Orthogonal basis functions in discrete least squares rational approximation

A. Bultheel, M. Van Barel, P. Van gucht [1,2]

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

Abstract

We consider a problem that arises in the field of frequency domain system identification. If a discrete-time system has an input-output relation $Y(z) = G(z)U(z)$, with transfer function G, then the problem is to find a rational approximation $\hat{G}_n$ for G. The data given are measurements of input and output spectra in the frequency points z_k: $\{U(z_k), Y(z_k)\}_{k=1}^N$ together with some weight. The approximation criterion is to minimize the weighted discrete least squares norm of the vector obtained by evaluating $G - \hat{G}_n$ in the measurement points.

If the poles of the system are fixed, then the problem reduces to a linear least squares problem in two possible ways: by multiplying out the denominators and hide these in the weight, which leads to the construction of orthogonal vector polynomials, or the problem can be solved directly using an orthogonal basis of rational functions. The orthogonality of the basis is important because if the transfer function $\hat{G}_n$ is represented with respect to a non-orthogonal basis, then this least squares problem can be very ill conditioned. Even if an orthogonal basis is used, but with respect to the wrong inner product (e.g., the Lebesgue measure on the unit circle) numerical instability can be fatal in practice.

We show that both approaches lead to an inverse eigenvalue problem, which forms the common framework in which fast and numerically stable algorithms can be designed for the computation of the orthonormal basis.

Key words: approximation, least squares, orthogonal rational functions, system identification

AMS Subject classification: 93B30 41A20 65F35 65D15 42C15

Preprint *17 October 2002*

1 Introduction

This paper wants to illustrate how two approaches to discrete least squares rational approximation can be placed in a unifying framework. It has been proposed in the literature to solve the rational approximation problem, which is obviously nonlinear, by choosing the nonlinear parameters, i.e., by fixing the denominator, and solving the remaining linear problem for the linear parameters. In this survey, we shall consider two techniques that are used to solve the linear problem in an efficient and numerically stable way. The idea is to represent the solution of the linear problem with respect to an appropriate orthogonal basis, much like in the classical Forsythe algorithm [16] for discrete least squares polynomial approximation. During the iteration for the nonlinear parameters, one has to generate this basis dynamically. In the first case, the basis will consist of vector polynomials that have to be orthogonal with respect to a weight that depends upon the choice of the nonlinear parameters. In the second case, the basis consists of rational functions taken from a space that will depend upon the choice of the denominator and which are orthogonal with respect to an arbitrary weight. We shall illustrate that there is some common framework in which both approaches can be formulated.

The overall nonlinear problem is not the main concern of this paper. In the identification literature, there are first of all several cost functions that could be minimized. Covariance matrices and cross-covariance matrices of different noise sources will play a role, several norm functions may be considered etc. This has all been described in the literature and we do not want to enter these problems and stick to the discrete least squares type of problem. Even the simplest nonlinear least squares problem in itself is a difficult problem because the objective function can have several local minima, and it will be impossible to guarantee a global minimum with simple optimization techniques. Much more complicated techniques exist for identification [3,2] that will guarantee to give optimal solutions for more difficult problems. Unfortunately, they are often computationally quite expensive, while for certain classes of problems, the techniques we propose may give rather good results.

Let us start with a motivating application. Consider a linear system

$$y(t) = G(z)u(t) + v(t),$$

with input $u(t)$ and output $y(t)$, while $v(t)$ is some noise with expected value equal

[1] The authors are partially supported by the Fund for Scientific Research (FWO), projects "CORFU: Constructive study of orthogonal functions", grant #G.0184.02 and, "SMA: Structured matrices and their applications", grant G#0078.01, "ANCILA: Asymptotic analysis of the convergence behavior of iterative methods in numerical linear algebra", grant #G.0176.02, the K.U.Leuven research project "SLAP: Structured linear algebra package", grant OT-00-16, and the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture, grant IPA V/22. The scientific responsibility rests with the authors.
[2] Email of corresponding author: Adhemar.Bultheel@cs.kuleuven.ac.be

to zero. In many cases it can be considered to be white noise pushed through some modelling filter. The notation z stands for the shift operator: $zu(t) = u(t+1)$. The variable t refers to time which may be in a continuous or a discrete set. In discrete time $G(z) = \sum_{\tau\in\mathbb{Z}} g(\tau)z^{-\tau}$, while in continuous time $G(z) = \int_{\tau\in\mathbb{R}} g(\tau)z^{-\tau}d\tau$, which leads in both cases to the well known convolution relation in the time domain $y = g * u + v$. This setting is standard. See [23,31] for a thorough treatment of the identification problem.

Let us denote by $F(z)$ the continuous or discrete Fourier transform of $f(t)$, where $z \in \mathbb{R}$ in the first case and $z \in \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ in the second case. If F is rational, it is defined in the whole complex plane $\mathbb{C}$. If it should represent a transfer function, then stability requires that all its poles are inside the open unit disk for the discrete time case, or in the strictly upper half plane for the continuous time case. Furthermore, let us assume that by $\|F\|_w$ we denote the discrete least squares norm $\|F\|_w^2 = \sum_k (w_k F(z_k))^H (w_k F(z_k))$, then, under appropriate assumptions, the problem in its simplest form is the following. Given some frequency information $G(z_k)$, or equivalently $U(z_k)$ and $Y(z_k)$, where $\{z_k\}$ is a set of points either in $\mathbb{R}$ or on $\mathbb{T}$, find a rational approximation $\hat{G}(z)$ such that $\|G - \hat{G}\|_w$ is minimized. The weights w_k can e.g. depend on the accuracy of the measurements. If for example the measurement noise is uncorrelated and if σ_k is the variance of the measurement $G(z_k)$, then we could take $w_k = 1/\sigma_k$. Note that if $\hat{Y} = \hat{G}U + V$, then minimizing $\|Y - \hat{Y}\|_{w^y}$ would be another valid criterion. However, this is equivalent to the previous one because $\|Y - \hat{Y}\|_{w^y} = \|(G - \hat{G})U\|_{w^y} = \|G - \hat{G}\|_{w^g}$ if $w_k^g = w_k^y U(z_k)$.

2 First approach: Vector orthogonal polynomials

In a first method, the problem is linearized by multiplying out the denominators. Let $\hat{G} = B/A$ where for consistency with the systems theory literature, *we assume that $A(z)$ and $B(z)$ are polynomials in the variable z^{-1}*. Hence

$$\begin{aligned}
\|G - \hat{G}\|_{w^g}^2 &= \sum_{k=0}^{N} \left| \Big(Y(z_k)A(z_k) - U(z_k)B(z_k) \Big) w_k \right|^2, \quad w_k = \frac{w_k^g}{A(z_k)U(z_k)} \\
&= \sum_{k=0}^{N} \left| w_k [Y(z_k), \ -U(z_k)] \begin{bmatrix} A(z_k) \\ B(z_k) \end{bmatrix} \right|^2 = \mathbf{P}^H \mathbf{W}^H \mathbf{W} \mathbf{P} \\
&= \sum_{k=0}^{N} |W_k P(z_k)|^2 = \|P\|_W^2,
\end{aligned}$$

where

$$\begin{aligned}
&P(z) = [A(z), B(z)]^T, \quad W_k = w_k[Y(z_k), \ -U(z_k)] \\
&\mathbf{P} = [P(z_0)^T, \dots, P(z_N)^T]^T, \quad \mathbf{W} = \operatorname{diag}(W_0, \dots, W_n).
\end{aligned}$$

Of course, the minimum has to be taken with a degree constraint. For example if the approximant should have a prefixed order n, then P should be of strict degree n, the degree of the vector polynomial P being $\max\{\deg(A), \deg(B)\}$. Thus we have to minimize $\|P\|_W$, that is the norm of the vector polynomial P with respect to a discrete 2×2 matrix valued weight represented by the block diagonal matrix $\mathbf{W}^H\mathbf{W}$. The minimum is given by a vector polynomial from $\mathbb{P}_n^{2\times 1} \setminus \mathbb{P}_{n-1}^{2\times 1}$, orthogonal to $\mathbb{P}_{n-1}^{2\times 1}$. That is a vector polynomial of strict degree n that is orthogonal to all vector polynomials of lower degree where orthogonality is with respect to the inner product

$$\langle f, g\rangle_W = \sum_{k=0}^{N} f(z_k)^H W_k^H W_k g(z_k), \quad f, g \in \mathbb{P}_n^{2\times 1}. \tag{2.1}$$

Note that this is a genuine positive definite inner product as long as $n \leq N/2$ and that the $Y(z_k), U(z_k)$ in the weight vector W_k are not the result of a rational function of degree n or less. Obviously $\|f\|_W^2 = \langle f, f\rangle_W$. The reader should keep in mind that there are two independent vector polynomials in $\mathbb{P}_n^{2\times 1}$ orthogonal to $\mathbb{P}_{n-1}^{2\times 1}$. We shall order them such that the first one has a leading coefficient of the form $[a\ 0]^T$ with $a > 0$ and the second one has a leading coefficient of the form $[c\ d]^T$ with $d > 0$. With this convention, the ith orthogonal vector polynomial (OVP) ϕ_i will have degree $\lfloor \frac{i}{2} \rfloor$, $i = 0, 1, \ldots$. It is therefore more convenient to consider pairs of OVPs with the same degree and denote by $\varphi_i = [\phi_{2i}|\phi_{2i+1}]$ the 2×2 orthogonal block polynomial (OBP) of degree i which contains the two OVPs of degree i as its two columns. If the OVPs are orthonormal, then the OBP are orthonormal in the sense that $\langle \varphi_1, \varphi_j\rangle = \delta_{i,j} I_2$ where the inner product is as in (2.1) where we allow f and g to be elements from $\mathbb{P}_n^{2\times 2}$. They are uniquely defined because their leading coefficient is an upper triangular matrix in $\mathbb{C}^{2\times 2}$ with positive diagonal elements. There may be an exception for the last OVP if N is even, but if we assume for simplicity that N is odd, we find that the optimal polynomial $P = \sum_{i=0}^{n} \varphi_i \vartheta_i$ with $\vartheta_i \in \mathbb{C}^{2\times 1}$ is given by $\varphi_n \vartheta_n$ for some $0 \neq \vartheta_n \in \mathbb{C}^{2\times 1}$ that minimizes $\|P\|_W^2 = \|\varphi_n \vartheta_n\|_W^2 = \|\vartheta_n\|^2$. Any normalisation with $0 \neq \vartheta_n$ will do because P contains numerator and denominator of the approximant and is therefore only defined up to a constant multiple. The construction of these OVPs can be done by a fast and numerically stable procedure.

The theory and computational aspects as well as the use of discrete rational approximation on the unit circle and the real line has been reported on in several papers to which we refer for further details [34–36,6]. For an application in system identification see [7]. We just summarize the main results for further reference.

The procedure can be formulated as an inverse eigenvalue problem.

Theorem 2.1 *[6] Let* $\mathbf{w}$ *be the* $(N+1) \times 2$ *matrix whose* k*th row is* W_k*,* $k = 0, 1, \ldots, N$*, and let* $\mathbf{w}_1$ *be the* $(N+1) \times 2$ *matrix whose first* 2×2 *block is an upper triangular Cholesky factor of* $\mathbf{w}^H\mathbf{w}$ *with positive diagonal elements and that has zeros everywhere else, and finally let* $\mathbf{Z}$ *be the diagonal matrix* $\operatorname{diag}(z_0, \ldots, z_N)$*.*

Then there exists a unique unitary matrix $\mathbf{Q}$ *such that*

$$\mathbf{Q}^H[\mathbf{w}|\mathbf{Z}^{-1}]\left[\begin{array}{c|c} I_2 & \\ \hline & \mathbf{Q} \end{array}\right] = [\mathbf{w}_1|\mathbf{H}]$$

with $[\mathbf{w}_1|\mathbf{H}]$ *a matrix whose lower triangular part is zero and which has positive elements on its main diagonal. Denote by* $\mathbf{\Phi}$ *the* $2(N+1)\times(N+1)$ *matrix* $[\phi_i(z_k)]$, $i = 0,\ldots,N$, $k = 0,\ldots,N$, *where* ϕ_i *is the* i*th orthogonal vector polynomial then* $\mathbf{Q} = \mathbf{W}\Phi$.

If $z_k \in \mathbb{R}$, $k = 0,\ldots,N$ *then* $\mathbf{H}$ *is Hermitian, hence pentadiagonal.*

If $z_k \in \mathbb{T}$, $k = 0,\ldots,N$ *then* $\mathbf{H}$ *is unitary.*

Remark 2.2 Several remarks are in place here.

(1) Note that this is an inverse eigenvalue problem. $\mathbf{Z}^{-1}$ are the eigenvalues of the matrix $\mathbf{H}$ and the columns of $\mathbf{Q}^H$ are the eigenvectors. The first two elements of the eigenvectors are fixed by the condition $\mathbf{Q}\mathbf{w}_1 = \mathbf{w}$. The problem is to find the matrix $\mathbf{H}$ and the eigenvectors.

(2) The unitary similarity transformations can be computed in a recursive way. I.e., we can add the data (W_k, z_k) for $k = 2, 3, \ldots$ and recursively update the matrix $\mathbf{H}$. Because this matrix has a well defined structure (Hermitian or unitary), the overall algorithm will require only $O(N^2)$ operations.

(3) In general, the degree n of the approximant will be much smaller than N, the number of measurements. In that case, we only need to compute the first $2n$ columns of $\mathbf{H}$, in which case the complexity reduces to $O(nN)$ operations.

(4) The matrix $\mathbf{H}$ is the key result because it contains the recurrence relation for the OVPs. It is most convenient to formulate this in terms of the OBP φ_i when we also think of the matrix $\mathbf{H}$ as consisting of 2×2 blocks. Because the OBP had upper triangular leading coefficients with positive diagonal elements, also the subdiagonal blocks of $\mathbf{H}$ have this property. The top block of $\mathbf{w}_1$ is φ_0^{-1}. The subsequent φ_i can then recursively be computed using the block upper Hessenberg matrix $\mathbf{H}$. In the continuous time case ($z_k \in \mathbb{R}$), $\mathbf{H}$ is a block tridiagonal Jacobi matrix, for the OBP. In the discrete time case ($z_k \in \mathbb{T}$), $\mathbf{H}$ is a block upper Hessenberg matrix that for computational efficiency should be stored in factored form. Each factor contains block Schur parameters for the OBP. The recursion is a generalization of the Szegő recurrence for polynomials orthogonal on $\mathbb{T}$. Note that there is no need to explicitly compute the vector polynomial that solves the identification problem. Evaluation through the recurrence relation implicitly or explicitly stored in $\mathbf{H}$ requires only $O(n)$ operations for every point in which we want it evaluated.

(5) In the identification problems, the measurements may be real, but for the discrete time case, the points z_k are complex. However, using $G(e^{-i\omega}) = \overline{G(e^{i\omega})}$, we can rearrange the complex pair $\mathrm{diag}(e^{i\omega_k}, e^{-i\omega_k})$ in the matrix $\mathbf{Z}$ by unitary

similarity transformation as a real 2×2 block with real entries. Hence, updating the associated weights W correspondingly, the computations can be performed using only real arithmetic.

(6) The algorithm is easily generalized for the case where the vector polynomials have size $p \times 1$ where p can be larger than 2. Moreover, it is possible to assign different maximum degrees for each of the components of the optimal solution that we are looking for. This generalization to $p > 2$ is important for the MIMO case. The idea to solve the MIMO case with these techniques is that the matrix numerator and denominator of the polynomial matrix fraction description of the approximant are stored column by column in a long vector. For a system with p inputs and q outputs, this results in vector polynomials of length $(p+q)q$ or $(p+q)p$ depending on whether a left or a right matrix fraction description is considered. Then the more general algorithm is applicable. This means that the technique will only be of interest for moderately small p and q. See [30].

To summarize this approach we reconsider the original nonlinear rational approximation problem. First some estimate for the denominator A is chosen. This fixes the weight W and a solution can be computed in a numerically stable and efficient way. The approximant itself is not computed explicitly, but it is stored in the form of a recurrence relation that allows to compute its value for any z very efficiently. With the new value of A, a new weight is defined and the process is repeated. In each iteration step, the weight is changed and the basis in which the solution is represented is optimal (condition number 1) and is recomputed on the fly. However, it is well known that this Sanathanan-Koerner type of iteration [9] does not guarantee convergence to a minimum. So it is hoped that it will converge to something that is close enough to the true minimum and that, with the last basis fixed, and with the ultimate solution as a starting point, a true maximum likelihood optimization algorithm can be applied to find the true minimum of the nonlinear problem. If the starting point is close enough to the true solution, then the weights will only change a bit and the loss of orthogonality will only be moderate. Note however, that there is no guarantee that this will work and give a global minimum that is hoped for. However, for moderately complicated problems, it may give quite good results in an efficient way. If it does not work out that way, more involved techniques are needed [3,2].

3 Second approach: orthogonal rational functions

In a second approach, the approximant $\hat{G}$ is represented with respect to an orthogonal basis of rational functions. Suppose again that we have some estimate of the denominator in the sense that we know all the poles of the approximant, say $\alpha_1, \ldots, \alpha_n$. Then the approximant should belong to the space

$$\mathcal{L}_n = \left\{ \frac{p_n(z)}{\prod_{j=1}^{n}(1 - \alpha_k/z)} : p_n \in \mathbb{P}_n \right\},$$

where $\mathbb{P}_n$ is the space of polynomials in $1/z$ of degree at most n. For stability, the α_k should be in $|z| < 1$ for the discrete time case and in $\mathrm{Im}(z) > 0$ for the continuous time case.

Assume this space is spanned by a set of rational basis functions $\{\phi_k\}_{k=0}^n$ so that we can write

$$\hat{G}(z) = \sum_{k=0}^{n} \theta_k \phi_k(z)$$

with the parameter vector $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_n]^T$ to be chosen in an optimal way.

For numerical (and many other theoretical [28]) reasons, one should choose the basis functions ϕ_k to be orthogonal. However, when considering the numerical problem, recalling that we have to minimize the least squares error $\|E\|_w$ for some weight w where $E = G - \hat{G}$, the orthogonality we need is with respect to the inner product

$$\langle f, g \rangle_w = \sum_{k=0}^{N} [w_k f(z_k)]^H [w_k g(z_k)], \tag{3.1}$$

since then $\|E\|_w^2 = \langle E, E \rangle_w$.

We shall discuss the matter of constructing an orthonormal basis in subsequent sections, but whatever the choice of the ϕ_k is, and whatever computational method is used, the idea of the nonlinear rational approximation problem is to solve the overdetermined system $\boldsymbol{\Phi\theta} = \mathbf{G}$ in a weighted least squares sense, where $\boldsymbol{\Phi}$ is the $(N+1) \times (n+1)$ matrix with entries $\phi_l(z_k)$, $k = 0, \ldots, N$, $l = 0, \ldots, n$ and $\mathbf{G} = [G(z_0), \ldots, G(z_N)]^T$. This solution is $\boldsymbol{\theta} = \boldsymbol{\Psi}^\dagger \mathbf{W}\mathbf{G}$ with $\boldsymbol{\Psi} = \mathbf{W}\boldsymbol{\Phi}$, $\mathbf{W} = \mathrm{diag}(w_0, \ldots, w_N)$ and $\boldsymbol{\Psi}^\dagger$ denotes the Moore-Penrose inverse. Since $\boldsymbol{\Phi}$ depends on the choice of the poles α, we get a residual vector $\mathbf{E} = \mathbf{G} - \boldsymbol{\Phi\theta}$ that will depend on $\boldsymbol{\alpha}$. Solving the nonlinear least squares problem $\min_{\boldsymbol{\alpha}} \mathbf{R}^H\mathbf{R}$ where $\mathbf{R} = \mathbf{W}\mathbf{E} = (\mathbf{I} - \boldsymbol{\Psi\Psi}^\dagger)\mathbf{W}\mathbf{G}$ will give a solution. It should be noted however that the problem is not convex, so that there is a real danger that a numerical algorithm will be trapped in a local minimum. Moreover, optimization should be done such that the α_k are restricted to the open unit disk. In cases where the poles should be real or complex conjugate, it is worthwhile to write $(z - \alpha)(z - \bar{\alpha}) = z^2 + bz + c$ with $b, c \in \mathbb{R}$ and solve the nonlinear routine with respect to the parameters (b, c) instead of $(\alpha, \bar{\alpha})$.

So it remains to design a numerically stable and efficient algorithm to solve the linear subproblem of each iteration step by constructing the ϕ_k as a set of appropriate Orthogonal Rational Functions (ORFs).

The idea of representing the transfer function G as a linear combination of ORFs with prescribed poles is quite old, and it has received much attention in the identification literature lately [1,10,11,13,19,21,22,26–29,32,38,41,42].

However, these papers almost always restrict the discussion to rational functions that are orthogonal with respect to a uniform continuous (Lebesgue) weight, mostly for discrete time systems. One of the reasons is probably that an explicit expression for these basis functions exists. Indeed, it is well known that the system $\{\tilde{\phi}_i\}_{i=0}^n$ with

$$\tilde{\phi}_0 = 1, \quad \tilde{\phi}_i(z) = \frac{\sqrt{1-|\alpha_i|^2}}{1-\alpha_i/z} B_{i-1}(z), \quad B_i(z) = \prod_{j=1}^{i} \frac{1/z - \overline{\alpha}_j}{1-\alpha_j/z}, \quad i = 1, 2, \ldots \tag{3.2}$$

is an orthonormal system for the inner product $\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{f(e^{i\omega})} g(e^{i\omega}) d\omega$. It is the so-called Takenaka-Malmquist basis [33,24]. It generalizes the Laguerre, Kautz and Hambo bases [20,18,12,11,27]. Another nice thing about these bases is that it can be rather directly generalized to MIMO systems if only a finite number of different poles is selected that are cyclically repeated. Indeed one can construct a system of ORFs for the Lebesgue measure on $\mathbb{T}$ starting from any rational inner function H (that is a matrix valued rational function satisfying $[H(z)]^H H(z) = I$ for $z \in \mathbb{T}$ and which has all its poles in the open unit disk). Taking a balanced minimal realization (that is writing $H(z) = C(zI - A)^{-1}B + D$, such that the size of A is minimal and such that $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is unitary) immediately gives an orthonormal basis.

Theorem 3.1 *[11,12] Let H be a scalar inner function and (A, B, C, D) its minimal balanced realization, then the elements of the vectors $V_k(z) = (zI - A)^{-1}B[H(z)]^k$, $k = 0, 1, \ldots$ form an orthonormal set of rational functions whose poles are the eigenvalues of A and hence the poles of H that are cyclically repeated.*

Given a vector of ORFs in the form, $V(z) = (zI - A)^{-1}B$, a vector of ORFs with additional poles of an inner function with minimal balanced state space realization (A', B', C', D') is obtained in the form $\tilde{V}(z) = (zI - \tilde{A})^{-1}\tilde{B}$ where

$$\tilde{A} = \left[\begin{array}{c|c} A & 0 \\ \hline B'C & A' \end{array}\right], \quad \tilde{B} = \left[\begin{array}{c} B \\ \hline B'D \end{array}\right].$$

Suppose we use this basis for finding the best discrete least squares approximant, then we have to solve the overdetermined system $\mathbf{W}\tilde{\mathbf{\Phi}}\boldsymbol{\theta} = \mathbf{WG}$ in (unweighted) least squares sense where $\mathbf{W} = \mathrm{diag}(w_0, \ldots, w_N)$, $\mathbf{G} = [G(z_0), \ldots, G(z_N)]^T$ and $\tilde{\mathbf{\Phi}}$ is the $(N+1) \times (n+1)$ matrix with elements $\tilde{\phi}_l(z_k)$, $k = 0, \ldots, N$, $l = 0, \ldots, n$,

and $\boldsymbol{\theta} = [\theta_0, \ldots, \theta_n]^T$, giving the optimal solution $\hat{G} = \sum_{i=0}^{n} \tilde{\phi}_i \theta_i$. Since the ORFs $\tilde{\phi}_i$ are not orthogonal with respect to the discrete inner product, it may not come as a surprise that the condition number of the matrix $\mathbf{W}\tilde{\mathbf{\Phi}}$ can be very large.

Example 3.2 [38] We consider the Takanaka-Malmquist basis with poles $.9 \pm .3i$ and $.7 \pm .2i$ repeated 5 times as in [38], the condition number of $\tilde{\mathbf{\Psi}} = \mathbf{W}\tilde{\mathbf{\Phi}}$ for $N = 200$, and $n = 20$ is $O(10^{17})$.

However, this need not be a disaster. A condition number that big means that when computations are done in IEEE standard double precision, the columns of $\tilde{\mathbf{\Psi}}$ are linearly dependent. The true least squares solution $\boldsymbol{\theta}$ and its computed value $\hat{\boldsymbol{\theta}}$ may differ considerably, but this does not mean that $\sum_l \tilde{\phi}_l \theta_l$ and $\sum_l \tilde{\phi}_l \hat{\theta}_l$ will be much different. For example, when solving the system in Matlab with the backslash operator, it will put the $\hat{\theta}_l$ that corresponds to the linearly dependent column equal to zero, and the approximation of G does not suffer from the ill conditioning. $\Diamond$

3.2 General orthogonal ORFs for a discrete measure

Although the previous basis may be used to solve the problem, it is also possible to construct an orthonormal basis of rational functions that is custom designed for the problem at hand, i.e., orthogonality with respect to the inner product (3.1). This will yield a Jacobian matrix with condition number 1, and will thus avoid any numerical instabilities. It is in general not possible to give an explicit expression for these ORFs, but the computational effort is comparable.

The theory of ORFs with prescribed poles orthogonal on $\mathbb{T}$ and $\mathbb{R}$ with respect to a general positive measure is described in [4]. The poles were located in $|z| > 1$ for the circle and in $\mathrm{Im}(z) < 0$ for the line, but switching the poles to be in $|z| < 1$ or $\mathrm{Im}(z) > 0$ is but a trivial exercise. See [39] for the discrete time case.

We intend however to give the explicit relation with the approach by OVPs that was summarized above. This will show that, at least in the scalar case, the present approach is conceptually simpler.

The idea is again that we assume that an estimate of the denominator is given. More precisely, suppose we know the poles of the approximant. Whereas in Section 2, we multiplied out the denominator (and absorbed it in the weight), to obtain a linear problem for a vector polynomial, this may be considered as an unnecessary complication. If the denominator is given, then the problem is linear as it is, and it can be solved with linear algebra techniques. To work with an orthonormal basis, we need to generalize the theory of orthogonal polynomials (on $\mathbb{R}$ or on $\mathbb{T}$) in one way or another. For the OVPs we needed a vector/matrix generalization. For the ORF case, we need a generalization from polynomials to rational functions. Both are generalizations in the sense that a scalar polynomial is a 1×1 polynomial and it

is also a rational function with all poles at the origin (if we work, as we do here, with polynomials in the variable z^{-1}). As we mentioned before, the theory of ORFs has been developed elsewhere, but we want to emphasize the specific relation with the OVP approach, and we shall therefore develop this view explicitly. The symmetric case was considered in [37] from a different point of view, but as far as we know, the relation has not been published explicitely yet.

We start with the following lemma.

Lemma 3.3 *Suppose $\phi_k(z) = p_k(z)/\pi_k(z)$ with $\pi_k(z) = \prod_{i=1}^{k}(z-\alpha_i)$, ($p_k(z)$ is a polynomial in z of degree at most k) and assume $p_k(\alpha_{k-1}) \neq 0$ for $k = 1, 2, \ldots, k+1$. Then, setting $\alpha_0 = 0$,*

$$\phi_k(z) \in \text{span}\{(z-\alpha_i)\phi_i(z) : i = 0, \ldots, k+1\}.$$

Thus there exist numbers $\eta_{i,k}$ such that

$$\phi_k(z) = (z-\alpha_0)\phi_0(z)\eta_{0k} + \cdots + (z-\alpha_{k+1})\phi_{k+1}(z)\eta_{k+1,k}.$$

Moreover $\eta_{i+1,i} \neq 0$, $i = 1, \ldots, k$.

PROOF. We give only a sketch in the case that all the α_k are mutually different. We have to prove that there exist some numbers η_{ik}, $i = 0, \ldots, k+1$ such that $\phi_k(z) = \sum_{i=0}^{k+1}(z-\alpha_i)\phi_i(z)\eta_{ik}$. If we multiply this relation by π_k, this becomes

$$p_k(z) = p_0\pi_{0:k}(z)\eta_{0k} + p_1(z)\pi_{1:k}(z)\eta_{1k} + \cdots + p_k(z)\pi_{k:k}(z)\eta_{kk} + p_{k+1}(z)\eta_{k+1,k}, \tag{3.3}$$

where $\pi_{i:j}(z) = \prod_{l=i}^{j}(z-\alpha_l)$. To show that these numbers exist, we note first that $p_k(z)$ is the unique polynomial in the set of polynomials of degree at most $k+1$ that has highest degree coefficient equal to zero and that takes the values $p_k(\alpha_i) = \lim_{z\to\alpha_i}[\phi_k(z)\pi_k(z)]$ in the points α_i for $i = 0, 1, \ldots, k$. Thus if we can find the numbers $\eta_{i,k}$, $i = 0, \ldots, k+1$ such that the polynomial in the right hand side of (3.3) has this property, it has to be equal to $p_k(z)$.

Replacing in (3.3) z by α_i for $i = 0, 1, \ldots, k$ gives

$$p_k(\alpha_i) = p_{i+1}(\alpha_i)\pi_{i+1:k}(\alpha_i)\eta_{i+1,k} + p_{i+2}(\alpha_i)\pi_{i+2:k}(\alpha_i)\eta_{i+2,k} + \cdots$$

Add to this an equation expressing the fact that the coefficient of z^{k+1} in the right hand side has to be zero, which is of the form $0 = \kappa_0\eta_{0,k} + \ldots + \kappa_{k+1}\eta_{k+1,k}$ where $p_i(z) = \kappa_i z^i + O(z^{i-1})$. Then we get a triangular system of $k+2$ linear equations whose diagonal elements are κ_0 and $p_{i+1}(\alpha_i)\pi_{i+1:k}(\alpha_i)$, $i = 0, 1, \ldots, k$ which are all nonzero by our assumption. So there exists a (unique) set of coefficients $\eta_{0k}, \ldots, \eta_{k+1,k}$ that solves the system.

In the case of coinciding poles α_i, a similar derivation can be made, where we evaluate $p_k(\alpha_i)$, $p_k'(\alpha_i)$, $p_k''(\alpha_i)$, etc., taking as many derivatives as the multiplicity of α_i requires. This leads to the same conclusion.

The last coefficient $\eta_{k+1,k}$ can not be zero, because then $p_k(\alpha_k) = p_{k+1}(\alpha_k)\eta_{k+1,k}$ would be zero, implying that in ϕ_k the factor $(z-\alpha_k)$ can be canceled in numerator and denominator which would mean that $\phi_k \in \mathcal{L}_{k-1}$, which is impossible. □

If the ORF system satisfies $p_k(\alpha_{k-1}) \neq 0$ for $k = 1, \ldots, N$ we shall say that it is *regular*. We assume from now on that the system is regular.

The choice $\alpha_0 = 0$ is not necessary. Any other choice for α_0 such that $p_1(\alpha_0) \neq 0$ will do.
Also the orthogonality is not used. Any set of rational functions $\phi_i \in \mathcal{L}_i \setminus \mathcal{L}_{i-1}$, $i = 1, \ldots, k$ with $\phi_0 \neq 0$ satisfying the regularity condition will give the same result.

Thus we may write

$$\Phi(z) = \Phi(z)(zI - \mathbf{A})\mathbf{H} + [0 \ \cdots \ 0 \ \eta_{n+1,n}(z-\alpha_{n+1})\phi_{n+1}(z)] \tag{3.4}$$

where

$$\Phi(z) = [\phi_0, \phi_1, \ldots, \phi_n], \qquad \mathbf{A} = \mathrm{diag}(\alpha_0, \alpha_1, \ldots, \alpha_n), \qquad \text{and} \quad \mathbf{H} = \begin{bmatrix} \eta_{00} & \cdots & \eta_{0,n-1} & \eta_{0,n} \\ \eta_{10} & \cdots & \eta_{1,n-1} & \eta_{1,n} \\ & \ddots & \vdots & \vdots \\ & & \eta_{n,n-1} & \eta_{n,n} \end{bmatrix}.$$

The ORFs will be uniquely defined if we impose that the subdiagonal elements $\eta_{k+1,k}$ are all positive (recall they are nonzero by Lemma 3.3).

Since the discrete measure has only $N+1$ mass points, the Gram-Schmidt procedure would find $\|\phi_{N+1}\|_w = 0$. This does not mean that $\phi_{N+1} = 0$. It only means that $\phi_{N+1}(z_k) = 0$ for $k = 0, 1, \ldots, N$. So we can choose an arbitrary α_{N+1} and set $\phi_{N+1}(z) = \prod_{j=0}^{N}(z - z_j)/\pi_{N+1}(z)$.

Therefore, we may write down the relation (3.4) for $n = N$ and letting $z = z_1, \ldots, z_N$ to get $\mathbf{\Phi} = (\mathbf{Z\Phi} - \mathbf{\Phi A})\mathbf{H}$ where $\mathbf{Z} = \mathrm{diag}(z_0, z_1, \ldots, z_N)$ and $\mathbf{\Phi}$ is the square matrix of size $N+1$ whose (k,l) element is $\phi_l(z_k)$.

Defining $\mathbf{W} = \mathrm{diag}(w_0, \ldots, w_N)$ and $\mathbf{Q} = \mathbf{W\Phi}$, we get

$$\mathbf{Q} = (\mathbf{ZQ} - \mathbf{QA})\mathbf{H}. \tag{3.5}$$

Lemma 3.4 *The upper Hessenberg matrix* $\mathbf{H}$ *is invertible.*

PROOF. We know that $\mathbf{Q}$ is a unitary matrix because the orthogonality of the ϕ_i implies $\mathbf{Q}^H\mathbf{Q} = \mathbf{\Phi}^H\mathbf{W}^H\mathbf{W}\mathbf{\Phi} = \mathbf{I}$. Therefore $|\det(\mathbf{Q})| = 1$.

Because $\mathbf{Q} = (\mathbf{ZQ} - \mathbf{QA})\mathbf{H}$, and thus $\det(\mathbf{Q}) = \det(\mathbf{ZQ} - \mathbf{QA})\det(\mathbf{H}) \neq 0$, $\mathbf{H}$ will be an invertible matrix. □

The inverse of an irreducible Hessenberg matrix, i.e., whose subdiagonal elements are nonzero, has been studied in the literature and is known to be a semiseparable matrix [17] (see also the proof in the appendix). That is a matrix whose lower triangular part is the lower triangular part of a rank one matrix. For the discussion of semiseparable matrices, it is convenient to introduce some Matlab-like notation: $\mathrm{tril}(\mathbf{M}, k)$ denotes the matrix that is the same as the matrix $\mathbf{M}$ on and below the kth diagonal ($k > 0$ is above the main diagonal and $k < 0$ is below the main diagonal. All its other elements are zero. $\mathrm{tril}(\mathbf{M}) = \mathrm{tril}(\mathbf{M}, 0)$. Similarly $\mathrm{triu}(\mathbf{M}, k)$ refers to the elements up and above the kth diagonal. Then $\mathbf{M}$ being semiseparable means that there exist vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\mathrm{tril}(\mathbf{M}) = \mathrm{tril}(\mathbf{uv}^H)$. The remaining part $\mathrm{triu}(\mathbf{M}, 1)$ has in general no specific structure.

Thus there exists a semiseparable matrix $\mathbf{S} = \mathbf{H}^{-1}$, and the relation (3.5), after multiplication with $\mathbf{Q}^H$ can be rewritten to arrive at

$$\mathbf{S} + \mathbf{A} = \mathbf{Q}^H\mathbf{ZQ}. \tag{3.6}$$

To completely define this inverse eigenvalue problem (the z_k are the eigenvalues of the semiseparable-plus-diagonal matrix $\mathbf{S} + \mathbf{A}$), we need an extra condition. This can be found as follows.

Lemma 3.5 *With the notation introduced before and with* $\mathbf{w} = [w_0, w_1, \ldots, w_N]^T$, *we have*

$$\mathbf{Q}^H\mathbf{w} = \mathbf{w}_1, \quad \mathbf{w}_1 = [\|\mathbf{w}\|, 0, \ldots, 0]^T.$$

PROOF. Because $\langle 1, 1\rangle_w = \|\mathbf{w}\|^2 = \phi_0^{-2}$ (note that we may consider ϕ_0 to be the element $\eta_{0,-1}$, so that in harmony with $\eta_{k+1,k} > 0$, we choose $\phi_0 > 0$ too), and because $\langle \phi_k, 1\rangle_w = \delta_{k0}\phi_0^{-1}$, we have

$$\mathbf{Q}^H\mathbf{w} = \mathbf{\Phi}^H\mathbf{W}^H\mathbf{W}[1, 1, \ldots, 1]^T = [\phi_0^{-1}, 0, \ldots, 0]^T,$$

which gives the result. □

Note that this relation prescribes the first column of $\mathbf{Q}$, which are the first components of the eigenvectors that are stored as the columns of $\mathbf{Q}^H$.

So we finally arrive at the complete inverse eigenvalue problem.

Theorem 3.6 *Let* $\mathbf{w}$ *be the vector whose* k*th element is* w_k*,* $k = 0, 1, \ldots, N$*, and let* $\mathbf{w}_1$ *be the vector whose first element is* $\|\mathbf{w}\|$ *followed by* N *zeros, let* $\mathbf{A} = \mathrm{diag}(\alpha_0, \ldots, \alpha_N)$ *(*$\alpha_0 = 0$*), and finally let* $\mathbf{Z} = \mathrm{diag}(z_0, \ldots, z_N)$*. Define the unitary matrix* $\mathbf{Q}$ *such that*

$$\mathbf{Q}^H[\mathbf{w}|\mathbf{Z}] \left[\begin{array}{c|c} 1 & \\ \hline & \mathbf{Q} \end{array}\right] = [\mathbf{w}_1|\mathbf{S} + \mathbf{A}].$$

with $\mathbf{S}$ *a semiseparable matrix. Denote by* $\mathbf{\Phi}$ *the square matrix* $[\phi_i(z_k)]$*,* $k = 0, \ldots, N$*,* $i = 0, \ldots, N$*, where* ϕ_i *is the* i*th orthonormal rational function, then* $\mathbf{Q} = \mathbf{W}\Phi$*.*

If $z_k \in \mathbb{R}$*,* $k = 0, \ldots, N$ *then* $\mathbf{S} + \mathbf{A}$ *is Hermitian.*

If $z_k \in \mathbb{T}$*,* $k = 0, \ldots, N$ *then* $\mathbf{S} + \mathbf{A}$ *is unitary.*

PROOF. All the ingredients for the proof were given above. The special properties for $\mathbf{S} + \mathbf{A}$ follow immediately from (3.6). □

Corollary 3.7 *If all* $z_k \in \mathbb{R}$*, then the strictly upper triangular part of the matrix* $\mathbf{S}$ *has rank 1, i.e.,* $\mathrm{triu}(\mathbf{S}, 1) = \mathrm{triu}(\mathbf{R}, 1)$ *with* $\mathbf{R}$ *of rank 1.*
If all $z_k \in \mathbb{T}$*, and if all* $\alpha_i \neq 0$*,* $i = 1, \ldots, N$*, then* $\mathrm{triu}(\mathbf{S}, 1) = \mathrm{triu}(\mathbf{R}, 1)$ *with* $\mathbf{R}$ *of rank 1. If however* m *of the* α_k*,* $k = 1, \ldots, n$ *are zero, then* $\mathrm{triu}(\mathbf{S}, 1) = \mathrm{triu}(\mathbf{R}, 1)$ *with* $\mathbf{R} = \mathrm{diag}(\mathbf{R}_0, \ldots, \mathbf{R}_m)$ *with all* $\mathbf{R}_i$ *of rank 1 for* $i = 0, \ldots, m$*.*

For a proof we refer to the Appendix.

Remark 3.8 The following comments can be made.

(1) Although $\mathbf{S} + \mathbf{A}$ has a special form that is easily characterized, the Hessenberg matrix $\mathbf{H} = \mathbf{S}^{-1}$ will have a corresponding structure, which is not easily described. For example, in the case of the real line, $\mathbf{H}$ will not be Hermitian, except when the α_k are real (which they can not be for stability reasons). Hence there will not be a three term recurrence relation in this case. Similarly for the case of the unit circle. If the α_k belong to the boundary ($\mathbb{R}$ or $\mathbb{T}$), we do have a three term recurrence, as given in [4, p. 261] and [5] (modulo a reflection in the boundary).

(2) Although the lack of a three term recurrence relation says that there is not a relation of the form $\varepsilon_k(z - \alpha_k)\phi_k(z) = (\beta_k + z\gamma_k)\phi_{k-1}(z) + \delta_k\phi_{k-2}(z)$, we do have a Szegő-type recurrence. That is, if we replace $\phi_{k-2}(z)$ by $(z - \alpha_k)\phi^*_{k-1}(z)$ in this expression, then these numbers $(\beta_k, \gamma_k, \delta_k, \varepsilon_k)$ do exist. These reciprocal functions are defined as

$$\phi_{k-1}^*(z) = \left(\prod_{i=1}^{k-1} \frac{z-\bar{\alpha}_i}{z-\alpha_i}\right) \overline{\phi_{k-1}(\bar{z})}, \quad \text{in the case of } \mathbb{R}$$

$$\phi_{k-1}^*(z) = \left(\prod_{i=1}^{k-1} \frac{1-\bar{\alpha}_i z}{z-\alpha_i}\right) \overline{\phi_{k-1}(1/\bar{z})}, \quad \text{in the case of } \mathbb{T}.$$

See [4, p. 75] where explicit expressions for the $(\beta_k, \gamma_k, \delta_k, \varepsilon_k)$ can be found. We have

$$e_k(z-\alpha_k)\phi_k(z) = (c_k z + d_k)\phi_{k-1}(z) + \bar{\rho}_k(z-\alpha_k)\phi_{k-1}^*(z), \tag{3.7}$$

where $c_k z + d_k$ is $z - \bar{\alpha}_{k-1}$ for $\mathbb{R}$ and $1 - \bar{\alpha}_{k-1} z$ for $\mathbb{T}$. The ρ_k are the generalizations of the Schur-Szegő parameters, and e_k is a normalizing constant. This recurrence does allow for a fast $O(nN)$ algorithm to compute the orthogonal functions up to degree n.

(3) The vector Φ of ORFs can also be represented in the form $\Phi^T(z) = (zI - A)^{-1}B$ like for the Takenaka-Malmquist basis and the A and B can be recursively updated using the ρ_k parameters when new poles are added, generalizing Theorem 3.1. See [40].

(4) The previous result says that both cases ($\mathbb{R}$ and $\mathbb{T}$) are similar, but that the case of the circle is "more general". We have indeed $\mathbf{S} = \text{tril}(\mathbf{u}\mathbf{v}^H) + \text{triu}(\mathbf{r}\mathbf{q}^H, 1)$ when $\alpha_i \neq 0$, for $i = 1, \ldots, N$. In the case of the real line, we have the extra symmetry that $\mathbf{r} = \mathbf{v}$ and $\mathbf{q} = \mathbf{u}$. In the case of the unit circle, also the $\mathbf{r}$ and $\mathbf{q}$ are related to the vectors $\mathbf{u}$ and $\mathbf{v}$ by the unitarity of $\mathbf{S} + \mathbf{A}$, but it is less explicit. So whereas the updating of the $\mathbf{H}$ matrix was quite different for $\mathbb{R}$ and for $\mathbb{T}$ (at least in the polynomial case, see [6,35,36]), the update of $\mathbf{S} + \mathbf{A}$ is quite similar for both cases. So we shall pursue the idea of working with $\mathbf{S}$ instead of $\mathbf{H}$ somewhat further and see how this can lead to a fast algorithm that is an alternative for the recurrence (3.7).

(5) The polynomials are a special case that is obtained by setting all $\alpha_k = 0$. In this case $\mathbf{S}$, and hence also $\mathbf{H}$ is Hermitian and thus tridiagonal in the case of all $z_k \in \mathbb{R}$, which gives the classical three term recurrence relation. In the case of all $z_k \in \mathbb{T}$, then $\mathbf{S}$ and thus also $\mathbf{H}$ is unitary, which corresponds to the classical Schur-Szegő recurrence relation, which is (3.7) where all α's are set to zero.

Also the OVPs that were discussed in Section 2 can be derived and treated in a similar way. The result would be as in Theorem 2.1 where instead of using $\mathbf{H}$, the same result is expressed using $\mathbf{S} = \mathbf{H}^{-1}$. This $\mathbf{S}$ is then the inverse of a block upper Hessenberg matrix and therefore semiseparable of order 2, i.e., such that $\text{tril}(\mathbf{S}) = \text{tril}(\mathbf{u_1}\mathbf{v}_1^H + \mathbf{u_2}\mathbf{v}_2^H)$. For a proof of this fact we refer to the appendix. The result is formally as in Theorem 3.6, but with $\mathbf{w}$ and $\mathbf{w}_1$ having the meaning as in Theorem 2.1 and $\mathbf{A} = \mathbf{0}$.

(6) The previous generalization to handle block Hessenberg matrices has also another application because like in the case of OVPs, if the poles are real or come in complex conjugate pairs, and the associated weights are also complex conjugate then the computations can be done in real arithmetic if the algorithm

sketched below is generalized to handle a block diagonal matrix $\mathbf{A}$.

To illustrate that fast algorithms can be designed, we give in the appendix an example of the computation of $\mathbf{S}+\mathbf{A}$ in the scalar ORF case. For simplicity we assume that we are in the case where triu$(\mathbf{S},1)$ is part of a rank 1 matrix, i.e., we are in the case of the real line, or we are in the case of the circle, but with all $\alpha_k \neq 0$ for $k \geq 1$. Then it is possible to design a fast, i.e., an $O(N^2)$ algorithm to compute the whole matrix $\mathbf{S}$. If we are in the circle case, with some of the $\alpha_k = 0$, $k \geq 0$, then other algorithms can be designed with an $O(N^2)$ complexity, but they are more involved.

An operation count shows that the overall algorithm to compute the $(N+1)\times(N+1)$ matrix $\mathbf{S}$ requires $O(N^2)$ operations. Of course, if we need only the first n ORFs, the computation can be reduced to $O(nN)$ complexity.

Although $\mathbf{SH} = \mathbf{I}$, this does not hold for the leading submatrices. Denoting by $\mathbf{M}_{i:j;k:l}$ the submatrix with elements from rows $i,\ldots,j$ and columns $k,\ldots,l$, of the matrix $\mathbf{M}$, and shortening $k:k$ to k, we define $\mathbf{S}_n = \mathbf{S}_{0:n;0:n}$, $\mathbf{s}_n = \mathbf{S}_{0:n;n+1} = \bar{q}_{n+1}\mathbf{r}_n$, $\mathbf{H}_n = \mathbf{H}_{0:n;0:n}$, $\mathbf{h}_n = \mathbf{H}_{0:n;n-1}$. Then we have

$$\mathbf{S}_n\mathbf{H}_n = \mathbf{I}_n - \mathbf{s}_n[0,\ldots,0,\eta_{n+1,n}].$$

If we write down the next to last column in this relation, we get

$$\mathbf{S}_n\mathbf{h}_n = [0,\ldots,0,1,0]^T, \tag{3.8}$$

from which $\mathbf{h}_n$ can be solved in $O(n)$ operations [8,25,15,14]. To compute the ORFs recursively, we start from $\phi_0 = 1/\|\mathbf{w}\|$ and use

$$\phi_{n-1}(z) = \Phi_n(z)(z\mathbf{I}_n - \mathbf{A}_n)\mathbf{h}_n,$$

where $\Phi_n = [\phi_0,\ldots,\phi_n]$ and $\mathbf{A}_n = \mathrm{diag}(\alpha_0,\ldots,\alpha_n)$, from which we can compute $\phi_n(z)$ if $\Phi_{n-1}(z)$ is known.

4 Numerical example

We consider a set of measurements with $z_j = \exp(i\omega_j)$, $j = 1,\ldots,100$ with $\omega_j \in [0.0038, 0.7632]$. The amplitude and phase of the measurements are given in Figure 1. The variance of the measurements is also shown. These data are symmetrically extended. For the poles we take the values $\alpha_0 = 0$ and $\alpha_{2k-1} = 0.9 * z_k$ and $\alpha_{2k} = \overline{\alpha}_{2k-1}$, $k = 1,2,\ldots,n$ where $n = 2,3,\ldots,30$.

The $(N+1)\times(n+1)$ matrix $\mathbf{\Phi}_n$, containing the first $n+1$ ORF, evaluated at the $N+1$ datapoints should be orthogonal with respect to the weight $w_i = 1/\sigma_i$ where σ_i is the given variance. Thus $\mathbf{Q}_n^H\mathbf{Q}_n - \mathbf{I}_{n+1}$ with $\mathbf{Q}_n = \mathbf{W}\mathbf{\Phi}_n$ should be zero. By rounding errors orthogonality may be lost. Therefore we compute the norm of this difference as a measure for the loss of orthogonality. This has been plotted in the last figure. In the case that the ORF are computed by the recurrence relation (3.7) it is

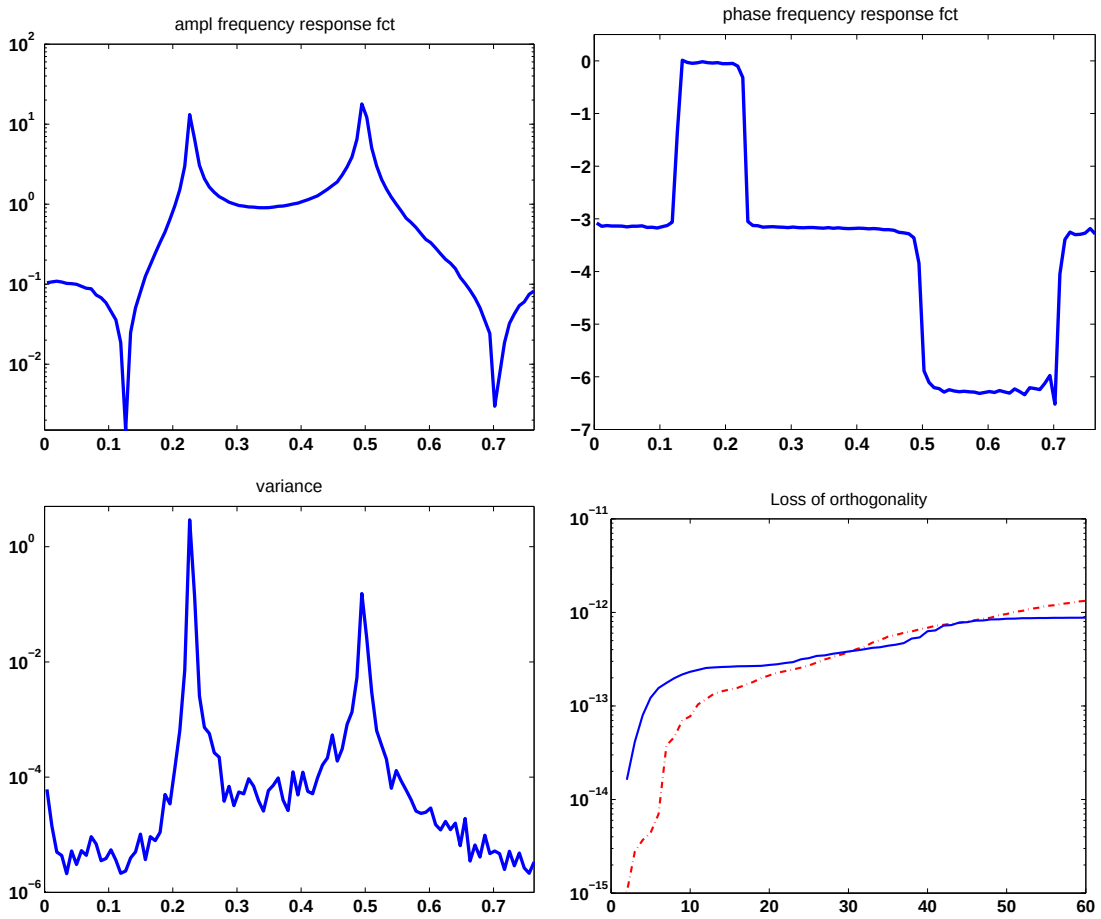

Fig. 1. *The loss of orthogonality: The amplitude and phase of the measured frequency response function and the corresponding variance of the measurements. As measure for the loss of orthogonality we compute* $\mathbf{\Phi}_n^H \mathbf{W}^H \mathbf{W} \mathbf{\Phi}_n - \mathbf{I}_{n+1}$ *for different values of* n*. The matrix* $\mathbf{\Phi}$ *has* N *rows and* $n+1$ *columns. This is plotted in the last figure for different values of* n*. The dashed line is for the ORF computed by the direct recurrence relation and the full line when they are computed by the technique of the semiseparable matrices.*

shown in dashed line, while for the case where they are computed by the technique of the semiseparable matrices, it is shown in full line.

Obviously both of the methods are very stable, and approximants up to degree 60 will not run into numerical problems. In fact, for this example, when only the first 6 poles are used, the approximation is already quite good for both methods.

For the Takenaka-Malmquist basis, the error is of the order 10^{14}, which is quite understandable because this basis was not designed to be orthogonal with respect to the weight that is used here. The condition number of $\mathbf{Q}_{60}$, that is the matrix $\mathbf{W}\mathbf{\Phi}_{60}$ in which $\mathbf{\Phi}_{60}$ contains the first 60 Takenaka-Malmquist basis functions evaluated in the $N+1$ data points, is approximately 10^6, while for the other bases considered here, it is $1 + O(10^{-12})$.

The numerical accuracy (in casu the loss of orthogonality, and the contition number

of $\mathbf{Q}_n$) depends strongly on the position of the poles. Choosing them closer to the unit circle will give a faster loss of orthogonality and multiple poles (e.g. cyclically repeating a finite number of poles) will make loss of orthogonality even more pronounced.

5 Appendix:

5.1 Proof of Corollary 3.7

We prove the following.

Corollary 5.1 *If all $z_k \in \mathbb{R}$, then the strictly upper triangular part of the matrix $\mathbf{S}$ has rank 1, i.e.,* $\mathrm{triu}(\mathbf{S},1) = \mathrm{triu}(\mathbf{R},1)$ *with $\mathbf{R}$ of rank 1.*
If all $z_k \in \mathbb{T}$, and if all $\alpha_i \neq 0$, $i = 1,\ldots,N$, then $\mathrm{triu}(\mathbf{S},1) = \mathrm{triu}(\mathbf{R},1)$ *with $\mathbf{R}$ of rank 1. If however m of the α_k, $k = 1,\ldots,n$ are zero, then* $\mathrm{triu}(\mathbf{S},1) = \mathrm{triu}(\mathbf{R},1)$ *with* $\mathbf{R} = \mathrm{diag}(\mathbf{R}_0,\ldots,\mathbf{R}_m)$ *with all $\mathbf{R}_i$ of rank 1 for $i = 0,\ldots,m$.*

PROOF. First consider the case $z_k \in \mathbb{R}$. Then, because $\mathbf{S}+\mathbf{A}$ is Hermitian, we have $\mathrm{triu}(\mathbf{S},1) = \mathrm{tril}(\mathbf{S},-1)^T = \mathrm{tril}(\mathbf{u}\mathbf{v}^H,-1)^H = \mathrm{triu}(\mathbf{v}\mathbf{u}^H,1)$.

For the case $z_k \in \mathbb{T}$, we use $\mathbf{S} = \mathbf{H}^{-1}$ and note that

$$(\mathbf{S}+\mathbf{A})^H = (\mathbf{S}+\mathbf{A})^{-1} = \mathbf{H}(\mathbf{I}+\mathbf{A}\mathbf{H})^{-1}.$$

Because $\mathbf{I}+\mathbf{A}\mathbf{H}$ is an upper Hessenberg matrix with nonzero subdiagonal elements, except for those subdiagonal elements corresponding to an $\alpha_i = 0$, its inverse $\mathbf{T}$ is weakly semiseparable, that is, it is of the form $\mathbf{R}+\mathbf{U}$ with $\mathbf{U}$ strictly upper triangular and $\mathbf{R} = \mathrm{diag}(\mathbf{R}_0,\ldots,\mathbf{R}_m)$ with all $\mathbf{R}_i$ of rank 1 (see Theorem 5.2 below). Because $\mathrm{triu}(\mathbf{S},1) = \mathrm{tril}((\mathbf{S}+\mathbf{A})^H,-1)^H$, we have to find the structure of $\mathrm{tril}(\mathbf{H}\mathbf{T},-1)$.

For the sake of simplicity, assume there is only one $\alpha_i = 0$, $i > 0$ (the generalisation should be obvious). So

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_0 & * \\ 0 & \mathbf{T}_1 \end{bmatrix}, \quad \mathbf{T}_k \text{ semiseparable}, \quad k = 0,1.$$

Define $\tilde{\mathbf{H}}$ as the upper triangular matrix which is obtained from $\mathbf{H}$ by deleting the first row and adding the row $[0,\ldots,0,1]$ at the bottom. We subdivide the matrix $\tilde{\mathbf{H}}$ in blocks having the same size as the blocks of $\mathbf{T}$. So

$$\begin{bmatrix} \tilde{\mathbf{H}}_0 & * \\ 0 & \tilde{\mathbf{H}}_1 \end{bmatrix} \begin{bmatrix} \mathbf{T}_0 & * \\ 0 & \mathbf{T}_1 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{H}}_0\mathbf{T}_0 & * \\ 0 & \tilde{\mathbf{H}}_1\mathbf{T}_1 \end{bmatrix} \equiv \mathbf{T}'.$$

Since the $\tilde{\mathbf{H}}_k$ is upper triangular with nonzero diagonal elements, and $\mathbf{T}_k = \mathbf{R}'_k + \mathbf{U}'_k$ with $\mathrm{rank}(\mathbf{R}'_k) = 1$ and $\mathrm{tril}(\mathbf{U}'_k) = 0$, it follows that $\tilde{\mathbf{H}}_k\mathbf{T}_k = \tilde{\mathbf{H}}_k\mathbf{R}'_k + \tilde{\mathbf{H}}_k\mathbf{U}'_k \equiv \mathbf{R}_k + \mathbf{U}_k$ with $\mathrm{rank}(\mathbf{R}_k) = 1$ and $\mathrm{tril}(\mathbf{U}_k) = 0$. It remains to note that the nontrivial part of $\mathrm{tril}((\mathbf{S}+\mathbf{A})^H, -1)$, that is the part without the zero main diagonal, is equal to $\mathrm{tril}(\mathbf{V})$ where $\mathbf{V}$ is the matrix obtained by deleting the last row and the last column of $\mathbf{T}'$. Indeed, $\mathbf{T}' = \tilde{\mathbf{H}}\mathbf{T}$ is the same as $\mathbf{HT} = (\mathbf{S}+\mathbf{A})^H$, except that the first row has been deleted and an extra row is added at the bottom. □

5.2 The inverse of a block upper Hessenberg matrix

We give a proof of the following general theorem.

Theorem 5.2 *Suppose* $\mathbf{H}$ *is an invertible block upper Hessenberg matrix with block size* k *and suppose that the subdiagonal blocks are invertible and upper triangular. If* $\mathbf{H}^{-1} = \mathbf{S}$*, then the lower triangular part of* $\mathbf{S}$ *is the lower triangular part of a rank* k *matrix. In other words if* $\mathbf{H}$ *has size* nk*, then there exist matrices* $\mathbf{u}$ *and* $\mathbf{v}$ *of size* $nk \times k$ *such that* $\mathrm{tril}(\mathbf{S}) = \mathrm{tril}(\mathbf{uv}^H)$*.*

If m *of the subdiagonal blocks in* $\mathbf{H}$ *are zero, then* $\mathrm{tril}(\mathbf{S}) = \mathrm{tril}(\mathbf{R})$ *with* $\mathbf{R} = \mathrm{diag}(\mathbf{R}_0, \ldots, \mathbf{R}_m)$*, where each* $\mathbf{R}_i$ *has rank* k*.*

PROOF. Suppose

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_3 & \mathbf{H}_4 \end{bmatrix}$$

with $\mathbf{H}_2$ of size $k \times k$. Note that $\mathbf{H}_3$ is upper triangular. Define the upper triangular matrix

$$\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{I}_k & \mathbf{H}_1 & \mathbf{H}_2 \\ 0 & \mathbf{H}_3 & \mathbf{H}_4 \\ 0 & 0 & \mathbf{I}_k \end{bmatrix}.$$

The matrices $\mathbf{I}_k$ are unit matrices of size k. Define

$$\mathbf{K} = \tilde{\mathbf{H}}^{-1} = \begin{bmatrix} \mathbf{I}_k & \mathbf{K}_1 & \mathbf{K}_2 \\ 0 & \mathbf{K}_3 & \mathbf{K}_4 \\ 0 & 0 & \mathbf{I}_k \end{bmatrix}.$$

Note that also $\mathbf{K}_3$ is upper triangular. Since $\tilde{\mathbf{H}}K = \mathbf{I}_{(n+1)k}$, we also have

$$\begin{bmatrix} \mathbf{I}_k & \mathbf{H}_1 & \mathbf{H}_2 \\ 0 & \mathbf{H}_3 & \mathbf{H}_4 \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & \mathbf{K}_1 \\ 0 & \mathbf{K}_3 \\ 0 & 0 \end{bmatrix} = \mathbf{I}_{nk}$$

or

$$\begin{bmatrix} \mathbf{I}_k \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & \mathbf{K}_1 \end{bmatrix} + \mathbf{H} \begin{bmatrix} 0 & \mathbf{K}_3 \\ 0 & 0 \end{bmatrix} = \mathbf{I}_{nk}$$

so that

$$\mathbf{H} \left(\mathbf{H}^{-1} \begin{bmatrix} \mathbf{I}_k \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & \mathbf{K}_1 \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{K}_3 \\ 0 & 0 \end{bmatrix} \right) = \mathbf{I}_{nk}. \tag{5.1}$$

Now define

$$\mathbf{S} = \mathbf{H}^{-1} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 \\ \mathbf{S}_3 & \mathbf{S}_4 \end{bmatrix},$$

with $\mathbf{S}_1$ of size $k \times k$, then from (5.1)

$$\mathbf{S} = \mathbf{S} \begin{bmatrix} \mathbf{I}_k \\ 0 \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & \mathbf{K}_1 \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{K}_3 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_3 \end{bmatrix} \begin{bmatrix} \mathbf{I}_k & \mathbf{K}_1 \end{bmatrix} + \begin{bmatrix} 0 & \mathbf{K}_3 \\ 0 & 0 \end{bmatrix} \equiv \mathbf{M}_1 + \mathbf{M}_2.$$

Obviously, the matrix $\mathbf{M}_1$ has rank k because $[\mathbf{S}_1^H \ \mathbf{S}_3^H]^H$ has full rank k, and since $\mathbf{K}_3$ is upper triangular, it follows that $\mathrm{tril}(\mathbf{S}) = \mathrm{tril}(\mathbf{M}_1)$, and this proves the theorem when all the subdiagonal blocks are invertible.

Suppose now that there is only one subdiagonal block equal to zero. Then the $\mathbf{H}$ is reducible, and we may write (note that this subdivision of $\mathbf{H}$ is different from the one we had before)

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_0 & * \\ 0 & \mathbf{H}_1 \end{bmatrix}$$

where $\mathbf{H}_0$ and $\mathbf{H}_1$ are block upper Hessenberg matrices of the form treated in the first part of this theorem. Thus we can apply the first part of this theorem to each of the blocks $\mathbf{H}_0$ and $\mathbf{H}_1$:

$$\mathbf{S} = \mathbf{H}^{-1} = \begin{bmatrix} \mathbf{S}_0 & * \\ 0 & \mathbf{S}_1 \end{bmatrix}$$

with for $i = 0, 1$, $\mathbf{S}_i = \mathbf{H}_i^{-1} = \mathbf{R}_i + \mathbf{U}_i$ where $\mathbf{R}_i$ is a rank k matrix and $\mathbf{U}_i$ is strictly upper triangular. This proves the theorem if there is only one zero subdiagonal block. The generalization to the case where there are more subdiagonal blocks equal to zero is obvious. □

5.3 An algorithm for the computation of $\mathbf{S} + \mathbf{A}$

As we have said, we will only look at the case where the strictly upper triangular part of the semiseparable matrix $\mathbf{S}$ is part of a rank 1 matrix. This depended on the fact that the poles were nonzero.

Recall that the rank 1 condition for the strictly upper triangular part depends on the fact that there are no poles equal to zero. If a pole is zero, this upper triangular part will have a zero block and two other blocks of rank 1. Thus the matrix is represented with the same amount of information, so that the algorithm has to be adapted, but it will still have the same complexity.

If the poles are not zero, but become small, the rank 1 condition will be almost violated, which will lead to numerical rounding errors. The numerically stable variant does exist, but it is however much more involved in general, so we will describe it elsewhere.

For a matrix $\mathbf{S} = \mathrm{tril}(\mathbf{u}\mathbf{v}^H) + \mathrm{triu}(\mathbf{r}\mathbf{q}^H, 1)$, the updating goes as follows. Suppose we have already the matrix $\mathbf{S} + \mathbf{A}$ of size 3 and we add new data (w_3, α_3, z_3):

$$\left[\begin{array}{c|c||ccc} w_3 & z_3 & & & \\ \hline \|\mathbf{w}_2\| & & u_0\bar{v}_0 + \alpha_0 & r_0\bar{q}_1 & r_0\bar{q}_2 \\ & & u_1\bar{v}_0 & u_1\bar{v}_1 + \alpha_1 & r_1\bar{q}_2 \\ & & u_2\bar{v}_0 & u_2\bar{v}_1 & u_1\bar{v}_2 + \alpha_2 \end{array}\right].$$

There are 4 vectors involved in the definition of the matrix $\mathbf{S}$: $\mathbf{u}, \mathbf{v}, \mathbf{q}, \mathbf{r}$. At this stage we have a vector $\mathbf{u} = [u_0, u_1, u_2]^T$, and similarly for the other vectors. We shall transform this into a new vector $\mathbf{u}' = [u'_0, u'_1, u'_2, u'_3]^T$ and similarly for the other vectors. This transformation will be done in several substeps. We shall have the successive transformations that build up $\mathbf{u}'$ from $\mathbf{u}$ from top to bottom:

$$\begin{bmatrix} \cdot \\ u_0 \\ u_1 \\ u_2 \end{bmatrix} \rightarrow \begin{bmatrix} u'_0 \\ u''_1 \\ u_1 \\ u_2 \end{bmatrix} \rightarrow \begin{bmatrix} u'_0 \\ u'_1 \\ u''_2 \\ u_2 \end{bmatrix} \rightarrow \begin{bmatrix} u'_0 \\ u'_1 \\ u'_2 \\ u''_3 \end{bmatrix} \rightarrow \begin{bmatrix} u'_0 \\ u'_2 \\ u'_2 \\ u'_3 \end{bmatrix}$$

and similarly for the other vectors. The first step is to define a unitary matrix $G_{1,2} = \begin{bmatrix} c & s \\ -\bar{s} & c \end{bmatrix}$ such that

$$G_{1,2}^H \begin{bmatrix} w_3 \\ \|\mathbf{w}_2\| \end{bmatrix} = \begin{bmatrix} \|\mathbf{w}_3\| \\ 0 \end{bmatrix}.$$

Define

$$\begin{bmatrix} a & b \\ e & d \end{bmatrix} = G_{1,2}^H \begin{bmatrix} z_3 & 0 \\ 0 & u_0\bar{v}_0 + \alpha_0 \end{bmatrix} G_{1,2}.$$

This is the principal 2×2 matrix in the following matrix which is the result of the similarity transformation

$$\left[\begin{array}{cc|cc} u_0'\bar{v}_0' + \alpha_0 & r_0'\bar{q}_1'' & r_0'\bar{q}_1 & r_0'\bar{q}_2 \\ u_1''\bar{v}_0' & u_1''\bar{v}_1'' + \alpha_1' & r_1''\bar{q}_1 & r_1''\bar{q}_2 \\ \hline u_1\bar{v}_0' & u_1\bar{v}_1'' & u_1\bar{v}_1 + \alpha_1 & r_1\bar{q}_2 \\ u_2\bar{v}_0' & u_2\bar{v}_1'' & u_2\bar{v}_1 & u_2\bar{v}_2 + \alpha_2 \end{array}\right]$$

where

$$\begin{aligned} &v_0' = -\bar{s}v_0, \quad v_1'' = \bar{c}v_0, \quad r_0' = -\bar{s}r_0, \quad r_1'' = \bar{c}r_0 \\ &u_0' = (a - \alpha_0)/\bar{v}_0', \quad u_1'' = e/\bar{v}_0', \quad \alpha_1' = d - u_1''\bar{v}_1'', \quad \bar{q}_1'' = b/r_0'. \end{aligned}$$

The next step is to do a unitary similarity transform on rows and columns 2 and 3 by a unitary matrix

$$G_{2,3} = \frac{1}{\sqrt{1 + |t|^2}} \begin{bmatrix} 1 & \bar{t} \\ -t & 1 \end{bmatrix}, \quad t = \frac{\alpha_1 - \alpha_1'}{\bar{v}_1 u_1'' - \bar{q}_1 r_1''},$$

which results in

$$\left[\begin{array}{c|ccc} u_0'\bar{v}_0' + \alpha_0 & r_0'\bar{q}_1' & r_0\bar{q}_2'' & r_0'\bar{q}_2 \\ \hline u_1'\bar{v}_0' & u_1'\bar{v}_1' + \alpha_1 & r_1'\bar{q}_2'' & r_1'\bar{q}_2 \\ u_2''\bar{v}_0' & u_2''\bar{v}_1' & u_2''\bar{v}_2'' + \alpha_2' & r_2''\bar{q}_2 \\ \hline u_2\bar{v}_0' & u_2\bar{v}_1' & u_2\bar{v}_2'' & u_2\bar{v}_2 + \alpha_2 \end{array}\right] \tag{5.2}$$

where

$$G_{2,3}^H \left[\begin{array}{c|c} u_1'' & r_1'' \\ \\ u_1 & r_1 \end{array}\right] = \left[\begin{array}{c|c} u_1' & r_1' \\ \\ u_2'' & r_2'' \end{array}\right], \quad \left[\begin{array}{c|c} \bar{v}_1'' & \bar{v}_1 \\ \hline \bar{q}_1'' & \bar{q}_1 \end{array}\right] G_{2,3} = \left[\begin{array}{c|c} \bar{v}_1' & \bar{v}_2'' \\ \hline \bar{q}_1' & \bar{q}_2'' \end{array}\right], \quad \alpha_2' = \alpha_1'.$$

This type of transformation is repeated for the next pair of rows and columns, transforming the pairs $(2,3),(3,4),\ldots,(n-1,n)$. The last step is in general to rename $u_n''\bar{v}_n'' + \alpha_n'$ as $u_n'\bar{v}_n' + \alpha_n$ which is obtained by setting $u_n' = u_n''$ and $v_n' = (\alpha_n' - \alpha_n + u_n''\bar{v}_n'')/u_n'$.

References

[1] H. Akçay and B. Ninness. Rational basis functions for robust identification from frequency and time domain measurements. *Automatica*, 34(9):1101–1117, 1998.

[2] L. Baratchart. Rational and meromorphic approximation in L^p of the circle: System-theoretic motivations, critical points and error rates. In N. Papamichael, editor, *Proceedings of the 3rd CMFT conference on computational methods and function theory 1997, Nicosia, Cyprus, October 13-17, 1997*, volume 11 of *Approx. Decompos.*, pages 45–78. World Scientific, 1999.

[3] L. Baratchart, J. Leblond, J.R. Partington, and N. Torkhani. Robust identification in the disc algebra from band-limited data. *IEEE Trans. Automat. Control*, 42(9):318–1325, 1997.

[4] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. *Orthogonal rational functions*, volume 5 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, 1999.

[5] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonal rational functions and tridiagonal matrices. *J. Comput. Appl. Math.*, 2001. Submitted.

[6] A. Bultheel and M. Van Barel. Vector orthogonal polynomials and least squares approximation. *SIAM J. Matrix Anal. Appl.*, 16(3):863–885, 1995.

[7] A. Bultheel, M. Van Barel, Y. Rolain, and J. Schoukens. Numerically robust rational approximation for identification. In IEEE, editor, *40th IEEE Conference on Decision and Control, Orlando FL, December 4-7, 2001*, pages 4770–4775. IEEE, 2001. Paper PrP06-4.

[8] S. Chandrasekaran and M. Gu. Fast and stable algorithms for banded plus semiseparable systems of linear equations. 2001. Submitted.

[9] Sanathanan C.K. and J. Koerner. Transfer function synthesis as a ratio of two complex polynomials. *IEEE Trans. Automat. Control*, 9(1):56–58, 1963.

[10] P.R. Clement. Laguerre functions in signal analysis and parameter identification. *J. Franklin Inst.*, 313, no. 2:85–95, 1982.

[11] T.J. de Hoog. *Rational orthonormal bases and related transforms in linear system modeling*. PhD thesis, T.U. Delft, 2001.

[12] T.J. de Hoog, P.S.C. Hueberger, and P.M.J. Van den Hof. A general transform theory of rational orthonormal basis function expansion. In *Proceedings of the 39th IEEE conference on Decision and Control*, pages 4649–4654, Sidney, Australia, 2000.

[13] D.K. de Vries and P. van den Hof. Frequency domain identification with generalized orthonormal basis functions. *IEEE Trans. Automat. Control*, AC-43(5):656–669, 1998.

[14] Y. Eidelman and I. Gohberg. Inversion formulas and linear complexity. *Comput. Math. Appl.*, 33:69–79, 1997.

[15] D. Fasino, N. Mastronardi, and M. Van Barel. Fast and stable algorithms for reducing diagonal plus semiseparable matrices to tridiagonal and bidiagonal form. Technical Report IRMA 3/2002, I.R.M.A. - C.N.R., Bari (It.), 2002.

[16] G.E. Forsythe. Generation and use of orthogonal polynomials for data fitting with a digital computer. *J. SIAM*, 5:74–88, 1957.

[17] F.R. Gantmacher and M.G. Kreĭn. *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme*. Akademie-Verlag, Berlin, 1960.

[18] P.S.C. Heuberger and P.M.J. Van den Hof. The Hambo transform: a signal and system transform induced by generalized orthonormal basis functions. *Selected Topics in Identification, Modelling and Control*, 8:85–94, 1995.

[19] P.S.C. Heuberger, P.M.J. Van den Hof, and O. Bosgra. A generalized orthogonal basis for linear dynamical systems. *IEEE Trans. Automat. Control*, 40:451–465, 1995.

[20] P.S.C. Hueberger. *On approximate system identification with system based orthonormal functions*. Phd, TU Delft, The Netherlands, 1991.

[21] Y.W. Lee. Synthesis of electrical networks by means of the Fourier transforms of Laguerre functions. *J. Math. Phys.*, 11:83–113, 1933.

[22] Y.W. Lee. *Statistical Theory of Communication*. Wiley, New York, 1960.

[23] L. Ljung. *System identification: Theory for the user*. Information and System Sciences Series. Prentice Hall, 2nd edition, 1999.

[24] F. Malmquist. Sur la détermination d'une classe de fonctions analytiques par leurs valeurs dans un ensemble donné de poits. In *C.R. 6ième Cong. Math. Scand. (Kopenhagen, 1925)*, pages 253–259, Copenhagen, 1926. Gjellerups.

[25] N. Mastronardi, S. Chandrasekaran, and S. Van Huffel. Fast and stable two-way algorithm for diagonal plus semi-separable systems of linear equations. *Numer. Linear Algebra Appl.*, 8(1):1–12, 2001.

[26] B. Ninness. Frequency domain estimation using orthonormal bases. In *Proceedings of the 13th IFAC World Congress, San Francisco 1996, Volume 8*, pages 381–386, 1996.

[27] B. Ninness and F. Gustafsson. A unified construction of orthogonal bases for system identification. *IEEE Trans. Automat. Control*, 42:515–522, 1997.

[28] B. Ninness, H. Hjalmarsson, and F. Gustafsson. The fundamental role of general orthogonal bases in system identification. *IEEE Trans. Automat. Control*, 44(7):1384–1407, 1999.

[29] Y. Nurges and Y. Yaaksoo. Laguerre state equations for a multivariable discrete system. *Automat. Remote Contr.*, 42:1601–1606, 1981.

[30] R. Pintelon, Y. Rolain, A. Bultheel, and M. Van Barel. Numerically robust frequency domain identification of multivariable systems. In P. Sas and B. Van Hal, editors, *Proceedings of the 2002 International Conference on Modal Analysis, Noise and Vibration Engineering, Leuven 2002, September 16-18*, pages 1316–1321, 2002.

[31] R. Pintelon and J. Schoukens. *System Identification: A Frequency Domain Approach.* IEEE Press, Piscataway, 2001.

[32] Z. Szabó, J. Bokor, and F. Schipp. Identification of rational approximate models in H_∞ using generalized orthogonal basis. *IEEE Trans. Automat. Control*, 44(1):153–158, 1999.

[33] S. Takenaka. On the orthogonal functions and a new formula of interpolation. *Japanese J. Math.*, 2:129–145, 1925.

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

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

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

[37] M. Van Barel, D. Fasino, L. Gemignani, and N. Mastronardi. Orthogonal rational functions and diagonal-plus-semiseparable matrices. In *Proceedings of the 47th Annual SPIE Meeting, Seattle, WA, USA*. SPIE, 2002. To appear.

[38] P.M.J. Van den Hof, P.S.C. Heuberger, and J. Bokor. System identification with generalized orthogonal basis functions. *Automatica*, 31:1821–1834, 1995.

[39] P. Van gucht and A. Bultheel. Computing orthogonal rational functions analytic outside the unit disc. Technical Report TW312, Department of Computer Science, K.U. Leuven, September 2000.

[40] P. Van gucht and A. Bultheel. State space representation for arbitrary orthogonal rational functions. *Systems Control Lett.*, 2001. Submitted.

[41] B. Wahlberg. System identification using Laguerre models. *IEEE Trans. Automat. Control*, AC-36(5):551–562, 1991.

[42] B. Wahlberg. System identification using Kautz models. *IEEE Trans. Automat. Control*, AC-39(6):1276–1282, 1994.