Optimization over Positive Polynomial Matrices

Y. Genin, Yu. Nesterov, P. Van Dooren
CESAME – Université catholique de Louvain
Av. G. Lemaître, 4, B-1348 Louvain-la-Neuve, Belgium
{genin,nesterov,vdooren}@csam.ucl.ac.be

Keywords: convex optimization, positive polynomials, Toeplitz matrices, Hankel matrices.

Abstract

Positive polynomial matrices play a fundamental role in systems and control theory. We give here a simplified proof of the fact that the convex set of positive polynomial matrices can be parameterized using block Hankel and block Toeplitz matrices. We also show how to derive efficient computational algorithms for optimization problems over positive pseudo polynomial matrices over the real line, over the imaginary axis and over the unit circle.

1 Introduction

Positive transfer functions play a fundamental role in systems and control theory: they represent e.g. spectral density functions of stochastic processes, show up in spectral factorizations, and are also related to the Riccati equations. When such transfer functions are *rational*, it is known since the work of Youla [11] that 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. [10]).

Positive transfer functions obviously form a convex set, and recently they were also being studied in the convex optimization literature [2], [7]. In order to optimize over the set of positive functions, it is important to have a compact (say "minimal") parameterization of these functions and recently such a parameterization was presented [7], [3]. In this paper we derive this result from simpler ideas and also develop the algorithmic aspects of fast algorithms in a more rigorous manner.

First of all, we recall the basic results of para-hermitian transfer functions, a concept we need when looking at the matrix case of positive transfer functions, since positive matrices inherently require some kind of symmetry. Then we recall the spectral factorization results for positive functions and derive the parameterization from this basic result. Connections with the positive real lemma are revisited in this perspective in a later chapter.

The optimization problem over the convex set of positive pseudo-polynomial matrices is considered. We show that this problem can be solved by fast algorithms for matrices with block Hankel or Toeplitz structure.

2 Para-hermitian transfer functions

Para-hermitian 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) or the unit circle (for discrete-time systems), but we will consider here as well the case of the real axis.

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). We denote the imaginary axis as $s \in j\Re$.

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). We denote the unit circle as $z \in e^{j\Re}$.

Real axis
This curve occurs in the standard treatment of the classical moment problem [1], [6]. We will choose the complex variable x in this case and denote the real axis as $x \in \Re$.

When we want to stress that a result holds for a particular curve we use the variable associated with that curve. Otherwise we use the variable p. We consider in this paper only the case of square *rational* transfer matrices $\Phi(p)$, i.e. $m \times m$ matrices $\Phi(p)$ whose entries are rational functions of the variable p.

Definition 1 *The para-conjugate transfer function $\Phi_*(p)$ of a given transfer matrix $\Phi(p)$ is defined as follows :*
$\Phi_*(s) = [\Phi(-\overline{s})]^*$ *for the imaginary axis*
$\Phi_*(z) = [\Phi(1/\overline{z})]^*$ *for the unit circle*
$\Phi_*(x) = [\Phi(\overline{x})]^*$ *for the real axis,*
where M^* is the conjugate transposed matrix of a matrix M.

We point out that the para-conjugate $\Phi_*(p)$ is again a rational transfer function of the complex variable p. From this definition we then define para-hermitian transfer functions as follows.

Definition 2 *A square transfer function $\Phi(p)$ is para-hermitian if it is equal to its para-conjugate : $\Phi_*(p) = \Phi(p)$.*

Notice that this definition depends on the choice of curve we are considering. But for each definition, a para-hermitian transfer function *evaluated* on the corresponding curve, is a Hermitian matrix. Indeed, $\Phi_*(p) = \Phi(p)$ implies for each case :
$\Phi_*(j\omega) = [\Phi(j\omega)]^*$ for $s = j\omega$ on the imaginary axis,
$\Phi_*(e^{j\omega}) = [\Phi(e^{j\omega})]^*$ for $z = e^{j\omega}$ on the unit circle,
$\Phi_*(\omega) = [\Phi(\omega)]^*$ for $x = \omega$ on the real axis,
where $\omega \in \Re$ is thus a real variable parameterizing the curve.

Since a para-hermitian transfer function is a Hermitian matrix when evaluated on the curve, it will have real eigenvalues and we can thus impose a condition of positivity on these eigenvalues. This leads to the following definition.

Definition 3 *A para-hermitian transfer function is* positive *(*non-negative*) if it is positive (*non-negative*) when evaluated on the curve : $\Phi(p) \succ 0$ ($\Phi(p) \succeq 0$).*

It turns out that non-negative para-hermitian transfer functions always possess a so-called *spectral factorization* :

$$\Phi(p) = G_*(p)G(p), \tag{1}$$

where the spectral factor $G(p)$ is again a square rational transfer function in p. This result was proven in the systems theory literature [11], [9].

3 Positive pseudo-polynomials

Pseudo-polynomial 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 pseudo-polynomial matrices must possess a certain symmetry.

Real axis
For a para-hermitian transfer function $\Phi(x)$ that is non-negative on the real axis $x \in \Re$ it follows from the para-hermitian nature that the coefficient matrices of the expansion

$$\Phi(x) = \sum_{k=-r}^{t} \Phi_k x^k \tag{2}$$

must all be Hermitian : $\Phi_k = \Phi_k^*$. Moreover, since x^2 is non-negative on the real axis $x \in \Re$, we can easily multiply or divide by a power of x^2 and then reduce such pseudo-polynomial matrices to polynomial matrices in x or in x^{-1}. If one chooses positive powers of x one has e.g.

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

and it is easy to see from the non-negativity 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 we use here for non-negative para-hermitian matrices on the real axis is thus

$$\Phi(x) = \sum_{k=0}^{2n} \Phi_k x^k, \quad \Phi_k = \Phi_k^* \tag{3}$$

Unit Circle
For a para-hermitian transfer function $\Phi(z)$ that is non-negative on the unit circle $z \in e^{j\Re}$ it follows from the para-hermitian nature that the coefficient matrices of the expansion

$$\Phi(z) = \sum_{k=-r}^{t} \Phi_k z^k \tag{4}$$

must satisfy the condition: $\Phi_{-k} = \Phi_k^*$ and thus that such a pseudo-polynomial matrix must have a symmetric expansion. The standard form we use here for non-negative para-hermitian matrices on the unit circle is thus

$$\Phi(z) = \sum_{k=-n}^{n} \Phi_k x^k, \quad \Phi_{-k} = \Phi_k^* \tag{5}$$

Imaginary axis
For a para-hermitian transfer function $\Phi(s)$ that is non-negative on the imaginary axis $s \in j\Re$ it follows from the para-hermitian nature that the coefficient matrices of the expansion

$$\Phi(s) = \sum_{k=-r}^{t} \Phi_k s^k \tag{6}$$

are Hermitian if k is even and 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 in the imaginary axis. One can again multiply by a power of $-s^2$ (which is non-negative 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 non-negativity 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 non-negative para-hermitian matrices on the imaginary axis is thus

$$\Phi(x) = \sum_{k=-n}^{n} \Phi_k x^k, \quad \Phi_{2k} = \Phi_{2k}^*, \quad \Phi_{2k+1} = -\Phi_{2k+1}^*. \tag{7}$$

4 Parameterization of positive transfer functions

In this section we derive a parameterization of non-negative pseudo-polynomial matrices in terms of constant Hermitian matrices.

Real axis
Let

$$P(x) = \sum_{k=0}^{2n} P_k x^k \tag{8}$$

be a $m \times m$ para-hermitian polynomial matrix with Hermitian coefficients : $P_k = P_k^*$. We consider the set of Hermitian matrices

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

with blocks of dimension $m \times m$. Define the array

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

then the relation

$$\Pi_*(x) Y \Pi(x) = P(x) \tag{9}$$

implies that

$$P_k = \sum_{i+j=k} Y_{i,j}, \quad k = 0, \dots, 2n, \tag{10}$$

where we assume the blocks $Y_{i,j} = 0$ for i and j outside their definition range. We observe that a simple choice for Y for obtaining this identity is

$$Y_0 = \begin{bmatrix} P_0 & \frac{1}{2}P_1 & & \\ \frac{1}{2}P_1 & P_2 & \ddots & \\ & \ddots & \ddots & \frac{1}{2}P_{2n-1} \\ & & \frac{1}{2}P_{2n-1} & P_{2n} \end{bmatrix}. \tag{11}$$

We also introduce two $(n+1)m \times (n+1)m$ block matrices. Let

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

be the block shift operator and define a matrix of the form

$$X = \left[\begin{array}{c|c} X_0 & 0 \\ \hline 0 & 0 \end{array}\right], \tag{12}$$

with $X_0 \in \mathcal{C}^{nm \times nm}$. Then we have the following characterization theorem.

Theorem 1 *A Hermitian matrix Y satisfies (9) iff it can be expressed as*

$$Y = Y_0 + Z^T X - XZ,$$

with X_0 skew Hermitian, i.e. $X_0 = -X_0^*$.

Proof : The if part is obvious since one has $\Pi_*(x)\left[Z^T X - XZ\right]\Pi(x) = 0$ for any matrix X of the form (12). Conversely, let Y be a solution, then solving (10) in terms of X one obtains

$$X = \sum_{k=0}^{n} (Z^{k+1})(Y - Y_0)(Z^k) \tag{13}$$

as well as

$$X = -\sum_{k=0}^{n} (Z^k)^T (Y - Y_0)(Z^{k+1})^T, \tag{14}$$

provided X has the form (12), what is readily verified from relations (10). Finally, one derives the skew symmetry of X_0 from comparing (13) and (14). □

Imposing the condition that (8) is also a non-negative transfer function leads to the following theorem.

Theorem 2 *A matrix polynomial $P(x)$ is non-negative on the real axis iff there exists a non-negative definite matrix Y satisfying (10).*

Proof : Because of the previous theorem, we need to prove the "only if" part only. We derive this from the existence of a spectral factorization

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

where $G(x)$ is polynomial in x : $G(x) = \sum_{k=0}^{n} G_n x^n$. Choose then

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

This matrix Y is non-negative and satisfies the constraints of the theorem. □

This thus proves the following theorem :

Theorem 3 *A pseudo-polynomial matrix of form (8) is non-negative definite on the real axis if and only if there exists a non-negative definite matrix Y with blocks $Y_{i,j}, i, j = 0, \dots, n$ such that (assuming $Y_{i,j} = 0$ for i and j outside their definition range) :*

$$P_k = \sum_{i+j=k} Y_{i,j} \qquad \textit{for } i = 0, \dots, 2n. \tag{15}$$

□

It turns out that this characterization of matrix polynomials non-negative on the real axis extends a result earlier obtained by Nesterov [7] for scalar polynomials.

Unit Circle

We now turn to the non-negative transfer functions on the unit circle. It follows from the finite expansion and from its para-hermitian character that such a pseudo-polynomial matrix :

$$P(z) = \sum_{k=-n}^{n} P_k z^k, \tag{16}$$

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

$$\Pi_*(z) Y \Pi(z) = \sum_{k=-n}^{n} P_k z^k, \tag{17}$$

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

$$P_k = \sum_{i-j=k} Y_{i,j}, \tag{18}$$

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

$$Y_0 = \begin{bmatrix} P_0 & P_1 & \dots & P_n \\ P_1^* & 0 & \vdots & 0 \\ \vdots & \vdots & & \vdots \\ P_n^* & 0 & \dots & 0 \end{bmatrix}, \tag{19}$$

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

Theorem 4 *A Hermitian matrix Y satisfies equation (17) iff it can be expressed as*

$$Y = Y_0 + X - Z^T X Z,$$

where X has the form (12) with X_0 an $mk \times mk$ Hermitian matrix, i.e. $X_0 = X_0^*$.

Proof :

By duplicating the argument used in the proof of Theorem 1, one shows that the solution X of equation (18) 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 (12). □

The positive pseudo-polynomial matrices on the unit circle can then be characterized as follows.

Theorem 5 *A pseudo-polynomial matrix $P(z)$ is non-negative definite on the unit circle iff there exists a non-negative definite matrix Y satisfying equations (18).*

The proof of this theorem is again based on the same spectral factorization argument as in Theorem 2 and is therefore omitted.

Imaginary axis

The third kind of non-negative pseudo-polynomial 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

$$P(s) = \sum_{k=0}^{2n} P_k s^k \tag{20}$$

with $s \in j\Re$. 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 3 to $\hat{P}(x)$, one obtains for $P(s)$ the following result :

Theorem 6 *A pseudo-polynomial matrix of the form (20) is non-negative on the imaginary axis iff there exists a non-negative matrix Y with blocks $Y_{i,j}, i,j = 0, \dots, n$ such that*

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

5 The optimization problem

Since the optimization problem will be defined in terms of scalar products, we are first recalling the appropriate definition when working on the space of complex matrices. For any couple of matrices X and Y we define the scalar product $\langle X, Y \rangle$ as follows

$$\langle X, Y \rangle = \mathrm{Re}\,(\mathrm{Trace}\, XY^*) = \mathrm{Re} \sum_i \sum_j x_{i,j} \overline{y}_{i,j}, \tag{21}$$

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

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

It seems appropriate to call this the Frobenius scalar product since it induces the Frobenius norm : $\langle X, X \rangle = \|X\|_F^2$. From the above relation it easily follows that if X and Y are partitioned conformably into blocks $X_{i,j}$ and $Y_{i,j}$ the we have the identity

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

Next we define on the set of pseudo-polynomial matrices a scalar product that is conformable with the above definition.

Real axis

We first start with non-negative matrices on the real axis. For a set of non-negative polynomials $P(x) = \sum_{k=0}^{2n} P_k x^k$ and $Q(x) = \sum_{k=0}^{2n} Q_k x^k$ we define a scalar product $\langle P, Q\rangle_{\Re}$ as follows :

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

It turns out that several important optimization problems can be formulated in the following standard form :

$$\min_{P\in\mathcal{K}_{\Re}} \{\langle C, P\rangle_{\Re} : \langle A_\ell, P\rangle_{\Re} = b_\ell, \ell = 1, \dots, q\}, \qquad (22)$$

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

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

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

$$P(x) \succeq 0, \quad x \in \Re.$$

As $P \in \mathcal{K}_{\Re}$ necessarily implies $P_k = P_k^*$ for all k, there is no restriction to assume all the $m \times m$ blocks C_k of C and blocks $A_{\ell,k}$ of A_ℓ to be 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 will be in the cone $\mathcal{K}_{\Re}$ iff there exists a non-negative block matrix Y with blocks $Y_{i,j}, i, j = 0, \dots n$ of dimension $m\times m$ satisfying

$$P_k = \sum_{i+j=k} Y_{i,j}, \quad k = 0, 1, \dots, 2n. \qquad (23)$$

By definition, the dual cone $\mathcal{K}^*_{\Re}$ is made of the matrix coefficients $Q \doteq [Q_0,\ Q_1, \dots,\ Q_{2n}\ ]$ of the matrix polynomials satisfying the constraint

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

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

$$H(Q) \doteq \begin{bmatrix} Q_0 & Q_1 & \dots & Q_n \\ Q_1 & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & Q_{2n-1} \\ Q_n & \dots & Q_{2n-1} & Q_{2n} \end{bmatrix} \qquad (24)$$

it turns out that one has the relation

$$\begin{aligned} \langle Q, P\rangle_{\Re} &= \sum_{k=0}^{2n} \langle Q_k, P_k\rangle_{\Re} = \sum_{k=0}^{2n} \sum_{i+j=k} \langle Q_k, Y_{i,j}\rangle_{\Re} \\ &= \langle H(Q), Y\rangle \end{aligned} \qquad (25)$$

because of (23) and the properties of the scalar product. Moreover, it can be shown that

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

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

As a consequence, the optimization problem (22) can be restated in its dual form

$$\max_{u_1,\dots,u_\ell} \{\sum_{\ell=1}^{q} b_\ell u_\ell : H(C - \sum_{\ell=1}^{q} u_\ell A_\ell) \succeq 0\}. \qquad (26)$$

From a numerical point of view, this dual formulation (26) has a considerable advantage over the primal form (22) since it involves an optimization in a space of variables of dimension q rather than $(n+1)m^2$. It is well known [8] that optimization problems of this type can be solved efficiently with the help of interior point methods and that their numerical implementation requires the calculation of the first and second derivatives of the barrier function

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

These derivatives can be expressed as follows. Denote

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

Then one derives that

$$\begin{aligned} \tfrac{\partial f(u)}{\partial u_\ell} &= \langle H(S)^{-1}, H(A_\ell)\rangle, \\ \tfrac{\partial^2 f(u)}{\partial u_\ell \partial u_s} &= \langle H(S)^{-1} H(A_\ell) H(S)^{-1}, H(A_s)\rangle. \end{aligned} \qquad (27)$$

Unit circle

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

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

The optimization problem now reads :

$$\min_{P\in\mathcal{K}_{\mathcal{C}}} \{\langle C, P\rangle_{\mathcal{C}} : \langle A_\ell, P\rangle_{\mathcal{C}} = b_\ell, \ell = 1, \dots, q\}, \qquad (28)$$

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

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

of non-negative pseudo-polynomial matrices

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

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

$$P_k = \sum_{i-j=k} Y_{i,j}, \quad k = -n, \dots, n. \qquad (29)$$

As before, there is no restriction to assume that the $m \times m$ blocks C_k of C and blocks $A_{\ell,k}$ of A_ℓ have the same type of symmetry as the blocks of P, since this will not affect the scalar products.

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

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

of the para-hermitian pseudo-polynomials satisfying the constraint

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

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

$$T(Q) \doteq \begin{bmatrix} Q_0 & Q_1 & \dots & Q_n \\ Q_1^* & Q_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & Q_1 \\ Q_n^* & \dots & Q_1^* & Q_0 \end{bmatrix} \tag{30}$$

one has the relations

$$\begin{aligned} \langle Q, P\rangle_{\mathcal{C}} &= \sum_{k=-n}^{n} \langle Q_k, P_k\rangle_{\Re} = \sum_{k=-n}^{n} \sum_{i-j=k} \langle Q_k, Y_{i,j}\rangle_{\Re} \\ &= \langle T(Q), Y\rangle \end{aligned} \tag{31}$$

so that the dual cone $\mathcal{K}_{\mathcal{C}}^*$ is characterized by $T(Q) \succeq 0$. Therefore the dual optimization problem of (28) becomes

$$\max_{u_1,\dots,u_\ell} \{\sum_{\ell=1}^{q} b_\ell u_\ell : T(C - \sum_{\ell=1}^{q} u_\ell A_\ell) \succeq 0\} \tag{32}$$

for which the appropriate barrier function is

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

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

$$\begin{aligned} \frac{\partial f(u)}{\partial u_\ell} &= \langle T(S)^{-1}, T(A_\ell)\rangle, \\ \frac{\partial^2 f(u)}{\partial u_\ell \partial u_s} &= \langle T(S)^{-1} T(A_\ell) T(S)^{-1}, T(A_s)\rangle, \end{aligned} \tag{33}$$

where

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

Imaginary axis

The imaginary case reformulation is left to the reader since, as shown in the previous section, it can be reduced to the real line problem in a trivial manner.

6 Computational aspects

In this section we consider Hermitian $(n+1)\times(n+1)$ block Toeplitz matrices with arbitrary $m \times m$ matrix blocks T_i :

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

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

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

Let us first define the block permutation matrix J :

$$J \doteq \begin{bmatrix} 0 & \dots & 0 & I_m \\ \vdots & ⋰ & ⋰ & 0 \\ 0 & ⋰ & ⋰ & \vdots \\ I_m & 0 & \dots & 0 \end{bmatrix}$$

The displacement theory of Toeplitz and Hankel matrices is well established [5] 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 :

$$\nabla_t T \doteq T - Z^T T Z, \quad \nabla_h H \doteq H - ZHZ.$$

It is easy to see that

$$\nabla_t T \doteq \begin{bmatrix} T_0 & T_1 & \dots & T_n \\ T_1^* & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ T_n^* & 0 & \dots & 0 \end{bmatrix}, \tag{34}$$

$$\nabla_h H = \begin{bmatrix} H_0 & 0 & \dots & 0 \\ H_1 & \vdots & & \vdots \\ \vdots & 0 & \dots & 0 \\ H_n & \dots & H_{2n-1} & H_{2n} \end{bmatrix}. \tag{35}$$

From the above expressions one sees that the original matrices T and H can be recovered from their respective displacement, and the inverse operator is easy to write down. One easily checks that :

$$T = \nabla_t T + Z^T \cdot \nabla_t T \cdot Z + \dots + Z^{nT} \cdot \nabla_t T \cdot Z^n, \tag{36}$$

and

$$H = \nabla_h H + Z \cdot \nabla_h H \cdot Z + \dots + Z^n \cdot \nabla_h H \cdot Z^n. \tag{37}$$

The proof of these inversion formulas is obtained by merely applying the displacement operator again to both sides of the equations.

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 yields indeed a block Toeplitz matrix JH, which we can define as T provided we choose $T_i = H_{i+n},\ i = -n, \dots, n$. Since $Z^T = JZJ$ we also have that the displacement operators are then related in a similar fashion :

$$T = JH \Leftrightarrow \nabla_t T = J\nabla_h H. \tag{38}$$

From the sparsity structure of the matrices in (34,35) 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 [5] tells us that the inverse of T or H (when it exists) has a displacement rank bounded by that of the matrix itself :

$$\text{rank } \nabla_t T^{-1} \leq \text{rank } \nabla_t T, \quad \text{rank } \nabla_h H^{-1} \leq \text{rank } \nabla_h H. \tag{39}$$

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 (34,35) it is very simple to construct low rank factorizations of $\nabla_t T$ or $\nabla_h H$:

$$\nabla_t T = F_t^*.G_t, \quad \nabla_h H = F_h^*.G_h, \tag{40}$$

where the number of rows of F_t and G_t equals $r_t \doteq \text{rank } \nabla_t T$ and the number of rows of F_h and G_h equals $r_h \doteq \text{rank } \nabla_h H$. But given such factorizations, there exist fast algorithms to derive from them the corresponding factorizations of the displacement of the inverses :

$$\nabla_t T^{-1} = A_t^*.B_t, \quad \nabla_h H^{-1} = A_h^*.B_h, \tag{41}$$

and these precise decompositions will be used in the sequel. We should point out that these factorizations are not unique and that for positive definite matrices T and H there exist particular choices of factorizations that indeed reflect these properties. In the sequel we will not worry about these aspects since they will only affect marginally the complexity results we want to stress.

Let us focus first on the case of Toeplitz displacement of a matrix X and suppose we have computed a rank r_t factorization of its Toeplitz displacement $\nabla_t X$:

$$\nabla_t X = A^* \cdot B \tag{42}$$

where A and B have dimensions $r_t \times m(n+1)$. We define an upper block triangular Toeplitz matrix $U(B)$ as a function of the partitioned matrix B, where each sub-block has dimensions $r_t \times m$:

$$B \doteq \left[\begin{array}{cccc} B_0 & B_1 & \dots & B_n \end{array}\right],$$

$$U(B) \doteq \left[\begin{array}{cccc} B_0 & B_1 & \dots & B_n \\ 0 & B_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_1 \\ 0 & \dots & 0 & B_0 \end{array}\right].$$

Doing the same for the matrix A we have

$$A \doteq \left[\begin{array}{cccc} A_0 & A_1 & \dots & A_n \end{array}\right],$$

$$U(A)^* \doteq \left[\begin{array}{cccc} A_0^* & 0 & \dots & 0 \\ A_1^* & A_0^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ A_n^* & \dots & A_1^* & A_0^* \end{array}\right].$$

It follows from the displacement equation $\nabla_t X = A^* \cdot B$ that

$$X = \sum_{j=0}^{n} (AZ^j)^*(BZ^j) = U(A)^* U(B) =$$

$$\left[\begin{array}{cccc} A_0^* & 0 & \dots & 0 \\ A_1^* & A_0^* & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ A_n^* & \dots & A_1^* & A_0^* \end{array}\right] \cdot \left[\begin{array}{cccc} B_0 & B_1 & \dots & B_n \\ 0 & B_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_1 \\ 0 & \dots & 0 & B_0 \end{array}\right]. \tag{43}$$

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

For a Hankel displacement $\nabla_h X$ of a matrix X we have a similar representation starting from the rank r_h factorization of $\nabla_h X$:

$$\nabla_h X = A^* \cdot B \tag{44}$$

If we now partition the matrix A in reversed order

$$A \doteq \left[\begin{array}{ccc} A_0 & \dots & A_n \end{array}\right] \Leftrightarrow AJ \doteq \left[\begin{array}{ccc} A_n & \dots & A_o \end{array}\right]$$

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

$$X = \sum_{j=0}^{n} (Z^j J A^*)(BZ^j) = JU(A)^* U(B) =$$

$$\left[\begin{array}{cccc} A_n^* & \dots & A_1^* & A_0^* \\ \vdots & \iddots & \iddots & 0 \\ A_1^* & A_0^* & \iddots & \vdots \\ A_0^* & 0 & \dots & 0 \end{array}\right] \cdot \left[\begin{array}{cccc} B_0 & B_1 & \dots & B_n \\ 0 & B_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & B_1 \\ 0 & \dots & 0 & B_0 \end{array}\right]. \tag{45}$$

When applied to a particular choice of displacement factors A and B for the inverse of a Hankel matrix, this formula is also known as the Christoffel-Darboux formula for $X = H^{-1}$. The complexity of the construction of the generators for the inverse of X is $O(rm^2 n \log^2 n)$..

Let us now see how to compute scalar products of the type

$$\langle X, Z^j \rangle, \quad \langle X, (Z^i)^T \rangle$$

where $\nabla_t X$ is given. For a Hermitian matrix X, it turns out that $\langle X, (Z^i)^T \rangle = \langle X, Z^i \rangle$ so that only one expression has to be evaluated. When $\nabla_h X$ is given we need to evaluate scalar products of the type

$$\langle X, JZ^j \rangle, \quad \langle X, J(Z^i)^T \rangle$$

but for Hermitian matrices we have $\langle X, J(Z^i)^T \rangle = \langle XJ, Z^i \rangle$ so that again only one expression has to be evaluated.

We first consider matrices X given by their Toeplitz displacement $\nabla_t X = A^* \cdot B$. Since we can write

$$U(A) = \sum_{k=0}^{n} \mathrm{diag}\,\{A_k\} Z^k, \quad U(B) = \sum_{k=0}^{n} \mathrm{diag}\,\{B_k\} Z^k.$$

and we have that

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

then we obtain

$$\langle U(A)^* U(B), Z^j \rangle =$$

$$\langle (n+1-j) B_j^* A_0 + \ldots + 2B_{n-1}^* A_{n-j-1} + B_n^* A_{n-j} \rangle.$$

One easily checks that

$$\langle U(B)^* U(A), Z^j \rangle =$$

$$\langle (n+1-j) A_j^* B_0 + \ldots + 2A_{n-1}^* B_{n-j-1} + A_n^* B_{n-j} \rangle.$$

Notice that since X is Hermitian, only one of these two expressions has to be computed since they are identical. These quantities clearly result from the convolution of the block vectors

$$[(n+1)B_0, nB_1, \cdots 2B_{n-1}, B_n], \ [A_0, A_1, \cdots, A_{n-1}, A_n],$$

and

$$[(n+1)A_0, nA_1, \cdots 2A_{n-1}, A_n], \ [B_0, B_1, \cdots, B_{n-1}, B_n],$$

which has a complexity of $O(r_t mn \log_2 n)$. For a matrix of displacement rank r_t, the overall complexity is thus $O(r_t n \log_2 n)$, *provided* that the matrices A and B are given.

We then consider matrices X given by their Hankel displacement $\nabla_h X = A^* \cdot B$. The considered inner products can in fact be rewritten in terms of JX as follows :

$$\langle X, JZ^j \rangle = \langle JX, Z^j \rangle, \quad \langle X, J(Z^i)^T \rangle = \langle (JX)^*, Z^i \rangle,$$

and since JX is block Toeplitz, we can again apply the same formulas as above. Our choice of relabeling the sub-blocks of the matrix A in reversed order in (45), actually yields exactly the same formulas for these inner products.

We also need the computation of inner products

$$< T(M)^{-1} T(Y) T(M)^{-1}, Z^j >, \quad j = 0, \ldots, n.$$

which can be obtained from

$$\begin{bmatrix} -T(Y) & T(M) \\ T(M) & 0 \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} 0 & T(M)^{-1} \\ T(M)^{-1} & T(M)^{-1} T(Y) T(M)^{-1} \end{bmatrix}.$$

The matrix on the left can also be permuted to a block Toeplitz matrix :

$$\hat{T} \doteq P \begin{bmatrix} -T(Y) & T(M) \\ T(M) & 0 \end{bmatrix} P =$$

$$\begin{bmatrix} \begin{bmatrix} -Y_0 & M_0 \\ M_0 & 0 \end{bmatrix} & \cdots & \cdots & \begin{bmatrix} -Y_n & M_n \\ M_n & 0 \end{bmatrix} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \begin{bmatrix} -Y_{-n} & M_{-n} \\ M_{-n} & 0 \end{bmatrix} & \cdots & \cdots & \begin{bmatrix} -Y_0 & M_0 \\ M_0 & 0 \end{bmatrix} \end{bmatrix}$$

but with $2m \times 2m$ blocks. In order to apply fast algorithms to $\hat{T}$ one needs to assume that certain submatrices of $\hat{T}$ are invertible but this follows easily from the positive definiteness of $T(M)$. So one can construct the factors $\hat{A}$ and $\hat{B}$ of the factorization $\nabla_t \hat{T} = \hat{T} - \hat{Z}^T \hat{T} \hat{Z} = \hat{A}^* \hat{B}$ at low cost. Selecting the appropriate rows of these matrices will yield a similar factorization for $\nabla_t (T(M)^{-1} T(Y) T(M)^{-1})$ and then we again apply the above formulas to compute the relevant inner products. The results for Hankel displacements are completely analogous.

7 Positive para-hermitian transfer functions

It is a well known result of state-space theory [9] that any proper transfer function of that type admits minimal realizations of the form

$$\Phi(s) = \begin{bmatrix} B^*(-sI_n - A^*)^{-1}, & I_m \end{bmatrix} Y_0 \begin{bmatrix} (sI_n - A)^{-1} B \\ I_m \end{bmatrix} \tag{46}$$

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 of an appropriate transformation of the variable s. Clearly, Y_0 is not uniquely defined from $\Phi(s)$. Indeed, let us replace Y_0 by $Y(X)$, defined as follows :

$$Y(X) = Y_0 + \begin{bmatrix} XA + A^*X & XB \\ B^*X & 0 \end{bmatrix} \tag{47}$$

and where X is any $n \times n$ Hermitian matrix. $\Phi(s)$ 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 [12], [9] says that the existence of a Hermitian matrix X such that $Y(X)$ is non-negative definite is a necessary and sufficient condition for $\Phi(s)$ to be a para-hermitian transfer function non-negative on the whole of the imaginary axis.

As the variable transformation $s = jx$ maps the imaginary axis onto the real axis, one can transform any para-hermitian transfer function into a hermitian transfer matrix and conversely. The corresponding realization then becomes :

$$\Phi(x) = \begin{bmatrix} B^*(xI_n - A^*)^{-1}, & I_m \end{bmatrix} Y_0 \begin{bmatrix} (xI_n - A)^{-1}B \\ I_m \end{bmatrix} \tag{48}$$

where Y is the same Hermitian matrix as before. Furthermore, if $\Phi(x)$ is non-negative definite for real x, one derives from previous case that there must exist skew-hermitian matrices $X = -X^*$ such that the Hermitian matrix

$$Y(X) = Y_0 + \begin{bmatrix} -XA + A^*X & -XB \\ B^*X & 0 \end{bmatrix} \tag{49}$$

is non-negative definite.

Similarly, the variable transformation $s = (z-1)/(z+1)$ maps the imaginary axis onto the unit circle. Therefore, one can transform a para-hermitian transfer function that is non-negative on the imaginary axis into a para-hermitian transfer function that is non-negative on the unit circle and conversely (see e.g. [3]). The realization then becomes

$$\Phi(z) = \begin{bmatrix} zB^*(I_n - zA^*)^{-1}, & I_m \end{bmatrix} Y_0 \begin{bmatrix} (zI_n - A)^{-1}B \\ I_m \end{bmatrix} \tag{50}$$

and $\Phi(z)$ is non-negative definite on the unit circle, iff there exist a Hermitian matrix X such that

$$Y(X) = Y_0 + \begin{bmatrix} A^*XA - X & A^*XB \\ B^*XA & B^*XB \end{bmatrix} \tag{51}$$

is non-negative definite.

From these conditions one can re-derive the results of section 4. This is done in the paper [3] starting from realizations of the pseudo-polynomial matrix $\Phi(p)$ for each of the three cases. The result follows by using $A = Z$, the block-shift matrix, and $B = [0, \dots, 0, I_m]^T$. Using these definitions we indeed have

$$(pI - A)^{-1}pB = \Pi(p)(p^{-n})$$

which links the realization to our pseudo-polynomials. Moreover, the linear matrix inequalities (49,51,47) then become the equations in X, Y and X_0 of section 4.

Acknowledgements

This paper presents research 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 author.

References

[1] N. I. Akhiezer, *The classical momernt problem*, Oliver & Boyd, London, 1965.

[2] S. Boyd, L. El-Ghaoui, E. Feron, V. Balakrishnan, *Linear Matrix Inequalities in Systems and Control Theory*, Studies in Applied Mathematics, **15**, SIAM, Philadelphia, PA, 1994.

[3] Y. Genin, Yu. Nesterov, P. Van Dooren, "Positive transfer functions and convex optimization", in *Proceedings ECC-99 Conference*, Karlsruhe, Paper F-143, 1999.

[4] V. Ionescu, C. Oara, M. Weiss, "Generalized Riccati theory and robust control. A Popov function approach", Wiley & Sons, New York, 1998.

[5] T. Kailath, A. Sayed, "Displacement structure : theory and applications", *SIAM Review*, **37**, pp. 297-386, 1995.

[6] S. Karlin, W. J. Studden. *Tchebycheff systems : with applications in analysis and statistics*, John Wiley and Sons, 1966.

[7] Yu. Nesterov, "Squared Functional Systems and optimization Problems", In H.Frenk et al. (Eds), *High Performance Optimization*, chapter 17, pp. 405-440. Kluwer Academic Publishers, 2000.

[8] Yu. Nesterov, A. Nemirovsky, *Interior Point Polynomial Methods in Convex Programming*, Studies in Applied Mathematics, SIAM, Philadelphia, PA, 1994

[9] V. Popov, *Hyperstability of Control Systems*, Springer Verlag, Berlin, 1973 (Roumanian version in 1966).

[10] J. C. Willems, "Least squares stationary Optimal Control and the algebraic Riccati equation", *IEEE Trans. Aut. Contr.*, **AC-16**, pp.621-634, 1971.

[11] D. Youla, "On the factorization of rational matrices", *IRE Trans. Inform. Theory*, **IT7**, pp. 172-189, 1961.

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