Brief paper

Spectral estimation by least-squares optimization based on rational covariance extension

Giovanna Fanizza\textsuperscript{a,∗}, Ryozo Nagamune\textsuperscript{b}

\textsuperscript{a}Department of Mathematics, Division of Optimization and System Theory, Royal Institute of Technology, SE 100 44 Stockholm, Sweden
\textsuperscript{b}Department of Mechanical Engineering, The University of British Columbia, Vancouver, BC, Canada V6T 1Z4

Received 9 December 2005; received in revised form 3 May 2006; accepted 4 September 2006

Abstract

This paper proposes a new spectral estimation technique based on rational covariance extension with degree constraint. The technique finds a rational spectral density function that approximates given spectral density data under constraint on a covariance sequence. Spectral density approximation problems are formulated as nonconvex optimization problems with respect to a Schur polynomial. To formulate the approximation problems, the least-squares sum is considered as a distance. Properties of optimization problems and numerical algorithms to solve them are explained. Numerical examples illustrate how the methods discussed in this paper are useful in stochastic model reduction and stochastic process modeling.

Keywords: Spectral estimation; Optimization; Rational covariance extension; Least-squares sum; Schur polynomial

1. Introduction

Spectral estimation is of great importance for various topics in systems theory; for instance, system identification, model reduction, speech and signal processing (Ljung, 1999; Söderström & Stoica, 1989; Stoica & Moses, 1997). In parametric spectral estimation, the main problem is to estimate the spectral density function that captures characteristics of a stochastic process, such as covariances, cepstrums, Markov parameters, and the frequency response of the process.

Traditionally, the maximum entropy (ME) solution for the covariance extension problem is one of the most popular spectral estimates, since it is rational of relatively low degree, and matches a given partial covariance sequence (Burg, 1975). However, there is no guarantee that the ME solution possesses favorable characteristics other than covariances matching (Byrnes, Gusev, & Lindquist, 2001). To overcome this drawback, several attempts have been made (Byrnes, Enqvist, & Lindquist, 2001; Byrnes, Georgiou, & Lindquist, 2000; Skelton & Anderson, 1986). Despite these attempts, the question of designing a spectral density that optimally approximates covariances and the frequency response of a process simultaneously has not been fully investigated yet, and this is the main topic of this paper.

In this paper, we will propose a new design technique of the spectral density function for a given covariance sequence. The core theory that we utilize in the technique is the rational covariance extension (RCE) theory with degree constraint in Byrnes, Lindquist, Gusev, and Matveev (1995). The theory parameterizes all the rational spectral densities of a bounded degree that match a covariance sequence in terms of a Schur polynomial, thus completing a line of study initiated in Georgiou (1983, 1987). Regarding the Schur polynomial as a design parameter, we will formulate and solve problems of approximating the frequency response of the spectral density function to given density data. Density data are pairs of gridded frequencies and “desired” amplitudes of a density function, and assumed to be obtained from experimental data or a priori information of a “desired” density shape. As a distance between a density function to density data, we will adopt the least-squares
sum (Nash & Sofer, 1996). We will minimize this distance under some constraint on covariances.

To be concrete about the constraint on covariances, we will consider the following two situations. The first situation is when we can trust the covariance accuracy. Such a situation occurs in model reduction, where we would like to reduce the order of a mathematical model maintaining the finite covariance sequence. In this case, we presume an allowable region of covariances from given ones, and approximate the density function to density data under this constraint. The “size” of the allowable region is at the discretion of designers. The corresponding problem becomes a “min–min” optimization problem, where the first and second minimizations are with respect to a Schur polynomial and a covariance sequence, respectively. The second case is when we assume that given covariances have uncertainty. Such situation is typical in using experimental data to estimate covariances, since the data length cannot be infinite in reality and the data are normally contaminated by noise. In this case, we approximate a density function to spectral density robustly, i.e., for any covariance sequence in the uncertainty region. The problem in this case is a “min–max” optimization problem, where we minimize, with respect to a Schur polynomial, the maximum (worst case) distance due to covariance uncertainty.

The “min–min” and “min–max” problems in this paper are nonconvex. However, since the cost functional is continuous and differentiable, and since the domain of the functional is connected, we can search for local optima with gradient-based algorithms, by selecting an appropriate initial point of a Schur polynomial. The optimization technique is the same as the one developed for sensitivity shaping in robust control in Nagamune and Blomqvist (2005). However, neither “min–min” nor “min–max” types of optimization problems mentioned above, have been dealt with in Nagamune and Blomqvist (2005). In that respect, this paper deals with a much wider and more interesting class of optimization problems.

2. Motivating applications

In system design, estimation of a spectral density is one of the major problems, and it has been studied in different settings (Desai & Pal, 1984; Stoica & Moses, 1997). In this section, through some examples in model reduction and spectral estimation, we will show the importance of introducing a new methodology for the estimation of the rational spectral density. These examples will be revisited in Section 6 to show the advantages of the proposed method.

2.1. Model reduction

In model reduction, for a given filter $W$, a reduced order system is computed (Obinata & Anderson, 2001). Two of the most common techniques of model reduction are the balanced-truncation (BT) (Moore, 1981; Mullis & Roberts, 1976), and the stochastically balanced truncation (SBT) (Desai & Pal, 1984; Lindquist & Picci, 1996). One advantage of BT and SBT model reduction is that it is easy to implement for small and medium order systems. As an example, let us consider the following system:

$$W(z) = \frac{(z - 0.9e^{1.2i})(z - 0.9e^{-1.2i})(z - 0.05)(z + 0.15)}{(z - 0.8e^{1.7i})(z - 0.8e^{-1.7i})(z - 0.1)(z + 0.1)}.$$ (1)

Using the model reduction techniques, the order of the system is reduced to two. Fig. 1 shows the spectral density computed by both BT model reduction and SBT model reduction. The rational spectral density obtained by SBT model reduction approximates the original one better that the one computed by BT model reduction at the “valley” and low frequency.

In this example, the computation of SBT reduced order system is easy. However, in large-scale settings, this model reduction technique can be numerically inefficient and can be ill-conditioned. Moreover, SBT model reduction does not guarantee the optimality in the approximation of the spectral density data.

2.2. Spectral estimation

The essence of spectral estimation is to estimate a spectral density from a finite record of stationary data sequence $\{y(t)\}_{t \in \mathbb{Z}}$. The stationary data sequence is modeled as an output signal of a filter $W$ whose input is a white noise $\{u(t)\}_{t \in \mathbb{Z}}$.

It is well known that the spectral density of the process $\{y(t)\}_{t \in \mathbb{Z}}$ has a Fourier representation

$$\Phi(z) = c_0 + \sum_{i=1}^{\infty} c_i (z^i + z^{-i}),$$ (2)

$$c_k := E[y(t+k)y(t)], \quad k = 0, 1, 2, \ldots ,$$ (3)

where $\{c_k\}_{k=1}^{\infty}$ is the covariance sequence and $E[\cdot]$ means the expected value. The covariance sequence is usually estimated...
from an approximation

\[ c_k = \frac{1}{N-k+1} \sum_{t=0}^{N-k} y_{t+k} y_t, \]

since only a finite record \( \{y_k\}_{k=0}^N \) of observation of the process is available.

In spectral estimation, we often face the problem of finding a spectral density, which is positive on the unit circle, and matches only a finite number of covariances \( c = (c_0, c_1, \ldots, c_n) \). One particular solution of this problem is the spectral density of the ME filter, whose spectral zeros are all at the origin. However, in some applications, a wider variety in the choice of spectral zeros of the spectral density \( \Phi(z) \) can be required. In fact, let us consider the following system taken from (Mari, Dahlén, & Lindquist, 2000):

\[ W(z) = \frac{z^5 + \sum_{i=1}^{5} s_i z^{-i}}{z^5 + \sum_{i=1}^{5} a_i z^{-i}}, \]

where the value of \( s_i \) and \( a_i \), are shown in Table 1. The output signal \( \{y_k\}_{k=0}^N \) of the filter (5) is sampled with \( N = 500 \).

From (4), we obtain the estimation of the covariance sequence \( (c_0, c_1, \ldots, c_5) \)

\[ (1.902, 0.861, -0.195, -0.505, -0.248, -0.500). \]

Using these covariances, the rational spectral density derived from a ME design of (5) is computed. In Section 6, we will show that by placing the spectral zeros in the unit disc other than at the origin, the corresponding rational spectral density may give a better estimation of the “true” one (Byrnes et al., 2001). There we are left with the task of estimating the spectral zeros such that the corresponding rational spectral density approximates the “true” one. Moreover, the estimation of the covariances (6), is affected by errors since only a finite number of observations of the process \( \{y(t)\}_{t \in \mathbb{Z}} \) are available, and the process itself is contaminated by noise (Stoica & Moses, 1997). In fact, the first six “true” covariances can be computed from (5) as

\[ \sigma = (1.862, 0.831, -0.215, -0.531, -0.22, -0.431). \]

Therefore, in the estimation of the rational spectral density, we should take into consideration the uncertainty in the covariance sequence.

3. Rational covariance extension problem

Modern spectral estimation is often based on a partial covariance sequence obtained from a stochastic process. The extension of this partial sequence is called the covariance extension.

In this section, we will review the rational covariance extension (RCE) problem and main known results for this problem (Byrnes et al., 2001).

Consider a covariance sequence of length \( n+1 \)

\[ c = (c_0, c_1, \ldots, c_n). \]

The sequence \( c \) is called positive when the corresponding Toeplitz matrix is positive definite (Georgiou, 1987). For a given positive covariance sequence (8), the RCE problem is to find a rational spectral density \( \Phi(z) \) of order at most \( n^1 \)

\[ \Phi(z) = \tilde{c}_0 + \sum_{k=1}^{\infty} \tilde{c}_i (z^i + z^{-i}) \]

that satisfies interpolation conditions

\[ \tilde{c}_i = c_i, \quad i = 0, 1, \ldots, n, \]

and a positivity condition

\[ \Phi(z) > 0 \quad \forall |z| \leq 1. \]

In Georgiou (1983, 1987), for the RCE problem, Georgiou conjectured that the class of all the spectral densities of degree at most \( n \) is completely characterized in terms of the class of Schur polynomials of degree \( n \). This conjecture was proven to be true in Byrnes and Lindquist (2000) and Byrnes et al. (1995) where Byrnes et al. stated the following theorem (we use notation \( z = [z^n, \ldots, 1]^T \) and \( z^a = [z^{-a}, z^{-(a-1)}, \ldots, 1]^T \)).

**Theorem 1.** Let \( c \) be a given positive covariance sequence. For each vector \( \sigma \) in

\[ \Xi := \{ \sigma \in \mathbb{R}^{n+1} : \sigma_0 > 0, z^\sigma \neq 0, \forall |z| \geq 1 \}, \]

there exists a unique vector \( \mathbf{a} \in \Xi \) such that the rational spectral density \( \Phi \) of order at most \( n \) satisfying (10) and (11), can be written as

\[ \Phi(z) = \frac{\sigma^T z(z^a)^T \sigma}{a^T z(z^a)^T a}. \]

Moreover, the map \( h_c \) from \( \Xi \) to \( \mathbf{a} \) is a diffeomorphism.

The concrete expression of the map \( h_c \) is shown in Fanizza and Nagamune (2006). Due to this theorem, the set of all spectral density functions of order at most \( n \), that match the given covariance \( c \), can be parameterized in terms of \( \sigma \) as

\[ \mathcal{P}_n(c) := \left\{ \Phi(z) : \frac{\sigma^T z(z^a)^T \sigma}{a^T z(z^a)^T a} = h_c(\sigma), \ \sigma \in \Xi \right\}. \]

Byrnes and Lindquist (2006) state a “dual” theorem of Theorem (1), “dual” in the sense that the vector \( \sigma \) is fixed, instead of the covariance sequence.

**Theorem 2.** Let \( \sigma \) be given in \( \Xi \). For each positive covariance \( c = (c_0, c_1, \ldots, c_n) \), there exists a unique vector \( \mathbf{a} \in \Xi \) such

\footnote{A spectral density is of order \( n \) if its outer spectral factor is of order \( n \).}
that the rational spectral density $\Phi$ of order at most $n$ satisfying (10) and (11), can be written as

$$\Phi(z) = \frac{\sigma^T z(z^*)^T \sigma}{a^T z(z^*)^T a}.$$}

Moreover, the map $g_\sigma$ from $c$ to $a$ is a diffeomorphism.

Due to this theorem, the set of all the rational spectral densities of order at most $n$ with fixed zero polynomials $\sigma^T z(z^*)^T \sigma$, can be parameterized in terms of the covariance sequence $c$ as

$$\mathcal{G}_n(\sigma) := \left\{ \Phi(z) := \frac{\sigma^T z(z^*)^T \sigma}{a^T z(z^*)^T a} \mid a = g_\sigma(c), \ c \text{ is positive} \right\}. \quad (14)$$

In this way, neither only covariance matching nor only the numerator of $\Phi$ determine the spectral density function uniquely, and there still is a freedom represented by the vector $\sigma$ and/or the vector $c$. Utilizing this freedom, we can do Markov parameter matching, cepstral matching (Stoica & Moses, 1997) or spectral density shaping, which are also important characteristics in stochastic processes. The aim of this paper is to exploit spectral density shaping, which are also important characteristics of a process. Next, we will formulate two optimization problems for spectral density matching with respect to $\sigma$ and $c$.

4. Spectral density approximation

Suppose that, for a given stochastic process, we have obtained a covariance sequence $c = (c_0, c_1, \ldots, c_n)$ and spectral density data $\varphi_k$ at a finite number of frequencies $\theta := \{\theta_k\}_{k=1}^N$. See Fig. 2.

We would like to find a rational spectral density of order at most $n$ that satisfies the matching of covariances and spectral density data. However, such a spectral density does not exist in most practical problems, since the actual process may not be finite dimensional, and the data are corrupted by noise. Therefore, we seek a rational spectral density of order at most $n$ that approximates the spectral density data under some requirements on the covariances. Later, we will consider two situations about these requirements: certain covariances and uncertain covariances.

To clarify the meaning of approximation, we need to introduce two distances: a distance between two covariance sequences, and a distance between the spectral data and a rational spectral density computed at each frequency.

4.1. Definition of distances

A distance between two covariance sequences of the same length is defined as

$$d_c(c, \hat{c}) := ||c - \hat{c}||, \quad (15)$$

where $|| \cdot ||$ is some (possibly weighted) norm.

As a distance between a rational spectral density function evaluated at the gridded frequencies, and spectral density data $\varphi = \{\varphi_k\}_{k=1}^N$, we will use a weighted least-squares sum $d_f$; for example,

$$d(\varphi, \hat{\varphi}) := \frac{1}{2} \sum_{k=1}^N w_k \left| 1 - \frac{\varphi_k}{\hat{\varphi}_k} \right|^2, \quad (16)$$

with $w = \{w_k\}_{k=1}^N$ positive scalars chosen by the designer and $\hat{\varphi} := \{\hat{\varphi}_k\}_{k=1}^N$, $\hat{\varphi}_k := \Phi(e^{i\theta_k})$.

4.2. Optimization for spectral density approximation

Suppose that $\Phi$ is taken from the set $\mathcal{P}_n(c)$. We seek such a $\Phi$ that approximates the spectral density data under some requirements on the covariance. Since $\Phi$ depends uniquely on both $\sigma$ and $c$, we write the element in the set $\mathcal{P}_n(c)$ as $\Phi(\sigma, c)$.

Thus, the distance $d$ can be written as functions of vectors $\sigma$ and $c$; for example, (16) can be written as

$$d(\varphi, \Phi(\sigma, c)) = \frac{1}{2} \sum_{k=1}^N w_k \left| 1 - \frac{\varphi_k}{\Phi_k} \right|^2,$$ \quad (17)

where $a = h_\sigma(\sigma)$ and $e_k := [e^{i\theta_1}, \ldots, e^{i\theta_n}]^T$.

Therefore, the spectral density approximation problem amounts to finding vector $\sigma \in \mathcal{Z}$ that minimizes the distance (17) under some requirements on the covariance sequence.

In this paper, we consider two different problems:

$$\min_{\sigma \in \mathcal{Z}} \min_{\varphi \in \mathcal{P}_n} d(\varphi, \Phi(\sigma, \hat{\varphi})),$$ \quad (18)

$$\min_{\sigma \in \mathcal{Z}} \max_{\varphi \in \mathcal{P}_n} d(\varphi, \Phi(\sigma, \hat{\varphi})), \quad (19)$$

where $\hat{\varphi}_n$ is defined, for a given scalar radius $r > 0$, as

$$\hat{\varphi}_n := \{\hat{\varphi} : \text{positive, } d_c(c, \hat{\varphi}) < r\}. \quad (20)$$

Both problems (18) and (19) are looking for the best $\sigma$ such that the corresponding rational spectral density $\Phi$ approximates the spectral density data. The difference is the interpretation of the set $\hat{\mathcal{P}}_n$. In (18), $\hat{\mathcal{P}}_n$ is interpreted as an “admissible” error on the covariances to improve the approximation of the spectral density data with respect to $\sigma$. On the other hand, in (19), $\hat{\mathcal{P}}_n$ is considered to be the uncertainty region of a given covariance sequence $c$. To maintain the approximation of
the spectral density against perturbation of \( c \), we minimize the worst-case discrepancy with respect to \( \sigma \).

**Remark 3.** The optimization problem (18) has been already studied from a theoretical point of view in Byrnes and Lindquist (2003), with the distance \( d \) being the Kullback–Leibler (KL) distance (Georgiou & Lindquist, 2003). Using the methodology suggested in this section, the problem studied in Byrnes and Lindquist (2003) can be solved numerically. The problem in Byrnes and Lindquist (2003) is approximately reformulated in our setting

\[
\min_{\sigma \in \mathbb{C}} \min_{\hat{c} \in \hat{S}^{KL}} d_{KL}(\sigma, \phi(\sigma, \hat{c})), \tag{21}
\]

with \( d_{KL} \), the discretized KL distance

\[
d_{KL}(\sigma, \phi(\sigma, c)) := \frac{1}{N} \sum_{k=1}^{N} (\phi_k \log \phi_k - \phi_k \log \phi_k). \tag{22}
\]

The positivity of (22) is guaranteed by the condition on the covariances \( c_0 = \int_{-\pi}^{\pi} \phi(e^{i\theta}) d\theta/2\pi \). Thus, the domain constraint \( \hat{S}_n^{KL} \) is equal to \( \hat{S}_n \) in (20) with one extra constraint \( c_0 = 0 \).

5. Properties and algorithms for optimization

The problems (18) and (19) are written compactly as

\[
\min_{\sigma \in \mathbb{C}} \min_{\hat{c} \in \hat{S}_n} f(\sigma, \hat{c}) \quad \text{and} \quad \min_{\sigma \in \mathbb{C}} \max_{\hat{c} \in \hat{S}_n} f(\sigma, \hat{c}), \tag{23}
\]

with \( f: \mathbb{C} \times \hat{S}_n \to \mathbb{R} \) defined by \( f(\sigma, c) := d(\sigma, c) \). The problems (23) have following properties:

1. (P1) \( f \) is nonconvex with respect to \( (\sigma, c) \). In addition, \( f(\cdot, c) : \mathbb{C} \to \mathbb{R} \) is nonconvex in the set \( \mathbb{C} \), and so is \( f(\sigma, \cdot) : \hat{S}_n \to \mathbb{R} \).
2. (P2) \( f \) is continuously differentiable on \( \mathbb{C} \times \hat{S}_n \).
3. (P3) The domain \( \mathbb{C} \times \hat{S}_n \) is nonconvex.

Due to (P1) and (P3), a unique global minimizer is not guaranteed. Thus, we try to find a local minimizer in \( \mathbb{C} \times \hat{S}_n \) by choosing a proper initial point. Due to (P2), we can use gradient-based algorithms to solve (23).

5.1. Initial points

The choice of the initial point depends on the problem that we have at hand. We will consider two estimation problems: the estimation of the optimal \( (\sigma, c) \) from experimental data, and in model reduction. Next, for the two situations, different candidates will be suggested as initial points.

In the first situation, the initial \( \sigma \) is taken to be \( [1, 0, \ldots, 0]^T \) which corresponds to the ME solution. The initial covariances \( c \) are the covariances estimated from experimental data. In the second one, the initial \( \sigma \) is taken such that \( \sigma^T \zeta(\varepsilon)^T \sigma \) is the zero polynomial of the spectral density of the BT model. The initial covariances \( c \) equal the first \( n_r \) element of the covariance sequence of the given spectral density, with \( n_r \) the degree of the reduced model. Besides these candidates of initial points, we can use whichever feasible points if we have a priori knowledge about ”good” spectral zeros.

5.2. Algorithms

Next, we will discuss algorithms that we adopt for solving the optimization problems. The choice of the algorithm depends on which problems we are solving, min–min or min–max.

For problem (18), there are two popular algorithms: Gauss–Newton and Levenberg–Marquardt (Nash & Sofer, 1996, Chapter 13). These two methods were originally developed for unconstrained nonlinear least-squares problems. In order to incorporate the constraints \( \sigma \in \mathbb{C} \) and \( c \in \hat{S}_n \), we modify these algorithms in line with Nagamune and Blomqvist (2005).

For problem (19), we solve two problems iteratively:

- In the first step, we design a \( \sigma \) by solving a min–max problem:

\[
\min_{\sigma \in \mathbb{C}} \max_{k=1,...,S} f(\sigma, c^{(k)}), \quad \tag{24}
\]

where \( \{c^{(k)}\}_{k=1}^{S} \) are samples of the covariances in the set \( \hat{S}_n \), and design parameters. The problem (24) is solved using a sequential quadratic programming method (Nash & Sofer, 1996, Chapter 15.5). By solving (24), we are able to find a worst-case \( c^{(k)} \) and a suboptimal \( \sigma \).

- In the second step, we analyze the designed \( \sigma \) by solving a maximization problem:

\[
\max_{l=1,...,N} f(\sigma, c^{(l)}), \quad \tag{25}
\]

where \( \{c^{(l)}\}_{l=1}^{N} \) is a set of \( N \) covariance sequences randomly chosen in \( \hat{S}_n \). In this way, the solution \( c \) of (25) affects adversely the density approximation for the designed \( \sigma \).

We redefine the sample of covariances by \( \{c^{(k)}\}_{k=1}^{S} \cup c \). We go back to the first step.

We will stop the iteration when the difference between the values of (24) and (25) are sufficiently small.

In any case, this algorithm guarantees an approximation of a density function only for a finite number of sampled covariances in \( \hat{S}_n \). Therefore, it is important to analyze the obtained \( \sigma \), for example, by sampling a number of \( c^{(k)} \in \hat{S}_n \). In the analysis, if we find some “bad” covariances, we should include them in the sampled set, and redo the iterations above.

6. Examples

6.1. An example in model reduction

Let us again consider the fourth order system (1). Suppose that we would like to reduce the order of this system to two.
For this purpose, we compute the first two covariances \((c_0, c_1)\) as
\[
W(z)W(z^{-1}) = c_0 + c_1(z + z^{-1}) + \cdots.
\]
(26)

We have tried several methods to reduce the model order, that is, the BT, and using the covariance \((c_0, c_1)\) computed in (26), our minimization (18) for two different distances: least-squares sum and KL distance. In the minimization problem (18), the spectral density data \(\varphi = \{\varphi_k\}_{k=1}^N\) are computed at the frequencies \(\theta = [0, 0.01, 0.02, \ldots, \pi]\), and the set \(\mathcal{C}_n\) is defined such that the distance \(d_c\) is:
\[
d_c(\hat{c}, c) = \|\hat{c} - c\|_{\infty} \leq 0.1.
\]

Fig. 3 shows the comparisons of the frequency responses, and as can be seen, the least-squares method approximates best the original system. In addition, Fig. 3 shows that the KL method approximates the original system worse than the BT does, even if the final value of the cost function for KL is smaller than the BT. This behavior can be easily explained. In fact, at each frequency \(\theta_k\), each element in the summation (22), can be either positive or negative. Thus, even if the total sum is small, this does not imply that the rational spectral density \(\Phi\) is close to the spectral density data \(\varphi\) at each sampled point of the frequency \(\theta_k\).

6.2. An example in spectral estimation

Let us again consider the fifth order system (5). The covariances are computed as explained in Section 2.2. Using the estimated covariances, we solve the min–max problem (19) for the least-squares distance (17).

In this example, the set \(\mathcal{C}_n\) is defined such that the distance \(d_c\) is
\[
d_c(\hat{c}, c) = \|\Omega(\hat{c} - c)\|_{\infty} < 1,
\]
where \(\Omega := \text{diag}([c_0 - c_0^*], \ldots, [c_5 - c_5^*])^{-1}\), and \(c^*\) and \(c^*\) are computed, respectively, in (6) and (7). Figs. 4 and 5 show, respectively, the nominal and robust performance of the designed \(\sigma\). To show the robustness of the solution, we chose randomly 30 covariance sequences in the set \(\mathcal{C}_n\), and compute the corresponding rational spectral density for a fixed zero polynomial \(\sigma^T(z^{*}) \hat{\sigma}\). From these figures, we can see that the designed \(\sigma\) provides a better approximation, both nominally and robustly, of the spectral density data than the maximum-entropy solution does.

6.3. Another example in spectral estimation

Finally, we will apply our proposed method to a spectral estimation example using real experimental data, which is taken from Byrnes et al. (2001, Example 3.4). In this example, speech data \(\{y_k\}_{k=0}^{250}\) measured during around 30 ms are given. From the data, we can compute the first seven covariance lags \((c_0, \ldots, c_6)\) as
\[
(1, 0.8689, 0.5813, 0.2540, -0.0031, -0.1299, -0.1662).
\]

In Byrnes et al. (2001), for the density matching to the periodogram, a manually tuned Schur polynomial has been proposed as \(\sigma(z) = z^5 + \sum_{k=1}^{5} \sigma_k z^{6-k}\), where the coefficients \((\sigma_1, \ldots, \sigma_6)\) were given by
\[
(-0.136, -0.301, -0.228, -0.507, 0.556, 0.429).
\]
(27)

In Fig. 6 (left), the robustness of design proposed in Byrnes et al. (2001) is shown by randomly choosing 30 covariance sequences in the set \(\mathcal{C}_n\) for a fixed Schur polynomial with coefficients (27), where we assume 0.1% uncertainty in each covariance lags. Although the corresponding spectral density approximates well the periodogram, the selected Schur polynomial is not optimal in whatever sense, and covariance uncertainty is not taken into account in their design.

Regarding the Schur polynomial in (27) as an initial point, we will solve the optimization problem (19) by the procedure

2 The first covariance is 1 due to data normalization.
presented in Section 4.2. Then, the optimization procedure has returned the following coefficients $(\sigma_1, \ldots, \sigma_6)$ of a Schur polynomial:

$(-0.167, -0.369, -0.279, -0.516, 0.657, 0.4445)$ \hspace{1cm} (28)

In Fig. 6 (right), the robustness of the solution (28) of the optimization problem is plotted with the periodogram, using the same 30 randomly chosen covariance sequences. This figure, as well as its zoom around particular frequencies in Fig. 7, shows that the proposed design approximates the periodogram around the frequencies 0.35 and 1.73 rad/s better than the previous design in Byrnes et al. (2001).

Finally, we would like to remark that the choice of the radius of $\hat{C}$ in the three examples presented in this section is arbitrary. However, it is not the purpose of this paper to suggest a systematic way of determining it. This will be exploited in detail in the future.

Fig. 5. Robustness of the ME solution (left) and of the solution of the min–max problem (right). The solid line represent the “true” system, and the dotted lines are for the 30 perturbed systems.

Fig. 6. Robustness of the previous and proposed designs.
7. Conclusions

In this paper, we have proposed a method for estimating a spectral density function from a given covariance sequence and spectral density data. The method is based on the covariance extension theory with degree constraint, where all bounded-order spectral densities matching the given covariance sequence are parameterized by the set of Schur polynomials. Using the Schur polynomials as design parameters, we have formulated the approximation problem of the frequency response of a spectral density to given density data. The distance between a density function and density data is measured with the least-squares sum evaluated at the gridded frequencies. To solve the optimization problems, we have used gradient-based method. Since all the problems are nonconvex, we have suggested some ways to select initial points for optimization. Numerical examples have illustrated that the rational spectral density estimated by the proposed technique approximates well the given spectral density data in the model reduction and spectral estimation settings.

Acknowledgments

The authors are grateful to Prof. A. Lindquist and Prof. A. Forsgren at the Royal Institute of Technology for many invaluable discussion and comments. In addition, the authors would like to thank Dr. P. Enqvist at the Royal Institute of Technology for providing them with experimental data. This work was supported by the Swedish Research Council (VR).

References


Giovanna Fanizza received M.S. degree in Mathematics from the University of Bari, Italy, in 1998. She spent the academic year 1999–2000 at the Department of Mathematics and Applications at the university Milano-Bicocca, Italy. She is currently a Ph.D. student at the Division of Optimization and System Theory at the Royal Institute of Technology (KTH), Stockholm, Sweden. Her research involves analytic interpolation with degree constraint, spectral estimation and model reduction.

Ryozo Nagamune received B.S. and M.S. degrees from the Department of Control Engineering at Osaka University, Japan, in 1995 and 1997, respectively, and Ph.D. degree from the Division of Optimization and Systems Theory at the Royal Institute of Technology, Stockholm, Sweden, in 2002. From January 2003 to June 2006, he was a postdoctoral researcher at the Mittag-Leffler Institute in Sweden, the University of California at Berkeley, and the Royal Institute of Technology in Sweden. He is currently an assistant professor at the Department of Mechanical Engineering, the University of British Columbia, Vancouver, Canada. His research interests are robust control theory and its application to mechanical control systems, spectral estimation and model reduction.