General block bi-orthogonal polynomials*

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

Workshop on Orthogonal Polynomials: non positive definite case
Rouen, France, April, 24–26, 1995

Abstract

We discuss formal orthogonal polynomials with respect to a moment matrix that has no structure whatsoever. In the classical case the moment matrix is often a Hankel or a Toeplitz matrix. We link this to block factorization of the moment matrix and its inverse, the block Hessenberg matrix of the recurrence relation, the computation of successive Schur complements and general subspace iterative methods. The Hankel case is elaborated as a special application of the general theory.

Keywords: formal orthogonal polynomials, Hankel matrix, Krylov subspace method.

1 Introduction

In a classical setting, formal orthogonal polynomials (OPs) are studied with respect to a linear functional μ. This functional is usually defined by its moments

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

and an "inner product" of two polynomials p and q is defined as

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

Such a situation immediately imposes a special structure on the moment matrix since $\langle z^i, z^j \rangle = \langle z^{i+j}, 1 \rangle = \langle 1, z^{i+j} \rangle = \mu_{i+j}$, we find that the moment matrix

$$M = \left[\left\langle z^i, z^j \right\rangle \right] = [\mu_{ij}] = [\mu_{i+j}]$$

is a Hankel matrix. A polynomial p_n of degree n (we suppose it is monic), is called a (true) orthogonal polynomial if it is uniquely defined by the orthogonality conditions $\left\langle z^k, p_n \right\rangle = 0$ for $k = 0, 1, \ldots, n-1$. Such polynomials are studied in the context of Padé approximation since they are related to the denominators of Padé approximants of the symbol of the Hankel matrix, i.e., of the formal series

*This research was performed as part of the project Lanczos of the National Fund for Scientific Research (NFWO), under contract #2.0042.93 and the project ROLLS under contract CHRX-CT93-0416.
Email to Adhemar.Bultheel@cs.kuleuven.ac.be

$f(z) = \sum_{k=0}^{\infty} \mu_k z^{-k-1}$. This situation is well understood and classical treatments are for example the books of Brezinski [2] and Draux [13]. For an arbitrary sequence $\{\mu_k\}$ of moments, it can not be guaranteed that all the (true) orthogonal polynomials exist, unless M is strongly regular, i.e., when all the principal leading submatrices of M are nonsingular. In general we can only find a subsequence of submatrices that is nonsingular and a corresponding subsequence of monic (true) OPs. Orthogonality has to be replaced by block-orthogonality. This means the following: let $M_0, M_1, \ldots$ be the sequence of successive nonsingular leading submatrices of M with M_k of size ν_{k+1}. Set $\alpha_k = \nu_k - \nu_{k-1}$, $(\nu_0 = 0)$, then M can be divided in blocks: $M = [M_{i,j}]$ with $M_{i,j}$ of size $\alpha_{i+1} \times \alpha_{j+1}$ and also the sequence of OPs will be divided in blocks: $[p_0, p_1, \ldots] = [\mathbf{p}_0, \mathbf{p}_2, \ldots]$ with

$$\mathbf{p}_k(z) = [p_k^0(z), \ldots, p_k^{\alpha_{k+1}-1}(z)] = [p_{\nu_k}(z), \ldots, p_{\nu_{k+1}-1}(z)].$$

Since in general, not all these polynomials exist, we shall have to relax the orthogonality (or the normalization condition, but we choose for the former). We require that the sequence of blocks $\mathbf{p}_k(z)$ is orthogonal, meaning that all the polynomials in the vector $\mathbf{p}_n$ are orthogonal to all polynomials in previous blocks, but not necessarily to all the other polynomials in the same block. This is called block orthogonality. Note that the first polynomial $p_n^0(z)$ in a block is a True Orthogonal Polynomial (as in the strongly regular case) and we shall therefore call it the TOP of block n. It exists and is uniquely defined.

The previous situation is a formal generalization of the real polynomials orthogonal with respect to a measure that lives on the real line (or an interval), thus polynomials orthogonal with respect to the inner product $\langle p, q\rangle = \int_a^b p(x)q(x)d\mu(x)$, $-\infty \leq a \leq b \leq \infty$. If μ is a finite positive Borel measure with infinitely many points of increase, then the moment matrix is a positive definite Hankel matrix and hence strongly regular. All the OPs exist and all blocks are of size 1.

A somewhat similar, yet different situation occurs for classical complex polynomials orthogonal with respect to a measure on the complex unit circle. In that case, the inner product is defined by

$$\langle p, q\rangle = \int_{-\pi}^{\pi} \overline{p(e^{i\theta})}q(e^{i\theta})d\mu(e^{i\theta}).$$

The corresponding moment matrix is

$$M = \left[\left\langle z^i, z^j\right\rangle\right] = [\mu_{j-i}],$$

a positive definite (strongly regular Hermitian) Toeplitz matrix. Thus the formal generalization here is to study arbitrary Toeplitz matrices. The situation is somewhat more complicated here because a Hankel matrix is by definition always symmetric, but a Toeplitz matrix can be non-symmetric and immediately we are confronted with the notion of bi-orthogonality, where the left and the right OPs need not be the same.

When considering orthogonality with respect to a positive definite measure, we still have a positive definite moment matrix, but the Hankel or Toeplitz structure is intimately related to the real line and the unit circle on which the measure lives. When it is supported on a more general set, the structure in the moment matrix will be lost.

In this paper, we shall give the formal generalization thereof and give algebraic aspects of formal orthogonal polynomials with respect to a completely arbitrary moment matrix.

Before we start our discussion, let us introduce the following notation. If p is a polynomial $p(z) = \sum_{i=0}^{n} p_i z^i$, then $\mathbf{p}$ will denote the stacking vector of its coefficients: $\mathbf{p} = [p_0, p_1, \ldots, p_n]^T$. If the context so requires, we suppose it to be extended with (infinitely many) zeros. For example, when $\mathbf{z} = [1, z, z^2, \ldots]^T$, then we shall write $p(z) = \mathbf{z}^T\mathbf{p}$. If the polynomial p is monic, we use a dot to delete the leading term, i.e., $\dot{p}(z) = p(z) - z^n$. Correspondingly, the stacking vector for $\dot{p}$ is $\dot{\mathbf{p}}$. The Frobenius or companion matrix of the polynomial p is

$$F(p) = \left[\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & \\ 1 & & & & \\ & 1 & & & -\dot{\mathbf{p}} \\ & & \ddots & & \\ & & & 1 & \end{array}\right], \quad p(z) = \dot{p}(z) + z^n.$$

By $Z = F(z^n)$ we denote the down shift matrix, but again here, its size will be assumed to be flexible, possibly infinite. Multiplying a matrix M with Z from the right will shift all the elements one place to the left (and deletes the first column). This is denoted as $M^< = MZ$. Similarly $Z^T M = M^\wedge$ is the matrix M shifted one place up.

We also need a notation for submatrices. If $M = [\mu_{ij}]$, then we set

$$M[k:l, m:n] = \begin{bmatrix} \mu_{km} & \cdots & \mu_{kn} \\ \vdots & & \vdots \\ \mu_{lm} & \cdots & \mu_{ln} \end{bmatrix}, \quad M[k,l] = M[0, k-1, 0:l-1],$$

while rows and columns are indicated by

$$M^{[i,*]} = M[i:i, 0:\infty] \quad \text{and} \quad M^{[*,j]} = M[0:\infty, j:j].$$

Consider the sequence of leading submatrices $M[k,k]$ of size k. Suppose that $M_{k-1} = M[\nu_k, \nu_k]$ for $k = 0, 1, \ldots$ denote the successive nonsingular elements in this sequence, all intermediate leading submatrices are singular. Correspondingly, the matrix M will be subdivided in blocks $M_{ij} = M[\nu_i : \nu_{i+1} - 1, \nu_j : \nu_{j+1} - 1]$ of size $\alpha_{i+1} \times \alpha_{j+1}$ where $\alpha_{k+1} = \nu_{k+1} - \nu_k$ (we set by definition $\nu_0 = 0$). The same sort of notation for submatrices and blocks will be used for all matrices that can be associated with the moment matrix M and can therefore be subdivided in similar blocks.

2 Formal orthogonality: general case

2.1 Moment matrix and generators

Consider the space $\mathbb{C}[w, z]$ of complex bivariate polynomials and define a linear functional μ on this space by the moments

$$\mu(w^i z^j) = \mu_{ij}, \quad i, j = 0, 1, 2, \ldots.$$

This defines the moment matrix $M = [\mu_{ij}]$ which is an arbitrary complex matrix. Next we define an "inner product" on the space of polynomials as $\langle p, q\rangle = \mu(\bar{p}(w)q(z))$ that is: the functional μ is applied to the product of the complex conjugate of the left polynomial with indeterminate replaced by w and the right polynomial with indeterminate z. Obviously $\langle \cdot, \cdot \rangle$ is completely defined by the moments $\mu_{ij} = \langle z^i, z^j \rangle = \mu(w^i z^j)$ and we have $\langle p, q\rangle = \mathbf{p}^H M \mathbf{q}$.

With the moment matrix M, we associate the formal bivariate series

$$M(x,y) = \mathbf{y}^H M \mathbf{x} = \sum_{i,j=0}^{\infty} \mu_{ij} \bar{y}^i x^j,$$

which is called the generating kernel or generator for M. If we write $(1-zx)^{-1}$ for the series $\sum_{k=0}^{\infty}(zx)^k$, then

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

is the generator for row i and similarly for the columns. Thus in terms of the inner product, the generator for M can be written as

$$M(x,y) = \left\langle (1-wy)^{-1}, (1-zx)^{-1} \right\rangle.$$

2.2 Block triangular factorizations

Let M be a matrix whose successive nonsingular leading submatrices are denoted as $M_0, M_1, \ldots$ with M_k of size ν_{k+1}, $k = 0, 1, \ldots$ and with block sizes $\alpha_{k+1} = \nu_{k+1} - \nu_k$. In that case, there exist unit upper triangular matrices A and B such that

$$B^H M A = D$$

with $D = \mathrm{diag}(D_{00}, D_{11}, \ldots)$ a block diagonal matrix with D_{kk} of size α_{k+1}. Since this relation also holds for finite sections, we have $\det M[\nu,\nu] = \det D[\nu,\nu]$ and thus are all D_{kk} nonsingular and all the leading submatrices of these blocks are singular. Note that $M^{-1} = AD^{-1}B^H$ (whenever M is invertible) which gives a block RDL factorization of M^{-1} (or of its leading submatrices). Since not all the leading submatrices of M are invertible, it is safer to leave this factorization in the form $B^H M A = D$. The block sizes in this factorization are fixed, but the blocks are not. All such factorizations are characterized as $(BD_b)^H M (AD_a) = D_b^H D D_a$ where D_a and D_b are block diagonals with unit upper triangular blocks. This shows that the columns ν_k, $k = 0, 1, \ldots$ of A and B are always the same. They are uniquely defined, being the first column in each block column of A and B. The other columns in the blocks are not uniquely defined. We can always factor the blocks D_{kk} as $[D_{kk}^b]^H D_{kk} [D_{kk}^a] = P_{kk}$ where D_{kk}^a and D_{kk}^b are unit upper triangular matrices and P_{kk} is a permuted diagonal matrix, which is in a sense the simplest possible form.

The computation of the factors B, A, D from M is essentially a block Gram-Schmidt algorithm and this factorization is immediately interpreted in terms of block orthogonal polynomials as we shall indicate below.

It is also possible to derive a block LDR factorization for the matrix M itself. Indeed, with $V = D^H B^{-1}$ and $U = DA^{-1}$, block upper triangular matrices, we can write

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

This is obtainable by a block Gaussian elimination applied to M. It is also related to the computation of successive Schur complements and as such, the block Gaussian algorithm is related to a block Schur algorithm. Also this shall be explained in a later section.

2.3 Block bi-orthogonal polynomials

We start with the interpretation of the inverse LDR factorization of M: $B^T M A = D$. Denote the kth block column as $A_{:,k}$ and set

$$\mathbf{a}_k(z) = \mathbf{z}^T A_{:,k} = [a_k^0(z), \ldots, a_k^{\alpha_{k+1}-1}(z)] = [a_{\nu_k}(z), \ldots, a_{\nu_{k+1}-1}(z)], \quad k = 0, 1, \ldots$$

and similarly for B. Then the previous factorization reads

$$\langle \mathbf{b}_i, \mathbf{a}_j \rangle = \left[\left\langle b_i^k, a_j^l \right\rangle \right]_{k=0,\ldots,\alpha_{i+1}-1}^{l=0,\ldots,\alpha_{j+1}-1} = \delta_{ij} D_{ij}.$$

We call the sequence of pairs of (monic) polynomials $\{(b_k, a_k)\}$ a sequence of block bi-orthogonal polynomials if the block sizes are dictated by the rank profile of the moment matrix, as explained before and if they satisfy the previous block biorthogonality condition with D_{jj} nonsingular. This means that a polynomial in block k will be orthogonal to all polynomials spanned by previous blocks, thus all polynomials of degree less than ν_k. The stacking vector of the first polynomial in each block satisfies

$$M_{n-1} \dot{\mathbf{a}}_n^0 = -[\mu_{0,\nu_n}, \ldots, \mu_{\nu_n-1,\nu_n}]^T. \tag{1}$$

Since M_{n-1} is invertible, it shows that it is uniquely defined. All the other polynomials in a block can be modified by adding multiples of previous polynomials in that block without violating the block bi-orthogonality conditions. Thus they are not unique. Therefore, we call the first polynomial of a block the True bi-Orthogonal Polynomial (TOP) of that block.

The other polynomials in a block are called inner polynomials. Obviously these are not uniquely defined by the previous definition.

The block two-sided modified Gram-Schmidt algorithm will compute the bi-orthogonal polynomials of block k by computing for $\nu = \nu_k, \nu_k + 1, \ldots$

$$a_\nu = z a_{\nu-1} - \sum_{i=0}^{k-1} \mathbf{a}_i D_{i,i}^{-1} \langle \mathbf{b}_i, z a_{\nu-1} \rangle$$

$$b_\nu = z b_{\nu-1} - \sum_{i=0}^{k-1} \mathbf{b}_i D_{i,i}^{-H} \langle z b_{\nu-1}, \mathbf{a}_i \rangle^H$$

and collect these polynomials in blocks $\mathbf{a}_k$ and $\mathbf{b}_k$. A block will be complete as soon as $D_{kk} = \langle \mathbf{b}_k, \mathbf{a}_k \rangle$ is invertible.

2.4 Schur complements

In the previous section, we were concerned with the factorization $B^H M A = D$ and in this section we consider the factorization $M = V^H D^{-1} U$. Like in the previous section, we denote the kth block row of U by $U_{k,:}$ and introduce the notation

$$\mathbf{u}_k(z) = [U_{k,:}]\mathbf{z} = [u_k^0(z), \ldots, u_k^{\alpha_{k+1}-1}(z)]^T = [u_{\nu_k}(z), \ldots, u_{\nu_{k+1}-1}(z)]^T, \quad k = 0, 1, \ldots$$

and similarly for V. It is then easy to see that the moment generator is

$$M(x, y) = \mathbf{y}^H M \mathbf{z} = \mathbf{y}^H V^H D^{-1} U \mathbf{x} = \sum_{k=0}^{m} \mathbf{v}_k(y)^H D_{kk}^{-1} \mathbf{u}_k(x).$$

The number m denotes the number of the last block one may construct. This m may be infinite, but it could also be finite, even when M is infinite, namely when the rank of M is finite, in which case block m will be infinitely large.

Recall that for a matrix

$$M = \begin{bmatrix} M_0 & M_{01} \\ M_{10} & M_{11} \end{bmatrix} \tag{2}$$

with M_0 invertible, the matrix

$$M_{\langle 0 \rangle} = M_{11} - M_{10} M_0^{-1} M_{01}$$

is called the Schur complement of M_0 in M. The Schur complement appears when computing a block Gaussian elimination

$$\begin{bmatrix} I & 0 \\ -M_{10}M_0^{-1} & I \end{bmatrix} \begin{bmatrix} M_0 & M_{01} \\ M_{10} & M_{11} \end{bmatrix} = \begin{bmatrix} M_0 & M_{01} \\ 0 & M_{\langle 0 \rangle} \end{bmatrix}.$$

Note also that this implies that $\det M = \det M_0 \det M_{\langle 0 \rangle}$. Another relation which can be immediately checked is

$$\begin{bmatrix} M_0 & M_{01} \\ M_{10} & M_{11} \end{bmatrix} = \begin{bmatrix} M_0 \\ M_{10} \end{bmatrix} M_0^{-1} \begin{bmatrix} M_0 & M_{01} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & M_{\langle 0 \rangle} \end{bmatrix}. \tag{3}$$

On the other hand

$$M(x,y) = \mathbf{v}_0(y)^H D_{00}^{-1} \mathbf{u}_0(x) + M'(x,y), \quad M'(x,y) = \sum_{k=1}^{m} \mathbf{v}_k(y)^H D_{kk}^{-1} \mathbf{u}_k(x),$$

thus

$$\begin{bmatrix} M_0 & M_{01} \\ M_{10} & M_{11} \end{bmatrix} = \begin{bmatrix} V_0^H \\ V_{01}^H \end{bmatrix} D_{00}^{-1} \begin{bmatrix} U_0 & U_{01} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & M'' \end{bmatrix}. \tag{4}$$

and therefore

$$\begin{bmatrix} M_0 \\ M_{10} \end{bmatrix} = \begin{bmatrix} V_0^H \\ V_{01}^H \end{bmatrix} D_{00}^{-1} U_0, \quad \begin{bmatrix} M_0 & M_{01} \end{bmatrix} = V_0^H D_{00}^{-1} \begin{bmatrix} U_0 & U_{01} \end{bmatrix}$$

and $V_0^H D_{00}^{-1} U_0 = M_0$. This leads to the conclusion that the first terms in the right hand sides of (3) and (4) are the same and thus $M_{\langle 0 \rangle} = M''$ or

$$M_{\langle 0 \rangle}(x,y) = (x\bar{y})^{-\alpha_1} M'(x,y) = (x\bar{y})^{-\alpha_1} [M(x,y) - \mathbf{v}_0(y)^H D_{00}^{-1} \mathbf{u}_0(x)].$$

This procedure can be repeated to compute Schur complements of Schur complements, so that one arrives at

$$M^{k+1}(x,y) = (x\bar{y})^{-\alpha_{k+1}} [M^k(x,y) - (x\bar{y})^{-\nu_k} \mathbf{v}_k(y)^H D_{kk}^{-1} \mathbf{u}_k(x)]$$

with $M^0 = M$,

$$M^k = \begin{bmatrix} M_0^k & M_{01}^k \\ M_{10}^k & M_{11}^k \end{bmatrix}$$

where M_0^k is the smallest invertible leading submatrix of M^k and $M^{k+1} = M_{\langle 0 \rangle}^k$ is the Schur complement of M_0^k in M^k. Taking out the leading zeros in $\mathbf{u}_k(z)$ and $\mathbf{v}_k(z)$ explicitly, we define

$$\mathbf{h}_k(z) = z^{-\nu_k} \mathbf{v}_k(z) \quad \text{and} \quad \mathbf{g}_k(z) = z^{-\nu_k} \mathbf{u}_k(z),$$

so that

$$M^{k+1}(x,y) = (x\bar{y})^{-\alpha_{k+1}} [M^k(x,y) - \mathbf{h}_k(y)^H D_{kk}^{-1} \mathbf{g}_k(x)].$$

The block triangular factors U and V can be computed from M by a block Gaussian elimination procedure without pivoting. To be more precise, U is the result of row elimination and V^H the result of column elimination. Doing both at the same time could be termed a block two-sided Gaussian elimination with the same effect as a block Gauss-Jordan diagonalization procedure. The procedure described above to compute successive Schur complements was also used by Jacobi to reduce a quadratic form. Therefore the transform $M^k \rightarrow M^{k+1}$ is sometimes called a Jacobi transform of M^k. Thus the Jacobi procedure is a method to compute a block LDR factorization of a matrix. It can be described as follows: start with $M^0 = M$, then for $k = 0, 1, \ldots$

Find the smallest invertible M_0^k in M^k
  let its size be α_{k+1}
Let $\mathbf{h}_k^H$ = the first α_{k+1} columns of M^k
Let $\mathbf{g}_k$ = the first α_{k+1} rows of M^k
Set $M^{k+1} = M^k - \mathbf{h}_k^H [M_0^k]^{-1} \mathbf{g}_k$

The Schur algorithm does exactly this, but stores and updates the Schur complements in factored form. Suppose there exists a scalar function $\sigma(x, y)$ such that $\sigma(x, y)M(x, y)$ is separable, in the sense that G and H can be found such that

$$\sigma(x, y)M(x, y) = H(y)^H G(x).$$

For any arbitrary factorization $M = H^H G$, we can of course apply this with $\sigma(x, y) = 1$, and $G(z)$ and $H(z)$ the generators for G and H respectively.

The factor $\sigma(x, y)$ plays a role when the matrix M has a certain structure, like for example when it has a low displacement structure. Let us briefly motivate this by the example of a Hankel matrix. Thus let us suppose that $\mu_{ij} = \mu_{i+j}$ are real numbers. In that case $\nabla M = ZM - MZ^T$ has an extremely simple structure, expressed by the generator relation

$$(\nabla M)(x, y) = (\bar{y} - x)M(x, y) = \bar{y}h(\bar{y}) - xh(x) \tag{5}$$

where $h(z) = \sum_{k=0}^{\infty} \mu_k z^k$ is the generator for the first column of M. Thus in this case, we have

$$\sigma(x, y) = \bar{y} - x, \quad H(y)^H = [\bar{y}h(\bar{y}) \quad -1], \quad G(x) = [1 \quad xh(x)]^T.$$

For a Toeplitz matrix $M = [\mu_{ij}]$ with $\mu_{ij} = \mu_{j-i}$, it is checked similarly that

$$\sigma(x, y) = 1 - x\bar{y}, \quad H(y)^H = [1 \quad -h(\bar{y})], \quad G(x) = [g(x) \quad 1]^T \tag{6}$$

with $h(z) = -\sum_{k=0}^{\infty} \mu_{-k} z^k$ and $g(z) = \sum_{k=1}^{\infty} \mu_k z^k$ the generators for the first column and the first row respectively.

More generally $\sigma(x, y)$ will stand for a certain structure which underlies the moment matrix. If we are treating a completely arbitrary matrix, we can not assume any structure and thus, in our case, $\sigma(x, y) = 1$. To return to the LDR factorization $M = V^H D^{-1} U$, we could consider computing $M = H^H G$ with $G = U$ and $H^H = V^H D^{-1}$, but any other choice will work as well. The Schur algorithm factors the Schur complements as $M^k = H_k^H G_k$ and computes updates of H_k and G_k. We shall only describe the first step, since all subsequent steps are repetitions thereof. So assume

$$M = H^H G, \quad H = \begin{bmatrix} H_{00} & H_{01} \\ H_{10} & H_{11} \end{bmatrix}, \quad G = \begin{bmatrix} G_{00} & G_{01} \\ G_{10} & G_{11} \end{bmatrix}. \tag{7}$$

with G_{00} and H_{00} of size α_1. First of all, we remark that G_{00} and H_{00} can always be assumed to be nonsingular.

Lemma 2.1 *Let M be as in (2) with M_0 invertible and let $M = \tilde{H}^H \tilde{G}$, then there exists a permutation matrix P such that with $H = P\tilde{H}$ and $G = P\tilde{G}$ the relation (7) holds with H_{00} and G_{00} invertible.*

Proof. Since $M_0 = [\tilde{H}_{00}^H \quad \tilde{H}_{10}^H][\tilde{G}_{00}^H \quad \tilde{G}_{10}^H]^H$ has full rank α_1, also $[\tilde{H}_{00}^H \quad \tilde{H}_{10}^H]$ and $[\tilde{G}_{00}^H \quad \tilde{G}_{10}^H]$ should have rank α_1. Thus, there exists a permutation matrix P such that $G = P\tilde{G}$ with G_{00} invertible. Then

$$QG = \hat{G} = \begin{bmatrix} G_{00} & G_{01} \\ 0 & \hat{G}_{11} \end{bmatrix}, \quad Q = \begin{bmatrix} I & 0 \\ \beta & I \end{bmatrix}, \quad \beta = -G_{10}G_{00}^{-1}.$$

Set

$$\hat{H} = Q^{-H}H = \begin{bmatrix} H_{00} & H_{01} \\ \hat{H}_{10} & \hat{H}_{11} \end{bmatrix}$$

then

$$\tilde{H}^H \tilde{G} = H^H G = (Q^{-H}H)^H (QG) = \hat{H}^H \hat{G}$$

and thus $M_0 = H_{00}^H G_{00}$. Since G_{00} and M_0 are invertible, also H_{00} is invertible. □

In view of this lemma, we assume henceforth that G_{00} and H_{00} are invertible. The Schur algorithm brings G and H simultaneously to a block upper triangular form (like the matrix $\hat{G}$ above). Therefore define

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

The matrix $S = I + \alpha^H \beta$ is nonsingular since

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

and because $\det(I + \beta\alpha^H) = \det(I + \alpha^H\beta)$, also $T = I + \beta\alpha^H$ is invertible. Let $S = S_L S_R$ and $T = T_L T_R$ be two arbitrary factorizations (with square invertible factors), then we are ready to define

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

It is directly verified that $\theta_G G$ and $H^H \theta_H$ have the form

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

while

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

Thus we have

$$M(x,y) = \mathbf{y}^H M \mathbf{x} = H(y)^H G(x) = H(y)^H \theta_H \theta_G g(x) = \tilde{H}(y)^H \tilde{G}(x).$$

Define now $\mathbf{g}_0(z)$ and $G_1(z)$ by

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

and similarly for $\tilde{H}$, then we get

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

Thus we have obtained the factorization $M_{\langle 0 \rangle}(x,y) = H_1(y)^H G_1(x)$ for the Schur complement while at the same time we obtained block rows and columns (the ones generated by $\mathbf{h}_0(z)$ and $\mathbf{g}_0(z)$) in a block LDR factorization of M.

2.5 Reproducing kernels

When a is a polynomial of degree less than ν_{n+1}, then it can be expanded in terms of the right or the left block OPs:

$$\begin{aligned} a(x) &= \sum_{k=0}^{n} \mathbf{a}_k(x)\mathbf{c}_k, \quad \mathbf{c}_k = D_{kk}^{-1} \langle \mathbf{b}_k, a \rangle \in \mathbb{C}^{\alpha_{k+1} \times 1} \\ \overline{a(y)} &= \sum_{k=0}^{n} \mathbf{d}_k^H \mathbf{b}_k(y)^H, \quad \mathbf{d}_k^H = \langle a, \mathbf{a}_k \rangle D_{kk}^{-1} \in \mathbb{C}^{1 \times \alpha_{k+1}}. \end{aligned}$$

Thus, if M_n is the leading submatrix of M of size ν_{n+1} and $K_n = M_n^{-T}$, then

$$K_n(x,y) = \mathbf{x}^T A_n D_n^{-1} B_n^H \bar{\mathbf{y}} = \sum_{k=0}^{n} \mathbf{a}_k(x) D_{kk}^{-1} \mathbf{b}_k(y)^H.$$

It follows from the last two relations that

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

Therefore, $K_n(x,y)$ is called the reproducing kernel for the space of polynomials of degree less than ν_{n+1}. We have the following determinant representation.

Theorem 2.2 *Let $K_n(x,y) = M_n^{-T}(x,y)$ be a reproducing kernel, then with $\nu+1 = \nu_{n+1}$,*

$$K_n(x,y) = -\frac{1}{\det M_n} \det \begin{bmatrix} & & & & 1 \\ & & & & \bar{y} \\ & & M_n & & \vdots \\ & & & & \bar{y}^\nu \\ 1 & x & \cdots & x^\nu & 0 \end{bmatrix}$$

Proof. Applying the relation $\det M = \det M_0 \det M_{\langle 0 \rangle}$ for the determinant of the large matrix above shows that it equals

$$\det M_n (0 - [1 \; x \; \cdots \; x^\nu] M_n^{-1} [1 \; y \; \cdots \; y^\nu]^H) = -K_n(x,y) \det M_n,$$

which proves the theorem. □

2.6 Hessenberg matrices

The block two-sided Gram Schmidt procedure generated the right OPs for $\nu_k \le \nu \le \nu_{k+1} - 1$ by a formula of the form

$$a_\nu(z) = z a_{\nu-1}(z) - \sum_{i=0}^{k-1} \mathbf{a}_i(z) \mathbf{t}_i^\nu, \quad \mathbf{t}_i^\nu = D_{ii}^{-1} \langle \mathbf{b}_i(z), z a_{\nu-1}(z) \rangle .$$

Thus

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

with T_{ik} of size $\alpha_{i+1} \times \alpha_{k+1}$. The diagonal block T_{kk} has the form of a companion matrix (its last column is the vector $\mathbf{t}_k^{\nu_{k+1}}$). The subdiagonal block $T_{k+1,k}$ is all zero except for a 1 in the right upper corner. The other blocks have no special structure.

We repeat this for $k = 0, 1, \ldots$, so that in terms of the unit upper triangular matrix A whose columns generate the right OPs, we can write this as $ZA = AT$

where Z is the down-shift matrix and T is a unit upper Hessenberg matrix whose kth block column is described above.

Combining the relations $M^< = MZ$, $B^H M A = D$ and $ZA = AT$, we find

$$B^H M^< A = B^H M Z A = B^H M A T = DT = J$$

where J is a block upper Hessenberg matrix, being the product of a block diagonal and an upper Hessenberg.

If we restrict this to matrices of size ν_{k+1}, we should rewrite this as

$$F_n A_n = A_n T_n$$

with $F_n = F(a^0_{n+1})$ the companion matrix of the TOP for block $n+1$, and $T_n = T[\nu_{n+1}, \nu_{n+1}]$ again unit upper Hessenberg with blocks T_{ij} as described above ($T_{ij} = 0$ for $i > j+1$). The relation (1), expressing the orthogonality of the TOP a^0_{n+1} gives

$$M_n F_n = M_n^<$$

where $M_n^< = M^<[\nu_{n+1}, \nu_{n+1}]$ (i.e., first shift and then truncate, not the other way around). Thus, similar to the relations for infinite matrices, we have for finite sections the relations $M_n^< = M_n F_n$, $F_n A_n = A_n T_n$ and $B_n^H M_n A_n = D_n$, which can be combined to give

$$B_n^H M_n^< A_n = B_n^H M_n F_n A_n = B_n^H M_n A_n T_n = D_n T_n = J_n. \tag{8}$$

As a direct consequence, we get the following determinant expressions for the TOP a^0_{n+1}.

Theorem 2.3 *The right TOP $a^0_{n+1}(z)$ for block $n+1$ is given by*

$$\begin{aligned} a^0_{n+1}(z) &= \det(zI_n - F_n) \\ &= \det M_n^{-1} \det(zM_n - M_n^<) \\ &= \det(zI_n - T_n) \\ &= \det M_n^{-1} \det \begin{bmatrix} & & & \mu_{0,\nu+1} \\ & & M_n & \vdots \\ & & & \mu_{\nu,\nu+1} \\ 1 & z & \cdots & z^{\nu+1} \end{bmatrix}, \quad \nu+1 = \nu_{n+1}. \end{aligned} \tag{9}$$

Proof. The first equality is evident, F_n being the companion matrix of the TOP. For the second relation, we multiply the matrix with M_n from the right and use $M_n F_n = M_n^<$. For the third one, we multiply the matrix of the second one from the left with B_n^H and from the right with A_n (both have determinant 1) and use $B_n^H M_n A_n = D_n$, $B_n^H M_n^< A_n = D_n T_n$ and $\det M_n = \det D_n$. The last line is proved similar to what has been done in the proof of Theorem 2.2 to find that the determinant of the big matrix equals

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

By relation (1) also the last equality follows. □

2.7 Rational approximation

We mentioned before that in the case where M is Hankel, the OPs are related to denominators of Padé approximants. Also in the general situation, we have some rational approximants. The generalization is that in the Hankel case, there is only one series namely the series $h(z)$ of (5). Therefore, all the approximating power is applied to this one series, leading to a genuine Padé approximant. A similar situation occurs for the Toeplitz case, where there are two series g and h as in (6) to characterize the moments, and the approximating ability of the rational form is distributed over the two series, giving rise to two-point Padé approximants. In the general case, all the rows (and all the columns) have generators which are completely independent from each other. Thus one may expect that the rational forms distribute their approximating power over all these series. For example the TOP $a^0_{n+1}(z)$ is the denominator of rational approximants for the series generating the first ν_{n+1} rows. The approximation is in a simultaneous Padé sense, i.e., in each of these series $\nu_{n+1}+1$ coefficients will be matched. If we go from n to $n+1$, there will be α_{n+1} extra rows and columns involved. Thus $a^0_{n+2}(z)$ will be the denominator of simultaneous Padé approximants, matching α_{n+1} extra coefficients of the series already approximated at the previous stage, but also approximating α_{n+1} new series corresponding to the α_{n+1} new rows that are considered. Let us make this more concrete.

We shall analyse the role of the right OPs in approximating the generators for the rows of M. Of course, a dual setup can be given for the left OPs and the generators of the columns. As we already mentioned, for $i = 0, 1, \ldots$, let $M^{[i,*]}(z) = e_i^T M\mathbf{z}$ be the generator for row i. Define the shifted series $g^{[i]}(z) = z^{-1}M^{[i,*]}(z^{-1})$. Thus

$$g^{[i]}(z) = \sum_{k=0}^{\infty} \frac{\mu_{ik}}{z^{k+1}} = \left\langle x^i, (z-x)^{-1} \right\rangle, \quad i = 0, 1, \ldots \tag{10}$$

These will be the series to be approximated. The approximants are described by

$$\begin{aligned} g_n^{[i]}(z) &= e_i^T M_n (zM_n - M_n^<)^{-1} M_n e_0 \\ &= \sum_{k=0}^{\infty} \frac{\mu_{ik}^n}{z^{k+1}}, i = 0, 1, \ldots, \nu_{n+1} - 1. \end{aligned} \tag{11}$$

The latter summation serves as a definition of μ_{ik}^n. The approximating property is described in the next theorem.

Theorem 2.4 *Let $g^{[i]}(z)$ and $g_n^{[i]}(z)$ be as defined above and let us use the notations A_n, B_n, D_n, T_n and V_n as before. Then*

$$\begin{aligned} g_n^{[i]}(z) &= e_i^T B_n^{-H} D_n (zI_n - T_n)^{-1} e_0 = e_i^T V_n^H (zI_n - T_n)^{-1} e_0 \\ &= c_{n+1}^{[i]}(z)/a_{n+1}^0(z). \end{aligned} \tag{12}$$

The numerator $c_{n+1}^{[i]}(z)$ is given by

$$c_{n+1}^{[i]}(z) = \left\langle x^i, \frac{a_{n+1}^0(z) - a_{n+1}^0(x)}{z-x} \right\rangle \tag{13}$$

$$= \det M_n^{-1} \det \begin{bmatrix} M_n e_0 & M_n^< \\ 0 & u_n(z) \end{bmatrix} \tag{14}$$

with $u_n(z)$ the polynomial part of

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

The $g_n^{[i]}(z)$ approximate the $g^{[i]}(z)$ for $i = 0, 1, \ldots, \nu_{n+1} - 1$ in the sense that

$$\mu_{ik} = \mu_{ik}^n \quad \text{for } k = 0, 1, \ldots, \nu_{n+1}. \tag{15}$$

This also implies that $c_{n+1}^{[i]}(z)$ is the polynomial part of $g^{[i]}(z)a_{n+1}^0(z)$.

Proof. Using (8) in (11) gives (12).

Using a formal expansion in (12) gives

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

Because $A_n^{-1} F_n A_n = T_n$, we have $T_n^k e_0 = A_n^{-1} F_n^k e_0$ so that

$$g_n^{[i]}(z) = e_i^T V_n^H A_n^{-1} \sum_{k=0}^{\infty} z^{-(k+1)} F_n^k e_0.$$

Now since $V_n^H A_n^{-1} = M_n$ and

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

it follows that

$$e_i^T M_n F_n^k e_0 = \mu_{ik}, \quad k = 0, 1, \ldots, \nu_{n+1}.$$

This proves (15).

With (10), we rewrite the right hand side of (13) as

$$\left\langle x^i, \frac{a_{n+1}^0(z) - a_{n+1}^0(x)}{z - x} \right\rangle = g^{[i]}(z) a_{n+1}^0(z) - \left\langle x^i, \frac{a_{n+1}^0(x)}{z - x} \right\rangle.$$

The left hand side is obviously a polynomial in z. Expanding $(z-x)^{-1} = \sum_0^\infty x^k z^{-(k+1)}$, we see that the second term on the right contains only negative powers of z. Thus equality is only possible if the left hand side is the polynomial part of $g^{[i]}(z)a_{n+1}^0(z)$ and the second term is the strictly proper part. Since (15) implies that the polynomial part of $g^{[i]}(z)a_{n+1}^0(z)$ and $g_n^{[i]}(z)a_{n+1}^0(z)$ are the same and the latter is $c_{n+1}^{[i]}(z)$ by definition, we have proved (13).

For (14), we note that

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

The first factor on the right is $\det M_n a_{n+1}^0(z)$ by (9), the second factor is $-g_n^{[i]}(z)$. Hence (14) follows. □

The previous theorem shows that we have approximated a vector of power series in z^{-1} by a vector of rational approximants with a common denominator. However, as n increases, not only do we match more coefficients in the series, but also the length of the vectors increases, i.e., more series are approximated by more rational functions.

It is an observation by B. Beckermann that these approximation properties can be interpreted by placing this in a framework of bivariate Padé approximation. Indeed,

$$G_n(z, w) = \sum_{i=0}^{\nu} g_n^{[i]}(z) w^{-(i+1)} = [w^{-1} \ \cdots \ w^{-\nu_{n+1}}] V_n^H (zI_n - T_n)^{-1} e_0,$$

will approximate $(zw)^{-1} M(z^{-1}, w^{-1})$ such that all the coefficients of $w^{-i} z^{-j}$ are matched in the rectangular grid $1 \le i \le \nu_{n+1}$, $1 \le j \le \nu_{n+1} + 1$.

2.8 Krylov subspace methods

The formal orthogonal polynomials can be related to Krylov subspace iteration in the following way. Given two arbitrary matrices P and Q of size p, we can construct the Krylov matrices

$$X = [x_0, x_1, \ldots] \quad \text{and} \quad Y = [y_0, y_1, \ldots]$$

with $x_k = P^k x_0$ and $y_k = [Q^H]^k y_0$ and x_0 and y_0 arbitrary. Consider $M = Y^H X = [\mu_{ij}]$ with $\mu_{ij} = y_0^H Q^i P^j x_0$ as a moment matrix with block indices ν_k and block sizes α_k. Note that $M = Y^H X$ provides us with a natural factorization $M(x,y) = Y(y)^H X(x)$, which was needed in the Schur algorithm. In this particular case

$$X(x) = X\mathbf{x} = (I - xP)^{-1} x_0 \quad \text{and} \quad Y(y) = Y\mathbf{y} = (I - yQ^H)^{-1} y_0$$

With the M we associate all the quantities we had before like A, B, D, V, U, T, J etc. A first observation is that with $\hat{Y} = YB$ and $\hat{X} = XA$,

$$\hat{Y}^H \hat{X} = B^H M A = (YB)^H (XA) = D.$$

This means that $\hat{Y}$ and $\hat{X}$ are block bi-orthogonal vectors. We also have that

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

Let X_n and Y_n denote the first ν_{n+1} columns of Y and X. The column spaces of X_n and Y_n are called Krylov subspaces (of order ν_{n+1}) and denoted as $\mathcal{X}_n$ and $\mathcal{Y}_n$. The matrix

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

is the representation of the projector onto $\mathcal{X}_n$ along $\mathcal{Y}_n$. The approximate solution of the system $Px = b$ with P very large can be obtained by solving the projected system $\Pi_n P z_n = \Pi_n P \hat{X}_n v_n = \Pi_n b$. This means that we project all the equations, which is obtained by multiplying with Π_n, and moreover restrict the solution z to a vector $z_n \in \mathcal{X}_n$, which is obtained by setting $z_n = \hat{X}_n v_n$. Thus we have to solve $\Pi_n P \hat{X}_n v_n = \Pi_n b$. Since

$$\Pi_n P \hat{X}_n = \hat{X}_n (\hat{Y}_n^H \hat{X}_n)^{-1} \hat{Y}_n^H P \hat{X}_n = \hat{X}_n D_n^{-1} D_n T_n = \hat{X}_n T_n,$$

we have to solve a small system $T_n v_n = c_n$, where $c_n = D_n^{-1} \hat{Y}_n^H b$, which eventually leads to the approximate solution $z_n = \hat{X}_n v_n$.

In the case where $Q = P = P^H$ is an Hermitian matrix, then M will be a Hankel matrix, and this procedure is the classical Lanczos iterative process for the solution of a system. In that case there is no need for block orthogonality. In the case where $Q = P$, but not necessarily Hermitian, we still have a Hankel matrix which need not be strongly regular. The previous setting generalizes this idea where the Krylov space $\mathcal{Y}_n$ can be chosen completely independent from the space $\mathcal{X}_n$.

Not only systems, but also eigenvalue problems can be solved in this way. Therefore set $b = \lambda x$, then obviously, one has to solve $T_n v_n = \lambda v_n$ as a small problem to approximate the large problem. The spectrum of the Hessenberg matrix T_n are called Ritz values and will be approximations for the eigenvalues of P.

Let $r_x = \operatorname{rank} X$, $r_y = \operatorname{rank} Y$, $r = \operatorname{rank} M \leq m = \min\{r_x, r_y\}$. Then we can compute the block OPs up to block n for which $\nu_{n+1} = r$. Since we have reached the largest invertible submatrix M_n at that point, the next block will be infinite and the procedure breaks down. We are then in the situation of the following theorem.

Theorem 2.5 *Suppose with the previous notation that $r = m = r_x$, then $\sigma(T_n) \subset \sigma(P)$ and $P\mathcal{X}_n \subset \mathcal{X}_n$. If $\hat{v}$ is an eigenvector of T_n, then $v = \hat{X}_n\hat{v}$ is an eigenvector of P, and the eigenvalues of T_n are zeros of $a^0_{n+1}(z)$.*

Proof. If X_n has reached its maximal rank say $X_n = [x_0, \ldots, x_{r-1}]$ with $r = \nu_{n+1}$, then it is obvious that $PX_n = [x_1, \ldots, x_r]$ generates vectors in the range of X_n, (x_r has got to be in the range of X_n ir the rank r is maximal) which proves the P-invariance of $\mathcal{X}_n$. Since $\hat{X}_n$ also generates the same space, there must be a matrix $\hat{T}_n$ such that $P\hat{X}_n = \hat{X}_n\hat{T}_n$. This $\hat{T}_n$ is equal to T_n because

$$D_nT_n = \hat{Y}_nP\hat{X}_n = \hat{Y}_n\hat{X}_n\hat{T}_n = D_n\hat{T}_n$$

and D_n is invertible. Now if $T_n\hat{v} = \lambda\hat{v}$ then

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

showing that $v = \hat{X}_n\hat{v}$ is an eigenvector of P. Finally, it has been shown before that $a^0_{n+1}(z) = \det(zI - T_n)$ so that the eigenvalues of T_n are zeros of $a^0_{n+1}(z)$. □

We note that if $r < m = r_x$, unlike the more classical case, then the eigenvalues of T_n may not be related to the eigenvalues of P nor to the eigenvalues of Q.

3 Case study: the Hankel matrix

As an example of the general theory, we apply it in the case where M is a Hankel matrix. Since it is usual to leave out the complex conjugate for this case, we shall do so in this section.

More details can be found in the literature [17, 9, 10, 11, 16, 1, 20, 22, 19, 24].

3.1 The moment matrix and its generator

Suppose $M = [\mu_{ij}]$ with $\mu_{ij} = \mu_{i+j}$, then

$$\mu_{ij} = \left\langle z^i, z^j \right\rangle = \mu_{i+j} = \left\langle z^{i+j}, 1 \right\rangle = \left\langle 1, z^{i+j} \right\rangle,$$

we do not really need to define the inner product through a linear functional on the space of bivariate polynomials, but it is sufficient, as one usually does, to define the functional on the space of univariate polynomials by $\mu_k = \mu(z^k)$, $k = 0, 1, \ldots$ and the inner product

$$\langle p, q \rangle = \mu(p(z)q(z)).$$

Another property which immediately follows is that

$$\langle zf(z), g(z) \rangle = \langle f(z), zg(z) \rangle,$$

which translates in the displacement property for the moment matrix as

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

or equivalently

$$\nabla M = ZM - MZ^T = \begin{bmatrix} 0 & -\mu_0 & -\mu_1 & \cdots \\ \mu_0 & & & \\ \mu_1 & & M^< - M^\wedge & \\ \vdots & & & \end{bmatrix} = \left[\begin{array}{c|c} 0 & -\mathsf{h} \\ \hline \mathsf{h}^T & 0 \end{array}\right].$$

Thus the moment generator for $M(x,y) = \mathbf{y}^T M \mathbf{x}$ satisfies

$$(y-x)M(x,y) = yh(y) - xh(x) \quad \text{with} \quad h(z) = \mathbf{hz} = \sum_{k=0}^{\infty} \mu_k z^k.$$

Because the Hankel matrix is symmetric, it also follows that the left and right OPs can be chosen to be the same. Thus $A = B$.

3.2 Block orthogonality

The Hankel structure implies that within a block of OPs, we may choose the inner polynomials to be shifted versions of the TOP for that block. Thus

$$\mathbf{a}_n(z) = [a_n^0(z), \ldots, a_n^{\alpha_{n+1}-1}(z)]$$

with $a_n^i(z) = z^i a_n^0(z)$ and the size of the blocks is fixed by

$$\alpha_{n+1} = \min\{\alpha : \left\langle z^{\nu_n+\alpha-1}, a_n^0(z) \right\rangle \neq 0\}.$$

As a consequence of this choice for the inner polynomials, the bottom block A_{nn} will be upper triangular Toeplitz and the bottom block R_{nn} in $M_n A_{.,n} = R_{.,n}$ will be lower triangular Hankel. The diagonal block $D_{nn} = A_{nn}^T R_{nn}$ is therefore lower triangular Hankel. It is well known, and it is easily proved, that the recurrence relation between the successive TOPs is given by

$$a_{n+1}^0(z) = [a_n^0(z)\ z a_n^0(z)\ \cdots\ z^{\alpha_{n+1}} a_n^0(z)]\mathbf{q}_n + a_{n-1}^0(z) p_n$$

where p_n and $\mathbf{q}_n = [\dot{\mathbf{q}}_n^T\ 1]^T$ satisfy

$$\left[\begin{array}{c|ccc|c} r_0^{n-1} & & & & r_0^n \\ r_1^{n-1} & & & r_0^n & r_1^n \\ \vdots & & \iddots & \vdots & \vdots \\ r_{\alpha_{n+1}}^{n-1} & r_0^n & \cdots & r_{\alpha_{n+1}-1}^n & r_{\alpha_{n+1}}^n \end{array}\right] \left[\begin{array}{c} p_n \\ \\ \dot{\mathbf{q}}_n \\ \\ 1 \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \end{array}\right]$$

with $r_i^k = \langle z^{\nu_{k+1}+i-1}, a_k^0(z)\rangle$, whence $r_0^k = \rho_k \neq 0$. The lower triangular middle block of the matrix in this system is R_{nn}. Thus $p_n = -\rho_n/\rho_{n-1}$ and $\dot{\mathbf{q}}_n$ is found by solving the system

$$R_{nn}\dot{\mathbf{q}}_n = -\mathbf{r}_n \tag{16}$$

with $\mathbf{r}_n = [r_1^n, \ldots, r_{\alpha_{n+1}}^n]^T + [r_1^{n-1}, \ldots, r_{\alpha_{n+1}}^{n-1}]^T$. This leads to the recurrence

$$a_{n+1}^0(z) = a_n^0(z) q_n(z) + a_{n-1}^0(z) p_n$$

with $q_n(z) = \mathbf{z}^T \mathbf{q}_n$, a monic polynomial of degree α_{n+1}.

Since the recurrence relation involves only three successive TOPs, the general relation $ZA = AT$ with T unit upper Hessenberg simplifies in the sense that T is now also block tridiagonal. The diagonal blocks are companion matrices as before: $T_{nn} = F(q_n)$, the blocks $T_{n+1,n}$ are again zero, except for the right top element which is 1, and the blocks $T_{n-1,n}$ are also zero except for the right top element, which is $-p_n$. Thus the matrix T has a very simple and sparse structure, which of course allows to give a very efficient block Gram-Schmidt algorithm. We only have to orthogonalize a suitably shifted TOP to the last $\alpha_{k+1}+1$ powers of z, the

other orthogonality relations come for free. The block tridiagonal structure of T is transferred to the matrix $J = DT$. One has

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

with $\alpha = \alpha_{k+1}$ and $\gamma_0^k = \rho_k \neq 0$. The matrix T is the block generalization of the classical tridiagonal Jacobi matrix for the classical orthogonal polynomials and J is its symmetric version.

Multiplying the defining equation (16) for q_n from the left with A_{nn}^T we get

$$\begin{bmatrix} & & \gamma_0^n \\ & \iddots & \vdots \\ \gamma_0^n & \cdots & \gamma_{\alpha-1}^n \end{bmatrix} \dot{\mathbf{q}}_n = - \begin{bmatrix} \gamma_1^n \\ \vdots \\ \gamma_\alpha^n \end{bmatrix}.$$

The matrix of this system is D_{nn}. Thus this means that $q_n(z)$ is the TOP for block 1 of a Hankel matrix whose leading block is D_{nn}.

3.3 Reproducing kernel and Christoffel-Darboux relation

The reproducing kernel can be described by the following Christoffel-Darboux relation.

Theorem 3.1 *The reproducing kernel* $K_n(x,y) = \mathbf{x}^T M_n^{-1} \mathbf{y}$ *satisfies*

$$K_n(x,y) = \frac{a_{n+1}^0(x) a_n^0(y) - a_{n+1}^0(y) a_n^0(x)}{\rho_n (x-y)}$$

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

Proof. If $\mathbf{a}(x) = \mathbf{x}^T A_n$, then

$$x\mathbf{a}(x) = \mathbf{a}(x) T_n + e_\nu^T a_{n+1}^0(x), \quad (\nu = \nu_{n+1} - 1).$$

It follows that

$$\begin{aligned} xK_n(x,y) &= x\mathbf{a}(x) D_n^{-1} \mathbf{a}(y)^T \\ &= \mathbf{a}(x) T_n D_n^{-1} \mathbf{a}(y)^T + a_{n+1}^0(x) e_\nu^T D_n^{-1} \mathbf{a}(y)^T. \end{aligned}$$

Clearly $e_\nu^T D_{nn}^{-1} = [0 \ \cdots \ 0 \ \rho_n^{-1} \ 0 \ \cdots \ 0]$ with ρ_n^{-1} at position ν_n. Therefore

$$xK_n(x,y) = \mathbf{a}(x) T_n D_n^{-1} \mathbf{a}(y)^T + \frac{a_{n+1}^0(x) a_n^0(y)}{\rho_n}.$$

Similarly, one obtains

$$yK_n(x,y) = \mathbf{a}(x) T_n D_n^{-1} \mathbf{a}(y)^T + \frac{a_{n+1}^0(y) a_n^0(x)}{\rho_n}.$$

The difference of the latter two relations yields the desired Christoffel-Darboux relation. □

Applying the Christoffel-Darboux relation for the case where q_n is the TOP for block 1 for a Hankel matrix with leading block equal to D_{nn}, we get

$$D_{nn}^{-1}(x,y) = \frac{q_n(y) - q_n(x)}{(y-x)\rho_n}.$$

3.4 Padé approximation

In the notation of the general case, we have $g^{[0]}(z) = z^{-1}h(z^{-1})$ where $h(z)$ is the generator for the first row of M. In the sequel we shall denote this series as $g(z)$. Because in the Hankel case

$$g^{[i+1]}(z) = zg^{[i]}(z) - \mu_i = z^{i+1}g(z) - s_{i+1}(z)$$

with $s_{i+1}(z) = zs_i(z) + \mu_i$, $s_0 = 0$, it follows that $[0\ s_1\ \cdots\ s_{\nu_{n+1}}]$ is the polynomial part of $g(z)[1\ z\ z^2\ \cdots\ z^{\nu_{n+1}}]$ and this is the bottom row $[0\ u_n]$ in the matrix of (14).

For the Padé approximation property we have

Theorem 3.2 *The approximant*

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

is a Padé approximant for $g(z)$ at infinity.

Proof. By the general theory, we know that $\mu_k^n = \mu_k$ for $k = 0, \ldots, \nu = \nu_{n+1} - 1$. Thus

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

Because $e_0^T M_n F_n^{\nu} = e_\nu^T M_n$, we find that

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

and thus $\mu_k^n = \mu_k$ for $k = 0, \ldots, 2\nu+1$, which shows that $g_n(z)$ is a Padé approximant for $g(z)$ at infinity. □

3.5 The Chebyshev algorithm and Schur complements

If the Padé approximant is $g_n(z) = c_{n+1}(z)/a_{n+1}^0(z)$, then

$$g(z)a_{n+1}^0(z) = c_{n+1}(z) + r_{n+1}^0(z)$$

where $c_{n+1}(z)$ is the polynomial part of $g(z)a_{n+1}^0(z)$ and $r_{n+1}^0(z)$ is the strictly proper part. It is easily checked that $r_{n+1}^0(z)$ is generating the first column in block $n+1$ of the matrix $R = MA$. More precisely

$$\mathbf{z}^T R = [\mathbf{r}_0(z^{-1})\ \ \mathbf{r}_1(z^{-1})\ \ \cdots\ ], \quad \mathbf{r}_n(z^{-1}) = [r_n^0(z^{-1}), \ldots, r_n^{\alpha_{n+1}-1}(z^{-1})].$$

Thus

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

The recurrence for the OPs transfers to the residuals and it follows that

$$\frac{p_n r_{n-1}^0(z)}{r_n^0(z)} = -q_n(z) + \frac{r_{n+1}^0(z)}{r_n^0(z)}.$$

Therefore, $-q_n(z)$ is the polynomial part in the division $p_n r_{n-1}^0(z)/r_n^0(z)$. The constant p_n is chosen to make $q_n(z)$ monic. This procedure of generating the polynomial $q_n(z)$ and the constant p_n is the well known Chebyshev algorithm and is equivalent with the Euclidean algorithm for power series in z^{-1}. Since it generates the columns

of the matrix R, we find by $M_n = R_n D_n^{-1} R_n^T$ that it gives a block Cholesky factorization of M_n and it should therefore be equivalent with the Schur algorithm and the computation of Schur complements. This equivalence is what we shall demonstrate in the remainder of this section.

The Schur complement of a Hankel matrix is not Hankel. Therefore we need to embed the Hankel matrices in a larger class of quasi Hankel matrices. These are matrices whose generator satisfies

$$M(x,y) = \frac{G(y)^T \Sigma G(x)}{y-x} = \frac{g_2(y)g_1(x) - g_1(y)g_2(x)}{y-x}.$$

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

with $G(0) \neq 0$. It is no restriction of the generality to assume that $g_1(0) \neq 0$ and $g_2(0) = 0$. The Hankel matrices form a subset of the quasi Hankels, obtained by setting $G(z)^T = [1 \quad zh(z)]$. From the Christoffel-Darboux relation, it follows immediately that the inverse of a Hankel matrix is a quasi Hankel.

As we did in the general case, we take out the redundant zeros in $r_k^0(z)$ and define

$$h_k(z) = z^{-\nu_k} r_k^0(z^{-1}) = \rho_k z^{\alpha_{k+1}} + \text{higher order terms.}$$

In this notation, the Chebyshev algorithm translates into the recurrence

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

If we set

$$G_k(z)^T = [\rho_{k-1}^{-1} z^{-\alpha_k} h_{k-1}(z) \quad h_k(z)] \tag{17}$$

then it holds that

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

The Chebyshev algorithm is thus equivalent with an update of the factor $G_k(z)$. To make the link with Schur complements, we show

Theorem 3.3 *Using our previous notation we set $M^0 = M$ and denote the successive Schur complements as $M^{k+1} = M^k_{\langle 0 \rangle}$, then these Schur complements are quasi Hankel and their generators are*

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

with $G_k(z)$ as in (17).

Proof. Since this is obviously true for $k = 0$ because $h_0(z) = zh(z)$, $\alpha_0 = 0$ and $h_{-1} = \rho_{-1} = -1$, we only prove the induction step.

Since

$$M^{k+1}(x,y) = (xy)^{-\alpha_{k+1}}[M^k(x,y) - \mathbf{y}^T R_{[\cdot,k]} D_{kk}^{-1} R_{[\cdot,k]}^T \mathbf{x}],$$

with $R_{[\cdot,k]}$ the kth block column of R where the leading zeros corresponding to blocks $0, \ldots, k-1$ are left out. Because $R_{[\cdot,k]}$ is Hankel, we may write

$$\mathbf{y}^T R_{[\cdot,k]} = y^{-1} h_k(y)[1 \; y^{-1} \; \cdots \; y^{-\alpha_{k+1}+1}]$$

so that

$$\mathbf{y}^T R_{[\cdot,k]} D_{kk}^{-1} R_{[\cdot,k]}^T \mathbf{x} = (xy)^{-1} h_k(y) h_k(x) d_k(x,y)$$

with $d_k(x,y) = D_{kk}^{-1}(x^{-1}, y^{-1})$. From the observation at the end of section 3.3, we know that

$$(xy)^{-1}(y-x)d_k(x,y) = -(q_k(y^{-1}) - q(x^{-1}))/\rho_k$$

and by induction hypothesis $(y-x)M^k(x,y) = G_k(y)^T \Sigma G_k(x)$. Thus

$$\begin{aligned}(y-x)M^{k+1}(x,y) &= (xy)^{-\alpha_{k+1}}\{G_k(y)^T\Sigma G_k(x) + \rho_k^{-1} h_k(y)h_k(x)[q_k(y^{-1} - q_k(x^{-1})]\} \\ &= (xy)^{-\alpha_{k+1}}\{G_k(y)^T(\Sigma + \Omega)G_k(x)\}\end{aligned}$$

with $\Omega = \rho_k^{-1}[q_k(y^{-1}) - q_k(x^{-1})][0 \quad 1]^T[0 \quad 1]$. Furthermore, it can be checked that $\theta_k(y)^T \Sigma \theta_k(x) = (xy)^{-\alpha_{k+1}}(\Sigma + \Omega)$ and thus

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

and this proves the theorem. □

3.6 Lanczos algorithm

The Krylov subspace iteration method reduces to the Hankel case when $Q = P$ in the general setting. The entries of the Hankel matrices are $\mu_k = y_0^T P^k x_0$. The symbol of the Hankel matrix is $h(z) = y_0^T(I - zP)^{-1}x_0$. The recurrence relation for the OPs with Hankel moment matrix is the one used in the Lanczos method. Since the recurrence is between the TOPs of successive blocks, this makes it possible to overcome the so called breakdown problem in the Lanczos method. There is a vast literature on this and related method, and we shall not elaborate it here.

4 Other cases

Parallel to the Hankel case, it is possible to develop the analog for a Toeplitz matrix. The inner product can be defined by a linear functional defined on the space of Laurent polynomials by $\mu(z^k) = \mu_k$, $k = 0, \pm 1, \pm 2, \ldots$ and $\langle p, q\rangle = \mu(\bar{p}(z^{-1})q(z))$. A Toeplitz matrix also has the displacement property that $\Delta M = I - ZMZ^T$ has rank 2, which shows in the form of the generator, see (6)

$$(1 - x\bar{y})M(x,y) = g(x) - h(\bar{y})$$

with $-h(z)$ generating the first column and $\mu_0 + g(z)$ generating the first row of M. Toeplitz matrices are embedded in the larger class of quasi Toeplitz matrices having a generator satisfying

$$M(x,y) = \frac{H(y)^H \Sigma G(x)}{1 - x\bar{y}}, \quad G(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix}, H(y) = \begin{bmatrix} h_1(x) \\ h_2(x) \end{bmatrix}.$$

The inverse of a Toeplitz matrix is quasi Toeplitz. It is also called a Bezoutian. Also the Schur complements appearing in the block LDR factorization of M are quasi Toeplitz with leading regular block being Toeplitz.

The Toeplitz case is however considerably more complicated than the Hankel case because we loose the symmetry of the Hankel matrices. We thus have to deal with left and right OPs again. We also have the choice between a row recurrence, which

is again a general three term recurrence, and the Szegő-Levinson type recurrence which mixes left and right OPs in the recurrence. The Hessenberg matrix T_n is still Hessenberg, but has a simple characterization. For example for the row recurrence, it can be written as $V_n U_n^{-1}$ with V_n and U_n both simple block bi-diagonal matrices. For further details we refer to the literature [7, 8, 18, 23, 21, 15, 14, 12].

Brezinski has developed a general theory of orthogonality [3, 4]. In this case the bi-orthogonality is described by considering a sequence of functionals $\{\mu^k\}$ which is applied to the sequence of polynomials $\{z^k\}$. This corresponds to our setting by assigning a different functional to each row of the moment matrix. In the case where there is essentially only a vector of d different linear functionals and the corresponding moment matrix satisfies $\mu_{i,j+1} = \mu_{i+d,j}$ then the corresponding orthogonal polynomials are called vector orthogonal polynomials of dimension d. Obviously, when $d = 1$ we have a Hankel matrix and when $d = -1$ we have a Toeplitz matrix. When $|d| > 1$, this is related to the rational approximation of d series simultaneously. For more details we refer to the above citations and [5, 6].

References

[1] R.P. Brent, F.G. Gustavson, and D.Y.Y. Yun. Fast solution of Toeplitz systems of equations and computation of Padé approximants. *J. Algorithms*, 1:259–295, 1980.

[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., 1992.

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

[5] C. Brezinski and M. Redivo Zaglia. Orhogonal polynomials of dimension -1 in the non definite case. *Rend. Mat. Roma, Ser. VII*, 14:127–133, 1994.

[6] C. Brezinski and J. van Iseghem. Vector orthogonal polynomials of dimension $-d$. In R.V.M. Zahar, editor, *Approximation and Computation*, volume 119 of *Internat. Ser. of Numer. Math.*, pages 29–39. Birkhäuser Verlag, 1994.

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

[8] A. Bultheel and M. Van Barel. Formal orthogonal polynomials and rational approximation for arbitrary bilinear forms. Technical Report TW163, Department of Computer Science, K.U. Leuven, December 1992.

[9] A. Bultheel and M. Van Barel. Euclid, Padé and Lanczos, another golden braid. *SIAM Rev.*, 1993. In preparation.

[10] A. Bultheel and M. Van Barel. Formal orthogonal polynomials for an arbitrary moment matrix and Lanczos type methods. In J.D. Brown, M.T. Chu, D.C. Ellison, and R.J. Plemmons, editors, *Proceedings of the Lanczos International Centenary Conference*, pages 273–275. SIAM, 1994.

[11] 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.

[12] T. F. Chan and Per C. Hansen. A look-ahead Levinson algorithm for general Toeplitz systems. *IEEE Trans. Sig. Proc.*, 40(5):1079–1090, 1992.

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

[14] R.W. Freund. A look-ahead Schur-type algorithm for solving general Toeplitz systems. *Z. Angew. Math. Mech.*, June 1993. AT&T Bell Labs. Numerical Analysis Manuscript 93-09.

[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] R.W. Freund and H. Zha. A look-ahead strategy for the solution of general Hankel systems. *Numer. Math.*, 64:295–322, 1993.

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

[18] M. Gutknecht and M. Hochbruck. Look-ahead Levinson- and Schur-type recurrences in the Padé table. Technical Report IPS 94-12, IPS, Zürich, 1994.

[19] 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.

[20] 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.

[21] M.H. Gutknecht. Stable row recurrences for the Padé table and generically superfast lookahead solvers for non-Hermitian Toeplitz systems. *Linear Algebra Appl.*, 188/189:351–422, 1993.

[22] 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.

[23] M.H. Gutknecht and M. Hochbruck. Look-ahead Levinson and Schur algorithms for non-Hermitian Toeplitz systems. *Numer. Math.*, 70:181–228, 1995.

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