Chapter One

Gems of Spectral Theory

The central theme of this monograph is the view of a remarkable 1915 theorem of Szegő as a result in spectral theory. We use this theme to present major aspects of the modern analytic theory of orthogonal polynomials. In this chapter, we bring together the major results that will flow from this theme.

1.1 WHAT IS SPECTRAL THEORY?

Broadly defined, spectral theory is the study of the relation of things to their spectral characteristics. By "things" here we mean mathematical objects, especially ones that model physical situations. Think of the brain modeled by a density function, or a piece of ocean with possible submarines again modeled by a density function. Other examples are the surface of a drum with some odd shape, a quantum particle interacting with some potential, or a vibrating string with a density function. To pass to more abstract mathematical objects, consider a differentiable manifold with Riemannian metric. To get into number theory, this manifold might have arithmetic significance, say, the upper half-plane with the Poincaré metric quotiented by a group of fractional linear transformations induced by some set of matrices with integral coefficients.

By spectral characteristics, mathematicians and physicists originally meant characteristic frequencies of the object—modes of vibration of the drum or, to state the example that gives the subject its name, the light spectrum produced by a chemical like Helium inside the sun.

Eventually, it was realized that besides the discrete set of frequencies associated with a drum, vibrating string, or compact Riemannian manifold, there were objects with continuous spectrum where the spectral characteristics become scattering or related data. For example, in the case of a brain, the spectral data is the raw output of a computer tomography machine. For quantum scattering on the line, it might be the reflection coefficient.

The process of going from the object to the spectral data or of going from some property of the object to some property of the data is called the *direct spectral problem* (or *direct problem*). The process of going from the spectral data to the object or from some aspect of the spectral data to some aspect of the object is the *inverse spectral problem* (or *inverse problem*).

The general wisdom is that direct problems are easier than inverse problems, and this is true on two levels: first, on the level of mere existence and/or even specifying the domain of definition; and second, in proving theorems that say if some property holds on one side, then some other property holds on the other.

Almost all these models (tomography is an exception) are described by a differential equation—ordinary or partial—or by a difference equation. In most cases, the object is a selfadjoint operator on some Hilbert space. In that case, the direct problem is usually solved via some variant of the spectral theorem, which says:

Theorem 1.1.1. *If A is a selfadjoint operator on a Hilbert space, $\mathcal{H}$, and $\varphi \in \mathcal{H}$, there is a measure $d\mu$ on $\mathbb{R}$ so that*

$$\langle \varphi, e^{-itA}\varphi \rangle = \int e^{-ixt}\, d\mu(x) \tag{1.1.1}$$

for all $t \in \mathbb{R}$.

Remarks. 1. All our Hilbert spaces are complex and $\langle \cdot, \cdot \rangle$ is linear in the second factor and antilinear in the first.

2. For a proof, see [14, 361, 369]. Also see Section 1.3 later for the case of bounded A.

3. I have ignored subtle points here when A is an unbounded operator (as happens for differential operators) concerning what it means to be selfadjoint, how e^{-itA} is defined, and so on. Because we look at difference equations in most of these notes, our A is bounded, and then for $n = 0, 1, 2, \ldots$, (1.1.1) is equivalent to

$$\langle \varphi, A^n \varphi \rangle = \int x^n\, d\mu(x) \tag{1.1.2}$$

4. We will also consider unitary operators, U, where $d\mu$ is now on $\partial\mathbb{D} = \{z \mid |z| = 1\}$ and

$$\langle \varphi, U^n \varphi \rangle = \int z^n\, d\mu(z) \tag{1.1.3}$$

for $n \in \mathbb{Z}$.

Notice that a spectral measure requires both an operator and a vector, φ. Sometimes there is a natural φ, sometimes not. Sometimes the full spectral measure is overkill—for example, the problem made famous by Mark Kac [212]: “Can you hear the shape of a drum?” asks about whether the eigenvalues of the Laplace–Beltrami operator of a (two-dimensional) compact surface determine the metric up to isometry. The spectral measure typically has point masses at the eigenvalues but also has weights for those masses so has more data than the eigenvalues alone.

It is worth noting that it is arguable whether the shape of a drum problem is a direct or an inverse problem. It asks if the direct map from isometry classes of manifolds to their eigenvalue spectrum is one-one. But on a different level, it asks if an inverse map exists!

By the way, the answer to Kac’s question is no (see [181]). For a review of more on this question and its higher-dimensional analogs, see [40, 64, 65, 180, 427].

Here is an example that shows we often do not understand the range of the direct map, and hence also the domain of the inverse map. Let $H_0 = -d^2/dx^2$ on $L^2(-\infty, \infty)$ and consider a function $V(x) \in L^1_{\text{loc}}(\mathbb{R})$ so that $(H_0 + 1)^{-1}(V + i)^{-1} \times (H_0+1)^{-1}$ is compact (e.g., this holds if $V(x) \to \infty$ as $|x| \to \infty$ but it also holds for $V = W^2 + W'$ with $W = x^2(2 + \sin(e^x))$ where V is unbounded below). Then

$$H = H_0 + V \tag{1.1.4}$$

has spectrum a set of eigenvalues $\{E_n\}_{n=1}^{\infty}$ where $E_n \to \infty$. It is well known that this is not sufficient spectral data to determine V.

Here is some additional data that is sufficient. Let H_D be H with a Dirichlet boundary condition at $x = 0$, that is,

$$H_D = H_D^+ \oplus H_D^- \tag{1.1.5}$$

where H_D^+ acts on $L^2(0, \infty)$ and H_D^- acts on $L^2(-\infty, 0)$, and selfadjointness is guaranteed by demanding $u(0) = 0$ boundary conditions.

Let E_n^D be the eigenvalues of H_D. It is not hard to prove the following:

(i) $E_n \le E_n^D \le E_{n+1}$

(ii) $E_n^D = E_n \Leftrightarrow u_n(0) = 0 \Leftrightarrow E_n^D = E_{n-1}^D$

Here u_n is the eigenfunction for H with eigenvalue E_n. Notice that (i) says each (E_n, E_{n+1}) contains at most one eigenvalue, and if there, it is simple. On the other hand, if $E_n^D \in \{E_j\}_{j=1}^{\infty}$, then it is a doubly degenerate eigenvalue.

If $E_n^D \in (E_n, E_{n+1})$, as noted E_n^D is simple, so we have a sign $\sigma_n^D \in \{\pm 1\}$, so E_n^D is an eigenvector of $H_D^{\sigma_n^D}$. If $E_n^D \in \{E_n, E_{n+1}\}$, σ_n^D is undefined. We will see shortly that $\{E_n\}_{n=1}^{\infty} \cup \{E_n^D, \sigma_n^D\}_{n=1}^{\infty}$ is a complete set of spectral data and that $\{V \mid E_n(V) = E_n(V_0)\}$ is an infinite-dimensional set of potentials. In a situation like this, where some set of the "spectral data" is distinguished but not determining, the set of objects whose spectral data in this subset is the same as for object$_0$ is called the *isospectral set* of object$_0$. It is usually a manifold, so we will often call it the *isospectral manifold* even if we have not proven it is a manifold!

Here is the theorem that describes what I have just indicated:

Theorem 1.1.2 ([165, 166])**.** *If $V, W \in L^1_{\text{loc}}$ and $E_n(V) = E_n(W)$, $E_n^D(V) = E_n^D(W)$, $\sigma_n^D(V) = \sigma_n^D(W)$, then $V = W$ (i.e., $V \mapsto \{E_n(V), E_n^D(V), \sigma_n^D(V)\}_{n=1}^{\infty}$ is one-one). Moreover, if $V \in L^1_{\text{loc}}$ and $N < \infty$ are given and $\tilde{E}_n$, $\tilde{E}_n^D$, σ_n^D are such that*

$$\begin{aligned} \tilde{E}_n &= E_n(V) && \textit{all } n \\ \tilde{E}_n^D &= E_n^D(V) && \textit{all } n > N \\ \tilde{\sigma}_n^D &= \sigma_n^D(V) && \textit{all } n > N \end{aligned}$$

$\{E_n, E_n^D\}$ *obey* (i) *and* (ii) *above, then there is a W with*

$$E_n(W) = \tilde{E}_n \qquad E_n^D(W) = \tilde{E}_n^D \qquad \sigma_n^D(W) = \sigma_n^D$$

for all n.

It is an interesting exercise to fix N and picture the topology of the allowed $\tilde{E}_n^D, \tilde{\sigma}_n^D$. Alas, it is not known precisely what direct data $\{\tilde{E}_n^D, \sigma_n^D\}$ can occur for a given V. It is definitely not all $\{\tilde{E}_n, \sigma_n^D\}$ obeying (i), (ii). For example, it cannot happen that

$$E_n^D = \tfrac{1}{4} E_n + \tfrac{3}{4} E_{n+1} \tag{1.1.6}$$

for all n.

Open Question 1. What is the range of the map $V \mapsto \{E_n(V), E_n^D(V), \sigma_n^D(V)\}$ as V runs through all L^1_{loc} functions with $(H_0+1)^{-1/2}(V+i)^{-1}(H_0+1)^{-1/2}$ compact, or through all continuous functions obeying $V(x) \to \infty$ as $|x| \to \infty$.

Even the most basic isospectral manifolds such as $V(x) = x^2$ where $E_n(V) = 2n+1$ are not understood.

Open Question 2. Prove that the isospectral manifold of continuous V's with $V(x) \to \infty$ as $x \to \infty$ and $E_n(V) = 2n+1$ is connected.

I have described this example in detail to emphasize how little we understand even some basic spectral problems.

Having set the stage with a very general overview, we are now going to focus in these notes on two classes of spectral problems: those associated with orthogonal polynomials on the real line (OPRL) and orthogonal polynomials on the unit circle (OPUC). These are the most simple and most basic of spectral setups for several reasons:

(a) As we will see, the construction of the inverse is not only simple and basic, but historically these problems appeared initially as what we will end up thinking of as an inverse problem.

(b) The objects are connected with difference—not differential—operators, so various technicalities that might cause difficulty concerning differentiability, unbounded operators, and so on are absent.

(c) They are, in essence, half-line problems; the parameters in the difference equation are indexed by $n = 1, 2, \ldots$ or $n = 0, 1, 2, \ldots$.

(c) is a virtue and a flaw. It is a virtue in that, as is typical for half-line problems, one can precisely describe the range of the direct map. It is a flaw in that the methods one develops are often not relevant to go to higher dimensions or, sometimes, even to whole-line problems.

OPRL appear initially in Section 1.2 and OPUC in Section 1.7.

Remarks and Historical Notes. The centrality of spectral theory to modern science can be seen by contemplating the variety of Nobel prizes that are related to the theory—from the 1915 physics prize awarded to the Braggs to the 1979 medicine prize for computer tomography.

1.2 OPRL AS A SOLUTION OF AN INVERSE PROBLEM

Let $d\rho$ be a measure on $\mathbb{R}$. All our measures will be positive with finite total weight. Normally, we will demand that ρ is a probability measure, that is, $\rho(\mathbb{R}) = 1$. But for now we only suppose $\rho(\mathbb{R}) < \infty$. ρ is called *trivial* if $L^2(\mathbb{R}, d\rho)$ is finite-dimensional; equivalently, if $\text{supp}(d\rho)$ is a finite set. Otherwise we call ρ *nontrivial*.

If

$$\int |x^n|\, d\rho(x) < \infty \tag{1.2.1}$$

for all n, we say $d\rho$ has *finite moments*. We will always suppose this. Indeed, we will soon mainly restrict ourselves to the case where ρ has bounded support.

If ρ is nontrivial with finite moments, every polynomial, P, obeys

$$0 < \int |P(x)|^2 \, d\rho(x) < \infty \tag{1.2.2}$$

since the integral can only be zero if ρ is supported on the finite set of zeros of P.

Thus, $\{x^n\}_{n=0}^\infty$ are independent in $L^2(\mathbb{R}, d\rho)$. They may or may not span L^2. If the support is bounded, they are spanning by the Weierstrass approximation theorem. In the case where the support is unbounded, there is a beautiful theory of when the polynomials span—it is presented in Section 3.8. One of the simplest examples of a case where they are not spanning is $\exp(-\sqrt{|x|}\,)\,dx$ (see Example 3.8.1 in Sections 3.8 and 3.9 for a discussion).

Thus, we can define monic orthogonal polynomials $\{P_n(x)\}_{n=0}^\infty$ of degree n by

$$P_n = \pi_n^\perp [x^n] \tag{1.2.3}$$

where π_n is the projection onto the n-dimensional space of polynomials of degree at most $n-1$ and

$$\pi_n^\perp = \mathbf{1} - \pi_n \tag{1.2.4}$$

So P_n is determined by

$$\begin{aligned} P_n(x) &= x^n + \text{lower order} \\ P_n \perp x^j \qquad & j = 0, \dots, n-1 \end{aligned} \tag{1.2.5}$$

By an obvious induction, we have

Proposition 1.2.1. $\{P_j\}_{j=0}^n$ *span* $\operatorname{Ran}(\pi_{n+1})$. *In particular,* $P_j/\|P_j\|$ *are an orthonormal basis of this* $n+1$*-dimensional space. So if* $Q \in \operatorname{Ran}(\pi_{n+1})$,

$$Q = \sum_{j=0}^n \langle P_j, Q\rangle \|P_j\|^{-2} P_j \tag{1.2.6}$$

One gets (1.2.6) by noting $Q-$rhs of (1.2.6) $\perp\ P_k$ for $k = 0, \dots, n$ since $\langle P_j, P_k\rangle = \|P_j\|^2 \delta_{jk}$. Here is a key fact for OPRL:

Proposition 1.2.2.

$$\langle P_j, x P_n\rangle = 0 \qquad \text{if } j < n-1 \tag{1.2.7}$$

Proof. $\langle P_j, x P_n\rangle = \langle x P_j, P_n\rangle = 0$ since $x P_j$ has degree $j+1 < n$. □

This leads to the recursion relation obeyed by OPRL:

Theorem 1.2.3. *For any nontrivial measure with finite moments, there exist* $\{b_j\}_{j=1}^\infty$ *in* $\mathbb{R}^\infty$ *and* $\{a_j\}_{j=1}^\infty$ *in* $(0,\infty)^\infty$ *so that for* $n \geq 0$,

$$x P_n(x) = P_{n+1}(x) + b_{n+1} P_n(x) + a_n^2 P_{n-1}(x) \tag{1.2.8}$$

where $P_{-1}(x) \equiv 0$ (*so we do not need* $a_n = 0$).

Proof. $xP_n(x) - P_{n+1}(x)$ is a polynomial of degree n (since the x^{n+1} terms cancel) and so orthogonal to P_{n+1}, that is,

$$\langle P_{n+1}, xP_n \rangle = \|P_{n+1}\|^2 \tag{1.2.9}$$

which means the coefficient of P_{n+1} in (1.2.6) with $Q = xP_n$ is 1. Moreover, the coefficient of P_{n-1} is

$$\langle P_{n-1}, xP_n \rangle \|P_{n-1}\|^{-2} = \langle P_n, xP_{n-1} \rangle \|P_{n-1}\|^{-2} \tag{1.2.10}$$

$$= \left(\frac{\|P_n\|}{\|P_{n-1}\|} \right)^2 \tag{1.2.11}$$

where (1.2.10) follows from the reality of P_j and x, and (1.2.11) uses (1.2.9) for n replaced by $n - 1$.

So we set

$$a_{n+1} = \frac{\|P_n\|}{\|P_{n+1}\|} \qquad b_{n+1} = \langle P_n, xP_n \rangle \|P_n\|^{-2} \tag{1.2.12}$$

and (1.2.6) becomes (1.2.8) on account of (1.2.7). □

The a_n's and b_n's are called *Jacobi parameters*. We start labeling with $n = 1$, but some authors start with $n = 0$ or even label b from $n = 0$ but a from $n = 1$. Also, some reverse the a's and b's or use other letters.

The formula (1.2.12) for a_n implies

Theorem 1.2.4. *We have that*

$$\|P_n\| = a_n \dots a_1 \rho(\mathbb{R}) \tag{1.2.13}$$

The orthonormal polynomials

$$p_n(x) = \frac{P_n(x)}{\|P_n\|} \tag{1.2.14}$$

obey

$$xp_n(x) = a_{n+1} p_{n+1}(x) + b_{n+1} p_n(x) + a_n p_{n-1}(x) \tag{1.2.15}$$

and multiplication by x in the orthonormal set $\{p_j\}_{j=0}^{\infty}$ has the matrix

$$J = \begin{pmatrix} b_1 & a_1 & 0 & & \\ a_1 & b_2 & a_2 & \ddots & \\ 0 & a_2 & b_3 & \ddots & \\ & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \tag{1.2.16}$$

Remarks. 1. Matrices of the form (1.2.16) are called *Jacobi matrices*.

2. When $\operatorname{supp}(d\rho)$ is bounded, $\{p_n\}_{n=0}^{\infty}$ is a basis, as we have seen. Shortly we will restrict to this case.

We now have our direct equation: $\{a_n, b_n\}_{n=1}^{\infty}$ defines a second-order difference equation for $n = 1, 2, 3, \dots,$

$$u_{n+1} = a_n^{-1}((\lambda - b_n)u_n - a_{n-1}u_{n-1}) \tag{1.2.17}$$

where a_0 is picked in a convenient way and λ is a parameter. The solution with

$$u_0 = 0 \qquad u_1 = 1 \tag{1.2.18}$$

is

$$u_n = p_{n-1}(\lambda) \tag{1.2.19}$$

In Section 1.3, we will turn to the direct problem of going from $\{a_n, b_n\}_{n=1}^{\infty}$ to $d\rho$, but we see that at the heart of OPRL is an inverse spectral problem! Central to this language is the idea that going from a difference/differential equation is a direct question.

We will eventually see (Section 3.2) that the inverse problem has a second method of solution. We note that the $P_n(x)$ for $d\rho$ and $c_0\, d\rho$ for any c_0 are the same and so also for Jacobi parameters. Thus, we will eventually mainly restrict to $\rho(\mathbb{R}) = 1$.

Before leaving this introduction, we want to discuss two other ways of understanding OPRL that actually work for positive measures on $\mathbb{C}$, so we pause to define OPs in that case. Let $d\zeta(z)$ be a positive measure on $\mathbb{C}$ so that

$$\int |z|^n \, d\zeta(z) < \infty \tag{1.2.20}$$

which is nontrivial (i.e., supp$(d\zeta)$ is not a finite set of points). Thus, we can form monic orthogonal polynomials $\Xi_n(z)$.

Unlike OPRL, $\Xi_n(z)$ do not obey a three-term recurrence relation because Proposition 1.2.2 uses reality (in general, $\langle \Xi_j, z\Xi_n\rangle = \langle \bar{z}\Xi_j, \Xi_n\rangle$). Indeed, only OPRL and OPUC (and polynomials for sets affinely related to $\mathbb{D}$ and $\partial\mathbb{D}$) are known to obey finite-order recursion relations, and so fit into the scheme of "spectral theory."

We note that $\{z^n\}_{n=0}^{\infty}$ may not span $L^2(\mathbb{C}, d\zeta)$ even if supp$(d\zeta)$ is bounded. For example, if there is an open set $U \subset \mathbb{C}$ and c so that

$$d\zeta \geq c\chi_U \, d^2z \tag{1.2.21}$$

then they are not dense since the closure of the set of polynomials is analytic on U (see the Notes). And, as we will see (Section 2.11, especially Theorem 2.11.5), for measures on $\partial\mathbb{D}$, the issue of density is subtle. But we can define $\{\Xi_n(z)\}_{n=0}^{\infty}$ in any event.

Theorem 1.2.5 (Christoffel Variational Principle)**.** *Let $\mathcal{M}_n$ be the monic polynomials of degree n, that is, $Q \in \mathcal{M}_n$ means*

$$Q(z) = z^n + \textit{lower order}$$

Then

$$\|\Xi_n\|^2 = \min_{Q \in \mathcal{M}_n} \|Q\|^2 \tag{1.2.22}$$

that is, for all $Q \in \mathcal{M}_n$,

$$\int |\Xi_n(z)|^2 \, d\zeta(z) \leq \int |Q(z)|^2 \, d\zeta(z) \tag{1.2.23}$$

with equality if and only if $Q = \Xi_n$.

Proof. This follows from the minimization property of orthogonalization, that is, if π is any orthogonal projection in a Hilbert space,

$$\|(1-\pi)u\|^2 = \min_{v\in \mathrm{Ran}(\pi)} \|u-v\|^2 \tag{1.2.24}$$

□

It is remarkable how powerful this principle is, given its simplicity.

The other general theorem concerns zeros.

Theorem 1.2.6. *Let $d\zeta$ be a measure obeying* (1.2.20) *and M_z multiplication by z on polynomials. Let π_n be the orthogonal projection in $L^2(\mathbb{C}, d\zeta)$ onto the n-dimensional space of polynomials of degree at most $n-1$. Let*

$$A = \pi_n M_z \pi_n \tag{1.2.25}$$

on $\mathrm{Ran}(\pi_n)$. *Then*

(i) *The eigenvalues of A are precisely the zeros of $\Xi_n(z)$.*
(ii) *Each eigenvalue of A has geometric multiplicity* 1.
(iii) *Each eigenvalue z_0 of A has algebraic multiplicity equal to the order of z_0 as a zero of $\Xi_n(z)$.*
(iv) *We have that*

$$\det(z-A) = \Xi_n(z) \tag{1.2.26}$$

Remark. Recall the geometric multiplicity of z_0 is the dimension of $\{v \mid (A-z_0)v = 0\}$. The algebraic multiplicity is the dimension of $\{v \mid (A-z_0)^\ell v = 0$ for some $\ell\}$. It is the order of the zero in $\det(z-A)$.

Proof. Let $v \in \mathrm{Ran}(\pi_{n+1})$. Then $\pi_n v = 0$ if and only if $v = c\,\Xi_n$. Thus, if $w \in \mathrm{Ran}(\pi_n)$, $w \neq 0$, then $(A-z_0)w = 0 \Leftrightarrow \pi_n(z-z_0)w = 0 \Leftrightarrow (z-z_0)w = c\,\Xi_n$. Moreover, $w \neq 0$ implies $(z-z_0)w \neq 0$, so $c \neq 0$.

$$\Xi_n(z) = c^{-1}(z-z_0)w \tag{1.2.27}$$

implies $\Xi_n(z_0) = 0$, so (i) is half proven. Conversely, if $\Xi(z_0) = 0$, (1.2.27) is solved precisely by

$$w(z) = \frac{c\,\Xi_n(z)}{z-z_0} \tag{1.2.28}$$

which lies in $\mathrm{Ran}(\pi_n)$. Thus, (i) is proven and so is (ii).

The same analysis shows $(A-z_0)^\ell w = 0$ with $(A-z_0)^{\ell-1}w \neq 0$ if and only if z_0 is a zero of $\Xi_n(z)$ of order at least ℓ, and this proves (iii).

(iv) holds since both sides are monic polynomials of degree n with the same zeros counting orders. □

Corollary 1.2.7 (Fejér's Theorem)**.** *Zeros of $\Xi_n(z)$ lie in the convex hull of* $\mathrm{supp}(d\zeta)$.

Proof. If $\Xi_n(z_0) = 0$, there is $w \in \operatorname{Ran}(\pi_n)$, with $\|w\|_{L^2} = 1$, so $\pi_n(z - z_0)w = 0$. Thus, $\langle w, (z - z_0)w\rangle = 0$, so

$$z_0 = \langle w, zw\rangle = \int z|w(z)|^2 \, d\zeta(z) \tag{1.2.29}$$

Since $\|w\| = 1$, $|w|^2 \, d\zeta$ is a probability measure, so the integral lies in the convex hull of $\operatorname{supp}(w^2 d\zeta)$, which lies in the convex hull of $\operatorname{supp}(d\zeta)$. □

Corollary 1.2.8. *Suppose that $d\rho$ is a measure on $\mathbb{R}$, with $a = \min \operatorname{supp}(d\rho)$, $b = \max \operatorname{supp}(d\rho)$. Then all the zeros of $P_n(x; d\rho)$ lie in $[a, b]$.*

Corollary 1.2.9. *Let $d\mu$ be a measure on $\partial\mathbb{D}$ and $\Phi_n(z; d\mu)$ the monic orthogonal polynomials. Then the zeros of Φ_n lie in $\overline{\mathbb{D}}$.*

Remark. One can show that if the convex hull of the support of $d\zeta$ does not lie in a straight line, then zeros lie in the interior of the convex hull of the support of the measure. In particular, in the case of Corollary 1.2.9, the zeros lie in $\mathbb{D}$, not merely $\overline{\mathbb{D}}$. We will prove this explicitly in Theorem 1.8.4.

Often, one has an explicit matrix representation of the operator A of (1.2.27), and so an explicit version of (1.2.24). For OPRL, one can take the basis $\{p_j\}_{j=0}^{n-1}$ and so get

Theorem 1.2.10. *Let $J_{n;F}$ be the $n \times n$ cutoff Jacobi matrix*

$$J_{n;F} = \begin{pmatrix} b_1 & a_1 & 0 & & & \\ a_1 & b_2 & a_2 & \ddots & & \\ 0 & a_2 & b_3 & \ddots & \ddots & \\ & & & \ddots & \ddots & \ddots \\ & & & & b_{n-1} & a_{n-1} \\ & & & & a_{n-1} & b_n \end{pmatrix} \tag{1.2.30}$$

Then

$$P_n(x) = \det(x - J_{n,F}) \tag{1.2.31}$$

Since $\det(x - A) = x^n - \operatorname{Tr}(A)x^{n-1} + O(x^{n-2})$ for $n \times n$ matrices, we see that

$$P_n(x) = x^n - \left(\sum_{j=1}^n b_j\right) x^{n-1} + O(x^{n-2}) \tag{1.2.32}$$

and, by (1.2.13)/(1.2.14),

$$p_n(x) = (a_1 \dots a_n)^{-1} \left[x^n - \left(\sum_{j=1}^n b_j\right) x^{n-1} \right] + O(x^{n-2}) \tag{1.2.33}$$

This provides another way of understanding the recursion (1.2.8). Expand $\det(x - J_{n+1,F})$ in minors in the last row. The minor of $x - b_{n+1}$ is $P_n(x)$ and the minor of $-a_n$ is $a_n P_{n-1}(x)$.

Remarks and Historical Notes. I would be remiss if I did not mention the "classical" OPRL: Jacobi, Laguerre, and Hermite associated, respectively, to the measures $(1+x)^\alpha(1-x)^\beta\,dx$ on $[-1,1]$ with $\alpha > -1$, $\beta > -1$, $x^\alpha e^{-x}$ on $[0,\infty)$ with $\alpha > -1$, and Hermite with $e^{-x^2}\,dx$. Jacobi polynomials with $\alpha = \beta = 0$ are Legendre, and with $|\alpha| = |\beta| = \frac{1}{2}$ are Chebyshev (of four kinds depending on the signs of α and β). Chebyshev with $\alpha = \beta = -\frac{1}{2}$ and $\alpha = \beta = \frac{1}{2}$ (of the first and second kind) will occur repeatedly later in these notes. They obey (up to normalization; U_n is normalized but not monic, while T_n is neither the normalized nor monic OP), respectively,

$$T_n(\cos\theta) = \cos(n\theta) \tag{1.2.34}$$

$$U_n(\cos\theta) = \frac{\sin((n+1)\theta)}{\sin\theta} \tag{1.2.35}$$

These and other specific examples are discussed in detail in Szegő [434] and Ismail [204].

The classical polynomials obey many other relations like the Rodriguez formula and second-order (in x) differential equations. This is specific to them; indeed, there is a theorem of Bochner (see [48, 188, 371] and [204, Section 20.1]) that says any set of orthogonal polynomials that obeys a second-order differential equation of the proper form is one of the classical ones!

The question of when $\{x^n\}_{n=0}^\infty$ are dense in $L^2(\mathbb{R}, d\rho)$ is intimately connected to the issue of determinacy of the moment problem discussed in the Notes to the next section. We will return to this issue in Sections 3.8 and 3.9.

Analyticity often places restrictions on the density of polynomials. If $U \subset \mathbb{C}$ is open and $d\zeta \geq c\chi_U\, d^2z$ for some measure on $\mathbb{C}$ for which (1.2.20) holds, then by the Cauchy integral formula, for any compact $K \subset U$, we have

$$\sup_{z\in K} |f(z)| \leq C_K \|f\|_{L^2(\mathbb{C},d\zeta)}$$

for any function analytic in U and in L^2. It follows that any f in the L^2-closure of the polynomials is analytic on U since the locally uniform limit of analytic functions is analytic. Thus, when (1.2.21) holds, the polynomials do not span L^2. A celebrated theorem of Mergelyan (for a proof, see Greene–Krantz [183, Ch. 12]) says that if K is compact, with $\mathbb{C}\setminus K$ connected, then the L^∞-closure of the polynomials is the functions continuous on K and analytic on K^{int}.

OPRL have their roots in work of Legendre, Gauss, and Jacobi. As a general abstract theory, the key figures were Chebyshev, Markov, Christoffel, and especially, Stieltjes. You can find more history in the books of Szegő [434], Chihara [82], Freud [141], Nevai [320], and Ismail [204].

Closely entwined to the history is the idea of continued fraction expansions of resolvents, an issue we return to in Sections 2.5 and 3.2 and which was pioneered by Jacobi for finite matrices (hence the name Jacobi matrix for (1.2.30)) and Stieltjes.

Variational principles like (1.2.22) for OPRL go back to Christoffel. Their use in OPUC with a twist (see Section 2.12 later) is due to Szegő [434]. As a spectral theory tool, they have been especially advocated and exploited by Freud [141] and Nevai [321].

That zeros of OPRL are eigenvalues of truncated Jacobi matrices is well known in the Schrödinger operator community. I am unsure who noted it first. The extension to measures on $\mathbb{C}$ where there is the complication of nontrivial algebraic multiplicity was arrived at in discussions I had with E. Brian Davies.

1.3 FAVARD'S THEOREM, THE SPECTRAL THEOREM, AND THE DIRECT PROBLEM FOR OPRL

What the orthogonal polynomial community calls Favard's theorem is the assertion that the map from measures on $\mathbb{R}$ (with finite moments) to Jacobi parameters is onto $\{a_n, b_n\}_{n=1}^{\infty}$ with $a_n > 0$ and $b_n \in \mathbb{R}$. It is intimately connected to the spectral theorem; indeed, we will prove the spectral theorem for bounded selfadjoint operators in this section (modulo some remarks in the Notes that go from Jacobi matrices to general operators). In the bounded case, we will see the map is also one-one if we restrict to probability measures.

Our discussion will be in three stages: first, finite Jacobi matrices, then bounded, and finally, unbounded (where we will assume, rather than prove, the spectral theorem).

Consider a trivial probability measure, that is,

$$d\rho = \sum_{j=1}^{N} \rho_j \delta_{x_j} \tag{1.3.1}$$

for

$$x_1 > x_2 > \cdots > x_N \tag{1.3.2}$$

and

$$\sum_{j=1}^{N} \rho_j = 1 \tag{1.3.3}$$

As usual, we can use Gram–Schmidt to define monic polynomials $P_0, \dots, P_{N-1}$ since our proof of independence of $\{x^j\}_{j=0}^{\infty}$ in the nontrivial case shows that $\{x^j\}_{j=0}^{N-1}$ are independent in this case. We can also use (1.2.3) to define $P_N(x)$ as the zero vector in $L^2(\mathbb{R}, d\rho)$, which, among monic Nth degree polynomials, is unique, namely,

$$P_N(x) = \prod_{j=1}^{N} (x - x_j) \tag{1.3.4}$$

The P's obey a recursion relation of the form (1.2.8) for $n = 0, 1, 2, \dots, N-1$ and so define $b_1, \dots, b_N, a_1, \dots, a_{N-1}$ and an $N \times N$ finite Jacobi matrix.

To go backward, we start with an $N \times N$ finite Jacobi matrix, that is, $\{b_j\}_{j=1}^{N}$ and $\{a_j\}_{j=1}^{N-1}$ are given with $a_j > 0$ and $b_j \in \mathbb{R}$, and we do not (yet) know they come from a measure.

We do not have a measure yet, so we cannot define P_j by orthogonality, but we do have recursion coefficients, so we define $\{P_j\}_{j=0}^N$ inductively by (1.2.8) with $P_0(x) \equiv 1$, $P_{-1}(x) \equiv 0$ (they could also be defined directly by (1.2.31)!), then p_j for $j = 0, 1, 2, \dots, N-1$ by $p_0(x) = 1$, and for $1 \le j \le N-1$,

$$p_j(x) = \frac{P_j(x)}{a_1 \dots a_j} \tag{1.3.5}$$

Then p_n obey (1.2.15) for $n = 0, 1, 2, \dots, N-2$ and

$$(b_N - x)p_{N-1}(x) + a_{N-1}p_{N-2}(x) = -(a_1 \dots a_{N-1})^{-1} P_N(x) \tag{1.3.6}$$

Proposition 1.3.1. *Let $J \equiv J_{N;F}$ be a finite Jacobi matrix given by* (1.2.30).

(a) *Define the vector $\vec{v}(x) \in \mathbb{C}^N$ by*

$$v_j(x) = p_{j-1}(x) \qquad j = 1, 2, \dots, N \tag{1.3.7}$$

Then

$$(J - x)\vec{v}(x) = -(a_1 \dots a_{N-1})^{-1}\delta_{jN} P_N(x) \tag{1.3.8}$$

(b) *If $\vec{w} \in \mathbb{C}^N$ obeys*

$$[(J - x)\vec{w}]_j = 0 \qquad j = 1, \dots, N-1 \tag{1.3.9}$$

then

$$w_j = w_1 p_{j-1}(x) \tag{1.3.10}$$

(c) *The eigenvalues of J are exactly the set of zeros of $P_N(x)$ and each zero has geometric multiplicity* 1.

(d) *The zeros of P_N are simple and real.*

(e) *If the zeros of P_N are labeled by* (1.3.2) *and*

$$(\varphi_\ell)_j = \frac{p_{j-1}(x_\ell)}{(\sum_{j=1}^N |p_{j-1}(x_\ell)|^2)^{1/2}} \tag{1.3.11}$$

then the φ_ℓ are an orthonormal basis of eigenvectors.

(f) *If*

$$\rho_\ell = |(\varphi_\ell)_1|^2 = \left(\sum_{j=1}^N |p_{j-1}(x_\ell)|^2 \right)^{-1} \tag{1.3.12}$$

then (1.3.3) *holds and $\{P_j(x)\}_{j=0}^N$ are the OPRL for the measure* (1.3.1).

Proof. (a) (1.3.8) is just (1.2.15) for $j = 1, \dots, N-1$ and (1.3.6) for $j = N$.

(b) (1.3.10) holds trivially for $j = 1$ and then inductively by subtracting (1.3.8) from (1.3.10), and noting this implies

$$(w_{j+1} - w_1 p_j(x)) = (a_j)^{-1}[(x - b_j)(w_j - w_1 p_{j-1}(x)) - a_{j-1}(w_{j-1} - w_1 p_{j-2}(x))] \tag{1.3.13}$$

for $j = 1, 2, \dots, N-1$ (with $a_{-1} \equiv 0$).

(c) Any eigenvector obeys (1.3.9) and so must be a multiple of $\vec{v}$. It obeys $[(J - x)\vec{v}(x)]_N = 0$ if and only if $P_N(x) = 0$ by (1.3.8). This argument shows any eigenvector is a multiple of φ_j given by (1.3.9), and so the geometric multiplicity is 1.

(d) Define φ_j by (1.3.9). Then $\langle \varphi_k, J\varphi_\ell \rangle = \langle J\varphi_k, \varphi_\ell \rangle$ implies, using $J\varphi_\ell = x_\ell \varphi_\ell$, that

$$(\bar{x}_k - x_\ell)\langle \varphi_k, \varphi_\ell \rangle = 0 \tag{1.3.14}$$

Taking $k = \ell$, we see x_k is real since $(\varphi_\ell)_1 \neq 0$ implies $\langle \varphi_\ell, \varphi_\ell \rangle \neq 0$.

To see that zeros are simple, suppose $P_N'(x_j) = 0$. Let

$$\vec{w} = \frac{\partial \vec{v}}{\partial x}\bigg|_{x=x_j} \tag{1.3.15}$$

(the components of v are polynomials, hence differentiable). Since $P_N'(x_1) = 0$, (1.3.8) implies

$$(J - x_j)w = v(x_j) \tag{1.3.16}$$

That cannot be since it implies

$$\begin{aligned} \langle v(x_j), v(x_j) \rangle &= \langle v(x_j), (J - x_j)w \rangle \\ &= \langle (J - x_j)v(x_j), w \rangle \\ &= 0 \end{aligned}$$

and $v_1(x_j) = 1$, so $\langle v(x_j), v(x_j) \rangle \neq 0$.

(e) $\|\varphi_\ell\|^2 = 1$ is immediate and $\langle \varphi_j, \varphi_\ell \rangle = 0$ for $j \neq \ell$ by (1.3.14). Since $P_N(x)$ has N zeros, the φ_ℓ's must span the space.

(f) Since $\{\varphi_\ell\}_{\ell=1}^N$ are an orthonormal basis, $U_{k\ell} = (\varphi_\ell)_k$ obeys

$$\sum_k \bar{U}_{k\ell} U_{kj} = \delta_{\ell j}$$

that is, $U^*U = \mathbf{1}$, so since it is finite-dimensional, $UU^* = 1$, that is (using $(\varphi_\ell)_j$ real to drop bars),

$$\sum_\ell (\varphi_\ell)_j (\varphi_\ell)_k = \delta_{jk} \tag{1.3.17}$$

This says, by the definitions (1.3.11) and (1.3.12),

$$\sum_\ell \rho_\ell \, p_{j-1}(x_\ell) p_{k-1}(x_\ell) = \delta_{jk} \tag{1.3.18}$$

Taking $j = k = 1$ and using $p_0(x) = 1$, we see that (1.3.3) holds and (1.3.18) implies that the $\{p_j\}_{j=0}^{N-1}$ are orthonormal polynomials for the measure (1.3.1), so $\{P_j\}_{j=0}^{N-1}$ are the monic OPRL. Since $P_N(x_j) = 0$, P_N is the monic OPRL for $d\rho$. $\square$

Remarks. 1. To be self-contained, we have given the standard argument that symmetric matrices have real eigenvalues and have algebraic multiplicities equal to geometric ones.

2. Notice that we have, in essence, just proven the spectral theorem for finite Jacobi matrices.

3. For a more conventional proof that the zeros of OPRL are all real and simple, see Subsection 5 of Section 1.2 of [399].

We have thus proven

Theorem 1.3.2 (Favard's Theorem for Trivial Measures)**.** *Every finite $N \times N$ Jacobi matrix is the Jacobi matrix of some measure supported on N points.*

Proof. (f) of the last theorem says the $\{P_j\}_{j=0}^N$ are the OPRL for $d\rho$ defined by (1.3.12) and $P_N(x_j) = 0$. The Jacobi parameters of P_j are the given Jacobi matrix since the polynomials alone obeying (1.2.8) determine a and b inductively by looking at the x^N and x^{N-1} terms on both sides of (1.2.8). For example, if $x_\ell^{(n)}$ are the roots of $P_n(x)$,

$$b_n = \left[\sum_{\ell=1}^{n} x_\ell^{(n)}\right] - \left[\sum_{\ell=1}^{n-1} x_\ell^{(n)}\right]$$

as will occur prominently in Section 8.5. □

Theorem 1.3.3. *The map from $d\rho$ of the form* (1.3.1)/(1.3.3) *to* $\{a_j\}_{j=0}^{N-1} \cup \{b_j\}_{j=0}^N$ *is one-one* (*and onto by Theorem 1.3.2*).

First Proof. Given the Jacobi matrix, J_N, of $d\rho$, following the construction of Theorem 1.3.2, construct a measure $d\rho' = \sum_{j=1}^N \rho_j' \delta_{x_j'}$. By construction, x_j' are the zeros of $P_N(x; d\rho)$, which are exactly the x_j's, that is, after renumbering $x_j' = x_j$. Moreover, the construction shows the normalized eigenvectors with positive first component are (1.3.11), so since φ_ℓ in $L^2(\mathbb{R}, d\rho)$ or $L^2(\mathbb{R}, d\rho')$ is the function $f(x) = \delta_{x_\ell x}$, we have

$$\begin{aligned} \rho_\ell &= \langle \varphi_\ell, 1 \rangle_{L^2(\mathbb{R}, d\rho)} \\ &= \langle \varphi_\ell, p_0 \rangle_{L^2(\mathbb{R}, d\rho)} \\ &= (\varphi_\ell)_1 \\ &= \text{given by (1.3.12)} \end{aligned}$$

showing $\rho_\ell = \rho_\ell'$. □

We want to give a second proof, not because this result is so important or so difficult, but because a slightly more involved proof will yield tools that are useful in the $N = \infty$ case.

Proposition 1.3.4. (a) *Two* (*probability*) *measures $d\rho, d\rho'$* (*supports can be infinite*) *have the same Jacobi parameters up to n, $\{a_j\}_{j=1}^{n-1} \cup \{b_j\}_{j=0}^n$, if and only if*

$$\int x^k \, d\rho = \int x^k \, d\rho' \tag{1.3.19}$$

$k = 0, 1, \dots, 2n - 1.$

(b) *Two measures, $d\rho, d\rho'$, each supported at n* (*possibly different*) *points are equal if and only if* (1.3.19) *holds for $k = 0, \dots, 2n - 1$.*

Proof. (a) By (1.2.8), we see that if Jacobi parameters are equal, then

$$P_j(x; d\rho) = P_j(x; d\rho') \tag{1.3.20}$$

Multiplying by x^ℓ, $\ell = 0, \dots, j-1$ and integrating, we see

$$\int x^{\ell+j} d\rho = \text{function of } \left\{ \int x^{\ell+k}\, d\rho \right\}_{k=0}^{j-1} \tag{1.3.21}$$

$$= \int x^{\ell+j}\, d\rho' \tag{1.3.22}$$

where the function is the same by (1.3.20), and (1.3.19) then follows by induction. As j runs from 0 to n and ℓ from 0 to $j-1$, $\ell + j$ goes from 0 to $2n-1$.

Conversely, if (1.3.19) for $k = 0, \dots, 2n-1$, the Gram matrices $\{\langle x^j, x^\ell \rangle\}_{0 \le j,\, \ell \le n-1}$ are equal, which, by the Gram–Schmidt process, implies $p_j(x; d\rho) = p_j(x; d\rho')$ for $0 \le j \le n-1$, and so

$$P_j(x; d\rho) = P_j(x; d\rho') \tag{1.3.23}$$

Since

$$P_n(x) = x^n - \sum_{j=0}^{n-1} \langle p_j, x^n \rangle p_j(x)$$

the moments $\int x^{n+\ell}\, d\rho$, $\ell = 0, \dots, n$, then also determine P_n so (1.3.23) also holds for $j = n$. As noted above, the polynomials determine the a's and b's in the recursion relation.

(b) As noted in (a), the stated moments determine $P_N(x)$ and so its zeros, and so $\{x_j\}_{j=1}^N$ and $\{x'_j\}_{j=1}^N$ are identical sets. Then the ρ's are determined by the equations

$$\sum_{j=1}^{N} \rho_j x_j^{\ell-1} = \int x^{\ell-1}\, d\rho \tag{1.3.24}$$

for $\ell = 1, 2, \dots, N$ since the Vandermonde determinant

$$\det(x_j^{\ell-1}) = \prod_{i<j} (x_i - x_j) \tag{1.3.25}$$

is nonzero. □

Second Proof of Theorem 1.3.3. By (a) of Proposition 1.3.4, the Jacobi parameters determine the first $2n$ moments and then, knowing the support is n points, the measures by (b) of the proposition. □

One can combine Theorems 1.3.2 and 1.3.3 and more in

Theorem 1.3.5. *Fix N. There is a one-one correspondence among each of*

(i) *Jacobi parameters $\{a_j\}_{j=1}^{N-1} \cup \{b_j\}_{j=1}^{N}$ with $b_j \in \mathbb{R}$ and $a_j > 0$.*
(ii) *Trivial measures of the form* (1.3.1) *where* (1.3.3) *holds and $\rho_j > 0$.*
(iii) *Unitary equivalence classes of symmetric $N \times N$ matrices A with a distinguished cyclic vector, φ.*

Remarks. 1. φ is called cyclic if $\{A^j\varphi\}_{j=0}^{\infty}$ span the space. For $N \times N$ matrices, we can instead take $\{A^j\varphi\}_{j=0}^{N-1}$ since if $P(A)$ is the (monic) characteristic polynomial of A, $P(A)A^\ell\varphi = 0$ shows inductively that $\{A^j\varphi\}_{j=N}^{\infty}$ are functions of $\{A^j\varphi\}_{j=0}^{N-1}$.

2. (A, φ) and (A', φ') are unitarily equivalent if and only if there is a unitary $U\colon \mathbb{C}^N \to \mathbb{C}^N$ so $UAU^{-1} = A'$ and $U\varphi = \varphi'$.

Proof. (i) $\Leftrightarrow$ (ii) is precisely the construction of Section 1.2 combined with Theorems 1.3.2 and 1.3.3.

It is easy to see that $\delta_1 = (1, 0, \dots, 0)^t$ is cyclic for a finite Jacobi matrix J. Indeed, if $\{p_\ell\}_{\ell=0}^{n-1}$ are the orthonormal polynomials, then $\delta_\ell = p_{\ell-1}(J)\delta_1$, so each Jacobi matrix with distinguished δ_1 is in an equivalence class.

Conversely, if φ is cyclic for A, $\{A^j\varphi\}_{j=0}^{N-1}$ must be independent (since they span $\mathbb{C}^N$). Thus, by Gram–Schmidt, we can find polynomials $\{p_j(A)\}_{j=0}^{N-1}$ with $p_0(A) = \mathbf{1}$ so $\varphi_j = p_{j-1}(A)\varphi$, $j = 0, \dots, N-1$, is an orthonormal basis. By the Gram–Schmidt construction, $\langle A^k\varphi, p_j(A)\varphi\rangle = 0$ if $k < j$. So by the same argument as in Section 1.2, there are constant $\{b_j\}_{j=1}^{N}$, $\{a_j\}_{j=1}^{N-1}$, so

$$A\varphi_j = a_{j+1}\varphi_{j+1} + b_{j+1}\varphi_j + a_j\varphi_{j-1} \tag{1.3.26}$$

for $j = 0, \dots, N-1$ where we interpret a_N and a_0 as 0. Thus, $\langle \varphi_j, A\varphi_k\rangle$ is a Jacobi matrix! The construction is unitarily invariant so the map is from equivalence classes to Jacobi matrices.

The two constructions are inverses showing the one-one correspondence. $\square$

Now we turn to the case of bounded semi-infinite Jacobi matrices.

Proposition 1.3.6. *A Jacobi matrix* (1.2.16) *is bounded on* ℓ^2 *if and only if*

$$\sup_n |a_n| + \sup_n |b_n| < \infty \tag{1.3.27}$$

Proof. Let δ_n be the vector with components δ_{nj}. $b_n = \langle \delta_n, J\delta_n\rangle$ while $a_n = \langle \delta_{n+1}, J\delta_n\rangle$ so $|b_n| \leq \|J\|$ and $|a_n| \leq \|J\|$. Thus, J bounded implies (1.3.27). A diagonal matrix $D = \{d_n\delta_{nm}\}$ has $\|D\| = \sup_n |d_n|$, and if A, B are the diagonal matrices with elements a and b, and if $S\delta_n = \delta_{n+1}$, then

$$J = AS^* + B + SA \tag{1.3.28}$$

so

$$\|J\| \leq 2\sup_n |a_n| + \sup_n |b_n| \tag{1.3.29}$$

We have thus proven

$$\sup_n |a_n| + \sup_n |b_n| \leq 2\|J\| \leq 4\big(\sup_n |a_n| + \sup_n |b_n|\big) \tag{1.3.30}$$

$\square$

We can now turn to the main theorem of this section (given our interest in the bounded support regime):

Theorem 1.3.7 (Favard's Theorem for Bounded Jacobi Matrices)**.** *Let* $\{a_n\}_{n=1}^\infty$, $\{b_n\}_{n=1}^\infty$ *be a set of Jacobi parameters obeying* (1.3.27). *Then there is a nontrivial measure, $d\rho$, of bounded support so that its Jacobi parameters are the given ones.*

Proof. Let J be a Jacobi matrix and $J_{n;F}$ its finite truncations. By Theorem 1.3.2, there are trivial n-point measures, $d\rho_n$, whose Jacobi parameters are $\{a_j\}_{j=0}^{n-1} \cup \{b_j\}_{j=0}^{n}$. By Proposition 1.3.4,

$$\int x^\ell \, d\rho_n = \int x^\ell \, d\rho_{n'} \tag{1.3.31}$$

for $\ell = 0, 1, \dots, 2\min(n, n') - 1$. In particular, for each ℓ, $\int x^\ell \, d\rho_n$ is constant for n large, so

$$\lim_{n\to\infty} \int x^\ell \, d\rho_n$$

exists for each ℓ.

By construction, $d\rho_n$ is supported on the eigenvalues of $J_{n;F}$ and so on $[-\|J_{n;F}\|, \|J_{n;F}\|]$, and so on $[-\|J\|, \|J\|]$. Thus, the $d\rho_n$'s are supported in a fixed compact set. Since the polynomials are dense in $C([-\|J\|, \|J\|])$, the probability measures, $d\rho_n$, have a weak limit $d\rho$. This weak limit, by (1.3.31), obeys

$$\int x^\ell \, d\rho_n = \int x^\ell \, d\rho \qquad \ell = 0, \dots, 2n-1 \tag{1.3.32}$$

By Proposition 1.3.4, the Jacobi parameters of $d\rho$ are J. □

Remark. Modulo discussion in the Notes, we have just proven the spectral theorem for bounded operators!

In the following, we could also discuss cyclic vectors, but we will not (see the Notes):

Theorem 1.3.8. *There is a one-one correspondence between bounded Jacobi matrices and nontrivial probability measures of bounded support under the map of measures to Jacobi parameters.*

Proof. Clearly, if $d\rho$ has support $[-C, C]$, then

$$|b_n| \le \int |x|\,|p_n(x)|^2 \, d\rho \le C$$

$$|a_n| \le \int |x|\,|p_n(x)|\,|p_{n-1}(x)| \, d\rho \le C$$

so J is bounded. By Favard's theorem, the map from measures of bounded support to bounded Jacobi parameters is onto. By Proposition 1.3.4, it is one-one. □

In this monograph, we are mainly interested in the bounded support case, so we will state Favard's theorem in the unbounded case without giving the proof for now. We will essentially prove it in Section 3.8; see Theorem 3.8.4.

Theorem 1.3.9 (Favard's Theorem)**.** *For any set of Jacobi parameters, there is a measure, $d\rho$, on $\mathbb{R}$ with $\int |x|^n\, d\rho(x) < \infty$ for all n, which has those Jacobi parameters.*

The measure may not be unique. This is discussed in Sections 3.8 and 3.9.

Remarks and Historical Notes. Favard's theorem is named after Favard [128] but goes back to Stieltjes [422]. The close connection to the spectral theorem also predates Favard in work of Stone [423] and Wintner [461]; see also Natanson [313], Perron [345], Sherman [382], and the discussion in Marcellán and Álvarez-Nodarse [295]. I am not aware of the approach here appearing elsewhere, but it will not surprise experts and I suspect is known to some.

Given any bounded selfadjoint operator, A, on a separable Hilbert space, $\mathcal{H}$, it is not hard to see that one can find $\{\varphi_j\}_{j=1}^N$ (N finite or infinite) so that for any $\ell, m, j \neq k$, $\langle A^\ell \varphi_j, A^m \varphi_k\rangle = 0$ and so that $\{A^\ell \varphi_j\}_{j,\ell}$ span $\mathcal{H}$. Thus, Theorem 1.3.7 and Gram–Schmidt imply there is a unitary U from $\mathcal{H}$ onto $\oplus_{j=1}^N L^2(\mathbb{R}, d\mu_j)$ so that $(UAU^{-1} f)_m(x) = x f_m(x)$. This is the spectral theorem for bounded operators.

The same idea shows that if A has a cyclic vector, φ, then applying Gram–Schmidt to $\{A^j \varphi\}_{j=0}^\infty$ yields an orthonormal basis in which J is a cyclic vector, allowing the two-part equivalence of Theorem 1.3.8 to extend to the three-part equivalence of Theorem 1.3.5.

1.4 GEMS OF SPECTRAL THEORY

In order to explain what I will mean by a gem of spectral theory, I begin by describing a pair of beautiful theorems in the spectral theory of OPRL:

Theorem 1.4.1 (Blumenthal–Weyl)**.** *Let J be a Jacobi matrix with Jacobi parameters $\{a_n, b_n\}_{n=1}^\infty$. If*

$$a_n \to 1 \qquad \text{and} \qquad b_n \to 0 \tag{1.4.1}$$

then

$$\sigma_{\text{ess}}(J) = [-2, 2] \tag{1.4.2}$$

Remarks. 1. Recall (see Reed–Simon [364, Section XIII.4]) that σ_{ess} is defined by $\sigma_{\text{ess}}(J) = \sigma(J) \setminus \sigma_d(J)$, where $\sigma(J)$, the spectrum of J, is $\{\lambda \mid (J - \lambda)$ does not have a bounded inverse$\}$, and $\sigma_d(J)$ are isolated points λ_0 of $\sigma(J)$, where $\oint_{|z-\lambda_0|=\varepsilon} (z - J)^{-1}\, dz$ is finite rank. For J's with cyclic vector (like Jacobi matrices) and spectral measure $d\rho$, $\sigma_{\text{ess}}(J)$ is the set of nonisolated points of $\text{supp}(d\rho)$.

2. See the Notes for a discussion of proof and history.

3. For any $a, b \in \mathbb{R}$ with $a > 0$, $N(a, b)$, the *Nevai class*, is the set of measures where $a_n \to a$, $b_n \to b$. By scaling, $\sigma_{\text{ess}}(J) = [b - 2a, b + 2a]$ if $J \in N(a, b)$.

Theorem 1.4.2 (Denisov–Rakhmanov). *Let J be a Jacobi matrix with measure $d\rho$ and Jacobi parameters $\{a_n, b_n\}_{n=1}^{\infty}$. Suppose* (1.4.2) *holds and*

$$d\rho(x) = f(x)\,dx + d\rho_{\mathrm{s}}(x) \tag{1.4.3}$$

where $d\rho_{\mathrm{s}}$ is singular and (*modulo sets of measure* 0)

$$\{x \mid f(x) > 0\} = [-2, 2]$$

Then (1.4.1) *holds.*

Remark. See the Notes for a discussion of proof and history. We will return to this theorem and prove it in Section 7.6.

These theorems are illuminated by the following:

Example 1.4.3. Let $a_n \equiv \frac{1}{2}$ and b_n be the sequence $(1, -1, 1, 1, -1, -1, 1, 1, 1, -1, -1, -1, \dots)$, that is, 1 k times followed by -1 k times for $k = 1, 2, \dots$. It is not hard to show $\sigma(J) = \sigma_{\mathrm{ess}}(J) = [-2, 2]$, so (1.4.2) is not sufficient for (1.4.1) to hold. □

Thus, we have a pair of deep theorems that go in opposite directions, but they do not set up equivalences. This leads us to:

Definition. By a *gem of spectral theory*, I mean a theorem that describes a class of spectral data and a class of objects so that an object is in the second class if and only if its spectral data lie in the first class.

This idea will be illuminated as we describe gems for OPUC and for OPRL in Sections 1.8 and 1.10 and a nongem in Section 1.9. In a sense, the overriding purpose of this book is to explore gems of OPRL/OPUC that depend on sum rules with positive coefficients. As we will see, the focus is somewhat narrower than that! And we will discuss some descendants of Szegő's theorem that are not gems (yet).

Remarks and Historical Notes. I find that some listeners object strongly to my use of the term "gem." I respond that it is a definition and I add that for a mathematician, a definition is not something that can be "wrong." But if I called them the "Jims of Spectral Theory," I wouldn't get the same reaction. And, of course, I used gems because of its connotation. Gems of spectral theory are typically beautiful and hard—but there can be beautiful and hard results that are not necessary and sufficient: Theorem 1.4.2 comes to mind.

The Blumenthal–Weyl theorem is named after contributions of Blumenthal [46] and Weyl [457]; Denisov–Rakhmanov after results of Rakhmanov [358, 359] and Denisov [107]; see Sections 9.1 and 9.2 of [400] for further history.

Theorem 1.4.1 is a consequence of Weyl's theorem (see Reed–Simon [364, Section XIII.4]) that if C is compact and selfadjoint and A bounded and selfadjoint, then $\sigma_{\mathrm{ess}}(A + C) = \sigma_{\mathrm{ess}}(A)$. In Theorem 1.4.1, $A = J_0$, the Jacobi matrix with $a_n \equiv 1$, $b \equiv 0$, and $C = J - J_0$ is compact when (1.4.1) holds.

Rakhmanov's theorem for OPUC is proven in Chapter 9 of [400]. Theorem 1.4.2 is proven in Section 13.4 of that book. As mentioned, we will provide a proof of a more general result in Chapter 7 of the present monograph.

1.5 SUM RULES AND THE PLANCHEREL THEOREM

The basic tool we will use is to establish sum rules with positive terms. In this section, we illustrate this with the granddaddy of all spectral sum formulae: the fact that if $A = \{a_{ij}\}_{1\le i,j\le N}$ is a finite matrix and $\{\lambda_j\}_{j=1}^N$ are its eigenvalues, then

$$\sum_{j=1}^N \lambda_j = \operatorname{Tr}(A) \equiv \sum_{j=1}^N a_{jj} \tag{1.5.1}$$

The left side is spectral theoretic and the right side involves the coefficients of the object.

One standard proof of (1.5.1) is to prove invariance of trace under similarity and the fact that there is a similarity taking A to upper triangular (even Jordan) form. But for us, the "right" proof is to note that the λ_j are the roots of the characteristic polynomials, so

$$\det(\lambda\mathbf{1} - A) = \prod_{j=1}^N (\lambda - \lambda_j) \tag{1.5.2}$$

Since, by expanding the determinant

$$\det(\lambda\mathbf{1} - A) = \lambda^n - \operatorname{Tr}(A)\lambda^{n-1} + \cdots \tag{1.5.3}$$

we get (1.5.1). The idea that sum rules occur as Taylor coefficients of suitable analytic functions recurs throughout this book.

In the infinite-dimensional case, there are convergence and other issues. Let X be a Banach space. A bounded linear map $A\colon X \to X$ is called *finite rank* if $\operatorname{Ran}(A)$ is finite-dimensional. Every such map has the form

$$Ax = \sum_{j=1}^N \ell_j(x)x_j \tag{1.5.4}$$

For some $\{\ell_j\}_{j=1}^N \subset X^*$ and $\{x_j\}_{j=1}^N \subset X$. It is not hard to show that

$$\operatorname{Tr}(A) = \sum_{j=1}^N \ell_j(x_j) \tag{1.5.5}$$

is independent of the ℓ's and x's used in the representation (1.5.4) (essentially by the invariance of trace in the finite-dimensional case). One defines the trace norm of a finite-rank operator by

$$\|A\|_1 = \inf\left\{ \sum_{j=1}^n \|\ell_j\|_{X^*}\|x_j\|_X \;\middle|\; A = \sum \ell_j(\cdot)x_j \right\} \tag{1.5.6}$$

The nuclear operators, $N(X)$, are the completion of the finite-rank operators in $\|\cdot\|_1$. It is not hard to see that every such object is associated to an operator and that one can define $\operatorname{Tr}(\cdot)$ on $N(X)$ since

$$|\operatorname{Tr}(A)| \le \|A\|_1 \tag{1.5.7}$$

If X is a Hilbert space, then $N(X)$ is called the *trace class operators*. A celebrated theorem of Lidskii says that

Theorem 1.5.1 (Lidskii's Theorem)**.** *If A is a trace class operator on a Hilbert space, $\mathcal{H}$, then $\sigma_{\text{ess}}(A) = \{0\}$ and A has nonzero eigenvalues $\{\lambda_j\}_{j=1}^N$ (counting algebraic multiplicity) so that*

$$\sum_{j=1}^{N} \lambda_j = \text{Tr}(A) \tag{1.5.8}$$

There are two limitations to note. First, on general Banach spaces, this result is false. Indeed, there is a Banach space, X, with a nuclear operator A so that $A^2 = 0$ (so any eigenvalue is 0) but $\text{Tr}(A) = 1$! (See the Notes.)

Second, consider the operator, C, on ℓ^2, which is a direct sum $C_1 \oplus C_2 \oplus \dots$ of 2×2 matrices

$$C_j = \begin{pmatrix} \alpha_j & \alpha_j \\ -\alpha_j & -\alpha_j \end{pmatrix} \tag{1.5.9}$$

$C_j^2 = 0$, so C has only eigenvalue zero. Indeed, it is easy to see that $\sigma(C) = \{0\}$. If $\sum_{j=1}^\infty |\alpha_j| = \infty$, but $\alpha_j \to 0$, then C is compact but not trace class. The sum of the eigenvalues is 0. As for the "trace," the sum of the diagonal matrix elements of C is conditionally convergent to zero, so it looks like a success. But conditionally convergent sums can be rearranged to any value! And rearranged sums are just rearranged bases. The moral is that, due to cancellations, (1.5.8) is subtle as soon as one leaves trace class, and it is unlikely that there is any kind of necessary and sufficient condition directly related to (1.5.8).

However, positivity can rescue something. It is not hard to prove

Theorem 1.5.2. *Let A be a bounded selfadjoint operator on a Hilbert space. Then A^2 is trace class if and only if A has a pure point spectrum with eigenvalues $\{\lambda_j(A)\}_{j=1}^\infty$ obeying*

$$\sum_{j=1}^{\infty} \lambda_j(A)^2 < \infty \tag{1.5.10}$$

In fact, if one writes $\text{Tr}(A^2) = \infty$ if A^2 is not trace class and $\sum \lambda_j(A)^2 = \infty$ if A has any nonpoint spectrum, Theorem 1.5.2 comes from a sum rule

$$\text{Tr}(A^2) = \sum_j \lambda_j(A)^2 \tag{1.5.11}$$

There are no cancellations because of positivity.

On $\ell^2(\partial\mathbb{D}, \frac{d\theta}{2\pi})$, one can specialize to operators of the form

$$(A_f g)(\theta) = \int f(\theta - \psi) g(\psi)\, \frac{d\psi}{2\pi} \tag{1.5.12}$$

where $\theta - \psi$ is computed mod 2π. Then $\lambda_j(A_f)$ are the Fourier coefficients, Theorem 1.5.2 is the Plancherel theorem, and the sum rule (1.5.11) is Parseval's equality. As we will see in Section 2.11, Szegő's theorem can be viewed as a kind of nonlinear Plancherel theorem.

Remarks and Historical Notes. The view of Theorem 1.5.2 as a sum rule with positivity, and so a model of Szegő's theorem as a sum rule, has been pushed especially by Killip [222].

For a proof of Lidskii's theorem, see, for example, [389], which obtains it from an equality for trace class operators

$$\det(1+zA) = \prod_{j=1}^{\infty}(1+z\lambda_j(A)) \tag{1.5.13}$$

An analog of (1.5.14) for Hilbert–Schmidt integral operators, namely,

$$\det[(1+zA)e^{-zA}] = \prod_{j=1}^{\infty}[(1+z\lambda_j(A))e^{-z\lambda_j(A)}] \tag{1.5.14}$$

goes back to Carleman [74] in 1921. One can regard him as the father of Theorem 1.5.2. Lidskii's theorem is named after [281], although the theorem was found somewhat earlier by Grothendieck [187]. Unaware of Grothendieck's work, Simon [389] rediscovered his approach to the problem.

For an introduction to nuclear operators on a general Banach space, see Chapter 10 of Simon [390]. (This book also discusses trace class, Lidskii's theorem, and proves (1.5.13) and (1.5.14); another reference on those subjects is Gohberg–Krein [170].) In particular, the example mentioned of a nuclear operator with $A^2 = 0$, but $\mathrm{Tr}(A) = 1$ is from Grothendieck [186].

1.6 PÓLYA'S CONJECTURE AND SZEGŐ'S THEOREM

Pólya and Szegő have linked names much like Hardy and Littlewood or Laurel and Hardy. This is most of all because of their great two-volume encyclopedia of analysis [353] and because, as part of Szegő's establishing of a great school of mathematics at Stanford, he brought Pólya to Palo Alto. But they are also linked in the initial history of the main theme of this monograph.

As we will see in Section 3.8, Hankel matrices, that is, finite matrices of the form $\{c_{j+k}\}_{jk=1}^{n}$ are fundamental to the theory of the moment problem on $\mathbb{R}$ (since they arise as Gram matrices for $\{x^j\}_{j=0}^{n-1}$). A *Toeplitz matrix*, T, is one of the form

$$t_{jk} = c_{j-k} \qquad 1 \le j,k \le n \tag{1.6.1}$$

Just as in the Hankel case, a situation of special interest is when c are the moments of a measure but now on $\partial\mathbb{D}$:

$$c_k = \int_0^{2\pi} e^{-ik\theta}\, d\mu(\theta) \tag{1.6.2}$$

We will, for now, restrict to the case $d\mu_{\mathrm{s}} = 0$ where

$$d\mu(\theta) = \frac{w(\theta)}{2\pi}\, d\theta + d\mu_{\mathrm{s}} \tag{1.6.3}$$

that is, to the case

$$c_k = \int e^{-ik\theta} w(\theta) \frac{d\theta}{2\pi} \tag{1.6.4}$$

Define $D_n(w)$ (more generally, $D_n(d\mu)$) to be the determinant of the $(n+1) \times (n+1)$ Toeplitz matrix

$$D_n(w) = \det \begin{vmatrix} c_0 & c_1 & \dots & c_n \\ c_{-1} & c_0 & \dots & c_{n-1} \\ \vdots & \vdots & & \vdots \\ c_{-n} & \dots & \dots & c_0 \end{vmatrix} \tag{1.6.5}$$

Because of a flurry of activity about moment problems on $\partial\mathbb{D}$ unleashed by Carathéodory in 1907 (see the Notes to Section 1.3 of [399]), Toeplitz matrices were all the rage from 1910–1915, and Pólya, a young postdoc, conjectured in [352] that if $w > 0$ and in L^1, then

$$\lim_{n\to\infty} D_n(w)^{1/n} = \exp\biggl(\int \log(w(\theta)) \frac{d\theta}{2\pi}\biggr) \tag{1.6.6}$$

In a visit back to his native Budapest, Pólya mentioned this conjecture to Szegő, then an undergraduate, and he proved the theorem below, published in 1915 [428]. At the time, Szegő was nineteen, and when the paper was published, he was serving in the Austrian Army in World War I! Here is the first version of Szegő's theorem:

Theorem 1.6.1 (Szegő's Theorem). *If* $w(\theta) \geq 0$ *and*

$$\int w(\theta) \frac{d\theta}{2\pi} < \infty \tag{1.6.7}$$

then (1.6.6) *holds.*

Remarks. 1. Since $\log_+(x) \equiv \max(0, \log(x)) \leq x$ (i.e., $x \leq e^x$ for $x \geq 1$), (1.6.7) implies $\int \log_+(w(\theta)) \frac{d\theta}{2\pi} < \infty$, so $\int \log(w(\theta)) \frac{d\theta}{2\pi}$ is either convergent or $-\infty$. In the latter case, we interpret the right side of (1.6.6) as 0.

2. Szegő (following a suggestion of Fekete) actually proved a stronger result, namely, that

$$\frac{D_{n+1}}{D_n} \to \text{RHS of (1.6.6)} \tag{1.6.8}$$

Since $D_n^{1/n} = D_0^{1/n} (\prod_{j=0}^{n-1} \frac{D_{j+1}}{D_j})^{1/n}$, (1.6.8) implies (1.6.6).

This theorem (in an extended form) is the subject of Chapter 2 where it is proven. For now, it does not appear to have a spectral content—its transformation to that form is the subject of the next two sections. But we note (1.6.6) is an equality (sum rule) with something involving a measure on one side and something rather different on the other, so my assertion that there is a gem lurking nearby should not be too surprising.

Remarks and Historical Notes. Not only did Szegő find the leading term in the asymptotics of $\log(D_n)$ in 1915, but he found the second term [433] thirty-seven years later! The strongest form of this second-term asymptotics is the following:

Theorem 1.6.2 (Sharp Form of the Strong Szegő Theorem)**.** *If* c_k *is given by* (1.6.4)*, if* $w(\theta) \geq 0$*, if* (1.6.7) *holds, and if*

$$\widehat{L}_k = \int_0^{2\pi} e^{-ik\theta} \log(w(\theta)) \, \frac{d\theta}{2\pi} \tag{1.6.9}$$

then

$$\lim_{n\to\infty} \frac{D_n}{e^{(n+1)\widehat{L}_0}} = \exp\left(\sum_{k=1}^{\infty} k|\widehat{L}_k|^2 \right) \tag{1.6.10}$$

Chapter 6 of [399] has six different proofs of this theorem, due in this strong form to Ibragimov [203]. There is a seventh proof in Section 9.10 of [400]; see also [401]. In Section 1.12, we show this implies a gem. We note that there are no general terms in an asymptotic series—for nonvanishing analytic w's, after the first two terms, the error is $O(e^{-cn})$.

Since $\log(\det(A)) = \mathrm{Tr}(\log(A))$ for positive matrices, A, if $T_{n+1}(w)$ is the matrix whose determinant is in (1.1.5), then (1.6.6) can be rewritten:

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Tr}(\log(T_n(w))) = \int \log(w(\theta)) \, \frac{d\theta}{2\pi} \tag{1.6.11}$$

More generally, one can prove that (see Theorem 2.7.13 of [399])

Theorem 1.6.3. *If* f *is a continuous function on* $[0, \infty)$ *with* $\lim_{x\to\infty} f(x)/x = 0$*, then*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Tr}(f(T_n(w))) = \int f(w(\theta)) \, \frac{d\theta}{2\pi} \tag{1.6.12}$$

We will focus on Szegő's theorem and its descendants within spectral theory, but it has given birth to many other children. In the period 1950–1970, it was a major theme in a program called Function Algebras looking at fairly abstract Banach algebras. This work is discussed in the Notes to Section 2.6 of [399].

1.7 OPUC AND SZEGŐ'S RESTATEMENT

In 1920, Szegő [430] revisited his theorem realizing, in part, that it could be restated in terms of orthogonal polynomials on the unit circle (OPUC). We will discuss that here. Another critical realization in that paper—rephrasing the theorem as a variational principle—will be discussed in Section 2.12. A third—asymptotics of OPUC—will appear in Section 2.9.

Let $\{f_j\}_{j=1}^N$ be a set of independent vectors in a Hilbert space. Let $\{g_j\}_{j=1}^N$ be the set obtained by unnormalized Gram–Schmidt, that is,

$$g_j = f_j + \sum_{k=1}^{j-1} h_{jk} f_k \tag{1.7.1}$$

and

$$\langle g_j, g_k \rangle = 0 \qquad \text{if } j \neq k \tag{1.7.2}$$

Let H be the $N \times N$ matrix

$$h_{jk} = \begin{cases} 1 & k = j \\ h_{jk} & k < j \\ 0 & k > j \end{cases} \tag{1.7.3}$$

Then the Gram matrices are clearly related by

$$G(f)_{jk} \equiv \langle f_j, f_k \rangle \qquad G(g)_{jk} \equiv \langle g_j, g_k \rangle \tag{1.7.4}$$

$$G(g) = H^* G(f) H \tag{1.7.5}$$

We thus conclude:

Theorem 1.7.1.

$$\det(G(f)) = \prod_{j=1}^{N} \|g_j\|^2 \tag{1.7.6}$$

Proof. By (1.7.3),

$$\det(H) = \det(H^*) = 1$$

and by (1.7.5),

$$\begin{aligned} \det(G(g)) &= |\det(H)|^2 \det(G(f)) \\ &= \det(G(f)) \end{aligned}$$

Since $G(g)_{jk} = \|g_j\|^2 \delta_{jk}$, (1.7.6) is immediate. □

Given a nontrivial measure, $d\mu$, on $\partial\mathbb{D}$, define the monic OPUC by

$$\Phi_n(z) = z^n + \text{ lower order} \tag{1.7.7}$$

$$\int \bar{z}^j \Phi_n(z)\, d\mu(z) = 0 \qquad j = 0, 1, \dots, n-1 \tag{1.7.8}$$

We will use $\Phi_n(z; d\mu)$ if we want the $d\mu$ dependence to be explicit.

Thus, if $f_j = z^{j-1}$, $j = 1, \dots, N$, and $g_j = \Phi_{j-1}$, $j = 1, \dots, N$, we see that f and g are related by (1.7.1)/(1.7.2). Recognizing $G(f)$ as the Toeplitz matrix with

$$c_{k-j} = \langle z^k, z^j \rangle = \int e^{-i(k-j)\theta}\, d\mu(\theta) \tag{1.7.9}$$

we obtain from (1.7.6) that

Corollary 1.7.2. *The* $(N+1) \times (N+1)$ *Toeplitz determinant,* $D_N(d\mu)$*, obeys*

$$D_N(d\mu) = \prod_{j=0}^{N} \|\Phi_j(z; d\mu)\|^2 \tag{1.7.10}$$

Szegő also realized a special feature of OPUC that makes Fekete's remark that D_{N+1}/D_N has the same limit as $D_N^{1/N}$ transparent, namely,

Proposition 1.7.3. *For each n,*

$$\|\Phi_n\| \leq \|\Phi_{n-1}\| \tag{1.7.11}$$

Thus, $\lim_{n\to\infty} \|\Phi_n\|^2$ *exists and equals*

$$\lim_{N\to\infty} D_N(d\mu)^{1/N} \tag{1.7.12}$$

Proof. Since Φ_n is orthogonal to any polynomial of degree $n-1$, it minimizes $\{\|\Phi_n + g\| \mid \deg(g) \leq n-1\}$. As a consequence,

$$\|\Phi_n\| = \min\{\|P_n\| \mid P_n(z) = z^n + \text{lower order}\}$$

Thus, since $z\Phi_{n-1}$ is monic and $|z| = 1$ on $\operatorname{supp}(d\mu)$,

$$\|\Phi_n\| \leq \|z\Phi_{n-1}\| = \|\Phi_{n-1}\|$$

proving (1.7.11). Since $\|\Phi_n\|$ is decreasing and positive, it has a limit and, of course, $(\|\Phi_0\|^2 \dots \|\Phi_n\|^2)^{1/n}$ then converges to $\lim \|\Phi_n\|^2$. $\square$

We thus see an equivalent form of Szegő's theorem, Theorem 1.6.1:

Theorem 1.7.4. (1.6.6) *is equivalent to*

$$\lim_{n\to\infty} \left\| \Phi_n\left(z, \frac{w(\theta)}{2\pi}\, d\theta\right)\right\|^2 = \exp\left(\int \log(w(\theta))\, \frac{d\theta}{2\pi}\right) \tag{1.7.13}$$

Remarks and Historical Notes. Szegő's great 1920–1921 paper [430] was the first systematic exploration of OPUC, although he had earlier discussed OPs on curves [429].

1.8 VERBLUNSKY'S FORM OF SZEGŐ'S THEOREM

In this section, we give the final reformulation of Szegő's theorem as a sum rule and see that it implies a gem of spectral theory. The first element we need is the recursion relation obeyed by the monic OPUC, $\Phi_n(z)$, that will give us the parameters of the direct problem.

We first define natural maps $\delta_n \colon L^2(\partial\mathbb{D}, d\mu)$ to itself by

$$(\delta_n f)(e^{i\theta}) = e^{in\theta}\, \overline{f(e^{i\theta})} \tag{1.8.1}$$

Proposition 1.8.1. (i) *δ_n is an anti-unitary map of L^2 to L^2.*

(ii) *If π_n is the orthogonal projection onto the span of $\{z^j\}_{j=0}^{n-1}$, then δ_n maps* $\operatorname{Ran}(\pi_{n+1})$ *to itself. Indeed, if $P \in \operatorname{Ran}(\pi_{n+1})$, then*

$$(\delta_n P)(z) = z^n\, \overline{P(1/\bar{z})} \tag{1.8.2}$$

Equivalently, if

$$P(z) = \sum_{j=0}^{n} c_j z^j \tag{1.8.3}$$

then

$$(\delta_n P)(z) = \sum_{j=0}^{n} \bar{c}_{n-j}\, z^j \tag{1.8.4}$$

(iii) *If* $f \in \mathrm{Ran}(\pi_{n+1})$ *and* $f \perp \{z, z^2, \dots, z^n\}$, *then* f *is a multiple of* $\delta_n(\Phi_n)$.

Proof. (i) Multiplication by $e^{in\theta}$ is unitary and $f \to \bar{f}$ is anti-unitary.

(ii) Immediate from $\delta_n(z^j) = z^n \overline{z^j} = z^{n-j}$.

(iii) Since δ_n is anti-unitary and $\delta_n(z^j) = z^{n-j}$, $\delta_n f \perp \{1, z, \dots, z^{n-1}\}$, so $\delta_n f = c\Phi_n$, and thus $f = \delta_n^2(f) = \bar{c}\delta_n(\Phi_n)$. □

We now shift to standard, albeit unfortunate, notation and use Φ_n^* for $\delta_n(\Phi_n)$ and, more generally, P^* for $\delta_n(P)$. It is hoped that in context, the value of n is clear. But for $d\mu = \frac{d\theta}{2\pi}$, $\Phi_n(z) = z^n$, $\Phi_n^* = 1$, and $(\Phi_n^*)^* = z^n$ becomes $1^* = z^n$. The notation is awful but, as I said, standard.

Theorem 1.8.2 (Szegő Recursion Relations)**.** *For any nontrivial measure* $d\mu$ *on* $\partial\mathbb{D}$, *there exist constants* $\alpha_n \in \mathbb{C}$ *so that*

$$\Phi_{n+1} = z\Phi_n - \bar{\alpha}_n \Phi_n^* \tag{1.8.5}$$

Proof. Since Φ_n and Φ_{n+1} are monic, $\Phi_{n+1} - z\Phi_n$ is a polynomial of degree n. Moreover, if $j = 1, \dots, n$,

$$\langle z^j, \Phi_{n+1} - z\Phi_n\rangle = \langle z^j, \Phi_{n+1}\rangle - \langle z^{j-1}, \Phi_n\rangle = 0$$

so, by (iii) of Proposition 1.8.1, there is α_n, so (1.8.5) holds. □

Applying * on $n+1$ degree polynomials, we obtain

$$\Phi_{n+1}^* = \Phi_n^* - \alpha_n z \Phi_n \tag{1.8.6}$$

The $\{\alpha_n\}_{n=0}^\infty$ are called *Verblunsky coefficients*. They are the analog of the Jacobi parameters for OPRL. The reason for the minus sign and complex conjugate will become clear later (see Theorem 2.5.2). Setting $z = 0$ in (1.8.5) and using the fact that Φ_n monic implies

$$\Phi_n^*(0) = 1 \tag{1.8.7}$$

we have

$$\alpha_n = -\overline{\Phi_{n+1}(0)} \tag{1.8.8}$$

The following is critical:

Theorem 1.8.3. *We have that*

$$\|\Phi_{n+1}\|^2 = (1 - |\alpha_n|^2)\|\Phi_n\|^2 \tag{1.8.9}$$

so that

$$|\alpha_n| < 1 \tag{1.8.10}$$

and if $\mu(\partial\mathbb{D}) = 1$, *then*

$$\|\Phi_n\|^2 = \prod_{j=0}^{n-1}(1 - |\alpha_j|^2) \tag{1.8.11}$$

Remark. Of course, more generally,

$$\|\Phi_n\|^2 = \left[\prod_{j=0}^{n-1}(1 - |\alpha_j|^2)\right]\mu(\partial\mathbb{D}) \tag{1.8.12}$$

Proof.

$$\begin{aligned}\|\Phi_n\|^2 = \|z\Phi_n\|^2 &= \|\Phi_{n+1} + \bar{\alpha}_n\Phi_n^*\|^2 \\ &= \|\Phi_{n+1}\|^2 + |\alpha_n|^2\|\Phi_n\|^2 \end{aligned}\tag{1.8.13}$$

since $\|\Phi_n^*\| = \|\Phi_n\|$ and $\Phi_n^* \perp \Phi_{n+1}$. (1.8.13) implies (1.8.9). $\|\Phi_n\|^2 > 0$ implies (1.8.10) and (1.8.11) follows by induction. □

From (1.8.11) and (1.8.5)/(1.8.6), we obtain

$$\varphi_{n+1} = \rho_n^{-1}(z\varphi_n - \bar{\alpha}_n\varphi_n^*) \tag{1.8.14}$$

$$\varphi_{n+1}^* = \rho_n^{-1}(\varphi_n^* - \alpha_n z\varphi_n) \tag{1.8.15}$$

where

$$\rho_n = (1 - |\alpha_n|^2)^{1/2} \tag{1.8.16}$$

The same calculation that led to (1.8.13) implies

Theorem 1.8.4. *If* $\Phi_n(z_0) = 0$, *then* $|z_0| < 1$. *If* $\Phi_n^*(z_0) = 0$, *then* $|z_0| > 1$.

Proof. Since $|z_0| < 1 \Leftrightarrow |1/z_0| > 1$, the first sentence implies the second. If $\Phi_n(z_0) = 0$, let $P(z) = \Phi_n(z)/(z - z_0)$, which is a polynomial of degree $n - 1$, so orthogonal to Φ_n. Then

$$\begin{aligned}\|P\|^2 = \|zP\|^2 &= \|(z - z_0)P + z_0P\|^2 \\ &= \|\Phi_n + z_0P\|^2 \\ &= \|\Phi_n\|^2 + |z_0|^2\|P\|^2 \end{aligned}\tag{1.8.17}$$

Since $\|\Phi_n\|^2 > 0$, $|z_0| < 1$. □

By Theorem 1.8.3, $d\mu \mapsto \{\alpha_n(d\mu)\}_{n=0}^\infty$ maps the nontrivial measure to $\mathbb{D}^\infty$. The following is fundamental to thinking of OPUC as a spectral problem:

Theorem 1.8.5 (Verblunsky's Theorem)**.** *The map of* $d\mu \mapsto \{\alpha_n(d\mu)\}_{n=0}^\infty$ *is a one-one map of nontrivial probability measures onto* $\mathbb{D}^\infty$.

We will prove this in Section 2.5 (see Theorem 2.5.3); see also the Notes to this section. We can now state Verblunsky's form of Szegő's theorem; by (1.8.11), the limit on the left of (1.7.13) is just an infinite product:

Theorem 1.8.6 (Verblunsky's Form of Szegő's Theorem)**.** *For any nontrivial probability measure $d\mu$ on $\partial\mathbb{D}$ with w given by* (1.6.3)*, we have*

$$\prod_{n=0}^{\infty}(1-|\alpha_n|^2) = \exp\biggl(\int \log(w(\theta))\,\frac{d\theta}{2\pi}\biggr) \tag{1.8.18}$$

This is the version we will prove in Chapter 2; see Section 2.7. We note that it has two differences from Szegő's theorem, even the variant in Theorem 1.7.4. First, we have written it in terms of Verblunsky coefficients, and second, unlike Szegő's original version, this allows $d\mu_{\rm s} \neq 0$. One has the remarkable fact that the left side of (1.8.18) is independent of $d\mu_{\rm s}$!

(1.8.18) *always* holds, although both sides can be zero connected with a "divergent product" on the left and a diverging integral on the right. The two sides are nonzero at the same time, so we get the following gem:

Corollary 1.8.7. *For nontrivial probability measures $d\mu$ on $\partial\mathbb{D}$ obeying* (1.6.3)*,*

$$\sum_{n=0}^{\infty}|\alpha_n|^2 < \infty \Leftrightarrow \int \log(w(\theta))\,\frac{d\theta}{2\pi} > -\infty \tag{1.8.19}$$

Remarks and Historical Notes. The Szegő recursion, (1.8.5)/(1.8.6), appeared first in 1939 in his famous book on orthogonal polynomials [434]. But at roughly the same time, they appeared in work of Geronimus [156, 157]. The history is murky, but especially as their proofs and presentations are different, it seems like Geronimus' work was independent but several months later. Interestingly enough, an equivalent form was rediscovered by Levinson [277] about ten years later, and the engineering literature sometimes calls it the Levinson or Levinson–Szegő algorithm.

Five years before Szegő, the α_n appeared in work of Verblunsky in two remarkable papers [452, 453] that were mainly ignored for almost seventy years! Verblunsky did not define the α_n via a recursion relation, but in [452], he proved there were rational functions $\zeta_n(c_0, c_1, \dots, c_{n-1}; \bar{c}_0, \dots, \bar{c}_{n-1}) \in \mathbb{C}$ and $R_n(c_0, c_1, \dots, c_{n-1}; \bar{c}_0, \dots, \bar{c}_{n-1}) \in (0,\infty)$ so that if $\{c_j\}_{j=0}^{n-1}$ were moments of some nontrivial measure on $\partial\mathbb{D}$, then the allowed values of c_n for nontrivial measures were all the possible values in the open disk of radius R_n in $\mathbb{C}$ centered at ζ_n. He then defined α_{n-1} by

$$c_n = \zeta_n + \alpha_{n-1}R_n \tag{1.8.20}$$

This is discussed in Section 3.1 of [399]. Interestingly enough, the analog of this approach for OPRL was rediscovered by Krein [252], Karlin–Studden [213], and Krein–Nudel'man [253], and codified in a book by Dette–Studden [112] who included the analysis of OPUC, thus reinventing [452]!

Theorem 1.8.4 goes back to Szegő [430]. The proof we give is due to Landau [263]. [399] has six proofs of the theorem.

In [452], Verblunsky also proved Theorem 1.8.5 using his definition of $\{\alpha_n\}_{n=0}^\infty$. Other proofs of this theorem are presented in [399] and [398]. In particular, we mention the spectral theory proof, the analog of the proof of Favard's theorem that we gave in Section 1.3. Of course, for that we need an analog of Jacobi matrices. The proper analog, the CMV matrix, will be discussed in Section 2.11. It is due to Cantero, Moral, and Velázquez [70] but essentially was discovered earlier by Amar, Gragg, Reichel, and Watson (see [403]) as a tool in numerical matrix analysis. See Chapter 4 of [399] and [403] for further discussions. Before [399, 400] introduced "Verblunsky coefficient," the α_n's had a wide variety of names: reflection coefficient, Schur parameter, Szegő parameter, and Geronimus coefficient.

In [453], Verblunsky proved Theorem 1.8.6. In particular, he had the sum rule (1.8.18) and he had a proof that allowed a singular part of the measure. Much of the literature since has attributed this singular-part-allowed result to Kolmogorov and Krein, whose work was later and which only proved $\sum|\alpha_n|^2=\infty \Leftrightarrow \int \log(w(\theta))\frac{d\theta}{2\pi} = -\infty$ with a singular part allowed. Others attributed the general result to Geronimus or Szegő—again based on later work.

It is also true that KdV sum rules should be viewed as analogs of Verblunsky's sum rule, but the connection was not realized until many years later. Indeed, the Killip–Simon sum rules discussed in Section 1.10 were discovered in a chain going back to KdV sum rules without knowing of Verblunsky's work. It was in tracking down the history of (1.8.18) that we uncovered [452, 453].

One of the consequences of Corollary 1.8.7 is the existence of mixed spectrum consistent with ℓ^2 decay: Given any measure $d\rho_{\rm s}$ with $\int d\rho_{\rm s} < 1$, there is a measure with a.c. support all of $\partial\mathbb{D}$, that $d\rho_{\rm s}$, and with $\sum_{j=0}^\infty |\alpha_j|^2 < \infty$. Not knowing of this, the existence of analogous mixed spectral results for Schrödinger operators was regarded as a significant problem around 2000.

1.9 BACK TO OPRL: SZEGŐ MAPPING AND THE SHOHAT–NEVAI THEOREM

We can translate the gem for OPUC to a result for OPRL using an interesting connection that Szegő found in 1922 [431, 434]. It is connected to the natural conformal bijection of $\mathbb{D} \to \mathbb{C} \cup \{\infty\} \setminus [-2,2]$ by

$$z \to E = z + z^{-1} \tag{1.9.1}$$

This maps $\partial\mathbb{D}$ two-one to $[-2,2]$ by

$$e^{i\theta} \xrightarrow{Q} 2\cos\theta \tag{1.9.2}$$

We can use this to map $C([-2,2])$, the continuous functions on $[-2,2]$, to $C(\partial\mathbb{D})$:

$$(Q_\sharp f)(e^{i\theta}) = f(Q(e^{i\theta})) = f(2\cos\theta) \tag{1.9.3}$$

Notice $\mathrm{Ran}(Q_\sharp)$ is exactly the set of all functions invariant under $e^{i\theta} \to e^{-i\theta}$. Duality then induces a map $Q_\sharp^*\colon \mathcal{M}_{+1,1}(\partial\mathbb{D}) \to \mathcal{M}_{+,1}([-2,2])$ between the probability measures by

$$\int f(x)[Q_\sharp^*(d\mu)](x) = \int (Q_\sharp f)(e^{i\theta})\, d\mu(\theta) \tag{1.9.4}$$

$Q_\sharp^*$ is onto $\mathcal{M}_{+,1}([-2,2])$, but it is not one-one. For example, if f is any nonnegative L^1 function with $f(\theta) + f(2\pi - \theta) = 1$ and $\int f(\theta)\frac{d\theta}{2\pi} = 1$, then $Q_\sharp(f\frac{d\theta}{2\pi}) = Q_\sharp(\frac{d\theta}{2\pi}) = \pi^{-1}(4-x^2)^{-1/2}\,dx$. However, restricted to measures invariant under $\theta \to -\theta$, $Q_\sharp$ is one-one, and we denote its restriction to even measures by Sz for Szegő mapping. Thus, $d\rho = \mathrm{Sz}(d\mu)$ if and only if $d\mu(\theta) = d\mu(-\theta)$ and

$$\int f(\theta)\, d\mu(\theta) = \int f\left(\arccos\left(\frac{x}{2}\right)\right) d\rho(x) \tag{1.9.5}$$

for any f obeying $f(-\theta) = f(\theta)$. Sz is a bijection between nontrivial even probability measures on $\partial\mathbb{D}$ and nontrivial probability measures on $[-2,2]$.

Because of the impact of symmetry on Szegő recursion, we see

$$d\mu \text{ even} \Leftrightarrow \overline{\Phi_n(z)} = \Phi_n(\bar{z}) \Leftrightarrow \alpha_n \in \mathbb{R} \text{ for all } n \tag{1.9.6}$$

Szegő [431, 434] proved the following:

Theorem 1.9.1. *Let $d\rho = \mathrm{Sz}(d\mu)$ for nontrivial probability measures on $[-2,2]$ and $\partial\mathbb{D}$. Let P_n, p_n be the monic and orthonormal OPRL for $d\rho$ and Φ_n, φ_n the monic and orthonormal OPUC for $d\mu$. Then*

$$P_n\left(z + \frac{1}{z}\right) = [1 - \alpha_{2n-1}(d\mu)]^{-1} z^{-n}[\Phi_{2n}(z) + \Phi_{2n}^*(z)] \tag{1.9.7}$$

$$\|P_n\|_{L^2(d\rho)}^2 = 2(1-\alpha_{2n-1})^{-1}\|\Phi_{2n}\|_{L^2(d\mu)}^* \tag{1.9.8}$$

$$p_n\left(z + \frac{1}{z}\right) = [2(1-\alpha_{2n-1})]^{-1/2} z^{-n}(\varphi_{2n}(z) + \varphi_{2n}^*(z)) \tag{1.9.9}$$

Sketch. (For details, see Theorem 13.1.5 of [400].) The right side of (1.9.7) is a Laurent polynomial of the form $\sum_{j=-n}^{n} c_j z^j$ invariant under $z \to \frac{1}{z}$ on account of (1.9.6). Every such Laurent polynomial has the form $Q_n(z + \frac{1}{z})$ for $Q_n(\cdot)$ of degree n.

Since $\Phi_{2n}(0) = -\bar{\alpha}_{2n-1}$, $\Phi_{2n}^*(z) = -\alpha_{2n-1}z^{2n} + \cdots$, so Q_n is monic. Moreover, by (1.9.5) for $\ell < n$,

$$\begin{aligned}\int Q_n(x)Q_\ell(x)\, d\rho(x) &= \int \overline{\Phi_{2n} + \Phi_{2n}^*}\,(z)\, z^{n-\ell}(\Phi_{2\ell} + \Phi_{2\ell}^*)\, d\mu(z) \\ &= 0\end{aligned} \tag{1.9.10}$$

since $\Phi_{2n} \perp \{z, \ldots, z^{2n-1}\}$ and $\Phi_{2n}^* \perp \{z, \ldots, z^{2n-1}\}$. Thus, the Q_n's are the monic OPRL for $d\rho$, that is, we have proven (1.9.7).

(1.9.8) follows from (1.9.7) and

$$\begin{aligned}\langle \Phi_{2n}, \Phi_{2n}^*\rangle &= \langle \Phi_{2n}, \Phi_{2n-1}^* - \alpha_{2n-1} z\Phi_{2n-1}\rangle \\ &= -\alpha_{2n-1}\langle \Phi_{2n}, \Phi_{2n} + \bar{\alpha}_{2n-1}\Phi_{2n-1}^*\rangle \\ &= -\alpha_{2n-1}\|\Phi_{2n}\|^2\end{aligned} \tag{1.9.11}$$

by using Szegő recursion and orthogonality. (1.9.9) is immediate from (1.9.7) and (1.9.8). $\square$

There are several other relations we want to note because we will need them in Section 3.11. First, (1.9.9) can be written

$$p_n\left(z+\frac{1}{z}\right) = [2(1-\alpha_{2n-1})]^{-1/2}\left(z^{-n}\varphi_{2n}(z) + z^n\varphi_{2n}\left(\frac{1}{z}\right)\right) \tag{1.9.12}$$

By the same method, one can see

$$p_n\left(z+\frac{1}{z}\right) = [2(1+\alpha_{2n-1})]^{-1/2}\left(z^{-(n-1)}\varphi_{2n-1}(z) + z^{(n-1)}\varphi_{2n-1}\left(\frac{1}{z}\right)\right) \tag{1.9.13}$$

Besides $d\rho = \mathrm{Sz}(d\mu)$, there is a second (nonprobability) measure one can associate to $d\mu$, namely,

$$d\rho_1(x) \equiv \mathrm{Sz}_1(d\mu)(x) = \tfrac{1}{4}(4-x^2)\,d\rho(x) \tag{1.9.14}$$

Its orthonormal polynomials are denoted by $q_n(x)$. As with the derivation of (1.9.9), one finds

$$\tfrac{1}{2}q_{n-1}\left(z+\frac{1}{z}\right) = [2(1+\alpha_{2n-1})]^{-1/2}\left(\frac{z^{-n}\varphi_{2n}(z) - z_n\varphi_{2n}(\frac{1}{z})}{z-z^{-1}}\right) \tag{1.9.15}$$

$$= [2(1-\alpha_{2n-1})]^{-1/2}\left(\frac{z^{-(n-1)}\varphi_{2n-1}(z) - z^{(n-1)}\varphi_{2n-1}(\frac{1}{z})}{z-z^{-1}}\right) \tag{1.9.16}$$

This leads to

$$z^{-n}\varphi_{2n}(z) = \left[\tfrac{1}{2}(1-\alpha_{2n-1})\right]^{1/2} p_n\left(z+\frac{1}{z}\right) + \left[\tfrac{1}{2}(1+\alpha_{2n-1})\right]^{1/2}\left(\frac{z-z^{-1}}{2}\right)q_{n-1}\left(z+\frac{1}{z}\right) \tag{1.9.17}$$

$$z^{-(n-1)}\varphi_{2n-1}(z) = \left[\tfrac{1}{2}(1+\alpha_{2n-1})\right]^{1/2} p_n\left(z+\frac{1}{z}\right) + \left[\tfrac{1}{2}(1-\alpha_{2n-1})\right]^{1/2}\left(\frac{z-z^{-1}}{2}\right)q_{n-1}\left(z+\frac{1}{z}\right) \tag{1.9.18}$$

When $z = e^{i\theta}$, $p_n(2\cos\theta)$ and $q_{n-1}(2\cos\theta)$ are real, but $(z-z^{-1})/2 = i\sin\theta$ is pure imaginary, so the absolute value square has no cross-term. Thus, we find the formula we will need in Section 3.11

$$|\varphi_{2n}(e^{i\theta})|^2 + |\varphi_{2n-1}(e^{i\theta})|^2 = |p_n(2\cos\theta)|^2 + \sin^2\theta|q_{n-1}(2\cos\theta)|^2 \tag{1.9.19}$$

where we used $([\frac{1}{2}(1+\alpha_{2n-1})]^{1/2})^2 + ([\frac{1}{2}(1-\alpha_{2n-1})]^{1/2})^2 = 1$ to miraculously have α_{2n-1} drop out!

From Theorem 1.9.1, we get the formula relating a_n, b_n and α_n:

Theorem 1.9.2 (Direct Geronimus Relations)**.** *Let $d\rho = \mathrm{Sz}(d\mu)$ for nontrivial probability measures on $[-2, 2]$ and $\partial\mathbb{D}$. Let $\{a_n, b_n\}_{n=1}^{\infty}$ be the Jacobi parameters for $d\rho$ and $\{\alpha_n\}_{n=0}^{\infty}$ the Verblunsky coefficients for $d\mu$. Then*

$$\text{(i)} \qquad (a_1 \dots a_n)^2 = 2(1 + \alpha_{2n-1}) \prod_{j=0}^{2n-2} (1 - \alpha_j^2) \tag{1.9.20}$$

$$\text{(ii)} \qquad a_{n+1}^2 = (1 + \alpha_{2n+1})(1 - \alpha_{2n}^2)(1 - \alpha_{2n-1}) \tag{1.9.21}$$

$$\text{(iii)} \qquad b_{n+1} = (1 - \alpha_{2n-1})\alpha_{2n} - (1 + \alpha_{2n-1})\alpha_{2n-2} \tag{1.9.22}$$

Remark. (i) holds for $n \geq 1$ and (ii)/(iii) for $n \geq 0$. For $n = 1$, (1.9.20) says $a_1^2 = 2(1 + \alpha_1)(1 - \alpha_0^2)$, so (1.9.21) holds for $n = 1$ if we define

$$\alpha_{-1} = -1 \tag{1.9.23}$$

While α_{-2} enters in (1.9.22) for $n = 0$, it is multiplied by $(1 + \alpha_{-1}) = 0$, so only the "boundary condition" (1.9.23) is needed.

Sketch. (For details, see Theorems 13.1.7 and 13.1.12 of [400].)

(i) Since

$$\frac{1 - \alpha_{2n-1}^2}{1 - \alpha_{2n-1}} = 1 + \alpha_{2n-1} \tag{1.9.24}$$

this is a rewriting of (1.9.8) using (1.8.11) and (1.2.13).

(ii) This follows from dividing (i) for $n + 1$ by (i) for n using (1.9.24).

(iii) This comes from (1.9.7) looking at the $O(z^{n-1})$ terms. By a simple induction from (1.2.8),

$$P_n(x) = x^n - \left(\sum_{j=1}^{n} b_j \right) x^{n-1} + O(x^{n-2}) \tag{1.9.25}$$

From (1.8.5) and (1.8.6), we get that if

$$\Phi_n(z) = z^n + C_n z^{n-1} + O(z^{n-2}) \tag{1.9.26}$$

$$\Phi_n^*(z) = -\alpha_{n-1} z^n + D_n z^{n-1} + O(z^{n-2}) \tag{1.9.27}$$

then, by induction,

$$C_n = \sum_{j=0}^{n-1} \bar{\alpha}_j \alpha_{j-1} \tag{1.9.28}$$

(where, as usual, $\alpha_{-1} = -1$) and

$$D_n = -\alpha_{n-2} - \alpha_{n-1} C_{n-1} \tag{1.9.29}$$

These formulae and (1.9.7) imply that

$$-\sum_{j=1}^{n} b_j = C_{2n-1} - \alpha_{n-2} \tag{1.9.30}$$

and this yields (1.9.22). □

This lets us "translate" Corollary 1.8.7 to OPRL:

Theorem 1.9.3 (Shohat–Nevai Theorem)**.** *Let*

$$d\rho(x) = f(x)\,dx + d\rho_{\mathrm{s}}(x)$$

be supported on $[-2, 2]$. *Then*

$$\int_{-2}^{2} (4 - x^2)^{-1/2} \log(f(x))\,dx > -\infty \tag{1.9.31}$$

if and only if

$$\limsup a_1 \dots a_n > 0 \tag{1.9.32}$$

If these conditions hold, then

$$\lim a_1 \dots a_n \tag{1.9.33}$$

exists in $(0, \infty)$ *and*

$$\sum_{n=1}^{\infty} (a_n - 1)^2 + b_n^2 < \infty \tag{1.9.34}$$

and

$$\sum_{n=1}^{N} (a_n - 1) \qquad \textit{and} \qquad \sum_{n=1}^{N} b_n \tag{1.9.35}$$

have limits in $(-\infty, \infty)$.

Remarks. 1. We emphasize (1.9.32) is lim sup, that is, it allows lim inf to be 0 so long as some subsequence stays away from 0.

2. This can be rephrased as saying $a_1 \dots a_n$ always has a limit when $\mathrm{supp}(d\rho) \subset [-2, 2]$ since the negation of (1.9.32) is $\lim a_1 \dots a_n = 0$. This is discussed further in Section 3.6.

Proof. Let μ be defined by $\mathrm{Sz}(d\mu) = d\rho$. By (1.9.30),

$$(a_1 \dots a_n)^2 \le 4 \prod_{j=0}^{2n-2} (1 - \alpha_j^2)$$

so (1.9.32) implies $\lim \prod_{j=0}^{\infty} (1 - \alpha_j^2)$ (the limit always exists) is strictly positive and thus, $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$. Conversely, if $\sum_j \alpha_j^2 < \infty$, then $\alpha_j \to 0$ and so, by (1.9.20), $\lim a_1 \dots a_n$ exists in $(0, \infty)$. We have thus proven that

$$(1.9.32) \Rightarrow \sum_{j=0}^{\infty} \alpha_j^2 < \infty \Rightarrow \lim a_1 \dots a_n \text{ exists in } (0, \infty) \tag{1.9.36}$$

On the other hand, if

$$d\mu = w(\theta)\,\frac{d\theta}{2\pi} + d\mu_{\mathrm{s}} \tag{1.9.37}$$

then, by (1.9.5),

$$w(\theta) = 2\pi\,|\sin\theta|\,f(2\cos\theta) \tag{1.9.38}$$

It follows that (changing variables, using $x = 2\cos\theta \Rightarrow dx = 2\sin\theta\,d\theta$ or $d\theta = (4-x^2)^{-1/2}\,dx$)

$$\int \log(w(\theta))\,\frac{d\theta}{2\pi} > -\infty \Leftrightarrow \int \log(f(x))(4-x^2)^{-1/2}\,dx > -\infty \tag{1.9.39}$$

Thus, (1.8.19), (1.9.36), and (1.9.39) imply

$$\limsup(a_1 \dots a_n) > 0 \Leftrightarrow (1.9.31)$$

and if this holds, then (1.9.33) has a limit.

Since b_{n+1} and $a_{n+1}^2 - 1$ are functions of α_{2n+j} $(j = -2, -1, 0, 1)$, we see that if $\sum_{j=0}^\infty \alpha_j^2 < \infty$, then $\sum b_n^2 < \infty$ and $\sum (a_n^2 - 1)^2 < \infty$. Since $(a_n + 1) \geq 1$, $(a_n - 1)^2 = (a_n^2 - 1)^2/(a_n + 1)^2 \leq (a_n^2 - 1)^2$, so (1.9.34) holds.

Finally, when $\sum_{j=0}^\infty \alpha_j^2 < \infty$, $a_{n+1}^2 - 1$ and b_{n+1} are the sum of an L^1 sequence and a telescoping sequence, so $a_{n+1}^2 - 1$ and b_{n+1} are summable. Since $(a_j^2 - 1) - 2(a_j - 1) = (a_j - 1)^2$ is summable, we see that so is $a_{n+1} - 1$. $\square$

We want to emphasize that while Corollary 1.8.7, on which Theorem 1.9.3 is based, is a gem (equivalence of purely spectral condition to purely sufficient condition), Theorem 1.9.3 is not. For it makes the a priori condition that $\mathrm{supp}(d\rho) \subset [-2, 2]$, that is, it is the equivalence of

$$(1.9.31) + \mathrm{supp}(d\rho) \subset [-2, 2] \tag{1.9.40}$$

to

$$(1.9.32) + \mathrm{supp}(d\rho) \subset [-2, 2] \tag{1.9.41}$$

(1.9.40) is purely spectral, but (1.9.41) is not a condition only about the Jacobi parameters. Indeed, $\mathrm{supp}(d\rho) \subset [-2, 2]$ is a very strong restriction if $\limsup(a_1 \dots a_n) > 0$. Indeed, it implies strong conditions on the b_n's ($\sum_{n=1}^\infty b_n^2 < \infty$ and $\sum_{n=1}^N b_n$ conditionally convergent).

Remarks and Historical Notes. The Szegő mapping was introduced by Szegő in [431] and further discussed by him in [434]. Its purpose was to carry over asymptotics of OPUC when the Szegő condition holds to asymptotics of OPRL when the OPRL Szegő condition holds (see Section 3.7).

$d\mu$ and $d\rho = \mathrm{Sz}(d\mu)$ can be related via their natural transforms

$$F(z) = \int \frac{e^{i\theta} + z}{e^{i\theta} - z}\,d\mu(\theta) \qquad m(z) = \int \frac{d\rho(x)}{x - z} \tag{1.9.42}$$

namely,

$$F(z) = 2(z - z^{-1})m(z + z^{-1}) \tag{1.9.43}$$

This formula is from Geronimus [159]; see also the proof of Theorem 13.1.2 in [400].

The map $z \to E = z + z^{-1}$ may seem miraculous, but it is canonical and uniquely determined. By the Riemann mapping theorem, there is an analytic bijection, g, of $\mathbb{D}$ to $\mathbb{C} \cup \{\infty\} \setminus [-2, 2]$ and it is uniquely determined by $g(0) = \infty$ and $\lim_{z\to 0} zg(z) > 0$. This unique map, abstractly guaranteed, is $g(z) = z + z^{-1}$. This will become a major theme in Chapter 9.

Geronimus [159, 160] found the relations (1.9.21)/(1.9.22). Other proofs can be found in Damanik–Killip [96], Killip–Nenciu [223], and Faybusovich–Gekhtman [129]. The latter two proofs are discussed in Section 13.2 of [400] and in Section 13.3 of the expected second edition of [400], which is posted online at `http://www.math.caltech.edu/opuc/newsection13-3.pdf`.

Szegő found a second natural map on nontrivial symmetric probability measures on $\partial\mathbb{D}$ to a large subset of measures on $[-2, 2]$, the map we called Sz_1 in (1.9.14). There are, in fact, four natural maps discussed in Section 13.2 of [400] and references therein. We note that all the original papers prior to 2000 use $[-1, 1]$ not $[-2, 2]$, and $z \to \frac{1}{2}(z + z^{-1})$. [400] discusses normalized measures (one needs to multiply $d\rho_1$ by $2[(1 - |\alpha_0|^2)(1 - \alpha_1)]^{-1}$ to normalize). For our purposes in Section 3.11, the unnormalized measure that leads to (1.9.19) is more convenient.

Szegő's book [434] includes (1.9.12)–(1.9.15) (in Section 11.5) and he noted their inverses (in Section 6 of his appendix). The compact consequence in (1.9.19) is from Máté–Nevai–Totik [302].

It is interesting to check these formulae in case $d\mu = \frac{d\theta}{2\pi}$. Then

$$\mathrm{Sz}(d\mu)(x) = \frac{1}{\pi} \frac{1}{\sqrt{4 - x^2}}\, dx \tag{1.9.44}$$

$$= \frac{1}{\pi}\, d\left(\arccos\left(\frac{x}{2}\right)\right) \tag{1.9.45}$$

and (Chebyshev polynomials of the first and second kinds)

$$p_n(2\cos\theta) = \sqrt{2}\,\cos(n\theta) \tag{1.9.46}$$

$$q_n(2\cos\theta) = \sqrt{2}\,\frac{\sin((n+1)\theta)}{\sin\theta} \tag{1.9.47}$$

$\alpha_{2n-1} = 0$ and, for example, (1.9.18) says

$$e^{-in\theta}e^{2ni\theta} = \frac{1}{\sqrt{2}}\sqrt{2}\,\cos(n\theta) + \frac{1}{\sqrt{2}}\, i\sin\theta\sqrt{2}\,\frac{\sin(n\theta)}{\sin\theta} \tag{1.9.48}$$

Theorem 1.9.3 first appeared in Nevai [320] using in part ideas in Shohat [384].

We will eventually see (Theorem 3.6.1) that Theorem 1.9.3 can be extended to situations where there is some point spectrum outside $[-2, 2]$, namely, we will need $\sigma_{\mathrm{ess}}(d\mu) = [-2, 2]$ and

$$\sum_{\substack{E\in\mathrm{supp}(d\mu)\\ E\notin[-2,2]}} \mathrm{dist}(E, \sigma_{\mathrm{ess}}(d\mu))^{1/2} < \infty \tag{1.9.49}$$

1.10 THE KILLIP–SIMON THEOREM

As we noted, Theorem 1.9.3 is a spectral result about OPRL related to Szegő's theorem, but not a gem as we defined it. Here is an OPRL gem that is related to Szegő's theorem.

It will involve the free Jacobi matrix, J_0, whose Jacobi parameters are

$$a_n \equiv 1 \qquad b_n \equiv 0 \tag{1.10.1}$$

The OPs for this case are (as is easy to check obey the recursion relations on account of trigonometric addition formulae; these are essentially the Chebyshev polynomials of the second kind; see (1.2.35))

$$P_n(2\cos\theta) = \frac{\sin(n+1)\theta}{\sin\theta} \tag{1.10.2}$$

The spectral measure is

$$d\rho_0(x) = \frac{1}{2\pi}\,(4-x^2)^{1/2}\,dx \tag{1.10.3}$$

so that

$$\sigma(J_0) = \sigma_{\text{ess}}(J_0) = \sigma_{\text{ac}}(J_0) = [-2, 2] \tag{1.10.4}$$

Theorem 1.10.1 (Killip–Simon Theorem)**.** *Let* $\{a_n, b_n\}_{n=1}^{\infty}$ *be the Jacobi parameters of a Jacobi matrix, J. Then*

$$\sum_{n=1}^{\infty} (a_n - 1)^2 + b_n^2 < \infty \tag{1.10.5}$$

if and only if

(a)

$$\sigma_{\text{ess}}(J) = \sigma_{\text{ess}}(J_0) \qquad (\textit{Blumenthal–Weyl}) \tag{1.10.6}$$

(b) *The eigenvalues* $E_n \notin \sigma_{\text{ess}}(J_0)$ *obey*

$$\sum_{n=1}^{\infty} \text{dist}(E_n, \sigma_{\text{ess}}(J_0))^{3/2} < \infty \qquad (\textit{Lieb–Thirring}) \tag{1.10.7}$$

(c) *The function f of* (1.4.3) *obeys*

$$\int_{\sigma(J_0)} \text{dist}(x, \mathbb{R}\setminus\sigma(J_0))^{1/2} \log(f(x))\,dx > -\infty \qquad (\textit{Quasi-Szegő}) \tag{1.10.8}$$

Remarks. 1. (1.10.5) is equivalent to $J - J_0$ being a Hilbert–Schmidt operator (see [170, 381]).

2. (1.10.8) is called "quasi-Szegő" because it looks like the Szegő condition (1.9.30) except $-\frac{1}{2}$ has become $\frac{1}{2}$, allowing a larger class of f's. Similarly, (1.10.7) looks like (1.9.49) except that $\frac{1}{2}$ has become $\frac{3}{2}$.

The proof of Theorem 1.10.1 will be the main topic of Chapter 3, but to set the stage we want to say something about it. As with Szegő's theorem, the key is a sum rule. It will involve two somewhat complicated-looking functions, F defined on $\mathbb{R} \setminus [-2, 2]$ and G on $(0, \infty)$:

$$F(\beta + \beta^{-1}) = \tfrac{1}{4}[\beta^2 - \beta^{-2} - \log(\beta^4)] \qquad \beta \in \mathbb{R} \setminus [-1, 1] \tag{1.10.9}$$

$$G(a) = a^2 - 1 - \log(a^2) \tag{1.10.10}$$

Notice that $\beta \mapsto \beta + \beta^{-1}$ is a bijection of $\mathbb{R} \setminus [-1, 1]$ to $\mathbb{R} \setminus [-2, 2]$ so (1.10.9) defines F.

We will eventually show that (Lemma 3.5.3)

$$F(E) = \tfrac{1}{2} \int_2^{|E|} (E^2 - 4)^{1/2} \, dE \tag{1.10.11}$$

which implies

$$F(E) > 0 \quad \text{on } \mathbb{R} \setminus [-2, 2] \tag{1.10.12}$$

and

$$F(E) = \tfrac{2}{3} (|E| - 2)^{3/2} + O((|E| - 2)^{5/2}) \tag{1.10.13}$$

We also see that (Lemma 3.5.2)

$$G(a) > 0 \quad \text{on } (0, \infty) \setminus \{1\} \tag{1.10.14}$$

$$G(a) = 2(a - 1)^2 + O((a - 1)^3) \tag{1.10.15}$$

We also need to define

$$Q(\rho) = \frac{1}{4\pi} \int_{-2}^{2} \log\left(\frac{\sqrt{4 - x^2}}{2\pi f(x)} \right) \sqrt{4 - x^2} \, dx \tag{1.10.16}$$

which, given (1.10.3), can be rewritten

$$Q(\rho) = -\tfrac{1}{2} \int \log\left[\left(\frac{d\rho}{d\rho_0} \right)^{-1} \right] d\rho_0 \tag{1.10.17}$$

whose integral is a relative entropy (see (2.2.1)). As we will show (Theorem 2.2.3), using Jensen's inequality, $Q(\rho) \geq 0$. The sum rule is

Theorem 1.10.2. *Let $d\rho$ be a nontrivial probability measure with associated Jacobi parameters $\{a_n, b_n\}_{n=1}^{\infty}$ and $\sigma_{\text{ess}}(d\rho) = [-2, 2]$. Then*

$$Q(\rho) + \sum F(E_n) = \sum_{n=1}^{\infty} [\tfrac{1}{4} b_n^2 + \tfrac{1}{2} G(a_n)] \tag{1.10.18}$$

This is called the P_2 sum rule. Notice that all terms on both sides are positive so the sums always make sense, but they may be infinite. Moreover, $\sigma_{\text{ess}}(d\rho) = [-2, 2]$ and the left-hand side of (1.10.18) $< \infty$ if and only if (a)–(c) of Theorem 1.10.1 holds, on account of (1.10.13) and (1.10.16). On the other hand, using Theorem 1.4.1 and (1.10.15), $\sigma_{\text{ess}}(d\rho) = [-2, 2]$ and the right-hand side of (1.10.18) $< \infty$ if and only if (1.10.5) holds. Thus, Theorem 1.10.2 implies Theorem 1.10.1.

Where will complicated objects like F and G come from? The sum rule of Verblunsky (1.8.18) is a form of Jensen's equality for analytic functions, hence the logs. In this case, the function is nonvanishing. The sum rule (1.10.18) will come from a Jensen–Poisson equality and involves two Taylor coefficients: the zeroth, which has logs, and the second without logs. There are terms from the zeros in this case, hence the logs in the sum involving F. These details will unfold in Chapter 3.

Remarks and Historical Notes. Theorems 1.10.1 and 1.10.2 are from Killip–Simon [225]. For historical context and the name "P_2," see the Notes to Sections 3.1 and 3.4.

1.11 PERTURBATIONS OF THE PERIODIC CASE

The material in Chapters 5, 6, and 8 is all connected with analyzing Szegő-like theorems for OPRL (and some related OPUC) where the $[-2, 2]$ of Theorem 1.10.1 is replaced by a union of a finite number of closed bounded intervals, especially the case of perturbations of periodic OPRL. Chapters 5 and 6 discuss periodic OPRL themselves, that is, Jacobi matrices, J_0, where

$$a_{n+p}^{(0)} = a_n^{(0)} \qquad b_{n+p}^{(0)} = b_n^{(0)} \tag{1.11.1}$$

for some $p \geq 2$ and all $n = 1, 2, \ldots$. (In Section 5.14, we also discuss OPUC when $\alpha_{n+p}^{(0)} = \alpha_n^{(0)}$, mainly with p even.) Rather than studying a_n, b_n, which approach $a_n \equiv 1$, $b_n \equiv 0$ in some sense, we want to discuss approach to J_0. J_0 is obviously parametrized by $\mathbb{R}^{2p} = \{(a_n^{(0)}, b_n^{(0)})_{n=1}^p\}$.

We begin the discussion by describing $\sigma(J_0)$, the spectrum of J_0 (see Sections 5.2, 5.3, and 5.4):

Theorem 1.11.1. *$\sigma_{\text{ess}}(J_0)$ is the disjoint union of $k + 1 \leq p$ distinct bounded intervals*

$$\sigma_{\text{ess}}(J_0) = \bigcup_{j=1}^{k+1} [c_j, d_j] \tag{1.11.2}$$

where

$$c_1 < d_1 < c_2 < \cdots < c_{k+1} < d_{k+1}$$

Each of the k gaps (d_j, c_{j+1}), $j = 1, \ldots, k$, has zero or one point mass.

Generically, $k = p - 1$. Indeed, $\{(a_n^{(0)}, b_n^{(0)}) \mid k < p - 1\}$ is a variety of codimension 2 in $\mathbb{R}^{2p}$. If $k = p - 1$, we say "all gaps are open."

While we will not say a lot about the proof now, we do want to mention one of the key tools. There is a natural polynomial in x, $\Delta(x; \{a_n^{(0)}, b_n^{(0)}\}_{n=1}^p) = \Delta(x; J_0)$ of exact degree p, so

$$\sigma_{\text{ess}}(J_0) = \Delta^{-1}([-2, 2]) \tag{1.11.3}$$

We are interested in the analog Theorem 1.10.1 when J_0 is a periodic Jacobi matrix. The conjectured analog of the spectral side is obvious: (1.10.6)–(1.10.8) were carefully stated in terms of $\sigma_{\text{ess}}(J_0)$ rather than $[-2, 2]$ precisely because they will be one side of the proper periodic theorem.

There is an obvious guess for an analog of (1.10.5), namely,

$$\sum_{n=1}^{\infty}(a_n - a_n^{(0)})^2 + (b_n - b_n^{(0)})^2 < \infty \tag{1.11.4}$$

This cannot be right for the following reason. The map

$$J_1 = \{(a_n^{(1)}, b_n^{(1)}) \mid a_{n+p}^{(1)} = a_n^{(1)},\ b_{n+p}^{(1)} = b_n^{(1)}\} \to \Delta(x, J_1) \tag{1.11.5}$$

is a map of $\mathbb{R}^{2p}$ to $\mathbb{R}^{p+1}$, since Δ has $p+1$ coefficients. As one would expect, generic inverse images of a fixed Δ are of dimension $2p - (p+1) = p - 1$. In fact, we will show (see Section 5.13):

Theorem 1.11.2. *For fixed periodic J_0, $\{J_1 \mid \Delta(x, J_1) = \Delta(x, J_0)\}$ is a torus of dimension k where*

$$k + 1 = \#\ \textit{of components of}\ \sigma_{\text{ess}}(J_0) \tag{1.11.6}$$

This set is called the *isospectral torus* of J_0, which we denote $\mathcal{T}_{J_0}$. By (1.11.3), if $J_1 \in \mathcal{T}_{J_0}$, $\sigma_{\text{ess}}(J_1) = \sigma_{\text{ess}}(J_0)$, and so J_1 also obeys (1.10.6)–(1.10.8), but J_1 does not obey (1.11.4). What we need is not ℓ^2 approach to a fixed J_0 but rather to $\mathcal{T}_{J_0}$. We define

$$d_m((a_n, b_n)_{n=1}^{\infty}, (a_n', b_n')_{n=1}^{\infty}) = \sum_{j=m}^{\infty} e^{-|j-m|}[|a_j - a_j'| + |b_j - b_j'|] \tag{1.11.7}$$

which measures the distances of the tails from each other. We also define

$$d_m((a_n, b_n)_{n=1}^{\infty}, \mathcal{T}_{J_0}) = \min_{(a_n', b_n') \in \mathcal{T}_{J_0}} d_m((a, b), (a', b')) \tag{1.11.8}$$

It can happen that the minimizing (a', b') is m-dependent and that $d_m((a, b), \mathcal{T}_{J_0}) \to 0$ as $m \to \infty$ without $d_m((a, b), J_1) \to 0$ for any J_1 (although, by compactness of $\mathcal{T}_{J_0}$, there will be J_1 and a subsequence for which $d_{m_\ell}((a, b), J_1) \to 0$ as $\ell \to \infty$).

Damanik–Killip–Simon [97] have proven:

Theorem 1.11.3 (DKS [97]). *Let J_0 be a fixed periodic Jacobi matrix of period p with all gaps open (i.e., $k = p - 1$). Let J be another bounded Jacobi matrix with Jacobi parameters $(a_n, b_n)_{n=1}^{\infty}$. Then the following are equivalent:*

(a) (1.10.6), (1.10.7), *and* (1.10.8) *hold.*

(b)

$$\sum_{m=1}^{\infty} d_m((a, b), \mathcal{T}_{J_0})^2 < \infty \tag{1.11.9}$$

The proof of this theorem is the main goal of Chapter 8. A key tool will be the study of the matrix $\Delta(J; J_0)$, that is, the matrix obtained by placing J for x in the polynomial $\Delta(x; J_0)$. Since Δ has degree p, $\Delta(J)$ will be a matrix of band width $2p+1$, that is, p diagonals strictly above, p strictly below, and on the main diagonal. Such a matrix can be thought of as "tridiagonal" if we replace a's and b's by $p \times p$ blocks. We will prove a Killip–Simon theorem for such block Jacobi matrices in Chapter 4, and that will be a main tool in proving Theorem 1.11.3.

In the periodic case, $\sigma_{\text{ess}}(J_0)$ is a disjoint union, (1.11.2). But not every such union is $\sigma_{\text{ess}}(J_0)$ for some periodic J_0. Basically, there is a natural map (harmonic measure),

$$\mathcal{M}\colon \{c_1 < d_1 < c_2 < \cdots < d_{k+1}\} \to \left\{ (\theta_j)_{j=1}^{k+1} \,\middle|\, \theta_j > 0;\ \sum_{j=1}^{k+1} \theta_j = 1 \right\}$$

which is continuous and onto. The allowed $\sigma_{\text{ess}}(J_0)$ for periodic J_0's with all gaps open is $\mathcal{M}((c,d)) = (\frac{1}{p}, \dots, \frac{1}{p})$, and if we drop the demand that all gaps are open, then the range is the set of rational θ's.

For other finite band sets, $\sigma_{\text{ess}}(J_0)$ can be that set if we allow certain almost periodic J_0's. There is no Killip–Simon-type theorem known in this case, but one-half of a Shohat–Nevai-type theorem is known due to work of Akhiezer, Widom, Aptekarev, and Peherstorfer–Yuditskii. It will be the subject of Chapter 9. Chapter 10 will discuss Killip–Simon-like theorems for perturbations of the graph Laplacian on a Bethe–Cayley tree.

Remarks and Historical Notes. As noted, Theorem 1.11.3 is from Damanik–Killip–Simon [97]. Prior results and historical context are discussed in the Notes to Section 8.1. The history of results mentioned in the last paragraph are in the Notes to Section 9.13.

1.12 OTHER GEMS IN THE SPECTRAL THEORY OF OPUC

While gems are the leitmotif of this chapter, our choice of topics is motivated by looking at relatives of Szegő's theorem. We will see that in this section by mentioning some other gems for OPUC (the Notes discuss OPRL) that will not be discussed further. Here are three theorems in particular:

Theorem 1.12.1 (Baxter's Theorem)**.** *Let μ be a probability measure on $\partial\mathbb{D}$ of the form* (1.6.3) *and let $\{\alpha_n\}_{n=1}^{\infty}$ be its Verblunsky coefficients. Then the following are equivalent:*

(i)

$$\sum_{n=0}^{\infty} |\alpha_n| < \infty \tag{1.12.1}$$

(ii) $d\mu_s = 0$,

$$\inf w(\theta) > 0 \tag{1.12.2}$$

$$\sum_{n=-\infty}^{\infty} |\widehat{w}_n| < \infty \tag{1.12.3}$$

where

$$\widehat{w}_n = \int e^{-in\theta} w(\theta) \, \frac{d\theta}{2\pi} \tag{1.12.4}$$

Remark. (1.12.3) implies w is continuous, so the inf in (1.12.2) is a min.

Theorem 1.12.2 (Ibragimov's Form of the Strong Szegő Theorem)**.** *Let μ be a probability measure on $\partial\mathbb{D}$ of the form* (1.6.3) *and let $\{\alpha_n\}_{n=1}^\infty$ be its Verblunsky coefficients. Then the following are equivalent:*

(i)

$$\sum_{n=0}^\infty n|\alpha_n|^2 < \infty \tag{1.12.5}$$

(ii) $d\mu_{\rm s} = 0$*, the Szegő condition* (1.8.19) *holds, and*

$$\sum_{n=1}^\infty n|\widehat{L}_n|^2 < \infty \tag{1.12.6}$$

where

$$\widehat{L}_n = \int e^{-in\theta} \log(w(\theta)) \, \frac{d\theta}{2\pi} \tag{1.12.7}$$

Theorem 1.12.3 (Nevai–Totik Theorem)**.** *Let μ be a probability measure on $\partial\mathbb{D}$ of the form* (1.6.3) *and let $\{\alpha_n\}_{n=1}^\infty$ be its Verblunsky coefficients. Let $R > 1$. Then the following are equivalent:*

(i) $\limsup|\alpha_n|^{1/n} \leq R^{-1}$

(ii) $\mu_{\rm s} = 0$ *and the Szegő function D, defined by* (2.9.14)*, has $D^{-1}(z)$ analytic in* $\{z \mid |z| < R\}$.

There are two distinctions between these results and Szegő's theorem. These only involve μ's with $\mu_{\rm s} = 0$ and with more rapid decay than just ℓ^2. If $\alpha_n \sim Cn^{-s}$; Szegő requires $s > \frac{1}{2}$, but these require $s > 1$ (and exponential decay in the case of the Nevai–Totik theorem).

Remarks and Historical Notes. Baxter's theorem is from Baxter [32] and is discussed in [399, Chapter 5]. Ibragimov's form is from Ibragimov [203] and related to Szegő's work on the second term in Toeplitz determinant asymptotics discussed in the Notes to Section 1.6 where references appear. The Nevai–Totik theorem is from Nevai–Totik [323] and discussed in [399, Chapter 7].

For analogs of Theorems 1.12.1 and 1.12.2 for OPRL, see Ryckman [375, 376]. For an OPRL analog of Theorem 1.12.3, see Damanik–Simon [100].