21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Spectral Estimation from Covariances and Inverse-Covariances

Per Enqvist

***Abstract*— In many practical applications second order moments are used for estimation of the power spectrum. The moments are estimated from the available data and then the inverse problem to seek a spectrum consistent with the moments is solved. The focus here is on the second step in this procedure. Usually, the second order moments considered are the covariances of the process, and well known methods are available for efficiently determining the autoregressive part of a generating model from them. The autoregressive part adds poles to the model and in order to also have zeros additional information needs to be determined from the data. Here, we consider the case where second order moments of the inverse of the power spectrum, so called inverse covariances, also are given. The exact inverse problem corresponding to autoregressive moving average models with one pole and one zero is studied explicitly, and the solvability of the inverse problem is considered. Then different exact and approximative interpolation problems are considered from a global optimization approach. The exact interpolation result generalizes a result by T.T. Georgiou on a maximally random process from a smoothing perspective. For the approximation a quadratic function of the interpolation error is minimized in one approach and the distance between interpolants of the covariances and the inverse covariances are minimized in the second approach.**

***Index Terms*— 93B15, 93B30, 93B50.**

Inverse covariances, robust spectral estimation, spectral distances, geometry of spectral measures, smoothing.

I. INTRODUCTION

Spectral estimation is an important tool used for example to determine noise models, in speech processing and for prediction and smoothing of signals. Covariances are widely used for estimation of the power spectral densities, and in particular for AR-models. Efficient algorithms exists for determining the models, *e.g.*, the Yule-Walker method and the least-squares method [11, Ch. 3.4]. Several other model structures for covariance interpolators have also been studied, for example in [8], [1], [6].

To improve the spectral estimation, the inverse-covariances [5] are also considered here (together with the covariances). Using more information about the process a better estimate of the power spectral density should be obtained. But this also raises questions of when the different types of information can be joined together, can the two different types of interpolation conditions be matched simultaneously, and does this depend on which models are considered. In Section II, these questions are considered for the exact interpolation problem.

In practice the second order moments have to be estimated from data. Standard methods for estimating covariances are well-known [11, Ch. 2.2]. By taking the inverse of a PSD-estimate and then the inverse Fourier transform estimates of the inverse covariances can be obtained [4]. However, the estimation of the moments are not considered here. Independently of how the moments are obtained, there will be some uncertainty on the real values of the moments and in practice it does not make sense to enforce exact matching, hence approximative matching will be considered.

One new approach in this paper is the fitting of two spectral densities to covariances and inverse-covariances, respectively, whilst minimizing the distance between the two estimates.

In Section II some basic properties of covariances and inverse-covariances are described. In Section III some different optimization approaches to the interpolation problem are considered, and in Section IV we draw the conclusions.

II. BACKGROUND

Consider the discrete time stochastic process $y(t), t \in \mathbb{Z}$ which is $m-$dimensional, zero-mean, and second-order stationary with power spectrum $d\mu$. The covariance (or, equivalently, autocorrelation) samples are defined as

$$c_k := \mathcal{E}\{y(t)y(t-k)^*\}, \text{ for } k = 0, \pm 1, \pm 2, \ldots,$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator. These are the Fourier coefficients of the power spectrum $d\mu$ of the process:

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta} d\mu(\theta) \text{ for } k = 0, \pm 1, \pm 2 \ldots, \quad (1)$$

and consequently $c_{-k} = c_k^*$. The power spectrum is thought of as a non-negative Hermitian measure on the unit circle $\mathbb{T} = \{z = e^{i\theta} : \theta \in (-\pi, \pi]\}$ which, for simplicity, is identified with the interval $(-\pi, \pi]$.

Non-negativity of the power spectrum can be characterized in terms of the covariances by the non-negativity of the Hermitian block-Toeplitz matrices [9], [10]

$$T = \begin{bmatrix} c_0 & c_{-1} & \cdots & c_{-p} \\ c_1 & c_0 & \cdots & c_{-p+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_p & c_{p-1} & \cdots & c_0 \end{bmatrix}, \quad (2)$$

for $p \geq 0$.

The inverse problem of determining a spectrum $d\mu$ from a set of finite covariances is known as the trigonometric moment problem. Whenever the corresponding Toeplitz matrix T is positive definite there is an infinite family of spectra satisfying the covariance constraints.

Assume that the measure $d\mu$ is absolutely continuous and that $d\mu = \Phi d\theta$, where Φ is the spectral density. The covariances (C) in (1) are then given by [11]

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta}\Phi(e^{i\theta})d\theta, \text{ for } k = 0, \pm 1, \pm 2 \ldots, \quad (3)$$

ISBN: 978-90-367-6321-9

and define the auto-correlations (AC) as $r_k := c_k/c_0$.

Assume also that the measure is coercive so that the spectral density is positive on the unit circle, hence Φ^{-1} exists. The inverse covariances (IC) are defined as [5]

$$c_k^i = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta}\left(\Phi(e^{i\theta})\right)^{-1} d\theta,\ \text{ for } k = 0, \pm 1, \pm 2 \ldots, \tag{4}$$

and define the inverse auto-correlations (IAC) as $r_k^i := c_k^i/c_0^i$.

Directly estimating a MA model from the AC is well known to not always be possible by exact interpolation [11, Ch. 3.6]. One of the advantages with the IAC is that a MA model can always be estimated by using the Yule-Walker equations to match an AR model to the inverse auto-correlation function (IACF). A tempting idea is therefore to estimate ARMA models from the AC and IAC by using the Yule-Walker algorithm to determine the AR-part from the AC and to determine the MA-part from the IAC. However, the combined ARMA model will then neither match the AC nor IAC moments.

Another approach is to determine an MA model by applying the Yule-Walker algorithm on the IAC and then using this MA model (zeros) and the C to determine an ARMA-model using [2]. This model will match the C but not the IAC. To match both C and IAC a simultaneous interpolation problem has to be considered.

However, not any given C and IAC data can be interpolated exactly. Consider the spectral density of an ARMA(1,1) system

$$\Phi(z) = \rho\frac{(z+p)(z^{-1}+p)}{(z+a)(z^{-1}+a)}, \tag{5}$$

where a and p have absolute value less than one. Then the positive real part of Φ is given by

$$\Phi_+(z) = \frac{d}{2} + \frac{b}{z+a}, \tag{6}$$

where

$$d = \rho\frac{1+p^2-2pa}{1-a^2}, \quad b = \rho\frac{(p-a)(1-ap)}{1-a^2} \tag{7}$$

Then the covariances are given by $c_0 = d$ and $c_1 = b$.

Similarly

$$\Phi^{-1}(z) = \rho^{-1}\frac{(z+a)(z^{-1}+a)}{(z+p)(z^{-1}+p)}. \tag{8}$$

The positive real part of Φ^{-1} is given by

$$\left(\Phi^{-1}\right)_+(z) = \frac{\delta}{2} + \frac{\beta}{z+p}, \tag{9}$$

where

$$\delta = \rho^{-1}\frac{1+a^2-2pa}{1-p^2}, \quad \beta = \rho^{-1}\frac{(a-p)(1-ap)}{1-p^2} \tag{10}$$

Then the IC are given by $c_0^i = \delta$ and $c_1^i = \beta$, and the IAC $r_1^i = \beta/\delta$.

The PSD Φ in (5) has three parameters, assume that we want to determine Φ so that $c_0 = 1$, and c_1 and r_1^i are matching some given numbers with absolute value less than one. In Figure 1, ρ has been eliminated and a and p are varied between -1 and 1 and the contour plot for c_1 and r_1^i are depicted. It can be seen that each red curve does not intersect all blue curves, hence it is not always possible to find a match to the given parameters. From (7) and (10),

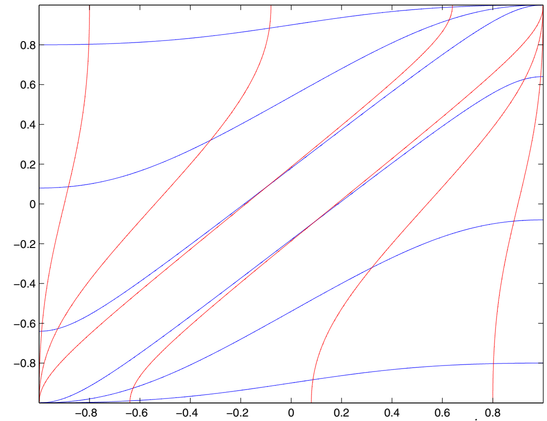

Fig. 1. Level curves for c_1 in "red", and level curves for r_1^i in "blue".

$$\frac{r_1}{r_1^i} = -\frac{(a-p)^2+1-p^2}{(a-p)^2+1-a^2} \leq 0, \tag{11}$$

hence r_1 and r_1^i have different signs. This was previously known for MA and AR processes [4], but does not hold in general.

For general spectral densities, there are similar restrictions. A consequence of the Cauchy-Schwarz inequality

$$1 = \left(\int \sqrt{\Phi}\sqrt{\Phi}^{-1} d\theta\right)^2 \leq \int \Phi d\theta \int \Phi^{-1} d\theta \tag{12}$$

is that the relation $c_0 c_0^i \geq 1$ always hold, and this also limits which covariance data that can be interpolated.

III. OPTIMIZATION APPROACHES TO THE INTERPOLATION PROBLEM

In this section we consider simultaneous interpolation of covariances and inverse-auto-correlations using an optimization approach.

A. *Exact simultaneous interpolation*

First assume that the C $c_0, c_1, \cdots, c_n$ and IAC $r_1^i, \cdots, r_m^i$ are given and that an exactly interpolating PSD Φ should be determined. This can be posed as an optimization problem where we choose to minimize the inverse variance, c_0^i, of the PSD Φ. The harmonic mean of the PSD is given by

$$\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi(e^{i\theta})^{-1} d\theta\right]^{-1},$$

and describes the minimal variance of the smoothing error [7, Sec.IV]. Maximizing the minimal variance is then equivalent to minimizing the function

$$f(\Phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi^{-1} d\theta \tag{13}$$

under the constraints

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi e^{ik\theta} d\theta,$$

for $k = 0, 1, \cdots, n$, and

$$r_k^i \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi^{-1} d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi^{-1} e^{ik\theta} d\theta,$$

for $k = 1, \cdots, m$. The objective function is convex and bounded from below by (12). The constraints on C are convex, but the ones on IAC are not. Lagrangian relaxation and duality theory leads to a spectral density of the form

$$\Phi = \sqrt{\frac{\beta_q - Q}{P}}, \tag{14}$$

where P and Q are pseudo-polynomials of the form

$$P(z) = \sum_{k=0}^{n} p_k \frac{z^k + z^{-k}}{2}, \quad Q(z) = \sum_{k=1}^{m} q_k \frac{z^k + z^{-k}}{2},$$

and $\beta_q = 1 + \sum_{k=1}^{m} q_k r_k^i$. This generalizes the result in [7] by adding zeros estimated from the IAC. Note that c_0^i is used for minimization in the objective function and therefore is not matched in the constraints. However, we can not guarrantee that a solution exists and therefore it may be necessary to consider approximative interpolation.

B. Approximative simultaneous interpolation

Given the same moments as before, the hard interpolation constraints are now replaced by quadratic penalties on the deviations from the nominal values of the moments. Let the weights $\alpha_k, \beta_k > 0$ be given, and add the weighted quadratic penalties to the previous objective function f in (13). Minimizing the function

$$f_{\approx}(\Phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi^{-1} d\theta + \sum_{k=0}^{n} \alpha_k \varepsilon_k^2 + \sum_{k=1}^{m} \beta_k \Delta_k^2 \tag{15}$$

under the constraints

$$\varepsilon_k - c_k + \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi e^{ik\theta} d\theta = 0,$$

for $k = 0, \cdots, n$, and

$$\Delta_k - r_k^i \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi^{-1} d\theta + \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi^{-1} e^{ik\theta} d\theta = 0,$$

for $k = 1, \cdots, m$, leads to a spectral density of the form as in (14) with interpolation errors $\varepsilon_k = -p_k/(2\alpha_k)$ and $\Delta_k = q_k/(2\beta_k)$. The derivation is similar to the one in [3].

C. Minimizing the distance between two interpolants

The previous optimization approaches led to spectral densities of a square root form. Alternative optimization problems are now considered to search for other power spectrum structures.

Consider now a simultaneous matching of two PSDs to the known data. Match the PSD Φ against the C $c_0, c_1, \cdots, c_n$ and the PSD Ψ against the IC $c_0^i, c_1^i, \cdots, c_m^i$, and then minimize the distance from Φ to Ψ. That is, minimize the function

$$f(\Phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} D(\Phi, \Psi) d\theta$$

under the constraints

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi e^{ik\theta} d\theta, k = 0, 1, \cdots, n$$

$$c_k^i = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Psi^{-1} e^{ik\theta} d\theta, k = 0, 1, \cdots, m.$$

This way two PSDs are estimated, Φ and Ψ. The distance between the optimal $\hat{\Phi}$ and $\hat{\Psi}$ is a measure of the discrepancy of the C and the IC ability to jointly describe the process.

Forming the Lagrangian function

$$\begin{aligned}
\mathcal{L} &= \int D(\Phi, \Psi) d\theta + \sum_{k=0}^{n} p_k \left(c_k - \int \Phi e^{ik\theta} d\theta \right) \\
&\quad + \sum_{k=0}^{n} q_k \left(c_k^i - \int \Psi^{-1} e^{ik\theta} d\theta \right) \qquad (16) \\
&= \int D(\Phi, \Psi) - P\Phi - Q_0 \Psi^{-1} d\theta \\
&\quad + \sum_{k=0}^{n} p_k c_k + \sum_{k=0}^{m} q_k c_k^i, \qquad (17)
\end{aligned}$$

where $Q_0(z) = \sum_{k=0}^{m} q_k \frac{z^k + z^{-k}}{2}$, and finding the first order optimality conditions reveals the structure of the optimal PSD estimates.

Next we consider a number of choices of D that are popular in interpolation theory.

1) Maximum entropy: Letting $D(\Phi, \Psi) = -\log \Phi/\Psi = -\log \Phi + \log \Psi$, then the problem is the maximum entropy problem for Φ and Ψ separately in the C and the IC, so

$$\Phi = \frac{1}{Q_0}, \quad \Psi = \frac{1}{P}.$$

Each of the two PSD can be determined by the Yule-Walker equations. However, D is not a distance and Φ and Ψ are not drawn to each other.

2) L_2 - distance: Letting $D(\Phi, \Psi) = (\Psi - \Phi)^2$, then

$$\Phi = P + \sqrt{\frac{Q_0}{P}}, \quad \Psi = \sqrt{\frac{Q_0}{P}}.$$

3) $1/\Psi$ - Weighted L_2 - distance: Letting $D(\Phi, \Psi) = (\Psi - \Phi)^2/\Psi$, then

$$\Phi = (P + 2)\sqrt{\frac{Q_0}{P(P + 4)}}, \quad \Psi = 2\sqrt{\frac{Q_0}{P(P + 4)}}.$$

4) Ψ - Weighted L_2 - distance: Letting $D(\Phi, \Psi) = (\Psi - \Phi)^2 \Psi$, then

$$\Phi = \frac{1}{2}\frac{3P^2 + 4Q_0}{\sqrt{P(P^2 + 4Q_0)}}, \quad \Psi = \frac{1}{2}\sqrt{\frac{P^2 + 4Q_0}{P}}.$$

5) Kullback-Leibler distance: Letting $D(\Phi,\Psi) = \Phi\log\frac{\Phi}{\Psi} + \Psi - \Phi$, then

$$\Phi = \sqrt{\frac{Q_0}{e^P-1}}e^P, \quad \Psi = \sqrt{\frac{Q_0}{e^P-1}}.$$

6) Itakura-Saito distance: Letting $D(\Phi,\Psi) = \frac{\Phi}{\Psi} - \log\frac{\Phi}{\Psi} - 1$, then

$$\Phi = \frac{1}{2}Q_0\left(1+\sqrt{1+4/(Q_0P)}\right),$$
$$\Psi = \frac{1}{2}Q_0\left(-1+\sqrt{1+4/(Q_0P)}\right).$$

7) Hellinger distance: Letting $D(\Phi,\Psi) = \left(\sqrt{\Phi}-\sqrt{\Psi}\right)^2$, then

$$\Phi = \sqrt{\frac{Q_0}{P}}\frac{1}{(1-P)^{3/2}}, \quad \Psi = \sqrt{\frac{Q_0(1-P)}{P}}.$$

8) α-divergence: Letting $D(\Phi,\Psi) = \Phi^\alpha\Psi^{1-\alpha} - \alpha\Phi - (1-\alpha)\Psi$, then

$$\Phi = \frac{\sqrt{Q_0}}{\sqrt{1-\alpha}}\frac{\tilde{P}^{1/(1-\alpha)}}{\sqrt{\tilde{P}^{\alpha/(\alpha-1)}-1}}$$
$$\Psi = \sqrt{1-\alpha}\sqrt{\frac{\tilde{P}^{\alpha/(\alpha-1)}-1}{Q_0}},$$

where $\tilde{P} = 1 - P/\alpha$.

Using for example the spectral distortion $D(\Phi,\Psi) = (\log\Phi - \log\Psi)^2$, there exists no closed form expression for the optimal form of the interpolant.

Note that the pseudo-polynomials P and Q_0 have to be determined simultaneously from the C and the IC data, and together determine the two estimates Φ and Ψ.

Note also that if $D(\Phi,\Psi) = D(\Psi^{-1},\Phi^{-1})$, as for the Itakura-Saito distance, then the structure of Φ and Ψ will be the same.

IV. CONCLUSIONS

The initial aim was to determine a power spectrum corresponding to an ARMA model, that interpolates the covariances and inverse-covariances. It has been shown that ARMA models can not always match the given covariance (C and IAC) data. Directly minimizing the quadratic interpolation deviations

$$\sum_{k=0}^{n}\left(c_k - \int\frac{P}{Q}e^{ik\theta}\right)^2 + \sum_{k=1}^{m}\left(r_k^i\int\frac{Q}{P} - \int\frac{Q}{P}e^{ik\theta}\right)^2$$

in the parameters of a PSD on the form $\Phi = P/Q$ leads to a difficult non-convex optimization problem. Here, a global approach has been applied using a general PSD and duality theory to derive the optimal shape of the PSD. If there exists a solution of the special form then it can be determined by solving a dual convex optimization problem. However, these approaches have not resulted in PSDs of rational form.

The natural form of the interpolant of the C and IAC parameters seems to be the square root of some rational function. This may depend on that locally most measures are similar when the two PSDs are close enough. The presence of the square root makes it more difficult to solve the optimization problems, and depending on the application, it may be computationally difficult to work with the determined PSD. What is interesting is that it seems to be the natural extension of the maximally random PSD determined in [7]. It determines the least-variance smoothing of the signal given the covariances $c_0, c_1, \cdots, c_n$.

We have only analyzed the existence question for ARMA(1,1) systems here. For the methods in Section III we do not expect that there is always a solution. Given C $c_0, c_1, \cdots, c_n$ and IAC $r_0^i, r_1^i, \cdots, r_m^i$, will there exist exact interpolants? If the dual problem, to the primal problem in Section III-A, has interior point solutions, then we know that there exist an exact interpolant of the form (14). Even if there exists no interpolants of the form (14), there may still exist interpolants of other forms but then the minimization problem will not have an optimal solution. The objective function (13) may approach some value (> 0) but never attain it.

REFERENCES

[1] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, “A new approach to Spectral Estimation: A tunable high-resolution spectral estimator,” *IEEE Trans. Signal Processing,* vol. 48, pp. 3189-3205, Nov. 2000.
[2] C.I. Byrnes, S.V. Gusev and A. Lindquist, “ From finite covariance windows to modeling filters: A convex optimization approach”, *SIAM Review* , 43 (Dec. 2001), 645-675.
[3] P. Enqvist, E. Avventi, “Approximative covariance interpolation with a quadratic penalty ,” *Proceedings of the IEEE Intern. Conf. on Decision and Control*, pp. 4275-4280, December 2007.
[4] C. Chatfield, ”Inverse Autocorrelations”, *Journal of Royal Statistical Society*, Vol. 142, No. 3, 1979, pp 363-377.
[5] W.S. Cleveland, ”The Inverse Autocorrelations of Time Series and their applicatons”, *Technometrics*, Vol. 14, No. 2, May 1972, pp 277-293.
[6] P. Enqvist, “Approximative Covariance Interpolation”, preprint arXiv:1104.1880.
[7] T.T. Georgiou, “The maximum entropy principle in the absence of a time-arrow: Fractional-pole models,” *IEEE Trans. on Information Theory*, vol. 53, no. 8, pp. 2841–2851, August 2007.
[8] A. Ferrante, M. Pavon, and F. Ramponi, “Hellinger vs. Kullback-Leibler multivariable spectrum approximation,” *IEEE Trans. Autom. Control,* bf 53(5): 954-967, May 2008.
[9] U. Grenander and G. Szegö, **Toeplitz Forms and their Applications**, Chelsea, 1958.
[10] S. Haykin, **Nonlinear Methods of Spectral Analysis,** Springer-Verlag, New York, 247 pages, 1979.
[11] P. Stoica and R. Moses, **Introduction to Spectral Analysis**, Prentice Hall, 1997.