Variance Results for Identification of Cascade Systems [★]

Bo Wahlberg ^aHåkan Hjalmarsson ^aJonas Mårtensson ^a

^aAutomatic Control Lab and ACCESS, School of Electrical Engineering, KTH, SE-100 44 Stockholm, Sweden.

Abstract

The objective of this contribution is to analyze statistical properties of estimated models of cascade systems. Models of such systems are important in for example cascade control applications. The aim is to present and analyze some fundamental limitations in the quality of an identified model of a cascade system under the condition that the true subsystems have certain common dynamics. The model quality is analyzed by studying the asymptotic (large data) covariance matrix of the Prediction Error Method parameter estimate. The analysis will focus on cascade systems with three subsystems. The main result is that if the true transfer functions of the first and second subsystem are identical, the output signal information from the second and third subsystems will not affect the asymptotic variance of the estimated model of the first subsystem. This result implies that for a cascade system with two subsystems, where the dynamics of the first subsystem is a factor of the dynamics of the second one, the output signal information from the second subsystem will not improve the asymptotic quality of the estimate of the first subsystem. The results are illustrated by some simple FIR examples.

Key words: System identification, cascade systems, variance analysis.

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 sensor and actuator. 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 problem setting in this paper has been motivated by the discussion of industrial applications of structural system identification given in [11]. The results

to be presented have been inspired by a recent geometric approach to variance analysis in system identification developed in [7,6,5]. This is a very powerful framework for predicting variance results, while the aim here is to use matrix linear algebra to give more direct proofs and extended results.

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 with three subsystems.

The corresponding transfer function description is

$$\begin{aligned} y_1(t) &= G_1(q)u(t) + e_1(t) \\ y_2(t) &= G_2(q)G_1(q)u(t) + e_2(t) \\ y_3(t) &= G_3(q)G_2(q)G_1(q)u(t) + e_3(t). \end{aligned}$$

We will restrict our attention to cascade systems with three output signals. This is enough to cover most of the interesting cases. The input signal is denoted by $u(t)$ and the three output signals are $y_1(t)$, $y_2(t)$ and $y_3(t)$, respectively. The

[★] Parts of this paper has been presented at the 17th IFAC World Congress. This work was partially supported by The Swedish Research Council and the Linnaeus Center ACCESS at KTH. Corresponding author Bo Wahlberg.

Email addresses: bo.wahlberg@ee.kth.se (Bo Wahlberg), hakan.hjalmarsson@ee.kth.se (Håkan Hjalmarsson), jonas.martensson@ee.kth.se (Jonas Mårtensson).

transfer functions $G_1(q)$, $G_2(q)$ and $G_3(q)$ are assumed to be stable. Here q is the shift operator, $q^{-1}u(t) = u(t-1)$ using normalized sampling time. The inputs to $G_2(q)$ and $G_3(q)$ are denoted by $u_2(t)$ and $u_3(t)$, respectively. The signals $e_1(t)$, $e_2(t)$ and $e_3(t)$ are the measurement noise processes. We assume that the dimensions of the input and the three output signals are one (the scalar case).

Cascade systems are very common in both process control and in control of servo mechanical systems. In process cascade control applications, the primary output $y_3(t)$ is often a quality variable such as a temperature or a level, while the secondary outputs $y_1(t)$ and $y_2(t)$ typically concern intermediate variables such as flows or pressures. In mechanical applications $y_1(t)$ and $y_2(t)$ are often rates while $y_3(t)$ is a position. The qualities of the sensors for measuring the outputs can be quite different. We will model this by the size of the variances of the measurement noise processes $\{e_1(t)\}$, $\{e_2(t)\}$ and $\{e_3(t)\}$. A high variance means a poor measurement quality.

Another important example of a cascade system is when $G_3(q)$ is the transfer function of an extra sensor used to measure $y_2(t)$. 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 of the form $\{u(t), y_1(t), y_2(t), y_3(t)\}$ obtained from a cascade system. 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. An asymptotically statistically optimal approach is to apply a Prediction Error Method (PEM), [4], to a constrained model structure that only allows for models of cascade form. We will in this contribution, in detail, analyze the variance of structured PEM model estimates. Because of the products $G_2(q)G_1(q)$ and $G_3(q)G_2(q)G_1(q)$ simple linear in the parameters model structures such as FIR or ARX models are not directly applicable. An alternative approach to find cascade structured estimates is to apply an Indirect Prediction Error Method, see [9], to unstructured FIR/ARX estimates. This method is further discussed in [10].

The outline of this paper is as follows. In Section 2, we will use a simple FIR example to analytically illustrate the basic variance properties of an identified model of a cascade system. In Section 3, this analysis is extended to and formalized for general models of three cascaded systems, which is enough to cover most of the important results. By letting the variance of the output from the second subsystem in a three subsystems structure tend to infinity, we remove the second output and obtain a two cascaded subsystems structure. This idea is used in Section 4 to obtain a more complete understanding of the variance properties when identifying a cascade system with two subsystems with certain common dynamics. The paper is finally concluded in Section 5.

2 Variance Analysis: Introductory FIR Example

To illustrate the statistical properties of PEM cascade model estimates, will study a simple example for a cascade systems with one input signal and two output signals. Consider the model structure

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

with two first order FIR transfer functions

$$\begin{aligned} G_1(q, \theta_1) &= 1 + b_1q^{-1}, \quad \theta_1 = b_1 \\ G_2(q, \theta_2) &= 1 + b_2q^{-1}, \quad \theta_2 = b_2. \end{aligned}$$

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

Let the input signal $u(t)$ be white noise with variance 1. Given a data set $\{u(t), y_1(t), y_2(t)\}$, $t = 1 \dots N$, the PEM estimates of the model parameters b_1 and b_2 are given by

$$\begin{aligned} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} &= \arg \min_{b_1, b_2} \left(\frac{1}{N} \sum_{t=1}^N \frac{[y_1(t) - u(t) - b_1u(t-1)]^2}{\lambda_1} \right. \\ &\quad \left. + \frac{1}{N} \sum_{t=1}^N \frac{[y_2(t) - u(t) - (b_1 + b_2)u(t-1) - b_1b_2u(t-2)]^2}{\lambda_2} \right). \end{aligned}$$

The asymptotic (large N) covariance matrix of the PEM estimates of b_1 and b_2 , which also in this case corresponds to the Cramér-Rao lower bound, is given by M^{-1} , where $M = NE\{\psi(t)\psi^T(t)\}$ and

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

This implies that

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

$$\text{Var} \hat{b}_2 \sim \frac{1}{N} \left(\frac{\frac{\lambda_1(1 + (b_2^o)^2)}{1 + (b_1^o)^2} + \frac{\lambda_2}{1 + (b_1^o)^2}}{1 + \frac{\lambda_1(b_1^o - b_2^o)^2}{\lambda_2(1 + (b_1^o)^2)}} \right). \quad (2)$$

We have used the notation $\sim$ to stress the asymptotic (large data records) relation. These two variance expressions reveal 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 PEM estimate equals

$$\text{Var} \hat{b}_1 \sim \frac{\lambda_1}{N}. \quad (3)$$

Recall that the input is white noise with variance 1. This corresponds to setting $\lambda_2 = \infty$ in the general expression (1), i.e. hiding the information in $y_2(t)$ in noise.

- If the quality of the first output $y_1(t)$ is much worse than for the second one $y_2(t)$, i.e. $\lambda_1 \gg \lambda_2$, we have (by letting $\lambda_1 \rightarrow \infty$)

$$\text{Var} \hat{b}_1 \sim \frac{\lambda_2}{N} \frac{(1 + (b_1^o)^2)}{(b_1^o - b_2^o)^2}, \quad (4)$$

which will be large if b_1^o and b_2^o are close.

- If the two true transfer functions are **identical**, i.e. $b_2^o = b_1^o$, Expression (1) shows that no improvement is obtained from the second output $y_2(t)$, and the variance of the parameter estimate reduces to

$$\text{Var} \hat{b}_1 \sim \frac{\lambda_1}{N} \quad (5)$$

which is exactly the same as when measuring only $y_1(t)$, c.f. (4). This is a special case of a recent result of [6], and, as will be shown in the next section, also holds for more general model structures.

- If $b_2^o = b_1^o$, then

$$\text{Var} \hat{b}_2 \sim \frac{1}{N} \left(\lambda_1 + \frac{\lambda_2}{1 + (b_1^o)^2} \right). \quad (6)$$

If $\lambda_1 = 0$, we exactly know the input $u_2(t)$ to the second subsystem. The variance of $u_2(t)$ is $1 + (b_1^o)^2$, which implies that the asymptotic variance of the estimate of b_2 based on $\{u_2(t), y_2(t)\}$ equals

$$\frac{1}{N} \frac{\lambda_2}{1 + (b_1^o)^2}. \quad (7)$$

If $\lambda_1 > 0$, Expression (6) shows that the variance of $\hat{b}_2$ will be the sum of the variance of $\hat{b}_1$, i.e. λ_1/N , and (7). Hence, if $b_2^o = b_1^o$ the asymptotical statistical quality of $\hat{b}_2$ is always worse than the quality of $\hat{b}_1$.

It is possible to estimate a second order FIR model

$$\begin{aligned} \bar{G}_2(q, \bar{\theta}_2) &= G_2(q, \theta_2) G_1(q, \theta_1) = (1 + b_2 q^{-1})(1 + b_1 q^{-1}) \\ &= 1 + \bar{b}_1 q^{-1} + \bar{b}_2 q^{-2}, \quad \bar{\theta}_2 = (\bar{b}_1 \ \bar{b}_2)^T, \end{aligned}$$

where $\bar{b}_1 = b_1 + b_2$ and $\bar{b}_2 = b_1 b_2$, from only $\{u(t), y_2(t)\}$ without any fundamental problems. It is, however, impossible to decide from only $y_2(t)$ which one of the two factors of this polynomial corresponds to $G_1(q, \theta_1)$ or $G_2(q, \theta_2)$. Furthermore, a first order perturbation analysis reveals that

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

This mapping is not invertible for $b_1 = b_2$. A small perturbation in coefficients $\bar{b}_1$ and $\bar{b}_2$ can then give a large perturbation in the b_1 and b_2 . This is a well known result in e.g. numerical analysis. See [8] for more recent results on the effect of coefficient perturbation for system functions having repeated poles and zeros. Notice that the FIR transfer function $\bar{G}_2$ also could have complex conjugate roots, while the transfer functions of the cascade system only can have real roots. The real double root case, $b_2 = b_1$, is on the border between real and complex roots (the breakaway point in a root locus).

The asymptotic covariance matrix of the estimates of $\bar{b}_1$ and $\bar{b}_2$ from measurement of only $\{u(t), y_2(t)\}$ 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 result 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, if only $y_2(t)$ is measured we again obtain variance result (4) for $\hat{b}_1$.

If it is known in advance that $G_2 = G_1$, one should use the apriori information to further constrain model structure, and in this example only estimate the parameter b_1 , since $b_2 = b_1$. The asymptotic variance of this constrained PEM estimate will be

$$\text{Var} \hat{b}_1 \sim \frac{1}{N} \frac{\lambda_1}{1 + \frac{\lambda_1(4 + 4(b_1^o)^2)}{\lambda_2}}, \quad (8)$$

which could be considerably lower than (5). Hence, using the information that $G_2 = G_1$ substantially improves the quality of the estimated model.

3 Variance Analysis: Three Subsystems

Consider the three subsystem cascade system in Fig. 1, and the corresponding model structure

$$\begin{aligned} y_1(t) &= G_1(q, \theta_1)u(t) + e_1(t) \\ y_2(t) &= G_2(q, \theta_2)G_1(q, \theta_1)u(t) + e_2(t) \\ y_3(t) &= G_3(q, \theta_3)G_2(q, \theta_2)G_1(q, \theta_1)u(t) + e_3(t), \end{aligned}$$

where $e_1(t)$, $e_2(t)$ and $e_3(t)$ are independent Gaussian white noises with known variances λ_1 , λ_2 and λ_3 , respectively. The three transfer functions are assumed to have independent parameterizations. The notation θ_1^o , θ_2^o and θ_3^o will be used for the parameters of the true underlying system to be identified. We will assume that the true system belongs to the model set. The analysis can be modified to also handle estimation of the noise variances. This will, however, not affect the variance results of the transfer function parameters. It also extends to multiple cascaded systems with more than three blocks.

Given a data set $\{u(t), y_1(t), y_2(t), y_3(t)\}$, $t = 1 \dots N$, the PEM estimates of the model parameters θ_1 , θ_2 and θ_3 are given by

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} = \arg \min_{\theta_1, \theta_2, \theta_3} \left(\frac{1}{N} \sum_{t=1}^N \frac{[y_1(t) - G_1(q, \theta_1)u(t)]^2}{\lambda_1} + \frac{1}{N} \sum_{t=1}^N \frac{[y_2(t) - G_2(q, \theta_2)G_1(q, \theta_1)u(t)]^2}{\lambda_2} + \frac{1}{N} \sum_{t=1}^N \frac{[y_3(t) - G_3(q, \theta_3)G_2(q, \theta_2)G_1(q, \theta_1)u(t)]^2}{\lambda_3} \right).$$

Define the 3×3 block matrix

$$\psi(t) = \begin{pmatrix} \frac{G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_1}} & \frac{G_2(q, \theta_2^o)G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} & \frac{G_3(q, \theta_3^o)G_2(q, \theta_2^o)G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_3}} \\ 0 & \frac{G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} & \frac{G_3(q, \theta_3^o)G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_3}} \\ 0 & 0 & \frac{G_3'(q, \theta_3^o)G_2(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_3}} \end{pmatrix},$$

where prime denotes differentiation with respect to the parameter vectors. The size of this block-partitioned matrix is $(n_1 + n_2 + n_3) \times 3$, where n_i denotes the number of parameters in θ_i (for $i = 1, 2, 3$). The asymptotic covariance matrix

of the PEM parameter estimates is then given by

$$\text{Cov} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} \sim M^{-1},$$

where $M = NE\{\psi(t)\psi^T(t)\}$. See *e.g.* [4] for details.

Consider the case when the following condition holds

Condition 1: $G_2(q, \theta_2^o)G_1'(q, \theta_1^o) = G_2'(q, \theta_2^o)G_1(q, \theta_1^o)$

Condition 1 holds if the true first transfer function $G_1(q, \theta_1^o)$ and second one $G_2(q, \theta_2^o)$ are **identical** and we have the same structure for these two submodels, while the third transfer function $G_3(q, \theta_3^o)$ could be different. Notice that we thus assume that the number of parameters in θ_1 and θ_2 are equal, i.e. $n_1 = n_2$.

The aim here is to generalize the key observation from two cascade subsystems FIR example in the previous section, i.e. under Condition 1 the asymptotic quality of the estimate of θ_1 should not be improved by measuring $y_2(t)$ and/or $y_3(t)$ (which just is a filtered version of the noise free version of $y_2(t)$ plus independent noise).

The matrix $\psi(t)$ now simplifies to

$$\psi(t) = \begin{pmatrix} \frac{G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_1}} & \frac{G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} & \frac{G_3(q, \theta_3^o)G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_3}} \\ 0 & \frac{G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} & \frac{G_3(q, \theta_3^o)G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_3}} \\ 0 & 0 & \frac{G_3'(q, \theta_3^o)G_2(q, \theta_2^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_3}} \end{pmatrix}.$$

Define the corresponding covariance matrices

$$\begin{aligned}
A &= \frac{N}{\lambda_1} \mathbb{E} \{ [G'_1(q, \theta_1^o) u(t)] [G'_1(q, \theta_1^o) u(t)]^T \}, \\
D &= \frac{N}{\lambda_2} \mathbb{E} \{ [G'_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)] \\
&\quad \times [G'_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)]^T \}, \\
F &= \frac{N}{\lambda_3} \mathbb{E} \{ [G'_3(q, \theta_3^o) G'_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)] \\
&\quad \times [G'_3(q, \theta_3^o) G'_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)]^T \}, \\
H &= \frac{N}{\lambda_3} \mathbb{E} \{ [G'_3(q, \theta_3^o) G'_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)] \\
&\quad \times [G'_3(q, \theta_3^o) G_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)]^T \}, \\
L &= \frac{N}{\lambda_3} \mathbb{E} \{ [G'_3(q, \theta_3^o) G_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)] \\
&\quad \times [G'_3(q, \theta_3^o) G_2(q, \theta_2^o) G_1(q, \theta_1^o) u(t)]^T \}. \quad (9)
\end{aligned}$$

to give

$$M = N \mathbb{E} \{ \psi(t) \psi^T(t) \} = \begin{pmatrix} A + D + F & D + F & H \\ D + F & D + F & H \\ H^T & H^T & L \end{pmatrix}. \quad (10)$$

Next we need to invert M to find the asymptotic covariance matrix of the PEM parameter estimate. Notice that block rows one and two of M are almost identical, and the same holds for block columns one and two. This observation can be used to block diagonalize M using the transformation matrix

$$T = \begin{pmatrix} I & -I & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix},$$

where I denotes an identity matrix. We then have

$$\tilde{M} = TMT^T = \begin{pmatrix} A & 0 & 0 \\ 0 & D + F & H \\ 0 & H^T & L \end{pmatrix}.$$

Introduce

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix} = \begin{pmatrix} D + F & H \\ H^T & L \end{pmatrix}^{-1}.$$

Now $M^{-1} = T^T \tilde{M}^{-1} T$, which gives

$$M^{-1} = \begin{pmatrix} A^{-1} & -A^{-1} & 0 \\ -A^{-1} & A^{-1} + S_{11} & S_{12} \\ 0 & S_{12}^T & S_{22} \end{pmatrix}. \quad (11)$$

This means that

$$\begin{aligned}
\text{Cov } \hat{\theta}_1 &\sim A^{-1} \\
\text{Cov } \hat{\theta}_2 &\sim A^{-1} + S_{11} \\
\text{Cov } \hat{\theta}_3 &\sim S_{22}.
\end{aligned}$$

Hence, the **key results under Condition 1** are:

- Since the matrix A^{-1} is the asymptotic covariance matrix of $\hat{\theta}_1$ when only $y_1(t)$ is available, the asymptotic quality of the estimate of θ_1 is not improved by also measuring $y_2(t)$ and/or $y_3(t)$.
- The covariance matrix of the estimate of $(\theta_2^T \theta_3^T)^T$ equals

$$\begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} D + F & H \\ H^T & L \end{pmatrix}^{-1}.$$

The second term in this expression is nothing else but the asymptotic covariance matrix of the estimate of $(\theta_2^T \theta_3^T)^T$ when $G_1(q)$ is known (or equally $\lambda_1 = 0$). Hence, for $\lambda_1 > 0$ the quality of the estimate of θ_2 is always worse than of the estimate of θ_1 . However, the quality of the estimate of θ_3 is independent of $y_1(t)$.

- By letting λ_3 tend to infinity, we remove the information from output $y_3(t)$ and obtain the variance expression for the two cascaded subsystems case. This extends the FIR model variance result presented in the example in Section 2 to general transfer function models $G_1(q, \theta_1)$ and $G_2(q, \theta_2)$. Under Condition 1 we here obtain

$$\begin{aligned}
\text{Cov } \hat{\theta}_1 &\sim A^{-1} \\
\text{Cov } \hat{\theta}_2 &\sim A^{-1} + D^{-1},
\end{aligned}$$

which correspond to (1) and (2) for the FIR models example.

- If it is known in advance that $G_2 = G_1$ ($\theta_2 = \theta_1$), this should be incorporated in the model structure. Let us study this for two cascaded subsystems

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

The asymptotic covariance matrix M^{-1} of $\hat{\theta}_1$ can then be calculated using

$$\psi(t) = \begin{pmatrix} \frac{G'_1(q, \theta_1^o) u(t)}{\sqrt{\lambda_1}} & \frac{2G'_1(q, \theta_1^o) G_1(q, \theta_1^o) u(t)}{\sqrt{\lambda_2}} \end{pmatrix}$$

in $M = N \mathbb{E} \{ \psi(t) \psi^T(t) \}$. This will result in $M = A + 4D$ and consequently

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

is the generalization of the FIR result (8). Since $(A + 4D)^{-1} < A^{-1}$ for positive definite matrices A and D , the variance can be considerably smaller than for the case with separate parameterizations (for which the corresponding covariance matrix equals A^{-1}).

We can further simplify the structural results by imposing

Condition 2:

$$\begin{aligned} G_2(q, \theta_2^o) G_1'(q, \theta_1^o) &= G_2'(q, \theta_2^o) G_1(q, \theta_1^o) \\ G_3(q, \theta_3^o) G_2'(q, \theta_2^o) &= G_3'(q, \theta_3^o) G_2(q, \theta_2^o) \end{aligned}$$

This is the case if we use the same model structures for all three submodels and the true subsystems are **identical**. Now $F = H = L$ in (9) and

$$S = \begin{pmatrix} D^{-1} & -D^{-1} \\ -D^{-1} & D^{-1} + L^{-1} \end{pmatrix}$$

in Expression (11) and thus

$$\begin{aligned} \text{Cov } \hat{\theta}_1 &\sim A^{-1} \\ \text{Cov } \hat{\theta}_2 &\sim A^{-1} + D^{-1} \\ \text{Cov } \hat{\theta}_3 &\sim D^{-1} + L^{-1}. \end{aligned}$$

We obtain exactly the same variance result $A^{-1} + D^{-1}$ of $\hat{\theta}_2$ as for the two subsystems case. Furthermore, in this case the quality of $\hat{\theta}_2$ is independent of the output $y_3(t)$.

A remaining question is if it is possible to obtain similar results when

$$G_3(q, \theta_3^o) G_2'(q, \theta_2^o) = G_3'(q, \theta_3^o) G_2(q, \theta_2^o) \quad (12)$$

i.e. the true second and third transfer functions are equal while the first one, $G_1(q, \theta_1^o)$, can be different. It turns out that the answer is no. This can be shown by calculating the corresponding matrix M under Condition (12). The intuition is as follows: The quality of $\hat{\theta}_1$ will here depend on all three output signals, but will also affect the estimates $\hat{\theta}_2$ and $\hat{\theta}_3$. This will, for example, create a dependence between the output $y_3(t)$ and the quality of $\hat{\theta}_2$.

The last remaining case is when

$$G_3(q, \theta_3^o) G_1'(q, \theta_1^o) = G_3'(q, \theta_3^o) G_1(q, \theta_1^o), \quad (13)$$

while $G_2(q, \theta_2^o)$ could be an arbitrary transfer function. This is the case when the true first and the third transfer function are identical. The answer here is that no general structural variance result holds. This can be verified by the following simple example.

Example: Let the input signal $u(t)$ be white noise with variance 1, and

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

Assume $G_3(q, \theta_3^o) = G_1(q, \theta_1^o) \neq G_2(q, \theta_2^o)$, and set $b_2^o = 0$ and thus $b_3^o = b_1^o \neq 0$ to further simplify the problem. For e.g. $b_3^o = b_1^o = 1$, direct calculations given in [12] give

$$\text{Cov } (\hat{b}_1) \sim \frac{1}{N} \frac{4/\lambda_2 + 3/\lambda_3}{4/(\lambda_1 \lambda_2) + 3/(\lambda_1 \lambda_3) + 2/\lambda_2^2 + 3/(\lambda_2 \lambda_3)},$$

which clearly is a function of the quality of all three output signals. ■

The example shows that the ordering of the systems is very important for the structural variance results to hold.

4 Variance Analysis: Two Cascaded Subsystems with Some Common Dynamics

The results for three cascaded subsystems in the previous section can be used to derive more advanced results for two cascaded subsystems as in Fig. 2. By letting the variance of $\{e_2(t)\}$, i.e. λ_2 , tend to infinity we remove the information from the second output signal, and the problem reduces to a two subsystems setup. If $y_2(t)$ is removed we have the model structure

$$\begin{aligned} y_1(t) &= G_1(q, \theta_1)u(t) + e_1(t) \\ y_3(t) &= G_3(q, \theta_3)G_2(q, \theta_2)G_1(q, \theta_1)u(t) + e_3(t), \end{aligned} \quad (14)$$

and the corresponding block diagram given in Fig. 2.

Fig. 2. Cascade system with two subsystems with possible common dynamics.

The PEM estimates are given by

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} = \arg \min_{\theta_1, \theta_2, \theta_3} \left(\frac{1}{N} \sum_{t=1}^N \frac{[y_1(t) - G_1(q, \theta_1)u(t)]^2}{\lambda_1} + \frac{1}{N} \sum_{t=1}^N \frac{[y_3(t) - G_3(q, \theta_3)G_2(q, \theta_2)G_1(q, \theta_1)u(t)]^2}{\lambda_3} \right).$$

Consider again

Condition 1: $G_2(q, \theta_2^o)G_1'(q, \theta_1^o) = G_2'(q, \theta_2^o)G_1(q, \theta_1^o)$

but for the mode structure (14). When $y_2(t)$ is removed this condition holds if the transfer function G_1 is also a factor of the transfer function of the second subsystem, i.e. the two subsystems have common dynamics and a suitable parametrization is used.

We do not have to redo all the calculations since we know that the result for this estimation problem can be obtained by setting $\lambda_2 = \infty$ in the three subsystem variance expression results. Now

$$D = \frac{N}{\lambda_2} E\{ [G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)] \times [G_2'(q, \theta_2^o)G_1(q, \theta_1^o)u(t)]^T \} \rightarrow 0, \quad \lambda_2 \rightarrow \infty,$$

while A, C, D and F in Expression (10) for M do not depend on λ_2 , and are thus unaffected. This means that

$$M^{-1} = \begin{pmatrix} A^{-1} & -A^{-1} & 0 \\ -A^{-1} & A^{-1} + R_{11} & R_{12} \\ 0 & R_{12}^T & R_{22} \end{pmatrix},$$

where

$$\begin{pmatrix} R_{11} & R_{12} \\ R_{12}^T & R_{22} \end{pmatrix} = \begin{pmatrix} F & H \\ H^T & L \end{pmatrix}^{-1}.$$

We have derived the following variance expressions for the parameter estimates corresponding to the model structure (14):

$$\begin{aligned} \text{Cov } \hat{\theta}_1 &\sim A^{-1}, \\ \text{Cov } \hat{\theta}_2 &\sim A^{-1} + R_{11}, \\ \text{Cov } \hat{\theta}_3 &\sim R_{22}. \end{aligned}$$

Hence, the **key results under Condition 1** for the model structure (14) are

- The asymptotic quality of the estimate of θ_1 is again independent of the information in the output signal $y_3(t)$.
- The covariance of the estimate of $(\theta_2^T \theta_3^T)^T$ equals

$$\begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} F & H \\ H^T & L \end{pmatrix}^{-1}.$$

The second term in this expression is the asymptotic covariance matrix of $(\theta_2^T \theta_3^T)^T$ when $G_1(q)$ is known (or equally $\lambda_1 = 0$). Hence, the quality of the estimate of θ_2

is for $\lambda_1 > 0$ always worse than the estimate of θ_1 . However, the quality of the estimate of θ_3 is independent of $y_1(t)$.

The result for the case when $y_2(t)$ and $y_3(t)$ are only measurable, i.e. a model structure of the form

$$\begin{aligned} y_2(t) &= G_2(q, \theta_2)G_1(q, \theta_1)u(t) + e_2(t) \\ y_3(t) &= G_3(q, \theta_3)G_2(q, \theta_2)G_1(q, \theta_1)u(t) + e_3(t), \end{aligned}$$

is more involved. It corresponds to setting λ_1 to infinity in the general result. If

$$G_3(q, \theta_3^o)G_2'(q, \theta_1^o) = G_3'(q, \theta_3^o)G_2(q, \theta_1^o),$$

i.e. when the dynamics of the second subsystem is a factor of the first subsystem, we, as noted in previous section, only have structural results if in addition

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

holds, i.e. under Condition 2.

5 Conclusion

The objective of this contribution has been to present some fundamental variance analysis results for identification of cascade systems. The analysis is based on asymptotic (large number of data) properties. Models of cascade systems are special in the sense that they contain products of transfer functions. It is well known that polynomials with double roots are sensitive to certain perturbations in the coefficients. The same problem occurs if there are common dynamics between two connected subsystems in a cascade structure. It will imply that the asymptotic quality of the estimate of the first subsystem only depends on the first output and no extra information is then obtained from the output from the second subsystem. This is very important to know, especially if the measurement quality of the two outputs are very different. A solution is then to constrain the model structure to only allow for models with common dynamics.

It would be interesting to study the identification problem for more complex interconnected block diagram structures, and in particular analysis of quality/variance properties. Many tools for modeling of physical systems are based on such representations. This information should then be incorporated in the model structure used in system identification. Identification techniques for such systems are by no means new. For example, [1] considered systems composed of cascade, feed-forward, feedback and multiplicative connections of linear dynamic and zero memory nonlinear elements, and showed that such systems can be identified in terms of the individual component subsystems from measurements of the system input and output only. This includes Wiener and Hammerstein models as special cases. More recent work on identification of general structured models can be found in [2,3].

Input design for structured systems is another interesting area for future research. Reconsider the FIR example in Section 2 with $b_2^o = b_1^o$ and $|b_1^o| < 1$, but replace the white noise input signal with

$$u(t) = \frac{K_u}{1 + b_1^o q^{-1}} e_u(t), \quad (15)$$

where $e_u(t)$ is white noise with variance 1. We then obtain

$$\begin{aligned} \text{Var } \hat{b}_1 &\sim \frac{1}{N} \frac{\lambda_1 (1 - (b_1^o)^2)}{K_u^2}, \\ \text{Var } \hat{b}_2 &\sim \frac{1}{N} \left(\frac{\lambda_1 (1 - (b_1^o)^2)}{K_u^2} + \frac{\lambda_2}{K_u^2} \right) \end{aligned}$$

By taking $K_u = \sqrt{1 - (b_1^o)^2}$, we obtain the same input variance as for the original example, i.e. $E\{u^2(t)\} = 1$. This choice will, however, lead to a poor estimate of b_2 since

$$\text{Var } \hat{b}_2 \sim \frac{1}{N} \left(\lambda_1 + \frac{\lambda_2}{1 - (b_1^o)^2} \right),$$

which could be considerably larger than expression (6) for the white input signal case. Another option is to constrain the output variance of the first subsystem, i.e. $u_2(t)$ in Fig. 1. For the white input case we have $E\{u_2^2(t)\} = 1 + (b_1^o)^2$, and $K_u = \sqrt{1 + (b_1^o)^2}$ gives the same output variance for the alternative input signal (15). This is a much better choice since

$$\text{Var } \hat{b}_2 \sim \frac{1}{N} \left(\frac{\lambda_1 (1 - (b_1^o)^2)}{1 + (b_1^o)^2} + \frac{\lambda_2}{1 + (b_1^o)^2} \right)$$

which is always lower than the result (6) for the white input signal. Notice that filtered input increases the input to noise signal ratio for the first subsystem, while keeping the same output signal energy levels. The reason is that we excite the system more without increasing the outputs by using the zero dynamics of the transfer function G_1 .

This example shows that there are considerable quality improvements to gain by proper choice of input signals when identifying cascade systems. It would be interesting to further investigate these potentials.

References

- [1] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.
- [2] K. Hsu, T. Vincent, and K. Poolla. A kernel based approach to structured nonlinear system identification part i: Algorithms. In *Proceedings of the 14th IFAC Symposium on System Identification, SYSID 2007*, Newcastle, Australia, 2006.
- [3] K. Hsu, T. Vincent, and K. Poolla. A kernel based approach to structured nonlinear system identification part ii: Convergence and consistency. In *Proceedings of the 14th IFAC Symposium on System Identification, SYSID 2007*, Newcastle, Australia, 2006.

- [4] L. Ljung. *System Identification - Theory For the User*, 2nd ed. Prentice Hall, Upper Saddle River, New Jersey, 1999.
- [5] J. Mårtensson. *Geometric analysis of stochastic model errors in system identification*. PhD thesis, KTH, October 2007. TRITA-EE 2007:061.
- [6] 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 2007.
- [7] 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 2007.
- [8] Anthony G. Place and Gregory H. Allen. Generalized pole sensitivity analysis due to parameter perturbation. *IEEE Transactions on Circuits and Systems – II: Analog and Digital Signal Processing*, 44(10), October 1997.
- [9] T. Söderström, P. Stoica, and B. Friedlander. An indirect prediction error method for system identification. *Automatica*, 27(1):183–188, 1991.
- [10] B. Wahlberg, H. Hjalmarsson, and J. Mårtensson. On identification of cascade systems. In *17th IFAC World Congress*, October 2007.
- [11] 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.
- [12] Bo Wahlberg, Håkan Hjalmarsson, and Jonas Mårtensson. Variance analysis of identification of cascade systems. In *Proceedings IEEE Conference on Decision and Control*, Cancun, Mexico, December 2008.