MAXIMUM-LIKELIHOOD IDENTIFICATION OF SAMPLED GAUSSIAN PROCESSES

Hagai Kirshner[1], *Simona Maggio*[2] *and Michael Unser*[1]

[1]Biomedical Imaging Group, EPFL, Lausanne
[2]Department of Electronics, Computer Science and Systems, University of Bologna
Emails: hagai.kirshner@epfl.ch, simona.maggio@unibo.it, michael.unser@epfl.ch

ABSTRACT

This work considers sampled data of continuous-domain Gaussian processes. We derive a maximum-likelihood estimator for identifying autoregressive moving average parameters while incorporating the sampling process into the problem formulation. The proposed identification approach introduces exponential models for both the continuous and the sampled processes. We construct a likelihood function from a digitally-filtered version of the available data which is asymptotically exact. This function has several local minima that originate from aliasing, plus a global minimum that corresponds to the maximum-likelihood estimator. We further compare the performance of the proposed algorithm with other currently available methods.

Keywords— Sampling, exponential splines, continuous-domain stochastic processes, maximum-likelihood estimation.

1. INTRODUCTION

Continuous-domain Gaussian processes are widely used in the fields of control theory and in signal/image processing and analysis. Such processes are often assumed to be the output of a linear time-invariant (LTI) system that is driven by white Gaussian noise. LTI systems with a rational transfer function give rise to autoregressive moving-average (ARMA) Gaussian processes. 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 modeling of physical phenomena, system identification, and 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. Recent works on this subject fall in two broad categories: direct and indirect [5]. Most direct methods specify some equivalent discrete-domain model that is characterized by the continuous-domain parameters. This discrete-domain model is then used for minimizing a cost function. In this way, the required continuous-domain parameters are directly estimated by the minimization process [4, 1, 8, 3, 7]. Another example of a direct method consists of power-spectrum parameterization [9, 2]. Indirect methods, on the other hand, rely on standard discrete-domain system identification methods such as the minimization of the prediction error. The discrete-domain system is then mapped to a continuous-domain one while preserving some required properties, such as the bi-linear transform. Such mappings are commonplace when designing digital filters.

Motivated by the deterministic theory of LTI systems, we exploit in this work the mathematical formulation of exponential splines. A first study of this property was recently suggested in [6] for the autoregressive model. Considering an ideal sampling procedure, 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. Another point we address in this work is the Cramér-Rao bound. This bound converges to zero for any sampling interval value with increasing number of data points where the sampling interval can take an arbitrary value. 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, as suggested here.

2. ARMA PROCESSES AND SPLINES

2.1. 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 variance innovation is a symmetric exponential

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

The spectral density function is then

$$\Phi(j\omega; s_1) = \frac{1}{(j\omega - s_1)(-j\omega - s_1)}. \tag{2}$$

The spectral density function of the sampled process is

$$\Phi_d(e^{j\omega}; s_1) = \frac{e^{2s_1} - 1}{2s_1} \cdot \frac{1}{(1 - e^{s_1}e^{-j\omega})(1 - e^{s_1}e^{j\omega})}. \tag{3}$$

The important observation is that one is able to link the continuous-domain and the discrete-domain autocorrelations via the Shannon-like interpolation formula

$$\varphi(t; s_1) = \sum_{n=-\infty}^{\infty} \varphi[n; s_1]\beta(t - n; s_1), \tag{4}$$

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

$$\hat{\beta}(\omega;s_1) = \frac{2s_1}{e^{2s_1}-1}\cdot\frac{(1-e^{s_1}e^{-j\omega})(1-e^{s_1}e^{j\omega})}{(j\omega-s_1)(-j\omega-s_1)}. \quad (5)$$

Observe that the latter expression is also equal to the ratio of (2) and (3). The key property here is that $\beta(t;s_1)$ is compactly supported, which is not directly apparent from the Fourier-domain expression (5). In fact, $\beta(t;s_1)$ is an exponential B-spline and the above method generalizes for higher-order systems.

2.2. General Case

A continuous-domain ARMA process is fully characterized by its autocorrelation function. The parameters of such a function are given by the following vector

$$\theta = \left\{\sigma^2, r_1, \ldots, r_q, s_1, \ldots, s_p\right\}, \quad (6)$$

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 variance is σ^2. Additionally, $p > q$. The Laplace transform of the corresponding autocorrelation function is given by

$$\Phi(s;\theta) = \sigma^2\frac{\prod_{k=1}^{q}(s-r_k)(-s-r_k)}{\prod_{k=1}^{p}(s-s_k)(-s-s_k)}. \quad (7)$$

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

$$\beta(t;\theta) = \mathcal{F}^{-1}\left\{\prod_{k=1}^{q}(j\omega-r_k)(-j\omega-r_k)\cdot \quad (8)\right.$$

$$\left.\cdot\prod_{k=1}^{p}\frac{(1-e^{j\omega+s_k})(1-e^{-j\omega+s_k})}{(j\omega-s_k)(-j\omega-s_k)}\right\}(t). \quad (9)$$

Definition 2 *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 $\hat{\beta}_d(\omega) = \mathcal{B}_d(e^{j\omega})$.*

Theorem 1 *A discrete-domain process that is defined by the ideal samples of the continuous-domain processes (7) on a uniform grid can be realized by a causal and stable digital filter applied to discrete-domain white Gaussian noise. The filter is*

$$H_d(z;\theta) = \sigma_d(\theta)\frac{\prod_{k=1}^{p-1}(1-\nu_k(\theta)z^{-1})}{\prod_{k=1}^{p}(1-\rho_k(\theta)z^{-1})}, \quad (10)$$

where

$$\rho_k(\theta) = e^{s_k} \quad (11)$$

$$\nu_k(\theta) = \text{roots of } \mathcal{B}_d(z;\theta) \text{ inside the unit circle} \quad (12)$$

$$\sigma_d^2(\theta) = \sigma^2\frac{\mathcal{B}_d(1;\theta)}{\prod_{k=1}^{p-1}(1-\nu_k(\theta))^2}. \quad (13)$$

The inverse filter is causal and stable, too.

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

$$\varphi(t;\theta) = \sum_{n=-\infty}^{\infty}\varphi(n;\theta)\cdot\eta(t-n;\theta), \quad (14)$$

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

$$\hat{\eta}(\omega;\theta) = \frac{\hat{\beta}(\omega;\theta)}{\hat{\beta}_d(\omega;\theta)}. \quad (15)$$

This is the generalization of (4) for arbitrary ARMA(p,q) processes.

3. MAXIMUM-LIKELIHOOD ESTIMATION

Let a continuous-domain ARMA(p,q) process be given by its uniform samples only. Considering a large number of data points $N \gg 1$, the Fisher information matrix is

$$I_{k,l}(\theta) \cong \frac{N}{4\pi}\int_0^{2\pi}\left(\frac{1}{\hat{\varphi}_d(\omega;\theta)}\frac{\partial\hat{\varphi}_d(\omega;\theta)}{\partial\theta_k}\right)\cdot \cdot\left(\frac{1}{\hat{\varphi}_d(\omega;\theta)}\frac{\partial\hat{\varphi}_d(\omega;\theta)}{\partial\theta_l}\right)d\omega, \quad (16)$$

where $k,l = 1,\ldots,p+q+1$. The CRB (Cramér-Rao Bound) is then the inverse of $I(\theta)$. It holds that $\hat{\varphi}_d(\omega;\theta)$ is not dependent upon N nor is the integrand of (16). Therefore, as the number of available samples becomes larger, the CRB becomes smaller regardless of T. This 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. Motivated by this observation, we approximate the log-likelihood function by means of a digital filter.

Definition 3 *Let θ be known and let $\mathbf{x}$ be N uniform ideal samples of the continuous-domain process (7) taken on a unit-interval grid. The log-likelihood function of $\mathbf{x}$, including a sign inversion, is*

$$l(\theta;\mathbf{x}) = \ln|\Sigma| + \mathbf{x}^T\Sigma^{-1}\mathbf{x}. \quad (17)$$

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

Definition 4 *Let θ be known. Then, the digital filter g_θ is given by $G_d(z) = 1/H_d(z)$ where $H_d(z)$ is given in (10).*

Definition 5 *Let θ be known. Then,*

$$\kappa(\theta) = \sum_{n=1}^{\infty} n\cdot c[n;\theta]^2, \quad (18)$$

where

$$c[n;\theta] = \frac{1}{n}\left\{\nu_1^n(\theta) + \ldots + \nu_{p-1}^n(\theta) - \rho_1^n(\theta) - \ldots - \rho_p^n(\theta)\right\}. \quad (19)$$

Theorem 2 *Let θ be known and let $\mathbf{x}$ be N uniform ideal samples of the continuous-domain process (7) taken on a unit-interval grid. Then,*

$$\lim_{N\to\infty} \mathbf{E}\Big(\mathrm{l}(\theta;\mathbf{x}) - \tilde{\mathrm{l}}(\theta;\mathbf{x})\Big) = 0, \tag{20}$$

where

$$\tilde{l}(\theta;\mathbf{x}) = N\ln\sigma_d^2(\theta) + \kappa(\theta) + \|\mathbf{x} * g_\theta\|_{\ell_2}^2. \tag{21}$$

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

The log-likelihood function (21) has several local minima. These local minima originate from aliasing and there exist several continuous-domain processes that result in similar discrete-domain power-spectrum upon sampling. These very processes generate the local minima. The peak response of these power spectra are distributed along the frequency axis in distinct bands that are π[rad/time-unit] wide. This property suggests that every local minimum can be obtained by minimizing (21) while allocating initial conditions that correspond to a peak response at the required band. Following Theorem 2, the global minimum of (21) corresponds to a ML estimator and we suggest here to minimize the likelihood function using several initial conditions.

4. EXPERIMENTAL RESULTS

The proposed approach was implemented in Matlab. It was compared with the polynomial B-spline estimator of [2]. Several sets of the parameters θ were considered. The Monte Carlo simulations were carried out for various sampling interval values. These 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 $N = 100,000$ sample values. The variance parameter was set to $\sigma^2 = 1$ and was unknown to the estimation algorithm. The value of $\kappa(\theta)$ (18) 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 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

$$\epsilon(a_0) = -10\log\left(\frac{\frac{1}{500}\sum_{n=1}^{500}(\hat{a}_{0,n} - a_0)^2}{(a_0)^2}\right), \tag{22}$$

where $\hat{a}_{0,n}$ is the estimation of a_0 at the n-th experiment. The parameters $\{a_k\}_{k=0}^{p}$ and $\{b_k\}_{k=0}^{q}$ are the coefficients of $\prod_{k=1}^{p}(s - s_k)$ and $\prod_{k=1}^{q}(s - r_k)$, respectively.

An ARMA(2,1) estimation comparison is given in Figure 1 and in Table 1. Our results indicate that the proposed approach outperforms the polynomial-based direct method. 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.

5. CONCLUSIONS

In this work, we have proposed a maximum-likelihood estimator for continuous-domain ARMA parameters from sampled data. It utilizes an exponential B-spline framework while introducing an exact zero-pole coupling for the sampled process. It was shown that the Cramér-Rao bound can be made arbitrarily small by increasing the number of samples while considering arbitrary sampling interval values. The likelihood function of the sampled process was approximated by means of a digital filtering operation and was shown be valid in the limit-in-the-mean sense. Experimental results indicate that the proposed exponential-based approach is a preferable choice over currently available methods that are restricted to relatively high sampling rates.

6. REFERENCES

[1] 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*, 47(5):1232–1244, May 1999.

[2] Jonas Gillberg and Lennart Ljung. Frequency-domain identification of continuous-time ARMA models from sampled data. *Automatica*, 45:1371–1378, 2009.

[3] 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.

[4] R. Johansson. Identification of continuous-time models. *IEEE Trans. signal processing*, 42(4):887–897, April 1994.

[5] Erik K. Larsson, Magnus Mossberg, and Torsten Söderström. An overview of important practical aspects of continuous-time ARMA system identification. *Circuits Systems Signal Processing*, 25(1):17–46, May 2006.

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

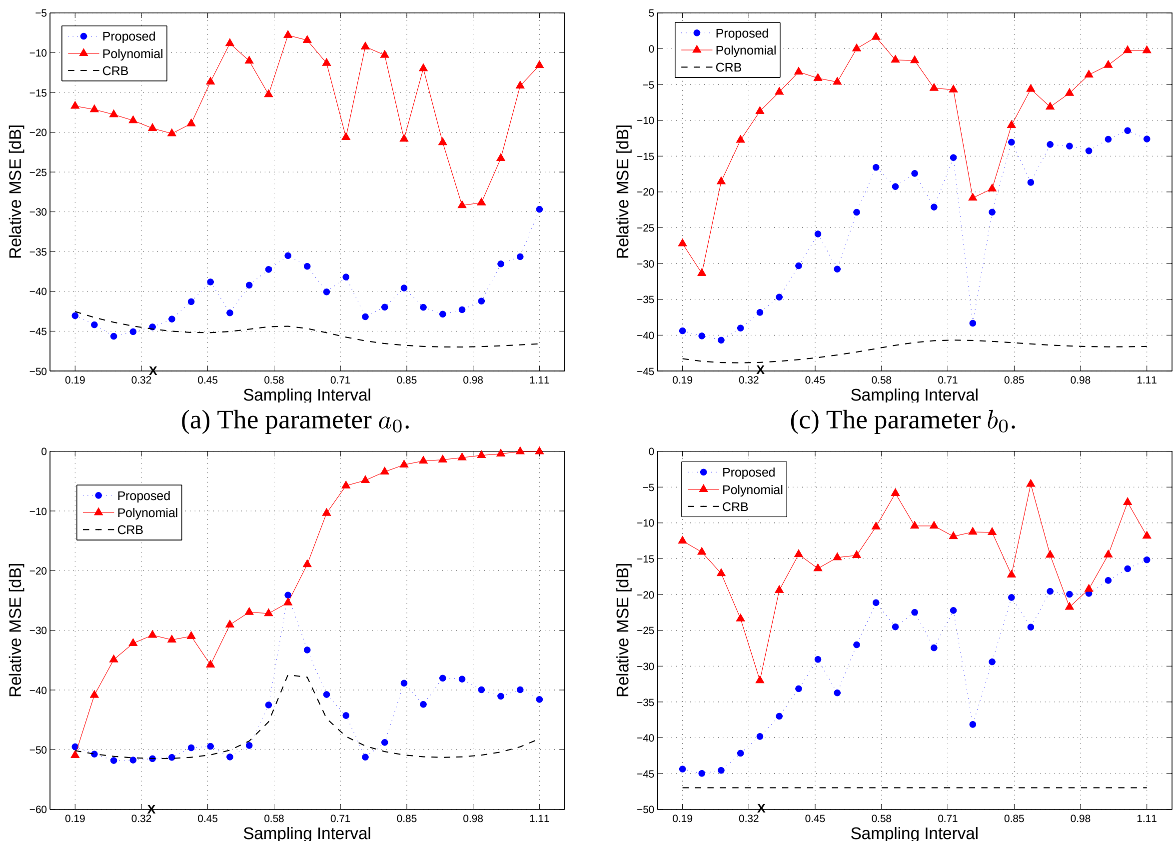
[7] M. Mossberg. Estimation of continuous-time stochastic signals from sample covariances. *IEEE Trans. Signal Processing*, 56(2):821–825, February 2008.

[8] Dinh-Tuan Pham. Estimation of continuous-time autoregressive model from finely sampled data. *IEEE Trans. Signal Processing*, 48(9):2576–2584, September 2000.

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

Table 1. Comparison of estimation errors for ARMA(2,1) processes.

Power Spectrum $\Phi(s;\theta)$	Sampling Interval [time-unit]	Estimation Error [dB]							
		$\epsilon(a_0)$		$\epsilon(a_1)$		$\epsilon(b_0)$		$\epsilon(\sigma^2)$	
		[2]	Proposed	[2]	Proposed	[2]	Proposed	[2]	Proposed
$\frac{(s-3)(s+3)}{(s^2+2s+26)(s^2-2s+26)}$	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
$(r_1=-3,\ s_{1,2}=-1\pm 5i)$	1.1111	-11.60	-29.69	-0.05	-41.59	-0.26	-12.61	-11.83	-15.16
$\frac{(s-4)(s+4)}{(s^2+2s+101)(s^2-2s+101)}$	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
$(r_1=-4,\ s_{1,2}=-1\pm 10i)$	0.7055	-19.96	-29.00	-0.05	-41.59	-0.26	-12.61	-11.83	-15.16



(a) The parameter a_0.

(c) The parameter b_0.

(b) The parameter a_1.

(d) The parameter σ^2.

Fig. 1. 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 $N = 100,000$.