SIAM J. MATRIX ANAL. APPL.
Vol. 25, No. 1, pp. 57–79

OPTIMIZATION PROBLEMS OVER POSITIVE PSEUDOPOLYNOMIAL MATRICES*

Y. GENIN†, Y. HACHEZ†, YU. NESTEROV†, AND P. VAN DOOREN†

Abstract. The Nesterov characterizations of positive pseudopolynomials on the real line, the imaginary axis, and the unit circle are extended to the matrix case. With the help of these characterizations, a class of optimization problems over the space of positive pseudopolynomial matrices is considered. These problems can be solved in an efficient manner due to the inherent block Toeplitz or block Hankel structure induced by the characterization in question. The efficient implementation of the resulting algorithms is discussed in detail. In particular, the real line setting of the problem leads naturally to ill-conditioned numerical systems. However, adopting a Chebyshev basis instead of the natural basis for describing the polynomial matrix space yields a restatement of the problem and of its solution approach with much better numerical properties.

Key words. convex optimization, positive polynomials, Toeplitz matrices, Hankel matrices

AMS subject classifications. 47A68, 65F30, 65Y20, 90C22

PII. S0895479803374840

1. Introduction.

This paper is concerned with a convex optimization problem over the set of polynomial matrices, which are nonnegative definite on distinguished contours of the complex plane, namely, the real line, the imaginary axis, and the unit circle. The set of such polynomial matrices is convex. Moreover, it has been shown by Nesterov that scalar polynomials of this type [12] admit a compact parametrization in terms of constant nonnegative definite matrices satisfying simple linear algebraic constraints.

The aim of this paper is to extend this parametrization to the matrix case and, with the help of this result, to discuss and to solve an important class of related convex optimization problems. In fact, the dual formulation of these optimization problems appears to be considerably more attractive from a computational viewpoint. On the one hand, it is stated in an optimization space of reduced dimension. On the other hand, this dual space is characterized by nonnegative definite matrices that have block Hankel or block Toeplitz structure.

A well-established technique for solving such optimization problems involves the introduction of a barrier function [13] whose differential characteristics have to be repeatedly evaluated along the numerical optimization process. Due to the Hankel or Toeplitz structure of the optimization space, fast, and even superfast, algorithms, based on displacement rank techniques, can be proposed for that purpose. The computational aspects of their implementation are discussed in some detail. In addition, as the real line formulation of the problem is shown to be inherently ill-conditioned, a change of polynomial basis is considered and discussed. This problem reformulation

*Received by the editors July 4, 2000; accepted for publication (in revised form) by M. Hyman January 3, 2003; published electronically May 15, 2003. This research was supported by NSF contract CCR-97-96315 and by the Belgian Programme on Inter-university Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors.

http://www.siam.org/journals/simax/25-1/37484.html

†Université Catholique de Louvain, Department of Mathematical Engineering, CESAME, avenue Georges Lemaître 4, B-1348 Louvain-la-Neuve, Belgium (genin@csam.ucl.ac.be, hachez@csam.ucl.ac.be, nesterov@csam.ucl.ac.be, vdooren@csam.ucl.ac.be). The research of the second author was supported by a research fellowship from the Belgian National Fund for Scientific Research.

exhibits much more interesting numerical prospects from this viewpoint.

The theory of positive transfer functions is well known for playing a fundamental role in systems and control theory. Such functions represent, e.g., spectral density functions of stochastic processes, appear in spectral factorizations, and also are related to the Riccati equations. It has been known since the work of Youla [17] that, when such transfer functions are rational, they possess rational spectral factorizations. Later on, it was shown that, using state-space models of positive transfer functions, one could express the condition of positivity in terms of linear matrix inequalities (see, e.g., [16]). Positive transfer functions obviously form a convex set, and they were recently studied in the convex optimization literature [4, 12]. The parametrization of pseudopolynomial matrices proposed in this paper fits naturally into that context. In particular, this parametrization can also be obtained as a straightforward application of the celebrated positive real lemma to an appropriate subset of positive paraconjugate transfer functions.

In section 2, the definition of positive paraconjugate transfer functions is given; in particular, such functions are well known for enjoying a remarkable spectral factorization property.

In section 3, positive pseudopolynomial matrices on the real line, the imaginary axis, and the unit circle are considered. In each case, the positivity constraint is shown to induce some form of symmetry on the pseudopolynomial matrix coefficients and to impose some restrictions on their formal degree.

In section 4, parametrizations of nonnegative pseudopolynomial matrices are derived in terms of appropriate subsets of nonnegative constant Hermitian matrices. These appropriate subsets are defined by linear algebraic relations and can be parametrized in terms of an arbitrary Hermitian or skew-Hermitian constant matrix of reduced dimension, depending on the particular contour of the real plane considered.

In section 5, an alternative proof of this parametrization is derived from the theory of positive paraconjugate transfer functions. In particular, with the help of the positive real lemma, any state-space realization of such a function is proved to involve some degree of freedom, which can be expressed in terms of a linear matrix inequality (LMI). This is precisely the characterization obtained in the preceding section.

In section 6, a class of important optimization problems is defined over the set of nonnegative pseudopolynomial matrices satisfying linear constraints. These constraints are assumed to be expressible in terms of Frobenius scalar products. Next, the dual form of these optimization problems is shown to be computationally much more attractive. The dimension of the optimization space appears to be reduced to the number of linear constraints instead of the pseudopolynomial matrix dimension, as in the primal form. In addition, this optimization space is characterized by nonnegative definite block Hankel or block Toeplitz matrices, depending on the particular complex plane contour considered. Furthermore, modern techniques for the numerical solution of the optimization problem involve the introduction of a barrier function. Since the appropriate barrier functions inherit the Hankel or Toeplitz structure of the optimization space, this paves the way for fast evaluations of the differential characteristics of the barrier function. Such fast evaluations are of paramount importance because they have to be made repeatedly in such optimization schemes [4, 5, 12, 13].

In section 7, the computational aspects of the fast algorithms which can be used to solve these optimization problems are considered and analyzed in detail. The optimization scheme mainly involves recurrent computations of the differential characteristics of the barrier function, namely, its gradient and its Hessian. These differential functions are evaluated by carrying out Frobenius scalar products of appropriate block Hermitian matrices with underlying Hankel or Toeplitz structure. Displacement rank techniques are especially suited for their fast evaluation. In particular, the required calculations can be broken down into fast, or even superfast, elementary numerical operations by exploiting the compact displacement rank representations resulting from the problem structure. It is also pointed out that the real line problem is inherently ill-conditioned. This fact is a well-known consequence of the Hankel structure.

In section 8, the real line optimization problem is reformulated to get around the above technical difficulty. As the Hankel structure is an obvious consequence of the expansion of polynomial matrices into the natural basis of their monomials $[I_m, x\, I_m, x^2\, I_m, \ldots]$, the remedy consists in a change of basis. In this light, it is proposed to substitute a basis of Chebyshev polynomials for the natural basis. Such a Chebyshev basis induces a Toeplitz-plus-Hankel structure to the problem with, in principle, a much better numerical conditioning. It is finally recalled how one can take advantage of the Toeplitz-plus-Hankel structure in fast algorithms based on appropriate displacement rank techniques [6].

2. Paraconjugate transfer functions. Paraconjugate transfer functions $\Phi(.)$ play an important role in systems theory. They are defined with respect to a curve in the complex plane, which is typically the imaginary axis (for continuous-time systems), the unit circle (for discrete-time systems), and the real axis $\mathbb{R}$ (for the moment problem).

Imaginary axis. This curve is the boundary of the stable region for continuous-time transfer functions in the complex variable s (which is also the variable of the Laplace transform of such dynamical systems): the imaginary axis is denoted $s \in j\mathbb{R}$.

Unit circle. This curve is the boundary of the stable region for discrete-time transfer functions in the complex variable z (which is also the variable of the so-called z-transform of such dynamical systems): the unit circle is denoted $z \in e^{j\mathbb{R}}$.

Real axis. This curve occurs in the standard treatment of the classical moment problem [1, 11]. In this case, the complex variable x will be used with the real axis and denoted $x \in \mathbb{R}$.

To stress that a result holds for a particular curve, the above particular variable notation will be adopted instead of the standard variable p. In this paper, only the case of square *rational* transfer matrices $\Phi(p)$ will be considered, i.e., $m \times m$ matrices $\Phi(p)$ whose entries are rational functions of the variable p.

DEFINITION 2.1. *The paraconjugate transfer function $\Phi_*(p)$ of a given transfer matrix $\Phi(p)$ is defined as follows:*

$$\Phi_*(s) = [\Phi(-\overline{s})]^* \text{ for the imaginary axis,}$$
$$\Phi_*(z) = [\Phi(1/\overline{z})]^* \text{ for the unit circle,}$$
$$\Phi_*(x) = [\Phi(\overline{x})]^* \text{ for the real axis,}$$

where M^* is the conjugate transposed matrix of a matrix M.

Let us point out that the paraconjugate $\Phi_*(p)$ is also a rational transfer function of the complex variable p. A para-Hermitian transfer function can then be defined as follows.

DEFINITION 2.2. *A square transfer function $\Phi(p)$ is para-Hermitian if it is equal to its paraconjugate: $\Phi_*(p) = \Phi(p)$.*

This definition depends on the choice of curve considered. However, a para-Hermitian transfer function *evaluated* on the corresponding curve is always a Hermitian matrix. Indeed, $\Phi_*(p) = \Phi(p)$ implies the following for each case:

$$\begin{aligned} \Phi_*(j\omega) &= [\Phi(j\omega)]^* \text{ for } s = j\omega \text{ on the imaginary axis,} \\ \Phi_*(e^{j\omega}) &= [\Phi(e^{j\omega})]^* \text{ for } z = e^{j\omega} \text{ on the unit circle,} \\ \Phi_*(\omega) &= [\Phi(\omega)]^* \text{ for } x = \omega \text{ on the real axis,} \end{aligned}$$

where $\omega \in \mathbb{R}$ is thus a real variable parametrizing the curve.

Since a paraconjugate transfer function is a Hermitian matrix when evaluated on the curve, all its eigenvalues are real. Therefore, a positivity constraint can be imposed on these eigenvalues. This leads to the following definition.

DEFINITION 2.3. *A paraconjugate transfer function is* positive *(*nonnegative*) if it is positive (*nonnegative*) when evaluated on the curve:* $\Phi(p) \succ 0$ ($\Phi(p) \succeq 0$).

Note that nonnegative paraconjugate transfer functions always possess a so-called *spectral factorization*,

$$\Phi(p) = G_*(p)G(p), \tag{2.1}$$

where the spectral factor $G(p)$ is again a square rational transfer function in p. This result is proven in the systems theory literature [17, 14].

3. Positive pseudopolynomial matrices. Pseudopolynomial matrices are matrices with a finite expansion in positive and negative powers of the independent variable p:

$$\Phi(p) = \sum_{k=-r}^{t} \Phi_k p^k.$$

Depending on the type of curve one considers, the coefficient matrices of such pseudopolynomial matrices must possess a certain symmetry.

Real axis. For a para-Hermitian transfer function $\Phi(x)$ that is nonnegative on the real axis $x \in \mathbb{R}$, it follows from the para-Hermitian nature that the coefficient matrices of the expansion

$$\Phi(x) = \sum_{k=-r}^{t} \Phi_k x^k \tag{3.1}$$

must all be Hermitian: $\Phi_k = \Phi_k^*$. Moreover, since x^2 is nonnegative on the real axis $x \in \mathbb{R}$, such pseudopolynomial matrices can be reduced to polynomial matrices in x or in x^{-1}; in particular, they reduce to the form

$$\Phi(x) = \sum_{k=0}^{t} \Phi_k x^k.$$

From the nonnegativity of $\Phi(x)$, it turns out that the highest degree coefficient must be of even degree $t = 2n$. For polynomial matrices in x^{-1}, the highest degree coefficient is also of even degree. The standard form used here for nonnegative para-Hermitian matrices on the real axis is

$$\Phi(x) = \sum_{k=0}^{2n} \Phi_k x^k, \quad \Phi_k = \Phi_k^*. \tag{3.2}$$

Unit circle. For a para-Hermitian transfer function $\Phi(z)$ that is nonnegative on the unit circle $z \in e^{j\mathbb{R}}$, it follows from the para-Hermitian nature that the coefficient matrices of the expansion

$$\Phi(z) = \sum_{k=-r}^{t} \Phi_k z^k \tag{3.3}$$

must satisfy the condition $\Phi_{-k} = \Phi_k^*$; thus such a pseudopolynomial matrix must have a symmetric expansion. The standard form used here for nonnegative para-Hermitian matrices on the unit circle is

$$\Phi(z) = \sum_{k=-n}^{n} \Phi_k z^k, \quad \Phi_{-k} = \Phi_k^*. \tag{3.4}$$

Imaginary axis. For a para-Hermitian transfer function $\Phi(s)$ that is nonnegative on the imaginary axis $s \in j\mathbb{R}$, it follows from the para-Hermitian nature that the coefficient matrices of the expansion

$$\Phi(s) = \sum_{k=-r}^{t} \Phi_k s^k \tag{3.5}$$

are Hermitian if k is even and are skew-Hermitian if k is odd:

$$\Phi_{2k} = \Phi_{2k}^*, \quad \Phi_{2k+1} = -\Phi_{2k+1}^*.$$

This follows easily from the change of variables $s = jx$ converting the real axis into the imaginary axis. One can again multiply by a power of $-s^2$ (which is nonnegative on the imaginary axis) to obtain a polynomial matrix in s or s^{-1},

$$\Phi(s) = \sum_{k=0}^{t} \Phi_k s^k,$$

and it is easy to see from the nonnegativity that the highest degree coefficient must be of even degree $t = 2n$. For polynomial matrices in s^{-1} the highest degree coefficient is also of even degree. The standard form we use here for nonnegative para-Hermitian matrices on the imaginary axis is

$$\Phi(x) = \sum_{k=0}^{2n} \Phi_k x^k, \quad \Phi_{2k} = \Phi_{2k}^*, \quad \Phi_{2k+1} = -\Phi_{2k+1}^*. \tag{3.6}$$

To end this section, let us observe that the pseudopolynomial matrices of interest have, in the above cases, $(2n+1)m^2$ degrees of freedom.

4. Parametrization of nonnegative pseudopolynomial matrices. The main result of this section highlights a parametrization of nonnegative pseudopolynomial matrices in terms of constant Hermitian or skew-Hermitian matrices.

To begin and, for further use, let us introduce two particular $(n+1)m \times (n+1)m$ block matrices: the standard block shift operator

$$Z \doteq \begin{bmatrix} 0 & I_m & & \\ & 0 & \ddots & \\ & & \ddots & I_m \\ & & & 0 \end{bmatrix}$$

on the one hand, and the degenerate matrix

$$X \doteq \left[\begin{array}{c|c} X_0 & 0 \\ \hline 0 & 0 \end{array}\right], \tag{4.1}$$

with X_0 any $nm \times nm$ complex matrix on the other hand.

4.1. Real axis. Let

$$P(x) = \sum_{k=0}^{2n} P_k x^k \tag{4.2}$$

be an $m \times m$ para-Hermitian polynomial matrix with Hermitian coefficients, i.e., $P_k = P_k^*$, and consider the set of Hermitian matrices

$$Y = \begin{bmatrix} Y_{0,0} & Y_{0,1} & \cdots & Y_{0,n} \\ Y_{1,0} & Y_{1,1} & \cdots & Y_{2,n} \\ \vdots & \vdots & & \vdots \\ Y_{n,0} & Y_{n,1} & \cdots & Y_{n,n} \end{bmatrix},$$

with blocks of dimension $m \times m$. If $\Pi(x)$ stands for

$$\Pi(x) = \begin{bmatrix} I_m & xI_m & \cdots & x^n I_m \end{bmatrix}^T,$$

the relation

$$\Pi_*(x) Y \Pi(x) = P(x) \tag{4.3}$$

implies that

$$P_k = \sum_{i+j=k} Y_{i,j}, \quad k = 0, \ldots, 2n, \tag{4.4}$$

within the convention that $Y_{i,j} = 0$ for i and j outside their definition range. A simple choice for Y so as to obtain this identity is found to be

$$Y_0 = \begin{bmatrix} P_0 & \frac{1}{2}P_1 & & \\ \frac{1}{2}P_1 & P_2 & \ddots & \\ & \ddots & \ddots & \frac{1}{2}P_{2n-1} \\ & & \frac{1}{2}P_{2n-1} & P_{2n} \end{bmatrix}. \tag{4.5}$$

Then, the following characterization theorem can be stated.

THEOREM 4.1. *A Hermitian matrix Y satisfies* (4.3) *if and only if it can be expressed as*

$$Y = Y_0 + Z^T X - XZ, \tag{4.6}$$

where X has the form (4.1) *and is skew-Hermitian, i.e., $X = -X^*$.*

Proof. The "if" part is obvious since one has $\Pi_*(x)\left[Z^TX - XZ\right]\Pi(x) = 0$ for any matrix X of the form (4.1). Conversely, let Y be a solution of (4.3) and let us set X as

$$X = \sum_{k=0}^{n}(Z^{k+1})(Y - Y_0)(Z^k). \tag{4.7}$$

It turns out that X has the structure (4.1) with $X = -X^*$ and satisfies (4.6). To see this, observe first that X has the structure (4.1) as an immediate consequence of relations (4.4). Next, inserting (4.7) in (4.6), one obtains successively

$$\begin{aligned} Y_0 + Z^TX - XZ &= Y_0 + Z^TZ\sum_{k=0}^{n} Z^k(Y - Y_0)Z^k - \sum_{k=0}^{n} Z^{k+1}(Y - Y_0)Z^{k+1} \\ &= Y_0 + Z^TZ(Y - Y_0) + (Z^TZ - I_{(n+1)m})\sum_{k=0}^{n-1} Z^{k+1}(Y - Y_0)Z^{k+1} \\ &= Y_0 + (Y - Y_0) \\ &= Y \end{aligned}$$

again in view of relations (4.4). Finally, one establishes the skew-Hermitian property of X from the fact that $Z^TX - XZ = X^*Z - Z^TX^*$ necessarily implies $X = -X^*$ for any matrix X of algebraic structure (4.1). ∎

Imposing the condition that $P(x)$ is also a nonnegative transfer function leads to the following theorem.

THEOREM 4.2. *A pseudopolynomial matrix $P(x) = \sum_{k=0}^{2n} P_kx^k$ is nonnegative definite on the real axis if and only if there exists a nonnegative definite Hermitian matrix Y with blocks $Y_{i,j}, i, j = 0, \ldots, n$, such that ($Y_{i,j} = 0$ for i and j outside their definition range)*

$$P_k = \sum_{i+j=k} Y_{i,j} \quad \text{for } k = 0, \ldots, 2n. \tag{4.8}$$

Proof. Because of the previous theorem, the "only if" part only needs a proof. It is obtained from the existence of a spectral factorization

$$P(x) = G_*(x)G(x),$$

where $G(x)$ is polynomial in x: $G(x) = \sum_{k=0}^{n} G_nx^n$. Indeed, choose

$$Y = \begin{bmatrix} G_0 & G_1 & \cdots & G_n \end{bmatrix}^* \begin{bmatrix} G_0 & G_1 & \cdots & G_n \end{bmatrix}.$$

This matrix Y is nonnegative and satisfies the constraints of the theorem. ∎

Let us point out that if $\det(P(x))$ has zeros, then Y cannot be strictly positive definite. This characterization of matrix polynomials nonnegative on the real axis extends a result obtained earlier by Nesterov [12] for scalar polynomials.

4.2. Unit circle. Let us now consider the case of the nonnegative transfer functions on the unit circle. It follows from its finite expansion and from its para-Hermitian character that such a pseudopolynomial matrix

$$P(z) = \sum_{k=-n}^{n} P_kz^k \tag{4.9}$$

has $m \times m$ coefficient matrices that satisfy $P_{-k} = P_k^*$. The set of Hermitian matrices of interest here is defined by the equation

$$\Pi_*(z) Y \Pi(z) = P(z), \tag{4.10}$$

where the same notation as above is used for the matrix Y and $\Pi(.)$. This is algebraically equivalent to the relations

$$P_k = \sum_{i-j=k} Y_{i,j}, \tag{4.11}$$

assuming $Y_{i,j} = 0$ for i and j outside their definition range. Clearly, the choice

$$Y_0 = \begin{bmatrix} P_0 & P_1 & \cdots & P_n \\ P_1^* & 0 & \vdots & 0 \\ \vdots & \vdots & & \vdots \\ P_n^* & 0 & \cdots & 0 \end{bmatrix} \tag{4.12}$$

is an admissible matrix Y. The characterization theorem now takes the following form.

THEOREM 4.3. *A Hermitian matrix Y satisfies* (4.10) *if and only if it can be expressed as*

$$Y = Y_0 + X - Z^T X Z, \tag{4.13}$$

where X has the form (4.1) *and is Hermitian, i.e., $X = X^*$.*

Proof. By duplicating the argument used in the proof of Theorem 4.1, one shows that the solution X of (4.13) is given by

$$X = \sum_{k=0}^{n} (Z^k)^T (Y - Y_0)(Z^k)$$

and that the resulting matrix X has the stated form because of (4.11). ☐

The positive pseudopolynomial matrices on the unit circle can then be characterized as follows.

THEOREM 4.4. *A pseudopolynomial matrix $P(z) = \sum_{k=-n}^{n} P_k z^k$ is nonnegative definite on the unit circle if and only if there exists a nonnegative definite Hermitian matrix Y with blocks $Y_{i,j}, i, j = 0, \ldots, n$, such that (assuming $Y_{i,j} = 0$ for i and j outside their definition range)*

$$P_k = \sum_{i-j=k} Y_{i,j} \quad \textit{for } k = -n, \ldots, 0, \ldots, n. \tag{4.14}$$

The proof of this theorem is again based on the same spectral factorization argument as in Theorem 4.2 and is therefore omitted. This characterization of pseudopolynomials nonnegative on the unit circle also extends a result previously obtained by Nesterov [12] for trigonometric polynomials.

4.3. Imaginary axis. The third kind of nonnegative pseudopolynomial matrices is that with respect to the imaginary axis. This formulation of the problem does not require any specific treatment since it can be reduced to the case of the real axis in a straightforward manner. Indeed, consider the para-Hermitian polynomial matrix

$$P(s) = \sum_{k=0}^{2n} P_k s^k \tag{4.15}$$

with $s \in j\mathbb{R}$. If $s = jx$, one derives from $P(s)$ the para-Hermitian polynomial matrix

$$\hat{P}(x) = \sum_{k=0}^{2n} (j^k P_k) x^k = \sum_{k=0}^{2n} \hat{P}_k x^k$$

with respect to the real line. In particular, this implies $P_k^* = (-1)^k P_k$ for all k. Therefore, applying Theorem 4.2 to $\hat{P}(x)$, one obtains for $P(s)$ the following result.

THEOREM 4.5. *A pseudopolynomial matrix $P(s) = \sum_{k=0}^{2n} P_k s^k$ is nonnegative on the imaginary axis if and only if there exists a nonnegative definite Hermitian matrix Y with blocks $Y_{i,j}, i, j = 0, \ldots, n$, such that ($Y_{i,j} = 0$ for i and j outside their definition range)*

$$P_k = (-j)^k \sum_{i+j=k} Y_{i,j} \quad \text{for } k = 0, \ldots, 2n.$$

5. Positive paraconjugate transfer functions. The parametrization of positive pseudopolynomial matrices, derived in the preceding section, can alternatively be obtained from the theory of positive paraconjugate transfer functions. More precisely, it follows from a straightforward application of the celebrated positive real lemma to the subclass of positive paraconjugate transfer functions that has a pseudopolynomial form.

To see this, let us start from a well-known result of state-space theory [14] that states that any proper paraconjugate transfer function admits minimal realizations of the form

$$\Phi(s) = \left[\begin{array}{cc} B^*(-sI_n - A^*)^{-1}, & I_m \end{array} \right] Y_0 \left[\begin{array}{c} (sI_n - A)^{-1}B \\ I_m \end{array} \right], \tag{5.1}$$

where Y_0 is some appropriate Hermitian matrix. Note that the assumption $\Phi(s)$ proper (i.e., $\Phi(s)$ bounded at $s = \infty$) is made for the sake of simplicity and could be lifted with the help of generalized state-space representations or with an appropriate transformation of the variable s. Clearly, Y_0 is not uniquely defined from $\Phi(s)$. Indeed, replace the matrix Y_0 with the matrix $Y(\tilde{X})$ defined as follows:

$$Y(\tilde{X}) = Y_0 + \left[\begin{array}{cc} \tilde{X}A + A^*\tilde{X} & \tilde{X}B \\ B^*\tilde{X} & 0 \end{array} \right], \tag{5.2}$$

where $\tilde{X}$ is any $n \times n$ block Hermitian matrix. The transfer function $\Phi(s)$ is easily verified by direct inspection not to be affected by this substitution, which clearly preserves the Hermitian property of the realization.

The well-known positive real lemma [8, 14, 18] states that the existence of a Hermitian matrix $\tilde{X}$ such that $Y(\tilde{X})$ is nonnegative definite is a necessary and sufficient condition for $\Phi(s)$ to be a para-Hermitian transfer function nonnegative on the whole of the imaginary axis. Let us apply this result to the transfer function

$$\Phi(s) = [-jE(-sI_n + jZ^T)^{-1}, I_m]\, Y_0 \left[\begin{array}{c} (sI_n - jZ)^{-1}jE^T \\ I_m \end{array} \right], \tag{5.3}$$

where $E = [0, \ldots, 0, I_m]$, and Y_0 is defined as in (4.5). Since $P_k = P_k^*$ for all k by assumption, $\Phi(s)$ is a well-defined paraconjugate transfer function. Moreover, one has by construction the relation

$$\Phi(jx) = x^{-2n} \sum_{k=0}^{2n} P_k x^k = x^{-2n} P(x).$$

Therefore, $\Phi(s)$ is a nonnegative paraconjugate transfer function if and only if $P(x)$ is a nonnegative polynomial matrix. In view of the positive real lemma, it finally appears that $P(x)$ is nonnegative if and only if there exists a Hermitian matrix $\tilde{X}$ such that the Hermitian matrix

$$Y(\tilde{X}) = Y_0 + \begin{bmatrix} j\tilde{X}Z - jZ^T\tilde{X} & j\tilde{X}E^T \\ -jE\tilde{X} & 0 \end{bmatrix}$$

is nonnegative definite. If one sets $X_0 \doteq -j\tilde{X}$, this is precisely the characterization provided by Theorems 4.1 and 4.2.

An alternative proof of Theorems 4.3 and 4.4 can be obtained on the basis of a similar argument. Consider a state-space realization of a paraconjugate transfer function of the form

$$\Phi(z) = \begin{bmatrix} zB^*(I_n - zA^*)^{-1}, & I_m \end{bmatrix} Y_0 \begin{bmatrix} (zI_n - A)^{-1}B \\ I_m \end{bmatrix} \tag{5.4}$$

with Y_0 some Hermitian matrix. Incidentally, this realization can also be deduced from (5.1) by means of the variable transformation $s = (z-1)/(z+1)$, which maps the unit circle onto the imaginary axis. The transfer function $\Phi(z)$ is nonnegative on the unit circle if the matrix $\Phi(e^{j\theta})$ is nonnegative definite for all θ in the interval $[0, 2\pi]$. In this setting, the positive real lemma states that $\Phi(z)$ will be a well-defined nonnegative paraconjugate transfer function if and only there exists a Hermitian matrix $\tilde{X}$ such that

$$Y(\tilde{X}) = Y_0 + \begin{bmatrix} A^*\tilde{X}A - \tilde{X} & A^*\tilde{X}B \\ B^*\tilde{X}A & B^*\tilde{X}B \end{bmatrix} \tag{5.5}$$

is nonnegative definite. With Y_0 as in (4.12), $A = Z$, and $B = E^T$, the following equality holds:

$$\Phi(z) = \sum_{k=-n}^{+n} P_k z^k.$$

Therefore, the pseudopolynomial matrix $P(z)$ is found to be nonnegative definite on the unit circle if and only if there exists a Hermitian matrix $\tilde{X}$ such that the matrix

$$Y(\tilde{X}) = Y_0 + \begin{bmatrix} Z^T\tilde{X}Z - \tilde{X} & Z^T\tilde{X}E^T \\ E\tilde{X}Z & E\tilde{X}E^T \end{bmatrix}$$

is nonnegative definite. Here again, this is exactly the characterization proposed in the previous section provided one substitutes $\tilde{X}$ for $-X_0$.

6. The optimization problem. The optimization problems considered in this paper are assumed to be stated in terms of appropriate scalar products defined over the space of complex matrices. For any couple of matrices X and Y let us set their scalar product as follows:

$$\langle X, Y\rangle \doteq \mathrm{Re}(\mathrm{Trace}\, XY^*) \equiv \mathrm{Re}\sum_i \sum_j x_{i,j}\overline{y}_{i,j}, \tag{6.1}$$

where $x_{i,j}$ and $y_{i,j}$ are the scalar entries of the matrices X and Y, respectively. It follows from this definition that

$$\langle X, Y\rangle = \langle \mathrm{Re}(X), \mathrm{Re}(Y)\rangle + \langle \mathrm{Im}(X), \mathrm{Im}(Y)\rangle.$$

Since this scalar product induces the Frobenius norm, i.e., $\|X\|_F^2 = \langle X, X\rangle$, it is called the Frobenius scalar product in what follows. If X and Y are partitioned conformably into blocks $X_{i,j}$ and $Y_{i,j}$, the above relation entails, in particular, the identity

$$\langle X, Y\rangle = \sum_i \sum_j \langle X_{i,j}, Y_{i,j}\rangle.$$

Let us now formulate several classes of optimization problems. Each class is defined on a particular curve of the complex plane and requires the definition of an inner product that is conformable with the above definition.

6.1. Real axis. For any couple of nonnegative polynomials $P(x) = \sum_{k=0}^{2n} P_k x^k$ and $Q(x) = \sum_{k=0}^{2n} Q_k x^k$, let us define their scalar product $\langle P, Q\rangle_{\mathbb{R}}$ as follows:

$$\langle P, Q\rangle_{\mathbb{R}} = \sum_{k=0}^{2n} \langle P_k, Q_k\rangle.$$

Several important optimization problems can be formulated in the following standard form:

$$\min_{P\in\mathcal{K}_{\mathbb{R}}} \{\langle C, P\rangle_{\mathbb{R}} : \langle A_\ell, P\rangle_{\mathbb{R}} = b_\ell, \ell = 1, \ldots, q\}, \tag{6.2}$$

for given C, A_ℓ, and b_ℓ, and where $\mathcal{K}_{\mathbb{R}}$ is the cone of matrix coefficients

$$P \doteq [P_0,\ P_1, \ldots,\ P_{2n}\ ]$$

of the polynomial matrix $P(x)$ which is nonnegative on the real axis, i.e.,

$$P(x) \succeq 0, \quad x \in \mathbb{R}.$$

As $P \in \mathcal{K}_{\mathbb{R}}$ necessarily implies $P_k = P_k^*$ for all k, we are not restricted to assuming that all the $m \times m$ blocks C_k of C and blocks $A_{\ell,k}$ of A_ℓ are Hermitian as well, since the anti-Hermitian part of these matrices would disappear anyway in the scalar products. As shown in the preceding section, P belongs to the cone $\mathcal{K}_{\mathbb{R}}$ if and only if there exists a nonnegative block matrix Y with blocks $Y_{i,j}, i, j = 0, \ldots n$, of dimension $m \times m$ satisfying

$$P_k = \sum_{i+j=k} Y_{i,j}, \quad k = 0, 1, \ldots, 2n. \tag{6.3}$$

By definition, the dual cone $\mathcal{K}_{\mathbb{R}}^*$ is the set of the matrix coefficients $Q \doteq [Q_0,\ Q_1, \ldots,\ Q_{2n}\ ]$ of the para-Hermitian matrix polynomials satisfying the constraint

$$\langle Q, P\rangle_{\mathbb{R}} \geq 0 \quad \forall P \in \mathcal{K}_{\mathbb{R}}.$$

If $H(Q)$ denotes the block Hankel matrix

$$H(Q) \doteq \begin{bmatrix} Q_0 & Q_1 & \cdots & Q_n \\ Q_1 & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & Q_{2n-1} \\ Q_n & \cdots & Q_{2n-1} & Q_{2n} \end{bmatrix}, \tag{6.4}$$

the properties of the scalar product and (6.3) allow one to write the equalities

$$\begin{aligned} \langle Q, P\rangle_{\mathbb{R}} &= \sum_{k=0}^{2n} \langle Q_k, P_k\rangle = \sum_{k=0}^{2n} \sum_{i+j=k} \langle Q_k, Y_{i,j}\rangle \\ &= \langle H(Q), Y\rangle. \end{aligned}$$

Moreover, the following equivalence is well known ("Fejer's theorem"; see [7]):

$$\langle H(Q), Y\rangle \geq 0 \quad \forall Y \succeq 0 \iff H(Q) \succeq 0.$$

Therefore the dual cone $\mathcal{K}_{\mathbb{R}}^*$ is characterized by $H(Q) \succeq 0$.

As a consequence, the optimization problem (6.2) can be restated in its dual form,

$$\max_{u_1,\ldots,u_q} \left\{ \sum_{\ell=1}^{q} b_\ell u_\ell : H\left(C - \sum_{\ell=1}^{q} u_\ell A_\ell \right) \succeq 0 \right\}. \tag{6.5}$$

From a numerical point of view, dual formulation (6.5) has a considerable advantage over the primal form (6.2) since it involves an optimization scheme in a space of variables of dimension q rather than $(2n+1)m^2$. Any optimization problem of this type can be solved efficiently with the help of interior-point methods [13]. Their numerical implementation requires the calculation of the first and second derivatives of the barrier function

$$f(u) = -\ln\det H\left(C - \sum_{\ell=1}^{q} A_\ell u_\ell \right).$$

These derivatives can be expressed as follows:

$$\begin{aligned} \frac{\partial f(u)}{\partial u_\ell} &= \langle H(S)^{-1}, H(A_\ell)\rangle, \\ \frac{\partial^2 f(u)}{\partial u_\ell \partial u_s} &= \langle H(S)^{-1} H(A_\ell) H(S)^{-1}, H(A_s)\rangle, \end{aligned} \tag{6.6}$$

where $S = C - \sum_{\ell=1}^{q} A_\ell u_\ell$.

6.2. Unit circle. The same property holds for optimization over the set of nonnegative pseudopolynomial matrices on the unit circle. The scalar product to be used for pseudopolynomials $P(z) = \sum_{k=-n}^{n} P_k z^k$ and $Q(z) = \sum_{k=-n}^{n} Q_k z^k$ is defined as follows:

$$\langle P, Q\rangle_{\mathbb{C}} \doteq \sum_{k=-n}^{n} \langle P_k, Q_k\rangle.$$

The optimization problem now reads

$$\min_{P\in\mathcal{K}_{\mathbb{C}}} \{\langle C, P\rangle_{\mathbb{C}} : \langle A_\ell, P\rangle_{\mathbb{C}} = b_\ell, \ell = 1, \ldots, q\}, \tag{6.7}$$

where $\mathcal{K}_{\mathbb{C}}$ is the cone of matrix coefficients

$$P \doteq [P_{-n}, \ldots, P_n\ ]$$

of nonnegative pseudopolynomial matrices

$$P(z) \succeq 0, \quad z \in e^{j\mathbb{R}},$$

on the unit circle. Note that the coefficients of such matrices satisfy $P_{-k} = P_k^*$ and that $P \in \mathcal{K}_{\mathbb{C}}$ necessarily implies

$$P_k = \sum_{i-j=k} Y_{i,j}, \quad k = -n, \ldots, n, \tag{6.8}$$

where Y is a nonnegative block matrix with blocks $Y_{i,j}, i, j = 0, \ldots n$, of dimension $m \times m$.

As before, we are not restricted to assuming that the $m \times m$ blocks C_k of C and $m \times m$ blocks $A_{\ell,k}$ of A_ℓ have the same type of symmetry as the blocks of P, since this does not affect the scalar products.

The dual cone $\mathcal{K}_{\mathbb{C}}^*$ is made of the matrix coefficients

$$Q \doteq [Q_{-n}, \ldots, Q_n\ ]$$

of the para-Hermitian pseudopolynomials satisfying the constraint

$$\langle Q, P\rangle_{\mathbb{C}} \geq 0 \quad \forall P \in \mathcal{K}_{\mathbb{C}}.$$

If $T(Q)$ denotes the block Toeplitz matrix

$$T(Q) \doteq \begin{bmatrix} Q_0 & Q_1 & \cdots & Q_n \\ Q_1^* & Q_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & Q_1 \\ Q_n^* & \cdots & Q_1^* & Q_0 \end{bmatrix}, \tag{6.9}$$

one has the relations

$$\begin{aligned} \langle Q, P\rangle_{\mathbb{C}} &= \sum_{k=-n}^{n} \langle Q_k, P_k\rangle = \sum_{k=-n}^{n} \sum_{i-j=k} \langle Q_k, Y_{i,j}\rangle \\ &= \langle T(Q), Y\rangle \end{aligned}$$

so that the dual cone $\mathcal{K}_{\mathbb{C}}^*$ is characterized by $T(Q) \succeq 0$.

Therefore the dual optimization problem (6.7) becomes

$$\max_{u_1,\dots,u_\ell} \left\{ \sum_{\ell=1}^{q} b_\ell u_\ell : T\left(C - \sum_{\ell=1}^{q} u_\ell A_\ell \right) \succeq 0 \right\} \tag{6.10}$$

for which the appropriate barrier function is

$$f(u) = -\ln\det T\left(C - \sum_{\ell=1}^{q} A_\ell u_\ell \right).$$

As in the block Hankel case, its derivatives can be expressed as follows:

$$\begin{aligned} \frac{\partial f(u)}{\partial u_\ell} &= \langle T(S)^{-1}, T(A_\ell)\rangle, \\ \frac{\partial^2 f(u)}{\partial u_\ell \partial u_s} &= \langle T(S)^{-1} T(A_\ell) T(S)^{-1}, T(A_s)\rangle, \end{aligned} \tag{6.11}$$

where $S = C - \sum_{\ell=1}^{q} A_\ell u_\ell$.

6.3. Imaginary axis. The imaginary case reformulation is left to the reader. As shown in the previous section, it is reducible to the real line situation in a trivial manner.

7. Computational aspects. Efficient numerical schemes to solve the optimization problems considered require repeated calculations of the differential characteristics of the barrier function, i.e., the gradient $\partial f(u)/\partial u_\ell$ and the Hessian $\partial^2 f(u)/\partial u_\ell \partial u_s$. The block Toeplitz or block Hankel structure underlying the optimization space allows one to carry out these computations in a fast, and even superfast, manner. The aim of this section is to explain this procedure in some detail.

7.1. Displacement structure. Let us first consider Hermitian $(n+1)\times(n+1)$ block Toeplitz matrices with arbitrary $m \times m$ matrix blocks T_i,

$$T \doteq \begin{bmatrix} T_0 & T_1 & \cdots & T_n \\ T_1^* & T_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & T_1 \\ T_n^* & \cdots & T_1^* & T_0 \end{bmatrix},$$

and $(n+1)\times(n+1)$ block Hankel matrices with Hermitian $m \times m$ matrix blocks H_i,

$$H \doteq \begin{bmatrix} H_0 & H_1 & \cdots & H_n \\ H_1 & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & H_{2n-1} \\ H_n & \cdots & H_{2n-1} & H_{2n} \end{bmatrix}.$$

Note that T and H are defined by $(2n+1)m^2$ parameters.

Also, let us set the block permutation matrix J,

$$J \doteq \begin{bmatrix} 0 & \cdots & 0 & I_m \\ \vdots & \iddots & \iddots & 0 \\ 0 & \iddots & \iddots & \vdots \\ I_m & 0 & \cdots & 0 \end{bmatrix},$$

that will play a special role in the subsequent developments.

The displacement theory of Toeplitz and Hankel matrices is well established [9, 10] and is the basis underlying most fast algorithms for decomposing such matrices. Using the block shift matrix one defines a "Toeplitz displacement operator" ∇_t and a "Hankel displacement operator" ∇_h as follows:

$$\nabla_t T \doteq T - Z^T T Z, \quad \nabla_h H \doteq H - ZHZ. \tag{7.1}$$

The reader may easily check that the following equalities hold:

$$\nabla_t T = \begin{bmatrix} T_0 & T_1 & \cdots & T_n \\ T_1^* & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ T_n^* & 0 & \cdots & 0 \end{bmatrix}, \tag{7.2}$$

$$\nabla_h H = \begin{bmatrix} H_0 & 0 & \cdots & 0 \\ H_1 & \vdots & & \vdots \\ \vdots & 0 & \cdots & 0 \\ H_n & \cdots & H_{2n-1} & H_{2n} \end{bmatrix}. \tag{7.3}$$

From the above expressions, ones notices that the original matrices T and H can be recovered from their respective displacement. The inverse operators are obtained by merely applying the displacement operator again and again to both sides of (7.2) to produce

$$T = \nabla_t T + Z^T \cdot \nabla_t T \cdot Z + \cdots + Z^{nT} \cdot \nabla_t T \cdot Z^n \tag{7.4}$$

and

$$H = \nabla_h H + Z \cdot \nabla_h H \cdot Z + \cdots + Z^n \cdot \nabla_h H \cdot Z^n. \tag{7.5}$$

It is also useful to point out that both displacements are closely related to each other. Permuting the block rows of a block Hankel matrix H indeed yields a block Toeplitz matrix JH, which can be defined as T by setting $T_i = H_{i+n},\ i = -n, \ldots, n$. Since $Z^T = JZJ$, the displacement operators are related in a similar fashion as follows:

$$T = JH \iff \nabla_t T = J \nabla_h H.$$

From the sparsity structure of matrices (7.2) and (7.3) it is obvious that the ranks of $\nabla_t T$ and $\nabla_h H$ cannot be larger than $2m$. This rank is called the "displacement rank" of the corresponding matrix.

The theory of displacement ranks [9, 10] tells us that the inverse of T or H (when it exists) has the same displacement as that of the matrix itself as follows:

$$\operatorname{rank} \nabla_t^* T^{-1} = \operatorname{rank} \nabla_t T, \quad \operatorname{rank} \nabla_h H^{-1} = \operatorname{rank} \nabla_h H,$$

where ∇_t^* stands for the transposed Toeplitz displacement operator, i.e., $\nabla_t^* T^{-1} = T^{-1} - Z T^{-1} Z^T$. Since the displacement rank of a block Toeplitz or block Hankel matrix is typically much lower than the dimensions of the corresponding matrix, and since the displacement operator can be inverted, it is economical to represent such a matrix by a rank factorization of its displacement. From the expressions (7.2), (7.3), it is simple to construct low rank factorizations of $\nabla_t T$ or $\nabla_h H$ as follows:

$$\nabla_t T = F_t^* \cdot G_t, \quad \nabla_h H = F_h^* \cdot G_h,$$

where the number of rows of F_t and G_t equals $r_t \doteq \operatorname{rank} \nabla_t T$, and the number of rows of F_h and G_h equals $r_h \doteq \operatorname{rank} \nabla_h H$.

Given such factorizations, fast generalized Schur-based algorithms can be used [9, 10] to derive from them the corresponding factorizations of the displacement of the inverses as follows:

$$\nabla_t^* T^{-1} = A_t^* \cdot B_t, \quad \nabla_h H^{-1} = A_h^* \cdot B_h,$$

and these precise decompositions are used in what follows. Moreover, as Schur algorithms can be implemented in a superfast manner by means of a divide-and-conquer strategy, the complexity of the above construction is found to be $\mathcal{O}(rm^2 n \log^2 n)$. Incidentally, let us note that these factorizations are not unique and that for positive definite matrices T and H there exist particular choices of factorizations that can benefit from these properties. For instance, one can choose in the Toeplitz case

$$G_t = \begin{bmatrix} T_0 & T_1 & T_2 & \cdots & T_n \\ 0 & -T_1 & -T_2 & \cdots & -T_n \end{bmatrix}, \tag{7.6}$$

$$F_t = \begin{bmatrix} T_0 & 0 \\ 0 & -T_0 \end{bmatrix}^{-1} G_t. \tag{7.7}$$

In what follows, these aspects will be disregarded since they only marginally affect the complexity results.

Let us focus first on the case of Toeplitz displacement of an $m(n+1) \times m(n+1)$ matrix X and suppose that a rank r_t factorization of its Toeplitz displacement $\nabla_t X$ has been computed,

$$\nabla_t X = F^* \cdot G,$$

where F and G have dimensions $r_t \times m(n+1)$. Let us also define an upper block triangular Toeplitz matrix $U(G)$ as a function of the partitioned matrix G, where each subblock has dimensions $r_t \times m$,

$$G \doteq \begin{bmatrix} G_0 & G_1 & \cdots & G_n \end{bmatrix},$$

$$U(G) \doteq \begin{bmatrix} G_0 & G_1 & \cdots & G_n \\ 0 & G_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & G_1 \\ 0 & \cdots & 0 & G_0 \end{bmatrix}.$$

Doing the same for the matrix F, one obtains

$$F \doteq \begin{bmatrix} F_0 & F_1 & \cdots & F_n \end{bmatrix},$$

$$U(F)^* \doteq \begin{bmatrix} F_0^* & 0 & \cdots & 0 \\ F_1^* & F_0^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ F_n^* & \cdots & F_1^* & F_0^* \end{bmatrix}.$$

It follows from the displacement equation $\nabla_t X = F^* \cdot G$ that

$$X = \sum_{j=0}^{n} (FZ^j)^*(GZ^j) = U(F)^* U(G)$$
$$= \begin{bmatrix} F_0^* & 0 & \cdots & 0 \\ F_1^* & F_0^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ F_n^* & \cdots & F_1^* & F_0^* \end{bmatrix} \cdot \begin{bmatrix} G_0 & G_1 & \cdots & G_n \\ 0 & G_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & G_1 \\ 0 & \cdots & 0 & G_0 \end{bmatrix}.$$

This formula, when applied to a particular choice of displacement factors F and G for the inverse of a Toeplitz matrix T, is also known as the Gohberg–Semencul formula for $X = T^{-1}$.

For the Hankel displacement $\nabla_h X$ of an $m(n+1) \times m(n+1)$ matrix X, there exists a similar representation starting based upon the rank r_h factorization of $\nabla_h X$,

$$\nabla_h X = F^* \cdot G,$$

where F and G have dimension $r_h \times m(n+1)$. If the matrix F is partitioned in reverse order,

$$F \doteq \begin{bmatrix} F_0 & \cdots & F_n \end{bmatrix} \iff FJ \doteq \begin{bmatrix} F_n & \cdots & F_0 \end{bmatrix},$$

then it follows from the relation $J\nabla_h X = \nabla_t (JX)$ that

$$X = J \sum_{j=0}^{n} (FJZ^j)^*(GZ^j) = JU(FJ)^* U(G) \tag{7.8}$$

$$= \begin{bmatrix} F_0^* & \cdots & F_{n-1}^* & F_n^* \\ \vdots & ⋰ & ⋰ & 0 \\ F_{n-1}^* & F_n^* & ⋰ & \vdots \\ F_n^* & 0 & \cdots & 0 \end{bmatrix} \cdot \begin{bmatrix} G_0 & G_1 & \cdots & G_n \\ 0 & G_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & G_1 \\ 0 & \cdots & 0 & G_0 \end{bmatrix}. \tag{7.9}$$

When applied to a particular choice of displacement factors F and G for the inverse of a Hankel matrix, this formula is also known as the Christoffel–Darboux formula for $X = H^{-1}$.

7.2. Implementation. The numerical solution of the optimization problem considered in section 6 requires evaluations of the gradient $\partial f(u)/\partial u_\ell$ and the Hessian $\partial^2 f(u)/\partial u_\ell \partial u_s$ as given by (6.6) or (6.11). Let us now focus on the fast computation of these elements using the displacement techniques mentioned above.

Consider the inner product $\langle X, T(A_s)\rangle$ which appears in (6.11) with $X = T(S)^{-1}$ or $X = T(S)^{-1}T(A_l)T(S)^{-1}$, and let $\operatorname{diag}\{W\}$ be the block diagonal matrix with all blocks equal to $W \in \mathbb{C}^{m\times m}$. Since

$$T(A_s) = \operatorname{diag}\{A_{s,0}\} + \sum_{k=0}^{n} [Z^k \operatorname{diag}\{A_{s,k}\} + (Z^k)^T \operatorname{diag}\{A_{s,k}^*\}],$$

the computation can be broken down into a summation of scalar products of the type

$$\langle X, Z^i \operatorname{diag}\{W\}\rangle, \quad \langle X, (Z^i)^T \operatorname{diag}\{W^*\}\rangle. \tag{7.10}$$

For Hermitian matrices X, it turns out that $\langle X, (Z^i)^T \operatorname{diag}\{W^*\}\rangle = \langle X, Z^i \operatorname{diag}\{W\}\rangle$ so that only one expression has to be evaluated.

Similarly, the inner product $\langle X, H(A_s)\rangle$ which appears in (6.6) with $X = H(S)^{-1}$ or $X = H(S)^{-1}H(A_l)H(S)^{-1}$ requires the evaluation of scalar products of the type

$$\langle X, JZ^i \operatorname{diag}\{W_1\}\rangle, \quad \langle X, J(Z^i)^T \operatorname{diag}\{W_2\}\rangle, \tag{7.11}$$

where W_1 and W_2 are Hermitian matrices of order m.

In addition, since the matrices X can be described by their Hankel or Toeplitz displacement, one can speed up the computation of (7.10) and (7.11). Let us first consider matrices X given by their Toeplitz displacement $\nabla_t X = F^* \cdot G$. Since

$$U(F) = \sum_{k=0}^{n} Z^k \operatorname{diag}\{F_k\}, \quad U(G) = \sum_{k=0}^{n} Z^k \operatorname{diag}\{G_k\},$$

and as

$$\langle Z^j \operatorname{diag}\{X\}, Z^i \operatorname{diag}\{Y\}\rangle = \delta_{i,j}(n+1-i)\langle X, Y\rangle,$$

one obtains the expression

$$\begin{aligned}\langle U(F)^*U(G), Z^j \operatorname{diag}\{W\}\rangle &= \langle (n+1-j)F_j^* G_0 + \cdots + 2F_{n-1}^* G_{n-j-1} + F_n^* G_{n-j},\, W\rangle \\ &\doteq \langle M_j, W\rangle.\end{aligned}$$

Since the matrix $X = U(F)^*U(G)$ is Hermitian, the roles of F_i and G_i can be interchanged in the above formula. Moreover, the quantities $\{M_j\}_{j=0}^n$ can be evaluated as the convolution of the block vectors

$$[(n+1)F_0, nF_1, \ldots, 2F_{n-1}, F_n], \quad [G_0, G_1, \ldots, G_{n-1}, G_n],$$

which has a complexity of $\mathcal{O}(r_t m^2 n \log_2 n)$ flops [10]. As the computation of the inner product $\langle M_j, W\rangle$ requires $\mathcal{O}(m^2)$ operations, the overall complexity of computing $\langle X, T(A_s)\rangle$ is thus found to be $\mathcal{O}(r_t m^2 n \log_2 n + m^2 n)$ flops for a matrix of displacement rank r_t, *provided that* the matrices F and G are given. If the matrix X is given by its transposed displacement $\nabla_t^* X = A^* \cdot B$, one can easily adapt the above formula and check that the overall complexity is also $\mathcal{O}(r_t m^2 n \log_2 n + m^2 n)$ flops, *provided that* the matrices A and B are given.

The calculations involving the Hessian, i.e., when $X = T(S)^{-1}T(A_\ell)T(S)^{-1}$, require some elaboration. With the matrix $\hat{T}$ defined by

$$\hat{T} = \begin{bmatrix} -T(A_\ell) & T(S) \\ T(S) & 0 \end{bmatrix},$$

note first that the following relation holds:

$$\hat{T}^{-1} = \begin{bmatrix} 0 & T(S)^{-1} \\ T(S)^{-1} & X \end{bmatrix}.$$

Furthermore, as $T(S)$ and $T(A_\ell)$ are block Toeplitz matrices, the rank of the matrix factors F and G in the block displacement equation

$$\nabla_t \hat{T} = \hat{T} - \begin{bmatrix} Z^T & 0 \\ 0 & Z^T \end{bmatrix} \hat{T} \begin{bmatrix} Z & 0 \\ 0 & Z \end{bmatrix} = F^* G$$

is equal to $4m$, as is easily verified. The corresponding factorization of the block displacement of the inverse can be achieved at low computational cost in the form

$$\nabla_t^* \hat{T}^{-1} = \hat{T}^{-1} - \begin{bmatrix} Z & 0 \\ 0 & Z \end{bmatrix} \hat{T}^{-1} \begin{bmatrix} Z^T & 0 \\ 0 & Z^T \end{bmatrix} = [A_1, A_2]^* \cdot [B_1, B_2].$$

Therefore, the expression of the transposed Toeplitz displacement of X is given by

$$\nabla_t^* X = A_2^* \cdot B_2.$$

The formalism described above for the fast computation of the relevant inner products can therefore be applied to construct the entries of the Hessian (6.11). If the displacement factors are computed using a superfast algorithm, the overall complexity of constructing the Hessian is therefore $\mathcal{O}(qr_t m^2 n \log^2 n + q^2 m^2 n)$.

Let us now consider matrices X given by their Hankel displacement $\nabla_h X = F^* \cdot G$. The inner products of interest can be rewritten in terms of JX as follows:

$$\langle X, JZ^i \operatorname{diag}\{W_1\}\rangle = \langle JX, Z^i \operatorname{diag}\{W_1\}\rangle,$$
$$\langle X, J(Z^i)^T \operatorname{diag}\{W_2\}\rangle = \langle (JX)^*, Z^i \operatorname{diag}\{W_2\}\rangle,$$

where W_1 and W_2 are Hermitian matrices of order m. Since JX is block Toeplitz, the above formulas could, in theory, be applied mutatis mutandis. From a practical viewpoint, however, this does not make much sense. As explained in the next section, the Hankel setting of the optimization problem considered is numerically ill-conditioned. Hence, the problem formulation itself needs to be redesigned so as to circumvent this inherent difficulty. This issue is addressed in the next section.

The actual solution of the optimization problem of section 6 is often achieved with the help of an iterative Newton scheme. In particular, this iterative process requires frequent evaluations of the so-called Newton directions, which involve the product of the inverse of the current Hessian by an appropriate given vector. From a practical viewpoint, this approach is efficient only if the Hessian dimension q is small. Otherwise, a conjugate gradient scheme could be more attractive since it does not require the inversion of the Hessian but rather its product with a vector. Such computations can be made at low cost with the help of the inner product formalism explained in the present section.

Let us briefly clarify this issue. Assume that the optimization problem is defined on the unit circle, and consider the product of the Hessian by a vector x to yield a vector y. By definition, one has in view of (6.11) that the sth component of y is given by

$$\begin{aligned} y_s &= \sum_\ell \frac{\partial^2 f(u)}{\partial u_\ell \partial u_s} x_\ell, \\ &= \sum_\ell \langle T(S)^{-1} T(A_\ell) T(S)^{-1}, T(A_s)\rangle x_\ell, \\ &= \langle T(S)^{-1} T(D) T(S)^{-1}, T(A_s)\rangle, \\ &= \langle T(S)^{-1} T(A_s) T(S)^{-1}, T(D)\rangle, \end{aligned}$$

where $T(D)$ stands for the block Toeplitz matrix $T(D) = \sum_\ell T(A_\ell x_\ell)$. Expressions of this type can be computed efficiently using the results derived above in this section. Performing k conjugate gradient steps at each Newton iteration therefore requires $\mathcal{O}(qr_t m^2 n \log^2 n + kqm^2 n)$ operations.

7.3. Complexity of the optimization scheme. Since interior-points methods require $\mathcal{O}(\sqrt{nm}\log\frac{1}{\epsilon})$ Newton steps to solve the optimization problems (6.5) and (6.10) up to an accuracy ϵ [13], the overall complexity of solving these problems depends on the method used to compute the Newton directions and is found to be

- $\mathcal{O}(\sqrt{nm}\log\frac{1}{\epsilon}[qr_t m^2 n\log^2 n + q^2m^2n + q^3])$ flops for the "inversion" of the Hessian;
- $\mathcal{O}(\sqrt{nm}\log\frac{1}{\epsilon}[qr_t m^2 n\log^2 n + kqm^2n])$ flops for the conjugate gradient scheme.

By solving the dual problem and using the matrix structures, we get a remarkable result for solving an optimization problem in a $(2n+1)m^2$-dimensional vector space, subject to q linear constraints and m semi-infinite inequality constraints (see (6.2) and (6.7)).

In particular, for nonnegative *scalar* polynomials, i.e., $m=1$, each Newton iteration requires $\mathcal{O}(qn(\log^2 n + q) + q^3)$ and $\mathcal{O}(qn(\log^2 n + k))$, respectively.

8. Chebyshev reformulation of the real line optimization problem. The formulation of the real line optimization problem exhibits a serious drawback: it involves positive definite Hankel matrices, which are numerically ill-conditioned [3, 15]. The celebrated Hilbert matrix is a good illustration of this fact. More generally, the Euclidean condition number $\kappa(H)$ of any positive definite Hankel matrix H of order $n+1$ was shown recently [3] to be bounded from below by

$$\kappa(H) \geq \frac{(1.792)^{2n}}{16(n+1)}, \quad n \geq 2.$$

Therefore, solving the real line optimization problem as considered in section 6 is inherently hazardous, and all the more so if the problem dimension is large. To get around this, let us first observe that the occurrence of the block Hankel structure originates from the choice of the natural powers $1, x, x^2, \ldots$ as a basis for describing the optimization space of the polynomial matrices $P(x) = \sum_{k=0}^{n} P_k x^k$, positive semidefinite on the real line. Obviously, other choices are possible. In this section, the alternative use of a basis of Chebyshev polynomials to describe the optimization is specifically investigated together with the consequences of this choice.

The first order Chebyshev polynomials $T_k(x)$ are well known to satisfy, for $k \geq 1$, the recurrence formula

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x)$$

initialized with $T_0(x) = 1$ and $T_1(x) = x$. In particular, one has the relation

$$T_i(x)T_j(x) = \frac{1}{2}[T_{i+j}(x) + T_{|i-j|}(x)] \quad \forall\, i, j \geq 0. \tag{8.1}$$

In order to emphasize our choice of the Chebyshev basis, let us denote by $\tilde{P}_k = \tilde{P}_k^*$ the matrix coordinates of any para-Hermitian polynomial matrix $P(x)$ in this basis, i.e.,

$$P(x) = \sum_{k=0}^{2n} \tilde{P}_k T_k(x).$$

Using the notation introduced in section 4, let us consider the set of Hermitian matrices Y such that one has the identity

$$\tilde{\Pi}^*(x) Y \tilde{\Pi}(x) = P(x) \tag{8.2}$$

with $\tilde{\Pi}(x) = [T_0(x)I_m,\ T_1(x)I_m, \ldots, T_n(x)I_m]^T$.

The algebraic constraints on Y implied by (8.2) can be expressed in terms of the Chebyshev basis in a simple manner. Using identities (8.1), one can easily check that the following relations hold:

$$\tilde{P}_k = \frac{1}{2}\left(\sum_{i+j=k} Y_{i,j} + \sum_{|i-j|=k} Y_{i,j}\right) \qquad \text{for } k = 0, \ldots, 2n. \tag{8.3}$$

If L stands for the block lower triangular matrix transforming $\Pi(x)$ into $\tilde{\Pi}(x)$, i.e., $\tilde{\Pi}(x) = L\,\Pi(x)$, a simple consequence of Theorem 4.1 is that the set of all solutions Y to (8.2) is parametrized by the relation

$$Y = Y_0 + L^{-T}(Z\,X - X\,Z^T)L^{-1},$$

where Y_0 is a particular solution of (8.2) and X is any skew-Hermitian matrix of form (4.1). Moreover, if Theorem 4.2 is applied to $L^T\,Y\,L$, the existence of a positive definite solution Y to (8.2) is also found to be the necessary and sufficient condition such that $P(x)$ is a well-defined positive polynomial matrix on the real line. Note incidentally that a particular matrix Y_0 satisfying (8.2) is provided by

$$Y_0 = \begin{bmatrix} \frac{1}{2}\tilde{P}_0 - \sum_{k=1}^{n}\tilde{P}_{2k} & \frac{1}{2}\tilde{P}_1 - \frac{1}{2}\sum_{k=2}^{n} -\tilde{P}_{2k-1} & & & \\ \frac{1}{2}\tilde{P}_1 - \frac{1}{2}\sum_{k=2}^{n}\tilde{P}_{2k-1} & 2\tilde{P}_2 & \tilde{P}_3 & & \\ & \tilde{P}_3 & 2\tilde{P}_4 & \ddots & \\ & & \ddots & \ddots & \tilde{P}_{2n-1} \\ & & & \tilde{P}_{2n-1} & 2\tilde{P}_{2n} \end{bmatrix}.$$

When such a Chebyshev basis is chosen, the optimization space is transformed into the convex cone $\tilde{\mathcal{K}}_{\mathbb{R}}$ of matrix coefficients

$$\tilde{P} \doteq [\tilde{P}_0, \tilde{P}_1, \ldots, \tilde{P}_{2n}]$$

of the polynomial matrices nonnegative definite on the real axis, i.e.,

$$P(x) \succeq 0, \quad x \in \mathbb{R}.$$

Furthermore and as shown above, $\tilde{P}$ belongs to the cone $\tilde{\mathcal{K}}_{\mathbb{R}}$ if and only if there exists a nonnegative block matrix Y with blocks $Y_{i,j}$ satisfying (8.3). By definition, the dual cone $\tilde{\mathcal{K}}^*_{\mathbb{R}}$ consists of the matrix coefficients $\tilde{Q} \doteq [\tilde{Q}_0, \tilde{Q}_1, \ldots, \tilde{Q}_{2n}]$ satisfying the constraints

$$\langle \tilde{Q}, \tilde{P}\rangle_{\mathbb{R}} \geq 0 \quad \forall \tilde{P} \in \tilde{\mathcal{K}}_{\mathbb{R}}.$$

Recall that we are not restricted to assuming that the matrix coefficients $\tilde{P}_k$ and $\tilde{Q}_k$ are Hermitian for all k. For any $\tilde{Q}$, let us set the block Toeplitz-plus-Hankel matrix

$$T_H(\tilde{Q}) \doteq \begin{bmatrix} \tilde{Q}_0 & \tilde{Q}_1 & \cdots & \tilde{Q}_n \\ \tilde{Q}_1 & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & \tilde{Q}_{2n-1} \\ \tilde{Q}_n & \cdots & \tilde{Q}_{2n-1} & \tilde{Q}_{2n} \end{bmatrix} + \begin{bmatrix} \tilde{Q}_0 & \tilde{Q}_1 & \cdots & \tilde{Q}_n \\ \tilde{Q}_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tilde{Q}_1 \\ \tilde{Q}_n & \cdots & \tilde{Q}_1 & \tilde{Q}_0 \end{bmatrix}. \tag{8.4}$$

In view of (8.3) and the scalar product definition, one derives the relation

$$\begin{aligned}\langle \tilde{P}, \tilde{Q}\rangle_{\mathbb{R}} &= \sum_{k=0}^{2n} \langle \tilde{Q}_k, \tilde{P}_k\rangle \\ &= \frac{1}{2}\sum_{k=0}^{2n}\left[\sum_{i+j=k}\langle \tilde{Q}_k, Y_{i,j}\rangle + \sum_{|i-j|=k}\langle \tilde{Q}_k, Y_{i,j}\rangle\right] \\ &= \frac{1}{2}\langle T_H(\tilde{Q}), Y\rangle,\end{aligned}$$

which shows that the dual cone $\mathcal{K}_{\mathbb{R}}^*$ is characterized by $T_H(\tilde{Q}) \succeq 0$.

Therefore, the dual form of the optimization problem (6.2) can be expressed in the present case as

$$\max_{u_1,\ldots,u_\ell}\left\{\sum_{\ell=1}^{q} b_\ell u_\ell : T_H\left(C - \sum_{\ell=1}^{q} u_\ell A_\ell\right) \succeq 0\right\}. \tag{8.5}$$

The corresponding barrier function is $f(u) = -\ln\det T_H(C - \sum_{\ell=1}^{q} u_\ell A_\ell)$ and the differential characteristics of interest now read

$$\begin{aligned}\frac{\partial f(u)}{\partial u_\ell} &= \langle T_H(S)^{-1}, TH(A_\ell)\rangle, \\ \frac{\partial^2 f(u)}{\partial u_\ell \partial u_s} &= \langle T_H(S)^{-1} T_H(A_\ell)\, T_H(S)^{-1}, T_H(A_s)\rangle,\end{aligned} \tag{8.6}$$

where $S = C - \sum_{\ell=1}^{q} A_\ell u_\ell$.

From a numerical viewpoint, this reformulation of the optimization problem on the real line exhibits a considerable advantage over its initial formulation in the sense that is not intrinsically ill-conditioned. Indeed, for all degrees n there exist nonnegative matrices $T_H(\tilde{Q})$ with a condition number equal to 2, as illustrated by the trivial example $\tilde{Q} = [I_m, 0, \ldots, 0]$. As a result, the numerical behavior of the computational optimization scheme is expected to be substantially improved.

Finally, let us point out that the differential characteristics of the Chebyshev basis reformulated barrier function (8.6) can also be computed in a fast way with the help of displacement techniques. This problem is not a straightforward generalization of the results presented in this paper. Nevertheless one expects to apply, as above, a divide-and-conquer strategy to get low complexity algorithms.

9. Conclusion. Cones of positive pseudopolynomial matrices are often encountered in practice as well as the corresponding dual cones, which are related to moment spaces. In this paper semidefinite representation of these cones is shown to be interesting from a computational viewpoint. In particular the dual optimization problems can be solved very efficiently using displacement-based factorizations as well as an appropriate divide-and-conquer strategy. These results are direct consequences of the Hankel or Toeplitz structure in the dual constraints.

During the review process Alkire and Vandenberghe [2] obtained an algorithm to solve optimization problems involving autocorrelation sequences. The associated cone consists of nonnegative cosine polynomials, which are particular pseudopolynomials. In their case the barrier function $f(u)$ is thus equal to the logarithmic barrier of a Toeplitz matrix $T(u)$. As the Levinson–Durbin algorithm is applied to factor the inverse Toeplitz matrix and DFT is then applied to assemble the gradient and the Hessian, the complexity of one iteration in their scheme is equal to $\mathcal{O}(n^3)$. Although this method is similar to the one proposed in this paper (if applied to this particular setting) the techniques presented above are more general. On the one hand, they can be applied to structured matrices with low displacement rank, in particular, block Hankel or Toeplitz. On the other hand, we consider the generic setting of conic optimization problems, for which the barrier function is more general.

REFERENCES

[1] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*, Oliver and Boyd, Edinburgh, 1965.

[2] B. Alkire and L. Vandenberghe, *Convex optimization problems involving finite autocorrelation sequences*, Math. Program., (2002), 93 (2002), pp. 331–359.

[3] B. Beckermann, *The condition number of real Vandermonde, Krylov and positive definite Hankel matrices*, Numer. Math., 85 (2000), pp. 553–577.

[4] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, Stud. Appl. Math. 15, SIAM, Philadelphia, PA, 1994.

[5] Y. Genin, Y. Nesterov, and P. Van Dooren, *Positive transfer functions and convex optimization*, in Proceedings of the European Control Conference ECC-99, Karlsruhe, Germany, 1999. Paper F-143.

[6] G. Heinig and K. Rost, *Algebraic Methods for Toeplitz-like Matrices and Operators*, Math. Res. 19, Akademie-Verlag, Berlin, 1984.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[8] V. Ionescu, C. Oară, and M. Weiss, *Generalized Riccati Theory and Robust Control: A Popov Function Approach*, John Wiley, Chichester, UK, 1999.

[9] T. Kailath and A. H. Sayed, *Displacement structure: Theory and applications*, SIAM Rev., 37 (1995), pp. 297–386.

[10] T. Kailath and A. H. Sayed, eds., *Fast Reliable Algorithms for Matrices with Structure*, SIAM, Philadelphia, PA, 1999.

[11] S. Karlin and W. J. Studden, *Tchebycheff Systems: With Applications in Analysis and Statistics*, Pure Appl. Math. 15, Wiley Interscience, New York, 1966.

[12] Y. Nesterov, *Squared functional systems and optimization problems*, in High Performance Optimization, Appl. Optim. 33, Kluwer, Dordrecht, 2000, pp. 405–440.

[13] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, Stud. Appl. Math. 13, SIAM, Philadelphia, PA, 1994.

[14] V.-M. Popov, *Hyperstability of Control Systems*, Grundlehren Math. Wiss. 204, Springer-Verlag, Berlin, 1973.

[15] E. E. Tyrtyshnikov, *How bad are Hankel matrices?* Numer. Math., 67 (1994), pp. 261–269.

[16] J. C. Willems, *Least squares stationary optimal control and the algebraic Riccati equation*, IEEE Trans. Automat. Control, 16 (1971), pp. 621–634.

[17] D. C. Youla, *On the factorization of rational matrices*, IRE Trans. Inform. Theory, 7 (1961), pp. 172–189.

[18] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice-Hall, Upper Saddle River, NJ, 1996.