Orthogonal rational functions on the unit circle: from the scalar to the matrix case

A. Bultheel[*], P. González-Vera[†], E. Hendriksen[‡], O. Njåstad[§]

Abstract

The purpose of these lecture notes is to give a short introduction to the theory of orthogonal rational functions (ORF) on the unit circle. We start with the classical problem of linear prediction of a stochastic process to give a motivation for the study of Szegő's problem and to show that in this context it will turn out that not as much the ORF but rather the reproducing kernels will play a central role. Another example of rational Krylov iteration shows that it might also be interesting to consider ORF on the real line, which we shall not discuss in these lectures.

In a second part we will show that most of the results of the scalar case thanslate easily to the case of matrix valued orthogonal rational functions (MORF).

There are however many aspects that are intimately related to these ideas that we do not touch upon like continued fractions, Nevanlinna-Pick interpolation, moment problems, and many other aspects of what is generally known as Schur analysis.

Contents

[*]Department of Computer Science, K.U.Leuven, Belgium. The work of this author is 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, the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with the author.

[†]Department Análisis Math., Univ. La Laguna, Tenerife, Spain. The work of this author was partially supported by the scientific research project PB96-1029 of the Spanish D.G.E.S.

[‡]Department of Mathematics, University of Amsterdam, The Netherlands.

[§]Department of Math. Sc., Norwegian Univ. of Science and Technology, Trondheim, Norway

Part 0: Motivation

1 Why orthogonal rational functions?

We want to give a summary of known results about orthogonal rational functions (ORF) with respect to some measure whose support is contained in the unit circle of the complex plane. But before we start, we want to give some motivation of why it may be interesting to generalize orthogonal polynomials to ORF for which the poles of the successive rational functions of increasing degree are taken from a preselected sequence of complex numbers.

1.1 Linear prediction

Consider a discrete time stationary stochastic process $\{y_n\}_{n\in\mathbb{Z}}$. The index n denotes a time instance. Stationarity means that $\mu_k = \mathsf{E}\{y_n\overline{y}_{n-k}\}$ is independent of n. We suppose that all the y_n are zero mean, identically distributed. A problem considered already by Wiener is to predict y_n at instant n from observation of the preceding y_{n-k}, $k \geq 1$. Thus we want to find coefficients a_k such that y_n is predicted by $\hat{y}_n = -\sum_{k=1}^{\infty} a_k y_{n-k}$ such that we minimize the energy of the prediction error or innovation process $e_n = \sum_{k=0}^{\infty} a_k y_{n-k}$ where $a_0 = 1$. That is we want to minimize the expected value $\mathsf{E}\{|e_n|^2\}$

Because in the space spanned by $\{y_n\}_{n\in\mathbb{Z}}$, we can consider $\mathsf{E}\{x\overline{y}\}$ as an inner product of x and y, we can formulate the previous problem by saying that we look for the orthogonal projection of the present onto its past. Modulo some technical details, it can be shown that this problem can be reformulated in an L_2 setting as an infinite dimensional least squares problem, i.e., finding the projection of 1 onto the space H_2^- spanned by $\{z^{-1}, z^{-2}, \ldots\}$ in the Hilbert space $L_2(\mu)$ of the unit circle where the orthogonality measure is the spectral measure of the process (the trigonometric moments of this measure are the μ_k introduced above). The result is that the optimal predictor is given by the spectral factor of the spectral measure. Suppose for simplicity that the spectral measure is absolutely continuous with weight $w(e^{i\omega})$, then the spectral factor is the function $\sigma(z)$ which is the outer spectral factor of w (σ and $1/\sigma$ are analytic outside the unit disk, and satisfy $|\sigma(t)|^2 = w(t)$ for $|t| = 1$).

There are two ways one can think of to solve this problem. Either we project 1 onto subspaces $\mathcal{L}_n^-$ of H_2^- of dimension $n = 1, 2, \ldots$ and let n tend to ∞, or we can solve the trigonometric moment problem to find the spectral measure and subsequently or simultaneously do a spectral factorization.

If we assume that the value of y_n is mostly defined by the near past and much less by a past that is longer ago, then a natural choice is to take $\mathcal{L}_n^- = \operatorname{span}\{z^{-1}, \ldots, z^{-n}\}$. So the optimal predictor of this finite dimensional problem is to find the polynomial $\hat{\varphi}_n(z) = \sum_{k=0}^{n} a_k z^{-k}$ with $a_0 = 1$ that has minimal norm in $L_2(\mu)$. This is known to be the reciprocal of the monic nth orthogonal polynomial. From the classical Szegő theory, it is known that (under appropriate conditions on the measure) these orthogonal polynomials converge (up to a constant) to the outer spectral factor of the orthogonality measure (whose inverse is often called Christoffel function in this context). Thus solving the projection problem leads asymptotically to a spectral factor of the measure. For practical applications though we do not want n to be very large, since that would need long and expensive filters. So we stop this process with a finite n, and hope to have a good approximation. It is not difficult however to find a simple example that requires a very long filter to get a reasonable approximation.

Suppose $y_n = -\sum_{k=1}^{\infty} a^k y_{n-k}$. Thus the predictor is $1 + a/z + a^2/z^2 + \cdots = 1/(1 - a/z)$. If $|a|$ is close to 1, then the sequence a^k will decay very slowly, and we shall need a high degree polynomial $\hat{\varphi}_n$ to obtain a good approximation. If however, we know from observing the FFT of the observations, an estimate $\hat{a}$ of a, then a filter expanded in terms of powers of $(1 - \hat{a}/z)$ might need only a couple of terms to give a good approximation. For a more complicated behaviour of the spectral density of the process, one might need more poles to get an accurate model and in such a case a rational approximant from $\text{span}\{1, 1/(1 - \alpha_1/z), 1/[(1 - \alpha_1/z)(1 - \alpha_2/z)], \ldots\}$ will be a better option to model the system.

1.2 Krylov subspace methods

If A is a linear operator on a Hilbert space H (e.g., a large $N \times N$ matrix operating on $\mathbb{R}^N$) and $v \in H$, then the space $\mathcal{K}_{n+1}(A, v) = \text{span}\{v_0, \ldots, v_n\}$ with $v_k = A^k v$, is called a Krylov subspace. To solve a linear equation $Ax = b$ or an eigenvalue problem for A, the problem is projected onto a Krylov subspace of finite (i.e., low) dimension ($n \ll N$ in the matrix example) and this low dimensional problem is solved to give an approximation to the original problem. To compute the projection, an orthogonal basis is constructed for the Krylov subspace. Clearly, the $(k+1)$st orthogonal vector q_k has to be a combination of the the first $k+1$ vectors in the Krylov subspace. Hence it is of the form $q_k = \varphi_k(A)v$ with $\varphi_k(z)$ a polynomial of degree k.

Exercise 1.1 Suppose that A is real self adjoint and positive definite, then prove that the orthonormality $q_k^T q_l = \delta_{k,l}$ is equivalent with the orthogonality of the polynomials $\langle \varphi_k, \varphi_k \rangle = \delta_{kl}$ with respect to the inner product defined by $\langle \varphi_k, \varphi_l \rangle = L(\varphi_k \varphi_l)$ where the linear functional L is defined on the space of polynomials by its moments $m_k = L(z^k) = v^T A^k v$. Note that the metric for the standard basis $\{1, z, z^2, \ldots\}$ is a Hankel matrix whose rank can not be larger than N, the size of A. ◇

Thus in the classical Lanczos method for symmetric matrices, the three-term recurrence relation for the orthogonal polynomials leads to a short recurrence between the successive vectors q_k, meaning that q_n can be computed from q_{n-1} and q_{n-2}, which does not need a full Gram-Schmidt orthogonalization.

However, computing the v_k is like an application of the power method and therefore, the v_k will quickly converge to an eigenvector corresponding to a dominant eigenvalue. Thus, if we want an eigenvalue in the neighborhood of α, then we should not iterate with A, but with $B = (A - \alpha I)^{-1}$. The rational Krylov method (RKS) of A. Ruhe [26, 27, 28] allows for a different shift α in every iteration step. Thus $v_k = (A - \alpha_k I)^{-1} v_{k-1}$, or even more generally $v_k = (A - \sigma_k I)(A - \alpha_k I)^{-1} v_{k-1}$, where α_k is used to enforce the influence of the eigenspaces of the eigenvalues in the neighborhood of α_k, while σ_k is used to suppress the influence of the eigenspaces of the eigenvalues in the neighborhood of σ_k. Anyway, this construction of v_k means that we may write v_k as $v_k = r_k(A)v$ with r_k a rational function of the form $p_k(z)/[(z - \alpha_1) \cdots (z - \alpha_k)]$ with p_k a polynomial of degree at most k, so that after orthogonalizing v_k with respect to the previous vectors, we obtain a vector $q_k = \varphi_k(A)v$ where $\varphi_k(z)$ is again a rational function of the same form as r_k. $p_k(z)/[(z - \alpha_1) \cdots (z - \alpha_k)]$ with p_k a polynomial of degree at most k. Since the classical moment matrix has a Hankel structure, this theory will be related to orthogonality on the real line.

Skipping all the technical details, it will be obvious that orthogonality of the q_k will lead to some orthogonality of the rational functions φ_k. Again, a simple recurrence of the ORF will lead to an efficient implementation of the RKS.

1.3 Numerical quadrature

One more example, before we dive into the general theory of ORF.

To compute an integral $\int_a^b f(x)w(x)dx$, it is well known that a quadrature formula of the form $\sum_{k=1}^n \lambda_k f(\xi_k)$ will be exact for all polynomials of degree up to $2n-1$ if the abscises of this quadrature formula are the zeros of the polynomial φ_n which is the nth orthogonal polynomial orthogonal with respect to the inner product $\langle f, g\rangle = \int_a^b f(x)g(x)w(x)dx$, and if the weights are given by the corresponding Christoffel numbers $\lambda_k = 1/\sum_{j=0}^{n-1} |\varphi_k(\xi_k)|^2$.

However, if the integrand f does not behave much like a polynomial, some other quadrature formula can be better. For example if we want to integrate $f(x) = \sin(x)/p_n(x)$ over the interval $[0, \pi/2]$ with $w = 1$. If $p_n(x)$ is a real polynomial with complex conjugate pairs of zeros that are close to the interval of integration, then it would be much better to have a quadrature formula that is exact for all rational functions of degree n that have poles close to the zeros of the polynomial p_n. Using ORF for the interval $[a, b]$ with respect to the weight w, with prescribed poles, it is possible to derive formulas that are similar to the Gaussian formulas, i.e., taking for the nodes the zeros of the ORF φ_n and as weights $\lambda_k = 1/\sum_{j=0}^{n-1} |\varphi_k(\xi_k)|^2$. These quadrature formulas will be exact in a certain space of dimension $2n-1$. It is clear that this should lead to much more accurate results in examples like we described above.

Part I: The scalar case

2 Orthogonal rational functions on the unit circle

We give in part I an introduction to the theory of ORF for a measure that is supported on the unit circle of the complex plane. All the results of this part (and more) can be found in the monograph [3].

2.1 Preliminaries

We denote

$$\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}, \quad \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}, \quad \mathbb{E} = \{z \in \mathbb{C} : |z| > 1\}.$$

For any function f we introduce the *parahermitian conjugate* f_*, defined by $f_*(z) = \overline{f(1/\overline{z})}$. Note that for $t \in \mathbb{T}$, $f_*(t) = \overline{f(t)}$.

Let μ be a probability measure on $\mathbb{T}$ with infinite support and $L_2(\mu)$ the Hilbert space with inner product $\langle f, g\rangle_\mu = \int f(t)\overline{g(t)}d\mu(t)$. If $d\mu = d\lambda$ is the normalized Lebesgue measure $d\lambda = dt/(2\pi)$ we drop μ from the notation. The trigonometric moments are denoted by $c_k = \int t^{-k} d\mu(t)$, $k \in \mathbb{Z}$.

Introducing the Riesz-Herglotz kernel

$$D(t, z) = \frac{t + z}{t - z}$$

we can associate with μ its Riesz-Herglotz transform

$$\Omega_\mu(z) = ic + \int D(t, z) d\mu(t)$$

with $\Omega_\mu(0) = 1 + ic$. This Ω_μ belongs to the class $\mathcal{C}$ of Carathéodory functions

$$\mathcal{C} = \{f \in H(\mathbb{D}) : \operatorname{Re} f(\mathbb{D}) > 0\}$$

where $H(\mathbb{D})$ denotes the functions analytic in $\mathbb{D}$.

Exercise 2.1 Prove that

$$\operatorname{Re} \Omega_\mu(z) = \int P(t, z) d\mu(t)$$

with Poisson kernel

$$P(t, z) = \frac{1}{2}[D(t, z) + \overline{D(1/\overline{t}, z)}] = \frac{t(1 - |z|^2)}{(t - z)(1 - \overline{z}t)}, \tag{2.1}$$

which can be simplified for $t \in \mathbb{T}$ as follows

$$P(t, z) = \operatorname{Re} D(t, z) = \frac{1 - |z|^2}{|t - z|^2}, \quad t \in \mathbb{T},\ z \in \mathbb{D}.$$

Moreover with $c_k = \int t^{-k} d\mu(t)$

$$\Omega_\mu(z) = c_0 + 2\sum_{k=1}^{\infty} c_k z^k, \quad (c_0 = 1)$$

and the nontangential limit of $\operatorname{Re} \Omega_\mu(z)$ when $|z| \to 1$ is a.e. $(d\lambda)$ equal to μ' which is the density function of the absolute continuous part of μ. ◇

Let $H_2(\mathbb{D})$ be the classical Hardy space of functions analytic in $\mathbb{D}$ that have a boundary value on $\mathbb{T}$ that is square integrable. Every function $f \in H_2(\mathbb{D})$ has a canonical inner-outer factorization. This means that it can be written an $f = UF$ with U inner and F outer. An inner function belongs to the class of bounded analytic functions

$$\mathcal{B}(\mathbb{D}) = \{f \in H(\mathbb{D}) : f(\mathbb{D}) \subset \mathbb{D}\}$$

and $|f(t)| = 1$ a.e. on $\mathbb{T}$. A Blaschke product is an example of an inner function. It is defined as $B(z) = \prod_n \zeta_n(z)$ with

$$\zeta_n(z) = z_n \frac{z - \alpha_n}{1 - \overline{\alpha}_n z}, \quad z_n = -\frac{|\alpha_n|}{\alpha_n}, \quad \alpha_n \in \mathbb{D} \setminus \{0\}$$

and $z_n = 1$ if $\alpha_n = 0$. It is well known that a Blaschke product converges (to a finite function not identically zero) if and only if

$$\sum_n (1 - |\alpha_n|) < \infty$$

which means that the $|\alpha_n|$ should approach 1 fast enough.

Any inner function is of the form

$$U(z) = e^{i\gamma} B(z) S(z), \quad \gamma \in \mathbb{R}, \quad S(z) = \exp\left\{-\int D(t,z) d\nu(t)\right\}$$

with ν a bounded positive singular ($\nu' = 0$ a.e.) measure and B a Blaschke product that catches all the zeros of U.

An outer function in $H_2(\mathbb{D})$ is a function of the form

$$F(z) = e^{i\gamma} \exp\left\{\int D(t,z) \log \psi(t) d\lambda(t)\right\}, \quad \gamma \in \mathbb{R}$$

where $\log \psi \in L_1$ and $\psi \in L_2$.

If $\log \mu' \in L_1$ (Szegő's condition), then we may define the spectral factor of μ

$$\sigma(z) = c \exp\left\{\frac{1}{2} \int D(t,z) \log \mu'(t) d\lambda(t)\right\}, \quad c \in \mathbb{T}, \quad z \in \mathbb{D}.$$

This constant can be fixed by requiring for example that $\sigma(0) > 0$. It is an outer function, uniquely defined up to a constant $c \in \mathbb{T}$. It has a nontangential limit to the boundary $\mathbb{T}$ that satisfies $|\sigma(t)|^2 = \mu'(t)$, a.e., $t \in \mathbb{T}$.

Exercise 2.2 Prove that

$$|\sigma(z)|^2 = \exp\left\{\int P(t,z) \log \mu'(t) d\lambda\right\}.$$

◇

More about the material in this section can be found in many textbooks a.o. [8, 16, 15, 14, 25, 30].

2.2 The fundamental spaces

We select a sequence $\{\alpha_k\}_{k=0}^{\infty} \subset \mathbb{D}$ with $\alpha_0 = 0$ and define the partial Blaschke products

$$B_0 = 1, \quad B_n(z) = B_{n-1}(z)\zeta_n(z) = \prod_{k=1}^{n} z_k \frac{z-\alpha_k}{1-\overline{\alpha}_k z}, \quad n \geq 1.$$

The functions $\{B_0, B_1, \ldots, B_n\}$ span the space

$$\mathcal{L}_n = \left\{ \frac{p_n}{\pi_n} : \pi_n(z) = \prod_{k=1}^{n}(1-\overline{\alpha}_k z), \quad p_n \in \mathcal{P}_n \right\} \tag{2.2}$$

where $\mathcal{P}_n$ is the space of polynomials of degree at most n.

If we set all $\alpha_k = 0$, then $\mathcal{L}_n = \mathcal{P}_n$. In that case the Gram matrix of $\mathcal{L}_n$ for the standard basis is a *Toeplitz matrix* with entries $\langle z^i, z^j \rangle_\mu = c_{j-i}$ where $i, j = 0, \ldots, n$. If all the α_k are mutually distinct, then the Gram matrix for the basis $\{1, 1/(1-\overline{\alpha}_1 z), \ldots, 1/(1-\overline{\alpha}_n z)\}$ has entries

$$\left\langle \frac{1}{1-\overline{\alpha}_i z}, \frac{1}{1-\overline{\alpha}_j z} \right\rangle_\mu = \frac{1}{2(1-\overline{\alpha}_i \alpha_j)} \int \left[\overline{D(t,\alpha_i)} + D(t,\alpha_j) \right] d\mu(t) = \frac{1}{2} \frac{\overline{\Omega_\mu(\alpha_i)} + \Omega_\mu(\alpha_j)}{1-\overline{\alpha}_j \alpha_i}.$$

Such a matrix is called a *Pick matrix*.

Exercise 2.3 If ∂_w^k denotes the kth derivative with respect to the variable w, prove that

$$\Omega_\mu^{(k)}(w) = \partial_w^k \Omega_\mu(w) = \int \partial_w^k D(t,w) d\mu(t) = 2(k!) \int \frac{t d\mu(t)}{(t-w)^{k+1}}$$

and

$$\overline{\Omega_\mu^{(k)}(w)} = \int \overline{\partial_w^k D(t,w)} d\mu(t).$$

◇

In the more general situation where we want to construct a Gram matrix for the space $\mathcal{L}_n$ for fixed n and where some of the of the α_k are repeated, then we can rearrange them so that equal α_k are grouped. We can then take the basis consisting of functions of the form $1/(1-\overline{\alpha}_k z), \ldots, 1/(1-\overline{\alpha}_k z)^{\nu_k}$ if α_k is repeated ν_k times, and similarly for the other α_i. Technically, it is rather difficult to write a general expression of an entry in the Gram matrix for such a basis, but with the help op the previous exercise, one can show the following.

Exercise 2.4 Suppose that there are $m+1$ different $\alpha_0, \ldots, \alpha_m$ (recall $\alpha_0 = 0$) which appear with multiplicity $\nu_0, \ldots, \nu_m$ respectively with $\sum_{i=0}^{m} \nu_i = n+1$, then the Gram matrix for the basis of $\mathcal{L}_n$ that we just mentioned will only depend on $\Omega_\mu^{(k)}(\alpha_i)$, $k = 0, \ldots, \nu_i - 1$, $i = 0, \ldots, m$. We could call this a generalized Pick matrix. ◇

Whatever the basis is that we choose for $\mathcal{L}_n$, it will always be similar to a Toeplitz matrix.

Theorem 2.1 *If $W = [w_0, \ldots, w_n]^T$ is a basis for $\mathcal{L}_n$, with Gram matrix $G_n(W) = \langle W, W \rangle_\mu = [\langle w_k, w_l \rangle_\mu]_{k,l=0}^{n}$, then there is a Toeplitz matrix T and an invertible matrix V such that $VG_n(W)V^H = T$.*

Proof. Note that we may always choose the basis $\{\ell_k(z) = z^k/\pi_n(z) : k = 0, \ldots, n\}$ with $\pi_n(z)$ as in (2.2). Note that the Gram matrix for this basis is Toeplitz, whose entries are trigonometric moments for the measure $d\mu(t)/|\pi_n(t)|^2$. Since every basis transformation is represented by an invertible matrix, the theorem follows. □

2.3 Reproducing kernels

Suppose $\{e_0,\ldots,e_n\}$ is a basis for $\mathcal{L}_n$, orthonormal with respect to μ. Then $k_n(z,w) = \sum_{k=0}^n e_k(z)\overline{e_k(w)}$ is a reproducing kernel for $\mathcal{L}_n$. I.e., $\langle f(\cdot), k_n(\cdot,w)\rangle_\mu = f(w)$ for all $f \in \mathcal{L}_n$ and $w \in \mathbb{D}$. If $f \in \mathcal{L}_m$ with $m > n$, then $\langle f(\cdot), k_n(\cdot,w)\rangle_\mu$ gives the orthogonal projection of f onto $\mathcal{L}_n$. Both observations follow immediately by writing f as a linear combination of the e_k and using the orthonormality. Note that the reproducing kernel is unique. The previous expression does not depend on a particular choice of the orthonormal basis. For example, it is immediately seen that if $\{e_k\}$ is an orthonormal basis for $\mathcal{L}_n$, then $\{B_n e_{k*}\}$ is also an orthonormal basis because $B_n e_{k*} \in \mathcal{L}_n$ for $k = 0,\ldots,n$ and

$$\langle B_n e_{k*}, B_n e_{l*}\rangle_\mu = \int |B_n(t)|^2 \overline{e_k(t)} e_l(t) d\mu(t) = \langle e_l, e_k\rangle = \delta_{k,l}.$$

Therefore, by using $B_n(z)e_{k*}(z) = B_n(z)\overline{e_k(1/\overline{z})}$ as orthonormal basis:

$$k_n(z,w) = B_n(z)\overline{B_n(w)} \sum_{k=0}^{n} \overline{e_k(1/\overline{z})} e_k(1/\overline{w}) = B_n(z)\overline{B_n(w)} k_n(1/\overline{w}, 1/\overline{z}).$$

By using a basis transformation, one may express $k_n(z,w)$ in terms of *any* basis for $\mathcal{L}_n$ as follows.

Theorem 2.2 *If $E_n(z) = [e_0(z),\ldots,e_n(z)]^T$ represents any basis for $\mathcal{L}_n$, then the reproducing kernel is given by*

$$k_n(z,w) = \frac{-1}{\det G_n} \det \begin{bmatrix} G_n & E_n(z) \\ E_n(w)^H & 0 \end{bmatrix}$$

where $G_n = \langle E_n, E_n\rangle_\mu$ is the Gram matrix of E_n. The superscript ^{H} refers to the complex conjugate (Hermitian) transpose.

Proof. Let $F_n = VE_n$ with V invertible be a column of orthonormal basis functions. Then $I = \langle F_n, F_n\rangle_\mu = V\langle E_n, E_n\rangle_\mu V^H = VG_nV^H$, so that $G_n^{-1} = V^HV$ and thus $k_n(z,w) = F_n(w)^H F_n(z) = E_n(w)^H V^H V E_n(z) = E_n(w)^H G_n^{-1} E_n(z)$, which proves the theorem. □

From now on, we shall use the notation ϕ_k for the orthonormal basis for $\mathcal{L}_n$ that is ordered such that $\phi_0 = 1$ and $\phi_k \in \mathcal{L}_k \backslash \mathcal{L}_{k-1}$ for $k = 1,2,\ldots$. Using the partial Blaschke products B_k, which also form a basis, we can write $\phi_n = a_{n,n}B_n + a_{n,n-1}B_{n-1} + \cdots + a_{n,1}B_1 + a_{n,0}$. We shall denote the leading coefficient $a_{n,n}$ (with respect to the basis B_k) by κ_n. Since ϕ_n is uniquely defined up to a constant of modulus 1, we can fix ϕ_n uniquely, for example by assuming that $\kappa_n > 0$, which we shall do throughout this lecture. Note that the trailing coefficient $a_{n,0}$ (with respect to the basis B_k) is given by $\phi_n(\alpha_1)$. To derive a similar notation for the leading coefficient, we need a generalization of the *reciprocal* function. We shall denote it by a superscript *. In general we set for any function $f \in \mathcal{L}_n$: $f^*(z) = B_n(z)f_*(z)$. Note that the superstar notation is ambiguous since it depends on the n that is considered. So a notation like $f^{[n]}$ instead of f^* would be more appropriate. However, in order not to overload the notation, if not explicitly mentioned, it should be clear from the context which n is intended. Note that with this notation we have $\phi_n^*(\alpha_n) = \kappa_n > 0$. We can now immediately write down as a consequence of the previous theorem:

Corollary 2.3 *If the $\phi_n = \kappa_n B_n + \cdots$ with $\kappa_n > 0$ are the orthonormal basis functions for $\mathcal{L}_n$ as introduced above and $k_n(z,w)$ is the reproducing kernel, then $k_n(z,\alpha_n) = \kappa_n \phi_n^*(z)$ and $k_n(\alpha_n,\alpha_n) = \kappa_n^2$.*

Proof. Exercise. □

The following Christoffel-Darboux (CD) relations hold:

Theorem 2.4 *With the notation just introduced we have*

$$\begin{aligned} k_n(z,w) &= \frac{\phi_{n+1}^*(z)\overline{\phi_{n+1}^*(w)} - \phi_{n+1}(z)\overline{\phi_{n+1}(w)}}{1-\zeta_{n+1}(z)\overline{\zeta_{n+1}(w)}} \\ &= \frac{\phi_n^*(z)\overline{\phi_n^*(w)} - \zeta_n(z)\overline{\zeta_n(w)}\phi_n(z)\overline{\phi_n(w)}}{1-\zeta_n(z)\overline{\zeta_n(w)}}. \end{aligned}$$

If $z = w = t \in \mathbb{T}$, then one may pass to the limit which will introduce derivatives:

$$P(t,\alpha_{n+1})k_n(t,t) = t[\phi_{n+1}'(t)\overline{\phi_{n+1}(t)} - \phi_{n+1}^{*'}(t)\overline{\phi_{n+1}^*(t)}],$$

with $P(t,z)$ the Poisson kernel (2.1).

Proof. Because the numerators and denominators of the expressions in the right hand side vanish for $z = 1/\overline{w}$, this zero cancels out and the right hand sides are in $\mathcal{L}_n$ as a function of z. Using $k_n(z,w) = B_n(z)\overline{B_n(w)}\sum_{k=0}^n \phi_{k*}(z)\overline{\phi_{k*}(w)}$ (which follows from Theorem 2.2) we may write

$$\frac{k_{n+1}(z,w)}{B_{n+1}(z)\overline{B_{n+1}(w)}} - \frac{k_n(z,w)}{B_n(z)\overline{B_n(w)}} = \phi_{(n+1)*}(z)\overline{\phi_{(n+1)*}(w)}.$$

Multiplying by the denominator gives

$$[k_n(z,w) + \phi_{n+1}(z)\overline{\phi_{n+1}(w)}] - \zeta_{n+1}(z)\overline{\zeta_{n+1}(w)}k_n(z,w) = \phi_{n+1}^*(z)\overline{\phi_{n+1}^*(w)}$$

which gives the first formula when z and w do not coincide on $\mathbb{T}$. The other formula is obtained when in the previous formula we replace n by $n-1$ so that

$$k_n(z,w) + \zeta_n(z)\overline{\zeta_n(w)}[\phi_n(z)\overline{\phi_n(w)} - k_n(z,w)] = \phi_n^*(z)\overline{\phi_n^*(w)}$$

from which the proof follows. The confluent case is left as an exercise. □

It can be seen that this relation does not depend on the fact that all $\alpha_k \in \mathbb{D}$.

Note that if we set $\phi_k = p_k/\pi_k$ with $\pi_k(z) = \prod_{i=0}^k (1-\overline{\alpha}_i z)$, then the p_k do not form an orthogonal polynomial sequence w.r.t. a positive measure, so that the theory of ORF is not quite the same as the theory of orthogonal polynomials w.r.t. a varying measure. With the CD formula, setting $z = w$, it is not difficult to derive the following property.

Theorem 2.5 *For all $n \geq 0$, $\phi_n^*(z) \neq 0$ for $z \in \mathbb{D}$ (hence, $\phi_n(z) \neq 0$ for $z \in \mathbb{E}$) and $|\phi_{n+1}(z)/\phi_{n+1}^*(z)| < 1$, $(= 1, > 1)$ for $z \in \mathbb{D}$, $(\mathbb{T}, \mathbb{E})$.*

Proof. Exercise. □

2.4 Recurrence relations

The kernels satisfy a fundamental recurrence relation

Theorem 2.6 *The kernels satisfy (superstar w.r.t. z)*

$$\begin{bmatrix} k_n^*(z,w) \\ k_n(z,w) \end{bmatrix} = t_n(z,w) \begin{bmatrix} k_{n-1}^*(z,w) \\ k_{n-1}(z,w) \end{bmatrix}$$

with

$$t_n(z,w) = c_n \begin{bmatrix} 1 & \overline{\rho}_n \\ \rho_n & 1 \end{bmatrix} \begin{bmatrix} \zeta_n(z) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \overline{\gamma_n} \\ \gamma_n & 1 \end{bmatrix}$$

where

$$\begin{aligned} c_n &= (1-|\rho_n|^2)^{-1} \\ \rho_n &= \rho_n(w) = \overline{\phi_n(w)}/\phi_n^*(w) \\ \gamma_n &= \gamma_n(w) = -\zeta_n(w)\rho_n(w). \end{aligned}$$

Proof. The CD relation implies

$$k_n(z,w) = \phi_n^*(z)\overline{\phi_n^*(w)} + \zeta_n(z)\overline{\zeta_n(w)}k_{n-1}(z,w).$$

Multiply this with $\overline{\rho_n(w)}$ and substitute (superstar w.r.t. z)

$$\phi_n^*(z)\overline{\phi_n^*(w)\rho_n(w)} = k_n^*(z,w) - \zeta_n(z)k_{n-1}^*(z,w)$$

to get

$$k_n(z,w)\overline{\rho_n(w)} = \zeta_n(z)\overline{\zeta_n(w)}k_{n-1}(z,w)\overline{\rho_n(w)} + k_n^*(z,w) - \zeta_n(z)k_{n-1}^*(z,w).$$

Take the superstar conjugate of this relation and solve for $k_n^*(z,w)$ and $k_n(z,w)$ and the result follows. □

The CD relation also implies a recurrence relation for the ϕ_n.

Theorem 2.7 *The orthonormal functions satisfy*

$$\begin{bmatrix} \phi_n(z) \\ \phi_n^*(z) \end{bmatrix} = \tilde{t}_n(z) \begin{bmatrix} \phi_{n-1}(z) \\ \phi_{n-1}^*(z) \end{bmatrix}$$

where

$$\begin{aligned} \tilde{t}_n(z) &= \frac{\kappa_n}{\kappa_{n-1}} \begin{bmatrix} 1 & 0 \\ 0 & z_n \end{bmatrix} \begin{bmatrix} \varepsilon_n & \delta_n \\ \overline{\delta}_n & \overline{\varepsilon}_n \end{bmatrix} \begin{bmatrix} \frac{z-\alpha_{n-1}}{1-\overline{\alpha}_n z} & 0 \\ 0 & \frac{1-\overline{\alpha}_{n-1}z}{1-\overline{\alpha}_n z} \end{bmatrix} \\ &= \frac{\kappa_n}{\kappa_{n-1}} \frac{1-\overline{\alpha}_{n-1}z}{1-\overline{\alpha}_n z} \begin{bmatrix} \overline{z}_{n-1}\varepsilon_n & 0 \\ 0 & z_n\overline{\varepsilon}_n \end{bmatrix} \begin{bmatrix} 1 & \overline{\lambda}_n \\ \lambda_n & 1 \end{bmatrix} \begin{bmatrix} \zeta_{n-1}(z) & 0 \\ 0 & 1 \end{bmatrix} \\ &= \sqrt{\frac{1-|\alpha_n|^2}{1-|\alpha_{n-1}|^2}} \frac{1}{\sqrt{1-|\lambda_n|^2}} \frac{1-\overline{\alpha}_{n-1}z}{1-\overline{\alpha}_n z} \begin{bmatrix} \eta_n^1 & 0 \\ 0 & \eta_n^2 \end{bmatrix} \begin{bmatrix} 1 & \overline{\lambda}_n \\ \lambda_n & 1 \end{bmatrix} \begin{bmatrix} \zeta_{n-1}(z) & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

with

$$\varepsilon_n = z_n \frac{1-\overline{\alpha}_{n-1}\alpha_n}{1-|\alpha_{n-1}|^2} \frac{\overline{\phi_n^*(\alpha_{n-1})}}{\kappa_n}, \quad \delta_n = \frac{1-\overline{\alpha}_n\alpha_{n-1}}{1-|\alpha_{n-1}|^2} \frac{\phi_n(\alpha_{n-1})}{\kappa_n},$$

$$\lambda_n = \overline{z}_{n-1}\frac{\overline{\delta}_n}{\overline{\varepsilon}_n} = \eta_n^1 \frac{\overline{\phi_n(\alpha_{n-1})}}{\phi_n^*(\alpha_{n-1})} \in \mathbb{D}, \quad \eta_n^1 = z_n\overline{z}_{n-1}\frac{1-\overline{\alpha}_{n-1}\alpha_n}{1-\overline{\alpha}_n\alpha_{n-1}} \in \mathbb{T},$$

$$\eta_n^2 = \overline{z}_{n-1}z_n\overline{\eta}_n^1 = \frac{1-\overline{\alpha}_n\alpha_{n-1}}{|1-\overline{\alpha}_n\alpha_{n-1}|}\frac{\phi_n^*(\alpha_{n-1})}{|\phi_n^*(\alpha_{n-1})|} \in \mathbb{T}.$$

Proof.(sketch) From the superstar conjugate (w.r.t. z) of the CD relation we find

$$\frac{\phi_n^*(z)\phi_n(\alpha_{n-1}) - \phi_n(z)\phi_n^*(\alpha_{n-1})}{\zeta_n(\alpha_{n-1}) - \zeta_n(z)} = \phi_{n-1}(z)\kappa_{n-1}.$$

The superstar conjugate of this relation is

$$\frac{\phi_n^*(z)\overline{\phi_n^*(\alpha_{n-1})} - \phi_n(z)\overline{\phi_n(\alpha_{n-1})}}{1-\zeta_n(z)\overline{\zeta_n(\alpha_{n-1})}} = \phi_{n-1}^*(z)\kappa_{n-1}.$$

Eliminate $\phi_n^*(z)$ between these two relations and the first form of the recurrence for $\phi_n(z)$ is obtained. The second one follows immediately from this because the CD relations imply $(|\phi_n^*(z)|^2 - |\phi_n(z)|^2)/(1-|\zeta_n(z)|^2) > 0$ so that $\phi_n^*(\alpha_{n-1}) \neq 0$ and $\lambda_n \in \mathbb{D}$. For the third one, recall that $\kappa_n = \phi_n^*(\alpha_n)$, and use again the CD relation with $z = w = \alpha_n$ to obtain the appropriate factor in front. □

Note that if all the α_k are zero, then we recover the polynomial case. The recurrence relation is just the Szegő recurrence and the λ_n being then equal to $\overline{\phi_n(0)}/\kappa_n$, are the Szegő parameters (and are sometimes called Schur or reflection coefficients).

Since this is derived from the CD relation, the first recurrence does not depend on the α_k being in $\mathbb{D}$. However, we can not allow $\overline{\alpha}_n\alpha_{n-1} = 1$, in which case λ_n is not defined. In such a case we call the system *degenerate*. If $\phi_n(\alpha_{n-1}) = 0$, then we call the system *exceptional*. If all the α_k are in $\mathbb{D}$ or all in $\mathbb{E}$, then the system is non-degenerate and $\lambda_n \in \mathbb{D}$.

The vector $[\phi_n, \phi_n^*]^T$ is a solution of the recurrence with initial condition $[\phi_0, \phi_0^*] = [1, 1]$. This is not the only solution. With the initial condition $[1, -1]$ we get another, independent solution. It will be formulated in terms of some functions $\psi_n \in \mathcal{L}_n$ which we shall call the ORF *of the second kind*. We introduce them as follows.

$$\psi_0 = 1, \quad \psi_n(z) = \int D(t,z)[\phi_n(t) - \phi_n(z)]d\mu(t), \quad n \geq 1.$$

Exercise 2.5 Let f be such that $f_* \in \mathcal{L}_{n-1}$, then prove by orthogonality that

$$\int D(t,z)[f(t)/f(z) - 1]\phi_n(t)d\mu(t) = 0.$$

From this derive that for such an f

$$\psi_n(z) = \int D(t,z)[\phi_n(t)f(t)/f(z) - \phi_n(z)]d\mu(t), \quad n \geq 1. \tag{2.3}$$

Similarly derive that for an $f \neq 0$ such that $f \in \mathcal{L}_n$ and $f(\alpha_n) = 0$ it holds that

$$\psi_n^*(z) = \int D(t,z)\left[\phi_n^*(t)f(t)/f(z) - \phi_n^*(z)\right]d\mu(t), \quad n \geq 1.$$

◇

We can now prove that $(\psi_n, -\psi_n^*)$ is another solution of the recurrence relation for (ϕ_n, ϕ_n^*), which corresponds to the initial condition $(1, -1)$. In other words (ψ_n, ψ_n^*) satisfies the same recurrence relation as (ϕ_n, ϕ_n^*), but with λ_n replaced by $-\lambda_n$.

Theorem 2.8 *With the notation of theorem 2.7 we have*

$$\begin{bmatrix} \phi_n(z) & \psi_n(z) \\ \phi_n^*(z) & -\psi_n^*(z) \end{bmatrix} = \tilde{t}_n(z) \begin{bmatrix} \phi_{n-1}(z) & \psi_{n-1}(z) \\ \phi_{n-1}^*(z) & -\psi_{n-1}^*(z) \end{bmatrix}. \tag{2.4}$$

Proof. We only have to prove the relation for the ψ_n. From the definition and the previous exercise we have for $n > 1$

$$\begin{bmatrix} \psi_{n-1}(z) \\ -\psi_{n-1}^*(z) \end{bmatrix} = -\Omega_\mu(z) \begin{bmatrix} \phi_{n-1}(z) \\ \phi_{n-1}^*(z) \end{bmatrix} + \int D(t,z) \begin{bmatrix} \phi_{n-1}(t) \\ \frac{\zeta_{n-1}(z)}{\zeta_{n-1}(t)} \phi_{n-1}^*(t) \end{bmatrix} d\mu(t).$$

Multiply from the left by $\tilde{t}_n(z)$ and the right hand side becomes

$$-\Omega_\mu(z) \begin{bmatrix} \phi_n(z) \\ \phi_n^*(z) \end{bmatrix} + \int D(t,z) Q(t,z) \begin{bmatrix} \phi_n(t) \\ \phi_n^*(t) \end{bmatrix} d\mu(t)$$

with

$$Q(t,z) = \frac{(z-\alpha_{n-1})(1-\overline{\alpha}_n t)}{(1-\overline{\alpha}_n z)(t-\alpha_{n-1})}.$$

Using a technique like in the previous exercise, it can be shown that the integral equals

$$\int D(t,z) \begin{bmatrix} \phi_n(t) \\ \frac{\zeta_n(z)}{\zeta_n(t)} \phi_n^*(t) \end{bmatrix} d\mu(t)$$

so that in the right hand side we get again the same expression as in the starting relation but with $n-1$ replaced by n. This proves the theorem for $n > 1$. Checking the theorem for $n = 1$ is left as an exercise. □

Far reaching generalizations of the CD relations can be obtained for any couple of solutions (x_n, x_n^+) and (y_n, y_n^+) of this recurrence. We give them without proof.

Theorem 2.9 *Given two solutions (x_n, x_n^+) and (y_n, y_n^+) of this recurrence, we can define*

$$F_n(z,w) = x_n^+(z) y_n(w) - x_n(z) y_n^+(w),$$

and then we have a Liouville-Ostrogradskii formula

$$\frac{F_n(z,w)}{1-\zeta_n(z)/\zeta_n(w)} - B_n(w) \frac{F_0(z,w)}{1-\zeta_0(z)/\zeta_0(w)} = -\sum_{k=0}^{n-1} x_k(z) y_k^+(w) \frac{B_n(w)}{B_k(w)}$$

and with the definition

$$G_n(z,w) = x_n^+(z)\overline{y_n^+(w)} - x_n(z)\overline{y_n(w)},$$

we have a Green formula

$$\frac{G_n(z,w)}{1-\zeta_n(z)\overline{\zeta_n(w)}} - \frac{G_0(z,w)}{1-\zeta_0(z)\overline{\zeta_0(w)}} = \sum_{k=0}^{n-1} x_k(z)\overline{y_k(w)}.$$

Choosing (x_n, x_n^+) and/or (y_n, y_n^+) equal to (ϕ_n, ϕ_n^*) or (ψ_n, ψ_n^*) gives several identities which we leave for the reader to discover. We just give one example:

$$\psi_n(z)\phi_n^*(z) + \psi_n^*(z)\phi_n(z) = 2B_n(z)P(z, \alpha_n) \tag{2.5}$$

with $P(z, w)$ the Poisson kernel (2.1). It is obtained by taking $z = w$ and $(x, x^+) = (\phi, \phi^*)$ and $(y, y^+) = (\psi, -\psi^*)$ in de Liouville-Ostrogradskii formula. It can also be obtained by taking determinants in (2.4). Therefore we refer to it as the determinant formula.

Note that the CD relation was crucial in our development so far. From this, we derived the recurrence relation for the kernels, and for the ORF which gave rise to the introduction of the second kind functions.

This can be inverted: if we have a recurrence relation for ϕ_n as given above, then they will be ORF with respect to some positive measure on $\mathbb{T}$. This is a Favard type theorem which reads as follows:

Theorem 2.10 (Favard) *Given a sequence $\{\alpha_0 = 0, \alpha_1, \alpha_2, \ldots\} \subset \mathbb{D}$ and suppose that with initial condition $\phi_0 = 1$, the ϕ_n are generated by a recurrence relation of the 3rd form given in Theorem 2.7 with all $\lambda_n \in \mathbb{D}$, and the unimodular constants such that $\phi_n^*(\alpha_n) > 0$, then these ϕ_n will form a sequence of ORF with respect to a probability measure on $\mathbb{T}$.*

We formulate the theorem here as a motivation for the introduction of quadrature formulas and will give its proof later in section 4.

Because the functions of the second kind satisfy a recurrence relation of this form (the λ_n have to be replaced by $-\lambda_n$), it follows by the Favard theorem that they also are a system of ORF, with respect to a positive measure on $\mathbb{T}$. Thus we can attribute to ψ_n the same properties as we can attribute to the ϕ_n (location of the zeros, CD relations, etc.). If $\Omega_\mu \in \mathcal{C}$ is the Riesz-Herglotz transform of μ, then clearly $1/\Omega_\mu$ is also in $\mathcal{C}$. So it can be written as the Riesz-Herglotz transform of some measure ν. It can be shown that if the ϕ_n are orthogonal w.r.t. μ, then the associated functions of the second kind ψ_n, will be orthogonal w.r.t. the associated ν.

To obtain a constructive proof for this theorem is closely related to the applications we mentioned in the beginning: We construct a measure that will generate an inner product that entails orthogonality in $\mathcal{L}_n$ and then let n tend to infinity. The first part is related to numerical quadrature and interpolation which we shall consider in the next section.

3 Quadrature and interpolation

3.1 Quadrature

One way to obtain a quadrature formula for the integral $I_\mu\{f\} = \int f(t)d\mu(t)$ is to interpolate f by a function f_n in a number of points $\{\xi_{ni}\}_{i=1}^n \subset \mathbb{T}$, and to approximate $I_\mu\{f\}$ by $I_n\{f\} = I_\mu\{f_n\}$.

Exercise 3.1 Consider the function $f_n \in \mathcal{L}_{p,q} = \{fg : f \in \mathcal{L}_{p*}, g \in \mathcal{L}_q\}$, $p, q \geq 0$, $p + q = n - 1$, where $\mathcal{L}_{p*} = \{f : f_* \in \mathcal{L}_p\}$. Show that this f_n shall interpolate f in the points $\{\xi_{ni}\}_{i=1}^n \subset \mathbb{T}$ if

$$f_n(t) = \sum_{k=1}^n L_{n,k}(t) f(\xi_{nk}), \quad L_{n,k}(t) = \ell_{n,k}(t) \left(\prod_{i=1}^q \frac{1 - \overline{\alpha}_i \xi_{nk}}{1 - \overline{\alpha}_i t} \right) \prod_{i=1}^p \frac{\xi_{nk} - \alpha_i}{t - \alpha_i},$$

where $\ell_{n,k}$ are the classical Lagrange polynomials, i.e., $\ell_{n,k}(\xi_{nj}) = \delta_{k,j}$. Note that the $L_{n,k} \in \mathcal{L}_{p,q}$ are rational generalizations for which also $L_{n,k}(\xi_{nj}) = \delta_{k,j}$. ◇

Thus the quadrature formula is $I_n\{f\} = \sum_{k=1}^{n} \lambda_{n,k} f(\xi_{nk})$, $\lambda_{n,k} = I_\mu\{L_{n,k}\}$. This is called an interpolating quadrature formula, which is obviously exact for all $f \in \mathcal{L}_{p,q}$.

To obtain the largest possible domain in which we get an exact quadrature formula, we have to choose the nodes ξ_{nk} in a particular way. Like in Gaussian quadrature formulas, we could try to choose them as the zeros of ϕ_n, but that is impossible because the ORF ϕ_n have all their zeros inside $\mathbb{D}$, and not on $\mathbb{T}$. However the following result holds.

Theorem 3.1 *The function $Q_n(z) = \phi_n(z) + \tau\phi_n^*(z)$, $\tau \in \mathbb{T}$ has n simple zeros on $\mathbb{T}$ and it is para-orthogonal which means that it is orthogonal to $\mathcal{L}_{n-1} \cap \mathcal{L}_n(\alpha_n)$ where $\mathcal{L}_n(\alpha_n) = \{f \in \mathcal{L}_n : f(\alpha_n) = 0\}$. Moreover $\langle 1, Q_n\rangle_\mu \neq 0 \neq \langle B_n, Q_n\rangle_\mu$.*

Proof. By the CD relations we know that $|\phi_n^*|^2 > |\phi_n|^2$ in $\mathbb{D}$ and the opposite inequality holds in $\mathbb{E}$. Thus Q_n can only have zeros in $\mathbb{T}$. These zeros are simple, otherwise, we would have $Q_n(\xi) = Q_n'(\xi) = 0$, which implies that in ξ we have $\phi_n/\phi_n^* = -\tau = -1/\overline{\tau} = \overline{\phi_n^*/\phi_n}$ or $\phi_n'\overline{\phi_n} - (\phi_n^*)'\overline{\phi_n^*} = 0$. But letting z and w approach $\xi \in \mathbb{T}$ in the CD formula leads to $\frac{(1-\overline{\alpha}_n\xi)(\xi-\alpha_n)}{1-|\alpha_n|^2}[\phi_n'(\xi)\overline{\phi_n(\xi)} - (\phi_n^*)'(\xi)\overline{\phi_n^*(\xi)}] = k_{n-1}(\xi,\xi) > 0$, which is a contradiction.

Concerning the para-orthogonality we note that $\langle f, \phi_n\rangle_\mu = 0$ if $f \in \mathcal{L}_{n-1}$ while if $f \in \mathcal{L}_n(\alpha_n)$, then it can be written as $B_n g_*$ with $g \in \mathcal{L}_{n-1}$, and thus $\langle f, \phi_n^*\rangle_\mu = \langle B_n g_*, B_n \phi_{n*}\rangle_\mu = \langle \phi_n, g\rangle_\mu = 0$. On the other hand $\langle 1, Q_n\rangle_\mu = \tau\langle 1, \phi_n^*\rangle_\mu = \tau\langle \phi_n, B_n\rangle_\mu \neq 0$, and similarly for the other inequality. □

Now take an arbitrary function $R \in \mathcal{L}_{n-1,n-1}$, suppose we interpolate it with a function $R_n \in \mathcal{L}_{0,n-1} = \mathcal{L}_{n-1}$, using the zeros $\{\xi_{n1}, \ldots \xi_{nn}\}$ of Q_n. Then simply writing it out shows that the interpolation error can be written as

$$E_n(z) = R(z) - R_n(z) = Q_n(z)S(z), \quad S \in \mathcal{L}_{n-1} \cap \mathcal{L}_n(\alpha_n).$$

Thus, because of the para-orthogonality of Q_n, we get an integration error $I_\mu\{E_n\} = 0$, so that the quadrature formula is exact in $\mathcal{L}_{n-1,n-1}$. This is the highest possible degree of exactness that can be obtained with n nodes in the sense that with n nodes on $\mathbb{T}$, it is impossible to integrate exactly all functions in $\mathcal{L}_{n,n-1}$ or $\mathcal{L}_{n-1,n}$. In this sense the quadrature formula is optimal and we shall refer to it as the *rational (or R-) Szegő quadrature formula.* We denote the (discrete) measure having masses λ_{nk} in the points ξ_{nk} by μ_n so that by definition $I_n\{f\} = I_{\mu_n}\{f\}$.

3.2 Interpolation

We have used interpolation to construct a quadrature formula and found that integration with respect to μ and μ_n is the same in $\mathcal{L}_{n-1,n-1}$. This happens if and only if their Riesz-Herglotz transforms take the same values in the points $\{\alpha_0, \ldots, \alpha_n\}$ (taking multiplicity into account) since the respective Gram matrices depend only on their (derivative) values in the poles $\{\alpha_k\}_{k=0}^{n-1}$ (see exercise 2.4). So let us first find out what the Riesz-Herglotz transform of μ_n is.

Theorem 3.2 *Set $\Omega_n(z) = \int D(t,z)d\mu_n(t) = -\frac{P_n(z)}{Q_n(z)}$. Then for $n \geq 1$*

$$P_n(z) = \int D(t,z)[Q_n(t)f(t)/f(z) - Q_n(z)]d\mu(t) = \psi_n(z) - \tau\psi_n^*(z)$$

for any f such that $f_* \in \mathcal{L}_{n-1} \cap \mathcal{L}_n(\alpha_n)$.

Proof. If we write the terms explicitly, then it becomes clear that the previous integrand is in $\mathcal{L}_{n-1,n-1}$ for $n \geq 2$, and thus we can replace the integral by the R-Szegő quadrature formula with nodes the zeros of Q_n, giving $P_n(z)f(z) = -f(z)Q_n(z)I_n\{D(\cdot,z)\} = -f(z)Q_n(z)\Omega_n(z)$. The case $n = 1$ is left as an exercise. The expression with ψ_k follows from (2.3) and its superstar. □

Because of this result we could call P_n a para-orthogonal function of the second kind. It is para-orthogonal with respect to the same measure as for which the ψ_n are orthogonal.

The previous theorem has an interesting corollary.

Corollary 3.3 *The weights of the R-Szegő formula are given by*

$$\lambda_{nk} = \frac{1}{2\xi_{nk}} \frac{P_n(\xi_{nk})}{Q'_n(\xi_{nk})} = \frac{1}{\sum_{i=0}^{n-1} |\phi_i(\xi_{nk})|^2} > 0.$$

Proof. Recall $\Omega_n(z) = -P_n(z)/Q_n(z) = I_n\{D(\cdot,z)\} = \sum_{k=1}^n \lambda_{nk} D(\xi_{nk}, z)$ and $Q_n(\xi_{nk}) = 0$. Thus after multiplying by $(z - \xi_{nk})$ we have

$$(z - \xi_{nk})\Omega_n(z) = \sum_{j\neq k} \lambda_{nj} D(\xi_{nj}, z)(z - \xi_{nk}) + \lambda_{nk} D(\xi_{nk}, z)(z - \xi_{nk}).$$

Taking the limit for $z \to \xi_{nk}$ gives the first formula. For the expression with the kernel, use the fact that $Q_n(\xi_{nk}) = 0$, thus $\tau = 1/\overline{\tau} = -\phi_n(\xi_{nk})/\phi_n^*(\xi_{nk})$, and the confluent CD formula to get

$$Q'_n(\xi_{nk}) = \frac{P(\xi_{nk}, \alpha_n)}{\xi_{nk}\overline{\phi_n(\xi_{nk})}} k_{n-1}(\xi_{nk}, \xi_{nk})$$

with $P(t,z)$ the Poisson kernel (2.1), while the determinant formula leads to $P_n(\xi_{nk}) = 2P(\xi_{nk}, \alpha_n)/\overline{\phi_n(\xi_{nk})}$, so the corollary is proved. □

We can now derive interpolation properties for $\Omega_n = -P_n/Q_n$:

Theorem 3.4 *Let Ω_n and Ω_μ be the Riesz-Herglotz transforms of μ_n and μ respectively, then $\Omega_\mu(z) - \Omega_n(z) = zB_{n-1}(z)h(z)$, for $n \geq 1$ with h analytic in $\mathbb{D}$.*

Proof. This follows because of exercise 2.4. □

Since we assumed $\tau \in \mathbb{T}$, we cannot set $\tau = 0$, but we do have the same type of interpolation with an extra interpolation in α_n.

Theorem 3.5 *For the ORF and the ORF of the second kind we have*

$$\Omega_\mu(z)\phi_n(z) + \psi_n(z) = zB_{n-1}(z)h(z), \quad \Omega_\mu(z)\phi_n^*(z) - \psi_n^*(z) = zB_n(z)g(z)$$

with h and g analytic in $\mathbb{D}$.

Proof. Since $[\phi_n\Omega_\mu + \psi_n]/B_{n-1} = \int D(t,z)\phi_n(t)/B_{n-1}(t)d\mu(t)$, we find a function analytic in $\mathbb{D}$ because it is a Cauchy-Stieltjes integral and setting $z = 0$, so that $D(t,0) = 1$, the integral becomes $\langle \phi_n, B_{n-1}\rangle_\mu = 0$. The second relation is similarly proved since now for $z = 0$ we find $\langle \phi_n^*, B_n\rangle_\mu = 0$. □

Note that by the determinant formula, we have for $R_n(z) = \psi_n^*/\phi_n^*$ that for $t \in \mathbb{T}$: $w_n(t) = \operatorname{Re} R_n(t) = \frac{1}{2}[R_n(t) + R_{n*}(t)] = P(t,\alpha_n)/|\phi_n(t)|^2$. Because of the extra interpolation in α_n, one might expect that the absolute continuous measure $w_n(t)d\lambda(t)$ gives the same integrals in $\mathcal{L}_{n,n}$, and it does indeed.

Theorem 3.6 *The inner product in $\mathcal{L}_n$ is the same for the measure μ and for the absolute continuous measure with weight $w_n = P(\cdot,\alpha_n)/|\phi_n|^2$.*

Proof. Since $\int |\phi_n(t)|^2 w_n(t)d\lambda(t) = \int P(t,\alpha_n)d\lambda(t) = 1 = \|\phi_n\|_\mu^2$, the norm is maintained. Moreover

$$\langle \phi_n, \phi_k\rangle_{w_n} = \int P(t,\alpha_n)\frac{\phi_{k*}(t)}{\phi_{n*}(t)}d\lambda(t) = \int P(t,\alpha_n)\left\{\frac{\phi_k^*(t)B_n(t)/B_k(t)}{\phi_n^*(t)}\right\}d\lambda(t) = 0,$$

which follows because the factor in curly brackets is analytic in $\mathbb{D}\cup\mathbb{T}$, so that we may apply Poisson's formula to find zero because $B_n(z)/B_k(z)$ is zero in α_n. Thus ϕ_n has norm 1 and is orthogonal to $\mathcal{L}_{n-1}$ for the weight w_n, and because the recurrence relation then defines all the previous ϕ_k uniquely (provided they have the proper normalization of positive leading coefficient), the theorem follows. □

We note that the para-orthogonality conditions alone do not define the functions Q_n completely. The para-orthogonal functions that we proposed are also $\overline{\tau}$-invariant, which means that $Q_n^* = \overline{\tau}Q_n$. The latter is essential to guarantee that they have n simple zeros on $\mathbb{T}$. It can be shown that this invariance property and the para-orthogonality completely defines the Q_n up to the parameter $\tau \in \mathbb{T}$ and a normalizing constant factor.

3.3 Interpolation and quadrature using the kernels

After we formulated the recurrence relation for the kernels in Theorem 2.4, we gave the recurrence for the ORF in Theorem 2.7, but from there on, the kernels were neglected. However, what has been developed for the ORF, can be repeated and generalized to the kernels. Since the methodology is completely analogous, we leave the proofs in this section as a major exercise.

First of all, we note that the kernels are produced by their recurrence relation when we give the initial conditions $(k_0^*, k_0) = (1,1)$. As in the case of the ORF, we obtain another independent solution when we start with $(l_0^*, -l_0) = (1,-1)$. The resulting kernels could be called kernels of the second kind. They satisfy

$$\begin{bmatrix} l_n^*(z,w) \\ -l_n(z,w) \end{bmatrix} = t_n(z,w)\begin{bmatrix} l_{n-1}^*(z,w) \\ -l_{n-1}(z,w) \end{bmatrix}$$

with $t_n(z,w)$ as in Theorem 2.4. By our remark after Theorem 2.10, it should be clear that $l_n(z,w) = \sum_{k=0}^n \psi_n(z)\overline{\psi_k(w)}$ is a reproducing kernel for the space $\mathcal{L}_n$ w.r.t. the same measure as for which the ψ_k are the ORF. The interpolation properties are a bit harder to obtain because they require a parameterized (in w) Riesz-Herglotz transform:

$$\Omega_\mu(z,w) = \int \frac{D(t,z)}{P(t,w)} d\mu(t) + c, \quad c = \frac{\overline{w}\int t d\mu(t) - w \int t^{-1} d\mu(t)}{1-|w|^2} \in i\mathbb{R}, \quad (z,w) \in \mathbb{D}\times\mathbb{D},$$

$D(t,z)$ the Dirichlet and $P(t,w)$ the Poisson kernel. The c is chosen to make $\Omega_\mu(w,w)$ real (and hence it equals 1). Then it is possible to formulate an interpolation property saying that $l_n(z,w) - k_n(z,w)\Omega_\mu(z,w)$ vanishes for all $z \in \{w, \alpha_1, \ldots, \alpha_n\}$. Thus the special role of $\alpha_0 = 0$ has been removed and is played by an arbitrary $w \in \mathbb{D}$. Since $k_n(z,w)$ does not vanish for any $z \in \mathbb{D}\cup\mathbb{T}$ if $w \in \mathbb{D}$, we may consider $\Omega_n(z,w) = l_n(z,w)/k_n(z,w)$, which is the parameterized Riesz-Herlotz transform of the absolutely continuous measure $P(t,w)k_n(w,w)/|k_n(t,w)|^2 d\lambda(t)$. The inner product in $\mathcal{L}_n$ w.r.t. the latter will not depend on w and will give the same results as the inner product w.r.t. $d\mu(t)$.

4 Density and the proof of Favard's theorem

4.1 Density

For the density of the Blaschke products in $L_p(\mathbb{T})$, one can easily adapt a result of [1, p.244] to find the following.

Theorem 4.1 *Define the Blachke products as before for $n \geq 0$ and $B_{-n} = 1/B_n$ for $n \geq 1$. Then the system $\{B_n\}_{n\in\mathbb{Z}}$ is complete in $L_p(\mathbb{T})$, for any $p \geq 1$ as well as in the space $C(\mathbb{T})$ of continuous functions on $\mathbb{T}$ if and only if $\sum(1-|\alpha_k|) = \infty$ (the Blaschke product diverges). Also the system $\{B_n\}_{n=0}^{\infty}$ is complete in H_p for any $p \geq 1$ if and only if $\sum(1-|\alpha_k|) = \infty$ (the Blaschke product diverges).*

Walsh [31, p.305-306] states

Theorem 4.2 *If the Blaschke product diverges and $f \in H_2$, then $f_n \in \mathcal{L}_n$ which interpolates f in $\{w, \alpha_1, \ldots, \alpha_n\} \subset \mathbb{D}$ (w is an arbitrary but fixed number in $\mathbb{D}$) will converge to f uniformly on compact subsets of $\mathbb{D}$, and if f is continuous up to $\mathbb{T}$, we also have uniform convergence on $\mathbb{T}$.*

For a general probability measure, Szegő's condition $\log \mu' \in L_1$ plays a role. For example, from Walsh [31, p.116,186,50,92,144] we may conclude that the polynomials are dense in $L_p(\mu)$ if and only if $\log \mu' \notin L_1$. For the rational case, the situation seems to be less simple.

Theorem 4.3 *(1) If $\sum_k(1-|\alpha_k|) = \infty$ then $\log \mu' \notin L_1 \Leftrightarrow \{B_n\}_{n\geq 0}$ is complete in $L_2(\mu)$*
(2) If $\log \mu' \in L_1$ then $\sum_k(1-|\alpha_k|) = \infty \Leftrightarrow \{B_n\}_{n\geq 0}$ is complete in $H_2(\mu)$
(3) If $\sum_k(1-|\alpha_k|) = \infty$ then $\{B_n\}_{n\in\mathbb{Z}}$ is complete in $L_2(\mu)$.

Note that we have no equivalence anymore. The divergence of the Blaschke product implies completeness of $\{B_n\}_{n=0}^{\infty}$ in $H_2(\mu)$ and of $\{B_n\}_{n\in\mathbb{Z}}$ in $L_2(\mu)$, but the converse need not be true. We leave it to the reader to look up the proofs of the above theorems. The precise characterization of completeness of the Blaschke products in not totally cleared out.

4.2 Proof of Favard theorem

It is easily proved by induction that the ϕ_n^* generated by the recursion are indeed superstar conjugates of the ϕ_n that it generates and the parameters λ_n are given by their expressions of Theorem 2.7. Moreover, the $\{\phi_k\}_{k=0}^n$ are a set of ORF with respect to the weight $w_n(t) = P(t,\alpha_n)/|\phi_n|^2$. This is shown as follows. Note that for $0 \le m \le n$

$$\langle \phi_n, \phi_m \rangle_{w_n} = \int \frac{\phi_{m*}(t)}{\phi_{n*}(t)} P(t,\alpha_n) d\lambda(t) = \left. \frac{B_n(z)}{B_m(z)} \frac{\phi_m^*(z)}{\phi_n^*(z)} \right|_{z=\alpha_n} = \delta_{nm}.$$

Thus $\phi_n \perp_{w_n} \mathcal{L}_{n-1}$, and because an inverse recurrence (obtained by multiplying (2.4) from the left with the inverse of the matrix $\tilde{t}_n(z)$), defines all the previous ORF uniquely, the orthonormality with respect to w_n is proved.

Now for $n \to \infty$, we have a sequence of weights w_n, uniformly bounded ($\int w_n d\lambda = 1$), so that there is a subsequence w_{n_k} which converges weakly. Thus there is some μ such that $\lim_{n\to\infty} \int f(t) w_{n_k}(t) d\lambda(t) = \int f(t) d\mu(t)$ for all functions f continuous on $\mathbb{T}$. If $\sum(1-|\alpha_k|) = \infty$, then a previous completeness result shows that the measure μ is unique because the Riesz representation of a linear functional on $C(\mathbb{T})$ is unique.

5 Convergence

5.1 Orthogonal polynomials w.r.t. varying measures

For the convergence of the ORF we shall rely on convergence results of orthogonal polynomials with respect to varying measures OPVM. Let $\mathcal{P}_m$ we the space of polynomials of at most degree m and construct a sequence of OPVM for measure $d\mu(t)/|\pi_n(t)|^2$ with $\pi_n(t) = \prod_{k=1}^n (1-\overline{\alpha}_k z)$. We denote the orthonormal ones by $\phi_{n,k}(z) = v_{n,k} z^k + \cdots$, $v_{n,k} > 0$. By our general theory they satisfy the recurrence $\phi_{n,m}(z) = e_{n,m}[z\phi_{n,m-1}(z) + \overline{\lambda}_{n,m} \phi_{n,m-1}^{\bar{*}}(z)]$ with $\lambda_{n,m} = \overline{\phi_{n,m}(0)}/v_{n,m}$ and $e_{n,m} = (1-|\lambda_{n,m}|^2)^{-1/2} = v_{n,m}/v_{n,m-1}$. Here $\phi_{n,m}^{\bar{*}}(z) = z^n \overline{\phi_{n,m}(1/\overline{z})}$. From these OP we can construct a set of ORF for $\mathcal{L}_n$, but as n increases, the whole set of ORF will change.

Exercise 5.1 Denote $f_{n,m}(z) = t_{n,m}\phi_{n,m}(z)/\pi_n(z)$ with $t_{n,m} \in \mathbb{T}$ for an appropriate normalization. Prove that the $\{f_{n,m} : m = 0,\ldots,n\}$ is a set of orthonormal rational functions for $\mathcal{L}_n$ with respect to μ. They are obtained by orthonormalizing the basis $\{z^k/\pi_n(z) : k = 0,\ldots,n\}$. Show also that

$$h_n(z) = \frac{z\phi_{n,n}(z) - \frac{\phi_{n,n}(\alpha_n)}{\phi_{n,n}^{\bar{*}}(\alpha_n)} \alpha_n \phi_{n,n}^{\bar{*}}(z)}{(z-\alpha_n)\pi_n(z)} \in \mathcal{L}_n \setminus \mathcal{L}_{n-1} \tag{5.1}$$

and that $\langle h_n, g \rangle_\mu = 0$ for all $g \in \mathcal{L}_{n-1}$, so that there is some constant c_n with $|c_n| = |\phi_n^*(0)|$ such that $h_n = c_n \phi_n$. ◇

We need the following conditions for the probability measure μ and the point set $A = \{\alpha_1, \alpha_2, \ldots\}$:

1. $\mu' > 0$ a.e. (λ) (Erdős-Turán condition)
2. $\sum_{n=1}^\infty (1-|\alpha_n|) = \infty$ (BD = Blaschke divergence condition)

The first condition is denoted as $\mu \in ET$. If it satisfies the stronger Szegő condition $\log \mu' \in L_1$, we denote it by $\mu \in SZ$. The second condition is denoted as $A \in BD$, while if it is replaced by the stronger condition that A is compactly contained in $\mathbb{D}$ (i.e., the α_k stay away from the boundary), we denote it as $A \in CC$. The following results are borrowed from the work of Guillermo Lopez [20, 21, 22].

Theorem 5.1 *If $\mu \in ET$ and $A \in BD$ then with our previous notation*

1. $\lim_{n\to\infty} \lambda_{n,n+k+1} = 0$.
2. $\lim_{n\to\infty} \upsilon_{n,n+k+1}/\upsilon_{n,n+k} = 1$.
3. $\lim_{n\to\infty} \phi_{n,n+k+1}(z)/\phi_{n,n+k}(z) = z$ *locally uniformly in* $\mathbb{E}$.
4. $\lim_{n\to\infty} \phi^{\bar{*}}_{n,n+k+1}(z)/\phi^{\bar{*}}_{n,n+k}(z) = 1$ *locally uniformly in* $\mathbb{D}$.
5. $\lim_{n\to\infty} \phi^{\bar{*}}_{n,n+k}(z)/\phi_{n,n+k}(z) = 0$ *locally uniformly in* $\mathbb{E}$.

If $\mu \in SZ$, $A \in CC$ and σ is the outer spectral factor of μ, then

6. $\lim_{n\to\infty} \phi^{\bar{*}}_{n,n+k}(z)/\pi_n(z) = 1/\sigma(z)$ *locally uniformly in* $\mathbb{D}$.

5.2 Szegő's condition and convergence

We can now show that the following holds.

Theorem 5.2 *If $\mu \in SZ$ and $A \in BD$, then we have the following (l.u. means locally uniformly)*

$$\lim_{n\to\infty} k_n(z,w) = \frac{1}{\overline{\sigma(w)}(1-\overline{w}z)\sigma(z)}, \quad \text{l.u. } (z,w) \in \mathbb{D}\times\mathbb{D},$$

$$\lim_{n\to\infty} (1-\overline{\alpha}_n z)\frac{\phi_n^*(z)}{\phi_n^*(0)} = \frac{\sigma(0)}{\sigma(z)}, \quad \text{l.u. } z \in \mathbb{D},$$

$$\lim_{n\to\infty} \phi_n(z) = \lim_{n\to\infty} \frac{\phi_n(z)\phi_n^*(0)}{\phi_n^*(z)} = \lim_{n\to\infty} \frac{\phi_n(z)}{\phi_n^*(z)} = 0, \quad \text{l.u. } z \in \mathbb{D}.$$

If $\mu \in SZ$, $\sigma(0) > 0$, and $A \in CC$ then

$$\lim_{n\to\infty} \eta_n \phi_n^*(z) \frac{1-\overline{\alpha}_n z}{\sqrt{1-|\alpha_n|^2}} = \frac{1}{\sigma(z)}, \quad \eta_n = \frac{|\phi_n^*(0)|}{\phi_n^*(0)} \in \mathbb{T}, \quad \text{l.u. } z \in \mathbb{D}.$$

Proof. For the first relation, note that since the $f_{n,k}$ form an orthonormal basis for $\mathcal{L}_n$, we have

$$k_n(z,w) = \sum_{k=0}^{n} f_{n,k}(z)\overline{f_{n,k}(w)} = \frac{\sum_{k=0}^{n} \phi_{n,k}(z)\overline{\phi_{n,k}(w)}}{\pi_n(z)\overline{\pi_n(w)}}.$$

Then use the CD relation for the ϕ_{nk} and divide by $f^*_{n,n}(z)\overline{f^*_{n,n}(w)}$:

$$\frac{k_n(z,w)}{f^*_{n,n}(z)\overline{f^*_{n,n}(w)}} = \frac{1}{1-\overline{w}z} - \frac{z\overline{w}}{1-\overline{w}z}\left(\frac{\phi_{n,n}(z)}{\phi^{\bar{*}}_{n,n}(z)}\right)\overline{\left(\frac{\phi_{n,n}(w)}{\phi^{\bar{*}}_{n,n}(w)}\right)}. \tag{5.2}$$

The last two factors go to 0, while for some $\eta_n \in \mathbb{T}$, $\lim_{n\to\infty} \eta_n f_{n,n}^*(z) = 1/\sigma(z)$, which proves the first result.
Since $h_n = c_n\phi_n$, we have by our expression for h_n

$$(1-\overline{\alpha}_n z)\phi_n^*(z) = \frac{1}{\overline{c}_n}\frac{\phi_{n,n}^{\bar{*}}(z)}{\pi_n(z)}\left[1-\overline{\eta}_n\overline{\alpha}_n z\frac{\phi_{n,n}(z)}{\phi_{n,n}^{\bar{*}}(z)}\right], \quad \eta_n = \frac{\phi_{n,n}(\alpha_n)}{\phi_{n,n}^{\bar{*}}(\alpha_n)}. \tag{5.3}$$

Recall that $\phi_{n,n}(z)/\phi_{n,n}^{\bar{*}}(z) \to 0$ l.u. in $\mathbb{D}$, and being a Blaschke product it is bounded by 1 in $\mathbb{D}\cup\mathbb{T}$ so that $|\eta_n| \le 1$. Hence the term between square brackets goes to 1 as $n \to \infty$. On the other hand, $\phi_{n,n}^{\bar{*}}(z)/\pi_n(z)$ converges to $1/\sigma(z)$ l.u. in $\mathbb{D}$, up to some normalizing constant. But this constant will cancel against the same constant in the denominator which is obtained by setting $z = 0$. This proves the second result.
Taking $z = w$ in the first relation implies $\sum_{k=1}^n |\phi_k(z)|^2$ converges for $n \to \infty$. Thus $\lim_n \phi_n(z) = 0$. When multiplying the inverse of the second relation of this theorem with $\phi_n(z)$, and noting that $(1-\overline{\alpha}_n z)^{-1}$ is uniformly bounded if z is in a compact subset of $\mathbb{D}$, we also get the second limit.
For the third one, we note that $|\phi_n^*(0)|$ is uniformly bounded away from 0, because the CD relation implies $|\phi_n^*(0)|^2 \ge 1$.
For the last limit, note that η_n is used to normalize ϕ_n^* to make it positive in $z = 0$. By the CD relation for $z = w = 0$

$$\frac{|\phi_n^*(0)|^2}{1-|\alpha_n|^2} = k_{n-1}(0,0) + \frac{|\phi_n(0)|^2}{1-|\alpha_n|^2}.$$

Because $\phi_n(0) \to 0$, and the α_k stay away from the circle, the last term goes to 0, and we know that the second one goes to $|\sigma(0)|^{-2}$. Thus $|\phi_n^*(0)|/\sqrt{1-|\alpha_n|^2} \to 1/\sigma(0)$. In combination with the second limit of this theorem, the result follows. □

When $\mu \notin SZ$, then σ need not be defined, and the previous relations can not be obtained. A way to avoid the σ is to consider ratio asymptotics, so that the σ cancel out. These are weaker results in the sense that if $\mu \in SZ$, then the ratio asymptotics are almost immediately obtained from the stronger asymptotics that were previously obtained. We use again the OPVM with measure $\mu/|\pi_n|^2$. Denoting the reproducing kernel for $\mathcal{P}_n$ w.r.t. this measure as $k_{n,n}(z,w)$, it is easily seen that $k_{n,n}(z,0) = v_n\phi_{n,n}^{\bar{*}}(z)$. Also the usual rational kernel for $\mathcal{L}_n$ satisfies $k_n(z,0) = k_{n,n}(z,0)/\pi_n(z)$.

Exercise 5.2 Applying Theorem 3.6 to the OPVM, prove that the inner product in $\mathcal{L}_n$ w.r.t. $d\mu(t)$ and w.r.t. $v_n^2/|k_n(t,0)|^2 d\lambda(t)$ is the same. Of course, this is also a direct consequence of the results in section 3.3. ◇

Theorem 5.3 *If $\mu \in ET$ and $A \in BD$, then*

$$\lim_{n\to\infty}\frac{k_n(z,0)}{k_{n+1}(z,0)} = 1, \quad l.u.\ z \in \mathbb{D}.$$

and with the superstar referring to z,

$$\lim_{n\to\infty} k_n^*(z,0)/k_n(z,0) = 0, \quad l.u.\ z \in \mathbb{D}.$$

Proof. Set $g_n(z) = k_n(z,0)/k_{n+1}(z,0) - v_n/v_{n+1}$, then by the previous exercise we see that

$$\int |g_n(t)|^2 d\lambda(t) = 2\frac{v_n^2}{v_{n+1}^2}\left(1 - \frac{v_n}{v_{n+1}}\right). \tag{5.4}$$

On the other hand, because g_n is analytic in $\mathbb{D}\cup\mathbb{T}$, we have by the Poisson formula $g_n(z) = \int P(t,z)g_n(t)d\lambda(t)$. For $t \in \mathbb{T}$, $|P(t,z)| \le M$ holds uniformly for z in a compact subset of $\mathbb{D}$. Thus we also have $|g_n(z)| \le M \int |g_n(t)|d\lambda(t)$ and also $|g_n(z)|^2 \le M^2 \int |g_n(t)|^2 d\lambda(t)$ so that

$$|g_n(z)|^2 \le 2M^2\frac{v_n^2}{v_{n+1}^2}\left(1 - \frac{v_n}{v_{n+1}}\right). \tag{5.5}$$

Thus $\frac{k_n(z,0)}{k_{n+1}(z,0)} \to 1$ iff $\frac{v_n}{v_{n+1}} \to 1$. Obviously $k_{n-1}(0,0)/k_n(0,0) \le 1$, but using the CD relation we also have

$$\frac{k_{n-1}(0,0)}{k_n(0,0)} = \frac{|\phi_n^*(0)|^2 - |\phi_n(0)|^2}{|\phi_n^*(0)|^2 - |\alpha_n|^2|\phi_n(0)|^2} \ge \frac{|\phi_n^*(0)|^2 - |\phi_n(0)|^2}{|\phi_n^*(0)|^2 - |\phi_n(0)|^2} = 1.$$

So that $\lim_{n\to\infty} v_n^2/v_{n+1}^2 = 1$, which proves the first part.

For the second relation, note that $k_n(z,0) = \phi_{n,n}^{\bar{*}}(z)v_n\eta_n/\pi_n(z)$ and hence $k_n^*(z,0) = \phi_{n,n}(z)v_n\overline{\eta}_n/\pi_n(z)$ with $\eta_n \in \mathbb{T}$. Take the ratio and use $\phi_{n,n}(z)/\phi_{n,n}^{\bar{*}}(z) \to 0$ l.u. in $\mathbb{D}$, to conclude the proof. □

Theorem 5.4 *If $\mu \in ET$ and $A \in CC$, then*

$$\lim_{n\to\infty} \frac{\zeta_n(z)\phi_n(z)}{\phi_n^*(z)} = \lim_{n\to\infty}\frac{\phi_n(z)}{\phi_n^*(z)} = \lim_{n\to\infty}\lambda_n = 0, \quad \textit{l.u. in } \mathbb{D}.$$

Denote $\Phi_n^*(z) = \phi_n^*(z)/\phi_n^*(0)$, then under the previous conditions

$$\lim_{n\to\infty}\frac{\Phi_{n+1}^*(z)(1-\overline{\alpha}_{n+1}z)}{\Phi_n^*(z)(1-\overline{\alpha}_n z)} = 1, \quad \textit{l.u. in } \mathbb{D}.$$

Proof. Repeat the proof of the second part in Theorem 5.2. I.e., use exercise 5.1 giving $\phi_n = c_n h_n$ and its superstar conjugate, to find expressions for $(z-\alpha_n)\phi_n(z)$ and $(1-\overline{\alpha}_n z)\phi_n^*(z)$ and take their ratio. Then using $\phi_{n,n}(z)/\phi_{n,n}^{\bar{*}}(z) \to 0$ from theorem 5.1 leads to the first conclusion. Note that we need $\phi_{n,n}(\alpha_n)/\phi_{n,n}^{\bar{*}}(\alpha_n)$ to go to zero, which can only be guaranteed when $A \in CC$. It is also needed for the convergence of $\phi_n(0)/\phi_n^*(0)$ because this will follow form Theorem 5.3 or the previous exercise if A is in a compact subset of $\mathbb{D}$. For the last equality note that $|\lambda_n| = |\phi_n(\alpha_{n-1})/\phi_n^*(\alpha_{n-1})|$.

For the second formula, write the second CD formula for $k_n(z,0)$ and its superstar conjugate and eliminate $\phi_n(z)$ to obtain

$$\Phi_n^*(z)(1-\overline{\alpha}_n z) = \frac{\phi_n^*(z)(1-\overline{\alpha}_n z)}{\phi_n^*(0)} = \frac{k_n(z,0)}{v_n^2}\left[1 - \overline{\alpha}_n z\frac{\overline{\phi_n(0)}}{\phi_n^*(0)}\frac{k_n^*(z,0)}{k_n(z,0)}\right].$$

Note that the term in square brackets goes to 1 l.u. in $\mathbb{D}$ as $n \to \infty$. Rewrite this for n replaced by $n+1$ and take their ratio. Because $k_{n+1}(z,0)/k_n(z,0)$ and v_{n+1}/v_n go to 1 as $n \to \infty$, the second result follows. □

6 Szegő's problem

This problem is essentially the problem of linear prediction. Recall that the linear prediction problem can be formulated as finding $f \in H_2(\mu)$ such that $\|f\|_\mu^2$ is minimal with the side condition that $f(0) = 1$. This f represented the prediction error. (Note that we have replaced the z of the first section by $1/z$ to be in line with our discussion of ORF, which is obviously a matter of convention.) We consider a slightly more general situation where we replace the constraint by $f(w) = 1$ for some $w \in \mathbb{D}$. This is a problem that can be solved in any reproducing kernel Hilbert space.

Theorem 6.1 *Let H be a reproducing kernel Hilbert space with kernel $k(z,w)$, then the minimum of $\|f\|_\mu^2$ with $f(w) = 1$ is obtained for $f(z) = K(z) \equiv k(z,w)/k(w,w)$ and the minimum is $1/k(w,w)$.*

Proof. Obviously $\|K(z)\|_\mu^2 = 1/k(w,w)$. For any other f satisfying $f(w) = 1$ for which $\|f\|_\mu^2 = m$ we have $0 \le \|f - K\|_\mu^2 = \|f\|_\mu^2 + 1/k(w,w) - 2\mathrm{Re}\,\langle f, K\rangle_\mu = m - 1/k(w,w)$. Thus $m \ge 1/k(w,w)$. □

From the polynomial case, it is known that the reproducing kernel for $H_2(\mu)$ is given by the Szegő kernel $s(z,w) = [(1 - \overline{w}z)\sigma(z)\overline{\sigma(w)}]^{-1}$ where σ is the outer spectral factor of μ, which we suppose to be normalized by $\sigma(0) > 0$. Thus we have a completely predictable sequence if the minimum $|\sigma(0)|^2$ of $\|f\|_\mu^2$, $f(0) = 1$ is zero, i.e. since $|\sigma(0)|^2 = \exp\{\int \log \mu'(t) d\lambda(t)\}$, if $\int \log \mu' d\lambda = -\infty$, and thus $\log \mu' \notin L_1$. If we do not have complete predictability, it is still a valuable objective to find the best possible predictor by minimizing the prediction error.

Instead of minimizing over the complete past (i.e. all of $H_2(\mu)$), we may be less ambitious and start by minimizing over a finite dimensional subspace, e.g. $\mathcal{L}_n$. There the minimum of the general problem is $k_n(z,w)/k_n(w,w)$.

The advantage of computing these approximants using ORF over the computation with OPVM is that they are easily computed recursively. If the approximation is not good enough for a certain n, then increasing n by 1 requires just one more step of the recurrence relation. For the OPVM, increasing n by 1, would mean that we have to start the computations all over again. Note also that we have at every step an estimate of the prediction error which is $1/k_n(0,0)$ or $1/k_n(w,w)$ for the general problem. Thus if the $|\phi_n(w)|$ do not go to zero fast enough, then $k_n(w,w)$ will go to ∞ and the error will go to zero, but under the conditions of our theorems, the $k_n(w,w)$ will converge, so that the error will not go to zero. It will be bounded from below by $(1 - |w|^2)|\sigma(w)|^2$.

We also note the following result

Theorem 6.2 *Let $\mu \in SZ$, $w \in \mathbb{D}$ a fixed number, and let $s_w(z) = s(z,w)$ be the Szegő kernel. Consider the problem $\min \|f - s_w\|_\mu^2$ over all $f \in \mathcal{L}_n$, then the solution is $k_n(z,w)$ and the minimum is $s_w(w) - k_n(w,w)$. If $A \in BD$ then $\lim_{n\to\infty} \|k_n(t,w) - s_w(t)\|_\mu = 0$.*

Proof. Suppose $f(w) = a$, then $\|s_w - f\|_\mu^2 = \|s_w\|_\mu^2 + \|f\|_\mu^2 - 2\mathrm{Re} f(w)$. Thus we have to minimize $\|f\|_\mu^2 - 2\mathrm{Re}\; a$ over $f \in \mathcal{L}_n$ with $f(w) = a$ and minimize the result over all possible a. In other words, we have to find the infimum over a of $|a|^2/k_n(w,w) - 2\mathrm{Re}\; a$ and this is given by $a = k_n(w,w)$. This proves the first part. If $A \in BD$ then as $n \to \infty$, $\mathcal{L}_n$ becomes dense in $H_2(\mu)$, which means that the error in $H_2(\mu)$ goes to 0. □

Part II: The matrix case

The results in this part can be found in [2, 5, 17, 19, 10, 13, 12, 11, 18]. For the polynomial case and Hilbert modules see also [32, 33, 23, 7, 4, 9, 24].

7 Hilbert modules and Hardy spaces

7.1 Inner products and norms

We now consider matrix valued functions: $f : \mathbb{C} \to \mathbb{C}^{N\times N}$. The space of these functions is a left and a right module over $\mathbb{C}^{N\times N}$. (The product is noncommutative.) We write $A \geq 0$ to mean that A is nonnegative definite and $A > 0$ means that A is positive definite, while $A \geq B$ is the same as $A - B \geq 0$. By A^H we mean the Hermitian conjugate and $\mathrm{tr}\,(A)$ is the normalized trace of A, i.e., $\mathrm{tr}\,(A) = \sum_{k=1}^N a_{kk}$. By $S = A^{1/2}$ we denote the Hermitian square root of $A \geq 0$, thus $A = S^2$ while $L = A^{L/2}$, resp. $R = A^{R/2}$ denote a left resp. right square root, meaning $A = LL^H$ resp. $A = R^H R$. Note that L is unique up to a right unitary factor and R up to a left unitary factor.

We introduce a matrix valued and a scalar valued inner product which will define $\mathbb{L}_2$ of square integrable functions that will be both a Hilbert module and a Hilbert space. Let μ be a nonnegative Hermitian measure on $\mathbb{T}$. This means that it is a square matrix whose entries are complex valued measures such that the whole matrix takes values that are nonegative definite. Because $0 \leq M \leq \mathrm{tr}\,(M)I$ holds for any nonnegative definite matrix, it holds for the scalar trace-measure $\tau = \mathrm{tr}\,(\mu)$ that $\tau(E) = 0$ implies $\mu(E) = 0$. This means that μ is absolutely continuous with respect to τ and we may define the trace-derivative μ'_τ as the matrix whose entries are $d\mu_{i,j}/d\tau$ and it holds that $\mu(E) = \int_E \mu'_\tau d\tau$. Furthermore it can be shown that $0 \leq \mu'_\tau \leq I$ a.e. τ and also $(\mu'_\tau)^{1/2}$ is measurable. More generally, we define $\int f d\mu g^H$ as $\int f \mu'_\tau g^H d\tau$, thus as a matrix whose entries are $\int [f\mu'_\tau g^H]_{ij} d\tau$. We denote this matrix valued "inner product" as $\langle\langle f, g\rangle\rangle_{\ell,\mu}$ where the ℓ stands for "left" since it is possible to define in a completely analogous way $\langle\langle f, g\rangle\rangle_{r,\mu} = \int f^H d\mu g$. In the sequel we treat only the left version and leave the right version to the reader. For many results however, left and right elements will be interacting. We shall always assume that $\int d\mu = I$.

The class of matrix valued functions for which $\langle\langle f, f\rangle\rangle_{\ell,\mu}$ exists forms a left module over $\mathbb{C}^{N\times N}$ that has some pre-Hilbert space-like properties. Indeed, for any constant square matrix a it holds that

- $\langle\langle f + g, h\rangle\rangle_{\ell,\mu} = \langle\langle f, h\rangle\rangle_{\ell,\mu} + \langle\langle g, h\rangle\rangle_{\ell,\mu}$ and $\langle\langle af, g\rangle\rangle_{\ell,\mu} = a\,\langle\langle f, g\rangle\rangle_{\ell,\mu}$ (linearity)
- $\langle\langle f, g\rangle\rangle_{\ell,\mu} = [\langle\langle g, f\rangle\rangle_{\ell,\mu}]^H$ (symmetry)
- $\langle\langle f, f\rangle\rangle_{\ell,\mu} \geq 0$ with tr $\langle\langle f, f\rangle\rangle_{\ell,\mu} = 0 \Leftrightarrow f = 0$ (positive definite)

To equip this class of functions with some topology for which it has to be complete, we also need a genuine scalar valued norm. This norm will be implied by a scalar valued inner product, which we define as follows: $\langle f, g\rangle_{\ell,\mu} = \mathrm{tr}\,\langle\langle f, g\rangle\rangle_{\ell,\mu} = \int \mathrm{tr}\,(f\mu'_\tau g^H)d\tau$. The corresponding scalar norm is $|f|_{\ell,\mu} = [\langle f, f\rangle_{\ell,\mu}]^{1/2}$. So we can now thefine the *Hilbert space* $\mathbb{L}_2(\mu)$ of square integrable matrix valued functions with respect to the scalar valued $\langle\cdot,\cdot\rangle_{\ell,\mu}$, that is complete with respect to the norm $|f|_{\ell,\mu}$. It is also a *Hilbert module* with respect to the matrix valued $\langle\langle\cdot,\cdot\rangle\rangle_{\ell,\mu}$ as we explained above.

Occasionally we also use a Euclidean inner product, namely $(f,g)_E = \mathrm{tr}\,(fg^H) = \mathrm{tr}\,(g^H f)$ and a Euclidean norm $|f|_E = (f,f)_E^{1/2} = |f|_F/\sqrt{N}$ where $|f|_F$ is the Frobenius norm and N the size of the matrix.

If $\mu = \lambda I$, we drop μ from the notation. For example $\mathbb{L}_2$ means $\mathbb{L}_2(\lambda I)$ and $\langle\!\langle f,g\rangle\!\rangle_\ell$ means $\langle\!\langle f,g\rangle\!\rangle_{\ell,\lambda}$ etc. We can define matrix valued Hardy spaces $\mathbb{H}_p$ as the matrix valued functions f, analytic in $\mathbb{D}$ and such that $|f|_p = \sup_{0\le r<1}[\int |f(rt)|_E^p d\lambda(t)]^{1/p} < \infty$ if $0<p<\infty$ while for $p=\infty$, $|f|_\infty = \sup_{0\le r<1, t\in\mathbb{T}} |f(rt)|_E < \infty$. As in the scalar case, any $f\in\mathbb{H}_p$ has a nontangential limit to the circle that belongs to $\mathbb{L}_p$ (which has an obvious definition viz., $f\in\mathbb{L}_p \Leftrightarrow |f|_E \in L_p$).

7.2 Carathéodory function and spectral factor

Carathéodory functions are functions analytic in $\mathbb{D}$ with positive real part. Thus with $\mathrm{Re}\,\Omega(z) = \frac{1}{2}[\Omega(z)+\Omega(z)^H] \ge 0$. There is an essentially unique relationship between positive measures and Carathéodory functions, which we represent as $\Omega_\mu(z) = \int D(t,z)d\mu(t) + i\mathrm{Im}\,\Omega_\mu(0)$. We shall assume that $\Omega_\mu(0)$ is real so that according to $\int d\mu = I$ we get $\Omega_\mu(0)=I$. Just like in the scalar case we have $\mathrm{Re}\,\Omega_\mu(z) = \int P(t,z)d\mu(t)$ and this has a nontangential limit to the circle $\mathbb{T}$ which equals $\mu' = d\mu/d\lambda$ a.e. (λ) and its Fourier series is $\Omega_\mu(z) = c_0 + 2\sum_{k=1}^\infty c_k z^k$ where the coefficients are the moments $c_k = \int t^{-k}d\mu(t)$.

For the spectral factor, we need to assume that $\log\det\mu' \in L_1$ (Szegő's condition). If that condition holds then there exist left and right outer spectral factors σ^L and σ^R such that $\mu' = \sigma^L\sigma^L_* = \sigma^R_*\sigma^R$ on $\mathbb{T}$. The substar for a matrix function is defined as $f_*(z) = f(1/\overline{z})^H$. These σ^L and σ^R are outer in $\mathbb{H}_2$ in the sense that their determinant is scalar outer in $H_{2/N}$. They are analytic in $\mathbb{D}$ and their determinant does not vanish in $\mathbb{D}$, so that their inverses are also outer in $\mathbb{H}_2$. These σ^L, resp. σ^R are uniquely defined up to a right, resp. left unitary constant factor.

8 MORF and reproducing kernels

8.1 Orthogonal rational functions

Matrix valued orthogonal rational functions (MORF) can now be obtained in much the same way as in the scalar case. Given a sequence of $\alpha_k \in \mathbb{D}$, we keep the notation of Blaschke factors $\zeta_k(z)$ and Blaschke products $B_n(z)$ from the scalar case and define $\mathcal{L}_n$ as the linear span over $\mathbb{C}^{N\times N}$ of $\{B_k I\}_{k=0}^n$. The block Gram matrix consists of the blocks $\langle\!\langle B_k, B_l\rangle\!\rangle_{\ell,\mu} = \langle\!\langle B_k, B_l\rangle\!\rangle_{r,\mu}$ and it depends only on $\Omega_\mu(\alpha_k)$ and its derivatives in these points up to an order depending on the multiplicity of α_k in the sequence.

By a block Gram-Schmidt algorithm we can orthormalize the previous basis as follows: assuming $\int d\mu = I$, set $\phi_0^L = I$, and for $n=1,2,\ldots$ we set $\tilde{\phi}_n^L = B_n I - \sum_{k=0}^{n-1}\left\langle\!\left\langle B_n I, \phi_k^L\right\rangle\!\right\rangle_{\ell,\mu}\phi_k^L$ and $\phi_n^L = m_n^{-1}\tilde{\phi}_n^L$ where m_n is a left square root of $\langle\!\langle\tilde{\phi}_n^L,\tilde{\phi}_n^L\rangle\!\rangle_{\ell,\mu}$. Note that the left square root need not be invertible in general if $\tilde{\phi}_n^L$ is not zero. Then the algorithm will break down, and we say that μ is degenerate. It can be shown that if $\log\det\mu' \in L_1$, then this degeneracy will not occur and all the ORF can be constructed. The basis ϕ_n for the Hilbert module is orthonormal in the sense that $\left\langle\!\left\langle \phi_k^L, \phi_j^L\right\rangle\!\right\rangle_{\ell,\mu} = \delta_{k,j}I$. The ϕ_n are defined up to a constant unitary factor from the left. To fix this, we shall assume that the leading coefficient κ_n^L of ϕ_n^L with respect to the basis $\{B_k I\}$ is Hermitian positive definite (every invertible matrix A can be written as $A = U^L B^L$ with U^L unitary and B^L positive definite [6, p. 22]). The leading coefficient is $\kappa_n^L = [\phi_n^{L*}(\alpha_n)]^H$ where the superstar for $f \in \mathcal{L}_n$ is defined as in the scalar case: $f^* = B_n f_*$. Note that for $f, g \in \mathcal{L}_n$ we have $\langle\langle f, g\rangle\rangle_{\ell,\mu} = \langle\langle f^*, g^*\rangle\rangle_{r,\mu}$.

8.2 Reproducing kernels

Given a left orthonormal system ϕ_k^L for $\mathcal{L}_n$, we can define a left reproducing kernel $k_n^L(z,w) = \sum_{k=0}^n \phi_k^L(w)^H \phi_k^L(z)$ which reproduces because $\left\langle\left\langle f, k_n^L(\cdot, w)\right\rangle\right\rangle_{\ell,\mu} = f(w)$ for all $f \in \mathcal{L}_n$. The right reproducing kernel is given by $k_n^R(w,z) = \sum_{k=0}^n \phi_k^R(z)[\phi_k^R(w)]^H$ and for any $f \in \mathcal{L}_n$ we have $\left\langle\left\langle k_n^R(w,\cdot), f\right\rangle\right\rangle_{r,\mu} = f(w)$.

Exercise 8.1 Mimic the scalar proof to derive the following CD-relations:

$$k_n^L(z,w) = \frac{[\phi_{n+1}^{R*}(w)]^H \phi_{n+1}^{R*}(z) - [\phi_{n+1}^L(w)]^H \phi_{n+1}^L(z)}{1 - \zeta_{n+1}(z)\overline{\zeta_{n+1}(w)}}$$

and

$$k_n^L(z,w) = \frac{[\phi_n^{R*}(w)]^H \phi_n^{R*}(z) - \zeta_n(z)\overline{\zeta_n(w)}[\phi_n^L(w)]^H \phi_n^L(z)}{1 - \zeta_n(z)\overline{\zeta_n(w)}}.$$

Consequently $k_n^L(z,\alpha_n) = \kappa_n^R \phi_n^{R*}(z)$ and $k_n^L(\alpha_n,\alpha_n) = \kappa_n^R[\kappa_n^R]^H$.

The right versions are obtained by symmetry.

$$k_n^R(w,z) = \frac{\phi_{n+1}^{L*}(z)[\phi_{n+1}^{L*}(w)]^H - \phi_{n+1}^R(z)[\phi_{n+1}^R(w)]^H}{1 - \zeta_{n+1}(z)\overline{\zeta_{n+1}(w)}}$$

and

$$k_n^R(w,z) = \frac{\phi_n^{L*}(z)[\phi_n^{L*}(w)]^H - \zeta_n(z)\overline{\zeta_n(w)}\phi_n^R(z)[\phi_n^R(w)]^H}{1 - \zeta_n(z)\overline{\zeta_n(w)}}.$$

From this right variant we have $k_n^R(\alpha_n, z) = \phi_n^{L*}(z)\kappa_n^L$ and $k_n^R(\alpha_n,\alpha_n) = [\kappa_n^L]^H \kappa_n^L$. ◇

Theorem 8.1 *The reproducing kernels satisfy the following recurrence relations*

$$s_n(w)\begin{bmatrix} k_n^{R*}(w,z) \\ k_n^L(z,w)\end{bmatrix} = t_n(z,w)\begin{bmatrix} k_{n-1}^{R*}(w,z) \\ k_{n-1}^L(z,w)\end{bmatrix}$$

where

$$s_n(w) = \begin{bmatrix} I - [\rho_n^R(w)]^H \rho_n^L(w) & 0 \\ 0 & I - \rho_n^L(w)[\rho_n^R(w)]^H \end{bmatrix},$$

$$t_n(z,w) = \begin{bmatrix} I & [\rho_n^R(w)]^H \\ \rho_n^L(w) & I \end{bmatrix}\begin{bmatrix} \zeta_n(z) I & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} I & [\gamma_n^R(w)]^H \\ \gamma_n^L(w) & I \end{bmatrix},$$

with

$$\rho_n^L(w) = [\phi_n^L(w)]^H[\phi_n^{L*}(w)]^{-1}, \quad \rho_n^R(w) = [\phi_n^{R*}(w)]^{-1}[\phi_n^R(w)]^H,$$

and $[\gamma_n^R(w)\ \gamma_n^L(w)] = -\zeta_n(w)[\rho_n^R(w)\ \rho_n^L(w)]$.

Proof. The proof is along the same lines as in the scalar case. One needs the relation $k_n^L(z,w) = B_n(z)\overline{B_n(w)}k_n^R(1/\overline{z}, 1/\overline{w})$. We leave it as an exercise to the reader. □

Note that it follows from the CD relations that

$$\{I - \rho_n^L(w)[\rho_n^R(w)]^H\}[\phi_n^{R*}(w)]^H \phi_n^{R*}(w) = [\phi_n^{R*}(w)]^H \phi_n^{R*}(w) - [\phi_n^L(w)]^H \phi_n^L(w) > 0.$$

Thus $I - \rho_n^L(w)[\rho_n^R(w)]^H$ can not be singular for $w \in \mathbb{D}$. Note that this implies also that $[\phi_n^{R*}(w)]^H \phi_n^{R*}(w)$ is not singular and hence positive definite, thus $\det \phi_n^{R*}(w) \neq 0$ for all $w \in \mathbb{D}$ and thus it is invertible in $\mathbb{D}$. A similar observation holds for $I - \rho_n^L(w)[\rho_n^R(w)]^H$ and for $\phi_n^{L*}(w)$.

Exercise 8.2 From this recurrence derive that

$$\Big(I - \gamma_n^L(w)[\gamma_n^R(w)]^H\Big)k_{n-1}^L(w,w) = \Big(I - \rho_n^L(w)[\rho_n^R(w)]^H\Big)k_n^L(w,w)$$

and symmetrically

$$\Big(I - [\gamma_n^R(w)]^H \gamma_n^L(w)\Big)k_{n-1}^R(w,w) = \Big(I - [\rho_n^R(w)]^H \rho_n^L(w)\Big)k_n^R(w,w).$$

◇

We can bring a bit more symmetry into the recurrence for the kernels by considering normalized versions of the kernels. Suppose $L_n(w) = [k_n^L(w,w)]^{L/2}$ is a left square root and $R_n(w) = [k_n^R(w,w)]^{R/2}$ is a right square root, then we call $K_n^L(z,w) = [L_n(w)]^{-1}k_n(z,w)$ and $K_n^R(w,z) = k_n^R(w,z)[R_n(w)]^{-1}$ the normalized kernels.

Exercise 8.3 Use the CD relations to show that $\rho_n^L(w)k_{n-1}^R(w,w) = k_{n-1}^L(w,w)\rho_n^R(w)$. With the left square root $L_n(w)$ and the right square roots $R_n(w)$ that we just introduced, use this result to prove that $[L_{n-1}(w)]^{-1}\rho_n^L(w)[R_{n-1}(w)]^H = [L_{n-1}(w)]^H\rho_n^R(w)[R_{n-1}(w)]^{-1}$. Thus we succeeded in symmetrizing the ρ_n^R and ρ_n^L and call the result $\rho_n(w)$ the *symmetrized recursion parameter* from now on. The CD relations also imply that $\rho_n(w)$ is strictly contractive in $\mathbb{D}$, i.e., $\rho_n^H\rho_n < I$ and hence also $\rho_n\rho_n^H < I$. Similar observations can be made when ρ is replaced by γ. ◇

We need one more lemma before we can state our symmetrized recurrence relation for the normalized kernels.

Lemma 8.2 *Define $F(\rho) = I - \rho\rho^H$ and $G(\rho) = I - \rho^H\rho$, then it is always possible to choose the square roots $L_n(w)$ and $R_n(w)$ in the definition of $\rho_n(w)$ such that (recall that $A^{1/2}$ denotes the Hermitian square root and $A^{-1/2}$ is its inverse)*

$$[L_{n-1}(w)]^{-1}L_n(w) = F(\rho_n(w))^{-1/2}F(\gamma_n(w))^{1/2}$$

and

$$R_n(w)[R_{n-1}(w)]^{-1} = G(\rho_n(w))^{-1/2}G(\gamma_n(w))^{1/2}.$$

Proof. First note that $F(\gamma_n) = L_{n-1}^{-1}(I - |\zeta_n|^2\rho_n^L[\rho_n^R]^H)L_{n-1}$. Hence

$$L_{n-1}F(\gamma_n)L_{n-1}^H = (I - |\zeta_n|^2\rho_n^L[\rho_n^R]^H)k_{n-1}^L(w,w),$$

which by the previous exercise 8.2 equals $(I - \rho_n^L[\rho_n^R]^H)k_n^L(w,w)$ while $I - \rho_n^L[\rho_n^R]^H = L_{n-1}F(\rho_n)L_{n-1}^{-1}$. Thus we may conclude that $F(\gamma_n)L_{n-1}^H L_n^{-H} = F(\rho_n)L_{n-1}^{-1}L_n$. Multiply by the inverse of $F(\gamma_n)$ to get $[F(\rho_n)]^{-1}F(\gamma_n) = L_{n-1}^{-1}L_nL_n^HL_{n-1}^{-H}$. This is obviously positive definite, so we can take its Hermitian square root $Q_n = L_{n-1}^{-1}L_nU_n$ with U_n some unitary matrix. This U_n can be included in the left square root L_n. This proves the first relation. The second one follows by symmetry. □

We have now a simplified recursion for the normalized kernels which follows immediately from the previous results.

Theorem 8.3 *Define $F(\rho) = I - \rho\rho^H$ and $G(\rho) = I - \rho^H\rho$ and the Halmos extension*

$$H(\rho) = \begin{bmatrix} G(\rho)^{-1/2} & 0 \\ 0 & F(\rho)^{-1/2} \end{bmatrix} \begin{bmatrix} I & \rho^H \\ \rho & I \end{bmatrix} = \begin{bmatrix} I & \rho^H \\ \rho & I \end{bmatrix} \begin{bmatrix} G(\rho)^{-1/2} & 0 \\ 0 & F(\rho)^{-1/2} \end{bmatrix},$$

the following recurrence holds

$$\begin{bmatrix} K_n^{R*}(w,z) \\ K_n^L(z,w) \end{bmatrix} = \theta_n(z,w) \begin{bmatrix} K_{n-1}^{R*}(w,z) \\ K_{n-1}^L(z,w) \end{bmatrix},$$

with

$$\theta_n(z,w) = H(\rho_n(w))Z_n(z)H(\gamma_n(w)), \quad Z_n(z) = \begin{bmatrix} \zeta_n(z)I & 0 \\ 0 & I \end{bmatrix}.$$

where $\rho_n(w)$ is as defined in the previous lemma and $\gamma_n(w) = -\zeta_n(w)\rho_n(w)$.

We can, like in the scalar case, define associated kernels $L_n^L(z,w)$ and $L^R(w,z)$, which start from the initial conditions $L_0^{R*} = I = L_0^L$ and then obey the recursion

$$\begin{bmatrix} L_n^{R*}(w,z) \\ -L_n^L(z,w) \end{bmatrix} = \theta_n(z,w) \begin{bmatrix} L_{n-1}^{R*}(w,z) \\ -L_{n-1}^L(z,w) \end{bmatrix},$$

with $\theta_n(z,w)$ like in the previous theorem.

The matrices θ_n have some special properties. For example, it is clear that they are para-J-unitary, i.e., they satisfy

$$\theta_n^* J \theta_n = \theta_n J \theta_n^* = J, \quad J = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}.$$

This immediately implies

Corollary 8.4 *For the normalized kernels K_n^L, K_n^R, L_n^L and L_n^R, the following relations hold*

- $K_n^L(z,w)L_n^{L*}(z,w) + L_n^L(z,w)K_n^{L*}(z,w) = 2B_n(z)I,$
- $K_n^{R*}(w,z)L_n^R(w,z) + L_n^{R*}(w,z)K_n^R(w,z) = 2B_n(z)I,$
- $K_n^{L*}(z,w)K_n^L(z,w) = K_n^R(w,z)K_n^{R*}(w,z),$
- $L_n^{L*}(z,w)L_n^L(z,w) = L_n^R(w,z)L_n^{R*}(w,z).$

Proof. This follows from the fact that

$$\begin{bmatrix} K_n^{R*} & L_n^{R*} \\ K_n^L & -L_n^L \end{bmatrix} = \Theta_n \begin{bmatrix} I & I \\ I & -I \end{bmatrix} \Leftrightarrow \Theta_n = \frac{1}{2} \begin{bmatrix} K_n^{R*} + L_n^{R*} & K_n^{R*} - L_n^{R*} \\ K_n^L - L_n^L & K_n^L + L_n^L \end{bmatrix}$$

where $\Theta_n = \theta_n \theta_{n-1} \cdots \theta_1$. Because all the θ_k are *-J-unitary, the same holds for the product Θ_n. Writing explicitly $\Theta_n J \Theta_n^* = J$ gives the previous results. □

The matrices θ_n, and therefore also the product Θ_n, are also *J-lossless*, which means that as a function of z, its entries are functions in the Nevanlinna class (ratios of H_∞ functions), it is J-contractive in $\mathbb{D}$ and J-unitary on $\mathbb{T}$, thus $\theta_n J \theta_n^H \le J$ for $z \in \mathbb{D}$ while $\theta_n J \theta_n^H = J$ for $z \in \mathbb{T}$.

Among other things, it is shown in [5] that this induces the following properties.

- $K_n^L(z,w)$, $L_n^L(w,z)$, $K_n^R(w,z)$, and $L_n^R(w,z)$ are invertible for $w \in \mathbb{D}$ and $z \in \mathbb{D} \cup \mathbb{T}$. Hence also ϕ_n^{R*} and ϕ_n^{L*} are invertible in $\mathbb{D} \cup \mathbb{T}$.
- $[K_n^L(z,w)]^{-1} L_n^L(z,w)$ and $L_n^R(w,z)[K_n^R(w,z)]^{-1}$ are Carathéodory functions for $w \in \mathbb{D}$.
- $[K_n^L(z,w)]^{-1}$ and $[K_n^R(w,z)]^{-1}$ are in $\mathbb{H}_2$.

9 Recurrence for the MORF

9.1 The recursion

The proof of the recurrence relation for the MORF is given in different steps which we formulate as lemmas. Note that in these lemmas the ϕ_n^L and ϕ_n^R are temporarily just orthogonal functions. They are not necessarily normalized in the particular way we agreed upon before. In the ultimate theorem 9.4 we will be back to the usual situation.

Lemma 9.1 *Starting with arbitrary invertible constant matrices ϕ_0^L and ϕ_0^R from $\mathbb{C}^{N \times N}$, define for $k = 1, \ldots, n$*

$$f_k^L(z) = \frac{z - \alpha_{k-1}}{1 - \overline{\alpha}_k z} \phi_{k-1}^L(z), \quad f_k^R(z) = \frac{z - \alpha_{k-1}}{1 - \overline{\alpha}_k z} \phi_{k-1}^R(z),$$

$$D_k^L = -C_k^L \left\langle\!\left\langle f_k^{R*}, \phi_0^L \right\rangle\!\right\rangle_{\ell,\mu} \left\langle\!\left\langle f_k^L, \phi_0^L \right\rangle\!\right\rangle_{\ell,\mu}^{-1}, \quad D_k^R = -\left\langle\!\left\langle \phi_0^R, f_k^R \right\rangle\!\right\rangle_{r,\mu}^{-1} \left\langle\!\left\langle \phi_0^R, f_k^{L*} \right\rangle\!\right\rangle_{r,\mu} C_k^R,$$

$$\phi_k^L(z) = D_k^L f_k^L(z) + C_k^L f_k^{R*}(z), \quad \phi_k^R(z) = f_k^R(z) D_k^R + f_k^{L*}(z) C_k^R,$$

then $\{\phi_k^L\}_{k=0}^n$ and $\{\phi_k^R\}_{k=0}^n$ form a left, resp. right orthogonal basis for $\mathcal{L}_n$. Moreover

$$D_k^L = -C_k^L \left\langle\!\left\langle f_k^{R*}, \phi_p^L \right\rangle\!\right\rangle_{\ell,\mu} \left\langle\!\left\langle f_k^L, \phi_p^L \right\rangle\!\right\rangle_{\ell,\mu}^{-1}, \quad D_k^R = -\left\langle\!\left\langle \phi_p^R, f_k^R \right\rangle\!\right\rangle_{r,\mu}^{-1} \left\langle\!\left\langle \phi_p^R, f_k^{L*} \right\rangle\!\right\rangle_{r,\mu}$$

for any $p = 0, \ldots, k-1$ and

$$C_k^L = \frac{1 - \overline{\alpha}_k \alpha_{k-1}}{1 - |\alpha_{k-1}|^2} \phi_k^L(\alpha_{k-1})[\phi_{k-1}^{R*}(\alpha_{k-1})]^{-1}, \quad C_k^R = \frac{1 - \overline{\alpha}_k \alpha_{k-1}}{1 - |\alpha_{k-1}|^2} [\phi_{k-1}^{L*}(\alpha_{k-1})]^{-1} \phi_k^R(\alpha_{k-1}),$$

$$D_k^L = -z_k \frac{1 - \overline{\alpha}_{k-1} \alpha_k}{1 - |\alpha_{k-1}|^2} \Big[[\phi_{k-1}^{L*}(\alpha_{k-1})]^{-1} \phi_k^{L*}(\alpha_{k-1}) \Big]^H$$

and

$$D_k^R = -z_k \frac{1 - \overline{\alpha}_{k-1} \alpha_k}{1 - |\alpha_{k-1}|^2} \Big[\phi_k^{R*}(\alpha_{k-1}) [\phi_{k-1}^{R*}(\alpha_{k-1})]^{-1} \Big]^H.$$

Proof. This is by induction on n. The matrix D_1^L is chosen in such a way that ϕ_1^L is left orthogonal to ϕ_0^L and similarly for ϕ_1^R. Suppose now that the theorem holds up to $n-1$. Then it is clear that $\phi_n^L \in \mathcal{L}_n \setminus \mathcal{L}_{n-1}$. Using $\langle\langle f, g\rangle\rangle_{\ell,\mu} = \langle\langle f^*, g^*\rangle\rangle_{r,\mu}$, it is not difficult to show that it is left orthogonal to $\frac{z-\alpha_n}{1-\overline{\alpha}_{n-1}z}\mathcal{L}_{n-2} \subset \mathcal{L}_{n-1}$, while D_n^L is chosen such that ϕ_n^L is left orthogonal to ϕ_0^L. Because for any $f \in \mathcal{L}_{n-1}$ clearly $f(z) - f(\alpha_n) \in \frac{z-\alpha_n}{1-\overline{\alpha}_{n-1}z}\mathcal{L}_{n-2}$, the left orthogonality to $\mathcal{L}_{n-1}$ follows.
The relation between C's and D's follows because $\left\langle\left\langle \phi_k^L, \phi_p^L \right\rangle\right\rangle_{\ell,\mu} = \left\langle\left\langle \phi_p^R, \phi_k^R \right\rangle\right\rangle_{r,\mu} = 0$ for all $p = 0, \ldots, k-1$. □

Lemma 9.2 *The following recursion also gives left and right MORF (we use the notation of the previous lemma)*

$$\begin{bmatrix} \phi_n^L(z) \\ \phi_n^{R*}(z) \end{bmatrix} = \frac{1-\overline{\alpha}_{n-1}z}{1-\overline{\alpha}_n z} \begin{bmatrix} I & [\lambda_n^L]^H \\ \lambda_n^R & I \end{bmatrix} \begin{bmatrix} \zeta_{n-1}(z)I & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} \phi_{n-1}^L(z) \\ \phi_{n-1}^{R*}(z) \end{bmatrix},$$

with

$$\lambda_n^L = -\overline{z}_{n-1} \left\langle\left\langle \phi_{n-1}^L, \phi_n^{R*} \right\rangle\right\rangle_{\ell,\mu}^{-1} \left\langle\left\langle \phi_{n-1}^L, \phi_n^L \right\rangle\right\rangle_{\ell,\mu}, \quad \lambda_n^R = -\overline{z}_{n-1} \left\langle\left\langle \phi_n^R, \phi_{n-1}^R \right\rangle\right\rangle_{r,\mu} \left\langle\left\langle \phi_n^{L*}, \phi_{n-1}^R \right\rangle\right\rangle_{r,\mu}^{-1}.$$

Proof. Setting

$$\lambda_n^L = -\overline{z}_{n-1}[[D_n^L]^{-1}C_n^L]^H, \quad \lambda_n^R = -\overline{z}_{n-1}[C_n^R[D_n^R]^{-1}]^H,$$

it readily follows from the previous relations that the right hand side of this theorem results in

$$\begin{bmatrix} -\overline{z}_{n-1}[D_n^L]^{-1}\phi_n^L(z) \\ -\overline{z}_{n-1}[D_n^R]^{-1}\phi_n^{R*}(z) \end{bmatrix}.$$

and because the ϕ_n^L are left orthogonal, so are the $-\overline{z}_{n-1}[D_n^L]^{-1}\phi_n^L(z)$ and a similar observation holds for the right versions. □

Lemma 9.3 *If we work with orthonormal MORF, then*

$$\lambda_n^L = \lambda_n^R = \overline{z}_{n-1}z_n \frac{1-\overline{\alpha}_{n-1}\alpha_n}{1-\overline{\alpha}_n\alpha_{n-1}}\rho_n(\alpha_{n-1})$$

with the $\rho_n(w)$ the symmetrized recursion parameter.

Proof. Recall $\kappa_n^R = \phi_n^{R*}(\alpha_n)$ and $\kappa_n^L = \phi_n^{L*}(\alpha_n)$ and the expressions for the λ's, the C's and the D's to find that

$$\lambda_n^L = \overline{z}_{n-1}z_n \frac{1-\overline{\alpha}_{n-1}\alpha_n}{1-\overline{\alpha}_n\alpha_{n-1}}[\kappa_{n-1}^R]^{-1}[\phi_n^L(\alpha_{n-1})]^H[\phi_n^{L*}(\alpha_{n-1})]^{-1}[\kappa_{n-1}^L]^H$$

and

$$\lambda_n^R = \overline{z}_{n-1}z_n \frac{1-\overline{\alpha}_{n-1}\alpha_n}{1-\overline{\alpha}_n\alpha_{n-1}}[\kappa_{n-1}^R]^H[\phi_n^{R*}(\alpha_{n-1})]^{-1}[\phi_n^R(\alpha_{n-1})]^H[\kappa_{n-1}^L]^{-1}.$$

We may now define matrices

$$N_n^L = \left\langle\left\langle \phi_n^L, \phi_n^L \right\rangle\right\rangle_{\ell,\mu}^{L/2}, \quad N_n^R = \left\langle\left\langle \phi_n^R, \phi_n^R \right\rangle\right\rangle_{r,\mu}^{R/2},$$

and multiply ϕ_n^L from the left by $[N_n^L]^{-1}$ and multiply ϕ_n^R from the right by $[N_n^R]^{-1}$, then we obtain orthonormal bases which we shall again denote by ϕ_n^L and ϕ_n^R. Note that the left and right square roots can be chosen such that the leading coefficients are positive definite. Furthermore, from the expression for λ_n^L and for ρ_n^L, we see that

$$\lambda_n^L = \overline{z}_{n-1} z_n \frac{1-\overline{\alpha}_{n-1}\alpha_n}{1-\overline{\alpha}_n \alpha_{n-1}} [\kappa_{n-1}^R]^{-1} \rho_n^L(\alpha_{n-1}) [\kappa_{n-1}^L]^H.$$

Finally use $\rho_n(w) = [K_{n-1}^L(w,w)]^{L/2} \rho_n^L(w) [[K_{n-1}^R(w,w)]^{R/2}]^H$ and $K_n^L(\alpha_n,\alpha_n) = \kappa_n^R[\kappa_n^R]^H$ and $K_n^R(\alpha_n,\alpha_n) = [\kappa_n^L]^H \kappa_n^L$ to find the expression for λ_n. □

Theorem 9.4 *The MORF have initial conditions $\phi_0^L = \phi_0^R = I$ and satisfy the recurrence*

$$\begin{bmatrix} \phi_n^L(z) \\ \phi_n^{R*}(z) \end{bmatrix} = t_n(z) \begin{bmatrix} \phi_{n-1}^L(z) \\ \phi_{n-1}^{R*}(z) \end{bmatrix},$$

with

$$t_n(z) = \frac{\sqrt{1-|\alpha_n|^2}}{\sqrt{1-|\alpha_{n-1}|^2}} \frac{1-\overline{\alpha}_{n-1} z}{1-\overline{\alpha}_n z} H(\lambda_n) Z_{n-1}(z)$$

where $H(\lambda_n)$ and $Z_n(z)$ are as defined before and

$$\lambda_n = \eta_n \rho_n(\alpha_{n-1}), \quad \eta_n = z_n \overline{z}_{n-1} \frac{1-\overline{\alpha}_{n-1}\alpha_n}{1-\overline{\alpha}_n \alpha_{n-1}} \in \mathbb{T}$$

with $\rho_n(w)$ the symmetrized recurrence parameter,

Proof. First note that

$$\begin{bmatrix} \hat{\phi}_n^L(z) \\ \hat{\phi}_n^{R*}(z) \end{bmatrix} = \frac{1-\overline{\alpha}_{n-1} z}{1-\overline{\alpha}_n z} \begin{bmatrix} I & \lambda_n^H \\ \lambda_n & I \end{bmatrix} Z_{n-1}(z) \begin{bmatrix} \phi_{n-1}^L(z) \\ \phi_{n-1}^{R*}(z) \end{bmatrix}$$

gives orthogonal functions, so that it remains to define the normalizing factors

$$N_n = \left[\left\langle\!\left\langle \hat{\phi}_n^L, \hat{\phi}_n^L \right\rangle\!\right\rangle_{\ell,\mu}^{L/2} \right]^{-1} \quad \text{and} \quad M_n = \left[\left\langle\!\left\langle \hat{\phi}_n^{R*}, \hat{\phi}_n^{R*} \right\rangle\!\right\rangle_{\ell,\mu}^{L/2} \right]^{-1}.$$

From the CD relation it follows that

$$\begin{bmatrix} \phi_n^L(\alpha_{n-1}) \\ \phi_n^{R*}(\alpha_{n-1}) \end{bmatrix}^H J \begin{bmatrix} \phi_n^L(\alpha_{n-1}) \\ \phi_n^{R*}(\alpha_{n-1}) \end{bmatrix} = -(1-|\zeta_n(\alpha_{n-1})|^2) \kappa_{n-1}^R [\kappa_{n-1}^R]^H.$$

On the other hand, the recursion gives

$$\begin{bmatrix} \phi_n^L(\alpha_{n-1}) \\ \phi_n^{R*}(\alpha_{n-1}) \end{bmatrix} = \frac{1-\overline{\alpha}_{n-1} z}{1-\overline{\alpha}_n z} \begin{bmatrix} N_n & 0 \\ 0 & M_n \end{bmatrix} \begin{bmatrix} \lambda_n^H [\kappa_{n-1}^R]^H \\ [\kappa_{n-1}^R]^H \end{bmatrix}.$$

Plug this into the previous relation and one gets (we set $\gamma_j = \sqrt{1-|\alpha_j|^2}$)

$$\lambda_n N_n^H N_n \lambda_n^H - M_n^H M_n = -\frac{\gamma_n^2}{\gamma_{n-1}^2} I.$$

For arbitrary unitary matrices U_n and V_n,

$$N_n = \frac{\gamma_n}{\gamma_{n-1}} U_n (I - \lambda_n^H \lambda_n)^{-1/2}, \quad M_n = \frac{\gamma_n}{\gamma_{n-1}} V_n (I - \lambda_n \lambda_n^H)^{-1/2}$$

are solutions. The U_n and V_n can be used to normalize the leading coefficients of ϕ_n^L and ϕ_n^R. □

Again it can be noted that for all $\alpha_k = 0$, the matrix versions of Szegő polynomials appear as a special case. The λ_n which are in this case given by $\rho_n(0)$ and are the block Szegő parameters.

9.2 Functions of the second kind

We can define functions of the second kind by setting $\phi_0^L = \psi_0^R = I$ as initial conditions for the recursion

$$\begin{bmatrix} \psi_n^L(z) \\ -\psi_n^{R*}(z) \end{bmatrix} = t_n(z) \begin{bmatrix} \psi_{n-1}^L(z) \\ -\psi_{n-1}^{R*}(z) \end{bmatrix},$$

with t_n as in the previous theorem. Along the same lines as in the scalar case, it can be proved that for $n \geq 1$

$$\psi_n^L(z) = \int D(t,z)[\phi_n^L(t) - \phi_n^L(z)] d\mu(t), \quad \psi_n^R(z) = \int D(t,z) d\mu(t) [\phi_n^R(t) - \phi_n^R(z)].$$

Because

$$T_n(z) = t_n(z) \cdots t_1(z) = \frac{1}{2} \begin{bmatrix} \phi_n^L(z) + \psi_n^L(z) & \phi_n^L(z) - \psi_n^L(z) \\ \phi_n^{R*}(z) - \psi_n^{R*}(z) & \phi_n^{R*}(z) + \psi_n^{R*}(z) \end{bmatrix}$$

and $(1 - \overline{\alpha}_n z) T_n(z)/\gamma_n$ is again a lossless and *-J-unitary matrix, we may conclude as in the case of the kernels that

- ϕ_n^{R*} and ψ_n^{R*} have no zeros in $\mathbb{D} \cup \mathbb{T}$ (a zero of a matrix means a zero of its determinant).
- $[\phi_n^{R*}]^{-1} \psi_n^{R*}$ and $\psi_n^{R*} [\phi_n^{R*}]^{-1}$ are Carathéodory functions.
- $[\phi_n^{L*}]^{-1}$ and $\psi_n^{R*}]^{-1}$ are in $\mathbb{H}_2$.
- $\phi_{n*}^R(t) \psi_n^R(t) + \psi_{n*}^R(t) \phi_n^R = 2P(t, \alpha_n) I$ for $t \in \mathbb{T}$
- $\phi_n^L(t) \psi_{n*}^L(t) + \psi_n^L(t) \phi_{n*}^L = 2P(t, \alpha_n) I$ for $t \in \mathbb{T}$
- $\phi_n^R \phi_n^{R*} = \phi_n^{L*} \phi_n^L$ and $\psi_n^R \psi_n^{R*} = \psi_n^{L*} \psi_n^L$

10 Interpolation and quadrature

10.1 The kernels

We can obtain interpolation properties for the (normalized) kernels like we did in the scalar case. This is related to polynomial kernels for varying measures in the sense that if $k_n^L(z,w)$ is a left reproducing kernel for the left inner product w.r.t. μ, then it can be written as $k_n^L(z,w) = k_{n,n}^L(z,w)/|\pi_n(z)|^2$ where $k_{n,n}^L(z,w)$ is the left polynomial reproducing kernel for the left inner product with respect to the measure $\mu/|\pi_n|^2$. We have

Theorem 10.1 *If $K_n^L(z,w)$ is the left normalized reproducing kernel for $\mathcal{L}_n$ with respect to μ, then for $d\mu_n^L(t) = [K_{n*}^L(t,0)K_n^L(t,0)]^{-1}d\lambda(t)$, we $\langle\langle f,g\rangle\rangle_{\ell,\mu} = \langle\langle f,g\rangle\rangle_{\ell,\mu_n^L}$ for all f and g in $\mathcal{L}_n$.*

Proof. It suffices to prove that for all $f \in \mathcal{L}_n$: $\langle\langle f, k_n^L(\cdot,w)\rangle\rangle_{\ell,\mu} = \langle\langle f, k_n^L(\cdot,w)\rangle\rangle_{\ell,\mu_n^L}$ because for a set of $n+1$ mutually different $\xi_k \in \mathbb{D}$, the $k_n^L(z,\xi_k)$ from a basis for $\mathcal{L}_n$. Therefore, it is sufficient to prove that for any polynomial p we have $\langle\langle p/\pi_n, k_n^L(\cdot,w)\rangle\rangle_{\ell,\mu_n^L} = p(w)/\pi_n(w)$. If $\phi_{k,n}^R$ is a set of right MOPVM for $\nu = \mu/|\pi_n|^2$, it follows that by using the polynomial CD relation that $K_n^L(z,0) = \phi_{nn}^{R*}(z)/\pi_n(z)$, and thus, $d\mu_n^L = |\pi_n|^2[\phi_{nn}^{R*}]_*\phi_{nn}^{R*}$. Therefore $\langle\langle p/\pi_n, k_n^L(\cdot,w)\rangle\rangle_{\ell,\mu_n^L} = I_1 - I_2$ with

$$I_1 = \frac{1}{\pi_n(w)}\left[\int p(t)[(\phi_{n,n}^{R*}(t))_*\phi_{n,n}^{R*}(t)]^{-1}(\phi_{n,n}^{R*}(t))_*\frac{t}{t-w}d\lambda(t)\right]\phi_{n,n}^{R*}(w) = \frac{p(w)}{\pi_n(w)}.$$

Because the integral is by Cauchy's theorem equal to $p(w)[\phi_{n,n}^{R*}(w)]^{-1}$.

$$I_2 = \frac{1}{\pi_n(w)}\left[\int p(t)[(\phi_{n,n}^{R*}(t))_*\phi_{n,n}^{R*}(t)]^{-1}(\phi_{n,n}^{L}(t))_*\frac{1}{t-w}d\lambda(t)\right]\phi_{n,n}^{L}(w).$$

This is zero because $(\phi_{n,n}^{R*})_*\phi_{n,n}^{R*} = \phi_{n,n}^{L*}(\phi_{n,n}^{L*})_*$ which reduces the integral to the Hermitian conjugate of $\int[\phi_{n,n}^{L*}(t)]^{-1}p^*(t)\frac{t}{1-\overline{w}t}d\lambda(t)$ and this is zero by Cauchy's theorem. This concludes the proof. □

Theorem 10.2 *With $\Omega_\mu(z) = \int D(t,z)d\mu(t)$ and $\Omega_n^L(z) = \int D(t,z)d\mu_n^L(t)$ we have $\Omega_n^L(z) = [K_n^L(z,0)]^{-1}L_n^L(z,0)$ and the left outer spectral factor of μ_n^L is $\sigma_n^L(z) = [K_n^L(z,0)]^{-1}$. Moreover we have the interpolation property*

$$\Omega_\mu(z) - \Omega_n^L(z) = zB_n(z)h(z)$$

with h analytic in $\mathbb{D}$.

Proof. We know that $[K_n^L(z,0)]^{-1}L_n^L(z,0)$ is a Carathéodory function and because $\frac{1}{2}[\Omega_n^L(t) + \Omega_{n*}^L(t)] = [K_{n*}^L(t,0)K_n^L(t,0)]^{-1}$, it is the Riesz-Herglotz transform of μ_n^L. Also, it is obvious that σ_n^L is a spectral factor. It is outer in $\mathbb{H}_2$ because $K_n^L(z,0)$ is a rational function in $\mathbb{H}_2$ and it does not vanish in $\mathbb{D}\cup\mathbb{T}$. For the interpolation property, note that with $\nu = \mu - \mu_n^L$

$$\frac{\Omega_\mu(z) - \Omega_n^L(z)}{B_n(z)} = \int\left[\frac{1}{B_n(z)} - \frac{1}{B_n(t)}\right]D(t,z)d\nu(t) + \int\frac{1}{B_n(t)}D(t,z)d\nu(t).$$

The first integral is zero because the term in square brackets is of the form f_* with $f = p/\pi_n$ where p is a scalar polynomial in t of degree less than $n+1$. Thus it equals $\langle\langle I, fI\rangle\rangle_{\ell,\nu}$. Because $fI \in \mathcal{L}_n$, this integral is 0. The integrand for the second term is analytic in $\mathbb{D}$ because $\int f(t)D(t,z)d\mu(t)$ is analytic in $\mathbb{D}$ for any $f \in \mathbb{L}_1(\nu)$ and any measure ν. Because $\langle\langle I,I\rangle\rangle_{\ell,\mu} = \langle\langle I,I\rangle\rangle_{\ell,\mu_n^L}$ we get the interpolation at the origin. □

Corollary 10.3

$$\Omega_n^L(z) = [K_n^L(z,0)]^{-1}L_n^L(z,0) = L_n^R(0,z)[K_n^R(0,z)]^{-1} = \Omega_n^R(z).$$

Proof. Because $[K_{n*}^L(t,0)K_n^L(t,0)]^{-1} = [K_n^R(0,t)K_{n*}^R(0,t)]^{-1}$, we have by the Riesz-Herglotz theorem that $\Omega_n^L = \Omega_n^R$ in $\mathbb{D}$, but because it are just rational functions, equality also holds on $\mathbb{C}$. □

10.2 The MORF

Similar properties can be derived for the MORF. Recall that $P(t,z) = \operatorname{Re} D(t,z)$ is the Poisson kernel. We have

Theorem 10.4 *Consider the measure* $d\hat{\mu}_n^L(t) = P(t,\alpha_n)[\phi_{n*}^L(t)\phi_n^L(t)]^{-1}d\lambda(t)$*, then in* $\mathcal{L}_n$*, the inner products* $\langle\!\langle \cdot,\cdot\rangle\!\rangle_{\ell,\mu}$ *and* $\langle\!\langle \cdot,\cdot\rangle\!\rangle_{\ell,\hat{\mu}_n}$ *are the same.*

Proof. A proof will be given if it can be shown that the ϕ_n^L, which is a left orthonormal basis for μ, is also a left orthonormal basis for $\hat{\mu}_n^L$. This can be shown by a backward recursion. Obviously $\langle\!\langle \phi_n^L, \phi_n^L\rangle\!\rangle_{\ell,\hat{\mu}_n^L} = I$. Also for $k = 0,\ldots,n-1$:

$$\langle\!\langle \phi_n^L, \phi_k^L\rangle\!\rangle_{\ell,\hat{\mu}_n^L} = \int [\phi_n^{L*}(t)]^{-1}\phi_k^{L*}(t)P(t,\alpha_n)B_n(t)/B_k(t)d\lambda(t),$$

which is zero by the Poisson formula. □

Theorem 10.5 *With* $\Omega_\mu(z) = \int D(t,z)d\mu(t)$ *and* $\hat{\Omega}_n^L(z) = \int D(t,z)d\hat{\mu}_n^L(t)$ *where* $\hat{\mu}_n^L$ *is as in the previous theorem, then* $\hat{\Omega}_n^L(z) = \psi_n^{L*}(z)[\phi_n^{L*}(z)]^{-1}$*, and* $\hat{\sigma}_n^R(z) = \frac{\sqrt{1-|\alpha_k|^2}}{1-\overline{\alpha}_n z}[\phi_n^{L*}(z)]^{-1}$ *is a right outer spectral factor of* $\hat{\mu}_n^L$*. Furthermore we have the interpolation property*

$$\Omega_\mu(z) - \hat{\Omega}_n^L(z) = zB_n(z)h(z)$$

with h *analytic in* $\mathbb{D}$*.*

Proof. The proof is complete parallel to the proof of theorem 10.2. □

Corollary 10.6

$$\hat{\Omega}_n^L(z) = \psi_n^{L*}(z)[\phi_n^{L*}(z)]^{-1} = [\phi_n^{R*}(z)]^{-1}\psi_n^{R*}(z) = \hat{\Omega}_n^R(z).$$

With these results, it is not difficult to derive a Favard type theorem.

Theorem 10.7 *Let* ϕ_n^L *and* ϕ_n^R *be generated by the recursion of Theorem 9.4. Then there is a positive definite measure on* $\mathbb{T}$ *for which they form a left, resp. a right MORF sequence.*

Proof. This is along the same lines as the scalar proof of Theorem 2.7. □

This measure will be unique if $A \in BD$ and $\log\det\mu' \in L_1$ which follows from the completeness of the basis of Blaschke products in $\mathbb{H}_2$ with respect to the norm $|\cdot|_{\ell,\mu}$.

11 Minimalisation and Szegő's problem

If $\Phi(s)$ is a function with Hermitian nonnegative definite values, then the problems $\inf_s \Phi(s)$ and $\inf_s \operatorname{tr}\Phi(s)$ have the same solutions. By a solution we mean an s such that $\Phi(s) \leq \Phi(t)$ for all t. We have the following.

Theorem 11.1 *The minimum of* $\langle\!\langle f,f\rangle\!\rangle_{\ell,\mu}$ *for all* $f \in \mathcal{L}_n$ *with* $f(w) = I$ *is* $[k_n^L(w.w)]^{-1}$ *given by and it is obtained for* $f = [k_n^L(w,w)]^{-1}k_n^L(\cdot,w)$*, provided that* $\det k_n^L(w,w) \neq 0$*.*

Proof. Similar to the proof of Theorem 6.1. □

We have formulated this theorem for $\mathcal{L}_n$ but it actually holds for any Hilbert module with reproducing kernel.

Define the left Szegő kernel $s^L(z,w) = [(1-\overline{w}z)\sigma^L(z)(\sigma^L(w))^H]^{-1}$ with $\sigma^L(0) > 0$ with $w \in \mathbb{D}$, then it is left reproducing in $\mathbb{H}_2$. We can conclude that, as in the scalar case, some vector valued stochastic process will be completely predictable if $\exp\{\int \log\det \mu'(t) d\lambda(t)\} = 0$, i.e. $\log\det\mu' \notin L_1$. If we assume that the Blaschke products diverge ($A \in BD$) and that $\log\det\mu' \in L_1$ ($\mu \in SZ$), then the set $\{B_n\}_{n\geq 0}$ is complete in the Banach space $\mathbb{H}_2(\mu)$ (thus with respect to the trace norm $|\cdot|_{\ell,\mu}$). So the analog of Theorem 6.2 is

Theorem 11.2 *Let $\mu \in SZ$, $w \in \mathbb{D}$ a fixed number, and let $s^L_w(z) = s^L(z,w)$ be the left Szegő kernel. Consider the problem $\min \|f - s^L_w\|^2_{\ell,\mu}$ over all $f \in \mathcal{L}_n$, then the solution is $k^L_n(z,w)$ and the minimum is $s^L_w(w) - k^L_n(w,w)$. If $A \in BD$ then $\lim_{n\to\infty} |k^L_n(t,w) - s^L_w(t)|_{\ell,\mu} = 0$.*

Proof. Because

$$\langle\langle s^L_w - f, s^L_w - f\rangle\rangle_{\ell,\mu} = \langle\langle s^L_w, s^L_w\rangle\rangle_{\ell,\mu} + \langle\langle f, f\rangle\rangle_{\ell,\mu} - 2\mathrm{Re}\,\langle\langle f, s^L_w\rangle\rangle_{\ell,\mu}$$

By Cauchy's theorem $\langle\langle s^L_w, s^L_w\rangle\rangle_{\ell,\mu} = s^L_w(w)$ and $\langle\langle f, s^L_w\rangle\rangle_{\ell,\mu} = f(w)$, so that we have to minimize $s^L_w(w) + \langle\langle f, f\rangle\rangle_{\ell,\mu} - 2\mathrm{Re}\, f(w)$. Like in the scalar case it can be seen that this minimum is obtained for f as claimed. Because $A \in BD$ implies completeness in the Banach space $\mathbb{H}_2$, we have convergence in the norm of this space. □

This problem is immediately related to the prediction of vector valued stochastic processes, but it can also be interpreted in the context of inverse scattering theory or network synthesis.

What we did not discuss

There are many things that were not discussed in this introduction.

First of all there is the scalar case with many things to be explored like further convergence results, characterization of the properties of the ORF in terms of the recurrence coefficients, moment problems, Nevanlinna-Pick and other classical interpolation problems. Some of these are included to some extend in [3].

There is a completely similar theory of ORF on the real line when all the poles are outside the real line. Interesting special cases are measures whose support is an interval or a half line. In the latter case results in the vein of what has been developped here can be obtained [29].

A slightly different theory emerges when poles are allowed to fall inside the support of the measure (see e.g. [3]).

In all those cases there is practically no result about an operator theoretical approach that generalizes the Jacobi matrices of classical polynomials.

For the matrix case, there is a huge literature on so-called Schur analysis. This involves matricial Nevanlinna-Pick interpolation problems and tangential or directional versions. A lot of energy is also spent on discussing generalizations of the Blaschke factors. Indeed instead of just using $B_k I$ as a basis function, one could consider products of factors of the form $\zeta_j U_j$ where U_j represents a rank one matrix. The recursion, which is in our treatment connected with J-lossless matrices, can be considerably generalized by replacing J by much more general matrices with a finer structure. And all this could be extended to the non-square matrix case. Etc.

References

[1] N.I. Achieser. *Theory of approximation.* Frederick Ungar Publ. Co., New York, 1956.

[2] A. Bultheel. Orthogonal matrix functions related to the multivariable Nevanlinna-Pick problem. *Bull. Soc. Math. Belg. Sér. B*, 32(2):149–170, 1980.

[3] 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.

[4] Ph. Delsarte, Y. Genin, and Y. Kamp. Orthogonal polynomial matrices on the unit circle. *IEEE Trans. Circuits and Systems*, 25:149–160, 1978.

[5] P. Dewilde and H. Dym. Lossless chain scattering matrices and optimum linear prediction : the vector case. *Internat. J. Circuit Th. Appl.*, 9:135–175, 1981.

[6] R.G. Douglas. *Banach algebra techniques in operator theory.* Academic Press, New York, 1972.

[7] V.K. Dubovoj, B. Fritzsche, and B. Kirstein. *Matricial version of the classical Schur problem*, volume 129 of *Teubner-Texte zur Mathematik.* Teubner Verlagsgesellschaft, Stuttgart, Leipzig, 1992.

[8] P.L. Duren. *The theory of H^p spaces*, volume 38 of *Pure and Applied Mathematics.* Academic Press, New York, 1970.

[9] H. Dym. *J-contractive matrix functions, reproducing kernel Hilbert spaces and interpolation*, volume 71 of *CBMS Regional Conf. Ser. in Math.* Amer. Math. Soc., Providence, Rhode Island, 1989.

[10] B. Fritzsche and B. Kirstein amd A. Lasarow. Orthogonal rational matrix-valued functions on the unit circle. *Math. Nachr.*, 2002. To appear.

[11] B. Fritzsche and B. Kirstein amd A. Lasarow. Further aspects of the theory of orthogonal rational matrix-valued functions on the unit circle. *Math. Nachr.*, 2003. Submitted.

[12] B. Fritzsche and B. Kirstein amd A. Lasarow. On a moment problem for rational matrix-valued functions. *Linear Algebra Appl.*, 345:1–31, 2003.

[13] B. Fritzsche and B. Kirstein amd A. Lasarow. On rank invariance of moment matrices of nonnegative Hermitian-valued Borel measures on the unit circle. *Math. Nachr.*, 263:1–30, 2003.

[14] J. B. Garnett. *Bounded analytic functions.* Academic Press, New York, 1981.

[15] K. Hoffman. *Banach spaces of analytic functions.* Prentice-Hall, Englewood Cliffs, 1962.

[16] P. Koosis. *Introduction to H_p spaces*, volume 115 of *Cambridge Tracts in Mathematics.* Cambridge University Press, 2nd edition, 1998.

[17] P. Kravanja. Een Szegő theorie voor rationale matrixfuncties. Master thesis, K.U.Leuven, Dept. Computerwetenschappen, 1994.

[18] H. Langer and A. Lasarow. Solution of a multiple Nevanlinna-Pick problem via orthogonal rational functions. *J. Comput. Appl. Math.*, 2004. Submitted.

[19] A. Lasarow. *Aufbau einer Szegő-Theorie für rationale Matrixfunktionen.* PhD thesis, Universität Leipzig, Fak. Mathematik Informatik, 2000.

[20] G.L. Lopes [López-Lagomasino]. On the asymptotics of the ratio of orthogonal polynomials and convergence of multipoint Padé approximants. *Math. USSR-Sb.*, 56:207–219, 1985.

[21] G.L. López [López-Lagomasino]. Szegő's theorem for orthogonal polynomials with respect to varying measures. In M. Alfaro et al., editors, *Orthogonal polynomials and their applications*, volume 1329 of *Lecture Notes in Math.*, pages 255–260. Springer, 1988.

[22] G.L. López [López-Lagomasino]. Asymptotics of polynomials orthogonal with respect to varying measures. *Constr. Approx.*, 5:199–219, 1989.

[23] P. Masani. The prediction theory of multivariate stochastic processes, III. *Acta Math.*, 104:141–162, 1960.

[24] M. Rosenblum and J. Rovnyak. *Hardy classes and operator theory.* Oxford University Press, New York, 1985.

[25] W. Rudin. *Real and complex analysis.* McGraw-Hill, New York, 3rd edition, 1987.

[26] A. Ruhe. Rational Krylov sequence methods for eigenvalue computation. *Linear Algebra Appl.*, 58:391–405, 1984.

[27] A. Ruhe. Rational Krylov: A practical algorithm for large sparse nonsymmetric matrix pencils. *SIAM J. Sci. Comput.*, 19(5):1535–1551, 1998.

[28] A. Ruhe. The rational Krylov algorithm for nonlinear matrix eigenvalue problems. *J. Math. Sci.*, 114:176–180, 2003.

[29] J. Van Deun. *Orthogonal rational functions: asymptotics and computational aspects.* PhD thesis, K.U.Leuven, dept. computer science, 2004.

[30] J.L. Walsh. Interpolation and functions analytic interior to the unit circle. *Trans. Amer. Math. Soc.*, 34:523–556, 1932.

[31] J.L. Walsh. *Interpolation and approximation*, volume 20 of *Amer. Math. Soc. Colloq. Publ.* Amer. Math. Soc., Providence, Rhode Island, 3rd edition, 1960. First edition 1935.

[32] N. Wiener and P. Masani. The prediction theory of multivariate stochastic processes, I. The regularity condition. *Acta Math.*, 98:111–150, 1957.

[33] N. Wiener and P. Masani. The prediction theory of multivariate stochastic processes, II. The linear predictor. *Acta Math.*, 99:93–139, 1958.