ELSEVIER

Modeling continuous-time processes via input-to-state filters☆

Kaushik Mahata*, Minyue Fu

Centre for Complex Dynamic Systems and Control, School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia

Abstract

A direct algorithm to estimate continuous-time ARMA (CARMA) models is proposed in this paper. In this approach, we first pass the observed data through an input-to-state filter and compute the state covariance matrix. The properties of the state covariance matrix are then exploited to estimate the half-spectrum of the observed data at a set of user-defined points on the right-half plane. Finally, the continuous-time parameters are obtained from the half-spectrum estimates by solving an analytic interpolation problem with a positive real constraint. As shown by simulations, the proposed algorithm delivers much more reliable estimates than indirect modeling approaches, which rely on estimating an intermediate discrete-time model.
© 2006 Elsevier Ltd. All rights reserved.

Keywords: Continuous-time processes; Identification; Input-to-state filtering; ARMA modeling

1. Introduction

The identification of continuous-time stochastic processes is a fundamental research issue which has received considerable interest recently. Since many natural processes are continuous-time, it is of interest in many practical applications to identify a continuous-time model instead of a discrete-time model (Larsson, 2004). Although the signal is in continuous-time, in practice one works with sampled data. One popular approach is to identify a discrete-time system from uniformly sampled data, as shown in Larsson (2004) and references therein. Subsequently, the estimated discrete-time model is converted back to a continuous-time model via a nonlinear transformation (Söderström, 1991). This approach will be referred to as the *indirect* approach. Apart from the obvious difficulty of solving nonlinear equations, this approach also

suffers from several other setbacks: (i) at a fast sampling rate, the poles and the zeros of the associated discrete-time system cluster close to the point $1 + i0$ in the complex plane, leading to a numerically ill-conditioned identification problem; (ii) the continuous-time parameters can be very sensitive to the sampled data. These issues have been discussed in detail in Fan, Söderström, Mossberg, Carlsson, and Zhou (1999), Fan, Söderström, and Zhou (1999), Larsson and Mossberg (2003), Larsson and Söderström (2002), Söderström (1991) and Söderström and Mossberg (2000). A second approach is to identify the continuous-time parameters *directly*. The basic idea here is to replace the differentiation operator with the *delta operator* (Feuer & Goodwin, 1996; Goodwin, Middleton, & Poor, 1992). Several methods have been developed using this approach for autoregressive models, see Fan, Söderström, Mossberg et al. (1999), Fan, Söderström, and Zhou (1999), Larsson and Söderström (2002) and Söderström and Mossberg (2000). This approach is advantageous in many cases as it is computationally efficient and avoids nonlinear transformations if the underlying model is autoregressive. This technique also benefits from nonuniform sampling (Larsson, 2004; Larsson & Söderström, 2002). However, it is not well understood how we can extend this technique for an ARMA model since the mapping from the continuous-time zeros to the equivalent discrete-time zeros

* Corresponding author.
E-mail addresses: Kaushik.Mahata@newcastle.edu.au (K. Mahata), Minyue.Fu@newcastle.edu.au (M. Fu).

is complicated. It is also not known how to guarantee the positivity of the estimated spectrum. Therefore, for ARMA models, the only available time domain approach seems to be the indirect method described above (Larsson & Mossberg, 2003; Söderström, 1991).

If we consider a discrete-time process, we can ensure the stability of the estimated autoregressive model in a fairly simple way (Ljung, 1999; Makhoul, 1975; Söderström & Stoica, 1989). But unlike the discrete-time counterpart, the mapping from the lagged covariance estimates to the system parameters for a continuous-time system is more complicated. Hence, the standard discrete-time algorithms cannot be extended directly. In this paper, we propose a direct approach for modeling continuous-time stochastic processes. Specifically, we provide an estimation algorithm with the following properties: (i) it is computationally efficient; (ii) the stability of the estimated model is guaranteed; (iii) it can handle irregularly sampled data; (iv) it is possible to circumvent the problems associated with the *sampling zeros* for an ARMA model.

Our approach uses the framework of input-to-state filtering (Georgiou, 2001, 2002a, 2002b) where we first estimate the half-spectrum and its derivatives evaluated at some pre-specified points in the right-half plane. This is achieved using a linear operation on the covariance matrix of the output of an input-to-state filter. Subsequently, we present an approach for estimating a stable rational model from the estimates of the half-spectrum. In this step, we apply linear interpolation combined with a regularization step similar to that proposed in Mari, Stoica, and McKelvey (2000) and Stoica, McKelvey, and Mari (2000). If the resulting model is unstable then a recent spectral zero assignment algorithm (Byrnes, Georgiou, & Lindquist, 2001), (see also Georgiou, 1999), is used to compute a stable model. The approach can also be used when the data are irregularly sampled. We provide additional insights into the estimation of the half-spectrum, and discuss the numerical and statistical issues involved in the estimation of such statistics. We also carry out an asymptotic second-order statistical analysis. The proposed algorithm is tested using numerical simulations.

In the framework of input-to-state filtering it is also possible to model the spectrum in terms of the generalized orthogonal basis functions (Heuberger, Van den Hof, & Wahlberg, 2005). This approach has been explored in Blomqvist and Fanizza (2003).

2. Input-to-state filters

In this section, we briefly state a few key results for any continuous-time wide-sense stationary stochastic process $u(t)$. As a special application, we will apply these results to continuous-time ARMA processes in the later sections. Analogous results for discrete-time processes have been derived in Georgiou (2001, Theorem 1), Georgiou (2002a) and Georgiou (2002b, Corollary 1). The extensions for continuous-time case can be found in Georgiou (2002a, Section V).

Consider a *scalar* and real-valued continuous-time stochastic process $u(t)$ having an autocorrelation function

$$r_\tau := \mathscr{E}\{u(t+\tau)u(t)\}.$$

Then, the spectrum of the process is defined as

$$\phi(s) := \int_{-\infty}^{\infty} d\tau\, r_\tau e^{-s\tau}, \quad \mathrm{Re}(s) = 0.$$

In this work, we use the so-called *half-spectrum* $f(s)$ of $u(t)$, which is defined as

$$f(s) := \int_{0}^{\infty} d\tau\, r_\tau e^{-s\tau}, \quad \mathrm{Re}(s) \geqslant 0.$$

This leads to

$$\phi(s) = f(s) + f(-s), \quad \mathrm{Re}(s) = 0. \tag{1}$$

In the following, we estimate $f(s)$ and its derivatives at a pre-defined set of points $\{s_k\}_{k=1}^{m}$ from the observed continuous-time signal $u(t)$. The points $\{s_k\}_{k=1}^{m}$ are referred to as the *interpolation points* and satisfy $\mathrm{Re}(s_k) > 0$, $\forall k$. The main idea here is to use an input-to-state filter

$$\dot{z}(t) = Fz(t) + gu(t), \tag{2}$$

where F has eigenvalues at $\{-s_k\}_{k=1}^{m}$ and the pair (F, g) is controllable. We assume that the filter in (2) has a pole of order n_k at $-s_k$, while the order of the filter is n, i.e., F is a $n \times n$ matrix, and $\sum_{k=1}^{m} n_k = n$. We show that the covariance matrix of the output $z(t)$ can be used to extract the estimates of $f(s_k)$ and its derivatives. In particular, a pole of order n_k at $-s_k$ enables us to extract the derivatives of $f(s)$ up to order $n_k - 1$ evaluated at s_k. The following proposition is the first step in that direction.

Proposition 1. *Assume that $f(\infty)$ is bounded. Let E be the unique positive definite solution to the Lyapunov equation*

$$FE + EF' + gg' = 0. \tag{3}$$

Then there exist scalar-valued functions $\{w_k\}_{k=0}^{n-1}$ of F and $f(s)$ such that

$$P := \mathscr{E}\{z(t)z'(t)\} = WE + EW', \tag{4}$$

where

$$W := \frac{1}{2\pi i} \oint_{C_R} ds\, f(-s)[sI - F]^{-1} = \sum_{k=0}^{n-1} w_k F^k \tag{5}$$

with C_R being the infinite semicircular contour encircling the entire right-half plane traversed in the clock-wise *direction. Moreover, the coefficients $\{w_k\}_{k=0}^{n-1}$ are invariant of the choice of the coordinates of $z(t)$.*

Proof. The proof of (4) and (5) is similar to Theorem 1 in Georgiou (2001); see also Mahata and Fu (2005) for a more direct proof. To show the invariance of $\{w_k\}_{k=0}^{n-1}$, consider the state sequence $z_1(t)$ of the input-to-state filter (F_1, g_1), where

$$F_1 = LFL^{-1}, \quad g_1 = Lg,$$

for some nonsingular matrix L. Let the covariance matrix of $z_1(t)$ be P_1. It follows that $P_1 = LPL'$ and the solution E_1 to the Lyapunov equation $F_1 E_1 + E_1 F_1' + g_1 g_1' = 0$ satisfies $E_1 = LEL'$. Then (4) gives

$$P_1 = W_1 E_1 + E_1 W_1',$$

where $W_1 = LWL^{-1}$. Now, from (5) we can verify our assertion that $W_1 = \sum_{k=0}^{n-1} w_k F_1^k$. $\quad\square$

Next, we state a method for computing $\{f(s_k)\}_{k=1}^m$ and its derivatives.

Theorem 1. *Let F be chosen such that its Jordan form has the block-diagonal structure*

$$\text{diag}\{J(-s_1, n_1), \ldots, J(-s_m, n_m)\}, \tag{6}$$

where for each $k \in \{1, 2, \ldots, m\}$, the matrix $J(-s_k, n_k)$ is an elementary $n_k \times n_k$ Jordan block having $-s_k$ along the main diagonal, ones *along the first upper sub-diagonal*, and *zeros elsewhere. Define the polynomial*

$$\varpi(s) := \sum_{k=0}^{n-1} w_k s^k.$$

Denote the rth derivative of $f(s)$ evaluated at $s = s_k$ by $f^{(r)}(s_k)$. Then

$$f^{(r)}(s_k) = (-1)^r \varpi^{(r)}(-s_k), \quad 0 \leqslant r < n_k, \quad 1 \leqslant k \leqslant m. \tag{7}$$

Proof. The proof is similar to Theorem 2 in Georgiou (2001); see also Mahata and Fu (2005) for a direct proof. $\quad\square$

Remarks.

(1) If we allow multiple Jordan blocks of F with the same eigenvalue $-s_k$, then the information present in the associated Jordan blocks is redundant, and we are unable to extract n pieces of statistics about $f(s)$. See the proof in Mahata and Fu (2005) for details.
(2) Extracting statistics from the data for modeling discrete-time processes is often accomplished by estimating the covariances of the observed data. The theory of input-to-state filtering is not necessary in this development (Ljung, 1999). However, Theorem 1 is vital for continuous-time processes.

Next, we express the vector of $f^{(j)}(s_k)$ as a linear function of the coefficients $\{w_k\}_{k=0}^{n-1}$. Define

$$\check{f}_k := \begin{bmatrix} f^{(0)}(s_k) \\ \vdots \\ f^{(n_k-1)}(s_k) \end{bmatrix}, \qquad w := \begin{bmatrix} w_0 \\ \vdots \\ w_{n-1} \end{bmatrix}. \tag{8}$$

From (7) we get $\check{f}_k = \mathfrak{D}_k w$ for $k = 1, \ldots, m$, where $\mathfrak{D}_k$ is a $n_k \times n$ matrix defined as

$$[\mathfrak{D}_k]_{ij} = (-1)^{i-1} \left[\frac{d^{i-1}}{ds^{i-1}} \{s^{j-1}\} \right]_{s=-s_k}.$$

A real-valued implementation of the input-to-state filter requires a self-conjugate set of interpolation points. However, if s_j is the complex conjugate of s_k, then $\check{f}_j$ is the complex conjugate of $\check{f}_k$. Hence, it is sufficient to consider the interpolation points having nonnegative imaginary parts. Let the number of real-valued interpolation points be n_r, and the number of interpolation points with strictly positive imaginary parts be n_c. Clearly, $m = n_r + 2n_c$. Without loss of generality, we assume that $\{s_k\}_{k=1}^{n_r}$ are real-valued, and $\{s_k\}_{k=n_r+1}^{n_r+n_c}$ have strictly positive imaginary parts. Define the vectors

$$f_R := \begin{bmatrix} \check{f}_1 \\ \vdots \\ \check{f}_{n_r} \end{bmatrix}, \qquad f_C := \begin{bmatrix} \check{f}_{n_r+1} \\ \vdots \\ \check{f}_{n_r+n_c} \end{bmatrix}, \qquad f := \begin{bmatrix} f_R \\ \text{Re}(f_C) \\ \text{Im}(f_C) \end{bmatrix}. \tag{9}$$

From the above discussion it follows that

$$f = \mathbf{D}w, \tag{10}$$

where

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 \\ \text{Re}(\mathbf{D}_2) \\ \text{Im}(\mathbf{D}_2) \end{bmatrix}, \qquad \mathbf{D}_1 = \begin{bmatrix} \mathfrak{D}_1 \\ \vdots \\ \mathfrak{D}_{n_r} \end{bmatrix}, \qquad \mathbf{D}_2 = \begin{bmatrix} \mathfrak{D}_{n_r+1} \\ \vdots \\ \mathfrak{D}_{n_r+n_c} \end{bmatrix}.$$

Using the results derived so far, we propose an algorithm below for the estimation of f. The justification behind the algorithm will be given in Section 3. In the algorithm, we implement the input-to-state filter in the controllable canonical form. Let the characteristic polynomial of F be

$$\Delta(s) := \prod_{k=1}^n (s + s_k) = s^n + \sum_{k=1}^n \delta_k s^{n-k}. \tag{11}$$

The state space matrices for the input-to-state filter in the controllable canonical form are then given by

$$F_\star = \begin{bmatrix} -\delta_1 & \cdots & -\delta_{n-1} & -\delta_n \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad g_\star = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The associated state vector and the state covariance are denoted by $z_\star(t)$ and $P_\star$, respectively.

$$\dot{z}_\star(t) = F_\star z_\star(t) + g_\star u(t), \qquad P_\star := \mathscr{E}\{z_\star(t) z_\star'(t)\}. \tag{12}$$

From (4) and (5) we have

$$P_\star = \sum_{k=0}^{n-1} w_k \{F_\star^k E_\star + E_\star (F_\star')^k\}, \tag{13}$$

where $E_\star$ is the solution to the Lyapunov equation

$$F_\star E_\star + E_\star F_\star' + g_\star g_\star' = 0. \tag{14}$$

Therefore, computing w from $P_\star$ amounts to solving a least-squares problem. In Section 3, we show that it is enough to consider only the diagonal elements in the matrix-valued (13). In the following, for a $n \times n$ real-valued matrix H we denote

$$\eth[H] = [[H]_{1,1} \quad \cdots \quad [H]_{n,n}]'.$$

Considering only the diagonal elements of (13) we get

$$\eth[P_\star] = X_\star w, \tag{15}$$

where

$$X_\star :=$$

$$[\eth[2E_\star]\eth[F_\star E_\star + E_\star F_\star'] \quad \cdots \quad \eth[F_\star^{n-1}E_\star + E_\star (F_\star')^{n-1}]].$$

Since $F_\star$ and $g_\star$ are user defined, $X_\star$ is assumed to be known in the following algorithm. We assume further that the observed process $u(t)$ is known for $t \in [0, T]$.

Algorithm.

(1) Compute the state output $z_\star(t)$ in (12), and obtain an estimate of $P_\star$ by computing

$$\hat{P}_\star := \frac{1}{T} \int_0^T dt \, z_\star(t) z_\star'(t). \tag{16}$$

(2) Compute the estimates of w and f as

$$\hat{w} = X_\star^{-1} \eth[\hat{P}_\star], \quad \hat{f} = \mathbf{D}\hat{w}. \tag{17}$$

From a practical point of view one might prefer to implement (16) in discrete-time, where the integral is approximated by a Riemannian summation

$$\check{P}_\star = \frac{1}{N} \sum_{k=0}^{N-1} z_\star(kt_s) z_\star'(kt_s),$$

where t_s is the *sampling interval* and $T = Nt_s$. In fact, $\check{P}_\star$ is a consistent estimate of $P_\star$. Provided t_s is sufficiently small, decreasing t_s any further has no effect on the accuracy of $\check{P}_\star$ (Wahlberg, 1990). Obviously, it is also possible to estimate $P_\star$ using nonuniform sampling of $z_\star(t)$.

It is often assumed that the observed data are the sampled version of the continuous-time process. Then, it is required to implement the input-to-state filter (12) using a discretization technique (Garnier & Young, 2004; Middleton & Goodwin, 1990, p. 33). The systematic errors introduced due to such approximation are not significant (Garnier & Young, 2004) for practical values of t_s. Analysis of such systematic errors is beyond the scope of this paper. However, if the continuous-time signal is available to the user, it is possible to avoid systematic errors by implementing (12) using analog devices for which several efficient architectures are readily available (Kailath, 1980, p. 35). The hardware implementation requires a bank of analog integrators and static gain circuits. Such circuits can be implemented reliably using operational amplifiers (Franco, 2001). The output $z_\star(t)$ from the input-to-state filter can then be sampled for the computation of $\hat{P}_\star$.

3. Some computational and statistical aspects

In this section, we focus on the computational and the statistical aspects involved in the estimation of the half-spectrum. The primary aim is to justify the algorithm for computing $\hat{w}$ and $\hat{f}$ proposed in Section 2. First, we examine the rank of the system of equations to be solved in order to determine $\hat{w}$. We also comment on the choice of the coordinates of the state of the input-to-state filter. Finally, we determine the second-order statistics of $\hat{w}$ and $\hat{f}$. To this end, we need some additional notation. Define

$$\mathbb{H} := \{H \in \mathbb{R}^{n \times n} : [H]_{i,j} = (-1)^j h_{i+j-1}\}$$

for some real-valued $\{h_k\}_{k=1}^{2n-1}$, and

$$\mathbb{H}_0 := \{H \in \mathbb{H} : [H]_{i,j} = 0 \text{ when } i + j \text{ is odd}\}.$$

A key property of the set $\mathbb{H}_0$ is that if $H \in \mathbb{H}_0$ and $i + j = 2\ell$ for some positive integer $\ell \leqslant n$. Then

$$[H]_{i,j} = (-1)^{j-\ell} [H]_{\ell,\ell}.$$

It is also straightforward to verify that $H \in \mathbb{H}$ implies $H + H' \in \mathbb{H}_0$.

Proposition 2. *The matrix $P_\star$ defined in (12) is a member of $\mathbb{H}_0$.*

Proof. Denote the differentiation operator by p. Since $(F_\star, g_\star)$ is in the controllable canonical form, we have

$$z_\star(t) = [z_1(t) \quad \cdots \quad z_n(t)]', \quad z_j(t) := \frac{p^{n-j}}{\Delta(p)} u(t),$$

where the polynomial $\Delta(\cdot)$ is defined in (11). By the definition of $P_\star$, we get

$$[P_\star]_{jk} = \mathscr{E}\{z_j(t) z_k(t)\}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega \, \phi(i\omega) \frac{(i\omega)^{n-j}(-i\omega)^{n-k}}{|\Delta(i\omega)|^2}. \tag{18}$$

Note that the spectrum $\phi(\omega)$ is an even function of ω. Therefore, if $j + k$ is an odd integer the integral vanishes because the integrand is an odd function of ω. If $j + k = 2\ell$ for some integer ℓ, then we have

$$(i\omega)^{n-j}(-i\omega)^{n-k} = \omega^{2n-2\ell}(i)^{2n-2\ell}(-1)^{n-k}$$

$$= \omega^{2n-2\ell}(-1)^{k-\ell}.$$

Consequently, it follows from (18) that

$$[P_\star]_{jk} = \frac{(-1)^{k-\ell}}{2\pi} \int_{-\infty}^{\infty} d\omega \frac{\omega^{2n-2\ell}\phi(i\omega)}{|\Delta(i\omega)|^2}$$

$$= (-1)^{k-\ell}\mathcal{E}\left\{\frac{p^{n-\ell}}{\Delta(p)}u(t) \cdot \frac{p^{n-\ell}}{\Delta(p)}u(t)\right\}$$

$$= (-1)^{k-\ell}[P_\star]_{\ell\ell}.$$

From the last equality, it is straightforward to verify that $P_\star \in \mathbb{H}_0$. $\square$

Remark. If $u(t)$ is a continuous-time white noise (i.e., derivative of the Weiner process) of unit variance, then $P_\star = E_\star$; see (14). Consequently, $E_\star \in \mathbb{H}_0$.

In the following proposition, we show that the matrices involved in (13) belong to $\mathbb{H}_0$.

Proposition 3. *The matrix* $X_k := F_\star^k E_\star \in \mathbb{H}$ *for any positive integer* k.

Proof. In this proof, we use the Matlab notation $H(j : k, :), j \leqslant k$ to denote the sub-matrix of H composed of the jth to kth rows of H. Similarly, $H(:, j : k), j \leqslant k$ denotes the sub-matrix of H composed of the jth to kth columns of H. From the structure of $F_\star$ we have for $k \geqslant 0$ that

$$X_{k+1}(2 : n, :) = F_\star(2 : n, :)X_k = [I_{n-1}0_{(n-1)\times 1}]X_k$$

$$= X_k(1 : n - 1, :). \tag{19}$$

Next, we show using mathematical induction that

$$X_{k+1}(:, 2 : n) = -X_k(:, 1 : n - 1). \tag{20}$$

Consider the case $k = 0$. Note that,

$$X_1 + X_1' = -g_\star g_\star'.$$

Since $[g_\star g_\star']_{jk} = 0$ unless $j = k = 1$, we get

$$X_1(:, 2 : n) = -X_1'(:, 2 : n) = -[X_1(2 : n, :)]'$$

$$= -[X_0(1 : n - 1, :)]' = -X_0'(:, 1 : n - 1)$$

$$= -X_0(:, 1 : n - 1),$$

where in the third equality we have used (19) and the last equality follows because $X_0 = E_\star$ is a symmetric matrix. Thus, (20) holds for $k = 0$. Now, assume that (20) holds for $k = 0, 1, \ldots, \ell - 1$. Then

$$X_{\ell+1}(:, 2 : n) = F_\star X_\ell(:, 2 : n) = -F_\star X_{\ell-1}(:, 1 : n - 1)$$

$$= -X_\ell(:, 1 : n - 1).$$

Hence, (20) follows by induction. Now $X_0 = E_\star \in \mathbb{H}$. Applying (19) and (20) recursively it is readily verified that $X_k \in \mathbb{H}$ for all $k \geqslant 0$. Hence the proposition follows. $\square$

It follows from Proposition 3 that $X_k + X_k' \in \mathbb{H}_0$ for all k. Since $P_\star \in \mathbb{H}_0$, the number of independent equations in the matrix (13) is only n. Even if the input-to-state filter is *not* implemented in the controllable canonical form, it is natural to expect that the number of independent equations is n. Indeed, if $z(t) = Lz_\star(t)$ for some nonsingular L, then using (13) and Proposition 1 we get

$$LP_\star L' = \sum_{k=0}^{n-1} w_k L(X_k + X_k')L'.$$

Hence

$$[L \otimes L]\text{vec}[P_\star] = \sum_{k=0}^{n-1} w_k[L \otimes L]\text{vec}[X_k + X_k']. \tag{21}$$

This is a weighted version of (13). Note that we denote the matrix Kronecker product by $\otimes$. Since the underlying system of equations is not overdetermined, the weighting has no effect on the final solution. However, in a practical scenario, the true covariance matrix $P_\star$ is unknown, and we work with an estimate $\hat{P}_\star$ not necessarily belonging to $\mathbb{H}_0$. Then we might expect that solving an overdetermined system with proper weighting (21) may give more accurate estimates of $\{w_k\}_{k=0}^{n-1}$. We explore the second-order statistical properties of $\hat{P}_\star$ to examine this aspect. We first state the following basic result.

Proposition 4. *Let* $u_A(t)$, $u_B(t)$, $u_C(t)$ *and* $u_D(t)$ *be jointly Gaussian real and scalar-valued continuous-time stationary stochastic processes. Let us define*

$$\hat{P}_{AB}(\tau) := \frac{1}{T}\int_0^T dt\, u_A(t + \tau)u_B(t),$$

$$P_{AB}(\tau) := \mathcal{E}\{u_A(t + \tau)u_B(t)\}.$$

The cross-spectrum $\Phi_{AB}(i\omega)$ *of* $u_A(t)$ *and* $u_B(t)$ *is given by*

$$\Phi_{AB}(i\omega) = \int_{-\infty}^{\infty} d\omega\, P_{AB}(\tau)e^{-i\omega\tau}.$$

Assume that each of the functions $P_{AC}(\tau)P_{BD}(\tau)$ *and* $P_{AD}(\tau)P_{BC}(\tau)$ *tend to zero as* $\tau \to \infty$ *at a rate faster than* τ^{-2}. *Then as* $T \to \infty$ *the asymptotic covariance between* $\hat{P}_{AB}(0)$ *and* $\hat{P}_{CD}(0)$ *is given by*

$$\mathcal{E}\{\hat{P}_{AB}(0) - P_{AB}(0)\}\{\hat{P}_{CD}(0) - P_{CD}(0)\}$$

$$= \frac{1}{2\pi T}\int_{-\infty}^{\infty} d\omega[\Phi_{AC}(i\omega)\Phi_{DB}(i\omega) + \Phi_{AD}(i\omega)\Phi_{CB}(i\omega)].$$

Proof. By straightforward algebra, we have

$$\mathcal{E}\{\hat{P}_{AB}(0) - P_{AB}(0)\}\{\hat{P}_{CD}(0) - P_{CD}(0)\}$$

$$= \mathcal{E}\{\hat{P}_{AB}(0)\hat{P}_{CD}(0)\} - P_{AB}(0)P_{CD}(0)$$

$$= \frac{1}{T^2} \int_0^T \int_0^T dt_1 dt_2 [\mathcal{E}\{u_A(t_1)u_B(t_1)u_C(t_2)u_D(t_2)\}$$

$$\quad - \mathcal{E}\{u_A(t_1)u_B(t_1)\}\mathcal{E}\{u_C(t_2)u_D(t_2)\}]$$

$$= \frac{1}{T^2} \int_0^T \int_0^T dt_1 dt_2 [\mathcal{E}\{u_A(t_1)u_C(t_2)\}\mathcal{E}\{u_B(t_1)u_D(t_2)\}$$

$$\quad + \mathcal{E}\{u_A(t_1)u_D(t_2)\}\mathcal{E}\{u_B(t_1)u_C(t_2)\}]$$

$$= \frac{1}{T^2} \int_0^T \int_0^T dt_1 dt_2 [P_{AC}(t_1 - t_2)P_{BD}(t_1 - t_2)$$

$$\quad + P_{AD}(t_1 - t_2)P_{BC}(t_1 - t_2)]$$

$$= \frac{1}{T^2} \int_{-T}^T dt (T - t)[P_{AC}(t)P_{BD}(t) + P_{AD}(t)P_{BC}(t)],$$

where in the second equality we have used a well-known formula for the fourth-order moment of a jointly Gaussian random variables (Janssen & Stoica, 1988). Since the functions $P_{AC}(\tau)P_{BD}(\tau)$ and $P_{AD}(\tau)P_{BC}(\tau)$ tend to zero as $\tau \to \infty$ at a rate faster than τ^{-2}, we can write as $T \to \infty$

$$\mathcal{E}\{\hat{P}_{AB}(0) - P_{AB}(0)\}\{\hat{P}_{CD}(0) - P_{CD}(0)\}$$

$$= \frac{1/T}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dt \, d\omega_1 \, d\omega_2 \, e^{i(\omega_1 + \omega_2)t}$$

$$\quad \times [\Phi_{AC}(i\omega_1)\Phi_{BD}(i\omega_2) + \Phi_{AD}(i\omega_1)\Phi_{BC}(i\omega_2)]$$

$$= \frac{1}{2\pi T} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\omega_1 \, d\omega_2 \, \delta(\omega_1 + \omega_2)$$

$$\quad \times [\Phi_{AC}(i\omega_1)\Phi_{BD}(i\omega_2) + \Phi_{AD}(i\omega_1)\Phi_{BC}(i\omega_2)]$$

$$= \frac{1}{2\pi T} \int_{-\infty}^{\infty} d\omega [\Phi_{AC}(i\omega)\Phi_{BD}(-i\omega)$$

$$\quad + \Phi_{AD}(i\omega)\Phi_{BC}(-i\omega)].$$

Hence, the proposition follows. □

Since $P_\star \in \mathbb{H}_0$, each element along an anti-sub-diagonal of $\hat{P}_\star$ gives the estimates of the same quantity (up to a sign factor). Our next proposition explores the correlation structure of the elements in $\{[\hat{P}_\star]_{ij} : i + j = 2\ell\}$ for a fixed ℓ.

Proposition 5. Let $\hat{P}_\star$ be given by (16). Then for $i + j = k + l = 2\ell$ it holds as $T \to \infty$ that

$$\text{cov}\{[\hat{P}_\star]_{ij}, [\hat{P}_\star]_{kl}\} = (-1)^{j-l}\text{cov}\{[\hat{P}_\star]_{\ell\ell}, [\hat{P}_\star]_{\ell\ell}\}, \quad (22)$$

where $\text{cov}\{x_1, x_2\}$ denotes the covariance between two random variables x_1 and x_2. Furthermore, for the diagonal elements of

$\hat{P}_\star$ it holds for $T \to \infty$ that

$$\text{cov}\{[\hat{P}_\star]_{ii}, [\hat{P}_\star]_{jj}\} = \frac{2}{T}\text{var}\left\{\frac{p^{2n-i-j}}{\Delta^2(p)}u_s(t)\right\}, \quad (23)$$

where $u_s(t)$ is a stationary stochastic process having a spectral density $\Phi_u^2(i\omega)$.

Proof. Recall that $[z_\star(t)]_i = \{p^{n-i}/\Delta(p)\}u(t)$, where p denotes the differentiation operator. Consider the case $i + j = k + l = 2\ell$. Then $(-1)^k = (-1)^{2\ell-l} = (-1)^{-l} = (-1)^l$. Proposition 4 gives

$$\text{cov}\{[\hat{P}_\star]_{ij}, [\hat{P}_\star]_{kl}\}$$

$$= \frac{1}{2\pi T} \int_{-\infty}^{\infty} d\omega \frac{|\Phi_u(i\omega)|^2}{|\Delta(i\omega)|^4}$$

$$\quad \times [(i\omega)^{2n-i-l}(-i\omega)^{2n-j-k} + (i\omega)^{2n-i-k}(-i\omega)^{2n-j-l}]$$

$$= \frac{1}{2\pi T} \int_{-\infty}^{\infty} d\omega \frac{|\Phi_u(i\omega)|^2}{|\Delta(i\omega)|^4}(i\omega)^{4n-i-j-k-l}$$

$$\quad \times [(-1)^{j+k} + (-1)^{j+l}]$$

$$= \frac{1}{2\pi T} \int_{-\infty}^{\infty} d\omega \frac{|\Phi_u(i\omega)|^2}{|\Delta(i\omega)|^4}(i\omega)^{4(n-\ell)}[(-1)^{j-l} + (-1)^{j-l}]$$

$$= \frac{2(-1)^{j-l}}{2\pi T} \int_{-\infty}^{\infty} d\omega \frac{\omega^{4(n-\ell)}|\Phi_u(i\omega)|^2}{|\Delta(i\omega)|^4}. \quad (24)$$

Now putting $j = l = \ell$ in (24) we get (22). The calculation leading to (23) is similar i.e.,

$$\text{cov}\{[\hat{P}_\star]_{ii}, [\hat{P}_\star]_{jj}\} = \frac{1}{2\pi T} \int_{-\infty}^{\infty} d\omega \frac{|\Phi_u(i\omega)|^2}{|\Delta(i\omega)|^4}$$

$$\quad \times [(i\omega)^{2n-i-j}(-i\omega)^{2n-i-j}$$

$$\quad + (i\omega)^{2n-i-j}(-i\omega)^{2n-i-j}],$$

which gives (23). □

Since $P_\star \in \mathbb{H}_0$, every element of the set $\mathbb{S}_\ell := \{(-1)^{k-\ell}[\hat{P}_\star]_{j,k} : j + k = 2\ell\}$ gives an estimate of $[P_\star]_{\ell,\ell}$. However, from (22) it follows that the asymptotic covariance matrix of the elements in $\mathbb{S}_\ell$ is of rank one. Hence, it is not possible to improve the statistical accuracy of estimated $[P_\star]_{\ell,\ell}$ by taking a linear combination of the elements in $\mathbb{S}_\ell$. In practice, T must be large enough to ensure the reliability of the extracted statistics. For such practical values of T, the inference made from the asymptotic analysis is still valid. Numerical simulation results also confirm this observation. Thus, we conclude that there is no improvement in statistical accuracy if we solve the overdetermined system of equations (21), and the coordinates of $z(t)$ have no effect on the statistical accuracy of $\hat{w}$. The observations so far justify the algorithm described in Section 2.

Proposition 6. As the observation time $T \to \infty$ the asymptotic covariance matrix $C_{\hat{w}}$ of $\sqrt{T}\hat{w}$ in (17) is given by

$$C_{\hat{w}} = X^{-1}C_P(X')^{-1}, \quad (25)$$

where $\hat{P}_\star$ is given in (16) and C_P is the asymptotic covariance matrix of $\sqrt{T}\mathfrak{d}[\hat{P}_\star]$, which is a Hankel matrix given by

$$[C_P]_{i,j} = 2\,\text{var}\left\{\frac{p^{2n-i-j}}{\Delta^2(p)}u_s(t)\right\}, \tag{26}$$

where $u_s(t)$ is a stationary stochastic process having spectral density $\Phi_u^2(\mathrm{i}\omega)$. Consequently, the asymptotic covariance matrix of $\sqrt{T}\hat{f}$ is given by

$$C_f = \mathbf{D}C_{\hat{w}}\mathbf{D}'. \tag{27}$$

Proof. The expression for C_P in (26) follows directly from (24), while (25) and (27) are consequences of (17). $\quad\square$

4. Fitting a rational model

The algorithm at the end of Section 2 gives the estimates of $f(s)$ at the interpolation points. In this section, we propose an approach to fit a rational model to the interpolation data. Assume that $u(t)$ has a strictly proper rational spectrum of order v, i.e.,

$$\phi(s) = \frac{c(s)c(-s)}{a(s)a(-s)}, \tag{28}$$

where

$$a(s) = s^v + \sum_{k=1}^v a_k s^{v-k}, \quad c(s) = \sum_{k=1}^v c_k s^{v-k}.$$

Then the half-spectrum $f(s)$ also admits a strictly proper rational representation as

$$f(s) = \frac{b(s)}{a(s)}, \quad b(s) = \sum_{k=1}^v b_k s^{v-k}$$

such that

$$\varphi(s) := c(s)c(-s) = a(s)b(-s) + b(s)a(-s). \tag{29}$$

Our approach in this work is to identify the *real-valued* parameters $\{a_k\}_{k=1}^v$ and $\{b_k\}_{k=1}^v$ from the data. Subsequently, we can evaluate the right-hand side of the equation (29). Then a spectral factorization of (29) will lead to the parameters $\{c_k\}_{k=1}^v$. Note that the right-hand side of (29) needs to be positive real in order to ensure the existence of a stable spectral factor. Another important issue is to ensure the stability of the estimated polynomial $a(s)$. The problem of computing $a(s)$ and $b(s)$ from the *interpolation conditions* originating from $f(s)$ and its derivatives evaluated at $\{s_k\}_{k=1}^m$ is in fact a linear problem. However, when we impose the stability constraint on $a(s)$ and positivity constraint on the right-hand side of (29), we have to solve a Nevanlinna–Pick interpolation problem with a degree constraint (Delsarte, Genin, Kamp, & Van Dooren, 1982; Kimura, 1987), which is more difficult.

In order to keep the description simple, we do not consider interpolation conditions involving first and higher order derivatives of $f(s)$. However, the following discussion can be generalized[1] to account for the interpolation conditions involving first and higher order derivatives of $f(s)$. Thus, from now on we consider the case $n_k = 1$ for $k \in \{1, \ldots, m\}$.

Let $\hat{f}_k$ be the estimate of $f(s_k)$. Once $n \geqslant 2v$, it is straightforward to solve

$$\hat{f}_k a(s_k) = b(s_k), \quad k \in \{1, \ldots, n\} \tag{30}$$

in the least-squares sense. It is also straightforward to incorporate weights and solve a weighted least-squares problem. In order to derive the statistical properties of the resulting estimates, we need to express the problem in terms of the real-valued vector $\hat{f}$ as defined in (17). For that we need some notation. Define

$$\psi_k := \begin{bmatrix} s_k^{v-1} \\ \vdots \\ s_k \\ 1 \end{bmatrix}, \quad \Psi_R := \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_{n_r} \end{bmatrix}, \quad \Psi_C := \begin{bmatrix} \psi_{n_r+1} \\ \vdots \\ \psi_{n_r+n_c} \end{bmatrix},$$

$$\gamma_R := \begin{bmatrix} s_1^v \\ \vdots \\ s_{n_r}^v \end{bmatrix}, \quad \gamma_C := \begin{bmatrix} s_{n_r+1}^v \\ \vdots \\ s_{n_r+n_c}^v \end{bmatrix}, \quad \gamma := \begin{bmatrix} \gamma_R \\ \text{Re}(\gamma_C) \\ \text{Im}(\gamma_C) \end{bmatrix},$$

$$\mathbf{F} := \begin{bmatrix} \text{diag}(f_R) & 0 & 0 \\ 0 & \text{diag}\{\text{Re}(f_C)\} & -\text{diag}\{\text{Im}(f_C)\} \\ 0 & \text{diag}\{\text{Im}(f_C)\} & \text{diag}\{\text{Re}(f_C)\} \end{bmatrix},$$

$$\Psi := [\Psi_R' \ \text{Re}(\Psi_C') \ \text{Im}(\Psi_C')]'.$$

Also, see (9) for definitions of f_R and f_C. In the following, we denote the estimate of $\mathbf{F}$ derived from $\hat{f}$ by $\hat{\mathbf{F}}$. Let us introduce

$$\theta_1 := [a_1 \ \cdots \ a_v \ b_1 \ \cdots \ b_v]',$$

$$G = [-\mathbf{F}\Psi \ \Psi], \quad \hat{G} = [-\hat{\mathbf{F}}\Psi \ \Psi].$$

It can be verified that $G\theta_1 = \mathbf{F}\gamma$. Consequently, the least-squares estimate of θ_1 is given by

$$\hat{\theta}_1 = [\hat{G}'\Delta\hat{G}]^{-1}\hat{G}'\Delta\hat{\mathbf{F}}\gamma,$$

where Δ is a positive-definite weighting matrix chosen by the user. The following proposition quantifies the asymptotic covariance matrix of $\hat{\theta}_1$.

---

[1] To illustrate how to set up a linear problem in terms of interpolation constraints involving derivatives of $f(s)$ note that

$$\frac{\mathrm{d}f(s)}{\mathrm{d}s}a(s) + \frac{\mathrm{d}a(s)}{\mathrm{d}s}f(s) = \frac{\mathrm{d}b(s)}{\mathrm{d}s}.$$

Here, we estimate $f(s)$ and $\mathrm{d}f(s)/\mathrm{d}s$ for known s values, giving the interpolation conditions. Thus, the above equation is still linear in $\{a_k\}_{k=1}^v$ and $\{b_k\}_{k=1}^v$. Now, the procedure can be repeated successively for higher order derivatives.

Proposition 7. *Define*

$$U = \begin{bmatrix} \operatorname{diag}(U_R) & 0 & 0 \\ 0 & \operatorname{diag}\{\operatorname{Re}(U_C)\} & -\operatorname{diag}\{\operatorname{Im}(U_C)\} \\ 0 & \operatorname{diag}\{\operatorname{Im}(U_C)\} & \operatorname{diag}\{\operatorname{Re}(U_C)\} \end{bmatrix},$$

where

$$U_R = \begin{bmatrix} a(s_1) \\ \vdots \\ a(s_{n_r}) \end{bmatrix}, \quad U_C = \begin{bmatrix} a(s_{n_r+1}) \\ \vdots \\ a(s_{n_r+n_c}) \end{bmatrix}.$$

Then the asymptotic covariance matrix of $\sqrt{T}\hat{\theta}_1$ is given by

$$C_{\hat{\theta}_1} = T\mathscr{E}\{(\hat{\theta}_1 - \theta_1)(\hat{\theta}_1 - \theta_1)'\}$$

$$= [G'\Delta G]^{-1} G'\Delta U C_f U'\Delta G[G'\Delta G]^{-1}, \tag{31}$$

the optimal choice of Δ is given by

$$\Delta = (U C_f U')^{-1},$$

and the associated covariance matrix of the optimum estimate of $\sqrt{T}\theta_1$ is $[G'(U C_f U')^{-1} G]^{-1}$.

Proof. In this proof, we denote $\theta_a = [a_1 \ \cdots \ a_v]'$. Using a standard technique of deriving the asymptotic accuracy of the least-squares-based estimates (Söderström & Stoica, 1989, p. 285), one can show that the asymptotic estimation error is given by

$$\hat{\theta}_1 - \theta_1 = [\hat{G}'\Delta\hat{G}]^{-1}\hat{G}'\Delta(\hat{\mathbf{F}}\gamma - \hat{G}\theta_1)$$

$$\approx [G'\Delta G]^{-1} G'\Delta(\hat{\mathbf{F}}\gamma - \hat{G}\theta_1)$$

$$= [G'\Delta G]^{-1} G'\Delta(\hat{\mathbf{F}} - \mathbf{F})(\gamma + \Psi\theta_a)$$

$$= [G'\Delta G]^{-1} G'\Delta U(\hat{f} - f).$$

Now, it is straightforward to derive (31). The remaining part of the proposition follows from the theory of the best linear unbiased estimates (Ljung, 1999, p. 555). □

The implementation of the optimally weighted least-squares estimator requires a bootstrapping procedure. First, we need an initial estimate of the system parameters obtained without any weighting. This knowledge is used to compute C_f, which is then used to implement the optimally weighted estimator.

Let $\hat{a}(s)$, $\hat{b}(s)$ and $\hat{\varphi}(s)$ denote the estimates of $a(s)$, $b(s)$ and $\varphi(s)$, respectively, derived from $\hat{\theta}_1$. The requirement for spectral factorization

$$\hat{\varphi}(s) > 0, \quad \operatorname{Re}(s) = 0 \tag{32}$$

may not hold, in general. This can be fixed by using a regularization procedure described below. The idea here is to perturb the coefficients of $\hat{\varphi}(s)$, so that (32) is enforced. This actually amounts to solving a linear matrix inequality (LMI). In fact, we can use the results in Stoica et al. (2000) (see also Mari

et al., 2000) for estimation of moving average processes. Using a bilinear transformation $s = (z - 1)/(z + 1)$, we define

$$\hat{\varphi}_d(z) := \frac{(z+1)^{2v}}{z^v}\hat{\varphi}\left\{\frac{z-1}{z+1}\right\}$$

$$= [(1+z)(1+z^{-1})]^v\hat{\varphi}\left\{\frac{z-1}{z+1}\right\} = \sum_{j=-v}^{v}\beta_j z^j.$$

It is straightforward to verify that $\beta_j = \beta_{-j}, \forall j$ and that $\hat{\varphi}_d(z)$ is real-valued for $|z| = 1$. Using the property of the bilinear transformation we can express (32) as

$$\hat{\varphi}_d(z) > 0, \quad |z| = 1. \tag{33}$$

However, (33) is equivalent to the constraint that $\hat{\varphi}_d(e^{-i\omega})$ is the spectral density function of a moving average process. This problem occurs naturally in identifying an order-v moving average process, where $\{\beta_j\}_{j=0}^v$ represent the covariances estimated from the data. Also, it is not guaranteed that (33) holds. Therefore, it is required to modify the vector $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \cdots \ \beta_v]'$. We may obtain the modified vector $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0 \ \hat{\beta}_1 \ \cdots \ \hat{\beta}_v]'$ using

$$\hat{\boldsymbol{\beta}} = \arg\min_{\bar{\boldsymbol{\beta}}} \|\boldsymbol{\beta} - \bar{\boldsymbol{\beta}}\|^2 \tag{34}$$

subject to

$$\sum_{j=-v}^{v} \bar{\beta}_j z^j > 0, \quad |z| = 1, \tag{35}$$

where $\bar{\boldsymbol{\beta}} = [\bar{\beta}_0 \ \cdots \ \bar{\beta}_v]'$. The above optimization problem can be cast as a semidefinite programming problem (Stoica et al., 2000), thus can be solved numerically in a computationally efficient way. To show this let us partition

$$\bar{\boldsymbol{\beta}} = [\bar{\beta}_0 \ \bar{\boldsymbol{\beta}}_1']', \quad \boldsymbol{\beta} = [\beta_0 \ \boldsymbol{\beta}_1']'$$

and define $A \in \mathbb{R}^{v \times v}$ and $b \in \mathbb{R}^v$ as

$$A = \begin{bmatrix} 0 & 1 & & 0 \\ \vdots & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then minimizing (34) subject to (35) is equivalent to solving the semidefinite programming problem (Stoica et al., 2000):

$$\min_{\bar{\beta},\mathscr{Q}} \quad \mu$$

$$\text{s.t.} \quad \begin{bmatrix} \mu & \bar{\boldsymbol{\beta}}' - \boldsymbol{\beta}' \\ \bar{\boldsymbol{\beta}} - \boldsymbol{\beta} & I \end{bmatrix} > 0,$$

$$\begin{bmatrix} \mathscr{Q} - A\mathscr{Q}A' & \bar{\boldsymbol{\beta}}_1 - A\mathscr{Q}b \\ \bar{\boldsymbol{\beta}}_1 - b'\mathscr{Q}A' & \bar{\beta}_0 - b'\mathscr{Q}b \end{bmatrix} > 0$$

and $\hat{\boldsymbol{\beta}}$ is given by the argument minimizer with respect to $\bar{\boldsymbol{\beta}}$.

Let $\check{\varphi}_d(z)$ be the spectral density computed from $\hat{\boldsymbol{\beta}}$. An inverse bilinear transformation gives a refined version $\check{\varphi}(s)$ of

$\hat{\varphi}(s)$ such that $\check{\varphi}(i\omega) > 0$ for all real-valued ω. We compute $\check{\varphi}(s)$ as

$$\check{\varphi}(s) = \left. \frac{z^v}{(z+1)^{2v}} \check{\varphi}_d(z) \right|_{z=(1+s)/(1-s)} .$$

From $\check{\varphi}(s)$, we can compute a consistent estimate $\check{c}(s)$ of $c(s)$ by solving the spectral factorization problem

$$\check{\varphi}(s) = \check{c}(s)\check{c}(-s).$$

For a large enough observation length the condition (32) is satisfied since $\hat{\theta}_1$ comes sufficiently close to θ_1 due to the consistency properties. Then the regularization procedure is not needed. Therefore, the asymptotic covariance expression in Proposition 7 remains valid. Also, by the same reasoning, we can extend the result to derive the asymptotic covariance of the estimated c-parameters. This is done in the following proposition. Here, for the simplicity of notation, we extend the definitions of a_k, b_k and c_k so that $a_k = b_k = c_k = 0$ for $k > v$ and $k < 0$. Also, we define $b_0 = c_0 = 0$ and $a_0 = 1$.

Proposition 8. *Denote* $\theta_a = [a_1 \ \cdots \ a_v]'$, $\theta_b = [b_1 \ \cdots \ b_v]'$ *and* $\theta_c = [c_1 \ \cdots \ c_v]'$. *Also define*

$$\theta = [\theta_a' \ \theta_b' \ \theta_c']'$$

and let $\hat{\theta}$ *denote the estimate of* θ *derived in above. Let us define the* $v \times v$ *matrix* $\mathcal{V}(\theta_a)$ *as*

$$[\mathcal{V}(\theta_a)]_{jk} = (-1)^{v-k} a_{2j-k}. \tag{36}$$

We also define $\mathcal{V}(\theta_b)$ *and* $\mathcal{V}(\theta_c)$ *similarly. Then the asymptotic covariance matrix of* $\sqrt{T}\hat{\theta}$ *is given by*

$$C_{\hat{\theta}} := T\mathcal{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)'\} = V C_{\hat{\theta}_1} V',$$

where $C_{\hat{\theta}_1}$ *is given in Proposition 7, and*

$$V = \begin{bmatrix} I & 0 \\ 0 & I \\ \mathcal{V}^{-1}(\theta_c)\mathcal{V}(\theta_b) & \mathcal{V}^{-1}(\theta_c)\mathcal{V}(\theta_a) \end{bmatrix}. \tag{37}$$

Proof. Denote the estimation error in θ_a, θ_b and θ_c by $\tilde{\theta}_a$, $\tilde{\theta}_b$ and $\tilde{\theta}_c$, respectively. Also $\tilde{a}(s)$, $\tilde{b}(s)$ and $\tilde{c}(s)$ denote the associated perturbations in the polynomials $a(s)$, $b(s)$ and $c(s)$, respectively. Then, we have the following asymptotic perturbation expansion of the first order:

$$a(s)\tilde{b}(-s) + a(-s)\tilde{b}(s) + b(s)\tilde{a}(-s) + b(-s)\tilde{a}(s)$$

$$= c(s)\tilde{c}(-s) + c(-s)\tilde{c}(s). \tag{38}$$

Now, we equate the coefficients of $\{s^k\}_{k=0}^{2v}$ on both sides of (38). In matrix notation, we have

$$\{\mathcal{S}_1(\theta_a) + \mathcal{S}_2(\theta_a)\} \begin{bmatrix} 0 \\ \tilde{\theta}_b \end{bmatrix} + \{\mathcal{S}_1(\theta_b) + \mathcal{S}_2(\theta_b)\} \begin{bmatrix} 0 \\ \tilde{\theta}_a \end{bmatrix}$$

$$= \{\mathcal{S}_1(\theta_c) + \mathcal{S}_2(\theta_c)\} \begin{bmatrix} 0 \\ \tilde{\theta}_c \end{bmatrix}, \tag{39}$$

where $\mathcal{S}_1(\theta_a)$, $\mathcal{S}_2(\theta_a)$ are $(2v+1) \times (v+1)$ matrices defined element-wise as

$$[\mathcal{S}_1(\theta_a)]_{jk} = (-1)^{v-j+k} a_{j-k},$$

$$[\mathcal{S}_2(\theta_a)]_{jk} = (-1)^{v+1-k} a_{j-k}.$$

Then for an even j we see that

$$[\mathcal{S}_1(\theta_a)]_{jk} + [\mathcal{S}_2(\theta_a)]_{jk} = \{(-1)^{v-2v+k} + (-1)^{v+1-k}\}a_{j-k}$$

$$= \{(-1)^{v-k} + (-1)^{v+1-k}\}a_{j-k}$$

$$= 0.$$

Similarly, for an odd j we have

$$[\mathcal{S}_1(\theta_a)]_{jk} = (-1)^{v-2v-1+k} a_{j-k}$$

$$= (-1)^{v+1-k} a_{j-k} = [\mathcal{S}_2(\theta_a)]_{jk}.$$

Therefore, every even numbered row of (39) vanishes. Also the first row of (39) vanishes, because the coefficient of s^{2v+1} vanishes on both sides of (38). Now, retaining only the relevant terms in (39) corresponding to the third, fifth, ..., $(2v+1)$th rows we get

$$\mathcal{V}(\theta_a)\tilde{\theta}_b + \mathcal{V}(\theta_a)\tilde{\theta}_b = \mathcal{V}(\theta_c)\tilde{\theta}_c \ \Rightarrow \ \hat{\theta} - \theta = V(\hat{\theta}_1 - \theta_1),$$

where $[\mathcal{V}(\theta_a)]_{jk} = [\mathcal{S}_2(\theta_a)]_{2j+1,k+1} = (-1)^{v-k} a_{2j-k}$. Now (37) is straightforward from the last equality, since $\mathcal{V}(\theta_c)$ is nonsingular by construction (36). $\quad\square$

If the roots of $\hat{a}(s)$ obtained from the linear interpolation approach are in the left-half plane, then we get a consistent estimate of the CARMA process transfer function as $\check{c}(s)/\hat{a}(s)$. However, $\hat{a}(s)$ may have roots in the right-half plane in some rare occasions. In that case a second regularization step is required. One popular but heuristic approach to handle this problem is to reflect the unstable roots of $\hat{a}(s)$ to the left-half plane about the imaginary axis. Another possibility is to use the spectral zero assignability approach in Georgiou (1999). Note that we have a reliable estimate $\check{c}(s)$ of $c(s)$. Therefore, the convex optimization algorithm in Byrnes et al. (2001) can be used to re-estimate $a(s)$ using $v + 1$ interpolation data. A natural way to choose this subset of size $v + 1$ is be to pick up the interpolation data with lower statistical variation. The expression (27) can be used for that purpose. Numerical simulations show that the heuristic strategy of reflecting the unstable poles of $\hat{a}(s)$ to the left-half plane works as good as the Byrnes–Georgiou–Lindquist algorithm in Byrnes et al. (2001).

5. An illustrative example

In this section, we illustrate the proposed direct modeling approach using numerical simulations. To conduct the simulation, we first need to simulate the sampled version of a continuous-time stochastic process $u(t)$. We do so by using the method in Larsson (2004) and Söderström (2002) as follows. First, we

express the continuous-time process $u(t)$ in state space:

$$\dot{\zeta}(t) = A_c \zeta(t) + b_c e(t),$$

$$u(t) = c_c \zeta(t), \tag{40}$$

where $\zeta(t)$ is the state vector, and $e(t)$ is a unity variance continuous-time white noise, i.e., $\mathscr{E}\{e(t + \tau)e(t)\} = \delta(\tau)$. Note that $\delta(\cdot)$ is the Dirac's delta function. Let the process be sampled at a sampling interval t_s. We want to seek for an *equivalent* discrete-time state space model such that the second-order statistics of the discrete-time model output are the same as those of the underlying continuous-time process at the sampling instants. This discrete-time model is given by Larsson (2004), Söderström (2002)

$$\zeta\{(k + 1)t_s\} = e^{A_c t_s} \zeta(kt_s) + e_d(kt_s),$$

$$u(kt_s) = c_c \zeta(kt_s), \tag{41}$$

where $e_d(kt_s)$ is a vector-valued discrete-time white noise sequence with $\mathscr{E}\{e_d(t)e_d'(t)\} = R_d$. The covariance matrix R_d is related to the underlying continuous-time model via two Lyapunov equations as follows:

$$A_c Q + Q A_c' + b_c b_c' = 0, \tag{42}$$

$$Q = e^{A_c t_s} Q e^{A_c' t_s} + R_d. \tag{43}$$

More precisely, we need to solve for Q in (42), and subsequently compute R_d from (43). We point out that $Q = \mathscr{E}\{\zeta(t)\zeta'(t)\}$, which is easy to verify from (42).

In order to compute the estimates of the half-spectrum at the selected interpolation points we need to compute the output of the input-to-state filter using the samples $u(kt_s)$ at the sampling instants. This is done by a popular discretization technique known as state variable filtering (SVF); see Garnier and Young (2004) and references therein. In the discretization of the input-to-state filter, the input signal is assumed to vary linearly in between the sampling instants (commonly referred to as the first-order hold).

In the simulations, we consider a CARMA model with (see also (28))

$$a(s) = s^3 + 0.3s^2 + 9s + 0.9, \quad c(s) = s^2 + 0.5s + 6.$$

The observation time is 500 s, and the sampling interval is 0.05 s. This means that the total number of samples is 10^4. We estimate the half-spectrum at the interpolation points $\{1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 1\pm2i, 1\pm3i, 1\pm4i\}$. No derivative constraint is used in the estimation process. The interpolation data are then used to fit a CARMA model using the procedure outlined in Section 4. The performance of the proposed algorithm is compared with that of the so-called prediction error method (PEM)-based approach (Larsson & Mossberg, 2003; Söderström, 1991). PEM is an indirect method. Although PEM is known as the most accurate estimator, it does not always have a solution (Larsson & Mossberg, 2003; Söderström, 1991). Approximately 25% of our simulations fail to give a PEM solution. The results presented here are based on only
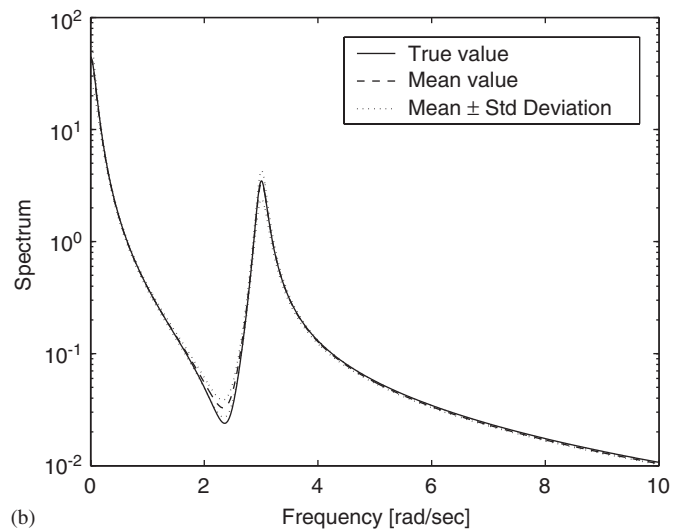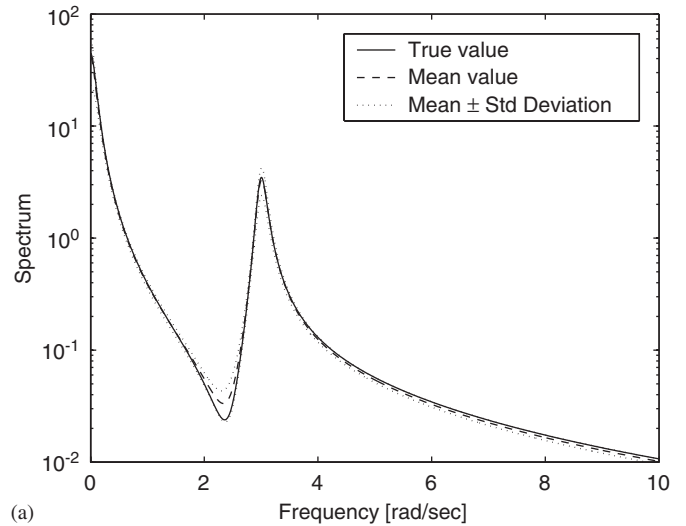

Fig. 1. Comparison of the mean of the estimated spectrum (dashed line) and the true spectrum (solid line). The mean $\pm$ standard deviation of the estimated spectrum is shown in dotted lines.

those realizations for which we get a PEM solution. The estimation results obtained from 100 Monte-Carlo simulations are shown in Fig. 1, where the true spectrum is compared with the estimated mean value $\pm$ standard deviation. As can be seen in Fig. 1, the estimation accuracy of the proposed approach is comparable with the prediction error method. For about 2% of the cases PEM deviates significantly from the true parameter vector (the optimization routine fails to reach the global minimum). These special cases are excluded in the evaluation of the statistical performance. In Table 1, we show the analytical and empirical standard deviations of the proposed estimates.

The performance of the proposed approach is similar to PEM as far as estimation of $\{a_k\}_{k=1}^3$ are concerned. However, PEM is slightly better than the proposed method in estimation of $\{c_k\}_{k=1}^3$ when it gives a solution. The analytical standard deviations of the proposed estimates match well with the empirical standard deviations in Table 1. The algorithms are implemented

Table 1
Parameter estimation performance of the proposed algorithm

	True value	Mean	Std.-Dev. (empirical)	Std.-Dev. (analytical)
a_1	0.3	0.3186	0.0396	0.0408
a_2	9	9.0331	0.1240	0.1126
a_3	0.9	0.9961	0.2393	0.1876
c_1	1	0.9740	0.0290	0.0224
c_2	0.5	0.5887	0.1201	0.1152
c_3	6	5.9368	0.2211	0.1792

using Matlab 6.1 on a 2.8 GHz Pentium IV processor with 1 GB RAM. The discrete-time PEM estimate is computed using the pem routine in the System Identification Toolbox (we do not provide any initial guess to the pem routine in our simulations; however the estimated model is constrained to be strictly proper). The average time required to compute the proposed estimate is 0.2 s, while it takes 1.0 s on average to compute the PEM estimate.

6. Conclusions

In this paper, we have proposed a novel direct approach for modeling continuous-time stochastic processes. The main idea is to use an input-to-state filter to compute the half-spectrum in some prescribed points in the right-half plane. The estimated samples of the half-spectrum are then used to obtain a rational model of the half-spectrum using linear interpolation with a positivity constraint. This is done by solving a semidefinite programming problem. The unique feature of the approach is twofold. Firstly, it is not required to estimate an intermediate discrete-time model. Thus, we can avoid many numerical difficulties associated with an indirect method. Secondly, it offers estimates which are comparable to PEM in terms of accuracy. However the PEM technique often fails to give a solution, which is not the case with the proposed method. One important open research question is to understand how the interpolation points affect the estimation results. More interpolation points tend to give better estimates (but at an expense of more computation). However, it is not clear how to best choose their locations.

References

Blomqvist, A., & Fanizza, V. (2003). Identification of rational spectral densities using orthonormal basis functions. In *IFAC symposium on system identification*, Rotterdam, The Netherlands, 2003.

Byrnes, C. I., Georgiou, T. T., & Lindquist, A. (2001). A generalized entropy criterion for Nevanlinna–Pick interpolation with degree constraint. *IEEE Transactions on Automatic Control*, 46(6), 822–839.

Delsarte, Ph., Genin, Y., Kamp, Y., & Van Dooren, P. (1982). Speech modeling and the trigonometric moment problem. *Philips Journal of Research*, 37, 277–292.

Fan, H., Söderström, T., Mossberg, M., Carlsson, B., & Zhou, Y. (1999). Estimation of continuous-time AR process parameters from discrete-time data. *IEEE Transactions on Signal Processing*, 47, 1232–1244.

Fan, H., Söderström, T., & Zhou, Y. (1999). Continuous-time AR process parameter estimation in presence of additive white noise. *IEEE Transactions on Signal Processing*, 3392–3398

Feuer, A., & Goodwin, G. C. (1996). *Sampling in digital signal processing and control*. Boston: Birkhäuser.

Franco, S. (2001). *Design with operational amplifiers and analog integrated circuits*. New York: McGraw-Hill.

Garnier, H., & Young, P. C. (2004). Time-domain approaches for continuous-time model identification from sampled data. In *Invited tutorial paper for the American control conference* (*ACC'2004*), Boston, MA, USA, June 2004.

Georgiou, T. T. (1999). The interpolation problem with a degree constraint. *IEEE Transactions on Automatic Control*, 44(3), 631–635.

Georgiou, T. T. (2001). Spectral estimation via selective harmonic amplification. *IEEE Transactions on Automatic Control*, 46(1), 29–42.

Georgiou, T. T. (2002a). Spectral analysis based on the state covariance: The maximum entropy spectrum and linear fractional parametrization. *IEEE Transactions on Automatic Control*, 47(11), 1811–1823.

Georgiou, T. T. (2002b). The structure of state covariances and its relation to the power spectrum of the input. *IEEE Transactions on Automatic Control*, 47(7), 1056–1066.

Goodwin, G. C., Middleton, R. H., & Poor, H. V. (1992). High-speed digital signal processing and control. *Proceedings of IEEE*, 80(2), 240–259.

Heuberger, P. S. C., Van den Hof, P. M. J., & Wahlberg, B. (Eds.). (2005). *Modelling and identification with rational orthogonal basis functions*. Berlin: Springer.

Janssen, P., & Stoica, P. (1988). On the expectation of the product of four matrix-valued Gaussian random variables. *IEEE Transactions on Automatic Control*, AC-33, 867–870.

Kailath, T. (1980). *Linear systems*. Englewood Cliffs, NJ: Prentice-Hall.

Kimura, H. (1987). Positive partial realization of covariance sequences. In: C.I. Byrnes, & A. Lindquist (Eds.), *Modeling, identification and robust control* (pp. 499–513). Amsterdam, The Netherlands: North-Holland.

Larsson, E. (2004). *Identification of stochastic continuous-time systems*. Ph.D. thesis, Uppsala dissertations from the Faculty of Science and Technology, Department of Information Technology, Uppsala University.

Larsson, E., & Mossberg, M. (2003). On possibilities for estimating continuous-time ARMA parameters. In *IFAC symposium on system identification*, Rotterdam, The Netherlands, August 27–29, 2003.

Larsson, E. K., & Söderström, T. (2002). Identification of continuous-time AR processes from unevenly sampled data. *Automatica*, 38(4), 709–718.

Ljung, L. (1999). *System identification—theory for the user*. 2nd ed., Upper Saddle River, NJ, USA: Prentice-Hall.

Mahata, K., & Fu, M. (2005). Modelling continuous-time stochastic processes using input-to-state filters. In *16th IFAC world congress*, Prague, 2005.

Makhoul, J. (1975). Linear prediction: A tutorial overview. *Proceedings of IEEE*, 63(4), 561–580.

Mari, J., Stoica, P., & McKelvey, T. (2000). Vector ARMA estimation: A reliable subspace approach. *IEEE Transactions on Signal Processing*, 48(7), 2092–2104.

Middleton, R. H., & Goodwin, G. C. (1990). *Digital control and estimation: A unified approach*. Englewood Cliffs, NJ: Prentice-Hall.

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*. 2nd ed., London: Springer.

Söderström, T., & Mossberg, M. (2000). Performance evaluation of methods for identifying continuous-time autoregressive processes. *Automatica*, 36, 53–59.

Söderström, T., & Stoica, P. (1989). *System identification*. Hemel Hempstead, UK: Prentice-Hall International.

Stoica, P., McKelvey, T., & Mari, J. (2000). MA estimation in polynomial time. *IEEE Transactions on Signal Processing*, 48(7), 1999–2012.

Wahlberg, B. (1990). The effect of rapid sampling in system identification. *Automatica*, 26, 167–170.

Kaushik Mahata received his M.E. degree in Signal Processing from Indian Institute of Science, Bangalore, India in 2000, and a Ph.D. degree in Signal Processing from Uppsala University, Sweden in 2003. Currently, he is a research academic in the University of Newcastle, Australia. His research interest includes estimation identification, and spectrum analysis.

Minyue Fu received his Bachelor's Degree in Electrical Engineering from the University of Science and Technology of China, Hefei, China, in 1982, and M.S. and Ph.D. degrees in Electrical Engineering from the University of Wisconsin-Madison in 1983 and 1987, respectively.

From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan. For the summer of 1989, he was employed by the Universite Catholoque de Louvain, Belgium, as a Maitre de Conferences Invited. He joined the Department of Electrical and Computer Engineering, the University of Newcastle, Australia, in 1989. Currently, he is a Chair Professor in Electrical Engineering and Head of School of Electrical Engineering and Computer Science. He was a Visiting Associate Professor at University of Iowa in 1995–1996 and a Visiting Professor at Nanyang Technological University, Singapore, 2002.

His main research interests include control systems, signal processing and communications. He has been an Associate Editor for the IEEE Transactions on Automatic Control, Automatica, and Journal of Optimization and Engineering. He is an Fellow of IEEE.