initially for $f \in C(X)$, but then for $f \in L^1(d\nu)$ since $C(X)$ is dense in L^1. Thus L has the form

$$L(f) = \int wf\,d\nu$$

with $w \in L^\infty$. Since $L(f) \geq 0$, $w \geq 0$ and $w = d\mu/d\nu$. Moreover, $s_\infty = \sup_{f\in C(X)}\{\int f\,d\mu / \int f\,d\nu\} = \sup_{f\in L^1} |L(f)|/\|f\|_{L^1(d\nu)} = \|w\|_\infty$. This is (2.2.115). □

Of course, with the leading behavior of the extreme eigenvalues, one can ask about corrections. Kac, Murdock, and Szegő [**603**] showed that if f has a single nondegenerate minimum, the low-lying eigenvalues occur at $\min_\theta(f(\theta)) + c\nu^2/n^2$ where n is the size of the Toeplitz matrix, ν is the eigenvalue label, and c an explicit constant ($\frac{1}{2}\pi^2 f''(\theta_0)$ if θ_0 is the unique minimum). This led to a huge industry; see [**91, 107, 240, 241, 269, 508, 519, 520, 527, 601, 621, 728, 773, 774, 775, 839, 840, 956, 1054, 1055, 1056, 1075, 1093, 1108**]. See [**602**] for applications of these ideas to Schrödinger operators.

Since we are discussing asymptotics of the lowest eigenvalues of Toeplitz matrices, we mention a beautiful result of Berg, Chen, and Ismail [**108**] that the lowest eigenvalue of a Hankel matrix has a nonzero limit if and only if the associated OPRL moment problem is indeterminate.

2.3. Entropy Semicontinuity and the First Proof of Szegő's Theorem

In this section, we will prove the following, which we call Szegő's theorem:

THEOREM 2.3.1. *Let*

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

and let $\{\alpha_j\}_{j=0}^\infty$ *be the Verblunsky coefficients of* μ. *Then*

$$\prod_{j=0}^\infty (1-|\alpha_j|^2) = \exp\biggl(\int_0^{2\pi} \log(w(\theta))\,\frac{d\theta}{2\pi}\biggr) \tag{2.3.1}$$

Remarks. 1. As noted, the integral on the right can be $-\infty$, which the theorem asserts happens if and only if $\sum|\alpha_j|^2 = \infty$.

2. It is remarkable that (2.3.1) is independent of $d\mu_{\rm s}$!

3. Szegő's ideas involve the equality of (2.3.1) with two other objects, as we have seen: $\lim_{n\to\infty} \sqrt[n]{D_n(d\mu)}$ and $\lambda_\infty(0, d\mu)$. Theorems 2.7.14 and 2.7.15 will summarize huge numbers of equivalences that form "Szegő's theorem."

The key to our proof will be to view the right side of (2.3.1) as a function of μ and to prove that in the weak topology on measures, the right side is upper semicontinuous (usc). It certainly cannot be continuous since $d\mu = d\theta/2\pi$ (i.e., $w = 1$ so the integral is 0) is a weak limit of pure point measures for which $w = 0$, so the integral is $-\infty$. In fact, it is a special case of the entropy where:

Definition. Let μ, ν be two (positive) measures on a compact metric space X. Define their *relative entropy* in $\mathbb{R} \cup \{-\infty\}$ by

$$S(\mu \mid \nu) = \begin{cases} -\infty & \text{if } \mu \text{ is not } \nu \text{ a.c.} \\ -\int \log\left(\frac{d\mu}{d\nu}\right) d\mu & \text{if } \mu \text{ is } \nu \text{ a.c.} \end{cases} \tag{2.3.2}$$

EXAMPLE 2.3.2. If

$$d\nu = w(\theta)\,\frac{d\theta}{2\pi} + d\nu_{\rm s}$$

and $d\mu = d\theta/2\pi$, then μ is ν a.c. so long as $w > 0$ a.e. θ $(d\mu)$ and $d\mu/d\nu = w^{-1}$ so

$$S\biggl(\frac{d\theta}{2\pi}\biggm| w(\theta)\,\frac{d\theta}{2\pi} + d\nu_{\rm s}\biggr) = \int_0^{2\pi} \log(w(\theta))\,\frac{d\theta}{2\pi} \tag{2.3.3}$$

Similarly, if $\zeta \in D$ with $\zeta = re^{i\varphi}$ and

$$d\mu_\zeta(\theta) = P_r(\theta,\varphi)\,\frac{d\theta}{2\pi} \tag{2.3.4}$$

then

$$S\biggl(d\mu_\zeta \biggm| w(\theta)\,\frac{d\theta}{2\pi} + d\nu_{\rm s}\biggr) = \int_0^{2\pi} \log\biggl(\frac{w(\theta)}{P_r(\theta,\varphi)}\biggr) P_r(\theta,\varphi)\,\frac{d\theta}{2\pi} \tag{2.3.5}$$

which is a quantity that arose in (2.2.7). □

LEMMA 2.3.3 (Linear Variational Principle for S). *Let $\mathcal{E}(X)$ be the family of strictly positive continuous functions on X. Then*

$$S(\mu \mid \nu) = \inf_{f\in\mathcal{E}(X)} \mathcal{S}(f;\mu,\nu) \tag{2.3.6}$$

where

$$\mathcal{S}(f;\mu,\nu) = \int f(x)\,d\nu(x) - \int [1 + \log(f(x))]\,d\mu(x) \tag{2.3.7}$$

PROOF. Suppose first that μ is ν-a.c. and let $d\mu = g\,d\nu$ and $A = \{x \mid g(x) \neq 0\}$. Define for $a, b > 0$,

$$Q(a,b) = ab^{-1} - 1 - \log(a) \tag{2.3.8}$$

For b fixed, Q is convex in a and takes its minimum at $a = b$ with $Q(a,b) = -\log(b)$. Thus

$$Q(a,b) \geq -\log(b) \tag{2.3.9}$$

and, by convexity, Q is monotone decreasing in a on $(0,b)$ and monotone increasing in a on (b,∞).

Since

$$d\nu = \chi_{X\setminus A}\,d\nu + g^{-1}\,d\mu \tag{2.3.10}$$

we can write

$$\mathcal{S}(f;\mu,\nu) = \int_{X\setminus A} f(x)\,d\nu(x) + \int Q(f(x),g(x))\,d\mu(x) \tag{2.3.11}$$

By (2.3.9) and $f \geq 0$, we have

$$\mathcal{S}(f;\mu,\nu) \geq S(\mu \mid \nu) \tag{2.3.12}$$

proving half of (2.3.6).

For the other direction, we need an approximation argument. Let $\mathcal{E}(X;\mu,\nu) = \{f \in L^1(d\mu + d\nu) \mid \varepsilon \leq f \leq \varepsilon^{-1}$ for some $\varepsilon > 0\}$. For $f \in \mathcal{E}(X;\mu,\nu)$, define $\mathcal{S}(f;\mu,\nu)$ by (2.3.7). For any such f, there exists $f_n \in \mathcal{E}(X)$ so $\varepsilon \leq f_n \leq \varepsilon^{-1}$ uniformly in n and $f_n(x) \to f(x)$ for a.e. $x(d\mu + d\nu)$. By (2.3.7) and the dominated convergence theorem, $\mathcal{S}(f_n;\mu,\nu) \to \mathcal{S}(f;\mu,\nu)$, so

$$\inf_{f\in\mathcal{E}(X)} \mathcal{S}(f;\mu,\nu) = \inf_{f\in\mathcal{E}(X;\mu,\nu)} \mathcal{S}(f;\mu,\nu) \tag{2.3.13}$$

Define

$$f_n(x) = \begin{cases} n^{-1} & \text{if } x \notin A \text{ or } g(x) \le n^{-1} \\ g(x) & \text{if } n^{-1} \le g(x) \le n \\ n & \text{if } g(x) \ge n \end{cases}$$

Then $\int_{X\backslash A} f_n(x)d\nu(x) = n^{-1}\nu(X\backslash A) \to 0$. For each n, $\int |Q(f_n(x), g(x))|\, d\mu(x) < \infty$ since $|Q(f_n(x), g(x))| \le ng^{-1}(x)+1+\log(n)$ and $g^{-1} \in L^1(d\mu)$. Moreover, by the monotonicity properties of Q, $Q(f_n(x), g(x))$ is monotone decreasing to $-\log(g(x))$. Since $-\int \log(g(x))\, d\mu = S(\mu \mid \nu)$ (which may be $-\infty$), $\mathcal{S}(f_n; \mu, \nu) \to S(\mu \mid \nu)$, proving (2.3.6) if μ is ν-a.c.

If μ is not ν-a.c., find A with $\nu(A) = 0$ and $\mu(A) > 0$, and then, given ε, K compact and U_ε open so $K \subset A \subset U_\varepsilon$ and $\mu(K) > 0$, $\nu(U_\varepsilon) < \varepsilon$. By Urysohn's lemma, find $F_\varepsilon \in C(X)$ so $1 \le F_\varepsilon \le \varepsilon^{-1}$ for all x, $F_\varepsilon = \varepsilon^{-1}$ on K, and $F_\varepsilon = 1$ on $X\backslash U_\varepsilon$. Then

$$\begin{aligned} \mathcal{S}(F_\varepsilon; \mu, \nu) &\le \nu(X\backslash U_\varepsilon) + \varepsilon^{-1}\nu(U_\varepsilon) - \log(\varepsilon^{-1})\mu(K) \\ &\le 2 - \log(\varepsilon^{-1})\mu(K) \to -\infty \end{aligned}$$

so the inf is $-\infty$. □

THEOREM 2.3.4. *$S(\mu \mid \nu)$ is jointly concave and jointly weakly upper semicontinuous in μ and ν. Moreover, if μ and ν are both probability measures, $S(\mu \mid \nu) \le 0$ with equality only if $\mu = \nu$.*

Remark. By jointly concave, we mean

$$S(\theta\mu_1 + (1-\theta)\mu_0 \mid \theta\nu_1 + (1-\theta)\nu_0) \ge \theta S(\mu_1 \mid \nu_1) + (1-\theta)S(\mu_0 \mid \nu_0)$$

PROOF. (2.3.6) says that S is an inf of jointly continuous linear functions; hence, it is automatically jointly concave and jointly weakly upper semicontinuous.

If μ is ν-a.c., let $A = \{x \mid d\mu/d\nu(x) \ne 0\}$ and $d\tilde\nu = \chi_A\, d\nu$. Then

$$\begin{aligned} S(\mu \mid \nu) &= \int \log\biggl(\frac{d\tilde\nu}{d\mu}\biggr)\, d\mu \\ &\le \log\biggl(\int \frac{d\tilde\nu}{d\mu}\, d\mu\biggr) \qquad \text{(by Jensen's inequality)} \\ &= \log(\nu(A)) \le 0 \end{aligned}$$

Since log is strictly concave, we have equality if and only if $\nu(A) = 1$ and $d\nu/d\mu$ is a constant, hence 1. □

Remark. One tends to think of the lack of full continuity as connected with zeros of $d\mu/d\nu$ or cases where μ_n are not ν-a.c. but $\mu = \text{w-lim}\,\mu_n$ is. But noncontinuity can happen in more benign-looking cases. On $\partial\mathbb{D}$, let $\mu_n = [1 - \alpha\cos(n\theta)]\frac{d\theta}{2\pi}$ where $|\alpha| < 1$, α real. By a simple change of variables,

$$\begin{aligned} S\biggl(\frac{d\theta}{2\pi}\,\bigg|\,\mu_n\biggr) &= \int \log(1 - \alpha\cos(\theta))\,\frac{d\theta}{2\pi} \\ &= \int \log\biggl(\biggl|z - \frac{\alpha}{2} - \frac{\alpha z^2}{2}\biggr|\biggr)\frac{dz}{2\pi} \\ &= \log\bigl[\tfrac{1}{2}\bigl(1 + \sqrt{1-\alpha^2}\,\bigr)\bigr] \end{aligned}$$

where we used Jensen's formula and the fact that the zero, z_0, of $f(z) = \frac{\alpha z^2}{2} + \frac{\alpha}{2} - z$ has $\frac{1}{|z_0|} = \frac{2}{\alpha}[\frac{1}{2}(1+\sqrt{1-\alpha^2})]$. Thus, if $\alpha \neq 0$,

$$\lim S\biggl(\frac{d\theta}{2\pi}\biggm| \mu_n\biggr) < S\biggl(\frac{d\theta}{2\pi}\biggm| \frac{d\theta}{2\pi}\biggr) = 0$$

Proof of Theorem 2.3.1. Let

$$d\mu_n = \frac{d\theta}{2\pi}\,|\varphi_n^*(e^{i\theta};d\mu)|^{-2} \tag{2.3.14}$$

be the Bernstein-Szegő approximation to $d\mu$. By (1.5.78) and Theorem 2.1.4,

$$\prod_{j=0}^{n-1}(1-|\alpha_j|^2) = \exp\biggl[S\biggl(\frac{d\theta}{2\pi}\,, d\mu_n\biggr)\biggr] \tag{2.3.15}$$

By (2.2.7) (for $\zeta = 0$) and (2.2.3),

$$\prod_{j=0}^{\infty}(1-|\alpha_j|^2) \geq \exp\biggl[S\biggl(\frac{d\theta}{2\pi}\,, d\mu\biggr)\biggr] \tag{2.3.16}$$

In (2.3.15), let $n\to\infty$. Clearly, $\prod_{j=0}^{n-1}(1-|\alpha_j|^2) \to \prod_{j=0}^{\infty}(1-|\alpha_j|^2)$. Since $d\mu_n \to d\mu$ weakly (Theorem 1.7.8), the weak usc of S implies

$$\limsup S\biggl(\frac{d\theta}{2\pi}\,, d\mu_n\biggr) \leq S\biggl(\frac{d\theta}{2\pi}\,, d\mu\biggr) \tag{2.3.17}$$

so (2.3.15) says

$$\prod_{j=0}^{\infty}(1-|\alpha_j|^2) \leq \exp\biggl[S\biggl(\frac{d\theta}{2\pi}\,, d\mu\biggr)\biggr] \tag{2.3.18}$$

(2.3.16) and (2.3.18) imply (2.3.1). □

Remark. By our reference to (2.2.7), a result for general ζ, the simplicity of (2.3.16) may be obscured. Basically, it comes from

$$\begin{aligned}\prod_{j=0}^{n-1}(1-|\alpha_j|^2) &= \|\Phi_n^*\|^2 \\ &\geq \int \exp(\log|\Phi_n^*(e^{i\theta})|^2 + \log w(\theta))\,\frac{d\theta}{2\pi} \qquad (2.3.19)\\ &\geq \exp\biggl(\int \log w(\theta)\biggr)\,\frac{d\theta}{2\pi} \qquad (2.3.20)\end{aligned}$$

where (2.3.19) comes from $d\mu \geq w(\theta)\frac{d\theta}{2\pi}$ and (2.3.20) from Jensen's inequality and $\int \log|\Phi_n^*(e^{i\theta})|\frac{d\theta}{2\pi} = \log|\Phi_n^*(0)| = 0$ since $\log|\Phi_n^*|$ is harmonic on account of the lack of zeros of Φ_n^* in $\mathbb{D}$. (2.3.20) implies (2.3.16) by taking limits.

Since (2.2.3) says $\prod_{j=0}^{\infty}(1-|\alpha_j|^2) = \lambda_\infty(0)$, Szegő's theorem can be regarded as the $\zeta = 0$ case of the following more general theorem:

Theorem 2.3.5 (Generalized Szegő Theorem). *Let $d\mu = w(\theta)\frac{d\theta}{2\pi} + d\mu_{\mathrm{s}}$. Then for any $\zeta \in \mathbb{D}$ with $\zeta = re^{i\varphi}$,*

$$\lambda_\infty(\zeta) = \exp\biggl(\int_0^{2\pi} P_r(\theta,\varphi)\log\biggl(\frac{w(\theta)}{P_r(\theta,\varphi)}\biggr)\frac{d\theta}{2\pi}\biggr) \tag{2.3.21}$$

Remarks. 1. (2.3.21) is intended to say $\lambda_\infty(\zeta) = 0$ if and only if $\int_0^{2\pi} \log(w(\theta))\, d\theta = -\infty$, proving Proposition 2.2.2 again.

2. (2.3.21) says $\lambda_\infty(\zeta)$ is independent of $d\mu_s$!

Proof. Let $d\mu^{(\zeta)}(\theta) = P_r(\theta,\varphi)\frac{d\theta}{2\pi}$ and define $d\mu_n$ by (2.3.14). Then (2.2.7) and (2.3.5) say that

$$\lambda_\infty(\zeta, d\mu) \geq \exp[S(d\mu^{(\zeta)} \mid d\mu)] \tag{2.3.22}$$

for any measure. Theorem 2.2.3(ii) says

$$\lambda_\infty(\zeta, d\mu_n) = \exp[S(d\mu^{(\zeta)} \mid d\mu_n)] \tag{2.3.23}$$

and Theorem 2.2.6 says

$$\lim_{n\to\infty} \lambda_\infty(\zeta, d\mu_n) = \lambda_\infty(\zeta, d\mu) \tag{2.3.24}$$

By usc of the entropy and Theorem 1.7.8,

$$\limsup S(d\mu^{(\zeta)} \mid d\mu_n) \leq S(d\mu^{(\zeta)} \mid d\mu) \tag{2.3.25}$$

(2.3.23)–(2.3.25) imply

$$\lambda_\infty(\zeta, d\mu) \leq \exp[S(d\mu^{(\zeta)} \mid d\mu)] \tag{2.3.26}$$

(2.3.22), (2.3.26), and (2.3.5) imply (2.3.21). □

We close this section by seeing another way to go from Theorem 2.3.1 to Theorem 2.3.5. We begin with

Lemma 2.3.6. *Let $\mathcal{A}(\mathbb{D})$ denote the set of continuous functions on $\bar{\mathbb{D}}$ analytic on $\mathbb{D}$. Then*

$$\lambda_\infty(\zeta) = \inf\biggl(\int |F(e^{i\theta})|^2\, d\mu(\theta) \biggm| F(\zeta) = 1,\ F \in \mathcal{A}(D)\biggr)$$

Proof. By replacing $F(e^{i\theta})$ by $F(re^{i\theta})$ and letting $r \uparrow 1$, we see the inf over $\mathcal{A}(\mathbb{D})$ is the same as the inf over functions analytic in a neighborhood of $\bar{\mathbb{D}}$ (since $F(re^{i\theta}) \to F(e^{i\theta})$ uniformly, integrals over μ converge even if μ has a singular part). Such functions can be uniformly approximated by polynomials, so the inf over polynomials is the same. □

Proof of Theorem 2.3.5 from Theorem 2.3.1. Fix $\zeta \in \mathbb{D}$. Let g be the conformal map of $\bar{\mathbb{D}} \to \bar{\mathbb{D}}$,

$$g(z) = \frac{z-\zeta}{1-\bar{\zeta}z}$$

Define $\tilde{g} : [0,2\pi) \to [0,2\pi)$ by $g(e^{i\theta}) = e^{i\tilde{g}(\theta)}$. If $\zeta = re^{i\varphi}$, a straightforward calculation shows that

$$\frac{d\tilde{g}}{d\theta} = P_r(\varphi,\theta) \tag{2.3.27}$$

We will also need another direct calculation,

$$P_r(\pi+\varphi, \tilde{g}(\theta)) = \frac{1}{P_r(\varphi,\theta)} \tag{2.3.28}$$

and that

$$g^{-1}(z) = \frac{z+\zeta}{1+\bar{\zeta}z}$$

so, by (2.3.28),

$$\frac{d\tilde{g}^{-1}}{d\theta} = P_r(\pi + \varphi, \theta) \tag{2.3.29}$$

Since $g(\zeta) = 0$,

$$\begin{aligned}
\lambda_\infty(\zeta, d\mu) &= \inf\biggl(\int |F(e^{i\theta})|^2 \, d\mu(\theta) \biggm| (F \circ g^{-1})(0) = 1,\, F \in \mathcal{A}(\mathbb{D})\biggr\} \\
&= \inf\biggl(|F \circ g(e^{i\theta})|^2 \, d\mu(\theta) \biggm| F(0) = 1,\, F \in \mathcal{A}(\mathbb{D})\biggr\} \\
&= \inf\biggl(\int |F(e^{i\theta})|^2 \, d\mu(\tilde{g}^{-1}(\theta)) \biggm| F(0) = 1,\, F \in \mathcal{A}(\mathbb{D})\biggr\} \\
&= \lambda_\infty(0, d\mu \circ \tilde{g}^{-1})
\end{aligned}$$

If $d\mu(\theta) = w(\theta)\frac{d\theta}{2\pi} + d\mu_{\rm s}(\theta)$, since $\tilde{g}^{-1}$ is C^∞ with C^∞ inverse,

$$d\mu \circ \tilde{g}^{-1} = w(\tilde{g}^{-1}(\theta)) \, \frac{d\tilde{g}^{-1}}{d\theta} \, \frac{d\theta}{2\pi} + d\tilde{\mu}_{\rm s}$$

Thus, by Theorem 2.3.1 and (2.3.29),

$$\begin{aligned}
\lambda_\infty(0, d\mu \circ \tilde{g}^{-1}) &= \exp\biggl(\int_0^{2\pi} \log[P_r(\pi + \varphi, \theta) w(\tilde{g}^{-1}(\theta))] \, \frac{d\theta}{2\pi}\biggr) \\
&= \exp\biggl(\int_0^{2\pi} \log[P_r(\pi + \varphi, \tilde{g}(\theta)) w(\theta)] \, \frac{d\tilde{g}}{d\theta} \, \frac{d\theta}{2\pi}\biggr) \\
&= \exp\biggl(\int_0^{2\pi} \log\biggl[\frac{w(\theta)}{P_r(\varphi, \theta)}\biggr] P_r(\varphi, \theta) \, \frac{d\theta}{2\pi}\biggr)
\end{aligned}$$

by (2.3.27) and (2.3.28). This proves (2.3.21). □

Remarks and Historical Notes. In his first pair of papers on OPUC, Szegő [**1018, 1019**] proved the equality of $\lim \sqrt[n]{D_n(d\mu)}$, $\exp(\int_0^{2\pi} \log(w(\theta)) \frac{d\theta}{2\pi})$, and $\lambda_\infty(0)$ for measures of the form $d\mu = w(\theta)\frac{d\theta}{2\pi}$. Verblunsky [**1067**] is responsible both for the fourth equality to $\prod_{j=0}^\infty (1 - |\alpha_j|^2)$ and for handling a singular continuous component.

In the West, Verblunsky's paper was not appreciated — and the standard story line is that Kolmogorov [**642**] had the first results which allowed singular components; he did not have the full equality. Rather he proved that the polynomials were dense in $L^2(\partial\mathbb{D}, d\mu)$ if and only if $\int \log(w(\theta))\frac{d\theta}{2\pi} = -\infty$. Krein [**652, 653**] extended this to $L^p(\partial\mathbb{D}, d\mu)$ (see Section 2.5 below), realized the connection to Szegő's work, and found analogs for $L^2(\mathbb{R}, d\mu)$ (see below). According to the standard story, the full theorem with the equality for general μ is due to Szegő in 1958 in [**479**] (while the book has two authors, the introduction says that the chapters on this theorem are due to Szegő). There are two problems with this story. First, Verblunsky [**1067**] had the full result 20+ years before Szegő and even five years before Kolmogorov. Second, Geronimus has the result in his 1958 book [**407**] and one of his students claims [**6**] he found it independently of Verblunsky. That is possible — but if so, Geronimus must have known of it very early since his sporadic quotations of Verblunsky began in the 1940's.

Szegő's theorem has been extended to a variety of other situations. Let $d\sigma$ be a finite measure on $(-\infty, \infty)$. Krein [**652, 653**] showed that $\{e^{i\alpha x}\}_{\alpha \geq 0}$ is dense in

$L^2(\mathbb{R}, d\sigma)$ if and only if $d\sigma = w\,dx + d\sigma_{\rm s}$ and $\int_{-\infty}^{\infty} |\log w|(1+x^2)^{-1}\,dx = \infty$. If this integral is finite, this family is not dense and, a fortiori, if $\int |x|^n\,d\sigma(x) < \infty$ for all n, $\{x^n\}_{n=0}^{\infty}$ is not dense so the moment problem is indeterminate. Krein also noted that for w supported on $[0,\infty)$, if

$$\int_0^\infty \frac{|\log w(x)|}{1+x}\,\frac{dx}{\sqrt{x}} < \infty$$

then the Stieltjes moment problem is indeterminate.

For a discussion of limit theorems for $D_{n+1}(c)/D_n(c)$ for c's which do not come from a measure (e.g., c's which are Fourier coefficients of an L^∞ function), see Böttcher-Silbermann [**146**, Section 5.4] and references therein.

In Section 13.8, we will discuss analogs of Szegő's theorem for OPRL using ideas close to the proof in this section. Indeed, our proof here is patterned after the Killip-Simon [**633**] proof of Theorem 13.8.6. The proof here, in turn, is related to Verblunsky's, as we will discuss shortly.

Semicontinuity of $\mu \mapsto S(\mu \mid \nu)$ for fixed ν is a standard part of the theory of entropy. Semicontinuity in ν and joint semicontinuity is not so commonly known but does appear sporadically; see, for example, Topsøe [**1048**].

Variational principles for thermodynamic quantities go back to Gibbs. While Gibbs focused on a principle for the free energy, there is an inverse Gibbs principle for the entropy, essentially by Fenchel's theorem [**347**] on double Legendre transforms. The modern era for such principles was ushered in by Lanford-Robinson [**674**]; see Simon [**967**]. The real analog of the Gibbs principle for a pair, μ, ν, of probability measures is not (2.3.6) but

$$S(\mu \mid \nu) = \inf_{g \in C(X)} \left[\log \int e^g\,d\nu - \int g\,d\mu \right] \tag{2.3.30}$$

If we set $\mathcal{G}(g) = \log \int e^g\,d\nu - \int g\,d\mu$ and $\mathcal{K}(f) = \int f(x)\,d\nu - 1 - \int \log(f(x))\,d\mu$, we have

$$\mathcal{K}(e^g) \ge \mathcal{G}(g) \ge S(\mu \mid \nu) \tag{2.3.31}$$

The first inequality in (2.3.31) comes from $y - 1 \ge \log(y)$ for $y > 0$ (use concavity of $\log(y)$). The second comes from

$$\int e^g\,d\nu \ge \int_A e^g \left(\frac{d\mu}{d\nu}\right)^{-1} d\mu \ge \exp\biggl(\int g\,d\mu - \int \log\left(\frac{d\mu}{d\nu}\right) d\mu\biggr)$$

via Jensen's inequality. (2.3.31) and (2.3.6) imply (2.3.30).

Remarkably, while he did not know he was using an entropy, a key element of Verblunsky's proof was the formula for $d\mu = w(\theta)\frac{d\theta}{2\pi} + d\mu_{\rm s}$:

$$\inf \left[\frac{\int e^g\,d\mu}{\exp\left(\int g \frac{d\theta}{2\pi}\right)} \right] = \exp\biggl(\int \log(w(\theta))\,\frac{d\theta}{2\pi} \biggr) \tag{2.3.32}$$

which is, of course, (2.3.30) for $S((d\theta/2\pi) \mid \mu)$! While he does not organize his proof as we do, there is clearly a close connection between his proof and the one in this section.

Verblunsky goes from (2.3.32) to (2.3.1) as follows: Let $\mathcal{L}_n$ be the positive Laurent polynomials spanned by $\{z^j\}_{j=-n}^{n}$. Since $\cup_n \mathcal{L}_n$ is dense in $C(X)$, (2.3.32) implies

$$\inf_n V_n(\mu) = \exp\biggl(\int \log(w(\theta))\,\frac{d\theta}{2\pi} \biggr) \tag{2.3.33}$$

where

$$V_n(\mu) = \inf_{P_n \in \mathcal{L}_n} \left[\frac{\int P_n \, d\mu}{\exp\left(\int \log(P_n) \frac{d\theta}{2\pi}\right)} \right] \tag{2.3.34}$$

Next, he used that Theorem 1.7.8 implies

$$\int P_n \, d\mu = \int P_n \, \frac{\prod_{j\equiv 0}^{n-1}(1-|\alpha_j|^2)}{|\Phi_n^*|^2} \, \frac{d\theta}{2\pi} \tag{2.3.35}$$

for any $P_n \in \mathcal{L}_n$, so in (2.3.34), we can replace $d\mu$ by the Bernstein-Szegő approximation.

Taking $P = |\Phi_n^*|^2 = \Phi_n^*(z)\,\overline{\Phi_n^*(z)} \in \mathcal{L}_n$, we see that the inf is no more than $\prod_{j=1}^{n-1}(1-|\alpha_j|^2)$ since $\int \log|\Phi_n^*(z)| \frac{d\theta}{2\pi} = 1$ by (2.1.16). On the other hand, using Jensen's inequality and (2.1.16), we see that $\prod_{j=1}^{n-1}(1-|\alpha_j|^2)$ is always a lower bound. Thus

$$V_n(\mu) = \prod_{j=1}^{n-1}(1-|\alpha_j|^2) \tag{2.3.36}$$

so (2.3.33) implies (2.3.1).

That $\int \log(w(\theta)) \frac{d\theta}{2\pi}$ is an entropy (although not the resulting semicontinuity) has been noted in the literature; see Foias-Frazho [**358**].

One consequence of Szegő's theorem is that among all measures μ with a given set $c_0 = 1, \dots, c_n$ of moments, $\int \log(w(\theta)) \frac{d\theta}{2\pi}$ is maximized by taking $d\mu$ to be the Bernstein-Szegő approximation $\frac{d\theta}{2\pi}|\varphi_n|^{-2}$ with φ_n the OPUC determined by $c_0, \dots, c_n$. For all such μ have the same $\alpha_0, \dots, \alpha_{n-1}$, so $\prod_{j=0}^{\infty}(1-|\alpha_j|^2) \le \prod_{j=0}^{n-1}(1-|\alpha_j|^2)$ = value for Bernstein-Szegő approximation. This maximum entropy property has been extensively discussed in the engineering literature [**282, 283, 313, 358, 749, 786**].

In a sense, the miracle of the proof of Verblunsky and the one in this section, as opposed to the one in Section 2.5, is that the fact that $d\mu_{\rm s}$ does not matter passes by without any explicit argument. In a sense, it is in the proof of (2.3.6). Making f small on a set of small μ measure decreases the ν integral without a cost in the μ integral. There is another way of putting this: The proof in Section 2.5 requires us to show that small sets do not matter in approximating by analytic functions, and that is a little tricky. In (2.3.6), only continuous functions enter.

For a history of ideas related to Theorem 2.3.4, see the discussion in Killip-Simon [**633**]. Theorem 2.3.5 is also due to Szegő in [**1018, 1019**]. The discussion under "Proof of Theorem 2.3.5 from Theorem 2.3.1" using conformal mapping fleshes out a remark in Máté-Nevai-Totik [**761**].

2.4. The Szegő Function

As a part of the analysis of what happens when the *Szegő condition*

$$\int_0^{2\pi} \log(w(\theta)) \, \frac{d\theta}{2\pi} > -\infty \tag{2.4.1}$$

holds, Szegő introduced a natural analytic function, $D(z)$. It will turn out to be $1/\Delta(z)$ with $\Delta(z)$ the function of (2.2.92), but that function is part of the Freudian parallel universe and its construction will *not* be assumed in the first part of this section.