EIGENVALUE ESTIMATES FOR NON-NORMAL MATRICES AND THE ZEROS OF RANDOM ORTHOGONAL POLYNOMIALS ON THE UNIT CIRCLE

E. B. DAVIES[1] AND BARRY SIMON[2]

Abstract. We prove that for any $n \times n$ matrix, A, and z with $|z| \geq \|A\|$, we have that $\|(z-A)^{-1}\| \leq \cot(\frac{\pi}{4n})\mathrm{dist}(z, \mathrm{spec}(A))^{-1}$. We apply this result to the study of random orthogonal polynomials on the unit circle.

1. Introduction

This paper concerns a sharp bound on the approximation of eigenvalues of general non-normal matrices that we found in a study of the zeros of orthogonal polynomials. We begin with a brief discussion of the motivating problem, which we return to in Section 7.

Given a probability measure $d\mu$ on $\mathbb{C}$ with

$$\int |z|^n \, d\mu(z) < \infty \tag{1.1}$$

we define the monic orthogonal polynomials, $\Phi_n(z)$, by

$$\Phi_n(z) = z^n + \text{lower order} \tag{1.2}$$

$$\int \overline{z^j}\, \Phi_n(z)\, d\mu(z) = 0 \qquad j = 0, 1, \dots, n-1 \tag{1.3}$$

If

$$\begin{aligned} P_n = {} & \text{orthogonal projection in } L^2(\mathbb{C}, d\mu) \\ & \qquad \text{onto polynomials of degree } n-1 \text{ or less} \end{aligned} \tag{1.4}$$

then

$$\Phi_n = (1 - P_n) z^n \tag{1.5}$$

Date: February 28, 2006.

[1] Department of Mathematics, King's College London, Strand, London WC2R 2LS, United Kingdom. E-mail: E.Brian.Davies@kcl.ac.uk. Supported in part by EPSRC grant GR/R81756.

[2] Mathematics 253-37, California Institute of Technology, Pasadena, CA 91125. E-mail: bsimon@caltech.edu. Supported in part by NSF grant DMS-0140592.

A key role is played by the operator

$$A_n = P_n M_z P_n \restriction \mathrm{Ran}(P_n) \tag{1.6}$$

where M_z is the operator of multiplication by z and A_n is an operator on the n-dimensional space $\mathrm{Ran}(P_n)$.

If z_0 is a zero of $\Phi_n(z)$ of order k, then $f_{z_0} \equiv (z-z_0)^{-k}\Phi_n(z)$ is in $\mathrm{Ran}(P_n)$ and

$$(A_n - z_0)^k f_{z_0} = 0 \qquad (A-z_0)^{k-1} f_{z_0} \neq 0 \tag{1.7}$$

which implies

$$\Phi_n(z) = \det(z - A_n) \tag{1.8}$$

Also, $\Phi_n(z)$ is the minimal polynomial for A_n.

In the study of orthogonal polynomials on the real line (OPRL), a key role is played by the fact that for any $y \in \mathrm{Ran}(P_n)$ with $\|y\|_{L^2} = 1$,

$$\mathrm{dist}(z_0, \{\text{zeros of } \Phi_n\}) \leq \|(A_n - z_0)y\| \qquad \text{(OPRL case)} \tag{1.9}$$

This holds because, in the OPRL case, A_n is self-adjoint. Indeed, for any normal operator, B, (throughout $\|\cdot\|$ is a Hilbert space norm; for $n \times n$ matrices, the usual matrix norm induced by the Euclidean inner product)

$$\mathrm{dist}(z_0, \mathrm{spec}(B)) = \|(B - z_0)^{-1}\|^{-1} \tag{1.10}$$

and, of course, for any invertible operator C,

$$\inf\{\|Cy\| \mid \|y\| = 1\} = \|C^{-1}\|^{-1} \tag{1.11}$$

We were motivated by seeking a replacement of (1.9) in a case where A_n is non-normal. Indeed, we had a specific situation of orthogonal polynomials on the unit circle (OPUC; see [17, 18]) where one has a sequence $z_n \in \partial\mathbb{D} = \{z \mid |z| = 1\}$ and corresponding unit trial vectors, y_n, so that

$$\|(A_n - z_n)y_n\| \leq C_1 e^{-C_2 n} \tag{1.12}$$

for all n with $C_2 > 0$. We would like to conclude that $\Phi_n(z)$ has zeros near z_n.

It is certainly not sufficient that $\|(A_n - z_n)y_n\| \to 0$. For the case $d\mu(z) = d\theta/2\pi$ has $\Phi_n(z) = \mathrm{dist}(1, \mathrm{spec}(A_n)) = 1$, but if $y_n = (1 + z + \cdots + z^{n-1})/\sqrt{n}$, then $\|(A_n - 1)y_n\| = \|P_n(z-1)y_n\| = n^{-1/2}\|P_n(z^n - 1)\| = n^{-1/2}\|1\| = n^{-1/2}$. As we will see later, by a clever choice of y_n, one can even get trial vectors with $\|(A_n - 1)y_n\| = O(n^{-1})$.

Of course, by (1.11), we are really seeking some kind of bound relating $\|(A_n - z_n)^{-1}\|$ to $\mathrm{dist}(z_n, \mathrm{spec}(A_n))$. At first sight, the prognosis for this does not seem hopeful. The $n \times n$ matrix,

$$N_n = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix} \tag{1.13}$$

has

$$\|(z - N_n)^{-1}\| \geq |z|^{-n} \tag{1.14}$$

since $(z - N_n)^{-1} = \sum_{j=0}^{n-1} z^{-j-1}(N_n)^j$ has z^{-n} in the $1, n$ position. Thus, as is well known, $\|(A_n - z)^{-1}\|$ for general $n \times n$ matrices A_n and general z cannot be bounded by better than $\operatorname{dist}(z, \operatorname{spec}(A_n))^{-n}$. Indeed, the existence of such bounds by Henrici [4] is part of an extensive literature on general variational bounds on eigenvalues. Translated to a variational bound, this would give $\operatorname{dist}(z_n, \{\text{zeros of } \Phi_n\}) \leq C\|(A_n - z_n)y\|^{1/n}$, which would not give anything useful from (1.12).

We note that as $n \to \infty$, there can be difficulties even if z_0 stays away from $\operatorname{spec}(A_n)$. For, by (1.14),

$$\|(1 - 2N_n)^{-1}\| \geq 2^{n-1} \tag{1.15}$$

diverges as $n \to \infty$ even though $\|2N_n\|$ is bounded in n.

Despite these initial negative indications, we have found a linear variational principle that lets us get information from (1.12). The key realization is that z_n and $\|A_n\|$ are not general. Indeed,

$$|z_n| = \|A_n\| = 1 \tag{1.16}$$

It is not a new result that a linear bound holds in the generality we discuss. In [11], Nikolski presents a general method for estimating norms of inverses in terms of minimal polynomials (see the proof of Lemma 3.2 of [11]) that is related to our argument in Subsection 6A. His ideas yield a linear bound but not with the optimal constant we find.

Our main theorem is

Theorem 1. *Let $\mathcal{M}_n$ be the set of pairs (A, z) where A is an $n \times n$ matrix, $z \in \mathbb{C}$ with*

$$|z| \geq \|A\| \tag{1.17}$$

and

$$z \notin \operatorname{spec}(A) \tag{1.18}$$

Then

$$c(n) \equiv \sup_{\mathcal{M}_n} \operatorname{dist}(z, \operatorname{spec}(A))\|(A - z)^{-1}\| = \cot\left(\frac{\pi}{4n}\right) \tag{1.19}$$

Of course, the remarkable fact, given (1.14), is that $c(n) < \infty$ when we only use the first power of $\operatorname{dist}(z, \operatorname{spec}(A))$. It implies that so long as (1.17) holds,

$$\operatorname{dist}(z, \operatorname{spec}(A)) \le c(n) \|(A - z)y\| \tag{1.20}$$

for any unit vector y. For this to be useful in the context of (1.12), we need only mild growth conditions on $c(n)$; see (1.21) below.

As an amusing aside, we note that

$$\begin{aligned} c(1) &= 1 = 0 + \sqrt{1} \\ c(2) &= 1 + \sqrt{2} \\ c(3) &= 2 + \sqrt{3} \end{aligned}$$

but the obvious extrapolation from this fails. Instead, because of properties of $\cot(x)$,

$$c(n) \le \frac{4}{\pi} n \tag{1.21}$$

$$\frac{c(n)}{n} \text{ is monotone increasing to } \frac{4}{\pi}$$

so, in fact, for $n \ge 3$,

$$\frac{2 + \sqrt{3}}{3} \le \frac{c(n)}{n} \le \frac{4}{\pi}$$

a spread of 2.3%.

We note that, by replacing A by A/z and z by 1, it suffices to prove

$$\sup_{\|A\| < 1} \operatorname{dist}(1, \operatorname{spec}(A)) \|(1 - A)^{-1}\| = \cot\left(\frac{\pi}{4n}\right) \tag{1.22}$$

and it is this that we will establish by proving three statements. We will use the special $n \times n$ matrix

$$M_n = \begin{pmatrix} 1 & 2 & \dots & 2 \\ 0 & 1 & \dots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{1.23}$$

given by

$$(M_n)_{k\ell} = \begin{cases} 2 & \text{if } k < \ell \\ 1 & \text{if } k = \ell \\ 0 & \text{if } k > \ell \end{cases}$$

Our three sub-results are

Theorem 2. $\|M_n\| = \cot(\pi/4n)$

Theorem 3. *For each $0 < a < 1$, there exist $n \times n$ matrices $A_n(a)$ with*

$$\|A_n(a)\| \leq 1 \qquad \operatorname{spec}(A_n) = \{a\} \tag{1.24}$$

and

$$\lim_{a\uparrow 1} (1-a)(1-A_n(a))^{-1} = M_n \tag{1.25}$$

Theorem 4. *Let A be an upper triangular matrix with $\|A\| \leq 1$ and $1 \notin \operatorname{spec}(A)$. Then*

$$\operatorname{dist}(1, \operatorname{spec}(A)) |(1-A)^{-1}_{k\ell}| \leq \begin{cases} 2 & \text{if } k < \ell \\ 1 & \text{if } k = \ell \\ 0 & \text{if } k > \ell \end{cases} \tag{1.26}$$

Proof that Theorems 2–4 $\Rightarrow$ Theorem 1. Any matrix has an orthonormal basis in which it is upper triangular: One constructs such a Schur basis by applying Gram-Schmidt to any algebraic basis in which A has Jordan normal form. In such a basis, (1.26) says that

$$\operatorname{dist}(1, \operatorname{spec}(A)) \|(1-A)^{-1} y\| \leq \|M_n y\| \leq \|M_n\| \, \|y\|$$

so Theorem 2 implies LHS of (1.22) $\leq \cot(\pi/4n)$.

On the other hand, using $A_n(a)$ in $\operatorname{dist}(1, \operatorname{spec}(A)) \|(1-A)^{-1}\|$ implies LHS of (1.22) $\geq \cot(\pi/4n)$. We thus have (1.22) and, as noted, this implies (1.19). ☐

To place Theorem 1 in context, we note that if $|z| > \|A\|$,

$$\|(z-A)^{-1}\| \leq \sum_{j=0}^{\infty} |z|^{-j-1} \|A\|^j = (|z| - \|A\|)^{-1} \tag{1.27}$$

So (1.19) provides a borderline between the dimension-independent bound (1.27) for $|z| > \|A\|$ and the exponential growth that may happen if $|z| < \|A\|$, essentially the phenomenon of pseudospectra which is well documented in [24]; see also [15].

The structure of this paper is as follows. In Section 2, we will prove Theorem 4, the most significant result in this paper since it implies $c(n) < \infty$ and, indeed, with no effort that $c(n) \leq 2n$. Our initial proofs of $c(n) < \infty$ were more involved — the fact that our final proof is quite simple should not obscure the fact that $c(n) < \infty$ is a result we find both surprising and deep.

In Section 3, we use upper triangular Toeplitz matrices to construct $A_n(a)$ and prove Theorem 3. Sections 4 and 5 prove Theorem 2; indeed, we also find that if

$$(Q_n(a))_{k\ell} = \begin{cases} 1 & \text{if } k < \ell \\ a & \text{if } k = \ell \\ 0 & \text{if } k > \ell \end{cases} \tag{1.28}$$

then

$$\|Q_n(1)\| = \frac{1}{2\sin(\frac{\pi}{4n+2})} \tag{1.29}$$

which means we can compute $\|Q_n(a)\|$ for $a = 0, \frac{1}{2}, 1$. While the calculation of $\|M_n\|$ and $\|Q_n(1)\|$ is based on explicit formulae for all the eigenvalues and eigenvectors of certain associated operators, we could just pull them out of a hat. Instead, in Section 4, we discuss the motivation that led to our guess of eigenvectors, and in Section 5 explicitly prove Theorem 2.

Section 6 contains a number of remarks and extensions concerning Theorem 1, most importantly to numerical range concerns. Section 7 contains the application to random OPUC.

Acknowledgments. This work was done while B. Simon was a visitor at King's College London. He would like to thank A. N. Pressley and E. B. Davies for the hospitality of King's College, and the London Mathematical Society for partial support. The calculations of M. Stoiciu [20, 21] were an inspiration for our pursuing the estimate we found. We appreciate useful correspondence/discussions with M. Haase, N. Higham, R. Nagel, N. K. Nikolski, V. Totik, and L. N. Trefethen.

2. The Key Bound

Our goal in this section is to prove Theorem 4. A is an upper triangular $n \times n$ matrix. Let $\lambda_1, \dots, \lambda_n$ be its diagonal elements. Since

$$\det(z - A) = \prod_{j=1}^{n} (z - \lambda_j) \tag{2.1}$$

the λ_j's are the eigenvalues of A counting algebraic multiplicity. In particular,

$$\sup_j |1 - \lambda_j|^{-1} = \operatorname{dist}(1, \operatorname{spec}(A))^{-1} \tag{2.2}$$

Define

$$C = (1 - A)^{-1} + (1 - A^*)^{-1} - 1 \tag{2.3}$$

Proposition 2.1. *Suppose* $\|A\| \leq 1$. *Then*

(a)

$$\begin{aligned} C_{jj} &= |1-\lambda_j|^{-2}(1-|\lambda_j|^2) \\ &\leq 2|1-\lambda_j|^{-1} \end{aligned} \tag{2.4}$$

(b)

$$C \geq 0$$

(c)

$$|C_{jk}| \leq |C_{jj}|^{1/2}|C_{kk}|^{1/2} \tag{2.5}$$

(d) *If* $j<k$, *then* $(1-A)^{-1}_{jk} = C_{jk}$.

Proof. (a) Since A is upper triangular,

$$[(1-A)^{-1}]_{jj} = (1-\lambda_j)^{-1} \tag{2.6}$$

so (2.4) comes from

$$(1-\lambda_j)^{-1} + (1-\bar{\lambda}_j)^{-1} - 1 = |1-\lambda_j|^{-2}(1-|\lambda_j|^2) \tag{2.7}$$

and the fact that for $|\lambda| \leq 1$,

$$\begin{aligned} |1-\lambda|^{-1}(1-|\lambda|^2) &= (1+|\lambda|)(1-|\lambda|)(|1-\lambda|^{-1}) \\ &\leq 2 \end{aligned}$$

since $1-|\lambda| \leq |1-\lambda|$.

(b) The operator analog of (2.7) is the direct computation

$$C = [(1-A)^{-1}]^*(1-A^*A)(1-A)^{-1} \geq 0 \tag{2.8}$$

since $\|A\| \leq 1$ implies $A^*A \leq 1$.

(c) This is true for any positive definite matrix.

(d) $(1-A^*)^{-1}$ is lower triangular and 1 is diagonal. ☐

Proof of Theorem 4. $(1-A)^{-1}$ is upper triangular so $[(1-A)^{-1}]_{k\ell} = 0$ if $k>\ell$. By (2.6) and (2.2),

$$|[(1-A)^{-1}]_{kk}| = |1-\lambda_k|^{-1} \leq \operatorname{dist}(1, \operatorname{spec}(A))^{-1} \tag{2.9}$$

By (a), (c), (d) of the proposition, if $k<\ell$,

$$\begin{aligned} |[(1-A)^{-1}]_{k\ell}| &\leq [|1-\lambda_k|^{-2}|1-\lambda_\ell|^{-2}(1-|\lambda_k|^2)(1-|\lambda_\ell|^2)]^{1/2} \\ &\leq 2[|1-\lambda_k|^{-1}|1-\lambda_\ell|^{-1}]^{1/2} \\ &\leq 2[\operatorname{dist}(1, \operatorname{spec}(A))]^{-1} \end{aligned}$$

by (2.2). ☐

3. Upper Triangular Toeplitz Matrices

A Toeplitz matrix [1] is one that is constant along diagonals, that is, A_{jk} is a function of $j-k$. An $n \times n$ upper triangular Toeplitz matrix (UTTM) is thus of the form

$$\begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_{n-1} \\ 0 & a_0 & a_1 & \dots & a_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_0 \end{pmatrix} \tag{3.1}$$

These concern us because M_n is of this form and because the operators, $A_n(a)$, of Theorem 3 will be of this form. In this section, after recalling the basics of UTTM, we will prove Theorem 3. Then we will state some results, essentially due to Schur [16], on the norms of UTTM that we will need in Section 5 in one calculation of the norm of M_n.

Given any function, f, which is analytic near zero, we write $T_n(f)$ for the matrix in (3.1) if

$$f(z) = a_0 + a_1 z + \dots + a_{n-1} z^{n-1} + O(z^n) \tag{3.2}$$

f is called a symbol for $T_n(f)$.

We note that

$$T_n(fg) = T_n(f) T_n(g) \tag{3.3}$$

This can be seen by multiplying matrices and Taylor series or by manipulating projections on ℓ^2 (see, e.g., Corollary 6.2.3 of [17]).

In addition, if f is analytic in $\{z \mid |z| < 1\}$, then

$$\|T_n(f)\| \le \sup_{|z|<1} |f(z)| \tag{3.4}$$

To see this well-known fact, associate an analytic function

$$v(z) = v_0 + v_1 z + \cdots \tag{3.5}$$

to the vector $\varphi_n(v) \in \mathbb{C}^n$ by

$$\varphi_n(v) = (v_{n-1}, v_{n-2}, \dots, v_0)^T \tag{3.6}$$

and note that with $\|\cdot\|_2$, the H^2 norm,

$$\|\varphi_n(v)\| = \inf\{\|v\|_2 \mid \varphi_n = \varphi_n(v)\} \tag{3.7}$$

$$T_n(f)\varphi_n(v) = \varphi_n(fv) \tag{3.8}$$

and

$$\|fv\|_2 \le \|f\|_\infty \|v\|_2 \tag{3.9}$$

If N_n is given by (1.13), then $T_n(f) = f(N_n)$, so an alternate proof of (3.4) may be based on von Neumann's theorem; see Subsection 6E.

Proof of Theorem 3. For a with $0 < a < 1$, define

$$f_a(z) = \frac{z+a}{1+az} \tag{3.10}$$

and define

$$A_n(a) = T_n(f_a) \tag{3.11}$$

Then $f_a(e^{i\theta}) = e^{i\theta}\,\overline{(1+ae^{i\theta})}/(1+ae^{i\theta})$ has $|f_a(e^{i\theta})| = 1$, so $\sup_{|z|<1}|f_a(z)| = 1$ and thus, by (3.4),

$$\|A_n(a)\| \leq 1 \tag{3.12}$$

By (3.1),

$$\operatorname{spec}(A_n(a)) = \{f_a(0)\} = \{a\} \tag{3.13}$$

By (3.5),

$$(1-A_n(a))^{-1} = T_n((1-f_a(z))^{-1}) \tag{3.14}$$

Now

$$(1-a)(1-f_a(z))^{-1} = \frac{z+a}{1-z} \tag{3.15}$$

so

$$\lim_{a\uparrow 1}(1-a)(1-f_a(z))^{-1} = \frac{1+z}{1-z} \tag{3.16}$$

Thus,

$$\lim_{a\uparrow 1}(1-a)(1-A_n(a))^{-1} = T_n\left(\frac{1+z}{1-z}\right) = M_n \tag{3.17}$$

since $(1+z)/(1-z) = 1+2z+2z^2+\cdots$. $\square$

We now want to refine (3.4) to get equality for a suitable f. A key role is played by

Lemma 3.1. *Let $\alpha \in \mathbb{D}$ and A an operator with $\overline{\alpha}^{\,-1} \notin \operatorname{spec}(A)$. Define*

$$B = (A-\alpha)(1-\overline{\alpha}A)^{-1} \tag{3.18}$$

Then

$$(1) \qquad \|B\| \leq 1 \Leftrightarrow \|A\| \leq 1 \tag{3.19}$$

$$(2) \qquad \|B\| = 1 \Leftrightarrow \|A\| = 1 \tag{3.20}$$

Proof. By a direct calculation,

$$1 - B^*B = (1-\alpha A^*)^{-1}[(1-|\alpha|^2)(1-A^*A)](1-\overline{\alpha}A)^{-1} \tag{3.21}$$

(3.19) follows since $1 - B^*B \geq 0 \Leftrightarrow 1 - A^*A \geq 0$, and (3.20) follows since (3.21) implies

$$\inf_{\|\varphi\|=1}(\varphi,(1-B^*B)\varphi) = 0 \Leftrightarrow \inf_{\|\varphi\|=1}(\varphi,(1-A^*A)\varphi) = 0 \qquad \square$$

Remark. This lemma is further discussed in Subsection 6E.

Theorem 3.2. *If A is an $n\times n$ UTTM with $\|A\|\leq 1$, then there exists an analytic function, f, on $\mathbb{D}$ such that*

$$\sup_{|z|<1} |f(z)| \leq 1 \tag{3.22}$$

and

$$A = T_n(f) \tag{3.23}$$

Proof. The proof is by induction on n. If $n=1$, $\|A\|\leq 1$ means $|a_0|\leq 1$ and we can take $f(z)\equiv a_0$. For general n, $\|A\|\leq 1$ means $|a_0|\leq 1$. If $|a_0|=1$, then $A=a_0\mathbf{1}$ and we can take $f(z)\equiv a_0$. If $a_0<1$, define B by (3.18) with $\alpha=a_0$. B is a UTTM with zero diagonal terms, so

$$B = \begin{pmatrix} 0 & & \tilde{B} \\ & \ddots & \\ 0 & & 0 \end{pmatrix} \tag{3.24}$$

where $\|\tilde{B}\| = \|B\| \leq 1$ by the lemma.

By the induction hypothesis, $\tilde{B} = T_{n-1}(g)$ where

$$\sup_{|z|<1} |g(z)| \leq 1 \tag{3.25}$$

Then (3.23) holds with

$$f = \frac{a_0 + zg}{1+\overline{a}_0 zg} \tag{3.26}$$

(3.25) and (3.26) imply (3.22). $\square$

Remarks. 1. By iterating $f\to g$, we see that one constructs f via the Schur algorithm; see Section 1.3 of [17].

2. Combining this and (3.4), one obtains Schur's celebrated result that $a_0 + a_1 z + \cdots + a_{n-1}z^{n-1}$ is the start of the Taylor series of a Schur function if and only if the matrix A of (3.1) obeys $A^*A\leq 1$. This result is intimately connected to Nehari's theorem on the norm of Hankel operators [8, 13]; see Partington [12].

3. This is classical; see [1, 10, 13].

To state the last result of this section, we need a definition:

Definition. A *Blaschke factor* is a function on $\mathbb{D}$ of the form

$$f(z,w) = \frac{z-w}{1-\overline{w}z} \tag{3.27}$$

where $w\in\mathbb{D}$. A (finite) *Blaschke product* is a function of the form

$$f(z) = \omega \prod_{j=1}^{k} f(z, w_k) \tag{3.28}$$

where $\omega \in \partial\mathbb{D}$. k is called the *order* of f. We allow $k = 0$, in which case $f(z)$ is a constant value in $\partial\mathbb{D}$.

Theorem 3.3. *An $n \times n$ UTTM, A, has $\|A\| = c$ if and only if $A = T_n(f)$ for an f so that $c^{-1}f$ is a Blaschke product of order $k \leq n-1$.*

Proof. (See as alternates: [10, 13].) Without loss, we can take $c = 1$. The proof is by induction on n. If $n = 1$, k must be 0, and the theorem says $|a_0| = 1$ if and only if $f(0) = \omega \in \partial\mathbb{D}$, which is true.

It is not hard to see that if f and f_1 are related by

$$f_1(z) = z^{-1}\, \frac{f(z) - f(0)}{1 - \overline{f(0)}\, f(z)}$$

then f is a Blaschke product of order $k \geq 1$ if and only if f_1 is a Blaschke product of order $k-1$.

Given A a UTTM with $\|A\| \leq 1$, $|a_0| = 1$ if and only if $A = T_n(a_0)$, that is, A is given by a Blaschke product of order 0. If $|a_0| < 1$, we define B by (3.18). $\|B\| = 1$ if and only if $\|A\| = 1$. $\tilde{B}$ given by (3.25) is related to A by $A = T_n(f)$ if and only if $\tilde{B} = T_{n-1}(f_1)$. Thus, by induction, $\|A\| = 1$ if and only if f is a Blaschke product of order $k \leq n-1$. □

4. Inverse of Differential/Difference Operators

In this section and the next, we will find explicit formulae for the norms of M_n and $Q_n \equiv Q_n(1)$ given by (1.28). Indeed, we will find all the eigenvalues and eigenvectors for $|M_n|$ and $|Q_n|$ where $|A| = \sqrt{A^*A}$. A key to our finding this was understanding a kind of continuum limit of M_n: Let K be the Volterra-type operator on $\mathcal{H} = L^2([0,1], dx)$ with integral kernel

$$K(x,y) = \begin{cases} 1 & 0 \leq x \leq y \leq 1 \\ 0 & 0 \leq y < x < 1 \end{cases}$$

In some formal sense, K is a limit of either M_n or Q_n, but in a precise sense, M_n is a restriction of K:

Proposition 4.1. *Let π_n be the projection of $\mathcal{H}$ onto the space of functions constant on each interval $[\frac{j}{n}, \frac{j+1}{n})$, $j = 0, 1, \dots, n-1$. Then*

$$\pi_n K \pi_n \tag{4.1}$$

is unitarily equivalent to $\frac{1}{2} M_n / n$. In particular,

$$\|M_n\| \leq 2n\|K\| \tag{4.2}$$

$$\lim_{n\to\infty} \frac{\|M_n\|}{n} = 2\|K\| \tag{4.3}$$

Proof. Let $\{f_j^{(n)}\}_{j=0}^{n-1}$ be the functions

$$f_j^{(n)}(x) = \begin{cases} \sqrt{n} & \frac{j}{n} \leq x < \frac{j+1}{n} \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

which form an orthonormal basis for $\mathrm{Ran}(\pi_n)$. Since

$$n\langle f_j^{(n)}, Kf_k^{(n)}\rangle = \tfrac{1}{2}\,(M_n)_{jk} \tag{4.5}$$

we have the claimed unitary equivalence. (4.2) is immediate from $\|\pi_n K \pi_n\| \leq \|K\|$. (4.3) follows if we note $\text{s-lim}_{n\to\infty} \pi_n = 1$, so $\lim \|\pi_n K \pi_n\| = \|K\|$. ☐

Notice that

$$(Kf)(x) = \int_x^1 f(y)\, dy \tag{4.6}$$

so

$$\frac{d}{dx}\,(Kf) = f \qquad Kf(1) = 0 \tag{4.7}$$

and K is an inverse of a derivative. That means K^*K will be the inverse of a second-order operator. Indeed,

$$\begin{aligned}(K^*K)(x,y) &= \int_0^1 \overline{K(z,x)}\, K(z,y)\, dz \\ &= \int_0^{\min(x,y)} dz \\ &= \min(x,y) \end{aligned} \tag{4.8}$$

which, as is well known, is the integral kernel of the inverse of $-\frac{d^2}{dx^2}$ with $u(0) = 0$, $u'(1) = 1$ boundary conditions.

We can therefore write down a complete orthonormal basis of eigenfunctions for K^*K:

$$\varphi_n(x) = \sin(\tfrac{1}{2}\,(2n-1)\pi x) \qquad n = 1, 2, \dots \tag{4.9}$$

$$(K^*K)\varphi_n = \frac{4}{(2n-1)^2\pi^2} \tag{4.10}$$

so

$$\|K\| = \|K^*K\|^{1/2} = \frac{2}{\pi} \tag{4.11}$$

By (4.2), (4.3), we have

Corollary 4.2.

$$\|M_n\| \leq \frac{4n}{\pi} \tag{4.12}$$

$$\lim_{n\to\infty} \frac{\|M_n\|}{n} = \frac{4}{\pi} \tag{4.13}$$

Of course, we will see this when we have proven Theorem 2, but it is interesting to have it now.

While M_n is related to differential operators via (4.5), we can compute the norm of Q_n by realizing it as the inverse of a difference operator. Specifically, let N_n be given by (1.13). Then

$$(1 - N_n)^{-1} = 1 + N_n + N_n^2 + \cdots + N_n^{n-1} = Q_n \tag{4.14}$$

Theorem 4.3. *Let*

$$D_n = (1 - N_n)(1 - N_n)^* \tag{4.15}$$

Then D_n has a complete set of eigenvectors:

$$v_j^{(\ell)} = \sin\left(\frac{\pi(2\ell+1)j}{2n+1}\right) \qquad j = 1, \dots, n;\ \ell = 0, \dots, n-1 \tag{4.16}$$

$$D_n v^{(\ell)} = 4\sin^2\left(\frac{\pi(2\ell+1)}{2(2n+1)}\right) v^{(\ell)} \tag{4.17}$$

$$\begin{aligned} \|Q_n\| &= (\min \textit{ eigenvalue of } D_n)^{-1/2} \\ &= \left[2\sin\left(\frac{\pi}{4n+2}\right)\right]^{-1} \end{aligned} \tag{4.18}$$

Proof. By a direct calculation,

$$D_n = \begin{pmatrix} 2 & -1 & 0 & & & & \\ -1 & 2 & -1 & & & & \\ 0 & -1 & 2 & & & & \\ & & & \ddots & & & \\ & & & & 2 & -1 & 0 \\ & & & & -1 & 2 & -1 \\ & & & & 0 & -1 & 1 \end{pmatrix} \tag{4.19}$$

is a discrete Laplacian with Dirichlet boundary condition at 0 and Neumann at n. Since

$$-\sin(q(j+1)) + 2\sin(qj) - \sin(q(j-1)) = 4\sin^2\left(\frac{q}{2}\right)\sin(qj)$$

(4.16)/(4.17) hold so long as q is such that $\sin(q(n+1)) = \sin(qn)$, that is,

$$\tfrac{1}{2}\left[q(n+1) + qn\right] = (\ell + \tfrac{1}{2})\pi$$

or $q = (2\ell+1)\pi/(2n+1)$. ∎

Remark. For OPUC with $d\mu = d\theta/2\pi$, in the basis $1, z, \dots, z^{n-1}$, A_n is given by the matrix, N_n, of (1.13), and so $\|(1-N_n)^{-1}\| = \|Q_n\| \sim 2n/\pi$. Thus, there are unit vectors, y_n, in this case with $\|(1-A_n)y_n\| \sim \pi/2n$.

5. The Norm of M_n

In this section, we will give two distinct but related proofs of Theorem 2. Both depend on a generating function relation:

Theorem 5.1. *For $\theta \in (0,\pi)$ and $z \in \mathbb{D}$, define*

$$S_\theta(z) = \sum_{j=0}^\infty \sin((2j+1)\theta) z^j \tag{5.1}$$

$$C_\theta(z) = \sum_{j=0}^\infty \cos((2j+1)\theta) z^j \tag{5.2}$$

Then

$$\frac{1+z}{1-z}\, C_\theta(z) = \cot(\theta) S_\theta(z) \tag{5.3}$$

Proof. Let $\omega = e^{i\theta}$ so, summing the geometric series,

$$\begin{aligned} S_\theta(z) &= (2i)^{-1} \sum_{j=0}^\infty (\omega^{2j+1} z^j - \bar\omega^{2j+1} z^j) \\ &= (2i)^{-1} \left[\frac{\omega}{1 - z\omega^2} - \frac{\bar\omega}{1 - z\bar\omega^2} \right] \qquad (5.4) \\ &= \frac{\sin(\theta)(1+z)}{(1 - z\omega^2)(1 - z\bar\omega^2)} \qquad (5.5) \end{aligned}$$

For $C_\omega(z)$, the calculation is similar; in (5.4), $(2i)^{-1}$ is replaced by $(2)^{-1}$ and the minus sign becomes a plus:

$$C_\omega(z) = \frac{\cos(\theta)(1-z)}{(1 - z\omega^2)(1 - z\bar\omega^2)} \tag{5.6}$$

(5.5) and (5.6) imply (5.3). $\square$

Our first proof of Theorem 2 depends on looking at the Hankel matrix [12, 13]

$$\widetilde{M_n} = \begin{pmatrix} 2 & 2 & \dots & 2 & 1 \\ 2 & 2 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix} \tag{5.7}$$

If W_n is the unitary permutation matrix

$$(Wv)_j = v_{n+1-j} \tag{5.8}$$

then

$$M_n = \widetilde{M}_n W \qquad \widetilde{M}_n = M_n W \tag{5.9}$$

and so

$$\|M_n\| = \|\widetilde{M}_n\| \tag{5.10}$$

Here is our first proof of Theorem 2:

Theorem 5.2. *Let*

$$c_j^{(n;\ell)} = \cos\left(\left(2\ell + \frac{1}{2}\right)\frac{\pi}{2n}\,(2j-1)\right) \qquad j = 1, 2, \dots, n;\ \ell = 0, \dots, n-1 \tag{5.11}$$

Then

$$\widetilde{M}_n c^{(n;\ell)} = \cot\left(\left(2\ell + \frac{1}{2}\right)\frac{\pi}{2n}\right) c^{(n;\ell)} \tag{5.12}$$

Thus,

$$\|M_n\| = \|\widetilde{M}_n\| = \cot\left(\frac{\pi}{4n}\right) \tag{5.13}$$

Proof. Let

$$c_j^{(n;\theta)} = \cos(\theta(2j-1)) \qquad j = 1, 2, \dots, n \tag{5.14}$$

and

$$s_j^{(n;\theta)} = \sin(\theta(2j-1)) \qquad j = 1, \dots, n \tag{5.15}$$

Then (5.3) implies that

$$M_n W c^{(n;\theta)} = \cot(\theta) W s^{(n;\theta)} \tag{5.16}$$

by looking at coefficients of $1, z, \dots, z^{n-1}$. The W comes from (3.6)/(3.8). If

$$\theta = \frac{\pi}{2} + 2\ell\pi \qquad \ell = 0, \dots, n-1 \tag{5.17}$$

then

$$W s^{(n;\theta)} = c^{(n;\theta)} \tag{5.18}$$

and (5.16) becomes (5.12).

Since $\widetilde{M}$ is self-adjoint, (5.13) follows from (5.12) either by noting that $\max|\cot((2\ell + \frac{1}{2})\frac{\pi}{2n})| = \cot(\frac{\pi}{4n})$ or by noting that $c^{(n;\theta=\pi/4n)}$ is a positive eigenvector of a positive self-adjoint matrix, so its eigenvalue is the norm by the Perron-Frobenius theorem. □

Our second proof relies on the following known result (see Milovanić et al. [5], page 272, and references therein; this result is called the Eneström-Kakeya theorem; see also Pólya-Szegő [14], problem 22 on pp. 107 and 301, who also mention Hurwitz):

Lemma 5.3. *Suppose*

$$0 < a_0 < a_1 < \cdots < a_n \tag{5.19}$$

Then

$$P(z) = a_0 + a_1 z + \cdots + a_n z^n \tag{5.20}$$

has all its zeros in $\mathbb{D}$*.*

Theorem 5.4. *Let*

$$S^{(n)}(z) = \sum_{j=0}^{n-1} \sin\biggl((2j+1)\frac{\pi}{4n}\biggr) z^j \tag{5.21}$$

$$C^{(n)}(z) = \sum_{j=0}^{n-1} \cos\biggl((2j+1)\frac{\pi}{4n}\biggr) z^j \tag{5.22}$$

Then

$$b^{(n)}(z) = \frac{S^{(n)}(z)}{C^{(n)}(z)} \tag{5.23}$$

is a Blaschke product of order $n-1$*. Moreover,*

$$\cot\biggl(\frac{\pi}{4n}\biggr) b^n(z) = 1 + 2\sum_{j=1}^{n-1} z^j + O(z^n) \tag{5.24}$$

and

$$\|M_n\| = \cot\biggl(\frac{\pi}{4n}\biggr) \tag{5.25}$$

Proof. The coefficients of $S^{(n)}$ obey (5.19) so, by the lemma, $S^{(n)}$ has all its zeros in $\mathbb{D}$. Moreover, by (5.18), $C^{(n)}(z) = z^n\, \overline{S^{(n)}(1/\bar{z})}$, which implies (5.23) is a Blaschke product.

(5.24) is just a translation of (5.3). (5.24) implies (5.25) by Theorem 3.3. □

6. Some Remarks and Extensions

In this section,we make some remarks that shed light on or extend Theorem 1, our main result.

A. An alternate proof. We give a simple proof of a weakened version of Theorem 4 but which suffices for applications like those in Section 7. This argument is related to ones in Section 3 of Nikolski [11].

Theorem 6.1. *If* $\|A\| \le 1$ *and* $1 \notin \operatorname{spec}(A)$*, then*

$$\operatorname{dist}(1, \operatorname{spec}(A))\|(1-A)^{-1}\| \le 2m \tag{6.1}$$

where m *is the degree of the minimal polynomial for* A*.*

Proof. We prove the result for $\|A\| < 1$. The general result follows by taking limits. We make repeated use of Lemma 3.1 which implies that if, for $\lambda \in \mathbb{D}$, and we define

$$B(\lambda) = \left(\frac{A-\lambda}{1-\overline{\lambda}A}\right)\left(\frac{1-\overline{\lambda}}{1-\lambda}\right) \tag{6.2}$$

then

$$\|B(\lambda)\| \leq 1 \tag{6.3}$$

By algebra,

$$(1-x)^{-1}\left[1 - \frac{x-\lambda}{1-\overline{\lambda}x}\left(\frac{1-\overline{\lambda}}{1-\lambda}\right)\right] = \frac{1}{1-\lambda}\left[1 + \overline{\lambda}\left(\frac{x-\lambda}{1-x\overline{\lambda}}\right)\right] \tag{6.4}$$

so, by Lemma 3.1 again,

$$\|(1-A)^{-1}(1-B(\lambda))\| \leq |1-\lambda|^{-1}(1+|\lambda|) \tag{6.5}$$

Now let $\prod_{j=1}^m (x-\lambda_j)$ be the minimal polynomial for A. Then

$$\prod_{j=1}^m B(\lambda_j) = 0$$

so

$$\begin{aligned}(1-A)^{-1} &= (1-A)^{-1}\left[1 - \prod_{j=1}^m B_j(\lambda)\right] \\ &= \sum_{j=1}^m (1-A)^{-1}[1-B_j(\lambda)] \prod_{k=j+1}^m B_k(\lambda) \end{aligned} \tag{6.6}$$

(the empty product for $j = m$ is interpreted as the identity operator) which, by (6.3) and (6.5), implies

$$\begin{aligned}\text{LHS of (6.1)} &\leq \sum_{j=1}^m \operatorname{dist}(1, \operatorname{spec}(A))|1-\lambda_j|^{-1}(1+|\lambda_j|) \\ &\leq 2m\end{aligned}$$

since $1+|\lambda_j| \leq 2$ and $\lambda_j \in \operatorname{spec}(A)$ so $\operatorname{dist}(1, \operatorname{spec}(A))|1-\lambda_j|^{-1} \leq 1$. $\square$

Remarks. 1. The factor $(1-\overline{\lambda})/(1-\lambda)$ is taken in (6.2) so $f_\lambda(z) = (z-\lambda)(1-\overline{\lambda}z)^{-1}(1-\overline{\lambda})(1-\lambda)^{-1}$ has $1 - f_\lambda(1) = 0$.

2. In place of the algebra (6.4), one can compute that the $\sup_{|z|<1}$ LHS of (6.4) is $|1-\lambda|^{-1}[1+|\lambda|]$ and use von Neumann's theorem as discussed in Subsection E below.

B. Minimal polynomials. While the constant 2 in (6.1) is worse than $4/\pi$ in (1.19)/(1.21), (6.1) appears to be stronger in that m, not n, appears, but we can also strengthen (1.19) in this way:

Theorem 6.2. *If $\|A\| \leq 1$, $1 \notin \operatorname{spec}(A)$, and m is the degree of the minimal polynomial for A, then*

$$\operatorname{dist}(1, \operatorname{spec}(A))\|(1-A)^{-1}\| \leq \cot\left(\frac{\pi}{4m}\right) \tag{6.7}$$

Proof. Let $\|y\| = 1$. Since $A^m y$ is a linear combination of $\{A^j y\}_{j=0}^{m-1}$, the cyclic subspace, V_y, has $\dim(V_y) \equiv m_y \leq m$. Since $A \restriction V_y$ is an operator of a space of dimension m_y, we have

$$\begin{aligned}\operatorname{dist}(1, \operatorname{spec}(A))\|(1-A)^{-1}y\| &\leq c(m_y) = \cot\left(\frac{\pi}{4m_y}\right) \\ &\leq \cot\left(\frac{\pi}{4m}\right)\end{aligned}$$

$\square$

C. Numerical range. For any bounded operator, A, on a Hilbert space, the numerical range, $\operatorname{Num}(A)$, is defined by

$$\operatorname{Num}(A) = \{\langle\varphi, A\varphi\rangle \mid \|\varphi\| = 1\} \tag{6.8}$$

It is a bounded convex set (see [3, p. 150]), and when A is a finite matrix, also closed. Theorem 1 can be improved to read:

Theorem 6.3. *Let $\widetilde{\mathcal{M}}_n$ be the set of pairs (A, z) where A is an $n \times n$ matrix, $z \in \mathbb{C}$ with*

$$z \notin \operatorname{spec}(A) \qquad z \notin \operatorname{Num}(A)^{\text{int}} \tag{6.9}$$

Then

$$\sup_{\widetilde{\mathcal{M}}_n} \operatorname{dist}(z, \operatorname{spec}(A))\|(A-z)^{-1}\| = \cot\left(\frac{\pi}{4n}\right) \tag{6.10}$$

Remarks. 1. Since $\operatorname{Num}(A) \subset \{z \mid |z| \leq \|A\|\}$, $\mathcal{M}_n \subset \widetilde{\mathcal{M}}_n$, and this is a strict improvement of (1.19).

2. We need only prove

$$\operatorname{dist}(z, \operatorname{spec}(A))\|(A-z)^{-1}\| \leq \cot\left(\frac{\pi}{4n}\right)$$

since the equality then follows from $\mathcal{M}_n \subset \widetilde{\mathcal{M}}_n$.

3. By replacing A by $e^{i\theta}(A-z)$ for suitable θ and z, we need only prove

$$\operatorname{Re}(A) \geq 0,\ 0 \notin \operatorname{spec}(A) \Rightarrow \operatorname{dist}(0, \operatorname{spec}(A))\|A^{-1}\| \leq \cot\left(\frac{\pi}{4n}\right) \tag{6.11}$$

for by convexity of $\mathrm{Num}(A)$, if $z \notin \mathrm{Num}(A)^{\mathrm{int}}$, there is a half-plane, P, with $\mathrm{Num}(A) \subset P$ and $z \in \partial P$. It is (6.11) we will prove below.

First Proof of Theorem 6.3. Let

$$C = A^{-1} + (A^*)^{-1} \tag{6.12}$$

$$= (A^*)^{-1} 2\,\mathrm{Re}(A)(A)^{-1} \geq 0 \tag{6.13}$$

Thus,

$$|C_{jk}| \leq |C_{jj}|^{1/2} |C_{kk}|^{1/2} \tag{6.14}$$

Now just follow the proof of Theorem 4 in Section 2. □

Second Proof of Theorem 6.3. We use Cayley transforms. For $0 < s$, define

$$B(s) = (1 - sA)(1 + sA)^{-1} \tag{6.15}$$

Since

$$\|(1 + sA)\varphi\|^2 - \|(1 - sA)\varphi\|^2 = 4s\ \mathrm{Re}(\varphi, A\varphi) \geq 0$$

we have that

$$\|B(s)\| \leq 1 \tag{6.16}$$

Because

$$1 - B(s) = 2sA(1 + sA)^{-1} \tag{6.17}$$

we have for s small that

$$\mathrm{dist}(1, \mathrm{spec}(B(s))) = 2s\,\mathrm{dist}(0, \mathrm{spec}(A)) + O(s^2) \tag{6.18}$$

Thus, by Theorem 1,

$$2s\,\mathrm{dist}(0, \mathrm{spec}(A)) \|(1 - B(s))^{-1}\| \leq \cot\left(\frac{\pi}{4n}\right) + O(s) \tag{6.19}$$

By (6.17),

$$(1 - B(s))^{-1} = (2s)^{-1}[A^{-1} + s]$$

so

$$\|A^{-1}\| \leq |s| + 2s\|(1 - B(s))^{-1}\| \tag{6.20}$$

This plus (6.18) implies (6.11) as $s \downarrow 0$. □

D. Bounded powers. We note that there is also a result if

$$\sup_{m\geq 0} \|A^m\| = c < \infty \tag{6.21}$$

We suspect the $3/2$ power in the following is not optimal. We note that one can also use this method if $\|A^m\|$ is polynomially bounded in m.

Theorem 6.4. *If* (6.21) *holds, then*

$$\|(1-A)^{-1}\| \leq c(3n)^{3/2}\text{dist}(1,\text{spec}(A))^{-3/2} \tag{6.22}$$

Proof. By the argument of Section 1 (using (1.11)), this is equivalent to

$$\text{dist}(1,\text{spec}(A)) \leq 3n(c\|(1-A)y\|)^{2/3} \tag{6.23}$$

for all unit vectors y.

Define for $1<r$,

$$\langle f,g\rangle_r = \sum_{m=0}^\infty r^{-2m}\langle A^m f, A^m g\rangle \tag{6.24}$$

By (6.21),

$$\|f\| \leq \|f\|_r \leq cr(r^2-1)^{-1/2}\|f\| \tag{6.25}$$

By (6.24),

$$\|Af\|_r^2 \leq r^2\|f\|_r^2 \tag{6.26}$$

so

$$\|A\|_r \leq r \tag{6.27}$$

so if $C = r^{-1}A$, then

$$\|C\|_r \leq 1 \tag{6.28}$$

Clearly, for $\|y\| = 1 \leq \|y\|_r$,

$$\begin{aligned}\|Cy-y\|_r &\leq |r^{-1}-1|\,\|y\|_r + r^{-1}\|(A-1)y\|_r \\ &\leq |r^{-1}-1|\,\|y\|_r + c(r^2-1)^{-1/2}\|(A-1)y\| \\ &\leq ((r-1)+c[2(r-1)]^{-1/2}\|(A-1)y\|)\|y\|_r \end{aligned} \tag{6.29}$$

It follows by Theorem 1 and the fact that $\text{spec}(A)$ is independent of $\|\cdot\|_r$ that

$$\text{dist}(1,r^{-1}\text{spec}(A)) \leq \frac{4n}{\pi}\left\{c\|(A-1)y\|(2(r-1))^{-1/2}+(r-1)\right\} \tag{6.30}$$

and thus

$$\text{dist}(1,\text{spec}(A)) \leq (r-1) + \frac{4\pi}{n}\left\{c\|(A-1)y\|(2(r-1))^{-1/2}+(r-1)\right\} \tag{6.31}$$

Choosing $r = 1 + \frac{1}{2}(c\|(A-1)y\|)^{2/3}$ and using $\frac{1}{2} + \frac{6n}{\pi} \leq 3n$, we obtain (6.23). $\square$

E. Von Neumann's theorem. Lemma 3.1 is a special case of a theorem of von Neumann. The now standard proof of this result uses Nagy dilations [23]; we have found a simple alternative that relies on

Lemma 6.5. *For any A, with $\|A\| < 1$ and $A = U|A|$, and U unitary, there exists an operator-valued function, g, analytic in a neighborhood of $\overline{\mathbb{D}}$ so that $g(e^{i\theta})$ is unitary and $g(0) = A$.*

Proof. Let

$$g(z) = U\left[\frac{z + |A|}{1 + z|A|}\right] \tag{6.32}$$

The factor in $[\dots]$ is unitary if $z = e^{i\theta}$, since

$$\begin{aligned}(e^{i\theta} + |A|)^*(e^{i\theta} + |A|) &= 1 + A^*A + 2\cos\theta|A| \\ &= (1 + e^{i\theta}|A|)^*(1 + e^{i\theta}|A|)\end{aligned}$$

$\square$

Theorem 6.6 (von Neumann [25])**.** *Let $f\colon \mathbb{D} \to \mathbb{D}$. If $\|A\| < 1$, define $f(A)$ by*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \qquad f(A) \equiv \sum_{n=0}^{\infty} a_n A^n \tag{6.33}$$

Then

$$\|f(A)\| \leq 1 \tag{6.34}$$

Proof of von Neumann's theorem, given the lemma. Suppose first that A obeys the hypotheses of the lemma. By a limiting argument, suppose f is analytic in a neighborhood of $\overline{\mathbb{D}}$. Applying the maximum principle to $f(g(z))$, we see

$$\begin{aligned}\|f(A)\| = \|f(g(0))\| &\leq \sup_{\theta} \|f(g(e^{i\theta}))\| \\ &= \sup_{\theta} |f(e^{i\theta})| \leq 1\end{aligned} \tag{6.35}$$

where (6.35) uses the spectral theorem for the unitary $g(e^{i\theta})$.

For general A, if $\tilde{A} = A \oplus 0$ on $\mathcal{H} \oplus \mathcal{H}$, then $\tilde{A} = U|\tilde{A}|$ with U unitary and we obtain $\|f(\tilde{A})\| \leq 1$. But $f(\tilde{A}) = f(A) \oplus 0$. $\square$

Remarks. 1. In general, $A = V|A|$ with V a partial isometry. We can extend this to a unitary U so long as $\dim(\mathrm{Ran}(V)^\perp) = \dim(\ker(V)^\perp)$. This is automatic in the finite-dimensional case and also if $\dim(\mathcal{H}) = \infty$ for $A \oplus 0$ since then both spaces are infinite-dimensional.

2. This proof is close to one of Nelson [9] who also uses the maximum principle and polar decomposition, but uses a different method for interpolating the self-adjoint part (see also Nikolski [10]).

7. Zeros of Random OPUC

In this section, we apply Theorem 1 to obtain results on certain OPUC. We begin by recalling the recursion relations for OPUC [17, 18, 19]. For each non-trivial probability measure, $d\mu$, on $\partial\mathbb{D}$, there is a sequence of complex numbers, $\{\alpha_n(d\mu)\}_{n=0}^\infty$, called Verblunsky coefficients so that

$$\Phi_{n+1}(z) = z\Phi_n(z) - \bar{\alpha}_n \Phi_n^*(z) \tag{7.1}$$

where

$$\Phi_n^*(z) = z^n\, \overline{\Phi_n(1/\bar{z})} \tag{7.2}$$

The α_n obey $|\alpha_n| < 1$ and Verblunsky's theorem [17, 19] says that $\mu \mapsto \{\alpha_n(d\mu)\}_{n=0}^\infty$ is a bicontinuous bijection from the non-trivial measures on $\partial\mathbb{D}$ with the topology of vague convergence to $\mathbb{D}^\infty$ with the product topology.

For each ρ in $(0,1)$, we define the ρ-model to be the set of random Verblunsky coefficients where α_n are independent, identically distributed random variables, each uniformly distributed in $\{z \mid |z| \leq \rho\}$. A point in the model space of α's will be denoted ω; $\Phi_n(z;\omega)$ will be the corresponding OPUC and $\{z_j^{(n)}(\omega)\}_{j=1}^n$ the zeros of Φ_n counting multiplicity. Our results here depend heavily on earlier results of Stoiciu [20, 21], who studied a closely related problem (see below). In turn, Stoiciu relied, in part, on earlier work on eigenvalues of random Schrödinger operators [7, 6].

We will prove the following three theorems:

Theorem 7.1. *Let $0 < \rho < 1$. Let $k \in \{1, 2, \dots\}$. Then for a.e. ω in the ρ-model,*

$$\limsup_{n\to\infty} \frac{\#\{j \mid |z_j^{(n)}(\omega)| < 1 - n^{-k}\}}{[\log(n)]^2} < \infty \tag{7.3}$$

Thus, the overwhelming bulk of zeros are polynomially close to $\partial\mathbb{D}$. If we look at a small slice of argument, we can say more:

Theorem 7.2. *Let $0 < \rho < 1$. Let $\theta_0 \in [0, 2\pi)$ and $a < b$ real. Let $\eta < 1$. Then with probability 1, for large n, there are no zeros in $\{z \mid \arg z \in (\theta_0 + \frac{2\pi a}{n}, \theta_0 + \frac{2\pi b}{n});\ |z| < 1 - \exp(-n^\eta)\}$.*

Finally and most importantly, we can describe the statistical distribution of the arguments:

Theorem 7.3. *Let $0 < \rho < 1$. Let $\theta_0 \in [0, 2\pi)$. Let $a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_\ell < b_\ell$ and let $k_1, \dots, k_\ell$ be in $\{0, 1, 2, \dots\}$. Then as $n \to \infty$,*

$$\text{Prob}\biggl(\#\biggl(j \biggm| \arg z_j^{(n)}(\omega) \in \biggl(\theta_0 + \frac{2\pi a_m}{n}, \theta_0 + \frac{2\pi b_n}{n}\biggr)\biggr) = k_m \text{ for } m = 1, \dots, \ell \tag{7.4}$$

converges to

$$\prod_{m=1}^{\ell} \frac{(b_m - a_m)^{k_m}}{k_m!} \, e^{-(b_m - a_m)} \tag{7.5}$$

This says the zeros are asymptotically Poisson distributed. As we stated, our proofs rely on ideas of Stoiciu, essentially using Theorem 1 to complete his program. To state the results of his that we use, we need a definition.

For $\beta \in \partial\mathbb{D}$, the paraorthogonal polynomials (POPUC) are defined by

$$\Phi_n^{(\beta)}(z) = \Phi_{n-1}(z) - \bar{\beta}\Phi_{n-1}^*(z) \tag{7.6}$$

These have zeros on $\partial\mathbb{D}$. Indeed, they are eigenvalues of a rank one unitary perturbation of the operator A_n of (1.6). We extend the ρ-model to include an additional set of independent parameters $\{\beta_j\}_{j=0}^\infty$ in $\partial\mathbb{D}$, each uniformly distributed on $\partial\mathbb{D}$. $\tilde{z}_j^{(n)}(\omega)$ denotes the zeros of $\Phi_n^{(\beta_n)}(z;\omega)$. Stoiciu [20, 21] completely analyzed these POPUC zeros. We will need three of his results:

Theorem 7.4 (= Theorem 6.1.3 of [21] = Theorem 6.3 of [20])**.** *Let I be an interval in $\partial\mathbb{D}$. Then*

$$\text{Prob}(2 \text{ or more } \tilde{z}_j^{(n)}(\omega) \text{ lie in } I) \leq \frac{1}{2}\biggl(\frac{n|I|}{2\pi}\biggr)^2 \tag{7.7}$$

where $|I|$ is the $d\theta$ measure of I.

For the next theorem, we need the fact that there is an explicit realization of A_n and the associated rank one perturbations as $n \times n$ complex CMV matrices (see [2, 17, 18, 19]), $\mathcal{C}_n$, whose eigenvalues are the z_j^n, and $\tilde{\mathcal{C}}_n^{(\beta_n)}$ whose eigenvalues are the $\tilde{z}_j^n$, so that

$$\|(\mathcal{C}_n - \mathcal{C}_n^{(\beta_n)})\varphi\| \leq |\varphi_{n-1}| + |\varphi_n| \tag{7.8}$$

The next theorem uses the components so (7.8) holds.

Theorem 7.5 (= Theorem 1.1.2 of [21] = Theorem 2.2 of [20])**.** *There exists a constant D_2 (depending only on ρ) so that for every eigenvector*

$\varphi^{(j,\omega;n)}$ of $\tilde{\mathcal{C}}_n^{(\beta_n)}$, we have for

$$|m - m(\varphi^{(j,\omega;n)})| \geq D_2(\log n) \tag{7.9}$$

that

$$|\varphi_m^{(j,\omega;n)}| \leq C_\omega e^{-4|m-m(\varphi^{(j,\omega;n)})|/D_2} \tag{7.10}$$

where C_ω is an a.e. finite constant and

$$m(\varphi) = \textit{first } k \textit{ so } |\varphi_k| = \max_m |\varphi_m| \tag{7.11}$$

We will also need the results that Stoiciu proves along the way that for each C_0,

$$\{\omega \mid C_\omega < C_0\} \equiv \Omega_{C_0} \tag{7.12}$$

is invariant under rotation of the measures $d\mu_\omega$, and that for each C_0 fixed and all $\omega \in \Omega_{C_0}$,

$$\#(j \mid m(\varphi^{(j,\omega;n)}) = m_0) \leq D_3(\log n) \tag{7.13}$$

where D_3 is only C_0-dependent and is independent of ω, m_0, and n. (7.13) comes from the fact that, by (7.10), for D_3 only depending on C_0,

$$\sum_{|m-m(\varphi)| \geq \frac{1}{4} D_3(\log n)} |\varphi_m|^2 \leq \tfrac{1}{2} \tag{7.14}$$

so, by (7.11), for φ's with $m(\varphi) = m_0$,

$$\tfrac{1}{2}\, D_3(\log n) |\varphi_{m_0}|^2 \geq \tfrac{1}{2} \tag{7.15}$$

which, given

$$\sum_\varphi |\varphi_{m_0}|^2 = 1 \tag{7.16}$$

implies (7.13).

The last of Stoiciu's results we will need is

Theorem 7.6 (= Theorem 1.0.6 of [21] = Theorem 1.1 of [20])**.** *For $\theta_0 \in [0, 2\pi)$ and $a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_\ell < b_\ell$ and $k_1, \dots, k_\ell$ in $\{0, 1, 2, \dots\}$, we have, as $n \to \infty$, that* (7.4) *with $z_j^{(n)}$ replaced by $\tilde{z}_j^{(n)}$ converges to* (7.5)*.*

With this background out of the way, we begin the proofs of the new Theorems 7.1–7.3 with

Theorem 7.7. *Fix $\rho \in (0, 1)$. Then for a.e. ω, there exists N_ω so if $n \geq N_\omega$, then*

$$\min_{j \neq k} |\tilde{z}_j^{(n)} - \tilde{z}_k^{(n)}| \geq 2n^{-4} \tag{7.17}$$

Remark. $n^{-3-\varepsilon}$ will work in place of n^{-4}.

Proof. For each n, cover $\partial\mathbb{D}$ by two sets of intervals of size $4n^{-4}$: one set non-overlapping, except at the end, starting with $[0, 4n^{-4}]$ and the other set starting with $[2n^{-4}, 6n^{-4}]$. If (7.17) fails for some n, then there are two zeros within one of these intervals. By (7.7), the probability of two zeros in one of these intervals is $O((nn^{-4})^2) = O(n^{-6})$. The number of intervals at order n is $O(n^4)$. Since $\sum_{n=1}^{\infty} n^4 n^{-6} < \infty$, the sum of the probabilities of two zeros in an interval is summable. By the Borel-Cantelli lemma [22] for a.e. ω, only finitely many intervals have two zeros. Hence, for large n, (7.17) holds. $\square$

Proof of Theorem 7.1. Obviously, if (7.3) holds for some k, it holds for all smaller k, so we will prove it for $k \geq 4$. We also need only prove it on any Ω_{C_0} given by (7.12) since $\cup\Omega_{C_0}$ has probability 1 by Theorem 7.5. Consider those $\varphi^{(j,\omega;n)}$ with

$$|m(\varphi^{(j,\omega;n)}) - n| \geq K(\log n) \tag{7.18}$$

By (7.13), the number of j for which (7.18) fails is $O((\log n)^2)$.

By (7.10) and (7.8) and the fact that φ is a unit eigenfunction, then

$$\|(\mathcal{C}_n - \tilde{z}_j^{(n)})\varphi^{(j,\omega;n)}\| \leq 2C_\omega n^{-4K/D_2} \tag{7.19}$$

so picking K large enough and n large enough that $\frac{4}{\pi}2C_\omega n^{-1} < 1$, we have

$$\|(\mathcal{C}_n - \tilde{z}_j^{(n)})\varphi^{(j,\omega;n)}\| \leq \frac{\pi}{4n}\, n^{-k} \tag{7.20}$$

Thus, by Theorem 1 and $\|\mathcal{C}_n\| = 1 = |\tilde{z}_j^{(n)}|$, we see that for each j obeying (7.18), there is a $z_j^{(n)}$ so

$$|z_j^{(n)} - \tilde{z}_j^{(n)}| \leq n^{-k} \tag{7.21}$$

By Theorem 7.7 and $k \geq 4$, the $z_j^{(n)}$ are distinct for n large, so we have $n - O((\log n)^2)$ zeros with $|z_j^{(n)}| \geq 1 - n^{-k}$. This is (7.3). $\square$

Proof of Theorem 7.2. In place of (7.18), we look for φ's so

$$|m(\varphi^{(j,\omega;n)}) - n| \geq \frac{D_2}{2}\, n^{1-\eta} \tag{7.22}$$

For such j's, using the above arguments, there are zeros $z_j^{(n)}$ with

$$|z_j^{(n)} - \tilde{z}_j^{(n)}| \leq C_\omega \exp(-2n^\eta) \tag{7.23}$$

$\square$

As in Stoiciu [20, 21], the distribution of $\tilde{z}_j^{(n)}$ for which (7.22) fails is rotation invariant. Since the number is $O(n^{1-\eta}\log n)$ out of $O(n)$

zeros, the probability of any of these had zeros lying in $\{z \mid \arg z \in (\theta_0 + \frac{2\pi a}{n}, \theta_0 + \frac{2\pi b}{n})\}$ goes to zero as $n \to \infty$.

Proof of Theorem 7.3. By the last proof, the zeros of Φ_n with the given arguments lie within $O(e^{-n^\eta})$ of those of $\Phi_n^{(\beta)}$ and, by Theorem 7.7, these zeros are distinct. Theorem 7.6 completes the proof if one gets upper and lower bounds by slightly increasing/decreasing the intervals on an $O(1/n)$ scale. □

We close with the remark about improving these theorems. While (7.13) is the best one can hope for as a uniform bound, with overwhelming probability the number should be bounded. Thus, we expect in Theorem 7.1 that one can obtain $O((\log n)^{-1})$ in place of $O((\log n)^{-2})$. It is possible in Theorem 7.2 that one can improve $O(e^{-n^\eta})$ for all $\eta \in 1$ to $O(e^{-An})$ for some A.

References

[1] A. Böttcher and B. Silbermann, *Analysis of Toeplitz Operators*, Springer, Berlin, 1990.

[2] M. J. Cantero, L. Moral, and L. Velázquez, *Five-diagonal matrices and zeros of orthogonal polynomials on the unit circle*, Linear Algebra Appl. **362** (2003), 29–56.

[3] E. B. Davies, *One-Parameter Semigroups*, London Mathematical Society Monographs, 15, Academic Press, London-New York, 1980.

[4] P. Henrici, *Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices*, Numer. Math. **4** (1962), 24–40.

[5] G. V. Milovanović, D. S. Mitrinović, and Th. M. Rassias, *Topics in Polynomials: Extremal Problems, Inequalities, Zeros*, World Scientific Publishing, River Edge, NJ, 1994.

[6] N. Minami, *Local fluctuation of the spectrum of a multidimensional Anderson tight binding model*, Comm. Math. Phys. **177** (1996), 709–725.

[7] S. A. Molchanov, *The local structure of the spectrum of the one-dimensional Schrödinger operator*, Comm. Math. Phys. **78** (1980/81), 429–446.

[8] Z. Nehari, *On bounded bilinear forms*, Ann. of Math. **65** (1957), 153–162.

[9] E. Nelson, *The distinguished boundary of the unit operator ball*, Proc. Amer. Math. Soc. **12** (1961), 994–995.

[10] N. K. Nikolski, *Operators, Functions, and Systems: An Easy Reading, Vol. 2: Model Operators and Systems*, Mathematical Surveys and Monographs, 93, American Mathematical Society, Providence, RI, 2002.

[11] N. K. Nikolski, *Condition numbers of large matrices, and analytic capacities*, to appear in St. Petersburg Math. J.

[12] J. R. Partington, *An Introduction to Hankel Operators*, London Mathematical Society Student Texts, 13, Cambridge University Press, Cambridge, 1988.

[13] V. V. Peller, *Hankel Operators and Their Applications*, Springer Monographs in Math., Springer, New York, 2003.

[14] G. Pólya and G. Szegő, *Problems and Theorems in Analysis. I*, reprint of the 1978 English translation, Classics in Mathematics, Springer, Berlin, 1998.

[15] Pseudospectra Gateway,
`http://web.comlab.ox.ac.uk/projects/pseudospectra/`

[16] I. Schur, *Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind, I, II*, J. Reine Angew. Math. **147** (1917), 205–232; **148** (1918), 122–145. English translation in "I. Schur Methods in Operator Theory and Signal Processing" (edited by I. Gohberg), pp. 31–59, 66–88, Operator Theory: Advances and Applications, 18, Birkhäuser, Basel, 1986.

[17] B. Simon, *Orthogonal Polynomials on the Unit Circle, Part 1: Classical Theory*, AMS Colloquium Series, American Mathematical Society, Providence, RI, 2005.

[18] B. Simon, *Orthogonal Polynomials on the Unit Circle, Part 2: Spectral Theory*, AMS Colloquium Series, American Mathematical Society, Providence, RI, 2005.

[19] B. Simon, *OPUC on one foot*, Bull. Amer. Math. Soc. **42** (2005), 431–460.

[20] M. Stoiciu, *The statistical distribution of the zeros of random paraorthogonal polynomials on the unit circle*, to appear in J. Approx. Theory.

[21] M. Stoiciu, *Zeros of Random Orthogonal Polynomials on the Unit Circle*, Ph.D. dissertation, 2005.
`http://etd.caltech.edu/etd/available/etd-05272005-110242/`

[22] D. Stroock, *A Concise Introduction to the Theory of Integration*, Series in Pure Math., 12, World Scientific Publishing, River Edge, NJ, 1990.

[23] B. Sz.-Nagy and C. Foias, *Harmonic Analysis of Operators on Hilbert Space*, North-Holland Publishing, Amsterdam-London; American Elsevier Publishing, New York; Akadémiai Kiadó, Budapest, 1970.

[24] L. N. Trefethen and M. Embree, *Spectra and Pseudospectra: The Behavior of Non-normal Matrices and Operators*, Princeton University Press, Princeton, NJ, expected 2005.

[25] J. von Neumann, *Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes*, Math. Nachr. **4** (1951), 258–281.