The distribution of zeros of the derivative of a random polynomial

Robin Pemantle and Igor Rivin

Abstract In this note we initiate the probabilistic study of the critical points of polynomials of large degree with a given distribution of roots. Namely, let f be a polynomial of degree n whose zeros are chosen IID from a probability measure μ on $\mathbb{C}$. We conjecture that the zero set of f' always converges in distribution to μ as $n \to \infty$. We prove this for measures with finite one-dimensional energy. When μ is uniform on the unit circle this condition fails. In this special case the zero set of f' converges in distribution to that of the IID Gaussian random power series, a well known determinantal point process.

1 Introduction

Since Gauss, there has been considerable interest in the location of the *critical points* (zeros of the derivative) of polynomials whose zeros were known – Gauss noted that these critical points were points of equilibrium of the electrical field whose charges were placed at the zeros of the polynomial, and this immediately leads to the proof of the well-known Gauss-Lucas Theorem, which states that the critical points of a polynomial f lie in the convex hull of the zeros of f (see, e.g. [18, Theorem 6.1]). There are too many refine-

Robin Pemantle
University of Pennsylvania, Department of Mathematics, 209 S. 33rd Street, Philadelphia, PA 19104, USA, e-mail: `pemantle@math.upenn.edu`

Igor Rivin
Temple University, Department of Mathematics, 1805 North Broad St Philadelphia, PA 19122, USA, e-mail: `rivin@math.temple.edu`

ments of this result to state. A partial list (of which several have precisely the same title!) is as follows: [1, 3, 5–9, 12, 14, 16, 17, 19, 20, 22–26]). Among these, we mention two extensions that are easy to state.

- Jensen's theorem: if $p(z)$ has real coefficients, then the non-real critical points of p lie in the union of the "Jensen Disks", where a Jensen disk J is a disk one of whose diameters is the segment joining a pair of conjugate (non-real) roots of p.
- Marden's theorem: Suppose the zeroes z_1, z_2, and z_3 of a third-degree polynomial $p(z)$ are non-collinear. There is a unique ellipse inscribed in the triangle with vertices z_1, z_2, z_3 and tangent to the sides at their midpoints: the Steiner inellipse. The foci of that ellipse are the zeroes of the derivative $p'(z)$.

There has not been any *probabilistic* study of critical points (despite the obvious statistical physics connection) from this viewpoint. There has been a very extensive study of random polynomials (some of it quoted further down in this paper), but generally this has meant some distribution on the coefficients of the polynomial, and not its roots [4]. Let us now define our problem:

Let μ be a probability measure on the complex numbers. Let $\{X_n : n \geq 0\}$ be random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ that are IID with common distribution μ. Let

$$f_n(z) := \prod_{j=1}^{n} (z - X_j)$$

be the random polynomial whose roots are $X_1, \ldots, X_n$. For any polynomial f we let $\mathcal{Z}(f)$ denote the empirical distribution of the roots of f, for example, $\mathcal{Z}(f_n) = \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j}$.

The question we address in this paper is:

Question 1.1 *When are the zeros of f_n' stochastically similar to the zeros of f_n?*

Some examples show why we expect this.

Example 1.1 *Suppose μ concentrates on real numbers. Then f_n has all real zeros and the zeros of f_n' interlace the zeros of f_n. It is immediate from this that the empirical distribution of the zeros of f_n' converges to μ as $n \to \infty$. The same is true when μ is concentrated on any affine line in the complex plane: interlacing holds and implies convergence of the zeros of f_n' to μ.[1] Once the support of μ is not contained in an affine subspace, however, the best we can say geometrically about the roots of f_n' is that they are contained in the convex hull of the roots of f_n; this is the Gauss-Lucas Theorem.*

---

[1] Even in this case there are interesting probabilistic questions concerning the distribution of critical points of f_n close to the edge of the support of μ, see [15]

Example 1.2 *Suppose the measure μ is atomic. If $\mu(a) = p > 0$ then the multiplicity of a as a zero of f_n is $n(p + o(1))$. The mulitplicity of a as a zero of f_n' is one less than the multplicity as a zero of f_n, hence also $n(p + o(1))$. This is true for each of the countably many atoms, whence it follows again that the empirical distribution of the zeros of f_n' converges to μ.*

Atomic measures are weakly dense in the space of all measures. Sufficient continuity of the roots of f' with respect to the roots of f would therefore imply that the zeros of f_n' always converge in distribution to μ as $n \to \infty$. In fact we conjecture this to be true.

Example 1.3 *Our first experimental example has the roots of f uniformly distributed in the unit disk. In the figure, we sample 300 points from the uniform distribution in the disk, and plot the critical points (see Figure 1). The reader may*
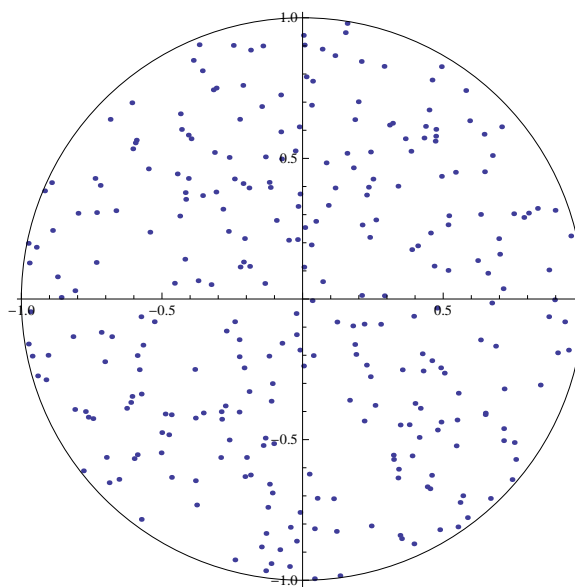

Fig. 1 Critical points of a polynomial whose roots are uniformly sampled inside the unit disk.

or may not be convinced that the critical points are uniformly distributed.

Example 1.4 *Our second example takes polynomials with roots uniformly distributed on the unit* circle, *and computes the critical points. In Figure 2 we do this with a sample of size 300. One sees that the convergence is rather quick.*

Remark 1 *The figures were produced with Mathematica. However, the reader wishing to try this at home should increase precision because Mathematica (and Maple, Matlab and R) do not use the best method of computing zeros of polynomials.*
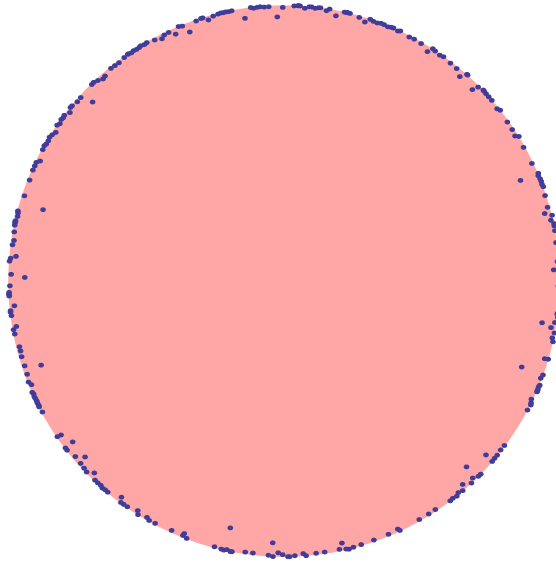
Fig. 2 Critical points of polynomial whose roots are uniformly sampled on the unit circle.

Conjecture 1. For any μ, as $n \to \infty$, $\mathcal{Z}(f')$ converges weakly to μ.

There may indeed be such a continuity argument, though the following counterexample shows that one would at least need to rule out some exceptional sets of low probability. Suppose that $f(z) = z^n - 1$. As $n \to \infty$, the distribution of the roots of f converge weakly to the uniform distribution on the unit circle. The roots of f'_n however are all concentrated at the origin. If one moves one of the n roots of f_n along the unit circle, until it meets the next root, a distance of order $1/n$, then one root of f'_n zooms from the origin out to the unit circle. This shows that small perturbations in the roots of f can lead to large perturbations in the roots of f'. It seems possible, though, that this is only true for a "small" set of "bad" functions f.

1.1 A little history

This circle of questions was first raised in discussions between one of us (IR) and the late Oded Schramm, when IR was visiting at Microsoft Research for the auspicious week of 9/11/2001. Schramm and IR had some ideas on how to approach the questions, but were somewhat stuck. There was always an intent to return to these questions, but Schramm's passing in September 2008 threw the plans into chaos. We (RP and IR) hope we can do justice to Oded's memory.

These questions are reminscent of questions of the kind often raised by Herb Wilf, that sound simple but are not. This work was first presented at a conference in Herb's honor and we hope it serves as a fitting tribute to Herb as well.

2 Results and Notations

Our goal in this paper is to prove cases of Conjecture 1.

Definition 2 *We definite the p-energy of μ to be*

$$\mathcal{E}_p(\mu) := \left(\int \int \frac{1}{|z-w|^p} \, d\mu(z) \, d\mu(w) \right)^{1/p}.$$

Since in the sequel we will only be using the 1*-energy, we will write $\mathcal{E}$ for $\mathcal{E}_1$.*

By Fubini's Theorem, when μ has finite 1-energy, the function V_μ defined by

$$V_\mu(z) := \int \frac{1}{z-w} \, d\mu(w)$$

is well defined and in $L^1(\mu)$.

Remark 2 *The potential function V_μ is sometimes called the* Cauchy transform *of the measure μ. Commonly it is implied that μ is supported on $\mathbb{R}$ or on the boundary of a region over which z varies, but this need not be the case and is not the case for us (except in Theorem 2).*

Theorem 1. *Suppose μ has finite 1-energy and that*

$$\mu \left\{ z : V_\mu(z) = 0 \right\} = 0. \tag{1}$$

Then $\mathcal{Z}(f'_n)$ converges in distribution to μ as $n \to \infty$.

A natural set of examples of μ with finite 1-energy is provided by the following observation:

Observation 1 *Suppose $\Omega \subset \mathbb{C}$ has Hausdorff dimension greater than one, and μ is in the measure class of the Hausdorff measure on Ω. Then μ has finite 1-energy.*

PROOF: This is essentially the content of [11][Theorem 4.13(b)].   □

In particular, if μ is uniform in an open subset (with compact closure) of $\mathbb{C}$, its 1-energy is finite.

A natural special case to which Theorem 1 does not apply is when μ is uniform on the unit circle; here the 1-energy is just barely infinite.

Theorem 2. *If μ is uniform on the unit circle then $\mathcal{Z}(f_n)$ converges to the unit circle in probability.*

This result is somewhat weak because we do not prove $\mathcal{Z}(f_n)$ has a limit in distribution, only that all subsequential limits are supported on the unit circle. By the Gauss-Lucas Theorem, all roots of f_n have modulus less than 1, so the convergence to μ is from the inside. Weak convergence to μ implies that only $o(n)$ points can be at distance $\Theta(1)$ inside the cirle; the number of such points turns out to be $\Theta(1)$. Indeed quite a bit can be said about the small outliers. For $0 < \rho < 1$, define $B_\rho := \{z : |z| \leq \rho\}$. The following result, which implies Theorem 2, is based on a very pretty result of Peres and Virag [21, Theorems 1 and 2] which we will quote in due course.

Theorem 3. *For any $\rho \in (0,1)$, as $n \to \infty$, the set $\mathcal{Z}(g_n) \cap B_\rho$ of zeros of g_n on B_ρ converges in distribution to a determinantal point process on B_ρ with the so-called Bergmann kernel $\pi^{-1}(1 - z_i \overline{z_j})^2$. The number $N(\rho)$ of zeros is distributed as the sum of independent Bernoullis with means ρ^{2k}, $1 \leq k < \infty$.*

2.1 Distance functions on the space of probability measures

If μ and ν are probability measures on a separable metric space S, then the **Prohorov**[2] distance $|\mu - \nu|_P$ is defined to be the least ε such that for every set A, $\mu(A) \leq \nu(A^\varepsilon) + \varepsilon$ and $\nu(A) \leq \mu(A^\varepsilon) + \varepsilon$. Here, A^ε is the set of all points within distance ε of some point of A. The Prohorov metric metrizes convergence in distribution. We view collections of points in $\mathbb{C}$ (e.g., the zeros of f_n) as probability measures on $\mathbb{C}$, therefore the Prohorov metric serves to metrize convergence of zero sets. The space of probability measures on S, denoted $\mathcal{P}(S)$, is itself a separable metric space, therefore one can define the Prohorov metric on $\mathcal{P}(S)$, and this metrizes convergence of laws of random zero sets.

The Ky Fan metric on random variables on a fixed probability space will be of some use as well. Defined by $K(X,Y) = \inf\{\varepsilon : \mathbb{P}(d(X,Y) > \varepsilon) < \varepsilon\}$, this metrizes convergence in probability. The two metrics are related (this is Strassen's Theorem):

$$|\mu - \nu|_P = \inf\{K(X,Y) : X \sim \mu, Y \sim \nu\}. \tag{2}$$

A good reference for the facts mentioned above is available on line [13]. We will make use of Rouché's Theorem. There are a number of formulations, of which the most elementary is probably the following statement proved as Theorem 10.10 in [2].

Theorem 4 (Rouché). *If f and g are analytic on a topological disk, B, and $|g| < |f|$ on ∂B, then f and $f + g$ have the same number of zeros on B.* □

---

[2] Also known as the Prokhorov and the Lévy-Pro(k)horov distance

3 Proof of Theorem 1

We begin by stating some lemmas. The first is nearly a triviality.

Lemma 1. *Suppose μ has finite 1-energy. Then*

(i)

$$t \cdot \mathbb{P}\left(|X_0 - X_1| \leq \frac{1}{t}\right) \to 0.$$

(ii) for any $C > 0$,

$$\mathbb{P}\left(\min_{1 \leq j \leq n} |X_j - X_{n+1}| \leq \frac{C}{n}\right) \to 0;$$

PROOF: For part (i) observe that $\limsup t \cdot \mathbb{P}(|X_0 - X_1| \leq 1/t) \leq 2 \limsup 2^j \cdot \mathbb{P}\left(|X_0 - X_1| \leq 2^{-j}\right)$ as t goes over reals and j goes over integers. We then have

$$\infty > \mathcal{E}(\mu)$$

$$= \mathbb{E}\frac{1}{|X_0 - X_1|}$$

$$\geq \frac{1}{2}\mathbb{E}\sum_{j \in \mathbb{Z}} 2^j \mathbf{1}_{|X_0 - X_1| \leq 2^{-j}}$$

$$= \frac{1}{2}\sum_{j} 2^j \mathbb{P}\left(|X_0 - X_1| \leq 2^{-j}\right)$$

and from the finiteness of the last sum it follows that the summand goes to zero. Part (ii) follows from part (i) upon observing, by symmetry, that

$$\mathbb{P}\left(\min_{1 \leq j \leq n} |X_j - X_{n+1}| \leq \frac{C}{n}\right) \leq n\mathbb{P}\left(|X_0 - X_1| \leq \frac{C}{n}\right).$$

$\square$

Define the nth empirical potential function $V_{\mu,n}$ by

$$V_{\mu,n}(z) := \frac{1}{n}\sum_{j=1}^{n} \frac{1}{z - X_j}$$

which is also the integral in w of $1/(z - w)$ against the measure $\mathcal{Z}(f_n)$. Our next lemma bounds $V'_{\mu,n}(z)$ on the disk $B := B_{C/n}(X_{n+1})$.

Lemma 2. *For all $\varepsilon > 0$,*

$$\mathbb{P}\left(\sup_{z \in B} |V'_{\mu,n}(z)| \geq \varepsilon n\right) \to 0$$

as $n \to \infty$.

PROOF: Let G_n denote the event that $\min_{1\le j\le n}|X_j - X_{n+1}| > 2C/n$. Let $S_n :=$ $\sup_{z\in B}|V'_{\mu,n}(z)|$. We will show that

$$\mathbb{E}S_n\mathbf{1}_{G_n} = o(n) \tag{3}$$

as $n \to \infty$. By Markov's inequality, this implies that $\mathbb{P}(S_n\mathbf{1}_{G_n} \ge \varepsilon n) \to 0$ for all $\varepsilon > 0$ as $n \to \infty$. By part (ii) of Lemma 1 we know that $\mathbb{P}(G_n) \to 1$, which then establishes that $\mathbb{P}(S_n \ge \varepsilon n) \to 0$, proving the lemma.

In order to show (3) we begin with

$$|V'_{\mu,n}(z)| = \left|\frac{1}{n}\sum_{j=1}^n \frac{-1}{(z - X_j)^2}\right| \le \frac{1}{n}\sum_{j=1}^n \frac{1}{|z - X_j|^2}.$$

Therefore,

$$S_n\mathbf{1}_{G_n} \le \frac{1}{n}\sum_{j=1}^n \frac{1}{(|X_{n+1} - X_j| - C/n)^2}\mathbf{1}_{G_n} \le \frac{1}{n}\sum_{j=1}^n \frac{4}{|X_{n+1} - X_j|^2}\mathbf{1}_{G_n}, \tag{4}$$

where we have used the triangle inequality, thus:

$$|z - X_j| = |(z - X_{n+1}) + (x_{n+1} - X_j)| \ge |X_{n+1} - X_j| - |z - X_{n+1}|.$$

Since we are in B, we know that $|z - X_{n+1}| \le C/n$, and since we are in G_n, we know that $C/n < |X_{n+1} - X_j|/2$.

Because S_n is the supremum of an average of n summands and the summands are exchangeable, the expectation of $S_n\mathbf{1}_{G_n}$ is bounded from above by the expectation of one summand. Referring to (4), and using the fact that G_n is contained in the event that $|X_{n+1} - X_1| > 2C/n$, this gives

$$\mathbb{E}S_n\mathbf{1}_{G_n} \le \mathbb{E}\frac{4}{|X_{n+1} - X_1|^2}\mathbf{1}_{|X_{n+1}-X_1|\ge 2C/n}.$$

A standard inequality for nonnegative variables (integrate by parts) is

$$\mathbb{E}W^2\mathbf{1}_{W\le t} \le \int_0^t 2s\mathbb{P}(W \ge s)\,ds.$$

When applied to $W = |X_{n+1} - X_1|^{-1}$ and $t = n/(2C)$, this yields

$$\mathbb{E}S_n\mathbf{1}_{G_n} \le \int_0^{n/(2C)} 2s\,\mathbb{P}\left(\frac{1}{|X_0 - X_1|} > s\right)\,ds.$$

The integrand goes to zero as $n \to \infty$ by part (i) of Lemma 1. It follows that the integral is $o(n)$, proving the lemma. $\square$

Define the lower modulus of V to distance C/n by

$$\underline{V}_n^C(z) := \inf_{w:|w-z|\le C/n} \left|V_{\mu,n}(w)\right| \ .$$

This depends on the argument μ as well as C and n but we omit this from the notation.

Lemma 3. *Assume μ has finite 1-energy. Then as $n \to \infty$, the random variable $\underline{V}_n^C(X_{n+1})$ converges in probability, and hence in distribution, to $|V_\mu(X_{n+1})|$.*

In the sequel we will need the Glivenko-Cantelli Theorem [10, Theorem 1.7.4]. Let $X_1, \ldots, X_n, \ldots$ be independent, identically distributed random variables in $\mathbb{R}$ with common cumulative distribution function F. The *empirical distribution function F_n* for $X_1, \ldots, X_n$ is defined by

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I_{(-\infty,x]}(X_i),$$

where I_C is the indicator function of the set C. For every fixed x, $F_n(x)$ is a sequence of random variables, which converges to $F(x)$ almost surely by the strong law of large numbers. Glivenko-Cantelli Theorem strengthen this by proving *uniform convergence* of F_n to F.

Theorem 5 (Glivenko-Cantelli).

$$\|F_n - F\|_\infty = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \longrightarrow 0 \quad \text{almost surely.}$$

The following Corollary is immediate:

Corollary 1. *Let f be a bounded continuous function on $\mathbb{R}$. Then*

$$\lim_{n\to\infty} \int_{\mathbb{R}} f dF_n = \int_{\mathbb{R}} f dF, \quad \text{almost surely.}$$

Another immediate Corollary is:

Corollary 2. *With notation as in the statement of Theorem 5, the Prohorov distance between F_n and F converges to zero almost surely.*

PROOF OF LEMMA 3: It is equivalent to show that $\underline{V}_n^C - |V_\mu(X_{n+1})| \to 0$ in probability, for which it sufficient to show

$$\sup_{u\in B} \left|V_{\mu,n}(u) - V_\mu(X_{n+1})\right| \to 0 \tag{5}$$

in probability. This will be shown by proving the following two statements:

$$\sup_{u\in B} \left|V_{\mu,n}(u) - V_{\mu,n}(X_{n+1})\right| \to 0 \text{ in probability ;} \tag{6}$$

$$\left|V_{\mu,n}(X_{n+1}) - V_\mu(X_{n+1})\right| \to 0 \text{ in probability .} \tag{7}$$

The left-hand side of (6) is bounded above by $(C/n) \sup_{u \in B} |V'_{\mu,n}(u)|$. By Lemma 2, for any $\varepsilon > 0$, the probability of this exceeding $C\varepsilon$ goes to zero as $n \to \infty$. This establishes (6).

For (7) we observe, using Dominated Convergence, that under the finite 1-energy condition,

$$\mathcal{E}^K(\mu) := \int \int \frac{1}{|z-w|} \mathbf{1}_{|z-w|^{-1} \geq K} \, d\mu(z) \, d\mu(w) \to 0$$

as $K \to \infty$. Define $\phi^{K,z}$ by

$$\phi^{K,z}(w) = \frac{1}{z-w} \frac{|z-w|}{\max\{|z-w|, 1/K\}}$$

in other words, it agrees with $1/(z-w)$ except that we multiply by a nonnegative real so as to truncate the magnitude at K. We observe for later use that

$$\left| \phi^{K,z}(w) - \frac{1}{|z-w|} \right| \leq \frac{1}{|z-w|} \mathbf{1}_{|z-w|^{-1} \geq K}$$

so that

$$\int \int \left| \phi^{K,z}(w) - \frac{1}{|z-w|} \right| d\mu(z) \, d\mu(w) \leq \mathcal{E}^K(\mu) \to 0. \qquad (8)$$

We now introduce the truncated potential and truncated empirical potential with respect to $\phi^{K,z}$:

$$V_\mu^K(z) := \int \phi^{K,z}(w) \, d\mu(w)$$

$$V_{\mu,n}^K(z) := \int \phi^{K,z}(w) \, d\mathcal{Z}(f_n)(w).$$

We claim that

$$\mathbb{E} \left| V_\mu^K(X_{n+1}) - V_\mu(X_{n+1}) \right| \leq \mathcal{E}^K(\mu). \qquad (9)$$

Indeed,

$$V_\mu(X_{n+1}) - V_\mu^K(X_{n+1}) = \int \left(\frac{1}{z - X_{n+1}} - \phi^{K,z}(X_{n+1}) \right) d\mu(z)$$

so taking an absolute value inside the integral, then integrating against the law of X_{n+1} and using (8) proves (9). The empirical distribution $V_{\mu,n}$ has mean μ and is independent of X_{n+1}, therefore the same argument proves

$$\mathbb{E} \left| V_{\mu,n}^K(X_{n+1}) - V_{\mu,n}(X_{n+1}) \right| \leq \mathcal{E}^K(\mu) \qquad (10)$$

independent of the value of n.

We now have two thirds of what we need for the triangle inequality. That is, to show (7) we will show that the following three expressions may all be made smaller than ε with probability $1 - \varepsilon$.

$$V_{\mu,n}(X_{n+1}) - V_{\mu,n}^K(X_{n+1})$$

$$V_{\mu,n}^K(X_{n+1}) - V_\mu^K(X_{n+1})$$

$$V_\mu^K(X_{n+1}) - V_\mu(X_{n+1})$$

Choosing K large enough so that $\mathcal{E}^K(\mu) < \varepsilon^2$, this follws for the third of these follows by (9) and for the first of these by (10). Fixing this value of K, we turn to the middle expression. The function $\phi^{K,z}$ is bounded and continuous. By the Corollary 1 to the Glivenko-Cantelli Theorem 5, the empirical law $\mathcal{Z}(f_n)$ converges weakly to μ, meaning that the integral of any bounded continuous function ϕ against $\mathcal{Z}(f_n)$ converges in probability to the integral of ϕ against μ. Setting $\phi := \phi^{K,z}$ and $z := X_{n+1}$ proves that $V_{\mu,n}^K(X_{n+1}) - V_\mu^K(X_{n+1})$ goes to zero in probability, establishing the middle statement (it is in fact true conditionally on X_{n+1}) and concluding the proof. $\qquad\square$

PROOF OF THEOREM 1: Suppose that $\underline{V}_n^C(X_{n+1}) > 1/C$. Then for all w with $|w - X_{n+1}| \leq C/n$, we have

$$f_n'(w) = \sum_{j=1}^{n} \frac{1}{w - X_j} = nV_{\mu,n}(w) \geq \frac{n}{C}$$

and hence

$$\left| f_n'(w) \right| = n \left| V_{\mu,n}(w) \right| \geq n\underline{V}_n^C(X_{n+1}) \geq \frac{n}{C}.$$

To apply Rouché's Theorem to the functions $1/f_n'$ and $z - X_{n+1}$ on the disk $B := B_{C/n}(X_{n+1})$ we note that $|1/f_n'| < C/n = |z - X_{n+1}|$ on ∂B and hence that the sum has precisely one zero in B, call it a_{n+1}. Taking reciprocals we see that a_{n+1} is also the unique value in $z \in B$ for which $f_n'(z) = -1/(z - X_{n+1})$. But $f_n'(z) + 1/(z - X_{n+1}) = f_{n+1}'(z)$, whence f_{n+1}' has the unique zero a_{n+1} on B.

Now fix any $\delta > 0$. Using the hypothesis that $\mu\{z : V_\mu(z) = 0\} = 0$, we pick a $C > 0$ such that $\mathbb{P}(|V_\mu(X_{n+1})| \leq 2/C) \leq \delta/2$. By Lemma 3, there is an n_0 such that for all $n \geq n_0$,

$$\mathbb{P}\left(\underline{V}^C(X_{n+1}) \leq \frac{1}{C} \right) \leq \delta.$$

It follows that the probability that f_{n+1}' has a unique zero a_{n+1} in B is at least $1 - \delta$ for $n \geq n_0$. By symmetry, we see that for each j, the probability is also at least $1 - \delta$ that f_{n+1}' has a unique zero, call it a_j, in the ball of radius C/n centered at X_j; equivalently, the expected number of $j \leq n+1$ for which there is not a unique zero of f_{n+1}' in $B_{C/n}(X_j)$ is at most δn for $n \geq n_0$.

Define x_j to equal a_j if f'_{n+1} has a unique root in $B_{C/n}(X_j)$ and the minimum distance from X_j to any X_i with $i \leq n+1$ and $i \neq j$ is at least $2C/n$. By convention, we define x_j to be the symbol Δ if either of these conditions fails. The values x_j other than Δ are distinct roots of f'_{n+1} and each such value is within distance C/n of a different root of f_{n+1}. Using part *(ii)* of Lemma 1 we see that the expected number of j for which $x_j = \Delta$ is $o(n)$. It follows that $\mathbb{P}(|\mathcal{Z}(f_{n+1}) - \mathcal{Z}(f'_{n+1})|_P \geq 2\delta) \to 0$ as $n \to \infty$. But also the Prohorov distance between $\mathcal{Z}(f_{n+1})$ and μ converges to zero by Corollary 2. The Prohorov distance metrizes convergence in distribution and $\delta > 0$ was arbitrary, so the theorem is proved.   □

4 Proof of remaining theorems

Let $\mathcal{G} := \sum_{j=0}^{\infty} Y_j z^j$ denote the standard complex Gaussian power series where $\{Y_j(\omega)\}$ are IID standard complex normals. The results we require from [21] are as follows.

Proposition 1 ([21]). *The set of zeros of $\mathcal{G}$ in the unit disk is a determinantal point process with joint intensities*

$$p(z_1, \ldots, z_n) = \pi^{-n} \det \left[\frac{1}{(1 - z_i \bar{z}_j)^2} \right] .$$

The number $N(\rho)$ of zeros of $\mathcal{G}$ on B_ρ is distributed as the sum of independent Bernoullis with means ρ^{2k}, $1 \leq k < \infty$.   □

To use these results we broaden them to random series whose coefficients are nearly IID Gaussian.

Lemma 4. *Let $\{g_n := \sum_{r=0}^{\infty} a_{nr} z^r\}$ be a sequence of power series. Suppose*

(i) for each k, the k-tuple $(a_{n,1}, \ldots, a_{n,k})$ converges weakly as $n \to \infty$ to a k-tuple of IID standard complex normals;
(ii) $\mathbb{E}|a_{nr}| \leq 1$ for all n and r.

Then on each disk B_ρ, the set $\mathcal{Z}(g_i) \cap B_\rho$ converges weakly to $\mathcal{Z}(\mathcal{G}) \cap \rho$.

PROOF: Throughout the proof we fix $\rho \in (0, 1)$ and denote $B := B_\rho$. Suppose an analytic function h has no zeros on ∂B. Denote by $||g - h||_B$ the sup norm on functions restricted to B. Note that if $h_n \to h$ uniformly on B then $\mathcal{Z}(h_n) \cap B \to \mathcal{Z}(h) \cap B$ in the weak topology on probability measures on B, provided that h has no zero on ∂B. We apply this with $h = \mathcal{G} := \sum_{j=0}^{\infty} Y_j z^j$ where $\{Y_j(\omega)\}$ are IID standard complex normals. For almost every ω, $h(\omega)$ has no zeros on ∂B. Hence given $\varepsilon > 0$ there is almost surely a $\delta(\omega) > 0$ such that $||g - \mathcal{G}||_B < \delta$ implies $|\mathcal{Z}(g) - \mathcal{Z}(\mathcal{G})|_P < \varepsilon$. Pick $\delta_0(\varepsilon)$ small enough so that $\mathbb{P}(\delta(\omega) \leq \delta_0) < \varepsilon/3$; thus $||g - \mathcal{G}||_B < \delta_0$ implies $|\mathcal{Z}(g) - \mathcal{Z}(\mathcal{G})| < \varepsilon$ for all $\mathcal{G}$ outside a set of measure at most $\varepsilon/3$.

By hypothesis (*ii*),

$$\mathbb{E}\left|\sum_{r=k+1}^{\infty} a_{nr}z^r\right| \leq \frac{\rho^{k+1}}{1-\rho}.$$

Thus, given $\varepsilon > 0$, once k is large enough so that $\rho^{k+1}/(1-\rho) < \varepsilon\delta_0(\varepsilon)/6$, we see that

$$\mathbb{P}\left(\left|\sum_{r=k+1}^{\infty} a_{nr}z^r\right| \geq \frac{\delta_0(\varepsilon)}{2}\right) \leq \frac{\varepsilon}{3}.$$

For such a $k(\varepsilon)$ also $|\sum_{r=k+1}^{\infty} Y_r z^r| \leq \varepsilon/3$. By hypothesis (*i*), given $\varepsilon > 0$ and the corresponding $\delta(\varepsilon)$ and $k(\varepsilon)$, we may choose n_0 such that $n \geq n_0$ implies that the law of $(a_{n1},\dots,a_{nk})$ is within $\min\{\varepsilon/3, \delta_0(\varepsilon)/(2k)\}$ of the product of k IID standard complex normals in the Prohorov metric. By the equivalence of the Prohorov metric to the minimal Ky Fan metric, there is a pair of random variables $\tilde{g}$ and $\tilde{h}$ such that $\tilde{g} \sim g_n$ and $\tilde{h} \sim \mathcal{G}$ and, except on a set of of measure $\varepsilon/3$, each of the first k coefficients of $\tilde{g}$ is within $\delta_0/(2k)$ of the corresponding coefficient of $\mathcal{G}$. By the choice of $k(\varepsilon)$, we then have

$$\mathbb{P}(||\tilde{g} - \tilde{h}||_B \geq \delta_0) \leq \frac{2\varepsilon}{3}.$$

By the choice of δ_0, this implies that

$$\mathbb{P}(|\mathcal{Z}(\tilde{g}) - \mathcal{Z}(\tilde{h})|_P \geq \varepsilon) < \varepsilon.$$

Because $\tilde{g} \sim g_n$ and $\tilde{h} \sim \mathcal{G}$, we see that the law of $\mathcal{Z}(g_n) \cap B$ and the law of $\mathcal{Z}(\mathcal{G}) \cap B$ are within ε in the Prohorov metric on laws on measures. Because $\varepsilon > 0$ was arbitrary, we see that the law of $\mathcal{Z}(g_n) \cap B$ converges to the law of $\mathcal{Z}(\mathcal{G}) \cap B$. $\qquad\square$

PROOF OF THEOREM 3: Let $\rho < 1$ be fixed for the duration of this argument and denote $B := B_\rho$. Let

$$g_n(z) := \frac{f_n'(z)}{f(z)} = \sum_{j=1}^{n} \frac{1}{z - X_j}.$$

Because $|X_j| = 1$, the rational function $1/(z - X_j) = -X_j^{-1}/(1 - X_j^{-1}z)$ is analytic on the open unit disk and represented there by the power series $-\sum_{r=0}^{\infty} X_j^{-r-1}z^r$. It follows that $-g_n/\sqrt{n}$ is analytic on the open unit disk and represented there by the power series $-g_n(z)/\sqrt{n} = \sum_{r=0}^{\infty} a_{nr}z^r$ where

$$a_{nr} = n^{-1/2}\sum_{j=1}^{n} X_j^{-r-1}.$$

The function $-g_n/\sqrt{n}$ has the same zeros on B as does f_n', the normalization by $-1/\sqrt{n}$ being inserted as a convenience for what is about to come.

We will apply Lemma 4 to the sequence $\{g_n\}$. The coefficients a_{nj} are normalized power sums of the variables $\{X_j\}$. For each $r \geq 0$ and each j, the variable X_j^{-r-1} is uniformly distributed on the unit circle. It follows that $\mathbb{E}a_{nr} = 0$ and that $\mathbb{E}a_{nr}\overline{a_{nr}} = n^{-1}\sum_{ij}X_i^{-r-1}\overline{X_j}^{-r-1} = n^{-1}\sum_{ij}\delta_{ij} = 1$. In particular, $\mathbb{E}|a_{nr}| \leq (\mathbb{E}|a_{nr}|^2)^{1/2} = 1$, satisfying the second hypothesis of Lemma 4. For the first hypothesis, fix k, let $\theta_j = \mathrm{Arg}(X_j)$, and let $\mathbf{v}^{(j)}$ denote the $(2k)$-vector $(\cos(\theta_j), -\sin(\theta_j), \cos(2\theta_j), -\sin(2\theta_j), \ldots, \cos(k\theta_j), -\sin(k\theta_j))$; in other words, $\mathbf{v}^{(j)}$ is the complex k-vector $(X_j^{-1}, X_j^{-2}, \ldots, X_j^{-k})$ viewed as a real $(2k)$-vector. For each $1 \leq s, t \leq 2k$ we have $\mathbb{E}\mathbf{v}_s^{(j)}\mathbf{v}_t^{(j)} = (1/2)\delta_{ij}$. Also the vectors $\{\mathbf{v}^{(j)}\}$ are independent as j varies. It follows from the multivariate central limit theorem (see, e.g., [10, Theorem 2.9.6]) that $\mathbf{u}^{(n)} := n^{-1/2}\sum_{j=1}^n \mathbf{v}^{(j)}$ converges to $1/\sqrt{2}$ times a standard $(2k)$-variate normal. For $1 \leq r \leq k$, the coefficient a_{nr} is equal to $\mathbf{u}_{2r-1}^{(n)} + i\mathbf{u}_{2r}^{(n)}$. Thus $\{a_{nr} : 1 \leq r \leq k\}$ converges in distribution as $n \to \infty$ to a k-tuple of IID standard complex normals. The hypotheses of Lemma 4 being verified, the theorem now follows from Proposition 1. □

References

1. Abdul Aziz. On the zeros of a polynomial and its derivative. *Bull. Austral. Math. Soc.*, 31(2):245–255, 1985.
2. J. Bak and D. Newman. *Complex Analysis*. undergraduate Texts in Mathematics. Springer-Verlag, Berlin, 1982.
3. Hubert E. Bray. On the Zeros of a Polynomial and of Its Derivative. *Amer. J. Math.*, 53(4):864–872, 1931.
4. A. Bharucha-Reid and M. Sambandham. *Random Polynomials*. Academic Press, Orlando, FL, 1986.
5. Branko Ćurgus and Vania Mascioni. A contraction of the Lucas polygon. *Proc. Amer. Math. Soc.*, 132(10):2973–2981 (electronic), 2004.
6. N. G. de Bruijn. On the zeros of a polynomial and of its derivative. *Nederl. Akad. Wetensch., Proc.*, 49:1037–1044 = Indagationes Math. 8, 635–642 (1946), 1946.
7. N. G. de Bruijn and T. A. Springer. On the zeros of a polynomial and of its derivative. II. *Nederl. Akad. Wetensch., Proc.*, 50:264–270 = Indagationes Math. 9, 458–464 (1947), 1947.
8. Dimitar K. Dimitrov. A refinement of the Gauss-Lucas theorem. *Proc. Amer. Math. Soc.*, 126(7):2065–2070, 1998.
9. Janusz Dronka. On the zeros of a polynomial and its derivative. *Zeszyty Nauk. Politech. Rzeszowskiej Mat. Fiz.*, (9):33–36, 1989.
10. R. Durrett. *Probability: Theory and Examples*. Duxbury Press, Belmont, CA, third edition, 2004.
11. Kenneth Falconer. *Fractal geometry*. John Wiley & Sons Inc., Hoboken, NJ, second edition, 2003. Mathematical foundations and applications.
12. A. W. Goodman, Q. I. Rahman, and J. S. Ratti. On the zeros of a polynomial and its derivative. *Proc. Amer. Math. Soc.*, 21:273–274, 1969.
13. Andreas Hofinger. The metrics of Prokhorov and Ky Fan for assessing uncertainty in inverse problems. *Österreich. Akad. Wiss. Math.-Natur. Kl. Sitzungsber. II*, 215:107–125 (2007), 2006.

14. André Joyal. On the zeros of a polynomial and its derivative. *J. Math. Anal. Appl.*, 26:315–317, 1969.

15. N.L. Komarova and I. Rivin. Harmonic mean, random polynomials and stochastic matrices. *Advances in Applied Mathematics*, 31(2):501–526, 2003.

16. K. Mahler. On the zeros of the derivative of a polynomial. *Proc. Roy. Soc. Ser. A*, 264:145–154, 1961.

17. S. M. Malamud. Inverse spectral problem for normal matrices and the Gauss-Lucas theorem. *Trans. Amer. Math. Soc.*, 357(10):4043–4064 (electronic), 2005.

18. M. Marden. *Geometry of Polynomials*, volume 3 of *Mathematical Surveys and Monographs*. AMS, 1949.

19. M. Marden. Conjectures on the critical points of a polynomial. *Amer. Math. Monthly*, 90(4):267–276, 1983.

20. Piotr Pawlowski. On the zeros of a polynomial and its derivatives. *Trans. Amer. Math. Soc.*, 350(11):4461–4472, 1998.

21. Y. Peres and B. Virag. Zeros of the i.i.d. Gaussian power series: a conformally invariant determinantal process. *Acta Math.*, 194:1–35, 2005.

22. Q. I. Rahman. On the zeros of a polynomial and its derivative. *Pacific J. Math.*, 41:525–528, 1972.

23. Bl. Sendov. Hausdorff geometry of polynomials. *East J. Approx.*, 7(2):123–178, 2001.

24. Bl. Sendov. New conjectures in the Hausdorff geometry of polynomials. *East J. Approx.*, 16(2):179–192, 2010.

25. È. A. Storozhenko. On a problem of Mahler on the zeros of a polynomial and its derivative. *Mat. Sb.*, 187(5):111–120, 1996.

26. Q. M. Tariq. On the zeros of a polynomial and its derivative. II. *J. Univ. Kuwait Sci.*, 13(2):151–156, 1986.