Formal orthogonal polynomials and Hankel/Toeplitz duality*

Adhemar Bultheel and Marc Van Barel
Department of Computer Science
K.U. Leuven

March 13, 2000

Abstract

For classical polynomials orthogonal with respect to a positive measure supported on the real line, the moment matrix is Hankel and positive definite. The polynomials satisfy a three term recurrence relation. When the measure is supported on the complex unit circle, the moment matrix is positive definite and Toeplitz. They satisfy a coupled Szegő recurrence relation but also a three term recurrence relation. In this paper we study the generalization for formal polynomials orthogonal with respect to an arbitrary moment matrix and consider arbitrary Hankel and Toeplitz matrices as special cases. The relation with Padé approximation and with Krylov subspace iterative methods is also outlined.

1 Introduction

In classical situations, formal orthogonal polynomials (OPs) are studied with respect to some linear functional. The "inner product" $\langle p,q\rangle$ of two polynomials is defined as

$$\langle p,q\rangle = \mu(pq)$$

where μ is a linear functional defined on the set of polynomials through the moments

$$\mu(z^k) = \mu_k, \quad k = 0,1,2,\ldots$$

This implies that $\langle z^i, z^j\rangle = \langle z^{i+j},1\rangle = \langle 1, z^{i+j}\rangle = \mu_{i+j}$ and therefore, we get a moment matrix for the basis $\{z^k : k = 0,1,\ldots\}$ which has a Hankel structure:

$$M = [\mu_{i,j}] = \left[\left\langle z^i, z^j\right\rangle\right] = [\mu_{i+j}], \quad i,j = 0,1,\ldots$$

In the so called normal situation, all leading submatrices M_n of size n are regular. (The matrix M is then called strongly regular.) This situation is well understood. A classical treatment is [2].

*This research was supported by the National Fund for Scientific Research (NFWO), project Lanczos, grant #2.0042.93.

In general there are singular leading submatrices of the moment matrix, but there is a certain structure in the grouping of singular submatrices of a Hankel matrix. This is usually reflected in a Block Structure Theorem of some kind [17]. The OPs will also be grouped in blocks which will be block orthogonal, that is, each polynomial in a block will be orthogonal to all polynomials in previous blocks, but they need not be orthogonal to other polynomials in the same block. Instead of speaking about formal block orthogonal polynomials, we shall go on calling them OPs for simplicity.

These formal OPs are generalizations of classical OPs with respect to a positive measure supported on (part of) the real line. See [13]. They still satisfy formal analogs of a three term recurrence relation and a Christoffel-Darboux type formula. These properties can be written in matrix notation and so they give rise to Hankel factorization properties, Hankel inversion formulas, Jacobi matrices etc. See [18]. They are also related to Padé approximation and continued fractions [39, Chap. 11],[40]. And they also deliver the formulas in Lanczos-type methods for the iterative solution of large (sparse) linear systems or eigenvalueproblems [33, 23, 24, 22].

On the other hand, there is a Szegő theory of OPs with respect to a measure supported on the complex unit circle. These Szegő polynomials also satisfy a three term recurrence relation, but also a coupled recurrence relation for the polynomials and their reciprocals. In this classical situation, the moment matrix is Toeplitz, Hermitian, strongly regular and positive definite. Here the formal generalization is to consider a moment matrix which is an arbitrary Toeplitz matrix. They will also satisfy a block orthogonality relation. There are many correspondences with the Hankel case, but also many differences. One thing that is lost is the symmetry and one should in fact speak about biorthogonal polynomials since left and right OPs are not the same. This was not the case for the Hankel matrix since a Hankel matrix is always symmetric. Again it is possible to generalize the recurrence relations, Christoffel-Darboux formulas, factorization and inversion formulas for Toeplitz matrices (usually referred to as Gohberg-Semencul formulas). The Jacobi matrix, which is tridiagonal (or block tridiagonal) in the Hankel case has however to be replaced by an upper Hessenberg matrix (which is tridiagonal again in the symmetric case). These OPs are related to Laurent-Padé and two-point Padé approximation.

The next step in the generalization is to consider a moment matrix which has no structure whatsoever. How much of the previous properties can be generalized and what is lost? This would be a generalization of OPs with respect to a measure with a support somewhere in the complex plane (e.g., somewhere on a Jordan curve or on some Julia set). Such formal orthogonalities were also considered by Brezinski [6, 4, 3]. Of course there is a big gap between the highly structured Hankel or Toeplitz matrices and arbitrary matrices. Several intermediate possibilities can be considered. There are other kinds of structured matrices (Sylvester, Bezoutian, Löwner, Cauchy, Vandermonde, etc.), there is the notion of displacement rank [28, 26], which covers the whole gap. Also block Hankel and Toeplitz matrices can be considered. The vector orthogonal polynomials of dimension d as considered in [6, 38], correspond to Hankel ($d > 0$) or Toeplitz ($d < 0$) matrices whose blocks are $|d| \times 1$.

Formal orthogonal polynomials appear implicitly or explicitly in different situations. They are known and used for a long time in connection with Padé approximation and continued fractions. Nowadays they became popular in connection with Lanczos type iterative methods of linear algebra. See in this connection the contribution of C. Brezinski in the na-digest [5], which contains several references.

The purpose of this paper is to bring together all these applications and at the same time

place them in a more general framework of formal orthogonal polynomials. This is done in Section 2. As far as we know, this has never been done before. It will be shown how much of the classical theory of Hankel and Toeplitz matrices can be generalized.
In section 3, these results are specialized to the Hankel case. Most of the results in this section are of course well known: the relation with Padé approximation, continued fractions and Krylov subspace methods in linear algebra. We include these results anyway to make the connection with the general case clear. The relation between Schur complements and the Euclidean algorithm seems to be new.
For the Toeplitz case, considered in Section 4, also the literature can provide several algorithms to solve the problems for a matrix which is not strongly regular. A survey of all these results and their interconnection is not readily available though. In particular, the results about the block factorization of the Hessenberg matrix as a product of a block bidiagonal and the inverse of a block bidiagonal has not been considered before.

One note of warning: although this paper ties up with many numerical problems (orthogonal polynomials, moment problems, continued fractions, solution of linear systems, eigenvalue computation, Gram-Schmidt orthogonalization, difference equations, Schur complements, rational approximation, ...) and algorithms (Euclidean algorithm and the Schur algorithm, fast algorithms for structured matrices and interpolation problems, iterative methods in linear algebra, etc.), we do not discuss their numerical aspects in detail and suppose that computations are performed in exact arithmetic. Our only objective is reveal their common backbone of formal orthogonal polynomials. Discussing the numerical aspects would lead us too far since each of these numerical problems has its own requirements for the way a practical implementation is realized and the technicalities may be divergent. We can however mention that an important tool to cope with stability problems is the design of look-ahead versions of the proposed algorithms. Such versions exist for the Hankel and Toeplitz case for which we refer to the existing literature. This look-ahead idea is adaptable for the general situation but we do not discuss it here.

We now introduce some notation. If p is a polynomial $p(z) = \sum_0^n p_i z^i$, then P will denote its coefficient vector, i.e., the column matrix of its coefficients $\mathrm{P} = [p_0, p_1, \ldots]^T$. If we denote $\mathbf{z} = [1, z, z^2, \ldots]^T$, then $p(z) = \mathbf{z}^T\mathrm{P}$. On occasion we also use $\mathbf{x}$, $\mathbf{w}$, $\mathbf{y}$ with a meaning similar to $\mathbf{z}$. If $p(z)$ is monic, then, we shall use the notation $\dot{p}(z)$ to denote the polynomial p without its highest degree term, i.e., if p is of degree n, then $\dot{p}(z) = p(z) - z^n$. Of course $\dot{\mathrm{P}}$ will be its coefficient vector. The coefficient vectors will be of flexible length, i.e., they can be extended with a number of zeros, possibly infinitely many, to match the dimensions in the formula where they appear. The Frobenius or companion matrix for the monic polynomial $p(z)$ is defined as

$$F(p) = \begin{bmatrix} 0 & 0 & \cdots & 0 & \\ 1 & & & & \\ & 1 & & & -\dot{\mathrm{P}} \\ & & \ddots & & \\ & & & 1 & \end{bmatrix}, \quad \mathrm{P}^T = [\dot{\mathrm{P}}^T \; 1].$$

The matrix Z will denote the down shift matrix. It will also be assumed to have a flexible length. The effect of the multiplication MZ is that all the elements of M are shifted one place to the left. We shall indicate this by the notation $M^< = MZ$. Similarly, Z^TM will shift the elements one place up and we shall denote it as $M^\wedge = Z^TM$.

For a matrix $M = [\mu_{ij}]$, we shall use the notation $M[k:l, m:n]$ to denote the submatrix

$$M[k:l,m:n] = \begin{bmatrix} \mu_{k,m} & \cdots & \mu_{k,n} \\ \vdots & & \vdots \\ \mu_{l,m} & \cdots & \mu_{l,n} \end{bmatrix}.$$

We abbreviate the leading $k \times l$ submatrix $M[0:k-1, 0:l-1]$ as $M[k,l]$ and we use $M^{[i,*]}$ to mean the i-th row $M[i:i, 0:\infty]$ and similarly, $M^{[*,j]}$ denotes column j.

2 Formal orthogonality: general moment matrix

2.1 Moment matrix and generator

Consider a complex bilinear form $\langle \cdot, \cdot \rangle$ defined on the set of complex polynomials $\mathbb{C}[z]$, which is linear in the second factor and conjugate linear in the first one:

$$\begin{aligned} \langle a_1 p_1 + a_2 p_2, q \rangle &= \overline{a}_1 \langle p_1, q \rangle + \overline{a}_2 \langle p_2, q \rangle \\ \langle p, a_1 q_1 + a_2 q_2 \rangle &= a_1 \langle p, q_1 \rangle + a_2 \langle p, q_2 \rangle . \end{aligned}$$

It is defined by its moments

$$\mu_{ij} = \left\langle z^i, z^j \right\rangle, \quad i,j = 0,1,2,\ldots \tag{2.1}$$

If it is Hermitian, then $\mu_{ji} = \overline{\mu}_{ij}$ but we shall not preassume this symmetry in general. The moments are collected in a moment matrix $M = [\mu_{ij}]$. In general we shall not suppose that M is positive definite nor that it is strongly regular. The latter means that the leading submatrices $M[\nu,\nu]$ are not necessarily regular for all ν.

We call the formal bivariate series

$$M(w,z) = \mathbf{w}^H M \mathbf{z} = \sum_{i,j=0}^{\infty} \mu_{ij} \overline{w}^i z^j$$

the generator for the moment matrix M. The generator for row i is given by

$$M^{[i,*]}(z) = \sum_{j=0}^{\infty} \mu_{ij} z^j = \left\langle \zeta^i, (1 - z\zeta)^{-1} \right\rangle$$

where we wrote $(1 - z\zeta)^{-1}$ for the formal series $\sum_{j=0}^{\infty} \zeta^j z^j$. A similar definition holds for the columns. Thus,

$$M(w,z) = \left\langle (1-w\zeta)^{-1}, (1-z\zeta)^{-1} \right\rangle = \sum_{i=0}^{\infty} M^{[i,*]}(z) \overline{w}^i = \sum_{j=0}^{\infty} M^{[*,j]}(w) z^j.$$

2.2 Block biorthogonality

Using the coefficient vectors of the polynomials and the moment matrix, we can express the inner product as follows

$$\langle p, q \rangle = \mathrm{P}^H M \mathrm{Q}, \quad p, q \in \mathbb{C}[z].$$

If the moment matrix were strongly regular, then we could use the two-sided Gram-Schmidt procedure to orthogonalize the system of polynomials $1, z, z^2, \ldots$ to generate the biorthogonal system $\{b_i, a_i\}_{i \geq 0}$, where a_i and b_i may be chosen to be monic polynomials of degree i. Biorthogonality means that the polynomials satisfy $\langle b_i, a_j \rangle = \delta_{ij} D_{ii}$, $D_{ii} \neq 0$. The set $\{b_i\}$ is the set of left orthogonal polynomials (LOP) and $\{a_i\}$ is the set of right orthogonal polynomials (ROP). Defining the unit upper triangular matrices A and B by

$$[a_0(z), a_1(z), \ldots] = \mathbf{z}^T A \quad \text{and} \quad [b_0(z), b_1(z), \ldots] = \mathbf{z}^T B$$

then biorthogonality means that we have

$$B^H M A = D, \quad D = \operatorname{diag}(D_{00}, D_{11}, \ldots) \tag{2.2}$$

where the entries D_{kk} are nonzero.

In general however, we can only guarantee the existence of a *block* biorthogonal system. This means the following. Suppose the submatrices $M[\nu, \nu]$ of the moment matrix M for $\nu = \nu_1, \nu_2, \ldots, \nu_N$ are the successive regular ones. If N is finite, we define $\nu_{N+1} = \infty$ and in any case set $\nu_0 = 0$. Then $\alpha_{n+1} = \nu_{n+1} - \nu_n$ for $n = 0, 1, \ldots, N$ will be the number of polynomials that will be grouped in block n.

So let block n consist of the α_{n+1} ROPs

$$\mathbf{a}_n = [a_{\nu_n}, a_{\nu_n+1}, \ldots, a_{\nu_n+\alpha_{n+1}-1}] = [a_n^1, a_n^2, \ldots, a_n^{\alpha_{n+1}}].$$

Thus we use the notation a_n^k to denote the kth polynomial of block n, i.e., $a_n^k = a_{\nu_n+k-1}$, $k = 1, \ldots, \alpha_{n+1}$. Using a similar notation for the LOPs, one has

$$\langle \mathbf{b}_i, \mathbf{a}_j \rangle = \left[\langle b_i^k, a_j^l \rangle \right]_{k=1,\ldots,\alpha_{i+1}}^{l=1,\ldots,\alpha_{j+1}} = \delta_{ij} D_{ij}$$

where D_{jj} is a square matrix of size α_{j+1}, and δ_{ij} represents the Kronecker delta. We have again the relation (2.2), but now with D_{jj} square matrices. The choice of the block sizes will guarantee that the D_{jj} are regular. In fact it guarantees that there is no factorization of the form (2.2) with smaller blocks that are regular. For further reference we state its finite dimensional version. Suppose that an index n for the matrices below means that we take the $\nu_{n+1} \times \nu_{n+1}$ leading submatrices[1], then we get from (2.2)

$$B_n^H M_n A_n = D_n. \tag{2.3}$$

In fact such a relation holds for any $\nu \times \nu$ leading submatrix (not only the regular ones). It then follows that also all the leading submatrices of each block D_{kk} of D are singular, except for the complete block which is regular. Indeed, B and A are unit lower triangular matrices. Hence any of its leading submatrices has determinant 1 and thus is $\det M[\nu, \nu] = \det D[\nu, \nu]$. Since $M[\nu, \nu]$ is nonsingular iff ν is equal to some index ν_n, the result follows.

In what follows, we shall concentrate on the right biorthogonal polynomials and call them ROPs for short. Of course, similar results can be obtained for the left biorthogonal polynomials.

The first ROP a_n^1 of block n has degree ν_n and it is right orthogonal to all polynomials of lower degree:

$$\left\langle z^i, a_n^1(z) \right\rangle = 0 \text{ for } i = 0, 1, \ldots, \nu_n - 1.$$

[1]We stick to the convention that an index ν will refer to a scalar numbering, while an index n will refer to a block numbering.

Note however, that $\langle z^{\nu_n}, a_n^1\rangle$ can be zero. Following Draux [13], we could call it a regular OP. It is also a True right Orthogonal Polynomial in the classical sense (i.e., not in block sense). Therefore we shall call it the TOP of block n. It is unique if monically normalized since its coefficient vector has to satisfy a linear system

$$M_{n-1}\dot{\mathrm{A}}_n^1 = -[\mu_{0,\nu_n}, \ldots, \mu_{\nu_n-1,\nu_n}]^T \tag{2.4}$$

where $\mathrm{A}_n^1 = [\dot{\mathrm{A}}_n^{1T}\ 1]^T$ and $M_{n-1} = M[\nu_n, \nu_n]$ is a regular submatrix of M. The remaining polynomials in block n will be called Internal OPs (IOPs). These IOPs are not uniquely defined in general. We do require that they have precise degree, i.e. a_n^k has degree $\nu_n + k - 1$ (we could, e.g. make them monic) but they need only be orthogonal to all polynomials in previous blocks but not to previous polynomials in the same block. The latter would be impossible for all polynomials in a block because of the singularity of the leading submatrices when one is inside a block.

There is a way to make them uniquely defined though, when apart from being monic, we require a_n^k to be not only right orthogonal to all $\mathbf{b}_j$, $j < n$ of corresponding previous left blocks, but also to all b_n^i polynomials in the corresponding left block, with the exception of one. Thus a_n^k is orthogonal to all b_m^i for $m = 0, 1, \ldots, n-1$, $i = 1, \ldots, \alpha_{m+1}$ and also to b_n^l, $l = 1, \ldots, \alpha_{n+1}$, $l \neq l_{k,n}$ with $\{l_{k,n} : k = 1, \ldots, \alpha_{n+1}\} = \{1, \ldots, \alpha_{n+1}\}$. That is, we require $D_{nn} = \langle \mathbf{b}_n, \mathbf{a}_n\rangle$ to be a permuted diagonal matrix. This is always possible since one Gaussian elimination with pivoting shows that for every regular matrix D, there exists a unit upper triangular matrix U and a unit lower triangular matrix L such that LDU is a permuted diagonal.

From the definition, it follows that whenever $[\mathbf{a}_0, \mathbf{a}_1, \ldots]$ is a set of monic block ROPs, then any other set of block ROPs can be obtained as $[\mathbf{a}_0, \mathbf{a}_1, \ldots]\,\mathrm{diag}(S_0, S_1, \ldots)$ where each S_k is unit upper triangular.

We shall however be liberal in the choice of the IOPs. When we say block orthogonal we mean that the polynomials in a block are of strict degree (monic) and orthogonal to the polynomials in the previous blocks where the beginning position and the size of each block is dictated by the rank structure of M through the indices ν_n (the block indices) and α_{n+1} (the block sizes).

2.3 Schur complements

The two-sided block Gram-Schmidt method produces the block biorthogonal polynomials, that is, it generates the unit upper triangular matrices A and B which satisfy $B^H M A = D$. Therefore it computes block UDL factorizations of M_n^{-1} for all $n = 0, 1, \ldots, N-1$. We can however also write $M = B^{-H} D A^{-1}$. Defining the unit upper triangular matrices $V = B^{-1}$ and $U = A^{-1}$, we also get

$$M = B^{-H} D A^{-1} = V^H D U. \tag{2.5}$$

The latter factorization gives a block LDU decomposition of M itself. Define

$$[\mathbf{v}_0(w)^H, \mathbf{v}_1(w)^H, \ldots] = \mathbf{w}^H V^H \quad \text{and} \quad [\mathbf{u}_0(z)^H, \mathbf{u}_1(z)^H, \ldots] = \mathbf{z}^H U^H$$

where $\mathbf{v}_i$ and $\mathbf{u}_i$ are series of size $\alpha_{i+1} \times 1$, then the matrix relation $M = V^H D U$ corresponds to the dyadic decomposition

$$M(w, z) = \sum_{i=0}^{N} \mathbf{v}_i(w)^H D_{ii} \mathbf{u}_i(z). \tag{2.6}$$

This decomposition is related to a Schur type algorithm and Schur complements. Consider a matrix

$$M = \begin{bmatrix} M_0 & M_{01} \\ M_{10} & M_{11} \end{bmatrix} \tag{2.7}$$

with M_0 invertible, then the Schur complement of M_0 in M is defined as the matrix $M_{11} - M_{10}M_0^{-1}M_{01}$. Following Parlett [30], we denote it by $M_{\langle 0\rangle}$.

Taking the previous LDU decomposition which is

$$M(w,z) = \mathbf{v}_0(w)^H D_{00}\mathbf{u}_0(z) + M'(w,z), \qquad M'(w,z) = \sum_{i=1}^{N} \mathbf{v}_i(w)^H D_{ii}\mathbf{u}_i(z).$$

We shall show that $(\overline{w}z)^{-\alpha_1}M'(w,z)$ is the generator $M_{\langle 0\rangle}(w,z)$ for the Schur complement $M_{\langle 0\rangle}$.

Theorem 2.1 *Let M be as in (2.7) with factorization (2.5), or equivalently (2.6). M_0 is invertible of size α_1. Define*

$$M_{\langle 0\rangle}(w,z) = (\overline{w}z)^{-\alpha_1}\left[M(w,z) - \mathbf{v}_0(w)^H D_{00}\mathbf{u}_0(z)\right], \tag{2.8}$$

then this is the generator for $M_{\langle 0\rangle}$, the Schur complement of M_0 in M.

Proof. We recall the factorization $M_0 = V_0^H D_{00} U_0$. Also

$$\mathbf{v}_0(w)^H = \mathbf{w}^H \begin{bmatrix} V_0^H \\ V_{10}^H \end{bmatrix} \quad \text{and} \quad \mathbf{u}_0(z) = \begin{bmatrix} U_0 & U_{01} \end{bmatrix}\mathbf{z}$$

where the associated matrices are the first α_1 columns of V^H and the first α_1 rows of U respectively, so that

$$\begin{bmatrix} V_0^H \\ V_{10}^H \end{bmatrix} D_{00}U_0 = \begin{bmatrix} M_0 \\ M_{10} \end{bmatrix} \quad \text{and} \quad V_0^H D_{00}\begin{bmatrix} U_0 & U_{01} \end{bmatrix} = \begin{bmatrix} M_0 & M_{01} \end{bmatrix}.$$

Therefore the matrix associated with the generator (2.8) is

$$M_{11} - V_{10}^H D_{00}U_{01} = M_{11} - V_{10}^H D_{00}U_0\left(U_0^{-1}D_{00}^{-1}V_0^{-H}\right)V_0^H D_{00}U_{01} = M_{11} - M_{10}M_0^{-1}M_{01}$$

and this is indeed the Schur complement $M_{\langle 0\rangle}$. □

This process can be repeated. Define $M^0 = M$ and recursively $M^{k+1} = M^k_{\langle 0\rangle}$, $k = 0,1,\ldots$ to be the Schur complement in M^k of the smallest invertible leading submatrix M_0^k of M^k. This corresponds to the recursion

$$\begin{aligned} M^{k+1}(w,z) &= (\overline{w}z)^{-\alpha_{k+1}}\left(M^k(w,z) - (\overline{w}z)^{-\nu_k}\mathbf{v}_k(w)^H D_{kk}\mathbf{u}_k(z)\right) \\ &= (\overline{w}z)^{-\alpha_{k+1}}\left(M^k(w,z) - \mathbf{h}_k(w)^H[M_0^k]^{-1}\mathbf{g}_k(z)\right) \end{aligned} \tag{2.9}$$

where M_0^k is of size α_{k+1}, $\mathbf{h}_k(w)^H$ is the generator for the first α_{k+1} columns of M^k and $\mathbf{g}_k(z)$ is the generator for the first α_{k+1} rows of M^k. By computing these successive Schur complements, we obtain a dyadic decomposition

$$M(w,z) = \sum_{i=0}^{N}(\overline{w}z)^{\nu_k}\mathbf{h}_i(w)^H[M_0^k]^{-1}\mathbf{g}_k(z), \tag{2.10}$$

which corresponds to a block LDU factorization of M. It is just a rescaled version of (2.6) and (2.5). In the case of an Hermitian matrix, this corresponds to the Jacobi transformation for diagonalizing an Hermitian form. The result is obtained by successively performing the steps (1 – 3) for $k = 0, 1, \ldots$: (1) Find M_0^k, the smallest leading submatrix in M^k and let its size be α_{k+1}, (2) Set $\mathbf{h}_k^H$ equal to the first α_{k+1} columns of M^k and $\mathbf{g}_k$ equal to the first α_{k+1} rows of M^k, (3) Compute $M^{k+1} = M_{\langle 0\rangle}^k$ by forming $M^k - \mathbf{h}_k^H [M_0^k]^{-1}\mathbf{g}_k$ and leaving out the first α_{k+1} rows and columns (which are zero).

This is also the basis of a general Schur type algorithm which performs the previous Jacobi transform for the M^k stored in factored form. This algorithm was first developed by Schur [32] in the context of positive definite Toeplitz matrices, but has been generalized to matrices with low displacement rank (see e.g., [27, 12]). For quasi Toeplitz matrices, $(1 - \overline{w}z)M(w, z)$ is factored as $G(w)^H H(z)$ with $G(w)$ and $H(z)$ of size $\delta \times 1$ where δ is the displacement rank of M. It is then a simple matter to update the factors G_k and H_k in the factorizations $(1 - \overline{w}z)M^k(w, z) = G_k(w)^H H_k(z)$ for $k = 1, 2, \ldots$

We now give the generalization of this idea for an arbitrary moment matrix. (See also [31].) We first recall the following lemma.

Lemma 2.2 *Let M be as in (2.6) and $M_{\langle 0\rangle}$ the Schur complement of M_0, then* $\det M = \det M_0 \det M_{\langle 0\rangle}$.

We shall describe the first step of the algorithm, the subsequent steps are similar. Let M_0 be the smallest (size α_1) leading submatrix of M. Suppose that we have a factorization for M of the form (see also the section on Krylov subspace iteration where such a factorization appears naturally)

$$M = H^H G \quad \text{with} \quad H = \begin{bmatrix} H_{00} & H_{01} \\ H_{10} & H_{11} \end{bmatrix}, \quad G = \begin{bmatrix} G_{00} & G_{01} \\ G_{10} & G_{11} \end{bmatrix},$$

where G_{00} and H_{00} are of size α_1. By row permutations on G (and corresponding column permutations on H^H) we can always assume that G_{00} is nonsingular. This implies that also H_{00} will be nonsingular. This can be seen as follows. By elementary row operations Q, we can bring G into the form

$$QG = \begin{bmatrix} G_{00} & G_{01} \\ 0 & \hat{G}_{11} \end{bmatrix} \quad \text{while} \quad H^H Q^{-1} = \begin{bmatrix} H_{00}^H & \hat{H}_{10}^H \\ H_{01}^H & \hat{H}_{11}^H \end{bmatrix}.$$

The left top block of M is M_0, which is supposed to be regular, thus

$$M_0 = \left[H^H G\right]_{00} = \left[H^H Q^{-1} Q G\right]_{00} = H_{00}^H G_{00}.$$

Since both M_0 and G_{00} are nonsingular, also H_{00} is nonsingular. Thus we can define

$$\beta = -G_{10} G_{00}^{-1} \quad \text{and} \quad \alpha = -H_{10} H_{00}^{-1}.$$

We remark that $S = I + \alpha^H \beta$ and $T = I + \beta\alpha^H$ are both regular matrices because

$$S = I + \alpha^H \beta = I + H_{00}^{-H} H_{10}^H G_{10} G_{00}^{-1} = H_{00}^{-H} M_0 G_{00}^{-1},$$

which proves that S is regular, but then also T is regular since by Lemma 2.2

$$\det \begin{bmatrix} I & -\alpha^H \\ \beta & I \end{bmatrix} = \det(I + \beta\alpha^H) = \pm \det(I + \alpha^H \beta) = \pm \det \begin{bmatrix} I & \beta \\ -\alpha^H & I \end{bmatrix}.$$

Then define

$$\theta_G = \begin{bmatrix} S_L^{-1} & 0 \\ 0 & T_L^{-1} \end{bmatrix} \begin{bmatrix} I & -\alpha^H \\ \beta & I \end{bmatrix} \quad \text{and} \quad \theta_H = \begin{bmatrix} I & \alpha^H \\ -\beta & I \end{bmatrix} \begin{bmatrix} S_R^{-1} & 0 \\ 0 & T_R^{-1} \end{bmatrix}$$

where $S_L S_R = S$ and $T_L T_R = T$ are factorizations of S and T. These definitions imply that

$$\theta_H \theta_G = \begin{bmatrix} I & \alpha^H \\ -\beta & I \end{bmatrix} \begin{bmatrix} S^{-1} & 0 \\ 0 & T^{-1} \end{bmatrix} \begin{bmatrix} I & -\alpha^H \\ \beta & I \end{bmatrix} = I.$$

Moreover, the definition of α and β was such that $\theta_G G$ and $H^H \theta_H$ have the form

$$\tilde{G} = \theta_G G = \begin{bmatrix} \tilde{G}_{00} & \tilde{G}_{01} \\ 0 & \tilde{G}_{11} \end{bmatrix} \quad \text{and} \quad \tilde{H}^H = H^H \theta_H = \begin{bmatrix} \tilde{H}_{00} & \tilde{H}_{01} \\ 0 & \tilde{H}_{11} \end{bmatrix}^H.$$

The matrices α and β generalize the reflection coefficient for the first step in the classical Schur algorithm for positive definite Toeplitz matrices. We can easily make the connection with formula (2.9) by taking the generating functions:

$$M(w,z) = H(w)^H G(z) = H(w)^H \theta_H \theta_G G(z) = \tilde{H}(w)^H \tilde{G}(z).$$

Setting

$$\tilde{G}(z) = \tilde{G}\mathbf{z} = \begin{bmatrix} \tilde{G}_{00} & \tilde{G}_{01} \\ 0 & \tilde{G}_{11} \end{bmatrix} \mathbf{z} = \begin{bmatrix} \tilde{H}_{00}^{-H} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{g}_0(z) \\ z^{\alpha_1} G_1(z) \end{bmatrix}$$

and similarly for $\tilde{H}(w)$, it follows that

$$\begin{aligned} M(w,z) &= \tilde{H}(w)^H \tilde{G}(z) = \begin{bmatrix} \mathbf{h}_0(w) \\ w^{\alpha_1} H_1(w) \end{bmatrix}^H \begin{bmatrix} M_0^{-1} & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{g}_0(z) \\ z^{\alpha_1} G_1(z) \end{bmatrix} \\ &= \mathbf{h}_0(w)^H M_0^{-1} \mathbf{g}_0(z) + (\overline{w} z)^{\alpha_1} H_1(w)^H G_1(z) \end{aligned}$$

and this can be rewritten in the form

$$H_1(w)^H G_1(z) = (\overline{w} z)^{-\alpha_1} [M(w,z) - \mathbf{h}_0(w)^H M_0^{-1} \mathbf{g}_0(z)]$$

which shows that $H_1(w)^H G_1(z)$ is a factorization for $M_{\langle 0 \rangle}(w,z)$ (see (2.9)), and the algorithm can proceed to the next step by following the same procedure. The Schur algorithm thus updates the successive Schur complements in factored form and at the same time one gets the block columns and rows (generated by $\mathbf{h}_0(w)$ and $\mathbf{g}_0(z)$) of a block LDU factorization of M. This procedure will break down when the maximal rank of M has been reached. There is no more regular leading submatrix of which to take the Schur complement.

2.4 Fourier series and reproducing kernels

The formal Fourier expansions with respect to a biorthogonal polynomial system with left/right blocks $\{\mathbf{b}_k, \mathbf{a}_k\}$, $k = 0, 1, \ldots$, can be given as

$$a(z) = \sum_{k=0}^{N} \mathbf{a}_k(z) c_k, \text{ with } c_k = D_{kk}^{-1} \langle \mathbf{b}_k, a \rangle.$$

Let M_n be a regular submatrix of M of size ν_{n+1} and suppose $K_n = M_n^{-T} = [m_{ij}]$. Then define the *kernel function* $K_n(y,x)$ to be the generator for K_n, i.e.,

$$K_n(y,x) = \mathbf{y}^H K_n \mathbf{x} = \sum_{i,j=0}^{\nu} \overline{y}^i m_{ij} x^j, \quad \nu = \nu_{n+1} - 1.$$

For this kernel, the following properties hold.

Theorem 2.3 *Let M be a moment matrix with some block index ν_{n+1} and M_n the corresponding regular submatrix. Let $K_n(y,x)$ be the kernel as defined above. Then with $\nu = \nu_{n+1} - 1$*

$$K_n(y,x) = \mathbf{y}^H \overline{B}_n D_n^{-T} A_n^T \mathbf{x} = \sum_{i=0}^{n} \mathbf{a}_i(x) D_{ii}^{-1} \mathbf{b}_i(y)^H \tag{2.11}$$

$$= -\frac{1}{\det M_n} \det \begin{bmatrix} & M_n & & & \begin{matrix} 1 \\ \overline{y} \\ \vdots \\ \overline{y}^\nu \end{matrix} \\ 1 & x & \cdots & x^\nu & 0 \end{bmatrix} \tag{2.12}$$

where $\{\mathbf{b}_k, \mathbf{a}_k\}$ denotes the kth blocks of left/right orthogonal polynomials and as before $\langle \mathbf{b}_i, \mathbf{a}_j \rangle = \delta_{ij} D_{ij}$.

Moreover, for any polynomial $a(x)$ of degree at most $\nu = \nu_{n+1} - 1$, the kernel $K_n(y,x)$ is reproducing:

$$\langle a(x), K_n(y,x) \rangle = \overline{a(y)}.$$

Proof. For (2.12), use the definition

$$K_n(y,x) = [1, x, \ldots, x^\nu] M_n^{-1} [1, y, \ldots, y^\nu]^H \tag{2.13}$$

and Lemma 2.2 to find that the determinant of the bordered matrix in (2.12) equals

$$\det M_n \det(0 - [1 \ldots x^\nu] M_n^{-1} [1 \ldots y^\nu]^H)$$

which proves the first result.

If the coefficient vectors for the ROPs in the blocks $\mathrm{A}_0, \mathrm{A}_1, \ldots, \mathrm{A}_n$ are collected in the unit upper triangular matrix A_n, (and similarly for the LOPs) then we know that (2.3) holds. Filling this into (2.13) gives (2.11).

For the reproducing property note that

$$\langle a(x), K_n(y,x) \rangle = \sum_{i=0}^{n} \langle a(x), \mathbf{a}_i(x) \rangle D_{ii}^{-1} \mathbf{b}_i(y)^H$$

which is the (formal) Fourier expansion of $\overline{a(y)}$ because if we set $a(y) = \sum_{i=0}^{n} \mathbf{b}_i(y) d_i$ with $d_i \in \mathbb{C}^{\alpha_{i+1} \times 1}$, then we find from the orthogonality that

$$d_i^H = \langle a(x), \mathbf{a}_i(x) \rangle D_{ii}^{-1},$$

which proves the theorem. □

Note the duality between (2.6) and (2.11). The first gives a dyadic decomposition of M and is in fact a (block) LDU decomposition of M which is related to unsymmetric Choleski and Schur algorithms. The second is a dyadic decomposition of M^{-1} and gives a so called inverse LDU decomposition of M (a UDL decomposition of M^{-1}). This is related to Gram-Schmidt type of algorithms, reproducing kernels, biorthogonal polynomials and, if there is some structure in M, it will also result in a Christoffel-Darboux type relation. See for example [27] for the Hermitian positive definite case of matrices with low displacement rank.

2.5 The Hessenberg matrix

The generalization of the Jacobi matrix for classical orthogonal polynomials is an upper Hessenberg matrix when the moment matrix has no special structure. Obviously, since all polynomials $a_k(z)$ in $\mathbf{a}(z) = [a_0(z), a_1(z), \ldots]$ are monic and of strict degree k, there should exist a unit upper Hessenberg matrix T such that

$$z\mathbf{a}(z) = \mathbf{a}(z)T.$$

To write down a matrix equivalent of this, define the unit upper triangular matrix A by $\mathbf{a}(z) = \mathbf{z}^T A$, and use Z for the downshift matrix, then the previous relation becomes

$$ZA = AT.$$

To give a finite dimensional analog and to show the block structure of T, we remark that T could be the result of a block two-sided Gram-Schmidt procedure. The latter transforms the set $1, z, z^2, \ldots$ into a set of block left/right orthogonal polynomials by the following procedure. Start with $a_0 = 1$ and then generate

$$a_{\nu+1}(z) = za_\nu(z) - \sum_{i=0}^{k-1} \mathbf{a}_i(z) \left\langle \mathbf{b}_i(\zeta), \mathbf{a}_i(\zeta) \right\rangle^{-1} \left\langle \mathbf{b}_i(\zeta), \zeta a_\nu(\zeta) \right\rangle$$

for the ROPs in block k, i.e. for $\nu_k \leq \nu + 1 < \nu_{k+1}$. (A similar relation holds for the LOPs). If we place all the generated polynomials of block k in the vector $\mathbf{a}_k = [a_{\nu_k}, \ldots, a_{\nu_{k+1}-1}]$ then we know that the block is complete as soon as the matrix $\langle \mathbf{b}_k, \mathbf{a}_k \rangle$ is regular. This defines the block size α_{k+1} and together with the previous ones also the block index $\nu_{k+1} = \sum_{i=1}^{k+1} \alpha_i$. For $n > k$ this can be summarized in the relation

$$z\mathbf{a}_k(z) = [\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_n][T_{0,k}^T \ldots T_{k,k}^T, T_{k+1,k}^T, 0 \ldots 0]^T. \tag{2.14}$$

The block $T_{i,k}$ is a matrix of size $\alpha_{i+1} \times \alpha_{k+1}$, the diagonal block is square and has the form of a companion matrix. Its last column is equal to $\langle \mathbf{b}_k, \mathbf{a}_k \rangle^{-1} \langle \mathbf{b}_k, za_{\nu_{k+1}-1} \rangle$. The subdiagonal block $T_{k+1,k}$ is zero except for its right top element which is 1. This last column is special because it relates the first polynomial a_{k+1}^1 of the next block with the previous blocks $\mathbf{a}_0, \ldots, \mathbf{a}_k$. Writing this out for $k = 0, \ldots, n$ and translating this to the matrices of coefficient vectors we get

$$F_n A_n = A_n T_n \tag{2.15}$$

where as before $A_n = A[\nu_{n+1}, \nu_{n+1}]$ is unit upper triangular, $F_n = F(a^1_{n+1})$ is the companion matrix for the TOP a^1_{n+1} of block $n+1$, and $T_n = T[\nu_{n+1}, \nu_{n+1}]$ is unit upper Hessenberg.

$$T_n = \begin{bmatrix} T_{0,0} & T_{0,1} & T_{0,2} & \cdots & T_{0,n} \\ T_{1,0} & T_{1,1} & T_{1,2} & & T_{1,n} \\ & T_{2,1} & T_{2,2} & \ddots & T_{2,n} \\ & & \ddots & \ddots & \vdots \\ & & & & T_{n-1,n} \\ & & & T_{n,n-1} & T_{n,n} \end{bmatrix}.$$

This is the finite analog of the relation $ZA = AT$. Note that this truncation changes Z into a Frobenius matrix. The infinite matrix T is the generalization of the Jacobi matrix (here written for monic polynomials). In the classical case it is tridiagonal, and as we shall see in the next section, it is block tridiagonal for a general Hankel moment matrix, reducing to a scalar tridiagonal when it is strongly regular.

Recall the definition $M^< = MZ$ and the relations $B^H MA = D$ and $ZA = AT$, so that we get immediately the following relations, defining J

$$B^H M^< A = B^H MZA = B^H MAT = DT = J.$$

Note that D was block diagonal and T unit upper Hessenberg, so that J shall be block upper Hessenberg.

For the truncation of this relation to submatrices of size ν_{n+1}, we should note the following. The orthogonality of the TOP a^1_{n+1} implies (2.4), and hence

$$M_n F_n = M_n^<$$

where $M_n^< = M^<[\nu_{n+1}, \nu_{n+1}]$. Thus we have (the index n means that we take the leading submatrices of size ν_{n+1}) $M_n^< = M_n F_n$, $F_n A_n = A_n T_n$ and $B_n^H M_n A_n = D_n$ and we get

$$B_n^H M_n^< A_n = B_n^H M_n F_n A_n = B_n^H M_n A_n T_n = D_n T_n = J_n. \tag{2.16}$$

As a direct consequence, we find the following determinant expressions for a^1_{n+1}.

Theorem 2.4 *The TOP a^1_{n+1} of block $n+1$ is given by the following determinant expressions*

$$\begin{aligned} a^1_{n+1}(z) &= \det(zI_n - F_n) \\ &= \det M_n^{-1} \det(zM_n - M_n^<) \\ &= \det(zI_n - T_n) \\ &= \det M_n^{-1} \det \begin{bmatrix} & & \mu_{0,\nu+1} \\ M_n & & \vdots \\ & & \mu_{\nu,\nu+1} \\ 1 \; z \; \cdots & & z^{\nu+1} \end{bmatrix}, \quad \nu + 1 = \nu_{n+1} \end{aligned}$$

where we have used the notation that was introduced above.

Proof. The first relation holds by definition of companion matrix. The second one follows by multiplying from the left with M_n and using $M_n^< = M_n F_n$. This result, multiplied from the left with B_n^H and from the right with A_n gives the next line because $\det B_n = \det A_n = 1$ and $zB_n^H M_n A_n - B_n^H M_n^< A_n = zD_n - D_n T_n$ while $\det M_n = \det D_n$. For the last line, note that the determinant on the right equals

$$\det M_n (z^{\nu+1} - [1\ z\ \cdots\ z^\nu] M_n^{-1} [\mu_{0,\nu+1} \cdots \mu_{\nu,\nu+1}]^T).$$

The rest follows from (2.4). □

2.6 Padé approximation

We now give some general rational approximations that can be obtained from the elements we have collected so far.

Let $M^{[i,*]}(z)$ be the generator for row i, and define $g^{[i]}(z) = z^{-1} M^{[i,*]}(z^{-1})$, i.e.,

$$g^{[i]}(z) = \sum_{k=0}^{\infty} \frac{\mu_{i,k}}{z^{k+1}} = \left\langle \zeta^i, (z-\zeta)^{-1} \right\rangle, \quad i = 0, 1, 2, \ldots \tag{2.17}$$

Furthermore, using the notation introduced above, define

$$\begin{aligned} g_n^{[i]}(z) &= e_i^T M_n (z M_n - M_n^<)^{-1} M_n e_0 \\ &= \sum_{k=0}^{\infty} \frac{\mu_{i,k}^n}{z^{k+1}}, \quad i = 0, 1, 2, \ldots, \nu_{n+1} - 1 \end{aligned} \tag{2.18}$$

where e_i is the ith unit vector. Then $\mu_{i,k} = \mu_{i,k}^n$ for $k = 0, 1, \ldots, \nu_{n+1}$. We formulate this as a theorem.

Theorem 2.5 *Let $g^{[i]}$ be defined by (2.17) and the rational function $g_n^{[i]}$ by (2.18). With the notation introduced before, the latter can also be expressed as*

$$g_n^{[i]}(z) = e_i^T B_n^{-H} D_n (zI - T_n)^{-1} e_0 \tag{2.19}$$

or in the rational form

$$g_n^{[i]}(z) = c_{n+1}^{[i]}(z) / a_{n+1}^1(z) \tag{2.20}$$

with the numerator, a polynomial of the second kind, given by

$$c_{n+1}^{[i]}(z) = \left\langle \zeta^i, \frac{a_{n+1}^1(z) - a_{n+1}^1(\zeta)}{z - \zeta} \right\rangle. \tag{2.21}$$

It has the determinant expression

$$c_{n+1}^{[i]}(z) = \det M_n^{-1} \det \begin{bmatrix} M_n e_0 & M_n^< \\ 0 & u_n \end{bmatrix} \tag{2.22}$$

with u_n the polynomial part of

$$[z\ z^2\ \ldots\ z^{\nu_{n+1}}] g^{[i]}(z).$$

Moreover, for $i = 0, 1, \ldots, \nu_{n+1} - 1$, $g^{[i]}$ and $g_n^{[i]}$ correspond in their initial terms:

$$\mu_{i,k} = \mu_{i,k}^n \text{ for } k = 0, 1, \ldots, \nu_{n+1}.$$

Thus $c_{n+1}^{[i]}(z)$ is the polynomial part of $g^{[i]}(z) a_{n+1}^1(z)$.

Proof. The alternative expression (2.19) for $g_n^{[i]}$ can be easily found from (2.18) when using (2.3) and (2.16) and the fact that $A_n^{-1}e_0 = e_0$.

To prove the correspondence of the expansions of $g^{[i]}$ and $g_n^{[i]}$, we can start from (2.19) and use a formal expansion of $(zI - T_n)^{-1}$.

$$g_n^{[i]}(z) = e_i^T B_n^{-H} D_n (\sum_{k=0}^{\infty} \frac{T_n^k}{z^{k+1}}) e_0.$$

Because $A_n^{-1} F_n A_n = T_n$, as we can see from (2.15), we also have

$$T_n^k e_0 = A_n^{-1} F_n^k A_n e_0 = A_n^{-1} F_n^k e_0$$

because A_n is unit upper triangular. Thus

$$\begin{aligned} g_n^{[i]}(z) &= e_i^T B_n^{-H} D_n A_n^{-1} \sum_{k=0}^{\infty} z^{-(k+1)} F_n^k e_0 \\ &= e_i^T M_n \sum_{k=0}^{\infty} z^{-(k+1)} F_n^k e_0 \qquad (2.23) \\ &= [\mu_{i,0}, \ldots, \mu_{i,\nu}] \sum_{k=0}^{\infty} z^{-(k+1)} F_n^k e_0 \end{aligned}$$

where $\nu = \nu_{n+1} - 1$. It is easily verified that

$$F_n^k e_0 = \begin{cases} e_k & \text{for } k = 0, \ldots, \nu = \nu_{n+1} - 1 \\ -\dot{\mathrm{A}}_{n+1}^1 & \text{for } k = \nu_{n+1} \end{cases}$$

This proves that

$$g_n^{[i]}(z) = \sum_{k=0}^{\nu} z^{-(k+1)} \mu_{i,k} - e_i^T M_n \dot{\mathrm{A}}_{n+1}^1 z^{-(\nu+2)} + \cdots$$

Since also

$$e_i^T M_n \dot{\mathrm{A}}_{n+1}^1 = -\mu_{i,\nu+1},$$

we have proved the correspondence.

The expression (2.21) follows from the formal equalities

$$\left\langle \zeta^i, \frac{a_{n+1}^1(z) - a_{n+1}^1(\zeta)}{z - \zeta} \right\rangle = g^{[i]}(z) a_{n+1}^1(z) - \left\langle \zeta^i, \frac{a_{n+1}^1(\zeta)}{z - \zeta} \right\rangle \qquad (2.24)$$

where we have used the expression (2.17) for $g^{[i]}$. Expanding $(z-\zeta)^{-1}$ as $\sum_0^\infty \zeta^k z^{-(k+1)}$, one sees that the second term is of the form $\sum_{k=0}^\infty \eta_k z^{-(k+1)}$ with η_k the coefficient of $z^{-(k+1)}$ in the expansion of $g^{[i]} a_{n+1}^1$. Therefore the left hand side of (2.24) is polynomial. Since a_{n+1}^1 has degree ν_{n+1} and $\mu_{ij} = \mu_{ij}^n$ for $j = 0, \ldots, \nu_{n+1}$, it follows that the polynomial part of $g_n^{[i]} a_{n+1}^1$ (which is $c_{n+1}^{[i]}$ by definition) and of $g^{[i]} a_{n+1}^1$ is the same. This proves (2.21).

The determinant expression (2.22) for the numerator can be found as follows. First note that

$$\det \begin{bmatrix} zM_n - M_n^< & M_n e_0 \\ e_i^T M_n & 0 \end{bmatrix} = \det(zM_n - M_n^<)\Big(-e_i^T M_n (zM_n - M_n^<)^{-1} M_n e_0\Big)$$

The first factor equals $a^1_{n+1} \det M_n$ and the second one equals $-g^{[i]}_n$. Hence the numerator $c^{[i]}_{n+1}$ is given by

$$c^{[i]}_{n+1}(z) = -\det M_n^{-1} \begin{bmatrix} zM_n - M_n^< & M_n e_0 \\ e_i^T M_n & 0 \end{bmatrix}.$$

This gives the desired result after some elementary matrix operations that do not change the determinant. □

The previous theorem shows a kind of vector Padé approximation. Indeed, when we consider the vector with components $g^{[i]}(z)$, $i = 0, \ldots, \nu - 1$ where $\nu = \nu_{n+1}$ and the vector of rational approximants $g^{[i]}_n(z)$, then, since $\deg c^{[i]}_{n+1} < \deg a^1_{n+1}$, the number of parameters in the latter is ν^2 (for the numerator coefficients) + ν (for the monic denominator), giving a total of $(\nu + 1)\nu$ parameters, which is precisely the number of coefficients that is fitted in the ν series. The block two-sided Gram-Schmidt procedure computes the denominators of the approximants for increasing n. This generates approximants of increasing degree ν_n, but also the number of series and the number of approximants is increasing! However, it is possible to keep this number constant when we choose a special moment matrix. We shall not go into the details here, but the idea is that if we want to have d different series to fit where d is independent of n, then one should consider a moment matrix which satisfies the relation $\mu_{i,j+1} = \mu_{i+d,j}$ for all $i, j \in \mathbb{N}$. This means that M is a block Hankel matrix with blocks of size $d \times 1$. An extreme example is the case $d = 1$, i.e. the moment matrix is Hankel, and then only one series is fitted, the other rows being shifted versions. This is also the idea behind the so called vector orthogonal polynomials of size d as considered in [37]. The qualification "vector" does not refer to the polynomials, since they are scalar, but to the fact that they are orthogonal with respect to several (i.e., a vector of) linear functionals, viz., each of the d series is a generator for the moments of a functional. This is in contrast with the classical Hankel case, where there is only one linear functional. An alternative way of describing the previous pattern in the moment matrix is to apply row permutations, so that we can think of the moment matrix as consisting of horizontal bands with a Hankel structure. Then the number of different bands corresponds to the number of series that are fitted. The number of rows in each band is related to the number of extra coefficients that are fitted (on top of the number that is always fitted in the general case) for the series that is associated with that band. The latter approach is more flexible because it allows to make bands with a different number of rows, hence allowing a different order of approximation per series. However, when one wants to compute the approximants recursively with an increasing order of approximation for the different series, then the rearrangement as in the vector orthogonal polynomials is more appropriate. It is possible to obtain a different order of approximation per series by including some irregularities in the d-periodicity of the moments.

If row i and row $i+1$ are two rows in the same Hankel band, then $e_{i+1}M = e_i M^<$ and one can use $M_n F_n = M_n^<$ in the formula (2.23) to show that $g^{[i]}_n$ fits an extra coefficient of the series $g^{[i]}$. One can also consult the section on the Hankel case below.

2.7 Krylov subspace iteration

We can make the following connection with Krylov subspace methods.

Consider the Krylov matrices

$$X = [x_0|x_1|\cdots], \quad x_k = P^k x_0 \quad \text{and} \quad Y = [y_0|y_1|\cdots], \quad y_k = [Q^H]^k y_0$$

where the x_0 and y_0 are arbitrary vectors of the vector space $\mathcal{V} = \mathbb{C}^p$ (with p finite or infinite) and P and Q some matrices.

The column spaces of the matrices X and Y are Krylov spaces. In general one denotes a Krylov subspace of order ν as

$$\mathcal{K}_\nu(x, P) = \text{span}\{x_k = P^k x : k = 0, \ldots, \nu - 1\}.$$

Only when the x_k are all linearly independent, the dimension of $\mathcal{K}_\nu(x, P)$ will be ν.

We set $M = Y^H X = [\mu_{ij}]$ where $\mu_{ij} = y_0^H Q^i P^j x_0$. Associate with M the quantities we introduced in the previous sections, then $\hat{X} = XA$ and $\hat{Y} = YB$ are matrices whose columns form a set of block bi-orthogonal vectors since $\hat{Y}^H \hat{X} = B^H M A = D$ while

$$\hat{Y}^H P \hat{X} = B^H M^< A = DT = J.$$

In the special case that $P = Q$, M will be a Hankel matrix and when $P = Q^{-1}$, M will be Toeplitz but it is an arbitrary matrix in general. Note also that the generator allows a natural factorization (recall the Schur algorithm of Section 2.3)

$$M(w, z) = Y(w)^H X(z) = y_0^H (I - wQ^H)^{-H} (I - zP)^{-1} x_0,$$

with

$$X(z) = X\mathbf{z} = (I - zP)^{-1} x_0, \quad Y(w) = Y\mathbf{w} = (I - wQ^H)^{-1} y_0.$$

Thus the generator for M is rational and of finite degree in z and w when the vector space $\mathcal{V}$ is finite dimensional.

Let $M_n = M[\nu_{n+1}, \nu_{n+1}]$ be regular, and denote the first ν_{n+1} columns of X and Y by X_n and Y_n with column spaces $\mathcal{X}_n = \mathcal{K}_{\nu_{n+1}}(x_0, P)$ and $\mathcal{Y}_n = \mathcal{K}_{\nu_{n+1}}(y_0, Q^H)$ respectively. It is well known that then there exists a unique oblique projection onto $\mathcal{X}_n$ along $\mathcal{Y}_n$ whose matrix representation is

$$\Pi_n = \hat{X}_n (\hat{Y}_n^H \hat{X}_n)^{-1} \hat{Y}_n^H.$$

Suppose we want to solve the square system $Px = b$, but the size of P is very large. One could then try to approximate it by "projecting" onto the lower dimensional space $\mathcal{X}_n$. That means we confine x to $\mathcal{X}_n$ by setting $x \approx z_n = \hat{X}_n v_n$ and we only satisfy the projected equations $\Pi_n P z_n = \Pi_n P \hat{X}_n v_n = \Pi_n b$. Since

$$\begin{aligned} \Pi_n P \hat{X}_n &= \hat{X}_n (\hat{Y}_n^H \hat{X}_n)^{-1} \hat{Y}_n^H P \hat{X}_n \\ &= \hat{X}_n D_n^{-1} D_n T_n = \hat{X}_n T_n \end{aligned}$$

we can solve the "small" system $T_n v_n = c_n$ with $c_n = D_n^{-1} \hat{Y}_n^H b$ and then take $z_n = \hat{X}_n v_n$ as an approximation for the true solution of the "large" system $Px = b$. This is the idea behind the Krylov subspace methods for the iterative solution of linear systems. By setting $b = \lambda x$, we can see that the solutions for the small eigenvalue problem $T_n v_n = \lambda v_n$ will give approximate solutions for the large one $Px = \lambda x$ and this is a justification that the same Krylov subspace methods can be used as iterative methods for approximating the eigenvalues of P. The elements in the spectrum $\sigma(T_n)$ of T_n are called Ritz values and they will appoximate the eigenvalues in the spectrum $\sigma(P)$.

A symmetric argumentation will give similar results with respect to Q.

In principle, when the dimension of $\mathcal{V}$ is N, then $\mathcal{K}_N(x_0, P)$ and $\mathcal{K}_N(y_0, Q^H)$ will be equal to the whole vector space $\mathcal{V}$ on condition that the rank of X and Y is N. This of course will depend upon the choice of the initial vectors x_0 and y_0 and the Jordan structure of P and Q. When the rank of M is smaller than N, the Krylov methods have a breakdown and in general it is not guaranteed that the spectrum of P and/or of Q will be recovered.

Set $r_x = \operatorname{rank} X$, $r_y = \operatorname{rank} Y$ and $r = \operatorname{rank} M \le m = \min\{r_x, r_y\}$. Then the Krylov method will break down when $\nu_{n+1} = r$. At that point, the maximal regular leading submatrix of M is reached and one enters an infinite block of OPs and there is no next TOP. Moreover, the following properties can be easily proved.

Theorem 2.6 *With the notation just introduced, we have the following. When $r = m = r_x$, then it is guaranteed that $\sigma(T_n) \subset \sigma(P)$, that the range of X_n is P-invariant and if $\hat{v}$ is an eigenvector of T_n i.e., $T_n\hat{v} = \lambda\hat{v}$, then $v = \hat{X}_n\hat{v}$ is an eigenvector of P: $Pv = \lambda v$. The eigenvalues λ of T_n are the zeros of the polynomial a^1_{n+1}. Similar results hold for $r = m = r_y$.*

Proof. If $X_n = [x_0, \ldots, x_r]$, then $PX_n = [x_1, \ldots, x_{r+1}]$. Since X_n has reached its maximal rank $r = m = r_x$, x_{r+1} will be in the column space of X_n. This means that this column space is P-invariant. The column space of X_n is equal to the column space of $\hat{X}_n$. Thus also the column space of $\hat{X}_n$ will be P-invariant. Hence, there must exist some matrix $\hat{T}_n$ such that $P\hat{X}_n = \hat{X}_n\hat{T}_n$. In fact, this matrix $\hat{T}_n$ is just T_n because

$$D_nT_n = \hat{Y}_n^H P\hat{X}_n = \hat{Y}_n^H\hat{X}_n\hat{T}_n = D_n\hat{T}_n$$

and D_n is nonsingular. Thus, if $\hat{v}$ is an eigenvector for T_n, i.e., $T_n\hat{v} = \lambda\hat{v}$, then

$$\hat{X}_nT_n\hat{v} = P\hat{X}_n\hat{v} = \lambda\hat{X}_n\hat{v}.$$

Because $\hat{v} \neq 0$ and the columns of $\hat{X}_n$ are linearly independent, also $\hat{X}_n\hat{v} \neq 0$. Hence we may conclude that if λ is an eigenvalue of T_n with corresponding eigenvector $\hat{v}$, then λ is also an eigenvalue of P with corresponding eigenvector $\hat{X}_n\hat{v}$.

By Theorem 2.3, we have

$$\det(zI - T_n) = a^1_{n+1}(z) = \det(zI - F_n),$$

which means that the eigenvalues of T_n are the zeros of the TOP $a^1_{n+1}(z)$.

This proves the theorem. □

When $r < m$, however, then it may happen that we do not find any eigenvalue not of P and not of Q.

An example of the latter situation is the choice

$$P = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 5 \end{bmatrix}, \quad x_0 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 3 \end{bmatrix}, \quad y_0 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

with

$$\sigma(P) = \{5, 1 \pm \sqrt{2}\}, \quad \sigma(Q) = \{1, 2 \pm \sqrt{2}\}.$$

The rank of $M = Y^HX$ is 1, while the rank of X and Y is 2. The algorithm breaks down after step $\nu = 1$ and will give $\lambda = 3$ as an approximation of $\sigma(P)$ and $\lambda = 2$ as an approximation of $\sigma(Q)$. Both values are wrong.

3 Formal orthogonality: the Hankel case

When the moment matrix has a Hankel structure, the theory is well known, and this section does not contain new results. We only give a brief survey to tie up with the general case and proofs are only given when they are short and clarify the link with the previous section. We are somewhat more precise to establish a less known equivalence between the Schur algorithm for Hankel matrices and the Euclidean slgorithm. For other details, the reader who is interested should consult the literature. See for example [13, 21, 22, 23, 24, 30].

3.1 The moment matrix and its generator

Now we shall suppose that the moment matrix has a Hankel structure, i.e., $\mu_{ij} = \mu_{i+j}$. Because $\mu_{ij} = \mu_{i+j} = \langle z^i, z^j\rangle = \langle z^{i+j}, 1\rangle = \langle 1, z^{i+j}\rangle$, it is clear that we do not really need a bilinear form. In fact, we can reduce the definition of the bilinear form to the definition of a linear form μ on the set of polynomials defined by

$$\langle p, q\rangle = \mu(\overline{p}q), \quad \text{with} \quad \mu(z^k) = \mu_k, \quad k = 0, 1, \ldots$$

Thus it holds in general that

$$\langle zf(z), g(z)\rangle = \langle f(z), zg(z)\rangle, \tag{3.1}$$

which translates in the matrix relation

$$Z^T M = M^\wedge = M^< = MZ$$

where Z is the downshift operator and $M^<(= M^\wedge)$ is the left (= up) shifted moment matrix obtained by deleting the first column (or the first row) in M. On the other hand, we also have

$$ZM - MZ^T = \left[\begin{array}{c|c} 0 & -\mu_0 \quad -\mu_1 \quad \ldots \\ \hline \mu_0 & \\ \mu_1 & M^< - M^\wedge \\ \vdots & \end{array}\right] = \left[\begin{array}{c|c} 0 & -H \\ \hline H^T & 0 \end{array}\right]$$

If we introduce the bivariate generator $M(w, z) = \mathbf{w}^T M\mathbf{z}$ as in the general case, (note that for notational simplicity we assumed w real) then the previous matrix relation yields the remarkable property

$$M(w, z) = \frac{wH(w) - zH(z)}{w - z}, \quad \text{where} \quad H(z) = M(0, z) = H\mathbf{z} = \sum_{k=0}^{\infty} \mu_k z^k. \tag{3.2}$$

Thus the bilinear form being defined through a linear form, a Hankel matrix being characterized by its first row (or column) and the bivariate generator $M(w, z)$ being characterized by the univariate $H(z)$ are three instances of the same phenomenon.

The Hankel structure also implies symmetry. Thus an Hermitian Hankel, it is automatically real. Therefore, classically, real Hankel moment matrices appear in the context of real orthogonal polynomials on a real (possibly infinite) interval. Here however, we allow the moments and the polynomials to be complex. The symmetry of the moment matrix then implies that if $a_k(z) = \mathbf{z}^T \mathrm{A}_k$ is the kth ROP, then we can choose $b_k(z) = \overline{a}_k(z) = \mathbf{z}^T \overline{\mathrm{A}}_k$ to be the kth LOP. Thus we may replace the matrix B^H from the previous formulas by A^T.

3.2 Block biorthogonality and block Jacobi matrix

The recurrence relation is well known for the Hankel case [13]. We include its derivation for completeness. We first determine the block sizes as follows.

Theorem 3.1 *Suppose ν_n is some block index for the moment matrix M and a_n^1, the TOP of block n which is by definition orthogonal to z^i for $i = 0, \ldots, \nu$ where $\nu + 1 = \nu_n$. Now suppose it is over-orthogonal in the sense that also $\langle z^i, a_n^1 \rangle = 0$ for $i = \nu + 1, \nu + 2, \ldots, \nu + \alpha - 1$ but nonzero for $i = \nu + \alpha$, then the size of block n will be α.*

Proof. We may indeed choose

$$a_{\nu+i}(z) = z^{i-1} a_{\nu+1}(z) \quad \text{for} \quad i = 1, \ldots, \alpha \tag{3.3}$$

since it follows from the relation (3.1) and the over-orthogonality of a_n^1 that $a_{\nu+i}$ is orthogonal to all z^j for $j = 0, 1, \ldots, \nu + \alpha - i$, hence to all polynomials in the previous blocks as long as $1 \leq i \leq \alpha$. Collecting these polynomials in a block $\mathbf{a}_n = [a_{\nu+1} \ldots, a_{\nu+\alpha}]$, we find $\langle \overline{\mathbf{a}}_n, \mathbf{a}_n \rangle = D_{nn}$ which has a lower triangular Hankel structure with nonzero diagonal element. The Hankel structure follows from the choice (3.3) and the property (3.1). The diagonal element is nonzero since $\langle a_{\nu+\alpha}, a_{\nu+1} \rangle = \langle z^{\nu+\alpha}, a_{\nu+1} \rangle = \rho_n \neq 0$. Thus we cannot take less than α elements in a block because then D_{nn} would not be regular. In other words the block size of block n is $\alpha_{n+1} = \alpha$. □

As a consequence of the above construction we can see that the orthogonality of $z^{\alpha_{n+1}} a_n^1(z)$ to z^i is automatic for $i = 0, \ldots, \nu_n - 2$. Therefore we have

Theorem 3.2 *For block n, internal polynomials are obtained by just shifting the TOP: $a_n^{i+1}(z) = z^i a_n^1(z)$, $i = 1, \ldots, \alpha_{n+1} - 1$, and for computing the TOP of block $n + 1$, one can use*

$$a_{n+1}^1(z) = q_n(z) a_n^1(z) + p_n a_{n-1}^1(z) \tag{3.4}$$

where $q_n(z)$ is a monic polynomial of degree α_{n+1} and p_n is a nonzero constant.

Proof. The internal polynomials were considered in the previous proof. The recurrence for a_{n+1}^1 is obtained as follows. Set $\nu + 1 = \nu_n$. Note that $a'(z) = z^{\alpha_{n+1}} a_n^1(z)$ is orthogonal to z^k, $k = 0, 1, \ldots, \nu - 1$ and $\langle z^\nu, a' \rangle = \langle z^{\nu_{n+1}-1}, a_n^1 \rangle = \rho_n \neq 0$. Since by definition, also $\left\langle z^\nu, a_{n-1}^1 \right\rangle = \rho_{n-1} \neq 0$, it is obvious that $p_n = -\rho_n / \rho_{n-1} \neq 0$ will make $\tilde{a} = a' + p_n a_{n-1}^1$ orthogonal to z^k, $k = 0, \ldots, \nu$. Orthogonality to $z^{\nu+i}$ for $i = 1, \ldots, \alpha_{n+1}$ can be obtained by subtracting suitable multiples of $z^i a_n^1$ from $\tilde{a}$. □

This implies that the general relation (2.14) reduces to

$$z \mathbf{a}_k(z) = [\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_n][0 \ldots 0, T_{k-1,k}^T, T_{k,k}^T, T_{k+1,k}^T, 0 \ldots 0]^T \tag{3.5}$$

while the relation (2.15) becomes

$$F_n A_n = A_n T_n$$

where as before F_n is the companion matrix of a_{n+1}^1 and T_n has now a block tridiagonal form

$$T_n = \begin{bmatrix} T_{0,0} & T_{0,1} & & & \\ T_{1,0} & T_{1,1} & T_{1,2} & & \\ & T_{2,1} & T_{2,2} & \ddots & \\ & & \ddots & \ddots & T_{n-1,n} \\ & & & T_{n,n-1} & T_{n,n} \end{bmatrix}.$$

If M is strongly regular, all the blocks have size 1×1, and then this is an ordinary tridiagonal matrix which is known as the (nonsymmetric) Jacobi matrix which reflects the three term recurrence relation. In this regular case each block of A_n consists of one polynomial $\mathbf{a}_k(z) = a_k(z) = a_k^1(z)$ and $a_{n+1}^1(z) = (z + \gamma_n)a_n^1(z) + p_n a_{n-1}^1(z)$ corresponds to $T_{n-1,n} = -p_n$ and $T_{n,n} = -\gamma_n$. For the block case, this generalizes to (3.4), so that $T_{n,n}$ is the companion matrix of q_n, $T_{n+1,n}$ is, as in the general case, zero except for its right top element, which is 1 and finally $T_{n-1,n}$ is zero except for its right top element, which is $-p_n$. For the initial conditions one may choose $a_{-1}^1 = 0$ and $a_0^1 = 1$.

Note that the block tridiagonal form of T is immediate from the symmetry of M. Indeed, $J = DT = A^T M^< A = A^T M^\wedge A = T^T D = J^T$, hence J is symmetric and block upper Hessenberg. Thus block tridiagonal. Therefore, also T is block tridiagonal. From $J = DT$, it is also easy to find that D and J have a nice lower triangular Hankel structure in their blocks [21, 30]. More precisely

Lemma 3.3 *The blocks of D are lower triangular Hankel, say*

$$D_{kk} = \begin{bmatrix} & & & \gamma_0^k \\ & & \iddots & \gamma_1^k \\ & \iddots & \iddots & \vdots \\ \gamma_0^k & \gamma_1^k & \cdots & \gamma_{\alpha-1}^k \end{bmatrix} \quad \text{and} \quad \left[\begin{array}{c|c} & J_{k-1,k} \\ \hline J_{k,k-1} & J_{k,k} \end{array}\right] = \left[\begin{array}{c|cccc} & & & & \gamma_0^k \\ \hline & & & \iddots & \gamma_1^k \\ & & \iddots & \iddots & \vdots \\ \gamma_0^k & \gamma_1^k & & \cdots & \gamma_\alpha^k \end{array}\right] \tag{3.6}$$

with $\alpha = \alpha_{k+1}$ and $\gamma_0^k = \rho_k$.

Proof. The Hankel structure of D_{kk} follows from the Hankel structure of M and the Hankel structure of the kth block column of A. The structure of J follows from its symmetry and the structure of T, especially that T_{kk} is a companion matrix. For more details, we refer to [21, 30]. □

3.3 Reproducing kernel and Christoffel-Darboux relation

We can also give a simple derivation of the Christoffel-Darboux relations. (Compare with [13, p. 81].)

Recall that the reproducing kernel is given by

$$K_n(y, x) = \sum_{k=0}^{n} \mathbf{a}_k(x) D_{kk}^{-1} \mathbf{a}_k(y)^T = \mathbf{y}^T M_n^{-T} \mathbf{x}.$$

We can derive the following Christoffel-Darboux relation.

Theorem 3.4 (Chistoffel-Darboux relation) *Suppose M is a Hankel matrix with block indices ν_k and block OPs $\mathbf{a}_k$, then the following Christoffel-Darboux relation holds*

$$K_n(y, x) = \frac{a_{n+1}^1(x) a_n^1(y) - a_{n+1}^1(y) a_n^1(x)}{(x - y)\rho_n} \tag{3.7}$$

where $\rho_n = \langle z^{\nu_{n+1}-1}, a_n^1 \rangle \neq 0$.

Proof. We compute first $xK_n(y,x)$. Set $\mathbf{a}(x) = [\mathbf{a}_0(x), \ldots, \mathbf{a}_n(x)]$ and use $x\mathbf{a}(x) = \mathbf{a}(x)T_n + e_\nu^T a_{n+1}^1(x)$ $(\nu = \nu_{n+1} - 1)$ to arrive at

$$\begin{aligned} xK_n(y,x) &= x\mathbf{a}(x)D_n^{-1}\mathbf{a}^T(y) \\ &= \mathbf{a}(x)T_nD_n^{-1}\mathbf{a}^T(y) + a_{n+1}^1(x)e_\nu^T D_n^{-1}\mathbf{a}^T(y) \end{aligned}$$

Note that $e_\nu^T D_n^{-1}$ selects the last row of D_n^{-1}. The latter is block diagonal with last block equal to D_{nn}^{-1}. Since D_{nn} is a Hankel matrix with zeros above the second diagonal and nonzero element ρ_n on this diagonal, its inverse has the form

$$\begin{bmatrix} & & \rho_n \\ & \iddots & \vdots \\ \rho_n & \cdots & \times \end{bmatrix}^{-1} = \begin{bmatrix} \times & \cdots & \rho_n^{-1} \\ \vdots & \iddots & \\ \rho_n^{-1} & & \end{bmatrix}.$$

Therefore $e_\nu^T D_n^{-1} = [0 \ \ldots \ 0 \ \rho_n^{-1} \ 0 \ \ldots \ 0]$ where the nonzero element is at position ν_n. Thus we have

$$xK_n(y,x) = \mathbf{a}(x)T_nD_n^{-1}\mathbf{a}^T(y) + \frac{a_{n+1}^1(x)a_n^1(y)}{\rho_n}. \tag{3.8}$$

For symmetry reasons, namely $T_nD_n^{-1} = D_n^{-1}J_nD_n^{-1}$ is symmetric, we also have

$$yK_n(y,x) = \mathbf{a}(x)T_nD_n^{-1}\mathbf{a}^T(y) + \frac{a_{n+1}^1(y)a_n^1(x)}{\rho_n}. \tag{3.9}$$

Subtraction of (3.8) and (3.9) leads to the given relation. □

The following results will be used in section 3.5. With the notation of Lemma 3.3, we define the Hankel matrix S^k whose generator for the first row is $\gamma_0^k z^\alpha + \gamma_1^k z^{\alpha+1} + \cdots + \gamma_\alpha^k z^{2\alpha}$ where $\alpha = \alpha_{k+1}$. The last element γ_α^k is the one in the lower right corner of the J-form in (3.6). It follows from Lemma 3.3, that if $q_k(z)$ is the monic polynomial appearing in the recursion for the FOPs $a_k^1(z)$ (with respect to M), then

$$\begin{bmatrix} & & \gamma_0^k \\ & \iddots & \vdots \\ \gamma_0^k & \cdots & \gamma_\alpha^k \end{bmatrix} \mathrm{Q}_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \times \end{bmatrix}.$$

Thus $q_k(z)$ is the TOP for block number 1 in the sequence of FOPs that can be constructed for the moment matrix S^k. Applying the Chistoffel-Darboux relation to this situation gives immediately

Lemma 3.5 *The generator for D_{kk}^{-1} is given by*

$$D_{kk}^{-1}(w,z) = \frac{q_k(w) - q_k(z)}{(w-z)\rho_k},$$

where $q_k(z)$ is the monic polynomial appearing in the recursion for the FOPs $a_k^1(z)$.

3.4 Continued fractions and Padé approximants

Also the relation with continued fractions and Padé approximation has been studied in [13]. Again we include a proof to show that it is a specialization of the general case of the previous section.

With the generalized 'three term recurrence relation' (3.4) we associate the continued fraction

$$\frac{p_1|}{|q_1(z)} + \frac{p_2|}{|q_2(z)} + \frac{p_3|}{|q_3(z)} + \cdots \tag{3.10}$$

whose convergents are c_k^1/a_k^1, $k = 1, 2, \ldots$ with the initial conditions

$$\begin{bmatrix} c_{-1}^1 & c_0^1 \\ a_{-1}^1 & a_0^1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We shall see that these convergents are the approximants that were discussed in Section 2.6 for the more general situation.

Recall that with the notation of the general case, $g^{[0]}(z) = z^{-1}H(z^{-1})$ which we denote in the sequel as $g(z)$. The other $g^{[i]}(z)$ are now shifted versions of $g(z)$ i.e.,

$$g^{[i+1]}(z) = zg^{[i]}(z) - \mu_i = z^{i+1}g(z) - s_{i+1}(z), \qquad s_{i+1}(z) = zs_i(z) + \mu_i, \quad s_0 = 0. \tag{3.11}$$

Note that $[0\ s_1\ \cdots\ s_{\nu_{n+1}}]$ is the polynomial part of $g(z)[1\ z\ \cdots\ z^{\nu_{n+1}}]$ and corresponds to the bottom row in the matrix of (2.22). The approximants (take $i = 0$ in the general situation) are now

$$g_n(z) = e_0^T M_n (zM_n - M_n^<)^{-1} M_n e_0 = \sum_{k=0}^{\infty} \mu_k^n z^{-(k+1)}. \tag{3.12}$$

We know that in general $\mu_i^n = \mu_i$ for $i = 0, 1, \ldots, \nu_{n+1}$. We verify that for the Hankel case the correspondence of the coefficients goes almost twice as far, which makes it an ordinary Padé approximant.

Theorem 3.6 *The approximant $g_n(z)$ is a Padé approximant at infinity for $g(z) = z^{-1}H(z^{-1})$.*

Proof. We already know from the general Theorem 2.5 for $i = 0$ that

$$g_n(z) = \sum_{k=0}^{\nu} \mu_k z^{-(k+1)} + \sum_{k=1}^{\infty} (e_0^T M_n F_n^{\nu+k} e_0) z^{-(\nu+k+1)}.$$

We use

$$M_n F_n = M_n^< \text{ and } e_j^T M_n^< = e_{j+1}^T M_n, \quad 0 \le j < \nu = \nu_{n+1} - 1$$

ν times to get $e_0^T M_n F_n^\nu = e_\nu^T M_n$. Then by (2.23), it holds for $1 \le k \le \nu + 1$ that

$$e_0^T M_n F_n^{\nu+k} e_0 = e_\nu^T M_n F_n^k e_0 = [\mu_\nu \cdots \mu_{2\nu}] F_n^k e_0 = \mu_{\nu+k}.$$

Hence

$$g_n(z) = \sum_{k=0}^{2\nu+1} \mu_k z^{-(k+1)} + \cdots$$

This proves that μ_k and μ_k^n correspond for at least $k = 0, 1, \ldots, 2\nu + 1 = 2\nu_{n+1} - 1$. Because a_{n+1}^1 is over-orthogonal by $\alpha_{n+2} - 1$, the fit is even better:

$$\mu_k = \mu_k^n, \text{ for } k = 0, 1, \ldots, 2\nu_{n+1} + \alpha_{n+2} - 2 = 2\nu_{n+2} - \alpha_{n+2} - 2.$$

Since a^1_{n+1} is the denominator of g_n, it has degree ν_{n+1}. The corresponding numerator is the polynomial part of $g(z)a^1_{n+1}(z)$ and it can be identified as a polynomial of the second kind (see Theorem 2.5). Its degree is therefore $\nu_{n+1} - \nu_1 \leq \nu_{n+1} - 1$. Thus, in view of the degrees of freedom we have ($\leq 2\nu_{n+1}$) and the number of coefficients that are matched ($\geq 2\nu_{n+1}$) we have proved that the rational function $g_n(z)$ is a Padé approximant at infinity for $g(z)$. □

Theorem 3.7 *The approximants $g_n(z)$ are the convergents of the continued fraction (3.10).*

Proof. The recurrence (3.4) gives

$$g(z)a^1_{n+1}(z) = g(z)q_n(z)a^1_n(z) + g(z)p_n a^1_{n-1}(z).$$

Each term has a polynomial part and a strictly proper part.

$$g(z)a^1_k(z) = c^1_k(z) + r^1_k(z) \tag{3.13}$$

where $c^1_{n+1}(z) = c_{n+1}(z)$ is the polynomial part (the numerator of $g_n(z)$) and $r^1_{n+1}(z)$ is strictly proper (the linearized residual of $g_n(z)$). In view of the degrees of the polynomials and the order of contact, the previous relation can be decoupled as

$$\begin{aligned} c^1_{n+1}(z) &= q_n(z)c^1_n(z) + p_n c^1_{n-1}(z) & (3.14)\\ r^1_{n+1}(z) &= q_n(z)r^1_n(z) + p_n r^1_{n-1}(z). & (3.15) \end{aligned}$$

Thus, after checking the initial conditions, we see that the numerator of g_n corresponds to the numerator of the nth convergent. We proved that $g_n(z)$ is the nth convergent of (3.10). □

Note that the recurrence relation for $r^1_n(z)$ starts with $r^1_0(z) = g(z)$ and $r^1_{-1}(z) = -1$. We now prove that these $r^1_n(z)$ can be associated with columns of the matrix $R = MA$.

Theorem 3.8 *Associate with the block lower triangular matrix $R = MA$ the strictly proper series*

$$[z^{-1}, z^{-2}, \ldots]R = [\mathbf{r}_0(z), \mathbf{r}_1(z), \ldots]$$

where the blocks $\mathbf{r}_k(z) = [r^1_k(z), \ldots, r^{\alpha_{k+1}}_k(z)]$ reflect the block structure of M. Then these $r^1_k(z)$ are the linearized residuals of (3.13).

Proof. From (3.11) and since $R = MA$,

$$\begin{aligned} [r_0(z), r_1(z), r_2(z), \ldots] &= [z^{-1}, z^{-2}, \ldots]MA \\ &= [g(z), zg(z), z^2 g(z), \ldots]A - [0, s_1(z), s_2(z), \ldots]A \\ &= g(z)[a_0(z), a_1(z), \ldots] - [0, t_1(z), t_2(z), \cdots], \end{aligned}$$

where $[0, t_1(z), t_2(z), \ldots] = [0, s_1(z), s_2(z), \ldots]A$. Since $s_k(z)$ is a polynomial, also $t_k(z)$ is a polynomial. Let us consider the first column of block n. Because $r^1_n(z)$ is strictly proper, the polynomial parts in the right hand side should cancel. Thus (using an obvious notation) $t^1_n(z)$ is the polynomial part of $g(z)a^1_n(z)$, which is the numerator $c^1_n(z)$ of $g_n(z)$. Therefore (3.13) holds, which proves the theorem. □

3.5 The Euclidean algorithm and Schur complements

If t and d are polynomials, then we say that q is the quotient and r the remainder in the division $t/d = q + r/d$ when q is polynomial and r/d is strictly proper (i.e., contains only negative powers in its expansion at ∞). We apply the same definition when t and d are formal series in z^{-1}. The quotient is the polynomial part in the expansion of t/d and r/d is the strictly proper part. This notion of division will be used to formulate the Euclidean algorithm.

Recall that q_n is a monic polynomial of degree α_{n+1} and $p_n = -\rho_n/\rho_{n-1}$ is a nonzero constant. Writing $R = A^{-T}D$, we get for the nth diagonal block in this relation:

$$R_{nn} = A_{nn}^{-T} D_{nn} = \begin{bmatrix} 1 & & \\ \vdots & \ddots & \\ \times & \cdots & 1 \end{bmatrix} \begin{bmatrix} & & \rho_n \\ & \cdot^{\cdot^{\cdot}} & \vdots \\ \rho_n & \cdots & \times \end{bmatrix} = \begin{bmatrix} & & \rho_n \\ & \cdot^{\cdot^{\cdot}} & \vdots \\ \rho_n & \cdots & \times \end{bmatrix}$$

which shows that

$$r_n^1(z) = \rho_n z^{-\nu_{n+1}} + \text{lower order terms}, \quad \rho_n \neq 0.$$

This confirms what we have found before in the proof of Theorem 3.6. The form of the $r_k^1(z)$ implies that $[r_{n-1}^1(z)]/[r_n^1(z)]$ has a polynomial part of degree α_{n+1} while $[r_{n+1}^1(z)]/[r_n^1(z)]$ is strictly proper. Thus from the relation (3.15)

$$\frac{p_n r_{n-1}^1(z)}{r_n^1(z)} = -q_n(z) + \frac{r_{n+1}^1(z)}{r_n^1(z)},$$

we see that $-q_n(z)$ is the quotient of the division $[p_n r_{n-1}^1(z)]/[r_n^1(z)]$ where the constant p_n is chosen to ensure that $q_n(z)$ is a monic polynomial. The series $r_{n+1}^1(z)$ is the remainder. Thus the p_n and $q_n(z)$ can be obtained by successive divisions of the series $r_k^1(z)$, normalized to give a monic polynomial quotient. This is precisely the Euclidean algorithm for power series in z^{-1} [9, 10].

Note that the R matrix is block lower triangular and its (nonsingular) leading submatrices R_n feature in

$$M_n = R_n D_n^{-1} R_n^T \tag{3.16}$$

so that this gives a Choleski type decomposition for all nonsingular leading submatrices of M. Therefore we can identify the Euclidean algorithm as a fast block Choleski algorithm for Hankel matrices.

The factorization (3.16) can be rewritten in a symmetric version of the form (2.10), which links this up with Schur complements and the Schur algorithm. In fact the Schur algorithm for Hankel matrices is essentially the same as the Euclidean algorithm. To clarify this statement, we start by embedding the Hankel matrices in the larger class of quasi Hankel matrices. That are matrices whose generators satisfy

$$M(w,z) = \frac{G(w)^T \Sigma G(z)}{w-z} = \frac{g_2(w)g_1(z) - g_1(w)g_2(z)}{w-z}, \quad \Sigma = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad G(z) = \begin{bmatrix} g_1(z) \\ g_2(z) \end{bmatrix}$$

where $G(0) \neq 0$. Without loss of generality, we can always choose $g_1(0) \neq 0$ and $g_2(0) = 0$. For example, the structure (3.2) shows that Hankel matrices are a special case of quasi Hankels obtained by choosing $G(z)^T = [1 \; zH(z)]$. Conversely, a quasi Hankel matrix is Hankel iff it has a generator of this form. Quasi Hankel matrices are congruent to Hankel matrices because we have

Lemma 3.9 *A matrix Q is quasi Hankel iff there exists a lower triangular Toeplitz matrix L such that $M = LQL^T$ is Hankel.*

Proof. Suppose that $g_1(0) \neq 0$ and $g_2(0) = 0$, say $g_2(z) = zH(z)$. Let the first column of L have generator $f(z) = 1/g_1(z)$. Then $f(w)Q(w,z)f(z)$ is the generator of M. This has the form $M(w,z) = \Big(wH(w) - zH(z)\Big)/(w-z)$, with $H(z) = g_2(z)/g_1(z)$, which proves that M is Hankel. The converse is even simpler to prove. □

There are several reasons to use this wider class of quasi Hankels. For example, the Chistoffel-Darboux formula shows that the inverse of a Hankel matrix is quasi Hankel. Also the Schur complement in a Hankel matrix is not Hankel but quasi Hankel. In fact this holds for any quasi Hankel matrix (see [30]). The latter will also follow from the next theorem which makes the link between the Schur algorithm and the Euclidean algorithm. Indeed the Euclidean algorithm computes recursive updates of the factors in the generators for the successive Schur complements. These generators are of the appropriate form to show that the Schur complements are quasi Hankel. We first need a change of variables. For $r_k^1(z)$ as defined in Theorem 3.8, set

$$h_k(z) = z^{-\nu_k} r_k^1(1/z) = \rho_k z^{\alpha_{k+1}} + \text{higher order terms.}$$

The Euclidean algorithm, which provided the recurrence for the $r_n^1(z)$ translates into

$$h_{k+1}(z) = z^{-\alpha_{k+1}}[z^{-\alpha_k} h_{k-1}(z) p_k + h_k(z) q_k(1/z)].$$

Define

$$G_k(z)^T = \left[\ \rho_{k-1}^{-1} z^{-\alpha_k} h_{k-1}(z) \,\middle|\, h_k(z)\ \right] \tag{3.17}$$

then the previous update leads to

$$G_{k+1}(z) = \theta_k(z) G_k(z), \quad \theta_k(z) = \begin{bmatrix} 0 & \rho_k^{-1} \\ -\rho_k & q_k(1/z) \end{bmatrix} z^{-\alpha_{k+1}}. \tag{3.18}$$

The Euclidean algorithm is thus equivalent with this recursive update of $G_k(z)$. We now show that these $G_k(z)$ are factors appearing in the generator for the successive Schur complements which were also considered in the general case.

Theorem 3.10 *Introduce the successive Schur complements $M^{k+1} = M^k_{\langle 0\rangle}$, $k \geq 0$, $M^0 = M$ as in Section 2.3. Then these Schur complements are quasi Hankel and their generators satisfy*

$$M^k(w,z) = \frac{G_k(w)^T \Sigma G_k(z)}{w-z},$$

with $G_k(z)$ as in (3.17).

Proof. This is obviously true for $k = 0$ since $M^0 = M$, $\alpha_0 = 0$, $h_0(z) = zH(z)$ and $h_{-1} = \rho_{-1} = -1$.
The induction step is proved as follows. It is clear that (see the proof of Theorem 2.1)

$$M^{k+1}(w,z) = (wz)^{-\alpha_{k+1}}[M^k(w,z) - \mathbf{w}^T R_{[*,k]} D_{kk}^{-1} R_{[*,k]}^T \mathbf{z}] \tag{3.19}$$

where $R_{[*,k]}$ is the kth block column of R, after leaving out all the zeros from block rows $0, 1, \ldots, k-1$. Recall that $R_{[*,k]}$ is Hankel and its generator can be written as

$$\mathbf{w}^T R_{[*,k]} = w^{-1} h_k(w) [1 \ w^{-1} \ \cdots \ w^{-\alpha_{k+1}+1}].$$

Therefore

$$\mathbf{w}^T R_{[*,k]} D_{kk}^{-1} R_{[*,k]}^T \mathbf{z} = (wz)^{-1} h_k(w) h_k(z) d_k(w,z) \tag{3.20}$$

where $d_k(w,z) = D_{kk}^{-1}(w^{-1}, z^{-1})$. From Lemma 3.5, we know that

$$(wz)^{-1}(w-z) d_k(w,z) = (z^{-1} - w^{-1}) D_{kk}(w^{-1}, z^{-1}) = -\Big(q_k(w^{-1}) - q_k(z^{-1})\Big)/\rho_k. \tag{3.21}$$

By induction hypothesis,

$$(w-z) M^k(w,z) = G_k(w)^T \Sigma G_k(z). \tag{3.22}$$

Multiply (3.19) by $(w-z)$ and plug in (3.20–3.22), then we get

$$\begin{aligned} (w-z) M^{k+1}(w,z) &= (wz)^{-\alpha_{k+1}} \left\{ G_k(w)^T \Sigma G_k(z) + \rho_k^{-1} h_k(w) h_k(z) [q_k(w^{-1}) - q_k(z^{-1})] \right\} \\ &= (wz)^{-\alpha_{k+1}} G_k(w)^T (\Sigma + \Omega) G_k(z) \end{aligned}$$

with $\Omega = \rho_k^{-1}[q_k(w^{-1}) - q_k(z^{-1})][0\ 1]^T[0\ 1]$. A straightforward computation gives $\theta_k(w)^T \Sigma \theta_k(z) = (wz)^{-\alpha_{k+1}}(\Sigma + \Omega)$. Thus, using (3.18), we arrive at

$$(w-z) M^{k+1}(w,z) = G_k(w)^T \theta_k(w)^T \Sigma \theta_k(z) G_k(z) = G_{k+1}(w)^T \Sigma G_{k+1}(z),$$

which proves the theorem. □

This theorem thus says that the Schur algorithm for Hankel matrices is basically the same as the Euclidean algorithm.

The previous theorem was given in the normal case in [29].

3.6 The Lanczos algorithm

When in the general setting of Krylov subspace methods, we take $Q = P$, then M will be Hankel with entries $\mu_k = y_0^H P^k x_0$, $k = 0, 1, \ldots$ Its symbol is $H(z) = y_0^H (I - zP)^{-1} x_0$ and the resulting Krylov iteration can be performed efficiently because of the Hankel structure of M. This is the well known Lanczos method for nonsymmetric matrices P. The recent literature is abundant on this subject and we shall not go into details here. See for example [14] for a survey.

4 Formal orthogonality: the Toeplitz case

Now we shall suppose that the moment matrix M is Toeplitz, i.e., $\mu_{i,j} = \mu_{i-j}$, but in general it need not be Hermitian or symmetric. Thus we should distinguish between left and right orthogonality. Again, for simplicity we shall only consider ROPs.

The most popular recurrence relation for polynomials orthogonal on the unit circle with respect to a positive measure is the coupled recursion for the polynomials and their reciprocals. This is the so called Szegő recurrence, also used in the Levinson algorithm. The moment matrix for these polynomials is positive Hermitian Toeplitz. It is possible to produce recurrence relations which are non symmetric block versions of these coupled recursions (see e.g. [25]). However, in our development, we shall concentrate on the three term type of recurrence relation.

4.1 The generator

The Toeplitz structure of the moment matrix allows to reduce the definition of the bilinear form for polynomials to a linear form defined on the space of Laurent polynomials by

$$\langle p, q \rangle = \mu\left(\overline{p}(z^{-1})q(z)\right)$$

where μ is defined by $\mu(z^{-k}) = \mu_k$, $k = 0, \pm 1, \pm 2, \ldots$ For convenience we introduce the parahermitian transformation as

$$f_*(z) = \overline{f}(z^{-1}) = \overline{f(1/\overline{z})}$$

for any formal series $f(z)$. Thus $\langle p, q \rangle = \mu(p_* q)$.

The generator of a Toeplitz matrix satisfies

$$M(w, z) = \mathbf{w}^T M \mathbf{z} = \frac{g^+(w) - g^-(z^{-1})}{1 - wz}, \quad g^+(w) = \sum_{k=0}^{\infty} \mu_k w^k, \quad g^-(z) = -\sum_{k=1}^{\infty} \mu_{-k} z^{-k}.$$

This is easily seen by noting that $(1 - wz)M(w, z)$ corresponds to the generator of the displacement difference $M - ZMZ^T$. This can be brought into the form

$$(1 - wz)M(w, z) = G(w)^T \Sigma H(z), \quad \Sigma = \mathrm{diag}(1, -1), \tag{4.1}$$

with

$$G(w)^T = [\tfrac{1}{2} + g^+(w), -\tfrac{1}{2} + g^+(w)], \quad H(z)^T = [\tfrac{1}{2} - g^-(z^{-1}), -\tfrac{1}{2} - g^-(z^{-1})].$$

A matrix with generator of the form (4.1) is called a quasi Toeplitz matrix. This class of matrices plays an important role in the recursive construction of Schur complements by the Schur algorithm. We do not consider this any further in this paper, but the reader who is interested can find more information for example in [29] for the Hermitian positive definite case, for the block Toeplitz case in [12], and for a time-varying notion of displacement rank in [31].

The generator for a Toeplitz matrix can be caught into one single Laurent series. Indeed,

$$M(z, z^{-1}) = g(z) = \sum_{k=-\infty}^{\infty} \mu_k z^k = g^+(z) - g^-(z).$$

The g^+ and g^- series (including its minus sign) appear naturally in the following context. Consider $\gamma(\zeta) = \zeta/(\zeta - z)$, then there exist two formal expansions for γ, namely

$$\gamma^+(\zeta) = \sum_{k=0}^{\infty} \frac{z^k}{\zeta^k} \quad \text{and} \quad \gamma^-(\zeta) = -\sum_{k=1}^{\infty} \frac{\zeta^k}{z^k}.$$

These are precisely the expansions which produce the generators g^+ and g^- since $g^+(z) = \mu(\gamma^+)$ and $g^-(z) = \mu(\gamma^-)$.

4.2 Christoffel-Darboux relation

For the derivation of a Christoffel-Darboux formula, we shall make use of the persymmetry of a Toeplitz matrix.

Denote by $\check{I}$ the permutation matrix (of appropriate size) which has ones on its second diagonal and is zero elsewhere. Its effect on a column vector is turning it upside down. A matrix M is persymmetric when $\check{I}\, M\, \check{I} = M^T$. Obviously, a Toeplitz matrix is persymmetric. Also its inverse is persymmetric and thus for $M_n = M[\nu_{n+1}, \nu_{n+1}]$, and using the same notation as before

$$M_n^{-T} = \check{I}\, M_n^{-1}\, \check{I} = \overline{B}_n D_n^{-T} A_n^T.$$

Consequently,

$$K_{n+1}(y,x) = \mathbf{y}^H M_{n+1}^{-T} \mathbf{x} = K_n(y,x) + \mathbf{a}_{n+1}(x) D_{n+1,n+1}^{-1} \mathbf{b}_{n+1}(y)^H \tag{4.2}$$

can also be written as

$$\begin{aligned} K_{n+1}(y,x) &= \mathbf{y}^H\, \check{I}\, M_{n+1}^{-1}\, \check{I}\, \mathbf{x} \\ &= \mathbf{y}^H\, \check{I}\, A_{n+1} D_{n+1}^{-1} B_{n+1}^H\, \check{I}\, \mathbf{x}^T \\ &= \sum_{k=0}^{n+1} (x\overline{y})^{\nu_{n+2}-\nu_{k+1}} \overline{\check{\mathbf{a}}_k(y)} D_{kk}^{-1} \check{\mathbf{b}}_k(x)^T \end{aligned}$$

where $\check{\mathbf{a}}_k(y)$ and $\check{\mathbf{b}}_k(x)$ are reversed block polynomials namely

$$\check{\mathbf{a}}_k(z) = z^{\nu_{k+1}-1} \overline{\mathbf{a}}_k(z^{-1}) = z^{\nu_{k+1}-1} \mathbf{a}_{k*}(z)$$

and similarly for $\check{\mathbf{b}}_k$. This leads to

$$\begin{aligned} K_{n+1}(y,x) &= (x\overline{y})^{\alpha_{n+2}} K_n(y,x) + \overline{\check{\mathbf{a}}_{n+1}(y)} D_{n+1,n+1}^{-1} \check{\mathbf{b}}_{n+1}(x)^T \\ &= (x\overline{y})^{\alpha_{n+2}} K_n(y,x) + \check{\mathbf{b}}_{n+1}(x) D_{n+1,n+1}^{-T} \check{\mathbf{a}}_{n+1}(y)^H. \end{aligned} \tag{4.3}$$

Subtracting (4.3) from (4.2) gives the following Christoffel-Darboux relation.

Theorem 4.1 (Christoffel-Darboux relation) *Suppose M is Toeplitz with block OPs $\{\mathbf{b}_k, \mathbf{a}_k\}$ of size α_{k+1} and $\langle \mathbf{b}_i, \mathbf{a}_j \rangle = \delta_{ij} D_{ij}$. Let $K_n(y,x)$ be the reproducing kernel and denote by $\{\check{\mathbf{b}}_k, \check{\mathbf{a}}_k\}$ the reversed block polynomials, then the following Christoffel-Darboux formula holds*

$$K_n(y,x) = \frac{\check{\mathbf{b}}_{n+1}(x) D_{n+1,n+1}^{-T} \check{\mathbf{a}}_{n+1}(y)^H - \mathbf{a}_{n+1}(x) D_{n+1,n+1}^{-1} \mathbf{b}_{n+1}(y)^H}{1 - (x\overline{y})^{\alpha_{n+2}}}.$$

4.3 Block structure and Iohvidov indices

As in the general case, if the matrix M is not strongly regular, we have to compute the ROPs in blocks. However, the blocks that will appear here in a natural way are *not* the same as the blocks defined in the general case, or as in the section on Hankel matrices.

The block structure is related to the block structure in the M-table. The M-table is an array whose (κ, ν) entry is equal to $\det M[\kappa; \nu]$, $\kappa = 0, \pm 1, \pm 2, \ldots$, $\nu = 0, 1, 2, \ldots$ where $M[\kappa; \nu] = M[\kappa : \kappa+\nu-1, \kappa : \kappa+\nu-1] = [\mu_{i-j-\kappa}]_{i,j=0}^{\nu-1}$. By convention we set $\det M[\kappa; 0] = 1$.

The block structure theorem for the M-table says that the zeros in the table will appear in square blocks (possibly infinite) and these blocks are separated by nonzero entries [17, 8]. Therefore we define an M-block as such a square block of zeros including its bottom border and its right border of nonzero entries when these borders are not at infinity. Then, when all the entries of the M-table are nonzero, all M-blocks will be of size 1. In general, the square M-blocks tile the complete M-table without holes.

The recurrence we propose moves along the row $\kappa = 0$ in the M-table (note that $M[0;\nu] = M[\nu,\nu]$) and the blocks of ROPs will correspond to the M-blocks one has to traverse in the M-table. Unfortunately, these blocks are not necessarily the same as the blocks defined by the regular leading submatrices of M. Our motivation for changing this concept lies in the way we shall compute the ROPs recursively.

Because in this section we shall always refer to blocks of ROPs, we shall use the same terminology (block index, block size, ...) and the same notation $(\nu_k, \alpha_k, \ldots)$ as in the previous sections since confusion is very unlikely.

Definition 4.1 (blocks, Iohvidov indices) *The block indices are defined as* $\nu_n = \sum_{k=1}^{n} \alpha_k$, $\nu_0 = 0$. *The block sizes* α_k *are obtained as* $\alpha_{n+1} = \alpha_{n+1}^{+} + \alpha_{n+1}^{-} + 1$. *The Iohvidov indices* $\alpha_{n+1}^{\pm}$ *of block* n *are defined by*

$$\begin{aligned} \alpha_{n+1}^{-} &= \min\{\alpha \in \mathbf{N} : \sigma = \langle 1, z^{\alpha+1} a_\nu \rangle \neq 0\} \\ \alpha_{n+1}^{+} &= \min\{\alpha \in \mathbf{N} : \rho = \langle z^{\nu+\alpha}, a_\nu \rangle \neq 0\} \end{aligned}$$

where $\nu = \nu_n$, *i.e.* $a_\nu = a_n^1$ *is the TOP of block* n.

An interpretation of the Iohvidov indices can be easily given. They indicate how far from the top (α_{n+1}^{-}) and from the bottom (α_{n+1}^{+}) the row $\kappa = 0$ will hit the nth M-block. Indeed, $\det M[0,0] = 1 \neq 0$ by convention and an induction which follows from the theorems below shows that the submatrices $M[\nu,\nu] = M[0;\nu]$ are regular whenever $\nu = \nu_n$, some block index so that $a_{\nu_n} = a_n^1$ is uniquely defined. The definition of the Iohvidov indices then says that

$$M^{[\kappa,*]}\mathrm{A}_n^1 \begin{cases} = 0 \text{ for } -\alpha_{n+1}^{-} \leq \kappa < \alpha_{n+1}^{+} + \nu \\ = \sigma_n \neq 0 \text{ for } \kappa = -\alpha_{n+1}^{-} - 1 \\ = \rho_n \neq 0 \text{ for } \kappa = \alpha_{n+1}^{+} + \nu. \end{cases}$$

This means that an M-block will start at column $\nu+1$ and that its top row is at position $\kappa = -\alpha_{n+1}^{-}$ and its (nonzero) bottom row is at $\kappa = \alpha_{n+1}^{+}$. Thus its size is $\alpha_{n+1} = \alpha_{n+1}^{-} + \alpha_{n+1}^{+} + 1$. As a special case we see that $M[0;\nu+1] = M[\nu+1,\nu+1]$ will be regular iff $\alpha_{n+1}^{+} = 0$. However whether α_{n+1}^{+} is zero or not, when $\alpha_{n+1}^{-} > 0$, then it is simple to define ROPs of degree $\nu_n + 1, \ldots, \nu_n + \alpha_{n+1}^{-}$ by just shifting $a_{\nu_n} = a_n^1$, i.e., $a_{\nu_n+i}(z) = a_n^{i+1}(z) = z^i a_n^1(z)$, $i = 1, \ldots, \alpha_{n+1}^{-}$. This follows from the definition of α_{n+1}^{-}. We shall in the case $\alpha_{n+1}^{+} = 0$ also consider these polynomials to belong to the same block as a_{ν_n}. Thus block n is in that case

$$\mathbf{a}_n(z) = [a_{\nu_n}(z), a_{\nu_n+1}(z), \ldots, a_{\nu_n+\alpha_{n+1}^{-}}(z)],$$

even though, according to the general definition of Section 2, it covers $\alpha_{n+1}^{-} + 1$ blocks of size 1.

When $\alpha_{n+1}^{+} > 0$, then with $\nu' = \nu_n + \alpha_{n+1}^{-}$, $M[\nu',\nu']$ will still be singular and we have to add α_{n+1}^{+} more polynomials in block n to reach the next regular submatrix $M_{n+1} = M[\nu_{n+1}, \nu_{n+1}]$.

4.4 Row recurrence

The further description of the blocks will be constructive and will involve the explicit computation of the recurrence.

We shall first exclude some trivial situations. First exclude the case where $\alpha_1^+ = \infty$ since this means that M is strictly upper triangular and then there do not exist (monic) orthogonal polynomials unless $\alpha_1^- = \infty$. The case $\alpha_1^- = \infty$ is also trivial, since then there is an infinite block of orthogonal polynomials which are just the powers of z.

Hence we suppose $\alpha_1^+ + \alpha_1^- < \infty$. Then block number 0 contains $\nu = \nu_1 = \alpha_1^+ + \alpha_1^- + 1$ polynomials $a_k = z^k$, $k = 0, 1, \ldots, \nu - 1$. The first monic polynomial of block number 1, i.e., $a_\nu = a_1^1$ has a coefficient vector that can be found by solving the system

$$M_0 \dot{\mathrm{A}}_1^1 = -[\mu_{-\nu} \cdots \mu_{-1}]^T, \quad M_0 = M[\nu_1, \nu_1].$$

Note that the matrix of this system is regular and has a band of zeros of width $\alpha_1^+ + \alpha_1^- = \nu - 1$ (which is 0 in the regular case). Hence this system can be solved very efficiently.

Now for the general situation of block n. Suppose we already know that $M_{n-1} = M[\nu, \nu]$ ($\nu = \nu_n$) is regular. It follows that the TOP of block n, i.e., a_n^1, will be uniquely defined. By definition of α_{n+1}^-, we know that the polynomials $a_{\nu_n+k}(z) = z^k a_n^1(z)$ are orthogonal to all polynomials of lower degree for $k = 0, 1, \ldots, \alpha_{n+1}^-$. This is a consequence of the Toeplitz structure of M which implies $\langle zf(z), zg(z)\rangle = \langle f(z), g(z)\rangle$. Now we define an auxiliary polynomial for block n as

$$\tilde{a}_n(z) = z^{\alpha_{n+1}^- + 1} a_n^1(z). \tag{4.4}$$

The auxiliary polynomial $\tilde{a}_n$ is the first one in block n (except for a_n^1 of course) which is obtained by shifting the previous one and is not orthogonal to all polynomials of lower degree. Since it is not an orthogonal polynomial, we shall refer to it as the NOP for block n. Note that we have

$$\left\langle z^k, \tilde{a}_n(z)\right\rangle = 0, \quad k = 1, 2, \ldots, \nu_{n+1} - 1$$

Before treating the generic case where both Iohvidov indices are finite, we shall deal with the cases of infinite α^+ and/or α^- first.

Theorem 4.2 *Let M be a Toeplitz matrix and suppose $\nu = \nu_n$ is a block index. Hence $M[\nu, \nu]$ is regular and the TOP $a_n^1 = a_\nu$ is uniquely defined. Let the Iohvidov index α_{n+1}^- for that block be infinite. Then $a_{\nu+k}(z) = z^k a_\nu(z)$, $k = 0, 1, \ldots$ is a set of ROPs, i.e., the block size $\alpha_{n+1} = \infty$.*

All the leading submatrices $M[k, k]$ for $k > \nu$ are regular iff the other Iohvidov index $\alpha_{n+1}^+ = 0$ and they are all singular iff $\alpha_{n+1}^+ > 0$.

Proof. The right orthogonality follows immediately from the definition of α_{n+1}^-.

When we express the fact that $[a_0, a_1, \ldots, a_{\nu-1} | a_\nu, za_\nu, \ldots]$ is right orthogonal to z^k, $k = 0, 1, \ldots$, in a matrix form (A is unit upper triangular and contains the coefficient vectors) we see that $MA = R$ with R a matrix of the form

$$R = \begin{bmatrix} R_{00} & 0 \\ R_{10} & R_{11} \end{bmatrix}$$

with R_{00} regular of size ν and R_{11} infinite lower triangular Toeplitz. The diagonal element of R_{11} is $\rho_n = \langle z^\nu, a_\nu(z)\rangle$ which is nonzero iff $\alpha_{n+1}^+ = 0$. This implies the regularity or singularity of the leading submatrices $M[k, k]$ for $k > \nu$ as claimed. □

The previous theorem treated the case where $\alpha_{n+1}^- = \infty$. The next one discusses $\alpha_{n+1}^+ = \infty$.

Theorem 4.3 *Let M be Toeplitz, with block index $\nu = \nu_n$ and corresponding Iohvidov indices $\alpha_{n+1}^{+} = \infty$ and $\alpha_{n+1}^{-} < \infty$, then all $M[k,k]$ are singular for $k > \nu$.*

An infinite block of ROPs can be computed as follows. Set $a_{\nu+k} = z^k a_\nu$ for $k = 0, 1, \ldots, \alpha_{n+1}^{-}$. Shifting the last one of these once more will give $\tilde{a}_n$, the NOP for block n. Setting $\omega = \nu + \alpha_{n+1}^{-} + 1$, we can find a nonzero constant p_n and monic polynomials d_j of degree j such that

$$a_{\omega+j}(z) = z^j \tilde{a}_{n-1}(z) p_n + \tilde{a}_n(z) d_j(z), \quad j = 0, 1, \ldots$$

fill up the remaining polynomials of (the infinite) block n.

Proof. With A_ν the coefficient vector of a_ν, it follows from the definition of $\alpha_{n+1}^{+} = \infty$, that $M\mathrm{A}_\nu = 0$. Hence, using the unit upper triangular matrix A defined by

$$[a_0(z), \ldots, a_\nu(z), z^{\nu+1}, z^{\nu+2}, \ldots] = \mathbf{z}^T A,$$

it is obvious from $R = MA$ that the singularity of $M[k,k]$ for $k > \nu$ follows immediately from the singularity of $R[k,k]$, since R contains a zero column viz. $M\mathrm{A}_\nu$.

The first α_{n+1}^{-} polynomials are indeed ROPs by definition of α_{n+1}^{-}.

The remaining polynomials can be obtained as indicated. For $j = 0$, we know that $\langle z^k, \tilde{a}_n \rangle = 0$ for $k = 1, 2, \ldots$, but it equals a nonzero constant σ_n for $k = 0$. A similar observation holds for $\tilde{a}_{n-1}$ namely $\langle z^k, \tilde{a}_{n-1} \rangle = 0$ for $k = 1, \ldots, \nu_n - 1$ and equals a nonzero constant σ_{n-1} for $k = 0$. Thus the constant $p_n = -\sigma_n / \sigma_{n-1} \neq 0$ exists so that $a_\omega = \tilde{a}_{n-1} p_n + \tilde{a}_n$ is right orthogonal to z^k for $k = 0, 1, \ldots, \nu_n - 1$.

For $j > 0$, suppose that $a_{\omega+i}$ for $i = 0, \ldots, j-1$ have been computed as ROPs, orthogonal to z^k, $k = 0, 1, \ldots, \nu_n - 1$. Hence $z^j a_\omega$ is right orthogonal to z^k, $k = j, \ldots, \nu_n + j - 1$. By adding multiples of $z^i \tilde{a}_n$, $i = 0, \ldots, j-1$, we can satisfy the orthogonality relation. The polynomial d_j has a coefficient vector D_j that can be found as the solution of the upper triangular Toeplitz system

$$\begin{bmatrix} t_0 & t_1 & \ldots & t_{j-1} \\ & t_0 & & t_{j-2} \\ & & \ddots & \vdots \\ & & & t_0 \end{bmatrix} \dot{\mathrm{D}}_j = -p_n \begin{bmatrix} t'_j \\ t'_{j-1} \\ \vdots \\ t'_1 \end{bmatrix} - \begin{bmatrix} t_j \\ t_{j-1} \\ \vdots \\ t_1 \end{bmatrix} \quad \text{where} \quad \begin{cases} t_k = \langle 1, z^k \tilde{a}_n \rangle, \\ t'_k = \langle 1, z^k \tilde{a}_{n-1} \rangle. \end{cases} \tag{4.5}$$

The polynomials $a_{\omega+j}$ thus obtained are monic and of the correct degree. This follows from $\deg \tilde{a}_{n-1} < \deg \tilde{a}_n$. The latter inequality follows from the definitions of ν_n and α_n and the Iohvidov indices being nonnegative. □

Now we are ready to treat the generic situation in the following theorem.

Theorem 4.4 *Let the moment matrix M be Toeplitz and let ν_n be a block index. Let the corresponding Iohvidov indices α_{n+1}^{+} and α_{n+1}^{-} be finite. Then the $\alpha_{n+1} = \alpha_{n+1}^{+} + \alpha_{n+1}^{-} + 1$ orthogonal polynomials for block n can be obtained as follows*

$$a_{\nu_n+i}(z) = z^i a_{\nu_n}(z), \text{ for } i = 0, \ldots, \alpha_{n+1}^{-} \tag{4.6}$$

and for $\omega = \nu_n + \alpha_{n+1}^{-} + 1 = \deg \tilde{a}_n$ (with $\tilde{a}_n$ the NOP for block n)

$$a_{\omega+j}(z) = z^j \tilde{a}_{n-1}(z) p_n + \tilde{a}_n(z) d_j(z), \quad j = 0, \ldots, \alpha_{n+1}^{+} - 1 \tag{4.7}$$

with $p_n \neq 0$, a constant, and d_j a monic polynomial of degree j.

The submatrices $M[k,k]$ are all singular for $k = \nu_n + 1, \ldots, \nu_{n+1} - 1$ while it is regular for $k = \nu_{n+1}$, unless $\alpha_{n+1}^{+} = 0$, in which case they are all regular.

Proof. The proof of the orthogonality and the recurrence relations is a repetition of the previous proofs and we shall not repeat it here. We only check the claim about the regularity of the leading submatrices.

The case where $\alpha_{n+1}^+ = 0$ has been considered before, so we only have to discuss $\alpha_{n+1}^+ > 0$. Define the unit upper triangular matrix A_n by

$$[a_0(z), \ldots, a_{\nu_n}(z), za_{\nu_n}(z), \ldots, z^{\alpha_{n+1}-1}a_{\nu_n}(z)] = \mathbf{z}^T A_n$$

then with $M_n = M[\nu_{n+1}, \nu_{n+1}]$, we define the matrix R_n as $R_n = M_n A_n$. This matrix will have the following form by definition of the Iohvidov indices and the Toeplitz structure of M

$$R_n = \left[\begin{array}{c|c|c} & 0 & T_1 \\ \cline{2-3} R_{n-1} & 0 & 0 \\ \cline{1-3} & 0 & 0 \\ \cline{2-3} & T_2 & 0 \end{array}\right] = \left[\begin{array}{c|c|c} & 0 & T_1 \\ \cline{1-3} R'_{n-1} & 0 & 0 \\ \cline{2-3} & 0 & 0 \\ \cline{1-3} & T_2 & 0 \end{array}\right]$$

with R_{n-1} and R'_{n-1} of size ν_n. The block T_1 is upper triangular Toeplitz with nonzero diagonal element $\sigma_n = \left\langle 1, z^{\alpha_{n+1}^- +1}a_{\nu_n}\right\rangle$ and of size α_{n+1}^+. The block T_2 is lower triangular Toeplitz with nonzero diagonal element $\rho_n = \left\langle z^{\nu_n+\alpha_{n+1}^+}, a_{\nu_n}\right\rangle$ and of size $\alpha_{n+1}^- + 1$. Since for all leading submatrices $\det M[k,k] = \det R[k,k]$, we have $\det M[k,k] = 0$ for $k = \nu_n + 1, \ldots, \nu_{n+1} - 1$ when $\alpha_{n+1}^+ > 0$.

Because $\det R_n = \pm \det T_1 \det T_2 \det R'_{n-1}$, R_n will be regular iff R'_{n-1} is regular. For $n = 0$, R'_{n-1} is empty and the theorem is proved. For $n > 0$, we use the block structure theorem for the M-table. Clearly $R'_{n-1} = M[\alpha_{n+1}^+; \nu_n]A_{n-1}$. As we know from the definition of α_{n+1}^+, this $M[\alpha_{n+1}^+; \nu_n]$ is nonsingular. This proves the theorem.

□

To construct all the ROPs we need a recurrence relation to start a new block. This is given in the next theorem.

Theorem 4.5 *Using the same notation as in the previous theorem, it holds that the TOP of block $n+1$ is given by*

$$a_{n+1}^1(z) = z^{\alpha_{n+1}^+}\tilde{a}_{n-1}(z)p_n + a_n^1(z)q_n(z) \tag{4.8}$$

where again $p_n \neq 0$ is a constant and q_n is a monic polynomial of degree α_{n+1}.

Proof. First we observe that $z^{\alpha_{n+1}^+}\tilde{a}_{n-1}(z)$ is orthogonal to z^k for $\alpha_{n+1}^+ < k \leq \nu_n - 1 + \alpha_{n+1}^+$. By adding multiples of $z^i\tilde{a}_n$, for $i = 0, 1, \ldots, \alpha_{n+1}^+$, we can construct a polynomial

$$\begin{aligned} \hat{a}(z) &= z^{\alpha_{n+1}^+}\tilde{a}_{n-1}(z)p_n + \tilde{a}_n(z)r(z) \\ &= z^{\alpha_{n+1}^+}\tilde{a}_{n-1}(z)p_n + z^{\alpha_{n+1}^- +1}a_n^1(z)r(z) \end{aligned} \tag{4.9}$$

which is orthogonal to z^k for $0 \leq k \leq \nu_n + \alpha_{n+1}^+ - 1$ and where p_n is a nonzero constant and $r(z)$ a monic polynomial of degree α_{n+1}^+. The remaining $\alpha_{n+1}^- + 1$ orthogonality conditions can be satisfied by adding multiples of $z^l a_n^1(z)$ for $l = 0, 1, \ldots, \alpha_{n+1}^-$ which are indeed orthogonal to z^k for $0 \leq k \leq \nu_n + \alpha_{n+1}^+ + l - 1$. This then leads to the form proposed in the theorem:

$a_{n+1}^1(z) = \hat{a}(z) + s(z)a_n^1(z)$ with $s(z)$ of degree at most α_{n+1}^- and where $q_n(z) = s(z) + z^{\alpha_{n+1}^-+1}r(z)$. Scrutinizing the formulas reveals that $p_n = -\sigma_n/\sigma_{n-1} \neq 0$ and $q_n(0) = s(0) = -p_n\rho_{n-1}/\rho_n \neq 0$ where $\sigma_k = \langle 1, \tilde{a}_k\rangle \neq 0$ and $\rho_k = \langle z^{\nu_{k+1}}, \tilde{a}_k\rangle \neq 0$. □

Note that the polynomials d_j of theorem 4.4 can be simply obtained by the recursion

$$d_j(z) = zd_{j-1}(z) + d_j(0).$$

The polynomial $r(z)$ in the previous theorem is obtained in the same way with $j = \alpha_{n+1}^+$.

4.5 Factoring the Hessenberg matrix

The recurrence relation that we have just derived for the block orthogonal polynomials gives rise just like in the general case to a relation of the form (2.15), i.e.,

$$F_nA_n = A_nT_n$$

but unlike in the Hankel case, the upper Hessenberg matrix T_n does not simplify considerably. Although it has a special structure, it is complicated to describe. There is however another formulation, which does give simple matrices. We should then replace the previous relation by

$$F_nA_nU_n = A_nV_n$$

where we have replaced T_n by $V_nU_n^{-1}$. The matrix U_n is a simple upper block bidiagonal matrix and V_n is a unit upper Hessenberg matrix which is block bidiagonal too with a very simple structure. Why is it that we have this extra complication with the expression $F_nA_n = A_nT_n$? In the left-hand side, F_n acts like a shift operator, i.e., a_k in the kth column of A_n is replaced by za_k. The right-hand side expresses these polynomials in terms of the a_j. However, by our recurrence relation, we got expressions which also involved the NOPs $\tilde{a}_k$-polynomials too. The latter do not correspond to columns of the A_n matrix. Thus we have to express them explicitly in terms of a_k's and this makes T_n so complicated. But the auxiliary polynomial $\tilde{a}_n$ is a shifted version of one of the a_k and it is therefore natural to bring that part of the recurrence which refers to one of these $\tilde{a}_k$ to the left-hand side, where these shifted polynomials are indeed available. This is the philosophy of the next manipulations of the formulas.

We use the notation ω as before, but we add an index to know which block it refers to: $\omega_n = \nu_n + \alpha_{n+1}^- + 1$ (this is the degree of the NOP of block n). We now rewrite the different updating schemes.

The expression (4.6) can be reformulated as

$$za_{\nu_n+i}(z) = a_{\nu_n+i+1} \text{ for } i = 0, \ldots, \alpha_{n+1}^- - 1. \tag{4.10}$$

The update (4.7) for $j = 0$ gives

$$z(a_{\omega_n-1} + a_{\omega_{n-1}-1}p_n) = a_{\omega_n} \tag{4.11}$$

(recall $d_0 = 1$ and $\tilde{a}_n = z^{\alpha_{n+1}^-+1}a_n^1 = za_{\omega_n-1}$).
For $j > 0$ in (4.7), use $d_j(z) = zd_{j-1}(z) + d_j(0)$ to get

$$a_{\omega_n+j}(z) = z\left(z^{j-1}\tilde{a}_{n-1}(z) + \tilde{a}_n(z)d_{j-1}(z)\right) + \tilde{a}_n(z)d_j(0).$$

The expression between big brackets is a_{ω_n+j-1}. With the definition of $\tilde{a}_n$, we can bring this into the form

$$z\left(a_{\omega_n+j-1}(z) + a_{\omega_n-1}d_j(0)\right) = a_{\omega_n+j}(z) \tag{4.12}$$

Finally, the update (4.8) will be rearranged. First note that the polynomial r in (4.9) is in fact the polynomial $d_{\alpha^+_{n+1}}$ that one would get by continuing (4.7) one step further. So, for simplicity of notation, set $d_{\alpha^+_{n+1}-1}(z) = t(z)$, so that $r(z) = zt(z) + r(0)$. We then get (recall $q_n(z) = s(z) + z^{\alpha^-_{n+1}+1}r(z)$)

$$\begin{aligned}
a^1_{n+1}(z) &= z^{\alpha^+_{n+1}}\tilde{a}_{n-1}(z)p_n + a^1_n(z)z^{\alpha^-_{n+1}+1}r(z) + a^1_n(z)s(z) \\
&= z\left(z^{\alpha^+_{n+1}-1}\tilde{a}_{n-1}(z)p_n + \tilde{a}_n(z)t(z)\right) + \tilde{a}_n(z)r(0) + a^1_n(z)s(z) \\
&= z\left(a_{\omega_n+\alpha^+_{n+1}-1}(z) + a_{\omega_n-1}(z)r(0)\right) + a^1_n(z)s(z)
\end{aligned}$$

($\tilde{a}_n = za_{\omega_n-1}$). This can be rearranged into the form

$$z\left(a_{\nu_{n+1}-1}(z) + a_{\omega_n-1}(z)r(0)\right) = a^1_{n+1}(z) - a^1_n(z)s(z). \tag{4.13}$$

Collecting now the relations (4.10–4.13) into one matrix relation, we get

$$F_nA_nU_n = A_nV_n.$$

F_n is still the companion matrix for the (monic) polynomial a^1_{n+1} of degree ν_{n+1} and A_n is the unit upper triangular matrix whose columns are the coefficient vectors of the polynomials a_k for $k = 0, \ldots, \nu_{n+1} - 1$. The unit upper Hessenberg matrix T_n is decomposed as $T_n = V_nU_n^{-1}$ where now V_n is a simplified unit upper Hessenberg and U_n is unit upper triangular. Puzzling the pieces together will reveal that U_n is block bi-diagonal

$$U_n = \begin{bmatrix} U_{00} & U_{01} & & & \\ & U_{11} & U_{12} & & \\ & & \ddots & \ddots & \\ & & & \ddots & U_{n-1,n} \\ & & & & U_{nn} \end{bmatrix}$$

with $U_{n-1,n}$ a typical off diagonal block of size $\alpha_n \times \alpha_{n+1}$ having the form

$$U_{n-1,n} = \begin{matrix} \\ \alpha^-_n \\ 1 \\ \alpha^+_n \end{matrix} \begin{matrix} \begin{matrix} \alpha^-_{n+1} & 1 & \alpha^+_{n+1} \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & p_n & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

and U_{nn}, a typical diagonal block of size $\alpha_{n+1} \times \alpha_{n+1}$ of the form

$$U_{nn} = \begin{matrix} \\ \alpha^-_{n+1} \\ 1 \\ \alpha^+_{n+1} \end{matrix} \begin{matrix} \begin{matrix} \alpha^-_{n+1} & 1 & \alpha^+_{n+1} \end{matrix} \\ \begin{bmatrix} I & 0 & 0 \\ 0 & 1 & \dot{\mathrm{R}}^T \\ 0 & 0 & I \end{bmatrix} \end{matrix}$$

where $\dot{\mathrm{R}}$ is the reversed coefficient vector of the polynomial $r(z)$, i.e., with $\alpha = \alpha_{n+1}^+$: $r(z) = [1\ \dot{\mathrm{R}}^T][z^\alpha\ \cdots\ z\ 1]^T$. As for the matrix V_n, this is also block bi-diagonal

$$V_n = \begin{bmatrix} V_{00} & & & \\ V_{10} & V_{11} & & \\ & \ddots & \ddots & \\ & & V_{n,n-1} & V_{nn} \end{bmatrix}$$

with a typical block V_{nn} of size $\alpha_{n+1} \times \alpha_{n+1}$ of the form

$$V_{nn} = \begin{array}{c} 1 \\ \\ \alpha_{n+1}^- \\ \\ \alpha_{n+1}^+ \end{array} \overset{\begin{array}{ccc} \alpha_{n+1}^- & \alpha_{n+1}^+ & 1 \end{array}}{\begin{bmatrix} 0 & 0 & -s_0 \\ & & -s_1 \\ I & 0 & \vdots \\ & & -s_{\alpha_{n+1}^-} \\ 0 & I & 0 \end{bmatrix}}$$

which is in fact the companion matrix for the polynomial $z^{\alpha_{n+1}} + s(z)$.

The subdiagonal blocks are zero everywhere except for the right top element which is 1.

We can now derive the following determinant formula for a_{n+1}^1:

$$\begin{aligned} a_{n+1}^1(z) &= \det(zI - F_n) \\ &= \det(zI - A_n V_n U_n^{-1} A_n^{-1}) \\ &= \det(zI - V_n U_n^{-1}) \\ &= \det(zU_n - V_n) \end{aligned}$$

because U_n and A_n are unit upper triangular. The latter relation gives a determinant expression in terms of the simple matrices U_n and V_n.

4.6 Krylov subspace iteration

The general setting specializes to the Toeplitz case when $Q = P^{-1}$. It seems that this is not a very practical situation because usually P^{-1} will not be available, except of course when $P^{-1} = P^H$, i.e., when P is a unitary matrix. However, in this case the situation is back to the normal positive definite case.

Theorem 4.6 *Suppose $P = Q^{-1}$ is unitary and $x_0 = y_0$. Then M is positive semidefinite. All the blocks are of size 1 and the normalized Hessenberg matrix $\tilde{T}_N = D_N^{1/2} T_N D_N^{-1/2}$ is unitary.*

Proof. Let us introduce the normalized matrices $\tilde{X}_N = \hat{X}_N D^{-1/2}$. Since $Q^H = P$, we have that $x_k = y_k$ and thus $M = X^H X$ is Hermitian positive semidefinite. This implies that all leading submatrices are regular until the rank of M is reached. Thus if M has rank N, then the columns of X_N are linearly independent and $N = \operatorname{rank} X_N = \operatorname{rank} M_N$. From the general Theorem 2.6 we also know that the column space of X_N is P-invariant and also its orthogonal complement is P-invariant (the eigenvectors of a unitary matrix are orthogonal). This is sufficient to deduce from $\tilde{X}_N^H \tilde{X}_N = I_N$ and $\tilde{X}_N^H P \tilde{X}_N = \tilde{T}_N$, that $P\tilde{X}_N = \tilde{X}_N \tilde{T}_N$

and $\tilde{X}_N^H P = \tilde{T}_N \tilde{X}_N^H$. Combining the latter two gives $\tilde{X}_N \tilde{T}_N^{-1} = P^{-1}\tilde{X}_N = P^H \tilde{X}_N = \tilde{X}_N \tilde{T}_N^H$. Multiplying from the left with $\tilde{X}_N^H$ gives $\tilde{T}_N^{-1} = \tilde{T}_N^H$, hence $\tilde{T}_N$ is unitary. □

If x_0 has only components along N eigenvectors of P, then X_N will generate the space of these N eigenvectors and the eigenvalues of $\tilde{T}_N$, which are also the eigenvalues of T_N, will give the corresponding eigenvalues. Since the moment matrix is positive semidefinite of rank N, we can associate it with a discrete measure with N mass points on the complex unit circle. The orthogonal polynomials will be the corresponding Szegő polynomials. In this case it is more appropriate to use a Szegő/Levinson type of coupled recurrence relations. Indeed, the Hessenberg matrix $\tilde{T}_N$ is unitary and it can be easily parametrized in terms of Schur parameters (or reflection coefficients). However taking a finite approximation $\tilde{T}_n$, $n < N$ we get eigenvalues which are zeros of the nth Szegő polynomial and these are inside the unit circle. Thus $\tilde{T}_n$ should be modified by forcing the last reflection coefficient to have modulus 1, in which case also the approximants will give eigenvalues on the complex unit circle. This has been worked out by Gragg and his coworkers [19, 20, 1]. It has also connections with discrete least squares approximation and matrix orthogonal polynomials for a discrete measure [34, 35, 36, 11]. A look ahead version of the Levinson and Schur algorithm can be found in [25, 15].

We remark however that the previous row recursion is an alternative for the most general Toeplitz case, which has the advantage that the eigenvalue problem $\det(zI - T_n) = 0$ for T_n can be replaced by a generalized eigenvalue problem $\det(zU_n - V_n) = 0$ with a particularly simple structure for U_n and V_n.

4.7 Two point Padé approximation

The rational approximants that can be obtained for the Toeplitz case are described as follows. Recall

$$\begin{aligned} g^+(z) &= \mu_0 + \mu_1 z + \mu_2 z^2 + \cdots \\ -g^-(z) &= \mu_{-1} z^{-1} + \mu_{-2} z^{-2} + \cdots \end{aligned}$$

From $M[\nu_n, \nu_n + 1]\mathrm{A}_n^1 = [0, \ldots, 0]^T$, it follows that we may define

$$\mathrm{C}_n^1 = \begin{bmatrix} \mu_0 & 0 & & 0 \\ \mu_1 & \mu_0 & & \\ & & \ddots & \vdots \\ \mu_{\nu_n - 1} & & \cdots \ \mu_0 & 0 \end{bmatrix} \mathrm{A}_n^1 = -\begin{bmatrix} 0 & \mu_{-1} & \mu_{-2} & \cdots & \mu_{-\nu_n} \\ & 0 & \mu_{-1} & & \\ & & \ddots & & \\ & & & 0 & \mu_{-1} \end{bmatrix} \mathrm{A}_n^1.$$

This C_n^1 is the coefficient vector of a polynomial of degree at most $\nu_n - 1$ which is obtained as the polynomial part of $g^-(z)a_n^1(z)$. This is the so called polynomial of the second kind and we can find it as

$$c_n^1(z) = \mu\left(\gamma^-(\zeta)[a_n^1(z) - a_n^1(\zeta)]\right) = \left\langle \gamma_*^-(\zeta), a_n^1(z) - a_n^1(\zeta)\right\rangle$$

where $\gamma_*^-(\zeta) = \overline{\gamma^-(1/\overline{\zeta})}$. Hence

$$g^-(z)a_n^1(z) - c_n^1(z) = -r_n^-(z) = -\sigma_n z^{-1-\alpha_{n+1}^-} + \text{lower order terms} \tag{4.14}$$

$$g^+(z)a_n^1(z) - c_n^1(z) = r_n^+(z) = \rho_n z^{\nu_n + \alpha_{n+1}^+} + \text{higher order terms.} \tag{4.15}$$

We can easily check from the recurrence that $a_n^1(0) \neq 0$, so that

$$\begin{aligned} g^-(z) - c_n^1(z)/a_n^1(z) &= *z^{-\nu_n-1-\alpha_{n+1}^-} + \text{lower order terms} \\ g^+(z) - c_n^1(z)/a_n^1(z) &= *z^{\nu_n+\alpha_{n+1}^+} + \text{higher order terms.} \end{aligned}$$

The rational approximant $g_n = c_n^1/a_n^1$ has a total of $2\nu_n$ parameters and fits in the series $g^\pm$ together a total of $2\nu_n + \alpha_{n+1} - 1$ coefficients. It is therefore called a two point Padé approximant for the two series $g^\pm$.

4.8 Schur algorithm and continued fractions

In the classical situation of a positive definite Hermitian Toeplitz matrix, the Levinson algorithm computes the inverse Choleski factorization of the moment matrix and the Schur algorithm computes the Choleski factorization itself. Both are related to the coupled recurrence relation for Szegő OPs on the unit circle. Since we worked here with the row (i.e., three term) recurrence relation for the ROPs only, we only find the L factor in the LDU factorization of the Toeplitz matrix. So, by a Schur algorithm, we shall mean here an algorithm that computes this L factor recursively. To derive such a Schur algorithm, we note that in the recurrence for the ROPs, we need to find the parameters p_n, and the coefficients of the polynomials q_n. The coefficients of the other polynomials used in the recurrence for the internal polynomials are contained in the $r(z)$-part of q_n. To compute p_n and q_n we need the values of

$$\left\{ \begin{array}{ll} \langle 1, z^l \tilde{a}_{n-1} \rangle, & l = 0,1,\ldots,\alpha_{n+1}^+ \\ \langle 1, z^l \tilde{a}_n \rangle, & l = 0,1,\ldots,\alpha_{n+1}^+ \end{array} \right\} \quad \text{for the } r(z) \text{ part of } q_n(z).$$

$$\left\{ \begin{array}{ll} \left\langle z^{\nu_{n-1}+\alpha_n^+ + l}, a_{n-1}^1 \right\rangle, & l = 0,1,\ldots,\alpha_{n+1}^- \\ \left\langle z^{\nu_n+\alpha_{n+1}^+ + l}, a_n^1 \right\rangle, & l = 0,1,\ldots,\alpha_{n+1}^- \end{array} \right\} \quad \text{for the } s(z) \text{ part of } q_n(z)$$

Thus we shall need for $k = 0,1,\ldots$

$$\langle 1, z^l \tilde{a}_k \rangle, \quad l = 0,1,\ldots,\max\{\alpha_{k+1}^+, \alpha_{k+2}^+\}$$

and

$$\left\langle z^{\nu_k+\alpha_{k+1}^+ + l}, a_k^1 \right\rangle, \quad l = 0,1,\ldots,\max\{\alpha_{k+1}^-, \alpha_{k+2}^-\}.$$

These correspond to the numbers

$$[\mu_{-l-\alpha_{k+1}^- -1}, \ldots, \mu_{-l-\nu_{k+1}+\alpha_{k+1}^+}]\mathrm{A}_k^1 \quad \text{and} \quad [\mu_{\nu_k+\alpha_{k+1}^+ + l}, \ldots, \mu_{\alpha_{k+1}^+ + l}]\mathrm{A}_k^1$$

respectively. Note that these are the lth coefficients in the series

$$s_k^-(z) = z^{\alpha_{k+1}^- +1} r_k^-(z) = \sigma_k + *z^{-1} + \cdots \quad \text{and} \quad s_k^+(z) = z^{-\nu_k-\alpha_{k+1}^+} r_k^+(z) = \rho_k + *z + \cdots$$

respectively where $r_k^\pm$ are as defined in (4.14-4.15). These series are related to block LDU factorizations of M and $M^<$ as follows.

Lemma 4.7 *Define $R^+ = MA$. This matrix is block lower triangular and $r_k^+(z)$ is the generator for the first column in block column number k. Define $\check{A}$ by $[\,\check{\mathbf{a}}_0(z), \check{\mathbf{a}}_1(z), \ldots] = \mathbf{z}^T \check{A}$, with $\check{\mathbf{a}}_k(z)$ as defined in section 4.2. Then the matrix $R^- = \check{A}^T M^<$ is block upper triangular and $z^{\nu_k-1} r_k^-(1/z)$ is the generator for the first row of block row number k.*

Proof. Clearly R^+ is block lower triangular by the orthogonality of the ROPs. From the inner product form for the r_k^+ coefficients, the generating property of $r_k^+(z)$ follows immediately.
Note that Ă is block lower triangular by construction. Again, by the orthogonality of the $a_k(z)$ polynomials, it follows that R^- is block upper triangular. Moreover, from the inner product form of the coefficients of the $r_k^-(z)$ series, it should be clear that these coefficients appear on the first row of block row number k of R^- and the lemma follows. □

Thus, if we want to compute the factors R^+ and R^- without computing the ROPs explicitly and without the evaluation of the inner products, we should compute the $r_k^\pm(z)$ recursively, and this is perfectly possible as shown by the previous deduction. The relation (4.5) expresses for example that the polynomial parts of $s_n^-(z)d_j(z)$ and of $-p_n z^j s_{n-1}^-(z)$ agree. We have indeed that with the notation of (4.5)

$$s_n^-(z) = t_0 + t_1 z^{-1} + t_2 z^{-2} + \cdots \quad \text{and} \quad s_{n-1}^-(z) = t'_0 + t'_1 z^{-1} + t'_2 z^{-2} + \cdots$$

Therefore we can obtain the successive coefficients of the polynomials $d_j(z)$ by long division of $-p_n s_{n-1}^-(z)/s_n^-(z)$. The constant $-p_n$ is to ensure that the first coefficient is 1. Thus $d_j(z)$ is the polynomial part of $-z^j p_n s_{n-1}^-(z)/s_n^-(z)$. Recall that for $j = \alpha_{n+1}^+$, d_j is the r-part of q_n, so that this part is settled. The s-part of q_n can be obtained similarly by computing the first $\alpha_{n+1}^- + 1$ terms in the long division of $-p_n s_{n-1}^+(z)/s_n^+(z)$.

These relations form the basis of a Schur type algorithm. We do not repeat the details here. These are available in e.g., [7]. Without further proof we formulate

Theorem 4.8 *The block triangular factors R^+ and R^- of the previous lemma can be computed recursively by a division algorithm as sketched above. The a_k^1, c_k^1, r_k^- and r_k^+ all satisfy the same recurrence relation, which is associated with the continued fraction expansions of the form*

$$\frac{c_{n+1}^1(z)}{a_{n+1}^1(z)} = \sum_{k=0}^{n} \left.\frac{p_k z^{\alpha_k^- + \alpha_{k+1}^+ + 1}}{q_k(z)}\right|$$

whose convergents are the two point Padé approximants.

For an interpretation in terms of Schur complements, we need the left and right OPs and the corresponding residual series. This is more natural in the context of a generalization of the coupled recurrence relation, which we did not consider here. This requires the imbedding of Toeplitz matrices in the larger class of quasi Toeplitz matrices.

References

[1] G.S. Ammar, W.B. Gragg, and L. Reichel. Constructing a unitary Hessenberg matrix from spectral data. In G. Golub and P. Van Dooren, editors, *Numerical linear algebra, digital signal processing and parallel algorithms*, volume 70 of *NATO-ASI Series, F: Computer and Systems Sciences*, pages 385–395, Berlin, 1991. Springer.

[2] C. Brezinski. *Padé-type approximants and general orthogonal polynomials*, volume 50 of *Internat. Ser. of Numer. Math.* Birkhäuser Verlag, Basel, 1980.

[3] C. Brezinski. *Biorthogonality and its applications to numerical analysis*, volume 156 of *Lecture Notes in Pure and Appl. Math.* Marcel Dekker Inc., 1991.

[4] C. Brezinski. A unified approach to various orthogonalities. *Ann. Fac. Sciences Toulouse, Ser. 3*, 1(3):277–292, 1992.

[5] C. Brezinski. Formal orthogonal polynomials. *NA Digest*, 95(08), February, 19, 1995.

[6] C. Brezinski and J. van Iseghem. Vector orthogonal polynomials of dimension $-d$. In R. Zahar, editor, *Approximation and Computation*, volume 119 of *Internat. Ser. of Numer. Math.* Birkhäuser Verlag, 1995. To appear.

[7] A. Bultheel. Division algorithms for continued fractions and the Padé table. *J. Comput. Appl. Math.*, 6(4):259–266, 1980.

[8] A. Bultheel. *Laurent series and their Padé approximations*, volume OT-27 of *Oper. Theory: Adv. Appl.* Birkhäuser Verlag, Basel-Boston, 1987.

[9] A. Bultheel and M. Van Barel. Euclid, Padé and Lanczos, another golden braid. Technical Report TW188, Department of Computer Science, K.U. Leuven, April 1993.

[10] A. Bultheel and M. Van Barel. Some applications of the Euclidean algorithm. In D. Bainov and V. Covachev, editors, *Proceedings of the Second International Colloquium on Numerical Analysis*, pages 45–54, Zeist, The Netherlands, 1994. VSP Inl. Sci. Publ.

[11] A. Bultheel and M. Van Barel. Vector orthogonal polynomials and least squares approximation. *SIAM J. Matrix Anal. Appl.*, 16(3), 1995. To appear.

[12] J. Chun and T. Kailath. Generalized displacement structure for block-Toeplitz, Toeplitz-block, and Toeplitz-derived matrices. In G. Golub and P. Van Dooren, editors, *Numerical linear algebra, digital signal processing and parallel algorithms*, volume 70 of *NATO-ASI Series, F: Computer and Systems Sciences*, pages 215–236, Berlin, 1991. Springer.

[13] A. Draux. *Polynômes orthogonaux formels – applications*, volume 974 of *Lecture Notes in Math.* Springer, Berlin, 1983.

[14] R.W. Freund, G.H. Golub, and N.M. Nachtigal. Iterative solution of linear systems. *Acta Numerica*, 1:57–100, 1992.

[15] R.W. Freund and H. Zha. Formally biorthogonal polynomials and a look-ahead Levinson algorithm for general Toeplitz systems. *Linear Algebra Appl.*, 188/189:255–303, 1993.

[16] I. Gohberg, editor. *I. Schur methods in operator theory and signal processing*, volume 18 of *Oper. Theory: Adv. Appl.* Birkhäuser Verlag, Basel, 1986.

[17] W.B. Gragg. The Padé table and its relation to certain algorithms of numerical analysis. *SIAM Rev.*, 14:1–62, 1972.

[18] W.B. Gragg. Matrix interpretations and applications of the continued fraction algorithm. *Rocky Mountain J. Math.*, 4(2):213–225, 1974.

[19] W.B. Gragg. Positive definite Toeplitz matrices, the Arnoldi process for isometric operators and Gaussian quadrature on the unit circle. In E.S. Nikolaev, editor, *Numerical methods in linear algebra*, pages 16–32. Moscow university press, 1982. (In Russian).

[20] W.B. Gragg. The QR algorithm for unitary Hessenberg matrices. *J. Comput. Appl. Math.*, 16:1–8, 1986.

[21] W.B. Gragg and A. Lindquist. On the partial realization problem. *Linear Algebra Appl.*, 50:277–319, 1983.

[22] M.H. Gutknecht. The unsymmetric Lanczos algorithms and their relations to Padé approximation, continued fractions, the QD algorithm, biconjugate gradient squared algorithms and fast Hankel solvers. In *Proceedings of the Copper Mountain Conference on Iterative Methods*, 1990.

[23] M.H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related algorithms. Part I. *SIAM J. Matrix Anal. Appl.*, 13(2):594–639, 1992.

[24] M.H. Gutknecht. A completed theory of the unsymmetric Lanczos process and related algorithms. Part II. *SIAM J. Matrix Anal. Appl.*, 15:15–58, 1994.

[25] M.H. Gutknecht and M. Hochbruck. Look-ahead Levinson and Schur algorithms for non-Hermitian Toeplitz systems. Technical Report 93-11, IPS, ETH Zürich, 1993.

[26] G. Heinig and K. Rost. *Algebraic methods for Toeplitz-like matrices and operators.* Akademie Verlag, Berlin, 1984. Also Birkhäuser Verlag, Basel.

[27] T. Kailath. Time-variant and time-invariant lattice filters for nonstationary processes. In *Outils et modèles mathématiques pour l'automatique, l'analyse de systèmes et le traitement du signal, Vol. 2*, pages 417–464, Paris, 1982. Editions du CNRS.

[28] T. Kailath, S.Y. Kung, and M. Morf. Displacement ranks of matrices and linear equations. *J. Math. Anal. Appl.*, 68:395–407, 1979.

[29] H. Lev-Ari and T. Kailath. Triangular factorization of structured Hermitian matrices. In I. Gohberg, editor, *I. Schur methods in operator theory and signal processing*, Oper. Theory: Adv. Appl., pages 301–324, Basel, 1986. Birkhäuser Verlag.

[30] B.N. Parlett. Reduction to tridiagonal form and minimal realizations. *SIAM J. Matrix Anal. Appl.*, 13:567–593, 1992.

[31] A.H. Sayed, T. Constantinescu, and T. Kailath. Time-variant displacement structure and interpolation problems. *IEEE Trans. Automat. Control*, 39:960–975, 1994.

[32] I. Schur. Über Potenzreihen die im Innern des Einheitskreises Beschränkt sind I. *J. Reine Angew. Math.*, 147:205–232, 1917. See also [16, p.31-59].

[33] E.L. Stiefel. Kernel polynomials in linear algebra and their numerical applications. In *Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues*, volume 49 of *NBS Appl. Math. Series*, pages 1–22. NBS, 1958.

[34] M. Van Barel and A. Bultheel. A parallel algorithm for discrete least squares rational approximation. *Numer. Math.*, 63:99–121, 1992.

[35] M. Van Barel and A. Bultheel. Discrete linearized least squares approximation on the unit circle. *J. Comput. Appl. Math.*, 50:545–563, 1994.

[36] M. Van Barel and A. Bultheel. Orthogonal polynomial vectors and least squares approximation for a discrete inner product. *Electron. Trans. Numer. Anal.*, 3:1–23, 1995.

[37] J. van Iseghem. *Approximants de Padé vectoriels.* PhD thesis, Lille, 1987.

[38] J. van Iseghem. Vector orthogonal relations. Vector QD-algorithm. *J. Comput. Appl. Math.*, 19:141–150, 1987.

[39] H.S. Wall. *Analytic theory of continued fractions.* Van Nostrand, Princeton, 1948.

[40] P. Wynn. A general system of orthogonal polynomials. *Quart. J. Math.*, 18:81–96, 1967.