A Continuous-Time Linear System Identification Method for Slowly Sampled Data

Damián Marelli and Minyue Fu, *Fellow, IEEE*

Abstract—Both direct and indirect methods exist for identifying continuous-time linear systems. A direct method estimates continuous-time input and output signals from their samples and then use them to obtain a continuous-time model, whereas an indirect method estimates a discrete-time model first. Both methods rely on fast sampling to ensure good accuracy. In this paper, we propose a more direct method where a continuous-time linear model is directly fitted to the available samples. This method produces an exact model asymptotically, modulo some possible aliasing ambiguity, even when the sampling rate is relatively slow. We also state conditions under which the aliasing ambiguity can be resolved, and we provide experiments showing that the proposed method is a valid option when a slow sampling frequency must be used but a large number of samples is available.

Index Terms—Continuous time systems, identification, parameter estimation, sampled data systems.

I. INTRODUCTION

DYNAMICAL systems encountered in the physical world are usually of a continuous-time nature (see [1] for a list of examples in biology, economics and physics). However, when measuring input and output signals, we obtain discrete-time signals formed by samples of possibly noisy and prefiltered versions of these continuous-time signals. Continuous-time linear system identification involves identifying a continuous-time linear model using these samples [2], [3]. Compared to its discrete-time counterpart where abundant techniques are available (see, e.g., [4] and [5]), relatively less research has been done in this area. Available methods for continuous-time linear system identification are summarized below.

1) Indirect Methods [6]: In this approach, the sampled signals are used to identify a discrete-time version of the linear system first. This is done by using any standard discrete-time linear system identification technique. The discrete-time model so obtained is then converted to a continuous-time model by using any standard conversion method (e.g., bilinear (Tustin) approximation, matched pole-zero method, zero-order hold (ZOH) method, etc., [7]). The advantage of this approach is that the actual identification is done using discrete-time techniques which are well developed and their behaviors well understood.

However, a drawback of this approach is that the accuracy of the standard linear model conversion methods depends on the sampling frequency, which therefore can not be too small. On the other hand, a very high sampling frequency causes the poles of the discretized version of the system to cluster at $1 + j0$, making them difficult to identify.

2) Direct Methods [8]: In this case, the parameters of a continuous-time linear model are tuned to minimize an objective function constructed using the sampled signals. Generally speaking, the construction of the objective function requires (explicitly or implicitly) the estimation of the continuous-time signals from their samples. Therefore, for these methods to be accurate, the sampling frequency needs to be high (this requirement can be replaced by assumptions on the inter-sample behavior of the input and output signals). Based on the definition of the cost function, the direct methods can be further divided into *time-domain* and *frequency-domain* methods.

3) Time-Domain Methods: These methods aim to transform the linear system's differential equation into an algebraic (linear) equation. In theory, this can be achieved by introducing some kind of preprocessing of the continuous-time signals before the sampling operation (e.g., state variable filtering, integration). In practice, however, the preprocessing is implemented in the discrete-time domain, after the sampling operation. For this to be done, the available samples can be used to approximate the continuous-time signals using polynomial methods (e.g., ZOH, first-order hold (FOH)) [9], [10] or to approximate the continuous-time signals' derivatives [11], [12]. A main advantage of this approach is that the continuous-time parameters can be solved via a linear least-squares solution.

4) Frequency-Domain Methods [13], [14]: Here, the available samples are used to estimate the spectrum of the input and the output signals. As with time-domain methods, this requires some assumption on the continuous-time signals. An advantage of this approach is that it provides a convenient framework for characterizing the inter-sample behavior of the output signal based on that of the input, as well as for dealing with the case when the input signal is assumed to be bandlimited. However, a disadvantage of this approach is that the continuous-time parameters are obtained by solving a nonlinear optimization problem.

A problem related to the identification of a continuous-time linear system is that of estimating a linear model for a continuous-time stochastic process based on sampled data [15]–[17]. An approach for estimating the parameters of a continuous-time autoregressive moving average (CARMA) model was recently proposed in [18] and [19], and consists of estimating the model by fitting the discrete-time autocorrelation function to

that obtained from the observed samples. In this paper we extend this idea to propose a novel continuous-time system identification method. More precisely, our proposed identification strategy consists of two steps. In the first step, an averaged autocorrelation of the (possibly nonstationary) continuous-time input signal is identified to fit the discrete-time autocorrelation function estimated from the input samples.[1] In the second step, the identified averaged input autocorrelation is used to identify the desired continuous-time system to fit the discrete-time cross-correlation between the input and output samples. In comparison with other direct methods, our method also aims to tune the continuous-time parameters to minimize an objective function constructed from the available samples. However, the key difference to the existing direct methods is that we do not need to estimate the continuous-time signals. It turns out that the proposed method produces, in theory, the exact system model asymptotically, modulo some possible aliasing ambiguity, as the number of samples tends to infinity, for any sampling frequency. However, numerical problems associated with the finite quantization of samples introduce a lower bound on the sampling frequency. Nevertheless, numerical simulations show that this lower bound can be smaller than the -3 dB bandwidth of the system in practice. Hence, as we show using experimental results in Section IX-A, our proposed scheme has advantages over existing schemes when the sampling rate is relatively low and the number of samples is relatively large. Many lightly damped mechanical structures (such as flexible beams) have the properties of relatively high bandwidth, which calls for relatively low sampling rate, and relatively stable dynamic structure, which allows a large number of samples to be collected. Example 2 in Section IX-A describes such a structure where the proposed scheme is applicable.

The aliasing ambiguity mentioned above is generated by the well-known fact that a continuous-time linear model produced using input–output data sampled at a given frequency, always has poles ambiguous in the sense that their imaginary parts can be shifted by any integer multiple of the sampling frequency without changing the dynamic response of the linear system at the sampling points. In Section VI, we state conditions under which aliasing ambiguity can be resolved asymptotically without *a prior* knowledge of the region of the poles.

The rest of the paper is organized as follows. In Section II, we introduce some necessary background material. In Section III, we describe the continuous-time linear system identification problem, and we motivate the proposed approach which is described in Section IV. In Section V, we state the resulting identification criterion, and in Section VI we state conditions under which the aliasing ambiguity can be resolved. In Section VII, we derive an adaptive parameter optimization algorithm for solving the proposed identification criterion. In Section VIII, we address the estimation of the averaged input autocorrelation function, which is required by the proposed

[1]While this first step follows an idea similar to the one in [18] and [19], it differs in three aspects: 1) We allow the input signal to have nonstationary second order statistics; 2) We consider the presence of an anti-alias filter within the sampling device (see Remark 4); and 3) we use a different parametrization of the input autocorrelation function, to deal with the presence of this filter (see Section VIII).

identification strategy. Finally, some simulation results are presented in Section IX, and concluding comments are given in Section X. For the ease of readability, all proofs are contained in the Appendix.

II. PRELIMINARIES

Throughout the paper we use the following:

Notation: Discrete-time expressions (i.e., signals and linear systems) are denoted using bold letters and their continuous-time counterparts using nonbold letters. Also, time-domain expressions are denoted in lowercase and their frequency-domain counterparts in uppercase. Finally, vectors are denoted with a single underbar (except when appearing as superscripts) and matrices with a double underbar.

1) The $\mathcal{Z}$-Transform of a Vector: Let $x(t)$ be a continuous-time signal, $X(s) = \mathcal{L}\{x(t)\}$ its Laplace transform and $\underline{\mathbf{x}} = [x(0), \ldots, x((m-1)\tau)]^T$ be a vector of m samples of $x(t)$. Consider the set $\mathcal{M} = \{e^{j2\pi k/m}, k = 0, \ldots, m-1\}$, and define the $\mathcal{Z}$-transform $\mathbf{X} = \mathcal{Z}\{\underline{\mathbf{x}}\}$ of the vector $\underline{\mathbf{x}}$, as the map $\mathbf{X} : \mathcal{M} \to \mathbb{C}$ given by

$$\mathbf{X}(z) = \sum_{l=0}^{m-1} x(l+1)z^{-l}. \tag{1}$$

We define the operator $\mathcal{Z}_{\tau,m} : X \mapsto \mathbf{X}$ (i.e., $\mathcal{Z}_{\tau,m}$ maps the Laplace transform of a continuous-time signal into the $\mathcal{Z}$-transform of the vector formed by the first m samples obtained using a sampling period τ).

Remark 1: Notice that in our definition of the $\mathcal{Z}$-transform of a vector we only allow z to take values in $\mathcal{M}$. This restriction is required for the result in Lemma 3 (in Appendix A) to hold. Due to this restriction, (1) is equivalent to the discrete Fourier transform (DFT) of the vector $\underline{\mathbf{x}}$. The advantage of using the z notation is that it permits in many cases expressing $\mathbf{X}(z)$ as a rational polynomial function in z. Although this is a subtlety, it facilitates the derivation of the proposed algorithm (in particular the result in Lemma 2).

III. CONTINUOUS-TIME LINEAR SYSTEM IDENTIFICATION

The setting for continuous-time linear system identification is depicted in Fig. 1. The input signal $u(t)$ and the output noise $v(t)$ are uncorrelated random processes. The discrete-time signals $\mathbf{u}(k)$ and $\mathbf{y}(k)$ are generated by first filtering the input signal and the noisy output signal $y(t)$ using the anti-alias filters $H(s)$ and $L(s)$, respectively, and then sampling them with a sampling period τ. It is assumed that $G(s)$ is modeled by

$$
\begin{aligned}
G^\theta(s) &= \frac{A^{\theta_a}(s)}{B^{\theta_b}(s)} \\
&= \frac{a_0 + a_1 s + \cdots + a_{n_a} s^{n_a}}{1 + b_1 s + \cdots + b_{n_b-1} s^{n_b-1} + b_{n_b} s^{n_b}}
\end{aligned} \tag{2}
$$

where $\underline{\theta} = [\underline{\theta}_a^T, \underline{\theta}_b^T]^T$ is the vector of parameters, with $\underline{\theta}_a = [a_0, \ldots, a_{n_a}]^T$ and $\underline{\theta}_b = [b_1, \ldots, b_{n_b}]^T$.

The research problem is to estimate the parameters θ based on the sampled signals $\mathbf{u}(k)$ and $\mathbf{y}(k)$. To this end there are different approaches as explained in Section I above. Notice that,
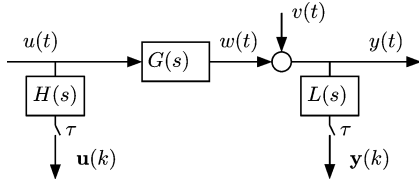
Fig. 1.   Continuous-time system identification setting.

generally speaking, these methods rely on the assumption that the signal $u(t)$ (resp. $y(t)$) is generated as follows:

$$u(t) = \sum_{k \in \mathbb{Z}} \mathbf{x}(k) f(t - k\tau) \tag{3}$$

for some function $f \in L_1(\mathbb{R})$ and some discrete-time sequence $\mathbf{x}(k)$. More precisely, assuming that the intersample behavior of $u(t)$ is ZOH or FOH is equivalent to assume that $f(t)$ is a zero or first order B-spline [20], [21], respectively. Also, assuming that $u(t)$ is band limited is equivalent to assume that $f(t)$ equals some properly scaled and dilated version of the $\mathrm{sinc}(t)$ function.

Under the assumption (3), the generating sequence $\mathbf{x}(k)$ can be reconstructed from the samples $\mathbf{u}(k) = u(\tau k)$, and in this way, the whole information about the continuous-time signal $u(t)$ (resp. $y(t)$) is contained in the samples $\mathbf{u}(k)$. However, a drawback of this assumption is that it is approximately correct only if the sampling frequency is very high. If this is not the case, a more suitable assumption is that $u(t)$ (resp. $y(t)$) is generated by

$$u(t) = \int_{\mathbb{R}} x(\xi) f_t(t - \xi) \, d\xi \tag{4}$$

where the generating function $x(t)$ is usually a zero-mean, white, continuous-time random process[2] and $f_t \in L_1(\mathbb{R})$ denotes the impulse response at time t of a linear time-varying filter. In this case, the function $x(t)$ cannot be reconstructed from the samples $\mathbf{u}(k)$, and therefore, $u(t)$ cannot be perfectly reconstructed. Hence, a different approach needs to be used.

IV. PROPOSED STOCHASTIC APPROACH

We use a stochastic approach based on second-order statistics. The second-order statistics of $u(t)$ (resp. $y(t)$) are determined by its time-varying correlation which is defined by

$$r_u(t', t) = \mathcal{E}\{u(t' + t)u(t')\} \tag{5}$$

with $\mathcal{E}\{\cdot\}$ denoting the expected value. Then, a characterization of the nonstationarity of $r_u(t', t)$ is given by the following condition.

1) Condition 1: There exist functions $\tilde{r}_u \in L_1(\mathbb{R})$ and $\mu : \mathbb{R} \to \mathbb{R}^+$ ($\mathbb{R}^+$ denotes the set of positive real numbers) such that, for all $t' \in \mathbb{R}$,

$$|r_u(t', t) - r_u(t' + \delta, t)| \leq \mu(\delta)\tilde{r}_u(t). \tag{6}$$

Now, we have the following result.

[2]Notice that, strictly speaking, a realization of such a random process may not be a Lebesgue measurable function, and therefore, (4) is not strictly correct. However, we ignore this technical subtlety which falls beyond the scope of this work.

Lemma 1: Let $u(t)$ be a continuous-time random process and let $a, b \in L_1(\mathbb{R})$ be two continuous-time functions. For each $k \in \mathbb{Z}$, define $\mathbf{v}(k) = (a * u)(\tau k)$ and $\mathbf{w}(k) = (b * u)(\tau k)$, and let

$$\mathbf{r}_{\mathbf{v}, \mathbf{w}}(k) = \lim_{N \to \infty} \frac{\tau}{N} \sum_{l=0}^{N-1} \mathcal{E}\{\mathbf{v}(l + k)\mathbf{w}(l)\}$$

be the (normalized) cross-correlation between $\mathbf{v}(k)$ and $\mathbf{w}(k)$. If $u(t)$ has uniformly bounded second moments (i.e., there exists $C > 0$ such that $\mathcal{E}\{u^2(t)\} < C$, for all $t \in \mathbb{R}$), then

$$\mathbf{r}_{\mathbf{v}, \mathbf{w}}(k) = (a * b^* * \bar{r}_u)(\tau k) + \boldsymbol{\varepsilon}(\tau k) \tag{7}$$

where $*$ denotes convolution and the superscript * denotes time-reversal (i.e., $b^*(t) = b(-t)$). The averaged autocorrelation function $\bar{r}_u(t)$ is defined by

$$\bar{r}_u(t) = \lim_{N \to \infty} \frac{\tau}{N} \sum_{n=0}^{N-1} r_u(n\tau, t) \tag{8}$$

and the nonstationary contribution $\boldsymbol{\varepsilon}(t)$ is defined in (41), Appendix A. Moreover, if $u(t)$ satisfies Condition 1, then, for all $t \in \mathbb{R}$,

$$|\boldsymbol{\varepsilon}(t)| \leq (|a| * |\mu \circ b|^* * \tilde{r}_u)(t), \tag{9}$$

where $\circ$ denotes pointwise multiplication (i.e., $\mu \circ b(t) = \mu(t)b(t)$), and $|\cdot|$ denotes absolute value.

Proof: See Appendix A. ∎

Remark 2: The function $\mu(t)$ in (6) states a bound on the rate of change of the second-order statistics of $u(t)$. Hence, for small values of t where these statistics remain approximately constant, we have that $\mu(t) \simeq 0$. Notice that, if this condition holds within a time interval longer than the settling-time of the impulse response $b(t)$, then $\mu(t)b(t) \simeq 0$. Hence, under this mild assumption, $\boldsymbol{\varepsilon}(t) \simeq 0$ and therefore (7) becomes

$$\mathbf{r}_{\mathbf{v}, \mathbf{w}}(k) \simeq (a * b^* * \bar{r}_u)(\tau k). \tag{10}$$

Now, in Fig. 1, let $\mathbf{r}_{\mathbf{u}}(k) = \mathbf{r}_{\mathbf{u}, \mathbf{u}}(k)$ be the autocorrelation of $\mathbf{u}(k)$ and $\mathbf{r}_{\mathbf{y}, \mathbf{u}}(k)$ be the cross-correlation between $\mathbf{y}(k)$ and $\mathbf{u}(k)$. Then, in view of Remark 2, if the statistics of $u(t)$ do not change within a time interval longer than the settling-time of $h(t) = \mathcal{L}^{-1}\{H(s)\}$, it follows that, for all $k \in \mathbb{Z}$,

$$\mathbf{r}_{\mathbf{u}}(k) \simeq (h * h^* * \bar{r}_u)(\tau k) \tag{11}$$

$$\mathbf{r}_{\mathbf{y}, \mathbf{u}}(k) \simeq (l * h^* * g * \bar{r}_u)(\tau k) \tag{12}$$

where $g(t)$ and $l(t)$ denote the inverse Laplace transforms of $G(s)$ and $L(s)$, respectively.

Suppose that $\bar{r}_u(t)$ is known. Then, (12) states a relationship between $g(t)$ and $\mathbf{r}_{\mathbf{y}, \mathbf{u}}(k)$. Since $\mathbf{r}_{\mathbf{y}, \mathbf{u}}(k)$ can be estimated from the available samples, (12) can be used to solve θ. In Section V, , we derive an identification criterion based on this idea, and in Section VII, we derive a parameter optimization algorithm for solving it.

The method outlined above relies on the knowledge of $\bar{r}_u(t)$. Depending on the application, this information can be known *a priori* (e.g., if $u(t)$ is a stationary random process with known

power spectrum), or otherwise estimated from the available samples. In Section VIII we use an argument similar to the one given above to show that $\bar{r}_u(t)$ can be estimated accurately from $\mathbf{r_u}(k)$ using (11), modulo some aliasing ambiguity.

V. PROPOSED IDENTIFICATION CRITERION

In this section we assume that the averaged autocorrelation function $\bar{r}_u(t)$ of the input signal $u(t)$ is known, and we use (12) to state a criterion to identify $\underline{\theta}$. We denote the impulse response of $G^\theta(s)$ by $g^\theta(t)$ and the parametric version of $\mathbf{r_{y,u}}(k)$ by $\mathbf{r}^\theta_{\mathbf{y,u}}(k)$, i.e.,

$$\mathbf{r}^\theta_{\mathbf{y,u}}(k) = (l * h^* * \bar{r}_u * g^\theta)(\tau k). \tag{13}$$

Also, we define an estimate $\mathbf{r}^{(N)}_{\mathbf{y,u}}(k)$ of $\mathbf{r_{y,u}}(k)$ by

$$\mathbf{r}^{(N)}_{\mathbf{y,u}}(k) = \frac{\tau}{\gamma(N)} \sum_{l=1}^{N} \lambda^{N-l} \mathbf{y}(l+k)\mathbf{u}(l) \tag{14}$$

where the *forgetting factor* λ is used to assign less weight to older measurements that are no longer representative of the (possibly slowly time-varying) parameters $\underline{\theta}$, and the scaling constant $\gamma(N) = \sum_{n=0}^{N-1} \lambda^{-n}$ is used so that, when the joint statistics of $\mathbf{x}(k)$ and $\mathbf{y}(k)$ are stationary, the expected value of $\mathbf{r}^{(N)}_{\mathbf{y,u}}(k)$ equals $\mathbf{r_{y,u}}(k)$ in (12). Hence, we can define the optimal vector of parameters $\underline{\theta}^{(N)}$ up to time N, as the one that makes $\mathbf{r}^\theta_{\mathbf{y,u}}(k)$ as close as possible to $\mathbf{r}^{(N)}_{\mathbf{y,u}}(k)$. Hence, in the frequency domain we have that [see (15), shown at the bottom of the page], where $\mathbf{R}^{(N)}_{\mathbf{y,u}}(z) := \mathcal{Z}\left\{\underline{\mathbf{r}}^{(N)}_{\mathbf{y,u}}\right\}(z)$ denotes the $\mathcal{Z}$-transform of the vector $\underline{\mathbf{r}}^{(N)}_{\mathbf{y,u}} = [\mathbf{r}^{(N)}_{\mathbf{y,u}}(0), \ldots, \mathbf{r}^{(N)}_{\mathbf{y,u}}(m)]$ and $\bar{R}_u(s) = \mathcal{L}\{\bar{r}_u(t)\}$ denotes the averaged power spectrum of $u(t)$. Now, if we assume that $L(s)$, $H(s)$, and $\bar{R}_u(s)$ have rational forms, we can write

$$L(s)H(-s)\bar{R}_u(s) = \frac{P(s)}{Q(s)}$$

and (15) becomes (16), shown at the bottom of the page.

Remark 3: To simplify the presentation above, we did not consider in (15) the values of $\mathbf{r}^{(N)}_{\mathbf{y,u}}(k)$ and $\mathbf{r}^\theta_{\mathbf{y,u}}(k)$ for negative values of k. This is a reasonable assumption since in most applications these functions vanish very quickly in the negative direction. However, if these values are to be considered, the analysis

above can be still applied by splitting $\mathbf{r}^\theta_{\mathbf{y,u}}(k)$ into its causal and (strictly) anticausal components.

Remark 4: Notice that the identification criterion (16) takes into account the presence of anti-alias filters within the input and output sampling devices. Although they are usually not considered in the continuous-time identification literature (e.g., [18], [19]), anti-alias filters are almost always present in real life, either implicitly or explicitly. These filters are often built with analog components which are subject to significant deviations from nominal values. For this reason, a mismatch between input and output filters often occurs, which if not taken into account, will induce a bias in the identified model. We give a numerical experiment illustrating this point in Section IX-D.

VI. CONSISTENCY

In this section, we address the question of consistency, i.e., whether the limit value of $\underline{\theta}^{(N)}$ tends to the true vector of parameters $\underline{\theta}^*$ as $N \rightarrow \infty$, i.e.,

$$\underline{\theta}^* = \lim_{N \rightarrow \infty} \underline{\theta}^{(N)}$$

or, $\lim_{N \rightarrow \infty} G^{\theta^{(N)}}(s) = G^{\theta^\star}(s)$. Unlike discrete-time system identification, the estimated continuous-time model based on sampled data is not always consistent. This is due to the problem of aliasing ambiguity. We will show this via an example of inconsistency later in this section.

The main result of this section is to provide conditions under which aliasing ambiguity can be resolved from the sampled data. Roughly speaking, we show that the estimate $\underline{\theta}^{(N)}$ is consistent when an anti-aliasing filter is used. To this end, we introduce the following result, which will also be used in Sections VII and VII-B to derive the proposed parameter optimization algorithm and its initialization, respectively.

Lemma 2: Let $N(s) = n_0 + n_1 s + \cdots + n_p s^p$, $D(s) = 1 + d_1 s + \cdots + d_q s^q$ and $X(s) = x_0 + x_1 s + \cdots + x_r s^r$ be polynomials on s of orders p, q and r, respectively. Let $\underline{\theta}_x = [x_0, x_1, \ldots, x_r]^T$ and for $c \in \mathbb{N}_0$ define

$$\underline{\boldsymbol{\Omega}}_c(z^{-1}) = [1, z^{-1}, \ldots, z^{-c}]. \tag{17}$$

If $r + p < q$, then

$$\mathcal{Z}_{\tau,m}\left\{\frac{N(s)X(s)}{D(s)}\right\}(z) = \frac{\underline{\boldsymbol{\Omega}}_{q-1}(z^{-1})\underline{\underline{M}}_{N,D}\,\underline{\theta}_x}{\mathbf{D}(z^{-1})}$$

$$\underline{\theta}^{(N)} = \arg\min_{\underline{\theta}} \sum_{k=0}^{m-1} \left| \mathbf{R}^{(N)}_{\mathbf{y,u}}(e^{j2\pi k/m}) - \mathcal{Z}_{\tau,m}\left\{L(s)H(-s)\bar{R}_u(s)G^\theta(s)\right\}(e^{j2\pi k/m}) \right|^2 \tag{15}$$

$$\underline{\theta}^{(N)} = \arg\min_{[\underline{\theta}_a, \underline{\theta}_b]} \sum_{k=0}^{m-1} \left| \mathbf{R}^{(N)}_{\mathbf{y,u}}(e^{j2\pi k/m}) - \mathcal{Z}_{\tau,m}\left\{\frac{P(s)}{Q(s)}\frac{A^{\theta_a}(s)}{B^{\theta_b}(s)}\right\}(e^{j2\pi k/m}) \right|^2. \tag{16}$$

where $\mathbf{D}(z^{-1})$ is the monic polynomial on z^{-1} of order q, whose roots $\mathbf{r}_i$, $i = 1, \ldots, q$ are obtained from the roots r_i, $i = 1, \ldots, q$ of $D(s)$ as follows:

$$\mathbf{r}_i = e^{-r_i \tau} \tag{18}$$

and the matrix $\underline{M}_{N,D} \in \mathbb{C}^{q \times (r+1)}$ (defined in Appendix B (50)) states a linear relationship between the coefficients $\underline{\theta}_x$ of $X(s)$ and those of the resulting discrete-time numerator.

Proof: See Appendix B. ∎

Now consider (16) and put $G(s) = A(s)/B(s)$. Let $V(s) = P(s)A(s)$, $W(s) = Q(s)B(s)$ and denote their orders by n_v and n_w, respectively. We have that

$$\lim_{N \to \infty} \mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(z) = \mathcal{Z}_{\tau,m} \left\{ \frac{V(s)}{W(s)} \right\}(z)$$
$$= \frac{\mathbf{V}(z^{-1})}{\mathbf{W}(z^{-1})} = \frac{\boldsymbol{\Omega}_{n_w-1}(z^{-1})\boldsymbol{\theta}_{\mathbf{v}}}{\mathbf{W}(z^{-1})} \tag{19}$$

where $\boldsymbol{\theta}_{\mathbf{v}}$ denotes the vector of coefficients of $\mathbf{V}(z^{-1})$. Also, using the notation $W^{\theta_b}(s) = Q(s)B^{\theta_b}(s)$, Lemma 2 gives

$$\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}(z) := \mathcal{Z}_{\tau,m} \left\{ \frac{P(s)A^{\theta_a}(s)}{W^{\theta_b}(s)} \right\}(z)$$
$$= \frac{\boldsymbol{\Omega}_{n_w-1}(z^{-1})\underline{M}_{P,W^{\theta_b}}\underline{\theta}_a}{\mathbf{W}^{\theta_b}(z^{-1})}. \tag{20}$$

From (19) and (20), the minimum of (16), as $N \to \infty$, is found when

$$\mathbf{W}^{\theta_b}(z^{-1}) = \mathbf{W}(z^{-1}) \tag{21}$$
$$M_{P,W^{\theta_b}}\underline{\theta}_a = \underline{\theta}_{\mathbf{v}}. \tag{22}$$

The vector $\underline{\theta}_b$ is not uniquely determined by (21). To see this notice that if $\underline{\theta}_b$ is a solution of (21), and the imaginary parts of the poles of its associated continuous-time denominator $B^{\theta_b}(s)$ are shifted by any integer multiple of the sampling frequency, to generate a new denominator $\tilde{B}^{\theta_b}(s)$, its corresponding vector of coefficients $\tilde{\underline{\theta}}_b$ is also a solution of (21). However, this aliasing ambiguity can be resolved if there is only one solution $\underline{\theta}_b$ to (21) which makes $\underline{\theta}_{\mathbf{v}}$ fall in the range of $\underline{M}_{P,W^{\theta_b}}$. Therefore, it can be resolved in general if $\underline{M}_{P,W^{\theta_b}}$ do not have full column range. Assuming that $G(s)$ is strictly proper (i.e., $n_a < n_b$), this condition is satisfied when $n_a < n_b - 1$ or $P(s)/Q(s) \neq 1$. Put in other words, aliasing ambiguity can be resolved when a nontrivial anti-aliasing filter is applied or $G(s)$ has relative degree greater than 1. If this is not the case, the ambiguity can only be resolved by *a prior* knowledge of the region of the poles of the true system $G(s)$.

A similar situation occurs when estimating the input autocorrelation $\bar{r}_u(t)$ using (11) (see Section VIII for more details).

To this end, we assume that $u(t)$ is generated by filtering a continuous-time white random process using a filter $F(s)$. By following a similar argument, aliasing ambiguity only exists if no input anti-alias filter is used and the numerator of $F(s)$ is one order smaller than its denominator. To illustrate this point, we consider the following example taken from [19], where the authors mistakenly concluded in [19, Section III-A] that the estimate $\underline{\theta}^{(N)}$ is always consistent, by stating that $\mathbf{R}_{\mathbf{u}}^{\theta}(z) = \mathbf{R}_{\mathbf{u}}^{\theta^*}(z) \Rightarrow \underline{\theta} = \underline{\theta}^*$ (recall that $\underline{\theta}^*$ denotes the true vector of parameters) without considering aliasing ambiguity. We will show that aliasing ambiguity exists in this case and the continuous-time model can not be uniquely identified.

1) Example 1: Let $F(s)$ be defined by

$$F(s) = \frac{2s+1}{s^2+2s+2}$$

which has poles at $-1 \pm j1$. Choose $\tau = 1$, $m = 4$ and suppose that no anti-alias filter is used when sampling $u(t)$. Using Lemma (3), it follows that [see (23), shown at the bottom of the page]. To find a continuous-time equivalent of $\mathbf{R}_{\mathbf{u}}^{(\infty)}(z)$ we do a residual-pole decomposition of (23), and we map each discrete-time pole to continuous-time by reversing the (18). Suppose that when inverting (18) we shift the imaginary parts of the continuous-time poles by $2\pi/\tau$, obtaining the poles $\pm 1 \pm j(1 + 2\pi)$ instead of $\pm 1 \pm j1$. After doing a spectral factorization of the result, we obtain

$$\tilde{F}(s) = \frac{2s+5.283}{s^2+2s+54.04}.$$

Since the only restriction on the numerator of $F(s)$ is that it is one order smaller than the denominator, all the $\tilde{F}(s)$ obtained by shifting the imaginary parts of the continuous-time poles by any integer multiple of the sampling frequency are valid solutions for $F(s)$. Therefore, the aliasing ambiguity cannot be resolved.

VII. PARAMETER OPTIMIZATION ALGORITHM

In this section we derive a parameter optimization algorithm for solving the proposed identification criterion (V). In Section VII-A, we derive an adaptive identification algorithm based on a quasi-Newton search method, whose initialization is presented in Section VII-B.

A. Adaptive Gradient Search Algorithm

The optimization problem (16) can be written as follows:

$$\underline{\theta}^{(N)} = \underset{\theta'}{\arg\min} V^{(N)}(\underline{\theta}'), \tag{24}$$

$$V^{(N)}(\underline{\theta}) = \sum_{k=0}^{m-1} \left| \mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(e^{j2\pi k/m}) - \mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}(e^{j2\pi k/m}) \right|^2 \tag{25}$$

$$\mathbf{R}_{\mathbf{u}}^{(\infty)}(z) = \lim_{N \to \infty} \mathbf{R}_{\mathbf{u}}^{(N)}(z) = \mathcal{Z}_{\tau,m} \left\{ |F(s)|^2 \right\}(z)$$
$$= \frac{-79.17z^{-1} + 249z^{-2} - 79.17z^{-3}}{7.389 - 26.64z^{-1} + 64.23z^{-2} + 24.64z^{-3} + 7.389z^{-4}}. \tag{23}$$

where

$$\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}(z) = \mathcal{Z}_{\tau,m}\left\{\frac{P(s)}{Q(s)}\frac{A^{\theta_a}(s)}{B^{\theta_b}(s)}\right\} \qquad (26)$$

is a function of the n-dimensional $(n = n_a + n_b + 1)$ vector $\underline{\theta}$. Using a quasi-Newton method, (24) and (25) is solved using the following iterative procedure:

$$\underline{\theta}^{(N+1)} = \underline{\theta}^{(N)} - \alpha^{(N)}\underline{\underline{B}}^{(N)}\underline{g}^{(N)}. \qquad (27)$$

The n-dimensional vector $\underline{g}^{(N)}$ denotes the gradient of $V(\theta)$ at $\underline{\theta}^{(N)}$, whose pth component is given by (θ_p denotes the pth entry of $\underline{\theta}$):

$$[\underline{g}^{(N)}]_p = 2\sum_{k=0}^{m-1}\frac{\partial\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta^{(N)}}}{\partial\theta_p}(e^{j2\pi k/m})$$
$$\times\left(\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}(e^{j2\pi k/m}) - \mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(e^{j2\pi k/m})\right). \quad (28)$$

The $n \times n$ matrix $\underline{\underline{B}}^{(N)}$ denotes an approximation of the inverse of the Hessian of $V(\theta)$ at $\underline{\theta}^{(N)}$, and is computed using the Broyden–Fletcher–Goldfarb–Shannon (BFGS) formula [22]:

$$\underline{\underline{B}}^{(N+1)} = \lambda\underline{\underline{B}}^{(N)} + \left(1 + \frac{\underline{q}_N^T\underline{\underline{B}}^{(N)}\underline{q}_N}{\underline{s}_N^T\underline{q}_N}\right)\frac{\underline{s}_N\underline{s}_N^T}{\underline{s}_N^T\underline{q}_N}$$
$$- \frac{\underline{s}_N\underline{q}_N^T\underline{\underline{B}}^{(N)} + \underline{\underline{B}}^{(N)}\underline{q}_N\underline{s}_N^T}{\underline{s}_N^T\underline{q}_N}$$
$$\underline{s}_N = \underline{\theta}^{(N+1)} - \underline{\theta}^{(N)}$$
$$\underline{q}_N = \underline{q}^{(N+1)} - \underline{q}^{(N)}. \qquad (29)$$

Finally, the scalar $\alpha^{(N)}$ denotes the step size at iteration N, which is obtained from a linear search algorithm in which, starting from the initial value $\alpha^{(N)} = 1$, the value of $\alpha^{(N)}$ is iteratively halved until

$$V(\underline{\theta}^{(N)} - \alpha^{(N)}\underline{\underline{B}}^{(N)}\underline{g}^{(N)}) < V(\underline{\theta}^{(N)}) \qquad (30)$$

or a maximum number of iterations is reached.

The computation of the gradient $\underline{g}^{(N)}$ in (28) requires the evaluation of the sample correlation $\mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(k)$ as well as $\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}(z)$ and its first-order derivatives

$$\nabla^{\theta}(z) = \left[\frac{\partial\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}}{\partial\theta_1}(z), \ldots, \frac{\partial\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}}{\partial\theta_{n_a+n_b}}(z)\right]. \qquad (31)$$

The sample correlation can be computed using the following update formula:

$$\mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(k) = \mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N-1)}(k)$$

$$+ \frac{1}{\gamma(N)}\left(\mathbf{y}(N+k)\mathbf{x}(N) - \mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N-1)}(k)\right) \quad (32)$$

(recall that $\gamma(N) = \sum_{n=0}^{N-1}\lambda^{-n}$). Also, as shown in Appendix C, using Lemma 2, we obtain

$$\mathbf{R}_{\mathbf{y},\mathbf{u}}^{\theta}(z) = \underline{\Sigma}_a^{\theta}(z)\underline{\theta}_a \qquad (33)$$

$$\underline{\nabla}^{\theta}(z) = \left[\underline{\Sigma}_a^{\theta}(z), -\underline{\Sigma}_b^{\theta}(z)\right] \qquad (34)$$

where the z-valued row vectors $\underline{\Sigma}_a^{\theta}(z) \in \mathbb{R}^{n_a+1}$ and $\underline{\Sigma}_b^{\theta}(z) \in \mathbb{R}^{n_b}$ are given by

$$\underline{\Sigma}_a^{\theta}(z) = \frac{\underline{\Omega}_{n_q+n_b-1}(z^{-1})\underline{\underline{M}}_{P,QB^{\theta_b}}}{\mathbf{Q}(z^{-1})\mathbf{B}^{\theta_b}(z^{-1})} \qquad (35)$$

$$\underline{\Sigma}_b^{\theta}(z) = \frac{\underline{\Omega}_{n_q+2n_b-1}(z^{-1})\underline{\underline{M}}'_{PA^{\theta_a},Q(B^{\theta_b})^2}}{\mathbf{Q}(z^{-1})\left(\mathbf{B}^{\theta_b}(z^{-1})\right)^2} \qquad (36)$$

with $\underline{\underline{M}}'$ denoting the matrix $\underline{\underline{M}}$ with its first column removed.

Remark 5: The memory of the adaptive algorithm (27) is determined by the forgetting factor λ in (29) (via the weight $\gamma(N) = \sum_{n=0}^{N-1}\lambda^{-n}$) and (32). This algorithm can be turned into batch one (i.e., nonadaptive) by choosing $\lambda = 1$.

B. Initialization

The recursive optimization algorithm (27) needs an initialization (or guess) of the unknown parameters. As concluded in Section IX-A below, sample-correlation fitting approaches like the one in [18] and [19] or our proposed continuous-time system identification method, are valid options when a slow sampling frequency must be used but a large number of samples is available. In this case, all other available methods fail, and hence they cannot be used to initialize (27). Therefore, the advantage of these approaches would be undermined if an initialization method is not provided. In the context of discrete-time system identification, a guess is often obtained using the method in [23]. In this section, we derive a method to obtain a guess for (16) by following the idea in [23].

Using Lemma 2, we can write (16) as (37), shown at the bottom of the page, where $\underline{\theta}_{\mathbf{b}}$ denotes the vector of coefficients of the discrete-time denominator $\mathbf{B}^{\theta_b}(z^{-1})$. Now, we can interpret (37) as a weighted least-squares problem with weighting factor $|\mathbf{B}^{\theta_b}(z^{-1})|^{-2}$. For the purposes of obtaining a guess, we can replace (37) by an iterative procedure in which, at iteration n, the discrete-time denominator $\mathbf{B}^{\theta_b}(z^{-1})$ is replaced by its value $\mathbf{B}^{(n-1)}(z^{-1})$ obtained using the parameters $\underline{\theta}_{\mathbf{b}}^{(n-1)}$ from the previous iteration. By doing so we can, in principle, jointly estimate the vector $[\underline{\theta}_a^T, \underline{\theta}_{\mathbf{b}}^T]^T$. However, a difficulty in doing so is that the matrix $\underline{\underline{M}}_{P,QB^{\theta_b}}$ depends (nonlinearly) on $\underline{\theta}_{\mathbf{b}}$

$$[\underline{\theta}_a, \underline{\theta}_{\mathbf{b}}] = \operatorname*{arg\,min}_{[\underline{\theta}_a, \underline{\theta}_{\mathbf{b}}]}\sum_{k=0}^{m-1}$$
$$\times\left|\frac{1}{\mathbf{B}^{\theta_b}(e^{-j2\pi k/m})}\left(\mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(e^{j2\pi k/m})(1 + \underline{\Omega}'_{n_b}(e^{-j2\pi k/m})\underline{\theta}_{\mathbf{b}}) - \frac{\underline{\Omega}_{n_q+n_b-1}(e^{-j2\pi k/m})\underline{\underline{M}}_{P,QB^{\theta_b}}\underline{\theta}_a}{\mathbf{Q}(e^{-j2\pi k/m})}\right)\right|^2 \qquad (37)$$

(via the mapping $\boldsymbol{\theta}_{\mathbf{b}} \mapsto \theta_b$), and therefore the problem, as opposite to that in [23], is still nonlinear in the parameters. In order to go around this issue, we define $\boldsymbol{\theta}_{\mathbf{P}} = \underline{M}_{P,QB^{\theta_b}} \underline{\theta}_a$ and we jointly estimate $[\boldsymbol{\theta}_{\mathbf{P}}^T, \boldsymbol{\theta}_{\mathbf{b}}^T]^T$, solving a linear least-squares problem. This results in the iterative procedure shown in (38) at the bottom of the page, which is initialized by choosing $\mathbf{B}^{(0)}(z^{-1}) = 1$. To retrieve the continuous-time parameters $[\theta_a^{(n)}, \theta_b^{(n)}]$ from $[\boldsymbol{\theta}_{\mathbf{P}}^{(n)}, \boldsymbol{\theta}_{\mathbf{b}}^{(n)}]$ we take $\underline{\theta}_b^{(n)}$ as the coefficients of the monic polynomial in s whose roots are obtained by reversing the pole-mapping (18) on the roots of the polynomial $\mathbf{B}^{(n)}(z^{-1}) = 1 + \underline{\boldsymbol{\Omega}}'_{n_b}(z^{-1})\underline{\theta}_{\mathbf{b}}^{(n)}$. Once $\underline{\theta}_b^{(n)}$ is obtained, in view of (16), (26) and (33), we can obtain $\underline{\theta}_a^{(n)}$ by solving the following linear least-squares problem:

$$\theta_a^{(n)} = \arg\min_{\underline{\theta}_a} \sum_{k=0}^{m-1} \left| \mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(e^{j2\pi k/m}) - \boldsymbol{\Sigma}_a^{\theta_b^{(n)}}(e^{j2\pi k/m})\underline{\theta}_a \right|^2 .$$

Notice that $\underline{\theta}_a^{(n)}$ and $\underline{\theta}_b^{(n)}$ need only be recovered at the end of the iteration process (38).

As its discrete-time counterpart [23], the recursive algorithm (38) is not guaranteed to converge, and the iterations are stopped when the minimization argument stops decreasing, which typically occurs after a small number of iterations.

Remark 6: It may happen that the polynomial $\mathbf{B}^{(n)}(z^{-1})$ have a real negative root. In this case, when reversing the pole-mapping (18), this pole needs to be split into two complex conjugated poles, in order to prevent $\underline{\theta}_b^{(n)}$ from having complex coefficients. By doing so, the order of $\underline{\theta}_b^{(n)}$ is increased by one for every real negative root of $\mathbf{B}^{(n)}(z^{-1})$. However, notice that in (27), the sizes of $\underline{\theta}_b^{(N-1)}$ and that of $\underline{\theta}_b^{(N)}$ do not need to be the same. Then, the order of $\underline{\theta}_b$ can be reduced after the first iteration of (27).

Remark 7: To resolve the aliasing ambiguity, following (22), we select the denominator coefficients $\underline{\theta}_b^{(n)}$ by shifting the identified poles by an integer multiple of the sampling (angular) frequency $\omega_s = 2\pi/\tau$, so that they minimize the difference between the vector of numerator coefficients of $\mathbf{R}_{\mathbf{y},\mathbf{u}}^\theta(z)$ and its projection onto the range of $\underline{M}_{P,W^{\theta_b^{(n)}}}$.

VIII. ESTIMATION OF AVERAGED INPUT AUTOCORRELATION

So far, we have assumed that the averaged autocorrelation function $\bar{r}_u(t)$ is known. However, this may not be a realistic assumption in certain applications. In this section we describe how to estimate $\bar{r}_u(t)$ using the same ideas described in Section V for estimating $g(t)$.

Recall that $\bar{R}_u(s) = \mathcal{L}\{\bar{r}_u(t)\}$ denotes the averaged power spectrum of $u(t)$. To estimate $\bar{R}_u(s)$ we use the following rational parametric model:

$$\begin{aligned} R_u^\zeta(s) &= \frac{U^{\zeta_u}(s)}{V^{\zeta_v}(s)} \\ &= \frac{u_0 + u_2 s^2 + \cdots + u_{2n_u} s^{2n_u}}{1 + v_2 s^2 + \cdots + v_{2n_v-2} s^{2n_v-2} + v_{2n_v} s^{2n_v}} \end{aligned}$$

where $\underline{\zeta} = [\underline{\zeta}_u^T, \underline{\zeta}_v^T]^T$ is the vector of parameters, with $\underline{\zeta}_u = [u_0, u_2, \ldots, u_{2n_u}]^T$ and $\underline{\zeta}_v = [v_2, \ldots, v_{2n_v}]^T$. Then, we consider the optimization criterion shown in (39) at the bottom of the page, where $\mathbf{R}_{\mathbf{u}}^{(N)}(z) = \mathcal{Z}\{\mathbf{r}_{\mathbf{u}}^{(N)}\}(z)$ denotes the $\mathcal{Z}$-transform of the vector $\mathbf{r}_{\mathbf{u}}^{(N)} = [\mathbf{r}_{\mathbf{y}}^{(N)}(0), \ldots, \mathbf{r}_{\mathbf{y}}^{(N)}(m)]$ of m samples of the estimate

$$\mathbf{r}_{\mathbf{u}}^{(N)}(k) = \frac{\tau}{\gamma(N)} \sum_{l=1}^N \lambda^{N-l} \mathbf{u}(l+k)\mathbf{u}(l)$$

and

$$\frac{S(s)}{T(s)} = H(s)H(-s).$$

The optimization problem (39) can be solved following the steps on Section VII, taking into account that the polynomials $U^{\zeta_u}(s)$ and $V^{\zeta_v}(s)$ have only terms of even powers.

IX. SIMULATIONS

In this section we analyze the performance of the proposed method. To this end, we use the Rao-Garnier benchmark system [6], which is given by

$$G(s) = \frac{-6400s + 1600}{s^4 + 5s^3 + 408s^2 + 416s + 1600}$$

$$\begin{aligned} [\boldsymbol{\theta}_{\mathbf{P}}^{(n)}, \boldsymbol{\theta}_{\mathbf{b}}^{(n)}] = \arg\min_{[\underline{\theta}_{\mathbf{P}}, \underline{\theta}_{\mathbf{b}}]} \sum_{k=0}^{m-1} \\ \times \left| \frac{1}{\mathbf{B}^{(n-1)}(e^{-j2\pi k/m})} \left(\mathbf{R}_{\mathbf{y},\mathbf{u}}^{(N)}(e^{j2\pi k/m})(1 + \underline{\boldsymbol{\Omega}}_{n_b}'(e^{-j2\pi k/m})\underline{\theta}_{\mathbf{b}}) - \frac{\underline{\boldsymbol{\Omega}}_{n_q+n_b-1}(e^{-j2\pi k/m})\underline{\theta}_{\mathbf{P}}}{\mathbf{Q}(e^{-j2\pi k/m})} \right) \right|^2 \end{aligned} \quad (38)$$

$$\underline{\zeta}^{(N)} = \arg\min_{[\underline{\zeta}_u, \underline{\zeta}_v]} \sum_{k=0}^{m-1} \left| \mathbf{R}_{\mathbf{u}}^{(N)}(e^{j2\pi k/m}) - \mathcal{Z}_{\tau,m}\left\{ \frac{S(s)}{T(s)} \frac{U^{\zeta_u}(s)}{V^{\zeta_v}(s)} \right\}(e^{j2\pi k/m}) \right|^2 \quad (39)$$
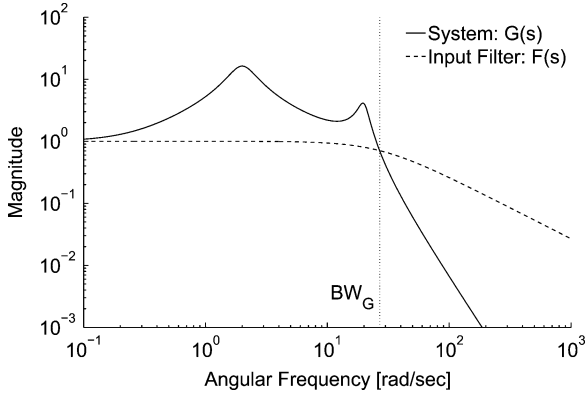
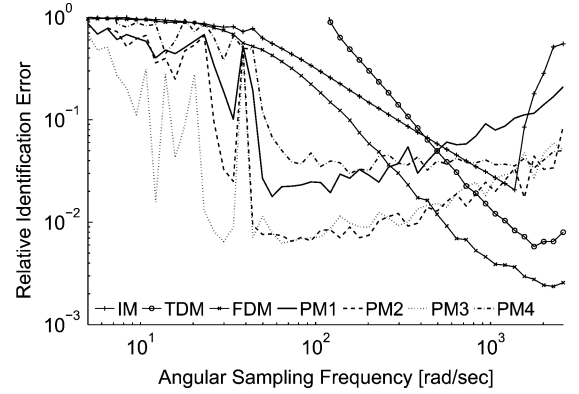Fig. 2.   Frequency responses of $G(s)$ and $F(s)$.

Fig. 3. Identification error versus sampling frequency for: indirect method (IM), time-domain method (TDM), frequency-domain method (FDM), and the proposed method with $N = 10^5$ (PM1), $N = 10^6$ (PM2), $N = 10^6$ and known $\bar{r}_u(t)$ (PM3), and a fixed sampling interval of $T = 2000$ s (PM4).
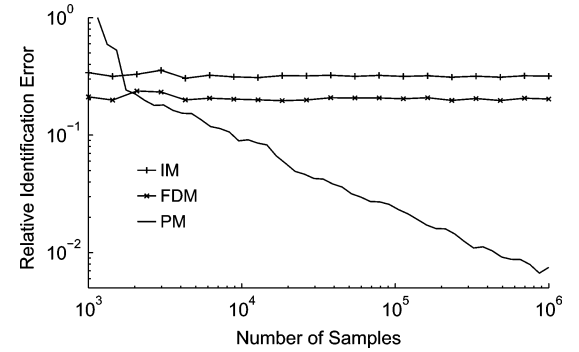

Fig. 4.   Identification error versus number of samples ($\omega_s = 100$).

and was used as a benchmark example in [2], [3], [9], [24], and [25]. This system has two pairs of complex conjugated poles at $-0.5 \pm j1.937$ and $-2 \pm j19.90$ and a -3 dB bandwidth of $BW_G = 26.85$ rad/s. As described in (4), the input signal $u(t)$ is generated by filtering white noise using

$$F(s) = \frac{26.85}{s + 26.85}$$

which has the same bandwidth as $G(s)$. The frequency responses of $G(s)$ and $F(s)$ are shown in Fig. 2. The performance comparison is done in terms of the identification error defined by

$$\varepsilon = \frac{\left\| G(s) - G^\theta(s) \right\|_2}{\left\| G(s) \right\|_2}. \tag{40}$$

A nontrivial issue is the generation of the discrete-time samples $\mathbf{u}(k)$ and $\mathbf{y}(k)$. To do so, let $\underline{\mathbf{\Phi}}(z)$ be the power spectrum of the discrete-time vector signal $[\mathbf{u}(k), \mathbf{y}(k)]^T$, and let $\underline{\mathbf{L}}(z)$ be a spectral factor (i.e., $\underline{\mathbf{\Phi}}(z) = \underline{\mathbf{L}}(z)\underline{\mathbf{L}}^*(z)$ with the superscript * denoting transpose conjugation) obtained using any of the methods described in [26]. Then, the vector $[\mathbf{u}(k), \mathbf{y}(k)]^T$ is generated by filtering a (vector) white noise using $\underline{\mathbf{L}}(z)$.

A. Comparative Performance Analysis

In this section we compare the performance of the proposed method with that of the methods described in Section I. For the indirect method, the discrete-time identification stage is done using the prediction error method [4], and the model conversion using the ZOH method. Notice that, in this way, the continuous-time model obtained will necessarily have its numerator one order smaller than its denominator (i.e., $n_a = n_b - 1 = 3$), which will make the comparison somehow "unfair" in the sense that the other methods use $n_a = 2$. For the direct time-domain method we use the state variable filter method [9] where the cutoff angular frequency of the state variable filters is chosen as $\lambda = BW_G = 26.85$ rad/s. For designing the frequency domain method we assume that the input signal is band limited (i.e., we use Approach 1 in [14]). For the proposed method, we choose m to satisfy $\tau m \geq 12$ s so that the sampling interval τm is longer than the settling-time of the impulse response of $\mathbf{F}(s)\mathbf{G}(s)$. We also use the method described in Section VIII to estimate the input autocorrelation $\bar{r}_u(t)$.

In the first experiment we compare the estimation error of the different methods as a function of the sampling angular frequency ω_s. To do so we eliminate the anti-alias filters $H(s)$ and $L(s)$ and the output noise $v(t)$ (see Fig. 1), and we average the identification error obtained over 10 runs. For the indirect method (IM), the time-domain method (TDM) and the frequency-domain method (FDM) we use $N = 10^5$ samples (as shown in Fig. 4 below, the performance of these methods does not significantly change with the number of samples). For the proposed method we use both $N = 10^5$ (PM1) and $N = 10^6$ samples (PM2). The results are shown in Fig. 3. We see that, when using $N = 10^5$, the proposed method is able to identify the system $\mathbf{G}(s)$ at frequencies above a threshold of about 40 rad/s. The reason for this is that the algorithm is unable to estimate $\bar{r}_u(t)$ below this threshold. This is to some extent overcome when using $N = 10^6$. However, the error of the proposed method still has peaks in the neighborhoods of $\omega_s = 40/k$ rad/s, $k = 1, 2, 3, \ldots$. The reason for this is that, at these angular sampling frequencies, due to aliasing, the pair of complex conjugated poles at $-2 \pm j19.90$ combines into a single real pole, preventing the proper identification of $G(s)$. This effect is more evident when using the proposed method assuming that $\bar{r}_u(t)$ is known (PM3). Also we see that, for frequencies above the 40 rad/s threshold, while $N = 10^6$ leads to a smaller identification error than the one obtained using $N = 10^5$, the error increases with an increase in ω_s. The reason

TABLE I
AVERAGED IDENTIFICATION ERROR AND
VARIANCE FOR $\omega_s = 300$ AND $N = 10^5$

	Identification Error	Error Variance
IM	99.09×10^{-3}	21.8×10^{-6}
TDM	139.6×10^{-3}	69.12×10^{-6}
FDM	32.86×10^{-3}	51.24×10^{-6}
PM	34.76×10^{-3}	220.0×10^{-6}

for this is that, increasing the sampling frequency while keeping constant the number of samples causes a reduction in the sampling interval $T = 2\pi N/\omega_s$. This is illustrated in Fig. 4 by using the proposed method with a fixed sampling interval of $T = 2000$ (PM4), showing that the error does not depend on the sampling frequency. As a side remark, we point out that the indirect method is unable to identify $\mathbf{G}(s)$ at sampling frequencies above a threshold of 1400 rad/s. This is due to the clustering of discretized poles at $1 + j0$, as mentioned in Section I.

In the second experiment, we evaluate the dependence of the identification error with the number of available samples N at $\omega_s = 100$. We do not consider the time-domain method in this simulation since the sampling frequency we use is too slow. We can see in Fig. 4 that the errors of the indirect and frequency-domain methods do not change significantly with N. On the other hand, the error of the proposed method diminishes with an increase on N, because it determines how well the sample correlation estimate $\mathbf{r}_{\mathbf{y},\mathbf{u}}^{(N)}(k)$ approaches the true correlation $\mathbf{r}_{\mathbf{y},\mathbf{u}}(k)$, and it becomes smaller that the errors of the other methods when $N \gtrsim 2000$.

We conclude from Figs. 3 and 4 that the proposed method outperforms the other methods when the sampling frequency is slow and the number of samples is large. In the opposite situation when the sampling frequency is high and the number of available samples is small, the other methods outperform the proposed method.

We show in Table I the identification error and its variance, averaged over 100 runs, obtained at $\omega_s = 300$ and using $N = 10^5$ samples. We see that the identification error variance produced by the proposed method is significantly larger than that produced by the other methods.

Example 2: Potential applications of the proposed method include lightly damped mechanical structures. For example, a flexible beam studied in [27, Fig. 11], which is a good generic model for many flexible structures, has three dominant modes and a bandwidth of 300 Hz. Suppose a sampling rate of 1 kHz is used, which is the typical rate used for active vibration control. Then, for collecting $N = 10 \times 10^3$ to $N = 60 \times 10^3$ samples, the system is required to be structurally stationary for ten seconds to one minute, which is a reasonable scenario.

B. Computational Cost Analysis

In this section we analyze the complexity of the proposed method. We use the number of real multiplications as the complexity index. More precisely, a real division counts as one real multiplication, a complex multiplication requires four real multiplications and a complex division requires eight. Also, the so-
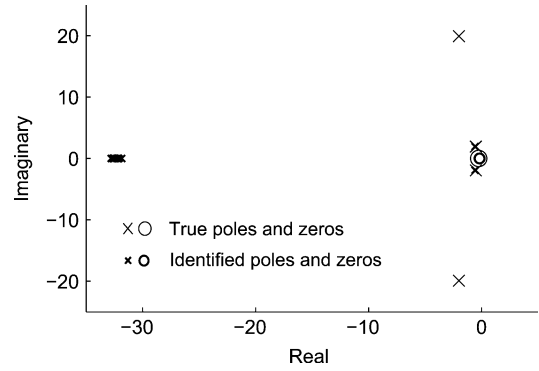

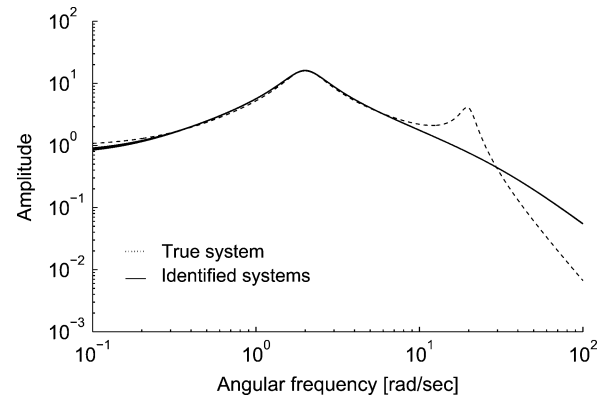Fig. 5. Effect of undermodeling in the proposed method: pole-zero map.

Fig. 6. Effect of undermodeling in the proposed method: frequency response.

lution of the system of m equations with n unknowns (e.g., to solve each iteration of (38)), requires approximately mn^2 multiplications [28, Ch. 5.3], and finding the roots of a polynomial or degree n requires about $10n^3$ multiplications, when using the QR algorithm to find the eigenvalues of a companion matrix [28, Ch. 7.5.6]. Using these approximations, and recalling that m denotes the number of frequency points in (16) and (38), and $n = n_a + n_b + 1$ denotes the total number of estimated parameters, each iteration of the initialization algorithm (38) requires $2mn^2$ multiplications. Also, each iteration of the parameter optimization algorithm (27) requires about $18mn_b^3$ multiplications, and each iteration of the linear search algorithm (30) requires $2mn_b^3$.

We evaluate the complexity of the proposed method, averaged over 100 runs, obtained at $\omega_s = 300$ and using $N = 10^5$ samples. We have that $n = 6$ and $m = 1200$. The initialization algorithm requires, on average, three iterations, and therefore requires about 259.2×10^3 multiplications. Also, the parameter optimization algorithm requires 11.52 main iterations and 113.8 linear search iterations, requiring 33.40×10^6 multiplications. Hence, on average, the proposed method has an overall complexity of about 33.66×10^6 multiplications.

C. Effects of Undermodeling and Overmodeling

In this section we study the effects of undermodeling and overmodeling in the proposed identification method. To study the effects of undermodeling, we reduce the number n_b of poles from 4 to 3. In Figs. 5 and 6 we see the locations of poles and
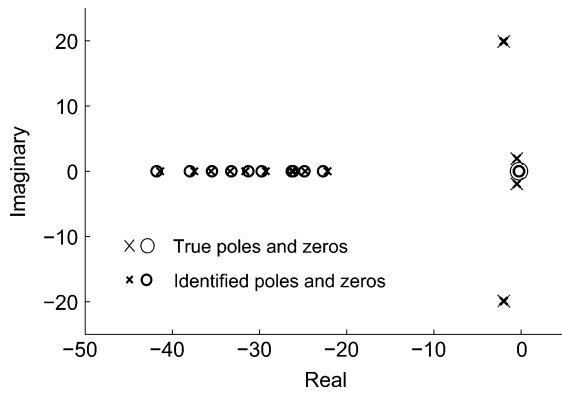
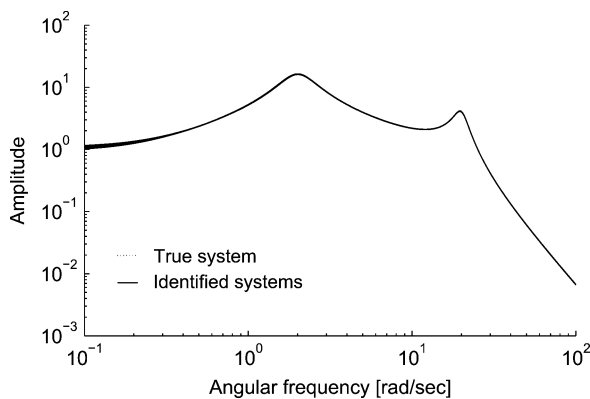Fig. 7.   Effect of overmodeling in the proposed method: pole-zero map.

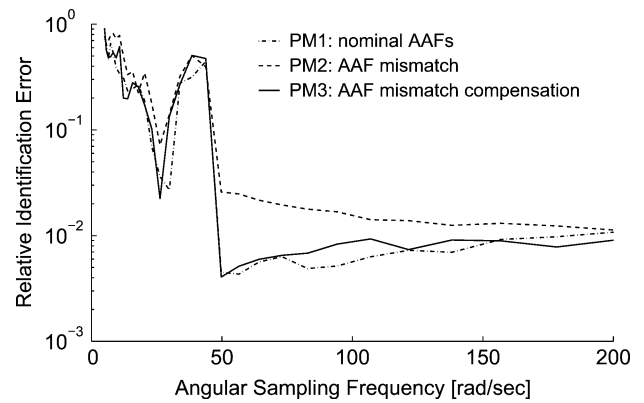Fig. 8.   Effect of overmodeling in the proposed method: frequency response.

Fig. 9.   Effect of using anti-alias filters (AAF) ($N = 10^6$ samples).
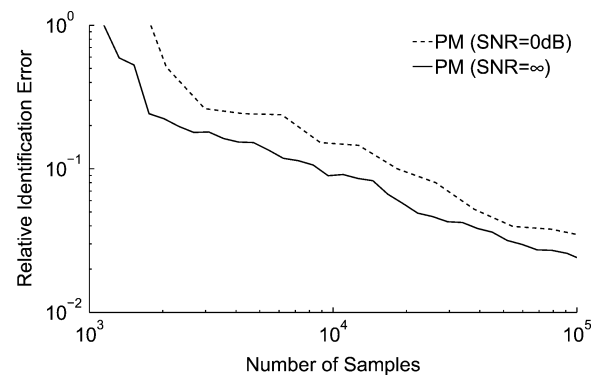
Fig. 10.   Identification error versus number of samples with output noise ($\omega_s = 100$).

zeros, as well as frequency responses, of the true system and of those systems identified using ten runs of the proposed method. We use $N = 10^6$ samples and $\omega_s = 100$. We see that the algorithm properly estimates the zero of the true system, as well as the low frequency poles. Also, the remaining pole is allocated at around 32 rad/s, in an attempt to approximate with a single pole, the slope change produced by the pair of complex conjugated poles at $-2 \pm j19.90$.

In order to see the effect of overmodeling, we increase the number n_b of poles to 5 and the number n_a of zeros from 1 to 2. We see in Figs. 7 and 8 that the method properly identifies the poles and zeros of the system, and the extra pole is placed near the extra zero so that their effects cancel each other.

D. Using Anti-Alias Filters

In this section we evaluate the identification performance when the anti-alias filters $H(s)$ and $L(s)$ are used. We consider three scenarios. In the first (PM1), we design the anti-alias filters using second order Butterworth low-pass filters with cutoff angular frequency $0.4 \times \omega_s$, and we consider them in the identification criterion (15). As mentioned in Remark 4, a mismatch between input and output anti-alias filters often occurs. To simulate this situation, we perturbed the poles of the output anti-alias filter by adding a zero mean Gaussian random variable with standard deviation 0.02 to its real and imaginary

components.[3] In the second scenario (PM2), we carry out the identification without reflecting this perturbation in identification criterion (15), which results in a degraded performance. On the other hand, we consider this perturbation within (15) in the third scenario (PM3), which exhibits a performance similar to that of scenario PM1. This indicates that the accuracy of the proposed method is not significantly affected by the presence of anti-alias filters, if they are properly taken into account within the identification criterion.

E. Presence of Output Noise

In this section, we analyze the performance of the proposed method in the presence of output noise $v(t)$. Output noise has the effect of corrupting the sample correlation estimate $\mathbf{r}_{\mathbf{y},\mathbf{u}}^{(N)}(k)$ in (14), so that a larger number of samples is required for $\mathbf{r}_{\mathbf{y},\mathbf{u}}^{(N)}(k)$ to properly approach the true correlation $\mathbf{r}_{\mathbf{y},\mathbf{u}}(k)$. However, its effect vanishes asymptotically as the number of samples N goes to infinity. This is shown in Fig. 10 where we compare the accuracy of the proposed method with an output signal-to-noise ratio (SNR) of 0 dB and without output noise (SNR $= \infty$).

[3]We have only perturbed the poles of the output anti-alias filter, and not those of the input filter, in order to avoid degrading the identification of the input power spectrum using VIII. By doing so, we are able to observe only the degradation of the system identification criterion (15).
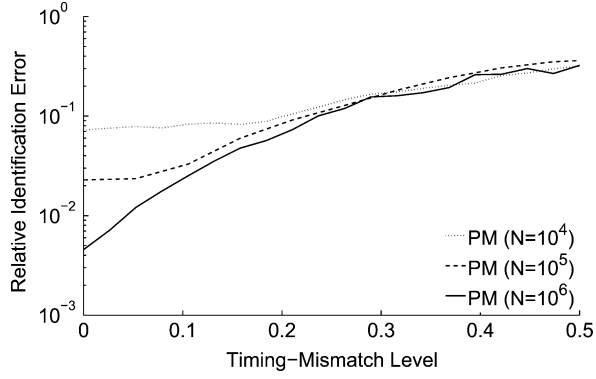
Fig. 11. Identification error versus timing mismatch level ($\omega_s = 100$, $N_s = 10^5$).

F. Presence of Sampling Timing Mismatch

In the last simulation we study the effect timing errors in the sampling devices. To this end we assume that the irregular sampling grid is periodic with a five-sample period. This corresponds to the sampling pattern which occurs when using time-interleaved analog-to-digital converters affected by timing mismatch errors [29]. We generate this mismatch by adding a zero mean Gaussian random variable with standard deviation σ to the ideal sampling time, and we define the *timing mismatch level* by σ/τ (recall that τ denotes the sampling time). In Fig. 11, we show the identification error for different timing mismatch levels and different number of samples. We see that the identification error of the proposed method decreases with the timing mismatch level, up to a threshold which is given by the number of samples.

X. CONCLUSION

We have proposed a novel identification method that directly fits a continuous-time linear model to the given sampled input and output signals. In this way, the proposed method is able to produce asymptotically, as the number of samples approaches infinity, the exact model of the linear system being identified, modulo a possible aliasing ambiguity which can be resolved under some given conditions. This is valid in theory, for any sampling frequency. In practice, however, numerical problems introduce a lower bound on the sampling frequency. Nevertheless, simulation results show that this lower bound can be smaller than the -3-dB bandwidth of the system. In comparison with the available methods in the literature, the proposed method is a valid option when a slow sampling frequency must be used but a large number of samples is available.

APPENDIX A
PROOF OF LEMMA 1

We have that

$$
\begin{aligned}
\mathbf{r}_{\mathbf{v},\mathbf{w}}(k) &= \lim_{N \to \infty} \frac{\tau}{N} \sum_{l=0}^{N-1} \mathcal{E}\{\mathbf{v}(l+k)\mathbf{w}(l)\} \\
&= \lim_{N \to \infty} \frac{\tau}{N} \sum_{l=0}^{N-1} \mathcal{E}\{(a*u)(\tau(l+k))(b*u)(\tau l)\} \\
&= \lim_{N \to \infty} \frac{\tau}{N} \sum_{l=0}^{N-1} \mathcal{E}\left\{ \int_{\mathbb{R}} a(t)u(\tau(l+k)-t)dt \right. \\
&\qquad\qquad \left. \int_{\mathbb{R}} b(t')u(\tau l - t')dt' \right\}.
\end{aligned}
$$

Now, since $a, b \in L_1(\mathbb{R})$ and $\mathcal{E}\{u^2(t)\} < C$ for all $t \in \mathbb{R}$, in view of Fubini's theorem we can exchange the expectation with the integrations and in view of Lebesgue's dominated convergence theorem we can exchange the limit with the integrations. By doing so,

$$
\begin{aligned}
\mathbf{r}_{\mathbf{v},\mathbf{w}}(k) &= \int_{\mathbb{R}} \int_{\mathbb{R}} a(t)b(t') \lim_{N \to \infty} \frac{\tau}{N} \\
&\quad \times \sum_{l=0}^{N-1} \mathcal{E}\{u(\tau(l+k)-t)u(\tau l - t')\} \, dt \, dt'.
\end{aligned}
$$

Using (5), we can write

$$
\begin{aligned}
\mathcal{E}\{u(\tau(l+k)-t)u(\tau l - t')\} &= r_u(\tau L, \tau k + t' - t) \\
&\quad + r_u(\tau l - t', \tau k + t' - t) - r_u(\tau l, \tau k + t' - t).
\end{aligned}
$$

Now, using (8) and defining (41), shown at the bottom of the page, we have that

$$
\begin{aligned}
\mathbf{r}_{\mathbf{v},\mathbf{w}}(k) &= \int_{\mathbb{R}} \int_{\mathbb{R}} a(t)b(t')\bar{r}_u(\tau k + t' - t)dt \, ds + \boldsymbol{\varepsilon}(\tau k) \\
&= \int_{\mathbb{R}} b(t')(a*\bar{r}_u)(\tau k + t')ds + \boldsymbol{\varepsilon}(\tau k) \\
&= (a*b^* * \bar{r}_u)(\tau k) + \boldsymbol{\varepsilon}(\tau k).
\end{aligned}
$$

To show (9), from (6) we have that

$$
\begin{aligned}
|r_u(\tau l - t', \tau k + t' - t) &- r_u(\tau l, \tau k + t' - t)| \\
&\leq \mu(t')\tilde{r}_u(\tau k + t' - t).
\end{aligned}
$$

Then,

$$
\begin{aligned}
|\boldsymbol{\varepsilon}(\xi)| &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} |a(t)||b(t')|\mu(t')\tilde{r}_u(\xi + t' - t)dt \, dt' \\
&= \int_{\mathbb{R}} |b(s)|\mu(t')(|a| * \tilde{r}_u)(\xi + t')|dt' \\
&= (a * |\mu \circ b|^* * \tilde{r}_u)(\xi).
\end{aligned}
$$

$$
\boldsymbol{\varepsilon}(\xi) = \int_{\mathbb{R}} \int_{\mathbb{R}} a(t)b(t') \lim_{N \to \infty} \frac{\tau}{N} \sum_{l=0}^{N-1} (r_u(\tau l - t', \xi + t' - t) - r_u(\tau l, \xi + t' - t)) \, dt \, dt', \tag{41}
$$

$$\frac{\partial \mathbf{R}^\theta_{\mathbf{y},\mathbf{u}}}{\partial \theta_i}(z) = \mathcal{F}_{\tau,m}\left\{\frac{P(s)}{Q(s)B^{\theta_b}(s)}\frac{\partial A^{\theta_a}(s)}{\partial \theta_i} - \frac{P(s)A^{\theta_a}(s)}{Q(s)\left(B^{\theta_b}(s)\right)^2}\frac{\partial B^{\theta_b}(s)}{\partial \theta_i}\right\}.$$

$$\underline{\nabla}^\theta(z) = \left[\mathcal{F}_{\tau,m}\left\{\frac{P(s)\underline{\Omega}_{n_a}(s)}{Q(s)B^{\theta_b}(s)}\right\}, -\mathcal{F}_{\tau,m}\left\{\frac{P(s)A^{\theta_a}(s)\underline{\Omega}'_{n_b}(s)}{Q(s)\left(B^{\theta_b}(s)\right)^2}\right\}\right]$$

APPENDIX B
PROOF OF LEMMA 2

We need the following lemma:

Lemma 3: For all $a \in \mathbb{C}$, the following relation holds:

$$\mathcal{Z}_{\tau,m}\left\{\frac{1}{s-a}\right\} = \frac{e^{-a\tau}(e^{a\tau m}-1)}{z^{-1}-e^{-a\tau}} \quad (42)$$

i.e., $\mathcal{Z}_{\tau,m}$ maps a single pole in the Laplace domain into a single pole in the $\mathcal{Z}$-transformed domain.

Proof: We have that

$$\mathcal{L}^{-1}\left\{\frac{1}{s-a}\right\} = e^{at}.$$

Hence, it is straightforward to verify that

$$\mathcal{Z}_{\tau,m}\left\{\frac{1}{s-a}\right\} = \sum_{l=0}^{m-1} e^{a\tau l}z^{-l} = \frac{e^{-a\tau}(e^{a\tau m}z^{-m}-1)}{z^{-1}-e^{-a\tau}}.$$

Finally, (42) follows since we only let z take values in $\mathcal{M} = \{e^{j2\pi k/m}, k=0,\ldots,m-1\}$. ■

To simplify the proof of Lemma 2 we assume that $D(s)$ has distinct roots. Define $Y(s) = N(s)X(s) = \underline{\Omega}_{q-1}(s)\theta_y$ ($\underline{\Omega}_{q-1}(s)$ is defined as in (17) but with s replacing z^{-1}), and do a residual-pole decomposition of $Y(s)/D(s)$ to obtain

$$\frac{N(s)X(s)}{D(s)} = \frac{Y(s)}{D(s)} = \sum_{i=1}^r \frac{k_i}{s-r_i}. \quad (43)$$

Defining $\underline{\boldsymbol{\xi}}_k = [k_1,\ldots,k_r]^T$ it follows that

$$\theta_y = \underline{\Gamma}_D \underline{\xi}_k \quad (44)$$

with the matrix $\underline{\Gamma}_D \in \mathbb{C}^{q\times q}$ given by $\underline{\Gamma}_D = [\underline{\Gamma}_{D,1},\ldots,\underline{\Gamma}_{D,q}]$, where for each $i=1,\ldots,q$, $\underline{\Gamma}_{D,i} = [\gamma_{D,i,0},\ldots,\gamma_{D,i,q}]^T$ is the vector of coefficients of the monic polynomial obtained from $D(s)$ by eliminating its ith root, i.e.,

$$\Omega_q(s)\Gamma_{D,i} = \gamma_{D,i,0} + \cdots + \gamma_{D,i,q}s^q = \prod_{j=1,j\neq i}^p (s-r_i).$$

Now, combining (43) and (42), we can write

$$\mathcal{Z}_{\tau,m}\left\{\frac{N(s)X(s)}{D(s)}\right\} = \sum_{i=1}^r \frac{\mathbf{k}_i}{z^{-1}-\mathbf{r}_i} \quad (45)$$

with $\mathbf{k}_i = e^{-r_i\tau}(e^{r_i\tau m}-1)$ and $\mathbf{r}_i = e^{-r_i\tau}$. Defining $\underline{\boldsymbol{\xi}}_\mathbf{k} = [\mathbf{k}_1,\ldots,\mathbf{k}_r]^T$ we have that

$$\underline{\boldsymbol{\xi}}_\mathbf{k} = \underline{\Lambda}_\mathbf{D}\underline{\xi}_k \quad (46)$$

with $\underline{\Lambda}_\mathbf{D} \in \mathbb{C}^{q\times q}$ given by

$$\underline{\Lambda}_\mathbf{D} = \text{diag}\left\{\mathbf{r}_1\left(\mathbf{r}_1^{-\tau m}-1\right),\ldots,\mathbf{r}_q\left(\mathbf{r}_q^{-\tau m}-1\right)\right\}.$$

Taking common denominator on the right-hand side of (45), we obtain

$$\mathcal{Z}_{\tau,m}\left\{\frac{N(s)X(s)}{D(s)}\right\} = \frac{\mathbf{Y}(z^{-1})}{\mathbf{D}(z^{-1})} = \frac{\underline{\Omega}_{q-1}(z^{-1})\underline{\theta}_\mathbf{y}}{\mathbf{D}(z^{-1})} \quad (47)$$

where $\underline{\theta}_\mathbf{y}$ is the vector of parameters of $\mathbf{Y}(z^{-1})$ (i.e., $\mathbf{Y}(z^{-1}) = \underline{\Omega}_{q-1}(z^{-1})\underline{\theta}_\mathbf{y}$). It follows that

$$\underline{\theta}_\mathbf{y} = \underline{\Gamma}_\mathbf{D}\underline{\xi}_\mathbf{k} \quad (48)$$

with the matrix $\underline{\Gamma}_\mathbf{D}$ constructed in a way similar to $\underline{\Gamma}_D$, but considering $\mathbf{D}(z^{-1})$ instead of $D(s)$.

Finally, it is easy to verify that

$$\underline{\xi}_y = \underline{\Psi}_N\underline{\xi}_x \quad (49)$$

with $\underline{\Psi}_N \in \mathbb{C}^{q\times(r+1)}$ given by

$$\left[\underline{\Psi}_N\right]_{i,j} = \begin{cases} n_{i-j}, & 0 \leq i-j \leq p \\ 0, & \text{Otherwise.} \end{cases}$$

Then, the result follows putting (44), (46), (48), and (49) into (47) and defining

$$\underline{M}_{N,D} = \underline{\Gamma}_\mathbf{D}\underline{\Lambda}_\mathbf{D}\underline{\Gamma}_D^{-1}\underline{\Psi}_N. \quad (50)$$

APPENDIX C
COMPUTATION OF $\mathbf{R}^\theta_{\mathbf{y},\mathbf{u}}$ AND $\underline{\nabla}^\theta$

From (26) and (17), we have that

$$\mathbf{R}^\theta_{\mathbf{y},\mathbf{u}}(z) = \mathcal{Z}_{\tau,m}\left\{\frac{P(s)}{Q(s)}\frac{\underline{\Omega}_{n_a}(s)}{B^{\theta_b}(s)}\right\}\theta_a$$

and (33) follows from Lemma 2. Also, from (26) we have that [see the first equation at the top of the page]. Then, from (31) it follows that [see the second equation at the top of the page] (recall that $\underline{\Omega}'_{n_b}$ denotes the row vector $\underline{\Omega}'_{n_b}$ with its first entry removed) and (34) follows from Lemma 2.

REFERENCES

[1] E. Larsson, M. Mossberg, and T. Soderstrom, "Estimation of continuous-time stochastic system parameters," in *Identification of Continuous-Time Models From Sampled Data*, ser. Advances in Industrial Control. New York: Springer, 2008, pp. 30–66.

[2] *Identification of Continuous-Time Models From Sampled Data*, ser. Advances in Industrial Control, H. Garnier and L. Wang, Eds., 1st ed. New York: Springer, 2008.

[3] G. Rao and H. Unbehauen, "Identification of continuous-time systems," *Proc. Inst. Electr. Eng.—Control Theory Applications*, vol. 153, no. 2, pp. 185–220, 2006.

[4] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[5] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[6] G. Rao and H. Garnier, "Numerical illustrations of the relevance of direct continuous-time model identification," in *Proc. 15th IFAC World Congr.*, Barcelona, Spain, 2002.

[7] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice-Hall Int., 1996.

[8] H. Garnier, L. Wang, and P. Young, "Direct identification of continuous-time models from sampled data: Issues, basic solutions and relevance," in *Identification of Continuous-Time Models From Sampled Data*, ser. Advances in Industrial Control. New York: Springer, 2008, pp. 1–29.

[9] H. Garnier, M. Mensler, and A. Richard, "Continuous-time model identification from sampled data: Implementation issues and performance evaluation," *Int. J. Control*, vol. 76, no. 13, pp. 1337–1357, 2003.

[10] H. Garnier and P. Young, "Time-domain approaches to continuous-time model identification of dynamical systems from sampled data," in *Proc. 2004 Amer. Control Conf.*, Boston, MA, 2004.

[11] T. Soderstrom, H. Fan, B. Carlsson, and S. Bigi, "Least squares parameter estimation of continuous-time ARX models from discrete-time data," *IEEE Trans. Autom. Control*, vol. 42, no. 5, pp. 659–673, 1997.

[12] P. De and H. Fan, "A delta least squares lattice algorithm for fast sampling," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2396–2406, 1999.

[13] R. Pintelon, J. Schoukens, and Y. Rolain, "Frequency-domain approach to continuous-time system identification: Some practical aspects," in *Identification of Continuous-Time Models from Sampled Data*, ser. Advances in Industrial Control. New York: Springer, 2008, pp. 214–248.

[14] J. Gillberg and L. Ljung, "Frequency-domain identification of continuous-time output error models from sampled data," in *Proc. 16th IFAC World Congr.*, Prague, 2005.

[15] D. Pham, "Estimation of continuous-time autoregressive model from finely sampled data," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2576–2584, 2000.

[16] E. Larsson, M. Mossberg, and T. Söoderström, "An overview of important practical aspects of continuous-time ARMA system identification," *Circuits, Syst., Signal Process.*, vol. 25, no. 1, pp. 17–46, 2006.

[17] M. Mossberg and E. Larsson, "Fast and approximative estimation of continuous-time stochastic signals from discrete-time data," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2004, vol. 2, pp. ii/529–ii/532.

[18] M. Mossberg, "Sequential estimation of stochastic continuous-time signals from sample covariances," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2005, vol. 4, pp. iv/277–iv/280.

[19] M. Mossberg, "Estimation of continuous-time stochastic signals from sample covariances," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 821–825, 2008.

[20] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. Part I. Theory," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 821–833, 1993.

[21] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. Part II. Efficient design and applications," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 834–848, 1993.

[22] R. Fletcher, *Practical Methods of Optimization*, 2nd ed. Chichester, U.K.: Wiley-Interscience, 1987.

[23] C. Sanathanan and J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Trans. Autom. Control*, vol. 8, no. 1, pp. 56–58, 1963.

[24] L. Ljung, "Initialisation aspects for subspace and output-error identification methods," in *Proc. Eur. Control Conf.*, 2003.

[25] G. Rao and H. Garnier, "Identification of continuous-time systems: direct or indirect," *Syst. Sci.*, vol. 30, no. 3, pp. 25–50, 2004.

[26] A. Sayed and T. Kailath, "A survey of spectral factorization methods," *Numer. Linear Algebra Appl.*, vol. 8, no. 6–7, pp. 467–496, 2001.

[27] S. Moheimani, A. Fleming, and S. Behrens, "On the feedback structure of wideband piezoelectric shunt damping systems," *Smart Mater. Structures*, vol. 12, no. 1, pp. 49–56, 2003.

[28] G. Golub and C. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.

[29] C. Vogel, "The impact of combined channel mismatch effects in time-interleaved ADCS," *IEEE Trans. Instrum. Meas.*, vol. 54, pp. 415–427, 2005.

Damián Marelli received the Bachelor's degree in electronics engineering from the Universidad Nacional de Rosario, Argentina, in 1995 and the Ph.D. degree in electrical engineering and the Bachelor's (Hons.) degree in mathematics from the University of Newcastle, Australia, in 2003.

In 2003, he was a Research Associate at the School of Electrical Engineering and Computer Science at the University of Newcastle, Australia. In 2004 and 2005, he held a postdoctoral research fellowship at the Laboratoire d'Analyse Topologie et Probabilités, CNRS / Université de Provence, France. Since 2006, he has been a Research Academic at the ARC Centre for Complex Dynamic Systems and Control at the University of Newcastle, Australia. He held an Intra-European Marie Curie Fellowship at the Faculty of Mathematics, University of Vienna, Austria, from 2007 to 2008. His main research interests include multirate signal processing, time-frequency analysis, system identification, and statistical signal processing.

Minyue Fu (S'84–M'87–SM'94–F'04) received the Bachelor's degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 1982 and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison in 1983 and 1987, respectively.

From 1983 to 1987, he was a Teaching Assistant and a Research Assistant at the University of Wisconsin-Madison. He worked as a Computer Engineering Consultant at Nicolet Instruments, Inc., Madison, WI, during 1987. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI. For summer 1989, he was employed by the Universite Catholoque de Louvain, Belgium, as a Maitre de Conferences Invited. In 1989, he joined the Department of Electrical and Computer Engineering, the University of Newcastle, Australia, where he is currently a Chair Professor in electrical engineering. In addition, he was a Visiting Associate Professor at the University of Iowa in 1995–1996, and a Senior Fellow/Visiting Professor at Nanyang Technological University, Singapore, in 2002. He holds a ChangJiang Visiting Professorship at Shandong University and visiting positions at South China University of Technology and Zhejiang University in China. His main research interests include control systems, signal processing and communications.

Dr. Fu has been an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, and the *Journal of Optimization and Engineering*.