Accuracy analysis of time domain maximum likelihood method and sample maximum likelihood method for errors-in-variables identification

Mei Hong, Torsten Söderström
Division of Systems and Control,
Department of Information Technology, Uppsala University,
P O Box 337, SE-75105 Uppsala, Sweden

Johan Schoukens and Rik Pintelon
Department ELEC, Vrije Universiteit Brussel
B-1050 Brussels, Belgium

September 4, 2007

Abstract

The time domain maximum likelihood (TML) method and the sample maximum Likelihood (SML) method are two approaches for identifying errors-in-variables models. Both methods may give the optimal estimation accuracy (achieve Cramér-Rao lower bound) but in different senses. In the TML method, an important assumption is that the noise-free input signal is modeled as a stationary process with rational spectrum. For SML, the noise-free input needs to be periodic. It is interesting to know which of these assumptions contain more information to boost the estimation performance. In this paper, the estimation accuracy of the two methods is analyzed statistically. Numerical comparisons between the two estimates are also done under different signal-to-noise ratios (SNRs). The results suggest that TML and SML have similar estimation accuracy at moderate or high SNR.

1 Introduction

The dynamic errors-in-variables (EIV) identification problem has been a topic of active research for several decades. Till now, many solutions have been proposed with different approaches. For example, the Koopmans-Levin (KL) method [3], the Frisch scheme [2], the Bias-Eliminating Least Squares methods [14], [15], the prediction error method [8], frequency domain methods [4], and methods based on higher order moments statistics [13], *etc*. See [10] and references therein for a comprehensive survey in this respect.

In system identification, besides system properties and method performances, experimental conditions also play an important role. For example, periodic input signals will give many interesting advantages in identification. The sample maximum likelihood (SML) method [7] works under the assumption that the noise-free signal is periodic, and it provides optimal estimation accuracy under that assumption. If periodic data are not available, among the possible methods for identifying EIV systems, the time domain maximum likelihood method (TML), also called the joint output approach, [8], will achieve the Cramér-Rao lower bound. This property is conditioned on the prior information that the true input is an ARMA process.

The comparison of the TML and SML methods is of general interest. When the input can freely be chosen it is important to know whether a random (filtered white noise) input or a periodic input will lead to the smallest uncertainty of the estimated plant model parameters. If there is no significant difference then other issues are important such as the ease of generating starting values, the optimization complexity, etc.

In general, the TML method and the SML method work under different experimental situations. An essential assumption for the TML method is that the noise-free input signal is a stationary stochastic process with rational spectrum, so that it can be described as an ARMA process. Also, in the TML method, the input and output noises are usually described as ARMA processes. In contrast, the SML method works under more general noise-free input signals and noise conditions, but with another necessary assumption: the noise-free signal is periodic. Further, for the SML method cross-correlation between the noise sources is allowed.

In this report, we focus on comparing the asymptotic covariance matrix of these two methods. The paper is organized as follows. In Section 2 we describe the EIV problem and introduce notations. The main idea of the TML and SML methods are reviewed in Section 3 and 4. In Section 5, we make a statistical comparison for TML and SML under high SNR cases. Numerical comparisons between the asymptotic covariance matrices of these two methods under different SNR are shown in Section 6. Further, discussions on how to optimally utilize the periodic data are given in Section 7 before we draw conclusions in Section 8.

2 Notations and setup

As a typical model example, consider the linear single-input single-output (SISO) system depicted in Figure 1 with noise-corrupted input and output measurements.

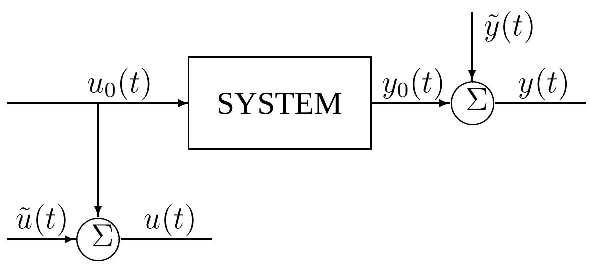

Figure 1: The basic setup for a dynamic errors-in-variables problem.

Let the noise-free input and output processes $u_0(t)$ and $y_0(t)$ be linked by a linear stable, discrete-time, dynamic system

$$A(q^{-1})\, y_o(t) = B(q^{-1})\, u_o(t), \tag{2.1}$$

where

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \ldots + a_n\, q^{-n} \\ B(q^{-1}) &= b_1\, q^{-1} + \ldots + b_n\, q^{-n} \end{aligned} \tag{2.2}$$

are polynomials[1] in the backward shift operator q^{-1}.

For errors-in-variables systems, the input and the output are measured with additive noises:

$$\begin{aligned} u(t) &= u_o(t) + \tilde{u}(t), \\ y(t) &= y_o(t) + \tilde{y}(t). \end{aligned} \tag{2.3}$$

The system has a transfer function

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}. \tag{2.4}$$

The unperturbed input is modeled as an ARMA process driven by white noise

$$u_o(t) = \frac{C(q^{-1})}{D(q^{-1})} v(t). \tag{2.5}$$

The noise $\tilde{u}(t)$ and $\tilde{y}(t)$ are assumed to be mutually independent zero mean white noise sequences both independent of $u_o(t)$ [2].

The noise variances are denoted

$$E\tilde{u}^2(t) = \lambda_u^2, \;\; E\tilde{y}^2(t) = \lambda_y^2, \;\; Ev^2(t) = \lambda_v^2. \tag{2.6}$$

Problem: The task is to consistently estimate the system parameter vector

$$\theta = (a_1 \ldots a_n \quad b_1 \ldots b_n)^T \tag{2.7}$$

from the measured noisy data $\{u(t), y(t)\}_{t=1}^{N}$.

3 Review of the TML method

In the TML method, we consider the EIV system as a multivariable system with both $u(t)$ and $y(t)$ as outputs. An important assumption for this method is that the signal $u_o(t)$ is stationary with rational spectrum, so that $u_o(t)$ can be described as an ARMA process of the type

$$u_o(t) = \frac{C(q^{-1})}{D(q^{-1})}\; e(t)\;, \tag{3.1}$$

where $e(t)$ is a white noise with variance λ_e and the polynomials $C(q^{-1})$, $D(q^{-1})$ are relatively prime and asymptotically stable, with known degrees.

---

[1]It can be generalized to include a b_0 term, or to allow different degrees of A and B.

[2]For SML, $\tilde{u}$ and $\tilde{y}$ might be correlated. TML can also be extended to accommodate cases with rather arbitrarily correlated noises. The assumptions here will simplify the analysis. However, it will be not crucial for the analysis.

In this way the whole errors-in-variables model can be rewritten as a system with a two-dimensional output vector $z(t) = (y(t)\, u(t))^T$ and three mutually uncorrelated white noise sources $e(t)$, $\tilde{u}(t)$ and $\tilde{y}(t)$:

$$z(t) = \begin{pmatrix} y(t) \\ u(t) \end{pmatrix} = \begin{pmatrix} \frac{B(q^{-1})C(q^{-1})}{A(q^{-1})D(q^{-1})} & 0 & 1 \\ \frac{C(q^{-1})}{D(q^{-1})} & 1 & 0 \end{pmatrix} \begin{pmatrix} e(t) \\ \tilde{u}(t) \\ \tilde{y}(t) \end{pmatrix} . \tag{3.2}$$

By transforming the model to a general state space model and then using the well-known techniques [1] to convert it into the innovations form, we will get

$$z(t) = S(q^{-1}, \vartheta)\, \varepsilon(t, \vartheta) \ , \tag{3.3}$$

where $S(q^{-1})$ is a stable transfer function matrix which can be computed from the Riccati equation for Kalman filters, and $\varepsilon(t, \vartheta)$ is the prediction error $\varepsilon(t, \vartheta) = z(t) - \hat{z}(t|t-1; \vartheta)$, which depends on the data and the model matrices. Note that the parameter vector ϑ contains not only the system parameters θ but also the noise parameters and parameters of $u_o(t)$, *i.e.* the coefficients of the polynomials C and D.

The parameter vector ϑ is consistently estimated from a data sequence $z(t)_{t=1}^{N}$ by minimizing the loss function:

$$\hat{\vartheta}_N = \arg\min_{\vartheta} \frac{1}{N} \sum_{t=1}^{N} \ell(\varepsilon(t, \vartheta), \vartheta, t), \tag{3.4}$$

with

$$\ell(\varepsilon(t,\vartheta), \vartheta, t) = \frac{1}{2} \log \det Q(\vartheta) + \frac{1}{2} \varepsilon^T(t, \vartheta) Q^{-1}(\vartheta) \varepsilon(t, \vartheta), \tag{3.5}$$

where $Q(\vartheta)$ denotes the covariance matrix of the prediction errors. For Gaussian distributed data, the covariance matrix of the TML estimates parameters turns out to be asymptotically ($N \to \infty$) equal to the Cramér-Rao bound [9].

4 Review of the SML method

The ML estimate can also be computed in the frequency domain [6]. Let $U(w_k)$ and $Y(w_k)$, with $w_k = 2\pi k/N$, $k = 1, \ldots, N$, denote the discrete Fourier transforms of the input and output measurements, respectively. Write the transfer function as $G(e^{iw_k}) = B(e^{iw_k})/A(e^{iw_k})$ (note that there is no need to assume that A is stable as long as the system has stationary input and output signals, e.g. an unstable plant captured in a stabilizing feedback loop is allowed). The ML criterion in the frequency domain can be written as

$$\begin{aligned} V(\theta) = \quad & \frac{1}{N} \sum_{k=1}^{N} |B(e^{iw_k}, \theta)U(w_k) - A(e^{iw_k}, \theta)Y(w_k)|^2 \\ & \times \{\sigma_U^2(w_k)|B(e^{iw_k}, \theta)|^2 + \sigma_Y^2(w_k)|A(e^{iw_k}, \theta)|^2 \\ & -2\mathrm{Re}\, [\sigma_{YU}(w_k)A(e^{iw_k}, \theta)B(e^{-iw_k}, \theta)]\}^{-1}, \end{aligned} \tag{4.1}$$

where $\sigma_U^2(w_k)$, $\sigma_Y^2(w_k)$ and $\sigma_{YU}(w_k)$ are the variance or covariance of the input and output noise at frequency w_k, respectively. If these (co)variances are known a priori, it is easy to minimize the cost function (4.1) to get good estimates. However, knowing exactly the noise model is not realistic in many practical cases. Then we have to consider the (co)variances of the noises as additional parameters which should also be estimated from the data. In this case, a high dimensional nonlinear optimization problem should be solved, which leads to infeasible situations. Instead of doing so, another way is to replace the exact covariance matrices of the disturbances by sample estimates obtained from a small number (M) of repeated experiments. This is the fundamental idea of the sample ML method. An important assumption is utilized in SML, namely to have periodic excitation signals, where each period plays the role of an independent repeated experiment. The definitions for the sample (co)variances of $\hat{\sigma}_U^2(w_k), \hat{\sigma}_Y^2(w_k)$, and $\hat{\sigma}_{YU}(w_k)$ are:

$$\begin{aligned}
\hat{\sigma}_U^2(w_k) &= \frac{1}{M-1}\sum_{l=1}^{M}|U_l(w_k)-\bar{U}(w_k)|^2,\\
\hat{\sigma}_Y^2(w_k) &= \frac{1}{M-1}\sum_{l=1}^{M}|Y_l(w_k)-\bar{Y}(w_k)|^2,\\
\hat{\sigma}_{YU}(w_k) &= \frac{1}{M-1}\sum_{l=1}^{M}\big(Y_l(w_k)-\bar{Y}(w_k)\big)\big(U_l(w_k)-\bar{U}(w_k)\big)^*;
\end{aligned}$$

where $*$ indicates the complex conjugate and $\bar{U}(w_k)$ and $\bar{Y}(w_k)$ denote the sample mean of input and output which are similarly defined as:

$$\bar{U}(w_k) = \frac{1}{M}\sum_{l=1}^{M}U_l(w_k), \quad \bar{Y}(w_k)=\frac{1}{M}\sum_{l=1}^{M}Y_l(w_k)\ .$$

The parameter vector θ is estimated by minimizing,

$$\begin{aligned}
\bar{V}(\theta) = \quad & \frac{1}{N}\sum_{k=1}^{N}|B(e^{iw_k},\theta)\bar{U}(w_k)-A(e^{iw_k},\theta)\bar{Y}(w_k)|^2\\
&\times\{\hat{\sigma}_{\bar{U}}^2(w_k)|B(e^{iw_k},\theta)|^2+\hat{\sigma}_{\bar{Y}}^2(w_k)|A(e^{iw_k},\theta)|^2\\
&-2\mathrm{Re}\,[\hat{\sigma}_{\bar{Y}\bar{U}}(w_k)A(e^{iw_k},\theta)B(e^{-iw_k},\theta)]\}^{-1}\ ,
\end{aligned}\tag{4.2}$$

where $\hat{\sigma}_{\bar{U}}^2 = \hat{\sigma}_U^2/M$, $\hat{\sigma}_{\bar{Y}}^2 = \hat{\sigma}_Y^2/M$ and $\hat{\sigma}_{\bar{Y}\bar{U}} = \hat{\sigma}_{YU}/M$. The cost function (4.2) is an approximation of (4.1) by replacing the exact covariances of the noise by their sample estimates. The major advantage of this approach is that the plant parameters remain as the only unknowns to be estimated, which leads to a low dimension of the nonlinear optimization problem.

It is clear that this SML estimator is no longer an exact ML estimator. However, it was proved in [7] that the estimator is consistent if the number of experiments $M \geq 4$. For $M \geq 6$ the covariance matrix of the model parameters $\mathrm{cov}(\hat{\theta}_{\mathrm{SML}})$ is related to the covariance matrix $\mathrm{cov}(\hat{\theta}_{\mathrm{ML}})$ of the estimates assuming known noise variances by

$$\mathrm{cov}(\hat{\theta}_{\mathrm{SML}}) = \frac{M-2}{M-3}\mathrm{cov}(\hat{\theta}_{\mathrm{ML}})\left(1+O((M\lambda_v^2)^{-1})\right)\tag{4.3}$$

i.e. for sufficiently large signal-to-noise ratios the loss in efficiency of SML is $(M-2)/(M-3)$[3] which is not a large factor even for small values of M.

[3] In Theorem 1 below, the factor $(M-2)/(M-3)$ is disregarded.

5 Statistical analysis of the accuracy of TML and SML estimates

In this section, we will analyze the normalized asymptotic covariance matrices of SML and TML and try to reveal the behaviors of the two methods for high SNRs. We will express in terms of the noise variances $\lambda_v^2,\ \lambda_u^2,\ \lambda_y^2$ rather than in the SNR values.

From the reviews above, it can be seen that the TML method and the SML method work under different assumptions. We assume here that NM periodic data are available, where M is the number of periods and N denotes the number of data points in each period. Also assume that in each period the noise-free input signal is the same realization of a stationary process. This experimental condition is suitable for both approaches. The TML method uses all data points and the information that the input signal is an ARMA process, but does not exploit the periodicity of the data. However, the SML method uses the periodic information but disregards that the input signal is an ARMA process and does not use any parametric models of the noise terms.

Consider the case when SNRs at both input and output are high. This is achieved by keeping λ_u^2 and λ_y^2 fixed, and letting λ_v^2 tend to infinity. We have the following theorem.

Theorem 1: The normalized asymptotic covariance matrices of the SML and TML estimators have the relation:

$$\lim_{\lambda_v^2\to\infty} \lambda_v^2 \mathrm{cov}(\hat{\theta}_{\mathrm{SML}}) = \lim_{\lambda_v^2\to\infty} \lambda_v^2 \mathrm{cov}(\hat{\theta}_{\mathrm{TML}}) = \lambda_v^2 \mathrm{CRBA}, \tag{5.1}$$

where CRBA is an asymptotic Cramér-Rao lower bound when data number tends to infinity. It holds

$$\mathrm{CRBA} = M_1^{-1}, \tag{5.2}$$

where the (j,k) elements of the matrix M_1 is defined as

$$M_1(j,k) \;\;=\;\; \frac{1}{2\pi i}\oint G_j G_k^* \phi_0 \frac{1}{\lambda_y^2 + GG^*\lambda_u^2}\frac{dz}{z}. \tag{5.3}$$

Here ϕ_0 is the spectrum of the noise-free input, and G_j denotes the derivative of the system transfer function $G(z)$ with respect to the system parameters θ_j.

Proof: See Appendix A. ■

The theorem states that, for large SNR's,

$$\mathrm{cov}(\hat{\theta}_{\mathrm{SML}}) \approx \mathrm{cov}(\hat{\theta}_{\mathrm{TML}}) \approx \mathrm{CRBA}. \tag{5.4}$$

The asymptotic estimation accuracy for SML and TML will be very similar when the SNRs at both the input and output sides are large. Both will be approximately equal to an asymptotical CRB, which is directly proportional to $1/\lambda_v^2$. As stated in [5], this result can be weakened for SML. It is enough that for SML, *either* the input SNR or the output SNR becomes large, the $\mathrm{cov}(\hat{\theta}_{\mathrm{SML}})$ will reduce to CRBA. The covariance expressions in Theorem 1 can be simplified. The details are given in Appendix B.

6 Numerical comparisons of TML and SML estimates

We will show some numerical experiments for the TML and SML methods for different signal-to-noise ratios (SNR). A second order system and a sixth order system are illustrated in this paper. The polynomials of this second order system are

$$\begin{aligned} A(q^{-1}) &= 1 - 1.5\,q^{-1} + 0.7\,q^{-2}, \\ B(q^{-1}) &= 2 + 1.0\,q^{-1} + 0.5\,q^{-2}, \end{aligned} \tag{6.1}$$

and the sixth order system is

$$\begin{aligned} A(q^{-1}) &= 1 - 1.1\,q^{-1} + 0.5\,q^{-2} - 0.12\,q^{-3} \\ &\quad +0.23\,q^{-4} - 0.235\,q^{-5} + 0.175\,q^{-6}, \\ B(q^{-1}) &= 0.1 + 1.0\,q^{-1} + 0.85\,q^{-2} + 0.06\,q^{-3} \\ &\quad -0.534\,q^{-4} + 0.504\,q^{-5} + 0.324\,q^{-6}. \end{aligned} \tag{6.2}$$

The polynomials of the noise-free input signal model are

$$\begin{aligned} C(q^{-1}) &= 1 + 0.5\,q^{-1}, \\ D(q^{-1}) &= 1 - 0.5\,q^{-1}. \end{aligned} \tag{6.3}$$

All the comparisons are based on the asymptotic case where the data number N is assumed to be large enough, and we assume $M = 6$ periods data are available. For comparison, we also give the asymptotic covariance matrix of the frequency domain maximum likelihood (FML) method calculated under the assumption of knowing the input-output noise variances and the period information. See [7] for details. The standard deviations (std) are calculated from the theoretical covariance matrices of the estimation parameters, which have been proved to well meet their relevant Monte-Carlo simulations. Details on these formulas can be found in [9], [6], [7] and [5].

In the following numerical analysis, we fix the noise variance at both input and output sides to be 1 (for white noises cases), and let the noise variance of $v(t)$ firstly be 10 and then decrease to 0.1. For these two cases, the estimation results for the second order system (6.1) and the sixth order system (6.2) are shown in Figure 2 and 3. In each figure, the upper part give the spectra of the noises and noise-free signals at the input and output sides. In the lower part of the figures, the standard deviation of the estimated system transfer function under different frequencies are plotted for different estimating methods. Besides, the amplitude of the true transfer function G are also given for convenience.

Comparison results show that when the SNR at both the input and the output side are high or moderate, the two methods always give very similar performance both for low and high order systems. See Figure 2. When the SNR becomes very low, the difference of the TML method and the SML method are observable only in the low SNR frequency regions especially for the high order dynamic systems. See Figure 3. It seems that, in regions where SNR is poor, the benefit of using periodic information in the SML method is more pronounced, which results in the SML having a lower covariance matrix than that of the TML. In Figure 4, new comparison results under the same condition as in Figure 3 are shown except adding the periodic information to TML by simple averaging

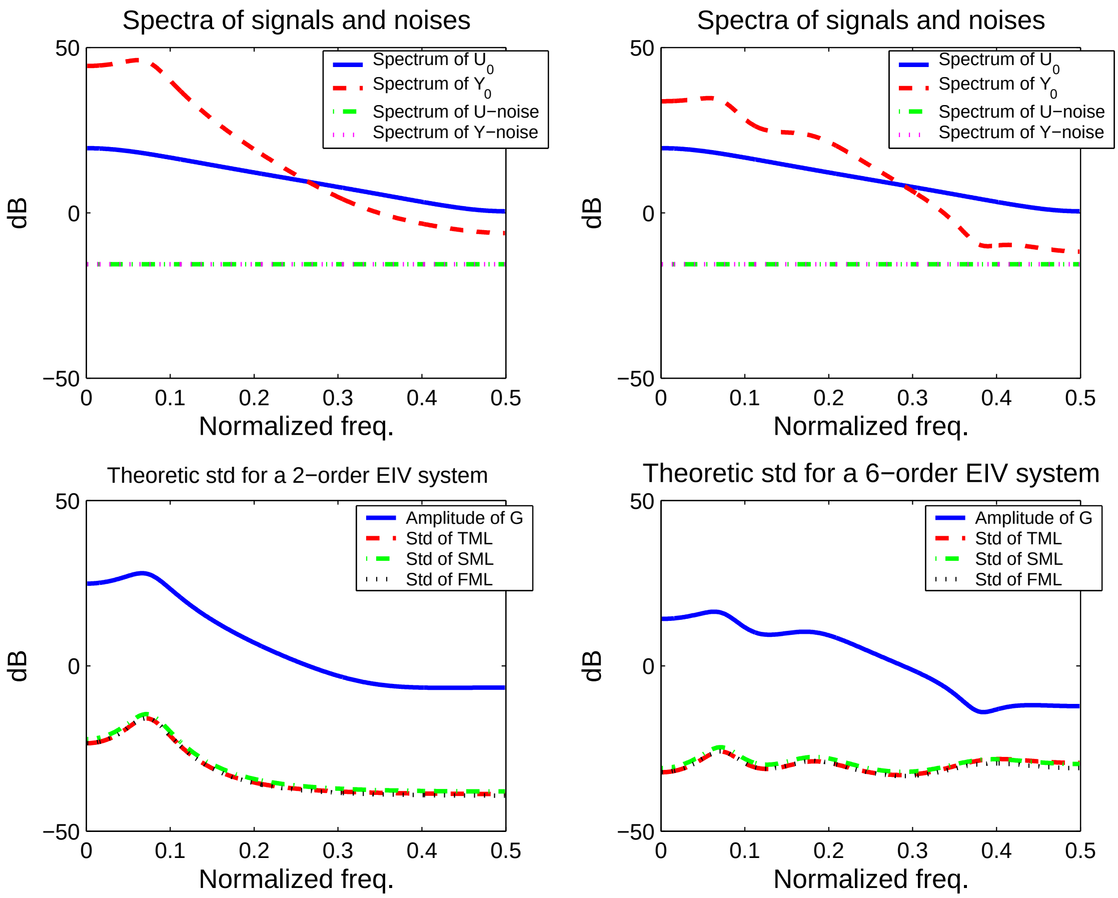

Figure 2: Comparison of the TML, SML and FML estimates for a second (left) and a sixth (right) order system with $\lambda_{\tilde{u}}^2 = 1$, $\lambda_{\tilde{y}}^2 = 1$, and $\lambda_v^2 = 10$.

of the data over the M periods. It can be seen that the difference between TML and SML estimation results in the low SNR area has disappeared.

Besides, several examples with colored output measurement noises were also studied. They give similar results as for the white noise cases.

The preceeding theoretical and numerical studies show that, when the SNR level is large or modest, the estimation accuracy of SML and TML method are quite similar. However, we should note that they are not identical, since the two methods are based on different conditions/assumptions. The matrices $\mathrm{cov}(\hat{\theta}_{\mathrm{SML}})$, $\mathrm{cov}(\hat{\theta}_{\mathrm{TML}})$ and CRBA are only approximately equal to each other. It means that the difference between the two covariance matrices is not necessarily positive definite, and there are not any order relation between these three matrices. For example, let us check the eigenvalues of the matrix difference between the asymptotic covariance matrices of TML and SML. Let the vector Λ denote the eigenvalues of the difference matrix of the covariance matrices of the two methods. For the second order system as (6.1) with the noise-free input signal model as (6.3), it holds

$$\Lambda = \begin{pmatrix} -1.96 \times 10^{-3}, & 1.13 \times 10^{-4}, & -1.48 \times 10^{-4}, & 1.48 \times 10^{-8}, & 2.97 \times 10^{-6} \end{pmatrix}.$$

It can be seen that the eigenvalues of the difference matrix have both positive and neg-

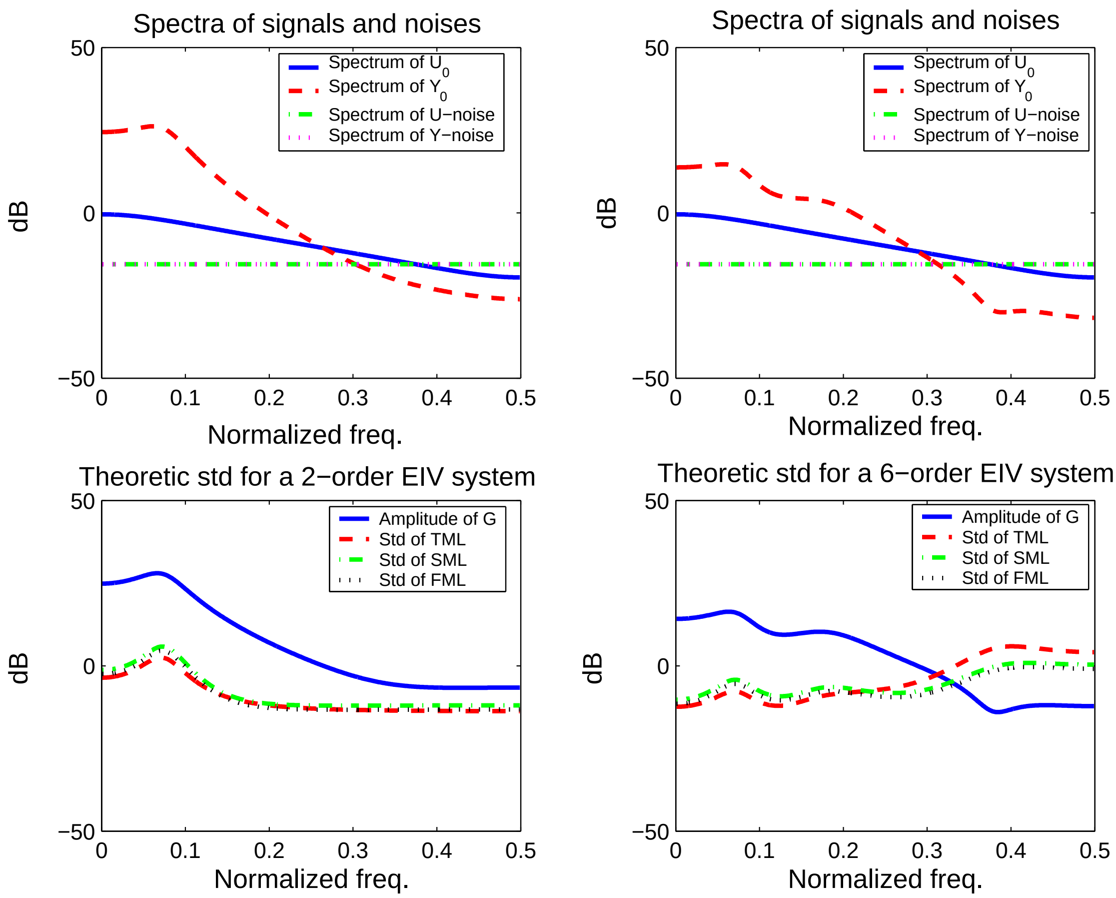

Figure 3: Comparison of the TML, SML and FML estimates for a second (left) and a sixth (right) order system with $\lambda_{\tilde{u}}^2 = 1$, $\lambda_{\tilde{y}}^2 = 1$, and $\lambda_v^2 = 0.1$.

ative values. Hence the difference between the two covariance matrices is indefinite. None of the two methods, SML and TML, is uniformly better than the other.

7 Using periodic data

When the unperturbed input is periodic, the way the estimation problem is treated so far, both for SML and for TML, is to average over the M periods. In this way we get a new data set, where the data length is N (not NM as for the original data series). The effect is also that the variance of the measurement noise decreases with a factor M, both on the input side and on the output side.

However, using the averaged data in this way to compute the covariance matrix of estimates does not give the true CRB. The true CRB is lower. The reason can be explained as follows. Let the measured data series be a long vector, that is partitioned into M blocks each of size N,

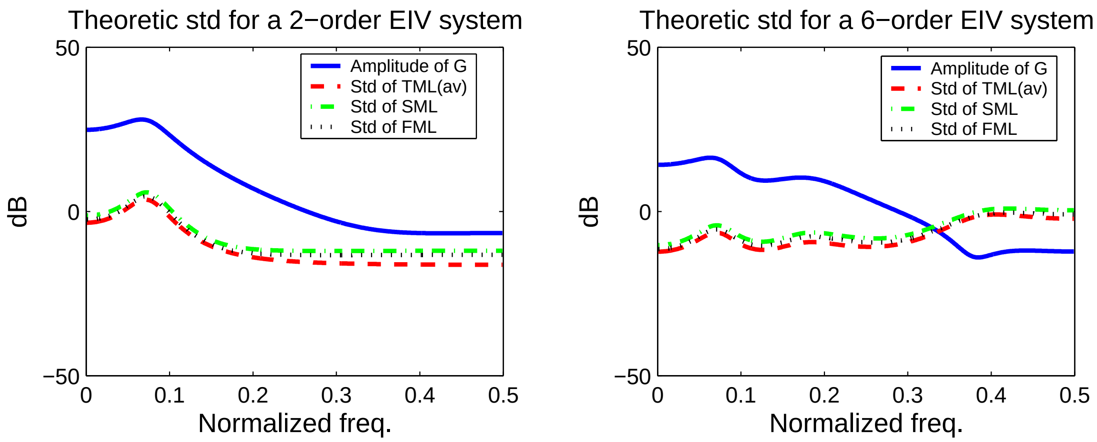

Figure 4: Comparison of the TML (with averaging of the data), SML and FML estimates for a second (left) and a sixth (right) order system with $\lambda_{\tilde{u}}^2 = 1$, $\lambda_{\tilde{y}}^2 = 1$, and $\lambda_v^2 = 0.1$.

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_M \end{pmatrix}. \tag{7.1}$$

Let us then make a linear (nonsingular) transformation of the full data vector as

$$W = \begin{pmatrix} \frac{1}{M}I_N & \frac{1}{M}I_N & \dots & \frac{1}{M}I_N \\ I_N & -I_N & & \\ & I_N & -I_N & \\ & \dots & I_N & -I_N \end{pmatrix} Z = \begin{pmatrix} \frac{1}{M}\sum_k Z_k \\ Z_1 - Z_2 \\ \dots \\ Z_{M-1} - Z_M \end{pmatrix} \triangleq \begin{pmatrix} W_1 \\ W_2 \\ \dots \\ W_M \end{pmatrix}. \tag{7.2}$$

To compute the CRB from Z must be the same as to compute the CRB from W. However, in the simplified form we use only W_1 for computing the CRB and neglect the remaining part of the data. The parts $W_2, \dots, W_M$ do not depend on the noise-free input, but on the noise statistics (say the variance λ_u^2 of the input noise $\tilde{u}(t)$ and the variance λ_y^2 of the output noise $\tilde{y}(t)$. As the CRB of the system parameters (that is, the A and B coefficients) and the noise parameters is **not** block-diagonal, it will be beneficial from the accuracy point of view, to make use of also the remaining data $W_2, \dots, W_M$.

To simplify things, the preceding sections consider only the case when W_1 is used. It is a reasonable thing to do and worthwhile enough to focus on this 'simplified or idealized CRB'. In this section, we will further examine the effect of additional data records $W_2, \dots W_M$. To this aim, it is useful to use the Slepian-Bang formula, [12], for the CRB. It holds for Gaussian distributed data

$$\text{FIM}_{j,k} = \frac{1}{2}\text{tr}\left(R^{-1}R_j R^{-1}R_k\right), \quad R_j = \frac{\partial R}{\partial \theta_j}, \tag{7.3}$$

and R denotes the covariance matrix of the full data vector W.

Now split the covariance matrix R as

$$R = EWW^T = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22}, \end{pmatrix} \tag{7.4}$$

where R_{11} corresponds to the data part W_1 and the block R_{22} is associated to $W_2 \dots W_M$.

It is easy to see that for large λ_v^2

$$R = \begin{pmatrix} O(\lambda_v^2) & O(1) \\ O(1) & O(1) \end{pmatrix}. \tag{7.5}$$

The previous analysis, where only the data part W_1 was used, lead to the conclusion

$$\mathsf{tr}\left[R_{11}^{-1}R_{11,j}R_{11}^{-1}R_{11,k}\right] = O(\lambda_v^2). \tag{7.6}$$

see Appendix A.1.1. We now show that the same type of result holds also when the full matrix R is considered. More specifically, we have the following result.

Lemma 1: It holds that

$$\mathsf{tr}\left[R^{-1}R_jR^{-1}R_k\right] - \mathsf{tr}\left[R_{11}^{-1}R_{11,j}R_{11}^{-1}R_{11,k}\right] = O(1), \quad \lambda_v^2 \to \infty. \tag{7.7}$$

Proof. See Appendix D. ∎

Lemma 1 can be made more precise. The details are given in Appendix C.

8 Conclusions

In this paper, the asymptotic covariance matrices of the TML method and the SML method for estimating the EIV systems have been theoretically and numerically compared. It was shown that, although these two estimates are based on the different assumptions, they have very similar estimation accuracy when the SNR values at both input and output sides are not low. When the SNR is very low (less than 0 dB), it seems that the benefit of using the periodic information is more important than knowing that both the signals and noises have rational spectra. A notable accuracy difference can be observed at low SNR regions especially for high order dynamic systems. From the efficiency point of view, SML and TML have similar estimation accuracy at moderate or high SNR cases.

A Proof of Theorem 1

A.1 Proof of $\mathrm{cov}(\hat{\theta}_{\mathrm{TML}}) \approx \mathrm{CRBA}$

The basis here is to use the Whittle formula for the information matrix, see [11]

$$\mathrm{FIM}_{j,k} = \frac{1}{4\pi i}\oint \mathrm{tr}\,\left(\Phi(z)\frac{\partial\Phi(z)}{\partial\theta_j}\Phi(z)\frac{\partial\Phi(z)}{\partial\theta_k}\right)\frac{dz}{z}. \tag{A.1}$$

Here, the spectrum of the measured input-output data is written as

$$\Phi(z) = \tilde{\Phi}(z) + H(z)\phi_0(z)H^*(z), \tag{A.2}$$

$$\tilde{\Phi}(z) = \begin{pmatrix} \lambda_y^2 & 0 \\ 0 & \lambda_u^2 \end{pmatrix}, \tag{A.3}$$

$$H(z) = \begin{pmatrix} G(z) \\ 1 \end{pmatrix} = \begin{pmatrix} B(z)/A(z) \\ 1 \end{pmatrix}, \tag{A.4}$$

and $\phi_0(z)$ is the spectrum of the noise-free input.

Note that this framework and formulation can (most likely) be generalized to cover

- MIMO case
- $\tilde{u}, \tilde{y}$ colored noise
- $\tilde{u}, \tilde{y}$ cross-correlated

A.1.1 Neglecting noise parameters

First we study the information matrix with respect to the components associated with the A and B parameters (and neglect for the time being the impact of the noise parameters). Such θ values enter only in H, not in $\tilde{\Phi}$ or ϕ_0.

Although $\Phi(z) = O(\lambda_v^2)$, it will be shown that $\Phi^{-1}(z) = O(1)$ because $H(z)\phi_0(z)H^*(z)$ in (A.2) is a rank one matrix. In fact, for large λ_v^2

$$\begin{aligned}
\Phi^{-1} &= \begin{pmatrix} GG^*\phi_0 + \lambda_y^2 & G\phi_0 \\ G^*\phi_0 & \phi_0 + \lambda_u^2 \end{pmatrix}^{-1} \\
&= \frac{1}{\phi_0\left(GG^*\lambda_u^2 + \lambda_y^2\right) + \lambda_u^2\lambda_y^2} \begin{pmatrix} \phi_0 + \lambda_u^2 & -G\phi_0 \\ -G^*\phi_0 & GG^*\phi_0 + \lambda_y^2 \end{pmatrix} \\
&\approx \frac{1}{GG^*\lambda_u^2 + \lambda_y^2} \begin{pmatrix} 1 & -G \\ -G^* & GG^* \end{pmatrix} = O(1),
\end{aligned} \tag{A.5}$$

We now have

$$\frac{\partial\Phi(z)}{\partial\theta_j} = \frac{\partial H(z)}{\partial\theta_j}\phi_0(z)H^*(z) + H(z)\phi_0(z)\frac{\partial H^*(z)}{\partial\theta_j} \triangleq H_j\phi_0H^* + H\phi_0H_j^*. \tag{A.6}$$

In the following, the argument z is dropped for convenience. Then,

$$\begin{aligned}
\Phi^{-1}\frac{\partial\Phi}{\partial\theta_j} &= \Phi^{-1}\left[H_j\phi_0H^* + H\phi_0H_j^*\right] \\
&= \Phi^{-1}H_j\phi_0H^* + \left[\tilde{\Phi} + H\phi_0H^*\right]^{-1} H\phi_0H_j^* \\
&= \Phi^{-1}H_j\phi_0H^* + \left[\tilde{\Phi}^{-1} - \tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1} H^*\tilde{\Phi}^{-1}\right] H\phi_0H_j^* \\
&= \Phi^{-1}H_j\phi_0H^* + \tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1} \\
&\quad \times \left[\left(\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right) - H^*\tilde{\Phi}^{-1}H\right]\phi_0H_j^* \\
&= \Phi^{-1}H_j\phi_0H^* + \tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1} H_j^*.
\end{aligned} \tag{A.7}$$

In case of a large SNR we have

$$\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H \approx H^*\tilde{\Phi}^{-1}H. \tag{A.8}$$

We also have

$$\begin{aligned}
H^*\Phi^{-1} &= H^*\left[\tilde{\Phi}^{-1} - \tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H^*\tilde{\Phi}^{-1}\right] \\
&= \left[\left(\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right) - H^*\tilde{\Phi}^{-1}H\right]\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H^*\tilde{\Phi}^{-1} \\
&= \phi_0^{-1}\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H^*\tilde{\Phi}^{-1} \\
&\approx \phi_0^{-1}\left[H^*\tilde{\Phi}^{-1}H\right]^{-1}H^*\tilde{\Phi}^{-1}.
\end{aligned} \tag{A.9}$$

Using (A.7) we get

$$\begin{aligned}
\mathrm{tr}\left[\Phi^{-1}\frac{\partial\Phi}{\partial\theta_j}\Phi^{-1}\frac{\partial\Phi}{\partial\theta_k}\right] &= \mathrm{tr}\left[\Phi^{-1}H_j\phi_0H^* + \tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H_j^*\right] \\
&\quad\times\left[\Phi^{-1}H_k\phi_0H^* + \tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H_k^*\right] \\
&= \mathrm{tr}\left[\Phi^{-1}H_j\phi_0H^*\Phi^{-1}H_k\phi_0H^*\right] \\
&\quad+\mathrm{tr}\left[\Phi^{-1}H_j\phi_0H^*\tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H_k^*\right] \\
&\quad+\mathrm{tr}\left[\Phi^{-1}H_k\phi_0H^*\tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H_j^*\right] \\
&\quad+\mathrm{tr}\left[\tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}\right. \\
&\quad\left.\times H_j^*\tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H_k^*\right].
\end{aligned} \tag{A.10}$$

Using the high SNR approximation (A.8) we get further

$$\begin{aligned}
\mathrm{tr}\left[\Phi^{-1}\frac{\partial\Phi}{\partial\theta_j}\Phi^{-1}\frac{\partial\Phi}{\partial\theta_k}\right] &\approx \mathrm{tr}\left[\phi_0^{-1}\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H^*\tilde{\Phi}^{-1}H_j\phi_0\right. \\
&\quad\left.\times\phi_0^{-1}\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H^*\tilde{\Phi}^{-1}H_k\phi_0\right] \\
&\quad+\mathrm{tr}\left[H_k^*\Phi^{-1}H_j\phi_0\right] \\
&\quad+\mathrm{tr}\left[H_j^*\Phi^{-1}H_k\phi_0\right] \\
&\quad+\mathrm{tr}\left[\tilde{\Phi}^{-1}H\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H_j^*\tilde{\Phi}^{-1}H\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H_k^*\right].
\end{aligned} \tag{A.11}$$

When $\lambda_v^2 \to \infty$, the second and the third term in (A.11) dominate, and give the same contribution to the Fisher information matrix. Hence

$$\begin{aligned}
F_{j,k} &\triangleq \{\mathrm{FIM}\}_{j,k} \\
&\approx \operatorname{tr}\left\{\frac{1}{2\pi i}\oint \begin{bmatrix} G_j \\ 0 \end{bmatrix} \phi_o \begin{bmatrix} GG^*\phi_0+\lambda_y^2 & G\phi_0 \\ G^*\phi_0 & \phi_0+\lambda_u^2 \end{bmatrix}^{-1} \begin{bmatrix} G_k^* & 0 \end{bmatrix} \frac{dz}{z}\right\} \\
&= \frac{1}{2\pi i}\oint G_j(z)G_k^*(z)\phi_0(z)\frac{\phi_0(z)+\lambda_u^2}{\phi_0(z)[G(z)G^*(z)\lambda_u^2+\lambda_y^2]+\lambda_u^2\lambda_y^2}\frac{dz}{z}. \qquad \text{(A.12)}
\end{aligned}$$

Assuming the SNR is large in the sense $\lambda_v^2 \to \infty$ we have

$$F_{j,k} \approx \frac{1}{2\pi i}\oint G_j(z)G_k^*(z)\phi_0(z)\frac{1}{G(z)G^*(z)\lambda_u^2+\lambda_y^2}\frac{dz}{z} \triangleq M_1(j,k). \qquad \text{(A.13)}$$

As shown in Appendix B, the calculations for (A.13) can be simplified.

A.1.2 Taking effects of the noise parameters into account

So far, we have considered a block of the Fisher information matrix rather than the Cramér-Rao lower bound. Let the total parameter vector be

$$\begin{pmatrix} \theta \\ \vartheta \end{pmatrix}, \qquad \text{(A.14)}$$

where the vector ϑ contains parameter describing the spectra of the noises and the noise-free input. The total information matrix has the form

$$F = \begin{pmatrix} F_{\theta\theta} & F_{\theta\vartheta} \\ F_{\vartheta\theta} & F_{\vartheta\vartheta} \end{pmatrix}. \qquad \text{(A.15)}$$

Our previous analysis implied that

$$F_{\theta\theta} = O(\lambda_v^2), \quad [F_{\theta\theta}]^{-1} = O(1/\lambda_v^2). \qquad \text{(A.16)}$$

However, the true CRB for the vector $\hat{\theta}$ is

$$\mathrm{CRB}_\theta = \left(F^{-1}\right)_{\theta\theta} = \left(F_{\theta\theta} - F_{\theta\vartheta}F_{\vartheta\vartheta}^{-1}F_{\vartheta\theta}\right)^{-1}. \qquad \text{(A.17)}$$

A more careful analysis will give the following lemma:

Lemma 2: When $\lambda_v^2 \to \infty$, $F_{\theta\vartheta} = O(1)$ and $F_{\vartheta\vartheta} = O(1)$.

Proof: See Appendix D. ■

This implies that

$$F_{\theta\theta} - F_{\theta\vartheta}F_{\vartheta\vartheta}^{-1}F_{\vartheta\theta} \approx F_{\theta\theta} = O(\lambda_v^2), \qquad \text{(A.18)}$$

and we can then (asymptotically for large SNR) neglect the influence of the blocks $F_{\theta\vartheta}$ and $F_{\vartheta\vartheta}$. A more sophisticated way of expressing this is to say that the canonical correlations between θ and ϑ which are the same as the singular values of

$$F_{\theta\theta}^{-1/2}F_{\theta\vartheta}F_{\vartheta\vartheta}^{-1/2} \qquad \text{(A.19)}$$

are all very small.

Then, the asymptotic covariance matrix of TML, $\mathrm{cov}(\hat{\theta}_{\mathrm{TML}})$ will approximate equal to $F_{\theta\vartheta}^{-1}$, *i.e.* $\mathrm{cov}(\hat{\theta}_{\mathrm{TML}}) \approx F_{\theta\vartheta}^{-1} \triangleq M_1^{-1} \triangleq \mathrm{CRBA}$. The proofs are done.

A.1.3 Numerical illustrations of canonical correlations

In the following, we use a numerical example to support the theoretical statements as in (A.18)-(A.19). Consider a system given by

$$
\begin{aligned}
(1 - 1.5q^{-1} + 0.7q^{-2})y_0(t) &= (2.0q^{-1} + 1.0q^{-2})u_0(t), && \text{(A.20)}\\
(1 - 0.5q^{-1})u_0(t) &= (1 + 0.7q^{-1})v(t). && \text{(A.21)}
\end{aligned}
$$

Set the noise variances at input and output sides as $\lambda_{\tilde{u}}^2 = 3$ and $\lambda_{\tilde{u}}^2 = 2$. The noise variance λ_v^2 was used as parameter and varied.

Firstly, in Figure 5, we show the variances (normalized with N) from the $\mathrm{cov}(\hat{\theta}_{\mathrm{TML}})$ of the system parameters θ as functions of λ_v^2. One can see that all variances decrease as $1/\lambda_v^2$ for large λ_v^2. The dashed lines correspond to the asymptotic results (CRBA).

Secondly, Figure 6 shows the corresponding part of the Fisher information matrix, more precisely the diagonal elements of $F_{\theta\theta}$. It is seen that these four elements increase linearly with λ_v^2 for high values of λ_v^2. Recall that Figure 5 and 6 illustrate different things, as Figure 5 refer to the diagonal elements of $(F_{\theta\theta} - F_{\theta\vartheta}F_{\vartheta\vartheta}^{-1}F_{\vartheta\theta})^{-1}$.

Finally, we show in Figure 7 that the canonical correlations, that is the singular values of the matrix $F_{\theta\theta}^{-1/2}F_{\theta\vartheta}F_{\vartheta\vartheta}^{-1/2}$. As expected from theory, the canonical correlations become small when λ_v^2 is large.

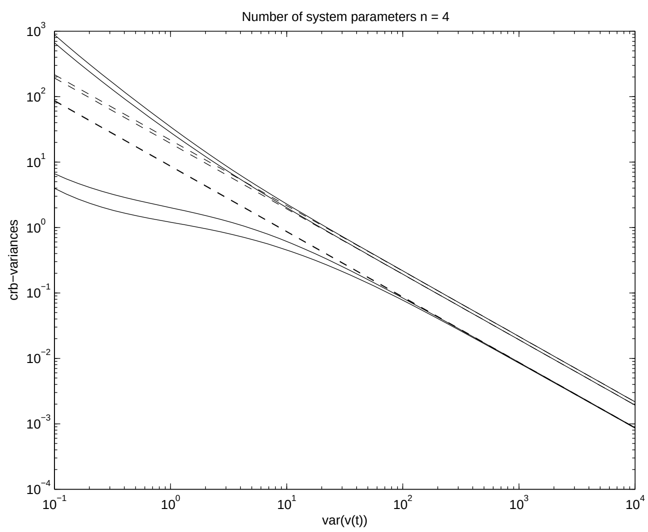

Figure 5: Variances as expressed by the CRBT.

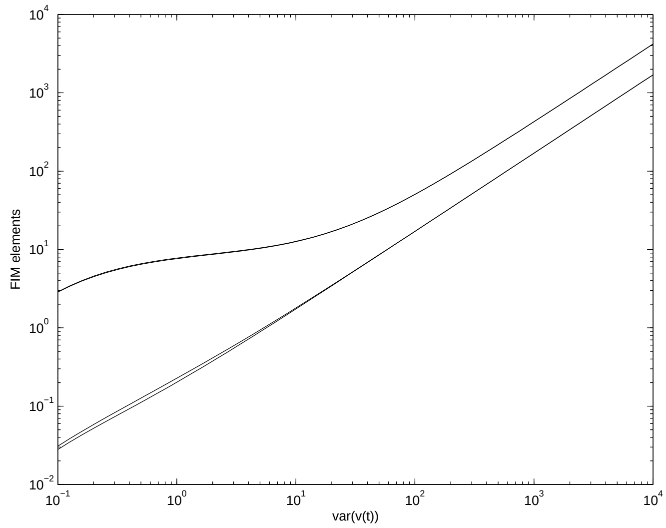

Figure 6: Diagonal elements of upper left block of the Fisher information matrix (for TML): $F_{\theta\theta}$.

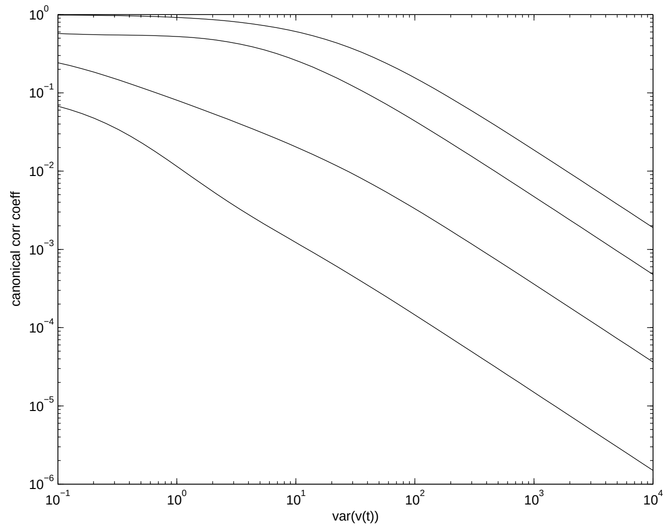

Figure 7: Canonical correlations.

A.2 Proof of $\mathrm{cov}(\hat{\theta}_{\mathrm{SML}}) \approx \mathrm{CRBA}$

The normalized asymptotic ($N \to \infty$) covariance matrices for the SML [4] methods are derived in [5] as

$$\mathrm{cov}(\hat{\theta}_{\mathrm{SML}}) = M_1^{-1} + M_1^{-1} M_2^{-1} M_1^{-1} = M_1^{-1}(M_1 + M_2) M_1^{-1}, \tag{A.22}$$

where

$$M_1 = \frac{1}{2\pi i} \oint G_j G_k^* \phi_0 \frac{1}{\lambda_y^2 + GG^* \lambda_u^2} \frac{dz}{z}, \tag{A.23}$$

[4]Here we use the SML that neglecting the noise parameters are to be estimated. In practice, when the noise parameters are unknown, there will be factors that are somewhat larger than 1 in front of the two terms in (A.22). The factors depend only on M. See [5] for a detailed formula.

and

$$M_2 \quad = \quad \frac{1}{2\pi i}\oint G_j G_k^* \frac{\lambda_u^2\lambda_y^2}{(\lambda_y^2 + GG^*\lambda_u^2)^2}\frac{dz}{z}. \tag{A.24}$$

So for large values of SNR, *for example* when λ_y^2 be fixed and λ_u^2 and λ_y^2 become very small, the first term in (A.22) will dominates and the second term can be neglected. Then $\mathrm{cov}(\hat{\theta}_{\mathrm{SML}}) \approx M_1^{-1} = \mathrm{CRBA}$ is proved.

B Simplifying the covariance expressions

The calculation for CRBA can be simplified by using the following way. Firstly, we make a spectral factorization (with $L(z)$ being a monic polynomial)

$$G(z)G^*(z)\lambda_u^2 + \lambda_y^2 = \frac{B(z)B^*(z)}{A(z)A^*(z)}\lambda_u^2 + \lambda_y^2 = \lambda_\varepsilon^2\frac{L(z)L^*(z)}{A(z)A^*(z)}. \tag{B.1}$$

This spectral factorization can easily be generalized to cases when the measurement noises are correlated and colored.

Next we write the noise-free input spectrum as

$$\phi_0(z) = \lambda_v^2\frac{C(z)C^*(z)}{D(z)D^*(z)}. \tag{B.2}$$

Then we find from (A.13) that

$$F_{j,k} \approx E\left[\frac{G_j(q^{-1})C(q^{-1})A(q^{-1})}{\lambda_\varepsilon D(q^{-1})L(q^{-1})}v(t)\right]\left[\frac{G_k(q^{-1})C(q^{-1})A(q^{-1})}{\lambda_\varepsilon D(q^{-1})L(q^{-1})}v(t)\right]. \tag{B.3}$$

We next elaborate more on $G_j(z)$. It apparently holds

$$\begin{pmatrix} \frac{\partial G(q^{-1})}{\partial a_1} \\ \vdots \\ \frac{\partial G(q^{-1})}{\partial b_n} \end{pmatrix} = \begin{pmatrix} 0 & -b_1 & \dots & b_n & & \\ 0 & 0 & \ddots & & \ddots & \\ 0 & \dots & & -b_1 & \dots & -b_n \\ 1 & a_1 & \dots & a_n & & \\ 0 & \ddots & & & \ddots & \\ 0 & \dots & 1 & & & a_n \end{pmatrix} \begin{pmatrix} \frac{q^{-1}}{A^2(q^{-1})} \\ \dots \\ \frac{q^{-2n}}{A^2(q^{-1})} \end{pmatrix}. \tag{B.4}$$

The matrix in (B.4) is a Sylvester one, and will be denoted by $\mathcal{S}(A, B)$. It is nonsingular if we assume that A and B are coprime. From (B.3) and (B.4) we find

$$\{F_{j,k}\}_{j,k=1}^{2n} = \mathcal{S}(A, B)P_0\mathcal{S}^T(A, B), \tag{B.5}$$

where

$$P_0 = \mathrm{cov}\left(\frac{C(q^{-1})}{\lambda_\varepsilon A(q^{-1})D(q^{-1})L(q^{-1})}\begin{pmatrix} v(t-1) \\ \vdots \\ v(t-2n) \end{pmatrix}\right). \tag{B.6}$$

The matrix P_0 can easily be determined numerically in an accurate way.

Similarity, for M_2 in the expression of $\mathrm{cov}(\hat{\theta}_{\mathrm{SML}})$, we can formulate it as:

$$\begin{aligned} M_2 &= \frac{1}{2\pi i}\oint G_j G_k^* \frac{\lambda_u^2\lambda_y^2}{(\lambda_y^2 + GG^*\lambda_u^2)^2}\frac{dz}{z} \\ &= \mathcal{S}(A,B)P_2\mathcal{S}^T(A,B), \end{aligned} \tag{B.7}$$

where the matrix P_2 is given by

$$P_2 = \mathrm{cov}\left(\frac{\lambda_u\lambda_y}{\lambda_v\lambda_\varepsilon^2}\frac{1}{L^2(q^{-1})}\left(\begin{array}{c} v(t-1) \\ \vdots \\ v(t-2n) \end{array}\right)\right). \tag{B.8}$$

C Proof of Lemma 1

Using the inversion lemma of partitioned matrices,

$$\begin{aligned} R^{-1}R_j &= \left(\begin{array}{cc} R_{11} & R_{12} \\ R_{21} & R_{22} \end{array}\right)^{-1}\left(\begin{array}{cc} R_{11,j} & R_{12,j} \\ R_{21,j} & R_{22,j} \end{array}\right) \\ &= \left[\left(\begin{array}{cc} R_{11}^{-1} & 0 \\ 0 & 0 \end{array}\right) + \left(\begin{array}{c} R_{11}^{-1}R_{12} \\ -I \end{array}\right)\left(R_{22} - R_{21}R_{11}^{-1}R_{12}\right)^{-1}\left(\begin{array}{cc} R_{21}R_{11}^{-1} & -I \end{array}\right)\right] \\ &\quad \times \left(\begin{array}{cc} R_{11,j} & R_{12,j} \\ R_{21,j} & R_{22,j} \end{array}\right) \\ &= \left(\begin{array}{cc} R_{11}^{-1}R_{11,j} & 0 \\ 0 & 0 \end{array}\right) \\ &\quad + \left(\begin{array}{c} R_{11}^{-1}R_{12} \\ -I \end{array}\right)\left(R_{22} - R_{21}R_{11}^{-1}R_{12}\right)^{-1} \\ &\quad \times \left[\begin{array}{cc} R_{21}R_{11}^{-1}R_{11,j} - R_{21,j} & R_{21}R_{11}^{-1}R_{12,j} - R_{22,j} \end{array}\right] \\ &= \left(\begin{array}{cc} R_{11}^{-1}R_{11,j} & 0 \\ 0 & 0 \end{array}\right) + \left(\begin{array}{c} O(1/\lambda_v^2) \\ O(1) \end{array}\right) \times O(1)\left[\begin{array}{cc} O(1) & O(1) \end{array}\right], \end{aligned} \tag{C.1}$$

which proves the claim.

D Proof of Lemma 2

To proceed, we need to examine the magnitude of the different block terms. First we note that the the filter H does not depend on the vector ϑ. In fact,

$$\Phi_{\vartheta_j} = \tilde{\Phi}_{\vartheta_j} + H\phi_{0\vartheta_j}H^*. \tag{D.1}$$

Next we recall that, see (A.9),

$$H^*\Phi^{-1} \approx \phi_0^{-1}\left[H^*\tilde{\Phi}^{-1}H\right]^{-1}H^*\tilde{\Phi}^{-1}. \tag{D.2}$$

We next get using (A.7), (A.8), (A.9) and letting λ_v^2 be large

$$
\begin{aligned}
\mathrm{tr}\left[\Phi^{-1}\frac{\partial\Phi}{\partial\theta_j}\Phi^{-1}\frac{\partial\Phi}{\partial\vartheta_k}\right] &= \mathrm{tr}\left[\Phi^{-1}H_j\phi_0H^* + \tilde{\Phi}^{-1}H\left[\phi_0^{-1} + H^*\tilde{\Phi}^{-1}H\right]^{-1}H_j^*\right] \\
&\quad \times\Phi^{-1}\left[\tilde{\Phi}_{\vartheta_k} + H\phi_{0\vartheta_k}H^*\right] \\
&\approx \mathrm{tr}\left[\Phi^{-1}H_j\phi_0H^*\Phi^{-1} + \tilde{\Phi}^{-1}H\left[H^*\tilde{\Phi}^{-1}H\right]^{-1}H_j^*\Phi^{-1}\right] \\
&\quad \times\left[\tilde{\Phi}_{\vartheta_k} + H\phi_{0\vartheta_k}H^*\right] \\
&\approx \mathrm{tr}\left[\Phi^{-1}H_j\phi_0H^*\Phi^{-1}\left[\tilde{\Phi}_{\vartheta_k} + H\phi_{0\vartheta_k}H^*\right]\right] \\
&\quad +\mathrm{tr}\left[\tilde{\Phi}^{-1}H\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H_j^*\Phi^{-1}\left[\tilde{\Phi}_{\vartheta_k} + H\phi_{0\vartheta_k}H^*\right]\right] \\
&= O(1) + \mathrm{tr}\left[H^*\tilde{\Phi}^{-1}H\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H_j^*\Phi^{-1}H\phi_{0\vartheta_k}\right] \\
&= O(1) + \mathrm{tr}\left[H_j^*\tilde{\Phi}^{-1}H\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}\phi_0^{-1}\right] \\
&= O(1). \qquad\qquad \text{(D.3)}
\end{aligned}
$$

Hence,

$$F_{\theta\vartheta} = O(1). \qquad\qquad \text{(D.4)}$$

Furthermore,

$$
\begin{aligned}
\mathrm{tr}\left[\Phi^{-1}\frac{\partial\Phi}{\partial\vartheta_j}\Phi^{-1}\frac{\partial\Phi}{\partial\vartheta_k}\right] &= \mathrm{tr}\left[\Phi^{-1}\left[\tilde{\Phi}_{\vartheta_j} + H\phi_{0\vartheta_j}H^*\right]\Phi^{-1}\left[\tilde{\Phi}_{\vartheta_k} + H\phi_{0\vartheta_k}H^*\right]\right] \\
&= \mathrm{tr}\left[\Phi^{-1}\tilde{\Phi}_{\vartheta_j}\Phi^{-1}\tilde{\Phi}_{\vartheta_k}\right] \\
&+ \mathrm{tr}\left[H^*\Phi^{-1}\tilde{\Phi}_{\vartheta_j}\Phi^{-1}H\phi_{0\vartheta_k}\right] \\
&+ \mathrm{tr}\left[H^*\Phi^{-1}\tilde{\Phi}_{\vartheta_k}\Phi^{-1}H\phi_{0\vartheta_j}\right] \\
&+ \mathrm{tr}\left[H^*\Phi^{-1}H\phi_{0\vartheta_j}H^*\Phi^{-1}H\phi_{0\vartheta_j}\right]. \qquad\qquad \text{(D.5)}
\end{aligned}
$$

Now recall (A.5). Hence the first term of (D.5) is $O(1)$. The second and third terms of (D.5) can be evaluated in the same way. Using (D.2) the second term is approximated by

$$\mathrm{tr}\left[\phi_0^{-1}\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H^*\tilde{\Phi}^{-1}\tilde{\Phi}_{\vartheta_j}\tilde{\Phi}^{-1}H\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}\phi_0^{-1}\phi_{0\vartheta_k}\right] = O(1). \qquad \text{(D.6)}$$

Finally, the last term of (D.5) is found to satisfy, for large λ_v^2,

$$
\begin{aligned}
&\mathrm{tr}\left[\phi_0^{-1}\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H^*\tilde{\Phi}^{-1}H\phi_{0\vartheta_j}\;\phi_0^{-1}\left(H^*\tilde{\Phi}^{-1}H\right)^{-1}H^*\tilde{\Phi}^{-1}H\phi_{0\vartheta_k}\right] \\
&= \left[\frac{\phi_{0\vartheta_j}\phi_{0\vartheta_k}}{\phi_0^2}\right] = O(1). \qquad\qquad \text{(D.7)}
\end{aligned}
$$

Hence

$$F_{\vartheta\vartheta} = O(1), \tag{D.8}$$

and the lemma 2 is proved.

E Further results for periodic data

We can in fact be more precise for Lemma 1. The blocks of the data in (7.1) can be written as

$$Z_k = Z_0 + \tilde{Z}_k, \quad k = 1, \dots, M, \tag{E.1}$$

where Z_0 denotes the effect of the noise-free input, and the noise contributions $\{\tilde{Z}_k\}_{k=1}^{M}$ are assumed to be uncorrelated between different periods. Introduce the notations

$$\begin{array}{rcl} R_0 & = & \mathrm{cov}(Z_0), \\ \tilde{R} & = & \mathrm{cov}(\tilde{Z}_k). \end{array} \tag{E.2}$$

Using the full data vector W, from (7.2) we have

$$R = EWW^T = \begin{pmatrix} R_0 + \frac{1}{M}\tilde{R} & 0 & & 0 & \\ 0 & 2\tilde{R} & -\tilde{R} & & \\ \vdots & -\tilde{R} & 2\tilde{R} & & \\ 0 & & & \ddots & \\ & & & & 2\tilde{R} \end{pmatrix} \tag{E.3}$$

$$= \begin{pmatrix} R_0 + \frac{1}{M}\tilde{R} & \mathbf{0} \\ \mathbf{0} & J \otimes \tilde{R} \end{pmatrix} \triangleq \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}, \tag{E.4}$$

where $\otimes$ denotes Kronecker product and

$$J = \begin{pmatrix} 2 & -1 & 0 & & \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ & & & \ddots & -1 \\ & & & -1 & 2 \end{pmatrix}. \tag{E.5}$$

As R in (E.4) is block diagonal, we find directly that taking the block R_2 into account means that we get an additional term in the Fisher information matrix. We get

$$\begin{aligned} \mathrm{FIM}_{j,k} &= \frac{1}{2}\mathsf{tr}\ \left[R^{-1}R_jR^{-1}R_k\right] \\ &= \frac{1}{2}\mathsf{tr}\ \left[R_1^{-1}R_{1,j}R_1^{-1}R_{1,k}\right] \\ &\quad + \frac{1}{2}\mathsf{tr}\ \left[R_2^{-1}R_{2,j}R_2^{-1}R_{2,k}\right]. \end{aligned} \tag{E.6}$$

The second, additional term in (E.6) can be expressed more explicitly as

$$
\begin{aligned}
\tilde{F}_{j,k} &= \frac{1}{2}\mathsf{tr}\ \left[R_2^{-1}R_{2,j}R_2^{-1}R_{2,k}\right] \\
&= \frac{1}{2}\mathsf{tr}\ \left[(J\otimes\tilde{R})^{-1}(J\otimes\tilde{R}_j)(J\otimes\tilde{R})^{-1}(J\otimes\tilde{R}_k)\right] \\
&= \frac{1}{2}\mathsf{tr}\ \left[I_{M-1}\otimes\left(\tilde{R}^{-1}\tilde{R}_j\tilde{R}^{-1}\tilde{R}_k\right)\right] \\
&= \frac{M-1}{2}\mathsf{tr}\ \left[\tilde{R}^{-1}\tilde{R}_j\tilde{R}^{-1}\tilde{R}_k\right].
\end{aligned} \tag{E.7}
$$

Using the partioning of the parameter vector into θ and ϑ as in (A.14) , we find easily that $\tilde{F}_{j,k}$ is nonzero only for the $\vartheta\vartheta$ block. In particular, when both $\tilde{y}(t)$ and $\tilde{u}(t)$ are white noise, we have

$$
\tilde{R} = \begin{pmatrix} \lambda_y^2 I_N & \mathbf{0} \\ \mathbf{0} & \lambda_y^2 I_N \end{pmatrix}. \tag{E.8}
$$

Then it is straightforward to derive

$$
\tilde{F}_{\lambda_y^2,\lambda_y^2} = \frac{(M-1)N}{2\lambda_y^4}, \tag{E.9}
$$

$$
\tilde{F}_{\lambda_u^2,\lambda_u^2} = \frac{(M-1)N}{2\lambda_u^4}, \tag{E.10}
$$

$$
\tilde{F}_{\lambda_y^2,\lambda_u^2} = 0. \tag{E.11}
$$

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, NJ, USA, 1979.

[2] S. Beghelli, R.P. Guidorzi, and U. Soverini. The Frisch scheme in dynamic system identification. *Automatica*, 26:171–176, 1990.

[3] K. V. Fernando and H. Nicholson. Identification of linear systems with input and output noise: the Koopmans–Levin method. *IEE Proceedings, Part D*, 132(1):30–36, January 1985.

[4] R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, and H. Van Hamme. Parametric identification of transfer functions in the frequency domain – a survey. *IEEE Transactions on Automatic Control*, 39(11):2245–2260, November 1994.

[5] R. Pintelon and M. Hong. Asymptotic uncertainty of transfer function estimates using nonparametric noise models. In *Proc. IEEE Instrumentation and Measurement Technology Conference*, Warzaw, Poland, May 01-03 2007.

[6] R. Pintelon and J. Schoukens. *System Identification. A Frequency Domain Approach*. IEEE Press, New York, NY, USA, 2001.

[7] J. Schoukens, R. Pintelon, G. Vandersteen, and P. Guillaume. Frequency domain system identification using non-parametric noise models estimated from a small number of data sets. *Automatica*, 33:1073–1086, 1997.

[8] T. Söderström. Identification of stochastic linear systems in presence of input noise. *Automatica*, 17:713–725, 1981.

[9] T. Söderström. On computing the Cramer-Rao bound and covariance matrices for PEM estimates in linear state space models. In *Proc. 14th IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

[10] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6):939–958, June 2007. Survey paper.

[11] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, Hemel Hempstead, UK, 1989.

[12] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Pearson Prentice-Hall, Upper Saddle River, NJ, 2005.

[13] J. K. Tugnait and Y. Ye. Stochastic system identification with noisy input-output measurement using polyspectra. *IEEE Transactions on Automatic Control*, AC-40:670–683, 1995.

[14] W. X. Zheng and C. B. Feng. Unbiased parameter estimation of linear systems in presence of input and output noise. *International Journal of Adaptive Control and Signal Processing*, 3:231–251, 1989.

[15] W. X. Zheng and C. B. Feng. Identification of a class of dynamic errors–in–variables models. *International Journal of Adaptive Control and Signal Processing*, 6:431–440, 1992.