Vector orthogonal polynomials and least squares approximation

Adhemar Bultheel and Marc Van Barel

Report TW 184, February 1993

Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

Vector orthogonal polynomials and least squares approximation

Adhemar Bultheel and Marc Van Barel

Report TW 184, February 1993

Department of Computer Science, K.U.Leuven

Abstract

We describe an algorithm for complex discrete least squares approximation, which turns out to be very efficient when function values are prescribed in points on the real axis or on the unit circle. In the case of polynomial approximation, this reduces to algorithms proposed by Rutishauser, Gragg, Harrod, Reichel, Ammar and others. The underlying reason for efficiency is the existence of a recurrence relation for orthogonal polynomials, which are used to represent the solution. We show how these ideas can be generalized to least squares approximation problems of a more general nature.

Keywords : vector orthogonal polynomials, discrete least squares approximation, QR factorization
MSC : Primary : 41A20 Secondary : 65F25, 65D05, 65D15, 30E20

VECTOR ORTHOGONAL POLYNOMIALS AND LEAST SQUARES APPROXIMATION*

ADHEMAR BULTHEEL AND MARC VAN BAREL†

Abstract. We describe an algorithm for complex discrete least squares approximation, which turns out to be very efficient when function values are prescribed in points on the real axis or on the unit circle. In the case of polynomial approximation, this reduces to algorithms proposed by Rutishauser, Gragg, Harrod, Reichel, Ammar and others. The underlying reason for efficiency is the existence of a recurrence relation for orthogonal polynomials, which are used to represent the solution. We show how these ideas can be generalized to least squares approximation problems of a more general nature.

Key words. orthogonal vector polynomials, discrete least squares.

AMS subject classifications. 41A20, 65F25, 65D05, 65D15, 30E10

1. Introduction. Let $\{z_k\}_{k=0}^m$ be a set of complex nodes and $\{w_k^2\}_{k=0}^m$ a set of positive weights (let us assume that $w_k > 0$).
We shall first solve the problem of finding the least squares polynomial approximant in the space with positive semi-definite inner product

$$\langle f, g\rangle = \sum_{k=0}^{m} \overline{f(z_k)} w_k^2 g(z_k). \tag{1}$$

Note that this is a positive definite inner product for the space of vectors $(f(z_0), \dots, f(z_m))$ representing the function values at the given nodes.
The polynomial $p \in \mathbb{P}_n$ of degree $n \leq m$ which minimizes

$$\|f - p\|, \quad \text{with} \quad \|v\| = \langle v, v\rangle^{1/2}$$

(note that this is a semi-norm) can be found as follows. Find a basis $\{\varphi_0, \dots, \varphi_n\}$ for $\mathbb{P}_n$ which is orthonormal with respect to $\langle\cdot,\cdot\rangle$. The solution p is the generalized Fourier expansion of f with respect to this basis, truncated after the term of degree n. An algorithm that solves the problem will compute implicitly or explicitly the orthonormal basis and the Fourier coefficients. As we shall see in the following sections, we can reduce the complexity of such an algorithm by an order of magnitude when a "short recurrence" exists for the orthogonal polynomials. We shall consider the case where all the z_i are on the real line, in which case a three-term recurrence relation exists, and the case where all the z_i are on the complex unit circle, in which case a Szegő type recurrence relation exists.

The above mentioned discrete least squares problem is closely related to many other problems in numerical analysis. For example, consider the quadrature formula

$$\int_a^b w(x) f(x) dx \approx \sum_{k=0}^{m} w_k^2 f(z_k)$$

where $w(x)$ is a positive weight for the real interval $[a, b]$. We get a Gaussian quadrature formula, exact for all polynomials of degree $2m+1$ by a special choice of the nodes

*This research was supported by the HCM project ROLLS of the European Community under contract ERBCHRXCT930416

†Department of Computer Science, K.U.Leuven, Celestijnenlaan 200A, B-3001 Leuven, Belgium

and weights. The nodes z_k are the zeros of the $(m+1)$st orthogonal polynomial with respect to $\langle f, g\rangle = \int_a^b f(x)g(x)w(x)dx$. These are also the eigenvalues of the truncated Jacobi matrix which is associated with this orthogonal system. The weights w_i^2 are proportional to q_{i0}^2 where q_{i0} is the first component of the corresponding eigenvector.

Another link can be made with inverse spectral problems. These come in several forms. One variant is precisely the inverse of the previous quadrature problem: find the Jacobi matrix, when its eigenvalues and the first entries of the normalized eigenvectors are given.

We shall call the computation of the quadrature formula or the eigenvalue decomposition of the Jacobi matrix direct problems, while the inverse spectral problem, and the least squares problem will be called inverse problems.

For a survey of inverse spectral problems, we refer to Boley-Golub [5]. One of the methods mentioned there is the Rutishauser-Gragg-Harrod algorithm. This algorithm can be traced back to Rutishauser [14] and was adapted by Gragg-Harrod [11] with a technique of Kahan-Pal-Walker for chasing a nonzero element in the matrix.

For a discrete least squares interpretation of these algorithms we refer to Reichel [12]. When the z_i are not on the real line, but on the unit circle, similar ideas lead to algorithms discussed by Ammar and He [4] and Ammar-Gragg-Reichel [2] for the inverse eigenvalue problem and to Reichel-Ammar-Gragg [13] for a least squares interpretation.

We shall first survey the general theory in the context of discrete least squares approximation where the z_k are arbitrary complex numbers in sections 2,3 and 4. In section 5, we shall explain how the complexity can be reduced with an order of magnitude when short recurrences exist.

The next step (section 6) is to generalize these results to the problem of minimizing

$$\min \sum_{k=0}^{m} |w_{0k}p_0(z_k) + \cdots + w_{\alpha k}p_\alpha(z_k)|^2 \tag{2}$$

where the $\{w_{0k}, \ldots, w_{\alpha k}\}_{k=0}^m$ are given complex numbers and the polynomials p_i of degree at most d_i, $i = 0, \ldots, \alpha$ have to be found, with the constraint that at least one of them is monic of strict degree.

When $\alpha = 1$, this generalization is related with rational approximation, in contrast with the previously described problem, which is related to polynomial approximation. We shall refer to the generalized problem as the matrix case, while the simpler polynomial case is referred to as the scalar case.

For the matrix case, we may distinguish between two levels of complication. When all the degrees d_i are equal, it will turn out (section 7) that the solution method can be described in terms of square matrix orthogonal polynomials of size $\alpha + 1$, and the previous theory of scalar orthogonal polynomials is readily generalized.

When not all the degrees d_i are equal, we are in the most general case that we shall consider here (section 8). The solution can be described now in terms of vector orthogonal polynomials, which allow to combine the scalar orthogonal polynomials of the first case and the matrix orthogonal polynomials of the second case which will show up both during the solution of the problem.

The break-down of the algorithm will only occur in the case of exact interpolation. This is discussed in section 9.

To avoid an unduly complicated notation, we shall mainly restrict ourselves in this paper to the case $\alpha = 1$, but the generalization to general $\alpha \geq 1$ should be obvious.

2. Polynomial least squares approximation. Discrete least squares approximation by polynomials is a classical problem in numerical analysis where orthogonal polynomials play a central role.
Given an inner product $\langle \cdot, \cdot \rangle$ defined on $\mathbb{P}_m \times \mathbb{P}_m$, the polynomial $p \in \mathbb{P}_n$ of degree at most $n \leq m$, which minimizes the error

$$\|f - p\|, \qquad p \in \mathbb{P}_n$$

is given by

$$p = \sum_{k=0}^{n} \varphi_k a_k, \qquad a_k = \langle f, \varphi_k \rangle$$

when the $\{\varphi_k\}_0^n$ form an orthonormal set of polynomials:

$$\varphi_k \in \mathbb{P}_k - \mathbb{P}_{k-1}, \qquad \mathbb{P}_{-1} = \varnothing, \qquad \langle \varphi_k, \varphi_l \rangle = \delta_{kl}.$$

The inner product we shall consider here is of the discrete form (1) where the z_i are distinct complex numbers.
Note that when $m = n$, the least squares solution is the interpolating polynomial, so that interpolation can be seen as a special case.
To illustrate where the orthogonal polynomials show up in this context, we start with an arbitrary polynomial basis $\{\psi_k\}$, $\psi_k \in \mathbb{P}_k - \mathbb{P}_{k-1}$. Setting

$$p = \sum_{k=0}^{n} \psi_k a_k^{\Psi}, \qquad a_k^{\Psi} \in \mathbb{C},$$

the least squares problem can be formulated as finding the weighted least squares solution of the system of linear equations

$$\sum_{k=0}^{n} \psi_k(z_i) a_k^{\Psi} = f(z_i), \quad i = 0, \ldots, m,$$

which is the same as the least squares solution of

$$W \Psi_n A_n^{\Psi} = WF$$

where $W = \mathrm{diag}(w_0, \ldots, w_m)$ and

$$\Psi_n = \begin{bmatrix} \psi_0(z_0) & \ldots & \psi_n(z_0) \\ \vdots & & \vdots \\ \psi_0(z_m) & \ldots & \psi_n(z_m) \end{bmatrix}, \quad A_n^{\Psi} = \begin{bmatrix} a_0^{\Psi} \\ \vdots \\ a_n^{\Psi} \end{bmatrix}, \quad F = \begin{bmatrix} f(z_0) \\ \vdots \\ f(z_m) \end{bmatrix}.$$

Note that when $\psi_k(z) = z^k$, the power basis, then Ψ_n is a rectangular Vandermonde matrix.
The normal equations for this system are

$$(\Psi_n^H W^2 \Psi_n) A_n^{\Psi} = \Psi_n^H W^2 F.$$

When the ψ_k are chosen to be the orthonormal polynomials φ_k, then $\Psi_n^H W^2 \Psi_n = I_{n+1}$ and the previous system gives the solution $A_n^{\Psi} = \Psi_n^H W^2 F$ immediately.

When the least squares problem is solved by QR factorization, i.e., when Q is an $m \times m$ unitary matrix such that $Q^H W \Psi_n = [R^T \; 0^T]^T$, is upper triangular, we have to solve the triangular system given by the first $n+1$ rows of

$$\begin{bmatrix} R \\ 0 \end{bmatrix} A_n^\Psi = Q^H W F + \begin{bmatrix} 0 \\ X \end{bmatrix}$$

where X is related to the residual vector r by

$$\begin{bmatrix} 0 \\ X \end{bmatrix} = Q^H r, \quad r = W \Psi_n A_n^\Psi - WF.$$

Note that the least squares error is $\|X\| = \|r\|$. Again, when the ψ_k are replaced by the orthonormal polynomials φ_k, we get the trivial system ($m \geq n$)

$$\begin{bmatrix} I_{n+1} \\ 0 \end{bmatrix} A_n^\Phi = Q^H W F + \begin{bmatrix} 0 \\ X \end{bmatrix}.$$

Note that a unitary matrix Q is always related to the orthonormal polynomials φ_k by

$$Q = W\Phi$$

where

$$\Phi = \Phi_m = \begin{bmatrix} \varphi_0(z_0) & \dots & \varphi_m(z_0) \\ \vdots & & \vdots \\ \varphi_0(z_m) & \dots & \varphi_m(z_m) \end{bmatrix}$$

since

$$Q^H Q = \Phi^H W^2 \Phi = I_{m+1}.$$

3. The Hessenberg matrix. From the previous discussion, it follows that the central problem is to construct the orthonormal basis $\{\varphi_k\}$. In general, the polynomial $z\varphi_{k-1}(z)$ can be expressed as a linear combination of the polynomials $\varphi_0, \dots, \varphi_k$, leading to a relation of the form

$$z\varphi_{k-1}(z) = \eta_{kk}\varphi_k(z) + \dots + \eta_{0k}\varphi_0(z), \quad k = 1, \dots, m+1.$$

We can express the previous relations as

$$z[\varphi_0(z), \dots, \varphi_m(z)] = [\varphi_0(z), \dots, \varphi_m(z)]H + e_{m+1}^T \varphi_{m+1}(z)\eta_{m+1,m+1}, \tag{3}$$

where H is an upper Hessenberg matrix

$$H = \begin{bmatrix} \eta_{01} & \cdots & \eta_{0m} & \eta_{0,m+1} \\ \eta_{11} & \cdots & \eta_{1m} & \eta_{1,m+1} \\ & \ddots & \vdots & \vdots \\ & & \eta_{mm} & \eta_{m,m+1} \end{bmatrix}$$

and $e_{m+1}^T = [0 \; 0 \; \dots \; 0 \; 1]$.

Note that a discrete inner product of the proposed form will cause a break-down in the generation of the polynomials at stage $m+1$. Indeed, we should identify a function with the $(m+1)$-vector of its function values in z_k, $k = 0, \dots, m$. Thus when we say the "polynomial p", we actually mean the vector $(p(z_0), \dots, p(z_m))$. Thus our "function space" is a space of $(m+1)$-vectors, which is inherently $(m+1)$-dimensional, and thus the $(m+1)$-st orthogonal polynomial will be orthogonal to the whole space, hence it must be zero. Thus, if φ_k are these orthogonal polynomials, then $[\varphi_{m+1}(z_0), \dots, \varphi_{m+1}(z_m)]^T$ will be the zero vector. This is equivalent with saying that φ_{m+1} is proportional to $(z - z_0) \dots (z - z_m)$.

Even when we use terms as "functions", and "polynomials", the problem considered is in fact a vectorial problem, which can be best formulated in terms of matrices, which we shall do below.

Setting $\Phi = \Phi_m$ as before, we rewrite the relation (referr) as

$$Z\Phi = \Phi H$$

with $Z = \mathrm{diag}(z_0, \dots, z_m)$.

Multiplying with the diagonal matrix W and using $WZ = ZW$, we are lead to

$$H = (W\Phi)^H Z(W\Phi) = Q^H ZQ,$$

which means that the diagonal matrix Z and the Hessenberg matrix H are unitarily similar.

The constant polynomial φ_0 is normalized when it is equal to η_{00}^{-1} with η_{00} given by

$$Q^H \mathbf{w}_1 = [\eta_{00}, 0, \dots, 0]^T$$

where $\mathbf{w}_1 = [w_0, \dots, w_m]^T$. Indeed, using $Q = W\Phi$ and supposing $\|\varphi_0\| = 1$, we see that all the entries in $Q^H \mathbf{w}_1$ are zero by orthogonality, except for the first one, which is $1/\varphi_0$.

This condition is not sufficient to characterize Q completely. We can fix it uniquely by making the φ_k to have positive leading coefficients. This will be obtained when all the η_{kk}, $k = 0, 1, \dots, m$ are positive. Since we assumed that all the weights w_i^2 are positive, the η_{kk} are nonzero and therefore this normalization can always be realized.

We thus obtained a 1-to-1 relation between the data $\{z_i, w_i\}_0^m$, the unitary matrix Q and the elements η_{ij}, $i = 0, \dots, m$, $j = 0, \dots, m+1$ of an extended (with η_{00}) Hessenberg matrix and this also fixes the orthonormal polynomials.

Since Z and H are unitarily similar, they have the same spectrum and the construction of H from Z by unitary similarity transformations is in fact an inverse spectral problem: given the spectrum Z and the first components of the eigenvectors, find the set of orthonormal eigenvectors (the columns of Q^H), such that $Q^H ZQ$ is the eigenvalue decomposition of some upper Hessenberg matrix with the normalization described above.

In the direct problem, one computes the eigenvalues $\{z_k\}_0^m$ and the eigenvectors Q from the Hessenberg matrix, e.g., with the QR algorithm. For the inverse problem, the Hessenberg matrix is reconstructed from the spectral data by an algorithm which could be called an inverse QR algorithm. This is the Rutishauser-Gragg-Harrod algorithm for the case of the real line [11, 12] and the unitary inverse QR algorithm described in [2] for the case of the unit circle. For the least squares problem, we add the function values $f(z_k)$ and when these are properly transformed by the similatity transformations of the inverse QR algorithm, this will result in the generalized Fourier coefficients of the approximant and some information about the corresponding residual. Indeed, the solution of the approximation problem is given by

$$p = [\varphi_0, \dots, \varphi_n] A_n^\Phi, \quad A_n^\Phi = \Phi_n^H W^2 F.$$

Note that the normal equations are never explicitly formed.

The whole scheme can be collected in one table giving the relations

$$Q^H \left[\ \mathbf{w}_0 \mid \mathbf{w}_1 \parallel Z\ \right] \left[\begin{array}{cc} I_2 & \\ & Q \end{array}\right] = \left[\begin{array}{c|c||cccc} & \eta_{00} & \eta_{01} & \cdots & \eta_{0m} & \eta_{0,m+1} \\ A_n^\Phi & 0 & \eta_{11} & \cdots & \eta_{1m} & \eta_{1,m+1} \\ -- & \vdots & & \ddots & \vdots & \vdots \\ X & 0 & & & \eta_{mm} & \eta_{m,m+1} \end{array}\right]$$

with $\mathbf{w}_0 = WF$ and $\mathbf{w}_1 = [w_0, \ldots, w_m]^T$ as before. The approximation error is $\|X\|$. For further reference we shall refer to the matrix of the right hand side as the extended Hessenberg matrix.

4. Updating. Suppose that A_n^Φ was computed by the last scheme for some data set $\{z_i, f_i, w_i\}_0^m$. We then end up with a scheme of the following form $(n = 3, m = 5)$

$$\left[\begin{array}{cc|cccccc} \times & \times & \times & \times & \times & \times & \times & \times \\ \times & & \times & \times & \times & \times & \times & \times \\ \times & & & \times & \times & \times & \times & \times \\ \times & & & & \times & \times & \times & \times \\ \hline \times & & & & & \times & \times & \times \\ \times & & & & & & \times & \times \end{array}\right]$$

A new data triple $(z_{m+1}, f_{m+1}, w_{m+1})$ can be added, for example, in the top line. The three crosses in the top line of the left scheme below represent $w_{m+1}f_{m+1}$, w_{m+1} and z_{m+1} respectively. The other crosses correspond to the ones we had in the previous scheme.

$$\left[\begin{array}{cc|c|cccccc} \times & \times & \times & & & & & & \\ \hline \times & \times & & \times & \times & \times & \times & \times & \times \\ \times & & & \times & \times & \times & \times & \times & \times \\ \times & & & & \times & \times & \times & \times & \times \\ \times & & & & & \times & \times & \times & \times \\ \hline \times & & & & & & \times & \times & \times \\ \times & & & & & & & \times & \times \end{array}\right] \Rightarrow \left[\begin{array}{cc|ccccccc} \times & \times & \times & \times & \times & \times & \times & \times & \times \\ \times & & \times & \times & \times & \times & \times & \times & \times \\ \times & & & \times & \times & \times & \times & \times & \times \\ \times & & & & \times & \times & \times & \times & \times \\ \hline \times & & & & & \times & \times & \times & \times \\ \times & & & & & & \times & \times & \times \\ \times & & & & & & & \times & \times \end{array}\right]$$

This left scheme has to be transformed by unitary similarity transformations into the right scheme which has the same form as the original one but with one extra row and one extra column. This result is obtained by eliminating the (2,2) element by an elementary rotation/reflection in the plane of the first two rows. The corresponding transformation on the columns will influence columns 3 and 4 and will introduce a nonzero element at position (3,3) which should not be there. This is eliminated by a rotation/reflection in the plane of rows 2 and 3 etc. We call this procedure chasing the elements down the diagonal. In the first column of the result, we find above the horizontal line the updated coefficients A_n^Φ. When we do not change n, it is sufficient to perform only the operations which influence these coefficients. Thus we could have stopped after we obtained the form

$$
\begin{array}{|cc|ccccccc|}
\hline
\times & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & & \times & \times & \times & \times & \times & \times & \times \\
\times & & & \times & \times & \times & \times & \times & \times \\
\times & & & & \times & \times & \times & \times & \times \\
\hline
\times & & & & & \times & \times & \times & \times \\
\times & & & & & \times & \times & \times & \times \\
\times & & & & & & & \times & \times \\
\hline
\end{array}
$$

This can be done with $O(n^2)$ operations per new data point. In the special case of data on the real line or on the unit circle, this reduces to $O(n)$ operations as we shall see in the next section.

5. Recurrence relations. The algorithm described above simplifies considerably when the orthogonal polynomials satisfy a particular recurrence relation. A classical situation occurs when the $z_i \in \mathbb{R}$, $i = 0, 1, \ldots, m$. Since also the weights w_i are real, the Q and H matrix will be real, which means that we can drop the complex conjugation from our notation. However, in view of the generalization to follow, where we shall have complex numbers instead of the w_i, we keep for the moment the bar, although it has no effect, being applied to real numbers. Thus we observe that for $z_i \in \mathbb{R}$, the Hessenberg matrix H satisfies

$$H^H = (Q^H Z Q)^H = Q^H Z Q = H.$$

This means that H is Hermitian and therefore tridiagonal. The matrix H reduces to the classical Jacobi matrix

$$H = \begin{bmatrix} a_0 & \bar{b}_1 & & \\ b_1 & a_1 & \ddots & \\ & \ddots & \ddots & \bar{b}_m \\ & & b_m & a_m \end{bmatrix}$$

containing the coefficients of the three term recurrence relation

$$\varphi_{-1} = 0, \quad z\varphi_k(z) = \bar{b}_k \varphi_{k-1}(z) + a_k \varphi_k(z) + b_{k+1}\varphi_{k+1}(z), \quad k = 0, 1, \ldots, m-1.$$

A similar situation occurs when the z_i are purely imaginary, in which case the matrix H is skew Hermitian. We shall not discuss this case separately.

The algorithm we described before now needs to perform rotations (or reflections) on vectors of length 3 or 4, which reduces the complexity of the algorithm by an order. This is the basis of the Rutishauser-Gragg-Harrod algorithm [14, 11]. See also [5, 12, 6].

In this context, it was observed only lately [9, 10, 2, 13, 3] that also the situation where the $z_i \in \mathbb{T}$ (the unit circle) leads to a simplification. It follows from

$$H^H H = Q^H Z^H Z Q = Q^H Q = I_{m+1}$$

that H is then a unitary Hessenberg matrix. The related orthogonal polynomials are orthogonal with respect to a discrete measure supported on the unit circle. The 3-term recurrence relation is replaced by a recurrence of Szegő-type

$$z\varphi_{k-1}(z) = \varphi_k(z)\sigma_k + \varphi^*_{k-1}(z)\gamma_k$$

with

$$\varphi_k^*(z) = z^k \overline{\varphi_k(1/\overline{z})} \in \mathbb{P}_k \quad \text{and} \quad \sigma_k^2 = 1 - |\gamma_k|^2, \quad \sigma_k > 0$$

where the γ_k are the so called reflection coefficients or Schur parameters. Just like in the case of a tridiagonal matrix, the Hessenberg matrix is built up from the recurrence coefficients γ_k, σ_k. However, the connection is much more complicated. For example, for $m = 3$, H has the form

$$H = \begin{bmatrix} -\gamma_1 & -\sigma_1\gamma_2 & -\sigma_1\sigma_2\gamma_3 & -\sigma_1\sigma_2\sigma_3\gamma_4 \\ \sigma_1 & -\bar{\gamma}_1\gamma_2 & -\bar{\gamma}_1\sigma_2\gamma_3 & -\bar{\gamma}_1\sigma_2\sigma_3\gamma_4 \\ & \sigma_2 & -\bar{\gamma}_2\gamma_3 & -\bar{\gamma}_2\sigma_3\gamma_4 \\ & & -\sigma_3 & \bar{\gamma}_3\gamma_4 \end{bmatrix}.$$

The Schur parameters can be recovered from the Hessenberg matrix by

$$\sigma_j = \eta_{jj}, \quad j = 1, \ldots, m, \quad \eta_{00} = 1/\varphi_0 = \sigma_0,$$

$$\gamma_j = -\eta_{0j}/(\sigma_1\sigma_2 \ldots \sigma_{j-1}), \quad j = 1, \ldots, m+1.$$

The complexity reduction in the algorithm is obtained from the important observation that any unitary Hessenberg matrix H can be written as a product of elementary unitary factors

$$H = G_1 G_2 \ldots G_m G'_{m+1}$$

with

$$G_k = I_{k-1} \oplus \begin{bmatrix} -\gamma_k & \sigma_k \\ \sigma_k & \bar{\gamma}_k \end{bmatrix} \oplus I_{m-k}, \quad k = 1, \ldots, m$$

and

$$G'_{m+1} = \text{diag}(1, \ldots, 1, -\gamma_{m+1}).$$

This result can be found e.g., in [9, 2].

Now an elementary similarity transformation on rows/columns k and $k+1$ of H, represented in this factored form, will only affect the factors G_k and part of the factors G_{k-1} and G_{k+1}. Again, these operations require computations on short vectors of length 3, making the algorithm very efficient again. For the details consult [9, 2, 13]. For example, the interpolation problem ($n = m$) is solved in $O(m^2)$ operations instead of $O(m^3)$.

6. Vector approximants. The previous situation of polynomial approximation can be generalized as follows.

Given $\{z_i; f_{0i}, \ldots, f_{\alpha i}; w_{0i}, \ldots, w_{\alpha i}\}_{i=0}^m$, find polynomials $p_k \in \mathbb{P}_{d_k}$, $k = 0, \ldots, \alpha$, such that

$$\sum_{i=0}^{m} |w_{0i} f_{0i} p_0(z_i) + \cdots + w_{\alpha i} f_{\alpha i} p_\alpha(z_i)|^2$$

is minimized. Now it doesn't really matter whether the w_{ji} are positive or not, since the products $w_{ji} f_{ji}$ will now play the role of the weights and the f_{ji} are arbitrary complex numbers. Thus, to simplify the notation, we could as well write w_{ji} instead of $w_{ji}f_{ji}$ since these numbers will always appear as products. Thus the problem is to minimize

$$\sum_{i=0}^{m} |w_{0i}p_0(z_i) + \cdots + w_{\alpha i}p_\alpha(z_i)|^2.$$

Setting $\mathbf{d} = (d_0, \ldots, d_\alpha)$, $\mathbb{P}_\mathbf{d} = [\mathbb{P}_{d_0}, \ldots, \mathbb{P}_{d_\alpha}]^T$,

$$w_i = [w_{0i}, \ldots, w_{\alpha i}], \quad p(z) = [p_0(z), \ldots, p_\alpha(z)]^T \in \mathbb{P}_\mathbf{d},$$

we can write this as

$$\min \sum_{i=0}^{m} |w_i p(z_i)|^2, \quad p \in \mathbb{P}_\mathbf{d}.$$

Of course, this problem has the trivial solution $p = 0$, unless we require at least one of the $p_i(z)$ to be of strict degree d_i, e.g., by making it monic. This, or any other normalization condition could be imposed for that matter.

We require in this paper that p_α is monic of degree d_α, and rephrase this as $p_\alpha \in \mathbb{P}^M_{d_\alpha}$.

To explain the general idea, we restrict ourselves to $\alpha = 1$, the case of a general α being a straightforward generalization which would only increase the notational burden. Thus we consider the problem

$$\min \sum_{i=0}^{m} |w_{0i}p_0(z_i) + w_{1i}p_1(z_i)|^2, \quad p_0 \in \mathbb{P}_{d_0}, p_1 \in \mathbb{P}^M_{d_1}.$$

Note that when $w_{0i} = w_i > 0$, $w_{1i} = -w_i f_i$, and $p_1 \equiv 1 \in \mathbb{P}^M_0$, (i.e., $d_1 = 0$), then we get the polynomial approximation problem discussed before.
When we set $w_{0i} = w_i f_{0i}$ and $w_{1i} = -w_i f_{1i}$ with $w_i > 0$, the problem becomes

$$\min \sum_{i=0}^{m} w_i^2 |f_{0i}p_0(z_i) - f_{1i}p_1(z_i)|^2$$

which is a linearized version of the rational least squares problem of determining the rational approximant p_0/p_1 for the data f_{1i}/f_{0i}, or equivalently the rational approximant p_1/p_0 for the data f_{0i}/f_{1i}. Note that in the linearized form, it is as easy to prescribe pole information ($f_{0i} = 0$) as it is to fix a finite function value ($f_{0i} \neq 0$).

The solution of the general case is partly parallel to the polynomial case $d_1 = 0$ discussed before, and partly parallel to another simple case, namely $d_0 = d_1 = n$, which we shall discuss first in section 7. In the subsequent section 8, we shall consider the general case where $d_0 \neq d_1$.

7. Equal degrees. We consider the case $\alpha = 1$, $d_0 = d_1 = n$. This means that $\mathbb{P}_\mathbf{d}$ is here equal to $\mathbb{P}^{2\times 1}_n$.

7.1. The optimization problem. We have to find

$$\min \sum_{i=0}^{m} |w_i p(z_i)|^2, \quad p_0 \in \mathbb{P}_n, \quad p_1 \in \mathbb{P}^M_n$$

where $w_i = [w_{0i} \;\; w_{1i}]$ and $p(z) = [p_0(z) \;\; p_1(z)]^T \in \mathbb{P}_n^{2\times 1}$. This problem was considered in [15, 17]. We propose a solution of the form

$$p(z) = \sum_{k=0}^{n} \varphi_k(z) a_k,$$

where

$$\varphi_k(z) \in \mathbb{P}_k^{2\times 2} - \mathbb{P}_{k-1}^{2\times 2}, \quad a_k \in \mathbb{C}^{2\times 1}, \quad k = 0, 1, \ldots, n.$$

Proposing $p(z)$ to be of this form assumes that the leading coefficients of the block polynomials φ_k are nonsingular. Otherwise this would not represent all possible couples of polynomials $(p_0, p_1) \in \mathbb{P}_n^{2\times 1}$. We shall call this the regular case and assume for the moment that we are in this comfortable situation. In the singular case, a break-down may occur during the algorithm, and we shall deal with that separately. Note that the singular case did not show up in the previous scalar polynomial case, unless at the very end when $n = m+1$, since the weights were assumed to be positive. We shall see below that in this block polynomial situation, the weights are not positive and could even be singular.

When we denote

$$\begin{aligned} W &= \mathrm{diag}(w_0, \ldots, w_m) \in \mathbb{C}^{(m+1)\times(2m+2)} \\ A_n^\Phi &= [a_0^T, \ldots, a_n^T]^T \in \mathbb{C}^{(2n+2)\times 1} \\ \Phi_n &= \begin{bmatrix} \varphi_0(z_0) & \ldots & \varphi_n(z_0) \\ \vdots & & \vdots \\ \varphi_0(z_m) & \ldots & \varphi_n(z_m) \end{bmatrix} \in \mathbb{C}^{(2m+2)\times(2n+2)} \end{aligned}$$

the optimization problem is to find the least squares solution of the homogeneous linear system

$$W \Phi_n A_n^\Phi = 0$$

with the constraint that p_1 should be monic of degree n.

For simplicity reasons, suppose that $m+1 = 2(m'+1)$ is even. If it were not, we would have to make a modification in our formulations for the index m'. The algorithm however, does not depend on m being odd or even as we shall see later.

By making the block polynomials φ_k orthogonal so that

$$\sum_{i=0}^{m} \varphi_k(z_i)^H w_i^H w_i \varphi_l(z_i) = \delta_{kl} I_2, \quad k, l = 0, 1, \ldots, m', \tag{4}$$

we can construct a unitary matrix $Q \in \mathbb{C}^{(m+1)\times(m+1)}$ by setting

$$Q = W\Phi$$

where $\Phi = \Phi_{m'}$ is a $(2m+2)\times(m+1)$ matrix, so that Q is a square matrix of size $m+1$.

We also assume that the number of data points $m+1$ is at least equal to the number of unknowns $2n+1$ (recall that one coefficient is fixed by the monic normalization). The unitarity of the matrix Q means that

$$Q^H Q = \Phi^H W^H W \Phi = I_{m+1}$$

and the optimization problem reduces to

$$\begin{aligned}\min\sum_{i=0}^{m} p(z_i)^H w_i^H w_i p(z_i) &= \min(A_{m'}^{\Phi})^H \Phi^H W^H W \Phi(A_{m'}^{\Phi}) \\ &= \min(A_{m'}^{\Phi})^H (A_{m'}^{\Phi}) \\ &= \min\sum_{k=0}^{m'} a_k^H a_k \\ &= \min\sum_{k=0}^{m'} (|a_{1k}|^2 + |a_{2k}|^2), \quad a_k = [a_{1k} \;\; a_{2k}]^T\end{aligned}$$

with the constraint that $p(z) \in \mathbb{P}_n^{2\times 1}$; thus $a_{n+1} = \cdots = a_{m'} = 0$, and $p_1 \in \mathbb{P}_n^M$. Since the leading term of p_1 is only influenced by $\varphi_n a_n$, we are free to choose $a_0, \ldots, a_{n-1}$, so that we can set them equal to zero, to minimize the error. Thus it remains to find

$$\min(|a_{1n}|^2 + |a_{2n}|^2)$$

such that

$$\varphi_n(z)\begin{bmatrix} a_{1n} \\ a_{2n} \end{bmatrix} = \begin{bmatrix} p_0(z) \\ p_1(z) \end{bmatrix} \in \begin{bmatrix} \mathbb{P}_n \\ \mathbb{P}_n^M \end{bmatrix}.$$

To monitor the degree of p_1, we shall require that the polynomials φ_k have an upper triangular leading coefficient:

$$\varphi_k(z) = \begin{bmatrix} \alpha_k & \gamma_k \\ 0 & \beta_k \end{bmatrix} z^k + \cdots$$

with $\alpha_k, \beta_k > 0$. Note that this is always possible in the regular case. The condition $p_1 \in \mathbb{P}_n^M$ then sets $a_{2n} = 1/\beta_n$ and a_{1n} is arbitrary, hence to be set equal to zero if we want to minimize the error.
As a conclusion, we have solved the approximation problem by computing the n-th block polynomial φ_n, orthonormal in the sense of (4) and with leading coefficient upper triangular. The solution is

$$p(z) = \begin{bmatrix} p_0(z) \\ p_1(z) \end{bmatrix} = \varphi_n(z)\begin{bmatrix} 0 \\ a_{2n} \end{bmatrix}, \quad a_{2n} = 1/\beta_n.$$

7.2. The algorithm. As in the scalar polynomial case, expressing $z\varphi_k(z)$ in terms of $\varphi_0, \ldots, \varphi_{k+1}$ for $z \in \{z_0, \ldots, z_m\}$ leads to the matrix relation

$$\mathbf{Z}\Phi = \Phi H,$$

where as before $\Phi = \Phi_{m'}$, $Z = \mathrm{diag}(z_0, \ldots, z_m)$, $\mathbf{Z} = Z \otimes I_2 = \mathrm{diag}(z_0 I_2, \ldots, z_m I_2)$, and H is a block upper Hessenberg matrix with 2×2 blocks. If the leading coefficient of φ_k is upper triangular, then the subdiagonal blocks of H are upper triangular. The computational scheme is compressed in the formula

$$Q^H[\mathbf{w}|Z]\begin{bmatrix} I_2 & \\ & Q \end{bmatrix} = \left[\begin{array}{c|cccc} \eta_{00} & \eta_{01} & \cdots & \eta_{0m'} & \eta_{0,m'+1} \\ & \eta_{11} & & \eta_{1m'} & \eta_{1,m'+1} \\ & & \ddots & \vdots & \vdots \\ & & & \eta_{m'm'} & \eta_{m',m'+1} \end{array}\right]$$

where $\mathbf{w} = [w_0^T, \ldots, w_m^T]^T$ and where all η_{ij} are 2×2 blocks and the η_{ii} are upper triangular with positive diagonal elements. Thus

$$\varphi_0 = \eta_{00}^{-1}; \quad z\varphi_{k-1}(z) = \varphi_0(z)\eta_{0k} + \cdots + \varphi_k(z)\eta_{kk}, \quad k = 1, \ldots, m'.$$

The updating after adding the data (z_{m+1}, w_{m+1}), where $w_{m+1} = (w_{0,m+1}, w_{1,m+1})$, makes the transformation with unitary similarity transformations from the left to the right scheme below. The three crosses in the top row of the left scheme represent the new data.

$$\left[\begin{array}{cc|c|cc|cc|cc}
\times & \times & \times & & & & & & \\
\hline
\times & \times & & \times & \times & \times & \times & \times & \times \\
 & \times & & \times & \times & \times & \times & \times & \times \\
\hline
 & & & \times & \times & \times & \times & \times & \times \\
 & & & & \times & \times & \times & \times & \times \\
\hline
 & & & & & \times & \times & \times & \times \\
 & & & & & & \times & \times & \times
\end{array}\right] \Rightarrow \left[\begin{array}{cc|cc|cc|cc|c}
\times & \times & \times & \times & \times & \times & \times & \times & \times \\
 & \times & \times & \times & \times & \times & \times & \times & \times \\
\hline
 & & \times & \times & \times & \times & \times & \times & \times \\
 & & & \times & \times & \times & \times & \times & \times \\
\hline
 & & & & \times & \times & \times & \times & \times \\
 & & & & & \times & \times & \times & \times \\
\hline
 & & & & & & \times & \times & \times
\end{array}\right]$$

The successive elementary transformations eliminate the crosses on the subdiagonal, chasing them down the matrix. This example also illustrates what happens at the end when m is an even number: the polynomial $\varphi_{m'} \in \mathbb{P}_{m'}^{2\times1}$ instead of $\varphi_{m'} \in \mathbb{P}_{m'}^{2\times2}$. Again, when finishing this updating after φ_n has been computed, it will require only $O(n^2)$ operations per data point introduced. In the special case of data on the real line or the unit circle, this reduces to $O(n)$ operations. For the details, we refer to [15, 17].

By the same arguments as in the scalar case, it is still true that H is Hermitian, hence block tridiagonal, when all the z_i are real. Taking into account that the subdiagonal blocks are upper triangular, we obtain in this case that H is pentadiagonal and the extended Hessenberg matrix has the form

$$\left[\begin{array}{c|cccc}
B_0 & A_0 & B_1^H & & \\
 & B_1 & A_1 & \ddots & \\
 & & \ddots & \ddots & B_{m'}^H \\
 & & & B_{m'} & A_{m'}
\end{array}\right]$$

with the 2×2 blocks B_k upper triangular and the A_k Hermitian. This leads to the following block 3-term recurrence

$$\varphi_0 = B_0^{-1}, \quad z\varphi_k(z) = \varphi_{k-1}B_k^H + \varphi_k(z)A_k + \varphi_{k+1}(z)B_{k+1}, \quad 0 \le k < m'.$$

This case was considered in [15].
Similarly, the case where all z_i lie on the unit circle $\mathbb{T}$, leads to a 2×2 block generalization of the corresponding polynomial case. For example, the extended unitary block Hessenberg matrix takes the form ($m' = 3$)

$$[H_0|H] = \left[\begin{array}{c|cccc}
\sigma_0 & -\gamma_1 & -\Sigma_1\gamma_2 & -\Sigma_1\Sigma_2\gamma_3 & \Sigma_1\Sigma_2\Sigma_3 \\
 & \sigma_1 & -\Gamma_1\gamma_2 & -\Gamma_1\Sigma_2\gamma_3 & \Gamma_1\Sigma_2\Sigma_3 \\
 & & \sigma_2 & -\Gamma_2\gamma_3 & \Gamma_2\Sigma_3 \\
 & & & \sigma_3 & \Gamma_3
\end{array}\right], \quad \gamma_i, \sigma_i, \Sigma_i, \Gamma_i \in \mathbb{C}^{2\times2}.$$

The matrices

$$U_k = \left[\begin{array}{cc} -\gamma_k & \Sigma_k \\ \sigma_k & \Gamma_k \end{array}\right]$$

are unitary: $U_k^H U_k = I_4$. Note that by allowing some asymmetry in the U_k we do not need a $-\gamma_4$ in the last column as we had in the scalar case. We have for $k = 1, \dots, m'$, the block Szegő recurrence relations

$$\begin{aligned} \varphi_k(z)\sigma_k &= z\varphi_{k-1}(z) + \varphi'_{k-1}(z)\gamma_k \\ \varphi'_k(z)\Sigma_k^H &= z\varphi_{k-1}(z)\gamma_k^H + \varphi'_{k-1}(z) \end{aligned}$$

which start with $\varphi_0 = \varphi'_0 = \sigma_0^{-1}$.
The block Hessenberg matrix can again be factored as

$$H = G_1 G_2 \dots G_{m'}$$

with

$$G_k = I_{2(k-1)} \oplus U_k \oplus I_{m-2k-1}, \quad k = 1, \dots, m'.$$

The proof of this can be found in [17]. This makes it possible to perform the elementary unitary similarity transformations of the updating procedure only on vectors of maximal length 5, very much like in the case of real points z_i. Thus also here, the complexity of the algorithm reduces to $O(m^2)$ for interpolation. More details can be found in [17]. For the case of the real line, the algorithm was also discussed in [1], solving an open problem in [5, p. 615]. The previous procedure now solves the problem also for the case of the unit circle.

7.3. Summary. The case $\alpha = 1$, $d_0 = d_1 = n$ and also the case $\alpha \geq 1$, $d_0 = d_1 = \cdots = d_\alpha = n$ for that matter, generalizes the polynomial approximation problem by constructing orthonormal polynomials φ_k which are $(\alpha+1) \times (\alpha+1)$ polynomial matrices and these are generated by a block 3-term recurrence relation when all $z_i \in \mathbb{R}$ and by a block Szegő recurrence relation when all $z_i \in \mathbb{T}$.
the computational algorithm is basically the same, since it reduces the extended matrix

$$[\mathbf{w}|Z] \in \mathbb{C}^{(m+1)\times(\alpha+m+2)}$$

by a sequence of elementary unitary similarity transformations to an upper trapezoidal matrix

$$Q^H[\mathbf{w}|Z]\begin{bmatrix} I_{\alpha+1} & \\ & Q \end{bmatrix} = [H_0|H]$$

with H block upper Hessenberg with $(\alpha+1) \times (\alpha+1)$ blocks and

$$H_0 = Q^H\mathbf{w} = [\eta_{00}^T, 0, \dots, 0]^T,$$

where $\eta_{00} \in \mathbb{C}^{(\alpha+1)\times(\alpha+1)}$ is upper triangular with positive diagonal elements, as well as all the subdiagonal blocks of H. For $n = m'$, where $(\alpha+1)(m'+1) - 1 = m+1$, (which implies that $\sigma_{m'}$ is of size $\alpha \times (\alpha+1)$), we solve an interpolation problem. It requires $O(m^2)$ operations when $z_i \in \mathbb{R}$ or $\in \mathbb{T}$, instead of $O(m^3)$ when the z_i are arbitrary in $\mathbb{C}$.

8. Arbitrary degrees. In this section we consider the case $\alpha = 1$ with $d_0 \neq d_1$. For more details we refer to [16].

8.1. The problem. We suppose without loss of generality that $d_0 = \delta$ and $d_1 = n + \delta$, $n, \delta \geq 0$. We have to find once more

$$\min \sum_{i=0}^{m} |w_i p(z_i)|^2, \quad p_0 \in \mathbb{P}_\delta, \quad p_1 \in \mathbb{P}_{n+\delta}^M$$

with $w_i = [w_{0i} \ \ w_{1i}]$ and $[p_0(z) \ \ p_1(z)]^T \in \mathbb{P}_{\mathbf{d}}$, $\mathbf{d} = (d_0, d_1)$.
The polynomial approximation problem is recovered by setting $\delta = 0$. The case $d_0 = d_1 = \delta$ is recovered by setting $n = 0$.
The simplest approach to the general problem is by starting with the algorithm. In the subsequent subsections, we propose a computational scheme involving unitary similarity transformations, next we give an interpretation in terms of orthogonal polynomials and finally we solve the approximation problem.

8.2. The algorithm. Comparing the cases $\delta = 0$ and $n = 0$, we see that the algorithm applies a sequence of elementary unitary similarity transformations on an extended matrix

$$[\mathbf{w}|Z], \quad \mathbf{w} = [w_0^T, \ldots, w_m^T]^T, \quad Z = \mathrm{diag}(z_0, \ldots, z_m)$$

to bring it in the form of an extended (block) upper Hessenberg

$$Q^H [\mathbf{w}|Z] \begin{bmatrix} I_2 & \\ & Q \end{bmatrix} = [H_0|H].$$

When $n = 0$, the transformations were aimed at chasing down the elements of $[\mathbf{w}|Z]$ below the main diagonal, making $[H_0|H]$ upper triangular. Therefore H turned out to be block upper Hessenberg.

When $\delta = 0$, the transformations had much the same objective, but now, there was no attempt to eliminate elements from the first column of $\mathbf{w}$, only elements from the second column were pushed to the SE part of the matrix. The matrix then turned out to be upper Hessenberg in a scalar sense.

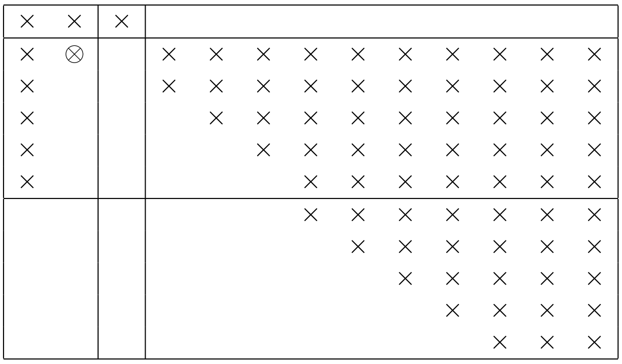
The general case can be treated by an algorithm that combines both of these objectives. We start like in the polynomial case ($n = 0$), chasing only elements from the second column of $\mathbf{w}$. However, once we reached row $n + 1$, we start eliminating elements in the first column too.

Applying this procedure shows that the extended Hessenberg $[H_0|H]$ has the form

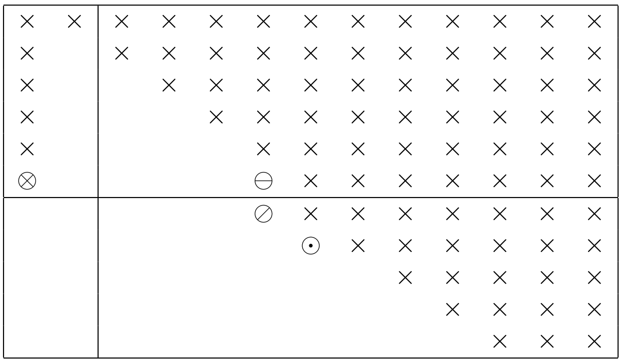
$$\begin{array}{r} \\ 0 \rightarrow \\ \\ \\ \\ n \rightarrow \\ \\ \\ \\ \\ m \rightarrow \end{array}
\begin{array}{|cc|ccccc|ccccc|}
\multicolumn{2}{c}{} & 0 & & & & n & & & & & m \\
\hline
\times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & & \times & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & & & \times & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & & & & \times & \times & \times & \times & \times & \times & \times & \times \\
\times & & & & & \times & \times & \times & \times & \times & \times & \times \\
\hline
 & & & & & \times & \times & \times & \times & \times & \times & \times \\
 & & & & & & \times & \times & \times & \times & \times & \times \\
 & & & & & & & \times & \times & \times & \times & \times \\
 & & & & & & & & \times & \times & \times & \times \\
 & & & & & & & & & \times & \times & \times \\
\hline
\end{array} = [H_0|H].$$

This means that the NW part of H, of size $(n + 1) \times (n + 1)$, will be scalar upper Hessenberg as in the case $n = 0$, while the SE part of size $(m - n + 2) \times (m - n + 2)$ has the block upper Hessenberg form of the case $\delta = 0$.

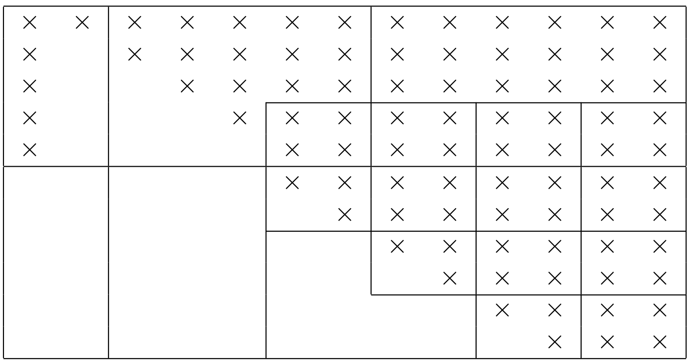
The updating procedure works as follows. Starting with (the new data are found in the first row)



the element $\otimes$ is chased down the diagonal by elementary unitary similarity transformations operating on two successive rows/columns until we reach the following scheme (where $\odot = 0$ and $\oslash$ and $\ominus$ are the last elements introduced which are in general nonzero)



Now the element $\otimes$ in row $n+1$ is eliminated by a rotation/reflection in the plane of this row and the previous one. The corresponding transformation on the columns will introduce a nonzero element at position $\odot$. Then $\oslash$ and $\odot$ are chased down the diagonal in the usual way until we reach the final situation



8.3. Orthogonal vector polynomials. The unitary matrix Q involved in the previous transformation was for the case $\delta = 0$ of the form $Q = W\Phi_m$ where W was a scalar diagonal matrix of the weights and Φ_m was the matrix with ij-element given by $\varphi_j(z_i)$, with φ_j the j-th orthonormal polynomial.

When $n = 0$, then $Q = W\Phi_{m'}$, where W is the block diagonal with blocks being the 2×1 "weights" w_i and $\Phi_{m'}$ is the block matrix with 2×2 blocks, where the ij-block is given by $\varphi_j(z_i)$, with φ_j the j-th block orthonormal polynomial.

For the general case, we shall have a mixture of both. For the NW part of the H matrix, we have the scalar situation and for the SE part we have the block situation.

To unify both situations, we turn to vector polynomials π_k of size 2×1. For the block part, we see a block polynomial φ_j as a collection of two columns and set

$$\varphi_j(z) = [\pi_{2j-1}(z)|\pi_{2j}(z)].$$

For the scalar part, we embed the scalar polynomial φ_j in a vector polynomial π_j by setting

$$\pi_j(z) = \begin{bmatrix} 0 \\ \varphi_j(z) \end{bmatrix}.$$

In both cases, the orthogonality of the φ_j translates into the orthogonality relation

$$\sum_{i=0}^{m} \pi_k(z_i)^H w_i^H w_i \pi_l(z_i) = \delta_{kl}$$

for the vector polynomials π_k. Let us apply this to the situation of the previous algorithm. For simplicity, we suppose that all $z_i \in \mathbb{R}$. For $z_i \in \mathbb{T}$, the situation is similar.

For column number $j = 0, 1, \ldots, n-1$, we are in the situation of scalar orthogonal polynomials: $Q_{ij} = w_{1i}\varphi_j(z_i) = w_i\pi_j(z_i)$. Setting

$$[H_0|H] = \left[\begin{array}{cc|ccc|c} \times & b_0 & a_0 & \bar{b}_1 & & \\ \vdots & & b_1 & a_1 & \ddots & \\ \times & & & \ddots & \ddots & \ddots \\ \hline & & & & \ddots & \ddots \end{array}\right]$$

we have for $j = 0, \ldots, n-2$ the 3-term recurrence relation

$$z\varphi_j(z) = \varphi_{j-1}(z)\bar{b}_j + \varphi_j(z)a_j + \varphi_{j+1}(z)b_{j+1}, \quad \varphi_{-1} = 0, \quad \varphi_0 = b_0^{-1}.$$

By embedding, this becomes

$$z\pi_j(z) = \pi_{j-1}(z)\bar{b}_j + \pi_j(z)a_j + \pi_{j+1}(z)b_{j+1}, \quad \pi_0 = [0 \;\; \varphi_0]^T.$$

Thus, setting

$$\Pi_j = [\pi_j(z_0)^T, \ldots, \pi_j(z_m)^T]^T$$

we have for the columns Q_j of Q the equality

$$Q_j = W\Pi_j, \quad j = 0, 1, \ldots, n-1.$$

For the trailing part of Q, i.e., for columns $(n+2j-1, n+2j)$, $j = 0, 1, \ldots$, we are in the block polynomial case. The block polynomials $\varphi_j(z)$ group two vector polynomials

$$\varphi_j(z) = [\pi_{n+2j-1}(z)|\pi_{n+2j}(z)],$$

which correspond to two columns of Q, namely

$$\mathbf{Q}_j = [Q_{n+2j-1}|Q_{n+2j}].$$

Observe that we have the following relation between $\mathbf{Q}_j$ and the block orthogonal polynomials

$$\mathbf{Q}_{ij} = w_i \varphi_j(z_i) = \begin{bmatrix} Q_{2i,n+2j-1} & Q_{2i,n+2j} \\ Q_{2i+1,n+2j-1} & Q_{2i+1,n+2j} \end{bmatrix},$$

where this time $w_i = [w_{0i} \;\; w_{1i}]$. As above, denote the vector of function values for π_j by Π_j. The block column of function values for φ_j is denoted by $\mathbf{\Phi}_j$. Then clearly

$$\mathbf{Q}_j = W\mathbf{\Phi}_j, \quad \mathbf{\Phi}_j = [\Pi_{n+2j-1}|\Pi_{n+2j}].$$

Denoting in the extended Hessenberg matrix

$$[H_0|H] = \left[\begin{array}{cc|cccccc} \times & b_0 & \ddots & \ddots & & & & \\ \times & & \ddots & \ddots & B_0^T & & & \\ \cline{5-8} 0 & & & B_0 & \multicolumn{1}{|c}{A_0} & B_1^T & & \\ 0 & & & & \multicolumn{1}{|c}{B_1} & A_1 & \ddots & \\ \vdots & & & & \multicolumn{1}{|c}{} & \ddots & \ddots & \end{array}\right], \quad B_0 = \begin{bmatrix} 0 & b_{n-1} \\ 0 & 0 \end{bmatrix},$$

we have the block recurrence

$$z\varphi_j(z) = \varphi_{j-1}(z)B_j^T + \varphi_j(z)A_j + \varphi_{j+1}(z)B_{j+1}, \quad j = 0, 1, \ldots$$

The missing link between the scalar and the block part is the initial condition for this block recurrence. This is related to columns $n-2, n-1$ and n of Q. Because columns $n-2$ and $n-1$ are generated by the scalar recurrence, we know that these columns are $Q_j = W\Pi_j$, $j = n-2, n-1$, where the Π_j are related to the embedded scalar polynomials. A problem appears in column Q_n where the 3-term recurrence of the leading (scalar) part migrates to the block 3-term recurrence of the trailing (block) part, i.e., from a 3-term to a 5-term scalar recurrence. We look at this column in greater detail. Because

$$\begin{bmatrix} d_0 \\ \vdots \\ d_n \\ 0 \\ \vdots \\ 0 \end{bmatrix} = Q^H \begin{bmatrix} w_{00} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ w_{0m} \end{bmatrix} = Q^H W E, \quad E = \begin{bmatrix} 1 \\ 0 \\ \hline 1 \\ 0 \\ \hline \vdots \\ \hline 1 \\ 0 \end{bmatrix},$$

we have

$$Q_0 d_0 + \cdots + Q_n d_n = WE;$$

thus

$$\begin{aligned} Q_n &= \frac{1}{d_n}(WE - [Q_0|\dots|Q_{n-1}]A_{n-1}^{\Phi}), \quad A_{n-1}^{\Phi} = [d_0,\dots,d_{n-1}]^T \\ &= \frac{1}{d_n}(WE - W[\Pi_0|\dots|\Pi_{n-1}]A_{n-1}^{\Phi}) \\ &= W\frac{1}{d_n}(E - [\Pi_0|\dots|\Pi_{n-1}]A_{n-1}^{\Phi}) \\ &= W\frac{1}{d_n}(E - P_{n-1}), \quad P_{n-1} = [\Pi_0|\dots|\Pi_{n-1}]A_{n-1}^{\Phi}. \end{aligned}$$

Setting $Q_n = W\Pi_n$, $\Pi_n = [\pi_n(z_0)^T,\dots,\pi_n(z_m)^T]^T$, we find that

$$\pi_n(z) = \frac{1}{d_n}\begin{bmatrix} 1 \\ p_{n-1}(z) \end{bmatrix} \tag{5}$$

where

$$p_{n-1}(z) = \varphi_0(z)d_0 + \dots + \varphi_{n-1}(z)d_{n-1}$$

is the polynomial least squares approximant of degree $n-1$ for the data (z_i, w_i), $i = 0,\dots,m$.

8.4. Solution of the general problem. Now we are ready to solve the general problem. We start with the degree structure of the polynomials $\pi_j(z)$. Suppose the j-th column of Q is Q_j, which we write as

$$Q_j = W\Pi_j, \quad \Pi_j = [\pi_j(z_0)^T,\dots,\pi_j(z_m)^T]^T$$

with $W = \text{diag}(w_0,\dots,w_m)$ and $\pi_j(z) = [\psi_j(z) \;\; \phi_j(z)]^T$. Then it follows from the previous analysis that the ϕ_j are the scalar orthogonal polynomials φ_j, and hence the degree of $\phi_j(z)$ is j, for $j = 0,1,\dots,n-1$. Moreover, the ψ_j are zero for the same indices (their degree is $-\infty$). For $j = n$, we just found that ψ_n is $1/d_n$, thus of degree 0 and ϕ_n is of degree at most $n-1$, since the latter is proportional to the polynomial least squares approximant of that degree. With the block recurrence relation, we now easily find that the degree structure of the block polynomials

$$\varphi_j = [\pi_{n+2j-1}|\pi_{n+2j}] = \begin{bmatrix} \psi_{n+2j-1} & \psi_{n+2j} \\ \phi_{n+2j-1} & \phi_{n+2j} \end{bmatrix} \quad \text{is} \quad \begin{bmatrix} j-1 & j \\ n+j-1 & n+j-1 \end{bmatrix}$$

for $j = 1, 2, \dots$, while φ_0 has degree structure

$$\begin{bmatrix} -\infty & 0 \\ n-1 & n-1 \end{bmatrix}.$$

It can be checked that in the regular case, that is when all the subdiagonal elements $b_0,\dots,b_{n-1}$ as well as d_n are nonzero and when also all the subdiagonal blocks $B_1,\dots,B'_m$ are regular (upper triangular), then the degrees of $\phi_k = \varphi_k$ are precisely k for $k = 0,1,\dots,n-1$ and in the block polynomials φ_j, the entries ψ_{n+2j} and ϕ_{n+2j-1} have the precise degrees that are indicated, i.e., j and $n+j-1$ respectively. Thus, if we propose a solution to our approximation problem of the form (suppose $m \geq n+2\delta$)

$$p(z) = \sum_{j=0}^{n+2\delta+1} \pi_j(z)a_j, \quad a_j \in \mathbb{C},$$

then $p(z) = [p_0(z) \ \ p_1(z)]^T$ will automatically satisfy the degree restrictions $d_0 \leq \delta$ and $d_1 \leq n+\delta$. We have to find

$$\min(A_n^\Pi)^H \mathbf{\Pi}_{n'}^H W^H W \mathbf{\Pi}_{n'}(A_n^\Pi), \quad n' = n + 2\delta + 1,$$

where

$$A_n^\Pi = [a_0, \ldots, a_{n'}]^T \quad \text{and} \quad \mathbf{\Pi}_{n'} = [\Pi_0|\ldots|\Pi_{n'}].$$

Since $W\mathbf{\Pi}_{n'}$ form the first $n'+1$ columns of the unitary matrix Q, this reduces to

$$\min(A_{n'}^\Pi)^H(A_{n'}^\Pi) = \min \sum_{j=0}^{n'} |a_j|^2.$$

If we require as before that $p_1(z)$ is monic of degree $n+\delta$, then $a_{n'} = 1/\beta_{n'}$ where β_j is the leading coefficient in ϕ_j. The remaining a_j are arbitrary. Hence, to minimize the error, we should make them all zero. Thus our solution is given by

$$p(z) = \pi_{n'}(z)a_{n'}, \quad n' = 2n + \delta + 1, \quad a_{n'} = 1/\beta_{n'}.$$

9. The singular case. Let us start by considering the singular case for $d_0 = d_1 = n$. We shall then generate a singular subdiagonal block η_{kk} of the Hessenberg matrix. The algorithm performing the unitary similarity transformations will not be harmed by this situation. However, the sequence of block orthogonal polynomials will break down. From the relation

$$z\varphi_{k-1}(z) = \varphi_0(z)\eta_{0k} + \cdots + \varphi_k(z)\eta_{kk}$$

it follows that if η_{kk} is singular, then this can not be solved for $\varphi_k(z)$. In the regular case, all the η_{jj} are regular and then the leading coefficient of φ_k is $\eta_{00}^{-1}\ldots\eta_{kk}^{-1}$. Thus, if all the η_{jj} are regular upper triangular, then also the leading coefficient of φ_k will be regular upper triangular. As we have said in the introduction, the singular situation will always occur, even in the scalar polynomial case with positive weights, but there only at the very end where $k = m+1$. That is exactly the stage where we reached the situation where the least squares solution becomes the solution of an interpolation problem. We show below that this is precisely what will also happen when some premature break-down occurs.

Suppose that the *scalar* entries of the extended block Hessenberg matrix are $[H_0|H] = [h_{ij}]_{i,j=0,1,\ldots}$. (We use h_{ij} to distinguish them from the block entries η_{ij}.) Suppose that the element h_{kk} is the first element on its diagonal that becomes zero and thus produces some singular subdiagonal block in H. Then it is no problem to construct the successive scalar columns of the matrix $\Phi = \Phi_{m'}$ until the recurrence relation hits the zero entry h_{kk}. If we denote for $j = 0, 1, \ldots, k-1$, the j-th column of Φ as Π_j, then we know from what we have seen, that Π_j represents the vector of function values at the nodes $z_0, \ldots, z_m$ of some vector polynomial $\pi_j(z) \in \mathbb{P}^{2\times 1}$. The problem in the singular case is that $\pi_k(z)$ can not be solved from

$$z\pi_{k-2} = \pi_k h_{kk} + \pi_{k-1}h_{k-1,k} + \cdots + \pi_0 h_{0k}$$

because $h_{kk} = 0$. However, from

$$Q \left[\begin{array}{c|c} H_0 & H \end{array}\right] = \left[\begin{array}{cc|c} \mathbf{w}_0 & \mathbf{w}_1 & Z \end{array}\right] \left[\begin{array}{c|c} I_2 & 0 \\ 0 & Q \end{array}\right]$$

it follows that

$$\mathbf{w}_0 = Q_0 h_{00}; \qquad \mathbf{w}_1 = Q_0 h_{01} + Q_1 h_{11}$$

and for $k \geq 2$

$$ZQ_{k-2} = Q_0 h_{0k} + \cdots + Q_k h_{kk}$$

where Q_j, $j = 0, 1, \ldots$ denotes the j-th column of Q. We shall discuss the case $h_{kk} = 0$ separately for $k = 0$, $k = 1$ and $k \geq 2$ separately.
If $h_{00} = 0$, then $\mathbf{w}_0 = 0$. This is a very unlikely situation because then there is only a trivial solution $(p_0, p_1) = (1, 0)$ which fits exactly.
Next consider $h_{11} = 0$; then define π'_1 as

$$\pi'_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \pi_0 h_{01}.$$

Then

$$\begin{aligned} W\Pi'_1 &= (\mathbf{w}_1 - W\Pi_0 h_{01}) \\ &= (\mathbf{w}_1 - Q_0 h_{01}) \\ &= Q_1 h_{11} = 0. \end{aligned}$$

This means that we get an exact approximation since $w_i \pi'_1(z_i) = 0$, $i = 0, \ldots, m$.
For the general case $h_{kk} = 0$, $k \geq 2$, we have that

$$ZQ_{k-2} - Q_0 h_{0k} - \cdots - Q_{k-1} h_{k-1,k} = Q_k h_{kk} = 0.$$

Since $Q_j = W\Pi_j$ for $j = 0, \ldots, k-1$, we also have

$$\begin{aligned} 0 &= ZW\Pi_{k-2} - W\Pi_0 h_{0k} - \cdots - W\Pi_{k-1} h_{k-1,k} \\ &= W\Big(\mathbf{Z}\Pi_{k-2} - \Pi_0 h_{0k} - \cdots - \Pi_{k-1} h_{k-1,k}\Big) \end{aligned} \tag{6}$$

where $\mathbf{Z} = Z \otimes I_2$. Define the polynomial

$$\pi'_k(z) = z\pi_{k-2}(z) - \pi_0(z) h_{0k} - \cdots - \pi_{k-1}(z) h_{k-1,k}$$

then, $W\Pi'_k = W[\pi'_k(z_0)^T, \ldots, \pi'_k(z_m)^T]^T$ will be zero since it is equal to the expression (6), which is zero. This means that

$$w_i \pi'_k(z_i) = 0, \quad i = 0, \ldots, m.$$

The latter relations just tell us that this π'_k is an exact solution of the approximation problem, i.e., it interpolates.

In the general situation where $d_0 \neq d_1$, we have to distinguish between the scalar and the block part. For the scalar part we can now also have a break-down in the sequence of orthogonal polynomials since the weights are not positive anymore but arbitrary complex numbers.

Using the notation

$$[H_0|H] = \left[\begin{array}{cc|ccc|c} \times & h_{00} & h_{01} & \ldots & h_{0,n+1} & \\ \vdots & & \ddots & \ddots & \vdots & \\ \times & & & h_{nn} & h_{n,n+1} & \ddots \\ \hline 0 & & & & \ddots & \ddots \end{array}\right]$$

for the NW part of the extended Hessenberg matrix, the situation is there as sketched above: whenever some h_{kk} is zero, we will have an interpolating polynomial solution. It then holds that

$$\pi'_k(z) = z\pi_{k-1}(z) - \pi_0(z)h_{0k} - \cdots - \pi_{k-1}(z)h_{k-1,k}$$

and because $W\Pi'_k = W[\pi'^T_k(z_0), \ldots, \pi'^T_k(z_m)]^T$ is zero, we get

$$w_i \pi'_k(z_i) = 0, \quad i = 0, \ldots, m,$$

identifying $\pi'_k(z)$ as a (polynomial) interpolant.

For the SE part, i.e., for the block polynomial part, a zero on the subsubdiagonal (i.e., when we get a singular subdiagonal block in the Hessenberg matrix), will imply interpolation as we explained above for the block case.

The remaining problem is the case where the bottom element in the first column of the transformed extended Hessenberg matrix becomes zero. That is the element that has previously been denoted by d_n. Indeed, if this is zero, then our derivation, which gave (5):

$$\pi_n(z) = \frac{1}{d_n} \begin{bmatrix} 1 \\ p_{n-1}(z) \end{bmatrix}$$

does not hold anymore. But again, here we will have interpolation, i.e., a least squares error equal to zero. It follows from the derivation in the previous section that when $d_n = 0$,

$$W(E - P_{n-1}) = d_n Q_n = 0.$$

Thus

$$w_i \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \sum_{k=0}^{n-1} \pi_k(z_i) d_k \right) = w_i \begin{bmatrix} 1 \\ p_{n-1}(z_i) \end{bmatrix} = 0$$

where $p_{n-1}(z) = \sum_{k=0}^{n-1} \varphi_k(z) a_k^\Phi$. This is the same as

$$w_{0i} - w_{1i} p_{n-1}(z_i) = 0, \quad i = 0, \ldots, m$$

which means that $(1, p_{n-1}(z))/d'$ with $d' \neq 0$ to normalize $p_{n-1}(z)$ as a monic polynomial, will fit the data exactly.

10. Conclusion. We have shown that the inverse QR algorithm for solving discrete polynomial approximation problems for knots on the real line or on the unit circle can be generalized to more general approximation problems of the form (2).

In the previous section, we only considered the problem of updating, i.e., how to adapt the approximant when one knot is added to the set of data. There also exists a possibility to consider downdating, i.e., when one knot is removed from the set of interpolation points. For the polynomial approximation problem, this was discussed in [6] for real data and in [3] for data on the unit circle. The procedure can be based on a direct QR algorithm which will "diagonalize" the Hessenberg matrix in one row and column (e.g., the last one). This means that the only nonzero element in the last row and the last column of the transformed Hessenberg matrix is z_m on the diagonal. The unitary similarity transformations on the rest of the extended

Hessenberg matrix brings out the corresponding weight in its first columns and the leading $m \times m$ part gives the solution for the downdated problem. Of course, just as the updating procedure can be generalized, also the downdating procedure can be adapted to our general situation. A combination of downdating and updating provides a tool for least squares approximation with a sliding window, i.e., where a window slides over the data, letting new data enter and simultaneously forgetting about the oldest data.

The inverse QR algorithm that we described in the previous sections is in principle applicable in the situation of arbitrary complex data. However its complexity can be reduced by an order of magnitude if the knots are real or located on the unit circle. The secret of this complexity reduction is the exploitation of a recurrence relation for the corresponding orthogonal polynomials and the parametrization of the Hessenberg matrix involved in terms of the recurrence coefficients.

The polynomial discrete least squares approximation problem discussed in the papers where the algorithm was first conceived gave rise to the construction of a sequence of polynomials orthogonal with respect to a discrete inner product. In the more general problem, these generalize to orthogonal block polynomials when all the degrees of the approximating polynomials are equal, or, in the more general case of arbitrary degrees, both scalar and block orthogonal polynomials appear which can be uniformly treated as vector orthogonal polynomials.

The algorithm has been reported to have excellent numerical stability properties [11, 12] and is to be preferred over the so called Stieltjes procedure [12]. See also [7, 8]. Moreover it is well suited for implementation in a pipeline fashion on a parallel architecture [17, 15].

REFERENCES

[1] G. Ammar and W. Gragg, *$O(n^2)$ reduction algorithms for the construction of a band matrix from spectral data*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 426–431.

[2] G. Ammar, W. Gragg, and L. Reichel, *Constructing a unitary Hessenberg matrix from spectral data*, in Numerical linear algebra, digital signal processing and parallel algorithms, G. Golub and P. Van Dooren, eds., vol. 70 of NATO-ASI Series, F: Computer and Systems Sciences, Berlin, 1991, Springer-Verlag, pp. 385–395.

[3] ———, *Downdating of Szegő polynomials and data-fitting applications*, Linear Algebra Appl., 172 (1992), pp. 315–336.

[4] G. Ammar and C. He, *On an inverse eigenvalue problem for unitary Hessenberg matrices*, Linear Algebra Appl., (1994). To appear.

[5] D. Boley and G. Golub, *A survey of matrix inverse eigenvalue problems*, in Inverse problems, vol. 3, Bristol, England, 1987, Physics Trust Publications, pp. 595–622.

[6] S. Elhay, G. Golub, and J. Kautsky, *Updating and downdating of orthogonal polynomials with data fitting applications*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 327–353.

[7] W. Gautschi, *On generating orthogonal polynomials*, SIAM J. Sci. Statist. Comput., 3 (1982), pp. 289–317.

[8] ———, *Computational problems and applications of orthogonal polynomials*, in Orthogonal polynomials and their applications, C. Brezinski, L. Gori, and A. Ronveaux, eds., vol. 9 of IMACS Annals on Computing and Applied Mathematics, 1991, pp. 61–71.

[9] W. Gragg, *The QR algorithm for unitary Hessenberg matrices*, J. Comput. Appl. Math., 16 (1986), pp. 1–8.

[10] W. Gragg and L. Reichel, *A divide and conquer method for unitary orthogonal eigenproblems*, Numer. Math., 57 (1990), pp. 695–718.

[11] W. B. Gragg and W. J. Harrod, *The numerically stable reconstruction of Jacobi matrices from spectral data*, Numer. Math., 44 (1984), pp. 317–335.

[12] L. Reichel, *Fast QR decomposition of Vandermonde-like matrices and polynomial least squares approximation*, SIAM J. Matrix Anal. Appl., 12 (1991), pp. 552–564.

[13] L. Reichel, G. Ammar, and W. Gragg, *Discrete least squares approximation by trigonometric polynomials*, Math. Comp., 57 (1991), pp. 273–289.
[14] H. Rutishauser, *On Jacobi rotation patterns*, in Proceedings of Symposia in Applied Mathematics, vol. 15 , Experimental Arithmetic, High Speed Computing and Mathematics, Providence, 1963, Amer. Math. Soc., pp. 219–239.
[15] M. Van Barel and A. Bultheel, *A parallel algorithm for discrete least squares rational approximation*, Numer. Math., 63 (1992), pp. 99–121.
[16] ———, *Discrete least squares approximation with polynomial vectors*, Technical Report TW190, Department of Computer Science, K.U. Leuven, May 1993.
[17] ———, *Discrete linearized least squares approximation on the unit circle*, J. Comput. Appl. Math., 50 (1994), pp. 545–563.