Orthogonal rational functions on the unit circle: theory, applications and generalizations*

Adhemar Bultheel†, Pablo González-Vera‡, Erik Hendriksen§, Olav Njåstad¶

Presented at the Workshop on
Orthogonal polynomials on the unit circle: Theory and applications,
Madrid, June 27-30, 1994

Abstract

We give a survey of the basic theory of orthogonal rational functions with poles outside the unit disk. We give a motivation by linear prediction and signal processing applications. We include the basic theory of recurrence relations, Christoffel-Darboux relations, reproducing kernels, associated functions and interpolation properties, para-orthogonal functions and quadrature, moment problems and Favard theorems, and convergence properties.

We touch only briefly on the following related and generalized topics. We give an indication of related results for the case where the poles (interpolation points) are on the boundary (i.e. the circle). We briefly consider the case where the functions are matrix valued and relate it to tangential interpolation problems. Some results are mentioned for the case where the inner product is taken w.r.t. a Gramm matrix which is quasi Toeplitz and finally we mention related results about time-varying systems.

1 Motivation (linear prediction)

Let $s = (s_n) \in \ell_2$ be a discrete signal. Its Fourier transform $S(e^{i\omega}) \in L_2$ is called its *spectrum*, while we call $R(e^{i\omega}) = |S(e^{i\omega})|^2$ the *power spectrum*. Defining $S_*(z) = \overline{S(1/\overline{z})}$, we can write this as $R(t) = S(t)S_*(t)$ with $t \in \mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$. When $R(z) = \sum_k r_k z^k$, then we call the sequence $r = (r_k)$ the *autocorrelation function* of the signal s. Obviously $r_n = \sum_k s_{n+k}\overline{s}_k = \overline{r}_{-k}$, $n \in \mathbb{Z}$. The *energy* of the signal corresponds to $r_0 = \|s\|^2 = \|S\|^2$. In the linear prediction problem, we want to predict the value of s_n from its past, i.e., by the estimate $\hat{s}_n = -\sum_{k\geq 1} a_k s_{n-k}$. Assume that the signal is in a stationary regime, so that the a_k do not depend on n. The corresponding prediction error is $e_n = s_n - \hat{s}_n = \sum_{k\geq 0} a_k s_{n-k}$ or in the frequency domain: $E(z) = A(z)S(z)$, with $a_0 = A(0) = 1$. The best possible such a model in least squares sense is the one which minimizes $\|e\|^2 = \|E\|^2 = \|AS\|^2 = \|A\|_R^2$

*This research was performed as part of the European project ROLLS under contract CHRX-CT93-0416.
†Department of Computer Science, K.U. Leuven, Belgium
‡Facultad de Matemáticas, Universidad de La Lagune, Tenerife, Spain
§Department of Mathematics, University of Amsterdam, The Netherlands
¶Department of Mathematical Sciences, NTH-Trondheim, Norway

where $R = |S|^2$ is the power spectrum which is to be used as a (positive) weight. The optimization has to be solved under the constraint $a_0 = A(0) = 1$.

The mathematical framework is thus the following. Consider the finite positive measure $d\mu(\omega) = R(e^{i\omega})d\lambda(\omega)$ where $d\lambda(\omega) = \frac{d\omega}{2\pi}$ is the normalized Lebesgue measure. Then we are working in a weighted Hilbert space on the unit circle with inner product $\langle f, g\rangle = \int f(e^{i\omega})\overline{g(e^{i\omega})}d\mu(\omega)$ and we have to find $\min_{a_i} \|1 + a_1 z + a_2 z^2 + \cdots\|_\mu^2$, that is, we have to project 1 onto the span$\{z, z^2, \ldots\}$. This is the optimization problem considered by Szegő and this leads to the study of orthogonal polynomials on the unit circle. Note that this problem is also equivalent to a spectral factorization problem, which means that we are given $R(t) > 0$ on $\mathbb{T}$ and we have to find $S(z)$ such that $R(z) = S(z)S_*(z)$ which is $R = \overline{|S|^2}$ on $\mathbb{T}$. The solution in terms of the OPS $\{\varphi_n\}$ is the following: Define $\varphi_n^*(z) = z^n\overline{\varphi_n(1/\overline{z})}$, then $1/\varphi_n^*$ will converge to S as $n \to \infty$. If $G^2 = \|e\|^2$ is the ultimate prediction error, then it turns out that the leading coefficient κ_n of $\varphi_n(z) = \kappa_n z^n + \cdots$ will give G as $n \to \infty$. Therefore, the modelling filter will be $G/A(z)$, which means that the signal is approximated by the output of a system whose input is a pulse (i.e., 1 at time zero and zero everywhere else) and whose transfer function is given by the optimal prediction filter $G/A(z)$. The latter is obtained as the limit of κ_n/φ_n^*.

However this filter is autoregressive, meaning that it does not contain finite zeros. Thus if there are some zeros in the actual filter which are close to $\mathbb{T}$, then this will be manifestly visible in the shape of the power spectrum R and the previous model will have difficulties in approximating (predicting) such a filter since the zeros have to be simulated by a large number of poles. That is why autoregressive-moving average filters may be more appropriate. The latter are filters which do allow some zeros in the numerator. Letting the denominator as well as the denominator free in the optimization problem leads to a highly nonlinear and ill conditioned problem in many situations [7]. Therefore, even rough but fixed estimates of the zeros of the filter may lead to better results for finite n and yet the problem will be linear.

Thus instead of finding the solutions to the minimization problems

$$\min\{\|f\|_\mu^2 : f \in \mathcal{L}_n, f(0) = 1\} \tag{1.1}$$

where $\mathcal{L}_n$ is the set $\mathcal{P}_n$ of polynomials of degree at most n, we could replace this by any other finite dimensional subspace $\mathcal{L}_n$ of H_2. In view of the previous discussion, a space of the form

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

where the $\alpha_i \in \mathbb{D}$ are prefixed, seems to be appropriate. Solving the problem (1.1) leads to a prediction filter of the form $G_n/A_n = [G_n\pi_n(z)]/p_n(z)$ with zeros $1/\overline{\alpha}_i$. Note however that the solutions are *not* given by the orthogonal rational functions (ORF) $\{\varphi_k\}$ which form a basis for $\mathcal{L}_n$. The solution is given by a reproducing kernel $k_n(z, w) = \sum_{i=0}^{n} \varphi_k(z)\overline{\varphi_k(w)}$. This kernel reproduces, which means that $\langle f(t), k_n(t, w)\rangle = f(w)$ for all $f \in \mathcal{L}_n$. It is very easy to show that if $\mathcal{L}_n$ is any finite dimensional subspace with reproducing kernel k_n, then the solution of the problem $\min\{\|f\|_\mu^2 : f \in \mathcal{L}_n, f(w) = 1\}$ is given by $f(z) = k_n(z, w)/k_n(w, w)$ with minimum equal to $1/k_n(w, w)$.

Therefore the reproducing kernels play a more prominent role in prediction problems, but the ORF are never far away. They deserve a study for their own interest and for other applications like quadrature etc. This is what we shall concentrate on in the next sections. For more information on this section see [38, 35, 30, 28]

2 Function classes

We start here with the mathematical framework. We still use $\mathbb{T}$ to denote the unit circle and use $\mathbb{D}$ for the open unit disk. We normalize the measure μ on $\mathbb{T}$ to be a probability measure: $\int d\mu(t) = 1$. $L_2(\mu)$ is the Hilbert space with inner product

$$\langle f, g \rangle = \int f(t)\overline{g(t)}d\mu(t), \quad t = e^{i\omega}.$$

λ denotes the normalized Lebesgue measure: $d\lambda(t) = (2\pi)^{-1}d\omega$.

The class $\mathcal{C}$ is the class of Carathéodory or positive real functions which are functions analytic in $\mathbb{D}$ and whose real part is positive in $\mathbb{D}$. It is well known that they are the Riesz-Herglotz transform of a positive measure:

$$\mathcal{C} \ni \Omega(z) = ic + \int D(t,z)d\mu(t), \quad c \in \mathbb{R}, \quad D(t,z) = \frac{t+z}{t-z}. \tag{2.1}$$

Define the Poisson kernel as $P(z,w) = \frac{1}{2}[D(z,w)+D(z,w)]$ where the substar is with respect to z. For $z = t \in \mathbb{T}$, this leads to the familiar formula $P(t,w) = (1-|w|^2)/(|t-w|^2) = \Re D(t,w)$. Hence

$$\Re\Omega(z) = \int P(t,z)d\mu(t) > 0.$$

Consider the Lebesgue decomposition of μ and the Fourier-Stieltjes coefficients (trigonometric moments) $r_k = \int t^{-k}d\mu(t) = \overline{r}_{-k}$, then

$$d\mu(t) = \mu'(t)d\lambda(t) + d\mu_s(t), \quad \mu'(t) = \sum_{k=-\infty}^{\infty} r_k t^k.$$

Taking $\Omega(0) \in \mathbb{R}$, i.e., setting $c = 0$ in (2.1), we get the following relation with Ω

$$\Omega(z) = 1 + 2\sum_{k=1}^{\infty} r_k z^k, \qquad \mu'(z) = \frac{1}{2}[\Omega(z) + \Omega_*(z)]. \tag{2.2}$$

Non-tangential limits as z approaches the boundary $\mathbb{T}$ exist and corresponding equalities are valid. Note that we assumed a normalization $\Omega(0) = 1$, which matches the normalization of μ.

The class $\mathcal{B}$ is the class of Schur or bounded analytic functions. These are functions analytic in $\mathbb{D}$ with modulus bounded by 1. They map the unit disk into itself and they form a unit ball in an appropriate function space.

By the *Cayley transform*, a disk is mapped into a half plane and back, so that it is no surprise that class $\mathcal{C}$ and class $\mathcal{B}$ functions are mapped into each other by such a transform. For example if $\Omega \in \mathcal{C}$, then its Cayley transform $\Gamma(z) = [\Omega(z) - \Omega(0)]/[\Omega(z) + \Omega(0)]$ will be in $\mathcal{B}$. Moreover, it is mapped such that $\Gamma(0) = 0$. We could as well have normalized it by making it zero in any other point $w \in \mathbb{D}$.

A *Moebius transform* $\mathcal{M}_\alpha : w \mapsto \frac{w-\alpha}{1-\overline{\alpha}w}$ maps the unit disk into itself, hence also $\mathcal{B}$ into itself. This forms the basis of the Schur algorithm which relies on the property

$$f \in \mathcal{B} \Leftrightarrow \{\rho = f(0) \in \mathbb{D} \quad \text{and} \quad \mathcal{M}_\rho(f) \in \mathcal{B}\}. \tag{2.3}$$

The *Schur algorithm* starts with $f_0 = f$ and constructs the reflection coefficients $\rho_i = f_{i-1}(0)$, where $f_i = M_{\rho_i}(f_{i-1})$. Obviously $f \in \mathcal{B}$ iff all the reflection coefficients are in $\mathbb{D}$.

The *Nevanlinna-Pick algorithm* (NP) is a straightforward generalization of the Schur algorithm, in fact it is often called the Schur algorithm in the literature. The generalization consists in observing that there is no need to choose the origin and set $\rho = f(0)$ in (2.3), but we could as well choose any other $\alpha \in \mathbb{D}$ and take $\rho = f(\alpha)$. Thus, given a sequence $(\alpha_i) \subset \mathbb{D}$, the NP algorithm constructs reflection coefficients $\rho_i = f_{i-1}(\alpha_i)$ with $f_i = \mathcal{M}_{\rho_i}(f_{i-1})$. Again, $f_0 \in \mathcal{B}$ iff all $\rho_i \in \mathbb{D}$.

Finally, the *spectral factor* is another function which can be associated with the measure μ in a unique way. If the Szegő condition $\int \log \mu'(t) d\lambda(t) > -\infty$ is satisfied, then we can define the outer spectral factor σ of the measure μ as

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

An appropriate normalization (e.g., $\sigma(w) > 0$ for some chosen $w \in \mathbb{D}$) fixes c and hence σ uniquely. This function is outer, which means that σ and $1/\sigma$ are both in the Hardy class of the unit disk, with non-tangential limit to the boundary: $\mu'(t) = |\sigma(t)|^2$, $t \in \mathbb{T}$.

To recapitulate: we have, with suitable normalizations and under appropriate conditions, established a relation between the following objects

- a probability measure μ on $\mathbb{T}$
- a function $\Omega \in \mathcal{C}$
- a function $\Gamma \in \mathcal{B}$
- an outer spectral factor σ.

On the boundary we have the relations

$$\mu'(t) = \Re\Omega(t) = |\sigma(t)|^2, \qquad t \in \mathbb{T}$$

and when not on $\mathbb{T}$, we have

$$\mu'(z) = \frac{1}{2}[\Omega(z) + \Omega_*(z)] = \sigma(z)\sigma_*(z)$$

wherever this makes sense.

For an absolutely continuous measure, the relation is even 1-1. In any case approximating one of these objects normally results in a corresponding approximation of the others. This will be illustrated in the next 4 sections. In later sections, we shall concentrate on the NP algorithm (approximation of Γ) and leave the approximantion of the other objects to the phantasy of the reader.

More about the material in this section can be found in classical books on complex analysis.

3 The spaces $\mathcal{L}_n$ and ORF

We shall fix a sequence of points $(\alpha_i) \subset \mathbb{D}$ and consider the nested spaces

$$\mathcal{L}_n = \left\{ \frac{p_n(z)}{\pi_n(z)} : p_n \in \mathcal{P}_n, \pi_n(z) = \prod_{i=1}^{n}(1 - \overline{\alpha}_i z) \right\}.$$

These spaces are generated by a basis of finite Blaschke products $\mathcal{L}_n = \text{span}\{B_k\}_{k=0}^n$ where $B_0 = 1$ and for $n \geq 1$, B_n is a finite product of Blaschke factors ζ_k: $B_n = \zeta_1 \ldots \zeta_n$. Such a Blaschke factor is a Moebius transform of the unit disk:

$$\zeta_k(z) = z_k \frac{z - \alpha_k}{1 - \overline{\alpha}_k z}, \quad z_k = -\frac{\overline{\alpha}_k}{|\alpha_k|}.$$

The factors z_k are needed to make it possible that the product converges when $n \to \infty$. Repetitions in the sequence α_k are allowed. When $\alpha_k = 0$, we set $z_k = 1$. Whenever it is needed, we shall set $\alpha_0 = 0$. Note that when $\alpha_k = 0$ for all k, then $\zeta_k(z) = z$, $B_n = z^n$ and thus $\mathcal{L}_n = \mathcal{P}_n$ so that we are then back in the classical polynomial case.

The superstar conjugate for polynomials will be generalized for a function in $\mathcal{L}_n \setminus \mathcal{L}_{n-1}$ by $f^*(z) = B_n(z) f_*(z)$. By orthonormalization we can obtain the set $\{\varphi_n\}$ of ORFs. We shall fix them by requiring that their leading coefficient $\kappa_n = a_{nn}$ in the expansion $\varphi_n(z) = \sum_{k=0}^n a_{kn} B_k(z)$ is positive. Note that $\kappa_n = \varphi_n^*(\alpha_n)$.

When $k_n(z, w) = \sum_{k=0}^n \varphi_k(z)\overline{\varphi_k(w)}$ is the reproducing kernel for $\mathcal{L}_n$, then it can be shown that $k_n(z, \alpha_n) = \kappa_n \varphi_n^*(z)$ and consequently that $k_n(\alpha_n, \alpha_n) = \kappa_n^2$.

The reproducing kernel and the ORF can be related by the previous summation, but they are also linked by the Christoffel-Darboux relations, which can take one of the following forms

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

The proof consists in first showing that the right hand side is in the correct space $\mathcal{L}_n$ and then, using the orthogonality of the φ_k, show that it reproduces the functions in $\mathcal{L}_n$.

In a signal processing context, the most direct generalization of the Levinson algorithm (which is equivalent with the Szegő recurrence relation for orthogonal polynomials) is not the recurrence relation for the ORF, but it is a recurrence relation for the reproducing kernels. This takes the following form, which follows from the Christoffel-Darboux relations

$$\begin{bmatrix} k_n^*(z, w) \\ k_n(z, w) \end{bmatrix} = t_n(z, w) \begin{bmatrix} k_{n-1}^*(z, w) \\ k_{n-1}(z, w) \end{bmatrix} \tag{3.1a}$$

(the superstars are with respect to z); with transition matrix

$$t_n(z, w) = c_n(w) \begin{bmatrix} 1 & \overline{\rho}_n(w) \\ \rho_n(w) & 1 \end{bmatrix} \begin{bmatrix} \zeta_n(z) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & \overline{\gamma}_n(w) \\ \gamma_n(w) & 1 \end{bmatrix} \tag{3.1b}$$

where $c_n(w) = (1 - |\rho_n(w)|^2)^{-1}$ is a normalizing factor and where

$$\begin{aligned} \rho_n(w) &= \overline{\varphi_n(w)}/\varphi_n^*(w) \\ \gamma_n(w) &= -\zeta_n(w)\rho_n(w) \end{aligned}$$

The ρ_n are the direct generalizations of the reflection coefficients. As you can see, they depend on w. Setting all $\alpha_k = w = 0$, one gets $\zeta_n(z) = z$, $\gamma_n(0) = 0$ and $k_n(z, 0) = \varphi_n^*(z)$ so that this collapses to the Szegő recurrence.

It is however perfectly possible to find recurrences for the ORF themselves. Using the Christoffel-Darboux relations again, one may obtain the following

$$\begin{bmatrix} \varphi_n(z) \\ \varphi_n^*(z) \end{bmatrix} = \theta_n(z) \begin{bmatrix} \varphi_{n-1}(z) \\ \varphi_{n-1}^*(z) \end{bmatrix} \tag{3.2a}$$

$$\theta_n = e_n \begin{bmatrix} \eta_{n1} & 0 \\ 0 & \eta_{n2} \end{bmatrix} \begin{bmatrix} 1 & \overline{\lambda}_n \\ \lambda_n & 1 \end{bmatrix} \begin{bmatrix} \zeta_{n-1}(z) & 0 \\ 0 & 1 \end{bmatrix} \frac{1-\overline{\alpha}_{n-1}z}{1-\overline{\alpha}_n z} \tag{3.2b}$$

where

$$e_n^2 = \frac{1-|\alpha_n|^2}{1-|\alpha_{n-1}|^2} \frac{1}{1-|\lambda_n|^2}$$

is a factor to make the functions have norm 1 and the $\eta_{ni} \in \mathbb{T}$ are unimodular constants to make the leading coefficients $\kappa_n > 0$. The role of the reflection coefficients is here played by the λ_n which are given by the following equivalent formulas

$$\begin{aligned} \lambda_n = \eta_n \rho_n(\alpha_{n-1}) &= \eta_n \overline{\frac{\varphi_n(\alpha_{n-1})}{\varphi_n^*(\alpha_{n-1})}}, \quad \eta_n \in \mathbb{T} \\ &= z_{n-1} \frac{\left\langle \varphi_k, \frac{z-\alpha_{n-1}}{1-\overline{\alpha}_n z} \varphi_{n-1} \right\rangle}{\left\langle \varphi_k, \frac{1-\overline{\alpha}_{n-1} z}{1-\overline{\alpha}_n z} \varphi_{n-1}^* \right\rangle}, \quad k \in \{0, \ldots, n-1\}. \end{aligned}$$

Note that the Christoffel-Darboux relations imply that all the zeros of φ_n are in $\mathbb{D}$ (and hence, those of φ_n^* are outside $\mathbb{D} \cup \mathbb{T}$). Moreover we also get $|\varphi_n(w)/\varphi_n^*(w)| < 1$ for $w \in \mathbb{D}$, so that it follows immediately that $\rho_n(w), \gamma_n(w) \in \mathbb{D}$ for $w \in \mathbb{D}$ and also $\lambda_n \in \mathbb{D}$, which is a well known property of the classical reflection coefficients.

4 Interpolation and finite moment problems

From our signal processing interpretation, we may expect that the $k_n(z,0)/k_n(0,0)$ will give an approximation of the optimal infinite predictor. Since $k_n(z,\alpha_n)/k_n(\alpha_n,\alpha_n) = \varphi_n^*(z)/\kappa_n$ we could choose the sequence α_n to tend to 0 and expect that we then get the optimal predictor in the end. Anyhow, whether we let α_n tend to 0 or not, we could expect that something should be approximated in some sense by expressions involving the ORF. So the question is in what sense do we obtain approximations for the measure μ', and for all the functions we have related to the measure like $\Omega \in \mathcal{C}$, $\Gamma \in \mathcal{B}$ and σ?

To answer this question, we have to introduce the so called functions of the second kind, which we shall denote by ψ_n. There are two ways of characterizing them. One way is to say that they are another, independent, solution of the recurrence for the ORFs. So the definition would then be

$$\begin{bmatrix} \psi_n(z) \\ -\psi_n^*(z) \end{bmatrix} = \theta_n(z) \begin{bmatrix} \psi_{n-1}(z) \\ -\psi_{n-1}^*(z) \end{bmatrix}; \quad \psi_0(z) = 1$$

where θ_n is as in (3.2b).

Another characterization is by the following definition

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

Both definitions are equivalent and whatever definition is taken, the other will follow.

From the first definition, we most easily obtain the determinant relation, by taking the determinants in the following expression which collects the recurrence for φ_n and ψ_n

$$\begin{bmatrix} \varphi_n & \psi_n \\ \varphi_n^* & -\psi_n^* \end{bmatrix} = \theta_n\theta_{n-1}\dots\theta_1 \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

This leads immediately to

$$\frac{1}{2}[\psi_n\varphi_{n*} + \psi_{n*}\varphi_n] = \frac{1-|\alpha_n|^2}{(z-\alpha_n)(z-\alpha_n)_*} = P(z,\alpha_n)$$

where we see the Poisson kernel appear on the right. Dividing by $\varphi_{n*}\varphi_n$ and setting $\Omega_n = \psi_n^*/\varphi_n^*$, we get

$$\frac{1}{2}[\Omega_n + \Omega_{n*}] = \frac{P(z,\alpha_n)}{\varphi_n\varphi_{n*}}.$$

This suggests [compare with (2.2)] that Ω_n is a (rational) function in class $\mathcal{C}$ which is the Riesz-Herglotz transform of the (rational) weight function $\mu_n'(t) = P(t,\alpha_n)/|\varphi_n(t)|^2$ on the right hand side. This is true indeed:

$$\mathcal{C} \ni \Omega_n(z) = \int D(t,z)d\mu_n(t), \quad d\mu_n(t) = \frac{P(t,\alpha_n)}{|\varphi_n^*(t)|^2}d\lambda(t).$$

Thus somehow, Ω_n approximates Ω and somehow μ_n approximates μ. Approximates means that Ω_n "interpolates" Ω in the points $0 = \alpha_0, \alpha_1, \dots, \alpha_n$ while μ_n and μ define the same inner product in $\mathcal{L}_n$. Let us make this more precise. Interpolation has to be understood in the following sense

$$\frac{\varphi_n^*(z)\Omega(z) - \psi_n^*(z)}{zB_n(z)} \in H(\mathbb{D}) \quad \text{and} \quad \frac{\varphi_n(z)\Omega(z) + \psi_n(z)}{zB_{n-1}(z)} \in H(\mathbb{D})$$

where $H(\mathbb{D})$ denotes the set of holomorphic functions in $\mathbb{D}$. Thus Ω_n interpolates Ω in $0, \alpha_1, \dots \alpha_n$ and $-\Omega_{n*}$ interpolates Ω in $0, \alpha_1, \dots, \alpha_{n-1}$ in a linearized Hermite sense. If some repetition of points α_i appear, we shall have matched derivatives in those points. Note a certain asymmetry, since in the first formula we have satisfied $n+1$ interpolation conditions, while in the second only n conditions were satisfied. This can be explained by our normalizations: since $\Omega(0) = 1$ and $\psi_0 = \varphi_0 = 1$, the interpolation in $z = 0$ is "free" in the first formula.

It can be shown by partial fraction decomposition that the Gram matrix for the inner product in $\mathcal{L}_n$ which is generated by the basis functions $\{B_k\}_{k=0}^n$ only depends on the values of $\Omega(\alpha_i)$ for $i = 0, \dots, n$, where function values have to be replaced by derivative values when some of the points are repeated. By our previous interpolation result, we see that these are precisely those values which are the same for Ω and Ω_n. Thus $\langle\cdot,\cdot\rangle_\mu = \langle\cdot,\cdot\rangle_{\mu_n}$ in $\mathcal{L}_n$. In the latter form, it is obvious that we have solved here a finite moment problem. For example, when all the $\alpha_k = 0$, then $\mathcal{L}_n = \mathcal{P}_n$ are just polynomials and the interpolation conditions state that $\Omega_n^{(j)}(0) = \Omega^{(j)}(0)$ for $j = 0, 1, \dots, n$. Recalling (2.2) it is seen that this means precisely that we solved here a Carathéodory-Fejér interpolation problem or equivalently that we have constructed a solution to a finite trigonometric moment problem where the moments $r_k, k = 0, 1, \dots, n$ are given.

When not all the α_k collapse in the origin, then we solve a Nevanlinna-Pick interpolation problem (which is usually formulated in terms of the corresponding Schur function Γ) but at the same time we solve a generalization of a moment problem. If the inner products are the same in $\mathcal{L}_n$, then this means that for example $\int B_k(t)d\mu_n(t) = \int B_k(t)d\mu(t)$ for $k = 0, \ldots, n$ Thus that we have constructed a measure μ_n which has some preassigned (generalized) moments $\int B_k d\mu_n$. Of course we can choose any other basis b_k for $\mathcal{L}_n$ instead of the Blaschke products and call $\int b_k d\mu$ generalized moments.

We can make this more clear as follows. Using the MacLaurin expansion for the kernel $D(t,z) = (t+z)/(t-z) = 1 + 2\sum_{k\geq 1}(z/t)^k$, we obtain

$$\Omega(z) = \int D(t,z)d\mu(t) = 1 + 2\sum_{k=1}^{\infty} r_k z^k, \quad r_k = \int t^{-k} d\mu(t).$$

If Ω_n is a Carathéodory-Fejér approximant for Ω, then it matches the moments $r_0, \ldots, r_n$ and thus a finite trigonometric moment problem has been solved, implying that the inner products w.r.t. μ and μ_n are the same in the subspace of polynomials $\mathcal{P}_n$.

If however, we use a Newton expansion for $D(t,z) = 1 + 2\sum_{k\geq 1} z\pi^*_{k-1}(z)/\pi^*_k(t)$ where $\pi^*_k(z) = (z-\alpha_1)\ldots(z-\alpha_k)$, then we obtain

$$\Omega(z) = \int D(t,z)d\mu(t) = 1 + 2\sum_{k=1}^{\infty} \tilde{r}_k z\pi^*_{k-1}(z), \quad \tilde{r}_k = \int \frac{d\mu(t)}{\pi^*_k(t)}. \tag{4.1}$$

Now the $\tilde{r}_k$ are generalized moments and interpolation of Ω in the points $0, \alpha_1, \ldots, \alpha_n$ corresponds to matching the moments $\tilde{r}_0, \ldots, \tilde{r}_n$ and therefore we have equality of the inner products w.r.t. μ and μ_n in the subspace $\mathcal{L}_n$.

Finally we remark that the φ^*_n also approximate the Szegő kernel, which is basically the spectral factor of $\mu(t)/P(t,\alpha)$ and it reproduces functions in $H_2(\mu)$, that is the analytic functions in $L_2(\mu)$. We shall come back to that in a later section on convergence.

The opening sentences of this section, which referred to the signal processing application suggested that it would be most natural to find that the reproducing kernels approximate in some sense the spectral factor of the measure (see also the first section). Also this is precisely as expected. Define the normalized kernels $K_n(z,w) = k_n(z,w)/k_n(w,w)$, then we can obtain the following interpolating property for the inverse of the spectral factor

$$\frac{K_n(z,0) - 1/\sigma(z)}{zB_n(z)} \in H(\mathbb{D}).$$

Along the same lines we have followed above, we could define "kernels of the second kind" (two approaches are possible), and construct interpolating functions for Ω, which is equivalent to the solution of a generalized moment problem as explained before. However, it can also be derived directly from the interpolating property for the spectral factor that we have

$$d\mu_n(t) = \frac{d\lambda(t)}{|K_n(t,0)|^2} \Rightarrow \langle\cdot,\cdot\rangle_\mu = \langle\cdot,\cdot\rangle_{\mu_n} \text{ in } \mathcal{L}_n.$$

As we said, the Nevanlinna-Pick algorithm is usually formulated in terms of a Schur function. We shall illustrate this now. Let μ be a probability measure and $\Omega \in \mathcal{C}$ its Riesz-Herglotz transform with $\Omega(0) = 1$, then we may pick a fixed $w \in \mathbb{D}$ and construct the Schur

function $\Gamma = \Gamma_0(z,w) = [\Omega(z) - \Omega(w)]/[\Omega(z) + \Omega(w)]$ (note $\Gamma(w) = 0$). We define $\Gamma_0 = \Gamma$ and construct the sequence Γ_n by the transformations

$$\Gamma_{n-1}(z,w) \to \Gamma'_n(z,w) = \frac{1}{\zeta_n(z)} \left(\frac{\Gamma_{n-1}(z,w) - \gamma_n(w)}{1 - \overline{\gamma}_n(w)\Gamma_{n-1}(z,w)} \right), \quad \gamma_n(w) = \Gamma_{n-1}(\alpha_n, w)$$

$$\Gamma'_n(z,w) \to \Gamma_n(z,w) = \frac{\Gamma'_n(z,w) - \rho_n(w)}{1 - \overline{\rho}_n(w)\Gamma'_n(z,w)}, \quad \rho_n(w) = \Gamma'_n(w,w)$$

Then, it can be shown that these γ_n and ρ_n are the same as those featuring in the recurrence relation (3.1) for the reproducing kernels. The idea behind these maps is that first a Moebius transform with γ_n will generate a zero in α_n and this zero is divided out with the factor $1/\zeta_n$. The second step renormalizes Γ'_n to have a zero at w. This renormalization step is not really necessary and does not occur in the original NP algorithm. However, it does allow to produce the $\gamma_n(w)$ and $\rho_n(w)$ needed in the recurrence (3.1) and it is an alternative to generate them. These reflection coefficients are essential in the signal processing context because one does not really need the predictor in the form of a polynomial or a rational function to realize the filter. It is even simpler to realize it in a lattice network which only needs the parameters γ_n and ρ_n.

A similar algorithm can be designed to generate the λ_n in the recurrence relation (3.2) for the ORF. Then we construct the Schur function Γ_0 as $\Gamma_0(z) = [\Omega(z) - 1]/[z(\Omega(z) + 1)]$, where we recall that $\Omega(0) = 1$, so that in this construction we have divided out immediately the zero in the origin. Since there is no need to renormalize at w in every step now, we may apply the classical algorithms to generate

$$\Gamma_n(z) = \frac{1}{\zeta_n(z)} \left(\frac{\Gamma_{n-1}(z) - \lambda'_n}{1 - \overline{\lambda'}_n\Gamma_{n-1}(z)} \right), \quad \lambda'_n = \Gamma_{n-1}(\alpha_n).$$

The plot is now simpler: generate a zero at α_n and divide it out and that's it. The λ'_n are given a prime here since they are only equal to the λ_n in (3.2b) up to a unimodular factor. This is because we have chosen a positive leading coefficient: $\kappa_n > 0$. Another normalization would have given exactly the same reflection coefficients.

5 Quadrature, moments and Favard

If we want to construct quadrature formulas on the unit circle, we would like the knots to be on $\mathbb{T}$ and the weights to be positive, just as we have for quadrature on the real axis. The zeros of the ORF are however all strictly inside $\mathbb{D}$ and those of φ_n^* are in $|z| > 1$. To generate knots on $\mathbb{T}$, we have to introduce para-orthogonal rational functions defined as

$$Q_n(z,\tau) = \varphi_n(z) + \tau\varphi_n^*(z), \qquad \tau \in \mathbb{T}.$$

These are also in $\mathcal{L}_n$ and it can be shown that they have precisely n simple zeros $\{\xi_{n1}, \ldots, \xi_{nn}\}$, all in $\mathbb{T}$. We call then para-orthogonal since they are orthogonal to the subspace $\mathcal{L}_{n-1} \cap \zeta_n\mathcal{L}_{n-1}$ while they are not orthogonal to 1 and B_n.

Constructing in our usual way functions P_n of the second kind associated with Q_n, it turns out that these are given by $P_n(z,\tau) = \psi_n(z) - \tau\psi_n^*(z)$, where the ψ_n are the second kind functions for the ORF. It follows easily from the interpolating properties we have seen

before that we obtain now the interpolating property

$$\frac{Q_n\Omega + P_n}{zB_{n-1}} \in H(\mathbb{D}).$$

In other words $\Omega_n = -P_n/Q_n$ interpolates Ω in the points $0 = \alpha_0, \alpha_1, \ldots, \alpha_{n-1}$. By our previous observations, this means that we can also construct a measure μ_n such that $\langle \cdot, \cdot \rangle_{\mu_n} = \langle \cdot, \cdot \rangle_{\mu}$ in $\mathcal{L}_{n-1}$. Because Ω_n has the ξ_{ni} as poles on $\mathbb{T}$, this measure is a discrete measure and it can be written as a weighted sum of Dirac delta's in those ξ_{ni} with weights w_{ni}. The latter can be computed and they are given by the following equivalent expressions

$$w_{ni} = \frac{1}{2\xi_{ni}} \frac{P_n(\xi_{ni})}{Q_n'(\xi_{ni})} = \left[\sum_{j=0}^{n-1} |\varphi_j(\xi_{ni})|^2 \right]^{-1} > 0$$

(prime means derivative now) which obviously shows that they are positive. Because the inner product w.r.t. μ and μ_n was equal in $\mathcal{L}_{n-1}$, it follows that the quadrature formula

$$\sum_{k=1}^{n} w_{nk} f(\xi_{nk}) \approx \int f(t) d\mu(t)$$

is exact for all $f \in \mathcal{R}_{n-1} = \mathcal{L}_{n-1} + \mathcal{L}_{(n-1)*}$ where $\mathcal{L}_{k*} = \{f_* : f \in \mathcal{L}_k\}$.

Again we have solved here a certain moment problem. In fact this quadrature formula can be used to construct a solution to the infinite moment problem where an infinite sequence of (generalized) moments are given and the question is to describe all its solutions. Of course this is in a sense equivalent to the study of convergence properties for the kernels, for the ORFs etc.

Under reasonable conditions (the moment matrices should be positive definite) the generalized moment problem as discussed in the previous section always has at least one solution which may be obtained as the limit of a convergent subsequence of e.g. the measures μ_n which describe the previous quadrature formulas. Following Akhiezer [2], we can show that for $z \in \mathbb{C}$, a fixed complex number, and for $\tau \in \mathbb{T}$ free, the values $R_n(z,\tau) = -P(z,\tau)/Q_n(z,\tau)$ describe a circle $K_n(z)$. Its radius is given by

$$r_n = \left| \frac{\psi_n^* \varphi_n + \psi_n \varphi_n^*}{|\psi_n^*|^2 - |\varphi_n|^2} \right| = \frac{2|z||B_{n-1}(z)|}{|1-|z|^2| k_{n-1}(z,z)}$$

from which we see immediately that if the Blaschke product goes to zero, then the radius goes to zero and hence the limiting circle is a point which means that the moment problem has a unique solution. It is well known that under the Blaschke condition $\sum_{k=1}^{\infty}(1-|\alpha_k|) = \infty$, the Blaschke product B_n converges locally uniformly to 0 in $\mathbb{D}$ which we denote as $B_n \rightrightarrows 0$. Note that this also shows the invariance property, that is, $K_\infty(z)$ is a point for some z, then it is a point fo any other z too.

If the Blaschke condition is not satisfied, then K_∞ can be a circle with positive radius. The circle and its interior are described by the values of all $\Omega_\mu(z) = \int D(t,z) d\mu(t)$ where μ ranges over all solutions of the moment problem. A solution μ for which Ω_μ is on the boundary K_∞ is called N-extremal. A solution will be N-extremal iff $\mathcal{L}_\infty$ is dense in $L_2(\mu)$. We have not definitely formulated the sharpest condition under which the generalized moment problem has a unique solution yet. However, we have a strong feeling that these can be expressed

in terms of a Carleman condition like $\sum_k (\tilde{r}_k)^{-1/2k} = \infty$ where the $\tilde{r}_k$ are the generalized moments.

The question of uniqueness of the solution of the moment problem is intimately related with the uniqueness problem in Favard's theorem. This theorem states that if the rational functions φ_n are generated by a recurrence relation of the form and with the properties of (3.2), then these will form an ORF system for some measure μ. Again, the quadrature formula, we have given above plays a central role in constructing this measure (see [13]). Whether this μ will be unique is equivalent to the previous uniqueness problem.

A question which has not received much attention in the literature, and which in this context seems natural is whether it is true that kernel functions which are generated by a recurrence relation of the form and with the properties of (3.1), are reproducing kernels for some measure μ. Some partial results in this direction were obtained in [14].

6 Convergence results

We recall the Blaschke condition (BC): $\sum(1-|\alpha_k|) = \infty$ which is equivalent with $B_n \rightrightarrows 0$ in $\mathbb{D}$ and also with the system $\{B_k\}_{k\geq 0}$ being complete in $H_2(\mu)$. This condition says that the α_k can approach $\mathbb{T}$ in the limit, but not too fast. A stronger condition is (SBC): $|\alpha_k| \leq r < 1$, which says that the α_k should be bounded away from $\mathbb{T}$.

Also recall the Szegő condition (SC): $\int \log \mu'(t) d\lambda(t) > -\infty$ which ensures the existence of the spectral factor σ and which prevents the system $\{t^k\}_{k\geq 0}$ from being complete in $L_2(\mu)$.

We might expect that if $\mathcal{L}_n$ approaches $H_2(\mu)$, then its reproducing kernel should approach the reproducing kernel for $H_2(\mu)$. The latter is called the Szegő kernel and is given by

$$s(z,w) = [(1-\overline{w}z)\overline{\sigma(w)}\sigma(z)]^{-1}$$

From our interpolation results, we recall that also the measure $d\mu(t)/P(t,w)$ plays a role where $P(z,w)$ is the Poisson kernel. The inverse of its spectral factor is given by

$$\begin{aligned}\pi(z,w) &= \exp\left\{-\frac{1}{2}\int D(t,z)\log\left[\frac{\mu'(t)}{P(t,w)}\right]d\lambda(t)\right\} \\ &= z_w \frac{\sqrt{1-|w|^2}}{(1-\overline{w}z)\sigma(z)}, \qquad z_w = \frac{\sigma(w)}{|\sigma(w)|}.\end{aligned}$$

Both kernels s and π are related by $s(z,w) = \pi(z,w)\overline{\pi(w,w)}$.

We are now ready to state some of the convergence results known so far. If the Blaschke condition is satisfied, then as $n \to \infty$

$$(\mathrm{BC}) \Longrightarrow \begin{cases} 1. & \|k_n(t,w) - s(t,w)\|_\mu \to 0 \text{ for } w \in \mathbb{D} \\ 2. & \Omega_n = \psi_n^*/\varphi_n^* \rightrightarrows 0 \text{ in } \mathbb{D} \end{cases}$$

If we assume the stronger condition (SBC) and assuming that the subsequence $\alpha_{n(s)} \to w \in \mathbb{D}$, then we have the following convergence results for the corresponding subsequences

$$(\mathrm{SBC}) \Longrightarrow \begin{cases} 1. & k_{n(s)}(z,w) \rightrightarrows s(z,w) \text{ for } z,w \in \mathbb{D} \\ 2. & \varphi_{n(s)} \rightrightarrows 0 \text{ in } \mathbb{D} \\ 3. & \kappa_{n(s)} \to s(w,w) = |\pi(w,w)|^2 \\ 4. & \varphi^*_{n(s)} \rightrightarrows \pi(z,w) \text{ for } z \in \mathbb{D} \end{cases}$$

For other convergence results (ratio asymptotics etc.) see [45, 46, 42, 39]. We also mention here the close relation with the study of orthogonal polynomials with respect to a varying measure [4, 44, 41, 40]. In these papers a more general situation is considered where the sequence of interpolation points (α_k) (the Newton case) is replaced by a triangular array of interpolation points $((\alpha_{nk})_{k=0}^n)_{n\geq 0}$. That is at step n, the interpolation is at the points $(\alpha_{nk})_{k=0}^n$ and this sequence can change at every step.

7 Boundary case

So far, we have assumed that all the interpolation points α_k were in $\mathbb{D}$. In this section we shall assume that they are on the boundary $\mathbb{T}$, but not equal to 1. A theory completely similar to the previous one can be developed. There are strong parallels, but there are also fundamental differences. We quickly go over the results, some of them can be found in [20, 24, 27, 26].

It is easily verified that in this case $\mathcal{L}_n = \mathcal{L}_{n*}$. The basis of Blaschke products is no good anymore, so we have to find another basis, which we take to be $\{b_k\}_{k=0}^n$ with $b_0 = 1$ and $b_n(z) = b_{n-1}(z)Z_n(z)$ where

$$Z_n(z) = \frac{i(1-z)(1-\overline{\alpha}_n)}{1-\overline{\alpha}_n z} = \frac{i(1-z)(1-\alpha_n)}{z-\alpha_n} = Z_{n*}(z)$$

We can now orthonormalize this basis and obtain again a system $\{\varphi_n\}$ of ORFs. However, there may be problems when such a φ_n is singular. We shall call φ_n singular when $\varphi_n(\alpha_{n-1}) = 0$ and regular otherwise. When we do not have a singular ORF, then these functions satisfy a three term recurrence relation of the form $(n \geq 2)$

$$\varphi_n(z) = \left[A_n Z_n(z) + B_n \frac{Z_n(z)}{Z_{n-2}(z)}\right] \varphi_{n-1}(z) + C_n \frac{Z_n(z)}{Z_{n-2}(z)} \varphi_{n-2}(z).$$

When we define the leading coefficient κ_n and the coefficient λ_n of φ_n with respect to the basis $\{b_n\}$ as follows $\varphi_n(z) = \kappa_n b_n(z) + \lambda_n b_{n-1}(z) + \cdots$, then it can be shown that

$$\begin{aligned} E_n &= A_n + B_n/Z_{n-2}(\alpha_{n-1}) = [\kappa_n + \lambda_n/Z_n(\alpha_{n-1})]/\kappa_{n-1} \neq 0 \\ C_n &= -E_n/E_{n-1} \neq 0. \end{aligned}$$

One can prove Christoffel-Darboux relations which now take the form

$$\frac{\varphi_n(w)}{Z_n(w)}\frac{\varphi_{n-1}(z)}{Z_{n-1}(z)} - \frac{\varphi_n(z)}{Z_n(z)}\frac{\varphi_{n-1}(w)}{Z_{n-1}(w)} = F(z,w)E_n \sum_{k=0}^{n-1} \varphi_k(z)\varphi_k(w)$$

with $F(z,w) = [-i(z-w)]/[(z-1)(w-1)] = 1/Z_{n-1}(z) - 1/Z_{n-1}(w)$.

The para-orthogonal functions of the previous case are now replaced by quasi-orthogonal functions, which are defined as

$$Q_n(z,\tau) = \varphi_n(z) + \tau \frac{Z_n(z)}{Z_{n-1}(z)} \varphi_{n-1}(z), \quad \tau \in \mathbb{R}.$$

Again one can show that with exception of at most $n+1$ values of τ (i.e. for a regular τ), $Q_n(z,\tau)$ has n simple zeros $\{\xi_{n1}, \ldots, \xi_{nn}\} \subset \mathbb{T} - \{\alpha_1, \ldots, \alpha_n\}$. After the corresponding

weights $w_{ni} > 0$ have been found, we have, like in the previous case that the quadrature formula

$$\int f(t)d\mu(t) \approx \sum_{k=1}^{n} w_{nk} f(\xi_{nk}) = \int f(t)d\mu_n(t)$$

is exact for all $f \in \mathcal{R}_{n-1} = \mathcal{L}_{n-1} \cdot \mathcal{L}_{(n-1)*}$.

This quadrature gives a solution for the finite moment problem where the moments $\tilde{r}_k$, $k = 0, \ldots, n$ of the same form as in (4.1) are given. It can be shown that the infinite moment problem has at least one solution and that we can always find a subsequence $\mu_{n(s)}$ of the sequence μ_n converging to a solution μ of the moment problem.

We can also define functions of the second kind ψ_n as we did before. The exactness of the quadrature is equivalent with interpolation properties of the function $\Omega_n = \psi_n/\varphi_n$, which have to be stated in a more delicate way. If $\pi_n^*(z) = (z - \alpha_1) \ldots (z - \alpha_n)$, then we can find polynomials f_n and g_n such that $\varphi_n = f_n/\pi_n^*$ and $\psi_n = g_n/\pi_n^*$. It then holds that

$$g_n(z) - f_n(z)\Omega(z) = z\pi_n^*(z)\pi_{n-1}^*(z)A_n(z), \quad \lim_{z\to\alpha_k} A_n(z) \text{ finite}$$
$$g_n(z) - f_n(z)\Omega(z) = [z\pi_n^*(z)\pi_{n-1}^*(z)]_*\hat{A}_n(z), \quad \lim_{z\to 1/\overline{\alpha}_k} \hat{A}_n(z) \text{ finite}$$

for $k = 0, \ldots, n$ ($\alpha_0 = 0$). This says essentially that Ω_n interpolates Ω in linearized Hermite sense in the points $\{0, \infty, \alpha_1, \alpha_1, \ldots, \alpha_{n-1}, \alpha_{n-1}, \alpha_n\}$.

If the subsequence $\mu_{n(s)}$ converges to the solution μ of the moment problem and Ω_μ is its Riesz-Herglotz transform, then we can prove that $\Omega_{n(s)} \rightrightarrows \Omega_\mu$ in $\mathbb{D}$.

8 Quasi-Toeplitz case

If we return to our signal processing motivation for a moment, we recall that we proposed in the first section to minimize the prediction error $\|e\|^2$. However, this infinite sum of squares is often replaced by a finite sum, so that we have to minimize $\sum_{k=0}^{N} |e_k|^2$. This means that we replace the infinite system which was to be solved in a least squares sense, by a finite one. Writing down the normal equations for the infinite system lead to a system whose matrix is the positive definite Toeplitz matrix of moments (autocorrelation coefficients). If we write down the normal equations for the finite system, we do not obtain a Toeplitz matrix anymore. However it is of the form $S_N^H S_N$ where S_N is a $(N+1) \times \infty$ Toeplitz matrix and it has therefore a special structure which is called quasi-Toeplitz, or equivalently, one says that it has a low (Toeplitz) displacement rank.

Mathematically, we have to consider the Hardy space H_2. Suppose $\mathcal{S}$ is a 1-1 transform of H_2 into itself, which maps the natural basis $\nu(z) = [1 \;\; z \;\; z^2 \;\; \cdots]$ to a new basis $\tilde{\nu}(z) = \mathcal{S}\nu(z)$. If $(\cdot, \cdot)$ denotes the classical inner product for H_2, then we define a new inner product as $\langle f, g \rangle = (\mathcal{S}f, \mathcal{S}g) = (f, \mathcal{S}^*\mathcal{S}g)$ with $\mathcal{M} = \mathcal{S}^*\mathcal{S}$ positive definite. Thus the Gram or moment matrix $M = [r_{kl}]$ of the natural basis is still Hermitian and positive definite, but it is no longer Toeplitz; that is, the moments $\langle z^k, z^l \rangle = r_{kl}$ need no longer be a function of $k - l$. However the special structure left can be displayed by a low displacement rank this means that if Z denotes the down shift matrix, then there exists a signature matrix, that is a matrix of the form $\Sigma = \text{diag}(s_i)$ with $s_i = \pm 1$, and a matrix G such that for a moderate integer d

$$M - ZMZ^H = G\Sigma G^H, \quad G \in \mathbb{C}^{\infty\times d}, \quad \Sigma \in \mathbb{C}^{d\times d}. \tag{8.1}$$

The spectral function that we can associate with M is now defined as $M(z,w) = \nu(z) M \nu(w)^H$. Let $\mathcal{F}$ be the set of spectral functions (i.e. for which M is positive definite Hermitian). In the frequency domain, the low displacement property (8.1) is expressed as $(1 - \overline{w}z)M(z,w) = G(z)\Sigma G(w)^H$. The pair (G, Σ) is called a generator for M. A generator defines M uniquely but the generator is not uniquely defined by M. For example if (G, Σ) is a generator for M, then also (GU, Σ) with U an arbitrary Σ-unitary matrix ($U^H \Sigma U = \Sigma$) will also be a generator for M.

A generator will be called proper at $\alpha \in \mathbb{D}$ if $G(\alpha) = [\gamma | 0 \cdots 0]$ with $\gamma > 0$. Any generator can be made proper by multiplication with a Σ-unitary factor.

If we set $G(z) = [\gamma | \Gamma]$ and $\Sigma = 1 \oplus (-J)$, then we may consider $g(z) = \Gamma(z)/\gamma(z)$ as a vectorial Schur function for the signature J since it follows from the positive definiteness of M that $g(z)^H J g(z) < 1$ for $z \in \mathbb{D}$.

Let $\alpha \in \mathbb{D}$ be fixed, and define

$$M_\alpha(z,w) = \frac{M(z,w) - M(z,\alpha)M(\alpha,\alpha)^{-1}M(\alpha,w)}{\zeta_\alpha(z)\overline{\zeta_\alpha(w)}}, \quad \zeta_\alpha(z) = \frac{z-\alpha}{1-\overline{\alpha}z}$$

then the matrix of this bilinear form corresponds to the Schur complement of r_{00} in M when $\alpha = 0$. Therefore we shall call this the (generalized) Schur complement of M at α.

The Schur algorithm is based on the property which says that $M(z,w) \in \mathcal{F}$ iff $M(0,0) > 0$ and $M_0(z,w) \in \mathcal{F}$. The Nevanlinna-Pick algorithm is based on the fact that $M(z,w) \in \mathcal{F}$ iff $M(\alpha,\alpha) > 0$ and $M_\alpha(z,w) \in \mathcal{F}$. Thus in the NP algorithm we choose a sequence of points in the disk $(\alpha_k) \subset \mathbb{D}$ and construct a sequence of Schur complements $M_k(z,w) \in \mathcal{F}$ at α_k by the previous transforms.

The Schur-invariance theorem says that the signature Σ and the displacement rank d of the sequence of Schur complements $M_k(z,w)$ is independent of k. The immediate consequence is that we do not need to update from M_k to M_{k+1}, but it is sufficient to update the generator matrix G_k to G_{k+1}. Thus the NP algorithm goes as follows: Assume that G_{k-1} is proper at α_{k-1}, then we can first divide out the zero α_k by the transform

$$G_{k-1}(z) \to G'_k(z) = G_{k-1}(z) \begin{bmatrix} 1 & 0 \\ 0 & \zeta_{k-1}(z) I_{d-1} \end{bmatrix}^{-1}$$

and then make G'_k proper at α_k by a Σ-unitary transformation

$$G'_k(z) \to G_k(z) = G'_k(z) V_k^{-1}$$

where

$$V_k = \begin{bmatrix} P_k & 0 \\ 0 & Q_k \end{bmatrix} \begin{bmatrix} 1 & \Lambda_k \\ \Lambda_k^H & J \end{bmatrix} \tag{8.2}$$

with Λ_k acting as $1 \times (d-1)$ vectorial reflection coefficients:

$$\Lambda_k = \Gamma'_k(\alpha_k)/\gamma'_k(\alpha_k) \in \mathbb{C}^{1 \times (d-1)}.$$

The property of the reflection coefficients being in $\mathbb{D}$ is now generalized to $P_k^2 = 1 - \Lambda_k J \Lambda_k^H > 0$, while the relation with Q_k is $Q_k^H J Q_k = J - \Lambda_k^H \Lambda_k$.

To introduce the ORF we need one more definition. A generator is called admissible if there exists a constant vector $a \in \mathbb{C}^{1 \times d}$ such that $G(z)a = 1$. It can be shown that such

an admissible generator always exists, but it may not be minimal. Define the matrix θ_n in terms of the V_n matrix (8.2) as

$$\theta_n(z) = \frac{\sqrt{1-|\alpha_n|^2}}{\sqrt{1-|\alpha_{n-1}|^2}} \frac{1-\overline{\alpha}_n z}{1-\overline{\alpha}_n z} \begin{bmatrix} \zeta_{n-1}(z) & 0 \\ 0 & I_{d-1} \end{bmatrix} V_n^H$$

then the scalar function φ_n and the $1 \times (d-1)$ function Φ_n, which are generated by

$$[\varphi_n(z)|\Phi_n^*(z)] = [\varphi_{n-1}(z)|\Phi_{n-1}^*(z)]\theta_n(z), \qquad [\varphi_0|\Phi_0^*] = a$$

will have entries which are in $\mathcal{L}_n$. If we assume that φ_n has all its zeros in $\mathbb{D}$ then it can be shown that the φ_n will be the ORF for the quasi-Toeplitz inner product.

Also here we have Christoffel-Darboux relations, which now look like

$$\begin{aligned} k_n(z,w) &= \sum_{k=0}^{n} \varphi_k(z)\overline{\varphi_k(w)} \\ &= \frac{a\Sigma a^H}{1-\overline{w}z} - \frac{[\varphi_{n+1}(z)|\Phi_{n+1}^*(z)]\Sigma[\varphi_{n+1}(w)|\Phi_{n+1}^*(w)]^H}{1-\zeta_{n+1}(z)\overline{\zeta_{n+1}(w)}} \\ &= \frac{a\Sigma a^H}{1-\overline{w}z} - \frac{\varphi_n(z)\zeta_n(z)\overline{\zeta_n(w)\varphi_n(w)} - \Phi_n^*(z)J\Phi_n^*(w)^H}{1-\zeta_n(z)\overline{\zeta_n(w)}} \end{aligned}$$

More details can be found in [21]. See also [48].

9 Matrix versions

If in our prediction problem, we have vector valued signals, i.e., each sample $s_k \in \mathbb{C}^p$, then the power spectrum $R(e^{i\omega})$ of the signal will be a function with values in the set of Hermitian positive definite $p \times p$ matrices. The moments $\int t^{-k} d\mu(t)$ are $p \times p$ matrices and the predictor $A(z)$ is a square matrix valued function with $A(0) = I_p$. For the matrix valued inner product, we could distinguish a left and a right version

$$\langle f,g\rangle_R = \int [f(t)]d\mu(t)[g(t)]^H \quad \text{or} \quad \langle f,g\rangle_L = \int [f(t)]^H d\mu(t)[g(t)]$$

which give rise to left and right ORFs which may be obtained by orthogonalization of the basis functions $B_n(z)I_p$. Except for the non-commutativity of the products and possible problems of degeneracy, in this setting, the theory is much the same as in the scalar case. See for example [50, 51, 29, 5, 6].

However, a much finer structure is possible here. Instead of the basis functions $B_n(z)I_p$, one could consider basis functions which are products of functions of the form

$$Z_n(z) = U_n[\zeta_n(z)I_{k_n} \oplus I_{p-k_n}]U_n^H, \qquad U_n^H \Sigma U_n = \Sigma$$

where Σ is some signature matrix. The basis functions and the spaces to be considered are then given by

$$B_0 = I_p, \quad B_n(z) = B_{n-1}(z)Z_n(z), \qquad \mathcal{L}_n = \operatorname{span}\{B_k\}_{k=0}^n.$$

Moreover, we should not only consider the case of square valued functions, but also of rectangular functions. This is needed when the filter is considered as the transfer function of a system with a q-dimensional input signal and producing a p-dimensional output signal. Then we should consider $p \times q$ valued functions.

A considerable simplification of the problem occurs if we do not consider the matricial problem directly, but split it up as a sequence of vectorial problems by "looking at the matrices in certain directions", for example by considering the problem row per row. Let us be more explicit. Suppose we want to solve a directional or tangential Nevanlinna-Pick problem for a Schur function $\Gamma(z) \in \mathcal{B}^{p\times q}$, which means that Γ is analytic in $\mathbb{D}$ and

$$\|\Gamma\|_\infty = \sup_{t\in\mathbb{T}} \sigma_{\max}(\Gamma(t)) < 1$$

where $\sigma_{\max}$ represents the maximal singular value. Suppose we are given the distinct points $\alpha_1, \ldots, \alpha_N$ in $\mathbb{D}$, the vectors $x_1, \ldots, x_N$ all in $\mathbb{C}^{1\times p}$ and the vectors $y_1, \ldots, y_N$ all in $\mathbb{C}^{1\times q}$. The problem is to find all functions $\Gamma \in \mathcal{B}^{p\times q}$ such that the directional interpolation conditions $y_i = x_i\Gamma(\alpha_i)$ are satisfied for $i = 1, \ldots, N$. By choosing for x_i the natural basis vectors, possibly repeated cyclically, we select in each interpolation condition one of the rows of Γ. When some points coincide, an adaptation for derivatives should been made.

It is easier to describe the algorithm if we write the function Γ as a ratio of two matrix valued functions, similar to what has been done in the previous section. Thus we consider $\Gamma(z) = U_0(z)^{-1}V_0(z)$ and set

$$G_0(z) = [U_0(z) \;\; V_0(z)], \quad U_0 : p \times p, \quad V_0 : p \times q$$

with U_0 and V_0 analytic in $\mathbb{D}$. The following form of the directional Nevanlinna-Pick algorithm can be found in [1]. A sequence of analytic $p\times(p+q)$-valued functions G_k is constructed, starting with G_0 where at each step the transition is described by the transformations

$$G_{k+1}(z) = \theta_k^L(z)G_k(z)\theta_k^R(z)$$

where

$$\theta_k^R(z) = [I_{p+q} + (\zeta_k(z) - 1)\Sigma_{pq}z_k^H(z_k\Sigma_{pq}z_k^H)^{-1}z_k]W_k$$
$$\theta_k^L(z) = [I_p + (\zeta_k^{-1}(z) - 1)x_k^H(x_kx_k^H)^{-1}x_k]$$

$$\Sigma_{pq} = I_p \oplus (-I_q), \quad W_k\Sigma_{pq}W_k = \Sigma_{pq}, \quad z_k = x_kG_k(\alpha_k)$$

The meaning is that first θ_k^R is applied such that $x_kG_k(\alpha_k)\theta_k^R(\alpha_k) = 0$, that is we produce a zero at α_k in the direction of x_k. Then the transformation θ_k^L is applied which divides out this zero. The latter may be more obvious if we write

$$[\theta_k^L]^{-1} = [I_p + (\zeta_k(z) - 1)x_k^H(x_kx_k^H)^{-1}x_k]$$

which is a Blaschke-Potapov factor.

If one is interested in the recurrence for the kernels, one could again introduce in each step an extra normalization which generates a zero at some fixed point $w \in \mathbb{D}$ as it was done in the scalar case.

What we presented here is a one-sided directional formulation. We could of course work on the columns instead of the rows, or even work on both sides simultaneously.

10 Time varying case

As a last generalization we mention the situation in which the signal is not stationary or close to stationary. This means that we consider systems which map a vector valued input $u_k \in \mathbb{C}^{p\times 1}$ into a vector valued output $y_k \in \mathbb{C}^{q\times 1}$ and which has a transfer function $H_k(z)$ which changes at every instance of time. A state space description of a time invariant system looks like

$$\begin{array}{rcl} x_{k+1} & = & Ax_{k+1} + Bu_k \\ y_k & = & Cx_k + Du_k \end{array} \rightarrow H(z) = D + C(I - zA)^{-1}zB$$

where u_k is the input vector, y_k is the output vector and $x_k \in \mathbb{C}^{n\times 1}$ is the state vector (all at time k). $H(z)$ is the transfer function, because in the frequency domain, the relation between input and output is given by $Y(z) = H(z)U(z)$. If the dimension n of the state space (the space in which the state vectors live) is minimal, the realization is called minimal. In this case, neither the transfer function, nor the state matrices A, B, C, D depend on time k. However, in the time varying case, these do depend on time and we get a state space description of the form

$$\begin{array}{rcl} x_{k+1} & = & A_k x_{k+1} + B_k u_k \\ y_k & = & C_k x_k + D_k u_k \end{array} \rightarrow H_k(z) = D_k + C_k(I - zA_k)^{-1}zB_k$$

where the state vectors $x_k \in \mathbb{C}^{n_k\times 1}$ live in a time varying state space of dimension n_k at time k. We shall indicate this as $(x_k) \in \ell^{(n_k)}$ while the dimension of input and output space is constant: $(u_k) \in \ell^q$ and $(y_k) \in \ell^p$. If we denote by $\vec{x}, \vec{u}, \vec{y}$ the columns whose entries are the vectors x_k, u_k and y_k respectively, then we can represent the behaviour of the system over all time by the global relations

$$\begin{array}{rcl} \mathcal{Z}^{-1}\vec{x} & = & \mathcal{A}\vec{x} + \mathcal{B}\vec{u} \\ \vec{y} & = & \mathcal{C}\vec{x} + \mathcal{D}\vec{u} \end{array} \rightarrow \vec{y} = \mathcal{H}\vec{u} = [\mathcal{D} + \mathcal{C}(I - \mathcal{Z}\mathcal{A})^{-1}\mathcal{Z}\mathcal{B}]\vec{u}$$

where $\mathcal{A} = \mathrm{diag}(A_k)$, $\mathcal{B} = \mathrm{diag}(B_k)$, $\mathcal{C} = \mathrm{diag}(C_k)$, $\mathcal{D} = \mathrm{diag}(D_k)$ are all block diagonals and $\mathcal{Z}$ is the block shift operator.

To illustrate the adaptations of the algorithms which we shall need in this situation, we consider again the Nevanlinna-Pick algorithm in its time varying appearance. First we reformulate the scalar time invariant case in a special way to see the generalization. This can be described as follows: given $\alpha_k \in \mathbb{D}$ and function values $w_k \in \mathbb{C}$ for $k = 1, 2, \ldots, N$, find all functions f analytic in $\mathbb{D}$ such that

$$\begin{array}{l} (1)\ f(\alpha_k) = w_k, \quad k = 1, \ldots, N \\ (2)\ \sup\limits_{z\in\mathbb{D}} |f(z)| < 1 \end{array}$$

or if we define the lower triangular Toeplitz matrix

$$T = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & f_0 & 0 & 0 & \ddots \\ \ddots & f_1 & \boxed{f_0} & 0 & \ddots \\ \ddots & f_2 & f_1 & f_0 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

then this is equivalent with finding the lower triangular matrix T such that

$$\begin{aligned} &(1)\ \mathrm{diag}((I - Z_k)^{-1}T) = W_k, \quad k = 1, \ldots, N \\ &(2)\ \|T\| < 1 \end{aligned}$$

where $Z_k = \alpha_k Z^H$ ($Z =$ is the shift operator) and $W_k = w_k I$. Note that

$$Z_k = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & \\ \cdots & 0 & \boxed{0} & \alpha_k & 0 & \cdots \\ & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

so that

$$(I - Z_k)^{-1} = I + Z_k + Z_k^2 + \cdots = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & & \\ \cdots & 0 & \boxed{1} & \alpha_k & \alpha_k^2 & \cdots \\ & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$

Thus $\mathrm{diag}((I - Z_k)^{-1}T) = f(\alpha_k)I$ is like evaluating the function f in the point α_k.

The last formulation is precisely what is meant by the time varying Nevanlinna-Pick problem except that the matrices T, Z_k and W_k now have the following meaning

$$T = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & f_{-1,-1} & 0 & 0 & \ddots \\ \ddots & f_{0,-1} & \boxed{f_{0,0}} & 0 & \ddots \\ \ddots & f_{1,-1} & f_{1,0} & f_{1,1} & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}, \qquad Z_k = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & 0 & \alpha_{-1}^{(k)} & 0 & \ddots \\ \ddots & 0 & \boxed{0} & \alpha_0^{(k)} & \ddots \\ \ddots & 0 & 0 & 0 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

and $W_k = \mathrm{diag}(\ldots w_{-1}^{(k)} \boxed{w_0^{(k)}} w_1^{(k)} \ldots)$. The basic idea of the theory developed here is that in the time invariant case, the scalar values are identified with a (constant) diagonal in an infinite dimensional operator, while in the time varying case one still works with such diagonals, but now the diagonals contain infinitely many different values.

This theory is still in full expansion and one may think of systems which have a time varying low displacement rank, or time varying directional interpolation problems etc. The discovery of the time varying ORFs behind all this is still unclear for the moment. [36, 31]

References

[1] R. Ackner, H. Lev-Ari, and T. Kailath. The Schur algorithm for matrix-valued meromorphic functions. *SIAM J. Matr. Anal. Appl.*, 15:140–150, 1994.

[2] N.I. Akhiezer. *The classical moment problem.* Oliver and Boyd, Edinburgh, 1969. Originally published Moscow, 1961.

[3] M. Bakonyi and T. Constantinescu. *Schur's algorithm and several applications*, volume 261 of *Pitman Research Notes in Mathematics*. Longman, Harlow, UK, 1992.

[4] T. Bloom, D.S. Lubinsky, and H. Stahl. Interpolatory integration rules and orthogonal polynomials with varying weight. *Numerical Algorithms*, 3(1-4):55–65, 1992.

[5] A. Bultheel. Orthogonal matrix functions related to the multivariable Nevanlinna-Pick problem. *Tijdschr. Belgisch Wisk. Genootschap, ser. B*, 32(2):149–170, 1980.

[6] A. Bultheel. Inequalities in Hilbert modules of matrix valued functions. *Proc. American Mathematical Society.*, 85(3):369–372, 1982.

[7] A. Bultheel. On the ill-conditioning of locating the transmission zeros in least squares ARMA filtering. *J. Comput. Appl. Math.*, 11(1):103–118, 1984.

[8] A. Bultheel and P. Dewilde. Orthogonal functions related to the Nevanlinna-Pick problem. In P. Dewilde, editor, *Proc. 4th Int. Conf. on Math. Theory of Networks and Systems at Delft*, pages 207–212, North-Hollywood, 1979. Western Periodicals.

[9] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. A Szegő theory for rational functions. Technical Report TW131, Department of Computer Science, K.U. Leuven, May 1990.

[10] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonal rational functions similar to Szegő polynomials. In C. Brezinski, L. Gori, and A. Ronveaux, editors, *Orthogonal polynomials and their applications*, volume 9 of *IMACS annals on computing and applied mathematics*, pages 195–204, Basel, 1991. J.C. Baltzer AG.

[11] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. The computation of orthogonal rational functions and their interpolating properties. *Numer. Alg.*, 2(1):85–114, 1992.

[12] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Convergence of modified approximants associated with orthogonal rational functions. *J. Comput. Appl. Math.*, 1992. Accepted, to appear Jan. 1995.

[13] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. A Favard theorem for orthogonal rational functions on the unit circle. *Numer. Alg.*, 3:81–89, 1992.

[14] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Favard theorem for reproducing kernels. *J. Comput. Appl. Math.*, 1992. Accepted, to appear Jan. 1995.

[15] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. A moment problem associated to rational Szegő functions. *Numer. Alg.*, 3:91–104, 1992.

[16] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonal rational functions and quadrature on the unit circle. *Numer. Alg.*, 3:105–116, 1992.

[17] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Quadrature formulas on the unit circle based on rational functions. *J. Comput. Appl. Math.*, 1992. To appear.

[18] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Asymptotics for orthogonal rational functions. *Trans. Amer. Math. Soc.*, 1993. Accepted.

[19] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Convergence of orthogonal rational functions. In *Computational methods and function theory, Malaysia, March 1994*, 1993. Submitted.

[20] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Moment problems and orthogonal functions. *J. Comput. Appl. Math.*, 48:49–68, 1993.

[21] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonal rational functions for a quasi-Toeplitz moment matrix. 1993. Manuscript, presented at the 2nd Int. Conf. Numer. Anal., Plovdiv, Bulg., 1993.

[22] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonal rational functions on the unit circle and the real line. In P. Dewilde, M.A. Kaashoek, and M. Verhaegen, editors, *Challenges of a generalized system theory*, volume 40 of *Essays of the Dutch Academy of Sciences, Sect. Physics, First series*, pages 241–252, Amsterdam, 1993. North Holland.

[23] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. On convergence of multipoint Padé approximants associated to the unit circle. *J. Comp. Appl. Math.*, 1994. Submitted.

[24] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonal rational functions with poles on the unit circle. *J. Math. Anal. Appl.*, 182:221–243, 1994.

[25] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonal rational functions and modified approximants. *Annals of Numerical Analysis*, 1994. Submitted, Proc. Luminy, Sept. 1994.

[26] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Orthogonality and boundary interpolation. In A.M. Cuyt, editor, *Nonlinear numerical methods and rational approximation II*, pages 37–48. Kluwer, 1994. To appear.

[27] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. Recurrence relations for orthogonal functions. In S.C. Cooper and W.J. Thron, editors, *Continued Fractions and Orthogonal Functions*, volume 154 of *Lecture Notes in Pure and Applied Mathematics Series*, pages 24–46. Marcel Dekker, 1994.

[28] A. Bultheel and M. Van Barel. Linear prediction: mathematics and engineering. *Bulletin of the Belgian Mathematical Society Simon Stevin*, 1:1–58, 1994.

[29] Ph. Delsarte, Y. Genin, and Y. Kamp. The Nevanlinna-Pick problem for matrix-valued functions. *SIAM J. Appl. Math.*, 36:47–61, 1979.

[30] Ph. Delsarte, Y. Genin, and Y. Kamp. On the role of the Nevanlinna-Pick problem in circuit and system theory. *Int. J. Circuit Th. Appl.*, 9:177–187, 1981.

[31] P. Dewilde, M.A. Kaashoek, and M. Verhaegen, editors. *Challenges of a generalized, non-stationary system theory.* Essays of the Dutch Academy of Arts and Sciences. Dutch Acad. Arts Sci., Amsterdam, 1993.

[32] M.M. Djrbashian. A survey on the theory of orthogonal systems and some open problems. In P. Nevai, editor, *Orthogonal polynomials: Theory and practice*, volume 294 of *Series C: Mathematical and Physical Sciences*, pages 135–146, Boston, 1990. NATO-ASI, Kluwer Academic Publishers.

[33] C. Foias and A. Frazho. *The commutant lifting approach to interpolation problems*, volume 44 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1990.

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

[35] I. Gohberg, editor. *I. Schur methods in operator theory and signal processing*, volume 18 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, Basel, 1986.

[36] I. Gohberg, editor. *Time-variant systems and interpolation*, volume 56 of *Operator Theory, Advances and Applications*. Birkhäuser Verlag, 1992.

[37] C.K. Koç and G.R. Chen. A fast algorithm for scalar Nevanlinna-Pick interpolation. *Numerische Math.*, 64(1):115–126, 1993.

[38] E. Landau, editor. *Outils et modèles mathématiques pour l'automatique, l'analyse de systèmes et le traitement du signal, Vol. 2*. Editions du CNRS, Paris, 1982.

[39] X. Li and K. Pan. Strong and weak convergence of rational functions orthogonal on the unit circle. 1993. Manuscript.

[40] G.L. López. 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 Mathematics*, pages 255–260. Springer Verlag, 1988.

[41] G.L. López. Asymptotics of polynomials orthogonal with respect to varying measures. *Constructive Approximation*, 5:199–219, 1989.

[42] K. Pan. Asymptotics for orthogonal systems of rational functions on the unit circle with a given set of poles. *J. Approx. Th.*, 1991. Submitted.

[43] K. Pan. On characerization theorems for measures associated with orthogonal systems of rational functions on the unit circle. *Approximation Theory*, 70:262–272, 1992.

[44] K. Pan. On orthogonal systems of rational functions on the unit circle and polynomials orthogonal with respect to varying measures. *J. Comp. Appl. Math.*, 47(3):313–322, 1993.

[45] K. Pan. On the convergence of rational interpolation approximant of Carathéodory functions. *Journ. Comp. Appl. Math.*, 1993. Submitted.

[46] K. Pan. Strong and weak convergence of orthogonal systems of rational functions on the unit circle. *J. Comp. Appl. Math.*, 46:427–436, 1993.

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

[48] A.H. Sayed. *Displacement structure in signal processing and mathematics*. PhD thesis, Stanford University, August 1992.

[49] J. L. Walsh. *Interpolation and approximation*, volume 20 of *AMS Colloquium publications*. American Mathematical Society, Providence, Rhode Island, 3rd edition, 1960. First edition 1935.

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

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