FROM RANDOM MATRICES TO RANDOM ANALYTIC FUNCTIONS[1]

BY MANJUNATH KRISHNAPUR

University of Toronto

We consider two families of random matrix-valued analytic functions: (1) $G_1 - zG_2$ and (2) $G_0 + zG_1 + z^2G_2 + \cdots$, where G_i are $n \times n$ random matrices with independent standard complex Gaussian entries. The random set of z where these matrix-analytic functions become singular is shown to be determinantal point processes in the sphere and the hyperbolic plane, respectively. The kernels of these determinantal processes are reproducing kernels of certain Hilbert spaces ("Bargmann–Fock spaces") of holomorphic functions on the corresponding surfaces. Along with the new results, this also gives a unified framework in which to view a theorem of Peres and Virág ($n = 1$ in the second setting above) and a well-known result of Ginibre on Gaussian random matrices (which may be viewed as an analogue of our results in the whole plane).

1. Leading up to the results. Singular points of random matrix-valued analytic functions are a common generalization of eigenvalues of random matrices and zeros of random polynomials. The setting is that we have an analytic function of z taking values in the space of $n \times n$ matrices. Singular points are those (random) z where the matrix becomes singular, that is, the zeros of the determinant. This notion was introduced in the Ph.D. thesis [12] of the author, where some basic facts were found. Of course, singular points are just the zeros of the (random analytic function) determinant, so in what sense is this concept novel?

In the case of random matrices as well as random analytic functions, the following features may be observed.

1. For very special models, usually with independent Gaussian coefficients or entries, one may solve *exactly* for the distribution of zeros or eigenvalues.
2. For more general models with independent coefficients or entries, under rather weak assumptions on moments, one can usually analyze the empirical measure of eigenvalues or zeros as the size of the matrix increases or the degree of the polynomial goes to infinity.
3. Substituting independence with assumptions of particular kinds of symmetry and dependence of entries or coefficients, eigenvalues and zeros have been studied with varying degrees of success.

The point here is that the determinant of a random matrix-valued analytic function has coefficients that are dependent in a very complicated way and one might not expect it to be tractable. Nevertheless,

- In this paper we demonstrate that certain special models of random matrix-valued analytic functions based on independent complex Gaussians have singular sets that turn out to be determinantal point processes in the two-dimensional sphere or the hyperbolic plane! This is the exactly solvable situation.
- In a subsequent paper [11], we shall study the asymptotics of the counting measure on the singular set of random matrix-valued analytic functions, under weak assumptions of independence (and some moment conditions), as the matrix size goes to infinity. This will be a generalization of the circular law for non-Hermitian random matrices with independent entries.
- Numerous questions suggest themselves, taking unitary or Hermitian matrix coefficients in a random polynomial, for example. Answers are yet to be found.

Despite its natural appeal, the concept of a random matrix-valued analytic function does not seem to have been considered in the literature, perhaps because the focus in random matrix theory has been mostly on eigenvalues in one dimension (real line or the circle). We do not know a way to force singular points to lie on the line (except the case of eigenvalues of Hermitian matrices).

A notion that is necessary to even state some of our results is that of a *determinantal point process*, first defined by Macchi [14]. For the reader not familiar with them, a brief introduction to determinantal processes is given in the Appendix. This is sufficient for the purposes of this paper, but to know more, the reader may consult the surveys [7, 19]. The reader interested merely in our results and proofs may now jump directly to the next section. The rest of this section is devoted to motivating the results and establishing the context and is not logically necessary to read the rest of the paper. First some notation.

Notation: $\mathbb{D}$ is the unit disk in the complex plane. The two-dimensional sphere $\mathbb{S}^2$ will always be identified with $\mathbb{C} \cup \{\infty\}$ via stereographic projection. m denotes Lebesgue measure. g and g_i, $i \geq 0$, are always independent standard complex Gaussian random variables, with density $\pi^{-1} \exp\{-|z|^2\}$ in the plane. G and G_i, $i \geq 0$, are $n \times n$ matrices whose entries are i.i.d. standard complex Gaussians. The group of permutations of a set T will be denoted by $\mathcal{S}(T)$. When $T = \{1, 2, \ldots, k\}$ we just write $\mathcal{S}_k$. We denote the set $\{1, 2, \ldots, k\}$ by $[k]$. And $\mathcal{U}(N)$ is the group of $N \times N$ unitary matrices. Lastly, we write "$\overset{d}{\to}$" for convergence in distribution.

The results of this paper are motivated by the following two results, one from the realm of random matrices [4] and another concerning zeros of random analytic functions [17].

RESULT 1 (Ginibre [4]).   The set of eigenvalues of the $n \times n$ matrix G with i.i.d. standard complex Gaussian entries is a determinantal point process with ker-

nel

$$\mathbb{K}(z, w) = \sum_{k=0}^{n-1} \frac{(z\overline{w})^k}{k!} \tag{1}$$

with respect to the reference measure $d\mu(z) = \frac{1}{\pi} e^{-|z|^2}$. The corresponding Hilbert space $\mathbf{H} = \text{span}\{1, z, \ldots, z^{n-1}\} \subset L^2(\mathbb{C}, \frac{e^{-|z|^2}}{\pi} dm(z))$.

RESULT 2 (Peres and Virág [17]). Let $\mathbf{f}(z) = g_0 + g_1 z + g_2 z^2 + \cdots$ be the random analytic function (the radius of convergence is 1) whose coefficients g_i are i.i.d. standard complex Gaussians. Then the zeros of $\mathbf{f}$ form a determinantal point process on the unit disk $\mathbb{D}$ with the kernel (the Bergman kernel of the unit disk)

$$\mathbb{K}(z, w) = \frac{1}{\pi(1 - z\overline{w})^2}$$

with respect to the background measure $d\mu(z) = \frac{1}{\pi} dm(z)$ on $\mathbb{D}$. The corresponding Hilbert space $\mathbf{H} = \text{span}\{1, z, z^2, \ldots\} \subset L^2(\mathbb{D}, \frac{dm(z)}{\pi})$ is the space of all analytic functions in $L^2(\mathbb{D})$.

Knowing that a point process is determinantal greatly facilitates studying its properties. This motivates us to ask whether these two are isolated results or whether they are part of a bigger picture. As a start, consider the following random analytic functions whose zero sets are known to be determinantal:

- $z - g = 0$: One zero, g, with standard complex Gaussian distribution on $\mathbb{C}$. $\mathbf{H} = \text{span}\{1\}$ in $L^2(\mathbb{C}, \frac{e^{-|z|^2}}{\pi} dm(z))$. Determinantal, of course!
- $zg_1 - g_2 = 0$: One zero, $\frac{g_2}{g_1}$, that has density $\frac{1}{\pi(1+|z|^2)^2} dm(z)$. This is just the push-forward of the uniform measure on $\mathbb{S}^2 = \mathbb{C} \cup \{\infty\}$ under stereographic projection. Again, a one-point process is determinantal. In this case, $\mathbf{H} = \text{span}\{1\}$ in $L^2(\mathbb{C} \cup \{\infty\}, \frac{1}{\pi(1+|z|^2)^2} dm(z))$.
- $g_0 + zg_1 + z^2 g_2 + \cdots = 0$. This is the i.i.d. power series of Result 2.

Our key observation is that Ginibre's result (Result 1) describes the law of zeros of the random analytic function $\det(zI - G)$, which may in turn be thought of as the matrix version of the analytic function $z - g$, the first of the three examples above. This suggests that we consider the matrix versions of the second and third examples. This leads us to two families of random matrix-valued analytic functions.

1. $zG_1 - G_2$, where G_1, G_2 are $n \times n$ independent matrices with i.i.d. standard complex Gaussian entries.
2. $G_0 + zG_1 + z^2 G_2 + \cdots$, where G_k are independent $n \times n$ matrices with each G_k having i.i.d. standard complex Gaussian entries.

The analogy with Ginibre's result strongly suggests that the singular points of these matrix-valued analytic functions might be determinantal point processes. But how to guess which determinantal processes?

Invariance to the rescue: The key feature that allows us to guess which determinantal processes is invariance under a large group of transformations. A point process $\mathcal{X}$ on the space Ω is said to be invariant (in distribution) under a transformation $T : \Omega \to \Omega$ if $T(\mathcal{X}) \overset{d}{=} \mathcal{X}$.

- First, consider the matrix-valued analytic function $zG_1 - G_2$. We claim that the singular points (which are just the eigenvalues of $G_1^{-1}G_2$) are invariant under all linear fractional transformations

$$\lambda \to \frac{\lambda\alpha - \overline{\beta}}{\lambda\beta - \overline{\alpha}}$$

with α, β complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$. These are precisely rotations of the two-dimensional sphere, when the sphere is identified with $\mathbb{C} \cup \{\infty\}$ via stereographic projection.

Let $|\alpha|^2 + |\beta|^2 = 1$ and define

$$C = \alpha G_1 - \beta G_2 \quad \text{and} \quad D = \overline{\beta} G_1 + \overline{\alpha} G_2.$$

Then

$$\det(zC - D) = \det\big((z\alpha - \overline{\beta})G_1 - (z\beta - \overline{\alpha})G_2\big)$$

$$= (z\beta - \overline{\alpha})^n \det\left(\frac{z\alpha - \overline{\beta}}{z\beta - \overline{\alpha}}G_1 - G_2\right).$$

Thus the zeros of $\det(zC - D)$ are precisely $\{\frac{\lambda_i\alpha - \overline{\beta}}{\lambda_i\beta - \overline{\alpha}}\}$, where $\{\lambda_i\}$ are the zeros of $\det(zG_1 - G_2)$. On the other hand, (C, D) has the same distribution as (G_1, G_2). This implies that

(2)
$$\left\{\frac{\lambda_i\alpha - \overline{\beta}}{\lambda_i\beta - \overline{\alpha}}\right\}_{1 \le i \le n} \overset{d}{=} \{\lambda_i\}_{1 \le i \le n}$$

for any α, β with $|\alpha|^2 + |\beta|^2 = 1$ which is the claimed invariance.
- Next consider the matrix-valued analytic function $M(z) := \sum_{k=0}^{\infty} G_k z^k$, where G_k are independent random matrices with i.i.d. standard complex Gaussian entries. We claim that the set of singular points is invariant in distribution under the isometries of the hyperbolic plane, namely the linear fractional transformations

$$\varphi(\lambda) = \frac{\alpha\lambda + \beta}{\overline{\beta}\lambda + \overline{\alpha}}, \qquad |\alpha|^2 - |\beta|^2 = 1,$$

that map the unit disk injectively onto itself. These are precisely the conformal automorphisms of the unit disk.

Let $M(z) = (M_{i,j}(z))$, so that $M_{i,j}$ are i.i.d. copies of the random power series of Result 2. It was discovered by Leboeuf [13] (or see [18]) that

$$(3) \qquad\qquad M_{1,1}(\varphi(\cdot)) \overset{d}{=} \varphi'(\cdot)^{-1/2} M_{1,1}(\cdot).$$

This is because $M_{1,1}$ is the Gaussian element of the Hilbert space of analytic functions with the boundary inner product $\int_0^{2\pi} f(e^{i\theta})\overline{g}(e^{i\theta})\frac{d\theta}{2\pi}$ (first define this inner product for analytic functions that extend continuously to the boundary, and then take the completion). On this Hilbert space, $f \to \varphi'(\cdot)^{-1/2} f(\varphi(\cdot))$ is a unitary transformation. When a unitary transformation is applied to the Gaussian element of the Hilbert space, we get again the Gaussian element of the Hilbert space, yielding (3) (of course, the Gaussian element is itself not an element of the Hilbert space, almost surely).

An alternate way is to just check that the centered Gaussian processes on the two sides of (3) have the same covariance kernel $\frac{1}{1-\varphi(z)\overline{\varphi(w)}}$. Anyhow, the independence of distinct $M_{i,j}$ and the fact that φ' is nonrandom shows that

$$\det(M(\varphi(\cdot))) \overset{d}{=} \varphi'(\cdot)^{-n/2} \det(M(\cdot)).$$

Since φ' is a nowhere vanishing analytic function on the unit disk it follows that the singular set of $M(\cdot)$ is invariant in distribution under the action of hyperbolic isometries.

These are two special cases of a large class of invariant zero sets introduced in [12]. The general situation is that one applies a homogeneous polynomial of several complex variables ("det" in our case) to a bunch of i.i.d. copies of a Gaussian analytic function ($zg_1 - g_2$ or $g_0 + zg_1 + z^2 g_2 + \cdots$ in the two cases). If the individual Gaussian analytic functions have invariant zero sets, then so will the homogeneous polynomial of copies of them. Thus constructions (applying a homogeneous polynomial to i.i.d. copies of an analytic function) that are simple at the level of functions are not simple at all at the level of zeros and may give something drastically new.

Let us return to our original question, which led to a digression into the issue of invariance: What determinantal processes might the singular sets of the two families of random matrix-analytic functions ($zG_1 - G_2$ and $G_0 + zG_1 + z^2 G_2 + \cdots$) be? Now we have obtained a strong restriction: *The determinantal process must be isometry-invariant in $\mathbb{S}^2$ or $\mathbb{D}$, respectively.* Further, from the fact that we are looking at zeros of analytic functions, we expect that these determinantal processes will be defined by Hilbert spaces of analytic functions. Such determinantal processes were studied first by Caillol [1] under the name "one-component plasma on the sphere" and by Jancovici and Téllez [8] on the hyperbolic plane. The underlying Hilbert spaces here, the Bargmann–Fock spaces, also came up in the context of quantum Hall effect [5] and in the earlier mentioned work of Leboeuf [13]. In [12] (see Theorem 3.0.5 therein) we showed that, under certain restrictive assumptions, these are the only possible invariant determinantal processes.

1. On the sphere ($\mathbb{C} \cup \{\infty\}$) we have for each $n \in \{1, 2, \ldots\}$ an invariant determinantal point process with kernel

$$\mathbb{K}(z, w) = (1 + z\overline{w})^{n-1}$$

with respect to the background measure $d\mu_n(z) = \frac{n}{\pi(1+|z|^2)^{n+1}}\, dm(z)$. Invariance of the point process under an analytic transformation T is equivalent to saying that the joint intensities (correlation functions) with respect to Lebesgue measure satisfy

$$\rho_k(Tz_1, \ldots, Tz_k)|T'(z_1)|^2 \cdots |T'(z_k)|^2 = \rho_k(z_1, \ldots, z_k).$$

For the case at hand, this is easily checked from the fact

$$\frac{T'(z)\overline{T'(w)}}{(1 + T(z)\overline{T(w)})^2} = \frac{1}{(1 + z\overline{w})^2}.$$

The parameter n is the total number of points in the point process, or equivalently, it denotes the first intensity of the point process with respect to the spherical area measure $\frac{1}{\pi(1+|z|^2)^2}$.

2. On the unit disk, we have for each $n > 0$, an invariant determinantal point process with kernel

$$\mathbb{K}(z, w) = \frac{1}{(1 - z\overline{w})^{n+1}}$$

with respect to the background measure $d\mu_n(z) = \frac{n}{\pi}(1 - |z|^2)^{n-1}\, dm(z)$. Again it is easy to check that these are invariant, now using

$$\frac{\varphi'(z)\overline{\varphi'(w)}}{(1 - \varphi(z)\overline{\varphi(w)})^2} = \frac{1}{(1 - z\overline{w})^2}.$$

The parameter n denotes the first intensity of the point process with respect to the hyperbolic measure $\frac{1}{\pi(1-|z|^2)^2}$.

Thus, on the sphere and the disk, we have a family of invariant singular sets and a family of invariant determinantal processes, respectively. Then by comparing the first intensities of these determinantal processes and the set of singular points of our matrix-analytic functions, we match the singular sets to determinantal processes.

2. Statements of results. We now state our results.

THEOREM 3. *Let G_1, G_2 be i.i.d. $n \times n$ matrices with i.i.d. standard complex Gaussian entries. The zeros of $\det(zG_1 - G_2)$ form a determinantal point process on $\mathbb{S}^2$ with kernel*

$$\mathbb{K}(z, w) = (1 + z\overline{w})^{n-1}$$

with respect to the background measure $d\mu_n(z) = \frac{n}{\pi}\frac{dm(z)}{(1+|z|^2)^{n+1}}$. *Equivalently, we may say that the defining Hilbert space is the subspace of analytic functions in* $L^2(\mathbb{C} \cup \{\infty\}, \mu_n)$.

THEOREM 4. *Let* G_k *be i.i.d.* $n \times n$ *matrices with i.i.d. standard complex Gaussian entries. Then for each* $n \geq 1$, *the zeros of* $\det(G_0 + zG_1 + z^2 G_2 + \cdots)$ *form a determinantal point process on* $\mathbb{D}$ *with kernel*

$$\mathbb{K}(z, w) = \frac{1}{(1 - z\overline{w})^{n+1}}$$

with respect to the background measure $d\mu_n(z) = \frac{n}{\pi}(1 - |z|^2)^{n-1}\, dm(z)$. *Equivalently, we may say that the defining Hilbert space is the subspace of analytic functions in* $L^2(\mathbb{D}, \mu_n)$.

Theorem 3 is proved in Section 3 via the Schur decomposition of the matrix $G_1^{-1}G_2$, along the lines of Ginibre's proof of Result 1. Theorem 4 will be proved in Section 5 as a corollary of the following more general theorem which appears to be of potential interest beyond the specific application to Theorem 4. Theorem 5 is proved in Section 4.

THEOREM 5. *Fix* $n \geq 1$ *and let* A_N *be a sequence of* $n \times n$ *matrices such that* $\sqrt{N} A_N \xrightarrow{d} X_0$ *as* $N \to \infty$, *for some random matrix* X_0. *Independently of* A, *pick* P, Q *independent matrices chosen from Haar measure on* $\mathcal{U}(N)$ *and define the* $N \times N$ *matrix* V *by*

$$(4) \qquad\qquad V = Q^* \begin{bmatrix} A_N & 0 \\ 0 & I_{N-n} \end{bmatrix} P^*.$$

Set $\mathbf{f}_N(z) := \frac{\det(zI+V)}{\det(I+zV^*)}$. *Let* X_0 *and* G_i, $i \geq 1$, *be independent* $n \times n$ *random matrices, where* G_i *have independent standard complex Gaussian entries. Then*

$$N^{n/2}\mathbf{f}_N(z) \xrightarrow{d} \det\left(X_0 + \sum_{k\geq 1} G_k z^k\right)$$

in the sense that any finite set of coefficients in the power series expansion of $N^{n/2}\mathbf{f}_N(\cdot)$ *converge jointly in distribution to the corresponding vector of coefficients in the power series expansion of the right-hand side.*

The relevance of this theorem to Theorem 4 is through a result of Życzkowski and Sommers [20], who found random matrix models whose eigenvalue distributions are determinantal processes with kernels that are truncated versions of the kernels in Theorem 4. Theorem 5 gives the distribution of the limiting random analytic function as the matrix size increases, while the result of Życzkowski and Sommers gives the limiting distribution of zeros. Putting the two together we get the distribution of zeros of the limiting random analytic function.

REMARK 6. The statements of Theorems 3 and 4 may already be found in the thesis [12]. Theorem 3 appeared there with a proof but is being published here for the first time. Theorem 4 was conjectured in [12] and a partial proof was given, showing that the first and second joint intensities (correlation functions) of the singular set of $G_0 + zG_1 + z^2G_2 + \cdots$ are as claimed. While in [12] we tried to prove Theorem 4 by starting with the random matrix-valued analytic function and then finding the distribution of its zeros, in this paper we take the opposite direction.

REMARK 7. It is natural to ask whether there are other (perhaps even many) singular sets that are also determinantal. Without claiming that there are not, we would like to emphasize that the determinantal processes in Result 1 together with those in Theorems 3 and 4 are *the most natural* determinantal point processes in the three canonical surfaces of constant curvature, namely, the plane, the sphere and the hyperbolic plane, respectively.

3. Spherical ensembles. Let $\mathcal{X}$ denote the set of singular points of $zG_1 - G_2$. Since the number of points is exactly n, Theorem 3 is equivalent (see the facts stated after Definition A.1 in the Appendix) to saying that the joint density of the singular points is proportional to

$$\prod_{i<j} |z_i - z_j|^2 \prod_{k=1}^{n} \frac{1}{(1 + |z_k|^2)^{n+1}}.$$

(If $\{P_1, \ldots, P_n\}$ are the points on the two-dimensional sphere obtained by stereographic projection of $z_1, \ldots, z_n$, then the density of these points with respect to Lebesgue measure on $(\mathbb{S}^2)^n$ turns out to be simply $\prod_{i<j} \|P_i - P_j\|_{\mathbb{R}^3}^2$, where $\|\cdot\|_{\mathbb{R}^3}$ is the Euclidean norm in $\mathbb{R}^3$.)

The following lemma will greatly simplify the job of integrating out auxiliary variables later.

LEMMA 8. *Let $\mathcal{X}$ be a point process on $\mathbb{C}$ with n points almost surely. Assume that the vector of points (in uniform random order) has density*

$$p(z_1, \ldots, z_n) = |\Delta(z_1, \ldots, z_n)|^2 V(|z_1|^2, \ldots, |z_n|^2).$$

Here $\Delta(z_1, \ldots, z_n)$ denotes the Vandermonde factor $\prod_{i<j}(z_j - z_i)$.

Suppose also that $\mathcal{X}$ has a distribution invariant under isometries of the sphere $\mathbb{S}^2$, that is, under the transformations $\varphi_{\alpha,\beta}(z) = \frac{\alpha z + \beta}{-\bar{\beta}z + \bar{\alpha}}$, for any α, β satisfying $|\alpha|^2 + |\beta|^2 = 1$. Then

$$(5) \qquad V(|z_1|^2, \ldots, |z_n|^2) = \text{Const.} \prod_{k=1}^{n} \frac{1}{(1 + |z_k|^2)^{n+1}}.$$

PROOF.   The claim is that the probability density of the n points of $\mathcal{X}$ (in uniform random order) with respect to Lebesgue measure is

$$q(z_1, \ldots, z_n) := \text{Const.} \, |\Delta(z_1, \ldots, z_n)|^2 \prod_{k=1}^{n} \frac{1}{(1 + |z_k|^2)^{n+1}}.$$

First let us check that the density q is invariant under the isometries of $\mathbb{S}^2$. For this let $\varphi = \varphi_{\alpha,\beta}$. Then,

$$(6) \qquad \qquad \varphi'(z) = \frac{1}{(-\overline{\beta}z + \overline{\alpha})^2},$$

$$(7) \qquad \qquad 1 + |\varphi(z)|^2 = \frac{1 + |z|^2}{|-\overline{\beta}z + \overline{\alpha}|^2},$$

$$(8) \qquad \qquad \varphi(z) - \varphi(w) = \frac{z - w}{(-\overline{\beta}z + \overline{\alpha})(-\overline{\beta}w + \overline{\alpha})}.$$

From (6), (7) and (8), it follows that

$$(9) \qquad \qquad q(\varphi(z_1), \ldots, \varphi(z_n)) \prod_{k=1}^{n} |\varphi'(z_k)|^2 = q(z_1, \ldots, z_n),$$

which shows the invariance of q.

Invariance of $\mathcal{X}$ under the transformation φ means that for every $z_1, \ldots, z_n$, we have

$$(10) \qquad \qquad p(\varphi(z_1), \ldots, \varphi(z_n)) \prod_{k=1}^{n} |\varphi'(z_k)|^2 = p(z_1, \ldots, z_n).$$

Then $W(z_1, \ldots, z_n) := \frac{p(z_1, \ldots, z_n)}{q(z_1, \ldots, z_n)}$ has the following properties:

- $W(z_1, \ldots, z_n)$ is a function of $|z_1|^2, \ldots, |z_n|^2$ only, by the assumption on p and the definition of q.
- $W(\varphi(z_1), \ldots, \varphi(z_n)) = W(z_1, \ldots, z_n)$ for every $z_1, \ldots, z_n$ from (10) and (9).

We claim that these two statements imply that W is a constant. To see this fix $z_1 = r_1 > 0$ and $z_k = r_k e^{i\theta_k}$ for $2 \leq k \leq n$. Let $\alpha = \frac{1}{\sqrt{1+r_1^2}}, \beta = -\frac{r_1}{\sqrt{1+r_1^2}}$. Then $|\alpha|^2 + |\beta|^2 = 1$ and hence $\varphi = \varphi_{\alpha,\beta}$ is an isometry of $\mathbb{S}^2$. From the above-stated properties of W, we deduce

$$W(z_1, \ldots, z_n) = W(\varphi(z_1), \ldots, \varphi(z_n))$$

$$= W\left(0, \frac{z_2 - z_1}{1 + z_2 \overline{z}_1}, \ldots, \frac{z_n - z_1}{1 + z_n \overline{z}_1}\right)$$

$$= W\left(0, \left|\frac{r_2 e^{i\theta_2} - z_1}{1 + r_2 e^{i\theta_2} \overline{z}_1}\right|, \ldots, \left|\frac{r_n e^{i\theta_n} - z_1}{1 + r_n e^{i\theta_n} \overline{z}_1}\right|\right).$$

Take $z_1 = 1$ and $1 < r_k < 1 + \varepsilon$. Then as θ_k, $2 \le k \le n$ vary independently over $[0, 2\pi]$, the quantities $|\frac{r_k e^{i\theta_k} - z_1}{1 + r_k e^{i\theta_k} \bar{z}_1}|$ vary over the intervals $[\frac{r_k - 1}{r_k + 1}, \frac{r_k + 1}{r_k - 1}]$. However, even as θ_ks vary, the quantity $W(z_1, \ldots, z_n)$ does not change because W is a function of r_ks only. By our choice of r_ks, this means that

$$W(0, t_2, \ldots, t_n) = \text{Const.} \qquad \forall t_k \in \left[\varepsilon, \frac{1}{\varepsilon}\right].$$

ε is arbitrary, hence $W(0, t_2, \ldots, t_n)$ is constant. This implies that $W(0, z_2, \ldots, z_n)$ is constant and therefore $W(z_1, \ldots, z_n)$ is constant.

This shows that $p(z_1, \ldots, z_n) = \text{Const.}\, q(z_1, \ldots, z_n)$. $\square$

PROOF OF THEOREM 3.   Recall (2) which asserts that $\mathcal{X}$ is invariant in distribution under the action of automorphisms of $\mathbb{S}^2$. By Lemma 8, it suffices to show that the density of points in $\mathcal{X}$ is of the form given in (5). We use the following well-known matrix decomposition (see the book [6], Section 2.3).

Schur decomposition: Any diagonalizable matrix $M \in GL(n, \mathbb{C})$ can be written as

$$(11) \qquad\qquad M = U(Z + T)U^*,$$

where U is unitary, T is strictly upper triangular and Z is diagonal. Moreover, if the eigenvalues of M are distinct, then the decomposition is almost unique in the following sense:

$M = V(W + S)V^*$ in addition to (11), with V, S, W being, respectively, unitary, strictly upper triangular, and diagonal, if and only if the following hold:

1. The entries of W are a permutation of the elements of Z.
2. If this permutation is identity, then $V = U\Theta$ and $\Theta S \Theta^* = T$ for some Θ that is both diagonal and unitary, that is, for Θ of the form $\text{Diagonal}(e^{i\theta_1}, \ldots, e^{i\theta_n})$.

Fix an arbitrary total order on $\mathbb{C}^n$ (e.g., lexicographic order). Then for any M with distinct eigenvalues, there is a unique diagonal matrix Z with $z_1 < z_2 < \cdots < z_n$ and a unitary U and strictly upper triangular T that are unique up to $(U, T) \rightarrow (U\Theta, \Theta^* T \Theta)$, such that (11) holds.

Corresponding to the matrix decomposition (11) there is the following measure decomposition due to Ginibre [4] and Dyson (see [15], Appendix 35). Note that the eigenvalues of M are a.s. [Lebesgue] distinct, whence an almost unique decomposition exists.

Ginibre's measure decomposition: If M is decomposed as in (11), with the elements of Z in a uniformly randomly chosen order, then

$$(12) \qquad \prod_{i,j} dm(M_{ij}) = \left(\prod_{i<j} |z_i - z_j|^2 \prod_k dm(z_k)\right)\left(\prod_{i<j} dm(T_{ij})\right) d\nu(U),$$

where ν is the Haar measure on the on the unitary group $\mathcal{U}(n)$. Equivalently, if we choose (Z, T, U) according to the product measure on the right of (12), and *define* M by (11), then M has the measure on the left of (12).

Conditional on G_1, the matrix $M := G_1^{-1} G_2$ has the density

$$e^{-\operatorname{tr}(M^* G_1^* G_1 M)} |\det(G_1)|^{2n}$$

with respect to the Lebesgue measure on $GL(n, \mathbb{C}) \subset \mathbb{C}^{n^2}$. From the measure decomposition (12) we get the density of Z, T, U, G_1 to be

$$\left(\prod_{i<j} |z_i - z_j|^2 \prod_{k=1}^{n} dm(z_k) \right) e^{-\operatorname{tr}(G_1^* G_1 (I + M M^*))} |\det(G_1)|^{2n}$$

with respect to the measure $d\nu(U) \prod_{i<j} dm(T_{ij}) \prod_{i,j} dm(G_1(i,j))$ (we have omitted constants entirely—they can be recovered at the end). Thus the density of Z is obtained by integrating over T, U, G_1. Now write $z_k = r_k e^{i\theta_k}$ so that $Z = \Theta R$ where $\Theta = \operatorname{Diagonal}(e^{i\theta_1}, \dots, e^{i\theta_n})$ and $R = \operatorname{Diagonal}(r_1, \dots, r_n)$. Then

$$M M^* = U \Theta (R + \Theta^* T)(R + \Theta^* T)^* \Theta^* U^*.$$

As ν is the Haar measure, $d\nu(U\Theta) = d\nu(U)$. The elements of $\Theta^* T$ are the same as elements of T, but multiplied by complex numbers of absolute value 1. Hence, $\Theta^* T$ has the same "distribution" as T. Thus replacing U by $\Theta^* U$ and T by $\Theta^* T$ we see that the density of Z is of the form $V(r_1, \dots, r_n) \prod_{i<j} |z_i - z_j|^2$. This is the form of the density required to apply Lemma 8. Thus we conclude that the eigenvalue density is

$$(13) \qquad \operatorname{Const.} \prod_{i<j} |z_i - z_j|^2 \prod_{k=1}^{n} \frac{1}{(1 + |z_k|^2)^{n+1}}.$$

To compute the constant, note that

$$\left\{ \sqrt{\frac{n}{\pi} \binom{n-1}{k}} \frac{z^k}{(1 + |z|^2)^{(n+1)/2}} \right\}_{0 \le k \le n-1}$$

is an orthonormal set. Projection on the Hilbert space generated by these functions gives a determinantal process whose kernel is as given in the statement of the theorem. $\square$

4. Proof of Theorem 5. We first find the coefficients in the power series expansion of f_N prior to taking limits using the following lemma. Randomness plays no role here.

LEMMA 9. *Let V be an $N \times N$ matrix and define $f(z) = \frac{\det(zI + V)}{\det(I + z V^*)}$. Then*

$$f^{(k)}(0) = \det(V) \sum_{\pi \in \mathcal{S}_k} \operatorname{sgn}(\pi) \prod_{c \in \pi} \left[\operatorname{Tr}(V^{-|c|}) - \operatorname{Tr}(V^{*|c|}) \right],$$

where we write $c \in \pi$ to mean that c is a cycle of π.

PROOF.   Let $\chi(z) = \det(I + zV^{-1})$ and let $\psi(z) = \det(I + zV^*)$. Then $f(z) = \frac{\det(V)\chi(z)}{\psi(z)}$. Hence,

$$(14) \qquad f^{(k)}(z) = \det(V) \sum_{p=0}^{k} \binom{k}{p} \chi^{(k-p)}(z) \left(\frac{1}{\psi}\right)^{(p)}(z).$$

First let us find the derivatives of ψ and χ. Let $V_{[j_1,\ldots,j_k]}$ be the $k \times k$ matrix obtained from V by deleting all rows and columns except the $j_1, \ldots, j_k$th ones.

$$\psi^{(k)}(0) = k! \sum_{j_1 < \cdots < j_k} \det\left(V^*_{[j_1,\ldots,j_k]}\right)$$

$$= \sum_{(j_1,\ldots,j_k)} \det\left(V^*_{[j_1,\ldots,j_k]}\right) \qquad \text{(summand vanishes if } j_1 = j_2\text{)}$$

$$= \sum_{(j_1,\ldots,j_k)} \sum_{\pi \in \mathscr{S}_k} \mathrm{sgn}(\pi) \prod_{i=1}^{k} V^*_{j_i, j_{\pi(i)}}$$

$$= \sum_{\pi \in \mathscr{S}_k} \mathrm{sgn}(\pi) \sum_{(j_1,\ldots,j_k)} \prod_{i=1}^{k} V^*_{j_i, j_{\pi(i)}}.$$

The inner sum factors over cycles of π. Let us write $c \in \pi$ to mean that c is a cycle of π and let $|c|$ denote the size of the cycle c. Then we may write

$$(15) \qquad \psi^{(k)}(0) = \sum_{\pi \in \mathscr{S}_k} \mathrm{sgn}(\pi) \prod_{c \in \pi} \mathrm{Tr}(V^{*|c|}).$$

Analogously, we have

$$(16) \qquad \chi^{(k)}(0) = \sum_{\pi \in \mathscr{S}_k} \mathrm{sgn}(\pi) \prod_{c \in \pi} \mathrm{Tr}(V^{-|c|}).$$

To compute the derivatives of f using (14), we need the derivatives of $\varphi := 1/\psi$ at 0. These will be given by the sequence that we shall provisionally call $\{b_k\}$. Set $b_0 = 1$ and for $k \geq 1$ define

$$b_k = \sum_{\pi \in \mathscr{S}_k} \mathrm{sgn}(\pi) \prod_{c \in \pi} [-\mathrm{Tr}(V^{*|c|})].$$

Then for any $k \geq 1$ we calculate using (15)

$$\sum_{j=0}^{k} \binom{k}{j} b_{k-j} \psi^{(j)}(0)$$

$$= \sum_{T \subset [k]} \left(\sum_{\pi \in \mathscr{S}(T^c)} \mathrm{sgn}(\pi) \prod_{c \in \pi} \mathrm{Tr}(V^{*|c|}) \right) \left(\sum_{\pi \in \mathscr{S}(T)} \mathrm{sgn}(\pi) \prod_{c \in \pi} [-\mathrm{Tr}(V^{*|c|})] \right).$$

Fix a subset $T \subset [k]$. A permutation of T and a permutation of T^c together give a permutation of $[k]$. Let $\pi = c_1 \cdots c_l$ be a permutation of $[k]$. Then it can arise from summands in which T is a union (possibly empty) of some of the cycles c_is. Thus for $k \geq 1$

$$\sum_{j=0}^{k} \binom{k}{j} b_{k-j} \psi^{(j)}(0) = \sum_{\substack{\pi \in \mathcal{S}_k \\ \pi = c_1 \cdots c_l}} \mathrm{sgn}(\pi) \sum_{B \subset [l]} \prod_{i \in B} \mathrm{Tr}(V^{*|c_i|}) \prod_{i \in B^c} [-\mathrm{Tr}(V^{*|c_i|})]$$

$$= \sum_{\pi \in \mathcal{S}_k} \mathrm{sgn}(\pi) \prod_{c \in \pi} [\mathrm{Tr}(V^{*|c|}) - \mathrm{Tr}(V^{*|c|})]$$

$$= 0.$$

However, the equation $\varphi \cdot \psi = 1$ implies that

$$\sum_{j=0}^{k} \binom{k}{j} \varphi^{(k-j)}(0) \psi^{(j)}(0) = \begin{cases} 1, & \text{if } k = 0, \\ 0, & \text{if } k \neq 0. \end{cases}$$

It is also clear that from these equations one may inductively recover $\varphi^{(k)}(0)$ in terms of the derivatives of ψ. This shows that $\varphi^{(k)}(0) = b_k$. That is,

$$(17) \qquad \varphi^{(k)}(0) = \sum_{\pi \in \mathcal{S}_k} \mathrm{sgn}(\pi) \prod_{c \in \pi} [-\mathrm{Tr}(V^{*|c|})].$$

Now we return to the derivatives of f. From (14), (16) and (17) we deduce that

$$f^{(k)}(0) = \det(V) \sum_{p=0}^{k} \binom{k}{p} \chi^{(k-p)}(0) \varphi^{(p)}(0)$$

$$= \det(V) \sum_{T \subset [k]} \chi^{|T^c|}(0) \varphi^{|T|}(0)$$

$$= \det(V) \sum_{T \subset [k]} \left(\sum_{\pi \in \mathcal{S}(T^c)} \mathrm{sgn}(\pi) \prod_{c \in \pi} \mathrm{Tr}(V^{-|c|}) \right)$$

$$\times \left(\sum_{\pi \in \mathcal{S}(T)} \mathrm{sgn}(\pi) \prod_{c \in \pi} [-\mathrm{Tr}(V^{*|c|})] \right).$$

Just as before, a permutation of T and a permutation of T^c together give a permutation of $[k]$ and a permutation $\pi \in \mathcal{S}_k$ can arise from summands in which T is a union (possibly empty) of some of the cycles of π. Therefore

$$f^{(k)}(0) = \det(V) \sum_{\substack{\pi \in \mathcal{S}_k \\ \pi = c_1 \cdots c_l}} \mathrm{sgn}(\pi) \sum_{B \subset [l]} \prod_{i \in B} \mathrm{Tr}(V^{-|c_i|}) \prod_{i \in B^c} [-\mathrm{Tr}(V^{*|c_i|})]$$

$$= \det(V) \sum_{\pi \in \mathcal{S}_k} \mathrm{sgn}(\pi) \prod_{c \in \pi} [\mathrm{Tr}(V^{-|c|}) - \mathrm{Tr}(V^{*|c|})]. \qquad \square$$

The probabilistic part of the theorem comes from the following lemma on Haar-distributed unitary matrices:

LEMMA 10. *Let U be an $N \times N$ random unitary matrix sampled from the Haar measure. Fix $n \geq 1$. After multiplication by $\sqrt{N}$, the principal $n \times n$ sub-matrices of U^p, $p \geq 1$, converge in distribution to independent matrices with i.i.d. standard complex Gaussian entries. In symbols,*

$$\sqrt{N}([U]_{i,j \leq n}, [U^2]_{i,j \leq n}, \ldots) \overset{d}{\to} (G_1, G_2, \ldots),$$

where G_i are independent $n \times n$ matrices with i.i.d. standard complex Gaussian entries. More precisely, any finite number of random variables $\sqrt{N}[U^p]_{i,j}$, $p \geq 1$, $i, j \leq n$, converge in distribution to independent standard complex Gaussians.

In the literature, there are many results which are similar in spirit to Lemma 10. For instance, Diaconis and Shahshahani [3] (a slight mistake in that paper was corrected in Diaconis and Evans [2]) showed that if U is sampled from Haar measure on $\mathcal{U}(N)$, then $(\text{Tr}(U), \text{Tr}(U^2), \ldots) \overset{d}{\to} (g_1, \sqrt{2}g_2, \ldots)$. Jiang [9], answering a question of Diaconis, proved that if p_N, q_N are negligible compared to $\sqrt{N}$, then the entries of the principal $p_N \times q_N$ submatrix of a unitary random matrix U sampled from Haar measure on $\mathcal{U}(N)$ are approximately independent complex Gaussians. Our requirement is somewhere between the two. We need only submatrices of fixed size, but of all powers of U. We give a complete proof of Lemma 10 in Section 6.

PROOF OF THEOREM 5. Define **f** as in the statement of the theorem. Lemma 9 asserts that

$$(18) \qquad \mathbf{f}^{(k)}(0) = \det(V) \sum_{\pi \in \mathcal{S}_n} \text{sgn}(\pi) \prod_{c \in \pi} [\text{Tr}(V^{-|c|}) - \text{Tr}(V^{*|c|})].$$

We want to find the limit distribution of $\{\mathbf{f}^{(k)}(0) : 0 \leq k < \infty\}$. First let us consider $\text{Tr}(V^{-p}) - \text{Tr}(V^{*p})$ for $p \geq 1$. Setting $P = [P_1 : P_2]$ and $Q^* = [Q_1^* : Q_2^*]$, where P_1, Q_1^* are $N \times n$ matrices, from (4) we get

$$(19) \qquad V^{-1} = P_1 A^{-1} Q_1 + P_2 Q_2 \quad \text{and} \quad V^* = P_1 A^* Q_1 + P_2 Q_2.$$

Then write

$$\text{Tr}(V^{-p}) - \text{Tr}(V^{*p})$$
$$= \sum_{i_1, \ldots, i_p} (V^{-1})_{i_1, i_2} \cdots (V^{-1})_{i_p, i_1} - (V^*)_{i_1, i_2} \cdots (V^*)_{i_p, i_1}$$
$$= \sum_{i_1, \ldots, i_p} \prod_{j=1}^{p} (P_1 A^{-1} Q_1 + P_2 Q_2)_{i_j, i_{j+1}} - \prod_{j=1}^{p} (P_1 A^* Q_1 + P_2 Q_2)_{i_j, i_{j+1}}.$$

Here it is implied that $i_{p+1} = i_1$. Expand each of the products to get a sum of 2^p terms. Each of these terms is identified uniquely by an integer $0 \le r \le p$ and a vector $\mathbf{q} = (q_1, \ldots, q_r)$ of integers $q_1 < q_2 < \cdots < q_r$ which are the values of j for which we choose $(P_1 A^{-1} Q_1)_{i_j, i_{j+1}}$ [or $(P_1 A^* Q_1)_{i_j, i_{j+1}}$], while for other j we choose $(P_2 Q_2)_{i_j, i_{j+1}}$ in both products.

A most important observation is that all summands with $r = 0$ cancel. What remains is

$$
(20) \quad \sum_{i_1, \ldots, i_p} \sum_{r \ge 1} \sum_{\mathbf{q}} \prod_{j \notin \mathbf{q}} (P_2 Q_2)_{i_j, i_{j+1}} \left(\prod_{l=1}^{r} (P_1 A^{-1} Q_1)_{i_{q_l}, i_{q_l+1}} - \prod_{l=1}^{r} (P_1 A^* Q_1)_{i_{q_l}, i_{q_l+1}} \right).
$$

We are using $\mathbf{q}$ to denote the vector $(q_1, \ldots, q_r)$ as well as the set $\{q_1, \ldots, q_r\}$ but this should not lead to any confusion. Now write for each $l = 1, \ldots, r$

$$
(P_1 A^{-1} Q_1)_{i_{q_l}, i_{q_l+1}} = \sum_{\alpha_l=1}^{n} \sum_{\beta_l=1}^{n} (P_1)_{i_{q_l}, \alpha_l} (A^{-1})_{\alpha_l, \beta_l} (Q_1)_{\beta_l, i_{q_l+1}},
$$

$$
(P_1 A^* Q_1)_{i_{q_l}, i_{q_l+1}} = \sum_{\alpha_l=1}^{n} \sum_{\beta_l=1}^{n} (P_1)_{i_{q_l}, \alpha_l} (A^*)_{\alpha_l, \beta_l} (Q_1)_{\beta_l, i_{q_l+1}}.
$$

Fix a choice of $r \ge 1$, $\mathbf{q}$ and α_l, β_l, $1 \le l \le r$. Sum over $i_1, \ldots, i_p$ in (20). When we sum over i_j for $q_1 < j \le q_2$, in both the summands corresponding to A^{-1} and A^*, we get a factor of (we have displayed only those factors that depend on i_j for $q_1 < j \le q_2$)

$$
\sum_{i_{q_1+1}, \ldots, i_{q_2}} (Q_1)_{\beta_1, i_{q_1+1}} \left[\prod_{j=q_1+1}^{q_2-1} (P_2 Q_2)_{i_j, i_{j+1}} \right] (P_1)_{i_{q_2}, \alpha_2}
$$

$$
= (Q_1 (P_2 Q_2)^{q_2 - q_1 - 1} P_1)_{\beta_1, \alpha_2}.
$$

Similarly we sum over i_j for j between $q_l + 1$ to q_{l+1} for every l (where $r + 1 = 1$). Write $\lambda_l = q_{l+1} - q_l - 1$ for $l \le r - 1$ and $\lambda_r = r - q_r + q_1 - 1$. Then for a fixed value of $r \ge 1$, $\mathbf{q}$ and α_l, β_l, $1 \le l \le r$, as we sum over all i_js in (20) we get

$$
\left[\prod_{l=1}^{r} (A^{-1})_{\alpha_l, \beta_l} - \prod_{l=1}^{r} (A^*)_{\alpha_l, \beta_l} \right] \cdot \prod_{l=1}^{r} (Q_1 (P_2 Q_2)^{\lambda_l - 1} P_1)_{\beta_l, \alpha_{l+1}}.
$$

Any choice of $(\lambda_1, \ldots, \lambda_r)$ comes from p different choices of $\mathbf{q}$ [by cyclically

rotating $(q_1, \ldots, q_r)$]. Therefore $\mathrm{Tr}(V^{-p}) - \mathrm{Tr}(V^{*p})$ is equal to

$$
p \sum_{\substack{r \geq 1, (\lambda_1, \ldots, \lambda_r) \\ \lambda_i \geq 1, \sum \lambda_i = p \ \ l \leq r}} \sum_{\alpha_l, \beta_l} \left[\prod_{l=1}^{r} (A^{-1})_{\alpha_l, \beta_l} - \prod_{l=1}^{r} (A^*)_{\alpha_l, \beta_l} \right]
$$

(21)

$$
\times \prod_{l=1}^{r} (Q_1 (P_2 Q_2)^{\lambda_l - 1} P_1)_{\beta_l, \alpha_{l+1}}.
$$

As before, here $r + 1 = 1$. Since PQ has Haar distribution, from Lemma 10 and the assumption on A, we know that

$$
\sqrt{N} (A, Q_1 (PQ)^0 P_1, Q_1 (PQ)^1 P_1, Q_1 (PQ)^2 P_1, \ldots) \xrightarrow{d} (X_0, G_1, G_2, \ldots),
$$

where G_i, $i \geq 1$ are independent $n \times n$ matrices with i.i.d. standard complex Gaussian entries and independent of X_0. [Pre-multiplication by Q_1 and post-multiplication by P_1 serve to pick out the first $n \times n$ principal sub-matrix of $(QP)^m$.]

Now consider $Q_1 (PQ)^m P_1$. For $m = 0$,

$$
Q_1 (PQ)^0 P_1 = Q_1 (P_2 Q_2)^0 P_1.
$$

Hence $\sqrt{N} Q_1 (P_2 Q_2)^0 P_1 \xrightarrow{d} G_1$. Next take $m = 1$. Since $PQ = P_1 Q_1 + P_2 Q_2$,

$$
Q_1 (PQ)^1 P_1 = Q_1 (P_2 Q_2)^1 P_1 + (Q_1 P_1)^2.
$$

From the $m = 0$ case, we know that the second summand is $O_p(N^{-1})$, whence, $\sqrt{N} (Q_1 (P_2 Q_2)^0 P_1, Q_1 (P_2 Q_2)^1 P_1) \xrightarrow{d} (G_1, G_2)$. Continuing inductively, for any m, we get $Q_1 (QP)^m P_1 = Q_1 (P_2 Q_2)^m P_1 + O_p(N^{-1})$. Thus

$$
\sqrt{N} (A, [Q_1 (P_2 Q_2)^0 P_1], [Q_1 (P_2 Q_2)^1 P_1], \ldots) \xrightarrow{d} (X_0, G_1, G_2, \ldots)
$$

in the sense that any finite subset of random variables on the left converge in distribution to the corresponding random variables on the right.

In (21) divide each of the r factors in the products inside the brackets by $\sqrt{N}$ and multiply each factor in the product outside the brackets by $\sqrt{N}$. A^* itself converges to 0 in probability and thus after dividing by $\sqrt{N}$; in the first product only A^{-1} survives in the limit. Thus we get

$$
\mathrm{Tr}(V^{-p}) - \mathrm{Tr}(V^{*p})
$$

$$
\xrightarrow{d} p \sum_{\substack{r \geq 1 \\ \lambda_i \geq 1, \sum \lambda_i = p}} \sum_{(\lambda_1, \ldots, \lambda_r)} \sum_{\alpha_l, \beta_l, l \leq r} \prod_{l=1}^{r} X_0^{-1}(\alpha_l, \beta_l) \cdot \prod_{l=1}^{r} G_{\lambda_l}(\beta_l, \alpha_{l+1})
$$

$$
= p \sum_{\substack{r \geq 1 \\ \lambda_i \geq 1, \sum \lambda_i = p}} \sum_{(\lambda_1, \ldots, \lambda_r)} \sum_{\alpha_l, l \leq r} \prod_{l=1}^{r} (X_0^{-1} G_{\lambda_l})(\alpha_l, \alpha_{l+1}).
$$

Use this in (18) and observe that $\det(V) = \det(A)\det(P^*Q^*) = \det(A)e^{i\theta}$, where θ is uniform on $[0, 2\pi]$ and independent of A. Absorb $e^{i\theta}$ into X_0 and denote $H_p = X_0^{-1}G_p$. Then we see that $N^{n/2}\mathbf{f}^{(k)}(0)$ converges in distribution to (jointly for $k \geq 0$, of course)

$$(22) \quad \det(X_0) \sum_{\pi \in \mathscr{S}_k} \mathrm{sgn}(\pi) \prod_{c \in \pi} \left(|c| \sum_{r \geq 1} \sum_{\substack{(\lambda_1, \ldots, \lambda_r) \\ \lambda_i \geq 1, \sum \lambda_i = |c|}} \sum_{\alpha_l, l \leq r} \prod_{l=1}^{r} (X_0^{-1}G_{\lambda_l})(\alpha_l, \alpha_{l+1}) \right),$$

where $|c|$ is the number of elements in the cycle c. We must reduce this further. When we completely expand the products in (22) we see that the right-hand side is equal to (as usual $r_i + 1 = 1$)

$$(23) \quad \sum_{\substack{\pi \in \mathscr{S}_k \\ \pi = c_1 \cdots c_m}} \mathrm{sgn}(\pi) \prod_{i=1}^{m} |c_i| \sum_{\substack{r_i \geq 1 \\ 1 \leq i \leq m}} \sum_{\substack{(\lambda_1^i, \ldots, \lambda_{r_i}^i) \\ \lambda_j^i \geq 1, \sum_j \lambda_j^i = |c_i|}} \sum_{\substack{\alpha_l^i \leq n \\ l \leq r_i, i \leq m}} \prod_{i=1}^{m} \prod_{j=1}^{r_i} H_{\lambda_j^i}(\alpha_j^i, \alpha_{j+1}^i).$$

The point is that many of the terms $\prod_i \prod_j H_{\lambda_j^i}(\alpha_j^i, \alpha_{j+1}^i)$ can arise from more than one permutation π and thus there is a lot of cancellation. This we investigate now.

Consider any term $\prod_{l=1}^{L} H_{\mu_l}(s_l, t_l)$ where $L \geq 1$, $\mu_l \geq 1$ for each $l \leq L$ and $1 \leq s_l, t_l \leq n$. We compute the coefficient of such a term in (23).

To organize the combinatorics that will emerge, for the term $\prod_{l=1}^{L} H_{\mu_l}(s_l, t_l)$ let us associate a directed multi-graph with edge weights as follows: We assume that $\sum_{l=1}^{L} \mu_l = k$ as only such terms can appear in (23).

The graph will have vertices $\{1, 2, \ldots, n\}$. For each $l \leq L$, put a directed edge from s_l to t_l and give it weight μ_l. Let us also put self loops with edge weight 0 at each vertex $v \in \{1, \ldots, n\} \setminus \{s_l, t_l : l \leq L\}$. Let us call this graph $\mathscr{G}$ (depends on μ_l, s_l etc., of course, but it would be horrifying to include that dependence in the notation). We group terms together by the graph they generate and find the total contribution for each graph.

The graph $\mathscr{G}$ can arise from a term in (23) only if the edges of $\mathscr{G}$ can be partitioned into edge-disjoint directed cycles. Note that $\mathscr{G}$ is a multi-graph and hence if $i \to j$ occurs twice in $\mathscr{G}$, the two instances will occur in two distinct cycles, but the cycles will be deemed disjoint. Also, a cycle may visit the same vertex more than once.

Furthermore, each such decomposition of $\mathscr{G}$ into disjoint cycles corresponds to some (usually more than one) choice of the permutation π in (23). Once π is fixed, the numbers r_i are just the sizes of cycles in this cycle decomposition of $\mathscr{G}$ and λ_j^i, α_j^i are also determined. An example is given below to elucidate the matter.

EXAMPLE 11.   Suppose $n = 6$ and let $k = 11$. Suppose we look at the term

$$(24) \qquad\qquad H_1(1, 3)H_2(3, 2)H_2(2, 1)H_1(2, 1)H_3(1, 4)H_2(4, 2).$$

This term can actually arise in (23) because the sum of the μ_ls is equal to k and the associated graph may be decomposed into disjoint cycles in two distinct ways: First as $\{(1, 3, 2), (1, 4, 2)\}$ and second as $(1, 3, 2, 1, 4, 2)$.

The first case, $\{(1, 3, 2), (1, 4, 2)\}$, can arise from any permutation $\pi \in \mathcal{S}_{11}$ that has two cycles of lengths 5 and 6 (these numbers come from adding the edge weights in each cycle). There are $\binom{11}{5}4!5!$ such permutations and they all have sign -1. Taking into account the weight $\prod |c_i|$ in (23), the contribution to the term (24) from all such permutations is $-11!$.

The second case, $(1, 3, 2, 1, 4, 2)$, can arise from any $\pi \in \mathcal{S}_{11}$ that is itself a cycle of length 11. There are 10! such permutations and they all have sign $+1$. Their total contribution is $+11!$. When put together, we see that the coefficient of (24) is zero. This is no coincidence and prepares the reader for what is stated next in general.

CLAIM. *If $\mathcal{G}$ can be decomposed into disjoint cycles in more than one way, then the coefficient of the corresponding term in (23) is zero. On the other hand, if $\mathcal{G}$ can be decomposed in a unique way into ℓ disjoint cycles, then the coefficient of the corresponding term is $(-1)^{k-\ell}k!$.*

PROOF. Suppose $\mathcal{G}$ can be decomposed into disjoint cycles in more than one way. Then some vertex, say 1 without losing generality, belongs to more than one cycle of $\mathcal{G}$. Then in $\mathcal{G}$ there are in-edges $i_1, \ldots, i_M$ leading to 1 and out-edges $j_1, \ldots, j_M$ leading away from 1, for some $M \geq 2$. In any decomposition of $\mathcal{G}$ into cycles, we have the obvious matching of in-edges with out-edges, by associating to each in-edge the out-edge that follows it in the cycle. Consider any cycle decomposition of $\mathcal{G}$, and suppose that in-edges i_1, i_2 are matched with j_1, j_2, respectively. We pair this cycle decomposition with a new cycle decomposition obtained by switching the matches $i_1 \rightarrow j_1, i_2 \rightarrow j_2$ to $i_1 \rightarrow j_2, i_2 \rightarrow j_1$ and leaving everything else intact. This leads to a pairing of all cycle decompositions of $\mathcal{G}$. We show that the total contribution from each pair is zero.

One can go from one cycle decomposition to its pair by splitting a cycle into two cycles or merging two cycles into one. Let us take the first one among them to have cycle sizes $\theta_1, \ldots, \theta_\ell$ and the second one to have cycle sizes $\theta_1 + \theta_2, \theta_3, \ldots, \theta_\ell$. Let the sums of edge weights along cycles in the first decomposition be $w_1, w_2, \ldots, w_\ell$ so that in the second cycle decomposition the sums of edge weights of cycles are $w_1 + w_2, w_3, \ldots, w_\ell$.

The permutations in $\mathcal{S}_k$ that respect the first cycle decomposition of $\mathcal{G}$ are precisely those with ℓ cycles of sizes $w_1, \ldots, w_\ell$. The number of such permutations is

$$(25) \qquad \frac{k!}{\prod_{i=1}^{\ell} w_i!} \prod_{i=1}^{\ell} (w_i - 1)!.$$

Each of these comes with the weight $\prod_{i=1}^{\ell} w_i$ in (23), whence the total contribution of these terms is $(-1)^{k-\ell} k!$.

The number of permutations that respect the second cycle decomposition of $\mathcal{G}$ is

$$\frac{k!}{(w_1 + w_2)! \prod_{i=3}^{\ell} w_i!} (w_1 + w_2 - 1)! \prod_{i=3}^{\ell} (w_i - 1)!.$$

Each of these comes with the weight $(w_1 + w_2) \prod_{i=3}^{\ell} w_i$ in (23), whence the total contribution of these terms is $(-1)^{k-\ell+1} k!$.

Thus the two cycle decompositions cancel each other out exactly and it is seen that the total contribution is zero. This proves the first part of the claim.

For the second part, there is only one cycle decomposition by assumption, and the same calculations that led to (25) show that the coefficient is $(-1)^{k-\ell} k!$. This completes the proof of the claim. $\square$

Now consider a $\mathcal{G}$ that has a unique cycle decomposition. Then that cycle decomposition may be regarded as a permutation $\tau \in \mathcal{S}_n$, where all those vertices that do not occur among $\{s_l, t_l\}$ are fixed points of τ (this is why we added self-loops to all these vertices when defining $\mathcal{G}$). Recall that the edge-weights of these self-loops is 0. It will be convenient to set $H_0 = I_n$. Observe that $\text{sgn}(\tau) = (-1)^{n-\ell}$. Then using the claim above to simplify (23) we finally have

$$(26) \quad N^{n/2} \mathbf{f}^{(k)}(0) \xrightarrow{d} (-1)^{k-n} k! \det(X_0) \sum_{\tau \in \mathcal{S}_n} \text{sgn}(\tau) \sum_{\substack{(w_1, \ldots, w_n) \\ w_i \geq 0, \sum_i w_i = k}} \prod_{i=1}^{n} (H_{w_i})_{i, \tau_i}.$$

Here $H_p = X_0^{-1} G_p$ for $p \geq 1$ and $H_0 = I_n$. If we forget the $(-1)^{k-n}$ factor, the right-hand side of (26) is precisely $k!$ times the coefficient of z^k in the power series expansion of $\det(X_0 + z G_1 + z^2 G_2 + \cdots)$ as may be seen by expanding the determinant as

$$\det(X_0) \sum_{\tau \in \mathcal{S}_n} \text{sgn}(\tau) \prod_{i=1}^{n} (I + z H_1 + z^2 H_2 + \cdots)_{i, \tau(i)}.$$

The factor $(-1)^{k-n}$ is rendered irrelevant by changing z to $-z$ and multiplying the whole function by $(-1)^n$.

Thus we have proved that any the power series coefficients of $N^{n/2} \mathbf{f}_N$ converge jointly in distribution to the coefficients in the power series of $\det(G_0 + z G_1 + z^2 G_2 + \cdots)$, in the sense that any finite number of coefficients in the former converge jointly in distribution to the corresponding coefficients in the latter. This completes the proof of the theorem. $\square$

5. Hyperbolic ensembles. In this section we prove Theorem 4. We shall make use of the following result of Życzkowski and Sommers [20]:

RESULT 12 (Życzkowski and Sommers [20]). Let U be an $(N+n) \times (N+n)$ unitary matrix sampled from Haar measure on $\mathcal{U}(N+n)$. Let V be the $N \times N$ principal sub-matrix obtained by deleting any n rows and n columns of U. Then the eigenvalues of V form a determinantal process in the unit disk $\mathbb{D}$ with kernel

$$\mathbb{K}_N^{(n)}(z, w) = \sum_{k=0}^{N-1} \binom{-n-1}{k} (-1)^k z^k \overline{w}^k$$

with respect to the reference measure $d\mu_n(z) = \frac{n}{\pi}(1 - |z|^2)^{n-1}\, dm(z)$.

REMARK 13. If V is a matrix, then $\frac{\det(zI-V)}{\det(I-zV^*)}$ is just the Blaschke product of the eigenvalues $\{\lambda_k\}$ of V. That is

(27)
$$\frac{\det(zI - V)}{\det(I - zV^*)} = \prod_{j=1}^{k} \frac{z - \lambda_k}{1 - z\overline{\lambda}_k}.$$

That being the case, the function $\mathbf{f}_N$ in Theorem 5 depends only on the eigenvalues and not the matrix V that we choose. Why then, do we use the truncated unitary matrix of Result 12 instead of directly using the diagonal matrix whose entries are a determinantal process with the truncated kernels? It may indeed be possible to prove Theorem 4 directly from the properties of determinantal processes without having to use Result 12. However, that would involve proving a bevy of central limit theorems (of nonlinear statistics) for determinantal processes that can substitute Lemma 10. We do not know if that is easy.

The advantages of the truncated unitary matrix over the diagonal matrix of its eigenvalues are: (1) The former is invariant under left and right multiplication by unitary matrices (which allows us to apply the rather easy Lemma 10). (2) The truncated unitary matrix has only n nondeterministic singular values, even as the matrix size goes to infinity. The cost is that we use the (far from trivial) result of Życzkowski and Sommers but this has the positive value of forging a direct link between random matrices and random analytic functions.

Applying Theorem 5 to truncated unitary matrices we almost get Theorem 4, but there is one snag. Theorem 5 gives convergence of coefficients in the power series, whereas to deduce convergence of zeros we need uniform convergence on compact sets. For instance, in the sequence $f_n(z) = n^n z^n$, all the power series coefficients converge to zero but $f_n(z)$ does not converge for any $z \neq 0$. The following lemma, deduced directly from properties of determinantal point processes, will establish the required tightness, a priori.

LEMMA 14.   *Fix $n \geq 1$. Let $\{\lambda_k : 1 \leq k \leq N\}$, be determinantal on the unit disk with kernel $\mathbb{K}_N$ with respect to the measure μ_n as in Result 12. Set*

$$\mathbf{f}_N(z) = \prod_{k=1}^{N} \frac{z - \lambda_k}{1 - z\overline{\lambda}_k}.$$

Then for any compact subset $K \in \mathbb{D}$, the set $\{\mathbf{f}_N(z) : z \in K\}$ is tight, uniformly in N.

We postpone the proof of Lemma 14 to Section 7 and proceed to prove Theorem 4 assuming the lemma.

PROOF OF THEOREM 4.   Let U be an $(N + n) \times (N + n)$ unitary matrix and write

$$U = \begin{bmatrix} A & C^* \\ B & V \end{bmatrix},$$

where A has size $n \times n$. By the unitarity of U, we have the following equations:

$$A^*A + B^*B = I_n \quad \text{and} \quad BB^* + VV^* = I_N.$$

As BB^* and B^*B have the same nonzero eigenvalues, it follows that VV^* has the same eigenvalues as A^*A, except that it has $N - n$ more eigenvalues all equal to 1. Thus there must exist unitary matrices $P, Q \in \mathcal{U}(N)$ such that

(28) $$V = P \begin{bmatrix} A & 0 \\ 0 & I_{N-n} \end{bmatrix} Q.$$

Now suppose U was sampled according to Haar measure on $\mathcal{U}(N + n)$. Then for any unitary matrices $P_0, Q_0 \in \mathcal{U}(N)$, we have

$$\begin{bmatrix} I_n & 0 \\ 0 & P_0 \end{bmatrix} \begin{bmatrix} A & C^* \\ B & V \end{bmatrix} \begin{bmatrix} I_n & 0 \\ 0 & Q_0 \end{bmatrix} \overset{d}{=} \begin{bmatrix} A & C^* \\ B & V \end{bmatrix},$$

because Haar measure is invariant under left and right multiplication by group elements. This shows that $P_0 V Q_0 \overset{d}{=} V$, which, together with (28) implies that

$$V = P \begin{bmatrix} A & 0 \\ 0 & I_{N-n} \end{bmatrix} Q,$$

where P, Q, A are independent, P, Q are distributed according to Haar measure on $\mathcal{U}(N)$ and A is the principal $n \times n$ submatrix of an $(N + n) \times (N + n)$ unitary matrix.

Lemma 10 shows that $\sqrt{N} A \overset{d}{\to} G_0$ where G_0 is an $n \times n$ matrix of i.i.d. standard complex Gaussians. Thus Theorem 5 applies and we get

$$N^{n/2} \frac{\det(zI - V)}{\det(I - zV^*)} \overset{d}{\to} \det\left(\sum_{k=0}^{\infty} G_k z^k \right),$$

where all G_k, $k \geq 0$ are i.i.d. matrices of i.i.d. standard complex Gaussians.

This convergence is only in the sense of pointwise convergence of coefficients in the power series. But in case of truncated unitary matrices, Result 12 and Lemma 14 together strengthen it to uniform convergence on compact subsets of $\mathbb{D}$. Therefore the zeros of $\mathbf{f}_N$ converge in distribution to the zeros of the limiting analytic function.

The upshot is that the point process of eigenvalues of V (which are exactly the zeros of $\mathbf{f}_N$) converge in distribution to the zeros of $\det(G_0 + zG_1 + z^2 G_2 + \cdots)$. Use the result of Życzkowski and Sommers and let $N \to \infty$. The kernels $\mathbb{K}_N$ increase (in the sense of operators, i.e., the associated Hilbert spaces increase) to the kernel $\mathbb{K}(z, w) = (1 - z\overline{w})^{-n-1}$. From the facts stated after Definition A.1 in the Appendix, it follows that the determinantal process with kernel $\mathbb{K}_N$ converges to the determinantal process with kernel $\mathbb{K}$ and the proof is complete. $\square$

6. Proof of Lemma 10.   In proving Lemma 10, we shall make use of the following result on the joint moments of entries of a unitary matrix from the book of Nica and Speicher [16], page 381 (we state a weaker form suited to our purpose):

RESULT 15.   Let $U = ((u_{i,j}))_{i,j \leq N}$ be chosen from Haar measure on $\mathcal{U}(N)$. Let $k \leq N$ and fix $i(\ell), j(\ell), i'(\ell), j'(\ell)$ for $1 \leq \ell \leq k$. Then

$$\mathbf{E}\left[\prod_{\ell=1}^{k} u_{i(\ell),j(\ell)} \prod_{\ell=1}^{k} \overline{u}_{i'(\ell),j'(\ell)} \right] = \sum_{\pi,\sigma \in \mathcal{S}_k} \mathrm{Wg}(N, \pi\sigma^{-1}) \prod_{\ell=1}^{k} \mathbf{1}_{i(\ell)=i'(\prod_\ell)} \mathbf{1}_{j(\ell)=j'(\sigma_\ell)},$$

where Wg (called "Weingarten function") has the property that as $N \to \infty$,

$$\mathrm{Wg}(N, \tau) = \begin{cases} N^{-k} + O(N^{-k-1}), & \text{if } \tau = e \text{ ("identity")}, \\ O(N^{-k-1}), & \text{if } \tau \neq e. \end{cases}$$

PROOF OF LEMMA 10.   We want to show that $\sqrt{N}(U^k)_{\alpha,\beta}$, $k \geq 1$, $1 \leq \alpha$, $\beta \leq n$ converge (jointly) in distribution to independent standard complex Gaussians. To use the method of moments consider two finite products of these random variables

$$S = \prod_{i=1}^{m} [(U^{k_i})_{\alpha_i,\beta_i}]^{p_i} \quad \text{and} \quad T = \prod_{i=1}^{m'} [(U^{k'_i})_{\alpha'_i,\beta'_i}]^{p'_i},$$

where $m, m', p_i, p'_i, k_i, k'_i \geq 1$ and $1 \leq \alpha_i, \beta_i, \alpha'_i, \beta'_i \leq n$ are fixed. We want to find $\mathbf{E}[S\overline{T}]$ asymptotically as $N \to \infty$.

The idea is simple. We expand each $(U^k)_{\alpha,\beta}$ as a sum of products of entries of U. Then we get a huge sum of products and we evaluate the expectation of each product using Result 15. Among the summands that do not vanish, most have the same contribution and the rest are negligible. We now delve into the details.

Let $\mathcal{P}_k(\alpha, \beta)$ denote all "paths" γ of length k connecting α to β. This just means that $\gamma \in [N]^{k+1}$, $\gamma(1) = \alpha$ and $\gamma(k+1) = \beta$. Then we write

$$(U^k)_{\alpha,\beta} = \sum_{\gamma \in \mathcal{P}_k(\alpha,\beta)} \prod_{j=1}^{k} u_{\gamma(j),\gamma(j+1)}.$$

Expanding each factor in the definition of S like this, we get

$$S = \sum_{\substack{\gamma_i^\ell \in \mathcal{P}_{k_i}(\alpha_i,\beta_i) \\ i \leq m; \ell \leq p_i}} \prod_{i=1}^{m} \prod_{\ell=1}^{p_i} \prod_{j=1}^{k_i} u_{\gamma_i^\ell(j),\gamma_i^\ell(j+1)}.$$

In other words, we are summing over a packet of p_1 paths of length k_1 from α_1 to β_1, a packet of p_2 paths of length k_2 from α_2 to β_2, etc. T may similarly be expanded as

$$T = \sum_{\substack{\Gamma_i^\ell \in \mathcal{P}_{k_i'}(\alpha_i',\beta_i') \\ i \leq m'; \ell \leq p_i'}} \prod_{i=1}^{m'} \prod_{\ell=1}^{p_i'} \prod_{j=1}^{k_i'} u_{\Gamma_i^\ell(j),\Gamma_i^\ell(j+1)}.$$

To evaluate $\mathbf{E}[S\overline{T}]$, for each pair of collections $\gamma = \{\gamma_i^\ell\}$ and $\Gamma = \{\Gamma_i^\ell\}$, we must find

$$(29) \qquad \mathbf{E}\left[\left(\prod_{i=1}^{m} \prod_{\ell=1}^{p_i} \prod_{j=1}^{k_i} u_{\gamma_i^\ell(j),\gamma_i^\ell(j+1)}\right)\left(\prod_{i=1}^{m'} \prod_{\ell=1}^{p_i'} \prod_{j=1}^{k_i'} \overline{u}_{\Gamma_i^\ell(j),\Gamma_i^\ell(j+1)}\right)\right].$$

Fix a collection of packets $\gamma_i^\ell \in \mathcal{P}_{k_i}(\alpha_i, \beta_i)$. For which collections $\Gamma_i^\ell \in \mathcal{P}_{k_i'}(\alpha_i', \beta_i')$ does (29) give a nonzero answer? For that to happen, the number of $u_{i,j}$s and the number of $\overline{u}_{i,j}$s inside the expectation must be the same (because $e^{i\theta}U \stackrel{d}{=} U$ for any $\theta \in \mathbb{R}$). Assume that this is the case.

It will be convenient to write $\gamma(i, \ell, j)$ in place of $\gamma_i^\ell(j)$. From Result 15, to get a nonzero answer in (29) we must have bijections

$$\{(i, \ell, j) : i \leq m, \ell \leq p_i, 1 \leq j \leq k_i\}$$
$$\stackrel{\pi}{\rightarrow} \{(i, \ell, j) : i \leq m', \ell \leq p_i', 1 \leq j \leq k_i'\},$$
$$\{(i, \ell, j) : i \leq m, \ell \leq p_i, 2 \leq j \leq k_i + 1\}$$
$$\stackrel{\sigma}{\rightarrow} \{(i, \ell, j) : i \leq m', \ell \leq p_i', 2 \leq j \leq k_i' + 1\}$$

such that

$$(\gamma(i, \ell, j))_{i \leq m, \ell \leq p_i, 1 \leq j \leq k_i} = (\Gamma(\pi(i, \ell, j)))_{i \leq m, \ell \leq p_i, 1 \leq j \leq k_i},$$
$$(\gamma(i, \ell, j))_{i \leq m, \ell \leq p_i, 2 \leq j \leq k_i + 1} = (\Gamma(\sigma(i, \ell, j)))_{i \leq m, \ell \leq p_i, 2 \leq j \leq k_i + 1}.$$

And for each such pair of bijections π, σ, we get a contribution of $\mathrm{Wg}(N, \pi\sigma^{-1})$.

Let us call the collection of packets γ *typical*, if all the paths γ_i^ℓ are pairwise disjoint (except possibly at the initial and final points) and also nonself-intersecting (again, if $\alpha_i = \beta_i$, the paths in packet i intersect themselves, but only at the end points).

If γ is typical, then it is clear that for Γ to yield a nonzero contribution, Γ must consist of exactly the same paths as γ. This forces $k_i = k_i'$, $p_i = p_i'$, $\alpha_i = \alpha_i'$ and $\beta_i = \beta_i'$ for every i. If this is so, then the only pairs of bijections (π, σ) that yield a nonzero contribution are those for which:

- $\pi = \sigma$ (from the disjointness of the paths).
- π permutes each packet of paths among itself. In particular there are $\prod_{i=1}^k p_i!$ such permutations.

This shows that for a typical γ, the expectation in (29) is equal to

$$(30) \qquad \mathbf{1}_{\Gamma=\gamma}\left(\prod_{i=1}^m p_i!\right)\mathrm{Wg}(N, e).$$

Here $\gamma = \Gamma$ means that the two sets of paths are the same. Now suppose γ is atypical. For any fixed γ, typical or atypical, the number of Γ for which (29) is nonzero is clearly bounded uniformly by m and $p_i, k_i, i \le m$. In particular it is independent of N. Therefore the expected value in (29) is bounded in absolute value by

$$(31) \qquad C \sup_\tau \mathrm{Wg}(N, \tau).$$

Now for an atypical γ, at least two of $\gamma_i^\ell(j)$, $1 \le i \le m$, $1 \le \ell \le p_i$, $2 \le j \le k_i$, must be equal (our definition of "typical" did not impose any condition on the initial and final points of the paths, which are anyway fixed throughout). Thus, if we set $r = p_1(k_1 - 1) + \cdots + p_m(k_m - 1)$, then it follows that the total number of atypical γ is less than $r^2 N^{r-1}$. Since the total number of γ is precisely N^r, this also tells us that there are at least $N^r - r^2 N^{r-1}$ typical γ. Put these counts together with the contributions of each typical and atypical path, as given in (30) and (31), respectively. Note that we get nonzero contribution from typical paths only if $S = T$. Also, the total number of factors in S is $r + \sum p_i$ (this is the "k" in Result 15). Hence

$$\mathbf{E}[S\overline{T}] = \mathbf{1}_{S=T} N^r(1 - O(1/N))\,\mathrm{Wg}(N, e)\prod_{i=1}^m p_i! + O(N^{r-1})\sup_{\tau\in\mathcal{S}_{r+\sum p_i}}\mathrm{Wg}(N, \tau)$$

$$= \mathbf{1}_{S=T} N^{-\sum p_i}\left(\prod_{i=1}^m p_i!\right)\left(1 + O\left(\frac{1}{N}\right)\right)$$

by virtue of the asymptotics of the Weingarten function, as given in Result 15.

The factor $N^{\sum p_i}$ is precisely compensated for, once we scale $(U^k)_{\alpha,\beta}$ by $\sqrt{N}$, as in the statement of the lemma. Since the moments of standard complex Gaussian

are easily seen to be $\mathbf{E}[g^p \overline{g}^q] = p! \mathbf{1}_{p=q}$, we have shown that $\sqrt{N}(U^k)_{\alpha, \beta}$, $k \geq 1$, $\alpha, \beta \leq n$, converge to independent standard complex Gaussians. $\quad\square$

7. Proof of Lemma 14. We prove Lemma 14 in this section. We shall make use of the following fact, which is a direct consequence of [7], Theorem 26.

RESULT 16. Fix $n > 0$. Let $\{\lambda_1, \ldots, \lambda_N\}$ be determinantal on the unit disk with kernel

$$\mathbb{K}_N(z, w) = \sum_{k=0}^{N-1} \binom{-n-1}{k} (-1)^k z^k \overline{w}^k$$

with respect to the background measure $d\mu_n(z) = \frac{n}{\pi}(1 - |z|^2)^{n-1} \, dm(z)$. Then the set $\{|\lambda_k|^2 : 1 \leq k \leq N\}$ has the same distribution as $\{Y_k : 0 \leq k \leq N-1\}$, where Y_ks are independent random variables and Y_k has distribution Beta$(k+1, n)$.

As a consequence of this result, it is very easy to see that $N^{n/2}\mathbf{f}_N(0)$ is tight. For,

$$\mathbf{E}[|\mathbf{f}_N(0)|^2] = N^n \prod_{k=0}^{N-1} \mathbf{E}[Y_k]$$

$$= N^n \prod_{k=0}^{N-1} \frac{k+1}{n+k+1}$$

$$= N^n \frac{n!}{(N+1)(N+2)\cdots(N+n)}$$

$$\to n!$$

as $N \to \infty$. For $z \neq 0$ it is not as simple, because for finite N, the distribution of $\{|\frac{z-\lambda_k}{1-z\overline{\lambda}_k}|\}$ is not the same as that of a set of independent random variables. It is in the $N \to \infty$ limit, but this is of no use to us.

PROOF OF LEMMA 14. Write $\varphi_z(\lambda) = \frac{z-\lambda}{1-z\overline{\lambda}}$. Write

$$\sum_{k=1}^{N} \log |\varphi_z(\lambda_k)|^2 = -\sum_{k=1}^{N}(1 - |\varphi_z(\lambda_k)|^2) + \sum_{k=1}^{N} h_z(\lambda_k),$$

where $h_z(\lambda) := \log |\varphi_z(\lambda)|^2 + (1 - |\varphi_z(\lambda)|^2)$. The lemma will be proved by showing that the following are tight (uniformly over z in compact sets, as N varies).

$$(1) \sum_{k=1}^{N} h_z(\lambda_k), \qquad (2) \ -n \log N + \sum_{k=1}^{N}(1 - |\varphi_z(\lambda_k)|^2).$$

We consider them one by one.

1. Consider $\sum_{k=1}^{N} h_z(\lambda_k)$. For each $z \in \mathbb{D}$, the function φ_z maps the unit disk onto itself and the unit circle onto itself. Therefore, given K, a compact subset of the unit disk, we may find $T < 1$ such that for any $z \in K$ and $|\lambda| > T$ we have $|\varphi_z(\lambda)|^2 > \frac{1}{2}$. From the power series expansion of $\log(1 - x)$, we then get

$$|h_z(\lambda)| < \left(1 - |\varphi_z(\lambda)|^2\right)^2 \qquad \text{for } z \in K, |\lambda| > T.$$

Observing that

$$1 - |\varphi_z(\lambda)|^2 = \frac{(1 - |z|^2)(1 - |\lambda|^2)}{|1 - z\overline{\lambda}|^2},$$

it follows that

$$\left|\sum_{k=1}^{N} h_z(\lambda_k)\right| \leq \sum_{|\lambda_k| < T} |h_z(\lambda_k)| + C \sum_{|\lambda_k| > T} (1 - |\lambda_k|^2)^2$$

for a constant C (does not depend on N or z, as long as $z \in K$). The first summand is tight because the set $\{\lambda_k : |\lambda_k| < T\}$ converges to the set of points in $T\mathbb{D}$ in the limiting determinantal process. The second summand may be stochastically bounded by $\sum_{k=0}^{N-1}(1 - Y_k)^2$ by Result 16. From the explicit distribution of Y_ks, we may compute the expected value of this sum as

$$\mathbf{E}\left[\sum_{k=0}^{N-1}(1 - Y_k)^2\right] = \sum_{k=0}^{N-1} \frac{n(n+1)}{(n+k+1)(n+k+2)}.$$

The random variables on the left are stochastically increasing in N and hence, a uniform bound on the expectations shows tightness.

2. Consider $-n \log N + \sum_{k=1}^{N}(1 - |\varphi_z(\lambda_k)|^2)$. We shall show tightness by proving that the expected value and variance of these random variables are bounded uniformly in N (and $z \in K$).

Expected value: As always, it is simpler for us to deal with $|\lambda_k|^2$. Therefore, write

$$1 - |\varphi_z(\lambda)|^2 = \frac{(1 - |z|^2)(1 - |\lambda|^2)}{|1 - z\overline{\lambda}|^2}$$

$$= (1 - |z|^2)(1 - |\lambda|^2) \sum_{p,q=0}^{\infty} z^p \overline{z}^q \lambda^q \overline{\lambda}^p.$$

Set $\lambda = \lambda_k$, sum over k and take expectations. Any term with $p \neq q$ vanishes, because of rotation invariance of $\{\lambda_k\}$. For $p = q$, we get terms with $|\lambda_k|^2$, which may be replaced by independent Beta random variables by Result 16.

Thus

$$\mathbf{E}\left[\sum_{k=1}^{N}\left(1-|\varphi_z(\lambda_k)|^2\right)\right]$$

$$= (1-|z|^2)\sum_{p=0}^{\infty}|z|^{2p}\mathbf{E}\left[\sum_{k=1}^{N}(1-|\lambda_k|^2)|\lambda_k|^{2p}\right]$$

$$= (1-|z|^2)\sum_{p=0}^{\infty}|z|^{2p}\sum_{k=1}^{N}\frac{\text{Beta}(k+p,n)-\text{Beta}(k+p+1,n)}{\text{Beta}(k,n)}$$

$$= (1-|z|^2)\sum_{p=0}^{\infty}|z|^{2p}\sum_{k=1}^{N}\frac{nk(k+1)\cdots(k+p-1)}{(k+n)\cdots(k+n+p)}$$

$$= (1-|z|^2)\sum_{p=0}^{\infty}|z|^{2p}\sum_{k=1}^{N}\left(\frac{n}{k}+\text{err}_k\right),$$

where in the last line, from the simple bounds

$$\frac{n}{k+n}\left(\frac{k}{k+n+1}\right)^{p+1} \le \frac{nk(k+1)\cdots(k+p-1)}{(k+n)\cdots(k+n+p)} \le \frac{n}{k+n},$$

we see that $|\text{err}_k| \le C(p)k^{-2}$ where $C(p)$ is at most a polynomial in p (in fact $|\text{err}_k| \le \frac{(p+2)n^2}{k(k+n)}$). Thus, it follows that

$$\mathbf{E}\left[\sum_{k=1}^{N}\left(1-|\varphi_z(\lambda_k)|^2\right)\right] = (n\log N)(1-|z|^2)\left(\sum_{p=0}^{\infty}|z|^{2p}\right)+O(1)$$

$$= n\log N + O(1).$$

This is exactly what we wanted to show about the expected value.

Variance: Now we want the variance of $X_N := \sum_{k=1}^{N}(1-|\varphi_z(\lambda_k)|^2)$. No really new ideas are needed, only the calculations are more tedious. Expand X_N in power series as before to get

(32)
$$E[X^2] = (1-|z|^2)^2$$

$$\times \sum_{p,q,r,s\ge0} z^{p+r}\overline{z}^{q+s}\mathbf{E}\left[\sum_{k,\ell=1}^{N}(1-|\lambda_k|^2)(1-|\lambda_\ell|^2)\lambda_k^q\overline{\lambda}_k^p\lambda_\ell^s\overline{\lambda}_\ell^r\right].$$

All terms in which $p+r \neq q+s$ vanish, by rotation invariance. Fix p,q,r,s so that $p+r = q+s$ and write the inner expectation as

$$\mathbf{E}\left[\sum_{k=1}^{N}(1-|\lambda_k|^2)^2|\lambda_k|^{2p+2r}\right] + \mathbf{E}\left[\sum_{k\neq\ell}(1-|\lambda_k|^2)(1-|\lambda_\ell|^2)\lambda_k^q\overline{\lambda}_k^p\lambda_\ell^s\overline{\lambda}_\ell^r\right].$$

The first one is already a function of the absolute values of λ_ks and hence, replacing them by independent Beta random variables, we get (details are similar to those in computing the expectation)

$$\mathbf{E}\left[\sum_{k=1}^{N}(1-|\lambda_k|^2)^2|\lambda_k|^{2p+2r}\right]$$

$$= \sum_{k=1}^{N}\text{Beta}(k+p+r,n) - 2\,\text{Beta}(k+p+r+1,n)$$

$$+ \text{Beta}(k+p+r+2,n)(\text{Beta}(k,n))^{-1}$$

$$= n(n+1)\sum_{k=1}^{N}\frac{k(k+1)\cdots(k+p+r-1)}{(k+n)\cdots(k+n+p+r+1)}$$

$$\leq n(n+1)\sum_{k=1}^{N}\frac{1}{k(k+1)},$$

which is bounded. Of course, we shall have to sum over p, q, r, s subject to $p+r = q+s$, but it is clear that because of the factor of $|z|^{2p+2r}$ in (32), the total contribution to (32) from this summand (all terms with $k = \ell$) is bounded as $N \to \infty$.

It remains to consider the sum over $k \neq \ell$. The two-point correlation is $\mathbb{K}_N(\lambda,\lambda)\mathbb{K}_N(\xi,\xi) - \mathbb{K}_N(\lambda,\xi)\mathbb{K}_N(\xi,\lambda)$. We consider

$$\mathbf{E}\left[\sum_{k\neq\ell}(1-|\lambda_k|^2)(1-|\lambda_\ell|^2)\lambda_k^q\overline{\lambda}_k^p\lambda_\ell^s\overline{\lambda}_\ell^r\right]$$

$$= \int_{\mathbb{D}^2}(1-|\lambda|^2)(1-|\xi|^2)\lambda^q\overline{\lambda}^p\xi^s\overline{\xi}^r\left(\mathbb{K}_N(\lambda,\lambda)\mathbb{K}_N(\xi,\xi) - |\mathbb{K}_N(\lambda,\xi)|^2\right).$$

Consider the first summand, where we choose the term $\mathbb{K}_N(\lambda,\lambda)\mathbb{K}_N(\xi,\xi)$ inside the brackets. This survives only if $p = q$ and $r = s$ and it is easily seen that this term when summed over $p = q$ and $r = s$ in (32) will give exactly $\mathbf{E}[X]^2$. When we compute the variance of X, we shall subtract $\mathbf{E}[X]^2$ from $\mathbf{E}[X^2]$ and this term gets cancelled.

Thus to show the boundedness of the variance, we only need to show the boundedness of

$$(33)\qquad \int_{\mathbb{D}^2}(1-|\lambda|^2)(1-|\xi|^2)\lambda^q\overline{\lambda}^p\xi^s\overline{\xi}^r\mathbb{K}_N(\lambda,\xi)\mathbb{K}_N(\xi,\lambda)\,d\mu_n(\lambda)\,d\mu_n(\xi).$$

Recall that

$$\mathbb{K}_N(\lambda,\xi) = \sum_{j=0}^{N-1}C_j\lambda^j\overline{\xi}^j,$$

where, $C_j = \binom{-n-1}{j}(-1)^j$. Moreover, for any j,

$$C_j \int_{\mathbb{D}} |\lambda|^{2j} \, d\mu_n(\lambda) = 1.$$

Therefore the integral in (33) is equal to

$$\sum_{i,j=0}^{N-1} C_i C_j \int_{\mathbb{D}} (1 - |\lambda|^2) \lambda^{q+i} \overline{\lambda}^{p+j} \, d\mu_n(\lambda) \int_{\mathbb{D}} (1 - |\xi|^2) \xi^{s+j} \overline{\xi}^{r+i} \, d\mu_n(\xi)$$

$$= \sum_{i,j=0}^{N-1} C_i C_j \delta_{q+i,p+j} \delta_{s+j,r+i} \left(\frac{1}{C_{p+j}} - \frac{1}{C_{p+j+1}} \right) \left(\frac{1}{C_{r+i}} - \frac{1}{C_{r+i+1}} \right).$$

We fixed p, q, r, s such that $p - q = s - r$. Therefore, there are $N - |p - q|$ choices for (i, j) which do not vanish (if $|p - q| \geq N$, there are no such terms). Without losing generality, let $p < q$, and write the above quantity as

$$\sum_{i=0}^{N-(q-p)} C_i C_{q-p-i} \left(\frac{1}{C_{p+j}} - \frac{1}{C_{p+j+1}} \right) \left(\frac{1}{C_{r+i}} - \frac{1}{C_{r+i+1}} \right).$$

Since $C_j^{-1} = \text{Beta}(j + 1, n)$ we compute

$$\frac{1}{C_j} - \frac{1}{C_{j+1}} = \frac{n\Gamma(j+1)}{\Gamma(n)\Gamma(j+n+2)}.$$

Therefore

$$C_i C_j \left(\frac{1}{C_{p+j}} - \frac{1}{C_{p+j+1}} \right) \left(\frac{1}{C_{r+i}} - \frac{1}{C_{r+i+1}} \right)$$

$$= \frac{\Gamma(i+n+1)\Gamma(j+n+1)}{\Gamma(i+1)\Gamma(j+1)} \frac{n^2 \Gamma(p+j+1)\Gamma(r+i+1)}{\Gamma(p+j+n+2)\Gamma(r+i+n+2)}$$

$$= n^2 \frac{(j+1)\cdots(j+p)(i+1)\cdots(i+r)}{(j+n+1)\cdots(j+n+p+1)(i+n+1)\cdots(i+n+r+1)}.$$

As before by the obvious bounds we may write this quantity as

$$\frac{n^2}{(i+1)(j+1)} + O(i^{-2}j^{-2}).$$

Now we take i, j with $j = i + q - p$ and sum over i as well as p, q, r, s (with the condition $p + r = q + s$). This yields a bounded quantity.

In summary, we wrote $E[X^2]$ as in (32). Terms with $k = \ell$ yielded a bounded quantity. Terms with $k \neq \ell$ were split into two sums. One of them is bounded while the other is exactly equal to $\mathbf{E}[X]^2$. Thus the variance is bounded as $N \to \infty$.

This completes the proof of Lemma 14. $\square$

8. Questions. We record here two among several natural questions that arise from the considerations of this paper.

1. In Theorem 4, the determinantal process exists for any positive real n, while the matrix analytic function makes sense only for integer values of n (size of the matrix). Is there a random zero set interpretation for hyperbolic determinantal processes for noninteger values of n?
2. For each z, the eigenvalues of the random matrix in Theorem 4 is a rescaled version of Ginibre's ensemble. Is it possible to find the joint law of the sets of eigenvalues as z varies over the unit disk?
3. Are there random matrices of the form given in Theorem 5 for which we can calculate the exact distribution of eigenvalues? Recall that these are random matrices for which: (1) The distribution is invariant under multiplication by unitary matrices. (2) The number of random singular values stays fixed even as the size of the matrix goes to infinity.

   If this can be done, then presumably we shall also get the distribution of zeros of $\det(X_0 + zG_1 + z^2G_2 + \cdots)$. We say "presumably," because we have not proved uniform convergence of $\mathbf{f}_N$ on compact sets, except in the special case of truncated Haar unitary matrices.

APPENDIX: DETERMINANTAL POINT PROCESSES

We give a brief introduction to determinantal processes, strictly limited to the context of this paper. More details, as well as proofs, may be found in the surveys [7, 19].

Let Ω be a region in the plane and let p be a positive continuous function on Ω. Define the measure μ by $d\mu(z) = p(z)\,dm(z)$. A simple point process on Ω is a random measure on Ω that takes values in the space of counting measures of locally finite subsets of Ω. If the number of points that fall in any compact set has exponential tails, then the distribution of the point process is determined by its correlation functions (joint intensities) with respect to μ,

$$
\rho_k(z_1, \ldots, z_k)
$$
$$
= \lim_{\varepsilon \downarrow 0} \frac{\mathbf{P}[\mathcal{X} \text{ has points in each of } D(z_i, \varepsilon), 1 \le i \le k]}{\prod_{i=1}^{k} \mu(D(z_i, \varepsilon))}
$$

for any $k \ge 1$ and any $z_1, \ldots, z_k \in \Omega$. There is also an integral version of this definition which is more appropriate in more general situations. Joint intensities need not exist in general.

Consider the space $L^2(\Omega, \mu)$ and its subspace $\mathbf{H}$ consisting of holomorphic functions. It is a fact that $\mathbf{H}$ is a closed subspace. For, suppose $f_n \in \mathbf{H}$ and $f_n \to f$ in $L^2(\mu)$. Then for any $z \in \Omega$, consider a disk $D(z, r)$ contained entirely in Ω. Convergence in L^2 and absolute continuity of μ shows that f_ns converge to f

in $L^2(m)$ on the annulus $D(z, r) \setminus D(z, \frac{r}{2})$. Apply Cauchy's integral formula to infer that $f_n \to f$ uniformly on $D(z, \frac{r}{4})$. Therefore f is itself analytic on Ω. This shows that **H** is a Hilbert space. This reasoning also shows that for any $z \in \Omega$, the evaluation $f \to f(z)$ is a bounded linear functional on **H**.

Switch notation and let **H** denote any closed subspace of $L^2(\mu)$ consisting of holomorphic functions. Then the evaluation $f \to f(z)$ is a bounded linear functional on **H**. As a consequence, if $\{\psi_n\}_{n \geq 1}$ is any orthonormal basis of **H**, then the series

$$\mathbb{K}(z, w) = \sum \psi_n(z) \overline{\psi}_n(w),$$

does converge and $\mathbb{K}$ (called the reproducing kernel of **H**) is independent of the choice of the basis. The integral operator on $L^2(\mu)$ defined by

$$\mathcal{K}f(z) = \int_\Omega f(w) \mathbb{K}(z, w) \, d\mu(w)$$

is precisely the projection operator on $L^2(\mu)$ onto the subspace **H**.

DEFINITION A.1.   Let **H** be a closed subspace of $L^2(\mu)$ consisting of complex analytic functions, and let $\mathbb{K}$ be the reproducing kernel of **H**. Then a point process $\mathcal{X}$ on Ω with joint intensities given by

$$\rho_k(z_1, \ldots, z_k) = \det(\mathbb{K}(z_i, z_j))_{i, j \leq k}$$

does exist and is called the determinantal point process on Ω with kernel $\mathbb{K}$ with respect to the measure μ. We also say that **H** is the associated Hilbert space.

We just state a few facts regarding these processes.

1. The number of points in $\mathcal{X}$ is almost surely equal to the dimension of **H**.
2. If subspaces $\mathbf{H}_n$ increase to **H**, then the corresponding determinantal processes converge in distribution, that is, $\mathcal{X}_n \overset{d}{\to} \mathcal{X}$. This was used tacitly in deducing Theorem 4 from Result 12.
3. If $\mathbf{H} = \mathrm{span}\{1, z, \ldots, z^{n-1}\}$, then by writing out the density (which is just $\frac{1}{n!}\rho_n(\cdot)$), one sees that the vector of n points of the process has density proportional to $\prod_{i<j} |z_i - z_j|^2 \prod_{i=1}^n p(z_i)$ with respect to the Lebesgue measure on Ω^n.

which finally enabled me to prove Theorem 4. Thanks are also due to the two reviewers for many comments, corrections and references and to Ron Peled for comments on the proof of Theorem 3.

Note added in proof: Just as the final version of this paper was about to be sent, Victor Katsnelson has sent us a preprint [10] of his joint work with Bernd Kirstein, on the Schur algorithm and system theory, using ideas with which it seems that the proof of Theorem 4 may be simplified. We leave this for the future.

REFERENCES

[1] CAILLOL, J. M. (1981). Exact results for a two-dimensional one-component plasma on a sphere. *J. Physique* **42** L–245–L–247.

[2] DIACONIS, P. and EVANS, S. N. (2001). Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.* **353** 2615–2633 (electronic). MR1828463

[3] DIACONIS, P. and SHAHSHAHANI, M. (1994). On the eigenvalues of random matrices. *J. Appl. Probab.* **31A** 49–62. Studies in applied probability. MR1274717

[4] GINIBRE, J. (1965). Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.* **6** 440–449. MR0173726

[5] GIRVIN, S. M. and JACH, T. (1984). Formalism for the quantum Hall effect: Hilbert space of analytic functions. *Phys. Rev. B* (3) **29** 5617–5625. MR743876

[6] HORN, R. A. and JOHNSON, C. R. (1985). *Matrix Analysis*. Cambridge Univ. Press, Cambridge. MR832183

[7] HOUGH, J. B., KRISHNAPUR, M., PERES, Y. and VIRÁG, B. (2006). Determinantal processes and independence. *Probab. Surv.* **3** 206–229 (electronic). MR2216966

[8] JANCOVICI, B. and TÉLLEZ, G. (1998). Two-dimensional Coulomb systems on a surface of constant negative curvature. *J. Statist. Phys.* **91** 953–977. MR1637270

[9] JIANG, T. (2006). How many entries of a typical orthogonal matrix can be approximated by independent normals? *Ann. Probab.* **34** 1497–1529. MR2257653

[10] KATSNELSON, V. and KIRSTEIN, B. The Schur algorithm in terms of system realization. Preprint.

[11] KRISHNAPUR, M. Singular points of random matrix-valued analytic functions: Asymptotics as the matrix size increses. Preprint.

[12] KRISHNAPUR, M. Zeros of random analytic functions. Ph.D. dissertation, U.C. Berkeley.

[13] LEBŒUF, P. (1999). Random analytic chaotic eigenstates. *J. Statist. Phys.* **95** 651–664. MR1700934

[14] MACCHI, O. (1975). The coincidence approach to stochastic point processes. *Adv. Appl. Probab.* **7** 83–122. MR0380979

[15] MEHTA, M. L. (2004). *Random Matrices*, 3rd ed. *Pure and Applied Mathematics* (*Amsterdam*) **142**. Elsevier, Amsterdam. MR2129906

[16] NICA, A. and SPEICHER, R. (2006). *Lectures on the Combinatorics of Free Probability. London Mathematical Society Lecture Note Series* **335**. Cambridge Univ. Press, Cambridge. MR2266879

[17] PERES, Y. and VIRÁG, B. (2005). Zeros of the i.i.d. Gaussian power series: A conformally invariant determinantal process. *Acta Math.* **194** 1–35. MR2231337

[18] SODIN, M. and TSIRELSON, B. (2004). Random complex zeroes. I. Asymptotic normality. *Israel J. Math.* **144** 125–149. MR2121537

[19] SOSHNIKOV, A. (2000). Determinantal random point fields. *Uspekhi Mat. Nauk* **55** 107–160. MR1799012

[20] ŻYCZKOWSKI, K. and SOMMERS, H.-J. (2000). Truncations of random unitary matrices. *J. Phys. A* **33** 2045–2057. MR1748745

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
40 ST. GEORGE STREET
TORONTO, ONTARIO
CANADA M5S 2E4
E-MAIL: manju@math.toronto.edu