Approximative Covariance interpolation

Per Enqvist

Abstract— **When methods of moments are used for identification of power spectral densities, a model is matched to estimated second order statistics such as,** *e.g.*, **covariance estimates. If the estimates are good there is an infinite family of power spectra consistent with such an estimate and in applications, such as identification, we want to single out the most representative spectrum. We choose a prior spectral density to represent** *a priori* **information, and the spectrum closest to it in a given quasi-distance is determined. However, if the estimates are based on few data, or the model class considered is not consistent with the process considered, it may be necessary to use an approximative covariance interpolation. Two different types of regularizations are considered in this paper that can be applied on many covariance interpolation based estimation methods.**

I. INTRODUCTION

Most system identification methods are based on an algorithm that is proven to give efficients estimates when the number of data goes to infinity. One such common estimate is the maximum likelihood method. However, in many cases only a small amount of data is available and the estimation method may give unexpected results. Here we will consider methods based on covariance interpolation instead. Depending on which model class is considered there are a number of different methods around now for matching AR, MA, ARMA and other models to covariances, such as the ones derived by Lindquist, Byrnes, Georgiou, Pavon, Ferrante, *et. al.* based on minimizing the Kullback-Leibler [1], Hellinger [2], the Itakura-Saito quasi-distance [3], [4], [5], and other distance concepts. However, also these methods depends on the amount of data that is available and also structural constraints. The covariances have to be estimated from the data and the errors in the estimates will increase the smaller the available data set is. Estimating the covariances from a short data sequence may generate a covariance matrix that is not non-negative definite, or does not have a supposed Toeplitz structure or the estimate does not correspond to a spectra in the supposed model class. So for short data sequences it is necessary to regularize the methods to obtain relevant model estimates. In this paper we compare two different approaches for dealing with these kinds of problems; the two different kinds of regularizations are based on quadratic penalties on the covariance estimation errors and extra entropy regularization of the determined spectrum. These approaches have been used before for the maximum entropy method for AR-models, the Kullback-Leibler method for the ARMA case with fixed MA-part and

a combined covariance and cepstrum interpolation problem, but here they will be used and compared in a more general setting.

The first kind of problem, with non-negative definite covariance matrices, is often "solved" by using a biased estimate of the covariance matrix. This bias is usually small and goes to zero as the number of data grows, but for small data sets it can be relevant. Another approach is to use a regularization of the first kind mentioned above, i.e., to find a spectrum within the model class which has a small quadratic distance to the estimated covariance matrix. By combining the covariance interpolation methods based on entropy maximization with a quadratic distance penalty the structure of the spectrum is taken into account when the best covariance sequence close to the estimates is determined.

The second kind of problem, with the estimate of the covariance matrix not having the supposed structure, is often solved using a projection onto the class of matrices with the desired structure. This problem is most obvious when a state-covariance interpolation approach is used; Then there is an imposed structure determined by the (A, b) matrices in the state-covariance definition. Again, another approach is to use the regularization of the first kind mentioned above. A small distance to a matrix with the desired structure is then obtained.

The third kind of problem, with a covariance estimate that can not be interpolated by a spectrum in the model class (but has the desired structure and is non-negative) as in MA-model covariance interpolation for some covariance estimates. Probably the most common approach to resolve this problem is to project the covariance estimates onto the set of covariances feasible for the desired model class. For the MA case, this would be the projection onto a positive cone, but to avoid having zeros on the unit circle a projection to a slightly smaller cone should be performed. Another approach is to use a regularization of either the first or second kind mentioned above. The amount of quadratic penalty regularization for the first method has to be determined recursively, and might fail for some cases as will be shown by some examples. The extra entropy regularization treats this case in an easier way and finds both the best approximating valid covariance and the interpolant with one optimization problem.

If we want to determine a MA-model estimate for a state-covariance estimated from a short data sequence, all of the three kinds of problems described above may occur. Then it would be necessary to use a combination of the two types of regularizations.

Here, a general approach to the covariance matching

problem is taken that holds for a large set of different quasi-distances and is inspired by the work in [6].

II. BACKGROUND

Let $(\ldots, y_{-1}, y_0, y_1, \ldots)$ be a scalar stationary stochastic real valued mean-zero process with covariances $r_k = \mathrm{E}\{y_{\ell+k}y_\ell\}$ and PSD Φ. The power spectral density Φ represents the energy content of the process across frequencies and has the covariances as Fourier coefficients,

$$\Phi(e^{i\theta}) \triangleq \sum_{k=-\infty}^{\infty} r_k e^{ik\theta}.$$

Consider the Hilbert space $L_2(-\pi, \pi]$ with the inner product

$$\langle a, b \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} a(e^{i\theta})b(e^{-i\theta})d\theta.$$

Then the covariances are given by $r_k = \langle \Phi, z^k \rangle$. Given a finite window of covariances $\mathbf{r} = \begin{pmatrix} r_0 & r_1 & \ldots & r_n \end{pmatrix}$, let $\mathfrak{F}_\mathbf{r}$ denote the set of PSD consistent with $\mathbf{r}$, i.e.,

$$\mathfrak{F}_\mathbf{r} = \left\{ \Phi \geq 0 \,|\, \langle \Phi, z^k \rangle = r_k, \quad k = 0, 1, \cdots, n \right\}.$$

In this paper, $\Phi \geq 0$ means that this inequality should hold on the unit circle, i.e., $\Phi(e^{i\theta}) \geq 0$ for $\theta \in (-\pi, \pi]$.

Furthermore, we assume initially that the symmetric Toeplitz matrix of the covariances $\mathbf{r}$,

$$T(\mathbf{r}) = \begin{bmatrix} r_0 & r_1 & \cdots & r_n \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_n & \cdots & r_1 & r_0 \end{bmatrix} \tag{1}$$

is positive definite, hence the set $\mathfrak{F}_\mathbf{r}$ contains an infinite number of PSDs [7, Sec.6.5]. Let $\mathcal{R} = \{\mathbf{r} \,|\, T(\mathbf{r}) > 0\}$.

III. MOMENT MATCHING

In many situations it is desired to fit a spectral density to data by finding one of a particular structure by matching moments. The most common PSD used to model stationary stochastic processes are the ones that correspond to Moving-Average (MA) and Auto-Regressive (AR) processes. Assume that $Q(z)$ is a pseudo-polynomial of degree n, i.e.,

$$Q(z) = q_0 + \frac{1}{2}q_1(z + z^{-1}) + \cdots + \frac{1}{2}q_n(z^n + z^{-n}). \tag{2}$$

Then, $\Phi = Q$ is the PSD of a MA-process and $\Phi = 1/Q$ is the PSD of an AR-process. It is well known that for an AR-process the coefficients $\{q_k\}_{k=0}^n$ of Q can always be tuned so that a window of covariances $\mathbf{r} \in \mathcal{R}$ is matched. On the other hand, it is also well known that for an MA-process there are some $\mathbf{r} \in \mathcal{R}$ (actually open subsets of such covariances) that are not matched for any choice of coefficients $\{q_k\}_{k=0}^n$. In both cases there are $n+1$ parameters that should be tuned to match $n+1$ constraints, but it is clearly the structure of the PSD that determines if solutions exists or not.

To evaluate the properties of different PSD structures, we let Φ depend on Q, and it will also be allowed to depend on some "prior estimate" PSD Ψ. Assuming now that $\Phi =$

$F(Q, \Psi)$, the moment matching constraint $\Phi \in \mathfrak{F}_\mathbf{r}$, can be expressed as

$$\langle F(Q, \Psi), z^k \rangle = r_k = \langle R, z^k \rangle, \quad k = 0, 1, \cdots, n, \tag{3}$$

where R is an arbitrary function in $\mathfrak{F}_\mathbf{r}$.

IV. EXACT AND APPROXIMATIVE INTERPOLATION

A. Exact interpolation

The distance measure will be assumed to be differentiable in the first argument, and it will be assumed to be a quasi-distance, i.e., it is assumed that $D(\Phi\|\Psi) \geq 0$ and $D(\Psi\|\Psi) = 0$ for any pair of PSD Φ and Ψ. Furthermore, we assume that

$$D(\Phi\|\Psi) = \int d(\Phi\|\Psi).$$

Note that D is not assumed to be symmetric, convex, to satisfy the triangle inequality or be zero if and only if $\Phi = \Psi$. However, these are certainly desired properties. Consider the optimization problem, to minimize the distance to Ψ for all $\Phi \in \mathfrak{F}_\mathbf{r}$, i.e.:

$$(\mathcal{P}_=) \quad \begin{bmatrix} \inf_{\Phi \geq 0} & D(\Phi\|\Psi) \\ \text{s.t.} & \langle \Phi, z^k \rangle - r_k = 0, \quad k = 0, 1, \cdots, n. \end{bmatrix} \tag{4}$$

Note that here that the PSD Φ is not constrained to be of a certain form, this form will be determined by the optimality conditions of the Lagrange relaxed functional, which in turn is determined by the geometry imposed by the distance measure.

The optimization problem $(\mathcal{P}_=)$ has no finite dimensional parametrization, but by considering the dual, an optimization problem with a finite number of variables is obtained. To this end, formal calculations are performed to determine the dual.

Form the Lagrangian function

$$L_0(\Phi; \mathbf{q}) \triangleq D(\Phi\|\Psi) + \sum_{k=0}^{n} q_k \left(r_k - \langle \Phi, z^k \rangle \right)$$

and since Φ is symmetric $\langle \Phi, \sum_{k=0}^{n} q_k z_k \rangle = \langle \Phi, Q(z) \rangle$, where Q is defined in (2). Let $R \in \mathfrak{F}_\mathbf{r}$ arbitrary. Then the Lagrangian function can be written as

$$L_0(\Phi; Q) = D(\Phi\|\Psi) + \langle R, Q \rangle - \langle \Phi, Q \rangle.$$

Assuming that a minimizer exists let

$$\hat{\Phi} := \arg \min_{\Phi \geq 0} L_0(\Phi, Q), \tag{5}$$

this defines the optimal PSD as a function of Q, i.e.,

$$\hat{\Phi} = F(Q; \Psi) \tag{6}$$

and determines the dual objective function

$$\Omega_0(Q; \Psi) \triangleq L_0(\hat{\Phi}, Q) = L_0(F(Q; \Psi), Q). \tag{7}$$

To ensure that the spectral densities $F(Q; \Psi)$ are non-negative the domain $\mathcal{Q}$ of feasible Q has to be specified, i.e.,

$$\mathcal{Q} = \{Q \,|\, F(Q; \Psi) \geq 0\}.$$

This leads to the dual problem to determine the maximizer of Ω_0 over all $Q \in \mathcal{Q}$, i.e.,

$$(\mathcal{D}_=) \quad \begin{bmatrix} \sup_Q & \Omega_0(Q; \Psi, R) \\ \text{s.t.} & F(Q; \Psi) \geq 0. \end{bmatrix}, \tag{8}$$

The derivative of Ω is (compare the proof of Proposition 4.1 in [6])

$$\frac{\partial \Omega}{\partial Q} = \int \left(\frac{\partial}{\partial Q} d(F \| \Psi) - Q \right) \frac{\partial F}{\partial Q} + \int (R - F)$$

and using that $F(Q, \Psi)$ minimizes L_0 it can be shown that the first integral is zero. The stationarity conditions for $(\mathcal{D}_=)$ are then

$$r_k - \langle F(Q; \Psi), z^k \rangle = 0,$$

for $k = 0, 1, \cdots, n$, which ensures that for an interior point solution the optimal $\Phi \in \mathfrak{F}_{\mathbf{r}}$.

B. Primal regularization

Consider now the approximative interpolation problem:

$$(\mathcal{P}_{\approx}^2) \quad \begin{bmatrix} \inf_{\Phi \geq 0} & D(\Phi \| \Psi) + \alpha \| \Delta \|^2 \\ \text{s.t.} & \langle \Phi, z^k \rangle - r_k = \Delta_k, \quad k = 0, 1, \cdots, n. \end{bmatrix} \tag{9}$$

Once again, the spectral density is not constrained to be of a certain form, this form will be determined by the optimality conditions of the Lagrange relaxed functional, which in turn is determined by the geometry imposed by the distance measure.

We show that the structure of the optimal Φ will be the same as for $(\mathcal{P}_=)$. Form the Lagrangian function

$$L(\Phi, \Delta; \mathbf{q}) \triangleq D(\Phi \| \Psi) + \alpha \| \Delta \|^2 + \sum_{k=0}^{n} q_k \left(\Delta_k + r_k - \langle \Phi, z^k \rangle \right)$$

$$= L_0(\Phi; \mathbf{q}) + \alpha \| \Delta \|^2 + \sum_{k=0}^{n} q_k \Delta_k$$

and since Φ is symmetric $\langle \Phi, \sum_{k=0}^{n} q_k z_k \rangle = \langle \Phi, Q(z) \rangle$, where Q is defined in (2).

The optimal $\hat{\Delta} = -\frac{1}{2\alpha} q$. So if α goes to infinity the approximation errors go to zero (if an exact solution exists).

The optimal PSD Φ is again determined by (5), hence the structure of Φ is preserved and $\hat{\Phi} = F(Q; \Psi)$, see (6).

The dual objective function is then given by

$$\Omega(Q; \Psi) = L(\hat{\Phi}, \hat{\Delta}, Q) = \Omega_0(Q; \Psi) - \frac{1}{4\alpha} \| Q \|^2. \tag{10}$$

This leads to a dual problem on the form

$$(\mathcal{D}_{\approx}^2) \quad \begin{bmatrix} \sup_Q & \Omega(Q; \Psi) \\ \text{s.t.} & F(Q; \Psi) \geq 0. \end{bmatrix}. \tag{11}$$

The stationarity conditions for $(\mathcal{D}_{\approx}^2)$ are

$$r_k - \langle F(Q; \Psi), z^k \rangle = \left\langle \frac{Q}{2\alpha}, z^k \right\rangle,$$

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

C. Dual regularization

Consider the dual regularized optimization problem:

$$(\mathcal{D}_{\approx}^1) \quad \begin{bmatrix} \sup_Q & \Omega(Q; \Psi, R) - \langle \lambda g(\Psi), \log Q \rangle, \\ \text{s.t.} & F(Q; \Psi) \geq 0. \end{bmatrix}, \tag{12}$$

where $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a function of the prior PSD Ψ, typically a constant or the identity operator.

The stationarity conditions are

$$r_k - \langle F(Q; \Psi), z^k \rangle = \left\langle \frac{\lambda g(\Psi)}{Q}, z^k \right\rangle, \quad k = 0, 1, \cdots, n.$$

The right hand side will be small for small λ. If Q is close to zero for some frequencies, the integral will still be bounded but have a derivative that goes to infinity as Q goes to zero.

D. Comparison of the two regularizations

We note that both the regularizations results in adding a concave function of Q to the dual objective function. In $(\mathcal{D}_{\approx}^1)$ it is a logaritmic term that works as a barrier function making sure that the optimum is in an interior point of $\mathcal{Q}$. If the optimum of the primal problem $(\mathcal{P}_=)$ is in an interior point, the regularization term is rather small and does not affect the solution much but tends to pull it slightly towards a spectrum with PSD $g(\Psi)$. If the optimum of the primal problem $(\mathcal{P}_=)$ is on the boundary, the unbounded derivative of the regularization term will push the solution towards the interior.

In $(\mathcal{D}_{\approx}^2)$ the regularization term is a quadratic function of the matching error. By allowing a slack in the covariance matching constraint the distance $D(\Phi \| \Psi)$ can be made smaller and a PSD closer to the prior is obtained. This means that more trust is put on the prior information and less is put on the covariances, which makes sense if the covariances are estimated from short data sequences. For the Kullback-Leibler distance it is shown in [8] that even if the covariances are not in $\mathcal{R}$, i.e. , corresponds to a positive definite Toeplitz matrix, an approximative solution is obtained if the α is chosen small enough. Note that the covariances can fail to correspond to a positive definite matrix and they can also fail to form a Toeplitz matrix, but an approximation is anyway guarranteed. But this does not hold for any choice of quasi-distance, as demonstrated by the following example.

E. Example

Consider now the approximative interpolation problem $(\mathcal{P}_{\approx}^2)$ for the special case that $d(\Phi \| \Psi) = \frac{1}{2} \frac{(\Phi - \Psi)^2}{\Psi}$:

Form the Lagrangian function

$$L(\Phi, \Delta; \mathbf{q}) \triangleq \frac{1}{2} \frac{(\Phi - \Psi)^2}{\Psi} + \alpha \| \Delta \|^2$$

$$+ \sum_{k=0}^{n} q_k \left(\Delta_k + r_k \right) - \langle \Phi, Q \rangle$$

The optimal $\hat{\Delta} = -\frac{1}{2\alpha}q$. The optimal PSD Φ is determined by

$$\int \left(\log \frac{\Phi}{\Psi} + 1 - 1 - Q \right) \delta\Phi \, d\theta = 0,$$

for all $\delta\Phi$, *i.e.* $\Phi = \Psi(Q+1)$. The dual objective function is then given by

$$\Omega(Q; \Psi) = -\frac{1}{2} \left\langle \Psi + \frac{1}{4\alpha}, (Q+1)^2 \right\rangle + \sum_{k=0}^{n} q_k r_k + \text{const.}$$

(13)

Therefore, the regularization term and α only changes the prior and no matter how small α is chosen it is not always possible to find an interior point solution satisfying the stationarity conditions

$$r_k - \left\langle \Psi(Q+1), z^k \right\rangle = \left\langle \frac{Q}{2\alpha}, z^k \right\rangle,$$

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

REFERENCES

[1] T.T. Georgiou and A. Lindquist, "Kullback-Leibler approximation of spectral density functions," *IEEE Transactions on Information Theory*, vol. 49, pp. 2910–2917, Nov 2003.

[2] A. Ferrante, M. Pavon, and F. Ramponi, "Hellinger versus Kullback-Leibler multivariable spectrum approximation," *IEEE Trans. Automatic Control*, vol. 53, no. 4, pp. 954–967, May 2008.

[3] J. Shore, "Minimum cross-entropy spectral analysis," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 230–237, Apr 1981.

[4] J. Shore and R. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Information Theory*, vol. 27, no. 4, pp. 472–482, Jul 1981.

[5] P. Enqvist and J.Karlsson, "Minimal Itakura-Saito distance and covariance interpolation," 2008, Conference on Decision and Control.

[6] P. Enqvist, "Covariance interpolation and geometry of power spectral densities," in *Proceeding ECC 2009*, 2009.

[7] B. Porat, *Digital Processing of Random Signals, Theory & Methods*, Prentice Hall, 1994.

[8] P. Enqvist and E. Avvent, "Approximative linear and logarithmic interpolation of spectra," Tech. Rep. TRITA-MAT 09 OS 02, KTH Mathematics, 2009, ISSN 1401-2294.