

Spectral Estimation by Least-Squares Optimization based on Rational Covariance Extension*

Giovanna Fanizza[†] and Ryozo Nagamune[†]

May 8, 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.

Key Words: 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 [25],[24],[17]. 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. In addition, rationality and degree of the density function are also crucial in practical applications.

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 [2]. However, there is no guarantee that the ME solution possesses favorable characteristics other than covariances matching [5]. To overcome this drawback, several attempts have been made; for example, simultaneous covariance and Markov parameter matching [23], simultaneous covariance and cepstrum matching [3], and estimation of the frequency

*This work was supported by grants from the Swedish Research Council (VR).

[†]The authors are with the Department of Mathematics, Division of Optimization and System Theory, Royal Institute of Technology, SE 100 44 Stockholm, Sweden. {vanna,ryozo}@math.kth.se

response of the density function using both filter banks and spectral zeros [4]. Despite these achievements, 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 theory with degree constraint* in [9]. 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 [12], [13]. 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 [21]. 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 deviation 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 situation 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 data *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. As an initial point, some candidates are, for example, the Schur polynomial corresponding to the ME solution or the stochastic balanced truncated model [11].

The basic idea in this paper is the same as the one developed for sensitivity shaping in robust control in [20]. However, neither “min-min” nor “min-max” types of optimization problems mentioned above, have been dealt with in [20]. In that respect, this paper deals with a much wider and more interesting class of problems.

The paper is organized as follows. In Section 2, we will explain, using some examples in the model reduction and spectral estimation, the motivation for suggesting a new technique for estimating a rational spectral density. The proposed

technique is based on the rational covariance extension theory which is reviewed in Section 3. Based on this theory, spectral density approximation problems are formulated as optimization problems in Section 4. The properties of the optimization problems and the procedure to solve these problems are presented in Section 5. To show the efficiency of the proposed methods, numerical examples are provided in Section 6. Computations of the cost functional and its gradient, which are necessary in numerical optimization, are explained in the Appendix.

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 [11],[25]. 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.

2.1 Model Reduction

In model reduction, for a given filter, a reduced order system is computed [22]. Two of the most common techniques to reduce the order of W are the *balanced-truncation (BT) model reduction* and the *stochastically balanced truncation (SBT) model reduction*. The BT model reduction was first introduced by [19], and later in the system and control literature by [18]. The SBT model reduction was first studied in [11] and later in the system and control context in [16]. One advantage of BT and SBT model reduction is that it is easy to implement for a relatively large order system. However, in some situations, BT and SBT model reduction can be insufficiently accurate in the approximation of the original spectral density. Next, we will show an example in which the BT model reduction and SBT model reduction do not approximate well the “true” system.

Let us consider the following system:

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

Using the model reduction techniques, the order of the system is reduced to two. Fig. 1 shows that the spectral density computed by both BT model reduction and SBT model reduction do not approximate well the original one in the “valley”, as well as low frequencies.

On the other hand, if we construct a system which is of order two, has spectral zeros at dominant spectral zeros of (1), $z = 0.9e^{\pm 1.2i}$, and matches the first two covariances of (1), then the corresponding rational spectral density will approximate the original one better than the one computed by BT [10] at the valley and low frequencies, at the expense of minor errors at high frequencies; see Fig. 2.

In this example the choice of the dominant spectral zeros is obvious. However, in more complex systems, such choice can be complicated, and a wrong one can affect adversely the estimation of the rational spectral density. Moreover, the choice of the dominant spectral zeros does not guarantee the *optimality* in the approximation of the spectral density data.

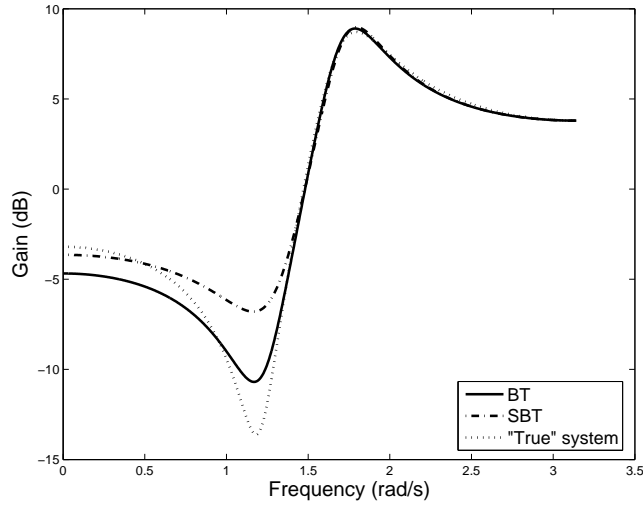


Figure 1: Frequency responses of the balanced truncated system, of the stochastically balanced truncated system, and of the “true” system

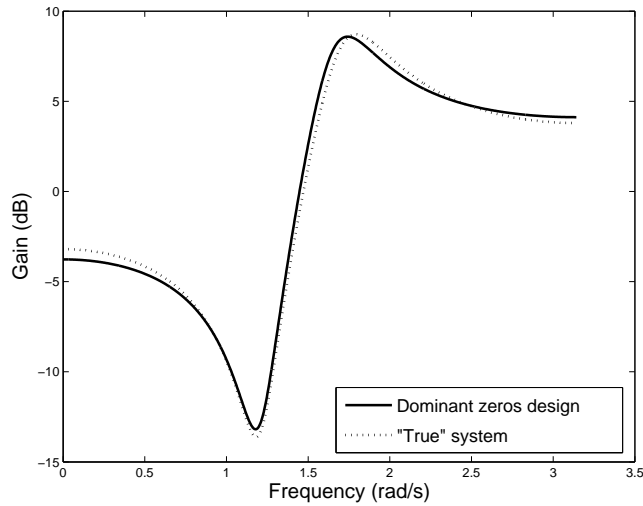


Figure 2: Frequency responses of the system by choice of the two dominant spectral zeros and the “true” system

In this paper, we propose a new technique, based on rational covariance extension theory, to estimate the optimal spectral zeros of a rational spectral density. After explaining this technique, in Section 6, we will revisit this example to show that the proposed method outperforms both BT/SBT model reduction and reduction by dominant spectral zeros.

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 $\{u(t)\}_{t \in \mathbb{Z}}$. See Fig. 3.

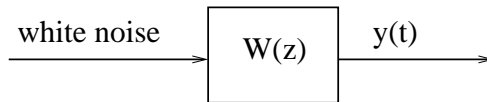


Figure 3: Filter, white noise and output data

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

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

where

$$c_k := E\{y(t+k)y(t)\}, \quad k = 0, 1, 2, \dots, \quad (3)$$

is the covariance sequence and $E\{\cdot\}$ means the expected value. 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

$$c = (c_0, c_1, \dots, c_n). \quad (4)$$

The covariance sequence is usually estimated from an approximation

$$c_k = \frac{1}{N-k+1} \sum_{t=0}^{N-k} y_{t+k} y_t, \quad (5)$$

since only a finite record

$$y_0, y_1, \dots, y_N \quad (6)$$

of observation of the process is available.

One particular solution of this problem is the spectral density of the *maximum-entropy* (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 Φ can be required. In fact, let us consider the following system taken from [15]

$$W(z) = \frac{z^5 + \sum_{i=1}^5 \sigma_i z^{5-i}}{z^5 + \sum_{i=1}^5 a_i z^{5-i}}, \quad (7)$$

where the value of σ_i and a_i , are shown in Table 1. The output signal of the filter (7) is sampled as in (6) with $N = 500$. From (5), we obtain the estimation of the covariance sequence:

$$(c_0, c_1, \dots, c_5) = (1.902, 0.861, -0.195, -0.505, -0.248, -0.500). \quad (8)$$

i	1	2	3	4	5
σ_i	-0.0550	-0.1497	-0.2159	0.1717	-0.0495
a_i	-0.7031	0.3029	0.1103	-0.1461	0.2845

Table 1: The value of σ_i and a_i .

Using these covariances, the rational spectral density derived from a ME design of (7) is computed. In Fig. 4 the frequency responses of the ME filter and the “true” system are plotted. Fig. 4 shows that the rational spectral density of the ME design does not approximate well the original one.

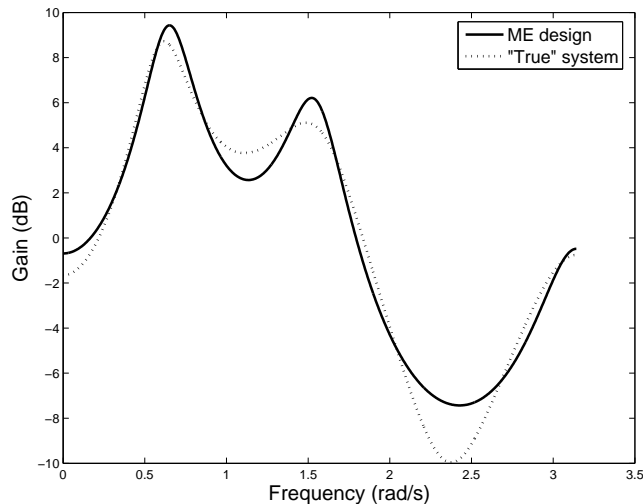


Figure 4: Frequency responses of the system of the maximum-entropy filter and the “true” system.

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 [3]. Therefore, we are left with the task of estimating the spectral zeros such that the corresponding rational spectral density approximates the “true” one better than the ME design does. Moreover, the estimation of the covariances (8), 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 [25]. In fact, the first six “true” covariances can be computed from (7) as

$$c^* = (1.862, 0.831, -0.215, -0.531, -0.224, -0.431). \quad (9)$$

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

In this paper, we propose a *robust* density shaping method with regard to covariance uncertainty. To show the robustness of the density function obtained by the proposed method, we will revisit this example in Section 6.

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 [5].

Consider a covariance sequence of length $n + 1$:

$$c = (c_0, c_1, \dots, c_n). \quad (10)$$

The sequence c is called *positive* when the corresponding Toeplitz matrix

$$T(c) := \begin{bmatrix} c_0 & c_1 & \dots & c_n \\ c_1 & c_0 & \dots & c_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ c_n & c_{n-1} & \dots & c_0 \end{bmatrix} \quad (11)$$

is positive definite. Let us denote \mathcal{C}_n^+ the space of all positive covariance sequences of length $n + 1$, that is,

$$\mathcal{C}_n^+ := \{c \in \mathbb{R}^{n+1} : T(c) > 0\}.$$

The set \mathcal{C}_n^+ is a nonempty, open, convex subset of \mathbb{R}^{n+1} .

For a given positive covariance sequence (10), the RCE problem is to find a rational spectral density $\Phi(z)$ of order at most n ¹

$$\Phi(z) = \tilde{c}_0 + \sum_{k=1}^{\infty} \tilde{c}_k (z^k + z^{-k}) \quad (12)$$

that satisfies interpolation conditions

$$\tilde{c}_i = c_i, \quad \forall i = 0, 1, \dots, n, \quad (13)$$

and a positivity condition

$$\Phi(z) > 0, \quad \forall |z| \leq 1. \quad (14)$$

In [12], [13], 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 [9],[7] where Byrnes et al. stated the following theorem (we use notation $\mathbf{z} = [z^n, \dots, 1]^T$ and $\mathbf{z}^* = [z^{-n}, z^{-(n-1)} \dots, 1]^T$):

Theorem 1 *Let c be given in \mathcal{C}_n^+ . For each vector $\boldsymbol{\sigma}$ in the set:*

$$\mathfrak{S} := \{ \boldsymbol{\sigma} := [\sigma_0, \sigma_1, \dots, \sigma_n]^T \in \mathbb{R}^{n+1} : \sigma_0 > 0, \mathbf{z}^T \boldsymbol{\sigma} \neq 0, \forall |z| \geq 1 \}, \quad (15)$$

there exists a unique vector $\mathbf{a} \in \mathfrak{S}$ such that the rational spectral density Φ of order at most n satisfying conditions (13) and (14), can be written as

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

Moreover, the map \mathbf{h}_c sending $\boldsymbol{\sigma}$ to \mathbf{a} is a diffeomorphism.

¹A spectral density is of order n if its outer spectral factor is of order n .

Remark 2 The concrete expression of the map h_c will be shown in the Appendix (see 43).

Due to this theorem, the set of all spectral densities functions of order at most n , that match the given covariance c , can be parameterized in terms of σ as

$$\mathcal{P}_n(c) := \left\{ \Phi(z) := \frac{\sigma^T \mathbf{z}(\mathbf{z}^*)^T \sigma}{\mathbf{a}^T \mathbf{z}(\mathbf{z}^*)^T \mathbf{a}}, \mathbf{a} = h_c(\sigma), \sigma \in \mathfrak{S} \right\}. \quad (16)$$

In [6], Byrnes and Lindquist state a “dual” theorem of Theorem (1), “dual” in the sense that the vector σ is fixed, instead of the covariance sequence.

Theorem 3 *Let σ be given in \mathfrak{S} . For each $c = (c_0, c_1, \dots, c_n)$ in the set \mathcal{C}_n^+ , there exists a unique vector $\mathbf{a} \in \mathfrak{S}$ such that the rational spectral density Φ of order at most n satisfying conditions (13) and (14), can be written as*

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

Moreover, the map g_σ sending c to \mathbf{a} is a diffeomorphism.

Remark 4 The diffeomorphic properties in Theorem 1 and Theorem 3 are important in Lemma 1 and Lemma 2 in the Appendix.

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

$$\mathcal{G}_n(\sigma) := \left\{ \Phi(z) := \frac{\sigma^T \mathbf{z}(\mathbf{z}^*)^T \sigma}{\mathbf{a}^T \mathbf{z}(\mathbf{z}^*)^T \mathbf{a}}, \mathbf{a} = g_\sigma(c), c \in \mathcal{C}_n^+ \right\}. \quad (17)$$

In this way, neither only covariance matching nor only the numerator of Φ determine the spectral density function uniquely, and there still is a freedom represented by the vector σ and/or the vector c . Utilizing this freedom, we can do Markov parameter matching, cepstral matching [25] or spectral density shaping, which are also important characteristics in stochastic processes. The aim of this paper is to exploit this freedom for taking into account such additional characteristics of a process. In the next section, we will formulate several optimization problems for spectral density matching with respect to σ and c .

4 Spectral density approximation

Suppose that, for a given stochastic process, we have obtained a covariance sequence $c = (c_0, c_1, \dots, c_n)$ and spectral density data φ_k at a finite number of frequencies $\boldsymbol{\theta} := \{\theta\}_{k=1}^N$. See Fig 5.

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

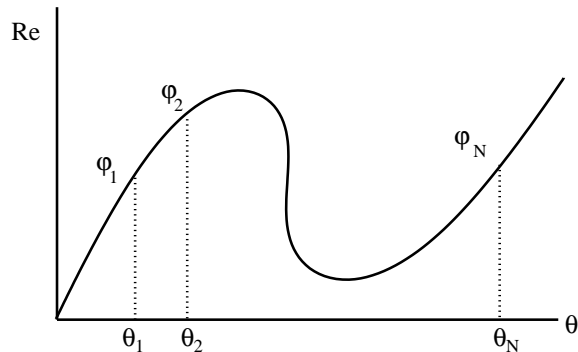


Figure 5: The spectral density data $(\varphi_k)_{k=1}^N$ at each frequency $(\theta)_{k=1}^N$

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 density 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:

$$d_c(c, \hat{c}) := \|c - \hat{c}\|, \quad (18)$$

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

As a distance between a rational spectral density function and spectral density data $\varphi = \{\varphi_k\}_{k=1}^N$, we will use the weighted squares sum and the weighted squares sum of logarithms

$$d_{LS}(\varphi, \phi) := \frac{1}{2} \sum_{k=1}^N w_k \left| 1 - \frac{1}{\varphi_k} \phi_k \right|^2, \quad (19)$$

$$d_{\log}(\varphi, \phi) := \frac{1}{2} \sum_{k=1}^N w_k |\log \varphi_k - \log \phi_k|^2, \quad (20)$$

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

4.2 Optimization for spectral density approximation

Suppose that Φ is taken from the set $\mathcal{P}_n(c)$. Since Φ depends uniquely on both σ and c , we write the element in the set $\mathcal{P}_n(c)$ as $\Phi(\sigma, c)$. Thus, the distance (19) and (20) can be written as functions of vectors σ and c :

$$d_{LS}(\varphi, \phi(\sigma, c)) = \frac{1}{2} \sum_{k=1}^N w_k \left| 1 - \frac{1}{\varphi_k} \frac{\sigma^T M_k \sigma}{\mathbf{a}^T M_k \mathbf{a}} \right|^2, \quad (21)$$

$$d_{\log}(\varphi, \phi(\boldsymbol{\sigma}, c)) = \frac{1}{2} \sum_{k=1}^N w_k \left| \log \varphi_k - \log \frac{\boldsymbol{\sigma}^T M_k \boldsymbol{\sigma}}{\mathbf{a}^T M_k \mathbf{a}} \right|^2, \quad (22)$$

where $\mathbf{a} = \mathbf{h}_c(\boldsymbol{\sigma})$ and M_k is the Toeplitz matrix

$$M_k := \begin{bmatrix} 1 & \cos \theta_k & \cdots & \cos n\theta_k \\ \cos \theta_k & 1 & \cdots & \cos(n-1)\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ \cos n\theta_k & \cos(n-1)\theta_k & \cdots & 1 \end{bmatrix}.$$

Note that the imaginary part of the product $\boldsymbol{\sigma}^T \mathbf{z}(\mathbf{z}^*)^T \boldsymbol{\sigma}$ evaluated at $z = e^{i\theta_k}$ disappears because it is a quadratic form of a skew symmetric matrix.

Remark 5 In the rest of the paper, the analysis will be done using only the distance function d_{LS} since the discussion for d_{\log} is analogous to that for d_{LS} .

Thus, the problem becomes to *find a vector $\boldsymbol{\sigma} \in \mathfrak{S}$ such that the corresponding rational spectral density Φ approximates the spectral density data under some requirements on the covariance sequence.*

In this paper, we consider two different optimization problems:

$$\min_{\boldsymbol{\sigma} \in \mathfrak{S}} \min_{\hat{c} \in \hat{\mathcal{C}}_n} d(\varphi, \phi(\boldsymbol{\sigma}, \hat{c})), \quad (23)$$

$$\min_{\boldsymbol{\sigma} \in \mathfrak{S}} \max_{\hat{c} \in \hat{\mathcal{C}}_n} d(\varphi, \phi(\boldsymbol{\sigma}, \hat{c})), \quad (24)$$

where d can be the least-squares discrepancy d_{LS} and $\hat{\mathcal{C}}_n$ is defined, for a given scalar radius $r > 0$, as

$$\hat{\mathcal{C}}_n := \{\hat{c} \in \mathcal{C}_n^+ : d_c(c, \hat{c}) < r\}. \quad (25)$$

Both problems (23) and (24) are looking for the *best* $\boldsymbol{\sigma}$ such that the corresponding rational spectral density Φ approximates the spectral density data. The difference is the interpretation of the set $\hat{\mathcal{C}}_n$. In (23), $\hat{\mathcal{C}}_n$ is interpreted as an “admissible” error on the covariances to improve the approximation of the spectral density data with respect to $\boldsymbol{\sigma}$. On the other hand, in (24), $\hat{\mathcal{C}}_n$ 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 $\boldsymbol{\sigma}$.

Remark 6 The optimization problem (23) has been already studied from a theoretical point of view in [8], with the distance d being the Kullback-Leibler distance [14]. Using the methodology suggested in this section, the optimization problem studied in [8] can be solved numerically. The problem formulated in [8] is reformulated in our setting:

$$\min_{\boldsymbol{\sigma} \in \mathfrak{S}} \min_{\hat{c} \in \hat{\mathcal{C}}_n^{KL}} d_{KL}(\varphi, \phi(\boldsymbol{\sigma}, \hat{c})), \quad (26)$$

with d_{KL} the discretized Kullback-Leibler distance

$$d_{KL}(\varphi, \phi(\boldsymbol{\sigma}, c)) := \frac{1}{N} \sum_{k=1}^N (\varphi_k \log \varphi_k - \varphi_k \log \phi_k). \quad (27)$$

The positivity of (27) is guaranteed by the condition on the covariances

$$c_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi(e^{i\theta}) d\theta.$$

Thus, the domain constraint $\hat{\mathcal{C}}_n^{KL}$ is equal to $\hat{\mathcal{C}}_n$ in (25) with one extra constraint $\hat{c}_0 = c_0$.

In the next section, we will describe properties of the two problems (23) and (24), and how to solve them numerically.

5 Properties and methodologies of the optimization problems

The optimization problems (23) and (24) can be written in a more compact form as

$$\min_{\boldsymbol{\sigma} \in \mathfrak{S}} \min_{\hat{c} \in \hat{\mathcal{C}}_n} \mathbf{f}(\boldsymbol{\sigma}, \hat{c}), \quad (28)$$

$$\min_{\boldsymbol{\sigma} \in \mathfrak{S}} \max_{\hat{c} \in \hat{\mathcal{C}}_n} \mathbf{f}(\boldsymbol{\sigma}, \hat{c}), \quad (29)$$

where $\mathbf{f} : \mathfrak{S} \times \hat{\mathcal{C}}_n \rightarrow \mathbb{R}$ is defined by

$$\mathbf{f}(\boldsymbol{\sigma}, c) := d(\boldsymbol{\varphi}, \boldsymbol{\phi}(\boldsymbol{\sigma}, c)). \quad (30)$$

5.1 Properties of the optimization problems

The optimization problems (28) and (29) have the following properties:

- The cost functional \mathbf{f} is nonconvex with respect to $(\boldsymbol{\sigma}, c)$.
- The cost functional \mathbf{f} is continuously differentiable on $\mathfrak{S} \times \hat{\mathcal{C}}_n$.
- The domain $\mathfrak{S} \times \hat{\mathcal{C}}_n$ is nonconvex.

In addition, the function $\mathbf{f}(\cdot, c) : \mathfrak{S} \rightarrow \mathbb{R}$ is nonconvex in the set \mathfrak{S} , and so is $\mathbf{f}(\boldsymbol{\sigma}, \cdot) : \hat{\mathcal{C}}_n \rightarrow \mathbb{R}$.

Due to the nonconvexity of the cost functional and the domain, a unique global minimizer is not guaranteed. Therefore, we are trying to find a *local minimizer* in $\mathfrak{S} \times \hat{\mathcal{C}}_n$ by choosing a proper initial point. We will explain how to choose the initial point in Section 5.2.

Due to the differentiability of the cost functional, we can use gradient-based algorithms, such as Newton's method and steepest descent method [21]. This will be explained in Section 5.3.

5.2 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 $(\boldsymbol{\sigma}, c)$ from experimental data, and the estimation of the optimal $(\boldsymbol{\sigma}, c)$ when we are interested in reducing the order of a given spectral density. Next, for the two situations, two different candidates will be suggested as initial points.

In the first situation, the initial $\boldsymbol{\sigma}$ is taken to be $[1, 0, \dots, 0]^T$ which corresponds to the maximum entropy solution. The initial covariances c are equal to the covariances estimated from experimental data. In the second one, the initial $\boldsymbol{\sigma}$ is taken such that $\boldsymbol{\sigma}^T \mathbf{z}(\mathbf{z}^*)^T \boldsymbol{\sigma}$ is the zero polynomial of the spectral density of the balanced truncated 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.

5.3 Algorithms

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

For problem (23), there are two popular algorithms: *Gauss-Newton* and *Levenberg-Marquardt* [21, Chapter 13]. These two methods were originally developed for unconstrained nonlinear least-squares problems. In order to incorporate the constraints $\boldsymbol{\sigma} \in \mathfrak{S}$ and $c \in \hat{\mathcal{C}}_n$, we modify these algorithms in line with [20].

For problem (24), we will solve two optimization problems iteratively:

- In the first step, we solve a min-max problem:

$$\min_{\boldsymbol{\sigma} \in \mathfrak{S}} \max_{k=1, \dots, S} d(\boldsymbol{\varphi}, \boldsymbol{\phi}(\boldsymbol{\sigma}, c^{(k)})) \quad (31)$$

where $\{c^{(k)}\}_{k=1}^S$ are samples of the covariances in the set $\hat{\mathcal{C}}_n$. The problem (31) is solved using a *sequential quadratic programming* method [21]. By solving (31), we are able to find a worst-case $c^{(k)}$ and a suboptimal $\boldsymbol{\sigma}$.

- In the second step, a finite number of covariance sequences $\{\tilde{c}^{(l)}\}_{l=0}^N$ are randomly chosen in the set $\hat{\mathcal{C}}_n$. For a fixed $\boldsymbol{\sigma}$, which is the solution of (31), a covariance sequence $c \in \{\tilde{c}^{(l)}\}_{l=0}^N$ is found that solves a maximization problem

$$\max_{l=1, \dots, N} d(\boldsymbol{\varphi}, \boldsymbol{\phi}(\boldsymbol{\sigma}, \tilde{c}^{(l)})) \quad (32)$$

In this way, the solution c of (32) affects adversely the density approximation for the fixed $\boldsymbol{\sigma}$.

- 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 (31) and (32) are sufficiently small.

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

6 Examples

In this section we will show the usefulness of our approach with the same numerical examples as in Section 2.

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

$$W(z)W(z^{-1}) = c_0 + c_1(z + z^{-1}) + \dots \quad (33)$$

We have tried several methods to reduce the model order, that is, the balanced truncation, and using the covariance (c_0, c_1) computed in (33), our minimization (23) for two different distances: least-squares sum and Kullback-Leibler distance. In the minimization problem (23), the spectral density data $\varphi = \{\varphi_k\}_{k=1}^N$ are computed at the frequencies $\theta = [0, 0.01, 0.02, \dots, \pi]$, and the set $\hat{\mathcal{C}}_n$ is defined such that the distance d_c is:

$$d_c(\hat{c}, c) = \|\hat{c} - c\|_\infty \leq 0.1.$$

Fig. 6 and Fig. 7 show the comparisons of the spectral zero locations and of the frequency responses, respectively. As can be seen from the figures, the least-squares method approximates best the original system.

Remark 7 If the distance in the optimization problem (23) is the Kullback-Leibler distance (27), the *steepest descent* algorithm [21], modified in order to take in consideration the constraints $\sigma \in \mathfrak{S}$ and $c \in \hat{\mathcal{C}}_n$, can be used.

Fig. 7 shows that the Kullback-Leibler (KL) method approximates the original system worse than the balanced truncation (BT) does, even if the final value of the cost function for KL is smaller than that for BT. This behavior can be easily explained. In fact, at each frequency θ_k , each element in the summation (27), can be either positive or negative. Thus, even if the total sum is small, this does not imply that the rational spectral density Φ is close to the spectral density data φ at each sampled point of the frequency θ_k .

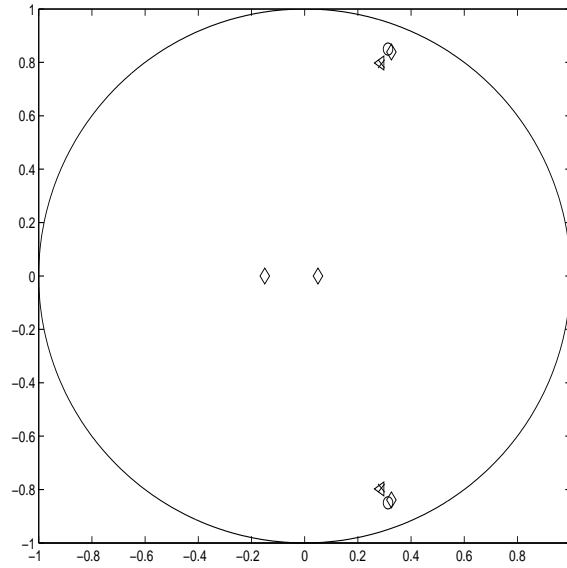


Figure 6: Spectral zero of the given system (\diamond), of balanced truncated model (\times), of the least-squares design (\circ) and the Kullback-Leibler design (\triangleleft).

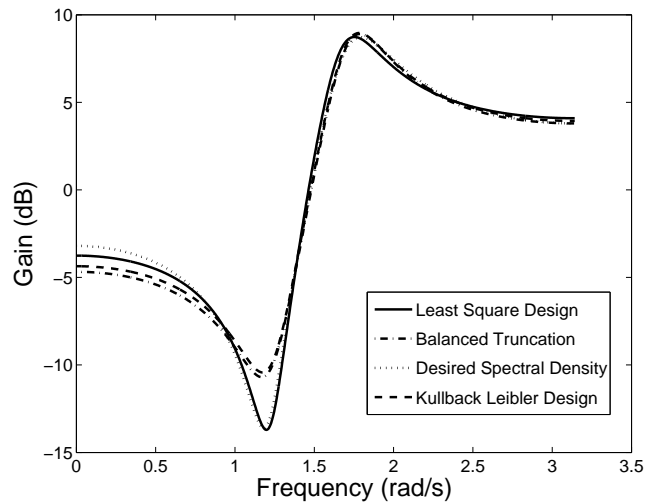


Figure 7: Frequency responses of the given system, of the balanced truncated system, of the least-squares design, and of the Kullback-Leibler design.

6.2 An example in Spectral Estimation

Let us again consider the fifth order system (7). The covariances are computed as explained in Section 2.2. Using the estimated covariances, we solve the min-max problem (24) for the least-squares distance.

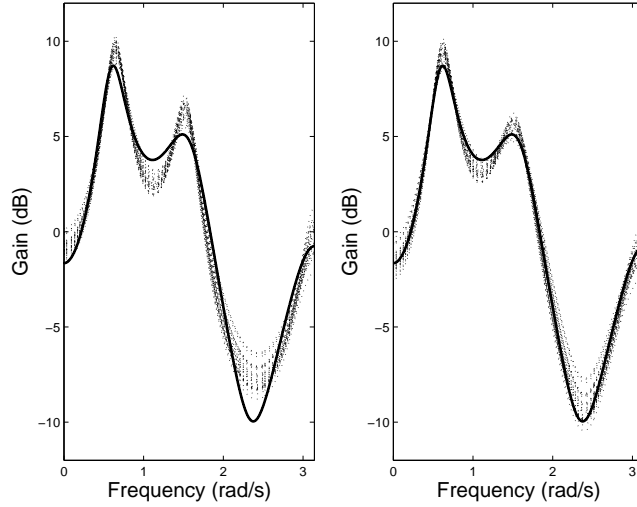


Figure 9: Robustness of the ME solution, dotted line (left) and of the solution of the min-max problem, dotted line (right). In both, left and right part, the solid line represent the “true” system.

$(\sigma_1, \dots, \sigma_6)$ were given by

$$(-0.136, -0.301, -0.228, -0.507, 0.556, 0.429). \quad (34)$$

In Fig. 10 (left), the robustness of design proposed in [3] is shown by choosing

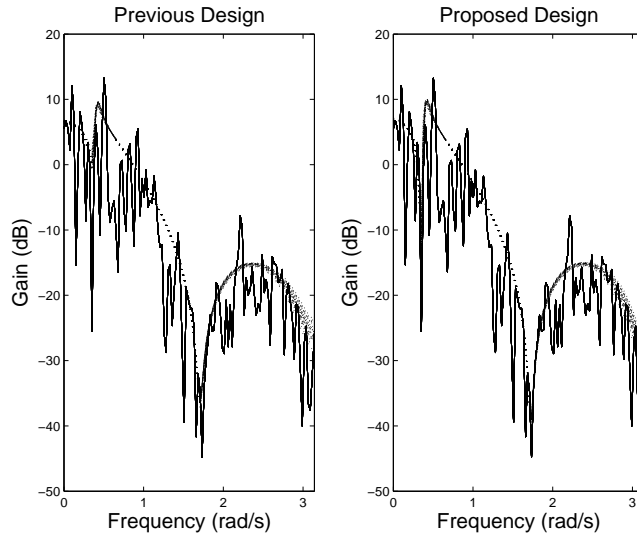


Figure 10: Robustness of the previous design, and of the proposed one.

randomly 30 covariance sequences in the set $\hat{\mathcal{C}}_n$ for a fixed Schur polynomial with coefficients (34), 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 (34) as an initial point, we will solve the optimization problem (24) by the procedure presented in Section 4.2.

Then, the optimization procedure has returned the following coefficients $(\sigma_1, \dots, \sigma_6)$ of a Schur polynomial:

$$(-0.167, -0.369, -0.279, -0.516, 0.657, 0.4445). \quad (35)$$

In Fig. 11, the optimal spectral zeros are plotted together with the zeros of the Schur polynomial with coefficients (34). In Fig. 10 (right), the robustness of the

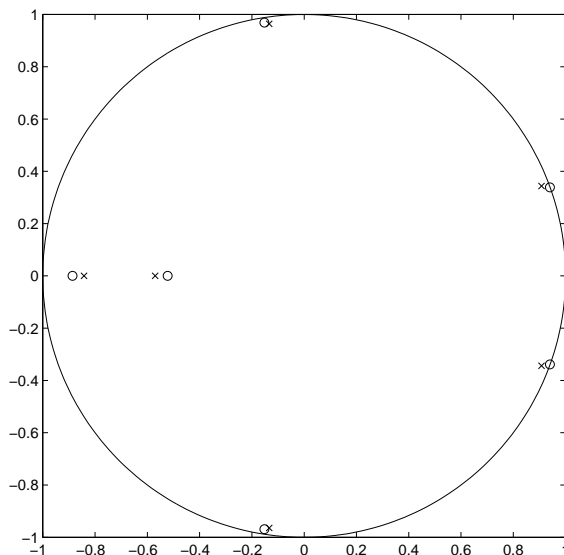


Figure 11: Optimal spectral zeros (o) and the zeros of the Schur polynomial (34).

solution (35) 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. 12, shows that the proposed design approximates the periodogram around the frequencies 0.35 rad/s and 1.73 rad/s better than the previous design in [3].

We remark that, as it has been done in this example, the proposed method can be used for “fine-tuning” of spectral zeros by selecting heuristically designed spectral zeros as initial points of the optimization.

Finally, we would like to remark that the choice of the radius of $\hat{\mathcal{C}}_n$ in the three examples presented in this section is arbitrarily. 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.

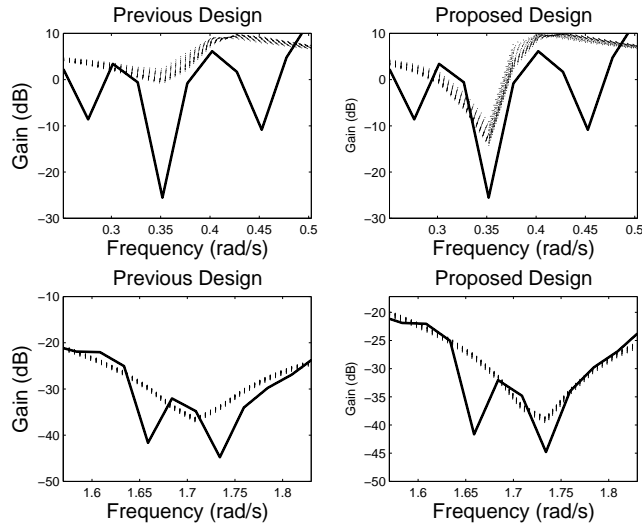


Figure 12: Zoom around the frequencies 0.35 rad/s and 1.73 rad/s of the robustness of the previous design and of the proposed one. Solid lines are the periodograms and dotted lines are the frequency responses of the system of the least-squares design and of the previous design.

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 and solved 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 methods (such as Newton's method and steepest decent 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.

Acknowledgment

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, we would like to thank Dr. P. Enqvist at the Royal Institute of Technology for providing them with experimental data.

A Computation of the gradients

In this appendix, we will derive the expression of gradients of the cost function d in (23) and (24) with respect to $\boldsymbol{\sigma}$ and c . The function d can be expressed in the following form:

$$d(\boldsymbol{\varphi}, \boldsymbol{\phi}(\boldsymbol{\sigma}, c)) := \frac{1}{2} \mathbf{F}(\boldsymbol{\sigma}, c)^T \mathbf{F}(\boldsymbol{\sigma}, c),$$

where $\mathbf{F} : \mathfrak{S} \times \mathcal{C}_n^+ \rightarrow \mathbb{R}^N$ is a vector-valued map and represented as

$$\mathbf{F}(\boldsymbol{\sigma}, c) := [f_1(\boldsymbol{\sigma}, c), \dots, f_N(\boldsymbol{\sigma}, c)]^T. \quad (36)$$

The gradient of the cost functional $d(\boldsymbol{\varphi}, \boldsymbol{\phi}(\boldsymbol{\sigma}, c))$ with respect to $(\boldsymbol{\sigma}, c)$ is computed by

$$\text{grad}(d(\boldsymbol{\varphi}, \boldsymbol{\phi}(\boldsymbol{\sigma}, c))) = [J_{\boldsymbol{\sigma}} \mathbf{F}, J_c \mathbf{F}]^T \cdot \mathbf{F}, \quad (37)$$

where

$$J_{\boldsymbol{\sigma}} \mathbf{F} := \left[\frac{\partial f_1}{\partial \boldsymbol{\sigma}}, \dots, \frac{\partial f_N}{\partial \boldsymbol{\sigma}} \right]^T, \quad (38)$$

$$J_c \mathbf{F} := \left[\frac{\partial f_1}{\partial c}, \dots, \frac{\partial f_N}{\partial c} \right]^T. \quad (39)$$

Remark 8 We will use the following notations: for the gradient of a functional $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with respect to x ,

$$\text{grad}f(x) := \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]^T, \quad x \in \mathbb{R}^n,$$

and for the Jacobian of a functional $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with respect to a vector $x \in \mathbb{R}^n$, $J_x F \in \mathbb{R}^{m \times n}$.

In the case of weighted least-squares sum d_{LS} ,

$$\begin{aligned} f_k(\boldsymbol{\sigma}, c) &:= \sqrt{w_k} \left(1 - \frac{1}{\varphi_k} \frac{\boldsymbol{\sigma}^T M_k \boldsymbol{\sigma}}{\mathbf{a}^T M_k \mathbf{a}} \right), \\ \frac{\partial f_k}{\partial \boldsymbol{\sigma}}(\boldsymbol{\sigma}, c) &= \frac{2\sqrt{w_k}}{\varphi_k \mathbf{a}^T M_k \mathbf{a}} \left[\frac{\boldsymbol{\sigma}^T M_k \boldsymbol{\sigma}}{\mathbf{a}^T M_k \mathbf{a}} \left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\sigma}} \right)^T M_k \mathbf{a} - M_k \boldsymbol{\sigma} \right], \\ \frac{\partial f_k}{\partial c}(\boldsymbol{\sigma}, c) &= \frac{2\sqrt{w_k} \boldsymbol{\sigma}^T M_k \boldsymbol{\sigma}}{\varphi_k (\mathbf{a}^T M_k \mathbf{a})^2} \left(\frac{\partial \mathbf{a}}{\partial c} \right)^T M_k \mathbf{a}. \end{aligned}$$

On the other hand, in the case of weighted least-squares sum of logarithms d_{log} ,

$$\begin{aligned} f_k(\boldsymbol{\sigma}, c) &:= \sqrt{w_k} \left(\log \varphi_k - \log \frac{\boldsymbol{\sigma}^T M_k \boldsymbol{\sigma}}{\mathbf{a}^T M_k \mathbf{a}} \right), \\ \frac{\partial f_k}{\partial \boldsymbol{\sigma}}(\boldsymbol{\sigma}, c) &= 2\sqrt{w_k} \left[\frac{\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\sigma}} \right)^T M_k \mathbf{a}}{\mathbf{a}^T M_k \mathbf{a}} - \frac{M_k \boldsymbol{\sigma}}{\boldsymbol{\sigma}^T M_k \boldsymbol{\sigma}} \right], \\ \frac{\partial f_k}{\partial c}(\boldsymbol{\sigma}, c) &= 2\sqrt{w_k} \frac{\left(\frac{\partial \mathbf{a}}{\partial c} \right)^T M_k \mathbf{a}}{\mathbf{a}^T M_k \mathbf{a}}. \end{aligned}$$

The computation of the vector \mathbf{a} is obtained by the continuation method developed in [1]. Next, the computation of $\partial \mathbf{a} / \partial \boldsymbol{\sigma}$ and $\partial \mathbf{a} / \partial c$ will be explained.

Lemma 1 For a given covariance sequence c , the derivative of the vector $\mathbf{a} = \mathbf{h}_c(\boldsymbol{\sigma})$ with respect to $\boldsymbol{\sigma}$ is computed by

$$\frac{\partial \mathbf{a}}{\partial \boldsymbol{\sigma}} = R(\boldsymbol{\sigma}) \left[R(\mathbf{a})K + \sum_{j=0}^n \mathbf{a}_j R(\mathbf{k}_j) \right]^{-1}, \quad (40)$$

where

- $R(\mathbf{v})$ is, for a vector $\mathbf{v} := [v_0, v_1, \dots, v_n]^T$, the Hankel + Toeplitz operator defined by

$$R(\mathbf{v}) := \begin{bmatrix} v_0 & \cdots & v_n \\ \vdots & \ddots & \\ v_n & & \end{bmatrix} + \begin{bmatrix} v_0 & \cdots & v_n \\ & \ddots & \vdots \\ & & v_0 \end{bmatrix}. \quad (41)$$

- K is derived from the covariance sequence (10) in the following way

$$K := \begin{bmatrix} c_0 & & & \\ 2c_1 & c_0 & & \\ \vdots & \ddots & \ddots & \\ 2c_n & \cdots & 2c_1 & c_0 \end{bmatrix}. \quad (42)$$

- \mathbf{a}_j is the $(j+1)$ -th element of the vector \mathbf{a} and \mathbf{k}_j is the $(j+1)$ -th column of the matrix K .

Proof It was established in [7],[9] that the map \mathbf{h}_c sending $\boldsymbol{\sigma}$ to \mathbf{a} is a diffeomorphism, thus the partial derivative (40) is well defined. Next, we will derive (40). See also [20, Section 3.2].

The nonlinear map \mathbf{h}_c can be expressed as a composition of two maps:

$$\mathbf{h}_c := \mathbf{h}_2 \circ \mathbf{h}_1. \quad (43)$$

The map \mathbf{h}_1 is defined in \mathfrak{S} as

$$\mathbf{h}_1(\boldsymbol{\sigma}) := \frac{1}{2}R(\boldsymbol{\sigma})\boldsymbol{\sigma}. \quad (44)$$

The map \mathbf{h}_2 is defined in \mathcal{D}

$$\mathcal{D} := \left\{ \mathbf{d} := [d_0, d_1, \dots, d_n]^T \in \mathbb{R}^{n+1}, d_0 + \sum_{j=1}^n d_j(z^j + z^{-j}) > 0 \forall |z| < 1 \right\},$$

as the inverse map of

$$\mathbf{g}_2(\mathbf{a}) := R(\mathbf{a})K\mathbf{a}, \quad \mathbf{a} \in \mathfrak{A}, \quad (45)$$

with \mathfrak{A} defined as

$$\mathfrak{A} := \left\{ \mathbf{a} \in \mathfrak{S} : \mathbf{a}^T \mathbf{z} \mathbf{z}^{-T} K \mathbf{a} + \mathbf{a}^T \mathbf{z}^{-1} \mathbf{z}^T K \mathbf{a} > 0, \forall |z| < 1 \right\}.$$

The map \mathbf{g}_2 was proven to be a diffeomorphism in [7], [9], its inverse $\mathbf{h}_2 := \mathbf{g}_2^{-1}$ is well-defined. In [7],[9] it was also proved the diffeomorphism of \mathbf{h}_1 . Thus, the following derivatives are well-defined:

$$\frac{\partial \mathbf{h}_1}{\partial \boldsymbol{\sigma}} = R(\boldsymbol{\sigma}), \quad (46)$$

$$\frac{\partial \mathbf{h}_2}{\partial \mathbf{a}} = \left[R(\mathbf{a})K + \sum_{j=0}^n \mathbf{a}_j R(\mathbf{k}_j) \right]^{-1}. \quad (47)$$

Finally, (40) is obtained from (46) and (47) using the chain rule.

Remark 9 In the next lemma, we will use the following notation

$$Z := \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}, \quad (48)$$

with $Z^0 = I_{n+1}$.

Lemma 2 For a given vector $\boldsymbol{\sigma} \in \mathfrak{S}$, the derivative of $\mathbf{a} = \mathbf{g}_{\boldsymbol{\sigma}}(c)$ with respect to c is computed by

$$\frac{\partial \mathbf{a}}{\partial c} = - \begin{bmatrix} P_0 \\ \vdots \\ P_n \end{bmatrix}^{-1} \begin{bmatrix} Q_0 \\ \vdots \\ Q_n \end{bmatrix} \quad (49)$$

Here $P_i \in \mathbb{R}^{(n+1) \times (n+1)^2}$ and $Q_i \in \mathbb{R}^{(n+1) \times 1}$, $i = 0, \dots, n$ are matrices defined respectively as:

$$P_i = 2(\mathbf{a}^T \otimes I_{n+1})(V_i \otimes I_{n+1}), \quad (50)$$

$$Q_i = (\mathbf{a}^T \otimes I_{n+1})(\nabla V_i) \mathbf{a}, \quad (51)$$

where I_{n+1} is the identity matrix, V_i is computed by

$$V_i = \frac{1}{2} [(Z^i + (Z^i)^T)K + K^T(Z^i + (Z^i)^T)],$$

and ∇ in (51) is a differential operator³ such that

$$\nabla V_i := \frac{1}{2} [((Z^i + (Z^i)^T) \otimes I_{n+1})L + M(Z^i + (Z^i)^T)],$$

³For a matrix-valued functional $V : \mathbb{R}^s \rightarrow \mathbb{R}^n \times \mathbb{R}^m$, the operator ∇ is defined as

$$\nabla V(x) := \begin{bmatrix} \text{grad}_x V_{11}(x) & \cdots & \text{grad}_x V_{1m}(x) \\ \vdots & & \vdots \\ \text{grad}_x V_{n1}(x) & \cdots & \text{grad}_x V_{nm}(x) \end{bmatrix},$$

where grad_x is the gradient with respect to x .

$$L := (I_{n+1} \otimes \mathbf{e}_1) + \sum_{k=1}^n \left((Z^k)^T \otimes 2\mathbf{e}_{k+1} \right),$$

$$M := (I_{n+1} \otimes \mathbf{e}_1) + \sum_{k=1}^n \left(Z^k \otimes 2\mathbf{e}_{k+1} \right),$$

with \mathbf{e}_k , for $k = 1, \dots, n+1$, the canonical basis in \mathbb{R}^{n+1} .

Proof In [6] it was proved that the map \mathbf{g}_σ sending the vector c in \mathbf{a} is a diffeomorphism, thus the derivative (49) is well-defined. Moreover, there exists a relation between the vector \mathbf{a} and the vector σ [4],[5]:

$$\mathbf{a}^T \mathbf{z}(\mathbf{z}^*)^T K \mathbf{a} + \mathbf{a}^T K^T \mathbf{z}(\mathbf{z}^*)^T \mathbf{a} = \sigma^T \mathbf{z}(\mathbf{z}^*)^T \sigma, \quad (52)$$

where K is defined in (42). Identifying coefficients of the same powers, we obtain

$$\begin{bmatrix} 2\mathbf{a}^T K \mathbf{a} \\ \mathbf{a}^T (Z + Z^T) K \mathbf{a} \\ \vdots \\ \mathbf{a}^T (Z^n + (Z^n)^T) K \mathbf{a} \end{bmatrix} = \frac{1}{2} R(\sigma) \sigma, \quad (53)$$

where the matrix Z is defined in (48). Each element of the column vector of the left hand side of (53) is a quadratic form, thus it can be written as

$$\mathbf{a}^T (Z^i + (Z^i)^T) K \mathbf{a} = \mathbf{a}^T S_i(c) \mathbf{a}, \quad \forall i = 0, \dots, n, \quad (54)$$

with $S_i(c)$ the symmetric part of $(Z^i + (Z^i)^T)K$ (the skew-symmetric part is zero for the properties of quadratic form). Inserting (54) in (53) yields

$$\begin{bmatrix} \mathbf{a}^T S_0(c) \mathbf{a} \\ \vdots \\ \mathbf{a}^T S_n(c) \mathbf{a} \end{bmatrix} = \frac{1}{2} R(\sigma) \sigma. \quad (55)$$

Derivating with respect to c the left and right side of (55), we obtain

$$\begin{bmatrix} P_0 \frac{\partial \mathbf{a}}{\partial c} + Q_0 \\ \vdots \\ P_n \frac{\partial \mathbf{a}}{\partial c} + Q_n \end{bmatrix} = 0, \quad (56)$$

with P_i and Q_i defined respectively in (50) and (51). The Jacobian $\partial \mathbf{a} / \partial c$ is computed from (56). The invertibility of the matrix $[P_0, \dots, P_n]^T$ is guaranteed by the diffeomorphic property of the map \mathbf{g}_σ in Theorem 3.

References

- [1] A. Blomqvist, G. Fanizza, and R. Nagamune. Computation of bounded N-P interpolants by solving nonlinear equations. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, pages 4511–4516, Maui, Hawaii, December 2003.

- [2] J.P. Burg. *Maximum entropy spectral analysis*. PhD thesis, Stanford University, Dept. Geophysics, Stanford CA, 1975.
- [3] C. I. Byrnes, P. Enqvist, and A. Lindquist. Cepstral coefficient, covariance lags and pole-zero models for finite data strings. *IEEE Trans. Signal Processing*, 49(4):677–693, April 2001.
- [4] 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.
- [5] C. I. Byrnes, S. V. Gusev, and A. Lindquist. From finite covariance windows to modeling filters: A convex optimization approach. *SIAM Review*, 43(4):645–675, 2001.
- [6] C. I. Byrnes and A. Lindquist. The generalized moment problem with complexity constraint. To be published in *Integral Equations and Operator Theory*.
- [7] C. I. Byrnes and A. Lindquist. On the duality between filtering and Nevanlinna-Pick interpolation. *SIAM J. Contr. and Optimiz.*, 39(3):757–775, 2000.
- [8] C. I. Byrnes and A. Lindquist. The uncertain generalized moment problem with complexity constrained. In M. Xiao W. Kang and C. Borges, editors, *New Trends in Nonlinear Dynamics and Control*, pages 267–278. Springer Verlag, 2003.
- [9] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Matveev. A complete parameterization of all positive rational extensions of a covariance sequence. *IEEE Trans. Automat. Control*, 40(11):1841–1857, November 1995.
- [10] C.I. Byrnes, G. Fanizza, and A. Lindquist. A homotopy continuation solution of the covariance extension equation. In W.P. Dayawansa, A. Lindquist, and Y. Zhou, editors, *New Directions and Applications in Control Theor.* Springer Verlag, 2005.
- [11] U.B Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Trans. Automat. Control*, 29:1097–1100, 1984.
- [12] T. T. Georgiou. *Partial Realization of Covariance Sequences*. PhD thesis, University of Florida, Gainesville, 1983.
- [13] T. T. Georgiou. Realization of power spectra from partial covariance sequences. *IEEE Trans. Acoustics, Speech and Signal Processing*, 35:438–449, 1987.
- [14] T. T. Georgiou and A. Lindquist. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. Information Theory*, 49(11):2910–2917, November 2003.
- [15] J.Mari, A. Dahlén, and A. Lindquist. A covariance extension approach to identification of time series. *Automatica*, 36(3):379–398, 2000.

- [16] A. Lindquist and G. Picci. Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica*, 32(5):709–733, 1996.
- [17] L. Ljung. *System Identification - Theory For the User*. PTR Prentice Hall, Upper Saddle River, N.J., 2 edition, 1999.
- [18] B.C. Moore. Principal component analysis in linear system controllability, observability, and model reduction. *IEEE Trans. Automat. Control*, 26(1):17–32, February 1981.
- [19] C.T. Mullis and R.A. Roberts. Synthesis of minimum roundoff noise fixed point digital filters. *IEEE Trans. Circuits and Systems*, 23(6):551–562, September 1976.
- [20] R. Nagamune and A. Blomqvist. Sensitivity shaping with degree constraint by nonlinear least-squares optimization. *Automatica*, 41(7):1219–1227, July 2005.
- [21] S. Nash and A. Sofer. *Linear and Nonlinear Programming*. McGraw-Hill, 1996.
- [22] G. Obinata and B.D.O. Anderson. *Model Reduction for Control System Design*. Springer-Verlag, London, 2001.
- [23] R.E. Skelton and B.D.O. Anderson. Q-markov covariance equivalent realizations. *Int. J. Control*, 44(5):1477–1490, 1986.
- [24] T. Söderström and P. Stoica. *System Identification*. Prentice Hall, 1989.
- [25] P. Stoica and R. Moses. *Introduction to Spectral Estimation*. Prentice-Hall, 1997.