Orthogonal Rational Functions
and
non-stationary stochastic processes: a Szegö Theory

Laurent Baratchart

(based on joint work with L. Golinskii, S. Kupin, M. Olivi and V. Lunot)

Abstract— **We present a generalization of Szegö theory of orthogonal polynomials on the unit circle to orthogonal rational functions. Unlike previous results, the poles of the rational functions may tend to the unit circle under smoothness assumptions on the density of the measure. Just like the Kolmogorov-Krein-Szegö theorem may be interpreted as an asymptotic estimate of the prediction error for stationary stochastic processes, the present theory yields an asymptotic estimate of the prediction error for certain, possibly nonstationary, stochastic processes. The latter admit a spectral calculus where the time-shift corresponds to multiplication by elementary Blaschke products of degree 1 (that reduce to multiplication by the independent variable in the stationnary case). When the poles of the best predictor tend to a point on the unit circle where the spectral density is nonzero, the prediction error goes to zero,** *i.e.* **the process is asymptotically deterministic.**

I. INTRODUCTION

The theory of orthogonal polynomials on the unit circle is a most classical piece of analysis [14] which is still the object of intensive studies; we refer the reader to [13] for an exposition of many recent developments in the field. The asymptotic behaviour of orthogonal polynomials is of special interest for many issues pertaining to approximation theory and the spectral theory of differential operators. Its connection with prediction theory of stationary stochastic processes has long been known [9]. Namely, the n-th orthonormal polynomial with respect to the spectral measure of the process yields the optimal regression coefficients of a linear one-step ahead predictor from the $n-1$-st last values, in the sense of minimum variance of the error. Likewise, the (inverse of) the dominant coefficient of the polynomial gives the prediction error. In particular, asymptotics for the dominant coefficient determine the asymptotically optimal prediction error from the past as time goes large.

As compared to orthogonal polynomials, orthogonal rational functions have not been much considered. They were apparently introduced by Dzrbasjan [7] and studied in particular by Pan [12], but the first systematic exposition seems to be the monograph by Bultheel et al. [5] which contains additional references. The emphasis has been more on the algebraic side of the theory and asymptotic analysis of orthogonal rational functions is still in its infancy. A connection with prediction theory of stationary processes

L. Baratchart is with INRIA Sophia-Antipolis-Méditerrannée, BP 93, 2004 route des Lucioles, 06902 Sophia-Antiolis Cedex, FRANCE. `baratcha@sophia.inria.fr`

is advocated in [5], but stops short of involvement in the subject.

The purpose of the present article is twofold. On the one hand, we present an analog of the Kolmogorov-Krein-Szegö theorem for orthogonal rational functions which is first of its kind in that it allows for the poles of these functions to approach the unit circle, generalizing the previously known results for compactly supported singular set [5]. On the other hand, we show that orthogonal rational functions are suited for prediction theory of certain, possibly *nonstationary* stochastic processes that we call *Blaschke varying* processes. These are characterized by a spectral calculus where time shift corresponds to multiplication by an elementary Blaschke product (that may depend on the time instant considered). This class of processes contains the familiar Gaussian stationary processes, but it contains many more that exhibit a much more varied behaviour. For instance, the process may be asymptotically deterministic along certain subsequences and nondeterministic along others.

We confine ourselves to scalar (complex-valued) processes. Most proofs are omitted and will appear elesewhere [3], [2].

II. ORTHOGONAL RATIONAL FUNCTIONS

A. BASIC DEFINITIONS

Set $\mathbb{D}$ to be the open unit disk and $\mathbb{T}$ the unit circle. Let (α_k), for $k \in \mathbb{N}$, be a sequence of points in $\mathbb{D}$. We set $\alpha_0 = 0$ by convention. Define the elementary Blaschke factor ζ_k as

$$\zeta_k(z) = \frac{z - \alpha_k}{1 - \bar{\alpha}_k z}. \tag{1}$$

Consider the "partial" Blaschke products $\mathcal{B}_k$ given by

$$\mathcal{B}_0(z) = 1, \quad \mathcal{B}_k(z) = \mathcal{B}_{k-1}(z)\zeta_k(z), \tag{2}$$

where ζ_k is given by (1) and $k \geq 1$. We also put

$$\mathcal{B}_{n,i} = \prod_{k=i}^{n} \zeta_k, \quad 1 \leq i \leq n, \tag{3}$$

so that $\mathcal{B}_{n,1} = \mathcal{B}_n$ for $n \geq 1$.

The functions $\{\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_n\}$ span the space

$$\mathcal{L}_n = \left\{ \frac{p_n}{\pi_n} : \pi_n(z) = \prod_{k=1}^{n}(1 - \bar{\alpha}_k z), \ p_n \in \mathcal{P}_n \right\}, \tag{4}$$

where $\mathcal{P}_n$ stands for the space of algebraic polynomials of degree at most n. In the classical case, that is when $\alpha_k = 0$ for all k, $\mathcal{L}_n$ coincides with $\mathcal{P}_n$.

Given a function g, we introduce its parahermitian conjugate g_* defined by $g_*(z) = \overline{g(1/\bar{z})}$. Observe that $|g_*| = |g|$ on $\mathbb{T}$ and that $\zeta_{n*} = \zeta_n^{-1}$, $\mathcal{B}_{k*} = \mathcal{B}_k^{-1}$. For $g \in \mathcal{L}_n$, we set $g^* = \mathcal{B}_n f_*$; clearly, $g^* \in \mathcal{L}_n$. Beware the upper and lower positioning for the star that have different meanings: $g^* \neq g_*$. Clearly $|g^*| = |g|$ on $\mathbb{T}$.

Each $g \in \mathcal{L}_n$ can be uniquely decomposed in the form

$$g = a_n \mathcal{B}_n + a_{n-1} \mathcal{B}_{n-1} + \cdots + a_1 \mathcal{B}_1 + a_0,$$

and then

$$g^* = \bar{a}_0 \mathcal{B}_{n,1} + \bar{a}_1 \mathcal{B}_{n,2} + \cdots + \bar{a}_{n-2} \mathcal{B}_{n,n-1} + \bar{a}_{n-1} \mathcal{B}_{n,n} + \bar{a}_n.$$

It is plain that $a_n = \overline{g^*(\alpha_n)}$ and $a_0 = g(\alpha_1)$.

Now, let μ be a probability measure on $\mathbb{T}$ with infinite support, and consider $\mathcal{L}_n$ as a subspace of $L^2(\mu)$. Let $(\phi_k)_{0 \leq k \leq n}$ be an orthonormal basis for $\mathcal{L}_n$ such that $\phi_0 = 1$ and $\phi_k \in \mathcal{L}_k \setminus \mathcal{L}_{k-1}$. Such a basis is easily obtained on applying the Gram-Schmidt orthonormalization process to $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_n$, and ϕ_n is unique up to a multiplicative unimodular constant. For definiteness, we normalize it so that $g^*(\alpha_n) = \overline{g^*(\alpha_n)} > 0$:

$$\phi_n = \kappa_n \mathcal{B}_n + a_{n,n-1} \mathcal{B}_{n-1} + \ldots + a_{n,1} \mathcal{B}_1 + a_{n,0} \mathcal{B}_0, \quad (5)$$

where $\kappa_n > 0$. To recap, ϕ_n is determined by the following three requirements:

- $\phi_n \in \mathcal{L}_n \setminus \mathcal{L}_{n-1}$ for $n \geq 1$;
- $\phi_n^*(\alpha_n) > 0$;
- $\int_{\mathbb{T}} \phi_n \bar{\phi}_m d\mu = \delta_{nm}$, where δ_{nm} is the Kronecker symbol.

Definition 2.1: The functions (ϕ_k) are called the orthogonal rational functions (of the first kind) associated with (α_k) and μ.

Generically, the dependence on the nodes (α_k) and the measure μ will be omitted. The words "orthogonal rational function" will be abbreviated as ORF or OR-function.

In the classical case where $(\alpha_k) \equiv (0)$, the orthogonal rational functions (ϕ_n) defined in (5) reduce to the familiar orthonormal polynomials [14]. The above definitions and the recurrence relations given below are presented with greater detail in [5], see also [11].

B. RECURRENCE RELATIONS

Clearly evaluation at $w \in \mathbb{D}$ is a continuous functional on $\mathcal{L}_n$ endowed with the $L^2(\mu)$-metric. Consequently $\mathcal{L}_n$ is a reproducing kernel Hilbert space. Let $k_n(., w)$ denote the reproducing kernel, *i.e.* the unique function in $\mathcal{L}_n$ such that

$$g(w) = \langle f, k_n(., w) \rangle_{L^2(\mu)}, \quad \forall g \in \mathcal{L}_n,$$

where $\langle ., . \rangle_{L^2(\mu)}$ indicates the $L^2(\mu)$ scalar product. It is a general fact from the theory that

$$k_n(z, w) = \sum_{k=0}^{n} e_n(z) \overline{e_n(w)} \quad (6)$$

for any orthonormal basis $(e_k)_{0 \leq k \leq n}$ of $\mathcal{L}_n$. Using this with $e_n = \phi_n$ and $e_n = \phi_n^*$ successively leads one after a short computation to the Christoffel-Darboux formulas:

$$k_{n-1}(z, w) = \frac{\phi_n^*(z) \overline{\phi_n^*(w)} - \phi_n(z) \overline{\phi_n(w)}}{1 - \zeta_n(z) \overline{\zeta_n(w)}}, \quad (7)$$

$$k_n(z, w) = \frac{\phi_n^*(z) \overline{\phi_n^*(w)} - \zeta_n(z) \overline{\zeta_n(w)} \phi_n(z) \overline{\phi_n(w)}}{1 - \zeta_n(z) \overline{\zeta_n(w)}}. \quad (8)$$

By induction, (7) and (8) quickly yields the following recurrence relations for orthogonal rational functions:

Theorem 2.2 ([5], Theorem 4.1.1): For $n \geq 1$, it holds that

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

where

$$T_n(z) = \sqrt{\frac{1 - |\alpha_n|^2}{1 - |\alpha_{n-1}|^2}} \frac{1}{\sqrt{1 - |\tilde{\gamma}_n|^2}} \frac{1 - \bar{\alpha}_{n-1} z}{1 - \bar{\alpha}_n z} \quad (9)$$

$$\times \begin{bmatrix} 1 & -\overline{\tilde{\gamma}_n} \\ -\tilde{\gamma}_n & 1 \end{bmatrix} \quad (10)$$

$$\times \begin{bmatrix} \lambda_n & 0 \\ 0 & \bar{\lambda}_n \end{bmatrix} \begin{bmatrix} \zeta_{n-1}(z) & 0 \\ 0 & 1 \end{bmatrix},$$

and

$$\tilde{\gamma}_n = -\frac{\overline{\phi_n(\alpha_{n-1})}}{\overline{\phi_n^*(\alpha_{n-1})}}, \qquad \eta_n = \frac{1 - \alpha_n \bar{\alpha}_{n-1}}{1 - \bar{\alpha}_n \alpha_{n-1}}, \quad (11)$$

$$\lambda_n = \frac{|1 - \bar{\alpha}_n \alpha_{n-1}|}{1 - \alpha_n \bar{\alpha}_{n-1}} \frac{\overline{\phi_n^*(\alpha_{n-1})}}{|\phi_n^*(\alpha_{n-1})|} \eta_n. \quad (12)$$

$(E|X_n - \hat{X}_n|^2)^{1/2}$ is $1/\kappa_n$

In the classical case where $(\alpha_k) \equiv (0)$, Theorem 2.2 reduces to the standard recurrence formulas for orthogonal polynomials on the unit circle [14]. From the theorem, it is easily checked by induction that ϕ_n has all its zeros inside $\mathbb{D}$. Thus ϕ_n^* has all its zeros outside $\overline{\mathbb{D}}$, for the zeros of ϕ_n^* are reflected from those of ϕ_n across $\mathbb{T}$..

III. AN ANALOG OF SZEGÖ THEORY

A. HARDY SPACES

Let m denote normalized Lebesgue measure on $\mathbb{T}$, so that $dm(t) = dt/(2\pi i t) = \frac{1}{2\pi} d\theta$ where $t = e^{i\theta} \in \mathbb{T}$. We write $L^p(\mathbb{T}) = L^p(\mathbb{T}, m)$ for the familiar Lebesgue spaces on $\mathbb{T}$. Whenever $E \subset \mathbb{T}$, the notation $C(E)$ indicates the space of complex continuous functions on E. We generically use the symbol $\mathcal{O}$, like in $\mathcal{O}(E)$, to designate an open neighborhood of E in $\mathbb{T}$.

The Hardy space H^2 of the unit disk is the closed subspace of $L^2(\mathbb{T})$ consisting of functions whose Fourier coefficients of strictly negative index do vanish. These are the nontangential limits of those functions f holomorphic in the unit disk having uniformly bounded L^2 means over all circles centered at 0 of radius less than 1:

$$\sup_{0 < r < 1} \int_0^{2\pi} |f(re^{i\theta})|^2 d\theta < +\infty, \quad (13)$$

and the above supremum is the squared $L^2(\mathbb{T})$-norm of the nontangential limit of f. The correspondence between the functions and their nontangential limits is one-to-one and, identifying them, we alternatively regard members of H^2 as holomorphic functions in the variable $z \in \mathbb{D}$. In fact, the extension to $\mathbb{D}$ is obtained from the values on $\mathbb{T}$ through a Cauchy as well as a Poisson integral [8]. The space H^∞ consists of those bounded holomorphic functions in $\mathbb{D}$ endowed with the *sup* norm. The nontangential limit of a H^∞-function lies in $L^\infty(\mathbb{T})$ and its norm is the supremum of the modulus of the function on $\mathbb{D}$. The disk algebra $A(\mathbb{D})$ is comprised of H^∞-functions that are continuous on $\overline{\mathbb{D}}$.

It is well-known [8] that a nonzero $f \in H^2$ can be uniquely factored as $f = jw$ where

$$w(z) = \exp\left\{ \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \log|f(e^{i\theta})| \, d\theta \right\} \quad (14)$$

belongs to H^2 and is called the *outer factor* of f, while $j \in H^\infty$ has modulus 1 a.e. on $\mathbb{T}$ and is called the *inner factor* of f. The latter may be further decomposed as $j = bS_\mu$, where

$$b(z) = e^{i\theta_0} z^k \prod_{z_l \neq 0} \frac{-\overline{z}_l}{|z_l|} \frac{z - z_l}{1 - \overline{z}_l z} \quad (15)$$

is the *Blaschke product*, with order $k \geq 0$ at the origin, associated to the sequence $z_l \in \mathbb{D} \setminus \{0\}$ and to the constant $e^{i\theta_0} \in \mathbf{T}$, while

$$S_\mu(z) = \exp\left\{ -\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \, d\mu(\theta) \right\} \quad (16)$$

is the *singular inner factor* associated with μ, a positive measure on $\mathbb{T}$ which is singular with respect to Lebesgue measure. The z_l are of course the zeros of f in $\mathbb{D} \setminus \{0\}$, counted with their multiplicities, while k is the order of the zero at 0. If there are infinitely many zeros, the convergence of the product $b(z)$ in $\mathbb{D}$ is ensured by the condition $\sum_l (1 - |z_l|) < \infty$ which holds automatically when $f \in H^2 \setminus \{0\}$. If there are only finitely many z_l, we say that (15) is a finite Blaschke product; note that a finite Blaschke product may alternatively be defined as a rational function of the form q/q^R, where q is an algebraic polynomial whose roots lie in $\mathbb{D}$ and q^R indicates the *reciprocal polynomial* given by $q^R(z) = z^n \overline{q(1/\overline{z})}$ if n is the degree of q. The integer n is also called the degree of the Blaschke product.

That $w(z)$ in (14) is well-defined rests on the fact that $\log|f| \in L^1$ if $f \in H^2 \setminus \{0\}$; this also entails that a H^2 function cannot vanish on a subset of strictly positive Lebesgue measure on $\mathbb{T}$ unless it is identically zero. For simplicity, we say that a function is outer (resp. inner) if it is equal to its outer (resp. inner) factor.

For $(\alpha_k)_{k \in \mathbb{N}}$ an infinite sequence in $\mathbb{D}$, we let $Acc(\alpha_k) = \overline{(\alpha_k)} \setminus (\alpha_k)$ be the set of its accumulation points, which is a closed subset of $\overline{\mathbb{D}}$. In connection with the asymptotic behaviour of ORFs introduced in section II-A, to be considered momentarily, we shall be interested in the case where the sequence (α_k) is hyperbolically non-separated, that is:

$$\sum_k (1 - |\alpha_k|) = +\infty. \quad (17)$$

Condition (17) means that the sequence of Blaschke products $\mathcal{B}_k$, defined in (2), diverges to zero locally uniformly in $\mathbb{D}$ or, equivalently, that no nonzero H^2-function can vanish at all α_k. Of great importance in this context is also the equivalence of (17) with the density of rational functions having poles at the points $(1/\overline{\alpha}_k)$ in H^2, as well as in $A(\mathbb{D})$ [1, App. A].

B. A SZEGÖ-TYPE THEOREM

We put $\mathcal{M}(\mathbb{T})$ to designate the set of Borel probability measures on $\mathbb{T}$. For $\mu \in \mathcal{M}(\mathbb{T})$, let μ_{ac} and μ_s be its absolutely continuous and singular components. We indicate by μ' the density of μ_{ac} with respect to m, so that $d\mu = d\mu_{ac} + d\mu_s = \mu' dm + d\mu_s$.

Recall that a measure μ is called *Szegő* (notation: $\mu \in$ (S)) if, and only if, $\log \mu' \in L^1(\mathbb{T})$. To $\mu \in$ (S), we associated Szegő function S is

$$S(z) = S[\mu](z) := \exp\left(\frac{1}{2} \int_\mathbb{T} \frac{t + z}{t - z} \log \mu' dm(t) \right). \quad (18)$$

The function S given by (18) is the outer function in H^2 such that $|S|^2 = \mu'$ a.e. on $\mathbb{T}$, normalized so that $S(0) > 0$.

The following assumptions on a measure μ will play an important role in connection with the asymptotic behaviour of ORFs:

$$\mu' \in \mathcal{C}(\mathcal{O}(Acc(\alpha_k) \cap \mathbb{T})), \quad (19)$$

$$\mu' > 0 \quad \text{on} \quad \mathcal{O}(Acc(\alpha_k) \cap \mathbb{T}), \quad (20)$$

$$\{Acc(\alpha_k) \cap \mathbb{T}\} \subset \mathbb{T} \setminus \text{supp}\, \mu_s, \quad (21)$$

where $\text{supp}\, \mu_s$ is the *closed* support of μ_s. When the sequence (α_k) accumulates nontangentially on $Acc(\alpha_k) \cap \mathbb{T}$, meaning that every convergent subsequence to $\xi \in \mathbb{T}$ tends to the latter nontangentially, then (19)-(20) can be substituted with a slightly property, namely:

$$\mu' \text{ is upper semicontinuous on } \mathcal{O}(Acc(\alpha_k) \cap \mathbb{T})$$

$$\text{with } 0 < \delta < \mu' < M < \infty \text{ on } \mathcal{O}(Acc(\alpha_k) \cap \mathbb{T}), \quad (22)$$

and each $\xi \in Acc(\alpha_k) \cap \mathbb{T}$ is a Lebesgue point of $\log \mu'$.

The next theorem and its corollaries shed light on the asymptotic behaviour of ORFs when the α_n do not tend too fast to the unit circle, and only accumulate there at points where the density μ' is regular and positive:

Theorem 3.1: Let (17), (19)-(21) be in force, and $\mu \in$ (S). Then

$$\lim_n |\phi_n^*(\alpha_n)|^2 |S(\alpha_n)|^2 (1 - |\alpha_n|^2) = 1. \quad (23)$$

If (α_k) accumulates nontangentially on $Acc(\alpha_k) \cap \mathbb{T}$, then it is enough to assume instead of (19)-(20) that (22) holds.

Remark 3.2: Theorem 3.1 deals only with densities μ' that are sufficiently smooth and nonvanishing at the accumulation points that (α_k) may have on the circle. It would be very interesting to know how much these assumptions can be relaxed. In particular, [3] gives an example where (23) holds although (20) fails.

In the classical case where $(\alpha_k) \equiv (0)$, Theorem 3.1 reduces to the well-known Kolmogorov-Krein-Szegö theorem for

polynomials orthonormal with respect to a Szegö measure [8, thm 3.1] (assumptions (17) and (19)-(21) are void in this case). When (α_k) is compactly embedded in $\mathbb{D}$, namely if no subsequence converges to a point on the unit circle, the result is [5, thm. 9.6.9]. The generalization stated here can be found in [3]. Note the following corollaries.

Corollary 3.3: Let (17), (19)-(21) be satisfied and $\mu \in$ (S). Then

$$\lim_n \left\| S\phi_n^*(z) - \beta_n \frac{\sqrt{1-|\alpha_n|^2}}{1-\overline{\alpha}_n z} \right\|_{L^2(\mathbb{T})} = 0, \quad (24)$$

where $\beta_n = (S\phi_n^*)(\alpha_n)/|(S\phi_n^*)(\alpha_n)|$. If (α_k) accumulates nontangentially on $Acc(\alpha_k) \cap \mathbb{T}$, then it is enough to assume instead of (19)-(20) that (22) holds.

Proof: Estimating the integral, we get

$$\left\| S\phi_n^*(z) - \beta_n \frac{\sqrt{1-|\alpha_n|^2}}{1-\overline{\alpha}_n z} \right\|^2$$

$$= \int_{\mathbb{T}} \left| S\phi_n^*(z) - \beta_n \frac{\sqrt{1-|\alpha_n|^2}}{1-\overline{\alpha}_n z} \right|^2 dm(z)$$

$$\leq \|\phi_n\|_\mu^2 - 2\mathrm{Re}\left(\frac{\overline{\beta}_n}{2i\pi} \int_{\mathbb{T}} S\phi_n^*(z) \frac{\sqrt{1-|\alpha_n|^2}}{z - \alpha_n} dz \right) + 1$$

$$= 2(1 - \sqrt{1-|\alpha_n|^2} \, |S(\alpha_n)| \, |\phi_n^*(\alpha_n)|),$$

and Theorem 3.1 yields (24). ∎

Corollary 3.4: Let (17), (19)-(21) be satisfied and $\mu \in$ (S). If $\alpha_{m(n)}$ is a subsequence of α_n such that $\lim_{n\to+\infty} |\alpha_{m(n)}| = 1$, then

$$\lim_{n\to+\infty} |\phi_{m(n)}^*(\alpha_{m(n)})| = +\infty. \quad (25)$$

More precisely, there are constants c, C such that

$$0 < c \leq |\phi_{m(n)}^*(\alpha_{m(n)})|(1 - |\alpha_{m(n)}|)^{-1/2} \leq C. \quad (26)$$

If (α_k) accumulates nontangentially on $Acc(\alpha_k) \cap \mathbb{T}$, then it is enough to assume instead of (19)-(20) that (22) holds.

Proof: For $f \in L^1(\mathbb{T})$, we let $P_z f$ designate the harmonic (Poisson) extension of f to $\mathbb{D}$ evaluated at $z \in \mathbb{D}$. From (14) we see that $|S(\alpha_n)| = \exp(P_{\alpha_n} \log \mu')$. Extracting a subsequence from $(\alpha_{m(n)})$ if necessary (we continue to denote this subsequence by $(\alpha_{m(n)})$ for simplicity), we may assume that $(\alpha_{m(n)})$ converges to $\xi \in Acc(\alpha_k) \cap \mathbb{T}$. Then, by standard properties of Poisson integrals if (19)-(20) hold or using the Fatou theorem in case (α_k) converges nontangentially to ξ, and (22) is in force, we get that $|S(\alpha_{m(n)})| \to \mu'(\xi)$. It is now immediate from (23) that $|\phi_{m(n)}^*(\alpha_{m(n)})| \sim (1 - |\alpha_{m(n)}|)^{-1/2}/(\sqrt{2}\mu'(\xi))$. By our assumptions, there are strictly positive constants δ, M such that $\delta < \mu'(\xi) < M$, hence (26) holds. ∎

Remark 3.5: Because $S \in H^2$, it is a result of Hardy and Littlewood [6, thm. 5.9] that $|S(z)| = o(1/(1 - |z|)^{1/2})$. Therefore (23) implies (40), no matter whether (19)-(20) or (22) do hold (compare Remark 3.2).

Corollary 3.6: Let (17) be satisfied and $\mu \in$ (S). If $\alpha_{m(n)}$ is a subsequence of α_n such that $\lim_{n\to+\infty} \alpha_{m(n)} = \alpha \in \mathbb{D}$, then

$$\lim_{n\to+\infty} |\phi_{m(n)}^*(\alpha_{m(n)})| = |S(\alpha)|^{-1}(1 - |\alpha|^2)^{-1/2}. \quad (27)$$

Proof: This is obvious from (23). ∎

IV. GENERALIZED MOMENT PROBLEMS

The theory of generalized moments is a far-reaching extension of the classical theory of polynomial moments. It was expounded in the monograph [10, Ch. 1-3] to which we refer the reader for a comprehensive discussion of the subject and many deep applications. Chapter 10 of [5] develops a special case of this topic in connection with orthogonal rational functions (without reference to [10], however). In the next section, we shall develop another special case aiming at the spectral calculus of certain nonstationary stochastic processes, of which ORFs can be regarded as optimal predictors. For the time being, we recall in this section some basic facts from the general theory and taylor them to our needs.

Let $\mathfrak{W} = \{w_k\}_{k=0,\ldots,\infty}$ be a system of continuous linearly independent complex-valued functions on $\mathbb{T}$. We denote by $\mathcal{L}_{n,\mathfrak{W}}$ the linear span of $w_0, \ldots, w_n$. We further define $w_{-k} = \bar{w}_k$, and we assume that the sum $\mathcal{L}_{n,\mathfrak{W}} \cap \bar{\mathcal{L}}_{n,\mathfrak{W}} = \{0\}$. Functions in $\mathcal{L}_{n,\mathfrak{W}}$ may be thought of as generalized analytic polynomials of degree at most n; in this vein, one can regard the space $\bar{\mathcal{L}}_{n,\mathfrak{W}} + \mathcal{L}_{n,\mathfrak{W}}$ as being comprised of generalized trigonometric polynomials of degree at most n.

A sequence of complex numbers $\{c_k\}_{k=0,\ldots,\infty}$ is called *positive* (w.r.t. to $\mathfrak{W}$) if

$$\left(\sum_{k=0}^n (a_k w_k + \overline{a_k w_k}) \geq 0 \right) \implies \left(\sum_{k=0}^n (a_k c_k + \overline{a_k c_k}) \geq 0 \right). \quad (28)$$

for all n and all complex numbers $a_0, \cdots, a_n$. To avoid trivial cases, we assume that $\mathfrak{W}$ is such that to each n there is at least one $n+1$-tuple $\mathfrak{a}_0, \cdots, \mathfrak{a}_n$ for which

$$\sum_{k=0}^n (\mathfrak{a}_k w_k + \overline{\mathfrak{a}_k w_k}) > 0. \quad (29)$$

The theorem below is the cornerstone of the additive part of generalized moment theory.

Theorem 4.1 ([10], Ch. 3, Theorem 1.3): Let the system of functions $\mathfrak{W}$ as above satisfy (29) for some coefficients $\{\mathfrak{a}_k\}, a_k \in \mathbb{C}$. Given a sequence $(c_k)_{k\in\mathbb{N}}$ of complex numbers, the following assertions on are equivalent:

- The sequence (c_k) is positive in the sense of (28),
- The c_k are generalized moments (w.r.t. to $\mathfrak{W}$), that is, there exists a measure $\sigma \in \mathcal{M}(\mathbb{T})$ with the property that

$$c_k = \int_{\mathbb{T}} w_k d\sigma. \quad (30)$$

We now bring some multiplicative structure into play.

Namely, for $\mathfrak{W} = \{w_k\}$ a system of functions as above satisfying (29), we shall suppose in addition that the following properties hold:

1) $w_0 = 1$,

2) One has $\bar{w}_j w_k \in \bar{\mathcal{L}}_{j,\mathfrak{W}} + \mathcal{L}_{k,\mathfrak{W}}$, or, equivalently,

$$\bar{w}_j w_k = \sum_{s=-j}^{k} \beta_{jk,s} w_s, \qquad (31)$$

where $\{\beta_{jk,s}\}_s$ are complex coefficients and, as ususal, $w_{-s} = \bar{w}_s$.

3) If for some n and some coefficients a_k it holds that

$$\sum_{k=0}^{n} (a_k w_k + \overline{a_k w_k}) \geq 0 \quad \text{on } \mathbb{T},$$

then there are some coefficients b_k such that

$$\sum_{k=0}^{n} (a_k w_k + \overline{a_k w_k}) = \left| \sum_{k=0}^{n} b_k w_k \right|^2. \qquad (32)$$

Notice that, by the linear independence of the functions w_k for $k \in \mathbb{Z}$, the coefficients $\{\beta_{jk,s}\}$ are uniquely determined by the system $\mathfrak{W}$ and that $\beta_{jk,s} = \beta_{kj,-s}$.

Now, to each probability measure $\sigma \in \mathcal{M}(\mathbb{T})$ and each $n \in \mathbb{N}$, we can associate a matrix $C_{n,\sigma} = [c_{jk}]_{j,k=0,\dots,n}$ by putting

$$c_{jk} = \int_{\mathbb{T}} \bar{w}_j w_k d\sigma. \qquad (33)$$

Clearly $C_{n,\sigma}$ is Hermitian and nonnegative. Since $w_0 = 1$ one has $c_{0k} = c_k$, the generalized moment of σ defined in (30) and, moreover,

$$c_{jk} = \sum_{s=-j}^{k} \beta_{jk,s} c_s, \qquad (34)$$

where the $\beta_{jk,s}$ are as in (31).

Conversely, we say that a Hermitian matrix $C = C^*$ is a *generalized Gram matrix* w.r.t. to $\mathfrak{W}$ (GGM or GG-matrix, for short) if its entries satisfy relations (34), where the coefficients $\beta_{jk,s}$ are those in (31). Note that we do not assume in this definition that $C \geq 0$. Subsequently, it is a natural question whether all nonnegative GG-matrices w.r.t. to $\mathfrak{W}$ come from some $\sigma \in \mathcal{M}(\mathbb{T})$ *via* (33). The answer (also explaining the terminology) is given by a slightly modified version of Theorem 4.1 which says, in particular, that relations (33) and (34) define the same object. Although not phrased exactly in this form, the essential part of the matter is in [10, Ch. 1,3]. Everything in the statement below is to be understood "w.r.t. to $\mathfrak{W}$".

Theorem 4.2: Let the system of functions $\mathfrak{W}$ satisfy (29) for some coefficients $\{\mathfrak{a}_k\}$, $\mathfrak{a}_k \in \mathbb{C}$. Assume moreover that assumptions 1), 2), and 3) above are met. Then, the following assertions are equivalent:

(1) the GG-matrix $C = [c_{jk}]_{j,k=0,\dots,n}$ is positive: $C \geq 0$,
(2) the sequence $\{c_k := c_{0k}\}_{k=0,\dots,n}$ is positive (in the sense of implication (28)),
(3) there is a measure $\sigma \in \mathcal{M}(\mathbb{T})$ with the property

$$c_{jk} = \int_{\mathbb{T}} \bar{w}_j w_k d\sigma.$$

Proof: It is plain that (3) and (30) are equivalent. Indeed, (3) implies (30) since $c_k = c_{0k} = \int_{\mathbb{T}} w_k d\sigma$.

Conversely if $c_k = \int_{\mathbb{T}} w_k d\sigma$ for all k, by (34) and (31), we have that

$$c_{jk} = \sum_{s=-j}^{k} \beta_{jk,s} c_s = \int_{\mathbb{T}} \sum_{s=-j}^{k} \beta_{jk,s} w_s d\sigma = \int_{\mathbb{T}} \bar{w}_j w_k d\sigma.$$

Now, (30) is equivalent to claim (2) by Theorem 4.1. Since (30) easily implies claim (1), it remains to show that (1) $\Rightarrow$ (30). By Theorem 4.1, we have to prove that if

$$\sum_{k=0}^{n} (a_k w_k + \overline{a_k w_k}) \geq 0$$

for some set of coefficients $\{a_k\}$, then

$$\sum_{k=0}^{n} (a_k c_k + \overline{a_k c_k}) \geq 0.$$

Denote by $T : \mathcal{L}_{n,\mathfrak{W}} \oplus \bar{\mathcal{L}}_{n,\mathfrak{W}} \to \mathbb{C}$ the linear map sending w_k to c_k for $k \geq 0$ and sending w_{-k} to $\bar{c}_k$. Note that T is well-defined by the linear independence of the w_k for $k \in \mathbb{Z}$. By (32), one has

$$\sum_{k=0}^{n} (a_k w_k + \overline{a_k w_k}) = \sum_{j,k=0}^{n} \bar{b}_j b_k \bar{w}_j w_k, \qquad (35)$$

and applying T to both sides of (35) while using (31) and (34), we get

$$
\begin{aligned}
\sum_{k=0}^{n} (a_k c_k + \overline{a_k c_k}) &= \sum_{j,k=0}^{n} \bar{b}_j b_k T(\bar{w}_j w_k) \\
&= \sum_{j,k=0}^{n} \bar{b}_j b_k \sum_{s=-j}^{k} \beta_{jk,s} T(w_s) \\
&= \sum_{j,k=0}^{n} \bar{b}_j b_k \sum_{s=-j}^{k} \beta_{jk,s} c_s \\
&= \sum_{j,k=0}^{n} \bar{b}_j b_k c_{jk} \geq 0.
\end{aligned}
$$

This achieves the proof. ∎

V. BLASCHKE-VARYING STOCHASTIC PROCESSES

A. SOME FAMILIES OF NONSTATIONARY PROCESSES

Let $\mathfrak{W} = \{w_0, w_1, w_2, \cdots\}$ be a system of linearly independent continuous functions on $\mathbb{T}$, linearly independent with their complex conjugates as well, like in the previous section. We set as before $w_{-k} = \bar{w}_k$. In addition to assumptions (29) and (1)-(3) in that section, we suppose that the family $\mathfrak{W}$ is separating, that is:

4) If, for some $\mu \in \mathcal{M}(\mathbb{T})$, one has

$$\int_{\mathbb{T}} w_k d\mu = 0$$

for all $k \in \mathbb{Z}$, then $\mu = 0$.

Notice that, in particular, the condition implies that the span of $\mathfrak{W} + \bar{\mathfrak{W}}$ is dense in $L^2(d\mu)$.

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space and $\{X_n\}_{n \in \mathbb{Z}}$ a family of complex-valued random variables, that is to say $X_n : \Omega \to \mathbb{C}$ is $\mathcal{F}$-measurable. In other words, $\mathbf{X} := \{X_n\}_n$ is a discrete stochastic process. We assume $E(X_n) = 0$ and $\sup_n E(|X_n|^2) < \infty$, where E indicates the mathematical expectation. When the random variables X_n have joint normal distribution, $\{X_n\}$ is called a Gaussian stochastic process. For $\{X_{n_1}, \ldots, X_{n_m}\}$, we define the covariance matrix by

$$\Gamma = \Gamma_{n_1,\ldots,n_m} = [\Gamma_{jk}], \quad \Gamma_{jk} = E(X_{n_k} \bar{X}_{n_j}),$$

where $j, k = 1, \ldots, m$. The process is stationary if the covariance matrix is invariant w.r.t. to the time shift, *i.e.*

$$\Gamma_{n_1,\ldots,n_m} = \Gamma_{n_1+l,\ldots,n_m+l}.$$

This results readily in the matrix $\Gamma_{n_1,\ldots,n_m}$ having Toeplitz structure when $n_1, \cdots, n_m$ are consecutive integers.

We say that $\mathbf{X}$ is a $\mathfrak{W}$-varying Gaussian process (abbreviations: $\mathfrak{W}$VSGP or $\mathfrak{W}$VSG-process) if $\Gamma = \Gamma_{n_1,\ldots,n_m}$ is a generalized Gram matrix w.r.t. to $\{w_{n_1}, \ldots, w_{n_m}\}$, in the sense of the preceding section. Note that such processes are *not* stationary in general.

If $\mathbf{X}$ is a $\mathfrak{W}$VSG-process, then Theorem 4.2 implies there exists a unique $\sigma \in \mathcal{M}(\mathbb{T})$ such that

$$\Gamma = \Gamma_{n_1,\ldots,n_m} = \left[\int_{\mathbb{T}} \bar{w}_{n_j} w_{n_k} d\sigma \right]_{j,k}.$$

The measure σ is called the spectral measure of $\mathbf{X}$. The name is justified by the following theorem.

Theorem 5.1: Let $\mathbf{X}$ be a $\mathfrak{W}$VSGP and σ be its spectral measure. Then there exists a unique family of random variables $Z_\xi = Z(., \xi), \xi \in \mathbb{T}$ defined on Ω such that

1) For $\{\xi_1, \ldots, \xi_k\}, \xi_j \in \mathbb{T}, \xi_i \neq \xi_j$, the random variables $\{Z_{\xi_j}\}_j$ are jointly normally distributed.
2) For $I = [\xi_1, \xi_2) \subset \mathbb{T}$, one writes $Z(I) = Z_{\xi_2} - Z_{\xi_1}$ and

$$E(|Z(I)|^2) = \sigma(I), \quad E(Z(I_1)\bar{Z}(I_2)) = 0,$$

for $I_1 \cap I_2 = \emptyset$.
3)
$$X_n = \int_{\mathbb{T}} w_n(\xi) dZ(., \xi). \quad (36)$$

Note that $\overline{\text{span}_{L^2(d\sigma)}\{w_k\}} = L^2(d\sigma)$ by property (4) and that (36) entails the map $U : L^2(\mathbf{X}) \to L^2(d\sigma)$ given by $U(X_n) = w_n$ is unitary. By conjugation under U, the shift defined by $\mathcal{Z}X_n = X_{n+1}$ goes to

$$U\mathcal{Z}U^*\left(\sum_k a_k w_k\right) = \sum_k a_k w_{k+1}$$

whenever $\sum_k a_k w_k$ and $\sum_k a_k w_{k+1}$ make sense in $L^2(\mathbf{X})$ and $L^2(d\sigma)$ respectively; in particular this holds for finite sums.

The proof of Theorem 5.1 dwells on the classical considerations leading to the spectral representation of stationary processs [4, Theorem 11.21]). We shall not produce this adaptation here, and details will be carried out in [2].

B. BLASCHKE VARYING PROCESSES

Given a sequence $(\alpha_k)_{k \in \mathbb{N}}$ in $\mathbb{D}$, a particular type of family $\mathfrak{W}$ satisfying the assumptions set out in the previous section is obtained upon letting $w_n = \mathcal{B}_n$ where $\mathcal{B}_n$ was defined in (2).

In this case $\mathcal{L}_{n,\mathfrak{W}} = \mathcal{L}_n$ defined in (4) and $\bar{w}_j w_k = \mathcal{B}_{k,j}$. The linear independence required for the theory are obvious. When the α_k are distinct, relation (31) follows from the identity

$$\mathcal{B}_{k,j} = \prod_{\ell=j}^{k}(-1/\bar{\alpha}_\ell) + \sum_{\ell=j}^{k} \frac{1}{\bar{\alpha}_\ell \prod_{j \leq s \leq k, \, s \neq \ell} \zeta_s(1/\bar{\alpha}_\ell)} \zeta_\ell, \quad (37)$$

which is easily checked by partial fraction expansion for $j \leq k$. The case of higher multiplicities generates more cumbersome formulas but can be carried out in a similar fashion. Relation (32) quickly follows from it polynomial analog, the Fejèr-Riesz factorization. Condition 4) holds because (17) implies density of $\mathcal{L}_n$ in $A(\mathbb{D})$.

A $\mathfrak{W}$VSG-process associated with such a family we call a *Blaschke-varying* Gaussian stochastic process. The terminology is intended to indicate that, thanks to Theorem 5.1, the time shift between instant j and instant k correspond to multiplication by the Blaschke product $\mathcal{B}_{k,j}$ in the spectral domain.

While Blaschke varying processes are clearly a strict generalization of classical stationnary processes (for which $(\alpha_k) \equiv (0)$ and $\zeta_k(z) = z$ for all k so that $\mathcal{B}_k = z^k$), their description in probabilistic terms is a very interesting issue which seems presently wide open. From a purely analytic viewpoint, the characterization of those generalized covariance matrices Γ associated with a system of functions ζ_k for some sequence (α_k) (*i.e.* the qualification of those nonnegative Hermitian matrices for which the coefficients $\beta_{jk,s}$ in (34) coincide with the coefficients in (37) for some (α_k) in $\mathbb{D}$) can be carried out using classical interpolation theory but goes beyond the scope of the present paper. One motivation to study the class of Blaschke varying processes lies with the fact that they are a nonstationary generalization of classical stationary Gaussian processes for which an interesting generalization of prediction theory exists as well. This is explained in the next and final section.

VI. PREDICTION THEORY OF BLASCHKE VARYING PROCESSES

Let $(\alpha_k)_{k \in \mathbb{N}}$ be a sequence of points in $\mathbb{D}$ and $\{\mathcal{B}_k\}$ be the sequence of Blaschke products defined in (2). Let further $\mathbf{X} := \{X_n\}_n$ be a Blaschke varying Gaussian stochastic process associated with the family $\mathfrak{W} = \{\mathcal{B}_0, \mathcal{B}_1, \mathcal{B}_2, \cdots\}$. Denote by σ its spectral measure and $Z(., \xi)$ the resolution of the identity given by Theorem 5.1, so that

$$X_n = \int_{\mathbb{T}} \mathcal{B}_n(\xi) dZ(., \xi).$$

We assume throughout that σ has infinite support. Put for simplicity $\mathcal{X}_{0,n-1} = \text{span}\{X_k : k = 0, \ldots, n-1\}$ The one-step ahead prediction problem is to compute $\hat{X}_n :=$

$(X_n | \mathcal{X}_{0,n-1}) = Pr_{\mathcal{X}_{0,n-1}} X_n$, where $Pr_{\mathcal{X}_{0,n-1}}$ indicates the orthogonal projection onto the subspace $\mathcal{X}_{0,n-1}$ in $L^2(\mathbf{X})$. The random variable $\hat{X}_n$ is the best (linear) predictor of X_n knowing $X_0, \cdots, X_{n-1}$, and $E|X_n - \hat{X}_n|^2$ is the square of the prediction error.

Since $\hat{X}_n$ lies in $\mathcal{X}_{0,n-1}$, we can write

$$\hat{X}_n = \sum_{k=0}^{n} \rho_{n,k} X_k$$

for some unique vector $(\rho_{n,n-1}, \cdots, \rho_{n,0})$ of complex co-efficients. These are called the *recursion* coefficients. Now, with the above notation and definitions, we have:

Proposition 6.1: Let ϕ_n be the normalized n-th orthogonal rational function associated with (α_k) and σ. Write ϕ_n as in (5):

$$\phi_n = \kappa_n \mathcal{B}_n + a_{n,n-1} \mathcal{B}_{n-1} + \ldots + a_{n,1} \mathcal{B}_1 + a_{n,0} \mathcal{B}_0.$$

Then the recursion coefficients are

$$(-a_{n,n-1}/\kappa_n, -a_{n,n-2}/\kappa_n \cdots, -a_{n,0}/\kappa_n). \qquad (38)$$

and the prediction error is $1/\kappa_n = 1/\phi_n^*(\alpha_n)$.

Proof: Because the map $U : L^2(\mathbf{X}) \to L^2(d\sigma)$ given by $U(X_n) = \mathcal{B}_n$ is unitary (see the remark after Theorem 5.1), and since the function

$$-(a_{n,n-1}/\kappa_n)\mathcal{B}_{n-1} - \ldots - (a_{n,1}/\kappa_n)\mathcal{B}_1 - (a_{n,0}/\kappa_n)\mathcal{B}_0$$

is the projection of $\mathcal{B}_n$ onto $\mathcal{L}_{n-1}$ by the very construction of ORFs, it is clear that the recursion coefficients are given by (38). Hence $\hat{X}_n$ is equal to

$$-(a_{n,n-1}/\kappa_n)X_{n-1} - \ldots - (a_{n,1}/\kappa_n)X_1 - (a_{n,0}/\kappa_n)X_0.$$

As $\mathcal{B}_n$ has unit $L^2(d\sigma)$-norm, it follows immediately from the unitary character of U that $(E|X_n - \hat{X}_n|^2)^{1/2}$ is $1/\kappa_n = 1/\phi_n^*(\alpha_n)$, as desired. ∎

We are now in position to obtain an asymptotic estimate of the prediction error as follows.

Theorem 6.2: Let $\mathbf{X} := \{X_n\}_n$ be a Blaschke varying Gaussian stochastic process associated with a sequence (α_k) of points in $\mathbb{D}$, and σ be its spectral measure. Denote by E_n the one-step ahead prediction error $\sqrt{E|X_n - \hat{X}_n|^2}$ of X_n knowing $X_0, \cdots, X_{n-1}$. Assume that $\sigma \in (S)$ and that (17) holds. For $m(n)$ be a sequence of integers such that $\lim_{n \to +\infty} \alpha_{m(n)} = \alpha \in \overline{\mathbb{D}}$, then

1) if $\alpha \in \mathbb{D}$

$$\lim_{n \to +\infty} E_{m(n)} = |S(\alpha)|\sqrt{1 - |\alpha|^2}; \qquad (39)$$

2) if $\alpha \in \mathbb{T}$ and (19)-(21) hold with μ replaced by σ,

$$\lim_{n \to +\infty} E_{m(n)} = 0, \qquad (40)$$

and if (α_k) accumulates nontangentially to α, it is enough to assume instead of (19)-(20) that (22) holds for σ.

Proof: The assertions follow at once from Proposition 6.1 and Corollaries 3.4-3.6. ∎

Theorem 6.2 indicates that the behaviour of Blaschke varying processes can be very varied: they may be asymptotically deterministic or else stochastic, or even some subsequences may be asymptotically deterministic and others not.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] Akhiezer, N. Theory of approximation. Dover Publications Inc., New York, 1992.
[2] Baratchart, L. and Kupin, S. and Golinskii, L. Multipoint Schur algorithm: generalized moment problems, Gaussian processes and prediction theory, *in preparation*
[3] Baratchart, L. and Kupin, S. and Olivi, M. and Lunot, V. Multipoint Schur algorithm and orthogonal rational functions: convergence prop-erties, http://arxiv.org/pdf/0812.2050 (submitted).
[4] L. Breiman, Probability. SIAM, Philadelphia, PA, 1992.
[5] Bultheel, A., González-Vera, P., Hendriksen, E., Njåstad, O. Orthogo-nal rational functions. Cambridge University Press, Cambridge, 1999.
[6] Duren, P. Theory of H^p spaces. Academic Press, New York, 1970.
[7] Dzrbasjan, M. Orthogonal systems of rational functions on the unit circle with given set of poles. Dokl. Akad. Nauk. SSSR 147 (1962), 1278–1281 (Russian); English transl. in Soviet Mat. Dokl., (1962) no. 3, 1794–1798.
[8] Garnett, J. Bounded analytic functions. Springer, New York, 2007.
[9] Grenander, U. and Szegö, G. Toeplitz forms. Chelsea, 2d edition (textually unaltered), 1984.
[10] M. Krein, A. Nudelman, The Markov moment problem and extremal problems. Translations of Mathematical Monographs, Vol. 50. AMS, Providence, RI, 1977.
[11] Langer, H., Lasarow, A. Solution of a multiple Nevanlinna-Pick problem via orthogonal rational functions. J. Math. Anal. Appl. 293 (2004), no. 2, 605–632.
[12] Pan, K. On the convergence of rational functions orthogonal on the unit circle. J. Comput. Appl. Math. (1996), no. 76, 315–324.
[13] Simon, B. Orthogonal polynomials on the unit circle, I, II. AMS Colloquium Publications, vol. 54, Providence, RI, 2005.
[14] Szegő, G. Orthogonal polynomials. AMS, Providence, RI, 1975.