APPROACHES FOR CONTINUOUS-TIME MODELING IN ERRORS-IN-VARIABLES IDENTIFICATION [1]

Torsten Söderström *, **Erik Larsson** *, **Kaushik Mahata** ** and **Magnus Mossberg** ***

* *Dept. of Information Technology, Uppsala University, P.O Box 337, SE 75105 Uppsala, Sweden. Email: Erik.K.Larsson@it.uu.se, Torsten.Soderstrom@it.uu.se*
** *CDSC,University of Newcastle, Callaghan, NSW 2308, Australia. Email: Kaushik.Mahata@newcastle.edu.au*
*** *Dept. of Electrical Engineering, Karlstad University SE-65188 Karlstad, Sweden. Email: Magnus.Mossberg@kau.se*

Abstract: Continuous-time identification is applied to an errors-in-variables setting. A continuous-time model is fitted to data consisting of discrete-time noise corrupted input and output measurements. The noise-free input is modelled as a continuous-time ARMA process. It is described how the Cramér-Rao lower bound for the estimation problem can be computed. Several parameter estimation approaches for the problem are presented, and also illustrated in a short numerical study.

Keywords: system identification, continuous-time, errors-in-variables

1. INTRODUCTION

Consider the SISO system depicted in Figure 1 with noise-corrupted input and output measurements.

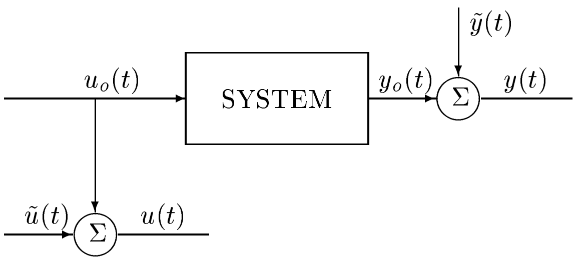

Fig. 1. The basic setup for a dynamic error-in-variables problem.

[1] **Research partially sponsored by the Swedish Research Council (contract 621-2002-4671) and the Australian Research Council.**

The noise–free input is denoted by $u_0(t)$ and the undisturbed output by $y_0(t)$. It is assumed that the input and the output signals are sampled uniformly at time instants $\{kh\}_{k=1}^{N}$. It is further assumed that the measurements are noise corrupted. The observed sampled data $\{u(kh)\}_{k=1}^{N}$ and $\{y(kh)\}_{k=1}^{N}$ are given by

$$\begin{aligned} u(kh) &= u_0(kh) + \tilde{u}(kh), \\ y(kh) &= y_0(kh) + \tilde{y}(kh), \end{aligned} \tag{1}$$

where $\{\tilde{u}(kh)\}_{k=1}^{N}$ and $\{\tilde{y}(kh)\}_{k=1}^{N}$ are zero mean, *discrete-time* white noise sequences with variances λ_u^2 and λ_y^2, respectively.

Normally it is not known how the noise-free input varies during the sampling intervals. Rather, it is natural to think of the unknown system as a continuous-time one, and that $u_0(t)$ varies continuously with time. Any discrete-time modeling will then be an approximation. For most identification approaches, though, one has to model also the noise-free input, not only the system. If $u_0(t)$

is assumed to be a stationary stochastic process with rational spectrum, then it can in fact equivalently be written as a continuous-time ARMA (CARMA) process:

$$D_c(p)u_0(t) = C_c(p)e_c(t), \tag{2}$$

where $e(t)_c$ is the time derivative of a Wiener process with unit incremental variance, $p = \frac{\mathrm{d}}{\mathrm{d}t}$, and

$$\begin{aligned} D_c(p) &= p^m + d_1 p^{m-1} + \cdots + d_{m-1}p + d_m, \\ C_c(p) &= c_1 p^{m-1} + \cdots + c_{m-1}p + c_m. \end{aligned} \tag{3}$$

Unless mentioned otherwise, the above assumption on the input process will be assumed throughout the paper.

The general task is to identify the system from the noise-corrupted data (1). We model the system as

$$A_c(p)y_0(t) = B_c(p)u_0(t). \tag{4}$$

The polynomials $A_c(p)$ and $B_c(p)$ are given by

$$\begin{aligned} A_c(p) &= p^n + a_1 p^{n-1} + \cdots + a_{n-1}p + a_n, \\ B_c(p) &= b_1 p^{n-1} + \cdots + b_{n-1}p + b_n. \end{aligned} \tag{5}$$

The relative degree of the system transfer function in (4) is assumed unity to keep notations simple, which can be relaxed if required.

The model under consideration is often referred to as the continuous-time errors-in-variables (EIV) model. Many methods have been proposed to solve the EIV problem in discrete-time. The popular approaches can be classified in a few broad categories (Söderström *et al.*, 2002); namely, the bias compensating least squares methods (Beghelli *et al.*, 1990; Zheng, 1998), prediction error and maximum likelihood methods (Söderström, 1981; Karlsson *et al.*, 2000), instrumental variable based approaches (Stoica *et al.*, 1995) and frequency-domain methods based on non-parametric noise modeling (Schoukens *et al.*, 1997). Unless we impose certain assumptions on the signal and noise models, it is well-known that the *general* EIV model is not uniquely identifiable from second order statistics (Anderson and Deistler, 1984).

Modeling of continuous-time EIV models, on the contrary, is a relatively unexplored area. Some results are described in (Sagara *et al.*, 1991; Mahata and Garnier, 2005). In this paper we explore several possible approaches to solve the problem. The results are also compared with the Cramér-Rao bound of the estimation problem (under Gaussian hypothesis).

2. BASIC ASSUMPTIONS AND NOTATIONS

In this section we introduce some basic assumptions and notations.

A1. The dynamic system (4) is asymptotically stable, *i.e.* $A_c(s)$ has all zeros in the left half plane. Further, all the system modes are observable and controllable, *i.e.* $A_c(s)$ and $B_c(s)$ have no common factors. Moreover, the order n of the system is *a priori* known.

A2. The polynomials $C_c(s)$ and $D_c(s)$ are coprime and have all zeros in the left half plane.

A3. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually uncorrelated, and uncorrelated with the noise–free signals $u_0(t)$ and $y_0(t)$.

A4. The processes $\tilde{u}(kh)$, $\tilde{y}(kh)$ and $e_c(t)$ are Gaussian distributed.

The parameter vector θ describes the system transfer function $B_c(s)/A_c(s)$:

$$\theta = \begin{bmatrix} a_1 & \ldots & a_n & b_1 & \ldots & b_n \end{bmatrix}^\top. \tag{6}$$

For future use we also introduce the extended parameter vector $\overline{\theta}$ as

$$\overline{\theta} = \begin{bmatrix} 1 & \theta^\top \end{bmatrix}^\top. \tag{7}$$

Another parameter vector, ψ, is used to describe the input model (2):

$$\psi = \begin{bmatrix} c_1 & \ldots & c_m & d_1 & \ldots & d_m \end{bmatrix}^\top. \tag{8}$$

Finally, we will use an augmented parameter vector ϑ which includes the unknown quantities:

$$\vartheta = \begin{bmatrix} \theta^\top & \psi^\top & \lambda_y^2 & \lambda_u^2 \end{bmatrix}^\top. \tag{9}$$

3. A STATE SPACE MODEL AND THE CRAMÉR-RAO LOWER BOUND

The Cramér-Rao lower bound (CRB) gives a lower bound on the covariance matrix of the parameter estimates for any unbiased estimate, see, for example (Söderström and Stoica, 1989). When applied to identification of dynamic systems this bound holds asymptotically for large data sets.

There are several ways to compute the CRB for an EIV problem, using, for example, a polynomial approach for a discrete-time model, see (Karlsson *et al.*, 2000), or using a general state-space model (with continuous or discrete time) with discrete-time measurements, see (Söderström, 2006). In both cases one has to include also a model of the noise-free input, and the CARMA model for $u_0(t)$ as in (2) is applied here. The state space approach for computing CRB is used here.

As a first step the noise-free system is written in state space form with the 2-dimensional vector $z_0(t) = \begin{pmatrix} y_0(t) & u_0(t) \end{pmatrix}^\top$ as output and the driving noise $e_c(t)$ as input. Such a model can for short be written as

$$\begin{aligned} \dot{x}(t) &= F_c(\theta, \psi)x(t) + G_c(\theta, \psi)e_c(t), \\ z_0(t) &= H_c(\theta, \psi)x(t). \end{aligned} \tag{10}$$

More details about the model (10) are given in Appendix A.

The sampling of the state-space model (10) with constant sampling interval h is uniquely given, see (Åström, 1970), (Söderström, 2002). The sampled state space model is written here as

$$\begin{aligned} x(kh+h) &= F_d(\theta,\psi)x(kh)+v(kh), \\ z_0(kh) &= H_d(\theta,\psi)x(kh), \end{aligned} \tag{11}$$

where $H_d = H_c$, $F_d = \mathrm{e}^{F_c h}$ and $v(kh)$ is a discrete-time white noise sequence with zero mean and covariance matrix

$$R_1(\theta,\psi) = \int_0^h \mathrm{e}^{F_c t} G_c G_c^\top \mathrm{e}^{F_c^\top t} \mathrm{d}t. \tag{12}$$

Finally, by adding the white measurement noise

$$e(kh) = \left[\, \tilde{y}(kh) \;\; \tilde{u}(kh) \,\right]^\top, \tag{13}$$

as in (1), one arrives at the state space model

$$\begin{aligned} x(kh+h) &= F_d(\vartheta)x(kh)+v(kh), \\ z(kh) &= H_d(\vartheta)x(kh)+e(kh), \end{aligned} \tag{14}$$

where the noise sequences $v(kh)$ and $e(kh)$ are white, zero mean, and with covariance matrices

$$\mathrm{E}\begin{bmatrix} v(kh) \\ e(kh) \end{bmatrix}\begin{bmatrix} v(kh) \\ e(kh) \end{bmatrix}^\top = \begin{bmatrix} R_1(\vartheta) & 0 \\ 0 & R_2(\vartheta) \end{bmatrix}. \tag{15}$$

For details how the CRB can be computed once the general state-space model (14), (15) is set up, we refer to (Söderström, 2006).

4. ESTIMATION ALGORITHMS

There are many ways to estimate the parameters of the system under study. We will first describe how maximum likelihood and prediction error estimates work for the treated problem. They both lead to rather complex optimization problems, that have some potential to cause numerical troubles, lack of finding the global optimum, etc. It is therefore of interest to examine also other schemes, even if these are a bit ad-hoc in nature. We describe some different approaches in the subsections to follow.

4.1 *Prediction error and maximum likelihood methods*

Both prediction error methods (PEM) and the maximum likelihood (ML) method are based on the state space model (14), (15). For both of these two methods, the optimal one-step prediction errors (which are the same as the innovations sequence of the state space model) have to be found. This is directly done by first finding the Kalman gain K by solving the usual algebraic Riccati equation

$$\begin{aligned} P &= F_d P F_d^\top + R_1 - F_d P H_d^\top \\ &\quad \times (H_d P H_d^\top + R_2)^{-1} H_d P F_d^\top, \end{aligned} \tag{16}$$

$$K = F_d P H_d^\top \Lambda^{-1}, \tag{17}$$

$$\Lambda = H_d P H_d^\top + R_2. \tag{18}$$

Note that the solution P, the Kalman gain K and the innovations covariance matrix Λ all depend on the augmented parameter vector ϑ although this dependence is not shown explicitly in (16) - (18) for convenience.

The prediction errors $\varepsilon(kh,\vartheta)$, $k = 1,\ldots N$ are then obtained by inverting the Kalman filter. In state space form, the filtering from measured data $\{z(kh)\}$ to the prediction errors is given by

$$\begin{aligned} \hat{x}(kh+h) &= F_d\hat{x}(kh) + K[z(kh) - H_d\hat{x}(kh)], \\ \varepsilon(kh,\vartheta) &= z(kh) - H_d\hat{x}(kh). \end{aligned} \tag{19}$$

The prediction error estimate is determined by minimizing a function $V_{\mathrm{PEM}}(\vartheta)$ that penalizes the prediction errors. The standard choice, as for example implemented in the Matlab System Identification Toolbox, is

$$V_{\mathrm{PEM}}(\vartheta) = \det\left(\frac{1}{N}\sum_{k=1}^{N}\varepsilon(kh,\vartheta)\varepsilon^\top(kh,\vartheta)\right). \tag{20}$$

Another possibility is to take

$$V_{\mathrm{PEM}}(\vartheta) = \frac{1}{N}\sum_{k=1}^{N}\varepsilon^\top(kh,\vartheta)Q\varepsilon(kh,\vartheta), \tag{21}$$

where Q is a user-chosen weighting matrix.

For the prediction error method one has in fact to modify the augmented parameter vector ϑ in an intermediate phase. The reason is that the prediction errors are determined uniquely from the Kalman gain $K(\vartheta)$ and that this gain is unaffected by scaling all noise intensities and variances (R_1 and R_2) by any constant amount. The fix is to set one of the noise parameters to a constant value during the optimization, and then afterwards to utilize also the information in the minimal value of the criterion to find the appropriate scaling factor, see (Söderström, 2006) for details.

The maximum likelihood estimate is minimizing the negative loglikelihood function, which can be shown for Gaussian distributed data to be equivalent to minimizing the criterion

$$\begin{aligned} V_{\mathrm{ML}}(\vartheta) &= \frac{1}{N}\sum_{k=1}^{N}\varepsilon^\top(kh,\vartheta)\Lambda^{-1}(\vartheta)\varepsilon(kh,\vartheta) \\ &\quad + \log\det\Lambda(\vartheta). \end{aligned} \tag{22}$$

See (Ljung, 1999), (Söderström and Stoica, 1989). Note that both the prediction errors $\varepsilon(kh,\vartheta)$ and the innovation covariance matrix $\Lambda(\vartheta)$ depend on the augmented parameter vector ϑ, in a coupled and rather complex way. It is not possible here to minimize the criterion V_{ML} first with respect to Λ only, as is the case when Λ consists of independent parameters, (Ljung, 1999), (Söderström and Stoica, 1989).

Both the prediction error estimate, $\hat{\vartheta}_{\mathrm{PEM}}$, and the maximum likelihood estimate, $\hat{\vartheta}_{\mathrm{ML}}$, will be asymptotically Gaussian distributed for large data sets (*i.e.* when $N \to \infty$). The maximum likelihood estimate will asymptotically have a covariance matrix that achieves the CRB,

$$\mathrm{cov}(\hat{\vartheta}_{\mathrm{ML}}) = \mathrm{CRB} \tag{23}$$

while the prediction error estimate will give a less accurate estimate in the sence

$$\mathrm{cov}(\hat{\vartheta}_{\mathrm{PEM}}) \geq \mathrm{CRB}. \tag{24}$$

The reason equality does not hold in (24) is in fact again the coupling between ε and Λ for the particular estimation problem under study.

Both the prediction error estimate and the maximum likelihood estimate are obtained from the solution of a nonlinear optimization problem. This implies that there are inherent risks for troubles in form of existence of local minima, divergence of the numerical search algorithms, etc. In fact, our numerical experimentations indicate that the numerical search may sometimes fail, in particular for low signal-to-noise ratios. (Typically one estimated noise variance, such as the estimated output noise variance, $\hat{\lambda}_y^2$ diverges and become very large during the numerical search). For this reason, it is of interest to also examine other estimation methods that are simpler computationally, even if they are more ad-hoc from a statistical point of view.

4.2 *Covariance matching method*

The covariance function of $z_0(t)$ in (10) is given by (Söderström, 2002)

$$\begin{aligned} R(\tau,\theta,\psi) &= \begin{bmatrix} R_{yy}(\tau,\theta,\psi) & R_{yu}(\tau,\theta,\psi) \\ R_{uy}(\tau,\theta,\psi) & R_{uu}(\tau,\psi) \end{bmatrix} \\ &= \mathrm{E}\{z_0(t+\tau)z_0^\top(t)\} = H_c e^{F_c\tau} P_0 H_c^T, \end{aligned} \tag{25}$$

for $\tau \geqslant 0$, where P_0 is the unique and nonnegative definite solution to the continuous-time Lyapunov equation

$$F_c P_0 + P_0 F_c^\top + G_c G_c^\top = 0. \tag{26}$$

In (25) the dependence on the parameter vectors θ, (6), and ψ, (8), that contain all continuous-time parameters, is emphasized. Note that $R_{uu}(\tau)$ depends only on ψ, not on θ, cf Appendix A.

If the mean values of $y(kh)$ and $u(kh)$ are zero, a possible estimator of $R(jh,\theta,\psi)$ is

$$\begin{aligned} \hat{R}(jh) &= \begin{bmatrix} \hat{R}_{yy}(jh) & \hat{R}_{yu}(jh) \\ \hat{R}_{uy}(jh) & \hat{R}_{uu}(jh) \end{bmatrix} \\ &= \frac{1}{N-j}\sum_{k=1}^{N-j} z(kh+jh)z^\top(kh), \end{aligned} \tag{27}$$

where j is a nonnegative integer. Under the given assumptions, the data are stationary and ergodic, so $\hat{R}(jh) \to R(jh)$ when $N \to \infty$ if $j > 0$ and h is fixed. When $j = 0$ the diagonal elements will also contain contributions due to the noise variances.

Following an approach similar to the one in (Mossberg, 2005), assume that $\hat{R}(jh)$ is available for $j = 0, \ldots, M$, and define the loss function

$$V_u(\psi) = \sum_{j=1}^{J} \left\{ \hat{R}_{uu}(jh) - R_{uu}(jh,\psi) \right\}^2, \tag{28}$$

where $J \leqslant M$, from which an estimate $\hat{\psi}$ is obtained as

$$\hat{\psi} = \arg\min_{\psi} V_u(\psi). \tag{29}$$

As a second step, consider the loss function

$$\begin{aligned} V_z(\theta) = &\sum_{j=0}^{K} \left\{ (\hat{R}_{yu}(jh) - R_{yu}(jh,\theta,\hat{\psi}) \right\})^2 \\ &+ \sum_{j=1}^{L} \left\{ \hat{R}_{yy}(jh) - R_{yy}(jh,\theta,\hat{\psi}) \right\}^2, \end{aligned} \tag{30}$$

where $K, L \leqslant M$, from which an estimate $\hat{\theta}$ is given as

$$\hat{\theta} = \arg\min_{\theta} V_z(\theta). \tag{31}$$

4.3 *Parallel input-output modeling*

Let us sample $u_0(t)$ and $y_0(t)$ as two independent processes. This is obviously a simplification; or rather, we discard some of the available information. It turns out that $u_0(kh)$ is an (discrete-time) ARMA(m,$m-1$) process, and $y_0(kh)$ is an (discrete-time) ARMA($n+m$,$n+m-1$) process:

$$D_d(q)u_0(kh) = C_d(q)e_1(kh), \tag{32}$$

$$A_d(q)D_d(q)y_0(kh) = B_d(q)e_2(kh), \tag{33}$$

where according to the assumption $e_1(kh)$ and $e_2(kh)$ are independent zero mean white noise sources with variances λ_1^2 and λ_2^2, respectively. Here and in the following q denotes the forward shift operator, $qx(kh) = x(kh+h)$.

From the relation (1) it follows that $u(kh)$ is an ARMA(m,m) process, and $y(kh)$ is an ARMA($n+m$,$n+m$) process:

$$D_d(q)u(kh) = C(q)v_1(kh), \quad (34)$$
$$A_d(q)D_d(q)y(kh) = B(q)v_2(kh), \quad (35)$$

where $v_1(kh)$ and $v_2(kh)$ are zero mean white noise sources with variances $\lambda_{v_1}^2$ and $\lambda_{v_2}^2$, respectively. Note that the spectra of $u_0(kh)$ and $y_0(kh)$ are strictly proper, while the spectra of $u(kh)$ and $y(kh)$ are not. The spectra are related as (32) and (33) to (34) and (35):

$$\phi_u^{(d)}(z) = \phi_{u_0}^{(d)}(z) + \lambda_u^2, \quad (36)$$
$$\phi_y^{(d)}(z) = \phi_{y_0}^{(d)}(z) + \lambda_y^2, \quad (37)$$

where

$$\phi_u^{(d)}(z) = \lambda_{v_1}^2 \frac{C(z)C(z^{-1})}{D_d(z)D_d(z^{-1})}, \quad (38)$$
$$\phi_{u_0}^{(d)}(z) = \lambda_1^2 \frac{C_d(z)C_d(z^{-1})}{D_d(z)D_d(z^{-1})}, \quad (39)$$
$$\phi_y^{(d)}(z) = \frac{\lambda_{v_2}^2 B(z)B(z^{-1})}{D_d(z)D_d(z^{-1})A_d(z)A_d(z^{-1})}, \quad (40)$$
$$\phi_{y_0}^{(d)}(z) = \frac{\lambda_2^2 B_d(z)B_d(z^{-1})}{D_d(z)D_d(z^{-1})A_d(z)A_d(z^{-1})}. \quad (41)$$

Next we develop a three step procedure for estimating ϑ.

Step 1

a) Estimate $D_d(q)$, $C(q)$ and $\lambda_{v_1}^2$ from the model (34) using measurements of $u(kh)$.
b) Let $y_F(kh) = \hat{D}_d(q)y(kh)$ and estimate $A_d(q)$, $B(q)$ and $\lambda_{v_2}^2$ from the model

$$A_d(q)y_F(kh) = B(q)v_2(kh). \quad (42)$$

Remark. Both these parts can be performed using the ARMAX routine in MATLAB. One of many alternatives is to use a Durbin scheme instead.

Remark. It is important that the estimation procedure takes into account that the processes $u(kh)$ and $y(kh)$ have common poles. This can be done in several ways, and here we do this by using the filtered measurements $y_F(kh)$.

Step 2 Estimate the measurement noise variances λ_y^2 and λ_u^2, and the spectra $\phi_{u_0}^{(d)}(z)$ and $\phi_{y_0}^{(d)}(z)$. A simple approach is to explore (36) and (37). For instance, equating equal powers of z in (36) immediately yields an expression for $\hat{\lambda}_u^2$. Then $\hat{\phi}_{u_0}^{(d)}(z) = \hat{\phi}_u^{(d)}(z) - \hat{\lambda}_u^2$. Similarly, $\hat{\phi}_{y_0}^{(d)}(z)$ and $\hat{\lambda}_y^2$ can be found using (37).

Remark. It is not guaranteed that these estimates will be positive. This can be cured, when needed, in several ways. See, for example, (Dumitrescu *et al.*, 2001).

Step 3 The final step is to convert the estimated discrete-time system parameters into the desired continuous-time system parameters. This type of problem appears in various branches of system science, and both exact and approximate algorithms exist, see (Söderström, 1991) and the references therein. We will use a method described in (Söderström, 1991), which is outlined in Appendix B for convenience.

4.4 *Shinbrot moment functionals*

The approach in this section does not need $u_0(t)$ to be a CARMA process. Here we consider the n-th order Fourier type Shinbort modulation function (Pearson and Lee, 1985; Pearson and Shen, 1993)

$$\phi_\ell(t) = e^{-i\omega_0 \ell t}\left(e^{-i\omega_0 t} - 1\right)^n = \sum_{k=0}^{n} c_k e^{-i\omega_0(\ell+k)t}, \quad (43)$$

where $\omega_0 = 2\pi/(Nh)$, and

$$c_k = (-1)^{n-k}\frac{n!}{k!(n-k)!}. \quad (44)$$

The property of $\phi_\ell(t)$ which makes it interesting in continuous-time model identification is

$$p^k\phi_\ell(t)\big|_{t=0} = 0, \qquad p^k\phi_\ell(t)\big|_{t=Nh} = 0 \quad (45)$$

for any $\ell \geq 0$, and $k \in \{0,1,\ldots,n\}$. Consequently, integrating by parts it is readily verified that the Shinbrot moment of $u(t)$ of order k with respect to the modulating function $\phi_\ell(t)$ is given by

$$\beta_u^{(k)}(\ell) := \int_0^{Nh} p^k u(t)\cdot\phi_\ell(t)\ \mathrm{d}t$$
$$= (-1)^k \int_0^{Nh} p^k\phi_\ell(t)\cdot u(t)\ \mathrm{d}t. \quad (46)$$

Similarly we define the Shinbrot moments $\beta_y^{(k)}(\ell)$ of the output $y(t)$. Approximating the last equality by the Riemannian summation, one can compute the Shinbrot moments very efficiently from the discrete Fourier transforms of the input-output data. The results are exact if the data are band-limited by $1/2h$ Hz. Let us define

$$\boldsymbol{\beta}(\ell) = [\,\beta_y^{(n)}(\ell)\ \cdots\ \beta_y^{(0)}(\ell)\ \beta_u^{(n-1)}(\ell)\ \cdots\ \beta_u^{(0)}(\ell)\,].$$
$$\bar{\boldsymbol{\beta}}(\ell) = [\,\beta_y^{(n)}(\ell)\ \cdots\ \beta_y^{(0)}(\ell)\ \beta_u^{(n)}(\ell)\ \cdots\ \beta_u^{(0)}(\ell)\,].$$

Let $\boldsymbol{\alpha} = [\,\mathrm{Re}\{\boldsymbol{\beta}^\top\}\ \mathrm{Im}\{\boldsymbol{\beta}^\top\}\,]^\top$. We define $\bar{\boldsymbol{\alpha}}$ similarly. From (4) and (6) we get

$$\boldsymbol{\alpha}_0(\ell)\overline{\theta} = 0, \qquad \ell \in \{0,1,\ldots,L-1\}, \quad (47)$$

where $\boldsymbol{\alpha}_0$ is the noise-free part of $\boldsymbol{\alpha}$. However, in an EIV scenario, (47) cannot be used directly since $\boldsymbol{\beta}_0$ is unknown. Nevertheless, for large enough L it holds that (see Section C.2 for a proof)

$$R := \frac{1}{L}\sum_{\ell=0}^{L-1}\bar{\boldsymbol{\alpha}}^\top(\ell)\boldsymbol{\alpha}(\ell) = \frac{1}{L}\sum_{\ell=0}^{L-1}\bar{\boldsymbol{\alpha}}_0^\top(\ell)\boldsymbol{\alpha}_0(\ell)$$
$$+\tilde{R}\mathrm{diag}\{\,\lambda 2_y I_{n+1},\ \lambda_u 2 I_n\,\}, \quad (48)$$

$$\Rightarrow R\theta = \tilde{R}\,\mathrm{diag}\{\,\lambda_u^2 I_{n+1},\ \lambda_y 2 I_n\}\,\theta,$$

where $\tilde{R}$ can be computed. Therefore, the bilinear least squares approach in (Ekman, 2005) can be used to derive consistent (as $N \to \infty$ for fixed h and L/N) estimates of θ. Next, we give the expression for $\tilde{R}$ (see Section C.2 for a proof). It turns out that

$$\tilde{R} = \begin{bmatrix} \bar{R} & 0 \\ 0 & \bar{R}(:, 2:n+1) \end{bmatrix},$$

where we use Matlab notation, and $\bar{R}$ is a $(n+1)\times(n+1)$ matrix such that $[\bar{R}]_{ij}$ is non-zero only if $i+j$ is even. Let $i+j = 2l$, then

$$[\bar{R}]_{ij} = (-1)^{i-l} \sum_{k=0}^{n} c_k^2 \sum_{\ell=k}^{L+k-1} \left(\frac{2\pi\ell}{Nh}\right)^{2(n-l+1)}. \tag{49}$$

Further comments on the implementation aspects of the Shinbrot functional method are given in Section C.1.

4.5 *Transformation from a discrete-time model*

A rather straighforward approach for estimating the transfer function of the system would be to first cast the estimation problem as a purely discrete-time one. Then any 'standard' method for estimating the discrete-time transfer function for an EIV problem may be applied, for examples see (Söderström *et al.*, 2002). As a second step the estimated discrete-time transfer function is converted to continuous-time by some method. This can be done in several ways, where assuming zero-order-hold sampling is just one possibility. Numerical problems may occur if the sampling interval h is very small in comparison to the time constants of the system.

5. NUMERICAL ILLUSTRATION

5.1 *Comparison of the methods*

The properties of the introduced methods are illustrated numerically in this section. None of the methods worked always well. The PEM (M1) had convergence problems in some cases, but gave very accurate estimates with variance close to CRB otherwise. The covariance matching method (M2) did not work well for some type of input signals. The method of parallel input-output modeling (M3) have sometimes problems in Step 1. The Shinbrot moment functional method (M4) showed best robustness in our preliminary tests. In this study we focussed on the estimation of the vector θ.

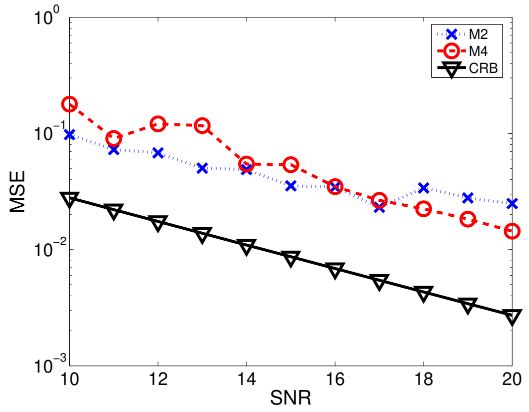

Fig. 2. The MSE for the estimates of θ by M2 and M4 for system (50), and the CRB as functions of the SNR.

The systems

$$\begin{aligned} &A_c(p) = p^2 + 2p + 2, \quad B_c(p) = 2p + 2, \\ &C_c(p) = 5, \quad D_c(p) = p^3 + 2p^2 + 2p + 2 \end{aligned} \tag{50}$$

and

$$\begin{aligned} &A_c(p) = p^2 + 2p + 2, \quad B_c(p) = p + 5, \\ &C_c(p) = 2, \quad D_c(p) = p + 2 \end{aligned} \tag{51}$$

are considered, and the variances λ_u^2 and λ_y^2 are chosen such that a certain signal-to-noise ratio (SNR) is fulfilled. Here we have $\mathrm{SNR} = \mathrm{SNR}_u = \mathrm{SNR}_y$, defined as $\mathrm{SNR}_u = \mathrm{var}\{u_0(t)\}/\lambda_u^2$ and $\mathrm{SNR}_y = \mathrm{var}\{y_0(t)\}/\lambda_y^2$.

Monte Carlo simulations with 50 realizations are carried out for the estimation of θ, *i.e.* the parameters in the polynomials $A_c(p)$ and $B_c(p)$ for systems (50) and (51), respectively, with $h = 0.2$, $N = 10000$, and with SNR varying between 10 and 20 dB. If M3 occasionally failed, a new realization data set was substituted and used for all methods. In this initial study, the user parameters J in (28), and K and L in (30) were chosen as $J = 30$, $K = 0$ and $L = 30$.

The mean square error (MSE)

$$\mathrm{MSE} = \frac{1}{50}\sum_{j=1}^{50} \|\,\hat{\theta}(j) - \theta\,\|^2,$$

$\hat{\theta}(j)$ being the estimate of θ from realization j, of the estimates given by methods M2 and M4 for system (50), and methods M1, M3 and M4 for system (51) are shown in Figures 2 and 3, respectively, together with the trace of the CRB matrix, as functions of the SNR.

5.2 *Illustration of the parallel input-output approach*

In this section we give some further examples illustrating the properites of the parallel input-output approach.

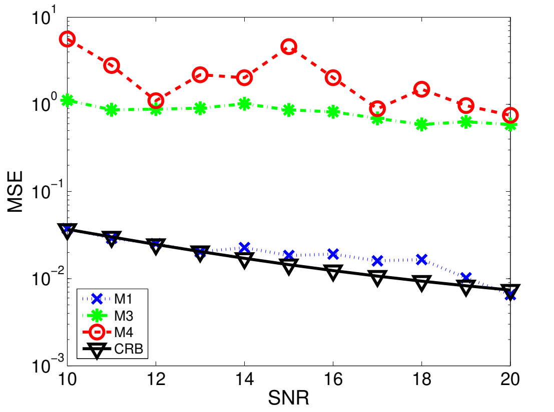

Fig. 3. The MSE for the estimates of θ by M1, M3 and M4 for system (51), and the CRB as functions of the SNR.

Process 1:

$$A_c(s) = s + 2 \quad B_c(s) = s + 5$$
$$C_c(s) = 2 \quad D_c(s) = s + 0.5$$

Process 2:

$$A_c(s) = s^2 + 2s + 2 \quad B_c(s) = s + 5$$
$$C_c(s) = 2 \quad D_c(s) = s + 2$$

The variances λ_u^2 and λ_y^2 are chosen such that a certain SNR is fulfilled. Here we have SNR $=$ SNR$_u =$ SNR$_y$.

The criteria selected for the performance evaluation are the normalized mean estimation error

$$\text{NMEE} \triangleq \frac{\|\vartheta - \hat{\vartheta}^*\|_2}{\|\vartheta\|_2} \tag{52}$$

where

$$\hat{\vartheta}^* = \frac{1}{M_c}\sum_{k=1}^{M_c} \hat{\vartheta}(k) \tag{53}$$

and the mean squared error

$$\text{MSE} \triangleq \frac{1}{M_c}\sum_{k=1}^{M_c} \|\hat{\vartheta}(k) - \vartheta\|_2^2, \tag{54}$$

where $\hat{\vartheta}(k)$ corresponds to the k:th estimate of the $M_c = 100$ number of Monte Carlo runs performed.

Furthermore, we show the quotient between the MSE and the CRB.

Remark. Realizations that yields strange results (non positive spectra etc) are discarded.

The results for Process 1 (sys1) are presented in Figures 4-6, whereas the results for Process 2 (sys2) are found in Figures 7-9.

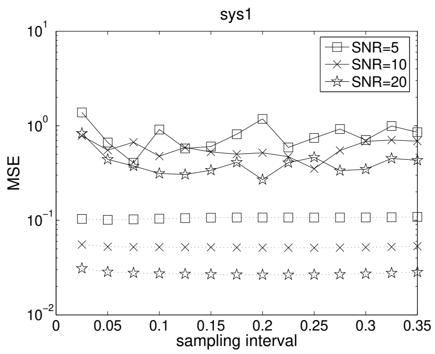

Fig. 4. The MSE and the CRB for estimating ϑ as functions of the sampling interval. Solid lines show estimates, whereas the dotted lines represent the CRB.

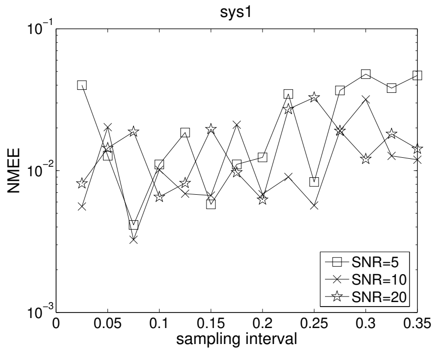

Fig. 5. The NMEE for estimating of ϑ as a function of the sampling interval.

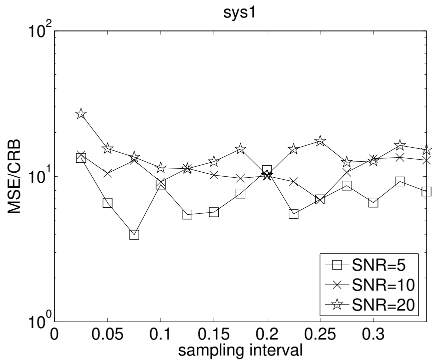

Fig. 6. The quotient between the MSE and the CRB.

6. CONCLUSIONS

The problem of fitting a continuous-time model to discrete-time input and output data corrupted by noise has been studied. The prediction error method, the covariance matching method, the method of parallel input-output modeling, and the Shinbrot moment functionals method have been presented for solving the problem. Initial simulation studies suggest that none of the methods is uniformly best.

7. REFERENCES

Anderson, B. D. O. and M. Deistler (1984). Identifiability of dynamic errors-in-variables models. *J. Time Series Analysis* **5**, 1–13.

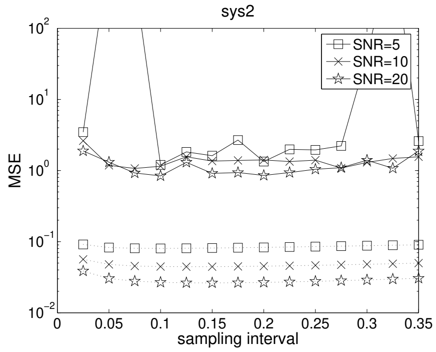

Fig. 7. The MSE and the CRB for estimating ϑ as functions of the sampling interval. Solid lines show estimates, whereas the dotted lines represent the CRB.

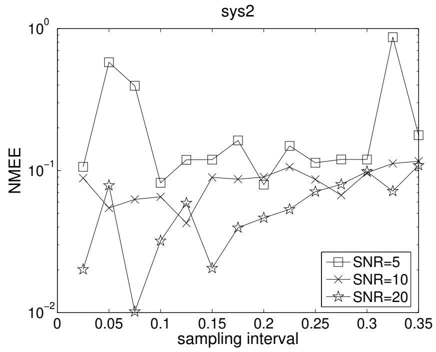

Fig. 8. The NMEE for estimating ϑ as a function of the sampling interval.

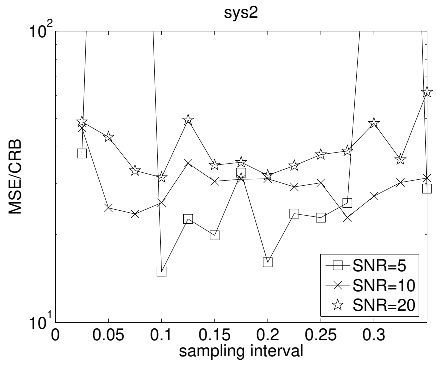

Fig. 9. The quotient between the MSE and the CRB.

Åström, K. J. (1970). *Introduction to Stochastic Control*. Academic Press. New York, NY.

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

Dumitrescu, B., I. Taăbuş and P. Stoica (2001). On the parameterization of positive real sequences and MA parameter estimation. *IEEE Transactions on Signal Processing* **49**, 2630–2639.

Ekman, M. (2005). Identification of linear systems with errors in variables using separable nonlinear least squares. In: *16th IFAC World Congress*. Prague, Czech Republic.

Karlsson, E, T Söderström and P Stoica (2000). Computing the Cramér-Rao lower bound for noisy input output systems. *Signal Processing* **80**(11), 2421–2447.

Ljung, L. (1999). *System Identification - Theory for the User, 2nd edition*. Prentice Hall. Upper Saddle River, NJ, USA.

Mahata, K. and H. Garnier (2005). Direct identification of continuous-time errors-in-variables models. In: *16th IFAC World Congress*. Prague, Czech Republic.

Mossberg, M. (2005). Identification of continuous-time ARX models using sample covariances. In: *American Control Conference*. Portland, OR. pp. 4766–4771.

Pearson, A. E. and F. C. Lee (1985). On the identification of polynomial input-output differential systems. *IEEE Transactions on Automatic Control* **AC-30**, 778–782.

Pearson, A. E. and Y. Shen (1993). Weighted least squares/MFT algorithms for linear differential system identification. In: *32nd IEEE confererence on Decision and Control*. San Antonio, Texas. pp. 2032–2037.

Sagara, S., Z.J. Yang and K. Wada (1991). Identification of continuous systems from noisy sampled input-output data. In: *IFAC Symposium on system identification and parameter estimation*. Budapest, Hungary. pp. 1622–1627.

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

Söderström, T. (1981). Identification of stochastic linear systems in presence of input noise. *Automatica* **17**, 713–725.

Söderström, T (1991). “Computing stochastic continuous-time models from ARMA models”. *International Journal of Control* **53**, 1311–1326.

Söderström, T (2002). *Discrete-Time Stochastic Systems - Estimation and Control, 2nd edition*. Springer-Verlag. London, UK.

Söderström, T. (2006). On computing the Cramer-Rao bound and covariance matrices for PEM estimates in linear state space models. In: *IFAC Symposium on System Identification*. Newcastle, Australia. Submitted.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice Hall International. Hemel Hempstead, UK.

Söderström, T., U. Soverini and K. Mahata (2002). Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing* **82**(8), 1139–1154.

Stoica, P., M. Cedervall and A. Eriksson (1995). Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems. *IEEE Transactions on Signal Processing* **43**, 2386–2397.

Zheng, W. X. (1998). Transfer function estimation form noisy input and output data. *International Journal of Adaptive Control and Signal Processing* **12**, 365–380.

Appendix A. DETAILS ABOUT THE STATE SPACE MODEL

The matrices in (10) can be written as $F_c \in \mathbb{R}^{\eta\times\eta}$, $G_c \in \mathbb{R}^{\eta\times 1}$ and $H_c \in \mathbb{R}^{2\times\eta}$ given as

$$F_c = \begin{bmatrix} F_{11} & F_{12} \\ \mathbf{0}^{m\times n} & F_{22} \end{bmatrix}, \quad G_c = \begin{bmatrix} \mathbf{0}^{n\times 1} \\ G_2 \end{bmatrix},$$

$$H_c = \begin{bmatrix} H_{11} & \mathbf{0}^{1\times m} \\ \mathbf{0}^{1\times n} & H_{22} \end{bmatrix},$$

where $\eta \triangleq n+m$ is introduced for notational convenience, $\mathbf{0}^{\ell_1\times\ell_2}$ denotes the zero matrix of dimension $\ell_1\times\ell_2$, and $F_{11}\in\mathbb{R}^{m\times m}$, $F_{21}\in\mathbb{R}^{n\times m}$, $F_{22}\in\mathbb{R}^{n\times n}$, $G_2\in\mathbb{R}^{m\times 1}$, $H_{11}\in\mathbb{R}^{1\times m}$ and $H_{22}\in\mathbb{R}^{1\times n}$ equal

$$F_{11} = \begin{bmatrix} -a_1 & 1 & & \\ \vdots & & \ddots & \\ \vdots & & & 1 \\ -a_n & & & \end{bmatrix}, F_{12} = \begin{bmatrix} b_1 & \\ \vdots & \mathbf{0}^{n\times(m-1)} \\ b_n & \end{bmatrix},$$

$$F_{22} = \begin{bmatrix} -d_1 & 1 & & \\ \vdots & & \ddots & \\ \vdots & & & 1 \\ -d_m & & & \end{bmatrix}, \quad G_2 = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix},$$

$$H_{11} = \begin{bmatrix} 1 & \mathbf{0}^{1\times(n-1)} \end{bmatrix}, \quad H_{22} = \begin{bmatrix} 1 & \mathbf{0}^{1\times(m-1)} \end{bmatrix}.$$

Note that a state space representation for only $u_0(t)$ is given as

$$\begin{aligned} \dot{\xi}(t) &= F_{22}\xi(t) + G_2 e_c(t), \\ u_0(t) &= H_{22}\xi(t). \end{aligned} \tag{A.1}$$

Remark. Note that the covariance function of $u_0(t)$ can be expressed as

$$\begin{aligned} R_{uu}(\tau,\psi) &= \mathrm{E}\{u_0(t+\tau)u_0^\top(t)\} \\ &= H_{22}e^{F_{22}\tau}QH_{22}^\top, \end{aligned} \tag{A.2}$$

for $\tau \geqslant 0$, where Q is the unique and nonnegative definite solution to the continuous-time Lyapunov equation

$$F_{22}Q + QF_{22}^\top + G_2G_2^\top = 0. \tag{A.3}$$

Appendix B. CONVERSION OF A STOCHASTIC DISCRETE-TIME MODEL TO CONTINUOUS-TIME

Consider the process u_0 which here is described in continuous-time by (4) and in discrete-time by (32). The mapping between the continuous-time and discrete-time poles is straightforward, and is given by the well-known relation, see *e.g.*, (Ljung, 1999),

$$\kappa_i = \mathrm{e}^{\rho_i h}, \quad i = 1,\ldots,n, \tag{B.1}$$

where ρ_i is a continuous-time pole and where κ_i is the corresponding discrete-time pole. This relation gives $\hat{D}_c(s)$ from $\hat{D}_d(q)$.

The relation between the continuous-time and discrete-time zeros is more difficult. One alternative is to proceed as follows:

- Introduce the coefficients $\{\gamma_j\}_{j=0}^{m-1}$ by

$$C_c(\mathrm{i}\omega)C_c(-\mathrm{i}\omega) \equiv \gamma_0+\gamma_1\omega^2+\ldots+\gamma_{m-1}\omega^{2(m-1)}. \tag{B.2}$$

Next define the processes

$$D_c(p)u_0^{(j)}(t) = p^j e_c^{(j)}(t), \quad j = 0,\ldots,m-1, \tag{B.3}$$

where $\mathrm{E}\{e_c^{(k)}(t)e_c^{(\ell)}(s)\} = \delta(t-s)\delta_{k,\ell}$. The spectral densities of $u_0^{(j)}(t)$ can be added together to give the spectral density of $u_0(t)$ as

$$\phi_{u_0}^{(c)}(\omega) = \frac{C_c(\mathrm{i}\omega)C_c(-\mathrm{i}\omega)}{D_c(\mathrm{i}\omega)D_c(-\mathrm{i}\omega)} = \sum_{j=0}^{m-1}\frac{\gamma_j\omega^{2j}}{D_c(\mathrm{i}\omega)D_c(-\mathrm{i}\omega)}. \tag{B.4}$$

Represent the systems (B.3) in controllable canonical state space forms as

$$\dot{x}(t) = \boldsymbol{F}_c x(t) + \boldsymbol{G}_c e_c^{(j)}(t), \tag{B.5}$$

$$u_0^{(j)}(t) = \boldsymbol{H}^{(j)}x(t), \tag{B.6}$$

which after sampling (Åström, 1970; Söderström *et al.*, 2002) gives

$$x(kh+h) = \boldsymbol{F}_d x(kh) + \boldsymbol{w}(kh), \tag{B.7}$$

$$u_0^{(j)}(kh) = \boldsymbol{H}^{(j)}x(kh), \tag{B.8}$$

where

$$\boldsymbol{F}_d = \mathrm{e}^{\boldsymbol{F}_c h}, \tag{B.9}$$

$$\mathrm{E}\{\boldsymbol{w}(kh)\boldsymbol{w}^\top(kh)\} = \int_0^h \mathrm{e}^{\boldsymbol{F}_c t}\boldsymbol{G}_c\boldsymbol{G}_c^\top \mathrm{e}^{\boldsymbol{F}_c^\top t}\mathrm{d}t \triangleq \boldsymbol{R}_d. \tag{B.10}$$

A sampled form of (B.4) can now be given as

$$\begin{aligned} \phi_{u_0}^{(d)}(\omega) &\equiv \sum_{j=0}^{m-1}\gamma_j\{\boldsymbol{H}^{(j)}(\mathrm{e}^{\mathrm{i}\omega h}I - \boldsymbol{F}_d)^{-1} \\ &\quad \times \boldsymbol{R}_d(\mathrm{e}^{-\mathrm{i}\omega h}I - \boldsymbol{F}_d^\top)^{-1}(\boldsymbol{H}^{(j)})^\top\} \\ &\triangleq \sum_{j=0}^{m-1}\gamma_j f_j(\omega). \end{aligned} \tag{B.11}$$

Notice that $f_j(\omega)$ can be calculated if $D_c(s)$ is known. Estimates of $\{\gamma_j\}$ can then be obtained by solving (B.11) in a least squares manner using $\hat{\phi}_{u_0}^{(d)}(\mathrm{i}\omega)$ and $\hat{f}_j(\mathrm{i}\omega)$, for $\omega = 0,\ldots$.

Remark. It might be better to use a total least squares approach instead because both ϕ and f_j contain errors.

Remark. As an alternative to solve (B.11) using a least squares approach, one can proceed as follows. Notice that the denominators of both sides of (B.11) are, by construction, equivalent. By equating the numerators of both sides, which are $(m-1)$th order polynomials in $\cos(\omega h)$, a linear system of equations in $\{\gamma_j\}_{j=0}^{m-1}$ is obtained and solved.

- Find $\hat{C}_c(s)$ from (B.2) by spectral factorization.

Finally, the same approach can be used to find $\hat{\theta}$ using the process y_0 described in continuous-time by (2), (4) and in discrete-time by (33).

Remark. One should use the fact that the systems generating $u_0(t)$ and $y_0(t)$ contain common zeros and poles.

Appendix C. FURTHER DETAILS ABOUT THE SHINBROT FUNCTIONAL METHOD

C.1 *Computation of Shinbrot moments*

In this section we comment briefly upon the computation of Shinbrot moments. Here we consider a general signal $x(t)$. The Shinbrot moment is given by

$$\begin{aligned}
\beta_x^{(k)}(\ell) &:= \int_0^{Nh} p^k x(t)\cdot\phi_\ell(t)\ \mathrm{d}t \\
&= (-1)^k \int_0^{Nh} p^k \phi_\ell(t)\cdot x(t)\ \mathrm{d}t \\
&= \sum_{j=0}^{n} [i\omega_0(\ell+j)]^k c_j \int_0^{Nh} e^{-i\omega_0(\ell+j)t}\cdot x(t)\ \mathrm{d}t \\
&= \sum_{j=0}^{n} c_j [i\omega_0(\ell+j)]^k X\{\omega_0(\ell+j)\},
\end{aligned}$$

where $X(\omega)$ is the Fourier transform of $x(t)$ at frequency ω. In practice a *scaled* version of $X(\ell\omega_0)$ is computed using the discrete Fourier transform

$$\check{X}(\ell) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x(kh)\, e^{-i\frac{2\pi}{N}\ell k}.$$

For small h this can be seen as a Riemannian approximation of the Fourier integral. However, if $x(t)$ is band-limited, this result is exact provided the signal is sampled faster than the Nyquist rate. With a slight abuse of notation from here on we write

$$\beta_x^{(k)}(\ell) = \omega_0^k \sum_{j=0}^{n} c_j (\ell+j)^k\ i^k \check{X}(\ell+j). \qquad \text{(C.1)}$$

C.2 *Statistical analysis*

Let $x_i(kh)$, $i\in\{1,2\}$ be sampled versions of two continuous-time processes such that

$$x_i(kh) = x_{i0}(kh) + \tilde{x}_i(kh), \qquad i\in\{1,2\}.$$

Here $x_{i0}(kh)$ is a bounded deterministic quantity, and $\tilde{x}_i(kh)$ is a *discrete-time* stationary stochastic process with bounded second and fourth order moments. Let $\beta_{x_i}^{(i)}(\ell)$ be the Shinbrot moment of $x_i(kh)$ with our usual notation. Moreover,

$$\beta_{x_i}^{(i)}(\ell) = \beta_{x_i0}^{(i)}(\ell) + \tilde{\beta}_{x_i}^{(i)}(\ell),$$

where $\beta_{x_i0}^{(i)}(\ell)$ and $\tilde{\beta}_{x_i}^{(i)}(\ell)$ denote the contributions of x_{i0} and $\tilde{x}_i$, respectively.

By keeping L/N and h fixed and taking $N\to\infty$, (hence $L\to\infty$) one can prove that

$$\begin{aligned}
S_L(k_1,k_2) &:= \frac{1}{L}\sum_{\ell=0}^{L-1} \Big[\mathrm{Re}\left\{\beta_{x_1}^{(k_1)}(\ell)\right\}\mathrm{Re}\left\{\beta_{x_2}^{(k_2)}(\ell)\right\} \\
&\quad +\ \mathrm{Im}\left\{\beta_{x_1}^{(k_1)}(\ell)\right\}\mathrm{Im}\left\{\beta_{x_2}^{(k_2)}(\ell)\right\}\Big] \\
&= S_{L0}(k_1,k_2) + \tilde{S}_L(k_1,k_2)
\end{aligned}$$

with probability one, where

$$\begin{aligned}
S_{L0}(k_1,k_2) &:= \frac{1}{L}\sum_{\ell=0}^{L-1} \Big[\mathrm{Re}\left\{\beta_{x_10}^{(k_1)}(\ell)\right\}\mathrm{Re}\left\{\beta_{x_20}^{(k_2)}(\ell)\right\} \\
&\quad +\mathrm{Im}\left\{\beta_{x_10}^{(k_1)}(\ell)\right\}\mathrm{Im}\left\{\beta_{x_20}^{(k_2)}(\ell)\right\}\Big],
\end{aligned}$$

and

$$\begin{aligned}
\tilde{S}_L(k_1,k_2) &:= \frac{1}{L}\sum_{\ell=0}^{L-1} \mathrm{E}\Big[\mathrm{Re}\left\{\tilde{\beta}_{x_1}^{(k_1)}(\ell)\right\}\mathrm{Re}\left\{\tilde{\beta}_{x_2}^{(k_2)}(\ell)\right\} \\
&\quad +\ \mathrm{Im}\left\{\tilde{\beta}_{x_1}^{(k_1)}(\ell)\right\}\mathrm{Im}\left\{\tilde{\beta}_{x_2}^{(k_2)}(\ell)\right\}\Big]. \qquad \text{(C.2)}
\end{aligned}$$

A detailed proof of this result is beyond the scope of this report.

Next we assume $\tilde{x}_1(kh) = \tilde{x}_2(kh) = x(kh)$, $k\in\mathbb{Z}$. Assume further that $x(kh)$ constitutes a zero-mean discrete-time white noise sequence of variance σ^2, then it can be shown that

$$\begin{aligned}
&\mathrm{E}\{\check{X}(\ell_1)\check{X}^\star(\ell_2)]\} = \sigma^2\delta_{\ell_1,\ell_2}, \\
&\mathrm{E}\{\check{X}(\ell_1)\check{X}(\ell_2)\} = 0,
\end{aligned} \qquad \text{(C.3)}$$

where $\check{X}^\star$ denotes the complex-conjugate of $\check{X}$.

Next we use the identities

$$\mathrm{Re}(a) = \frac{1}{2}\{a+a^\star\}, \qquad \mathrm{Im}(a) = \frac{1}{2i}\{a-a^\star\}.$$

Then from (C.1) and (C.3) we get

$$\begin{aligned}
\mu_\ell(k_1,k_2) &:= \mathrm{E}\left[\mathrm{Re}\left\{\tilde{\beta}_x^{(k_1)}(\ell)\right\}\mathrm{Re}\left\{\tilde{\beta}_x^{(k_2)}(\ell)\right\}\right] \\
&= \frac{\sigma^2}{4}\sum_{j=0}^{n} c_j^2 [\omega_0(\ell+j)]^{k_1+k_2}[i^{k_1-k_2}+i^{k_2-k_1}].
\end{aligned} \qquad \text{(C.4)}$$

Clearly, $\mu_\ell(k_1,k_2)=0$ if k_2-k_1 is an odd number. If k_2-k_1 is even, then $k_1+k_2=2l$, for some $l\in\mathbb{Z}$. Thus, then $(k_2-k_1)/2=k_2-l$, and

$$\mu_\ell(k_1,k_2)=(-1)^{k_2-l}\frac{\sigma^2}{2}\sum_{j=0}^{n}c_j^2[\omega_0(\ell+j)]^{2l}. \tag{C.5}$$

A similar calculation gives

$$\mathrm{E}\left[\mathrm{Im}\left\{\tilde{\beta}_x^{(k_1)}(\ell)\right\}\mathrm{Im}\left\{\tilde{\beta}_x^{(k_2)}(\ell)\right\}\right]=\mu_\ell(k_1,k_2). \tag{C.6}$$

Combining (C.2), (C.4), (C.5) and (C.6) we get

$$\tilde{S}_L(k_1,k_2)=\frac{(-1)^{k_2-l}}{L}\sigma^2\sum_{j=0}^{n}c_j^2\sum_{\ell=j}^{L-j+1}[\omega_0\ell]^{2l}$$

where $k_1+k_2=2l$, $l\in\mathbb{Z}$. Furthermore, $\tilde{S}_L(k_1,k_2)=0$ of k_1+k_2 is an odd number.

Now consider the vectors

$$\begin{aligned}\boldsymbol{\beta}_x(\ell)&=[\,\beta_x^{(n)}(\ell)\ \cdots\ \beta_x^{(0)}(\ell)\,],\\ \boldsymbol{\alpha}_x(\ell)&=[\,\mathrm{Re}\{\boldsymbol{\beta}_x^\top(\ell)\}\ \mathrm{Im}\{\boldsymbol{\beta}_x^\top(\ell)\}\,]^\top,\end{aligned}$$

and

$$R_x:=\frac{1}{L}\sum_{\ell=0}^{L-1}\boldsymbol{\alpha}_x^\top(\ell)\boldsymbol{\alpha}_x(\ell).$$

It follows right away that

$$R_x=R_{x0}+\tilde{R}_x,$$

where

$$R_{x0}=\frac{1}{L}\sum_{\ell=0}^{L-1}\boldsymbol{\alpha}_{x0}^\top(\ell)\boldsymbol{\alpha}_{x0},$$

and

$$[\tilde{R}_x]_{k_1,k_2}=\tilde{S}_L(n-k_1+1,n-k_2+1).$$

and consequently $\tilde{R}_x=\sigma^2\bar{R}$, see (49). Now the derivation of (48) is straightforward since $\tilde{u}(kh)$ and $\tilde{y}(kh)$ are mutually independent.