case but never published it. The discrete (i.e., difference equation) result goes back at least to Bôcher [**130**]. See Simon [**980**] for further discussion.

1.3. Carathéodory and Schur Functions

An analytic function, F, on $\mathbb{D}$ is called a *Carathéodory function* if and only if $F(0) = 1$ and $\operatorname{Re} F(z) > 0$ on $\mathbb{D}$. An analytic function, f, on $\mathbb{D}$ is called a *Schur function* if and only if $\sup_{z\in\mathbb{D}}|f(z)| \leq 1$. The association

$$F(z) = \frac{1 + zf(z)}{1 - zf(z)} \tag{1.3.1}$$

$$f(z) = \frac{1}{z}\,\frac{F(z) - 1}{F(z) + 1} \tag{1.3.2}$$

sets up a one-one correspondence between the two types of functions. These classes play a major role in OPUC, so we include here a sketch of the high points of their theory and their relation to measures on $\partial\mathbb{D}$.

We call F trivial if it is a rational function whose poles all lie on $\partial\mathbb{D}$ and which is pure imaginary at any regular point on $\partial\mathbb{D}$. We call f trivial if it is a finite Blaschke product

$$f(z) = e^{i\theta} \prod_{j=1}^{m} \frac{z - z_j}{1 - \bar{z}_j z} \tag{1.3.3}$$

for $z_1, \dots, z_m \in \mathbb{D}$. $f(z) \equiv w_0$, a constant, is considered a finite Blaschke product (so, trivial) if $|w_0| = 1$ and nontrivial if $|w_0| < 1$. It is easy to see that F is trivial if and only if the associated f is.

Some authors define a Schur function as an analytic map of $\mathbb{D}$ to $\mathbb{D}$. The difference with the definition we give is the functions $f(z) = w_0 \in \partial\mathbb{D}$, which we will call *degenerate* Schur functions. Thus *nondegenerate* Schur functions are precise analytic maps of $\mathbb{D}$ to $\mathbb{D}$. Note that degenerate functions are also trivial.

That (1.3.1)/(1.3.2) set up a one-one correspondence between the two classes follows if one notes that

$$\varphi(w) = \frac{1 + w}{1 - w}$$

is a bijection of $\mathbb{D}$ to $\{z \in \mathbb{C} \mid \operatorname{Re} z > 0\}$ with $\varphi(0) = 1$ and that the Schwarz lemma asserts that g is a Schur function with $g(0) = 0$ if and only if $g(z) = zf(z)$ for f a Schur function.

Given (1.3.1) with $|f| \leq 1$, we have the universal bound on Carathéodory functions:

$$\frac{1 - |z|}{1 + |z|} \leq |F(z)| \leq \frac{1 + |z|}{1 - |z|} \tag{1.3.4}$$

A *Laurent polynomial* is an analytic function, $g(z)$, on $\mathbb{C}\backslash\{0\}$ (meromorphic on $\mathbb{C}$), which is a finite linear combination of z^k with $k \in \mathbb{Z}$, that is, positive or negative. If $g \not\equiv 0$, there is a unique factorization

$$g(z) = z^\ell R(z) \tag{1.3.5}$$

with $\ell \in \mathbb{Z}$ and R a polynomial with $R(0) \neq 0$. The *degree* of g is the degree of the polynomial R. Thus, if $g(z) = \sum_{k=\ell}^{m} a_k z^k$ with $a_\ell \neq 0 \neq a_m$, $\deg(g) = m - \ell$. With these preliminaries in hand, we turn to the basic facts:

1. Fejér-Riesz Theorem

This result says that any Laurent polynomial f which is nonnegative on $\partial\mathbb{D}$ can be factored

$$f(e^{i\theta}) = |P(e^{i\theta})|^2 \tag{1.3.6}$$

where P is a polynomial whose zeros all lie in $\bar{\mathbb{D}}$ (or if one prefers, one can find a P whose zeros are all in $\mathbb{C}\backslash\mathbb{D}$). By analyticity, one then has

$$f(z) = P(z)\,\overline{P(1/\bar{z})} \tag{1.3.7}$$

This result is proven by noting first that nonnegativity in $\partial\mathbb{D}$ implies any zero f has on $\partial\mathbb{D}$ must be of even order. Since f is real on $\partial\mathbb{D}$, we first have

$$f(z) = z^{-n}Q(z) \tag{1.3.8}$$

where $\deg(Q) = 2n$, $Q(0) \neq 0$, and then that

$$f(z) = \overline{f(1/\bar{z})} \tag{1.3.9}$$

since both sides are analytic in $\mathbb{C}\backslash\{0\}$ and agree on $\partial\mathbb{D}$. Thus zeros in $\mathbb{C}\backslash\partial\mathbb{D}$ come in $z, 1/\bar{z}$ pairs of equal multiplicity. It follows from

$$z^{-1}(z - z_\ell)(z - \bar{z}_\ell^{-1}) = -\bar{z}_\ell^{-1}(z - z_\ell)(z^{-1} - \bar{z}_\ell)$$

that (1.3.6) holds where P is a constant times $\prod_{\ell=1}^n (z - z_\ell)$ with $\{z_\ell\}$ the set of zeros of Q in $\mathbb{D}$ union the zeros in $\partial\mathbb{D}$ with half their even multiplicity.

If we require that P have no zeros in $\mathbb{D}$, then the above argument determines P uniquely up to a multiplicative $e^{i\theta}$ factor. If we demand $P(0) > 0$, that factor is determined. Thus, we have the sharp form of the Fejér-Riesz theorem: If f is a Laurent polynomial that is nonnegative on $\partial\mathbb{D}$, then there exists a unique polynomial $P(z)$ with $P(0) > 0$ and $P(z) \neq 0$ for $z \in \mathbb{D}$ so that (1.3.6) holds. Below we will write an explicit formula for P (see (1.3.23)).

2. Toeplitz Matrices and the Carathéodory-Toeplitz Theorem

Given a sequence $\{c_n\}_{n=0}^\infty$ of complex numbers, when is there a nontrivial measure $d\mu$ on $\partial\mathbb{D}$ so that

$$c_n = \int e^{-in\theta}\, d\mu(\theta) \tag{1.3.10}$$

Define c_n for $n < 0$ by $c_n = \bar{c}_{-n}$ and form the $n \times n$ *Toeplitz matrix*

$$T_{ij}^{(n)} = c_{j-i} \qquad 0 \leq i, j \leq n-1 \tag{1.3.11}$$

that is,

$$T^{(n)} = \begin{pmatrix} c_0 & c_1 & \dots & c_{n-1} \\ c_{-1} & c_0 & \dots & c_{n-2} \\ \dots & \dots & \dots & \dots \\ c_{-n+1} & \dots & \dots & c_0 \end{pmatrix}$$

We will also define for $n = 0, 1, \dots$ (note the $n+1$ on the right),

$$D_n(d\mu) = \det(T^{(n+1)}) \tag{1.3.12}$$

Note. There are two sign conventions in the last few formulae. Some define c_n with $e^{in\theta}$ and some use c_{i-j}, not c_{j-i}, in (1.3.11). Either change causes a transpose in $T^{(n)}$, and using both (as, e.g., Szegő does in his 1920 papers [**1018, 1019**]) leaves $T^{(n)}$ as we define it. The most common conventions use (1.3.10), but $T_{ij}^{(n)} = c_{i-j}$. In any event, the reader needs to be aware that our $T^{(n)}$ may be others' $(T^{(n)})^t$

(but that does not change $\det(T^{(n)})$ nor the set of real numbers $\sum_{i,j=1}^n \bar\alpha_i \alpha_j T_{ij}^{(n)}$). In addition, some use $D_n = \det(T^{(n)})$ and/or define $T^{(n)}$ to be $(n+1)\times(n+1)$ rather than $n \times n$.

Let $P(e^{i\theta}) = \sum_{j=0}^{n-1} \alpha_j e^{ij\theta}$. If L is the linear functional on Laurent polynomials given by

$$L(z^{-n}) = c_n \tag{1.3.13}$$

then

$$L(P(z)\,\overline{P(1/\bar z)}\,) = \sum_{ij=0}^{n-1} T_{ij}^{(n)} \bar\alpha_j \alpha_i$$

It follows from the Fejér-Riesz theorem that L is strictly positive on all nonzero Laurent polynomials which are nonnegative on $\partial\mathbb{D}$ if and only if each $T^{(n)}$ is strictly positive definite. Since the Laurent polynomials are dense in $C(\partial\mathbb{D})$ by Weierstrass' theorem, the Riesz-Markov theorem [**896**, p. 111] says there exists a nontrivial measure $d\mu$ obeying (1.3.10) if and only if each $T^{(n)}$ is strictly positive. We thus have the *Carathéodory-Toeplitz theorem*: c_n are moments of a nontrivial measure on $\partial\mathbb{D}$ if and only if $D_n(c) > 0$ for all n.

3. Poisson Representation

The real and complex Poisson kernel are defined by

$$P_r(\theta,\varphi) = \frac{1-r^2}{1+r^2-2r\cos(\theta-\varphi)} \tag{1.3.14}$$

for $\theta,\varphi \in [0,2\pi)$ and $r \in [0,1)$ and by

$$C(z,w) = \frac{w+z}{w-z} \tag{1.3.15}$$

for $w \in \partial\mathbb{D}$, $z \in \mathbb{D}$. The connection is

$$P_r(\theta,\varphi) = \operatorname{Re} C(e^{i\theta}, re^{i\varphi}) = \operatorname{Re}\biggl[\frac{e^{i\theta}+re^{i\varphi}}{e^{i\theta}-re^{i\varphi}}\biggr] \tag{1.3.16}$$

which shows that $P_r(\theta,\varphi)$ for θ fixed is a harmonic function of $re^{i\varphi}$.

The *Poisson representation* says that if g is analytic in a neighborhood of $\bar{\mathbb{D}}$ with $g(0)$ real, then for $z \in \mathbb{D}$,

$$g(z) = \int \biggl(\frac{e^{i\theta}+z}{e^{i\theta}-z}\biggr) \operatorname{Re}(g(e^{i\theta}))\,\frac{d\theta}{2\pi} \tag{1.3.17}$$

The easiest way to see this is to note first that

$$\frac{e^{i\theta}+z}{e^{i\theta}-z} = \frac{1+ze^{-i\theta}}{1-ze^{-i\theta}} = 1 + 2\sum_{n=1}^\infty z^n e^{-in\theta} \tag{1.3.18}$$

uniformly convergent for $e^{i\theta} \in \partial\mathbb{D}$ and z running through compact subsets of $\mathbb{D}$. Secondly, if $g(z) = \sum_{n=0}^\infty a_n z^n$, then

$$\begin{aligned}\int e^{-in\theta} \operatorname{Re}(g(e^{i\theta}))\,\frac{d\theta}{2\pi} &= \frac12 \int e^{-in\theta}[g(e^{i\theta}) + \overline{g(e^{i\theta})}\,]\,\frac{d\theta}{2\pi} \\ &= \begin{cases} \operatorname{Re}(a_0) & \text{if } n=0 \\ \frac12 a_n & \text{if } n=1,2,\dots \end{cases}\end{aligned} \tag{1.3.19}$$

proving (1.3.17) since $a_0 = g(0)$ is assumed real.

An alternate proof uses the fact that $P_r(\theta,\varphi)$ is an approximate delta function as $r \uparrow 1$, so the real parts of the two sides of (1.3.17) are harmonic functions on $\mathbb{D}$, continuous on $\bar{\mathbb{D}}$, agreeing on $\partial\mathbb{D}$, and so equal by the maximum principle. Thus (1.3.17) holds up to an imaginary additive constant, which is zero since $g(0)$ is real.

Taking real parts of (1.3.17), we obtain for any h analytic or harmonic in a neighborhood of $\mathbb{D}$:

$$h(re^{i\theta}) = \int P_r(\theta,\varphi) h(e^{i\varphi})\, \frac{d\varphi}{2\pi} \tag{1.3.20}$$

In particular, taking $h(re^{i\theta}) = P_r(\theta,\varphi)$, we obtain

$$\int |P_r(\theta,\varphi)|^2\, \frac{d\varphi}{2\pi} = P_r(\varphi,\varphi) = \frac{1+r}{1-r} \tag{1.3.21}$$

If g is an analytic function in a neighborhood of $\bar{\mathbb{D}}$ with $g(0)=1$ and g nonvanishing on $\bar{\mathbb{D}}$, then applying (1.3.17) to $\log g$, we find

$$g(z) = \exp\biggl(\frac{1}{2\pi} \int \frac{e^{i\theta}+z}{e^{i\theta}-z}\, \log(|g(e^{i\theta})|)\, d\theta\biggr) \tag{1.3.22}$$

This extends by a limiting argument to cases where g is only nonvanishing on $\mathbb{D}$ (but still analytic in a neighborhood of $\bar{\mathbb{D}}$, although that can be weakened sometimes; see Subsection 9 below). In particular, the polynomial, P, that solves (1.3.6) with $P(0)>0$ and P nonvanishing on $\mathbb{D}$ is given by

$$P(z) = \exp\biggl(\frac{1}{4\pi} \int \frac{e^{i\theta}+z}{e^{i\theta}-z}\, \log(f(e^{i\theta}))\, d\theta\biggr) \tag{1.3.23}$$

4. Herglotz Representation

Let F be a Carathéodory function. For $0 \le r < 1$, let $d\mu_r(\theta)$ be the measure

$$d\mu_r(\theta) = \operatorname{Re} F(re^{i\theta})\, \frac{d\theta}{2\pi} \tag{1.3.24}$$

Define the Taylor coefficients of $F(z)$ at $z=0$ by

$$F(z) = 1 + 2\sum_{n=1}^\infty c_n z^n \tag{1.3.25}$$

(the reason for the 2 will be clear momentarily; see (1.3.18) and (1.3.29)). Then

$$\operatorname{Re} F(re^{i\theta}) = 1 + \sum_{n=1}^\infty (c_n r^n e^{in\theta} + \bar{c}_n r^n e^{-in\theta}) \tag{1.3.26}$$

Define $c_n = \bar{c}_{-n}$ for $n<0$ and $c_0 \equiv 1$. Then for $n = 0, \pm 1, \pm 2, \dots$,

$$\int e^{-in\theta} d\mu_r(\theta) = c_n r^{|n|} \tag{1.3.27}$$

and the Poisson representation for $g(z) = F(rz)$ says

$$F(rz) = \int \frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\mu_r(\theta) \tag{1.3.28}$$

Now take $r \uparrow 1$. By (1.3.27), $\int h(\theta)\, d\mu_r(\theta)$ has a limit for any Laurent polynomial h. Since such polynomials are dense in $C(\partial\mathbb{D})$ and $|\int h(\theta)\, d\mu_r(\theta)| \le \|h\|_\infty$, the $d\mu_r$'s converge weakly to a measure $d\mu$ with

$$\int e^{-in\theta}\, d\mu(\theta) = c_n \tag{1.3.29}$$

Since $C(z, e^{i\theta})$ is continuous in θ for each $z \in \mathbb{D}$, the weak convergence can be applied to (1.3.28) to conclude the *Herglotz representation theorem*: Any Carathéodory function has a representation

$$F(z) = \int \frac{e^{i\theta} + z}{e^{i\theta} - z}\, d\mu(\theta) \tag{1.3.30}$$

for a unique probability measure μ on $\partial\mathbb{D}$. F has a Taylor series (1.3.25) at $z = 0$ where c_n are the moments of $d\mu$ given by (1.3.29).

By combining this result and the Carathéodory-Toeplitz theorem, one obtains a solution to Carathéodory's problem: $\{2c_j\}_{j=1}^n$ are the Taylor coefficients of a nontrivial Carathéodory function if and only if the Toeplitz determinants $D_0(c), \dots, D_n(c)$ are strictly positive.

(1.3.30) suggests that F is the analog of (1.2.6) and, in many ways, it is (see the next subsection and also Theorem 3.2.11), but there are other analogs of the m-function for OPRL [**978**]:

(i) the relative Szegő function, $\delta_0 D$ (see Section 2.9)
(ii) the Schur function (see (1.1.15) and (1.2.57))
(iii) $zf(z)$ (see (9.2.30))
(iv) the function $m^+(z)$ (see (10.11.5) and (10.11.7))
(v) the function, $M(z)$, of Section 11.7 (see (11.7.76))

This is discussed further in Appendix B.2.

5. Boundary Values of the Carathéodory Function

There are four basic facts relating F to μ that we will need:

(i) $\frac{1}{2\pi} \operatorname{Re} F(re^{i\theta})\, d\theta$ converges weakly to $d\mu$. This is the basic construction above.

(ii) For Lebesgue a.e. θ, $\lim_{r\uparrow 1} F(re^{i\theta}) \equiv F(e^{i\theta})$ exists, and if $d\mu = w(\theta)\frac{d\theta}{2\pi} + d\mu_\text{s}$ with $d\mu_\text{s}$ singular, then

$$w(\theta) = \operatorname{Re} F(e^{i\theta}) \tag{1.3.31}$$

Since $e^{-F(z)}$ is a bounded analytic function, general principles (see, e.g., Rudin [**924**, p. 340]) imply it has a.e. boundary values which are nonvanishing. (1.3.31) is also a general fact (see Rudin [**924**, p. 244]). By (1.3.1),

$$\operatorname{Re} F(z) = \frac{1 - |zf|^2}{|1 - zf(z)|^2}$$

and thus, using (1.3.2) to see that f also has a.e. boundary values on $\partial\mathbb{D}$, we see that

$$d\mu_\text{ac}(\theta) = \frac{1 - |f(e^{i\theta})|^2}{|1 - e^{i\theta} f(e^{i\theta})|^2}\, \frac{d\theta}{2\pi} \tag{1.3.32}$$

In particular, the essential support of $d\mu_\text{ac}$ is $\{\theta \mid |f(e^{i\theta})| < 1\}$.

(iii) θ_0 is a pure point of $d\mu$ if and only if

$$\mu(\{\theta_0\}) = \lim_{r\uparrow 1} \left(\frac{1-r}{2}\right) [F(re^{i\theta_0})] \tag{1.3.33}$$

is nonzero. This follows from the dominated convergence theorem, the estimate (when $|z| < 1$) $(1-|z|)|e^{i\theta}+z|\,|e^{i\theta}-z|^{-1} \le 2$, and $\lim_{r\uparrow 1}(1-r)(e^{i\theta}+re^{i\theta_0})(e^{i\theta} - re^{i\theta_0}) = 0$ if $\theta \ne \theta_0$, $= 2$ if $\theta = \theta_0$.

(iv) $d\mu_\text{s}$ is supported on $\{\theta \mid \lim_{r\uparrow 1} \operatorname{Re} F(re^{i\theta}) = \infty\}$. This is also a standard fact about measures; it follows, for example, from Rudin [**924**, p. 243]. We note that just because this set is nonempty, one cannot conclude that $d\mu_\text{s} \ne 0$. For example, if the set is countable but at points where $|F(re^{i\theta_0})| \to \infty$, we have that $(1-r)|F(re^{i\theta_0})| \to 0$, then (iii) implies $d\mu_\text{s}$ is zero. We will see generalizations of this idea in Section 10.8.

One consequence of (iv) is that if F is analytic in a neighborhood of $\bar{\mathbb{D}}$, equivalently, by (1.3.25), if the c_n decay exponentially, then $d\mu_\text{s} = 0$.

See Lemma 3.2.15 for a discussion of boundary values of F when $\text{supp}(d\mu)$ has a gap.

6. Schur Parameters and the Schur Algorithm

Nondegenerate Schur functions are maps of $\mathbb{D}$ onto $\mathbb{D}$. Schur's algorithm exploits two ways of mapping a (suitable) Schur function to another:

(i) Since, for $\gamma \in \mathbb{D}$,
$$T_\gamma(w) = \frac{w-\gamma}{1-\bar{\gamma}w}$$
is an invertible analytic homeomorphism of $\mathbb{D}$ to $\mathbb{D}$ which maps $\gamma \to 0$, $T_{f(0)} \circ f$ is a nondegenerate Schur function which vanishes at zero.

(ii) By the Schwarz lemma, if f is a Schur function and $f(0) = 0$, then either $z^{-1}f(z)$ is a nondegenerate Schur function or else a constant in $\partial\mathbb{D}$.

Combining the two, we see that if f_0 is a nondegenerate Schur function and we let
$$\gamma_0 = f_0(0) \tag{1.3.34}$$
$$f_1(z) = \frac{1}{z}\,\frac{f_0(z)-\gamma_0}{1-\bar{\gamma}_0 f_0(z)} \tag{1.3.35}$$
then we have an algorithm for mapping one Schur function to another. By iterating, we get a sequence of numbers $\gamma_0, \gamma_1, \dots, \gamma_n, \dots$ and Schur functions $f = f_0, f_1, \dots, f_n, \dots$, called the *Schur iterates* of f. The only snag is if at some stage, $f_n(z) = \gamma_n \in \partial\mathbb{D}$. In that case, we stop. If $f_n(z) \equiv \gamma_n \in \mathbb{D}$, we continue, which means $f_{n+1} = f_{n+2} = \cdots = 0$ and $\gamma_{n+1} = \gamma_{n+2} = \cdots = 0$. Thus, any Schur function, f, is associated to either an infinite sequence $\gamma_0, \gamma_1, \dots$ in $\mathbb{D}$ or a finite sequence $\gamma_0, \dots, \gamma_{n-1}, \gamma_n$ with $\gamma_j \in \mathbb{D}$ for $j < n$ and $\gamma_n \in \partial\mathbb{D}$. These numbers are called the Schur parameters for f. It is not hard to see that the case $\gamma_n \in \partial\mathbb{D}$ holds if and only if f is trivial, that is, a finite Blaschke product.

More formally, we can describe the association as follows: Let $\mathbb{D}^{\infty,c}$ be the set of all sequences $\{\alpha_n\}_{n=0}^N$ in $\bar{\mathbb{D}}$ which either have $N = \infty$ and all $\alpha_n \in \mathbb{D}$ or else have $N < \infty$ with $\alpha_0, \dots, \alpha_{N-1} \in \mathbb{D}$ and $\alpha_N \in \partial\mathbb{D}$. $\mathbb{D}^{\infty,c}$ with the topology of pointwise convergence is compact and has $\mathbb{D}^\infty$ as a dense set. The Schur algorithm is a map $\mathcal{S}$ from the set of Schur functions to $\mathbb{D}^{\infty,c}$ with finite sequences corresponding to trivial Schur functions (finite Blaschke products). By (1.3.43) below, $\mathcal{S}$ is a homeomorphism of the unit ball in H^∞ to $\mathbb{D}^{\infty,c}$. We will let $\mathcal{S}^{-1}$ be the inverse of $\mathcal{S}$.

One can invert the transformation $f_j \to (\gamma_j, f_{j+1})$. Define
$$S_{\gamma,z}(w) = \frac{\gamma + zw}{1+\bar{\gamma}zw} \tag{1.3.36}$$
Then
$$f_j(z) = S_{\gamma_j,z}(f_{j+1}(z)) \tag{1.3.37}$$
or
$$f(z) = S_{\gamma_0,z}(S_{\gamma_1,z}(\dots S_{\gamma_j,z}(f_{j+1}(z))) \tag{1.3.38}$$

In particular,

$$f(z) = \frac{\gamma_0 + zf_1(z)}{1 + \bar{\gamma}_0 z f_1(z)} \tag{1.3.39}$$

This can also be written

$$f(z) = \gamma_0 + \frac{(1 - |\gamma_0|^2) z f_1}{1 + \bar{\gamma}_0 z f_1(z)} \tag{1.3.40}$$

We define the *Schur approximants*, $f^{[j]}(z)$, by replacing f_{j+1} in (1.3.38) by 0. Thus, $f^{[j]}(z)$ is that Schur function with

$$\gamma_\ell(f^{[j]}) = \begin{cases} \gamma_\ell(f) & \ell = 0, 1, \dots, j \\ 0 & \ell \geq j + 1 \end{cases} \tag{1.3.41}$$

Since $S_{\gamma,1}$ maps $\mathbb{D}$ to $\mathbb{D}$ and $S_{\gamma,z}(w) = S_{\gamma,1}(wz)$, $S_{\gamma,z}$ map $\mathbb{D}$ onto $\mathbb{D}$ if $|z| < 1$. Thus, each $f^{[j]}$ is a Schur function.

In the next subsection, we will show that the n-th Taylor coefficients of a Schur function f,

$$f(z) = \sum_{n=0}^{\infty} s_n(f) z^n \tag{1.3.42}$$

only depend on $\gamma_0, \dots, \gamma_n$. Thus, if f, g are Schur functions with $\gamma_\ell(f) = \gamma_\ell(g)$ for $\ell = 0, \dots, n$, we have that $\frac{1}{2}(f - g)$ is a Schur function which vanishes to order $n+1$ and so, by the Schwarz lemma, $|f(z) - g(z)| \leq 2|z|^{n+1}$, that is,

$$\gamma_j(f) = \gamma_j(g) \text{ for } j = 0, \dots, n \Rightarrow |f(z) - g(z)| \leq 2|z|^{n+1} \tag{1.3.43}$$

In particular,

$$|f(z) - f^{[n]}(z)| \leq 2|z|^{n+1} \tag{1.3.44}$$

and thus $f^{[n]} \to f$ uniformly on compact subsets of $\mathbb{D}$.

This also shows that if $\gamma_0, \gamma_1, \dots$ is an arbitrary sequence in $\mathbb{D}$ and $f^{[j]}$ is defined by (1.3.38) with $f_{j+1}(z) = 0$, then

$$|f^{[j]}(z) - f^{[\ell]}(z)| \leq 2|z|^{1+\min(\ell,j)} \tag{1.3.45}$$

so $f^{[j]}$ is Cauchy in local uniform norm. Thus there is a Schur function, f, with $\gamma_\ell(f) = \gamma_\ell$. Therefore, we have *Schur's criterion*: The association $f \to \{\gamma_j(f)\}_{j=0}^{\infty}$ sets up a one-one correspondence between nontrivial Schur functions and $\times_{j=0}^{\infty} \mathbb{D}$.

7. Taylor Coefficients for Schur Functions

(1.3.39) implies that

$$(1 + \bar{\gamma}_0 z f_1) f = \gamma_0 + z f_1 \tag{1.3.46}$$

Recall, (1.3.42), that $s_n(f)$ are the Taylor coefficients of f. Using $s_0(f) = \gamma_0$ and identifying the powers of z^n on both sides of (1.3.46) leads to

$$s_n(f) = (1 - |\gamma_0|^2) s_{n-1}(f_1) - \bar{\gamma}_0 \sum_{j=1}^{n-1} s_j(f) s_{n-1-j}(f_1) \tag{1.3.47}$$

if $n \geq 1$. This formula plus induction in n implies *Schur's recurrence relation*,

$$s_n(f) = \prod_{j=0}^{n-1} (1 - |\gamma_j|^2) \gamma_n + r_n(\gamma_0, \bar{\gamma}_0, \gamma_1, \bar{\gamma}_1, \dots, \gamma_{n-1}, \bar{\gamma}_{n-1}) \tag{1.3.48}$$

where r_n is a polynomial in its arguments.

This formula was used in the proof of (1.3.43), and the following consequence of it, (1.3.50), will play an important role in Section 3.1.

Suppose now that μ is a measure, F its m-function (given by (1.3.30)), and f is a Schur function (given by (1.3.2)). Since (1.3.1) says

$$F(z) = 1 + 2\sum_{n=1}^{\infty}(zf)^n \tag{1.3.49}$$

and the Taylor coefficients of F are given by (1.3.25), we have identifying coefficients of z^n on both sides of (1.3.49):

$$c_n = s_{n-1}(f) + \text{polynomial in } (s_0(f), \dots, s_{n-2}(f))$$

Thus we have that the moments of μ and the Schur coefficients of the associated Schur functions are related by (for $n \geq 1$)

$$c_n(f) = \prod_{j=0}^{n-2}(1-|\gamma_j|^2)\gamma_{n-1} + \tilde{r}_{n-1}(\gamma_0, \bar{\gamma}_0, \dots, \gamma_{n-2}, \bar{\gamma}_{n-2}) \tag{1.3.50}$$

where $\tilde{r}_{n-1}$ is a polynomial. Note that up to shift of index, the first terms on the right of (1.3.48) and (1.3.50) are the same, but the precise polynomials r_n and $\tilde{r}_n$ are different.

The explicit formulae quickly become very complicated. For example (and once one has Geronimus' theorem $\gamma_n = \alpha_n$),

$$c_1 = \gamma_0 \tag{1.3.51}$$

$$c_2 = \gamma_0^2 + \gamma_1(1-|\gamma_0|^2) \tag{1.3.52}$$

$$c_3 = (\gamma_0 - \gamma_1\bar{\gamma}_0)[\gamma_0^2 + \gamma_1(1-|\gamma_0|^2)] + \gamma_1\gamma_0 + \gamma_2(1-|\gamma_0|^2)(1-|\gamma_1|^2) \tag{1.3.53}$$

Later, we will have a compact formula (1.5.80) for γ_n as a ratio of determinants built from the c's.

(1.3.46) has an interesting consequence. It can be rewritten

$$f - \gamma_0 - zf_1 = -\bar{\gamma}_0 z f_1 f \tag{1.3.54}$$

which implies that

$$|f - \gamma_0 - zf_1| \leq |\gamma_0|\,|z| \tag{1.3.55}$$

Thus, by induction,

$$\left| f - \sum_{n=0}^{N}\gamma_n z^n - z^{N+1}f_{N+1}\right| \leq \sum_{n=0}^{N}|\gamma_n|\,|z|^{n+1} \tag{1.3.56}$$

which implies that

$$|f(z)| \leq 2\sum_{n=0}^{N}|\gamma_n|\,|z|^n + |z|^{N+1} \tag{1.3.57}$$

and thus, taking $N \to \infty$,

$$|f(z)| \leq 2\sum_{n=0}^{\infty}|\gamma_n|\,|z|^n \tag{1.3.58}$$

An interesting application of this is to note that if $\sum_{n=0}^{\infty}|\gamma_n| < \frac{1}{2}$, then $\sup_{z\in\mathbb{D}}|f(z)| < 1$, which implies that the associated Carathéodory function, F, is bounded on $\mathbb{D}$ and so the measure associated to F is purely absolutely continuous. Given Geronimus' theorem, this is only a few steps short of the easy half of Baxter's theorem.

8. Wall Polynomials

The Schur approximants are, by construction, rational functions. Indeed, since $f_{j+1}^{[j]} = 0$, by induction and (1.3.36)/(1.3.37), $f_{j+1-k}^{[j]}(z)$ is a ratio of polynomials of degree $k-1$. Thus, $f^{[j]}(z) = f_0^{[j]}$ is a ratio of polynomials of degree j.

Calculations with compositions of fractional linear transformations are done most easily using matrix multiplication: In fancy language, fractional linear transformations are projective automorphisms. More prosaically, if $S(w) = (aw + b)/(cw+d)$, then

$$\begin{pmatrix} S(w) \\ 1 \end{pmatrix} = \frac{1}{c+dw} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} w \\ 1 \end{pmatrix}$$

Since the $(c+dw)^{-1}$ factors commute with matrices, it is convenient to write

$$z = \frac{aw+b}{cw+d}$$

by the shorthand

$$z \doteq Aw \qquad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{1.3.59}$$

We clearly have

$$z \doteq Aw \quad \text{and} \quad w \doteq By \quad \Rightarrow \quad z \doteq ABy \tag{1.3.60}$$

and, in particular, if A is invertible,

$$z \doteq Aw \Leftrightarrow w \doteq A^{-1} z \tag{1.3.61}$$

Thus $f_j = (zf_{j+1} + \gamma_j)/(\bar{\gamma}_j z f_{j+1} + 1)$ and $f = f_0$ becomes

$$f \doteq \begin{pmatrix} z & \gamma_0 \\ \bar{\gamma}_0 z & 1 \end{pmatrix} \begin{pmatrix} z & \gamma_1 \\ \bar{\gamma}_1 z & 1 \end{pmatrix} \cdots \begin{pmatrix} z & \gamma_n \\ \bar{\gamma}_n z & 1 \end{pmatrix} f_{n+1}$$

Defining the matrix product on the right to be $\left(\begin{smallmatrix} C_n & A_n \\ D_n & B_n \end{smallmatrix}\right)$, we have the pair of preliminary formulae:

$$\begin{pmatrix} C_n & A_n \\ D_n & B_n \end{pmatrix} = \begin{pmatrix} C_{n-1} & A_{n-1} \\ D_{n-1} & B_{n-1} \end{pmatrix} \begin{pmatrix} z & \gamma_n \\ \bar{\gamma}_n z & 1 \end{pmatrix} \tag{1.3.62}$$

$$f = \frac{A_n + C_n f_{n+1}}{B_n + D_n f_{n+1}} \tag{1.3.63}$$

with initial condition $\left(\begin{smallmatrix} C_0 & A_0 \\ D_0 & B_0 \end{smallmatrix}\right) = \left(\begin{smallmatrix} z & \gamma_0 \\ \bar{\gamma}_0 z & 1 \end{smallmatrix}\right)$.

A_n and B_n are called the *Wall polynomials*; as we will see momentarily, C_n and D_n can be expressed in terms of A_n and B_n. By (1.3.62), z divides C_n and D_n, so we can define X_n and Y_n by $C_n = zX_n$, $D_n = zY_n$. Writing

$$\begin{pmatrix} C_n & A_n \\ D_n & B_n \end{pmatrix} \begin{pmatrix} z^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} X_n & A_n \\ Y_n & B_n \end{pmatrix}$$

and using $\left(\begin{smallmatrix} z^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1} \left(\begin{smallmatrix} z & \gamma_n \\ \bar{\gamma}_n z & 1 \end{smallmatrix}\right) \left(\begin{smallmatrix} z^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right) = \left(\begin{smallmatrix} z & \gamma_n z \\ \bar{\gamma}_n & 1 \end{smallmatrix}\right)$ shows

$$\begin{pmatrix} X_n & A_n \\ Y_n & B_n \end{pmatrix} = \begin{pmatrix} X_{n-1} & A_{n-1} \\ Y_{n-1} & B_{n-1} \end{pmatrix} \begin{pmatrix} z & \gamma_n z \\ \bar{\gamma}_n & 1 \end{pmatrix} \tag{1.3.64}$$

If one writes out the equations obeyed by A_n and X_n: $A_n = A_{n-1} + \gamma_n z X_{n-1}$; $X_n = zX_n + \bar{\gamma}_n A_{n-1}$, and notes the initial conditions, one finds inductively that $X_n = B_n^*$, and similarly, $Y_n = A_n^*$ where $A_n^*(z) = z^n \overline{A_n(1/\bar{z})}$ as in (1.1.7). We thus have the equations:

$$A_n(z) = A_{n-1}(z) + \gamma_n z B_{n-1}^*(z) \tag{1.3.65}$$
$$B_n(z) = B_{n-1}(z) + \gamma_n z A_{n-1}^*(z) \tag{1.3.66}$$
$$A_n^*(z) = zA_{n-1}^*(z) + \bar{\gamma}_n B_{n-1}(z) \tag{1.3.67}$$
$$B_n^*(z) = zB_{n-1}^*(z) + \bar{\gamma}_n A_{n-1}(z) \tag{1.3.68}$$
$$A_0(z) = \gamma_0 \tag{1.3.69}$$
$$B_0(z) = 1 \tag{1.3.70}$$

These last four equations can be nicely summarized in a single matrix equation:

$$W_n = \begin{pmatrix} zB_{n-1}^* & -A_{n-1}^* \\ -zA_{n-1} & B_{n-1} \end{pmatrix} \tag{1.3.71}$$
$$W_{n+1} = \begin{pmatrix} z & -\bar{\gamma}_n \\ -\gamma_n z & 1 \end{pmatrix} W_n \tag{1.3.72}$$

with initial condition

$$W_1 = \begin{pmatrix} z & -\bar{\gamma}_0 \\ -\gamma_0 z & 1 \end{pmatrix} \tag{1.3.73}$$

The equations hold if all minus signs are dropped; following Pintér-Nevai [**868**], we put them in to facilitate the comparison with the transfer matrix in Section 3.2.

Since

$$\begin{pmatrix} z & -\bar{\gamma}_n \\ -z\gamma_n & 1 \end{pmatrix}^{-1} = z^{-1}(1-|\gamma_n|^2)^{-1} \begin{pmatrix} 1 & \bar{\gamma}_n \\ z\gamma_n & z \end{pmatrix}$$

(1.3.71)/(1.3.72) imply the inverse recursion relations for A and B:

$$A_{n-1} = (1-|\gamma_n|^2)^{-1}(A_n - \gamma_n B_n^*) \tag{1.3.74}$$
$$B_{n-1} = (1-|\gamma_n|^2)^{-1}(B_n - \gamma_n A_n^*) \tag{1.3.75}$$

This can also be checked directly from (1.3.65)–(1.3.68).

Making the γ-dependence explicit, (1.3.72) also implies

$$W_{n+1}(z;\gamma_0,\dots,\gamma_n) = W_n(z;\gamma_1,\dots,\gamma_n) \begin{pmatrix} z & -\bar{\gamma}_0 \\ -\gamma_0 z & 1 \end{pmatrix} \tag{1.3.76}$$

which implies, given (1.3.71),

$$B_n(z;\gamma_0,\dots,\gamma_n) = B_{n-1}(z;\gamma_1,\dots,\gamma_n) + \bar{\gamma}_0 z A_{n-1}(z;\gamma_1,\dots,\gamma_n) \tag{1.3.77}$$
$$A_n(z;\gamma_0,\dots,\gamma_n) = zA_{n-1}(z;\gamma_1,\dots,\gamma_n) + \gamma_0 B_{n-1}(z;\gamma_1,\dots,\gamma_n) \tag{1.3.78}$$

Note that, by (1.3.68) and the initial condition, B_n^* is a monic polynomial of degree n. This means that if $\gamma_n \neq 0$, A_n is also a polynomial of exact degree n.

(1.3.63) now reads

$$f(z) = \frac{A_n(z) + zB_n^*(z) f_{n+1}(z)}{B_n(z) + zA_n^*(z) f_{n+1}(z)} \tag{1.3.79}$$

In particular, the Schur approximant $f^{[n]}$, given by replacing f_{n+1} by 0 in (1.3.38), is given by

$$f^{[n]}(z) = \frac{A_n(z)}{B_n(z)} \tag{1.3.80}$$

Thus (1.3.44) implies

$$\left| f(z) - \frac{A_n(z)}{B_n(z)} \right| \le 2|z|^{n+1} \tag{1.3.81}$$

so A_n/B_n converges to f uniformly on compact sets.

Taking determinants of (1.3.71), (1.3.72), and (1.3.73) implies

$$B_n(z)B_n^*(z) - A_n(z)A_n^*(z) = z^n \prod_{j=0}^{n} (1 - |\gamma_j|^2) \tag{1.3.82}$$

In particular, B_n and A_n have no common zeros away from $z = 0$. Since $B_n(0) = 1$ (by induction from (1.3.66)), they have no common zero. Thus, analyticity of $f^{[n]}$ in $\mathbb{D}$ implies that all the zeros of B_n lie in $\mathbb{C}\backslash\overline{\mathbb{D}}$. In fact, B_n has no zeros on $\partial\mathbb{D}$, either, for (1.3.82) on $\partial\mathbb{D}$ (using $P^*(z) = z^n\, \overline{P(z)}$ if $z \in \partial\mathbb{D}$) becomes

$$z \in \partial\mathbb{D} \Rightarrow |B_n(z)|^2 - |A_n(z)|^2 = \prod_{j=0}^{n} (1 - |\gamma_j|^2) \tag{1.3.83}$$

Thus on $\partial\mathbb{D}$, $|B_n(z)| > 0$.

(1.3.83) implies for $z \in \partial\mathbb{D}$,

$$1 - |f^{[n]}(z)|^2 = |B_n(z)|^{-2} \prod_{j=0}^{n} (1 - |\gamma_j|^2) \tag{1.3.84}$$

so, by the maximum principle,

$$|z| \le 1 \Rightarrow |f^{[n]}(z)| \le 1 \Rightarrow |A_n(z)| \le |B_n(z)| \tag{1.3.85}$$

Since $|A_n^*(z)/B_n(z)| = |A_n(z)/B_n(z)|$ on $\partial\mathbb{D}$, the maximum principle implies that

$$|z| \le 1 \Rightarrow |A_n^*(z)| \le |B_n(z)| \tag{1.3.86}$$

9. Aleksandrov Measures

If F is a Carathéodory function and Q is analytic from $\mathbb{C}_r = \{z \mid \mathrm{Re}(z) > 0\}$ to itself with $Q(1) = 1$, then $Q(F(z))$ is also a Carathéodory function. It is particularly interesting to look at this for Q which are bijections of $\mathbb{C}_r$ to itself.

To find the allowed Q's, it is useful to conformally map $\mathbb{C}_r$ to $\mathbb{D}$ with 1 to 0. For if f is a bijection of $\mathbb{D}$ to itself with $f(0) = 0$, then $f(z) = \lambda z$ for some $\lambda \in \partial\mathbb{D}$. To see this well-known fact, note $|f(z)z^{-1}| \le r^{-1}$ on $\{z \mid |z| \le r\}$ for $0 < r < 1$. So taking $r \to 1$, $|f(z)| \le |z|$ on $\mathbb{D}$. But f has a two-sided functional inverse g, so $|z| = |g(f(z))| \le |f(z)|$. Thus, $|f(z)| = |z|$ or $|f(z)/z| = 1$, which implies, by the maximum principle, that $f(z)/z = \lambda$.

If

$$H(z) = \frac{1+z}{1-z} \qquad H^{-1}(w) = \frac{w-1}{w+1} \tag{1.3.87}$$

then H maps $\mathbb{D}$ to $\mathbb{C}_r$ and 0 to 1. Thus, if $R_\lambda(z) = \lambda z$, the maps

$$Q_\lambda = H \circ R_\lambda \circ H^{-1} \tag{1.3.88}$$

are all the analytic maps of $\mathbb{C}_r$ to itself taking 1 to 1. By a direct calculation,

$$Q_\lambda(w) = \frac{(1-\lambda)+(1+\lambda)w}{(1+\lambda)+(1-\lambda)w} \tag{1.3.89}$$

Given F, a Carathéodory function on $\mathbb{D}$, we define the *associated family* to $\{F^{(\lambda)}(\cdot)\}_{\lambda\in\partial\mathbb{D}}$ by

$$F^{(\lambda)}(z) = \frac{(1-\lambda)+(1+\lambda)F(z)}{(1+\lambda)+(1-\lambda)F(z)} \tag{1.3.90}$$

Since each $F^{(\lambda)}(z)$ is a Carathéodory function, there are measures $d\mu_\lambda(\theta)$ on $\partial\mathbb{D}$ so that

$$F^{(\lambda)}(z) = \int \frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\mu_\lambda(\theta) \tag{1.3.91}$$

The family of measures $\{d\mu_\lambda\}_{\lambda\in\partial\mathbb{D}}$ is called the family of *Aleksandrov measures* associated to the measure $d\mu$ assigned to the measure $d\mu \equiv d\mu_1$ defining F. We have thus associated to each probability measure on $\partial\mathbb{D}$ a natural family of measures.

By (1.3.88), the map H will simplify Q. Indeed, it obviously shows that $zf^{(\lambda)} = \lambda z f$ with f the Schur function of $d\mu$. For $z \neq 0$, one can divide by z and then recover $z = 0$ by continuity, that is,

$$f^{(\lambda)}(z) = \lambda f(z) \tag{1.3.92}$$

Setting $z = 0$, we see

$$\gamma_0^{(\lambda)} = \lambda\gamma_0 \tag{1.3.93}$$

Plugging (1.3.92) and (1.3.93) into (1.3.35), one finds $f_1^{(\lambda)}(z) = \lambda f_1(z)$. We can now iterate and find that

$$f_n^{(\lambda)}(z) = \lambda f(z) \qquad \gamma_n^{(\lambda)} = \lambda\gamma_n \tag{1.3.94}$$

This is just framework, but we will see it occur in the study of rank one decompositions in Subsection 1.4.16 and also in the analysis of second kind polynomials in Section 3.2.

10. Inner and Outer Functions and Factors

Any Schur function f has a natural factorization

$$f = f_I f_O \tag{1.3.95}$$

where an *inner function* is a Schur function, g, where $g^*(e^{i\theta}) \equiv \lim_{r\uparrow 1} g(re^{i\theta})$ obeys $|g^*(e^{i\theta})| = 1$ for a.e. θ and f_O is an outer function (defined below).

The inner part has a further factorization

$$f = f_B f_{SI} \tag{1.3.96}$$

where f_B is a Blaschke product, that is,

$$f_B(z) = e^{i\theta} \prod_{j=1}^{\infty} \frac{z-z_j}{1-z\bar{z}_j} \tag{1.3.97}$$

where the z_j's obey

$$\sum_{j=1}^{\infty} (1-|z_j|) < \infty \tag{1.3.98}$$

(which guarantees convergence of (1.3.97) in $\mathbb{D}$). The *singular inner function*, f_{SI}, is an inner function with no zeros in $\mathbb{D}$.

An *outer function* is one of the form

$$f_O(z) = \exp\biggl(-\int g(e^{i\theta})\,\frac{e^{i\theta}+z}{e^{i\theta}-z}\,\frac{d\theta}{2\pi}\biggr) \tag{1.3.99}$$

with $g \in L^1$. If f_O is a Schur function, $g \geq 0$. Outer Schur functions are exactly those Schur functions, f, with the property that $f = hk$ with h, k Schur functions and h inner have $h(z) = e^{i\gamma}$ for some $e^{i\gamma} \in \partial\mathbb{D}$.

The factorization

$$f = f_B f_{SI} f_O \tag{1.3.100}$$

can be constructed as follows. First ([**924**, p. 311]), one shows that the zeros of a Schur function obey (1.3.98), so the product in (1.3.97) converges ([**924**, p. 310]). Next, one shows $h \equiv f/f_B$ is also a Schur function ([**924**, p. 338]). It follows that $q(z) = -\log h(z)$ is a positive multiple of a Carathéodory function and so

$$q(z) = \int \frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\nu(\theta)$$

by the Herglotz representation theorem. Writing $d\nu = g(e^{i\theta})\frac{d\theta}{2\pi} + d\nu_{\text{s}}$, we obtain (1.3.100) where

$$f_{SI}(z) = \exp\biggl(-\int \frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\nu_{\text{s}}(\theta)\biggr) \tag{1.3.101}$$

and f_O is given by the right side of (1.3.99).

Note that the measure ν constructed here is only tenuously related to the measure μ connected to the Carathéodory function associated to f. There is one important connection: f is inner if and only if μ is purely singular. For $\operatorname{Re} F(e^{i\theta}) = 0$ if and only if $|f^*(e^{i\theta})| = 1$. Thus, to an analyst interested in f, it is interesting (as we will prove sometimes; see, e.g., Theorems 4.3.4, 10.11.3, 10.11.4, 12.5.2, 12.6.1, and Corollary 9.3.4) that $d\mu_{\text{ac}} = 0$, but whether $d\mu_{\text{s}}$ is pure point or singular continuous or a mixture is not so interesting because its connection to ν, the measure of interest for f, is unclear.

In Section 13.8, we will need a generalization of this factorization to a more general class of analytic functions. We say f, a function on $\mathbb{D}$, is a *Nevanlinna function* if and only if

$$\sup_{0<r<1} \int \log_+(|f(re^{i\theta})|)\,\frac{d\theta}{2\pi} < \infty \tag{1.3.102}$$

where, for $x \geq 0$,

$$\log_+(x) = \max(0, \log(x))$$

Every Nevanlinna function, f, has a factorization of the form (1.3.100), except the function g in (1.3.99) need not be positive. An equivalent fact is that any Nevanlinna function, f, can be written $f = h_1/h_2$ where h_1 and h_2 are Schur functions and h_2 is nonvanishing on $\mathbb{D}$. For proofs of these facts, see [**924**, pp. 342–346].

Remarks and Historical Notes. The theory of analytic functions on $\mathbb{D}$ with positive real part goes back to a paper of Carathéodory [**183**] in 1907, with an explosion of followup papers in 1911 by Carathéodory himself [**184**], Toeplitz [**1045**], Carathéodory-Fejér [**185**], Fischer [**350**], Herglotz [**504**], and F. Riesz [**903**]. With second papers of Fejér [**344**] and Riesz [**904**] and Schur's great papers [**948, 949**], the classical era on the subject was closed; see Duren [**316**], Hoffman [**537**], and Garnett [**378**] for more modern issues in this arena.

Carathéodory [**183**] asked the following question: Given a finite sequence of numbers $c_0, c_1, \dots, c_n$, when is there an analytic function on $\mathbb{D}$ with positive real part whose Taylor series at 0 is $c_0 + 2\sum_{j=1}^n c_n z^n + O(z^{n+1})$ (he did not have the 2 but it is useful to normalize this way, as we have seen). Given his then recent work on convex sets, not surprisingly, he attacked the problem from that point of view. Essentially, he noted that the set of such c_j's is a convex set in $\mathbb{R} \times \mathbb{C}^n$, and he identified the extreme points of the subset with $c_0 = 1$ as associated with what we have called trivial C-functions with at most n poles (equivalently, so that the Toeplitz matrix, $T^{(n+1)}$, is nonnegative but singular, i.e., $\det(T^{(j)}) \geq 0$ for $j = 1, 2, \dots, n$ and $\det(T^{(n+1)}) = 0$).

The Fejér-Riesz theorem is due to them in [**344**] and [**904**]. Its analog on the real axis, that is, $P(x) \geq 0$ on $\mathbb{R}$ implies $P(x) = \overline{Q(\bar{x})}Q(x)$ where Q has all its zeros in $\bar{\mathbb{C}}_+$, is well-known and must long predate their work. Daubechies uses the Fejér-Riesz theorem as a part of her construction of smooth wavelets of compact support (see [**237**, p. 172]). Geronimo-Woerdeman [**400**] discuss versions of the Fejér-Riesz theorem for two variables (it is not true without additional restrictions on the positive polynomial).

Toeplitz matrices and the Carathéodory-Toeplitz theorem are from their papers [**184**] and [**1045**]. They worked in terms of Taylor coefficients of Carathéodory functions rather than moments of measures. It can be viewed as Bochner's theorem for the group, $\mathbb{Z}$, since the positivity of $\{T^{(n)}\}$ is precisely the assertion that $n \to c_n$ is a positive definite function. Interestingly enough, most discussions of the history of Bochner's theorem (Lax [**687**] is an exception!) mention Herglotz's precursor for the circle group (and Weil's extension to general LCA groups and Raikov's to Banach algebras), but not Carathéodory-Toeplitz! Of course, it was Bochner who realized the centrality of this result to Fourier analysis.

Instead of appealing to the Riesz-Markov theorem, the measure can be constructed as a limit of explicit measures as follows: If $T^{(n)}$ is singular, there is a trivial measure (with at most $n-1$ points, the zeros of the polynomial P of degree $n-1$ with $\langle P, T^{(n)}P\rangle = 0$) whose moments are $\{c_j\}_{j=0}^n$. Decreasing c_0, we can write $T^{(n)} = a_n \mathbf{1} + \widetilde{T}^{(n)}$ where $\widetilde{T}^{(n)}$ is a positive singular Toeplitz matrix, and $a_n > 0$ and so get the c_j's as moments of a linear combination of $a_n \frac{d\theta}{2\pi}$ and the point measures associated with $\widetilde{T}^{(n)}$.

While Poisson had the formula for the electrostatic potential of a charge distribution on the sphere, which can be viewed as the modern Poisson formula for that case, the representation (1.3.17) is due to Fatou [**340**].

The Herglotz representation is due to Herglotz [**504**] and Riesz [**903**]. Its analog for the upper half complex plane has been used especially by Pick and Nevanlinna, so that the analogs of Carathéodory functions on $\mathbb{C}_+$ with positive imaginary part are called variously, Herglotz, Nevanlinna, or Pick functions. The standard textbook presentations of the proof use compactness (a.k.a. Helly selection theorem) to get a limit point of the $d\mu_r$'s, which is silly because, as noted, it is easy to see the moments all converge and the Laurent polynomials are dense.

The history of the various limit theorems for the boundary values of functions with representation (1.3.30) is involved, in part because much of the theory was developed for $\mathbb{R}$, not $\partial\mathbb{D}$, and in terms of limits of $\varepsilon^{-1}\mu(\theta_0 - \varepsilon, \theta_0 + \varepsilon)$ rather than of $F((1-\varepsilon)e^{i\theta})$. The names de la Vallée Poussin, Lebesgue, and Fatou should certainly be mentioned.

With one exception, all the results discussed in Subsections 6–8 are from Schur's great papers [**948, 949**] (parts I and II with a single section numbering). It is remarkable that in an area that had already been heavily studied by Riesz and many other very good analysts, Schur not only found something new, but had stunning insights. His results are very extensive, and include some determinant equalities that have been frequently used since. (Actually, these equalities appear in a different form in Sylvester [**1012**], but his way of looking at them was new; see [**30**].)

Because of the time lag from Schur's work until further study, some of his results have been ignored and are credited to others. For example, Khrushchev [**625**] regards the convergence on compact subsets of $\mathbb{D}$ of A_n/B_n to f as a significant result and focuses on the 1943 submission dates of papers of Wall [**1077**] and Geronimus [**404**]. But the result was found twenty-five years earlier by Schur! The convergence of the Schur approximants is Theorem 3.I in [**948**] and the identification of the Schur approximants with a ratio of polynomials with recursion relations (1.3.65)–(1.3.70) are in Section 14 of [**949**]. Schur even notes that two of his four polynomials are related via what we would now call the process of reversal. Thus what are called Wall polynomials predates Wall's work and are found in Schur [**949**]. That said, Wall and Geronimus fully appreciated the connection between Schur's algorithm and continued fractions and, in particular, that the Schur approximants are only half the natural continued fraction approximants. In particular, they realized the recursion formulae (1.3.65)–(1.3.70) can be viewed as cases of the Euler-Wallis formulae for continued fractions. Further discussion of OPUC from the point of view of continued fraction approximation can be found in a sequence of papers by Jones, Njåstad, and Thron [**577, 579, 580, 581, 582**].

(1.3.40) is used in one way of viewing the iterated Schur algorithm as a continued fraction, and from this point of view, A_n/B_n are only half the approximants; see [**625**].

There is an interesting derivation of Schur's recurrence relation, (1.3.48), from the point of view of Wall polynomials. By using (1.3.65) and (1.3.66), one sees the standard continued fraction formula that

$$\begin{aligned}\frac{A_n(z)}{B_n(z)}-\frac{A_{n-1}(z)}{B_{n-1}(z)}&=\frac{\gamma_n z[B_{n-1}(z)B_{n-1}^*(z)-A_{n-1}(z)A_{n-1}^*(z)]}{B_{n-1}(z)B_n(z)}\\&=\frac{z^n\gamma_n\prod_{j=0}^{n-1}(1-|\gamma_j|^2)}{B_{n-1}(z)B_n(z)}\end{aligned}$$

by (1.3.82). By induction and (1.3.66), $B_n(0)=1$. (1.3.48) follows immediately from this formula, (1.3.81), and the fact that A_{n-1}/B_{n-1} is a function of $\gamma_1,\dots,\gamma_{n-1}$.

The recursion relation (1.3.47) and the formula (1.3.48) for the relation of the Schur parameters to Taylor coefficients is from Section 2 of Schur's paper [**948**]. It is surprising that Verblunsky did not realize the connection to his work, which is discussed in Section 3.1 below.

The one refinement not in Schur [**948, 949**] is the precise estimate (1.3.44), (1.3.45), and (1.3.81). Schur, who estimated Taylor coefficients, had an additional factor of $(1-|z|)^{-1}$ on the right. The versions we give and their elegant proof via the Schwarz lemma are taken from Dym-Katsnelson [**320**].

Blaschke products were introduced by Blaschke [**126**]. Riesz [**906**] proved that if $f \in H^p(\mathbb{D})$ and f_B is the Blaschke product of its zeros, then $f = f_B h$ with h also in $H^p(\mathbb{D})$. The factorization (1.3.100) is due to Smirnov [**990**]. Beurling [**119**] coined the terms "inner" and "outer."

1.4. An Introduction to Operator and Spectral Theory

Operator and spectral theory on a Hilbert space are vast subjects which have parts of great importance to the study of OPUC. Our goal in this section is to sketch the most important aspects of the spectral theorem for unitary operators and the theory of trace ideals. This is, of course, no replacement for book-length treatments, of which I recommend Akhiezer-Glazman [**20, 21**], Dunford-Schwartz [**314**], Reed-Simon [**896, 897**], or Riesz-Sz.-Nagy [**908**] for spectral theory and Gohberg-Krein [**440**] or Simon [**962**] for trace ideals. In particular, we will not discuss unbounded selfadjoint operators. As in the last two sections, we eschew set-out definitions, theorems, and proofs. Throughout this section, operators are assumed to act on a separable Hilbert space, $\mathcal{H}$.

1. Selfadjoint, Normal, and Unitary Operators

Given an everywhere defined bounded operator, A, on a separable Hilbert space, $\mathcal{H}$, its *adjoint* is defined by $\langle A^*\varphi, \psi\rangle = \langle \varphi, A\psi\rangle$ for all $\varphi, \psi \in \mathcal{H}$. A is called *selfadjoint* if $A = A^*$ and *unitary* if $AA^* = A^*A = \mathbf{1}$. If $AA^* = A^*A$, we call A *normal*. Obviously, selfadjoint and unitary operators are normal.

We will occasionally need to use *antilinear operators* which obey $A(a\varphi + b\psi) = \bar{a}A\varphi + \bar{b}A\psi$. An *anti-unitary operator* is an antilinear operator, A, with $\|A\varphi\| = \|\varphi\|$ and $\operatorname{ran}(A) = \mathcal{H}$. It obeys $\langle A\varphi, A\psi\rangle = \langle \psi, \varphi\rangle$.

2. Spectrum, Resolvent, and Green's Functions

If B is a bounded operator, we say B is *invertible* if and only if there exists a bounded operator C with $BC = CB = \mathbf{1}$. We then write $B^{-1} = C$. Unlike the case for finite matrices, it is *not* sufficient that a one-sided inverse exist. For example, if L is the operator on $\ell^2(\mathbb{Z}_+)$ given by $L\delta_n = \delta_{n-1}$ $(= 0$ if $n = 0)$ and R is given by $R\delta_n = \delta_{n+1}$, then $LR = \mathbf{1}$, but neither L nor R has a two-sided inverse.

The set of invertible operators is open in the $\|\cdot\|$-topology of operators. Indeed, if $\|X\| < \|B^{-1}\|^{-1}$, then

$$(B+X)^{-1} = B^{-1}\sum_{j=0}^{\infty}(-1)^j (XB^{-1})^j \tag{1.4.1}$$

where the series is absolutely convergent since $\|XB^{-1}\| \leq \|X\|\,\|B^{-1}\| < 1$ by assumption.

If B is a bounded operator, the *resolvent set* of B, denoted by $\rho(B)$, is the set of $z \in \mathbb{C}$ with $B - z$ invertible. The *spectrum*, $\sigma(B)$, of B is the complement of $\rho(B)$. By (1.4.1), $\rho(B)$ is open, and so $\sigma(B)$ is closed. If $z > \|B\|$, $(B-z)^{-1} = -z^{-1}\sum_{j=0}^{\infty}(z^{-1}B)^j$ so $z \in \rho(B)$, that is,

$$\sigma(B) \subset \{z \mid |z| \leq \|B\|\} \tag{1.4.2}$$

If C is invertible, then BC is invertible if and only if B is invertible and $(BC)^{-1} = C^{-1}B^{-1}$. In particular, writing $(B - z) = -zB(B^{-1} - z^{-1})$, we see