MATRIX-VALUED NEVANLINNA-PICK INTERPOLATION WITH COMPLEXITY CONSTRAINT: AN OPTIMIZATION APPROACH

A. BLOMQVIST, A. LINDQUIST
and R. NAGAMUNE

REPORT No. 15, 2002/2003, spring

ISSN 1103-467X
ISRN IML-R- -15-02/03- -SE+spring

INSTITUT MITTAG-LEFFLER
THE ROYAL SWEDISH ACADEMY OF SCIENCES

MATRIX-VALUED NEVANLINNA-PICK INTERPOLATION WITH COMPLEXITY CONSTRAINT: AN OPTIMIZATION APPROACH*

ANDERS BLOMQVIST†, ANDERS LINDQUIST†, AND RYOZO NAGAMUNE†

Abstract. Over the last several years a new theory of Nevanlinna-Pick interpolation with complexity constraint has been developed for scalar interpolants. In this paper we generalize this theory to the matrix-valued case, also allowing for multiple interpolation points. We parameterize a class of interpolants consisting of "most interpolants" of no higher degree than the central solution in terms of spectral zeros. This is a complete parameterization, and for each choice of interpolant we provide a convex optimization problem for determining it. This is derived in the context of duality theory of mathematical programming. To solve the convex optimization problem, we employ a homotopy continuation technique previously developed for the scalar case. These results can be applied to many classes of engineering problems, and, to illustrate this, we provide some examples. In particular, we apply our method to a benchmark problem in multivariate robust control. By constructing a controller satisfying all design specifications but having only half the McMillan degree of conventional H^∞ controllers, we demonstrate the advantage of the proposed method.

1. Introduction

Applications of Nevanlinna-Pick interpolation abound in robust control [15, 21, 23, 31, 38, 39, 55, 56], signal processing [4, 5, 6, 9, 28, 37] and maximal power transfer [59] in circuit theory, to mention a few. Since the interpolant has a specific interpretation as a transfer function in all these applications, it is important to restrict its degree, and at the same time satisfy all design specifications. The lack of insight provided by the classical techniques of Nevanlinna-Pick interpolation into questions regarding the degree of various interpolants has therefore been a limiting factor in this approach. In fact, the designer has generally been confined to the so called central solution, or the essentially equivalent solution derived by Nehari approach, the only type of solution for which algorithms have been available.

Over the last several years a new theory of analytic interpolation with complexity constraint has been developed for scalar interpolants [8, 7, 11, 9, 10, 5]. The basic idea is to parameterize complete classes of interpolants of at most a given degree in a smooth fashion, providing tuning parameters for modifying the design without increasing the complexity. This is done in the context of duality theory of mathematical programming, providing convex optimization problems for determining any interpolant belonging to such a class. In this context, new paradigms for spectral estimation [6, 4, 5] and robust controller synthesis [7, 48, 46, 49, 47, 50] have been developed in the single-input/single-output case.

∗ This research was supported by the Swedish Research Council and Institut Mittag-Leffler.

† Division of Optimization and Systems Theory, Department of Mathematics, Royal Institute of Technology, SE 100 44 Stockholm, Sweden.

However, all these results are for scalar interpolants, while the multivariable case is clearly more interesting and important in most of the applications mentioned above. For example, while our design procedures in robust control [7, 48, 46, 49, 47, 50] compare very favorably to H^∞ control methods in the scalar case, it is in the multivariable case that they have a chance to outperform classical control methods in general.

Motivated by this, in this paper we generalize the theory of [7] to the matrix-valued case, also allowing for multiple interpolation points. This generalization introduces new nontrivial and challenging issues, compelling us to take special care in formulating the appropriate complexity constraint. In fact, we parameterize a class of interpolants consisting of "most interpolants" of no higher degree than the central solution in terms of spectral zeros. This is a complete parameterization, and for each choice of interpolant we provide a convex optimization problem for determining it. This is derived in the context of a duality theory, generalizing that of [7, 8]; also see the survey in [9]. To do this, we regard the Nevanlinna-Pick interpolation as a generalized moment problem, to proceed along the lines of [10, 12].

The outline of this paper is as follows. In Section 2 we provide some motivation examples, introducing the reader to matrix interpolation in the context of signal processing and control. Section 3 is a preliminary in which we formulate the matrix-valued interpolation problem, first defining a corresponding class of rational strictly positive real functions with complexity constraint. We reformulate the problem as a generalized moment problem and provide a necessary and sufficient condition for existence of solutions, which we then interpret as a generalized Pick condition. The main theorems are presented in Section 4 and proved in Section 5. Generalizing results in [8, 7, 9, 10, 12] to matrix-valued analytic interpolation theory, we present a smooth, complete parameterization of the set of matrix-valued interpolants with complexity constraint in the context of duality theory of mathematical programming. In fact, to each choice of parameters, there is a pair of dual optimization problems, the optima of which uniquely determine the interpolant. The primal problem amounts to maximizing a generalized entropy gain subject to the interpolation conditions, while the dual problem is a convex optimization problem with a unique minimum. In Section 6, an algorithm for solving the dual problem is provided. Here we generalize to the matrix setting an approach first applied to the covariance extension problem in [22] and then extended in [49, 2] to Nevanlinna-Pick interpolation. Since the dual problem is ill-behaved close to the boundary, we reformulate the optimization problem to eliminate this property. This is done at the expense of global convexity, but the new functional is still locally strictly convex in a neighborhood of a unique minimizing point so that we can solve the problem by a homotopy continuation method. In Section 7, finally, a numerical example in robust control is presented. We consider a popular benchmark problem and show that our design achieves the design specifications with a controller of much lower degree than that of the H^∞ design with weighting functions.

2. Motivating examples from signal processing and control

To justify the problem formulation of this paper, we begin by briefly considering some motivating examples.

2.1. **Multivariate covariance extension.** Suppose that we are given a sequence $C_0, C_1, \dots, C_n$ of matrix-valued covariance lags

$$C_k = \mathrm{E}\{y_{t+k} {y_t}^{\mathsf{T}}\}, \quad k = 0, 1, \dots, n,$$

of some real ℓ-dimensional stationary stochastic process $\{y_t;\ t \in \mathbb{Z}\}$ with the property that the block Toeplitz matrix

$$\begin{bmatrix} C_0 & {C_1}^{\mathsf{T}} & \cdots & {C_n}^{\mathsf{T}} \\ C_1 & C_0 & \cdots & {C_{n-1}}^{\mathsf{T}} \\ \vdots & \vdots & \ddots & \vdots \\ C_n & C_{n-1} & \cdots & C_0 \end{bmatrix}$$

is positive definite. Such covariance lags can be determined from observations of $\{y_t\}$ via an ergodic estimate (see, e.g., [54]). The problem is to estimate the spectral density $\Phi(e^{i\theta}), \theta \in [-\pi, \pi]$, of $\{y_t\}$ by matching the given covariance sequence:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\theta} \Phi(e^{i\theta}) d\theta = C_k, \quad k = 0, 1, \dots, n. \tag{2.1}$$

Often one is more interested in the *minimum-phase*[1] spectral factor of Φ, i.e., a solution V of

$$V(z^{-1})^{\mathsf{T}} V(z) = \Phi(z) \tag{2.2}$$

with no poles and zeros in the closed unit disc. In fact, $V(z^{-1})^{\mathsf{T}}$ represents a filter that shapes white noise into a process $\{y_t\}$ with the spectral density Φ.

The problem of determining a $\Phi(z)$ that is positive on the unit circle and satisfies the finite number of moment equations (2.1) has infinitely many solution. However, for design purposes, we are interested in solutions that are rational of reasonably low degree. A favorite solution is the one that maximizes the entropy gain

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

The shaping filter V corresponding to this *maximum-entropy solution* has the form

$$V(z) = R(z)^{-1},$$

where $R(z)$ is a matrix polynomial of degree at most n whose coefficients are the unique solution of the *normal equations*, which are linear and can be solved by means of a matrix-version of the Levinson algorithm [54]; for some earlier papers, see [57, 58, 43]. Clearly, this V has McMillan degree at most $n\ell$.

As a first step toward generalizing this, one might ask whether there is a solution of the form

$$V(z) = \rho(z) R(z)^{-1}, \tag{2.3}$$

where $\rho(z)$ is an arbitrary scalar polynomial of degree at most n having no zeros in the closed unit disc and the property that $\rho(0) \neq 0$. This is a matrix version of a question answered in the affirmative in [28], the question of uniqueness left open and finally settled in [13]. In this paper, we shall prove that, for each $\rho(z)$, there is one and

[1]This is a somewhat nonstandard use of the term minimum-phase caused by having the the unit disc as the region of analyticity. From a mathematical point of view, the term *outer* might be more appropriate.

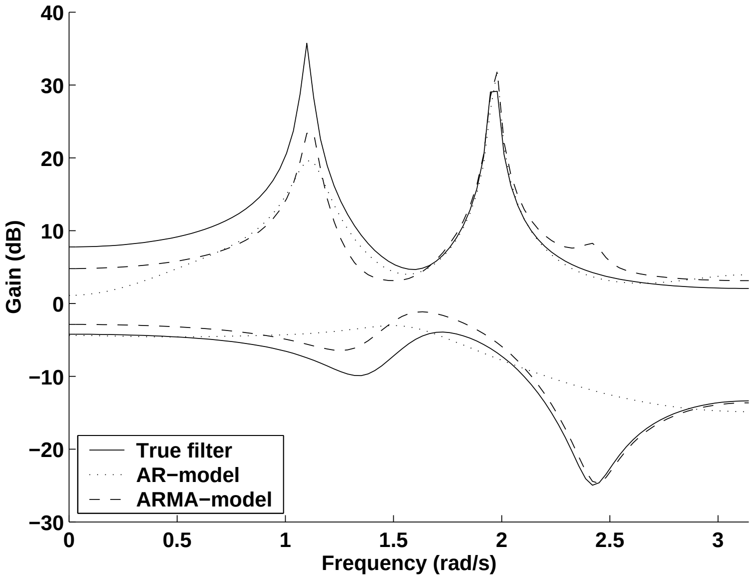

Figure 1: Spectral estimates compared to the true spectral density.

only one $R(z)$ so that Φ defined by (2.2) and (2.3) satisfies the moment conditions (2.1), and it is the Φ maximizing the generalized entropy gain

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |\rho(e^{i\theta})|^2 \log\det\Phi(e^{i\theta})d\theta.$$

This generalizes the corresponding scalar result in [8]. We shall also prove that this parameterization is smooth, forming a family of covariance extensions having a complexity no higher than the maximum entropy solution.

This spectral estimation problem can also be formulated as a matrix-valued Nevanlinna-Pick interpolation problem. In fact, as we shall see in Section 3, a strictly positive real $\ell \times \ell$ matrix-valued function F satisfies the interpolation condition

$$F(0) = \frac{1}{2}C_0, \quad \frac{1}{k!}F^{(k)}(0) = C_k, \quad k = 1, 2, \dots, n,$$

if and only if the spectral density

$$\Phi(z) = F(z) + F(z^{-1})^{\mathsf{T}}$$

satisfies (2.1).

Example 2.1. Consider a two-dimensional stationary stochastic process generated by passing white noise through a known shaping filter. Observing a sample sequence of this process, we want to recover the true shaping filter from a finite window $C_0, C_1, \dots, C_n$ of 2×2 covariance lags obtained from this observed data via ergodic estimates, while restricting the model order. The singular values of the estimated spectral densities for two different solutions are plotted in Figure 1, together with those of the true spectral density. The maximum-entropy solution, i.e., the AR-model determined by the matrix-version of the Levinson algorithm is depicted with a dotted line. By choosing the tuning-parameter polynomial $\rho(z)$ appropriately, we obtain instead the ARMA model, depicted with a dashed line. Note that this method also works for generic data. Hence, the existence of a "true model" is not required.

2.2. **Multivariable sensitivity shaping.** Let P be a linear control system with a vector-valued input u and a vector-valued output y, having a rational transfer function $P(s)$ with unstable poles and nonminimum-phase zeros; these are the poles and zeros, including multiplicities, of $P(s)$ that are located in the right half plane $\{s : \text{Re}\, s > 0\}$. We want to design a compensator C of low complexity so that the closed-loop system depicted in Figure 2 is internally stable, attenuates the effect of the disturbance d, tracks the reference signal r, and reduces the effect of the noise n.

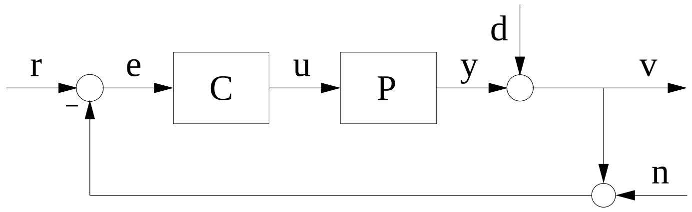

Figure 2: A feedback system.

This problem is standard in the robust control literature (see, e.g., [23, 21, 60]). Internal stability is achieved by requiring that the *sensitivity function*, i.e., the closed-loop transfer function

$$S(s) = \left[I + P(s)C(s)\right]^{-1} \tag{2.4}$$

from the disturbance d to the output v, is stable (all poles in the open left half plane) and satisfies certain interpolation conditions, to stated below. Substituting the Youla-parameterization into (2.4) yields a model matching form:

$$S(s) = T_1(s) - T_2(s)Q(s)T_3(s), \tag{2.5}$$

where T_j, $j = 1, 2, 3$ and Q are stable rational matrices with Q arbitrary. To avoid some technical complications and simplify notation, let us assume that the plant P is square and full rank (i.e, $\det P(s) \not\equiv 0$). Then both T_2 and T_3 are square and full rank.

Now, the (transmission) zeros of T_2 and T_3 are located at the zeros respectively the poles of the plant P. By inner-outer factorizations $T_2 = \Theta_2\tilde{T}_2$ and $T_3 = \tilde{T}_3\Theta_3$, respectively, the nonminimum-phase zeros of the plant are thus transferred to the inner function Θ_2 and the unstable poles to the inner function Θ_3. Moreover, the outer factor $\tilde{T}_2$ contains relevant information about "relative degree" of P. In particular, $\tilde{T}_2(\infty)$ has the same rank as $P(\infty)$. Then, following the procedure in [16], we define $\tilde{S} := \phi\Theta_2^* S\Theta_3^*$ and $\tilde{T}_1 := \phi\Theta_2^* T_1\Theta_3^*$, where $\phi := \det\Theta_2 \det\Theta_3$. Hence (2.5) can be transformed into

$$\tilde{S}(s) = \tilde{T}_1(s) - \phi(s)\tilde{T}_2(s)Q(s)\tilde{T}_3(s), \quad \|S\|_\infty = \|\tilde{S}\|_\infty, \tag{2.6}$$

where ϕ is a scalar inner function having zeros at the unstable poles and zeros of P. If these poles and zeros, denoted by $s_0, s_1, \dots, s_n$, are distinct and $P(\infty)$ has full rank, the interpolation conditions required for internal stability become

$$\tilde{S}(s_k) = \tilde{T}_1(s_k), \quad k = 0, 1, \dots, n, \tag{2.7}$$

whereas any multiple point has to be handled in a separate way. If s_k is an interpolation point of multiplicity ν so that $s_k = s_{k+1} = \cdots = s_{k+\nu-1}$, then the equations in (2.7) corresponding to $s_{k+1} = \cdots = s_{k+\nu-1}$ are replaced by

$$\tilde{S}^{(j)}(s_k) = \tilde{T}_1^{(j)}(s_k), \ j = 1, \dots, \nu - 1. \tag{2.8}$$

If $P(\infty)$ is rank deficient, we also need to add interpolation conditions at infinity to ensure that the controller is proper. To see this, recall that $P(\infty)$ has the same rank as $\tilde{T}_2(\infty)$. Therefore, if $P(\infty)$ is rank deficient, then $v^{\mathsf{T}}\tilde{T}_2(\infty) = 0$ for some v, and hence, in view of (2.6), we have the interpolation condition

$$v^{\mathsf{T}}\tilde{S}(\infty) = v^{\mathsf{T}}\tilde{T}_1(\infty). \tag{2.9}$$

If $\tilde{T}_2$, and thus P, is strictly proper, this interpolation condition becomes

$$\tilde{S}(\infty) = U_0 := \tilde{T}_1(\infty). \tag{2.10}$$

More generally, if in addition the first $k-1$ Markov parameters are zero, i.e., $A_1 = \cdots = A_{k-1} = 0$ in the expansion

$$\tilde{T}_2(s^{-1}) = A_1 s + A_2 s^2 + A_3 s^3 + \cdots, \tag{2.11}$$

and A_{k+1} is full rank, a similar argument shows that

$$\frac{d^j}{ds^j}\tilde{S}(s^{-1})\Big|_{s=0} = U_j := \frac{d^j}{ds^j}\tilde{T}_1(s^{-1})\Big|_{s=0}, \quad j = 1, \dots, k-1, \tag{2.12}$$

and

$$v^{\mathsf{T}}\frac{d^k}{ds^k}\tilde{S}(s^{-1})\Big|_{s=0} = v^{\mathsf{T}}\frac{d^k}{ds^k}\tilde{T}_1(s^{-1})\Big|_{s=0} \tag{2.13}$$

for any v such that $v^{\mathsf{T}}A_k = 0$.

We would like to express all these conditions as interpolation conditions involving some analytic function and its derivatives. To this end, introduce the modified sensitivity function

$$Z(s) := \tilde{S}(s^{-1}), \tag{2.14}$$

which has the same analyticity properties as S (and as $\tilde{S}$), i.e., Z is analytic in the right half of the complex plane. Then, to avoid tangential conditions, we replace conditions (2.10), (2.12) and (2.13) by

$$Z^{(j)}(0) = U_j, \quad j = 0, 1, \dots, k. \tag{2.15}$$

Likewise, (2.7) becomes

$$Z(s_k^{-1}) = \tilde{T}_1(s_k), \tag{2.16}$$

whereas (2.8) corresponds to easily computed but somewhat more complicated expressions for $Z^{(j)}(s_k^{-1})$, $j = 1, \dots, \nu - 1$.

Remark 2.2. These interpolation conditions in terms of Z are sufficient but may not be necessary. In fact, the tangential conditions (2.9) and (2.13) have been allowed to hold in all directions v. The reason for this is that tangential interpolation is not covered by the theory developed in this paper.

Remark 2.3. In our problem formulation, to be given in Section 3, we do not allow for interpolation points on the boundary of the analyticity region. Therefore we shall move the interpolation point $s = 0$ in (2.15) slightly into the open right half plane.

Next, we turn to disturbance attenuation and reference tracking, which are achieved by bounding the H^∞ norm of the sensitivity function, i.e.,

$$\|S\|_\infty = \|Z\|_\infty < \gamma. \tag{2.17}$$

The lowest such bound, i.e., the infimum of $\|Z\|_\infty$ over all stable Z satisfying the interpolation conditions, will be denoted by γ_{opt}. There are optimal solutions achieving this bound, and their largest singular values are uniform over the spectrum. However, in general one would like to shape the sensitivity function to obtain low sensitivity in designated part of the spectrum, which, due to the water-bed effect [52], is done at the expense of higher sensitivity in some other part of the spectrum. To achieve this, it is customary to use weighting functions, which however could increase the degree of the sensitivity function considerably, and hence the compensator.

However, we prefer sensitivity functions of low complexity, and therefore we would like to avoid weighting functions. To this end and to allow for greater design flexibility, we consider suboptimal solutions, of which there are infinitely many. Given some $\gamma > \gamma_{\text{opt}}$, we consider the whole class of stable Z satisfying the required interpolation conditions and some complexity constraint. In this class we would like to choose the one that best satisfies the additional specifications of sensitivity shaping. In this paper, we shall give a smooth, complete parameterization of such a class.

To bring this problem in conformity with the problem formulation in Section 3, we transform first the interpolation points in the right half plane to $z_0, z_1, \dots, z_n$ in the unit circle, via the linear fractional transformation $z = (s-1)(s+1)^{-1}$, and then the function Z to

$$F(z) := \left[\gamma I - Z\left(\frac{1+z}{1-z}\right)\right]\left[\gamma I + Z\left(\frac{1+z}{1-z}\right)\right]^{-1}.$$

For each Z satisfying (2.17), the new function F is analytic in the unit disc and has the property that $F(e^{i\theta}) + F(e^{-i\theta})^{\mathsf{T}} > 0$ for all θ. Let us call such a function a (matrix-valued) *Carathéodory function.* The problem is then reduced to finding a rational Carathéodory function F that has low complexity and satisfies the corresponding interpolation condition

$$F(z_k) = W_k \tag{2.18}$$

for each k such that z_k has multiplicity one and

$$\frac{1}{j!}F^{(j)}(z_k) = W_{k+j}, \quad j = 0, 1, \dots, \nu - 1, \tag{2.19}$$

whenever z_k has multiplicity ν and $z_k = z_{k+1} = \cdots = z_{k+\nu-1}$. It is straightforward, but tedious in the multiple-point case, to determine the interpolation values $W_0, W_1, \dots, W_n$.

Example 2.4. To illustrate the design flexibility of our approach, we consider an example in control, namely the double inverted pendulum depicted in Figure 3. The

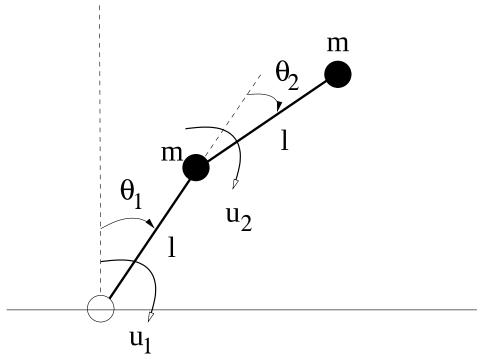

Figure 3: A double inverted pendulum.

linearized model for $m = 1$ (kg) and $l = 1$ (m) is given in [17, p. 37] as

$$\begin{aligned} \dot{x} &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ g & 0 & -g & 0 \\ 0 & 0 & 0 & 1 \\ -g & 0 & 3g & 0 \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 1 & -2 \\ 0 & 0 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \\ y &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} x, \end{aligned}$$

where $x := \begin{bmatrix} \theta_1 & \dot{\theta}_1 & \theta_2 & \dot{\theta}_2 \end{bmatrix}^{\mathsf{T}}$ is the state. This is our plant P. The goal is to design a stabilizing controller C which is robust against low-frequency disturbances d and high-frequency noise n for a zero reference signal ($r = 0$). (See Figure 2.) The plant transfer function has two unstable real poles and is of relative degree two, thus yielding four interpolation conditions.

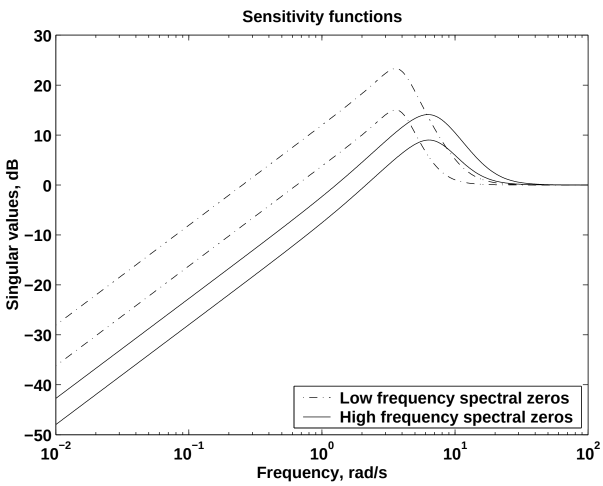

Figure 4: Frequency responses for various tunings of S.

Using the methods of this paper, we can now compute an arbitrary strictly proper controller in a class of controllers of degree at most eight, satisfying the specifications, by choosing the tuning parameters appropriately. Figure 4 shows the (singular-value) frequency responses of two sensitivity functions in this class. One, plotted with dashed-dotted lines, gives a small bandwidth but large robustness against measurement noise, whereas the other, plotted with solid lines, provides a large bandwidth and lower peak gain but a small robustness to noise. Therefore, using the methods of this paper, the controller with the appropriate frequency response can be determined by tuning certain design parameters to satisfy the specifications.

3. The interpolation problem

To formulate the interpolation problem we need first to define a class of positive real functions of low complexity.

3.1. **The class $\mathcal{F}_+(n)$.** An $\ell \times \ell$ matrix-valued, proper, rational function F that is analytic in the closed unit disc $\overline{\mathbb{D}} := \{z \in \mathbb{C} : \ |z| \leq 1\}$ is called *strictly positive real* if the spectral density function

$$\Phi(e^{i\theta}) := \Re\{F(e^{i\theta})\} \tag{3.1}$$

is positive definite for all $\theta \in [-\pi, \pi]$. Here,

$$\Re\{F(z)\} := \frac{1}{2}\left[F(z) + F^*(z)\right], \text{ where } F^*(z) := \overline{F(\bar{z}^{-1})}^{\mathsf{T}},$$

is the Hermitian generalization of the real part in the scalar case. Let $\mathcal{C}_+$ be the class of all such functions. If F belongs to $\mathcal{C}_+$, then so does F^{-1}. In particular, F is *outer*, i.e., all its poles and zeros are located in $\overline{\mathbb{D}}^c$, the complement of $\overline{\mathbb{D}}$.

Strictly positive real functions abound in control, circuit theory and signal processing, where they often represent transfer functions of filters or closed-loop control systems. Since design limitations require such devises to be of bounded complexity, the class $\mathcal{C}_+$ needs to be restricted to accommodate appropriate complexity constraints. Typically, the McMillan degree needs to be bounded.

To this end, first note that, to each $F \in \mathcal{C}_+$, there corresponds an outer $\ell \times \ell$ matrix-valued function V such that

$$V^*(z)V(z) = \Phi(z) := \Re\{F(z)\}, \tag{3.2}$$

which is unique modulo an orthogonal transformation. Determining V from F is a spectral factorization problem, which can be solved by determining the stabilizing solution of an algebraic Riccati equation (see, e.g., [14]). Conversely, if

$$V(z) = zC(I - zA)^{-1}B + D \tag{3.3}$$

is any minimal realization of V, appealing to the equations of the Kalman-Yakubovich-Popov Lemma, there is a unique F satisfying (3.2), and it is given by

$$F(z) = 2z(B^*XA + D^*C)(I - zA)^{-1}B + B^*XB + D^*D, \tag{3.4}$$

where X is the unique solution to the Lyapunov equation

$$X = A^*XA + C^*C. \tag{3.5}$$

Moreover, V is a proper rational function of the same McMillan degree as F, and so is the inverse V^{-1}.

Let the polynomial ρ be the least common denominator of all entries in V^{-1}. Then there is a matrix polynomial R of the same degree as ρ such that $V^{-1} = R/\rho$, and consequently

$$V(z) = \rho(z)R(z)^{-1}. \tag{3.6}$$

In this representation, the degree $r := \deg \rho$ is uniquely determined by F; to emphasize this we write $r(F)$. Now, define the class

$$\mathcal{F}_+(n) := \{F \in \mathcal{C}_+ \mid r(F) \leq n\}. \tag{3.7}$$

All functions $F \in \mathcal{F}_+(n)$ have McMillan degree at most ℓn, but, although this is a nongeneric situation, there are $F \in \mathcal{C}_+$ of McMillan degree at most ℓn that do not belong to $\mathcal{F}_+(n)$. In fact, the standard observable (standard reachable) realization of V^{-1} has dimension ℓr (see, e.g., [3, p. 106]), and consequently V^{-1}, and hence F, has McMillan degree at most ℓr. Moreover, the standard observable realization may not be minimal, so there is a thin set of $F \in \mathcal{C}_+$ of McMillan degree at most ℓn for which $r(F) > n$.

3.2. **Problem formulation.** Suppose that we are given a set

$$\mathcal{Z} := \{z_0, z_1, \dots, z_n\} \subset \mathbb{D} \tag{3.8}$$

of $n + 1$ *interpolation points* in the open unit disc $\mathbb{D}$. These points need not be distinct, but, if a certain number is repeated, it occurs in sequence. We say that z_k has *multiplicity* ν if $z_k = z_{k+1} = \cdots = z_{k+\nu-1}$ and no other point takes this value. Moreover, suppose we have a set of $n + 1$ matrix-valued *interpolation values*

$$\mathcal{W} := \{W_0, W_1, \dots, W_n\} \subset \mathbb{C}^{\ell\times\ell}. \tag{3.9}$$

We assume for convenience that $z_0 = 0$ and that W_0 is real and symmetric.

Now, consider the problem to find a function $F \in \mathcal{F}_+(n)$ that satisfies the interpolation condition

$$F(z_k) = W_k \tag{3.10}$$

for each k such that z_k has multiplicity one and

$$\frac{1}{j!}F^{(j)}(z_k) = W_{k+j}, \quad j = 0, 1, \dots, \nu - 1, \tag{3.11}$$

whenever z_k has multiplicity ν and $z_k = z_{k+1} = \cdots = z_{k+\nu-1}$.

This is a matrix-valued Nevanlinna-Pick interpolation problem with a nonclassical complexity constraint, namely the condition that the interpolant F must belong to the set $\mathcal{F}_+(n)$. In the scalar case $\ell = 1$, this is a degree constraint, and the problem has been studied in [6, 7, 10, 11, 29, 30]. In the present multivariable setting, this complexity constraint is not merely a degree constraint, as pointed out above. In fact, although all $F \in \mathcal{F}_+(n)$ have degree at most ℓn, $\mathcal{F}_+(n)$ does not contain all such functions.

This problem could be reformulated as a generalized moment problem. To see this, note that, by the matrix version of the Herglotz Theorem [18], any $F \in \mathcal{F}_+(n)$ could be represented as

$$F(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z}\Phi(e^{i\theta})d\theta, \tag{3.12}$$

where Φ is given by (3.1). Since therefore

$$\frac{1}{j!}F^{(j)}(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{2e^{i\theta}}{(e^{i\theta}-z)^{j+1}}\Phi(e^{i\theta})d\theta, \quad j = 1, 2, \dots,$$

the interpolation conditions (3.10) and (3.11) can be combined to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\alpha_k(e^{i\theta})\Phi(e^{i\theta})d\theta = W_k, \quad k = 0, 1, \dots, n, \tag{3.13}$$

where α_k is defined as

$$\alpha_k(z) = \frac{z+z_k}{z-z_k} \tag{3.14}$$

when z_k has multiplicity one, and as

$$\alpha_k(z) = \frac{z+z_k}{z-z_k}, \qquad \alpha_{k+j}(z) = \frac{2z}{(z-z_k)^{j+1}}, \quad j = 1, 2, \dots, \nu-1, \tag{3.15}$$

when z_k has multiplicity ν and $z_k = z_{k+1} = \dots = z_{k+\nu-1}$. In particular, since $z_0 = 0$, $\alpha_0 = 1$. Consequently, the Nevanlinna-Pick interpolation problem with complexity constraint formulated above is equivalent to finding an $F \in \mathfrak{F}_+(n)$ satisfying (3.13).

3.3. **A necessary and sufficient condition for existence of solutions.** Clearly the problem posed above does not have a solution for all choices of $\mathcal{W}$. Next, we shall therefore determine what conditions need to be imposed on the interpolation values $W_0, W_1, \dots, W_n$. To this end, we first introduce the class $\mathfrak{Q}(\ell, n)$ of $\ell \times \ell$ matrix-valued generalized pseudo-polynomials

$$Q(z) = \Re\left\{\sum_{k=0}^{n} Q_k\alpha_k(z)\right\} \tag{3.16}$$

with coefficients $Q_k \in \mathbb{C}^{\ell\times\ell}$ and Q_0 real and symmetric, and then we define the subset

$$\mathfrak{Q}_+(\ell, n) := \left\{Q \in \mathfrak{Q}(\ell, n) \mid Q(e^{i\theta}) > 0 \text{ for all } \theta \in [-\pi, \pi]\right\} \tag{3.17}$$

consisting of those $Q \in \mathfrak{Q}(\ell, n)$ that are positive on the unit circle.

Definition 3.1. Given the interpolation points $\mathcal{Z}$, the sequence $\mathcal{W}$ of interpolation values is *positive* if

$$\mathrm{Re}\left\{\sum_{k=0}^{n}\mathrm{tr}(Q_kW_k)\right\} > 0, \tag{3.18}$$

for all matrix sequences $Q_0, Q_1, \dots, Q_n$ such that the pseudo-polynomial Q defined by (3.16) belongs to $\mathfrak{Q}_+(\ell, n)$. Let $\mathfrak{W}_+(\ell, n)$ be the class of all such positive sequences. Here $\mathrm{tr}\{A\}$ denotes the trace of the square matrix A.

Theorem 3.1. *There exists an $F \in \mathfrak{F}_+(n)$ satisfying the interpolation condition* (3.13) *if and only if $\mathcal{W}$ is positive.*

The proof that positivity of $\mathcal{W}$ is necessary is classical. To see this, just note that, by the calculation of Proposition 3.2, (3.13) implies that

$$\mathrm{Re}\left\{\sum_{k=0}^{n}\mathrm{tr}(Q_kW_k)\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\{Q(e^{i\theta})\Phi(e^{i\theta})\}d\theta, \tag{3.19}$$

which is positive whenever $Q \in \mathcal{Q}_+(\ell, n)$. In Section 5 we shall prove that this condition is also sufficient.

Now, it will be useful to represent (3.18) in terms of the inner product

$$\langle A, B\rangle := \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr} A^*(e^{i\theta})B(e^{i\theta})d\theta \tag{3.20}$$

between two $\ell \times \ell$ matrix-valued $\mathcal{L}^2$ functions A and B.

Proposition 3.2. *Let $W : \mathbb{T} \to \mathbb{C}^{\ell\times\ell}$ be an arbitrary function defined on the unit circle $\mathbb{T}$ and satisfying the moment condition*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha_k(e^{i\theta})W(e^{i\theta})d\theta = W_k, \quad k = 0, 1, \dots, n. \tag{3.21}$$

Then, if Q is given by (3.16),

$$\operatorname{Re}\left\{\sum_{k=0}^{n} \operatorname{tr}(Q_k W_k)\right\} = \langle Q, W\rangle. \tag{3.22}$$

Proof. Given any W defined as in the proposition, we obtain

$$\begin{aligned}\langle Q, W\rangle &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{tr}(Q(e^{i\theta})W(e^{i\theta}))d\theta \\ &= \operatorname{Re}\sum_{k=0}^{n} \operatorname{tr}\left(Q_k \frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha_k(e^{i\theta})W(e^{i\theta})d\theta\right) \\ &= \operatorname{Re}\left\{\sum_{k=0}^{n} \operatorname{tr}(Q_k W_k)\right\},\end{aligned}$$

which establishes (3.22). □

3.4. **The generalized Pick condition.** The positivity condition in Theorem 3.1 is a generalized Pick condition. To see this, let $\Gamma(z)$ be any minimum-phase solution of the spectral factorization problem

$$\Gamma(z)\Gamma^*(z) = Q(z). \tag{3.23}$$

Then, introducing the vector of Cauchy type kernels

$$G(z) := \begin{bmatrix} g_0^*(z) & g_1^*(z) & \cdots & g_n^*(z)\end{bmatrix}, \tag{3.24}$$

where

$$g_k(z) = \frac{1}{2}\big(\alpha_k(z) + 1\big)$$

for those k for which α_k is given by (3.14) and

$$g_k(z) = \frac{1}{2}\alpha_k(z)$$

for all other k, $\Gamma(z)$ has a representation

$$\Gamma(z) = \big(G(z) \otimes I_\ell\big)\mathbf{\Gamma} \tag{3.25}$$

for some matrix $\mathbf{\Gamma} \in \mathbb{C}^{\ell(n+1)\times\ell}$, where $A \otimes B$ is the Kronecker product of A and B. Now, let W be defined as in Proposition 3.2. Then (3.23) yields

$$\langle Q, W\rangle = \langle \Gamma, W\Gamma\rangle = \operatorname{tr}\{\mathbf{\Gamma}^*\mathbf{\Pi}\mathbf{\Gamma}\}, \tag{3.26}$$

where $\mathbf{\Pi}$ is the generalized Pick matrix

$$\mathbf{\Pi} := \frac{1}{2\pi}\int_{-\pi}^{\pi} \big(G^*(e^{i\theta}) \otimes I_\ell\big) W(e^{i\theta}) \big(G(e^{i\theta}) \otimes I_\ell\big) d\theta. \tag{3.27}$$

Hence we have the following corollary of Theorem 3.1.

Corollary 3.3. *The sequence $\mathcal{W}$ is positive if and only if the Pick matrix* (3.27) *is positive definite.*

The Pick matrix $\mathbf{\Pi}$ can be computed in terms of interpolation data. In fact, when z_k has multiplicity ν and $z_k = z_{k+1} = \cdots = z_{k+\nu-1}$, we have

$$\frac{1}{j!}F^{(j)}(z_k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} g_{k+j}(e^{i\theta})F(e^{i\theta})d\theta, \quad j = 0, 1, \ldots, \nu - 1,$$

for any function F that is analytic in the unit disc $\mathbb{D}$. Using this Cauchy integral formula, a straight-forward, but tedious, calculation yields

$$\mathbf{\Pi} = \frac{1}{2}\left[\boldsymbol{W}(\boldsymbol{S} \otimes I_\ell) + (\boldsymbol{S} \otimes I_\ell)\boldsymbol{W}^*\right], \tag{3.28}$$

where $\boldsymbol{S}$ is the Gramian

$$\boldsymbol{S} := \frac{1}{2\pi}\int_{-\pi}^{\pi} G^*(e^{i\theta})G(e^{i\theta})d\theta, \tag{3.29}$$

and $\boldsymbol{W}$ is a block diagonal matrix consisting of one block

$$\begin{bmatrix} W_k & & \\ \vdots & \ddots & \\ W_{k+\nu-1} & \cdots & W_k \end{bmatrix}$$

for each distinct point in $\mathcal{Z}$ taken in order. The Gramian (3.29) can be determined by solving the Lyapunov equation

$$\boldsymbol{S} - \boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^* = \boldsymbol{b}\boldsymbol{b}^{\mathsf{T}}, \tag{3.30}$$

where $\boldsymbol{A}$ is a block diagonal matrix formed from the $\nu \times \nu$ blocks

$$A_k := \begin{bmatrix} z_k & & & \\ 1 & z_k & & \\ & \ddots & \ddots & \\ & & 1 & z_k \end{bmatrix},$$

and $\boldsymbol{b}$ is a column vector of ones and zeros in which the ones occur for those k for which (3.14) holds.

Specializing to the case when all interpolation points have multiplicity one, we obtain the classical Pick matrix

$$\mathbf{\Pi} = \frac{1}{2}\left[\frac{W_i + W_j^*}{1 - z_i\bar{z}_j}\right]_{i,j=0}^{n}.$$

On the other hand, when there is only one interpolation point with multiplicity $n+1$ located at the origin, as in the classical Carathéodory extension problem, the Pick matrix is the block Toeplitz matrix

$$\mathbf{\Pi} = \frac{1}{2}\begin{bmatrix} W_0 + W_0^* & W_1^* & \cdots & W_n^* \\ W_1 & W_0 + W_0^* & \cdots & W_{n-1}^* \\ \vdots & \vdots & \ddots & \vdots \\ W_n & W_{n-1} & \cdots & W_0 + W_0^* \end{bmatrix}.$$

See, e.g., [19, 20].

4. Main theorems

To motivate the approach taken in this paper, we first consider the special case when $z_0 = z_1 = \cdots = z_n = 0$, i.e.,

$$\alpha_0(z) = 1, \quad \alpha_k(z) = 2z^{-k},\ k = 1, \dots, n,$$

which is of particular interest in signal processing and identification. In this case the generalized Pick condition reduces to a Toeplitz condition, as described above. In particular, the interpolant that maximizes the entropy gain

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\det\Phi(e^{i\theta})d\theta, \tag{4.1}$$

is the maximum entropy solution discussed in Section 2. Like $\mathcal{W}$, the *cepstral coefficients* [51],

$$c_k := \frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha_k(e^{i\theta})\log\det\Phi(e^{i\theta})d\theta, \quad k = 0, 1, \dots, n, \tag{4.2}$$

can be observed. In the scalar case $\ell = 1$, it was noted in [4, 5] that the entropy gain (4.1) is precisely the zeroth cepstral coefficient c_0 and that the cepstral coefficients (4.2) together with the covariance data $\mathcal{W}$ form local coordinates of $\mathcal{F}_+(n)$. This observation led to maximizing linear combinations of the cepstral coefficients instead.

In this paper we shall apply the same strategy to the multivariable Nevanlinna-Pick problem when $\alpha_0, \alpha_1, \dots, \alpha_n$ are given by (3.14) and (3.15). Accordingly, we consider the problem of maximizing some linear combination

$$\mathrm{Re}\left\{\sum_{k=0}^{n} p_k c_k\right\} \tag{4.3}$$

of the coefficients (4.2), which, in this more general setting, will be referred to as the *generalized cepstral coefficients*. Introducing the generalized pseudo-polynomial

$$P(z) := \Re\left\{\sum_{k=0}^{n} p_k\alpha_k(z)\right\}, \tag{4.4}$$

(4.3) can be written as the *generalized entropy gain*

$$\mathbb{I}_P(\Phi) := \frac{1}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta})\log\det\Phi(e^{i\theta})d\theta, \tag{4.5}$$

which we want to maximize over the class $\mathcal{S}_+^{\ell\times\ell}$ of (not necessarily rational) bounded, coercive spectral densities Φ, i.e., bounded Φ such that Φ^{-1} is also bounded. Just as in [4, 5] we must require $P(z)$ to be positive on the unit circle, i.e., $P \in \mathcal{Q}_+(1, n)$, in order for a maximum of $\mathbb{I}_P(\Phi)$ to exist. In fact, the following theorem establishes a complete parameterization of all interpolants $F \in \mathcal{F}_+(n)$ in terms of the generalized pseudo-polynomial $P \in \mathcal{Q}_+(1,n)$.

Theorem 4.1. *Suppose that the positivity condition* (3.18) *holds. Then, given any* $P \in \mathcal{Q}_+(1,n)$*, the maximization problem*

$$\max_{\Phi \in \mathcal{S}_+^{\ell\times\ell}} \mathbb{I}_P(\Phi) \quad \text{subject to} \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha_k(e^{i\theta})\Phi(e^{i\theta})d\theta = W_k,\ k = 0,1,\dots,n, \tag{4.6}$$

has a unique optimal solution, and it takes the form

$$\Phi(z) = P(z)Q(z)^{-1}, \tag{4.7}$$

where $Q \in \mathcal{Q}_+(\ell,n)$*. Via* (3.12) *this establishes a one-one correspondence between interpolants* $F \in \mathcal{F}_+(n)$ *and* $P \in \mathcal{Q}_+(1,n)$*.*

This is a constrained optimization problem over the infinite-dimensional space $\mathcal{S}_+^{\ell\times\ell}$, which is hard to solve directly. In analogy with [7] we observe that the optimization problem has only finitely many constraints and thus a finite-dimensional dual. In fact, in Section 5, we shall demonstrate that Q in (4.7) can be determined by solving the dual optimization problem, namely the problem to find a $Q \in \mathcal{Q}_+(\ell,n)$ that minimizes the functional

$$\mathbb{J}_P(Q) := \langle Q, W\rangle - \frac{1}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta}) \log\det Q(e^{i\theta})d\theta. \tag{4.8}$$

This will be formalized in the next theorem. To this end, we recall from Definition 3.1 and Theorem 3.1 that the sequence $\mathcal{W}$ of interpolation values is positive if and only if

$$\langle Q, W\rangle > 0 \text{ for all } Q \in \mathcal{Q}_+(\ell,n). \tag{4.9}$$

Theorem 4.2. *Suppose that the positivity condition* (4.9) *holds. Then, given any* $P \in \mathcal{Q}_+(1,n)$*, the minimization problem*

$$\min_{Q\in\mathcal{Q}_+(\ell,n)} \mathbb{J}_P(Q) \tag{4.10}$$

has a unique optimal solution. Given the optimal solution $\hat{Q}$*, the unique interpolant* $F \in \mathcal{F}_+(n)$ *corresponding to* P*, mentioned in Theorem 4.1, is given by*

$$F(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} P(e^{i\theta})\hat{Q}(e^{i\theta})^{-1}d\theta. \tag{4.11}$$

The optimal solution $\hat{Q}$ *depends smoothly on the interpolation data* $\mathcal{W}$*. In particular, the map* $\mathcal{I} : \mathcal{Q}_+(\ell,n) \to \mathfrak{W}_+(\ell,n)$ *with components*

$$\mathcal{I}_k(Q) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha_k(e^{i\theta})P(e^{i\theta})Q(e^{i\theta})^{-1}d\theta, \quad k = 0,1,\dots,n, \tag{4.12}$$

is a diffeomorphism.

It is easy to see that, modulo sign change, any $P \in \mathcal{Q}_+(1,n)$ has a unique representation of the form

$$P(z) = \frac{\rho(z)\rho^*(z)}{\tau(z)\tau^*(z)}, \tag{4.13}$$

where

$$\tau(z) := \prod_{k=1}^{n}(1-\bar{z}_k z) \tag{4.14}$$

belongs to the class $\mathfrak{S}_+$ of polynominals with all roots in $\overline{\mathbb{D}}^c$, and where

$$\rho(z) = \rho_0 + \rho_1 z + \cdots + \rho_n z^n \tag{4.15}$$

is an arbitrary polynomial in $\mathfrak{S}_+$. The parameters $\rho_0, \rho_1, \dots, \rho_n$ can serve as "tuning parameters" in robust control and other applications. In the scalar case, rules of thumb for choosing these tuning parameters are given in [46, 47, 50] for sensitivity shaping and in [6] for high-resolution spectral estimation. Noting that ρ is still scalar, these rules of thumb essentially also apply to the present matrix case. In sensitivity shaping, the most effective rule is to place a root of (4.15) close to the unit circle at a frequency where a peak is desired. The interpolant F can be determined from the solution to the dual optimization problem (4.10) in a fashion to be described in Section 6.

Similarly, any $Q \in \mathfrak{Q}_+(\ell, n)$ has a representation (3.23), i.e. $Q(z) = \Gamma(z)\Gamma^*(z)$, unique up to an orthogonal transformation, where

$$\Gamma(z) = \tau(z)^{-1}R(z) \tag{4.16}$$

and the $\ell \times \ell$ matrix polynomial

$$R(z) = R_0 + R_1 z + \cdots + R_n z^n \tag{4.17}$$

are outer. In Section 6 we assume that the interpolation data $\mathcal{Z}$, $\mathcal{W}$ are self-conjugate, and thus the matrix coefficients are real. We also show that the dual optimization problem can be reformulated in terms of $R(z)$ so that, in particular, the spectral factorization step and complex number calculations are avoided.

Consequently, for each choice of tuning parameters $\rho_0, \rho_1, \dots, \rho_n$, the dual optimization problem provides an essentially unique matrix polynomial (4.17) so that

$$V(z) := \rho(z)R(z)^{-1} \tag{4.18}$$

is an outer spectral factor of $\Phi = PQ^{-1}$. Forming a minimal realization (3.3) of (4.18), the corresponding interpolant $F \in \mathfrak{F}_+(n)$ is given by (3.4).

5. Proofs in the context of duality theory

To solve the problem (4.6), we form the Lagrangian

$$L(\Phi, Q) := \mathbb{I}_P(\Phi) + \operatorname{Re}\left\{\sum_{k=0}^{n}\sum_{i=1}^{\ell}\sum_{j=1}^{\ell} q_k^{ji}\left[w_k^{ij} - \frac{1}{2\pi}\int_{-\pi}^{\pi}\alpha_k(e^{i\theta})\Phi_{ij}(e^{i\theta})d\theta\right]\right\},$$

where w_k^{ij} and Φ_{ij} are the matrix components of W_k and Φ respectively, and then solve the dual problem to minimize

$$\sup_{\Phi \in \mathcal{S}_+^{\ell\times\ell}} L(\Phi, Q)$$

with respect to the Lagrange multipliers q_k^{ij}, which are complex numbers except when $k = 0$ when they are real and $q_0^{ji} = q_0^{ij}$. Here, Q is the generalized pseudo-polynomial

(3.16) formed by taking Q_k to be the $\ell \times \ell$ matrix $[q_k^{ij}]_{i,j=1}^{\ell}$ for $k = 0, 1, \dots, n$. Then, using the identity (3.22), the Lagrangian can be written by

$$L(\Phi, Q) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta}) \log\det\Phi(e^{i\theta})d\theta + \langle Q, W\rangle - \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\{Q(e^{i\theta})\Phi(e^{i\theta})\}d\theta. \tag{5.1}$$

Clearly, the Lagrangian will be unbounded if Q is allowed to have negative eigenvalues on the unit circle. Hence, we determine the supremum for each $Q \in \mathcal{Q}_+(\ell, n)$. To this end, we want to determine a Φ such that the directional derivative

$$\begin{aligned}\delta L(\Phi, Q; \delta\Phi) \;&:=\; \lim_{\varepsilon\to 0} \frac{L(\Phi + \varepsilon\delta\Phi, Q) - L(\Phi, Q)}{\varepsilon} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta}) \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\log\left[\frac{\det(\Phi+\varepsilon\delta\Phi)}{\det\Phi}\right]d\theta - \frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\{Q\delta\Phi\}d\theta\end{aligned}$$

equals zero in all directions $\delta\Phi$ such that $\Phi + \varepsilon\delta\Phi \in \mathcal{S}_+^{\ell\times\ell}$ for some $\varepsilon > 0$. However, since

$$\log\left[\frac{\det(\Phi+\varepsilon\delta\Phi)}{\det\Phi}\right] = \log\det(I + \varepsilon\Phi^{-1}\delta\Phi) = \log\prod_{j=1}^{\ell}(1+\varepsilon\lambda_j) = \sum_{j=1}^{\ell}\log(1+\varepsilon\lambda_j),$$

where $\lambda_1(e^{i\theta}), \lambda_2(e^{i\theta}), \dots, \lambda_\ell(e^{i\theta})$ are the eigenvalues of $\Phi(e^{i\theta})^{-1}\delta\Phi(e^{i\theta})$, and $\log(1 + \varepsilon\lambda_j) = \varepsilon\lambda_j + O(\varepsilon^2)$, we have

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\log\left[\frac{\det(\Phi+\varepsilon\delta\Phi)}{\det\Phi}\right] = \sum_{j=1}^{\ell}\lambda_j = \mathrm{tr}(\Phi^{-1}\delta\Phi). \tag{5.2}$$

Consequently, in terms of the inner product the directional derivative can be written as

$$\delta L(\Phi, Q; \delta\Phi) = \langle\delta\Phi, P\Phi^{-1} - Q\rangle, \tag{5.3}$$

which equals zero for all $\delta\Phi$ if and only if

$$\Phi = PQ^{-1}. \tag{5.4}$$

Inserting this into (5.1) we obtain

$$\mathbb{J}_P(Q) + \frac{\ell}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta})\big(\log P(e^{i\theta}) - 1\big)d\theta,$$

where

$$\mathbb{J}_P(Q) = \langle Q, W\rangle - \frac{1}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta})\log\det Q(e^{i\theta})d\theta. \tag{5.5}$$

Hence, modulo an additive constant, $\mathbb{J}_P$ is precisely the dual function.

We want to show that this functional is strictly convex and that it has a unique minimum in $\mathcal{Q}_+(\ell, n)$. To this end, we form the directional derivative

$$\begin{aligned}\delta\mathbb{J}_P(Q; \delta Q) \;&:=\; \lim_{\varepsilon\to 0}\frac{\mathbb{J}_P(Q+\varepsilon\delta Q) - \mathbb{J}_P(Q)}{\varepsilon} \\ &= \langle\delta Q, W\rangle - \frac{1}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta})\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\log\left[\frac{\det(Q+\varepsilon\delta Q)}{\det Q}\right]d\theta \\ &= \langle\delta Q, W - PQ^{-1}\rangle, \qquad (5.6)\end{aligned}$$

where we have performed the same calculation as in (5.2). We need to determine a $Q \in \mathcal{Q}_+(\ell, n)$ such that

$$\delta\mathbb{J}_P(Q; \delta Q) = 0 \tag{5.7}$$

for all δQ of the form

$$\delta Q(e^{i\theta}) = \Re\left\{\sum_{k=0}^{n} \delta Q_k \alpha_k(e^{i\theta})\right\}, \tag{5.8}$$

where δQ_k, $k = 0, 1, \dots, n$, are arbitrary complex $\ell \times \ell$ matrices, except for δQ_0 that is real and symmetric. Inserting (5.8) into (5.6), we obtain

$$\begin{aligned}
\delta\mathbb{J}_P(Q; \delta Q) &= \operatorname{Re}\left\{\sum_{k=0}^{n} \operatorname{tr}\left(\delta Q_k \frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha_k(e^{i\theta})\left[W(e^{i\theta}) - P(e^{i\theta})Q(e^{i\theta})^{-1}\right] d\theta\right)\right\} \\
&= \operatorname{Re}\left\{\sum_{k=0}^{n} \operatorname{tr}\left(\delta Q_k \big[W_k - \mathcal{I}_k(Q)\big]\right)\right\}
\end{aligned}$$

where $\mathcal{I}_0(Q), \mathcal{I}_1(Q), \dots, \mathcal{I}_n(Q)$ are defined as in (4.12).

Lemma 5.1. *The stationarity condition* (5.7) *holds for all δQ of the form* (5.8) *if and only if $\mathcal{I}_k(Q) = W_k$, $k = 0, 1, \dots, n$.*

Proof. For an arbitrary (k, i, j) with $k \neq 0$, take all components of $\delta Q_0, \delta Q_1, \dots, \delta Q_n$ equal to zero except δq_k^{ji}, which we take to be $\lambda + i\mu$ with λ and μ arbitrary. Then, letting u_k^{ij} be the real part and v_k^{ij} the imaginary part of $w_k^{ij} - \mathcal{I}_k^{ij}(Q)$, we obtain

$$\delta\mathbb{J}_P(Q; \delta Q) = \operatorname{Re}\{(\lambda + i\mu)(u_k^{ij} + iv_k^{ij})\} = \lambda u_k^{ij} - \mu v_k^{ij},$$

and hence $w_k^{ij} = \mathcal{I}_k^{ij}(Q)$, as claimed. If $k = 0$, μ and v_k^{ij} equal to zero, so the same conclusion follows. The reverse statement is trivial. □

It remains to show that there is a $Q \in \mathcal{Q}_+(\ell, n)$ such that (5.7) holds.

Theorem 5.2. *Let $P \in \mathcal{Q}_+(1, n)$, and suppose that the positivity condition* (3.18) *holds. The dual functional $\mathbb{J}_P : \mathcal{Q}_+(\ell, n) \to \mathbb{R}$ is strictly convex and has a unique minimum $\hat{Q}$. Moreover,*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \alpha_k(e^{i\theta})P(e^{i\theta})\hat{Q}(e^{i\theta})^{-1} d\theta = W_k, \quad k = 0, 1, \dots, n. \tag{5.9}$$

Proof. To prove that $\mathbb{J}_P$ is strictly convex we form

$$\begin{aligned}
\delta^2\mathbb{J}_P(Q; \delta Q) &:= \lim_{\varepsilon \to 0} \frac{\delta\mathbb{J}_P(Q + \varepsilon\delta Q; \delta Q) - \delta\mathbb{J}_P(Q; \delta Q)}{\varepsilon} \\
&= -\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left\langle \delta Q, P\left[(Q + \varepsilon\delta Q)^{-1} - Q^{-1}\right]\right\rangle \\
&= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left\langle \delta Q, P\left[I - (I + \varepsilon Q^{-1}\delta Q)^{-1}\right] Q^{-1}\right\rangle
\end{aligned}$$

However,

$$(I + \varepsilon Q^{-1}\delta Q)^{-1} = I - \varepsilon Q^{-1}\delta Q + O(\varepsilon^2)$$

for sufficiently small $\varepsilon > 0$, and hence

$$\delta^2 \mathbb{J}_P(Q; \delta Q) = \langle \delta Q, PQ^{-1} \delta Q Q^{-1} \rangle.$$

Now, since $Q \in \mathcal{Q}_+(\ell, n)$ is positive definite on the unit circle, there is a nonsingular matrix function S such that $Q^{-1} = SS^*$. Then, using the commuting property of the trace, we have

$$\operatorname{tr}(\delta Q Q^{-1} \delta Q Q^{-1}) = \operatorname{tr}(S^* \delta Q S S^* \delta Q S),$$

and hence

$$\delta^2 \mathbb{J}_P(Q; \delta Q) = \langle S^* \delta Q S, P(S^* \delta Q S) \rangle \geq 0,$$

taking the value zero if and only if $S^* \delta Q S = 0$ or, equivalently, $\delta Q = 0$. Consequently, the Hessian of $\mathbb{J}_P(Q)$ is positive definite for all $Q \in \mathcal{Q}_+(\ell, n)$, implying that $\mathbb{J}_P$ is strictly convex, as claimed.

The rest of the proof is the same *mutatis mutandis* as the one in [8]. (Also see [7, 9, 10].) Since the linear term $\langle Q, W \rangle$ is positive and linear growth is faster than logarithmic, the function $\mathbb{J}_P$ is proper, i.e., the inverse images of compact sets are compact. In particular, if we extend the function $\mathbb{J}_P$ to the boundary of $\mathcal{Q}_+(\ell, n)$, it has compact sublevel sets. Consequently, $\mathbb{J}_P$ has a minimum, $\hat{Q}$, which is unique by strict convexity. We need to rule out that $\hat{Q}$ lies on the boundary. To this end, note that the boundary of $\mathcal{Q}_+(\ell, n)$ consists of the Q for which $\det Q$ has a zero on the unit circle, and for which the directional derivative $\delta \mathbb{J}_P(Q; \delta Q) = -\infty$ for all δQ pointing into $\mathcal{Q}_+(\ell, n)$. See [9, Section 4] for details.

Therefore, since $\mathcal{Q}_+(\ell, n)$ is an open set, $\delta \mathbb{J}_P(\hat{Q}; \delta Q) = 0$ for all δQ of the form (5.8), and therefore (5.9) follows from Lemma 5.1. □

Theorem 5.3. *Let $P \in \mathcal{Q}_+(1, n)$, and suppose that the positivity condition* (3.18) *holds. The primal functional $\mathbb{I}_P : \mathcal{S}_+^{\ell \times \ell} \to \mathbb{R}$ is strictly concave, and there is a unique optimal solution $\hat{\Phi} \in \mathcal{S}_+^{\ell \times \ell}$ to the problem* (4.6). *The maximum $\hat{\Phi}$ takes the form*

$$\hat{\Phi} = P \hat{Q}^{-1},$$

where $\hat{Q} \in \mathcal{Q}_+(\ell, n)$ is the optimal solution of the dual problem.

Proof. To show that $\mathbb{I}_P$ is strictly concave, we proceed as above. The calculation leading to (5.3) yields

$$\delta \mathbb{I}_P(\Phi; \delta \Phi) = \langle \delta \Phi, P \Phi^{-1} \rangle,$$

and, following the lines of the corresponding proof in Theorem 5.2,

$$\delta^2 \mathbb{I}_P(\Phi; \delta \Phi) \leq 0$$

with equality if and only if $\delta \Phi = 0$. Hence $\mathbb{I}_P$ is strictly concave, as claimed.

Let $\hat{Q}$ be the optimal solution of the dual problem. Then, since $\mathbb{I}_P$ is strictly concave, then so is $\Phi \mapsto L(\Phi, \hat{Q})$. Clearly $\hat{\Phi} := P \hat{Q}^{-1}$ belongs to $\mathcal{S}_+^{\ell \times \ell}$, and, by (5.3), it is a stationary point of the map $\Phi \mapsto L(\Phi, \hat{Q})$. Hence

$$L(\hat{\Phi}, \hat{Q}) \geq L(\Phi, \hat{Q}), \quad \text{for all } \Phi \in \mathcal{S}_+^{\ell \times \ell}. \tag{5.10}$$

However, by Theorem 5.2, $\hat{\Phi}$ satisfies the interpolation condition (3.13), and consequently

$$L(\hat{\Phi}, \hat{Q}) = \mathbb{I}_P(\hat{\Phi}).$$

Therefore, it follows from (5.10) that

$$\mathbb{I}_P(\Phi) \leq \mathbb{I}_P(\hat{\Phi})$$

for all $\Phi \in \mathcal{S}_+^{\ell\times\ell}$ that satisfies the interpolation condition (3.13), establishing optimality of $\hat{\Phi}$. □

Consequently, we have proved Theorem 4.1. To finish the proof of Theorem 4.2 it remains to establish that the map $\mathcal{I} : \mathcal{Q}_+(\ell, n) \to \mathfrak{W}_+(\ell, n)$ is a diffeomorphism. To this end, first note that $\mathcal{Q}_+(\ell, n)$ and $\mathfrak{W}_+(\ell, n)$ are both convex, open sets in $\mathbb{R}^{2n\ell^2+\frac{1}{2}\ell(\ell+1)}$ and hence diffeomorphic to $\mathbb{R}^{2n\ell^2+\frac{1}{2}\ell(\ell+1)}$. Moreover, the Jacobian of $\mathcal{I}$ is the Hessian of $\mathbb{J}_P$, which is positive definite on $\mathcal{Q}_+(\ell, n)$, as shown in the proof of Theorem 5.2. Hence, by Hadamard's global inverse function theorem [33], $\mathcal{I}$ is a diffeomorphism.

Finally, Theorem 3.1 is an immediate consequence of Theorem 4.1.

6. Solving the dual optimization problem

Recall that, by Theorem 4.2, for each choice of $P \in \mathcal{Q}_+(1, n)$, there is a unique solution to the basic interpolation problem of this paper, and this solution is obtained by determining the unique minimizer over $\mathcal{Q}_+(\ell, n)$ of the dual functional

$$\mathbb{J}_P(Q) := \langle Q, W\rangle - \langle \log\det Q, P\rangle\,. \tag{6.1}$$

This functional has the property that its gradient is infinite on the boundary of $\mathcal{Q}_+(\ell, n)$. This is precisely the property that buys us properness of the functional (4.12), and therefore it is essential in the proof of Theorem 4.2. However, from a computational point of view, this property is undesirable, especially if the minimum is close to the boundary. In fact, it adversely affects the accuracy of any Newton-type algorithm. For this reason, following [22, 49], we first reformulate the optimization problem to eliminate this property. This is done at the expense of global convexity, but the new functional is still locally strictly convex in a neighborhood of a unique minimizing point. Thus, if we were able to choose the initial point in the convexity region, a Newton method would work well. However, finding such an initial point is a highly nontrivial matter. Therefore, again following [22, 49], we want to design a homotopy continuation method that determines a sequence of points converging to the minimizing point.

6.1. **Reformulating the optimization problem.** In Section 3.4 we replaced the first term in (6.1) with a quadratic form by first defining the spectral factor $\Gamma(z)$ satisfying (3.23). Consequently, for each $Q = \Gamma\Gamma^*$, the right hand side of (6.1) can also be written as

$$\operatorname{tr}\boldsymbol{\Gamma}^*\boldsymbol{\Pi}\boldsymbol{\Gamma} - \langle \log\det\Gamma\Gamma^*, P\rangle\,,$$

where $\boldsymbol{\Pi}$ is the generalized Pick matrix defined by (3.27) or, alternatively, by (3.28). Let us now assume that the interpolation data $(\mathcal{Z}, \mathcal{W})$ is self-conjugate so that space $\mathcal{Q}_+(\ell, n)$ has dimension $\ell^2 n + \frac{1}{2}\ell(\ell+1)$ and the matrix coefficients $R_0, R_1, \dots, R_n$ in

$$R(z) := \tau(z)\Gamma(z) = R_0 + R_1 z + \cdots + R_n z^n \tag{6.2}$$

are real. We also assume that R_0 is upper triangular. Then, the space $\mathcal{R}_+(\ell,n)$ of all

$$\boldsymbol{R} := \begin{bmatrix} R_0 \\ \vdots \\ R_n \end{bmatrix} \in \mathbb{R}^{\ell(n+1)\times\ell}$$

such that (6.2) is outer and $R(e^{i\theta})R(e^{i\theta})^* > 0$ for all $\theta\in[-\pi,\pi]$ also has dimension $\ell^2 n + \frac{1}{2}\ell(\ell+1)$. In view of (3.25),

$$R(z) = \tau(z)(G(z)\otimes I_\ell)\boldsymbol{\Gamma}, \tag{6.3}$$

which defines a nonsingular linear transformation T such that

$$\boldsymbol{\Gamma} = T\boldsymbol{R}. \tag{6.4}$$

Under this change of coordinates, the Pick matrix becomes

$$\boldsymbol{K} = T^*\boldsymbol{\Pi}T, \tag{6.5}$$

and, since $\arg\det R(e^{-i\theta}) = -\arg\det R(e^{i\theta})$, (6.1) can be written

$$\mathbb{J}_P(Q) = \boldsymbol{J}_P(\boldsymbol{R}) - 2\left\langle \log\tau, P\right\rangle, \tag{6.6}$$

where the new cost functional

$$\boldsymbol{J}_P(\boldsymbol{R}) = \operatorname{tr}\boldsymbol{R}^{\mathsf{T}}\boldsymbol{K}\boldsymbol{R} - 2\left\langle \log\det R, P\right\rangle \tag{6.7}$$

is defined on the space $\mathcal{R}_+(\ell,n)$.

Proposition 6.1. *The functional $\boldsymbol{J}_P : \mathcal{R}_+(\ell,n)\to\mathbb{R}$ has a unique stationary point and is locally strictly convex about this point.*

Proof. Since $\Gamma(z) := R(z)/\tau(z)$ is a uniquely defined (outer) spectral factor of $Q(z)$, the map $\Psi : \mathcal{R}_+(\ell,n)\to\mathcal{Q}_+(\ell,n)$ sending $\boldsymbol{R}$ to $Q(z) = \Theta(z)\boldsymbol{R}\boldsymbol{R}^*\Theta^*(z)$, where

$$\Theta(z) := \frac{1}{\tau(z)}\begin{bmatrix} I_\ell & zI_\ell & \cdots & z^n I_\ell\end{bmatrix},$$

is a bijection with first and second directional derivatives

$$\delta\Psi(\boldsymbol{R};\delta\boldsymbol{R}) = \Theta(z)\big(\boldsymbol{R}(\delta\boldsymbol{R})^* + (\delta\boldsymbol{R})\boldsymbol{R}^*\big)\Theta^*(z)$$
$$\delta^2\Psi(\boldsymbol{R};\delta\boldsymbol{R}) = 2\Theta(z)\big((\delta\boldsymbol{R})(\delta\boldsymbol{R})^*\big)\Theta^*(z).$$

Now, $\delta\boldsymbol{R}\mapsto\delta\Psi(\boldsymbol{R};\delta\boldsymbol{R})$ is an injective linear map between Euclidean spaces of the same dimension, and hence it is bijective. In fact, since $\det R(z)$ has all its roots in the complement of the closed unit disc, the homogeneous equation

$$R(z)\Delta^*(z) + \Delta(z)R^*(z) \equiv 0, \quad \Delta(z) := \Theta(z)\delta\boldsymbol{R},$$

has a unique solution $\Delta(z)\equiv 0$. (See Lemma A.1 in Appendix A.) Therefore, since

$$\boldsymbol{J}_P(\boldsymbol{R}) = \mathbb{J}_P(\Psi(\boldsymbol{R})) + 2\langle\log\tau, P\rangle,$$

the directional derivative

$$\delta\boldsymbol{J}_P(\boldsymbol{R};\delta\boldsymbol{R}) = \delta\mathbb{J}_P(\Psi(\boldsymbol{R});\delta\Psi(\boldsymbol{R};\delta\boldsymbol{R}))$$

is zero for all $\delta\boldsymbol{R}$ if and only if $\delta\mathbb{J}_P(Q;\delta Q) = \langle\delta Q, W - PQ^{-1}\rangle$ is zero for all δQ. Consequently, $\boldsymbol{J}_P$ has a stationary point at $\hat{\boldsymbol{R}}$ if and only if $\mathbb{J}_P$ has a stationary point at $\Psi(\hat{\boldsymbol{R}})$. However, $\mathbb{J}_P$ has exactly one such point, and hence the same holds for $\boldsymbol{J}_P$. Moreover, since $\delta^2\mathbb{J}_P(Q;\delta Q) = \langle \delta Q, PQ^{-1}\delta Q Q^{-1}\rangle > 0$ for all $\delta Q \neq 0$ and $\delta\mathbb{J}_P(\hat{Q};\delta Q) = 0$ at the minimum $\hat{Q}$, the second directional derivative

$$\delta^2\boldsymbol{J}_P(\hat{\boldsymbol{R}};\delta\boldsymbol{R}) = \delta^2\mathbb{J}_P(\Psi(\hat{\boldsymbol{R}});\delta\Psi(\hat{\boldsymbol{R}};\delta\boldsymbol{R})) + \delta\mathbb{J}_P(\Psi(\hat{\boldsymbol{R}});\delta^2\Psi(\hat{\boldsymbol{R}};\delta\boldsymbol{R}))$$

is positive for sufficiently small $\delta\boldsymbol{R} \neq 0$. Therefore, $\boldsymbol{J}_P$ is strictly convex in some neighborhood of $\hat{\boldsymbol{R}}$. □

6.2. **The gradient and the Hessian of the new functional.** In order to use Newton's method to solve the new optimization problem, we need to determine the gradient and the Hessian of $\boldsymbol{J}_P$. We begin with the gradient.

Proposition 6.2. *Given the real $\ell \times \ell$ matrix-valued Fourier coefficients*

$$C_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\theta}P(e^{i\theta})\big(R^*(e^{i\theta})R(e^{i\theta})\big)^{-1}d\theta, \quad k = 0,1,\dots,n, \tag{6.8}$$

and the modified Pick matrix $\boldsymbol{K}$, given by (6.5)*, the gradient of $\boldsymbol{J}_P$ is given by*

$$\frac{\partial \boldsymbol{J}_P}{\partial \boldsymbol{R}}(\boldsymbol{R}) = 2\big(\boldsymbol{K} - C(\boldsymbol{R})\big)\boldsymbol{R}, \tag{6.9}$$

where the $(n+1)\ell \times (n+1)\ell$ matrix $C(\boldsymbol{R})$ is the Toeplitz matrix

$$C(\boldsymbol{R}) := \begin{bmatrix} C_0 & C_1 & \cdots & C_n \\ {C_1}^{\mathsf{T}} & C_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & C_1 \\ {C_n}^{\mathsf{T}} & \cdots & {C_1}^{\mathsf{T}} & C_0 \end{bmatrix}. \tag{6.10}$$

The proof of Proposition 6.2 is given in Appendix B, while the proof of the following proposition, describing the Hessian of $\boldsymbol{J}_P$, is given in Appendix C.

Proposition 6.3. *The Hessian of $\boldsymbol{J}_P$ is given by*

$$\frac{\partial^2}{(\partial \operatorname{vec}\boldsymbol{R})^2}\boldsymbol{J}_P(\boldsymbol{R}) = 2(I_\ell \otimes \boldsymbol{K}) - 2\frac{\partial^2}{(\partial \operatorname{vec}\boldsymbol{R})^2}\langle \log\det R, P\rangle. \tag{6.11}$$

Here the component of the second term are obtained by rearranging the elements in

$$\left(\frac{\partial}{\partial R_j} \otimes \frac{\partial}{\partial R_k}\right)\langle \log\det R, P\rangle = -{S_{j+k}}^{\mathsf{T}}, \quad j,k = 0,1,\dots,n, \tag{6.12}$$

where $S_0, S_1, \dots, S_{2n}$ are defined via the expansion

$$P(z)\Big(\operatorname{vec} R(z)^{-1}\Big)\Big(\operatorname{vec} R(z)^{-\mathsf{T}}\Big)^{\mathsf{T}} = \sum_{-\infty}^{\infty} S_k z^{-k}. \tag{6.13}$$

Remark 6.4. Since the left hand side of (6.13) is the product of three factors, two of which have Laurent expansions with infinitely many terms, one might wonder how to determine the coefficients $S_0, S_1, \dots, S_{2n}$ in a finite number of operations. As we shall see in Appendix C, this can be achieved by observing that $P(z)\big(R(z)^{\mathsf{T}} \otimes R(z)^{-1}\big)$ has the same elements as (6.13), appropriately rearranged, and can be factored as the product of two finite and one infinite Laurent expansion.

6.3. **The central solution.** The optimization problem to minimize $\boldsymbol{J}_P$ is particularly simple if $P = 1$. In this case, and only in this case, the problem can be reduced to one of solving a system of linear equations. This solution is generally called the *central solution.* In fact, since $\det R(z)$ has no zeros in $\overline{\mathbb{D}}$, by the mean-value theorem of harmonic functions,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log|\det R(e^{i\theta})| d\theta = \log|\det R(0)|.$$

Consequently, since $\arg\det R(e^{-i\theta}) = -\arg\det R(e^{i\theta})$,

$$\boldsymbol{J}_1(\boldsymbol{R}) = \operatorname{tr} \boldsymbol{R}^{\mathsf{T}}\boldsymbol{K}\boldsymbol{R} - 2\log\det R_0.$$

Since $\det R(z)$ has no zeros in the unit disc, R_0 is nonsingular. Therefore, setting the gradient of $\boldsymbol{J}_1(\boldsymbol{R})$ equal to zero, we obtain

$$\boldsymbol{K}\boldsymbol{R} = \boldsymbol{E}R_0^{-\mathsf{T}}, \quad \boldsymbol{E} = \begin{bmatrix} I_\ell & 0 & \cdots & 0\end{bmatrix}^{\mathsf{T}}, \tag{6.14}$$

and therefore $R_0 = \boldsymbol{E}^{\mathsf{T}}\boldsymbol{R} = \boldsymbol{E}^{\mathsf{T}}\boldsymbol{K}^{-1}\boldsymbol{E}R_0^{-\mathsf{T}}$, which yields

$$R_0 R_0^{\mathsf{T}} = \boldsymbol{E}^{\mathsf{T}}\boldsymbol{K}^{-1}\boldsymbol{E}. \tag{6.15}$$

First solving (6.15) for the unique Cholesky factor and inserting into (6.14), (6.14) reduces to a linear system of equations that has a unique solution $\boldsymbol{R}$ since $\boldsymbol{K}$ is positive definite.

6.4. **The continuation method.** Now, we would like to find the minimizer of $\boldsymbol{J}_P$ for an arbitrary $P \in \mathcal{Q}_+(1,n)$. To this end, we construct a homotopy between the gradient of $\boldsymbol{J}_1$ and the gradient of $\boldsymbol{J}_P$ along the lines of [22, 49], allowing us to pass from the central solution to the solution of interest.

Now, for any $\lambda \in [0,1]$, define

$$P_\lambda(z) := 1 + \lambda\big(P(z) - 1\big).$$

Then, since $\mathcal{Q}_+(1,n)$ is convex, $P_\lambda \in \mathcal{Q}_+(1,n)$. By Proposition 6.1, the functional

$$\boldsymbol{J}_{P_\lambda}(\boldsymbol{R}) = \operatorname{tr} \boldsymbol{R}^{\mathsf{T}}\boldsymbol{K}\boldsymbol{R} - 2\,\langle \log\det R, P_\lambda\rangle$$

has a unique minimum at $\hat{\boldsymbol{R}}(\lambda)$ and is locally strictly convex in some neighborhood of $\hat{\boldsymbol{R}}(\lambda)$. This point is the unique solution in $\mathcal{R}_+(\ell,n)$ of the nonlinear equation

$$h(\boldsymbol{R},\lambda) := \frac{\partial \boldsymbol{J}_{P_\lambda}(\boldsymbol{R})}{\partial \operatorname{vec}\boldsymbol{R}} = 0.$$

Then the function $h:\ \mathcal{R}_+(\ell,n)\times[0,1] \to \mathbb{R}^{(n+1)\ell^2}$ is a homotopy from the gradient of $\boldsymbol{J}_1$ to the gradient of $\boldsymbol{J}_P$. In particular, $\hat{\boldsymbol{R}}(0)$ is the central solution.

In view of the strict local convexity of $\boldsymbol{J}_{P_\lambda}$ in a neighborhood of $\hat{\boldsymbol{R}}(\lambda)$, the Jacobian of $h(\boldsymbol{R},\lambda)$ is positive definite at $\hat{\boldsymbol{R}}(\lambda)$. Consequently, by the implicit function theorem, the function $\lambda \to \hat{\boldsymbol{R}}(\lambda)$ is continuously differentiable on the interval $[0,1]$, and

$$\frac{d}{d\lambda}\operatorname{vec}\hat{\boldsymbol{R}}(\lambda) = -\left(\frac{\partial \boldsymbol{h}}{\partial \operatorname{vec}\boldsymbol{R}}(\boldsymbol{R},\lambda)\right)^{-1}\left(\frac{\partial \boldsymbol{h}}{\partial \lambda}(\boldsymbol{R},\lambda)\right)\Bigg|_{\boldsymbol{R}=\hat{\boldsymbol{R}}(\lambda)},$$

where the inverted matrix is the Hessian of $\boldsymbol{J}_{P_\lambda}$ that can be determined as in Proposition 6.3. We want to follow the trajectory $\hat{\boldsymbol{R}}(\lambda)$ defined by the solution of this differential equation with the central solution as the initial condition.

To this end, we construct an increasing sequence of numbers $\lambda_0, \lambda_1, \dots, \lambda_N$ on the interval $[0,1]$ with $\lambda_0 = 0$ and $\lambda_N = 1$. Then, for $k = 1, 2, \dots, N$, we solve the nonlinear equation $h(\boldsymbol{R}, \lambda_k) = 0$ for $\operatorname{vec} \hat{\boldsymbol{R}}(\lambda_k)$ by Newton's method with initial condition

$$\operatorname{vec} \boldsymbol{R}_0(\lambda_k) = \operatorname{vec} \hat{\boldsymbol{R}}(\lambda_{k-1}) + \frac{d}{d\lambda} \operatorname{vec} \hat{\boldsymbol{R}}(\lambda_k)(\lambda_k - \lambda_{k-1}).$$

The numbers $\lambda_0, \lambda_1, \dots, \lambda_N$ have to be chosen close enough so that, for each $k = 1, 2, \dots, N$, $\boldsymbol{R}_0(\lambda_k)$ lies in the local convexity region of $\boldsymbol{J}_{P_{\lambda_k}}$, guaranteeing that Newton's method converges to $\hat{\boldsymbol{R}}(\lambda_k)$. Strategies for choosing $\lambda_0, \lambda_1, \dots, \lambda_N$ are given in [22, 49]. This choice is a trade-off between convergence and staying in the locally convex region.

Remark 6.5. A MATLAB implementation of this algorithm is available [1].

7. An application to a benchmark problem in robust control

During the last two decades it has been discovered that analytic interpolation theory is closely related to several robust control problems, for example, the gain-margin maximization problem [55, 56, 38], the robust stabilization problem [39], sensitivity shaping in feedback control, simultaneous stabilization [31], the robust regulation problem [15], the general H^∞ control problem [23], and, more generally, the model matching problem. In this section we apply the theory of this paper to a benchmark problem in sensitivity shaping for a MIMO plant from a popular textbook on multivariable control by Maciejowski [44]. We refer the reader to Section 2 for notation.

The control system in [44] describes the vertical-plane dynamics of an airplane and can be linearized to yield a linear system P with three inputs, three outputs and five states, namely

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du, \end{aligned}$$

where

$$A = \begin{bmatrix} 0 & 0 & 1.1320 & 0 & -1.000 \\ 0 & -0.0538 & -0.1712 & 0 & 0.0705 \\ 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0.0485 & 0 & -0.8556 & -1.013 \\ 0 & -0.2909 & 0 & 1.0532 & -0.6859 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 \\ -0.12 & 1.0000 & 0 \\ 0 & 0 & 0 \\ 4.4190 & 0 & -1.665 \\ 1.5750 & 0 & -0.0732 \end{bmatrix} \quad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This system is not asymptotically stable due to the pole at the origin. It is strictly proper ($D = 0$) and the first Markov coefficient

$$CB = \begin{bmatrix} 0 & 0 & 0 \\ -0.12 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is rank deficient.

To compare our result with that of [44], we want to design a one-degree-of-freedom controller[2] C as in Figure 2 in Section 2 that renders the closed-loop system robust against various disturbances. More precisely, the specifications are

- Bandwidth about 10 rad/s
- Zero sensitivity at zero frequency; $S(0) = 0$
- Well-damped step responses

By exploiting the design freedom offered by choosing the *design parameters*, namely an upper limit γ of the gain, the tuning parameters $\rho_0, \rho_1, \dots, \rho_n$, and additional interpolation constraints, we shape the sensitivity function to meet the specifications, while limiting the degree of the controller.

First we deal with the pole at origin. By perturbing the A matrix we move the pole into the open right half-plane, generating an interpolation point as described in Section 2. More precisely, we move the pole to 10^{-6} by increasing A_{11} to 10^{-6}. This will ensure the sensitivity S to be zero near zero frequency. In terms of the modified sensitivity function Z introduced in Section 2, this yields the interpolation condition $Z(10^6) = 0$.

Since the plant is strictly proper and the first Markov parameter CB is rank deficient, we need to add interpolation conditions for Z and Z' at zero in accordance with (2.15). Moving the interpolation condition slightly into the open right half-plane (Remark 2.3), these conditions become

$$Z(10^{-8}) = I, \qquad Z'(10^{-8}) = U_1 := \begin{bmatrix} 0 & 0 & 0 \\ 0 & -2\cdot 10^{-6} & 0 \\ 0 & 0 & -2\cdot 10^{-6} \end{bmatrix}.$$

To force the controller to be strictly proper and create a steep "roll-off" of the complementary sensitivity function, we also add the condition $Z''(10^{-8}) = 0$. Then the class of bounded interpolants becomes

$$\left\{ Z \in RH^\infty : \begin{array}{c} Z(10^{-8}) = I,\ Z'(10^{-8}) = U_1, Z''(10^{-8}) = 0, \\ Z(10^6) = 0,\ \|Z\|_\infty < \gamma \end{array} \right\},$$

where γ is a bound to be selected in the design.

By means of a linear fractional transformation and an appropriate scaling, we transform the problem to the form considered in this paper, yielding the family

$$\left\{ F \in \mathfrak{F}_+(3) :\ F(0) = 1.9250 I,\ F'(0) = F''(0) = 0,\ F(0.9997) = I \ \right\},$$

for the particular choice of γ described next.

We now tune the design parameters to meet the design specifications. First we pick the upper bound $\gamma = 3.16$ (10 dB). However, the actual maximal norm of the sensitivity will be smaller. Furthermore, we want to peak the sensitivity function somewhat above 10 rad/s. We can achieve this by choosing spectral zeros close to the imaginary axis in the corresponding region. Here, we first pick the points $\{60, \pm 40i\}$ and transform them to the unit disc by the same linear fractional transformation as for the interpolation points. By rescaling each resulting root to have absolute value less than 0.95, if necessary, we avoid numerical difficulties and prevent the peak of $|S|$ from becoming too high. In this way, we obtain the spectral zeros $\{0.3969, 0.4936 \pm 0.4998i\}$,

[2]A control design that does not allow for a prefilter.

which we use in the algorithm of Section 6 to determine the corresponding unique interpolant F. Then we transform back to S and calculate $C(s) = P(s)^{-1}\big(S(s)^{-1} - I\big)$.

In Table 1 we compare our control design with the H^∞ design using the weighting functions of [44, pp. 306-315]. In Figure 5 the (singular-value) frequency responses of the sensitivity and the complementary sensitivity of both designs are plotted, and in Figure 6 the step responses are depicted. Clearly, both designs meet the design specifications. We emphasize that although our design meets the specifications at least as well as does the H^∞ design, the McMillan degree of our controller is only half of that of the H^∞ controller.

	Method of this paper	H^∞ **design**
Controller degree	8	17
Peak $\|S\|_\infty$ (dB)	1.3419	1.3582
Peak $\|T\|_\infty$ (dB)	0.9984	1.2328
Bandwidth S (rad/s)	7.3938	4.6202
Bandwidth T (rad/s)	16.1141	16.4140

Table 1. Summary of the designs.

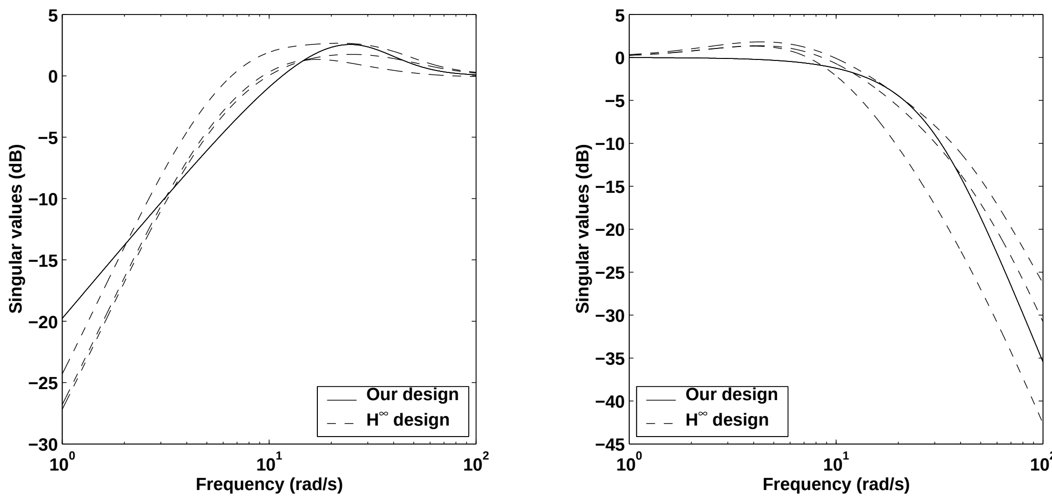

Figure 5: Singular value plots of the sensitivity and the complementary sensitivity.

H^∞ control design often leads to controllers of high degree, and it is therefore customary to apply some method of model reduction. This is typically done by balanced truncation [45], where states that correspond to relatively small entries on the diagonals of the balanced observability/controllability Gramian are removed. Although such procedures are quite *ad hoc*, a certain reduction in degree can often be done without unacceptable degradation in performance.

An interesting question is now whether the H^∞ design in the present example can be reduced to the same degree as our design, namely eight, without unacceptable degradation. The answer is "No". To see this we have used the DC gain matching function in MATLAB's Control Toolbox. Successively removing states in the H^∞ design, we found that the controller can be reduced to degree eleven without loss of internal stability and without undue degradation in performance, whereas reduction

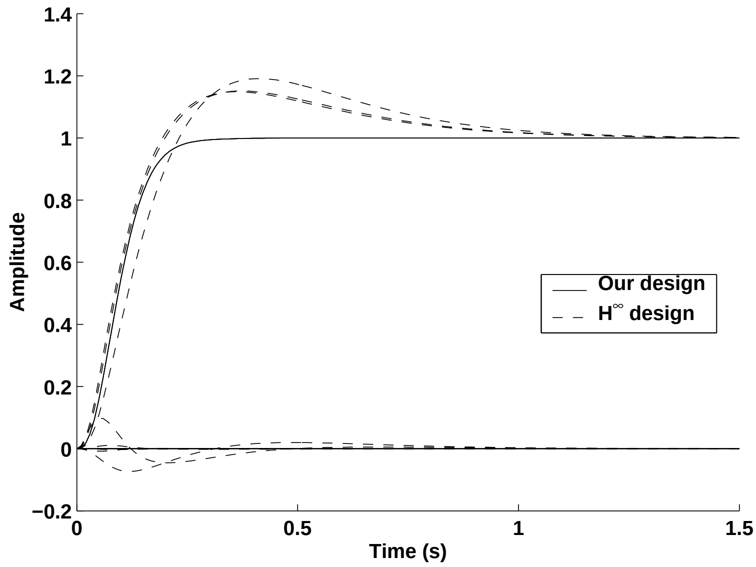

Figure 6: The step responses for the H^∞-design and our design.

to ten leads to an unacceptable design. The results, given in Table 2, still demonstrate the advantages of our new design.

	Method of this paper	**H^∞ design + model reduction**
Controller degree	8	11
Peak $\|S\|_\infty$ (dB)	1.3419	1.3593
Peak $\|T\|_\infty$ (dB)	0.9984	1.2327
Bandwidth S (rad/s)	7.3938	4.6202
Bandwidth T (rad/s)	16.1141	16.4140

Table 2. Our design compared with model-reduced H^∞ design.

Of course, model reduction could also be applied to our design. In fact, the degree of our controller can be reduced to six without unacceptable degradation in performance. restoring the ratio in the control degree between the two methods. The results are displayed in Table 3.

	Method of this paper + model reduction	**H^∞ design + model reduction**
Controller degree	6	11
Peak $\|S\|_\infty$ (dB)	1.3419	1.3593
Peak $\|T\|_\infty$ (dB)	0.9984	1.2327
Bandwidth S (rad/s)	7.3938	4.6202
Bandwidth T (rad/s)	16.1141	16.4140

Table 3. Comparison between model-reduced controllers.

The corresponding (singular-value) frequency responses of the sensitivity and the complementary sensitivity are displayed in Figure 7, and the step responses are depicted in Figure 8. Our design still is of considerably smaller McMillan degree while meeting the design specifications at least as well as the H^∞ design.

Remark 7.1. In interpreting these model-reduction results we need to observe that the interpolation conditions used in our procedure to ensure internal stability are in fact only sufficient (Remark 2.2). Modifying our procedure to handle tangential interpolation would allow us to use necessary and sufficient interpolation conditions. This would reduce the total number of interpolation conditions imposed on the sensitivity function, thus very likely leading to a lower degree controller. In fact, modifying our approach in this way, it is quite possible that the sixth-degree controller obtained above after model reduction of our design could then be obtained directly (without model reduction) using appropriate tuning.

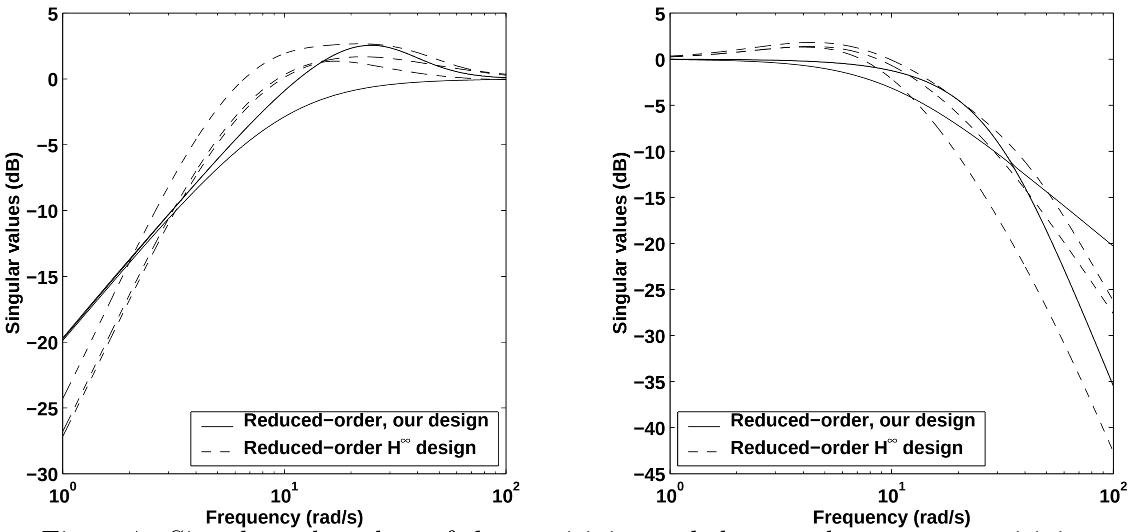

Figure 7: Singular value plots of the sensitivity and the complementary sensitivity corresponding to the model reduced controllers.

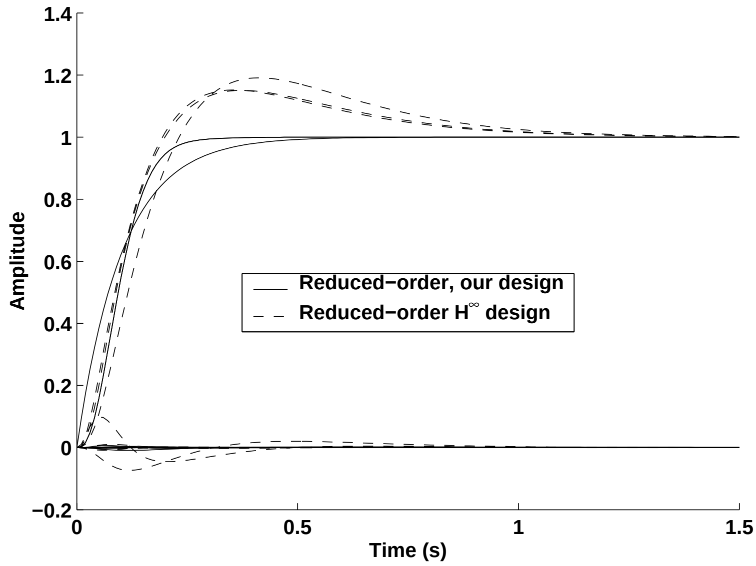

Figure 8: The step responses for the reduced-order designs.

8. Conclusions

In this paper we have developed a theory for matrix-valued Nevanlinna-Pick interpolation with complexity constraint. We have shown that the spectral zeros characterize completely a class of interpolants of a bounded complexity. We have devised a numerically stable algorithm based on homotopy continuation to compute any such interpolants. The potential advantage of the theory and the algorithm was illustrated by a benchmark multivariable control example.

The standard H^∞ control problem can be reduced to not only matrix-valued interpolation but also tangential interpolation. As pointed out in Remark 7.1, we expect that the reduction to the tangential Nevanlinna-Pick interpolation problem will be more natural in the sense that the degree bound can be much lower than the one in this paper (also, see [40, 42]). Therefore, it will be important to modify our theory to tangential Nevanlinna-Pick interpolation. This is the subject of future study.

As pointed out above, and as discussed in detail in [10, 12] in the scalar setting, the Nevanlinna-Pick interpolation problem of this paper can be regarded as a generalized moment problem with complexity constraint. In a different context, Lasserre [41] has recently developed an approach that connects certain moment problems to optimization. An optimization problem of type (4.10) was first introduced in [8] in the context of the covariance extension problem. This approach was originally motivated by the effectiveness of interior point methods; indeed, the logarithmic term in (4.10) was formed in analogy with a barrier term. Therefore, it is interesting to note that similar paradigms appear in recent work on positive polynomials and convex optimization over linear matrix inequalities (LMI); see, in particular, [26, 27]. There are also efficient methods of H^∞ control based on LMI techniques [35, 24, 53]. It would be interesting topics of future study to investigate possible connections between our work and that of [26, 27, 41, 35, 24].

Acknowledgements. The authors would like to thank Professor Per-Olof Gutman, the referees, and the associate editor for many good suggestions for improving the paper.

Appendix A. Nonsingularity of the Jacobian matrix $\delta\Psi$

To show that the Jacobian matrix of Ψ in Proposition 6.1 is bijective, we prove a somewhat more general statement.

Lemma A.1. *Let $\mathcal{V}(\ell, n)$ be the class of real $\ell \times \ell$ matrix polynomials*

$$V(z) = V_0 + V_1 z + \cdots + V_n z^n$$

such that V_0 is upper triangular, and let $R \in \mathcal{V}(\ell, n)$ have the properties that the constant term R_0 is nonsingular and that $\det R$ *and* $\det R^*$ *have no roots in common. Then, the linear map $S(R)$ sending $V \in \mathcal{V}(\ell, n)$ to*

$$S(R)V := R(z)V^*(z) + V(z)R^*(z)$$

is nonsingular.

In the proof of Proposition 6.1, $\det R$ has all its roots in the complement of the closed unit disc. A proof of Lemma A.1 restricted to this case can be found in [36]. (See also [34, Theorem 3.1], which refers to [36].) Nevertheless, we shall provide an independent proof of the more general statement of Lemma A.1. Indeed, our proof is short and straight-forward. Moreover, the general statement given here was left as an open problem in [36, p. 28].

Proof. Since $S(R)$ is a linear map between Euclidean spaces of the same dimension, it suffices to prove that $S(R)$ is injective. Without restriction we may assume that $R(z)$ is upper triangular. In fact, let $T(z)$ be a unimodular matrix polynomial with $T(0)$ upper triangular such that $T(z)R(z)$ is upper triangular. Such a T indeed exists due to the procedure deriving the Smith form [25]. Then

$$TS(R)VT^* = (TR)(TV)^* + (TV)(TR)^* = 0$$

if and only if $S(R)V = 0$. Moreover, the new V_0, i.e., $T(0)V(0)$, is still upper triangular. In this formulation

$$\det R(z) = r_{11}(z)r_{22}(z)\cdots r_{\ell\ell}(z),$$

where $r_{11}, r_{22}, \cdots, r_{\ell\ell}$ are the diagonal elements in R. In particular, by assumption, no r_{ii} can have zeros in common with any r_{jj}^*. It then remains to prove that

$$RV^* + VR^* = 0 \tag{A.1}$$

implies $V = 0$.

The proof is by induction. The statement clearly holds for $\ell = 1$. In fact, if $R(z_j) = 0$, then, by assumption, $R^*(z_j) \neq 0$, and hence, by (A.1), $V(z_j) = 0$. Consequently, we must have $V(z) = \lambda(z)R(z)$ for some real polynomial λ, which inserted into (A.1) yields

$$(\lambda + \lambda^*)RR^* = 0.$$

This implies that $\lambda = 0$ and hence that $V = 0$, as claimed.

Now, suppose that (A.1) implies $V = 0$ for $\ell = k - 1$. Then, for $\ell = k$, (A.1) can be written

$$\left[\begin{array}{c|ccc} r_{11} & r_{12} & \cdots & r_{1k} \\ \hline 0 & & & \\ \vdots & & \hat{R} & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} v_{11}^* & v_{21}^* & \cdots & v_{k1}^* \\ \hline v_{12}^* & & & \\ \vdots & & \hat{V}^* & \\ v_{1k}^* & & & \end{array}\right] + \left[\begin{array}{c|ccc} v_{11} & v_{12} & \cdots & v_{1k} \\ \hline v_{21} & & & \\ \vdots & & \hat{V} & \\ v_{k1} & & & \end{array}\right] \left[\begin{array}{c|cccc} r_{11}^* & 0 & \cdots & 0 & 0 \\ \hline r_{12}^* & & & & \\ \vdots & & \hat{R}^* & & \\ r_{1k}^* & & & & \end{array}\right] = 0, \tag{A.2}$$

which, in particular, contains the $(k-1) \times (k-1)$ matrix relation $\hat{R}\hat{V}^* + \hat{V}\hat{R}^* = 0$ of type (A.1). Consequently, by the induction assumption, $\hat{V} = 0$, so, to prove that $V = 0$, it just remain to show that the border elements $v_{11}, v_{12}, \ldots, v_{1k}, v_{21}, \ldots, v_{k1}$ are all zero. To this end, let us begin with the corner elements v_{1k} and v_{k1}. From the $(1, k)$ and $(k, 1)$ elements in (A.2), we have

$$r_{11}v_{k1}^* + v_{1k}r_{kk}^* = 0 \tag{A.3}$$

$$v_{1k}^*r_{kk} + v_{k1}r_{11}^* = 0. \tag{A.4}$$

In the same way as in the case $\ell = 1$, (A.3) implies that $v_{1k} = \lambda_{1k} r_{11}$ for some real polynomial λ_{1k}, and (A.4) implies that $v_{k1} = \lambda_{k1} r_{kk}$ for some real polynomial λ_{k1}, which inserted into (A.3) yields

$$(\lambda_{k1} + \lambda_{1k}^*) r_{11} r_{kk}^* = 0.$$

This implies that λ_{k1} and λ_{1k} are real numbers such that $\lambda_{1k} = -\lambda_{k1}$. However, by assumption, $V(0)$ is upper triangular, and $R(0)$ is upper triangular and nonsingular. Hence $v_{k1}(0) = 0$ and $r_{kk}(0) \neq 0$, implying that $\lambda_{k1} = v_{k1}(0)/r_{kk}(0) = 0$ and, consequently, $\lambda_{1k} = 0$. Since, therefore, $v_{1k} = 0$ and $v_{k1} = 0$, (A.2) now takes the form

$$\left[\begin{array}{c|c} \tilde{R} & * \\ \hline 0 & * \end{array}\right] \left[\begin{array}{c|c} \tilde{V}^* & 0 \\ \hline 0 & 0 \end{array}\right] + \left[\begin{array}{c|c} \tilde{V} & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c|c} \tilde{R}^* & 0 \\ \hline * & * \end{array}\right] = 0,$$

which only yields the $(k-1) \times (k-1)$ matrix relation $\tilde{R}\tilde{V}^* + \tilde{V}\tilde{R}^* = 0$ of type (A.1). However, by the induction assumption, $\tilde{V} = 0$. Therefore, $V = 0$ in the case $\ell = k$ also, so, by induction, $V = 0$ for all k. □

Appendix B. Computing the gradient

To establish the expression (6.9) in Proposition 6.2 for the gradient

$$\frac{\partial \boldsymbol{J}_P}{\partial \boldsymbol{R}}(\boldsymbol{R}) = 2\left(\boldsymbol{K}\boldsymbol{R} - \frac{\partial}{\partial \boldsymbol{R}} \langle \log\det R, P\rangle\right) \tag{B.1}$$

of (6.7), we need to determine

$$\begin{aligned} \frac{\partial}{\partial R_k} \langle \log\det R, P\rangle &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\partial}{\partial R_k} \log\det R^*(e^{i\theta}) P(e^{i\theta}) d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta} R^*(e^{i\theta})^{-\mathsf{T}} P(e^{i\theta}) d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} P(e^{i\theta}) \big(R^*(e^{i\theta}) R(e^{i\theta})\big)^{-\mathsf{T}} R^{\mathsf{T}}(e^{i\theta}) e^{-ik\theta} d\theta \\ &= \left[\begin{array}{ccccc} C_k^{\mathsf{T}} & \cdots & C_0 & \cdots & C_{n-k} \end{array}\right] \boldsymbol{R} \end{aligned}$$

where C_k is defined by (6.8). This completes the proof of Proposition 6.2.

Next, we explain how to actually compute $C_0, C_1, \dots, C_n$. In view of (4.13),

$$P\big(R^* R\big)^{-1} = \rho\rho^* \Big[(\tau R)^* (\tau R)\Big]^{-1}.$$

We can determine $\hat{C}_0, \hat{C}_1, \dots, \hat{C}_{2n}$ in the expansion

$$\Big[(\tau R)^* (\tau R)\Big]^{-1} = \hat{C}_0 + \hat{C}_1 z + \hat{C}_1^{\mathsf{T}} z^{-1} + \cdots + \hat{C}_{2n} z^{2n} + \hat{C}_{2n}^{\mathsf{T}} z^{-2n} + \cdots.$$

by solving a system of linear equations. Now, defining

$$\mu(z) := \mu_0 + \sum_{s=1}^{n} \mu_\ell (z^s + z^{-s}) = \rho(z)\rho^*(z),$$

we can identify matrix coefficients of equal powers in z in

$$\mu \Big[(\tau R)^* (\tau R)\Big]^{-1} = C_0 + C_1 z + C_1^{\mathsf{T}} z^{-1} + \cdots + C_n z^n + C_n^{\mathsf{T}} z^{-n} + \cdots,$$

to obtain

$$\begin{bmatrix} C_0 \\ C_1 \\ \vdots \\ C_n \end{bmatrix} = \left(\begin{bmatrix} \hat{C}_0 & \hat{C}_1{}^{\mathsf{T}} & \cdots & \hat{C}_n{}^{\mathsf{T}} \\ \hat{C}_1 & \hat{C}_0 & \cdots & \hat{C}_{n-1}{}^{\mathsf{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{C}_n & \hat{C}_{n-1} & \cdots & \hat{C}_0 \end{bmatrix} + \begin{bmatrix} \hat{C}_0 & \hat{C}_1 & \cdots & \hat{C}_n \\ \hat{C}_1 & \hat{C}_2 & \cdots & \hat{C}_{n+1} \\ \vdots & \vdots & \iddots & \vdots \\ \hat{C}_n & \hat{C}_{n+1} & \cdots & \hat{C}_{2n} \end{bmatrix} \right) \begin{bmatrix} \mu_0 I/2 \\ \mu_1 I \\ \vdots \\ \mu_n I \end{bmatrix}.$$

Appendix C. Computing the Hessian

We begin by proving Proposition 6.3. Since

$$\frac{\partial^2 \big(\operatorname{tr} \boldsymbol{R}^{\mathsf{T}} \boldsymbol{K} \boldsymbol{R} \big)}{\partial \operatorname{vec} \boldsymbol{R}^2} = 2(I_\ell \otimes \boldsymbol{K}),$$

it remains to establish (6.12). Since

$$\frac{\partial}{\partial R_j} \otimes \frac{\partial}{\partial R_k} \langle \log \det R, P \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial R_j} \otimes R^*(e^{i\theta})^{-\mathsf{T}} P(e^{i\theta}) e^{-ik\theta} d\theta,$$

(6.12) would follow if we could show that

$$\text{(C.1)} \qquad \frac{\partial}{\partial R_j} \otimes R^*(z)^{-\mathsf{T}} = -z^{-j} \operatorname{vec}(R^*)^{-\mathsf{T}} (\operatorname{vec} R^{-*})^{\mathsf{T}}.$$

Since $R^{-*}(z)^{\mathsf{T}} R^*(z)^{\mathsf{T}} \equiv I$, denoting the (s,t) element of R_j by R_j^{st} we obtain

$$\frac{\partial}{\partial R_j^{st}} R^*(z)^{-\mathsf{T}} = -(R^*)^{-\mathsf{T}} \frac{\partial (R^*)^{\mathsf{T}}}{\partial R_j^{st}} (R^*)^{-\mathsf{T}} = -z^{-j} (R^*)^{-\mathsf{T}} e_s e_t{}^{\mathsf{T}} (R^*)^{-\mathsf{T}},$$

and therefore

$$\begin{aligned} \frac{\partial}{\partial R_j} \otimes R^*(z)^{-\mathsf{T}} &:= \begin{bmatrix} \frac{\partial}{\partial R_j^{11}} R^*(z)^{-\mathsf{T}} & \cdots & \frac{\partial}{\partial R_j^{1\ell}} R^*(z)^{-\mathsf{T}} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial R_j^{\ell 1}} R^*(z)^{-\mathsf{T}} & \cdots & \frac{\partial}{\partial R_j^{\ell\ell}} R^*(z)^{-\mathsf{T}} \end{bmatrix} \\ &= -z^{-j} \operatorname{vec}(R^*)^{-\mathsf{T}} (\operatorname{vec} R^{-*})^{\mathsf{T}}, \end{aligned}$$

establishing (C.1), and hence proving Proposition 6.3.

Next, we answer the question in Remark 6.4. To compute $S_0, S_1, \dots, S_{2n}$, we first expand

$$P(z)(R^{-\mathsf{T}}(z) \otimes R^{-1}(z)) = \cdots + \tilde{S}_{2n} z^{-2n} + \cdots + \tilde{S}_1 z^{-1} + \tilde{S}_0 + \cdots,$$

and transform $\tilde{S}_k$ to the coefficient matrices of $P(\operatorname{vec} R^{-1})(\operatorname{vec} R^{-\mathsf{T}})^{\mathsf{T}}$ by comparing the elements of $R^{-\mathsf{T}} \otimes R^{-1}$ with those of $(\operatorname{vec} R^{-1})(\operatorname{vec} R^{-\mathsf{T}})^{\mathsf{T}}$. The computation of $\tilde{S}_k$ can be done by first observing that

$$\begin{aligned} R^{-\mathsf{T}} \otimes R^{-1} &= (R^*)^{\mathsf{T}} (R^{-*})^{\mathsf{T}} R^{-\mathsf{T}} \otimes (R^* R^{-*} R^{-1}) \\ &= (R^*)^{\mathsf{T}} (R^* R)^{-\mathsf{T}} \otimes (R^* (R R^*)^{-1}) \\ &= ((R^*)^{\mathsf{T}} \otimes R^*)((R^* R)^{-\mathsf{T}} \otimes (R R^*)^{-1}) \\ &= ((R^*)^{\mathsf{T}} \otimes R^*)((R^* R)^{\mathsf{T}} \otimes (R R^*))^{-1} \\ &= ((R^*)^{\mathsf{T}} \otimes R^*)((R^{\mathsf{T}} \otimes R)((R^*)^{\mathsf{T}} \otimes R^*))^{-1} \\ &= ((R^*)^{\mathsf{T}} \otimes R^*)((R^{\mathsf{T}} \otimes R)(R^{\mathsf{T}} \otimes R)^*)^{-1}, \end{aligned}$$

where we have used properties of the Kronecker product that may be found in, e.g., [32]. Multiplying this by P then yields

$$\begin{aligned} P(R^{-\mathsf{T}} \otimes R^{-1}) &= \mu\big((R^*)^\mathsf{T} \otimes R^*\big)\big((\tau R^\mathsf{T} \otimes R)(\tau R^\mathsf{T} \otimes R)^*\big)^{-1} \\ &= \underbrace{(\mu_0 + \mu_1(z + z^{-1}) + \cdots + \mu_n(z^n + z^{-n}))}_{\mu} \\ &\quad \times \underbrace{(U_0 + U_1 z^{-1} + \cdots + U_{2n} z^{-2n})}_{(R^*)^\mathsf{T} \otimes R^*} \underbrace{(T_0 + T_1 z + {T_1}^\mathsf{T} z^{-1} + \cdots)}_{\big((\tau R^\mathsf{T} \otimes R)(\tau R^\mathsf{T} \otimes R)^*\big)^{-1}}, \end{aligned}$$

from which we can compute $\tilde{S}_k$.

References

[1] The matlab codes are available at `http://www.math.kth.se/~andersb/software.html`.

[2] A. Blomqvist and R. Nagamune. An extension of a Nevanlinna-Pick interpolation solver to cases including derivative constraints. In *Proceeding of the 41st IEEE Conference on Decision and Control*, pages 2552–2557, Las Vegas, Nevada, December 2002.

[3] R. W. Brockett. *Finite Dimensional Linear Systems.* John Wiley & Sons, New York, 1970.

[4] C. I. Byrnes, P. Enqvist, and A. Lindquist. Cepstral coefficients, covariance lags and pole-zero models for finite data strings. *IEEE Trans. Signal Processing*, 49(4):677–693, April 2001.

[5] C. I. Byrnes, P. Enqvist, and A. Lindquist. Identifiability and well-posedness of shaping-filter parameterizations: A global analysis approach. *SIAM J. Contr. and Optimiz.*, 41(1):23–59, 2002.

[6] 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.

[7] C. I. Byrnes, T. T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 46(6):822–839, June 2001.

[8] C. I. Byrnes, S. V. Gusev, and A. Lindquist. A convex optimization approach to the rational covariance extension problem. *SIAM J. Contr. and Optimiz.*, 37(1):211–229, 1998.

[9] 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.

[10] C. I. Byrnes and A. Lindquist. Interior point solutions of variational problems and global inverse function theorems. Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden, Report TRITA/MAT-01-OS13, 2001.

[11] 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.

[12] C. I. Byrnes and A. Lindquist. A convex optimization approach to generalized moment problems. In Y. Oishi K. Hashimoto and Y. Yamamoto, editors, *Control and Modeling of Complex Systems: Cybernetics in the 21st Century*, pages 3–21. Birkhäuser, Boston, 2003.

[13] 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.

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

[15] M. K. K. Cevik and J. M. Schumacher. The Robust Regulation Problem with Robust Stability. Technical Report MAS-R9827, CWI, Amsterdam, 1999.

[16] B.-C. Chang and J. B. Pearson. Optimal Disturbance Reduction in Linear Multivariable Systems. *IEEE Trans. Automat. Control*, 29(10):880–887, October 1984.

[17] E. J. Davison, editor. *Benchmark problems for control system design*, May 1990. Report of the IFAC theory committee.

[18] Ph. Delsarte, Y. Genin, and Y. Kamp. Orthogonal Polynomial Matrices on the Unit Circle. *IEEE Trans. Circuits and Systems*, 25(3):149–160, March 1978.

[19] Ph. Delsarte, Y. Genin, and Y. Kamp. The Nevanlinna-Pick Problem for Matrix-valued Functions. *SIAM J. Appl. and Math.*, 36:47–61, Feb 1979.

[20] Ph. Delsarte, Y. Genin, and Y. Kamp. Schur Parametrization of Positive Definite Block-Toeplitz Systems. *SIAM J. Appl. and Math.*, 36:34–46, Feb 1979.

[21] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum. *Feedback Control Theory.* Macmillan Publishing Company, New York, 1992.

[22] P. Enqvist. A homotopy approach to rational covariance extension with degree constraint. *Int. J. Applied Mathematics and Computer Science*, 11(5):1173–1201, 2001.

[23] B. A. Francis. *A Course in H_∞ Control Theory.* Lecture Notes in Control and Information Sciences. Springer-Verlag, 1987.

[24] P. Gahinet and P. Apkarian. Robustness analysis using singular value sensitivities. *Int. J. Robust and Nonlinear Control*, 4:421–448, 1994.

[25] F. R. Gantmacher. *The Theory of Matrices.* Chelsea, New York, 1959.

[26] Y. Genin, Y. Hachez, Y. Nesterov, and P. Van Dooren. Optimization problems over positive pseudo-polynomial matrices. Submitted to the SIAM Journal on Matrix Analysis and Applications, 2000. See http://www. auto. ucl. ac. be/ vdooren.

[27] Y. Genin, Y. Hachez, Y. Nesterov, R. Stefan, P. Van Dooren, and S. Xu. Positivity and linear matrix inequalities. *European Journal of Control*, 8:275–298, 2002.

[28] T. T. Georgiou. Realization of power spectra from partial covariance sequences. *IEEE Trans. Acoustics, Speech and Signal Processing*, 35:438–449, 1987.

[29] T. T. Georgiou. A Topological Approach to Nevanlinna-Pick Interpolation. *SIAM J. Math. and Anal.*, 18(5):1248–1260, 1987.

[30] T. T. Georgiou. The Interpolation Problem with a Degree Constraint. *IEEE Trans. Automat. Control*, 44(3):631–635, March 1999.

[31] B. K. Ghosh. Transcendental and interpolation methods in simultaneous stabilization and simultaneous partial pole placement problems. *SIAM J. Contr. and Optimiz.*, 24:1091–1109, 1986.

[32] A. Graham. *Kronecker products and matrix calculus with applications.* John Wiley & Sons, 1981.

[33] J. Hadamard. Sur les correspondances ponctuelles. In *Oeuvres, Editions du Centre Nationale de la Researche Scientifique*, pages 383–384. Paris, 1968.

[34] D. Henrion and M. Šebek. Efficient numerical method for the discrete-time symmetric matrix polynomial equation. *IEE Proc. Control Theory Appl.*, 145(5):443–447, 1998.

[35] T. Iwasaki and R. E. Skelton. All controllers for the general h_∞ control problem: LMI existence conditions and state space formulas. *Automatica J. IFAC*, 30(8):1307–1317, 1994.

[36] Jan Ježek. Symmetric matrix polynomial equations. *Kybernetika (Prague)*, 22(1):19–30, 1986.

[37] R. E. Kalman. Realization of covariance sequences. In *Proc. Toeplitz Memorial Conference*, Tel Aviv, Israel, 1981.

[38] P. P. Khargonekar and A. Tannenbaum. Non-Euclidian Metrics and the Robust Stabilization of Systems with Parameter Uncertainty. *IEEE Trans. Automat. Control*, 30(10):1005–1013, October 1985.

[39] H. Kimura. Robust stabilizability for a class of transfer functions. *IEEE Trans. Automat. Control*, 29(10):788–793, October 1984.

[40] H. Kimura. Conjugation, interpolation and model-matching in H^∞. *Int. J. Control*, 49:269–307, 1989.

[41] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01.

[42] D. J. N. Limebeer and B. D. O. Anderson. An Interpolation Theory Approach to H^∞ Controller Degree Bounds. *Linear Algebra and Its Application*, 98:347–386, 1988.

[43] A. Lindquist. A new algorithm for optimal filtering of discrete-time stationary processes. *SIAM J. Control*, 12(4):736–746, 1974.

[44] J. M. Maciejowski. *Multivariable Feedback Design.* Addison-Wesley, Wokingham U.K., 1989.

[45] Bruce C. Moore. Principal component analysis in linear systems: controllability, observability, and model reduction. *IEEE Trans. Automat. Control*, 26(1):17–32, 1981.

[46] R. Nagamune. Closed-loop shaping based on the Nevanlinna-Pick interpolation with a degree bound. submitted to IEEE Trans. Automatic Control.
[47] R. Nagamune. Sensitivity reduction for SISO systems using the Nevanlinna-Pick interpolation with degree constraint. In *Proceedings of 14th International Symposium of Mathematical Theory of Networks and Systems*, Perpignan, France, 2000.
[48] R. Nagamune. Simultaneous robust regulation and robust stabilization with degree constraint. In *Proceedings of 15th International Symposium of Mathematical Theory of Networks and Systems*, University of Notre Dame, Indiana, August 2002.
[49] R. Nagamune. A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 48(1):113–117, January 2003.
[50] R. Nagamune and A. Lindquist. Sensitivity shaping in feedback control and analytic interpolation theory. In J.L. Medaldi et al, editor, *Optimal Control and Partial Differential Equations*, pages 404–413. IOS Press, Amsterdam, 2001.
[51] A. V. Oppenheim and R. W. Shafer. *Digital Signal Processing*. Prentice Hall, London, 1975.
[52] M. M. Seron, J. H. Braslavsky, and G. C. Goodwin. *Fundamental Limitations in Filtering and Control*. Springer-Verlag, 1997.
[53] R. E. Skelton, T. Iwasaki, and K. Grigoriadis. *A Unified Algebraic Approach to Linear Control Design*. Taylor & Francis, 1998.
[54] T. Söderström and P. Stoica. *System Identification*. Prentice Hall, 1989.
[55] A. Tannenbaum. Feedback stabilization of linear dynamical plants with uncertainty in the gain factor. *Int. J. Control*, 32(1):1–16, 1980.
[56] A. Tannenbaum. Modified Nevanlinna-Pick interpolation and feedback stabilization of linear plants with uncertainty in the gain factor. *Int. J. Control*, 36(2):331–336, 1982.
[57] P. Whittle. On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika*, 50:129–134, 1963.
[58] R. A. Wiggins and E. A. Robinson. Recursive solution to the multichannel filtering problem. *Journal Geophysical Research*, 70:1885–1891, 1966.
[59] D. C. Youla and M. Saito. Interpolation with positive-real functions. *J. Franlkin Institute*, 284:77–108, 1967.
[60] K. Zhou. *Essentials of Robust Control*. Prentice-Hall, New Jersey, 1998.