

On Identification of Cascade Systems¹

Bo Wahlberg* Håkan Hjalmarsson* Jonas Mårtensson*

*Automatic Control and ACCESS, School of Electrical Engineering, KTH,
SE-100 44 Stockholm, Sweden. (bo.wahlberg@ee.kth.se)

Abstract: The objective of this contribution is to discuss some aspects of system identification of cascade systems. Models of such systems are important in for example cascade control applications. We will restrict our attention to systems with one input signal and two output signals. First, we will analyze some fundamental limitations regarding the statistical properties of such estimated models and in particular why it can be difficult to identify cascade systems where the sub-transfer functions are close to each other. We will then show how an unstructured SIMO model estimate can be used to find a cascade system model using an indirect prediction error method or balanced model reduction.

Keywords: System identification, cascade systems, variance analysis, model reduction.

1. INTRODUCTION

System identification concerns the construction and validation of mathematical models of dynamical systems from experimental data. Important issues when designing the experiment are the choices and locations of the measurement sensors. For example, if all internal states of a linear dynamical system are measurable, the state-space equations can be effectively estimated using a standard least squares method. Most classical system identification methods concern, however, single-input single-output (SISO) systems, where the input signal and corresponding output signal are pre-specified by the choice of sensors and actuators. Many of these results can be generalized to multi-input multi-output (MIMO) systems. In particular, subspace system identification methods have shown very useful when dealing with the MIMO case. It is, however, important to take the structure of the underlying system into account when specifying the model structure. The objective of this contribution is to analyze identification of systems with a cascade or series structure as illustrated in Fig. 1.

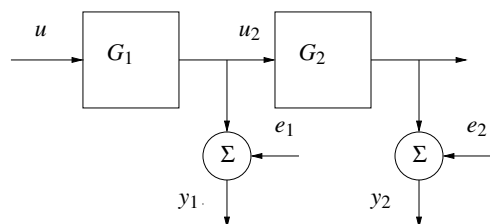


Fig. 1. Cascade system.

We will in more detail study systems with one input signal and two output signals. The following notation will be used.

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

Here the input signal is denoted by $u(t)$ and the two output signals are $y_1(t)$ and $y_2(t)$, respectively. The transfer functions are $G_1(q)$ and $G_2(q)$, and $e_1(t)$ and $e_2(t)$ are the measurement noise processes. The input to $G_2(q)$ is denoted by $u_2(t)$, and sometimes we will use the notation $G_3(q) = G_2(q)G_1(q)$ for the transfer function from $u(t)$ to $y_2(t)$.

¹ This work was partially supported by The Swedish Science Foundation and the Linnaeus Center ACCESS at KTH.

We assume that the dimensions of the input and the two output signals are all one (the scalar case).

Cascade systems are very common in both process control and in control of servo mechanical systems. In process control application, the primary output $y_2(t)$ is often a quality variable such as temperature or levels, while the secondary output $y_1(t)$ typically concerns an intermediate variable such as flows or pressures. In mechanical application $y_1(t)$ is often a rate while $y_2(t)$ is a position. The quality of the sensors for measuring the two outputs can be quite different. We will model this by the choice of the variances of $e_1(t)$ and $e_2(t)$. High variance means a poor measurement quality.

Another important example of a cascade system is when $G_2(q)$ is the transfer function of an extra sensor used to measure $y_1(t)$. This sensor may have some unknown characteristics/parameters that have to be estimated.

There are several questions that have to be answered and important user choices to be made when applying system identification methods to a data set obtained from a cascade system of the form $\{u(t), y_1(t), y_2(t)\}$. Any single-input-multi-output (SIMO) system identification method, such as subspace system identification, can be applied, but it is often not straightforward to impose the cascade model structure. For example, a subspace system identification method would return an estimate of the form

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t) \\ y_1(t) &= C^1x(t) \\ y_2(t) &= C^2x(t) \end{aligned}$$

where the specific state-space realization is indirectly given by the method. A natural estimate of $G_1(q)$ is $C^1(qI - A)^{-1}B$. Due to uncertainty in the estimate of $C^1(qI - A)^{-1}B$, it is, however, not trivial to find a minimal state space realizations of $G_1(q)$. It is also rather complicated how to find the transfer function $G_2(q)$ from estimates of the state space matrices (A, B, C^1, C^2) . This is a topic that will be further studied in this paper.

From an engineering point of view it is common to first estimate $G_1(q)$ from the 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 $\hat{u}_2(t)$ is an estimate of

the input signal $u_2(t)$ to the second subsystem $G_2(q)$. There are several options for determining this input signal. If the model estimate $\hat{G}_1(q)$ is reliable one can use $\hat{u}_2(t) = \hat{G}_1(q)u(t)$. The other extreme would be to take $\hat{u}_2(t) = y_1(t)$. It is of course also possible to use an optimal predictor of $u_2(t)$ based on the statistical properties of $e_1(t)$. In all these cases this introduces input noise/uncertainty, which may cause systematic estimation errors if a standard system identification method is applied to estimate $G_2(q)$ from the data $\{\hat{u}_2(t), y_2(t)\}$. A solution would be to use a more advanced errors-in-variables method, see Söderström [2006] for a recent overview.

It is of course possible to apply a Prediction Error Method (PEM) or the Maximum Likelihood (ML) method, Ljung [1999], to a constrained model structure that only allows for models of cascade form. Because of the product product $G_1(q)G_2(q)$ simple linear in the parameters model structures such as FIR or ARX models are not directly applicable. However, structured PEM and ML are asymptotically statistically optimal methods to solve the structured cascade system identification problem.

For state-space models one can use model structures with a constrained cascade state-space realization, e.g.

$$\begin{aligned} x_1(t+1) &= A_1x_1(t) + B_1u(t) \\ x_2(t+1) &= B_2C_1x_1(t) + A_2x_2(t) \\ y_1(t) &= C_1x_1(t) \\ y_2(t) &= C_2x_2(t) \end{aligned}$$

Here (A_1, B_1, C_1) are the matrices of a state-space realization of $G_1(q)$, and (A_2, B_2, C_2) correspond to a state-space description of $G_2(q)$.

The outline of this paper is as follows. First we will analyze the statistical properties of PEM estimates of cascade systems. This will be inspired by a recent geometric approach to variance analysis in system identification developed in Mårtensson and Hjalmarsson [2007a,b], Mårtensson [2007].

In the second part of the paper we will discuss how to find cascade model structures using indirect prediction error methods proposed in Söderström et al. [1991]. We will also discuss how this relates to model reduction.

2. VARIANCE ANALYSIS: FIR EXAMPLE

To illustrate the statistical properties of PEM cascade model estimates we will start with a very simple example. Consider the model structure

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

with two first order FIR transfer functions

$$\begin{aligned} G_1(q) &= 1 + b_1q^{-1} \\ G_2(q) &= 1 + b_2q^{-1} \end{aligned}$$

Here the parameters b_1 and b_2 need to be estimated from data $\{u(t), y_1(t), y_2(t)\}$. Let the true values of the FIR parameters be denoted by b_1^o and b_2^o . Furthermore, assume that the measurement noise processes $\{e_1(t)\}$ and $\{e_2(t)\}$ are independent white noise stochastic processes with variances λ_1 and λ_2 , respectively. Let the input signal $u(t)$ be white noise with variance 1.

The asymptotic covariance of the PEM estimate of b_1 and b_2 given N measurements of $\{u(t), y_1(t), y_2(t)\}$, which also in this case corresponds to the Cramér-Rao lower bound, is then given by M^{-1} , where

$$M = NE \{ \psi(t) \psi^T(t) \}$$

$$\psi(t) = \begin{pmatrix} \frac{u(t-1)}{\sqrt{\lambda_1}} & \frac{u(t-1) + b_2^o u(t-2)}{\sqrt{\lambda_2}} \\ 0 & \frac{u(t-1) + b_1^o u(t-2)}{\sqrt{\lambda_2}} \end{pmatrix}$$

This gives that the asymptotic variance of the parameter estimate of the first subsystem G_1 equals

$$\text{Var} \hat{b}_1 \sim \frac{1}{N} \frac{\lambda_1}{1 + \frac{\lambda_1(b_1^o - b_2^o)^2}{\lambda_2(1 + (b_1^o)^2)}}$$

We have used the notation \sim to stress the asymptotic (large data records) relation. This variance expression reveals some well known properties, but also some more novel results:

- In case only the output from the first system $y_1(t)$ is used to estimate b_1 , the asymptotic variance of the FIR estimate equals

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

This corresponds to setting $\lambda_2 = \infty$ in the general expression.

- If the quality of the first measurement $y_1(t)$ is much worse than for the second one $y_2(t)$, i.e.

$$\lambda_1 \gg \lambda_2$$

we have (by letting $\lambda_1 \rightarrow \infty$)

$$\text{Var} \hat{b}_1 \sim \frac{\lambda_2 (1 + (b_1^o)^2)}{N (b_1^o - b_2^o)^2}$$

Notice that it is not possible to separate b_1^o from b_2^o using only $y_2(t)$. Notice that $b_1^o = b_2^o$ gives further problems, as will be discussed below.

- Unless $b_1^o = b_2^o$, adding an additional sensor $y_2(t)$ strictly improves the quality of the estimate. However, if the two transfer functions are **identical**, i.e. $b_1^o = b_2^o$, no improvement is obtained from the second output $y_2(t)$, and the variance will equal

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

which is the same as for the only $y_1(t)$ case discussed above. This is a special case of a recent result of Mårtensson and Hjalmarsson [2007b], and as will be shown in next section holds for more general model structures. We will study in detail why this is the case.

It is of course possible to estimate a second order FIR model

$$\begin{aligned} G_3(q) &= G_1(q)G_2(q) \\ &= (1 + b_1q^{-1})(1 + b_2q^{-1}) = 1 + \bar{b}_1q^{-1} + \bar{b}_2q^{-2} \end{aligned}$$

from only $\{u(t), y_2(t)\}$ without any problem. It is, however, impossible to decide from only $y_2(t)$ which of the two roots of this polynomial that corresponds to G_1 . Furthermore, a first order perturbation analysis reveals that

$$\begin{pmatrix} \Delta b_1 \\ \Delta b_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ b_2 & b_1 \end{pmatrix}^{-1} \begin{pmatrix} \Delta \bar{b}_1 \\ \Delta \bar{b}_2 \end{pmatrix}$$

Hence, this mapping is not invertible if $b_1 = b_2$ and even a small perturbation in coefficient \bar{b}_1 or \bar{b}_2 can give a large perturbation in the roots. This is of course a well known result in e.g. numerical analysis. Notice that the FIR model of G_3 is more general in the sense that it also covers complex valued roots. The asymptotic covariance matrix of the estimates of \bar{b}_1 and \bar{b}_2 from measurement of only $y_2(t)$ equals

$$\text{Cov} \begin{pmatrix} \hat{\bar{b}}_1 \\ \hat{\bar{b}}_2 \end{pmatrix} \sim \frac{\lambda_2}{N} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

which by the perturbation analysis gives

$$\begin{aligned} \text{Cov} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} &\sim \frac{\lambda_2}{N} \begin{pmatrix} 1 & 1 \\ b_2^o & b_1^o \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 \\ b_2^o & b_1^o \end{pmatrix}^{-T} \\ &= \frac{\lambda_2}{N} \frac{1}{(b_1^o - b_2^o)^2} \begin{pmatrix} 1 + (b_1^o)^2 & -(1 + b_1^o b_2^o) \\ -(1 + b_1^o b_2^o) & 1 + (b_2^o)^2 \end{pmatrix} \end{aligned}$$

Hence, we again obtain

$$\text{Var} \hat{b}_1 \sim \frac{\lambda_2}{N} \frac{(1 + (b_1^o)^2)}{(b_1^o - b_2^o)^2}$$

if only $y_2(t)$ is measured.

If it is known in advance that $G_2 = G_1$, one should of course use this constraint in the model structure and in this example only estimate one parameter b_1 , since $b_2 = b_1$. The asymptotic variance of estimate will then be

$$\text{Var} \hat{b}_1 \sim \frac{1}{N} \frac{\lambda_1}{1 + \frac{\lambda_1(4 + 4(b_1^o)^2)}{\lambda_2}}$$

If $\lambda_1 \gg \lambda_2$ we obtain

$$\text{Var} \hat{b}_1 \sim \frac{1}{N} \frac{\lambda_2}{4 + 4(b_1^o)^2},$$

which of course behaves well. Hence, using the information that $G_2 = G_1$ really improves the estimated model quality.

3. VARIANCE ANALYSIS: GENERAL CASE

The observation and results from the simple example above can be extended to the general case using the geometric variance analysis framework of Mårtensson and Hjalmarsson [2007a,b]. It is also possible to use simple direct matrix computations, as will shown in this section.

Let $G_1(q, \theta_1)$ and $G_2(q, \theta_2)$ be arbitrary models of the cascade system with independent parameterizations

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

and define

$$\psi(t) = \begin{pmatrix} \frac{G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_1}} & \frac{G_2(q, \theta_2^o)G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} \\ 0 & \frac{G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} \end{pmatrix}$$

where prime denotes differentiation with respect to the parameter vectors. The asymptotic covariance matrix of the parameter estimate is then given by

$$\text{Cov} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \sim M^{-1}$$

where

$$M = NE \{ \psi(t) \psi^T(t) \}$$

See e.g., Ljung [1999] for details.

Assume now that the two true transfer functions are **identical**

$$G_2(q, \theta_2^o) = G_1(q, \theta_1^o) = G(q, \theta^o)$$

and that we are using the same model structure for both $G_1(q, \theta_1)$ and $G_2(q, \theta_2)$ and hence

$$G_2'(q, \theta_1^o) = G_1'(q, \theta_2^o) = G'(q, \theta^o)$$

This means that M will have the block structure

$$M = \begin{pmatrix} A+B & B \\ B & B \end{pmatrix}$$

where

$$A = \frac{N}{\lambda_1} E \{ [G'(q, \theta^o)u(t)][G'(q, \theta^o)u(t)]^T \},$$

$$\begin{aligned} B &= \frac{N}{\lambda_2} E \{ [G'(q, \theta^o)G(q, \theta^o)]u(t) \\ &\quad \times [G'(q, \theta^o)G(q, \theta^o)]u(t)]^T \}, \end{aligned}$$

- The matrix A^{-1} is the asymptotic covariance matrix of $\hat{\theta}_1$ if only $y_1(t)$ is available,
- The matrix B^{-1} is the asymptotic covariance matrix of $\hat{\theta}_2$ if $G_1(q)$ is known (or equally $\lambda_1 = 0$).

It is now easy to verify that

$$M^{-1} = \begin{pmatrix} A^{-1} & -A^{-1} \\ -A^{-1} & A^{-1} + B^{-1} \end{pmatrix}$$

and hence

$$\text{Cov}(\hat{\theta}_1) \sim A^{-1}$$

$$\text{Cov}(\hat{\theta}_2) \sim A^{-1} + B^{-1}$$

Key Results: If $G_2(q, \theta_2^o) = G_1(q, \theta_1^o) = G(q, \theta^o)$, we can make the following observation:

- Since the asymptotic covariance matrix of $\hat{\theta}_1$ equals A^{-1} , the quality of the estimate of θ_1 is never improved by measuring $y_2(t)$. This is independent of the quality of the second measurement!
- The covariance of the estimate of θ_2 equals $A^{-1} + B^{-1}$, and is thus always larger than or equal to the covariance of $\hat{\theta}_1$, since

$$A^{-1} + B^{-1} \geq A^{-1}$$

This is true even if one has noise free measurements of $y_2(t)$. It is also always larger than or equal to B^{-1} , which corresponds to setting $\lambda_1 = 0$, i.e. a perfect sensor for $y_1(t)$.

If it is known in advance that $G_2 = G_1$, this should of course be used in the model structure,

$$\begin{aligned} y_1(t) &= G(q, \theta)u(t) + e_1(t) \\ y_2(t) &= G(q, \theta)G(q, \theta)u(t) + e_2(t) \end{aligned}$$

The asymptotic covariance matrix M^{-1} of $\hat{\theta}$ can then calculated using

$$\psi(t) = \begin{pmatrix} \frac{G'(q, \theta^o)u(t)}{\sqrt{\lambda_1}} & \frac{2G'(q, \theta^o)G(q, \theta^o)u(t)}{\sqrt{\lambda_2}} \end{pmatrix}$$

in $M = NE \{ \psi(t) \psi^T(t) \}$. This means that $M = A + 4B$ with the notation above. Hence, the variance can be considerably smaller than for the case with separate parameterizations for which $M = A$, since

$$(A + 4B)^{-1} < A^{-1}$$

for positive definite matrices A and B .

Standard model validation tools can be used to test the hypothesis $\theta_1^o = \theta_2^o$ from the estimates. If the answer is affirmative then, in a second step, constrain the parametrization to $G_2 = G_1 = G$.

4. INDIRECT PEM

The cascade system identification problem fits well into the framework of indirect prediction error methods for system identification developed in Söderström et al. [1991]. Ordinary PEM and indirect PEM have the same asymptotic statistical properties.

Let us illustrate this idea using the previous simple FIR example. Assume that we estimate two independent model structures

$$\begin{aligned} y_1(t) &= u(t) + b_1 u(t-1) + e_1(t) \\ y_2(t) &= u(t) + \bar{b}_1 u(t-1) + \bar{b}_2 u(t-2) + e_2(t) \end{aligned}$$

For a cascade system we then have

$$\bar{b}_1 = b_1 + b_2, \quad \bar{b}_2 = b_1 b_2$$

Let us in a first step estimate the three parameters $(b_1, \bar{b}_1, \bar{b}_2)$. This can be done solving two independent least squares problems. Assume that the input signal $u(t)$ is white noise with variance 1. Then

$$\text{Cov}(\hat{b}_1, \hat{\bar{b}}_1, \hat{\bar{b}}_2)^T \sim M^{-1} = \frac{1}{N} \text{diag}(\lambda_1, \lambda_2, \lambda_2)$$

The idea of the indirect PEM approach now is to solve equations

$$b_1 = \hat{b}_1, \quad b_1 + b_2 = \hat{\bar{b}}_1, \quad b_1 b_2 = \hat{\bar{b}}_2$$

with respect to b_1 and b_2 in a weighed least squares sense,

$$\begin{aligned} V(b_1, b_2) &= [(b_1 - \hat{b}_1) (b_1 + b_2 - \hat{\bar{b}}_1) (b_1 b_2 - \hat{\bar{b}}_2)] \frac{1}{N} M \\ &\quad \times [(b_1 - \hat{b}_1) (b_1 + b_2 - \hat{\bar{b}}_1) (b_1 b_2 - \hat{\bar{b}}_2)]^T \end{aligned}$$

with respect to b_1 and b_2 . In the example with a white input signal, we obtain the cost function

$$V(b_1, b_2) = \frac{(b_1 - \hat{b}_1)^2}{\lambda_1} + \frac{(b_1 + b_2 - \hat{\bar{b}}_1)^2}{\lambda_2} + \frac{(b_1 b_2 - \hat{\bar{b}}_2)^2}{\lambda_2}$$

This is still a non-quadratic minimization problem, but is rather straightforward to handle. This idea can of course be extended to the general case.

5. MODEL REDUCTION

As illustrated by the indirect PEM example the problem of estimating two SISO models or an unstructured SIMO model and then in a second step find a structured cascade model is closely related to model approximation. It is, however, important to take the statistical properties of the model into account when doing the model reduction. A simple approach is to first estimate G_1 from $u(t)$ and $y_1(t)$ and then the series transfer function $G_3 = G_2 G_1$ from $u(t)$ and $y_2(t)$. Denote the corresponding estimates by \hat{G}_1 and \hat{G}_3 , respectively. To find the cascade transfer function G_1 and G_2 we can minimize the cost function

$$\begin{aligned} \bar{V}(G_1, G_2) &= \frac{1}{\lambda_1} \int_{-\pi}^{\pi} |G_1(e^{i\omega}) - \hat{G}_1(e^{i\omega})|^2 \Phi_u(\omega) d\omega \\ &\quad + \frac{1}{\lambda_2} \int_{-\pi}^{\pi} |G_2(e^{i\omega}) G_1(e^{i\omega}) - \hat{G}_3(e^{i\omega})|^2 \Phi_u(\omega) d\omega \end{aligned}$$

where $\Phi_u(\omega)$ is the input spectral density. This approach is called the asymptotic ML approach in Wahlberg [1989], and is closely related to PEM methods. It can also be generalized

to cover the colored noise case. For a white input signal the spectral density $\Phi_u(\omega)$ is a constant and it is easy to verify that then

$$\bar{V}(G_1, G_2) = V(b_1, b_2)$$

for the FIR example. Hence, the asymptotic ML estimate and the indirect PEM estimate will for the white input case coincide.

The idea of using model reduction to find structured models, which be discussed below, has been motivated by the discussion of applications of subspace methods in process industry given in Wahlberg et al. [2007]. Assume now that we have estimated a SIMO state-space model, using e.g. a subspace approach, resulting in an unstructured state-space model estimate

$$\begin{aligned} x(t+1) &= \hat{A}x(t) + \hat{B}u(t) \\ y_1(t) &= \hat{C}^1 x(t) \\ y_2(t) &= \hat{C}^2 x(t) \end{aligned}$$

We will assume that the true system and the corresponding estimate are both stable. For an exact estimate it would then be possible to use a state-space vector transformation to the transform estimated model to a cascade state-space representation

$$\begin{aligned} x_1(t+1) &= A_1 x_1(t) + B_1 u(t) \\ x_2(t+1) &= B_2 C_1 x_1(t) + A_2 x_2(t) \\ y_1(t) &= C_1 x_1(t) \\ y_2(t) &= C_2 x_2(t) \end{aligned}$$

However, due to model mismatch this would not be possible for an estimated model.

The following approximation approach is then possible:

Step 1 Find a reduced order model approximation of

$$\begin{aligned} x(t+1) &= \hat{A}x(t) + \hat{B}u(t) \\ y_1(t) &= \hat{C}^1 x(t) \end{aligned}$$

Since the corresponding transfer function is an estimate of G_1 it should be possible to find a good n_1 -order model with reduced number of states using e.g. a balanced model reduction approach Moore [1981]. The idea is to transform the system to a balanced realization using state variable transformation

$$\begin{pmatrix} x_1(t) \\ \bar{x}_2(t) \end{pmatrix} = T_b x(t)$$

We will not give any details on balanced model reduction and how to find e.g. T_b , but this is in principle straightforward solving linear Lyapunov equations. This will give the approximation

$$\begin{aligned} x_1(t+1) &= \hat{A}_1 x_1(t) + \hat{B}_1 u(t) \\ \bar{x}_2(t+1) &= \bar{A}_{21} x_1(t) + \hat{A}_2 \bar{x}_2(t) + \bar{B}_2 u(t) \\ y_1(t) &= \hat{C}_1 x_1(t) \\ y_2(t) &= \bar{C}_{21} x_1(t) + \hat{C}_2 \bar{x}_2(t) \end{aligned}$$

where

$$\begin{aligned} T_b \hat{A} T_b^{-1} &= \begin{pmatrix} \hat{A}_1 & \bar{A}_{12} \\ \bar{A}_{21} & \hat{A}_2 \end{pmatrix} \\ T_b \hat{B} &= \begin{pmatrix} \hat{B}_1 \\ \bar{B}_2 \end{pmatrix}, \quad \begin{pmatrix} \hat{C}_1 \\ \hat{C}_2 \end{pmatrix} T_b^{-1} = \begin{pmatrix} \hat{C}_1 & \bar{C}_{12} \\ \bar{C}_{21} & \hat{C}_2 \end{pmatrix} \end{aligned}$$

The balanced realization G_1 -approximation has been obtained by neglecting the contributions from $\bar{A}_{12} \bar{x}_2(t)$ in the update

equation of $x_1(t+1)$ and removing $\bar{C}_{12}\bar{x}_2(t)$ from the $y_1(t)$ -equation.

Step 2: The difficulty is now how to find $x_2(t)$ to give a cascade realization. Let us try to use the state-space transformation

$$x_2(t) = Tx_1(t) + \bar{x}_2(t)$$

Then

$$\begin{aligned} x_2(t+1) &= (T\hat{A}_1 + \bar{A}_{21})x_1(t) \\ &\quad + \hat{A}_2(x_2(t) - Tx_1(t)) + (T\hat{B}_1 + \bar{B}_2)u(t) \\ &= (T\hat{A}_1 + \bar{A}_{21} - \hat{A}_2T)x_1(t) \\ &\quad + \hat{A}_2x_2(t) + (T\hat{B}_1 + \bar{B}_2)u(t) \\ y_2(t) &= (\bar{C}_{21} - \hat{C}_2T)x_1(t) + \hat{C}_2x_2(t) \end{aligned}$$

Hence, if we could find T and \hat{B}_2 , such that

$$\begin{aligned} (T\hat{A}_1 + \bar{A}_{21} - \hat{A}_2T) &= \hat{B}_2\hat{C}_1 \\ T\hat{B}_1 + \bar{B}_2 &= 0 \\ \bar{C}_{21} - \hat{C}_2T &= 0 \end{aligned} \quad (1)$$

the the matrices $(\hat{A}_2, \hat{B}_2, \hat{C}_2)$ would give a state-space realization of G_2 . This relations form a set of linear equations in the elements of T and \hat{B}_2 . The dimension of the matrix T is $n_2 \times n_1$, where n_1 is the order of the G_1 approximation, n_2 is the order of G_2 and $n = n_1 + n_2$ is the dimension of the original system. We also have n_2 free parameters in \hat{B}_2 . The number of equations in (1) is $n_1 + n_2 + n_1n_2$, which should be compared with the number of free parameters $n_2 + n_1n_2$. This means that we in general due to uncertainty cannot expect to exactly solve these equations, and have instead to confine with e.g. a least squares solution. Notice that not all SIMO system can be transformed to a cascade realization! This is the reason why we have more equations than parameters.

Another approach closely related to cascade control is to find a high gain controller for the inner loop, e.g.

$$u(t) = -K(y_1(t) - r_2(t))$$

such that the closed loop system is stable and

$$\frac{K\hat{G}_1}{1 + K\hat{G}_1} \approx 1$$

Here $r_2(t)$ is the reference signal to the second system. The interpretation of the approximation is that the dynamics of closed loop inner system should be much faster than G_2 . Next do model reduction of

$$\begin{aligned} x(t+1) &= (\hat{A} - K\hat{B}\hat{C}^1)x(t) + K\hat{B}r_2(t) \\ y_2(t) &= \hat{C}^2x(t) \end{aligned}$$

to find a reduced order state-space model of G_2 . This is closely related to using the inverse

$$\frac{K}{1 + K\hat{G}_1} \approx \frac{1}{\hat{G}_1}$$

and performing model reduction of

$$\frac{K}{1 + K\hat{G}_1} \hat{G}_3$$

to find G_2 . Here \hat{G}_3 is an estimate of G_2G_1 obtained from the data $\{u(t), y_2(t)\}$. An advantage of using this inverse is that stability of the transfer functions is preserved. See Markusson [2002] for more ideas of using feedback inversion in system identification.

6. CONCLUSION

The objective of this contribution has been to discuss some important issues in identification of cascade systems. We have used a simple analytic FIR example to explain the fundamental quality problems when the two transfer functions of the subsystems to be estimated are almost identical.

Cascade systems correspond to SIMO system identification. The problem of identification of MISO has recently been thoroughly investigated by Gevers et al. [2006] and Mårtensson [2007]. Many of these results can be modified to the SIMO case. It would be interesting to further study this connection.

We have also shown how to use model reduction as a tool to find models with a cascade structure from unstructured estimates. The proposed method is quite ad hoc and the problem of cascade structured model reduction is more or less open. It would for example be interesting to find error bounds on the approximation error.

An important remaining problem is input design for identification of cascade systems. Notice that the input signal to the second subsystem is

$$u_2(q) = G_1(q)u(t)$$

which consequently is directly colored by the the first unknown subsystem. Hence, a good input for identification of G_1 may give a bad input to G_2 and vice versa.

REFERENCES

- M. Gevers, L. Miskovic, D. Bonvin, and A. Karimi. Identification of multi-input systems: variance analysis and input design issues. *Automatica*, 42(4):559–572, March 2006.
- L. Ljung. *System Identification - Theory For the User*, 2nd ed. Prentice Hall, Upper Saddle River, New Jersey, 1999.
- O. Markusson. *Model and System Inversion with Applications in Nonlinear System Identification and Control*. PhD thesis, KTH, January 2002.
- J. Mårtensson. *Geometric analysis of stochastic model errors in system identification*. PhD thesis, KTH, October 2007. TRITA-EE 2007:061.
- J. Mårtensson and H. Hjalmarsson. A geometric approach to variance analysis in system identification: Theory and nonlinear systems. In *46th IEEE Conference on Decision and Control*, New Orleans, USA, December 2007a.
- J. Mårtensson and H. Hjalmarsson. A geometric approach to variance analysis in system identification: Linear time-invariant systems. In *46th IEEE Conference on Decision and Control*, New Orleans, USA, December 2007b.
- B. Moore. Principal component analysis in linear systems: Controllability, observability and model reduction. *IEEE Transactions on Automatic Control*, 20(1):17–31, September 1981.
- T. Söderström. Errors-in-variables methods in system identification. In *Proceedings of the 14th IFAC Symposium on System Identification*, Newcastle, Australia, 2006.
- T. Söderström, P. Stoica, and B. Friedlander. An indirect prediction error method for system identification. *Automatica*, 27(1):183–188, 1991.
- B. Wahlberg. Model reduction of high-order estimated models: The asymptotic ML approach. *International Journal of Control*, January 1989.
- B. Wahlberg, M. Jansson, T. Matsko, and M. Molander. *Experiences from Subspace System Identification - Comments from Process Industry Users and Researchers*, pages 315–327. Springer, September 2007. ISBN 9783540735694.