number include Fisher-Hartwig [**351**], Böttcher-Silbermann [**145**], Widom [**1093**], Day [**243**], Trench [**1053**], Berg [**109**], Høholdt-Justesen [**538**], Gorodetsky [**474**], Bart-Gohberg-Kaashoek [**85**], Gohberg-Kaashoek-van Schagen [**438**], Basor-Chen [**89**], and Carey-Pincus [**186**].

In [**1098, 1099**], Widom discussed Szegő's theorem in case the c's in a Toeplitz matrix are themselves matrices.

Kac [**599**] discussed the relevance of the strong Szegő theorem to the asymptotics of $\frac{1}{n}\text{Tr}(T_n^\ell)$, finding explicit two-term formulae; see Section 6.5 and our discussion of Libkind [**700**] in the Notes to Section 2.7.

For extensions of the strong Szegő theorem to general manifolds, see the discussion at the end of the Notes to Section 2.7.

The ideas behind our Example 6.1.14 are motivated by Weierstrass' construction of nowhere differentiable functions [**923**, Example 7.18], the theory of Haar and Walsh functions [**824**], and the theory of wavelets [**237, 555, 1079, 1110**]. They are surely well-known to harmonic analysts.

6.2. The Borodin-Okounkov Formula

In this section, we will provide a second proof of Ibragimov's theorem that depends neither on the calculations in Theorem 2.1.3 nor on Baxter's theorem nor on the calculations in Lemma 6.1.4. Instead it will depend on a remarkable exact formula for

$$\frac{D_n(w\frac{d\theta}{2\pi})}{\exp((n+1)\hat{L}_0 + \sum_{k=1}^\infty k|\hat{L}_k|^2)} \tag{6.2.1}$$

whose form is such that it will immediately imply the ratio goes to 1 as $n \to \infty$. Exact formulae are often more algebraic than analytic, and that is true here: While there will be a few analytic steps (i.e., estimates), this section will largely involve some remarkable algebra.

The exact formula for (6.2.1) will involve Hankel operators, so we begin by discussing them, first on the level of sequences. Given any sequence $\{a_n\}_{n\in\mathbb{Z}}$, one defines the matrices $t(a)$, $h(a)$, and $h(\tilde{a})$ by

$$t(a) = \begin{pmatrix} a_0 & a_1 & a_2 & \dots \\ a_{-1} & a_0 & a_1 & \dots \\ a_{-2} & a_{-1} & a_0 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \qquad h(a) = \begin{pmatrix} a_1 & a_2 & a_3 & \dots \\ a_2 & a_3 & a_4 & \dots \\ a_3 & a_4 & a_5 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$
$$h(\tilde{a}) = \begin{pmatrix} a_{-1} & a_{-2} & a_{-3} & \dots \\ a_{-2} & a_{-3} & a_{-4} & \dots \\ a_{-3} & a_{-4} & a_{-5} & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \tag{6.2.2}$$

that is, with $i, j \geq 0$,

$$t(a)_{ij} = a_{i-j} \qquad h(a)_{ij} = a_{i+j+1} \qquad h(\tilde{a})_{ij} = a_{-i-j-1} \tag{6.2.3}$$

t is a *Toeplitz matrix*, h a *Hankel matrix*. Toeplitz matrices are constant along the usual matrix diagonals, Hankel matrices along the opposite diagonals.

This notation is consistent with the definition

$$\tilde{a}_j = a_{-j} \tag{6.2.4}$$

Notice that t depends on all a_j but $h(a)$ only on a_j for $j \geq 1$.

$t(a)$ and $h(a)$ define potentially unbounded operators on $\ell^2(\mathbb{Z}_+)$ by taking as domain $\mathcal{F} = \{\{a_n\} \mid a_n \equiv 0 \text{ for every } n \text{ large}\}$. If $t(a)$ (or $h(a)$) is bounded, that is, $\|t(a)\psi\| \le C\|\psi\|$ for all $\psi \in \mathcal{F}$, we can extend the operator to all of ℓ^2 and we use the same symbol.

Given f, a distribution on $\partial\mathbb{D}$, let

$$a_j = \int e^{-ij\theta} f(\theta)\, d\theta \tag{6.2.5}$$

and define

$$T(f) = t(a) \qquad H(f) = h(a) \qquad H(\tilde{f}) = h(\tilde{a}) \tag{6.2.6}$$

T and H are called *Toeplitz operators* and *Hankel operators*; f is called the *symbol* of T or H. In this section, our symbols will be $L^1(\partial\mathbb{D}, \frac{d\theta}{2\pi})$ functions.

Here is one reason why it is useful to employ symbols:

THEOREM 6.2.1. *If $f \in L^\infty(\partial\mathbb{D}, \frac{d\theta}{2\pi})$, $T(f)$ and $H(f)$ are bounded operators. Indeed,*

$$\|T(f)\| \le \|f\|_\infty \tag{6.2.7}$$

$$\|H(f)\| \le \|f\|_\infty \tag{6.2.8}$$

Remark. As we will discuss in the Notes, equality holds in (6.2.7), so $T(f)$ is bounded if and only if $f \in L^\infty$. It is for this reason that we will want to approximate f's by ones for which $f \in L^\infty$. Again, as we will discuss in the Notes, equality in (6.2.8) is more subtle. $H(f)$ only depends on $\{a_n\}_{n\ge 1}$ so $f \mapsto H(f)$ has a large kernel. Equality holds in (6.2.8) if one replaces $\|f\|_\infty$ by $\|f\|_{L^\infty/\ker(H)}$, that is, by

$$\inf_g \left\{ \|g\|_\infty \;\middle|\; \int (g-f) e^{-ij\theta}\, d\theta = 0;\ j = 1, 2, \dots \right\}$$

This refined result is *Nehari's theorem.*

PROOF. Let $\mathcal{H} = \ell^2(\mathbb{Z})$. On $\mathcal{H}$, define P_+, P_-, J by

$$(P_+x)_n = \chi_{[0,\infty)}(n) x_n \qquad (P_-x)_n = \chi_{(-\infty,-1]}(n) x_n \qquad Jx_n = x_{-n-1} \tag{6.2.9}$$

Also, for any f with a given by (6.2.5), define the convolution operator $C(f)$ by

$$(C(f)x)_n = \sum_m a_{n-m} x_m$$

initially for x's that have finitely many nonzero elements. If $f \in L^\infty$, then

$$C(f)x = (f\check{x})^\wedge \tag{6.2.10}$$

where $^\wedge$ is Fourier transform from $L^2(\partial\mathbb{D}, \frac{d\theta}{2\pi})$ to $\ell^2(\mathbb{Z})$ and $^\vee$ is its inverse. (6.2.10) and unitarity of Fourier transform implies that

$$\|C(f)x\|_{\mathcal{H}} \le \|f\|_\infty \|x\|_{\mathcal{H}}$$

so

$$\|C(f)\| \le \|f\|_\infty \tag{6.2.11}$$

To obtain (6.2.7) and (6.2.8), we use these important, easy-to-check, identities:

$$T(f) = P_+ C(f) P_+ \restriction P_+\mathcal{H} \tag{6.2.12}$$

$$H(f) = P_+ C(f) J P_+ \restriction P_+\mathcal{H} \tag{6.2.13}$$

$$H(\tilde{f}) = P_+ J C(f) P_+ \restriction P_+\mathcal{H} \tag{6.2.14}$$

Since J and P_+ are contractions, (6.2.14) immediately implies (6.2.7) and (6.2.8). $\square$

Since

$$JP_+ = P_- J \qquad JP_- = P_+ J \tag{6.2.15}$$

we can rewrite (6.2.13) and (6.2.14) as

$$H(f) = P_+ C(f) P_- J \tag{6.2.16}$$

$$H(\tilde{f}) = JP_- C(f) P_+ \tag{6.2.17}$$

Notice that

$$t(a)^t = t(\tilde{a})$$

and

$$h(a)^t = h(a)$$

so

$$t(a)^* = t(\bar{\tilde{a}}) \tag{6.2.18}$$

$$h(a)^* = h(\bar{a}) \tag{6.2.19}$$

where $\bar{a}_j = \overline{a_j}$. If $a(f)$ is given by (6.2.5), then $a(\bar{f}) = (a(f))^{\bar{\sim}}$, so (6.2.18) becomes

$$T(f)^* = T(\bar{f}) \tag{6.2.20}$$

Moreover, (6.2.19) becomes

$$H(f)^* = H(\bar{\tilde{f}}) \tag{6.2.21}$$

One might wonder why Hankel operators arise in studying Toeplitz operators. The reason is the following simple but fundamental fact:

THEOREM 6.2.2. *Let $f, g \in L^\infty$. Then*

$$T(fg) = T(f)T(g) + H(f)H(\tilde{g}) \tag{6.2.22}$$

PROOF. We have that

$$\begin{aligned} T(fg) &= P_+ C(f) C(g) P_+ \\ &= P_+ C(f)(P_+ + P_-) C(g) P_+ \\ &= (P_+ C(f) P_+)(P_+ C(g) P_+) + (P_+ C(f) P_- J)(JP_- C(g) P_+) \end{aligned}$$

since $P_- J^2 P_- = P_-^2 = P_-$. (6.2.12), (6.2.16), and (6.2.17) complete the proof. $\square$

$H^\infty(\mathbb{D})$ is the set of functions, f, in $L^\infty(\partial\mathbb{D})$ with a given by (6.2.5) obeying $a_j = 0$ for $j < 0$. $\overline{H^\infty(\mathbb{D})}$, the set of f's with $\bar{f} \in H^\infty$, clearly obeys $a_j = 0$ if $j > 0$. Thus

$$f \in H^\infty \Rightarrow H(\tilde{f}) = 0 \tag{6.2.23}$$

$$f \in \overline{H^\infty} \Rightarrow H(f) = 0 \tag{6.2.24}$$

This implies

COROLLARY 6.2.3. *Let $g \in L^\infty$. If $f \in H^\infty$, then $T(fg) = T(g)T(f)$. If $f \in \overline{H^\infty}$, then $T(fg) = T(f)T(g)$.*

Proof. If $f \in H^\infty$, by (6.2.22) and (6.2.23),

$$T(fg) - T(g)T(f) = H(g)H(\tilde{f}) = 0$$

and if $f \in \overline{H^\infty}$,

$$T(fg) - T(f)T(g) = H(f)H(\tilde{g}) = 0 \qquad \square$$

These formulae provide another way of seeing the Wiener-Hopf theorem:

Corollary 6.2.4. *Let $x_+ \in H^\infty$, $x_- \in \overline{H^\infty}$ with $\inf_{z\in\bar{\mathbb{D}}}|x_+(z)| > 0$ and $\inf_{z\in\mathbb{C}\setminus\mathbb{D}}|x_-(z)| > 0$. Then $T(x_+x_-)$ is invertible and*

$$T(x_+x_-)^{-1} = T(x_+^{-1})T(x_-^{-1}) \tag{6.2.25}$$

Proof. By hypothesis, $x_+^{-1} \in H_\infty$ and $x_-^{-1} \in \overline{H_\infty}$. Thus, by Corollary 6.2.3,

$$\begin{aligned} T(x_+x_+^{-1}) &= T(x_+)T(x_+^{-1}) \\ &= T(x_+^{-1})T(x_+) \end{aligned}$$

so $T(x_\pm)$ are invertible and

$$T(x_\pm)^{-1} = T(x_\pm^{-1}) \tag{6.2.26}$$

By Corollary 6.2.3 again,

$$T(x_+x_-) = T(x_-)T(x_+)$$

so, since $T(x_\pm)$ are invertible and

$$\begin{aligned} T(x_+x_-)^{-1} &= T(x_+)^{-1}T(x_-)^{-1} \\ &= T(x_+^{-1})T(x_-^{-1}) \end{aligned}$$

by (6.2.26). $\square$

Now let $w \geq 0$ with $\log w \in L^1(\partial\mathbb{D}, \frac{d\theta}{2\pi})$ and define D by (2.4.2) and define on $\partial\mathbb{D}$,

$$b = \frac{\bar{D}}{D} \qquad c = \frac{D}{\bar{D}} \tag{6.2.27}$$

in terms of the a.e. boundary values of D. Clearly, $|b| = |c| = 1$, so $b, c \in L^\infty$, and so are symbols of bounded Toeplitz and Hankel operators.

As in Section 5.1, we define Q_n and R_n on $\ell^2(\mathbb{Z}_+)$ by

$$(Q_n a)_j = \chi_{[0,n]}(j)a_j \qquad R_n = 1 - Q_n \tag{6.2.28}$$

Here are the two facts that we will prove below that are the sharp Borodin-Okounkov formula:

Fact 1. If $\log w = \sum_k \hat{L}_k e^{ik\theta}$ and

$$\sum_k |k|\,|\hat{L}_k|^2 < \infty \tag{6.2.29}$$

then $H(b)$ and $H(\tilde{c})$ are Hilbert-Schmidt, so $H(b)H(\tilde{c})$ is trace class.

Fact 2. If (6.2.29) holds, then

$$(6.2.1) = \det\nolimits_{R_n\mathcal{H}}(1 - R_nH(b)H(\tilde{c})R_n) \tag{6.2.30}$$

Remarks. 1. By (6.2.21) and $b = \bar{c}$, we have

$$H(b)^* = H(\tilde{c}) \tag{6.2.31}$$

and by (6.2.20),

$$T(b)^* = T(c)$$

2. $\det_{R_n\mathcal{H}}$ is the Fredholm determinant for 1+ trace class operators; see Subsection 1.4.12.

3. (6.2.30) is the *Borodin-Okounkov formula.* As we will discuss in the Notes, it was found by Geronimo-Case almost twenty-five years before Borodin-Okounkov! We say this is the sharp form because earlier versions had stronger conditions than (6.2.29), which is not only sufficient for the formula to hold, but necessary for (6.2.1) to make sense.

The proof will depend on the following sequence of steps:

Step 1. Introduce the space $H^{1/2}$ of functions, f, on $\partial\mathbb{D}$ with $\frac{1}{2}$ derivative and prove that if $\log w \in H^{1/2}$, then $b, c \in H^{1/2}$ and the map $\log w \to b$ (or c) is continuous in $H^{1/2}$-norm.

Step 2. Prove that if $f \in H^{1/2}$, $H(f)$ is Hilbert-Schmidt and $f \mapsto H(f)$ is continuous. Because of this, it will suffice to prove (6.2.2) for a dense set of $\log w$ in $H^{1/2}$; we will take those w's with $L = \log w$ obeying

$$\sum_{k=0}^{\infty} |\hat{L}_k| < \infty \tag{6.2.32}$$

Step 3. Let $\mathcal{H}_+ = P_+\mathcal{H} = \ell^2(0,1,2,\dots)$. Prove that if (6.2.15) and (6.2.32) hold, then

$$\det_{Q_n\mathcal{H}_+}(Q_n(T(b)T(c))^{-1}Q_n) = \frac{\det_{P_n\mathcal{H}_+}(1 - R_nH(b)H(\tilde{c})R_n)}{\det_{\mathcal{H}_+}(1 - H(b)H(\tilde{c}))} \tag{6.2.33}$$

This will essentially be Jacobi's relation (3.1.5).

Step 4. Prove that if (6.2.32) holds, then

$$\det(Q_n(T(b)T(c))^{-1}Q_n) = e^{-(n+1)\hat{L}_0} D_{n+1}\biggl(w\,\frac{d\theta}{2\pi}\biggr) \tag{6.2.34}$$

Step 5. Prove that if (6.2.32) holds (and then, by a limiting argument whenever (6.2.29) holds), one has *Widom's formula*,

$$\det(1 - H(b)H(\tilde{c})) = \exp\biggl(-\sum_{k=1}^{\infty} k|\hat{L}_j|^2\biggr) \tag{6.2.35}$$

Definition. $H^{1/2}(\partial\mathbb{D})$ is the Sobolev space of functions, f, on $L^2(\partial\mathbb{D}, \frac{d\theta}{2\pi})$ with

$$\sum_{k=-\infty}^{\infty} |k|\,|\hat{f}_k|^2 < \infty \tag{6.2.36}$$

We define an inner product on $H^{1/2}$ by

$$\langle f, g\rangle_{1/2} = \sum_{k=-\infty}^{\infty} (1 + |k|)\bar{\hat{f}}_k \hat{g}_k \tag{6.2.37}$$

and norm

$$\|f\|_{1/2} = \sqrt{\langle f, f\rangle_{1/2}} \tag{6.2.38}$$

$H^{1/2}_{\mathbb{R}}$ will denote the set of real-valued functions in $H^{1/2}$.

Definition. For $f \in H^{1/2}_{\mathbb{R}}$,

$$I(f) = \sum_{k>0} \hat{f}_k e^{ik\theta} - \sum_{k<0} \hat{f}_k e^{ik\theta} \tag{6.2.39}$$

and

$$B(f) = \exp(I(f)) \tag{6.2.40}$$

Remark. $-iI(f)$ is the conjugate function to f. It will also enter in Section 6.4.

Proposition 6.2.5. *If $w \in L^2(\partial\mathbb{D}, \frac{d\theta}{2\pi})$ is positive with $\log w \in H^{1/2}$, then*

$$B(\log w) = \frac{D_w}{\bar{D}_w} \tag{6.2.41}$$

Remark. This, of course, says $b(w) = \overline{B(\log w)}$ and $c(w) = B(\log w)$ in terms of the functions of (6.2.27).

Proof. By definition of D and (1.3.18), if $\log w \equiv L$,

$$D(z) = \exp\biggl(\tfrac{1}{2}\,\hat{L}_0 + \sum_{k=1}^{\infty} \hat{L}_k z^k\biggr)$$

so

$$D(e^{i\theta}) = \exp\biggl(\tfrac{1}{2}\,\hat{L}_0 + \sum_{k=1}^{\infty} \hat{L}_k e^{ik\theta}\biggr)$$

with the sum intended in $L^2(\partial\mathbb{D}, \frac{d\theta}{2\pi})$ sense. Thus, since $L_{-k} = \bar{L}_k$,

$$\frac{D}{\bar{D}} = \exp\biggl(\sum_{k=1}^{\infty} \hat{L}_k e^{ik\theta} - \sum_{k=1}^{\infty} \hat{L}_{-k} e^{-ik\theta}\biggr) = B(\log w) \qquad \square$$

We are now ready for the critical continuity result that implements Step 1.

Proposition 6.2.6 (Deift-Killip). *The map B on $H^{1/2}_{\mathbb{R}}$ has range in $H^{1/2}$ and B is continuous in $H^{1/2}$ norm.*

Proof. Clearly, I, which multiplies Fourier coefficients by $+1$, -1, or 0, is a contraction from $H^{1/2}_{\mathbb{R}}$ to $H^{1/2}$. Moreover, $I(f)$ is pure imaginary since f real implies $\hat{f}_{-k} = \bar{\hat{f}}_k$. Thus, $B(f) \in L^\infty$ and, by the fact that $ix \to e^{ix}$ is Lipschitz from $i\mathbb{R}$ to $\mathbb{C}$, Proposition 6.1.11 implies that $B(f) \in H^{1/2}$.

To check continuity, suppose $f_n \to f$ in $H^{1/2}_{\mathbb{R}}$. Note first that $I(\cdot)$ is continuous in L^2, so since $\exp(\cdot)$ is uniformly Lipschitz on $i\mathbb{R}$, $\|B(f) - B(f_n)\| \to 0$, and so we need only focus on the piece of the norm given by Devinatz's formula. Write

$$\begin{aligned} F_n &= e^{I(f_n)} - e^{I(f)} \\ &= (e^{I(f_n - f)} - 1)e^{I(f)} \\ &= G_n H \end{aligned}$$

with $H = e^{I(f)}$ and $G_n = (e^{I(f_n - f)} - 1)$. Thus

$$\left|\frac{F_n(\theta) - F_n(\varphi)}{e^{i\theta} - e^{i\varphi}}\right| \le \left|\frac{H(\theta) - H(\varphi)}{e^{i\theta} - e^{i\varphi}}\right| |G_n(\varphi)| + \left|\frac{G_n(\theta) - G_n(\varphi)}{e^{i\theta} - e^{i\varphi}}\right|$$

since $|H(\theta)| = 1$. Since $|e^{ix} - e^{iy}| \le |x - y|$, the second term integrated over θ and φ is bounded by $\|f_n - f\|_{1/2}$ by Devinatz's formula (6.1.58). The first term is a product of a function in $L^2(\partial\mathbb{D}^2, \frac{d\theta\, d\varphi}{(2\pi)^2})$ and a sequence uniformly bounded in L^∞ and converging in L^2 and so, by the lemma below, it goes to zero in $L^2(\partial\mathbb{D}^2, \frac{d\theta\, d\varphi}{(2\pi)^2})$. Thus, by Devinatz's formula, $\|F_n\|_{1/2} \to 0$. $\square$

LEMMA 6.2.7. *If $h \in L^2(M, d\mu)$, $f_n \in L^\infty \cap L^2(M)$, and $\sup_n \|f_n\|_\infty < \infty$, $\|f_n - f\|_2 \to 0$, then $\|h(f_n - f)\|_2 \to 0$.*

PROOF. Write h as $h_1 + h_2$ with $h_1 \in L^\infty$. Then

$$\|h(f_n - f)\|_2 \le \|h_1\|_\infty \|f_n - f\|_2 + 2\|h_2\|_2 \sup_n \|f_n\|_\infty$$

For each h_1, the first term goes to zero so

$$\limsup \|h(f_n - f)\|_2 \le 2\|h_2\|_2 \sup_n \|f_n\|_\infty$$

Since we can take $\|h_2\|_2$ arbitrarily small, the lim sup is 0. $\square$

That completes Step 1. For Step 2, notice that

THEOREM 6.2.8. *If $f \in H^{1/2}$, then $H(f) \in \mathcal{I}_2$, the Hilbert-Schmidt operators, and*

$$\|H(f)\|_{\mathcal{I}_2} \le \|f\|_{1/2} \tag{6.2.42}$$

PROOF. $H(f)$ has in its matrix representation $\hat{f}_1$ once, $\hat{f}_2$ twice, ..., so

$$\|H(f)\|_{\mathcal{I}_2}^2 = \sum_{k=1}^\infty k|\hat{f}_k|^2 \le \|f\|_{1/2}^2 \tag{6.2.43}$$

$\square$

For Step 3:

PROPOSITION 6.2.9. *Let K be a trace class operator on a separable Hilbert space, $\mathcal{H}$, so that $(1 - K)$ is invertible. Let P be an orthonormal projection and $Q = 1 - P$. Then*

$$\det_{P\mathcal{H}}(P(1 - K)^{-1}P) = \frac{\det_{Q\mathcal{H}}(Q(1 - K)Q)}{\det(1 - K)} \tag{6.2.44}$$

PROOF. Pick a basis $\{\varphi_j\}_{j=1}^\infty$ with each $\varphi_j \in P$ or $\varphi_j \in Q$. Let R_n be the projection onto the span of $\{\varphi_j\}_{j=1}^n$. Since $R_n K R_n$ is finite-dimensional, Jacobi's identity on minors (see Proposition 3.1.6) says that

$$\det_{PR_n\mathcal{H}}(PR_n(1 - K)^{-1}R_nP) = \frac{\det_{QR_n\mathcal{H}}(QR_n(1 - K)R_nQ)}{\det_{R_n\mathcal{H}}(P_n(1 - K)R_n)} \tag{6.2.45}$$

By (1.4.66) and (1.4.69), if $A \in \mathcal{I}_1$, $\det_{R_n\mathcal{H}}(R_n(1 - A)R_n) \to \det(1 - A)$ since

$$\det_{R_n\mathcal{H}}(R_n(1 - A)R_n) = \det_{\mathcal{H}}(1 - R_nAR_n)$$

so (6.2.44) follows from (6.2.45) by taking limits. $\square$

PROPOSITION 6.2.10. *If* (6.2.29) *and* (6.2.32) *hold, then*

$$\det_{Q_n\mathcal{H}_+}(Q_n(T(b)T(c))^{-1}Q_n) = \frac{\det_{R_n\mathcal{H}_+}(1 - R_nH(b)H(\tilde{c})R_n)}{\det_{\mathcal{H}_+}(1 - H(b)H(\tilde{c}))} \tag{6.2.46}$$

Remark. Once we have (6.2.52) below, this result extends to cases when only (6.2.29) holds. For (6.2.52) and (6.2.53) below imply that for $f \in H^{1/2}$, $T(b)T(c)$ is invertible and the inverse is continuous in $H^{1/2}$, which implies that both sides of (6.2.46) are continuous.

PROOF. Let $K = H(b)H(\tilde{c})$ so, by (6.2.22),

$$1 - K = T(b)T(c)$$

By (6.2.32), $b = x_+x_-$ with $x_+ \in H^\infty$ and $x_- \in \overline{H^\infty}$ both invertible in H^∞ and $\overline{H^\infty}$, respectively. Thus, by Corollary 6.2.4, $T(b)T(c)$ is invertible, and so, (6.2.44) holds, which is just (6.2.46), □

That gives us Step 3. To get Step 4, note that

PROPOSITION 6.2.11. *If* (6.2.32) *holds, then*

$$\det(Q_n(T(b)T(c))^{-1}Q_n) = e^{-(n+1)\hat{L}_0}D_n\biggl(w\,\frac{d\theta}{2\pi}\biggr) \tag{6.2.47}$$

Remark. Again, once we have (6.2.52) below, continuity will imply this if (6.2.29) holds even if (6.2.32) does not.

PROOF. We have that

$$\begin{aligned}
(T(b)T(c))^{-1} &= T(c)^{-1}T(b)^{-1} \\
&= T(D^{-1})T(\bar{D})T(D)T(\bar{D}^{-1}) &\quad (6.2.48)\\
&= T(D^{-1})T(w)T(\bar{D}^{-1}) &\quad (6.2.49)
\end{aligned}$$

where (6.2.48) follows from (6.2.25) and (6.2.49) from Corollary 6.2.3 and $D \in H^\infty$ (when (6.2.32) holds).

Since $D^{-1} \in H^\infty$, $Q_nT(D^{-1})R_n = 0$, and similarly, $R_nT(\bar{D}^{-1})Q_n = 0$. Thus

$$Q_nT(D^{-1})T(w)T(\bar{D}^{-1})Q_n = Q_nT(D^{-1})Q_nQ_nT(w)Q_nP_nT(\bar{D}^{-1})Q_n$$

So, by (6.2.49),

$$\begin{aligned}
&\det(Q_n(T(b)T(c))^{-1}Q_n) \\
&\qquad = \det(Q_n(T(D^{-1})Q_n)\det(Q_nT(w)Q_n)\det(Q_nT(\bar{D}^{-1})Q_n)
\end{aligned} \tag{6.2.50}$$

By definition, $\det(Q_nT(w)Q_n) = D_n(w\frac{d\theta}{2\pi})$. Since $T(D^{-1})$ is upper triangular, $\det(Q_nT(D^{-1})Q_n) = [D^{-1}(0)]^{n+1} = \exp(-\frac{1}{2}(n+1)\hat{L}_0)$ by Szegő's theorem (Theorem 2.7.14). Similarly, $\det(Q_nT(\bar{D}^{-1})Q_n) = \exp(-\frac{1}{2}(n+1)\hat{L}_0)$. Thus (6.2.50) implies (6.2.47). □

To prove Step 5, we need a theorem, whose proof we defer:

THEOREM 6.2.12 (Helton-Howe Theorem). *Let A, B be bounded operators on a Hilbert space $\mathcal{H}$ so that $[A,B]$ is trace class. Then $e^Ae^Be^{-A}e^{-B} - 1$ is trace class and*

$$\det(e^Ae^Be^{-A}e^{-B}) = \exp(\mathrm{Tr}[A,B]) \tag{6.2.51}$$

Given this, we have

THEOREM 6.2.13 (Widom's Formula). *If* (6.2.29) *holds, then*

$$\det(1 - H(b)H(\tilde{c})) = \exp\biggl(-\sum_{j=1}^{\infty} j|\hat{L}_j|^2\biggr) \tag{6.2.52}$$

PROOF. Suppose first w obeys (6.2.29) and (6.2.32). By (6.2.22) and (6.2.48),

$$\begin{aligned} 1 - H(b)H(\tilde{c}) &= T(b)T(c) && (6.2.53)\\ &= T(\bar{D})T(D)^{-1}T(\bar{D})^{-1}T(D) && (6.2.54)\end{aligned}$$

Letting $L_+ = \frac{1}{2}\hat{L}_0 + \sum_{k=1}^{\infty} \hat{L}_k e^{ik\theta}$ and $L_- = \bar{L}_+$ so $D = \exp(L_+)$, and using Corollary 6.2.3,

$$T(D) = T(e^{L_+}) = e^{T(L_+)}$$

so (6.2.54) becomes

$$\begin{aligned} \det(1 - H(b)H(\tilde{c})) &= \det(e^{T(L_-)}e^{-T(L_+)}e^{-T(L_-)}e^{T(L_+)})\\ &= \exp(-\mathrm{Tr}[T(L_-), T(L_+)]) && (6.2.55)\end{aligned}$$

by the Helton-Howe formula if we prove that $[T(L_-), T(L_+)]$ is trace class.

By Corollary 6.2.3, $T(L_-)T(L_+) = T(L_+L_-)$, and then, by (6.2.22),

$$\begin{aligned} [T(L_-), T(L_+)] &= T(L_+L_-) - T(L_+)T(L_-)\\ &= H(L_+)H(\bar{L}_-)\\ &= H(L_+)H(L_+)^* && (6.2.56)\end{aligned}$$

By Proposition 6.2.6, $H(L_+)$ is Hilbert-Schmidt so, by (6.2.56), $[T(L_-), T(L_+)]$ is trace class and, by (6.2.43),

$$\begin{aligned} \mathrm{Tr}(H(L_+)H(L_-)^*) &= \|H(L_+)\|_2^2\\ &= \sum_{j=1}^{\infty} j|\hat{L}_j|^2 && (6.2.57)\end{aligned}$$

By (6.2.55), we have (6.2.52) if (6.2.32) holds. Since Theorem 6.2.2 says that $w \mapsto b, c$ is continuous in $H^{1/2}$, $w \mapsto H(b)H(\tilde{c})$ is continuous in trace norm, and so, $w \mapsto \det(1 - H(b)H(\tilde{c}))$ is continuous in $H^{1/2}$-norm. Obviously, the left side of (6.2.52) is continuous. Thus (6.2.52) holds in $H^{1/2}$. □

We are now ready for the main result of this section:

THEOREM 6.2.14 (Borodin-Okounkov Formula). *Let $d\mu$ be a nontrivial probability measure on $\partial\mathbb{D}$. Suppose*

$$\sum_{k=1}^{\infty} k|\hat{L}_k|^2 < \infty \tag{6.2.58}$$

Then

$$\frac{D_{n+1}(w\frac{d\theta}{2\pi})}{\exp((n+1)\hat{L}_0 + \sum_{k=1}^{\infty} k|\hat{L}_k|^2)} = \det{}_{R_n\mathcal{H}}(1 - R_nH(b)H(\tilde{c})R_n) \tag{6.2.59}$$

PROOF. If w also obeys $\sum|\hat{L}_k| < \infty$, then this follows (6.2.47) and (6.2.52). The left side is trivially continuous in w in $\|\cdot\|_{1/2}$-norm since $D_{n+1}(\cdot)$ only depends on finitely many Fourier coefficients. By Proposition 6.2.6 and Theorem 6.2.9, the right side is continuous. Since the w's obeying $\sum|\hat{L}_k| < \infty$ are dense in $H_{1/2}$, the result is proven. □

THEOREM 6.2.15 (Second Proof of Ibragimov's Theorem). *Let $d\mu = w\frac{d\theta}{2\pi}$ be a nontrivial probability measure. If w obeys $\sum_{k=1}^\infty k|\hat{L}_k|^2 < \infty$, then*

$$\lim_{n\to\infty} e^{-(n+1)\hat{L}_0} D_n(d\mu) = \exp\biggl(\sum_{k=1}^\infty k|\hat{L}_k|^2\biggr) \tag{6.2.60}$$

PROOF. This is immediate from (6.2.59) since $H(b)H(\tilde{c})$ is trace class, so $R_nH(b)H(\tilde{c})R_n \to 0$ in trace norm and thus, $\det(1 - R_nH(b)H(\tilde{c})R_n) \to 1$ by (1.4.66). □

The proof shows that

THEOREM 6.2.16. *Under the hypothesis of Theorem 6.2.15,*

$$|e^{-(n+1)\hat{L}_0 - \sum_1^\infty k|\hat{L}_k|^2} D_n(d\mu) - 1| \le \sum_{k=1}^\infty k|\hat{L}_{k+n}|^2 \exp\biggl(\sum_{k=1}^\infty k|\hat{L}_{k+n}|^2\biggr) \tag{6.2.61}$$

and

$$|e^{-(n+1)\hat{L}_0 - \sum_1^\infty k|\hat{L}_k|^2} D_n(d\mu) - 1 - A_n| \le \tfrac{1}{2}\, A_n^2 e^{A_n} \tag{6.2.62}$$

where $A_n = \sum_{k=1}^\infty k|\hat{L}_{k+n}|^2$.

Remark. This shows once again (see Theorem 2.1.3) that if w is C^∞ and $w > 0$, then the error in the strong Szegő asymptotics is $O(n^{-\ell})$ for all ℓ.

PROOF. By (1.4.64) and (1.4.63),

$$|\det(1 + A) - 1| \le e^{\|A\|_1} - 1 \le \|A\|_1\, e^{\|A\|_1} \tag{6.2.63}$$

and

$$\begin{aligned} |\det(1 + A) - 1 - \mathrm{Tr}(A)| &\le e^{\|A_1\|_1} - 1 - \|A_1\|_1 \\ &\le \tfrac{1}{2}\, \|A\|_1^2\, e^{\|A\|_1} \end{aligned} \tag{6.2.64}$$

where we used Taylor's theorem with remainder. If we note that $R_nH(b)H(\tilde{c})R_n \ge 0$ and $\mathrm{Tr}(R_nH(b)H(\tilde{c})R_n) = \|H(\tilde{c})R_n\|^2 = \sum_{k=1}^\infty k|\hat{L}_{k+n}|$ since $H(\tilde{c})R_n$ is a Hankel matrix based on the sequence $\hat{L}_{n+1}, \hat{L}_{n+2}, \dots$, then (6.2.61) is (6.2.63) and (6.2.62) is (6.2.64). □

Finally, we turn to a proof of the Helton-Howe theorem:

PROOF OF THEOREM 6.2.12. If $[A, B]$ is trace class, by a simple induction, $[A^n, B] = A^{n-1}[A, B] + [A^{n-1}, B]A$ is trace class and

$$\|[A^n, B]\|_1 \le n\|A\|^{n-1}\|[A, B]\|_1$$

By a second induction,

$$\|[A^n, B^m]\|_1 \le nm\|A\|^{n-1}\|B\|^{m-1}\|[A, B]\|_1$$

Thus, by summing Taylor series, $[e^A, e^B]$ is trace class and

$$\|[e^A, e^B]\|_1 \le \|[A, B]\|_1\, e^{\|A\|+\|B\|}$$

It follows that $e^Ae^Be^{-A}e^{-B} - 1 = [e^A, e^B]e^{-A}e^{-B}$ is trace class and

$$\|e^Ae^Be^{-A}e^{-B} - 1\|_1 \le \|[A, B]\|_1\, e^{2(\|A\|+\|B\|)}$$

Define

$$F(s) = e^{sA}e^{sB}e^{-sA}e^{-sB} \tag{6.2.65}$$

and

$$X(s) = 1 - F(s) \tag{6.2.66}$$

For s small, $X(s)$ is small in $\|\cdot\|_1$, so

$$\log(F(s)) = -\sum_{n=1}^{\infty} \frac{X(s)^n}{n} \tag{6.2.67}$$

converges in $\|\cdot\|_1$.

By Lidskii's theorem (see (1.4.72) and (1.4.73)),

$$\log(\det(F(s))) = \mathrm{Tr}(\log F(s)) \tag{6.2.68}$$

so, by (6.2.67),

$$\begin{aligned} \frac{d}{ds}\log(\det(F(s))) &= -\sum_{n=1}^{\infty} \mathrm{Tr}\biggl(\sum_{j=0}^{n-1} X(s)^j X'(s) X(s)^{n-1-j}\biggr)\Big/ n \\ &= \mathrm{Tr}(F'(s)F(s)^{-1}) \end{aligned} \tag{6.2.69}$$

Since $F(s)$ is clearly invertible for all $s \in \mathbb{C}$, $\det(F(s))$ is nonvanishing, so both sides of (6.2.69) are entire in s, and thus their equality holds for all s.

Next, note that

$$F'(s) = e^{sA}e^{sB}C(s)e^{-sA}e^{-sB} \tag{6.2.70}$$

where

$$C(s) = B + e^{-sB}Ae^{sB} - A - e^{-sA}Be^{sA} \tag{6.2.71}$$

If $Y(s) = e^{sR}Qe^{-sR}$, with Q, R bounded operators, then $Y(s)$ is an entire function of s and

$$Y^{(n)}(s)\Big|_{s=0} = (Ad_R)^n(Q) \tag{6.2.72}$$

where

$$Ad_R(Q) = [R, Q] \tag{6.2.73}$$

so

$$C(s) = \sum_{n=1}^{\infty} \frac{(-s)^n}{n!}\,[Ad_B^n(A) - Ad_A^n(B)] \tag{6.2.74}$$

Thus $C(s)$ is trace class, so by (1.4.60),

$$\begin{aligned} \mathrm{Tr}(F'(s)F(s)^{-1}) &= \mathrm{Tr}(e^{sA}e^{sB}C(s)e^{-sA}e^{-sB}e^{sB}e^{sA}e^{-sB}e^{-sA}) \\ &= \mathrm{Tr}(C(s)) \\ &= 2s[A, B] \end{aligned} \tag{6.2.75}$$

where we used (1.4.60) again to conclude $\mathrm{Tr}(Ad_B^n(A)) = 0$ if $n \geq 2$.

By (6.2.69) and (6.2.75), $\frac{d}{ds}\log(\det(F(s))) = 2s[A, B]$ so

$$\det(F(s)) = \exp(s^2[A, B]) \tag{6.2.76}$$

which is (6.2.51). $\square$

Remarks and Historical Notes. The Borodin-Okounkov formula was proven by them in [**135**]. They were motivated by a question of Deift and Its. Alternate proofs are due to Böttcher [**138, 139**] and to Basor-Widom [**90**]. Our proof follows that of Böttcher [**139**] with one important technical condition; his proof is only for $L^\infty \cap H^{1/2}$, not all of $H^{1/2}$ (see below).

While the paper of Borodin-Okounkov evoked great interest and has become a standard name, the result was proven already many years earlier by Geronimo-Case [**395**]. While Geronimo-Case do not mention Hankel operators by name and are not explicit about what conditions are needed, their kernel G (their eqn. (V.11)) is a product of Hankel operators. What we call b, they call S, the S-matrix in their way of discussing OPUC. As we will see in Section 10.7, $S = \bar{D}/D$ (see the last sentence in that section!), so their formula is equivalent to the Borodin-Okounkov formula in the form we write it. It is unfortunate that this aspect of [**395**] seems to have escaped the notice of the experts in the field.

Borodin-Okounkov have an explicit trace class integral operator on $L^2(\partial\mathbb{D})$ where we use $H(\tilde{b})H(c)$. That their operator can be written in this form is an important observation of Basor-Widom [**90**].

The use of Hankel operators to study Toeplitz operators goes back to Gohberg-Krein [**439**]. This point of view, exploiting Theorem 6.2.2 and an analog for T_n, was developed for the study of Toeplitz determinants by Widom [**1099**]. In particular, that paper has a variant of what we have called Widom's formula (6.2.52): In place of $\det(1 - H(\tilde{b})H(c)) = \det(T(b)T(c))$, he has $\det(T(w)T(w^{-1}))$, which is the same if w is nice enough. It should be emphasized that Widom [**1099**] was the first paper to use operator methods on the study of the strong Szegő theorem, and the extensive later works on that subject all begin from his work.

The Helton-Howe formula appeared first in Helton-Howe [**501**]. They remarked their result should be connected to the strong Szegő theorem and this motivated Widom. Recently, Ehrhardt [**329**] proved a generalization of the Helton-Howe formula that if $[A,B]$ is trace class, then

$$\det e^A e^B e^{-(A+B)} = \exp(\tfrac{1}{2}\operatorname{Tr}([A,B]))$$

It is easy to prove this formula using either the method of proof we give for the Helton-Howe formula (close to the proof in [**329**]) or the original proof using the Baker-Campbell-Hausdorff (BCH) formula. Since $e^A e^B e^{-A} e^{-B} = (e^A e^B e^{-(A+B)})(e^B e^A e^{-(A+B)})^{-1}$, Ehrhardt's formula implies the Helton-Howe formula.

Helton-Howe use the BCH formula which says $e^A e^B = e^C$ with $C = A + B + \frac{1}{2}[A,B] + \cdots$ where $\cdots$ involves higher-order commutators [**554**]. Rather than rely on this result, whose first-principle proof is lengthy, we present the derivative argument — new here — which is motivated by one proof of the BCH formula as discussed in Magnus [**730**], Feynman [**349**], Oteo-Ros [**834**], and Dragt-Finn [**306**]. This proof has some overlap with Ehrhardt [**329**].

Prior discussions of the Borodin-Okounkov formula made the stronger assumption that $\log w \in L^\infty \cap H^{1/2}$. This space is called the *Krein algebra* since Krein [**659**] discovered this space is an algebra. That it is indeed an algebra is a simple consequence of Devinatz's formula (6.1.58). It can also be proven (Krein's original proof) using the analog of (6.2.22),

$$H(fg) = T(f)H(g) + H(f)T(\bar{g})$$

The idea we use to avoid $\log w \in L^\infty$ that, by Devinatz's formula, $b, c \in H^{1/2}$, is due to Killip (unpublished). The continuity result and that it should allow using the Borodin-Okounkov formula to prove the full Ibragimov theorem is a subsequent observation of Deift (unpublished).

The converse of Theorem 6.2.1 for T is easy. If $V_n : \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ by $(V_n u)_j = u_{j-n}$ and $u_j = 0$ if $j \notin [-n, n]$, then

$$V_n c(a) u = t(a) V_n u$$

so if $\|t(a)u\| \leq D\|u\|$ for all u, then $\|c(a)u\| \leq D\|u\|$ for all u, and thus, $t(a)$ is bounded if and only if $c(a)$ is bounded and the norms are the same.

On the other hand, if $c(a)$ is bounded with $\|c(a)u\| \leq D\|u\|$, then first $c(a)\delta_0 \equiv a$, so $a \in \ell^2$. Thus $c(a)u = (f\hat{u})^\vee$ for f in ℓ^2. If $|\{\theta \mid |f| > D\}| > 0$, and if $\hat{u} = \chi_{\{\theta \mid |f|>0\}} f/|f|$, then $\|c(a)u\| > D\|u\|$. We conclude that if $\|c(a)u\| \leq D\|u\|$, then $c(a) = C(f)$ for $f \in L^\infty$ with $\|f\|_\infty \leq D$.

The modern theory of Hankel operators started with the following result of Nehari [**803**]:

THEOREM 6.2.17 (Nehari's Criterion). *Let $\{a_n\}_{n=1}^\infty$ be a sequence so that $h(a)$ is bounded from $\ell^2(\mathbb{Z}_+)$ to $\ell^2(\mathbb{Z}_+)$, that is,*

$$\|h(a)u\|_2 \leq D\|u\|_2 \tag{6.2.77}$$

Then there exists $f \in L^\infty(\partial\mathbb{D}, \frac{d\theta}{2\pi})$, $\|f\|_\infty = D$ so that (6.2.5) *holds for $j \geq 1$. Conversely, if $f \in L^\infty$ and a is given by* (6.2.5), *then $h(a)$ is bounded and* (6.2.8) *holds.*

PROOF. We proved the converse statement in (6.2.8), so it suffices to show (6.2.77) implies the existence of $f \in L^\infty$ with $H(f) = h(a)$ and with $\|f\|_\infty \leq D$.

Since $a_n = \langle \delta_n, h(a)\delta_1 \rangle$, $\|a\|_\infty \leq D$. For $0 < r < 1$, let $(a^{(r)})_n = a_n r^n$ so $a^{(r)} \in \ell^1$ and thus, there is $f^{(r)} \in L^\infty$ with (6.2.5) relating $f^{(r)}$ and $a^{(r)}$. Let $g \in H^1$ and write $g = k\ell$, with $k, \ell \in H^2$ and $\|k\|_2 = \|\ell\|_2 = \|g\|_1$. Define $\mathcal{L}^{(r)}$ on H^1 by

$$\begin{aligned} \mathcal{L}^{(r)}(g) &= \sum_{m\geq 0} a_m^{(r)} \hat{g}(m) \\ &= \sum_{m\geq 0} a_m^{(r)} \sum_{j=0}^m \hat{k}(j)\hat{\ell}(m-j) \\ &= \langle M_r \hat{k}, h(a) M_r \hat{\ell} \rangle \end{aligned}$$

where $M_r : \ell^2 \to \ell^2$ by $(M_r b)_n = r^n b_n$. Thus, $\mathcal{L}^{(r)}(g)$ is well-defined and obeys

$$\begin{aligned} |\mathcal{L}^{(r)}(g)| &\leq D\|M_r\hat{k}\|_{\ell^2}\|M_r\hat{\ell}\|_{\ell^2} \\ &\leq D\|\hat{k}\|_{\ell^2}\|\hat{\ell}\|_{\ell^2} \\ &= D\|k\|_2\|\ell\|_2 = D\|g\|_1 \end{aligned}$$

Thus if g is a polynomial and we define

$$\mathcal{L}(g) = \sum a_n \hat{g}(m)$$

then

$$|\mathcal{L}(g)| \leq D\|g\|_1 \tag{6.2.78}$$

By the Hahn-Banach theorem, $\mathcal{L}$ extends to a continuous function on L^1 with the same norm, that is, there is $f \in L^\infty$ with $\|f\|_\infty \le D$ and $\int f(\theta)e^{-ik\theta}\frac{d\theta}{2\pi} = a_k$. $\square$

Remark. We use the extra complication of $a^{(r)}$ because even if f is a polynomial, k and ℓ may not be.

Shortly after Nehari's paper, Hartman [**493**] proved a nice complement to Nehari's theorem, namely, $h(a)$ is compact if and only if there is a continuous function f with $\hat{f}_n = a_n$. For a comprehensive study of Hankel operators, see Peller [**863**].

For finite invertible matrices X, Y, we have $\det(XYX^{-1}Y^{-1}) = 1$ so the Helton-Howe theorem is a result specific to infinite dimensions (of course, related to the fact that $\mathrm{Tr}([A,B]) = 0$ for finite matrices). Related to this, as noted above, is Widom's formula that

$$\det(T(w)T(w^{-1})) = \exp\biggl(\sum_{k=1}^\infty k|\hat{L}_k|^2\biggr) \tag{6.2.79}$$

to be compared with the consequence of the strong Szegő theorem that

$$\lim_{n\to\infty} \det(T_n(w)T_n(w^{-1})) = \exp\biggl(2\sum_{k=1}^\infty k|\hat{L}_k|^2\biggr) \tag{6.2.80}$$

The extra 2 in (6.2.80) is at first surprising but, in essence, the strong Szegő term, $\exp(\sum_{k=1}^\infty k|\hat{L}_k|^2)$, is an edge effect. $T(w)$ is semi-infinite and has only one edge while, for any n, T_n has two edges.

6.3. Representations of $\mathbb{U}(n)$ and the Bump-Diaconis Proof

In this section and Section 6.5, we prove the strong Szegő theorem without the weakest possible hypotheses used in the last two sections and in Section 6.4. They are here because they provide interesting insights into where the terms come from. The starting point in this section is (1.5.89) and (1.5.88), which can be combined to say that if $d\mu = w(\theta)\frac{d\theta}{2\pi}$ (i.e., $d\mu_{\mathrm{s}} = 0$), then

$$D_n\biggl(w\,\frac{d\theta}{2\pi}\biggr) = \int_{\mathbb{U}(n+1)} e^{F_{n+1}(g)}\,dg \tag{6.3.1}$$

where dg is Haar measure on $\mathbb{U}(n+1)$, the group of $(n+1)\times(n+1)$ unitary matrices and

$$F_{n+1}(g) = \sum_{j=0}^n \log w(\theta_j(g)) \tag{6.3.2}$$

with $\{e^{i\theta_j(g)}\}$ the eigenvalues of g.

Let

$$L(\theta) \equiv \log w(\theta) = \sum_{k=-\infty}^\infty \hat{L}_k e^{ik\theta} \tag{6.3.3}$$

and we suppose that

$$\sum_{k=-\infty}^\infty |\hat{L}_k| < \infty \tag{6.3.4}$$

Since

$$\sum_{j=0}^n e^{ik\theta_j(g)} = \mathrm{Tr}(g^k) \tag{6.3.5}$$