A Sampling Theory Approach for Continuous ARMA Identification

Hagai Kirshner, Simona Maggio, and Michael Unser

Abstract

The problem of estimating continuous-domain autoregressive moving-average processes from sampled data is considered. The proposed approach incorporates the sampling process into the problem formulation while introducing exponential models for both the continuous and the sampled processes. We derive an exact evaluation of the discrete-domain power-spectrum using exponential B-splines and further suggest an estimation approach that is based on digitally filtering the available data. The proposed functional, which is related to Whittle's likelihood function, exhibits several local minima that originate from aliasing. The global minimum, however, corresponds to a maximum-likelihood estimator, regardless of the sampling step. Experimental results indicate that the proposed approach closely follows the Cramér-Rao bound for various aliasing configurations.

I. Introduction

Continuous-domain ARMA (Auto Regressive Moving Average) processes are widely used in control theory and in signal/image processing and analysis. Typical examples of applications are system identification and adaptive filtering [1], [2]; speech analysis and synthesis [3]; stochastic differential equations and image modeling [4]–[6]. Linear estimation theory for ARMA processes is closely related to Sobolev spaces [7]; the reproducing kernel of a Sobolev space has a similar role in signal interpolation as the ARMA autocorrelation function has in linear minimum-norm estimator design. Further, Sobolev norms provide the continuous-domain regularization term in numerous inverse problems [8]–[12] and their relation to ARMA modeling suggests a parameterized modeling of continuous-domain signals. In practice, the available data is discrete and one is usually required to estimate the underlying continuous-domain parameters from sample values. Potential examples are continuous-domain structure modeling of physical phenomena, linear time invariant (LTI) system identification, as well as numerical analysis of differential operators.

The sampled version of a Gaussian ARMA process is a discrete-domain ARMA process whose zeros and poles are coupled in a non-trivial way [13]–[16]. Recent works on this subject fall in two broad categories: direct and indirect [17], [18]. Direct methods consist first of parameterizing a discrete-domain model by the continuous-domain parameters. The discrete-domain model is then used to minimize a cost function that involves the available data. In this way, the required continuous-domain parameters are directly estimated by the minimization process. An example of such a method would be the replacement of derivative operators by finite-difference operations [19]–[23]. In [24], a continuous-time AR model was recast into a discrete-time linear regression formulation rather than into a

discrete-time ARMA process. The regressor elements were then shown to be linear combinations of the discrete-time output measurements. As the least square solution of this regression might result in a biased estimation, the authors of [24] suggest two methods for reducing the bias effect: imposing constraints on the finite difference weights, or, alternatively, compensating for the bias as the final stage of the estimation process. The advantage of these two methods is that they require no shifting of the data when approximating the derivative values, giving rise to a reduced computational complexity over other least square methods. In [25], it is suggested to parameterize the autocorrelation sequence of the sampled process by applying the numerical decomposition method of Schur to obtain a state-space representation of the discrete-domain process. The cost function that has been proposed there minimizes the ℓ_2 norm of the difference between a sampled version of the autocorrelation model and the autocorrelation sample values. Another example of a direct method consists of power-spectrum parameterization [26], [27]. Indirect methods, on the other hand, rely on standard discrete-domain system identification methods such as the minimization of the prediction error variance. The discrete-domain system is then mapped to a continuous-domain one. The bilinear transform is one possible way of doing so, while alternative transformations are available, too [28].

Motivated by the deterministic theory of LTI systems, we exploit in this work the mathematical formulation of exponential splines. These functions provide a formal link between continuous-domain convolution operators and their discrete-domain counterparts [29], [30] and they will be shown to be suitable for describing sampled ARMA processes, too. A first study of this property was recently suggested in [31] for the autoregressive model. Considering an ideal sampling procedure, also known as instantaneous sampling, the autocorrelation sequence of the sampled process corresponds to sample values of the autocorrelation function of the original continuous-domain process. It then follows that both autocorrelation measures are of an exponential type, suggesting an exponential spline framework for describing the relation between an ARMA process and its sampled version. Another point is the Cramér-Rao bound. This bound converges to zero for any sampling interval value with increasing number of data points. In [32], the use of an anti-aliasing filter was suggested prior to low-rate sampling for systems of high bandwidth. It was shown there that pole ambiguity can be resolved in certain cases while minimizing a cost function that is based on approximating the autocorrelation function. Maximum-likelihood estimators, however, were not investigated in this context. While many of the currently available estimation algorithms are focused on base-band power spectra, it seems possible to derive an estimator that overcomes aliasing. Such an estimator could prove useful to optical imaging when the acquisition device has limited resolution, and to compressed sensing in the context of a reduced number of measurements. Another potential application is resolution conversion in which low-resolution digital images are displayed on a high-resolution device [33]. From a numerical perspective, Whittle's likelihood function plays an important role in deriving frequency-based estimation algorithms. This function is often approximated by means of discrete Fourier transform values and by Riemann sums. Such an approximation is not necessarily optimal and there may exist better numerical schemes.

This work provides a rigorous derivation of a maximum-likelihood-based estimator of continuous-domain ARMA parameters from sampled data. It utilizes the exponential B-spline framework while introducing an exact zero-pole

coupling for the sampled process. For that purpose, the relation between the autocorrelation function and the autocorrelation sequence is investigated in both time- and frequency domains. Based on this relation, it is shown that the Cramér-Rao bound can be made arbitrarily small by considering more sample values where the sampling interval can take an arbitrary value. The likelihood function of the sampled process is investigated, too. In particular, it is shown that this function possesses local minima that originate from aliasing. The global minimum, however, corresponds to the maximum-likelihood value. The only assumption that is made throughout this study is that the number of available samples is relatively large, allowing one to replace the whitening matrix by a digital filter. This approximation is shown to be valid when considering expected values of the likelihood function.

The paper is organized as follows. In Section II, we provide the mathematical conventions and notations that will be used throughout this work. In Section III, we describe the autocorrelation property of continuous-domain ARMA processes and its relation to the autocorrelation sequence of the sampled process. We then introduce in Section IV the Cramér-Rao bound for such processes. We propose a maximum-likelihood-based estimation approach in Section V and provide a detailed description of the estimation algorithm in Section VI. Experimental results are given in Section VII.

II. Notations

The bilateral Laplace transform of a scalar function $\varphi(t)$ is

$$\Phi(s) = \mathcal{L}\left\{\varphi(t)\right\}(s) = \int_{-\infty}^{\infty} \varphi(t) e^{-st}\, dt, \tag{1}$$

where s takes complex values that satisfy $\varphi(t)e^{-st} \in L_1$. The inverse Laplace transform is denoted $\mathcal{L}^{-1}$. The Fourier transform of this function is $\hat{\varphi}(\omega) = \mathcal{F}\left\{\varphi(t)\right\}(\omega) = \Phi(j\omega)$. The bilateral z-transform of the sequence $\left\{\varphi[n]\right\}_{n\in\mathcal{Z}}$ is

$$\Phi_d(z) = \mathcal{Z}\left\{\varphi[n]\right\}(z) = \sum_{n=-\infty}^{\infty} \varphi[n]\, z^{-n}, \tag{2}$$

where z takes complex values for which the infinite sum converges. The DTFT (Discrete Time Fourier Transform) of the same sequence is $\hat{\varphi}_d(\omega) = \Phi_d(e^{j\omega})$ and for sequences of length N, the DFT (Discrete Fourier Transform) is given by $\hat{\varphi}_d[k] = \Phi_d(e^{2\pi jk/N})$. The argument ω denotes either radial frequency in units of [rad/time-unit] when used with the Fourier transform, or normalized frequency in units of $[rad/sample]$ when used with the DTFT. Continuous-domain and discrete-domain convolution operations are denoted $*$. The transpose of a matrix Σ is Σ^T, its inverse is Σ^{-1}, and its determinant is $|\Sigma|$. The operator $\mathbf{E}$ denotes expected value.

III. ARMA Processes and Splines

A. ARMA Processes and Linear System Theory

A differential LTI system is fully described by a rational transfer function ($p > q$)

$$\hat{h}(\omega) = \frac{\prod_{k=1}^{q}(j\omega - r_k)}{\prod_{k=1}^{p}(j\omega - s_k)}, \tag{3}$$

where $\{r_k\}$ and $\{s_k\}$ are its zeros and poles, respectively. The system is causal and stable iff $\mathrm{Re}\left\{s_k\right\} < 0\ \forall k$.

If such a system is driven by continuous-domain white Gaussian noise, its output is a Gaussian ARMA process. The spectral density function of such a process is $\hat{\varphi}(\omega) = \sigma^2|\hat{h}(\omega)|^2$ and the autocorrelation function is $\varphi(t) = \sigma^2 \{h(\tau) * h(-\tau)\}(t)$, where σ^2 is the variance of the noise and where $h(t) = \mathcal{F}^{-1}\left\{\hat{h}(\omega)\right\}(t)$ is the impulse response of the system. Exponential splines provide a mathematical framework for relating continuous-domain LTI systems with their discrete-domain counterparts [29], [30]. The dependence of $\varphi(t)$ on $h(t)$ suggests that these splines may be equally helpful for relating continuous-domain and discrete-domain ARMA processes.

B. Motivating Example: First-Order AR Process

The autocorrelation function of a continuous-domain AR(1) process that has a pole at $s = s_1$ and a unit intensity innovation is a symmetric exponential

$$\varphi(t; s_1) = -\frac{1}{2s_1}\, e^{s_1|t|}. \tag{4}$$

The spectral density function is then

$$\Phi(j\omega; s_1) = \mathcal{F}\{\varphi(t; s_1)\}(\omega) = \frac{1}{(j\omega - s_1)(-j\omega - s_1)}. \tag{5}$$

Upon ideal, i.e. instantaneous, sampling, the autocorrelation sequence of the corresponding discrete-domain process is

$$\varphi[n; s_1] = -\frac{1}{2s_1}\, e^{s_1|n|}, \tag{6}$$

where a unit-time sampling interval was assumed for simplicity. The spectral density function of the sampled process is then

$$\Phi_d(e^{j\omega}; s_1) = \mathcal{F}_d\{\varphi[n; s_1]\}(e^{j\omega}) = \frac{e^{2s_1} - 1}{2s_1} \cdot \frac{1}{(1 - e^{s_1}e^{-j\omega})(1 - e^{s_1}e^{j\omega})}. \tag{7}$$

The important observation is that one is able to link the continuous-domain and the discrete-domain autocorrelations via the Shannon-like interpolation formula (see also Figure 1)

$$\varphi(t; s_1) = \sum_{n=-\infty}^{\infty} \varphi[n; s_1]\beta(t - n; s_1), \tag{8}$$

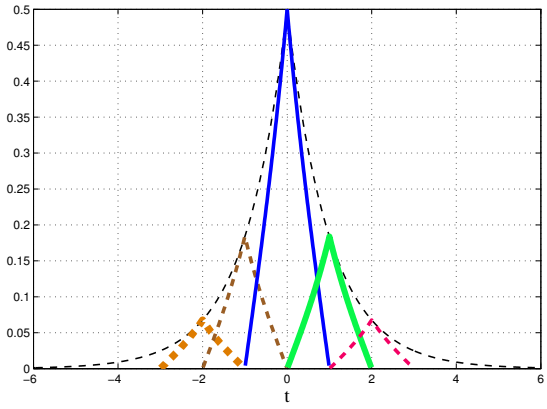
where $\beta(t; s_1)$ is an interpolating basis function whose Fourier expression is

$$\hat{\beta}(\omega; s_1) = \frac{2s_1}{e^{2s_1} - 1} \cdot \frac{(1 - e^{s_1}e^{-j\omega})(1 - e^{s_1}e^{j\omega})}{(j\omega - s_1)(-j\omega - s_1)}. \tag{9}$$

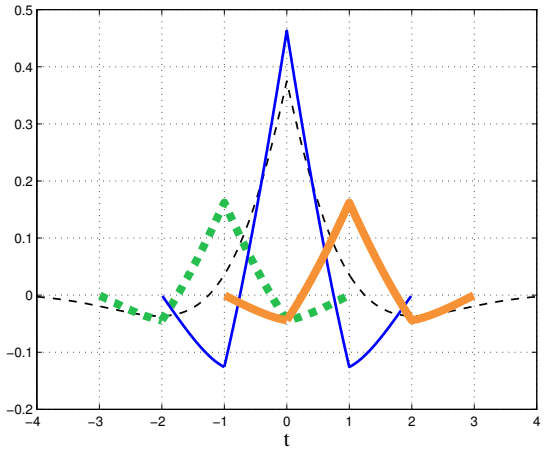
Observe that the latter expression is also equal to the ratio of (5) and (7). The corresponding time-domain expression is

$$\beta(t; s_1) = \begin{cases} \frac{\sinh[s_1(1-|t|)]}{\sinh s_1} & |t| \le 1 \\ 0 & |t| > 1. \end{cases} \tag{10}$$

The key property that will be exploited in this work is that $\beta(t; s_1)$ is compactly supported, which is not directly apparent from the Fourier-domain expression (9). In fact, $\beta(t; s_1)$ is an exponential B-spline and the above method generalizes for higher-order systems.



(a)



(b)

Fig. 1. Exponential B-spline decomposition of ARMA models. The autocorrelation function (black, dashed line) can be decomposed into a weighted sum of exponential B-spline functions (various colors, various widths, dashed and solid lines). Shown are only few B-spline functions. The parameters of the ARMA processes are (a) $\sigma = 1, s = -1$ and (b) $\sigma = 1, r = -1, s = -1 \pm i$.

C. General Case

A continuous-domain ARMA process is fully characterized by its parameters vector

$$\theta = \left\{\sigma^2, r_1, \ldots, r_q, s_1, \ldots, s_p\right\}, \tag{11}$$

where $\{s_k\}$ and $\{r_k\}$ are the poles and the zeroes of the process, respectively. The poles are assumed to have a strictly negative real part. The continuous-domain innovation process is assumed to be Gaussian and its intensity is σ^2. Additionally, $p > q$.

TABLE I

CONITNUOUS-TO-DISCRETE DOMAIN MAPPING OF MULTIPLE POLES.

Continuous Domain		Discrete Domain	
Laplace	Time	index	z-transform
$\frac{1}{s-s_k}+\frac{1}{-s-s_k}$	$e^{s_k\lvert t\rvert}$	$e^{s_k\lvert n\rvert}$	$-\frac{e^{-s_k}}{z^{-1}-e^{-s_k}}+\frac{e^{s_k}}{z^{-1}-e^{s_k}}$
$\frac{1}{(s-s_k)^2}+\frac{1}{(-s-s_k)^2}$	$\lvert t\rvert\cdot e^{s_k\lvert t\rvert}$	$\lvert n\rvert\cdot e^{s_k\lvert n\rvert}$	$\left[\frac{e^{-s_k}}{z^{-1}-e^{-s_k}}+\frac{e^{-2s_k}}{(z^{-1}-e^{-s_k})^2}\right]+\left[\frac{e^{s_k}}{z^{-1}-e^{s_k}}+\frac{e^{2s_k}}{(z^{-1}-e^{s_k})^2}\right]$
$\frac{1}{(s-s_k)^3}+\frac{1}{(-s-s_k)^3}$	$\frac{1}{2}t^2\cdot e^{s_k\lvert t\rvert}$	$\frac{1}{2}n^2\cdot e^{s_k\lvert n\rvert}$	$-\left[\frac{1}{2}\frac{e^{-s_k}}{z^{-1}-e^{-s_k}}+\frac{3}{2}\frac{e^{-2s_k}}{(z^{-1}-e^{-s_k})^2}+\frac{e^{-3s_k}}{(z^{-1}-e^{-s_k})^3}\right]+$ $+\left[\frac{1}{2}\frac{e^{s_k}}{z^{-1}-e^{s_k}}+\frac{3}{2}\frac{e^{2s_k}}{(z^{-1}-e^{s_k})^2}+\frac{e^{3s_k}}{(z^{-1}-e^{s_k})^3}\right]$

The Laplace transform of the corresponding autocorrelation function is given by

$$\Phi(s;\theta)=\sigma^2\frac{\prod_{k=1}^{q}(s-r_k)(-s-r_k)}{\prod_{k=1}^{p}(s-s_k)(-s-s_k)}. \tag{12}$$

By performing the partial-fraction decomposition of $\Phi(s;\theta)$ (we are assuming for simplicity that the poles are simple), we find that

$$\Phi(s;\theta)=\sigma^2\sum_{k=1}^{p}\alpha_k(\theta)\left\{\frac{1}{s-s_k}-\frac{1}{s+s_k}\right\}, \tag{13}$$

and we deduce that the autocorrelation function is a sum of exponentials

$$\varphi(t;\theta)=\sigma^2\sum_{k=1}^{p}\alpha_k(\theta)\cdot e^{s_k|t|}. \tag{14}$$

In cases where $\Phi(s;\theta)$ introduces pole multiplicity, $\varphi(t;\theta)$ would involve polynomial multiplications as given in Table I.

Ideally sampling a continuous-domain ARMA process yields a discrete-domain ARMA process. The autocorrelation sequence of the discrete-domain process is then given by the ideal samples of the autocorrelation function. For the partial decomposition of (14) and for a unit-time sampling interval, this sequence is given by

$$\varphi\left[n;\theta\right]=\sigma^2\sum_{k=1}^{p}\alpha_k(\theta)\cdot e^{s_k|n|}. \tag{15}$$

Autocorrelation functions that consist of multiple poles can be discretized in a similar manner as given in Table I.

ARMA models are closely related to generalized exponential B-splines. These finite-support functions stem from Green's functions of rational operators [30]. In our case, the Green's function is the autocorrelation function itself. Unlike [30], however, this work introduces non-causal symmetric B-splines.

Definition 1 (Symmetric Exponential B-spline): The exponential B-spline $\beta(t;\theta)$ with parameters θ is specified by the following inverse Fourier transform:

$$\beta(t;\theta) = \mathcal{F}^{-1}\left\{\prod_{k=1}^{q}(j\omega - r_k)(-j\omega - r_k)\cdot\prod_{k=1}^{p}\frac{(1-e^{j\omega+s_k})(1-e^{-j\omega+s_k})}{(j\omega - s_k)(-j\omega - s_k)}\right\}(t). \tag{16}$$

Observe that this function corresponds to the ratio between the ARMA power-spectrum and a discrete-domain AR power-spectrum with poles at e^{s_k}. The exponential B-spline kernels have the following properties:

- Compactly supported functions within the interval $[-p, p]$.
- Bounded and symmetric functions.
- Smooth functions: the first $2(p-q)-1$ derivatives are L_2 functions.
- Integer shifts of B-spline kernels form a Riesz basis [34].
- Weighted sum of shifted B-spline kernels can reproduce exponential functions of the type (14).
- The convolution of two exponential B-spline kernels yields another B-spline kernel of an augmented order. This property allows one to iteratively construct an exponential B-splines of any order.

Definition 2 (Localization filter): The localization filter with parameters θ is specified by the Laplace transform

$$\Delta(s;\theta) = \prod_{k=1}^{p}\left(1-e^{s+s_k}\right)\left(1-e^{-s+s_k}\right). \tag{17}$$

Proposition 1: The autocorrelation function of an ARMA process with parameters θ can be written as

$$\varphi(t;\theta) = \sigma^2\sum_{m=-\infty}^{\infty}\mathbf{p}[m;\theta]\cdot\beta(t-m;\theta), \tag{18}$$

where $\beta(t;\theta)$ is the exponential B-spline with parameters θ and the sequence $\mathbf{p}$ is given in the z-domain by

$$P_d(z;\theta) = \frac{1}{\prod_{k=1}^{p}\left(1-e^{s_k}z\right)\left(1-e^{s_k}z^{-1}\right)} = \frac{1}{\Delta(\ln z;\theta)}. \tag{19}$$

Proof: Take the Fourier transform of (18) and substitute (16) and (12). □

Definition 3: The discrete exponential B-spline kernel with parameters θ is given by $\mathcal{B}_d(z;\theta) = \sum_{n=-p}^{p}\beta[n;\theta]z^{-n}$ and $\hat{\beta}_d(\omega) = \mathcal{B}_d(e^{j\omega})$.

We rely on the fact that ideal sampling preserves the autocorrelation values of the continuous-domain ARMA process. In such a case, also known as instantaneous sampling, the values of the discrete-domain process are given by the point wise values of the continuous-domain process at the sampling points. The power spectrum of the sampled process is then related to the continuous-domain power spectrum through aliasing and it can be described by means of a rational transfer function in the z-domain [16]. The following Theorem utilizes the discrete exponential B-spline kernel $\mathcal{B}_d(z;\theta)$ of Definition 3 and provides a novel direct formula for extracting discrete-domain power spectrum from continuous-domain parameters.

Theorem 1: The discrete-domain ARMA process that is given by the ideal unit-interval samples of the continuous-domain process (12) has the following parameters:

$$H_d(z;\theta) = \sigma_d(\theta)\frac{\prod_{k=1}^{p-1}\left(1-\nu_k(\theta)z^{-1}\right)}{\prod_{k=1}^{p}\left(1-\rho_k(\theta)z^{-1}\right)}, \tag{20}$$

where

$$\rho_k(\theta) = e^{s_k} \tag{21}$$

$$\nu_k(\theta) = \text{roots of } \mathcal{B}_d(z;\theta) \text{ inside the unit circle,} \tag{22}$$

and where

$$\sigma_d^2(\theta) = \sigma^2 \frac{\mathcal{B}_d(1;\theta)}{\prod_{k=1}^{p-1}(1-\nu_k(\theta))^2}. \tag{23}$$

is the variance of the discrete-domain innovation process.

Proof: As $P_d(z;\theta)$ is an all-pole filter, the zeroes of

$$\Phi_d(z;\theta) = \sigma^2 P_d(z;\theta) \cdot \mathcal{B}_d(z;\theta) \tag{24}$$

originate from the zeroes of $\mathcal{B}_d(z;\theta)$ only. Those zeroes appear in reciprocal pairs due to the symmetry property of $\beta(t;\theta)$. Also, the DFT of the sampled version of the exponential B-spline satisfies

$$\hat{\beta}_d(\omega;\theta) = \sum_{k=-\infty}^{\infty} \hat{\beta}\left(\omega + 2\pi k;\theta\right) = \mathcal{B}_d(e^{j\omega}), \tag{25}$$

and does not vanish on $\omega \in [0, 2\pi)$ due to (16). The function $\beta(t;\theta)$ is of finite support so that $\mathcal{B}_d(z;\theta)$ has no poles. The symmetry property of $\Delta(s;\theta)$ indicates that the poles of $P_d(z;\theta)$ appear in reciprocal pairs. It then follows that $\Phi_d(z;\theta)$ can be described by a minimum-phase filter. The finite-support property of $\beta(t;\theta)$ and the structure of $\Delta(s;\theta)$ also guarantee that this minimum-phase filter has a rational transfer function. We further observe that

$$\mathcal{B}_d(z;\theta) = \mathcal{B}_d^-(z;\theta) \cdot \mathcal{B}_d^+(z;\theta) \tag{26}$$

$$P_d(z;\theta) = P_d^-(z;\theta) \cdot P_d^+(z;\theta), \tag{27}$$

where the superscript '$-$' denotes a causal filter and where $\mathcal{B}_d^+(z;\theta) = \mathcal{B}_d^-(z^{-1};\theta)$ and $P_d^+(z;\theta) = P_d^-(z^{-1};\theta)$. In particular, $P_d^-(z;\theta) = 1/\prod_{k=1}^{p}\left(1 - e^{s_k} z^{-1}\right)$ and $\mathcal{B}_d^-(z;\theta)$ has zeroes only; those zeroes originate from the roots of $\mathcal{B}_d(z;\theta)$ inside the unit circle. It then follows that

$$\Phi_d(z;\theta) = H_d(z;\theta) \cdot H_d(z^{-1};\theta), \tag{28}$$

where $H_d(z;\theta) = \sigma \cdot P_d^-(z;\theta) \cdot \mathcal{B}_d^-(z;\theta)$ is the required filter. Imposing the structure of (20) results in the expression for $\sigma_d^2(\theta)$. □

***Corollary** 1:* Let θ be known. Then, the autocorrelation function of a continuous-domain ARMA process is uniquely defined by its samples. Further,

$$\varphi(t;\theta) = \sum_{n=-\infty}^{\infty} \varphi(n;\theta) \cdot \eta(t-n;\theta), \tag{29}$$

where the interpolation kernel $\eta(t;\theta)$ is specified by its Fourier transform,

$$\hat{\eta}(\omega;\theta) = \frac{\hat{\beta}(\omega;\theta)}{\hat{\beta}_d(\omega;\theta)}. \tag{30}$$

This is the generalization of (8) for arbitrary ARMA(p,q) processes.

Proof: It was shown that $\Phi_d(z;\theta) = \sigma^2 P_d(z;\theta) \cdot \mathcal{B}_d(z;\theta)$ and that $\Phi(s;\theta) = \sigma^2 \mathcal{B}(s;\theta)/\Delta(s;\theta)$. Dividing the two equations while recalling that $P_d(z;\theta) = \frac{1}{\Delta(\ln z;\theta)}$ yields the required result. □

The exponential interpolation function $\eta(t;\theta)$ provides a means for interpolating continuous-domain ARMA models. It can also be interpreted as a spectral weighting function that relates the power-spectrum of the discrete-domain sampled process with the power-spectrum of the continuous-domain process it originates from. Unlike the polynomial-based weighting function of [27], this weighting function is parameterized, allowing one to describe band-pass power-spectrum, too.

It is also possible to consider the mapping of continuous-to-discrete parameters for non-ideal sampling procedures such as averaged sampling. In such cases, the continuous-domain process undergoes a continuous-domain filtering operation prior to the point-wise evaluation stage. The non-ideal samples are given by

$$c[n] = \int_{-\infty}^{\infty} x(t) s(t-n)\, dt, \tag{31}$$

where $x(t)$ is the stochastic process, $s(t)$ is a function that characterizes the acquisition device, and $n \in \mathbb{Z}$.

***Theorem** 2:* Let $s(t)$ describe a non-ideal sampling procedure. If (i) $s(t) \in L_1$ and (ii) $\sum_k |\hat{s}(\omega + 2\pi k)| > 0$ for all $\omega \in [0, 2\pi)$, then the discrete-domain process that originates from uniform non-ideal samples of the continuous-domain processes (12) can be realized by a causal and stable digital filter, applied to discrete-domain white Gaussian noise. The inverse filter is causal and stable, too.

Proof: By Young's inequality [35]

$$s(t) * s(-t) \in L_1. \tag{32}$$

The autocorrelation sequence of the non-ideal samples is

$$r[n;\theta,s(t)] = \Big\{ \varphi(t;\theta) * s(t) * s(-t) \Big\}(n), \tag{33}$$

where $\varphi(t;\theta)$ is the autocorrelation function of the continuous-domain ARMA process. It then follows that $r[n;\theta,s(t)] \in l_1$ as

$$\begin{aligned} \sum \Big| r[n;\theta,s(t)] \Big| = & \\ = & \sum \Big| \int \varphi(t-n;\theta) \cdot [s(t) * s(-t)]\, d\tau \Big| \\ \leq & \int \Big[\sum \Big| \varphi(t-n;\theta) \Big| \Big] \cdot \Big| s(t) * s(-t) \Big| \, d\tau \\ \leq & M \cdot \| s(t) * s(-t) \|_{L_1} . \end{aligned} \tag{34}$$

Because $\varphi(t;\theta)$ is composed of a finite sum of exponentially decaying functions, we have that $M < \infty$. The z-transform of $r[n;\theta,s(t)]$ has then an absolutely convergent Fourier series on the unit circle and it does not vanish there due to (ii). It has also real and strictly positive values there. For this reason, the function $R_d(z;\theta,s(t))$ does not reach a value of zero when evaluated on the unit circle. According to Krein [36], $R_d(z;\theta,s(t))$ has a canonical factorization on the unit circle. That is,

$$R_d(z;\theta,s(t)) = H_d^+(z;\theta,s(t)) \cdot H_d^-(z;\theta,s(t)), \tag{35}$$

where $H_d^+(z;\theta,s(t))$ is holomorphic (i.e., complex-valued analytic function) in $|z|<1$, continuous on $|z|\leq 1$, and does not have zeroes in $|z|\leq 1$; and where $H_d^-(z;\theta,s(t))$ is holomorphic in $|z|>1$, continuous on $|z|\geq 1$, and does not have zeros in $|z|\geq 1$. By a theorem of Wiener and Levy [36], [37], the function $\ln R_d(z;\theta,s(t))$ has an absolutely convergent Fourier series and one can write

$$H_d^+(z;\theta,s(t)) \;=\; \exp\Big(\frac{\alpha_0(\theta,s(t))}{2}+\sum_{n=1}^{\infty}\alpha_n(\theta,s(t))z^n\Big) \tag{36}$$

$$H_d^-(z;\theta,s(t)) \;=\; \exp\Big(\frac{\alpha_0(\theta,s(t))}{2}+\sum_{n=1}^{\infty}\alpha_{-n}(\theta,s(t))z^{-n}\Big), \tag{37}$$

where $\ln R_d(z;\theta,s(t))=\sum_{n=-\infty}^{\infty}\alpha_n(\theta,s(t))\cdot z^n$. The symmetry property of the autocorrelation sequence implies that $R_d(z;\theta,s(t))$ is symmetric on the unit circle and that $\ln R_d(z;\theta,s(t))$ is symmetric there, too. It then holds that $\alpha_n(\theta,s(t))=\alpha_{-n}(\theta,s(t))$, and that $H_d^-(z;\theta,s(t))=H_d^+(z^{-1};\theta,s(t))$. The expression of $H_d^-(z;\theta,s(t))$ corresponds then to a digital filter that is causal and stable. The inverse filter $1/H_d^-(z;\theta,s(t))$ is stable and causal too, as its region of convergence includes both the unit circle and infinity. This stems from the fact that $H_d^-(z;\theta,s(t))$ is non-zero and continuous on the unit circle as well as non-zero and holomorphic at infinity. □

Theorem 2 allows for a relatively large class of sampling models to be considered. Among them are polynomial and exponential B-splines, truncated and non-truncated Gaussian functions, and continuous-domain derivative filters such as the Laplacian of a Gaussian.

IV. Cramér-Rao Bound

Let a continuous-domain ARMA(p,q) process be given by its uniform samples only. Larsson and Larsson [38] provide closed-form expressions for the CRB of estimated continuous-domain parameters which utilize the state-space representation of the process. These expressions involve matrix inversion and eigenvalue decomposition, and when considering large data sets, Friedlander's approximate approach can be used instead [39]. The Fisher information matrix is given by

$$\begin{aligned} I_{k,l}(\theta) \;\cong\; \frac{N}{4\pi}\int_0^{2\pi}&\left(\frac{1}{\hat\varphi_d(\omega;\theta)}\frac{\partial\hat\varphi_d(\omega;\theta)}{\partial\theta_k}\right)\cdot \\ &\cdot\left(\frac{1}{\hat\varphi_d(\omega;\theta)}\frac{\partial\hat\varphi_d(\omega;\theta)}{\partial\theta_l}\right)d\omega, \end{aligned} \tag{38}$$

where $k,l=1,\ldots,p+q+1$. The CRB (Cramér-Rao Bound) is then the inverse of $I(\theta)$. Sampling interval dependency can be incorporated in the localization filter (17) by

$$\Delta(s;\theta,T)=\prod_{k=1}^{p}\big(1-e^{s+s_kT}\big)\big(1-e^{-s+s_kT}\big), \tag{39}$$

where T is the sampling interval. It then follows that $\hat\varphi_d(\omega;\theta)$ is not dependent upon N nor is the integrand of (38). As the number of available samples becomes larger, the CRB becomes smaller regardless of T. Such inverse proportionality with respect to N was already pointed out in [38] for sampled AR processes. This observation implies that aliasing effects can be compensated for by taking more measurements. It further suggests that there

exists an ML (Maximum-Likelihood) estimator that overcomes aliasing. Such an estimator has to take into account the fact that the discrete-domain poles and the zeroes of $\Phi_d(z;\theta)$ are coupled, ensuring that it corresponds to the samples of the autocorrelation function $\varphi(t;\theta)$.

V. MAXIMUM-LIKELIHOOD ESTIMATION

Motivated by the CRB, we approximate the log-likelihood function by means of a digital filter. The proposed approximation relies on the discrete-domain parameters of (20), namely, $\left\{\sigma_d^2, \nu_1, \ldots, \nu_q, \rho_1, \ldots, \rho_p\right\}$. These parameters can be numerically calculated from θ in a straightforward manner as was shown in Section III. We, however, do not allow the continuous-domain poles of θ to differ by $2\pi j\mathbb{Z}$ as such poles yield the same discrete-domain model upon sampling.

***Definition** 4:* Let θ be known and let $\mathbf{x}$ be N uniform ideal samples of the continuous-domain process (12) taken on a unit-interval grid. The probability density function of $\mathbf{x}$ is

$$f(\mathbf{x};\theta) = \frac{1}{(2\pi)^{\frac{N}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}\mathbf{x}^T\Sigma^{-1}\mathbf{x}\right\}, \tag{40}$$

where $\Sigma[m,n] = \varphi[m-n;\theta]$ is the autocorrelation matrix that corresponds to θ.

***Definition** 5:* The corresponding log-likelihood function, including a sign inversion, is

$$l(\theta;\mathbf{x}) = \ln|\Sigma| + \mathbf{x}^T\Sigma^{-1}\mathbf{x}. \tag{41}$$

***Definition** 6:* Let θ be known. Then, the digital filter g_θ is given by its z-transform $G_d(z) = 1/H_d(z)$ where $H_d(z)$ is given in (20).

***Definition** 7:* Let θ be known. Then,

$$\kappa(\theta) = \sum_{n=1}^{\infty} n \cdot c[n;\theta]^2, \tag{42}$$

where

$$c[n;\theta] = \frac{1}{n}\left\{\nu_1^n(\theta) + \ldots + \nu_{p-1}^n(\theta) - \rho_1^n(\theta) - \ldots - \rho_p^n(\theta)\right\}, \tag{43}$$

are the Fourier coefficients of $\ln\hat{\varphi}_d(\omega)$.

The constant $\kappa(\theta)$ can be interpreted by means of an inner product operation. Considering a discrete-domain ARMA power spectrum, we define the Fourier coefficients of its logarithm:

$$\ln\hat{\varphi}_d(\omega;\theta) = \sum_{n=-\infty}^{\infty} c[n;\theta]e^{-j\omega n}. \tag{44}$$

Recalling (20),

$$\ln H_d(e^{j\omega};\theta) = \frac{c[0;\theta]}{2} + \sum_{n=1}^{\infty} c[n]e^{-j\omega n}. \tag{45}$$

It then follows that

$$\kappa(\theta) = \left|\left\langle \frac{d}{d\omega}(\ln H_d(e^{j\omega})), \ln H_d(e^{j\omega};\theta)\right\rangle_{L_2}\right|. \tag{46}$$

***Theorem** 3:* Let θ be known and let $\mathbf{x}$ be N uniform ideal samples of the continuous-domain process (12) taken on a unit-interval grid. Then,

$$\lim_{N\to\infty} \mathbf{E}_{\mathbf{x}} \Big| l(\theta;\mathbf{x}) - \tilde{l}(\theta;\mathbf{x}) \Big| = 0, \tag{47}$$

where

$$\tilde{l}(\theta;\mathbf{x}) = N \ln \sigma_d^2(\theta) + \kappa(\theta) + \|\mathbf{x} * g_\theta\|_{\ell_2}^2 . \tag{48}$$

Here, $*$ denotes discrete-domain convolution of an N-length output sequence.

Proof: According to Szegő theorem for infinite Toeplitz matrices [40]–[42],

$$\lim_{N\to\infty} \Big\{ \ln|\Sigma| - N \cdot c[0;\theta] \Big\} = \sum_{n=1}^{\infty} n\, c[n;\theta]\, c[-n;\theta], \tag{49}$$

where c corresponds to the Fourier coefficients of $\ln \hat{\varphi}_d(\omega;\theta)$. In our case, these coefficients are given by (43) and the right-hand-side of the equation can be replaced by $\kappa(\theta)$. Also, by the Residue Theorem,

$$c[0;\theta] = \frac{1}{2\pi} \int_0^{2\pi} \ln \hat{\varphi}_d(\omega;\theta)\, d\omega = \ln \sigma_d^2(\theta). \tag{50}$$

It then holds that

$$\lim_{N\to\infty} \Big\{ \ln|\Sigma| - N \ln \sigma_d^2(\theta) - \kappa(\theta) \Big\} = 0. \tag{51}$$

The constant $\kappa(\theta)$ provides also a means for describing the limiting behavior of the determinant. Writing (51) in a different form [42],

$$|\Sigma(\theta)| \propto e^{\kappa(\theta)} \cdot (e^{c[0;\theta]})^N, \tag{52}$$

we observe that the right-hand-side of the equation is equal to the determinant up to multiplication by a constant; the important thing here is that this constant does not depend on N or on θ.

As for the term $\mathbf{x}^T \cdot \Sigma^{-1} \cdot \mathbf{x}$ that appears in $l(\theta;\mathbf{x})$, it may be approximated by digitally filtering $\mathbf{x}$ and by calculating the energy of the output. A possible filter one can use is g_θ. This filter is guaranteed by Theorem 1 to exist, to be stable, and to be causal. The N-length output of such a filter can be described by means of the lower triangular matrix

$$L[m,n] = g_\theta[m-n],\ m,n = 0,\ldots,N-1, \tag{53}$$

and it holds that $\|\mathbf{x} * g_\theta\|_{\ell_2}^2 = \mathbf{x}^T L L^T \mathbf{x}$. Following [43], the two matrices Σ^{-1} and LL^T are asymptotically equivalent as N grows. This stems from the fact that they both originate from the same power-spectrum $\hat{\varphi}_d(\omega;\theta)$. Such a power-spectrum gives rise to an infinite Toeplitz matrix T and the $N \times N$ matrix Σ^{-1} is defined by 1) truncating T and 2) taking the inverse. The matrix L is defined by 1) inverting T, 2) finding its Cholesky decomposition, and 3) truncating the lower triangle matrix.

Asymptotic equivalence implies that the norm of the difference between the two matrices converges to zero with increasing values of N. Focusing on a Toeplitz matrix $M\,[m,n] = t\,[m-n]$, one can associate a discrete-time Fourier transform to the sequence $\{t[k]\}$, and for autocorrelation matrices, this transform corresponds to the power spectrum function. Asymptotic equivalence of two covariance matrices implies that the power spectrum of

one matrix converges uniformly to the power spectrum of the other. In the context of this proof, the matrix L describes a truncated version of the filter g_θ. As $g_\theta \in \ell_1$, its discrete-time Fourier transform uniformly converges to $1/H_d(e^{j\omega})$. This means that with increasing numbers of samples, the expression $\mathbf{x}^T L$ converges in the ℓ_2 sense to $\mathbf{x} * g_\theta$. This fact stems from Parseval's property of the discrete-time Fourier transform. It then follows that the output is a stochastic process with a power spectrum that converges to a constant function. This constant is $\sigma_d^2(\theta)$. The term $\mathbf{x}^T LL^T\mathbf{x}$ is then an estimation for this value. The expression $\mathbf{x}^T\Sigma^{-1}\mathbf{x}$ amounts to de-correlating the process $\mathbf{x}$ and to estimating the variance of the uncorrelated process. The value of this variance is $\sigma_d(\theta)$, too. It then follows that $\mathbf{E}\mathbf{x}^T LL^T\mathbf{x} = \mathbf{E}\mathbf{x}^T\Sigma^-1\mathbf{x} = N\sigma_d^2(\theta)$. □

The likelihood function approximation $\tilde{l}(\theta, \mathbf{x})$ is related to Whittle's approximation of the log-likelihood function (41) [44]. Whittle's approximation is given by

$$\tilde{l}_W(\theta; \mathbf{x}) = \frac{N}{2\pi}\int_0^{2\pi} \ln \hat{\varphi}_d(\omega;\theta)d\omega + \frac{1}{2\pi}\int_0^{2\pi} \frac{\left|\hat{\mathbf{x}}_d(\omega)\right|^2}{\hat{\varphi}_d(\omega;\theta)}d\omega, \tag{54}$$

and it holds that

$$N\ln\sigma_d^2 = \frac{N}{2\pi}\int_0^{2\pi} \ln \hat{\varphi}_d(\omega;\theta)d\omega \tag{55}$$

$$\lim_{N\to\infty} \|\mathbf{x} * g_\theta\|_{\ell_2}^2 = \frac{1}{2\pi}\int_0^{2\pi} \frac{\left|\hat{\mathbf{x}}_d(\omega)\right|^2}{\hat{\varphi}_d(\omega;\theta)}d\omega, \tag{56}$$

where the limit is required due to the use of finite-length convolution in $\tilde{l}(\theta;\mathbf{x})$. It then follows that $\lim_{N\to\infty}\tilde{l}(\theta;\mathbf{x}) = \tilde{l}_W(\theta;\mathbf{x}) + \kappa(\theta)$. The integrals of Whittle's likelihood function are often approximated by Riemann sums that involve DFT values. DFT values of a finite-length signal, in this case $\mathbf{x}$, indeed coincide with the samples of its z-transform on the unit circle. Nevertheless, this is not the case for the infinite sequences that are described by $\Phi_d(z;\theta)$ and by $1/\Phi_d(z;\theta)$. The proposed approach, on the other hand, does provide an accurate evaluation of these integrals and the truncation error of the convolution operation is exponentially decreasing, providing a higher convergence rate than the Riemann sum method.

We propose in this work to estimate the ARMA parameters by minimizing the approximated likelihood function of Theorem 3,

$$\hat{\theta} = \arg\min_\theta \tilde{l}(\theta, \mathbf{x}). \tag{57}$$

Our approach differs from currently available methods in several aspects. It considers exponential autocorrelation models for both the continuous- and the discrete-domain processes whereas several previous works considered polynomial models. It also establishes a link between the continuous- and the discrete-domain models by incorporating the sampling process into the problem formulation. This, in turn, allows for the discrete model to stem naturally from the continuous-domain formulation, while no a priori assumptions are made on the digital data. Further, the log-likelihood function suggested here holds true for any value of sampling interval, rather than describing the limiting case of $T \to 0$. Additionally, the log-likelihood function considers discrete-domain data for determining continuous-domain statistics while no approximation of continuous-domain frequency spectrum or impulse responses is required.

The log-likelihood function (48) has several local minima, as demonstrated by Figure 2 and by Table II. These local minima originate from aliasing and there exist several continuous-domain processes that result in similar discrete-domain power-spectrum upon sampling (Figure 3). These very processes generate the local minima. The peak response of these power spectra are distributed along the frequency axis in distinct bands as shown in Figure 4; these frequency bands are π[rad/time-unit] wide. This property suggests that every local minimum can be obtained by minimizing (48) while allocating initial conditions that correspond to a peak response at the required band, for instance, at $(k + {}^1\!/_2)\pi$ where k is the band index [31]. Another way of determining initial conditions will be described later. Following Theorem 3, the global minimum of (48) corresponds to a ML estimator and we suggest here to minimize the likelihood function using several initial conditions.

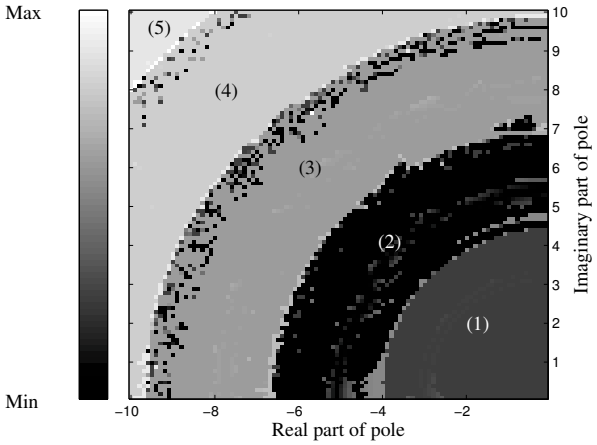


Fig. 2. Local minima of the likelihood function. Shown here are simulation results for a continuous-domain AR(2) process having poles $s_{1,2} = -1 \pm 5i$. Given a sampled version of such a process, the log-likelihood function (48) was repeatedly minimized using different initial conditions; the initial conditions are the real and imaginary part of the poles one starts with, and they are indicated by the x and y axes respectively. A detailed description of the five ring-like regions is given in Table II.

VI. THE ESTIMATION ALGORITHM

A. Proposed approach

The algorithm is described in Figure 5 while complementary information is given in Figures 6 and 11. The vector of parameters θ, which consists of the zeroes and the poles of (12), can also be represented by the polynomial coefficients of its numerator and of its denominator. These coefficients are real numbers and the numerical optimization was carried out using this type of parameterization. For example, the power-spectrum of an AR(2) process having two poles s_1, s_2 is given by

$$\Phi(s;\theta) = \frac{\sigma^2}{(s^2 - a_1 s + a_0)(s^2 + a_1 s + a_0)}, \tag{58}$$

TABLE II

DETAILED DESCRIPTION OF FIGURE 2.

Region	Log-likelihood Value	Estimated Poles	Frequency[b] [rad/time-unit]
1	-3656	$-2.4 \pm 2.1i$	0
2[a]	**-3675**	$\mathbf{-0.96 \pm 5.0}i$	**4.9**
3	-3557	$-0.46 \pm 7.8i$	7.8
4	-3067	$-0.25 \pm 11.1i$	11.1
5	-2212	$-0.16 \pm 14.0i$	14.0

[a] Global minimum of the log-likelihood function.

[b] Frequency of maximum response of $\Phi_L(j\omega)$.

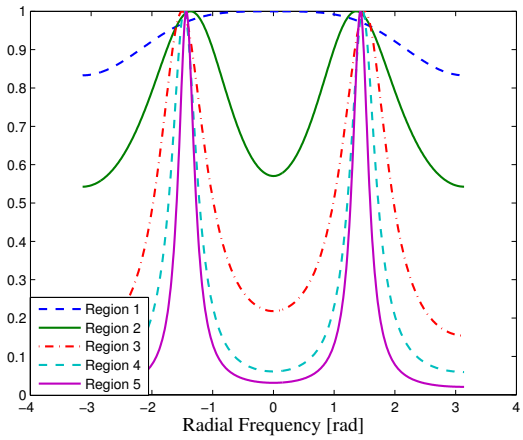


Fig. 3. Discrete-domain spectrum of sampled AR signals. Shown here is $\Phi_d(e^{j\omega})$ for the AR processes of Table II while considering a unit-sampling interval. These processes correspond to the local minima of Figure 2. The aliased spectrum shown here resembles each other in terms of their bandpass nature and in terms of the frequency of maximum response. Region 1, however, exhibits a low-pass signal, emphasizing the difference between the proposed ML-based estimation approach and other base-band methods.

Fig. 4. Frequencies of maximum response. Shown here by '$\times$' are the frequencies of maximum response of each local minimum of Figure 2. Theses frequencies are located within bands of width π [rad/time-unit]. Additional information is given in Table II.

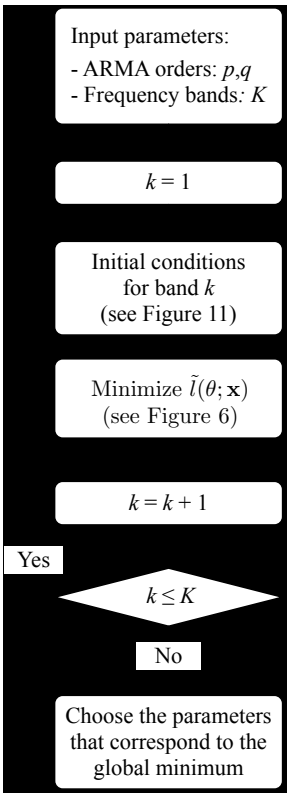


Fig. 5. Main stages of the proposed estimation algorithm.

where $a_0 = s_1 \cdot s_2$ and $a_1 = s_1 + s_2$. The proposed estimation algorithm aims at finding the coefficients a_0, a_1, and the variance of the innovation process σ^2. The dominant coefficient in this example (i.e., the one that changes substantially as the values of the poles change) is a_0. It has the most prominent effect on the geometric distance between the coefficients one starts with (i.e., the initial conditions for extracting every local minimum) and the coefficients of the local minimum one ends up with. In cases where s_1 and s_2 are mutual complex conjugate, the coefficient a_0 corresponds to their module value, as demonstrated in Figure 2. Every location in the image corresponds to a different set of initial conditions. This set was then used for minimizing the proposed approximation of the likelihood function and the color at that point indicates the value it converged to upon minimization. Initial conditions were determined by choosing various complex values for the pole s_1 and the radial appearance of the image stems from its module. For the general case of an ARMA(p,q) process, the parameters that are being optimized are $\theta = \left\{\sigma^2, a_0, a_1, \ldots, a_{p-1}, b_0, b_1, \ldots, b_{q-1}\right\}$, which gives rise to the following power-spectrum:

$$\Phi(s;\theta) = \sigma^2 \frac{\left[s^q + \sum_{n=0}^{q-1} b_n s^n\right]\left[(-s)^q + \sum_{n=0}^{q-1} b_n (-s)^n\right]}{\left[s^p + \sum_{m=0}^{p-1} a_m s^m\right]\left[(-s)^p + \sum_{m=0}^{p-1} a_m (-s)^m\right]}. \tag{59}$$

The proposed algorithm finds K local minima of the approximated likelihood function at consecutive frequency

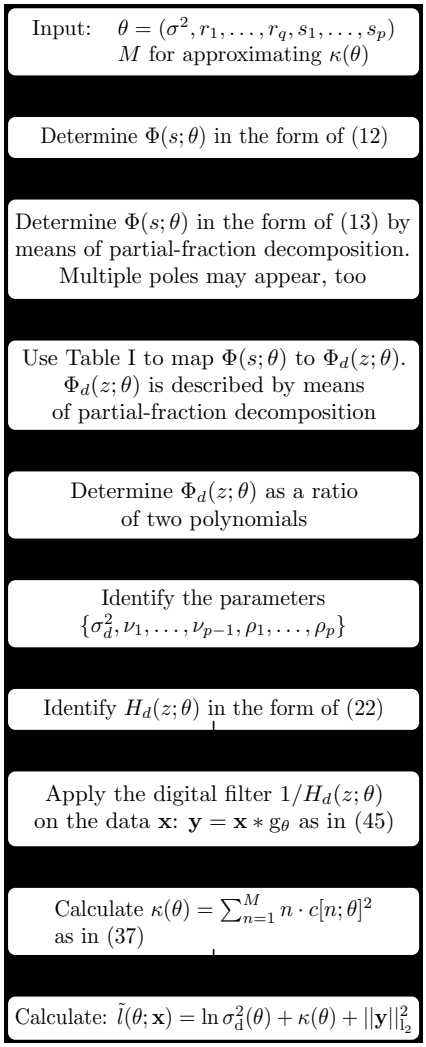


Fig. 6. A detailed description of the log-likelihood function $\tilde{l}(\theta;\mathbf{x})$ calculation. This workflow is part of the estimation algorithm of Figure 5.

bands and chooses the minimum value among them. Obtaining K different local minima requires K sets of initial conditions, and we suggest here to choose bandpass power spectra that have peak responses at $\omega_0 + \pi k$ [rad/time-unit], where k is the band index and where ω_0 is determined by fitting the available data with a discrete-domain AR(p) process and by extracting the frequency that corresponds to the maximum value of the power-spectrum. For arbitrary sampling-interval values, one should use $\frac{\pi k}{T}$ instead of πk. While there are many ways of obtaining such a band-pass spectrum, we provide a constructive way of doing so in the appendix. Initial values for σ can then be obtained by evaluating (14) at $t = 0$ and equating it to the variance of the available data. The value of K may be derived from known physical constraints of the problem at hand. When there is no such knowledge, this value can be determined

during the execution of the estimation algorithm: following Table II, likelihood values of local minima exhibit monotonicity; they decrease towards the global minimum and then start to monotonically increase as a function of the band index k. Such monotonicity then provides a decision rule for setting a value for K. It is noted that this value does not depend on the size of the data or on the ARMA order. It is related to the sampling rate value. A single sampling interval value partitions the frequency axis into non-overlapping segments which are $\frac{\pi}{T}$ wide, and for different sampling interval values a given continuous-domain model might be associated with different segments. Associating a segment with a power spectrum amounts to allocating the frequency of peak response to that segment. Every local minimum corresponds to a single segment and the choice of K, which is the number of segment the algorithm examines, should take this partitioning into account.

B. Discussion and relation with prior work

A precursor of Theorem 1 has been available for a while using the state-space representation. Several papers by Wahlberg, Söderström and Ljung [14], [15], [28] provide closed-form expressions for extracting discrete-time power spectrum from a given continuous-time model. These expressions involve exponentials of a matrix and integration of matrix elements; they also involve polynomial factorization that includes a parameterized inversion of a matrix. Larsson [45] derives explicit formulae for the first terms of the Maclaurin series of the variance and zeros of the sampled process as a function of the sampling interval variable T. These formulae allow one to analyze convergence for the limiting case of $T \to 0$. Autocorrelation functions of continuous-time ARMA processes are exponential splines, and the motivation for using such a parameterization stems from the fact that autocorrelation sequences of sampled processes are given by point wise values of these functions

Properties of the ARMA likelihood function have been investigated by Åström and Söderström [46]. Relying on the equivalence of the maximum-likelihood estimator with the prediction error variance minimization estimator, they showed that the likelihood function of a discrete-domain ARMA process has a unique global minimum when enough parameters are used in the fitted model. Our approximated likelihood function (48) has a global minimum, too, although it does not necessarily coincide with the global minimum of [46]. The reason for that is the zero-pole coupling that is introduced by the sampling process.

The large data sets we are considering in the present work require approximation of the likelihood function, and several authors addressed this problem in the past. Tsai and Chan [47] express the likelihood function in terms of a product of independent Gaussian variables that correspond to the discrete-time innovation process. The innovation process corresponds to the prediction error values and it is obtained by a linear predictor. This predictor is recursively defined and it depends on point-wise evaluations of the autocorrelation function. Jones and Vecchia [48] rely on the Gaussian property of the sampled process and they express the likelihood function in terms of the autocorrelation matrix. Their approximated likelihood function uses the sample variance for estimating the innovation process while assuming distinct roots of the power spectrum. The autocorrelation matrix is them inverted by the nearest-neighbor approximation method. The likelihood formulation we consider in our work takes advantage of the fact that the samples are taken at fixed intervals and finds a digital filter that can be directly applied to the sampled data. The

digital filter is then given in the z-domain, allowing one to use a direct form realization.

The estimation algorithm we suggest in the present work is related to the works of Söderström [28] and of Larsson, Mossberg and Söderström [17]. The author of [28] suggests three algorithms for computing the continuous-time counterpart of a known discrete-domain model. These algorithms can be utilized in indirect estimation approaches, for which the sampled data is used for estimating the discrete-domain model first. Two safeguards are noted in [28], though: existence of a continuous-domain model is not guaranteed for all discrete-domain models, and uniqueness of the continuous-domain poles is not guaranteed, either. If one, however, restricts the imaginary part of the continuous-domain poles to be in the interval $(-\frac{\pi}{T}, \frac{\pi}{T})$, as suggested in [28], the poles can then be uniquely resolved. This uniqueness condition imposes an upper bound on the sampling interval T. The authors of [17] introduce two estimation algorithms. The first algorithm relies on the indirect approach of [28] whereas the second algorithm consists of two steps: (1) replacing the differentiation operator by a difference operation and approximating the continuous-domain innovation by a discrete-domain white noise, (2) optimizing the discrete-domain model to minimize the variance of the discrete-domain innovation process.

Our approach to discretizing the continuous-domain model involves partial fraction decomposition and finite weighted sums. It provides a discrete-domain process whose autocorrelation sequence is in a perfect match with the point-wise values of the autocorrelation function. Additionally, the discrete-domain model we obtain by Theorem 1 is always guaranteed to have a continuous-domain counterpart. Our approach to the uniqueness problem does not impose restrictions on the sampling interval values. Instead, we are allowing for arbitrary values to be considered while exploiting the local minima properties of the likelihood function. In particular, we find several sets of estimated parameters, each corresponding to a local minimum, and choose the one that corresponds to the global minimum.

In terms of computational complexity, the proposed algorithm minimizes the likelihood function using several sets of initial conditions. Numerical optimization was also used in [27], [47], [48] and our algorithm repeats the optimization procedures several times. The number of repetitions, denoted K in Figure 5, is determined by the number of the local minima one wishes to find and it is a user-dependent parameter. Our spline formulation allows for fast implementation as it involves the parameter mapping of Theorem 1 and a direct form implementation of the digital filter (20).

VII. EXPERIMENTAL RESULTS

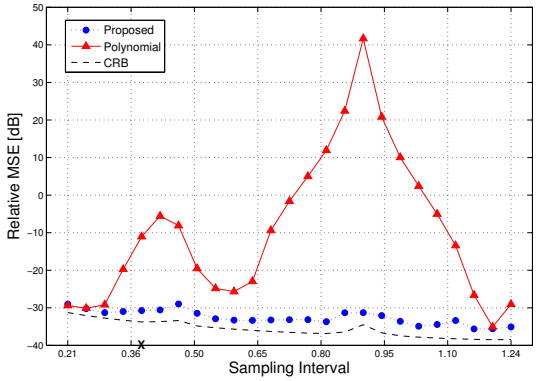
The proposed approach was implemented in Matlab using non-constrained numerical optimization. It was compared with the polynomial B-spline estimator of [27]. Sampled signals were generated by filtering a discrete-domain Gaussian white noise with the digital filter (20). Several sets of the parameters θ were considered. For every set, several Monte Carlo simulations were conducted; every simulation corresponds to a different sampling interval. Sampling-interval values were chosen so as to capture different aliasing configurations. A single Monte Carlo simulation involved 500 experiments and every experiment was carried out using $N = 10,000$ sample values. The variance parameter was set to $\sigma^2 = 1$ and was unknown to the estimation algorithm. The value of $\kappa(\theta)$ (42) was calculated using the first 500 terms in the infinite sum. The number of local minima that were examined was $K = 2$.

The estimation error for a Monte Carlo simulation is the relative MSE (Mean Square Error) between the estimated parameter and the correct one. For example, the estimation error of the single parameter a_0 is given by

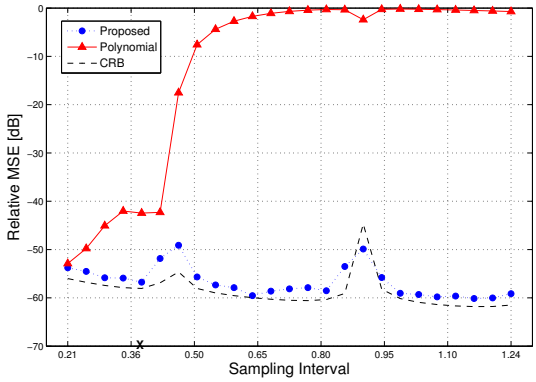
$$\epsilon(a_0) = -10 \log \left(\frac{\frac{1}{500} \sum_{n=1}^{500} (\hat{a}_{0,n} - a_0)^2}{(a_0)^2} \right), \tag{60}$$

where $\hat{a}_{0,n}$ is the estimation of a_0 at the n-th experiment.

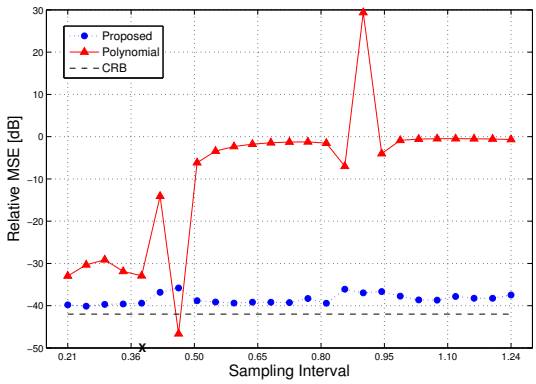
Experimental results for an AR(2) process are shown in Figure 7 and in Table III. The sampling interval values of the Figure were chosen so as to introduce various aliasing configurations. The maximum value of the continuous power spectrum of the AR model of Figure 7 is located at 5.05 [rad/time-unit] and the 3[dB] frequencies are 4.09 and 6.10 [rad/time-unit]. A sampling interval value of $T = 0.21$ [time-unit] corresponds to a sampling frequency of 29.92 [rad/time-unit] and it introduces minor aliasing effects. The value of $T = 1.24$ [time-unit] corresponds to a sampling frequency of 5.07 [rad/time-unit] and it introduces substantial aliasing effects. The various aliasing configurations are depicted in Figure 8. The CRB values were calculated according to [49]. We further describe in Figure 9 the proposed exponential-based approach from a frequency-domain point of view, emphasizing the flexibility of the exponential B-spline model to adapt to band-pass power spectra. An ARMA(2,1) estimation comparison is given in Figure 10 and in Table IV while considering various sampling aliasing configurations, too. We note that more sample values are required for the proposed estimation algorithm when the number of parameters increases. Also, the proposed algorithm relies on the fact that the number of samples N is relatively large and there is a trade off between N and ϵ. Our results indicate that the proposed approach outperforms the polynomial-based direct method while following the CRB for various aliasing configurations. It also guarantees that the estimated parameters correspond to a valid continuous-domain model whereas this property does not necessarily hold true for other discrete-domain-based methods, such as the minimization of the prediction error variance.



(a) The parameter a_0.



(b) The parameter a_1.



(c) The parameter σ^2.

Fig. 7. Comparison of estimation errors. Shown here are Monte Carlo simulation results for a continuous-domain AR(2) process. Sampling interval values were chosen so as to capture different aliasing configurations. The proposed exponential-based approach was compared with the polynomial model of [27]. The poles of the process are $s_{1,2} = -0.2 \pm 7i$ and the variance is $\sigma^2 = 1$. The number of samples is $N = 10,000$. The '$\mathbf{x}$' mark that appears on the x-axis corresponds to the sampling interval T_{max} for which $w_{max} = \pi/T_{max}$ is the frequency that corresponds to the maximum value of the continuous-domain power-spectrum. The CRB is depicted for comparison purposes.

TABLE III

COMPARISON OF ESTIMATION ERRORS FOR AR(2) PROCESSES.

Power Spectrum	**Sampling Interval**	**Estimation Error [dB]**					
		$\epsilon(a_0)$		$\epsilon(a_1)$		$\epsilon(\sigma^2)$	
$\Phi(s;\theta)$	**[time-unit]**	**[27]**	**Proposed**	**[27]**	**Proposed**	**[27]**	**Proposed**
$\frac{1}{(s^2+0.4s+49.04)(s^2-0.4s+49.04)}$ $(s_{1,2}=-0.2\pm 7i)$	0.3370	-19.76	-30.98	-42.03	-55.90	-31.89	-39.60
	0.6827	-9.40	-33.25	-1.09	-58.61	-1.46	-39.20
	1.1150	-13.42	-33.39	-0.40	-59.63	-0.48	-37.81
$\frac{1}{(s^2+s+9.25)(s^2-s+9.25)}$ $(s_{1,2}=-0.5\pm 3i)$	0.6059	-34.93	-34.97	-34.94	-45.89	-27.92	-36.99
	1.3050	-16.68	-35.79	-4.75	-43.77	-3.73	-35.81
	2.2370	6.93	-32.64	-1.86	-35.63	-8.16	-33.92
$\frac{1}{(s^2+6s+5)(s^2-6s+5)}$ $(s_{1,2}=-1,-5)$	0.9753	-21.09	-22.38	-17.95	-23.70	-11.73	-17.05
	2.3507	-3.65	-4.17	-3.44	-3.68	-1.19	-1.23
	3.6011	-1.83	-2.28	-1.31	-1.60	-0.27	-0.34

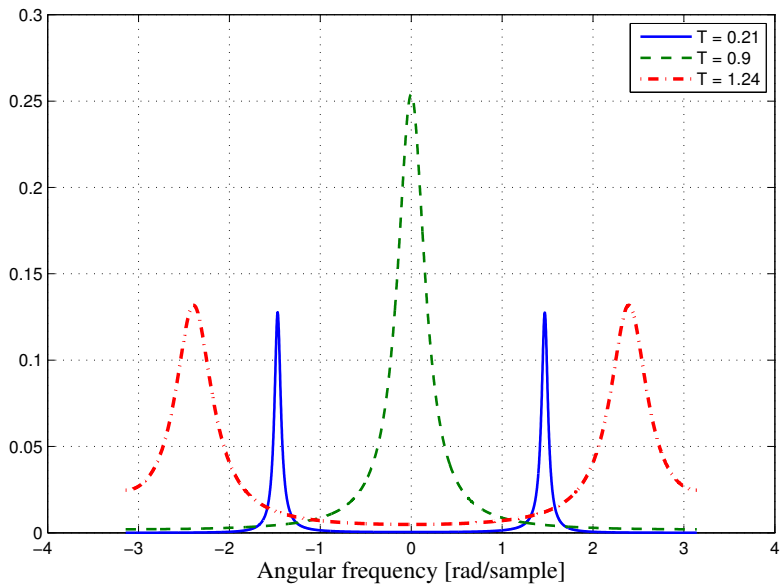


Fig. 8. Power spectra of sampled processes. Shown here are power spectra of discrete-domain processes that originate from the same continuous-domain process. Every power spectrum corresponds to a different sampling interval value. The poles of the continuous-domain AR process are $s_{1,2}=-0.2\pm 7i$.

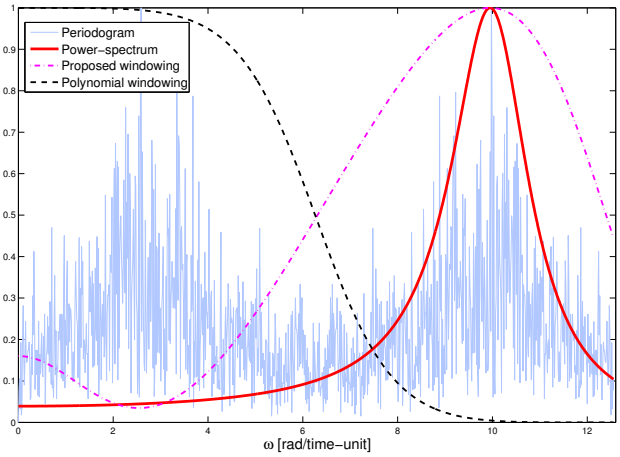
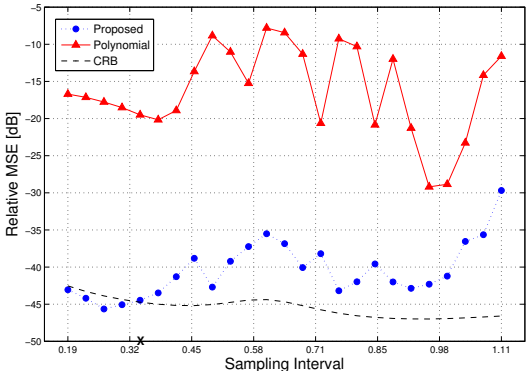


Fig. 9. A frequency-domain description of the proposed estimation approach. Shown here is a periodogram of sample values (absolute square value of the DFT), periodized in the range $[0, 4\pi]$. The power-spectrum of the AR(2) process is depicted in a solid red line. While the polynomial-based estimator captures a base-band signal (black dashed line), the proposed approach adequately identifies the proper frequency band (magenta dashed-dotted line). The proposed windowing function corresponds to the Fourier transform of the interpolator $\eta(t;\theta)$. The poles of the process are $s_{1,2} = -1 \pm 10i$ and the sampling interval takes a unit value, introducing prominent aliasing conditions.

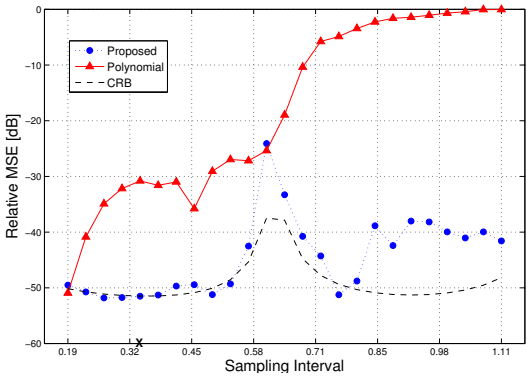
TABLE IV

COMPARISON OF ESTIMATION ERRORS FOR ARMA(2,1) PROCESSES.

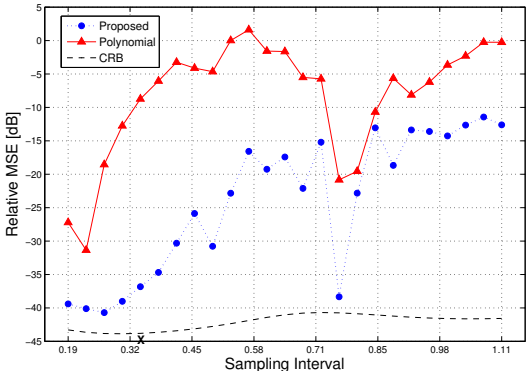
Power Spectrum	**Sampling Interval**	**Estimation Error [dB]**							
		$\epsilon(a_0)$		$\epsilon(a_1)$		$\epsilon(b_0)$		$\epsilon(\sigma^2)$	
$\Phi(s;\theta)$	**[time-unit]**	**[27]**	**Proposed**	**[27]**	**Proposed**	**[27]**	**Proposed**	**[27]**	**Proposed**
$\frac{(s+3)(-s+3)}{(s^2+2s+26)(s^2-2s+26)}$ $(r_1=-3,\ s_{1,2}=-1\pm 5i)$	0.3009	-18.52	-45.07	-32.18	-51.37	-12.76	-39.02	-23.37	-42.17
	0.6481	-8.44	-36.85	-18.96	-33.29	-1.65	-17.42	-10.43	-22.50
	1.1111	-11.60	-29.69	-0.05	-41.59	-0.26	-12.61	-11.83	-15.16
$\frac{(s+4)(-s+4)}{(s^2+2s+101)(s^2-2s+101)}$ $(r_1=-4,\ s_{1,2}=-1\pm 10i)$	0.1911	-19.51	-33.71	-32.18	-51.37	-12.76	-39.02	-23.37	-42.17
	0.4115	-14.48	-34.78	-18.96	-33.29	-1.65	-17.42	-10.43	-22.50
	0.7055	-19.96	-29.00	-0.05	-41.59	-0.26	-12.61	-11.83	-15.16
$\frac{(s+3)(-s+3)}{(s^2+8s+7)(s^2-8s+7)}$ $(r_1=-3,\ s_{1,2}=-1,-7)$	0.3250	-16.42	-21.08	-18.24	-17.26	-10.73	-20.36	1.30	-24.55
	1.3412	-4.63	-5.77	-3.49	-4.44	-4.56	-7.66	-15.31	-7.33
	1.8831	-2.97	-3.96	-2.00	-2.51	-3.65	-4.73	-9.11	-5.76



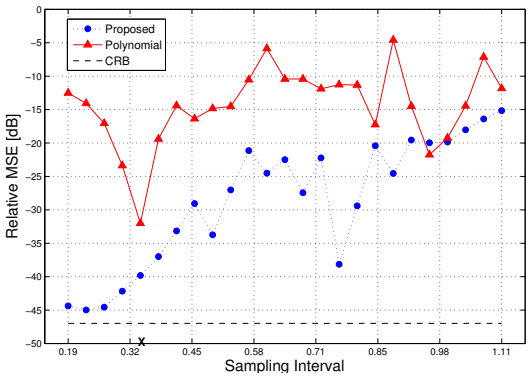
(a) The parameter a_0.



(b) The parameter a_1.



(c) The parameter b_0.



(d) The parameter σ^2.

Fig. 10. Comparison of estimation errors. Shown here are Monte Carlo simulation results for a continuous-domain ARMA(2,1) process. The poles of the process are $s_{1,2} = -1 \pm 5i$, the zero is $r_1 = -3$, and the variance is $\sigma^2 = 1$. The number of samples is $N = 100,000$.

VIII. CONCLUSIONS

In this work, we have proposed an estimator for continuous-domain ARMA parameters from sampled data that is based on the likelihood function. It utilizes an exponential B-spline framework while introducing an exact zero-pole coupling for the sampled process. The relation between the autocorrelation function and the autocorrelation sequence of the sampled process was investigated in both time- and frequency domains. Our approach relies on the known fact that the Cramér-Rao bound can be made arbitrarily small by increasing the number of samples while fixing the sampling interval at an arbitrary value. The likelihood function of the sampled process was then investigated and it was shown to posses local minima that originate from aliasing. The global minimum was experimentally shown to corresponds to the maximum-likelihood value. The only assumptions that were made throughout this work are that the number of available samples is relatively large and that the model orders are known, allowing one to replace the whitening matrix by a digital filter. This approximation was then shown to be valid when considering expected values of the likelihood function. Experimental results indicate that the proposed exponential-based approach closely follows the Cramér-Rao bound for various aliasing configurations, while imposing no restrictions on sampling rate values.

APPENDIX

ALLOCATION OF INITIAL CONDITIONS

The algorithm of Section VI repeatedly requires the allocation of band-pass power-spectrum. Each such allocation corresponds to a single local minimum extraction, and is based on equating the derivative of the power-spectrum $\Phi(s;\theta)$ to zero while substituting $s = j\omega_k$, where ω_k is the required frequency of maximum response. The coefficients of $\Phi_L(s)$ are then obtained by numerically minimizing the derivative value at ω_k. The initial value for the innovation variance is determined by the autocorrelation function of the model, which is the inverse Laplace transform of the power-spectrum, and by the variance of the available data. A description of the allocation workflow is given in Figure 11.

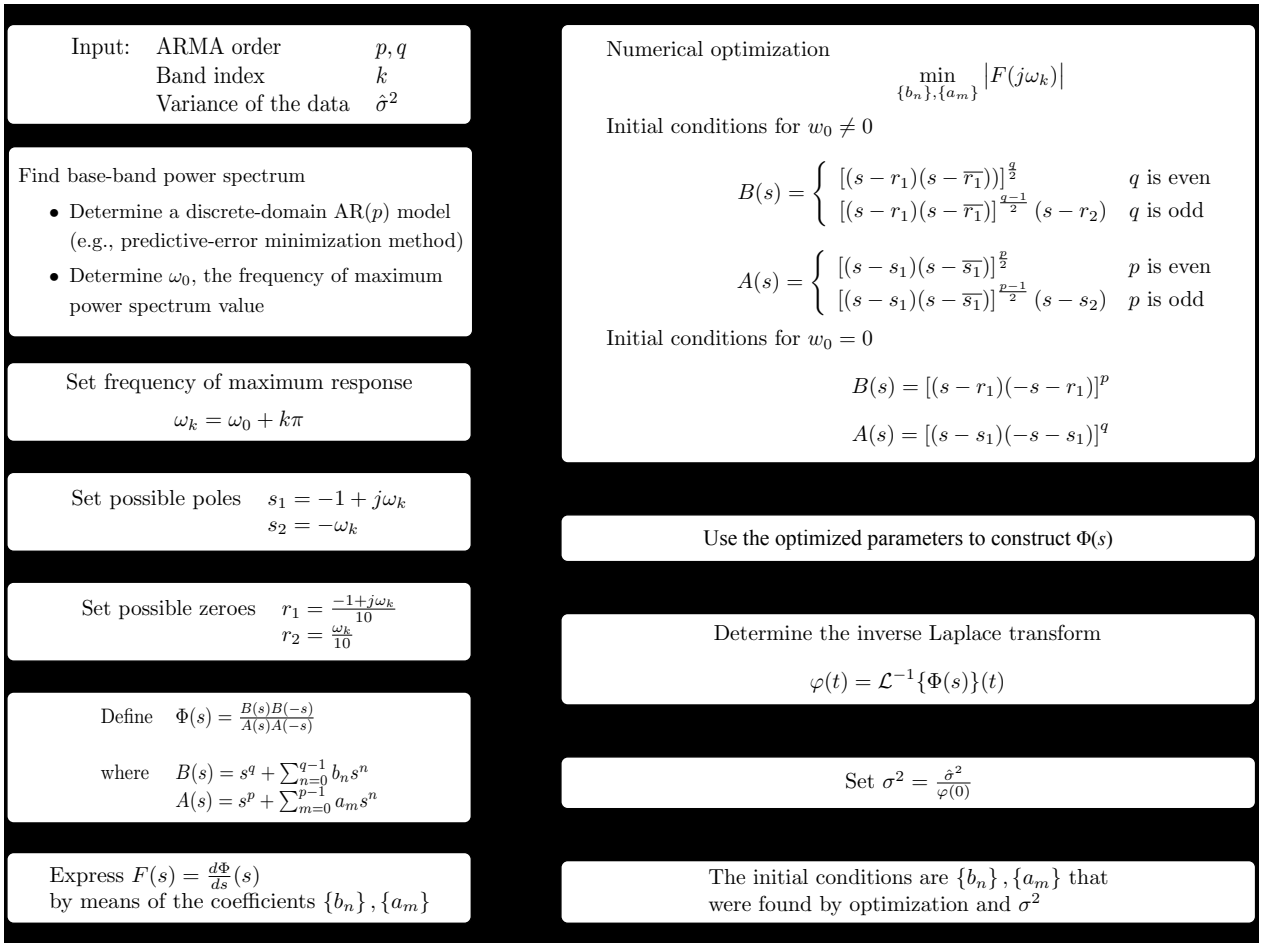


Fig. 11. Allocating initial conditions. Every local minimum of the likelihood function (48) is obtained by a different set of initial conditions. This workflow is part of the estimation algorithm of Figure 5.

REFERENCES

[1] L. Ljung, *System Identification*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, 1999.

[2] S. Haykin, *Adaptive Filter Theory*, 4th ed. Upper Saddle River, NJ: Prentice-Hall, 2002.

[3] I.-T. Lim and B. G. Lee, "Lossless pole-zero modeling for speech signals," *IEEE Trans. on Speech and Audio Processing*, vol. 1, no. 3, pp. 269–276, July 1993.

[4] H. Kirshner and M. Porat, "On the role of exponential spline in image interpolation," *IEEE Trans. Signal Processing*, vol. 18, no. 10, pp. 2198–2208, October 2009.

[5] J. M. Francos, A. Z. Meiri, and B. Porat, "A unified texture model based on a 2-D Wold like decomposition," *IEEE Trans. Signal Process.*, vol. 41, pp. 2665–2678, August 1993.

[6] S. Nemirovski and M. Porat, "On texture and image interpolation using Markov models," *Signal Processing: Image Communication*, vol. 24, pp. 139–157, May 2009.

[7] R. A. Adams, *Sobolev spaces*. New York, NY: Academic Press, 1975.

[8] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*. Kluwer, 2000.

[9] T. F. Chan and J. Shen, *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, 2005.

[10] S. Angenent, E. Pichon, and A. Tannenbaum, "Mathematical methods in medical image processing," *Bulletin of the American Mathematical Society*, pp. 365–396, 2006.

[11] S. Ramani, P. Thévenaz, and M. Unser, "Regularized interpolation for noisy images," *IEEE Trans. on Medical Imaging*, vol. 29, no. 2, pp. 543–558, February 2010.

[12] J.-C. Baritaux, K. Hassler, and M.Unser, "An efficient numerical method for general l_p regularization in fluorescence molecular tomography," *IEEE Trans. on Medical Imaging*, 2010, in press.

[13] K. J. Åström, *Introduction to Stochastic Control Theory*. Academic Press, 1970.

[14] B. Wahlberg, "Limit results for sampled systems," *International Journal of Control*, vol. 78, no. 3, pp. 1267–1283, 1988.

[15] B. Wahlberg, L. Ljung, and T. Söderström, "Sampling of continuous time stochastic processes," *Control –Theory and Advanced Technology*, vol. 9, no. 1, pp. 99–112, march 1993.

[16] T. Söderström, *Discrete-Time Stochastic Systems*, 2nd ed. London: Springer-Verlag, 2002.

[17] E. K. Larsson, M. Mossberg, and T. Söderström, "An overview of important practical aspects of continuous-time ARMA system identification," *Circuits Systems Signal Processing*, vol. 25, no. 1, pp. 17–46, May 2006.

[18] H. Garnier and L. Wang, Eds., *Identification of Continuous-Time Models from Sampled Data*. London: Springer-Verlag, 2008.

[19] R. Johansson, "Identification of continuous-time models," *IEEE Trans. signal processing*, vol. 42, no. 4, pp. 887–897, April 1994.

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

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

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

[23] G. C. Goodwin, J. I. Yuz, and H. Garnier, "Robustness issues in continuous-time system identification from sampled data," in *Proc. 16th IFAC World Congress*, Prague, Czech Republic, July 2005.

[24] H. Fan, T. Söderström, M. Mossberg, B. Carlsson, and Y. Zou, "Estimation of continuous-time AR process parameters from discrete-time data," *IEEE Trans. Signal Processing*, vol. 47, no. 5, pp. 1232–1244, May 1999.

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

[26] R. Pintelon, J. Schoukens, and Y. Rolain, *Frequency-Domain Approach to Continuous-Time System Identification: Some Practical Aspects*. London: Springer-Verlag, 2008, pp. 214–248.

[27] J. Gillberg and L. Ljung, "Frequency-domain identification of continuous-time ARMA models from sampled data," *Automatica*, vol. 45, pp. 1371–1378, 2009.

[28] T. Söderström, "Computing stochastic continuous-time models from ARMA models," *International Journal of Control*, vol. 53, no. 6, pp. 1311–1326, May 1991.

[29] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. on Signal Processing*, vol. 53, no. 4, pp. 1425–1438, April 2005.

[30] M. Unser, "Cardinal exponential splines: Part II—Think analog, act digital," *IEEE Trans. on Signal Processing*, vol. 53, no. 4, pp. 1439–1449, April 2005.

[31] S. Maggio, H. Kirshner, and M. Unser, "Continuous-time AR model identification: Does sampling rate really matter?" in *European Signal Processing Conference*, Denmark, 2010.

[32] D. Marelli and M. Fu, "A continuous-time linear system identification method for slowly sampled data," *IEEE Trans. Signal Processing*, vol. 58, no. 5, pp. 2521–2533, May 2010.

[33] H. Kirshner, M. Porat, and M. Unser, "A stochastic minimum-norm approach to image and texture interpolation," in *European Signal Processing Conference*, Denmark, 2010.

[34] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA: Birkhäuser, 2002.

[35] J. J. F. Fournier, "Sharpness in Young's inequality for convolution," *Pacific Journal of Mathematics*, vol. 72, no. 2, 1977.

[36] C. K. Chui, J. D. Ward, and P. W. Smith, "Cholesky factorization of positive definite bi-infinite matrices," *Numer. Func. Anal. and Optimiz.*, vol. 5, no. 1, pp. 1–20, 1982.

[37] Y. Katznelson, *An introduction to harmonic analysis*, 3rd ed. Cambridge University Press, 2004.

[38] E. K. Larsson and E. G. Larsson, "The CRB for parameter estimation in irregularly sampled continuous-time ARMA system," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 197–200, February 2004.

[39] B. Friedlander, "On the computation of the Cramér-Rao bound for ARMA parameter estimation," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-32, no. 4, pp. 721–727, August 1984.

[40] G. Szegő, "On certain Hermitian forms associated with the Fourier series of a positive function," *Comm. Sém. Math. Univ. Lund*, pp. 228–238, 1952.

[41] R. E. Hartwig and M. E. Fisher, "Asymptotic behavior of Toeplitz matrices and determinants," *Archive for Rational Mechanics and Analysis*, vol. 32, no. 3, pp. 190–225, January 1969.

[42] E. Basor, "Asymptotic formulas for Toeplitz determinants," *Trans. of the American Mathematical Society*, vol. 239, pp. 33–65, May 1978.

[43] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[44] P. Whittle, "The analysis of multiple stationary time series," *Journal of the Royal Statistical Society, Serise B (Methodological)*, vol. 15, no. 1, pp. 125–139, 1953.

[45] E. K. Larsson, "Limiting sampling results for continuous-time ARMA systems," *International Journal of Control*, vol. 78, no. 7, pp. 461–473, May 2005.

[46] K. J. Åström and T. Söderström, "Uniqueness of the maximum likelihood estimates of the parameters of an ARMA model," *IEEE Trans. on Automatic Control*, vol. AC-19, no. 6, pp. 769–773, december 1974.

[47] H. Tsai and K. S. Chan, "Maximum likelihood estimation of linear continuous time long memory processes with discrete time data," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, vol. 67, no. 5, pp. 703–716, 2005.

[48] R. H. Jones and A. V. Vecchia, "Fitting continuous ARMA models to unequally spaced spatial data," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 947–954, 1993.

[49] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. New Jesey: Prentice Hall, 1997.