Formal orthogonal polynomials for an arbitrary moment matrix and Lanczos type methods*

Adhemar Bultheel† Marc Van Barel†

Abstract

We give a framework for formal orthogonal polynomials with respect to an arbitrary moment matrix. When the moment matrix is Hankel, this simplifies to the classical framework. The relation with Padé approximation and with Krylov subspace methods is given.

1 Formal block orthogonal polynomials

We consider a linear functional defined on the space of polynomials in two variables, defined by the moments $\mu_{ij} = \mu(w^i z^j)$, $i, j \in \mathbf{N}$. Let $M = [\mu_{ij}]$ be the (infinite) moment matrix, then for two polynomials $p(w) = \mathbf{wp}$ and $q(z) = \mathbf{zq}$ ($\mathbf{w} = [1, w, w^2, \ldots]$ and $\mathbf{z} = [1, z, z^2, \ldots]$, $\mathbf{p}, \mathbf{q} \in \mathbf{C}^{\infty\times 1}$), we define a formal inner product by $\langle p, q\rangle = \mu\left(p^*(w)q(z)\right) = \mathbf{p}^* M\mathbf{q}$. We call $g(\omega, \zeta) = \mu\left(1/[(\omega - w)(\zeta - z)]\right)$ the generator for the matrix M, since at least formally:

$$g(\omega, \zeta) = \sum_{i,j} \frac{\mu_{ij}}{\omega^{i+1}\zeta^{j+1}} = \sum_i \frac{g^{[i]}(\zeta)}{\omega^{i+1}}$$

with $g^{[i]}(\zeta) = \mu\left(w^i/(\zeta - z)\right) = \sum_j \mu_{ij}/\zeta^{j+1}$, the generator for row i.

Any infinite matrix M can be factored as $B^*MA = D$ where $D = \mathrm{diag}(D_{00}, D_{11}, \ldots)$ is a block diagonal matrix with D_{ii} regular and of minimal size α_{i+1}, while A and B are unit upper triangular. The block sizes are determined by the fact that the $\nu \times \nu$ leading submatrices $M_{\nu\nu}$ of M are regular for $\nu = \nu_1, \nu_2, \ldots$ exactly when $\nu_n = \sum_{k=1}^n \alpha_k$, and they are singular for all other values of ν. When M is strongly regular, then all leading submatrices of M are regular and hence all the block sizes are equal to 1.

Let us define the monic polynomials

$$\mathbf{a}(z) = [\mathbf{a}_0, \mathbf{a}_1, \ldots] = \mathbf{z}A \quad \text{and} \quad \mathbf{b}(w) = [\mathbf{b}_0, \mathbf{b}_1, \ldots] = \mathbf{w}B.$$

The $\mathbf{a}_i(z)$ is a $1 \times \alpha_{i+1}$ polynomial vector whose elements are denoted as $\mathbf{a}_i = [a_i^1, \ldots, a_i^{\alpha_{i+1}}]$ and similarly for $\mathbf{b}_i$. If $\langle \mathbf{b}_i, \mathbf{a}_j\rangle$ represents a matrix of size $\alpha_{i+1} \times \alpha_{j+1}$ with (k, l) entry $\left\langle b_i^k, a_j^l \right\rangle$, then the polynomials are block orthogonal in the sense that $\langle \mathbf{b}_i, \mathbf{a}_j\rangle = \delta_{ij} D_{ii}$. Note that a_n^1 is uniquely defined by the regularity of $M_{n-1} = M_{\nu_n,\nu_n}$, but the "internal" polynomials a_n^k for $k > 0$ are not. The block orthogonal polynomials can be computed by a block two-sided Gram-Schmidt procedure.

Obviously there must exist a unit upper Hessenberg matrix T such that $z\mathbf{a}(z) = \mathbf{a}(z)T$. If Z denotes the down-shift matrix, then this translates in the matrix relation $ZA = AT$. For similar reasons, there also exists a unit upper Hessenberg matrix S such that $ZB = BS$.

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

†Department of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, B-3001 Heverlee, Belgium.

Let $M^< = MZ$, that is the matrix M shifted one place to the left, then the matrix J appearing in

$$B^*M^<A = B^*MZA = B^*MAT = DT = J$$

will be a block upper Hessenberg matrix. Similarly, with $Z^*M = M^\wedge$, we find that K appearing in $B^*M^\wedge A = S^*D = K^*$ will be block lower Hessenberg.

2 The Hankel case and approximation

When M is a real Hankel matrix, we find by symmetry considerations that it is possible to choose $A = B$, hence, $S = T$ while $D = A^*MA$ is symmetric. Moreover $A^*M^<A = A^*M^\wedge A = DT = J$ implies that also J is symmetric, hence block tridiagonal. Also T will be block tridiagonal. The relation $ZA = AT$ or $z\mathbf{a}(z) = \mathbf{a}(z)T$ corresponds to a block version of the three-term recurrence relation and T is the corresponding Jacobi matrix. This recurrence is obtained by the Euclidean algorithm applied to the couple $(1, g(\zeta))$ where $g(\zeta) = \sum_k \mu_k/\zeta^{k+1}$, the generator for the Hankel matrix $M = [\mu_{i+j}]$. The Euclidean algorithm generates a continued fraction expansion $g(z) = \sum_k \frac{\beta_k}{\mid \pi_k(z)}$ where the β_k are constants and the π_k are monic polynomials of degree α_{k+1}. The n-th convergent $g_n(z)$ is a Padé approximant for $g(z)$ of type $(\nu_{n+1} - 1/\nu_{n+1})$ and it can be expressed as $g_n = c_{n+1}/a^1_{n+1}$ where a^1_{n+1} is the first component of $\mathbf{a}_{n+1}$ and c_{n+1} is the polynomial of the second kind defined by $c_{n+1}(\zeta) = \mu_H\left([a^1_{n+1}(\zeta) - a^1_{n+1}(z)]/[\zeta - z]\right)$ where μ_H is the linear functional defined on the space of univariate polynomials by $\mu_H(z^k) = \mu_k$.

For a general moment matrix M, these approximation properties can be generalized as follows (see [2])

$$g^{[i]}(\zeta) - g^{[i]}_n(\zeta) = O_+(\zeta^{\nu_{n+1}+1}), \quad \text{for} \quad i = 0, 1, \ldots, \nu_{n+1} - 1$$

where $g^{[i]}$ is the generator for row i of M and $g^{[i]} = c^{[i]}_{n+1}/a^1_{n+1}$ with a^1_{n+1} as above and

$$c^{[i]}_{n+1}(\zeta) = \mu\left(w^i[a^1_{n+1}(\zeta) - a^1_{n+1}(z)]/[\zeta - z]\right).$$

When the index n for a matrix indicates that the leading sub-matrix of size ν_{n+1} is taken, then we may also express these approximants as

$$g^{[i]}_n(\zeta) = e_i^* M_n(\zeta M_n - M_n^<)^{-1} M_n e_0 = e_i^* B_n^{-*} D_n(\zeta I_n - T_n)^{-1} e_0$$

where e_i denotes the ith column of the unit matrix I_n. We also have

$$a^1_{n+1}(\zeta) = \det[M_n^{-1}(\zeta M_n - M_n^<)] = \det(\zeta I_n - T_n).$$

3 Krylov subspace methods

Consider the Krylov matrices

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

where the x_0 and y_0 are arbitrary vectors and P and Q some operators. We set $M = Y^*X = [\mu_{ij}]$ where $\mu_{ij} = y_0^* Q^i P^j x_0$. 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 $\hat{Y}^*\hat{W} = B^*MA = D$ while

$$\hat{Y}^* P \hat{X} = B^* M^< A = DT = J \quad \text{and} \quad \hat{Y}^* Q X = B^* M^\wedge A = S^* D = K^*.$$

In the special case that $P = Q$, M will be a Hankel matrix and the Euclidean algorithm will correspond to the Lanczos method for nonsymmetric matrices [3]. Note that the generator for

M is $g(\omega,\zeta) = y_0^*(\omega I - Q)^{-1}(\zeta I - P)^{-1}x_0$, which is rational and of finite degree in ζ and ω when the vector space is finite dimensional.

Set $r_x = \text{rank } X$, $r_y = \text{rank } Y$ and $r = \text{rank } M \leq m = \min\{r_x, r_y\}$. Then the Krylov method will break down when $\nu_{n+1} = r$. 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 $v = \hat{X}\hat{v}$ is an eigenvector of P: $Pv = \lambda v$. The eigenvalues λ are the zeros of the polynomial a_{n+1}^1. Similar results hold for $r = m = r_y$. When $r < m$, however, then it may happen that we do not find any eigenvalue not of P and not of Q.

4 Toeplitz and other cases

For a Toeplitz moment matrix, the fast versions of the orthogonalization procedure are often formulated as generalizations of the Levinson or Schur algorithm. However also generalizations of the three term recurrence relation can be obtained. This is sometimes referred to as a row recurrence, since one then moves in a Padé table along a row. The series approximated can be written as a formal Laurent series $g(\zeta) = \sum_{-\infty}^{\infty} \mu_k \zeta^k$, when $M = [\mu_{i-j}]$. So one gets a Laurent-Padé approximant or, when the series g is split as $g = g_+ - - g_-$, with g_+ having the nonnegative and $-g_-$ the negative powers, then one may reformulate it as a two-point Padé approximant.

In general, the row recurrence will require some Hessenberg matrix T with a special structure which is rather complicated to describe. However by bringing a part of the right hand side in $\mathbf{za} = \mathbf{a}T$ to the left hand side, we obtain a much simpler description [2]. In matrix terms, this means that we write $T = VU^{-1}$, which produces $\mathbf{za}U = \mathbf{a}V$, where U is upper block bidiagonal and V is lower block bidiagonal, with very simple blocks. With respect to the Krylov subspace interpretation, this means that in order to have a Toeplitz matrix M, we should have $Q = P^{-1}$. The eigenvalue problem for P is then reduced to a generalized eigenvalue problem since $a_{n+1}^1(\zeta) = \det(\zeta U_n - V_n)$.

The gap between a general M and Toeplitz or Hankel, can be filled by matrices of low displacement rank as introduced by Kailath and his coworkers. More recently, Brezinski and coworkers introduced a general concept of orthogonality [1], which is similar to ours. Their concept of vector orthogonal polynomials of dimension d corresponds to M being block Hankel ($d > 0$) or Toeplitz ($d < 0$) with $|d| \times 1$ blocks. For $d = \pm 1$, one gets the classical Hankel and Toeplitz cases. See e.g., [4].

References

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

[2] 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.

[3] ——, *Euclid, Padé and Lanczos, another golden braid*, Technical Report TW188, Department of Computer Science, K.U. Leuven, April 1993.

[4] J. van Iseghem, *Vector orthogonal relations. Vector QD-algorithm*, J. Comp. Appl. Math., 19 (1987), pp. 141–150.