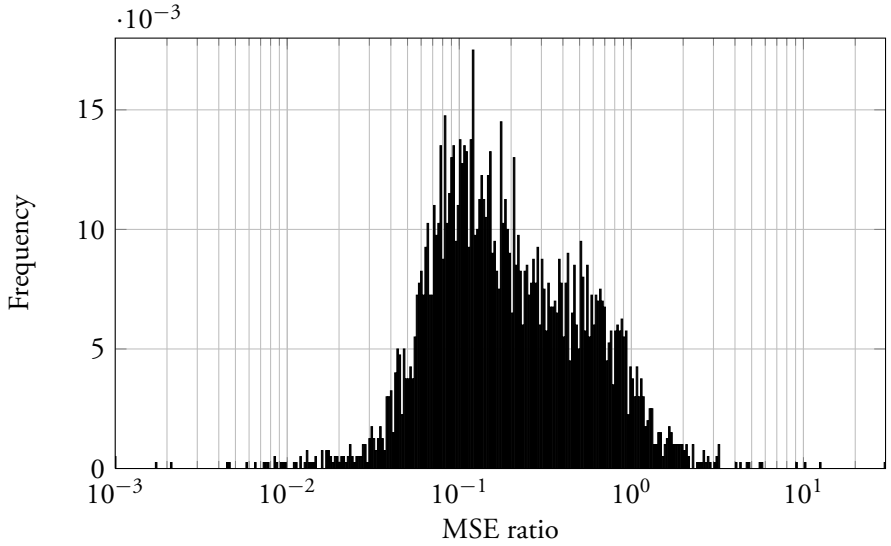


Figure 6.5 Histogram of the ratio between the MSE of TRIMM and LPM for the 4000 simulations.

Consider the continuous time system

$$G_0(s) = \frac{\omega_1^2}{s^2 + 2\xi\omega_1 s + \omega_1^2} + \frac{\omega_2^2}{s^2 + 2\xi\omega_2 s + \omega_2^2}$$

with $\omega_1 = 5$, $\omega_2 = 3\omega_1$ and $\xi = 0.1$. The system is sampled with sampling period $T_s = 0.1$.

The input is chosen as white Gaussian noise with unit variance and the output of the system is disturbed with white Gaussian noise with variance λ_e. The estimation error will be assessed via Monte-Carlo simulations. In each simulation new realizations of both the noise and the input are generated. $N = 100$ samples are collected of the output.

The used settings are the recommended $n_1 = n_2 = 20$, $L = 10$ for the proposed method and $R = 2$ and $n = 3$ for the local polynomial method. For the Blackman-Tukey method a Hann windows is used. By trial and error a good window width for this case is found to be 45 lags. Figure 6.6 shows $\left|G_0(e^{j\omega_k})\right|$ together with the resulting mean of the estimation errors over 500 Monte Carlo simulations in the noise free case ($\lambda_e = 0$) for the three methods. Since there is no noise it should be possible to increase the number of parameters estimated by TRIMM to enhance the performance. By means of a search in the parameter space the values $n_1 = n_2 = n_3 = 36$ and $L = 30$ are found

to give good performance for this setup. In Figure 6.6 the result with this particular choice of parameters is also shown.

The resulting Mean Square Errors are 0.24 for TRIMM with standard settings, 0.37 for the local polynomial method and 0.42 for the Blackman-Tukey method. The "optimized" parameters for TRIMM yield a MSE of 0.06.

Figure 6.7 shows the result under the same conditions as before, save for that the noise variance is $\lambda_e = 0.3$. This gives a MSE of 0.35 for TRIMM, 0.75 for the local polynomial method and 0.53 for the Blackman-Tukey method. The parameters optimized for the noise free case, $n_1 = n_2 = 36$ and $L = 30$ yield a MSE of 0.43.

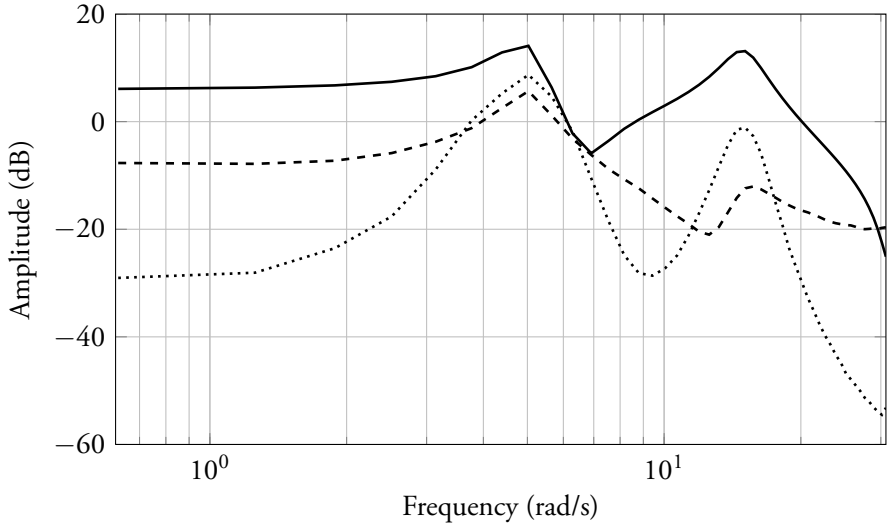
6.4 Summary

A good nonparametric estimate of the frequency response is useful in many engineering applications. It can also be used to gain insight into the identification problem at hand.

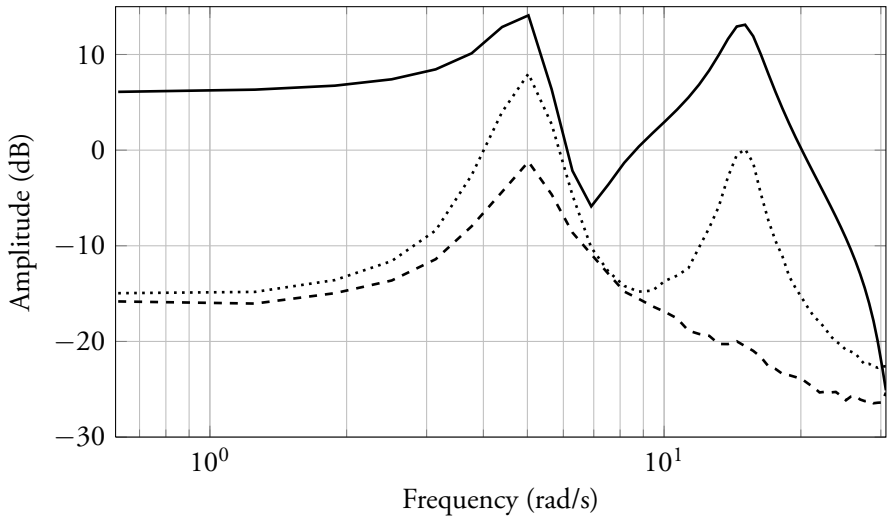
In this chapter TRIMM was introduced as a new method for estimating the frequency response of a linear system. It uses a parametrization of the known structure of the transient or leakage error. To increase the number of equations, a basic relation between the frequency responses at two different frequencies was used. The frequency response, together with the extra parameters, is then found by solving a global least squares problem.

A first attempt to analyze the algorithm was also performed. Two extreme cases were considered; $L = 0$ when none of the information from neighboring frequencies were used and $L = N/2$ when for each frequency, all available neighbors are used. The result seems to suggest that the choice of the design parameter L, the number of neighboring frequencies, is a trade-off between bias and variance.

Finally the proposed method was compared to some existing methods by simulations with promising results.

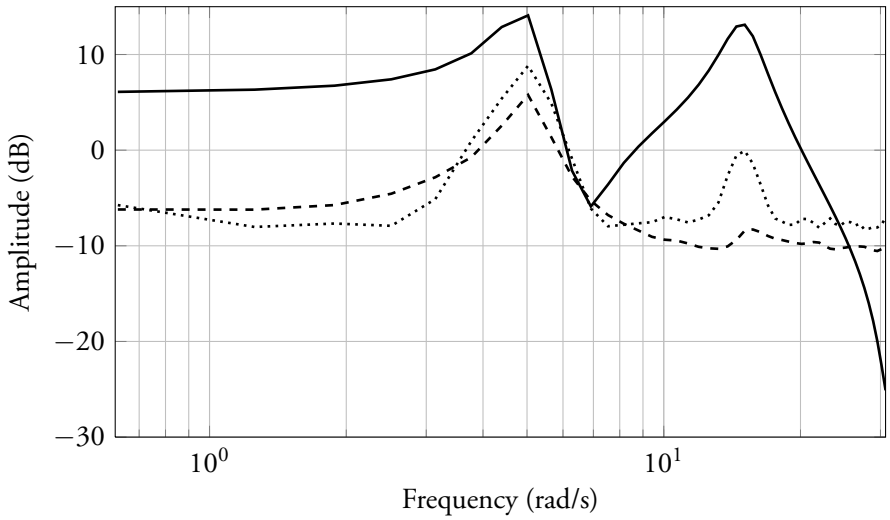


(a) TRIMM with $n_1 = n_2 = 20$ and $L = 10$ (- - -) and LPM (·····)

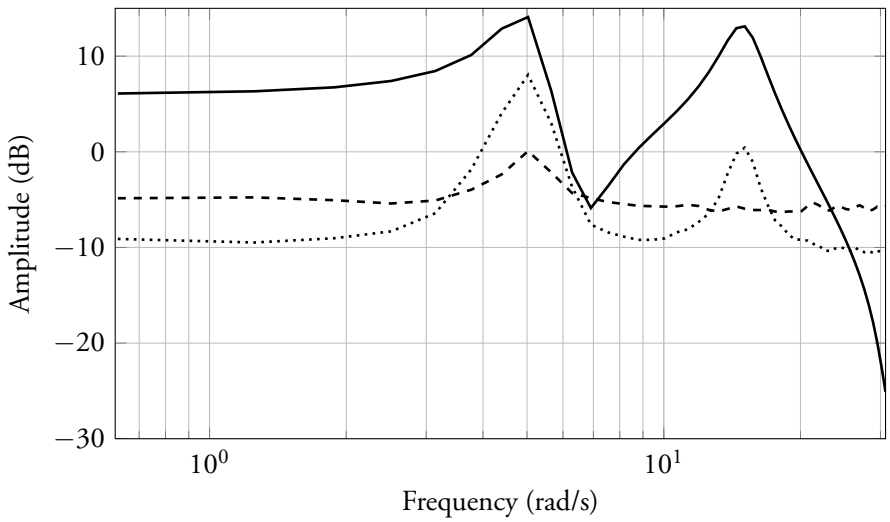


(b) TRIMM with $n_1 = n_2 = 36$ and $L = 30$ (- - -) and Blackman-Tukey with lag-window width of 45 (·····).

Figure 6.6 Frequency by frequency mean square error of the estimation with no noise ($\lambda_e = 0$). True frequency response is shown as (——).



(a) TRIMM with $n_1 = n_2 = 20$ and $L = 10$ (- - -) and LPM (·····)



(b) TRIMM with $n_1 = n_2 = 36$ and $L = 30$ (- - -) and Blackman-Tukey with lag-window width of 45 (·····).

Figure 6.7 Frequency by frequency mean square error of the estimation with noise ($\lambda_e = 0.3$). True frequency response is shown as (——).

Chapter 7

Conclusions

Chapters 3-5 considered system identification of systems with known structure while in Chapter 6, a novel method for nonparametric identification of the frequency response using the structure of the transient error was presented. Here the main results from the preceding chapters are summarized and some future directions for research are discussed.

7.1 Structures in System Identification

Structures systems are common in many industrial applications, they appear in everything from process industry to large scale control systems over networks. In Chapter 3 the fundamental limitations of the accuracy of the identified model of structured systems were studied. In particular, a special structure was studied, the parallel cascade system. The motivation for looking at this structure was two folded; it is common in industrial applications as shown by a few examples and secondly, this structure together with the cascade structure considered in Wahlberg et al. (2009a) can be used to build up most feedforward interconnected systems.

Simple fir-systems were used to demonstrate the effect of sensor placement, input signals and common dynamics of the subsystems on the asymptotic properties when a Prediction Error Method (pem) is used in the identification. The main result when the subsystems have the same dynamics, and only one of the outputs from the parallel subsystems is measured, is that the quality of the estimate of the parameters belonging to the measured subsystem is not improved by also measuring the output from the final subsystem. This result was also proven for subsystems with common, general dynamics. This result is an extension of the work in Wahlberg et al. (2009a) where the same property was proved for cascade systems. Furthermore it was shown that a priori knowledge of the structure, *i.e.,* if it is known that the subsystems have the same

dynamics, should always be included in the estimation if possible. This can improve the quality of the estimates considerably.

Using PEM to identify structured systems leads to a non-convex optimization problem even for simple FIR model structures. To be able to solve the optimization problem, a good initial estimate is needed, hopefully lying in the region of attraction for the global optimum. Subspace methods have proven useful for finding such initial estimates. Subspace methods are also popular in the industry mainly due to their simplicity for the user and they can easily handle MIMO systems. In Chapter 4, two methods for identifying systems with cascade structures using subspace methods were proposed. The first, indirect method, utilizes that the state-space matrices of a system on cascade form have a certain structure. The idea is to find a transform that takes the identified system back to this form that also minimizes the mean square error between the measured output and the predicted output from the model. This leads to a non-convex optimization problem. The second, direct method, uses the fact that the structure of the extended observability matrix is known for cascade systems. The matrices for the subsystems can then be found by solving a linear least squares problem. This only works if the second subsystem is a first order system. However, this is a common case in practice. In Chapter 5 the two methods were tested on a lab process, the two tank system. Both proposed methods showed promising results.

Future Work

The results presented in this thesis for parallel cascade systems and the previous work on cascade systems covers most of the interesting cases and can easily be extended to more general structures.

In the analysis of the parallel cascade system only output error type of models were considered, *i.e.,* the measurement noise is white. Looking at more general model structures, with non-white noise, where also a noise model is estimated, would be an interesting extension. How are the asymptotic properties affected by this for example in the case of common dynamics?

The two subspace based methods for identification of cascade systems might not be practically usable. The indirect method involves a solving a non-convex optimization problem whiles the indirect method only works for systems where the second subsystem has order one. They show, however, interesting ideas how structure could be implemented into subspace methods. Especially the indirect method utilizing the structure of the extended observability matrix is an interesting idea and should be further studied.

7.2 Nonparametric Frequency Response Identification

Good knowledge of the frequency response function of a system is vital in many applications. In Chapter 6, a new method for estimating the frequency response of a linear system, TRIMM, the Transient and Impulse Response Modeling Method, was proposed. The new method is inspired by the recent developments in nonparametric identification sparked by the introduction of the Local Polynomial Method. The method utilizes the known structure of the transient or leakage error. The parameters in the parametrization of this error are the same for all frequencies in the grid. This leads to a global least squares problem. To increase the number of equations, a basic relation between the frequency responses at two different frequencies was used.

A first attempt to analyze the method was also carried out in Chapter 6. Two extreme cases were considered; $L = 0$ when none of the information from neighboring frequencies were used and $L = N/2$ when for each frequency, all available neighbors are used. The result seems to suggest that the choice of the design parameter L, the number of neighboring frequencies, is a trade-off between bias and variance. For small values of L the smoothing is low. This gives a reduced bias but at a cost of higher variance and vice versa for large L.

The feasibility of the method was tested in simulations. For the two cases considered, one with a large amount of random systems, the second with a resonant system, TRIMM showed good performance compared to current methods. From this it is hard to say anything about the general performance of the method, but it shows promising results.

Future Work

The proposed method, TRIMM, is still quite new and the understanding of the method is far from complete. The first step is to do a more thorough analysis of the method under more general conditions to be able to compare to existing methods. From this theoretical analysis it might be possible to get a better understanding on how the design parameters should be chosen. Another interesting topic that should be further studied is if, and how, one can obtain a nonparametric estimate of the noise model using TRIMM. A good nonparametric noise model can for example be used to simplify parametric identification. With a nonparametric noise model the user avoids the hard problem of selecting a good noise model structure and model order, greatly simplifying the identification process. See Schoukens et al. (2011) for a more thorough discussion on this topic. In this thesis it was assumed that the input signal was a realization of a stochastic process. However the use of periodic signals could improve the estimate

of the nonparametric frequency response function. It would be interesting to see how TRIMM performs in this case.

To improve TRIMM further, more a priori information could be utilized. In addition to the frequency response, basically two impulse responses are estimated. Still the information that the estimated parameters are impulse responses is not used in the method. If this could be incorporated somehow, using for example regularization, the quality of the estimates might improve. As an extra bonus the use of regularization could simplify the choice of the design parameters.

Many interesting questions also remain in the area of nonparametric system identification. The new ideas from related areas, such as machine learning and statistics, could maybe be used in this setting. The papers by Pillonetto and Nicolao (2010); Pillonetto et al. (2011); Chen et al. (2012) try to estimate the impulse response, utilizing for example Gaussian processes and regularization. As these new methods require little user-interaction and yields good results compared to many classical methods, it would be interesting to compare and evaluate them in the nonparametric frequency response estimation context.

Abbreviations

ARMAX	Autoregressive Moving Average model with eXogenous input
ARX	AutoRegressive model with eXogenous input
BIBO	Bounded Input Bounded Output
DAQ	Data acquisition system
DFT	Discrete Fourier Transform
EIV	Errors In Variables
ETFE	Empirical Transfer Function Estimate
FIR	Finite Impulse Response
FRF	Frequency Response Function
IDFT	Inverse Discrete Fourier Transform
LPM	Local Polynomial Method
MIMO	Multiple-Input Multiple-Outputs
ML	Maximum Likelihood
MSE	Mean Square Error
OE	Output Error
PEM	Prediction Error Method
SISO	Single-Input Single-Output
SIMO	Single-Input Multiple-Outputs
SVD	Singular Value Decomposition
TRIMM	Transient Impulse Response Modeling Method

Appendix A

Appendix for Frequency Response Function Estimate

A.1 Matrices for the regressor problem

$$U_k = \begin{bmatrix} U_N(\omega_{k-L}) \\ \vdots \\ U_N(\omega_{k+L}) \end{bmatrix}$$

$$\Phi_1 = \begin{bmatrix} U_0 & 0 & \cdots & 0 \\ 0 & U_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & U_{N-1} \end{bmatrix}$$

$$\Psi_{k,1} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & e^{-j\omega_{k-L}} & \dots & e^{-j\omega_{k-L}(n_1-1)} \\ 1 & e^{-j\omega_{k-L+1}} & \dots & e^{-j\omega_{k-L+1}(n_1-1)} \\ \vdots & \vdots & & \vdots \\ 1 & e^{-j\omega_{k+L}} & \dots & e^{-j\omega_{k+L}(n_1-1)} \end{bmatrix}$$

$$\Psi_{k,2} = \begin{bmatrix} \varphi_1(\omega_{k-L},\omega_k)U_N(\omega_{k-L}) & \cdots & \varphi_{n_2}(\omega_{k-L},\omega_k)U_N(\omega_{k-L}) \\ \vdots & & \vdots \\ \varphi_1(\omega_{k+L},\omega_k)U_N(\omega_{k+L}) & \cdots & \varphi_{n_2}(\omega_{k+L},\omega_k)U_N(\omega_{k+L}) \end{bmatrix}$$

$$\Phi_2 = \begin{bmatrix} \Psi_{0,1} & \Psi_{0,2} \\ \vdots & \vdots \\ \Psi_{N-1,1} & \Psi_{N-1,2} \end{bmatrix}$$

A.2 Proof of Theorem 6.1

It will now be showed that for the special case when $L = N/2$ and the frequencies given by (6.12) are used; solving the resulting least squares problem (6.6) is equivalent to instead solving N local problems.

First it is noted that $\Phi^{\perp}_{2,k+1}$, defined in (6.7) is just $\Phi^{\perp}_{2,k}$ circular shifted one row. The same applies for Y_k, Y_{k+1} is Y_k circular shifted one row. The terms in (6.8) can then be rewritten as

$$\Phi_2^{\perp} Y = \frac{N}{2}\Phi^{\perp}_{2,k} Y_k,$$
$$(\Phi_2^{\perp})^*(\Phi_2^{\perp}) = \frac{N}{2}(\Phi^{\perp}_{2,k})^*\Phi^{\perp}_{2,k}.$$

Equation (6.8) then becomes

$$\frac{N}{2}\Phi^{\perp}_{2,k} Y_k = \frac{N}{2}(\Phi^{\perp}_{2,k})^*\Phi^{\perp}_{2,k}\theta \iff \Phi^{\perp}_{2,k} Y_k = (\Phi^{\perp}_{2,k})^*\Phi^{\perp}_{2,k}\theta. \tag{A.1}$$

Calculating $\hat{\theta}$ from (6.8) is thus equivalent to solve the local problem (A.1). Furthermore the obtained $\hat{\theta}^k$ from the local problem will be equal for all frequencies.

Since U_k also possess the shift property, the estimate of the FRF, $\hat{G}$, from (6.10) can be calculated locally. Reordering the equations, the following least squares problem can be solved for each frequency, k, separately.

$$\arg \min_{G(e^{j\omega_k}),\theta_1^k,\theta_2^k} \| Y^k - \hat{Y}^k \|_2^2, \quad k = 0, \ldots, N-1$$

where

$$\hat{Y}^k = \Phi_1^k G(e^{j\omega_k}) + \Phi_2^k \theta_1^k + \Phi_3^k \theta_2^k,$$

with

$$Y^k = \begin{bmatrix} Y_N(\omega_k) & Y_N(\omega_{k+1}) & \cdots & Y_N(\omega_{k+N-1}) \end{bmatrix}^T$$
$$\Phi_1^k = \begin{bmatrix} U_N(\omega_k) & U_N(\omega_{k+1}) & \cdots & U_N(\omega_{k+N-1}) \end{bmatrix}^T$$

$$\Phi_2^k = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & e^{-j\omega_k} & \cdots & e^{-j(n_1-1)\omega_k} \\ \vdots & \vdots & & \vdots \\ 1 & e^{-j\omega_{k+N-1}} & \cdots & e^{-j(n_1-1)\omega_{k+N-1}} \end{bmatrix}$$

$$\Phi_3^k = \begin{bmatrix} \varphi_1(\omega_k,\omega_k)U_N(\omega_k) & \cdots & \varphi_{n_2}(\omega_k,\omega_k)U_N(\omega_k) \\ \vdots & & \\ \varphi_1(\omega_{k+N-1},\omega_k)U_N(\omega_{k+N-1}) & \cdots & \varphi_{n_2}(\omega_{k+N-1},\omega_k)U_N(\omega_{k+N-1}) \end{bmatrix}$$

$$\theta_1^k = \begin{bmatrix} y_0^{\text{tra}} & y_1^{\text{tra}} & \cdots & y_{n_1-1}^{\text{tra}} \end{bmatrix}^T$$
$$\theta_2^k = \begin{bmatrix} g_1 & \cdots & g_{n_3} \end{bmatrix}^T.$$

□

A.3 Proof of Theorem 6.2

With the setup of Theorem 6.2 the parameter estimates have the asymptotic distribution

$$\sqrt{N}\left(\begin{bmatrix} \hat{G}(e^{j\omega}) \\ \hat{\theta}_2^k \end{bmatrix} - \begin{bmatrix} G_0(e^{j\omega}) \\ \theta_{20}^k \end{bmatrix}\right) \in \text{As}\,\mathcal{N}(0, \mathbf{P}_\theta)$$

where

$$\mathbf{P}_\theta = \lambda_e \left[E\left\{ \begin{bmatrix} \phi_1 & \phi_3^k \end{bmatrix}^* \begin{bmatrix} \phi_1 & \phi_3^k \end{bmatrix} \right\} \right]^{-1},$$

see for example Ljung (1999) for details. Since ϕ_2 is not persistently exciting the effects of the transient can be neglected.

Since $u(t)$ is white $E\left\{\varphi(t)\varphi(t)^T\right\}$ can be written as

$$E\left\{ \begin{bmatrix} \phi_1 & \phi_3^k \end{bmatrix}^* \begin{bmatrix} \phi_1 & \phi_3^k \end{bmatrix} \right\} = \lambda_u \begin{bmatrix} 1 & -\Delta^* \\ -\Delta & I + \Delta\Delta^* \end{bmatrix}$$

where

$$\Delta = \begin{bmatrix} e^{-j\omega} & e^{-j2\omega} & \cdots & e^{-jn_2\omega} \end{bmatrix}^T.$$

The inverse can be calculated using blockwise inversion (Horn and Johnson, 1990)

$$\mathbf{P}_\theta = \frac{\lambda_e}{\lambda_u} \begin{bmatrix} 1 & -\Delta^* \\ -\Delta & I + \Delta\Delta^* \end{bmatrix}^{-1} = \ldots = \frac{\lambda_e}{\lambda_u} \begin{bmatrix} (n_2+1) & \Delta^* \\ \Delta & I \end{bmatrix}$$

□

A.4 Proof of Theorem 6.3

With the assumptions in Theorem 6.3 the direct solution to the least squares problem (6.13) is given by

$$\begin{aligned}\hat{\theta} &= \left[\frac{1}{N}\sum_{t=0}^{N-1}\varphi_k(t)\varphi_k(t)^T\right]^{-1}\frac{1}{N}\sum_{t=0}^{N-1}\varphi_k(t)y(t)\\ &\xrightarrow[N\longrightarrow\infty]{} \left[E\left\{\varphi_k(t)\varphi_k(t)^T\right\}\right]^{-1}E\left\{\varphi_k(t)y(t)\right\}\\ &= \theta_0 + \left[E\left\{\varphi_k(t)\varphi_k(t)^T\right\}\right]^{-1}\\ &\times E\left\{\varphi_k(t)\sum_{k=n_2+1}^{\infty}g_k u(t-k) + \varphi_k e(t)\right\}\end{aligned}$$

where

$$\varphi_k(t) = \begin{bmatrix}u(t)\\ u(t-1)\\ u(t-2)\\ \vdots\\ u(t-n_2)\end{bmatrix} - \begin{bmatrix}0\\ e^{-j\frac{2\pi}{N}k}\\ e^{-j\frac{4\pi}{N}k}\\ \vdots\\ e^{-j\frac{2n_2\pi}{N}k}\end{bmatrix}u(t)$$

and

$$\theta^k = \begin{bmatrix}G(e^{j\omega_k}) & g_1 & g_2 & \cdots & g_{n_3}\end{bmatrix}^T.$$

The expected value can be calculated as

$$E\left\{\varphi_k(t)\sum_{k=n_2+1}^{\infty}g_k u(t-k) + \varphi_k(t)e(t)\right\} = \lambda_u\begin{bmatrix}-1\\ \Delta\end{bmatrix}\sum_{k=n_2+1}^{\infty}g_k e^{-j\omega k}$$

since $u(t)$ and $e(t)$ are independent.

Hence the asymptotic bias will be

$$\begin{aligned}\hat{\theta}^k - \theta_0^k &= \left[E\left\{\varphi_k(t)\varphi_k(t)^T\right\}\right]^{-1}\lambda_u \begin{bmatrix}-1\\ \Delta\end{bmatrix}\sum_{t=n_2+1}^{\infty} g_t e^{-j\omega k}\\ &= \begin{bmatrix}(n_2+1) & \Delta^*\\ \Delta & \mathbf{I}\end{bmatrix}\begin{bmatrix}-1\\ \Delta\end{bmatrix}\sum_{t=n_2+1}^{\infty} g_t e^{-j\omega t}\\ &= \begin{bmatrix}-1\\ 0\\ \vdots\\ 0\end{bmatrix}\sum_{t=n_2+1}^{\infty} g_t e^{-j\omega t}.\end{aligned}$$

□

Bibliography

A. Alenany, H. Shang, M. Soliman, and I. Ziedan. Improved subspace identification with prior information using constrained least squares. *IET Control Theory & Applications*, 5(13): 1568–1576 (2011).

K. Åström and T. Hägglund. *Advanced PID control*. ISA-The Instrumentation, Systems, and Automation Society (2006).

L. Baramov, M. Beneš, and V. Havlena. Model for estimating stress in pressurized boiler components based on interconnections. In *Decision and Control, 2007 46th IEEE Conference on*, 1603–1608 (2007).

M. S. Bartlett. Smoothing periodograms from time series with continious spectra. *Nature*, 161: 686–687 (1948).

J. Bendat and A. Piersol. *Engineering applications of correlation and spectral analysis*. Wiley (1980).

R. B. Blackman and J. W. Tukey. The measurement of power spectra from the point of view of communications engineering - part i. *Bell System Technical Journal*, 37: 185–282 (1958).

P. Broersen. A comparison of transfer function estimators. *Instrumentation and Measurement, IEEE Transactions on*, 44(3): 657 –661 (1995).

P. Caines. Weak and strong feedback free processes. *Automatic Control, IEEE Transactions on*, 21(5): 737 – 739 (1976).

T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and gaussian processes – revisited. *Automatica* (2012). To appear.

M. Cox. *Reliable numerical computation*, chapter 13 The leasts-squares solution of linear equations with block-angular observation matrix, 227–240. Oxford Univ. Press (1990).

M. Gevers, P. Hägg, H. Hjalmarsson, R. Pintelon, and J. Schoukens. The transient impulse response modeling method and the local polynomial method for nonparametric system identification. In *16th IFAC Symposium on System Identification* (2012). To appear.

M. Gevers, L. Mišković, D. Bonvin, and A. Karimi. Identification of multi-input systems: variance analysis and input design issues. *Automatica*, 42(4): 559–572 (2006).

M. Gevers, R. Pintelon, and J. Schoukens. The local polynomial method for nonparametric system identification: Improvements and experimentation. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 4302 –4307 (2011).

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3): pp. 424–438 (1969).

P. Hägg and H. Hjalmarsson. Non-parametric frequency function estimation using transient impulse response modelling. In *16th IFAC Symposium on System Identification* (2012). To appear.

P. Hägg, H. Hjalmarsson, and B. Wahlberg. A least squares approach to direct frequency response estimation. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 2160 –2165 (2011a).

P. Hägg, B. Wahlberg, and H. Sandberg. On subspace identification of cascade structured systems. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2843 –2848 (2010).

P. Hägg, B. Wahlberg, and H. Sandberg. On identification of parallel cascade serial systems. In *Proceedings of the 18th World Congress*, 9978–9983 (2011b).

W. Heath. Choice of weighting for averaged nonparametric transfer function estimates. *Automatic Control, IEEE Transactions on*, 52(10): 1914 –1920 (2007).

R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press (1990).

S. Lacy and D. Bernstein. Subspace identification with guaranteed stability using constrained optimization. *Automatic Control, IEEE Transactions on*, 48(7): 1259 – 1263 (2003).

C. A. Larsson and P. Hägg. Generation of excitation signals with prescribed autocorrelation for input and output constrained system. In *Decision and Control, 2012. CDC 2012. 51st IEEE Conference on* (2012). Submitted.

L. Ljung. *System Identification: Theory for the User, 2nd Edition*. Prentice Hall, Upper Saddle River, New Jersey (1999).

L. Ljung. *System Identification Toolbox: User´s Guide*. The Mathworks Inc., Natick, MA (2011).

C. Lyzell, M. Enqvist, and L. Ljung. Handling certain structure information in subspace identification. In *Proceedings of the 15th IFAC Symposium on System Identification* (2009).

J. M. Maciejowski. Guaranteed stability with subspace methods. *Systems & Control Letters*, 26(2): 153 – 156 (1995).

P. Massioni and M. Verhaegen. Subspace identification of circulant systems. *Automatica*, 44(11): 2825 – 2833 (2008).

D. Materassi and G. Innocenti. Topological identification in networks of dynamical systems. *Automatic Control, IEEE Transactions on*, 55(8): 1860 –1871 (2010).

Matlab. *Control System Toolbox: User´s Guide*. Mathworks Inc., Natick, MA. (2011).

T. McKelvey. Frequency domain identification. In *Preprints of the 12th IFAC Symposium on System Identification, Santa Barbara, USA.* (2000).

T. McKelvey and G. Guérin. Non-parametric frequency response estimation using a local rational model. In *16th IFAC Symposium on System Identification, SYSID 2012* (2012). To appear.

M. Okada and T. Sugie. Subspace system identification considering both noise attenuation and use of prior knowledge. In *Decision and Control, 1996., Proceedings of the 35th IEEE*, volume 4, 3662 –3667 vol.4 (1996).

G. Pillonetto, A. Chiuso, and G. D. Nicolao. Prediction error identification of linear systems: A nonparametric gaussian regression approach. *Automatica*, 47(2): 291 – 305 (2011).

G. Pillonetto and G. D. Nicolao. A new kernel-based approach for linear system identification. *Automatica*, 46(1): 81 – 93 (2010).

R. Pintelon and J. Schoukens. *System identification : a frequency domain approach*. IEEE Press, New York (2001).

R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbé. Estimation of nonparametric noise and FRF models for multivariable systems–Part I: Theory. *Mechanical Systems and Signal Processing*, 24(3): 573 – 595 (2010a).

R. Pintelon, J. Schoukens, G. Vandersteen, and K. Barbé. Estimation of nonparametric noise and FRF models for multivariable systems–Part II: Extensions, applications. *Mechanical Systems and Signal Processing*, 24(3): 596 – 616 (2010b).

Quanser. Product information sheet S13-1 rev.C - Coupled Water Tanks (2010).

B. Sanandaji, T. Vincent, and M. Wakin. Exact topology identification of large-scale interconnected dynamical systems from compressive observations. In *American Control Conference (ACC), 2011*, 649 –656 (2011).

J. Schoukens, Y. Rolain, and R. Pintelon. Leakage reduction in frequency-response function measurements. *Instrumentation and Measurement, IEEE Transactions on*, 55(6): 2286 –2291 (2006).

J. Schoukens, Y. Rolain, G. Vandersteen, and R. Pintelon. User friendly box-jenkins identification using nonparametric noise models. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, 2148 –2153 (2011).

J. Schoukens, G. Vandersteen, K. Barbé, and R. Pintelon. Nonparametric preprocessing in system identification: a powerful tool. *European Journal of Control*, 3-4: 260–274 (2009).

T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6): 939 – 958 (2007).

T. Söderström and P. Stoica. *System Identification*. Prentice Hall International (2001).

P. Stoica and R. Moses. *Spectral analysis of signals*. Pearson Prentice Hall, Upper Saddle River, N.J. (2005).

C. Sturk, H. Sandberg, P. Trnka, V. Havlena, and J. Rehor. Structured model order reduction of boiler-header models. In *Proceedings of the 18th IFAC World Congress*, volume 18 (2011).

P. Trnka and V. Havlena. Subspace like identification incorporating prior information. *Automatica*, 45(4): 1086 – 1091 (2009).

T. Van Gestel, J. Suykens, P. Van Dooren, and B. De Moor. Identification of stable models in subspace identification by using regularization. *Automatic Control, IEEE Transactions on*, 46(9): 1416 –1420 (2001).

P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems : Theory - Implementation - Applications*. Kluwer Academic Publishers, Boston/London/Dordrecht (1996).

B. Wahlberg, H. Hjalmarsson, and J. Mårtensson. On identification of cascade systems. In *17th IFAC World Congress*, 5036–5040 (2008).

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

B. Wahlberg and H. Sandberg. Cascade structural model approximation of identified state space models. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, 4198 –4203 (2008).

B. Wahlberg, P. Stoica, and P. Babu. On estimation of cascade systems with common dynamics. In *15th IFAC Symposium on System Identification, SYSID 2009* (2009b).

P. D. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15: 70–73 (1967).

"Human beings, who are almost unique in having the ability to learn from the experience of others, are also remarkable for their apparent disinclination to do so."

Douglas Adams