IDENTIFICATION OF RATIONAL SPECTRAL DENSITIES USING ORTHONORMAL BASIS FUNCTIONS

A. Blomqvist [*,1] **G. Fanizza** [*,1,2]

[*] *Division of Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden*

Abstract: This paper gives an algorithm for identifying spectral densities using orthonormal basis functions. Mathematically, this amounts to identifying a time-invariant linear SISO system with the additional constraint that the transfer function should be positive-real. Thus, we solve the long-standing problem of how to incorporate this positivity constraint while using orthonormal basis functions. The procedure is a variant of the THREE algorithm introduced by Byrnes, Georgiou and Lindquist. The relation between and numerical properties of the proposed and the THREE algorithms are discussed. The orthonormal basis functions are better scaled for a concentrated pole selection in the basis, which increases the accuracy of the estimates. A numerical example which highlights this phenomenon and illustrates the algorithm is given.

Keywords: System identification, Spectral estimation, Filter banks, Spectral density function, Linear filters

1. INTRODUCTION

Estimating rational spectral densities is equivalent to identifying time-invariant linear spectral factors under the additional constraint that spectral density should be real and positive on the unit circle. Then the spectral density can be written as the real part of a positive-real function. Therefore, the identification problem is an instance of identification of *passive* systems (see Caines (1988)). The nontrivial positivity constraint has (so far) not been possible to include in the identification procedures of Ho-Kalman type.

Different types of orthonormal basis functions for identification of general time-invariant linear systems have been thoroughly studied. They provide the means to incorporate a priori information of the system dynamics in the basis. But they also have advantages in terms of computational complexity and numerical conditioning. de Hoog et al. (2002) considers basis functions similar to the ones here for identifying general time-invariant linear systems; also see Ninness and Gustafsson (1997). These also give an introduction to orthonormal basis functions and their use in system identification.

Here the problem is treated from a quite different starting point leading up to an analytic interpolation problem. The algorithm uses the framework of Georgiou (2001) and is therefore closely related to the THREE algorithm proposed in Byrnes et al. (2000). The mathematical foundation is given in Georgiou and Lindquist (2002). The main contribution is the demonstration of how to identify passive systems using orthonormal basis functions. Furthermore, the paper contains a comparison to THREE basis functions, in particular for concentrated pole sets.

First the algorithm is described, then a comparison to the THREE algorithm is made and finally a numerical example is given. The notation is standard.

[1] This research was supported by the Swedish Research Council (VR)
[2] This research was supported by TMR

2. IDENTIFICATION ALGORITHM

The proposed identification procedure has four components: an orthonormal basis, a filter bank, an analytic interpolation problem and a homotopy continuation method. In the following each component is discussed and it is shown how to bring the components together.

2.1 *The Generalized Orthonormal Basis*

The orthonormal basis considered in this paper is generated from a finite set of points in the open unit disc. These points will be the poles of the basis functions. The basis is generalized in the sense that Laguerre, Kautz and other bases are special cases.

Let $\{\xi_k\}_{k=1}^n \subset \mathbb{D}$ be given. Then define the functions

$$G_k(z) := \frac{\sqrt{1-|\xi_k|^2}}{z-\xi_k} \prod_{j=1}^{k-1} \left(\frac{1-\xi_j^* z}{z-\xi_j} \right) \ \forall k. \quad (1)$$

These basis functions date back to early work in the 1920's by Takenaka and Malmquist. They are constructed by all-pass factors with the balanced state-space realization:

$$\frac{1-\xi_k^* z}{z-\xi_k} \sim \left[\begin{array}{c|c} \xi_k & \sqrt{1-|\xi_k|^2} \\ \hline \sqrt{1-|\xi_k|^2} & -\xi_k \end{array} \right], \quad (2)$$

Due to the well-known recursive relationship between two such function $G_k(z)$ and $G_l(z)$:

$$G_k(z)G_l(z) \sim \left[\begin{array}{cc|c} A_k & 0 & B_k \\ B_l C_k & A_l & B_l D_k \\ \hline D_l C_k & C_l & D_l D_k \end{array} \right], \quad (3)$$

the finite product of all-pass functions allows a balanced, minimal state-space realization. This will prove to be very useful in the filter bank construction in the next section.

Remark 1. If the poles are complex-conjugated, i.e., $\bar{\xi_k}$ is in the pole set whenever $\xi_k \notin \mathbb{R}$ is, we can get a real-valued state-space realization using e.g. a 2-parameter Kautz model.

2.2 *The Filter Bank Construction*

A filter bank constituted of the basis functions will put the problem into the desired form. Define the $(n+1)\times 1$ vector-valued analytic function

$$G(z) := \begin{bmatrix} G_0(z) \\ G_1(z) \\ \vdots \\ G_n(z) \end{bmatrix}, \quad (4)$$

where $G_0(z) \equiv 1$. It can be viewed as a bank of filters as showed in Figure 1. Now, construct a minimal balanced state-space realization $\{A, B, C, D\}$ as in Section 2.1 for the last basis function, $G_n(z)$. Then, a minimal, i.e., $(n+1)$-dimensional, Input-to-State (IS) realization, $\{A, B\}$, for the filter bank is given by the same A and B. The pair (A, B) will be a controllable pair.

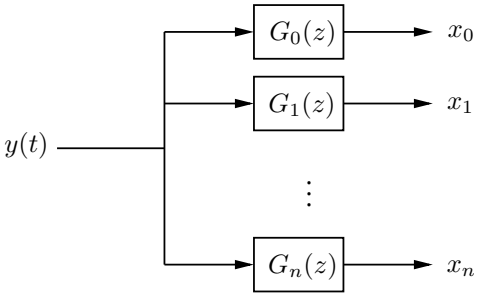

Fig. 1. The filter bank

The Linear Predictive Filter, a maximum entropy filter, is one example that falls under this framework. As pointed out in Georgiou (2001) the corresponding basis functions are $\{z^{-k}\}_{k=0}^n$. Thus, they can be interpreted as being generated as in (1) from a set with all poles at origin.

2.3 *The Analytic Interpolation Problem*

Given the IS realization of the filter bank, the identification problem can be stated as an analytic interpolation problem; see Georgiou (2001). The interpolant whose spectral density is closest to an a priori estimate $\Psi(z)$ in terms of the Kullback-Leibler distance is given by Georgiou and Lindquist (2002) and is shown to be the unique minimizer of a certain convex optimization problem.

The analytic interpolation problem is derived in Georgiou (2001). First, a characterization of the state-covariance matrices $P := \epsilon\{x^k x^{k*}\}$ (where ϵ denotes the expectation) for an IS filter is given. A state-covariance matrix should be positive definite and fulfill the algebraic condition

$$P = \frac{1}{2}(WE + EW^*), \quad (5)$$

where E is the reachability Gramian and thus the unique positive definite solution to the Lyaponov equation $E - AEA^* = BB^*$. For the orthonormal basis, E is the identity matrix, so P is a Toeplitz matrix. Furthermore, W and A commutes and W admits the representation

$$W = w(A) = w_0 I + w_1 A + \cdots + w_n A^n. \quad (6)$$

With this notation, the positive-real part of the spectral density, $f_y(z) : \Phi(z) = f_y(z) + f_y^*(z)$, should fulfill the interpolation condition

$$f_y(A^*) = W^*. \quad (7)$$

This leads to a generalization of the Nevanlinna-Pick interpolation problem. The state covariance matrix plays the role of a generalized Pick matrix.

Now assume that given data we estimate W consistent with (5) and want to minimize the so-called Kullback-Leibler distance to a given spectral density while meeting the interpolation condition (7). The solution is given by the following theorem of Georgiou and Lindquist (2002):

Theorem 2. Given $P \succ 0$ and (A, B) as earlier and a spectral density $\Psi(z)$, there is a unique $\Phi(z) = f_y(z) + f_y^*(z)$ that minimizes Kullback-Leibler distance

$$\mathbb{S}(\Psi\|\Phi) := \int \Psi \log \frac{\Psi}{\Phi} \tag{8}$$

subjected to $f_y(A^*) = W^*$. The minimizer takes the form

$$\hat{\Phi} = \frac{\Psi}{G^*\hat{\Lambda}G}$$

where $\hat{\Lambda}$ is the unique interior minimizer to the convex functional

$$\mathbb{J}_\Psi(\Lambda) := \text{trace}(\Lambda P) - \int \Psi \log(G^*\Lambda G). \tag{9}$$

Here and later the integration limits and variables are suppressed:

$$\int g := \frac{1}{2\pi}\int_{-\pi}^{\pi} g(e^{i\theta})d\theta.$$

If $\Psi(z) \equiv 1$, that is white noise with unit covariance, the solution is called the *maximum entropy solution*.

Remark 3. In de Hoog et al. (2002) a corresponding interpolation problem is also derived. That interpolation problem apparently look the same as the one of Theorem 2 but they are fundamentally different: here the interpolant is the positive-real part of the spectral density while it is the spectral factor itself in de Hoog et al. (2002).

2.4 *The Homotopy Continuation Method*

The optimization problem (9) in Theorem 2 may suffer from numerical problems, as pointed out in Georgiou and Lindquist (2002). In Blomqvist and Nagamune (2002) a numerical algorithm which seems to have better numerical properties is developed along the lines of Nagamune (2001).

The original, convex, functional $\mathbb{J}_\Psi(\Lambda)$ has the property that the gradient is infinite on the boundary of the feasible region. This causes numerical problems when the minimizer is close to the boundary. To avoid this, change variables to $\alpha(z) := C(z)\tau(z)$, where $C^*G^*(z)G(z)C = G^*(z)\Lambda G(z)$, and let $\boldsymbol{\alpha}$ be a vector with the n coefficients of $\alpha(z)$. The functional can then be written

$$J_\Psi(\alpha) = \boldsymbol{\alpha}^T K \boldsymbol{\alpha} - 2\int \Psi \log \alpha, \tag{10}$$

where

$$K := L_n^{-T}\Gamma^{-1}P\Gamma^{-T}L_n^{-1}, \tag{11}$$

with L_n nonsingular and given from the poles along with Γ as the controllability matrix of the IS realization of the filter bank. Thus, the problem is of the type in Nagamune (2001) and the same homotopy continuation method can be applied. The condition number of the matrix K will be of importance and will be discussed in Section 4.

3. ESTIMATING P AND W FROM DATA

The state covariance matrix from an IS filter as in Section 2.2 is necessarily positive semi-definite but it also fulfills the algebraic condition (5). However, experimental data typically does not. This section deals with the problem how to enforce the algebraic condition.

Given measurements $\{x^k\}_{k=1}^N$ (each x^k is a vector with components $x_0^k \dots x_n^k$) an estimate of the state covariance matrix P can be computed, e.g., as

$$\hat{P} = \frac{1}{N}\sum_{k=1}^{N} x^k x^{k*}.$$

For this estimate there is typically no solution to (5). Georgiou proposes a least-squares solution in terms of the coefficients $\{w_k\}_{k=0}^n$ in (6). However, this will not guarantee the modified P to be positive semi-definite.

To circumvent the problem of non-positive semi-definite state covariances, the least-squares program can be stated as a semi-definite program. This will guarantee the modified P to be positive at the cost of computational effort. More precisely, P is taken to be the solution of

$$(P) \quad \begin{array}{ll} \min & \|\hat{P} - P\|_{\text{Frob}} \\ \text{s.t.} & P = \dfrac{1}{2}(WE + EW) \succeq 0\,, \\ & AW = WA \end{array} \tag{12}$$

where E is as in (5). This can be rewritten as the semi-definite program

$$(SDP) \quad \begin{array}{ll} \min & t \\ \text{s.t.} & \|\operatorname{Vec}(\hat{P}) - \operatorname{Vec}(P)\|_2 \le t\,, \\ & -P \preceq 0 \end{array} \tag{13}$$

where feasible matrices are parameterized by (5) and (6). The program (SDP) is on standard form and there are several software packages available.

Remark 4. The program (SDP) is computationally considerably more expensive than a least-squares solution. Thus, an implementation should first compute the least-squares solution and check the positivity before proceeding to solve the (SDP).

Remark 5. This also allows the possibility, by requiring $P \succeq \varepsilon I$ for some ε, to avoid a high condition number in the optimization problem (10) caused by inverting a nearly singular matrix.

4. COMPARISON TO THE THREE ALGORITHM

The proposed algorithm resembles the THREE algorithm of Byrnes et al. (2000) and they will be compared in the subsequent. Firstly, the convergence rate of the expansion coefficients is discussed. Secondly, the numerical properties of the optimization problem (10) are analyzed. They turn out to be essentially the same, even though the orthonormal approach is internally better scaled. Finally, the accuracies of the state covariance estimates are considered. Generally, the orthonormal construction will be less noise-sensitive when the poles are concentrated, which is an interesting case when the system dynamics is concentrated to one frequency region. This will cause the main difference in the performance for the methods.

4.1 *Convergence rate of expansion coefficients*

A fast rate of convergence in terms of the expansion coefficients for the system to be identified has been a major reason for using orthonormal basis functions in identification, see for instance Wahlberg (1991). The same arguments can be used for the subclass of systems studied in this paper. If the poles of the basis functions are close to the dominant ones of the spectral density to be identified, it will require fewer coefficients to catch most of the dynamics. Thus this motivates both algorithms.

4.2 *Numerical properties of the optimization problem*

The THREE and orthonormal basis functions span the same space, so there is a non-singular coordinate transformation, T, between the basis functions:

$$G^{\text{THREE}}(z) = TG^{\text{Orth}}(z).$$

This gives a relation between the IS realizations

$$\begin{aligned} G^{\text{THREE}}(z) &= T(I - zA^{\text{Orth}})^{-1}B^{\text{Orth}}, \\ &= (I - zTA^{\text{Orth}}T^{-1})^{-1}TB^{\text{Orth}}, \\ &= (I - zA^{\text{THREE}})^{-1}B^{\text{THREE}}, \end{aligned}$$

which in turn gives relations between the controllability matrices and the true state covariances matrices respectively:

$$\Gamma^{\text{THREE}} = T\Gamma^{\text{Orth}}, \tag{14}$$

$$\begin{aligned} P^{\text{THREE}} &= \int G^{\text{THREE}}\Phi G^{\text{THREE}\ *}, \\ &= \int TG^{\text{Orth}}\Phi G^{\text{Orth}\ *}T^T, \\ &= TP^{\text{Orth}}T^T. \end{aligned} \tag{15}$$

In particular the accuracy of the estimates and condition number of the matrix K in (11) are important; for the maximum entropy solution the K matrix is actually inverted. For the true state covariance matrices, i.e., not noise-corrupt, (15) hold. Then K is invariant under coordinate changes:

$$\begin{aligned} K^{\text{THREE}} &= L_n^{-T}\Gamma^{\text{THREE}-1}P^{\text{THREE}}\Gamma^{\text{THREE}-T}L_n^{-1}, \\ &= L_n^{-T}\Gamma^{\text{Orth}-1}TT^{-1}P^{\text{Orth}}T^{-T}T^T\Gamma^{\text{Orth}-T}L_n^{-1}, \\ &= K^{\text{Orth}} := K. \end{aligned}$$

Therefore the matrices K, and thus the optimization problems, will converge when the data sequence goes to infinity.

Thus we can conclude that for reasonably long data sequences the numerical properties are essentially the same for the two problems. Even so, it is instructive to consider the internal conditioning of the K matrices. Firstly, consider the controllability matrix Γ. If the pole set is concentrated, i.e., if many poles are closely located, the condition number of Γ will be higher for the THREE filter bank. A two-pole example illustrates this:

Example 6. Given the poles $\{a, a+\varepsilon\}$ the IS realization of the filter banks for the THREE algorithm and the orthonormal basis functions are easily determined. The corresponding controllability matrices are

$$\Gamma^{\text{THREE}} = \begin{bmatrix} 1 & a \\ 1 & a+\varepsilon \end{bmatrix},$$

$$\Gamma^{\text{Orth}} = \begin{bmatrix} \tilde{b} & a\tilde{b} \\ -a\hat{b} & \hat{b}(1-|a|^2-a^2-\varepsilon a) \end{bmatrix},$$

respectively and where $\hat{b} = \sqrt{1-|a+\varepsilon|^2}$ and $\tilde{b} = \sqrt{1-|a|^2}$. If the absolute value of ε is small compared to a, the condition numbers under the Euclidean norm can be computed as

$$\begin{aligned} \kappa_2(\Gamma^{\text{THREE}}) &= \kappa_2\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & a \\ 0 & \varepsilon \end{bmatrix}\right), \\ &= \kappa_2\left(\begin{bmatrix} 1 & a \\ 0 & \varepsilon \end{bmatrix}\right) = \frac{1}{\varepsilon}, \\ \kappa_2(\Gamma^{\text{Orth}}) &= \kappa_2\left(\begin{bmatrix} \tilde{b} & 0 \\ -a\hat{b} & \hat{b}(1-|a|^2-\varepsilon a) \end{bmatrix}\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}\right), \\ &= \kappa_2\left(\begin{bmatrix} \tilde{b} & 0 \\ -a\hat{b} & \hat{b}(1-|a|^2-\varepsilon a) \end{bmatrix}\right), \\ &= \frac{\sqrt{1-|a|^2}}{\sqrt{1-|a+\varepsilon|^2}(1-|a|^2-\varepsilon a)}, \end{aligned}$$

respectively. Here the condition number for the THREE controllability matrix is considerably larger. Having several poles close to each other enforces this behavior.

Secondly, consider state covariance estimate $\hat{P}$. Its condition number will depend on the condition num-

ber of the corresponding controllability matrix. For infinite data sequences we have the bounds:

$$\frac{1}{\kappa_2(\Gamma)^2} \leq \frac{\kappa_2(P)}{\kappa_2(L_n^T K L_n)} \leq \kappa_2(\Gamma)^2. \qquad (16)$$

This gives a relation for the condition number of the state-covariance matrices:

$$\kappa_2(P^{\text{Orth}}) \leq \frac{\kappa_2(\Gamma^{\text{Orth}})^2}{\kappa_2(\Gamma^{\text{THREE}})^2} \kappa_2(P^{\text{THREE}}). \qquad (17)$$

Thus the condition number of the state covariance matrix will generally be higher when the corresponding controllability matrix is. The conclusion is that the optimization problem (10) will have approximately the same conditioning for the two sets of basis functions. However, will the problem with the orthonormal basis function be internally better scaled.

Here it is interesting to consider the basis functions z^{-k}. They are generically well-scaled since the corresponding controllability matrix is the identity matrix.

4.3 *Accuracy in estimating the state covariance*

For relatively short data sequences there are two effects that will determine the accuracy. Firstly, the orthonormal basis functions span the space so that each state contain information that is not in the prior states. From (17) we see that this will reduce the condition number of the state covariance matrix. This makes it less sensitive to noise. Secondly, the time lags in the filter banks will decrease the accuracy. As pointed out in Byrnes et al. (2000) one of the merits of the THREE algorithm is the time lag is one for all filters in the filter bank. These two effects work in different directions and which that is dominant will depend on the choice of poles for the filter banks.

5. A NUMERICAL EXAMPLE

Consider the shaping filter

$$W(z) = \frac{z - 0.9}{z - 0.8}.$$

Driven by Gaussian white noise, the output of the system is measured as depicted in Figure 2. The task is to recover the original system as accurately as possible. The shaping filter $W(z)$ is chosen to a simple first order filter. The zero relatively close to the pole will make it fairly hard to identify.

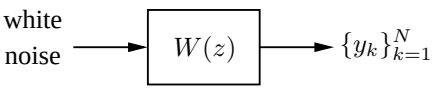

Fig. 2. The shaping filter in the numerical example

Given the artificial measurements $\{y_k\}_{k=1}^N$, the shaping filter (and correspondingly the spectral density) is identified using a Linear Prediction Filter (LPC), the THREE algorithm (THREE) and the orthonormal basis function procedure of this paper (Orth). The poles are chosen to be the same for the THREE and orthonormal algorithms and they are chosen to be in the vicinity of the pole at 0.80, but not necessarily identically to it. Since the THREE algorithm does not allow for repeated poles in the basis the case of distinct poles is consider. Also note that since the proposed algorithm is developed in the same framework as the THREE, the optimization problem (10) is also solved using the THREE algorithm.

In order to compare two estimates of the shaping filters, we compute the Kullback-Leibler distance, defined in (8), between the true and the estimated normalized spectral densities: $\mathbb{S}(\hat{\Phi}^{true} \| \hat{\Phi}^{estimate})$. The normalization guarantees that the distance will be nonnegative and that it is zero exactly when the normalized densities coincide.

The system is driven for a while so that the stationary assumption of the filter is at least approximatively valid. For each parameter set the 100 Monte Carlo simulations are performed and the average Kullback-Leibler distance is computed. Here the maximum entropy solution is computed, for simpler comparison. This means that the accuracy of the estimates can be significantly increased if a good initial estimate is given.

As discussed in Section 4 both the number of basis functions in the filter bank, this is the order of the identified system, and the length of the data sequence affects procedure differently. In the Table 1 the results are given for all combinations of the data length $N = 20, 200 \,\&\, 2000$ and the number of basis functions $n = 1, 3 \,\&\, 5$. In addition a few estimate for the case $n = 1$ and $N = 2000$ is plotted in Figure 3. The basis function poles are chose to be $\{0.75\}$, $\{0.75, 0.70 \,\&\, 0.80\}$, and $\{0.75, 0.70, 0.80, 0.72 \;\&\; 0.78\}$ for the different values of n. Note that for the orthonormal basis functions the ordering of the poles is vital (see Bodin et al. (2000) for a discussion in the case without positivity constraint).

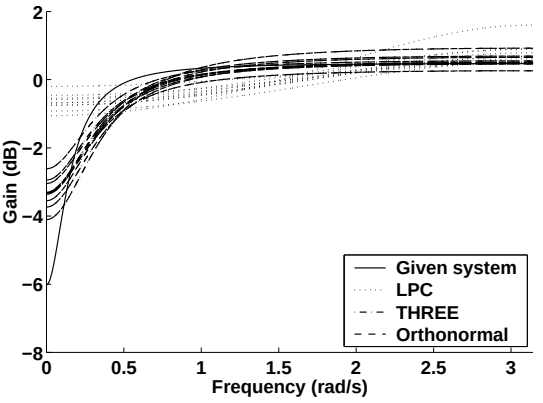

Fig. 3. The true and some identified spectral density for $n = 1$ and $N = 2000$

Table 1. Kullback-Leibler distances between true and estimated spectral densities

N		20			200			2000	
n	LPC	THREE	Orth	LPC	THREE	Orth	LPC	THREE	Orth
1	0.0633	0.0350	0.0392	0.0189	0.0052	0.0053	0.0153	0.0032	0.0032
3	0.1779	0.0354	0.1367	0.0235	0.0311	0.0097	0.0099	0.0288	0.0010
5	0.2704	0.0308	0.2136	0.0300	0.0324	0.0177	0.0076	0.0321	0.0025

The results are as expected from the discussion in Section 4. It is clear that the non-default basis choice can increase the accuracy in the estimates. The THREE algorithm gets problems in estimating the state covariance matrix in more cases than the orthonormal algorithm. For very short data sequences the THREE algorithm produces a better result when i works. Both the THREE and the orthonormal algorithms seem to perform worse when the number of basis functions increases. Thus, they are better suited for directly identifying a low-order model rather than first identifying a high-order model that is to be model-order reduced.

Remark 7. There is no comparison to the approximative identification algorithm of de Hoog et al. (2002) made since that algorithm is designed for deterministic identification, that is when the input signal is available. However, it would be possible to replace the least square estimates of Van den Hof et al. (1995) with estimate given from the interpolation problem and translated to the spectral factor. Applying the method directly to the transfer function $f_y(z)$ would not guarantee positive-realness.

Remark 8. In Byrnes et al. (2000) and Georgiou (2001) the increased precision in certain frequency ranges is emphasized, but in this example the whole frequency region is considered.

6. CONCLUSIONS

This paper provides a procedure for spectral estimation using orthonormal basis functions. The filter bank framework allow for combination of different types of basis functions including the orthonormal of this paper. It is discussed and illustrated how the orthonormality in the basis can be important; in particular this seems to be important for concentrated poles in the basis. An extension to vector processes seems fairly straight forward in the light of Blomqvist et al. (2002).

ACKNOWLEDGMENT

The authors want to thank their advisor Anders Lindquist and Ryozo Nagamune for suggestion of research topic and guidance.

References

A. Blomqvist, A. Lindquist, and R. Nagamune. Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimization approach. Preprint, 2002.

A. Blomqvist and R. Nagamune. A numerical algorithm for Kullback-Leibler approximations. Preprint, 2002.

P. Bodin, L. Villemoes, and B. Wahlberg. Selection of best orthonormal rational basis. *SIAM J. Contr. and Optimiz.*, 38(4):995–1032, 2000.

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*, 48(11):3189–3205, November 2000.

P. E. Caines. *Linear Stochastic Systems*. John Wiley & Sons, New York, 1988.

T. de Hoog, Z. Szabó, P. Heuberger, P. Van den Hof, and J. Bokor. Minimal partial realization from generalized orthonormal basis function expansions. *Automatica*, 38:655–669, 2002.

T. T. Georgiou. Spectral estimation via selective harmonic amplification. *IEEE Trans. Automat. Control*, 46(1):29–42, January 2001.

T. T. Georgiou and A. Lindquist. Kullback-Leibler approximation of spectral density functions. Preprint, 2002.

R. Nagamune. A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint. In *Proceeding of the 40th IEEE Conference on Decision and Control*, pages 1119–1124, Orlando, Florida, December 2001.

B. Ninness and F. Gustafsson. A unifying construction of orthonormal bases for system identification. *IEEE Trans. Automat. Control*, 42(4):515–521, 1997.

P. Van den Hof, P. Heuberger, and J. Bokor. System identification with generalized orthonormal basis functions. *Automatica*, 31:1821–1834, 1995.

B Wahlberg. System identification using Laguerre models. *IEEE Trans. Automat. Control*, 36(5):551–562, 1991.