ON THE LOCAL PAIRING BEHAVIOR OF CRITICAL POINTS AND ROOTS OF RANDOM POLYNOMIALS

SEAN O'ROURKE AND NOAH WILLIAMS

Abstract. We study the pairing between zeros and critical points of the polynomial $p_n(z) = \prod_{j=1}^n (z - X_j)$, whose roots $X_1, \dots, X_n$ are complex-valued random variables. Under a regularity assumption, we show that if the roots are independent and identically distributed, the Wasserstein distance between the empirical distributions of roots and critical points of p_n is on the order of $1/n$, up to logarithmic corrections. The proof relies on a careful construction of disjoint random Jordan curves in the complex plane, which allow us to naturally pair roots and nearby critical points. In addition, we establish asymptotic expansions to order $1/n^2$ for the locations of the nearest critical points to several fixed roots. This allows us to describe the joint limiting fluctuations of the critical points as n tends to infinity, extending a recent result of Kabluchko and Seidel. Finally, we present a local law that describes the behavior of the critical points when the roots are neither independent nor identically distributed.

1. Introduction

This paper concerns the nature of the pairing between the critical points and roots of random polynomials in a single complex variable. In particular, we consider polynomials of the form

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

where $X_1, \dots, X_n$ are complex-valued random variables (not necessarily independent or identically distributed). While much is known about the locations of the critical points of p_n when the roots are deterministic (see for example Marden's book [21] which contains the Gauss–Lucas theorem and Walsh's two circle theorem among other results), Pemantle and Rivin [25], Hanin [12, 13, 14], and Kabluchko [17] first demonstrated that the random version of this problem admits greater precision, especially when the degree n is large.

In particular, Pemantle and Rivin conjectured that when $X_1, \dots, X_n$ are chosen to be independent and identically distributed (iid) with distribution μ, then the empirical distribution constructed from the critical points of p_n converges weakly in probability to μ. They proved their conjecture in [25] for measures satisfying some technical assumptions, and Subramanian [30] refined their work for $\{X_j\}_{j=1}^n$ on the unit circle. Kabluchko first proved the conjecture in full generality in [17] to obtain the following result.

S. O'Rourke has been supported in part by NSF grants ECCS-1610003 and DMS-1810500.

Theorem 1.1 (Kabluchko [17]). *Let $X_1, X_2, \dots$ be iid complex-valued random variables with distribution μ. Then for any bounded and continuous function $\varphi : \mathbb{C} \to \mathbb{C}$,*

$$\frac{1}{n-1}\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) \longrightarrow \int \varphi(z) d\mu(z)$$

in probability as $n \to \infty$, where $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ are the critical points of the polynomial $p_n(z) = \prod_{j=1}^{n}(z - X_j)$.

Inspired by such results, the first author established several versions of Theorem 1.1 for random polynomials with dependent roots that satisfy some technical conditions [22]. For example, the conclusion of Theorem 1.1 holds for characteristic polynomials of certain classes of matrices from the classical compact matrix groups. Additionally, in [23], the authors adapted Kabluchko's strategy to the situation where p_n is perturbed to have $o(n)$ deterministic roots. Two other relevant works include Reddy's thesis [28] and the recent paper of Byun, Lee, and Reddy [3], who showed that under some mild assumptions, Kabluchko's result holds when p_n has mostly deterministic roots and several (potentially dependent) random ones. Byun, Lee, and Reddy proved several additional results including that the sequence of empirical measures constructed from the zeros of $p_n^{(k)}$ converges weakly in probability to the distribution μ, for *any* fixed choice of k, as well as a version of Theorem 1.1 when the roots $X_1, \dots, X_n$ are given by a $2D$ Coulomb gas density.

Theorem 1.1 and most of the cited works above focus on the macroscopic, or global, behavior of the critical points of p_n. For example, by combining Theorem 1.1 with the Law of Large Numbers, one obtains that, for any bounded and continuous function $\varphi : \mathbb{C} \to \mathbb{C}$,

$$\sum_{j=1}^{n} \varphi(X_j) = \sum_{j=1}^{n-1} \varphi(w_j^{(n)}) + o(n) \tag{2}$$

with high probability[1]. In contrast to Theorem 1.1, this paper focuses on describing the local behavior of the critical points.

One important aspect of the local critical point behavior is that the critical points and roots of p_n appear to pair with one another. Theorem 1.1 and (2) describe this phenomenon at the macroscopic level by comparing the global behaviors of the critical points and roots. However, a glance at Figures 1 and 2 suggests that a stronger pairing phenomenon exists. In particular, one sees that nearly every critical point is paired closely with a root of p_n, an indication that the local behavior of the critical points should be extremely similar to the local behavior of the roots.

Hanin investigated the pairing phenomenon between roots and critical points for several classes of random functions [12, 13, 14], including random polynomials with independent roots. He proved that the distance between a fixed, deterministic root and its nearest critical point is roughly $1/n$ in the case where μ has a bounded density supported on the Riemann sphere [14]. The root-and-critical point pairing for random polynomials was also explored in [23, 24], and Dennis and Hannay gave an electrostatic explanation of the phenomenon in [7]. Most recently, Steinerberger showed that the pairing phenomenon also holds for some classes of deterministic polynomials [29], and Kabluchko and Seidel determined the asymptotic fluctuations

[1]See Section 1.1 for a complete description of the asymptotic notation used here and in the sequel.

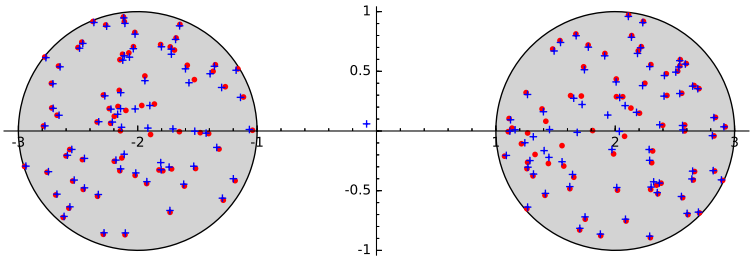

FIGURE 1. The roots (red dots) and critical points (blue crosses) of a random, degree 150 polynomial, where all 150 roots are chosen independently and uniformly on two disks. See Example 2.7.

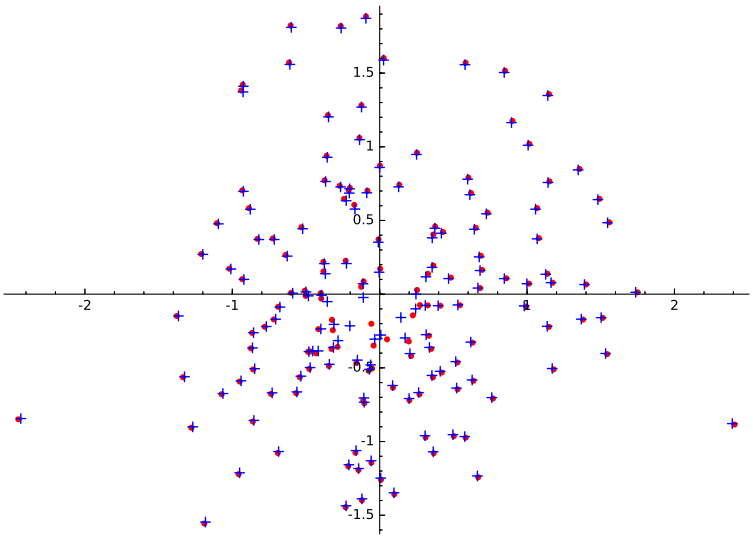

FIGURE 2. The roots (red dots) and critical points (blue crosses) of a random, degree 150 polynomial, where all 150 roots are chosen independently according to a standard complex normal distribution. See Example 2.6.

of the critical point of p_n that is nearest a given root [18]. Kabluchko and Seidel's results are similar to some of our conclusions below and appear to have been concurrently derived using different methods. We present a detailed comparison between [18] and our work in the next section.

In this paper, we refine the results mentioned above to obtain a more complete picture of the pairing that occurs between zeros and critical points of p_n. We begin by exhibiting a bound on the Wasserstein, or "transport," distance between the collections of roots and critical points of p_n. While this result explains the nearly 1-1 pairing between roots and critical points in Figures 1 and 2, it does not allow one to describe the behavior near any particular root. We accomplish this feat in the next section of the paper, where we discuss the joint fluctuations for a fixed

number of critical points of p_n. We conclude our analysis by establishing a local law that describes the mesoscopic behavior of the critical points of p_n. Many of our results focus on the cases where the roots $X_1, \ldots, X_n$ of p_n are iid, but for some of our results, we do not even require that the roots be independent.

1.1. **Notation.** Throughout the paper, we use asymptotic notation, such as O and o, under the assumption that $n \to \infty$. We write $X_n = O(Y_n)$, $Y_n = \Omega(X_n)$, $X_n \ll Y_n$, or $Y_n \gg X_n$ to denote the bound $|X_n| \leq CY_n$ for some constant $C > 0$ and for all $n > C$. If the implicit constant depends on a parameter k, e.g., $C = C_k$, we denote this with subscripts, e.g., $X_n = O_k(Y_n)$ or $X_n \ll_k Y_n$. By $X_n = o_k(Y_n)$, we mean that for any $\varepsilon > 0$, there is a natural number $N_{\varepsilon,k}$ depending on k and ε for which $n \geq N_{\varepsilon,k}$ implies $|X_n| \leq \varepsilon Y_n$. In general, C, c, K are constants which may change from one occurrence to the next. We often use subscripts, such as $C_{P_1,P_2,\ldots}$, to denote that the constant depends on some parameters $P_1, P_2, \ldots$.

We use the following set-theoretic conventions. For $z_0 \in \mathbb{C}$ and $r \geq 0$, we define

$$B(z_0, r) := \{z \in \mathbb{C} : |z - z_0| < r\}$$

to be the open ball of radius r centered at z_0, and $\overline{B}(z_0, r)$ to be its closure. The notations $\#S$ and $|S|$ denote the cardinality of the finite set S. The natural numbers, $\mathbb{N}$, do not include zero.

For a probability measure μ, we use $X \sim \mu$ to mean that the random variable X has distribution μ and $\mathrm{supp}(\mu)$ to denote its support. We say that a probability measure μ on $\mathbb{C}$ has *density* f if μ is absolutely continuous with respect to Lebesgue measure on $\mathbb{C}$ and the Radon–Nikodym derivative of μ with respect to Lebesgue measure is f. The random variable $\mathbb{1}_E$ is the indicator supported on the event E, and we say an event E (which depends on n) holds with *overwhelming probability* if for every $\alpha > 0$, $\mathbb{P}(E) \geq 1 - O_\alpha(n^{-\alpha})$.

Finally, we use d^2z to denote integration with respect to the Lebesgue measure on $\mathbb{C}$ to avoid confusion with complex line integrals, where we integrate against dz. We use $\sqrt{-1}$ to denote the imaginary unit and reserve i as an index.

Acknowledgements. The authors thank Boris Hanin for calling their attention to this line of research and for many useful conversations. The authors also thank the anonymous referees for useful feedback and corrections.

2. Main Results

We begin by introducing the Wasserstein metric in order to discuss the pairing between the roots and critical points of p_n that one sees in Figures 1 and 2.

2.1. **Wasserstein distance.** For probability measures μ and ν on $\mathbb{C}$, let $W_1(\mu, \nu)$ denote the L_1-Wasserstein distance between μ and ν defined by

$$W_1(\mu, \nu) := \inf_{\pi} \int |x - y| d\pi(x, y),$$

where the infimum is over all probability measures π on $\mathbb{C} \times \mathbb{C}$ with marginals μ and ν (see e.g. [37], Chapter 6). Theorem 2.3 below gives a bound on the Wasserstein distance between the empirical measures constructed from the roots $X_1, \ldots, X_n$

and the critical points $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ of the polynomial p_n defined in (1). We denote these empirical measures by

$$\mu_n := \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j} \quad \text{and} \quad \mu_n' := \frac{1}{n-1}\sum_{j=1}^{n-1} \delta_{w_j^{(n)}}, \tag{3}$$

respectively. Before we state Theorem 2.3, we motivate some regularity assumptions μ must satisfy in the hypothesis.

Consider that

$$\frac{1}{n}p_n'(z) = \prod_{j=2}^{n}(z - X_j)\left((z - X_1)\frac{1}{n}\sum_{j=2}^{n}\frac{1}{z - X_j} + \frac{1}{n}\right), \tag{4}$$

where the sum on the right-hand side is an empirical mean of iid random variables. Provided μ is sufficiently nice, the Law of Large Numbers implies $\frac{1}{n}\sum_{j=2}^{n}\frac{1}{z-X_j}$ converges in distribution to the Cauchy–Stieltjes transform of μ, which is given by

$$m_\mu(z) := \int_{\mathbb{C}} \frac{d\mu(x)}{z - x}, \tag{5}$$

and defined for those values of $z \in \mathbb{C}$ for which the integral exists. Heuristically speaking, if $m_\mu(z)$ is finite and bounded away from zero near $z = X_1$, then $p_n'(z) \approx 0$ for some z satisfying $|z - X_1| = O(1/n)$. If, on the other hand, $|m_\mu(z)|$ is close to 0 for z near X_1, we have $p_n'(z) \approx \prod_{j=2}^{n}(z - X_j)$, so $p_n'(z)$ need not have any zeros near X_1. Similar heuristic intuition applies if we replace X_1 in turn with $X_2, \dots, X_n$.

In light of the discussion above, conditions on the Cauchy–Stieltjes transform of μ feature prominently in this paper, particularly in the hypothesis of Theorem 2.3, which requires at least one of the assumptions below.

Assumption 2.1. Suppose there are positive constants, C_1, C_2, so that the following conditions hold when $X_1, \dots, X_n$ are iid complex-valued random variables with common distribution μ:

(i) for any $\varepsilon > 0$, $\mathbb{P}(|m_\mu(X_1)| < \varepsilon) \le C_1\varepsilon^2$;
(ii) the random variable $\eta_n := \max_{1\le j\le n}|X_j|$ satisfies $\mathbb{P}\left(\eta_n \ge n^{C_2}\right) = o(1)$.

Assumption 2.2 (Alternative to Assumption 2.1 for radially symmetric distributions)**.** Suppose μ has two finite absolute moments and a continuous density, f, that is radially symmetric about $z = z_0$ and that satisfies $f(z_0) > 0$.

We can now state the main result of this subsection.

Theorem 2.3. *Let $X_1, \dots, X_n$ be iid, complex random variables whose distribution, μ, has a bounded density and satisfies either Assumption 2.1 or Assumption 2.2. Then, there is a positive constant C, depending on μ, so that with probability $1 - o(1)$,*

$$W_1(\mu_n, \mu_n') \le \frac{C\eta_n(\log n)^9}{n}, \tag{6}$$

where $\eta_n := \max_{1\le j\le n}|X_j|$, and μ_n, μ_n' (defined in (3)) are the empirical measures constructed from the roots and critical points of $p_n(z) = \prod_{j=1}^{n}(z - X_j)$.

In the case where μ has sub-exponential tails, one can show that with probability tending to 1, $\eta_n = O(\log n)$. Consequently, Theorem 2.3 immediately implies the following corollary.

Corollary 2.4. *Let $X_1, \dots, X_n$ be iid, complex random variables whose distribution, μ, has a bounded density and satisfies Assumption 2.1 part* (i) *in addition to the following condition:*

(ii') *there exist $C, c > 0$ such that if $X \sim \mu$, then, $\mathbb{P}(|X| > t) \leq Ce^{-ct}$ for every $t > 0$.*

Then, there is a positive constant C_μ, depending only on μ, so that with probability $1 - o(1)$,

$$W_1(\mu_n, \mu_n') \leq \frac{C_\mu (\log n)^{10}}{n},$$

where μ_n, μ_n' (defined in (3)*) are the empirical measures constructed from the roots and critical points of $p_n(z) = \prod_{j=1}^n (z - X_j)$.*

Theorem 2.3 and Corollary 2.4 show that the roots and critical points can be paired in such a way that the typical spacing between a critical point and its paired root is $O(n^{-1})$, up to logarithmic corrections. This precisely describes the phenomenon observed in Figures 1 and 2, and the authors believe that these bounds are optimal (up to logarithmic factors) based on the theorems of Section 2.3 below and the results in [18].

A couple of remarks concerning Theorem 2.3 and its corollary are in order. Due to the heuristic that motivates our proof of Theorem 2.3 (see Figure 5), the authors conjecture that Assumption 2.1 part (i) can be weakened to require that for some fixed $\delta > 0$, $\mathbb{P}(|m_\mu(X_1)| < \varepsilon) \leq C_1 \varepsilon^{1+\delta}$. At present, we require $\delta = 1$ to obtain some technical bounds in the proof. An examination of the proof reveals exactly where this condition is needed. The second remark concerns the appearance of η_n on the right-hand side of (6). The authors believe this term is at least partially necessary. Indeed, based on numerical experiments, the Wasserstein distance $W_1(\mu_n, \mu_n')$ appears larger for distributions μ with extremely heavy tails. The precise dependence of η_n on the Wasserstein distance remains an open question.

2.2. **Examples of Theorem 2.3 and Corollary 2.4.** The assumptions of Theorem 2.3 and Corollary 2.4 are rather technical, so this subsection is devoted to several specific examples worked out in detail.

Example 2.5 (μ is uniform on a disk)**.** If μ has a uniform distribution on the disk of radius R centered at z_0, then, μ has density $f(z) = \frac{1}{\pi R^2} \mathbb{1}_{|z - z_0| \leq R}$ and Cauchy–Stieltjes transform

$$m_\mu(z) = \begin{cases} \frac{1}{R^2} (\overline{z - z_0}) & \text{if } |z - z_0| \leq R, \\ \frac{1}{z - z_0} & \text{if } |z - z_0| \geq R. \end{cases}$$

(Lemma A.1 facilitates the computation of $m_\mu(z)$ when μ is radially symmetric. For this example, apply Lemma A.1 when $z = 0$, $R = 1$, and apply a linear transformation.) It follows that if $X \sim \mu$, then for any $\varepsilon < 1$,

$$\mathbb{P}\left(|m_\mu(X)| < \varepsilon\right) \leq \mathbb{P}\left(|X - z_0| < R^2 \varepsilon\right) = R^2 \varepsilon^2,$$

so μ satisfies Assumption 2.1, and by Theorem 2.3, with probability $1 - o(1)$, $W_1(\mu_n, \mu_n') = O((\log n)^9 / n)$. (Note that almost surely, $\eta_n \leq |z_0| + R$).

Example 2.6 (μ is supported on all of $\mathbb{C}$)**.** Assumption 2.2 is easy to verify for a large class of measures that do not necessarily have compact support. For example, suppose μ has a standard complex normal distribution with density $f(z) = \frac{1}{\pi} e^{-|z|^2}$.

Clearly, μ is radially symmetric about the origin, and $f(z)$ is continuous with $f(0) = \pi^{-1} > 0$. Furthermore, μ has sub-exponential tails, so by Corollary 2.4, with probability tending to 1, $W_1(\mu_n, \mu_n') \leq O((\log n)^{10}/n)$. Figure 2 illustrates this example.

Example 2.7 (μ is not radially symmetric)**.** In this last example, we consider a situation where μ does not exhibit radial symmetry. Suppose μ is uniform on the two disks $B(-2,1)$ and $B(2,1)$ with density

$$f(z) = \frac{1}{2\pi}\left(\mathbb{1}_{|z+2|<1}(z) + \mathbb{1}_{|z-2|<1}(z)\right),$$

which is depicted in Figure 1. By separately considering the cases $|z+2| < 1$, $|z-2| < 1$, and $|z \pm 2| \geq 1$, we can use the calculations from Example 2.5 to obtain the Cauchy–Stieltjes transform:

$$m_\mu(z) = \begin{cases} \frac{1}{2}\left(\overline{z+2} + \frac{1}{z-2}\right) & \text{if } |z+2| < 1, \\ \frac{z}{z^2-4} & \text{if } |z \pm 2| \geq 1, \\ \frac{1}{2}\left(\overline{z-2} + \frac{1}{z+2}\right) & \text{if } |z-2| < 1. \end{cases} \tag{7}$$

Since μ has compact support, Assumption 2.1 part (ii) holds trivially. In Appendix A, we establish part (i), so by Theorem 2.3, with probability $1-o(1)$, $W_1(\mu_n, \mu_n') = O((\log n)^9/n)$.

2.3. **Fluctuations of the critical points.** While Theorem 2.3 describes the typical distance between a root and its paired critical point, it does not allow one to study any particular root or critical point. Toward this end, we now fix several of the roots and treat them as deterministic: consider the polynomial

$$p_n(z) := \prod_{l=1}^{s}(z - \xi_l) \prod_{j=1}^{n+1-s}(z - X_j),$$

where $X_1, \ldots, X_{n+1-s}$ are iid complex-valued random variables with distribution μ, and $\vec{\xi} = (\xi_1, \ldots, \xi_s)$ is a deterministic vector in $\mathbb{C}^s$. Our goal is to simultaneously study the behavior of the critical points closest to ξ_l, $1 \leq l \leq s$.

Our first result, Theorem 2.8, covers the situation where some of the values $\xi_1, \ldots, \xi_s$ are allowed to be inside the support of μ. In particular, for each $1 \leq l \leq s$, equation (9) locates the critical point, $w_l^{(n)}$, that is near ξ_l to within $O(n^{-2})$ (up to logarithmic corrections). This bound indicates that each $w_l^{(n)}$ is centered at

$$\hat{w}_l^{(n)} := \xi_l - \frac{1}{n+1}\frac{n}{\sum_{j\neq l}\frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s}\frac{1}{\xi_l - X_j}}, \tag{8}$$

rather than ξ_l. This observation allows us to express the fluctuations of each critical point as a sum of independent random variables (up to some lower order error terms), and we use this to show that the fluctuations of the vector $(w_1^{(n)}, \ldots, w_s^{(n)})$ converge in distribution to a multivariate normal distribution. See Figure 3.

In order to state Theorem 2.8 we need the following definitions. Let

$$\mathrm{M}_\mu := \{z \in \mathbb{C} : m_\mu(z) = 0\}$$

denote the set of zeros of m_μ. We say that a measure μ has a *density in a neighborhood of* z_0 if there exists a $\rho > 0$ so that the restriction of μ to the open ball $B(z_0, \rho)$ is absolutely continuous with respect to the Lebesgue measure on $B(z_0, \rho)$.

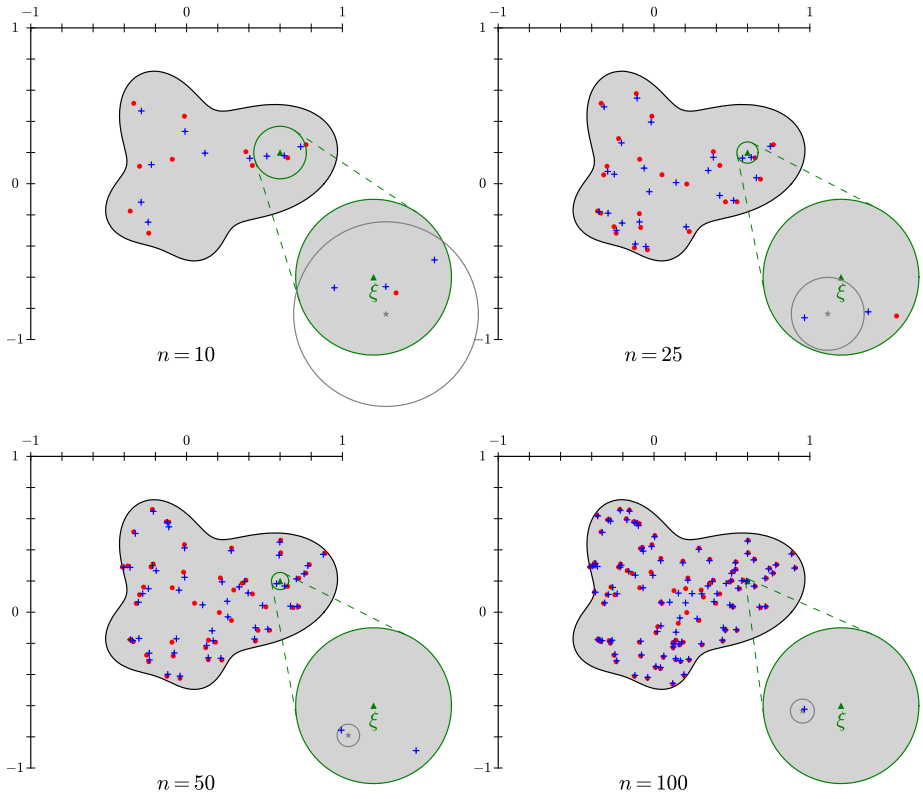

FIGURE 3. Simulation to illustrate Theorem 2.8. The roots (red dots) and critical points (blue crosses) of $p_n(z) = (z-\xi_1)\prod_{j=1}^{n}(z-X_j)$ for increasing values of n, where the roots, $X_1, \ldots, X_{100}$, are chosen independently and uniformly on the outlined region. The green circle centered at ξ_1 is of radius $\frac{2n}{n+1}\left(\sum_{j=1}^{n}\frac{1}{\xi_1 - X_j}\right)^{-1}$ and the gray circle has radius $\frac{20}{n^2}$ and center $\hat{w}_1^{(n)}$ (see (8)).

Theorem 2.8 (Locations and fluctuations of critical points when p_n has several deterministic roots)**.** *Let $X_1, X_2, \ldots$ be iid complex-valued random variables with distribution μ, fix s and the distinct, deterministic values $\xi_1, \ldots, \xi_s \notin \mathrm{M}_\mu$, and suppose that in a neighborhood of each ξ_l, $1 \le l \le s$, μ has a bounded density, f. Then, with probability $1 - o(1)$, the polynomial*

$$p_n(z) = \prod_{l=1}^{s}(z-\xi_l)\prod_{j=1}^{n+1-s}(z-X_j)$$

has s critical points, $w_1^{(n)}, \ldots, w_s^{(n)}$, such that for $1 \le l \le s$, $w_l^{(n)}$ is the unique critical point of p_n that is within a distance of $\frac{3}{|m_\mu(\xi_l)|n}$ of ξ_l, and

$$\left| w_l^{(n)} - \xi_l + \frac{1}{n+1}\frac{n}{\sum_{j\neq l}\frac{1}{\xi_l-\xi_j} + \sum_{j=1}^{n+1-s}\frac{1}{\xi_l - X_j}} \right| = O_{\mu,\vec{\xi}}\left(\left(\frac{\log n}{n}\right)^2\right). \quad (9)$$

In addition, if f is continuous at $\xi_1, \dots, \xi_s$, then we have

$$\left(\frac{n^{3/2}}{\sqrt{\log n}} \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right) \right)_{l=1}^{s} \longrightarrow (N_1, \dots, N_s) \tag{10}$$

in distribution as $n \to \infty$, where $(N_1, \dots, N_s)$ is a vector of complex random variables whose real and imaginary components $(\mathrm{Re}(N_1), \mathrm{Im}(N_1), \dots, \mathrm{Re}(N_s), \mathrm{Im}(N_s))$ have a multivariate normal distribution with mean zero and covariance structure characterized by

$$\begin{aligned} \mathrm{Cov}(\mathrm{Re}(N_j), \mathrm{Re}(N_l)) &= \begin{cases} \frac{\pi f(\xi_j)}{2} & \text{if } l = j, \\ 0 & \text{else} \end{cases} \\ \mathrm{Cov}(\mathrm{Im}(N_j), \mathrm{Im}(N_l)) &= \begin{cases} \frac{\pi f(\xi_j)}{2} & \text{if } l = j, \\ 0 & \text{else} \end{cases} \\ \mathrm{Cov}(\mathrm{Re}(N_j), \mathrm{Im}(N_l)) &= 0. \end{aligned} \tag{11}$$

Remark 2.9. Theorem 2.8 can also be extended to the case where $\xi_1, \dots, \xi_s$ are independent random variables (rather than deterministic values). This can be seen by conditioning on $\xi_1, \dots, \xi_s$ and applying Theorem 2.8; a similar argument was used in [18].

Compare Theorem 2.8 to Theorem 2.2 of [18], which describes the same phenomenon when $s = 1$. Both theorems identify the same fluctuations of $w_1^{(n)}$ about ξ_1, however, the two results locate the critical point $w_1^{(n)}$ on different scales. While Theorem 2.2 from [18] shows that $w_1^{(n)}$ is the unique critical point of p_n within a distance of order $o(1/\sqrt{n})$ of ξ_1, Theorem 2.8 refines the location of $w_1^{(n)}$ to within order $O(n^{-2})$ up to logarithmic corrections. In fact, since $\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_1 - X_j}$ converges almost surely to $m_\mu(\xi_1)$, the results of the two theorems can be combined to give a stronger picture of the local behavior of $w_1^{(n)}$. Note that in contrast to the method of proof used by Kabluchko and Seidel in [18], our approach is based on a deterministic argument (see Theorem 3.1).

For values of $\xi_1, \dots, \xi_s$ outside the support of μ, (10) and (11) demonstrate that the scaling factor $n^{3/2}/\sqrt{\log n}$ is too small to achieve a meaningful result. (Indeed, f may be chosen to be identically zero outside $\mathrm{supp}(\mu)$, so the random vector $(N_1, \dots N_s)$ is almost surely the zero vector.) The following result refines the analysis in this situation and is depicted in Figure 4.

Theorem 2.10 (Locations and Fluctuations of critical points when p_n has several roots outside $\mathrm{supp}(\mu)$)**.** *Let $X_1, X_2, \dots$ be iid complex-valued random variables with common distribution μ, fix $s \in \mathbb{N}$, and suppose $\xi_1, \dots, \xi_s \notin \mathrm{supp}(\mu) \cup \mathrm{M}_\mu$ are distinct, fixed deterministic values. Then, there exist constants $C, c_{\mu,\vec{\xi}}, C_{\mu,\vec{\xi}} > 0$, so that with probability at least $1 - C\exp(-c_{\mu,\vec{\xi}} n)$, the polynomial*

$$p_n(z) = \prod_{l=1}^{s} (z - \xi_l) \prod_{j=1}^{n+1-s} (z - X_j)$$

has s critical points, $w_1^{(n)}, \dots, w_s^{(n)}$, such that for $1 \leq l \leq s$, $w_l^{(n)}$ is the unique critical point of p_n that is within a distance of $\frac{3}{|m_\mu(\xi_l)|n}$ of ξ_l, and

$$\left| w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{n}{\sum_{j \neq l} \frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j}} \right| < \frac{C_{\mu,\vec{\xi}}}{n^2}. \tag{12}$$

In addition, we have

$$\left(n^{3/2} \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right) \right)_{l=1}^{s} \longrightarrow (N_1, \dots, N_s) \tag{13}$$

in distribution as $n \to \infty$, where $(N_1, \dots, N_s)$ is a vector of complex random variables whose real and imaginary components $(\mathrm{Re}(N_1), \mathrm{Im}(N_1), \dots, \mathrm{Re}(N_s), \mathrm{Im}(N_s))$ have a multivariate normal distribution with mean zero and covariance structure

$$\begin{aligned} \mathrm{Cov}(\mathrm{Re}(N_j), \mathrm{Re}(N_l)) &= \mathrm{Cov}\left(\mathrm{Re}\left(\frac{1}{\xi_j - X_1} \right), \mathrm{Re}\left(\frac{1}{\xi_l - X_1} \right) \right) \\ \mathrm{Cov}(\mathrm{Im}(N_j), \mathrm{Im}(N_l)) &= \mathrm{Cov}\left(\mathrm{Im}\left(\frac{1}{\xi_j - X_1} \right), \mathrm{Im}\left(\frac{1}{\xi_l - X_1} \right) \right) \\ \mathrm{Cov}(\mathrm{Re}(N_j), \mathrm{Im}(N_l)) &= \mathrm{Cov}\left(\mathrm{Re}\left(\frac{1}{\xi_j - X_1} \right), \mathrm{Im}\left(\frac{1}{\xi_l - X_1} \right) \right). \end{aligned} \tag{14}$$

Remark 2.11. After an application of the Borel–Cantelli lemma, Theorem 2.10 can be combined with Theorem 2.9 of [23] to establish the following: when μ has compact support, almost surely, for n sufficiently large, $w_1^{(n)}, \dots, w_s^{(n)}$, characterized by (12), are the only critical points of p_n outside an ε-neighborhood of $\mathrm{supp}(\mu) \cup M_\mu$.

In Section 3, we provide a generalization of Theorem 2.10 to a situation where p_n has a number of deterministic roots that may depend on n (see Theorem 3.4 below). The proofs of Theorems 2.8 and 2.10 are based on a technical, deterministic argument that applies to cases where $X_1, \dots, X_n$ are random variables that are not independent (see Theorem 3.1). To illustrate this point, we conclude the subsection with a result that demonstrates pairing between individual roots and critical points of p_n when p_n is the characteristic polynomial of a random matrix.

Theorem 2.12. *Fix $\varepsilon > 0$ and $\lambda \in \mathbb{C}$ with $|\lambda| \geq 1 + 3\varepsilon$. Let M be an $n \times n$ random matrix whose entries are iid copies of a random variable with mean zero, unit variance, and finite fourth moment. Let A be an $n \times n$ deterministic matrix with operator norm $\|A\| = O(1)$, $\mathrm{rank}(A) = O(1)$[2], and whose only nonzero eigenvalue is λ. Then almost surely, for n sufficiently large, the characteristic polynomial*[3]

$$p_n(z) := \det\left(zI - \frac{1}{\sqrt{n}} M - A \right)$$

of $\frac{1}{\sqrt{n}} M + A$ has a factorization $p_n(z) = (z - \xi) \prod_{i=1}^{n-1} (z - X_i)$, where

(i) The roots $X_1, \dots, X_{n-1}$ lie inside the disk $B(0, 1 + 2\varepsilon)$.
(ii) The root ξ lies outside the disk $B(0, 1 + 2\varepsilon)$ and satisfies $\xi = \lambda + o(1)$.

[2]We continue to use asymptotic notation, such as O and o, under the assumption that $n \to \infty$. In this theorem, n represents the dimension of the matrices M and A.

[3]Here, I denotes the identity matrix.

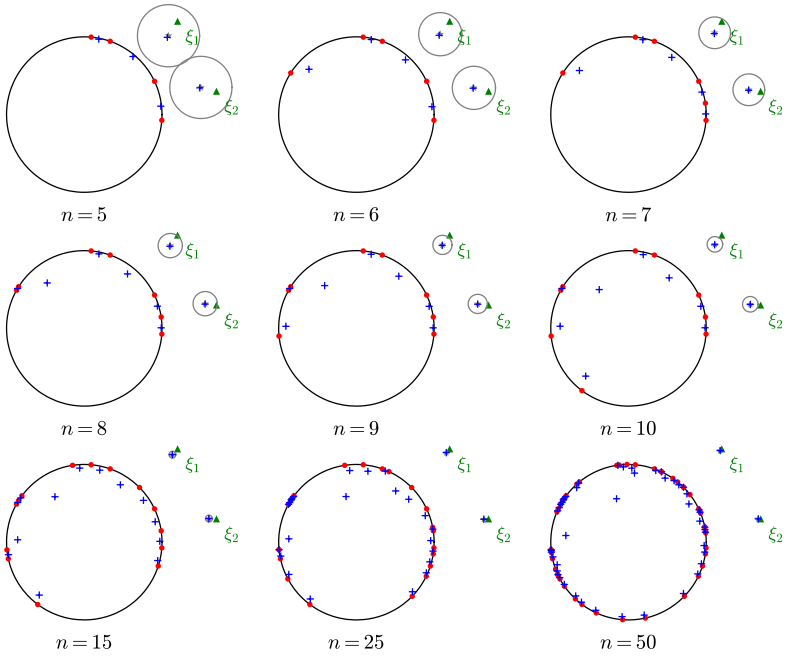

FIGURE 4. Simulation to illustrate Theorem 2.10. The roots (red circles) and critical points (blue crosses) of $p_n(z) = \prod_{l=1}^{2}(z - \xi_l)\prod_{j=1}^{n-1}(z - X_j)$ for increasing values of n, where the roots, $X_1, \ldots, X_{49}$, are chosen independently and uniformly on the unit circle. The gray circles are of radius $\frac{10}{n^2}$ and are centered at $\hat{w}_1^{(n)}$, $\hat{w}_2^{(n)}$ defined in (8).

(iii) p_n contains a unique critical point, $w_\xi^{(n-1)}$, which satisfies

$$\left| w_\xi^{(n-1)} - \xi + \frac{1}{n} \cdot \frac{1}{\frac{1}{n-1}\sum_{i=1}^{n-1} \frac{1}{\xi - X_i}} \right| = O_{\lambda,\varepsilon}\left(\frac{1}{n^2}\right). \tag{15}$$

and hence

$$w_\xi^{(n-1)} = \lambda + o(1) \tag{16}$$

Remark 2.13. The conclusion in (16) can be deduced from properties (i) and (ii) and Walsh's two circle theorem (see, for example, [27, Theorem 4.1.1]). However, the conclusion in (15) cannot be deduced from Walsh's two circle theorem and instead follows from Theorem 3.1. We prove Theorem 2.12 in Appendix A.

2.4. **A local law for the critical points.** In this subsection, we consider a local law that describes the behavior of the critical points of

$$p_n(z) := \prod_{i=1}^{n}(z - X_i).$$

We begin with the case where $X_1, \dots, X_n$ are arbitrary random variables (not assumed to be independent nor identically distributed) and then specialize our main result to several applications and examples.

Theorem 2.14 (Local law)**.** *Fix $C > 0$, and let $X_1, \dots, X_n$ be complex-valued random variables (not necessarily independent nor identically distributed) which satisfy the following axioms.*

(i) (Upper bound) With overwhelming probability, $\max_{1 \le i \le n} |X_i| \le e^{n^C}$,
(ii) (Anti-concentration) For every $a > 0$, there exists $b > 0$ such that

$$\left| \sum_{i=1}^{n} \frac{1}{Z - X_i} \right| \ge n^{-b} \tag{17}$$

with probability $1 - O_a(n^{-a})$, where Z is uniformly distributed on $B(0, n^C)$, independent of $X_1, \dots, X_n$.

Let $\varphi : \mathbb{C} \to \mathbb{R}$ be a twice continuously differentiable function (possibly depending on n) which is supported on $B(0, n^C)$ and which satisfies the pointwise bound

$$|\Delta \varphi(z)| \le n^C \tag{18}$$

for all $z \in \mathbb{C}$. Then, for every fixed $c > 0$ and every $\alpha > 0$,

$$\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) = \sum_{i=1}^{n} \varphi(X_i) + O_\alpha(\| \Delta \varphi \|_1 \log n) + O_\alpha(n^{-c})$$

with probability $1 - O_\alpha(n^{-\alpha})$, where $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ are the critical points of the polynomial

$$p_n(z) := \prod_{i=1}^{n} (z - X_i)$$

and $\| \Delta \varphi \|_1$ is the L_1-norm of $\Delta \varphi$. Here, the implicit constants in our asymptotic notation depend on C, c, and α.

Remark 2.15. Condition (ii) on the random variables $X_1, \dots, X_n$ from Theorem 2.14 is implied by the following:

(ii') for every $a > 0$, there exists $b > 0$ such that, for almost every $z \in B(0, n^C)$,

$$\left| \sum_{i=1}^{n} \frac{1}{z - X_i} \right| \ge n^{-b}$$

with probability $1 - O_a(n^{-a})$.

Indeed, the implication follows by simply conditioning on the random variable Z (which avoids a set of Lebesgue measure zero with probability 1).

The assumptions of Theorem 2.14 are fairly technical, and we derive some simpler conditions that guarantee when the hypotheses of Theorem 2.14 are met in Section 2.5. We now specialize Theorem 2.14 to the case where $X_1, \dots, X_n$ are independent random variables.

Theorem 2.16 (Local law for independent roots)**.** *Fix $C > 0$, and let $X_1, \dots, X_n$ be independent complex-valued random variables which satisfy $\max_{1 \le i \le n} \mathbb{E}|X_i| \le n^C$. In addition, assume X_1 is absolutely continuous (with respect to Lebesgue measure on $\mathbb{C}$) and has density bounded by n^C. Let $\varphi : \mathbb{C} \to \mathbb{R}$ be a twice continuously*

differentiable function (possibly depending on n) which is supported on $B(0, n^C)$ *and which satisfies the pointwise bound given in* (18) *for all* $z \in \mathbb{C}$*. Then, for every fixed* $c > 0$ *and every* $\alpha > 0$*,*

$$\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) = \sum_{i=1}^{n} \varphi(X_i) + O_\alpha(\| \Delta\varphi\|_1 \log n) + O_\alpha(n^{-c})$$

with probability $1 - O_\alpha(n^{-\alpha})$*, where* $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ *are the critical points of the polynomial* $p_n(z) := \prod_{i=1}^n (z - X_i)$ *and* $\| \Delta\varphi\|_1$ *is the* L_1*-norm of* $\Delta\varphi$*. Here, the implicit constants in our asymptotic notation depend on* C, c*, and* α*.*

Theorem 2.16 can be viewed as a local version of Theorem 1.1 and (2). Indeed, since the functions in the theorem above can depend on n, one can approximate an indicator function of Borel sets which changes with n. In addition, the error bound in Theorem 2.16 is significantly better then the error term from (2).

Interestingly, Theorem 2.16 only requires a single root (X_1) to actually be random; the rest may be deterministic. In particular, since the density of X_1 is bounded by n^C, X_1 can itself be quite close to deterministic. Obviously, though, the result fails for deterministic polynomials. For example, consider $q_n(z) := z^n - 1$. The conclusion of Theorem 2.16 fails for this polynomial since all of the critical points are located at the origin while the roots are the n-th roots of unity, located on the unit circle. However, Theorem 2.16 does apply to $p_n(z) := q_n(z)(z - X)$, where X is uniformly distributed on $B(z_0, n^{-C/2})$ for any fixed $z_0 \in \mathbb{C}$. Theorem 2.16 strengthens Theorem 1.6 of [3] for the empirical distribution associated with the zeros of p_n' by providing a rate of convergence. As a consequence of Theorem 2.16, we have the following central limit theorem (CLT).

Theorem 2.17 (Central limit theorem for linear statistics)**.** *Let* $X_1, X_2, \dots$ *be iid random variables which are absolutely continuous (with respect to Lebesgue measure on* $\mathbb{C}$*) and have a bounded density. In addition, assume* $\mathbb{E}|X_1| < \infty$*. Let* $\varphi : \mathbb{C} \to \mathbb{R}$ *be a twice continuously differentiable function with compact support which does not depend on n. Then,*

$$\frac{1}{\sqrt{n}} \left(\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) - \mathbb{E}\left[\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) \right] \right) \longrightarrow N(0, v^2)$$

in distribution as $n \to \infty$*, where* $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ *are the critical points of the polynomial* $p_n(z) := \prod_{i=1}^n (z - X_i)$ *and* v^2 *is the variance of* $\varphi(X_1)$*.*

We now state a version of Theorem 2.16 that applies when the function φ is analytic. While analyticity is a much more rigid assumption, the next result does not contain the extra factor of $\log n$ present in the error term from Theorem 2.16.

Theorem 2.18 (Local law for analytic test functions)**.** *Fix* $C, c, \varepsilon > 0$*. Let* μ *be a probability measure on* $\mathbb{C}$ *supported on* $B(0, C)$*, and assume*

$$|m_\mu(z)| \geq c \tag{19}$$

for all $z \in \Gamma$*, where* Γ *is the boundary of* $B(0, C + \varepsilon)$*. Then for any function* φ *(possibly depending on n), analytic in a neighborhood containing the closure of*

$B(0, C+\varepsilon)$, one has

$$\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) = \sum_{i=1}^{n} \varphi(X_i) + O\left(\oint_\Gamma |\varphi(z)||dz| \right),$$

where $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ are the critical points of the polynomial $p_n(z) := \prod_{i=1}^n (z - X_i)$ and $X_1, \dots, X_n$ are iid random variables with distribution μ. Here, the implicit constants in our asymptotic notation depend on C, c, and ε.

2.5. **Guaranteeing the assumptions in the local law.** In this section, we provide some criteria for assuring the assumptions in Theorem 2.14 are met.

Lemma 2.19 (Simple criterion for an upper bound)**.** *Fix $C, \varepsilon > 0$, and suppose $X_1, \dots, X_n$ are complex-valued random variables (not necessarily independent nor identically distributed). If $\max_{1 \le i \le n} \mathbb{E}|X_i|^\varepsilon \le n^C$, then $\max_{1 \le i \le n} |X_i| \le e^{n^C}$ with overwhelming probability.*

Proof. As $\mathbb{P}\left(\max_{1 \le i \le n} |X_i| > e^{n^C} \right) \le \sum_{i=1}^n \mathbb{P}(|X_i| > e^{n^C})$, the claim follows from a simple application of Markov's inequality. □

Lemma 2.20 (Criterion for anti-concentration)**.** *Fix $C > 0$, and let $X_1, \dots, X_n$ be complex-valued random variables such that X_1 is independent of $X_2, \dots, X_n$. In addition, assume X_1 is absolutely continuous (with respect to Lebesgue measure on $\mathbb{C}$) with density bounded by n^C, and suppose that $\mathbb{E}|X_1| \le n^C$. Then for every $a > 0$, there exists $b > 0$ such that*

$$\left| \sum_{i=1}^{n} \frac{1}{Z - X_i} \right| \ge n^{-b}$$

with probability $1 - O_a(n^{-a})$, where Z is uniformly distributed on $B(0, n^C)$ and independent of $X_1, \dots, X_n$.

We prove Lemma 2.20 in Appendix A.

2.6. **Overview and outline.** The remainder of the paper is devoted to proving our main results. In Section 3, we establish Theorems 2.8, 2.10, and 2.12 of Subsection 2.3 by way of Theorem 3.1 for deterministic polynomials, which we also use to prove a generalization to Theorem 2.10. Section 4 contains the proofs of the local laws from Subsection 2.4 including those for Theorems 2.14, 2.16, 2.17, and 2.18. We conclude the paper with a proof of Theorem 2.3 in Section 5.

There are two appendices that contain minor lemmata and supporting calculations. In Appendix A, we provide Lemma A.1 to simplify the computation of m_μ for radially symmetric distributions, we include calculations related to Example 2.7, and we justify Lemma 2.20. Appendix B contains some classical arguments that establish a Lindeberg CLT that we use to prove part of Theorem 2.8.

3. Proof of results in Section 2.3

The proofs of Theorems 2.8, 2.10, and 2.12 rely on the following theorem for deterministic polynomials.

Theorem 3.1. *Suppose ξ is a complex number, $\vec{X} = (X_1, X_2, \dots, X_n)$ is a vector of complex numbers, and C_1, C_2, k_{Lip} are positive values for which the following three conditions hold:*

(i) $C_1 \le \left|\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi - X_j}\right| \le C_2$;

(ii) *The function* $z \mapsto \frac{1}{n}\sum_{j=1}^{n}\frac{1}{z-X_j}$ *is Lipschitz continuous with constant* k_{Lip} *on the set* $\left\{z \in \mathbb{C} : |z-\xi| \le \frac{2}{C_1 n}\right\}$;

(iii) $\min_{1\le j\le n} |\xi - X_j| > \dfrac{3}{C_1 n}$.

Then, if $C > 0$ *and* $n \in \mathbb{N}$ *satisfy*

$$C > \frac{8(1+2C_2^2)}{C_1^3} \quad \textit{and} \quad n > 4C_2 \max\left\{\frac{1}{C_1}, C(k_{Lip}+1)\right\}, \tag{20}$$

the polynomial $p_n(z) := (z-\xi)\prod_{j=1}^{n}(z-X_j)$ *has exactly one critical point,* $w_\xi^{(n)}$, *that is within a distance of* $\frac{3}{2C_1 n}$ *of* ξ, *and*

$$\left| w_\xi^{(n)} - \xi + \frac{1}{n+1}\frac{1}{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi - X_j}} \right| < \frac{C(k_{Lip}+1)}{n^2}. \tag{21}$$

We remark that criteria (i) and (ii) appear relevant in view of (4) and its accompanying heuristic. Assumption (iii) helps to guarantee that $p_n(z)$ has only one critical point that is within order $O(1/n)$ of ξ, but with respect to establishing equation (21), (iii) is likely an artificial constraint related to the use of Rouché's theorem in the proof. We prove Theorem 3.1 in the next subsection.

3.1. **Proof of Theorem 3.1.** Our strategy is to compare $p_n(z)$ to the simpler polynomial

$$\widetilde{p}(z) = (z-\xi)(z-Y_n)^n,$$

where

$$Y_n := \xi - \frac{1}{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi - X_j}}$$

is chosen so that near ξ, the logarithmic derivatives

$$L_n(z) := \frac{1}{z-\xi} + \sum_{j=1}^{n}\frac{1}{z-X_j} \quad \text{and} \quad \widetilde{L}_n(z) := \frac{1}{z-\xi} + \frac{n}{z-Y_n}$$

of p_n and $\widetilde{p}_n$, respectively, are close to each other. In particular, we will use Rouché's theorem to show that L_n and $\widetilde{L}_n$ both have exactly one zero in each of the nested open balls

$$D_n^{\mathrm{sm}} := B\left(c_n, \frac{C(k_{\mathrm{Lip}}+1)}{n^2}\right) \quad \text{and} \quad D_n^{\mathrm{lg}} := B\left(\xi, \frac{3}{2C_1 n}\right),$$

where

$$c_n := \xi - \frac{1}{n+1}\frac{1}{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi - X_j}}$$

can be easily verified to be a root of $\widetilde{L}_n$. By "clearing the denominators" we will conclude that p_n has exactly one critical point in each of the two balls. The lemma below establishes a few key facts that we frequently reference throughout the proof.

Lemma 3.2. *Under the assumptions of Theorem 3.1:*

(i) For $|z - c_n| \leq \frac{C(k_{Lip}+1)}{n^2}$*:*

$$\frac{C(k_{Lip}+1)}{n^2} < |z-\xi| < \frac{5}{4C_1 n}, \text{ so } D_n^{sm} \subset D_n^{lg};$$
$$\frac{C(k_{Lip}+1)}{n^2} < |z - Y_n| < \frac{2}{C_1};$$
$$\frac{C(k_{Lip}+1)}{n^2} < \frac{1}{C_1 n} < |z - X_j| \text{ for } 1 \leq j \leq n.$$

(ii) For $|z - \xi| \leq \frac{3}{2nC_1}$*:*

$$\frac{3}{2C_1 n} < |z - Y_n| < \frac{5}{2C_1};$$
$$\frac{3}{2C_1 n} < |z - X_j| \text{ for } 1 \leq j \leq n;$$
$$\frac{1}{2C_1 n} < |z - c_n| \text{ if } |z-\xi| = \frac{3}{2nC_1}.$$

Proof. To prove (i), suppose $|z - c_n| \leq \frac{C(k_{\text{Lip}}+1)}{n^2}$. By the triangle inequality, we have

$$|z-\xi| \geq |c_n - \xi| - |z - c_n| \geq \frac{1}{(n+1)C_2} - \frac{C(k_{\text{Lip}}+1)}{n^2} \geq \frac{1}{2nC_2} - \frac{C(k_{\text{Lip}}+1)}{n^2},$$

and by the hypothesis that $n > 4C_2 C(k_{\text{Lip}}+1)$, it follows that

$$|z-\xi| > \frac{1}{2nC_2} - \frac{1}{4nC_2} = \frac{1}{4nC_2} > \frac{C(k_{\text{Lip}}+1)}{n^2}.$$

On the other hand, we have

$$|z-\xi| \leq |z - c_n| + |c_n - \xi| \leq \frac{C(k_{\text{Lip}}+1)}{n^2} + \frac{1}{(n+1)C_1},$$

and the assumption $n > 4C_2 C(k_{\text{Lip}}+1)$ guarantees that

$$|z-\xi| < \frac{1}{4nC_1} + \frac{1}{(n+1)C_1} < \frac{5}{4C_1 n}$$

(note: $C_1 \leq C_2$). This establishes the first inequality. The second follows from nearly identical reasoning; we omit the details. To achieve the inequalities $\frac{1}{C_1 n} < |z - X_j|$, we use $|z-\xi| < \frac{5}{4C_1 n}$, which we just proved, and the assumption that $\min_{1\leq j\leq n} |\xi - X_j| > \frac{3}{C_1 n}$. Indeed, for $1 \leq j \leq n$, the triangle inequality yields

$$|z - X_j| \geq |\xi - X_j| - |z-\xi| > \frac{3}{C_1 n} - \frac{5}{4C_1 n} > \frac{1}{C_2 n} > \frac{C(k_{\text{Lip}}+1)}{n^2}.$$

This completes the proof of part (i). Part (ii) follows from nearly identical reasoning. Note that the assumption $n > 4C_2/C_1$ is useful for achieving the lower bound on $|z - Y_n|$. We omit the remaining details. □

The lower bounds in Lemma 3.2 imply that under the assumptions of Theorem 3.1, $L_n(z)$ and $\widetilde{L}_n(z)$ are holomorphic on the domain D_n^{sm} and that $(z-\xi)L_n(z)$ and $(z-\xi)\widetilde{L}_n(z)$ are holomorphic on the domain D_n^{lg}. We will show that under the same assumptions, $\left|L_n(z) - \widetilde{L}_n(z)\right| < \left|\widetilde{L}_n(z)\right|$ for z in the boundaries of D_n^{sm} and

D_n^{lg} in order to justify Rouché's theorem. To that end, assume the hypotheses of Theorem 3.1 and let $z \in \partial D_n^{\mathrm{sm}} \cup \partial D_n^{\mathrm{lg}}$. Then, the triangle inequality implies

$$\begin{aligned}\left|L_n(z) - \widetilde{L}_n(z)\right| &= \left|\sum_{j=1}^{n} \frac{1}{z - X_j} - \frac{n}{z - Y_n}\right| \\ &\leq \left|\sum_{j=1}^{n} \frac{1}{z - X_j} - \sum_{j=1}^{n} \frac{1}{\xi - X_j}\right| + \left|\sum_{j=1}^{n} \frac{1}{\xi - X_j} - \frac{1}{\frac{1}{n}z - \frac{1}{n}Y_n}\right| \\ &\leq nk_{\mathrm{Lip}}\,|z - \xi| + \left|\sum_{j=1}^{n} \frac{1}{\xi - X_j} - \frac{1}{\frac{1}{n}(z-\xi) + \frac{1}{\sum_{j=1}^{n} \frac{1}{\xi - X_j}}}\right|,\end{aligned}$$

where we have used hypothesis (ii) of Theorem 3.1 to bound the first term on the left. By factoring $\left|\sum_{j=1}^{n} \frac{1}{\xi - X_j}\right|$ from both terms in the right summand, we obtain

$$\left|L_n(z) - \widetilde{L}_n(z)\right| \leq nk_{\mathrm{Lip}}\,|z - \xi| + n \cdot \left|\frac{1}{n}\sum_{j=1}^{n} \frac{1}{\xi - X_j}\right| \cdot \left|1 - \frac{1}{(z-\xi)\frac{1}{n}\sum_{j=1}^{n} \frac{1}{\xi - X_j} + 1}\right|,$$

and then, combining the fractions, factoring out another $\left|\sum_{j=1}^{n} \frac{1}{\xi - X_j}\right|$, and applying hypothesis (i) of Theorem 3.1 twice yields

$$\begin{aligned}\left|L_n(z) - \widetilde{L}_n(z)\right| &\leq nk_{\mathrm{Lip}}\,|z - \xi| + nC_2 \cdot \left|\frac{(z-\xi)\frac{1}{n}\sum_{j=1}^{n} \frac{1}{\xi - X_j}}{(z-\xi)\frac{1}{n}\sum_{j=1}^{n} \frac{1}{\xi - X_j} + 1}\right| \\ &\leq nk_{\mathrm{Lip}}\,|z - \xi| + nC_2^2 \cdot \frac{|z - \xi|}{\left|(z-\xi)\frac{1}{n}\sum_{j=1}^{n} \frac{1}{\xi - X_j} + 1\right|}.\end{aligned}$$

Finally, we can use the reverse triangle inequality and hypothesis (i) of Theorem 3.1 to show

$$\begin{aligned}\left|L_n(z) - \widetilde{L}_n(z)\right| &\leq nk_{\mathrm{Lip}}\,|z - \xi| + nC_2^2\,|z - \xi| \cdot \frac{1}{1 - \left|(z-\xi)\frac{1}{n}\sum_{j=1}^{n} \frac{1}{\xi - X_j}\right|} \\ &\leq n\,|z - \xi| \left(k_{\mathrm{Lip}} + \frac{C_2^2}{1 - |z - \xi|\,C_2}\right).\end{aligned} \tag{22}$$

At this point, we split the argument into two cases: $|z - c_n| = \frac{C(k_{\mathrm{Lip}}+1)}{n^2}$ and $|z - \xi| = \frac{3}{2nC_1}$. In the first case, Lemma 3.2 guarantees that $|z - \xi| < \frac{2}{nC_1}$, and the hypotheses of Theorem 3.1 require that $\frac{1}{2} > \frac{2C_2}{nC_1}$, so we obtain

$$\left|L_n(z) - \widetilde{L}_n(z)\right| < \frac{2}{C_1}\left(k_{\mathrm{Lip}} + 2C_2^2\right) \leq \frac{2}{C_1}(k_{\mathrm{Lip}} + 1)(1 + 2C_2^2). \tag{23}$$

On the other hand,

$$\begin{aligned}\left|\widetilde{L}_n(z)\right| &= \left|\frac{1}{z-\xi}+\frac{n}{z-Y_n}\right| \\ &= \left|\frac{z-Y_n+n(z-\xi)}{(z-\xi)(z-Y_n)}\right| \\ &= (n+1)\cdot|z-\xi|^{-1}\cdot|z-Y_n|^{-1}\cdot|z-c_n| \\ &> n\cdot\frac{nC_1}{2}\cdot\frac{C_1}{2}\cdot\frac{C(k_{\text{Lip}}+1)}{n^2},\end{aligned}$$

where the last inequality follows from Lemma 3.2. One of the assumptions in Theorem 3.1 is that $C > \frac{8(1+2C_2^2)}{C_1^3}$, so

$$\left|\widetilde{L}_n(z)\right| > \frac{C_1^2}{4}(k_{\text{Lip}}+1)\frac{8(1+2C_2^2)}{C_1^3} = \frac{2}{C_1}(k_{\text{Lip}}+1)(1+2C_2^2). \tag{24}$$

Combining (23) and (24) yields $\left|L_n(z)-\widetilde{L}_n(z)\right| < \left|\widetilde{L}_n(z)\right|$ for z in the boundary of D_n^{sm}. In addition, recall (Lemma 3.2 part (ii)) that $L_n(z)$ and $\widetilde{L}(z)$ are holomorphic on the domain D_n^{sm}, so Rouché's theorem guarantees that $L_n(z)$ and $\widetilde{L}_n(z)$ have the same number of zeros inside D_n^{sm}. Since c_n is the unique zero of $\widetilde{L}_n(z)$ in D_n^{sm}, we conclude that $L_n(z)$ has exactly one zero, $w_\xi^{(n)}$, in D_n^{sm}. Furthermore, $L_n(z) = \frac{p_n'(z)}{p_n(z)}$ (which is analytic for $z \in D_n^{\text{sm}}$ by (i) of Lemma 3.2), so the zeros of $L_n(z)$ in D_n^{sm} are the same as the critical points of $p_n(z)$ in D_n^{sm}. We conclude that $p_n(z)$ has exactly one critical point in D_n^{sm}.

Lemma 3.2 shows that $D_n^{\text{sm}} \subset D_n^{\text{lg}}$, so it remains to establish that $p_n(z)$ also has exactly one critical point in D_n^{lg}, for then, the critical point in both domains must be the same one. Continuing from (22), in the case where $|z-\xi| = \frac{3}{2C_1 n}$, we obtain

$$\left|L_n(z)-\widetilde{L}_n(z)\right| < \frac{3}{2C_1}(k_{\text{Lip}}+2C_2^2) \le \frac{3}{2C_1}(k_{\text{Lip}}+1)(1+2C_2^2), \tag{25}$$

where we have once again used the assumption that $\frac{1}{2} \ge \frac{2C_2}{nC_1}$. Similarly to above, we also have

$$\begin{aligned}\left|\widetilde{L}_n(z)\right| &= \left|\frac{1}{z-\xi}+\frac{n}{z-Y_n}\right| \\ &= \left|\frac{z-Y_n+n(z-\xi)}{(z-\xi)(z-Y_n)}\right| \\ &= (n+1)\cdot|z-\xi|^{-1}\cdot|z-Y_n|^{-1}\cdot|z-c_n| \\ &> n\cdot\frac{2C_1 n}{3}\cdot\frac{2C_1}{5}\cdot\frac{1}{2C_1 n} \\ &= \frac{2C_1 n}{15}\end{aligned}$$

where the inequality follows from Lemma 3.2, (ii). From the assumptions on n and C in Theorem 3.1, it follows that

$$n > 4C_2C(k_{\text{Lip}}+1) > \frac{32(1+2C_2^2)(k_{\text{Lip}}+1)}{C_1^2}\cdot\frac{C_2}{C_1} \ge \frac{32(1+2C_2^2)(k_{\text{Lip}}+1)}{C_1^2}$$

(recall $C_1 \leq C_2$), so in the case when $|z-\xi| = \frac{3}{2C_1 n}$,

$$\left|\widetilde{L}_n(z)\right| > \frac{64}{15C_1}(k_{\text{Lip}}+1)(1+2C_2^2). \tag{26}$$

Combining (25) and (26) yields $\left|L_n(z) - \widetilde{L}_n(z)\right| < \left|\widetilde{L}_n(z)\right|$ for z in the boundary of D_n^{lg}. Consequently, for $z \in \partial D_n^{\text{lg}}$,

$$\left|(z-\xi)L_n(z) - (z-\xi)\widetilde{L}_n(z)\right| < \left|(z-\xi)\widetilde{L}_n(z)\right|,$$

and since $(z-\xi)L_n(z)$, $(z-\xi)\widetilde{L}_n(z)$ are holomorphic in D_n^{lg} by Lemma 3.2, (ii), Rouché's theorem guarantees that $(z-\xi)L_n(z)$, $(z-\xi)\widetilde{L}_n(z)$ have the same number of zeros in D_n^{lg}. In fact, $(z-\xi)\widetilde{L}_n(z)$ has exactly one zero in D_n^{lg}, namely c_n, so

$$(z-\xi)L_n(z) = \frac{p_n'(z)}{\prod_{j=1}^n (z-X_j)}$$

has exactly one zero in D_n^{lg}, too. (Note: by Lemma 3.2, (i), $D_n^{\text{sm}} \subset D_n^{\text{lg}}$.) Hence, $p_n'(z)$ has exactly one root in D_n^{lg}, and as we showed above, this root lies in D_n^{sm}. The proof of Theorem 3.1 is complete.

In the remainder of this section, we use Theorem 3.1 to prove Theorems 2.8, 2.10 and 2.12. We also include a subsection where we sketch how the arguments could be modified to prove Theorem 3.4, which generalizes part of Theorem 2.10 to situations where p_n has many deterministic roots. When $\xi \in \text{supp}(\mu)$, it is difficult to control $\frac{1}{n}\sum_{j=1}^n \frac{1}{\xi - X_j}$, so we start with the proof of Theorem 2.10, which is more straightforward than the justification of Theorem 2.8.

3.2. **Proof of Theorem 2.10.** We begin by establishing equation (12) via Theorem 3.1. To that end, we consider $\{\xi_l\}_{l=1}^s$, one at a time, letting each in turn play the role of ξ in the statement of Theorem 3.1. Fix ξ_l, $1 \leq l \leq s$. We will show that for large n, on the complement of the "bad" event

$$E_n^l := \left\{ \left| \frac{1}{n+1-s} \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right| \geq \frac{|m_\mu(\xi_l)|}{3} \right\},$$

the hypotheses of Theorem 3.1 are satisfied with $\xi = \xi_l$,

$$\vec{X} = (\xi_1, \ldots, \xi_{l-1}, \xi_{l+1}, \ldots, \xi_s, X_1, \ldots, X_{n+1-s}),$$

and the positive constants

$$C_{1,l} := \frac{|m_\mu(\xi_l)|}{2},\ C_{2,l} := \frac{3\,|m_\mu(\xi_l)|}{2},\ k_{\text{Lip},l} := \frac{9}{\text{dist}(\xi_l, \text{supp}(\mu) \cup \{\xi_j : j \neq l\})^2}. \tag{27}$$

(Here, $\text{dist}(z, D) := \inf_{w \in D} |z-w|$ is the distance from $z \in \mathbb{C}$ to a set $D \subset \mathbb{C}$.)

For large n, on the complement of E_n^l,

$$\begin{aligned}
&\left| \frac{1}{n} \left(\sum_{\substack{j=1 \\ j \neq l}}^{s} \frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j} \right) \right| \\
&\qquad \leq \left| \frac{1}{n} \sum_{\substack{j=1 \\ j \neq l}}^{s} \frac{1}{\xi_l - \xi_j} \right| + \frac{n+1-s}{n} \left| \frac{1}{n+1-s} \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j} \right| \qquad (28) \\
&\qquad \leq o_l(1) + \left| \frac{1}{n+1-s} \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right| + |m_\mu(\xi_l)| \leq C_{2,l}
\end{aligned}$$

(The last inequality holds for large n.) Similarly, for large n, on the event $\left(E_n^l\right)^c$,

$$\left| \frac{1}{n} \left(\sum_{\substack{j=1 \\ j \neq l}}^{s} \frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j} \right) \right| \geq C_{1,l}, \qquad (29)$$

and condition (i) of Theorem 3.1 follows from equations (28) and (29). If n is chosen large enough that

$$\varepsilon_l := \operatorname{dist}(\xi_l, \operatorname{supp}(\mu) \cup \{\xi_j : j \neq l\}) > \frac{3}{C_{1,l} n},$$

then condition (iii) of Theorem 3.1 holds, and for $|z - \xi_l| \leq \frac{2}{C_{1,l} n}$,

$$\begin{aligned}
\min_{1 \leq j \leq n+1-s} |z - X_j| &\geq \min_{1 \leq j \leq n+1-s} |\xi_l - X_j| - |z - \xi_l| \geq \varepsilon_l - \frac{2}{C_{1,l} n} > \frac{\varepsilon_l}{3}, \\
\min_{j \neq l} |z - \xi_j| &\geq \min_{j \neq l} |\xi_l - \xi_j| - |z - \xi_l| \geq \varepsilon_l - \frac{2}{C_{1,l} n} > \frac{\varepsilon_l}{3}.
\end{aligned}$$

In particular, this shows that for positive integers $n > 3(C_1 \varepsilon_l)^{-1}$ and complex numbers $z, w \in \left\{ z : |z - \xi| \leq \frac{2}{C_{1,l} n} \right\}$,

$$\begin{aligned}
&\left| \frac{1}{n} \left(\sum_{\substack{j=1 \\ j \neq l}}^{s} \frac{1}{z - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{z - X_j} \right) - \frac{1}{n} \left(\sum_{\substack{j=1 \\ j \neq l}}^{s} \frac{1}{w - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{w - X_j} \right) \right| \\
&\qquad = \frac{1}{n} \left| \sum_{\substack{j=1 \\ j \neq l}}^{s} \frac{w - z}{(z - \xi_j)(w - \xi_j)} + \sum_{j=1}^{n+1-s} \frac{w - z}{(z - X_j)(w - X_j)} \right| \\
&\qquad \leq |w - z| \cdot \frac{9}{\varepsilon_l^2} = k_{\mathrm{Lip},l} \cdot |w - z| ,
\end{aligned}$$

which implies condition (ii) of Theorem 3.1.

Now, fix any $C > \max_{1\le l\le s} \frac{8(1+2C_{2,l}^2)}{C_{1,l}^3}$. If n is a natural number large enough to guarantee inequalities (28) and (29) for $1 \le l \le s$ and that satisfies

$$n > \max\left\{ \frac{4C_{2,l}}{C_{1,l}}, 4C_{2,l}C(k_{\mathrm{Lip},l}+1), \frac{3}{C_{1,l}\varepsilon_l} : 1 \le l \le s \right\}, \tag{30}$$

Theorem 3.1 guarantees that on the complement of $\cup_{l=1}^s E_n^l$, the polynomial p_n has s critical points, $w_1^{(n)}, \ldots, w_s^{(n)}$, such that for $1 \le l \le s$, $w_l^{(n)}$ is the unique critical point of p_n that is within a distance of $\frac{3}{|m_\mu(\xi_l)|n}$ of ξ_l, and

$$\left| w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{n}{\sum_{j\ne l} \frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j}} \right| < \frac{C(k_{\mathrm{Lip},l}+1)}{n^2}. \tag{31}$$

(Note that for large n, $w_1^{(n)}, \ldots, w_s^{(n)}$ are distinct because $\xi_1, \ldots, \xi_s$ are distinct and (31) implies $w_l^{(n)} \to \xi_l$ for $1 \le l \le s$.) We complete our justification of (12) from Theorem 2.10 by choosing $C_{\mu,\vec{\xi}}$ larger than $\max_l C(k_{\mathrm{Lip},l}+1)$ and applying Hoeffding's inequality to the bounded random variables $(\xi_l - X_j)^{-1}$ to achieve the desired control over $\mathbb{P}(\cup_l E_n^l)$. More specifically, since $\xi_l \notin \mathrm{supp}(\mu)$ for $1 \le l \le s$, the random variables $Y_j^l := (\xi_l - X_j)^{-1}$ are almost surely uniformly bounded by $K_l := \mathrm{dist}(\xi_l, \mathrm{supp}(\mu))^{-1}$, and the following version of Hoeffding's inequality applies with $t_l := \frac{|m_\mu(\xi_l)|}{3}$.

Lemma 3.3 (Hoeffding's inequality for complex-valued random variable; Lemma 3.1 from [23])**.** *Let $Y_1, \ldots, Y_n$ be iid complex-valued random variables which satisfy $|Y_j| \le K$ almost surely for some $K > 0$. Then there exist absolute constants $C, c > 0$ such that*

$$\mathbb{P}\left(\left| \frac{1}{n}\sum_{j=1}^n Y_j - \frac{1}{n}\mathbb{E}\left[\sum_{j=1}^n Y_j\right] \right| \ge t \right) \le C\exp\left(-cnt^2/K^2\right),$$

for every $t > 0$.

By Lemma 3.3, we can find $C, c_{\mu,\vec{\xi}} > 0$ such that $\cup_l E_n^l$ occurs with probability at least $1 - C\exp(-c_{\mu,\vec{\xi}} n)$ as is desired.

We have established with overwhelming probability the existence of the critical points $w_1^{(n)}, \ldots, w_s^{(n)}$ characterized by (12). It remains to show that they satisfy the convergence in (13). To that end, apply the Borel–Cantelli Lemma to the events $\cup_l E_n^l$ to see that almost surely, for large enough n, $w_l^{(n)}$ satisfies (12) for $1 \le l \le n$.

It follows that with probability 1, for sufficiently large n and any l, $1 \leq l \leq s$,

$$\begin{aligned}
&\sqrt{n}(n+1) \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right) \\
&\quad = m_\mu(\xi_l)^2 \cdot \sqrt{n} \left(\frac{1}{m_\mu(\xi_l)} - \frac{n}{\sum_{j \neq l} \frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j}} \right) + O_{\mu,\vec{\xi}}(n^{-1/2}) \\
&\quad = m_\mu(\xi_l)^2 \cdot \sqrt{n} \left(\frac{1}{m_\mu(\xi_l)} - \frac{1}{\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} + O_{\mu,\vec{\xi}}(1/n)} \right) + O_{\mu,\vec{\xi}}(n^{-1/2}) \\
&\quad = \frac{m_\mu(\xi_l)}{\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} + O_{\mu,\vec{\xi}}(1/n)} \cdot \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right) + O_{\mu,\vec{\xi}}(n^{-1/2}).
\end{aligned} \tag{32}$$

In the case $s > 1$, we have used that

$$\max_{1 \leq l \leq s} \left| \sum_{j \neq l} \frac{1}{\xi_l - \xi_k} - \sum_{j=n+2-s}^{n} \frac{1}{\xi_l - X_j} \right| = O_{\mu,\vec{\xi}}(1).$$

Now, we will use the Cramér–Wold device (see e.g. Theorem 29.4 in [2]) to show the convergence (13). To start, let $t_1, \ldots, t_s, r_1, \ldots, r_s$ be arbitrary real numbers and define the random variables

$$\begin{aligned}
Y_{n,l} &:= n^{3/2} \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right), \\
Z_{l,j} &:= \operatorname{Re} \left(\frac{1}{\xi_l - X_j} \right), \\
W_{l,j} &:= \operatorname{Im} \left(\frac{1}{\xi_l - X_j} \right),
\end{aligned}$$

for $1 \leq l \leq s$. By (32), we have, with probability tending to 1,

$$\begin{aligned}
Y_{n,l} &= \frac{n}{n+1} \frac{m_\mu(\xi_l)}{\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} + o(1)} \cdot \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right) + O\left(\frac{1}{\sqrt{n}} \right) \\
&= (1 + o(1)) \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right) + O\left(\frac{1}{\sqrt{n}} \right) \\
&= \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right) + o(1),
\end{aligned} \tag{33}$$

where all of the implied constants depend on $\xi_1, \ldots, \xi_s$ and μ, and we have made ample use of Slutsky's theorem (see e.g. Theorem 11.4 from [11]). To obtain the last line, we also used the classical CLT (see e.g. Theorem 29.5 from [2]) in conjunction with Slutsky's theorem. If we take linear combinations of the real and imaginary

parts of $Y_{n,l}$, we obtain that with probability at least $1 - o(1)$,

$$\sum_{l=1}^{s} t_l \operatorname{Re}(Y_{n,l}) + \sum_{l=1}^{s} r_l \operatorname{Im}(Y_{n,l})$$
$$= \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\sum_{l=1}^{s}\left[t_l Z_{l,j} + r_l W_{l,j} - t_l \operatorname{Re}\left(m_\mu(\xi_l)\right) - t_l \operatorname{Im}\left(m_\mu(\xi_l)\right)\right]\right) + o(1),$$

which converges by the classical CLT (and Slutsky's theorem) in distribution to a normally distributed random variable with mean 0 and variance $\operatorname{Var}\left(\sum_{l=1}^{s}\left[t_l Z_{l,1} + r_l W_{l,1}\right]\right)$. This limiting distribution is the same as the distribution of the random variable $\sum_{l=1}^{s}\left[t_l \operatorname{Re}(N_l) + r_l \operatorname{Im}(N_l)\right]$ with covariance structure given by (14), so by the Cramér–Wold strategy, the proof of Theorem 2.10 is complete.

The next subsection illustrates how to modify the argument above to prove a generalization of Theorem 3.4 to the case where p_n has a number of deterministic roots that may grow with n.

3.3. **Generalization of Theorem 2.10.** The following result shows how Theorem 3.1 could be used to locate the critical points near a number of outlying deterministic roots that is allowed to depend on n. Compare the following theorem to Theorem 2 in [14]. Both theorems discuss the pairing between s_n roots and critical points of p_n, where $s_n = o(n)$ is allowed to depend on n. Theorem 3.4 describes the locations of the critical points with higher precision than Theorem 2 of [14], however our theorem requires that the deterministic roots $\xi_1, \ldots, \xi_{s_n}$ be outside the support of μ, while Theorem 2 in [14] doesn't make this restriction.

Theorem 3.4 (Locations of critical points when p_n has many deterministic roots.)**.** *Suppose $X_1, X_2, \ldots$ are iid complex-valued random variables with distribution μ, let $\xi_1, \xi_2, \ldots$ be fixed deterministic values, let s_n, l_n, a_n be positive integers less than n, and fix $\varepsilon, L > 0$, so that all of these together satisfy:*

(i) $1 \le s_n \le l_n = o(n)$, $a_n l_n = o(n)$, $a_n = o(\sqrt{n})$;
(ii) $\min\left\{|m_\mu(\xi_l)| : 1 \le l \le s_n\right\} \ge \varepsilon$ *and* $\max\left\{|m_\mu(\xi_l)| : 1 \le l \le s_n\right\} \le L$;
(iii) $\min\left\{|\xi_l - x| : 1 \le l \le s_n,\ x \in \operatorname{supp}(\mu) \cup \{\xi_j\}_{j=1, j\ne l}^{l_n}\right\} > \frac{6}{\varepsilon \cdot a_n}$.

Then, there exist constants $C, c_{\mu,\varepsilon,L}, C_{\mu,\varepsilon,L} > 0$ so that with probability at least $1 - C \cdot s_n \exp(-c_{\mu,\varepsilon,L} \cdot n/a_n^2)$, the polynomial

$$p_n(z) = \prod_{l=1}^{l_n}(z - \xi_l) \prod_{j=1}^{n+1-l_n}(z - X_j)$$

has s_n critical points, $w_1^{(n)}, \ldots, w_{s_n}^{(n)}$, such that for $1 \le l \le s_n$, $w_l^{(n)}$ is the unique critical point of p_n within $\frac{3}{2\varepsilon n}$ of ξ_l and

$$\max_{1 \le l \le s_n}\left|w_l^{(n)} - \xi_l + \frac{1}{n+1}\frac{n}{\sum_{k=1,k\ne l}^{l_n}\frac{1}{\xi_l - \xi_k} + \sum_{j=1}^{n+1-l_n}\frac{1}{\xi_l - X_j}}\right| < \frac{C_{\mu,\varepsilon,L} \cdot a_n^2}{n^2}. \quad (34)$$

Theorem 3.4 follows from an argument quite similar to the one provided in the previous subsection. We outline the main differences in the following proof sketch.

Argue as in Subsection 3.2 for each l, $1 \leq l \leq s_n$, separately but in place of the definitions in equation (27) choose

$$C_1 := \frac{\varepsilon}{2},\ C_2 := \frac{3L}{2},\ \text{and } k_{\text{Lip}} := \frac{\varepsilon^2 a_n^2}{4}.$$

Also, modify the events E_n^l into the events

$$E_n^l := \left\{ \left| \frac{1}{n+1-l_n} \sum_{j=1}^{n+1-l_n} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right| \geq \frac{|m_\mu(\xi_l)|}{3} \right\},\ 1 \leq l \leq s_n.$$

Notice that condition (i) from Theorem 3.1 now holds for n sufficiently large (depending on the rate of convergence of $a_n l_n/n \to 0$) on the complement of E_n^l because

$$\left| \frac{1}{n} \sum_{k=1, k\neq l}^{l_n} \frac{1}{\xi_l - \xi_k} \right| \leq \frac{\varepsilon a_n l_n}{6n} = o(1),$$

and this limit is uniform with respect to $1 \leq l \leq s_n$. The requirements (30) on n now become

$$n > \max\left\{ \frac{4C_2}{C_1}, 4C_2 C \left(\frac{\varepsilon^2 a_n^2}{4} + 1 \right), a_n \right\},$$

which hold uniformly for $1 \leq l \leq s_n$ by assumption (i) in the statement of Theorem 3.4. By Hoeffding's inequality (Lemma 3.3), with $Y_j^l := \frac{1}{\xi_l - X_j}$, $K_l := \frac{\varepsilon a_n}{6}$, and $t_l := \frac{|m_\mu(\xi_l)|}{3} \geq \frac{\varepsilon}{3}$, there are constants $C, c_{\mu,\varepsilon} > 0$, independent of l, ξ_l, and s_n, so that for large n

$$\mathbb{P}\left(E_n^l\right) \leq C \exp\left(-c_{\mu,\varepsilon}(n+1-l_n)/a_n^2\right).$$

Taking a union over l, $1 \leq l \leq s_n$ establishes the desired result.

3.4. **Proof of Theorem 2.8.** We now proceed to prove Theorem 2.8. In order to control the behavior of $\frac{1}{n}\sum_{j=1}^n \frac{1}{\xi - X_j}$, we will rely on the Law of Large Numbers. Lemma 3.5 below justifies this approach by establishing some regularity properties for $\mathbb{E}(\xi - X_1)^{-1} = m_\mu(\xi)$ that we will continue to use throughout the remainder of the paper. We note that Lemma 3.5 is similar to Lemma 5.7 in [18].

Lemma 3.5 (Regularity properties of the Cauchy–Stieltjes transform)**.** *Suppose that on $B(\xi, \rho) \subset \mathbb{C}$, μ has a density with respect to the Lebesgue measure that is bounded by $C_{\mu,\xi,\rho}$. Then,*

(i) for any $z \in B(\xi, \rho/2)$,

$$|m_\mu(z)| \leq \int_{\mathbb{C}} \frac{1}{|z-w|}\, d\mu(w) \leq 2\pi C_{\mu,\xi,\rho} \min\{\rho/2, 1\} + \max\{2/\rho, 1\};$$

(ii) if $\rho = \infty$ so that μ has a density bounded by C_μ on all of $\mathbb{C}$, then there exist constants $\kappa_\mu, \varepsilon_\mu > 0$, depending on μ, so that the following holds. If $x, y \in \mathbb{C}$ with $|x - y| < \varepsilon_\mu$, then

$$|m_\mu(x) - m_\mu(y)| \leq \kappa_\mu |x-y| \log\left(|x-y|^{-1}\right).$$

Proof. To prove the first inequality, observe that for any $z \in B(\xi, \rho/2)$,

$$\begin{aligned}|m_\mu(z)| &\le \int_{\mathbb{C}} \frac{1}{|z-w|} \mathbb{1}_{|z-w|<\min\{\rho/2,1\}}\, d\mu(w) + \int_{\mathbb{C}} \frac{1}{|z-w|} \mathbb{1}_{|z-w|\ge\min\{\rho/2,1\}}\, d\mu(w) \\ &\le 2\pi C_{\mu,\xi,\rho} \int_0^{\min\{\rho/2,1\}} \frac{1}{r} \cdot r\, dr + \max\{2/\rho, 1\} \\ &\le 2\pi C_{\mu,\xi,\rho} \min\{\rho/2, 1\} + \max\{2/\rho, 1\},\end{aligned}$$

where we have used polar coordinates in the integral. To prove (ii), let $Z \sim \mu$ and fix $x, y \in \mathbb{C}$ with $|x-y| \le 1$. We will compute the difference

$$|m_\mu(x) - m_\mu(y)| = \left| \mathbb{E}\left[\frac{1}{x-Z}\right] - \mathbb{E}\left[\frac{1}{y-Z}\right] \right| \le |x-y|\, \mathbb{E}\left[\frac{1}{|x-Z|\,|y-Z|}\right] \quad (35)$$

by splitting the expectations over each of the events

$$\begin{aligned}\mathcal{A} &:= \{|x-Z| \ge |x-y| \text{ and } |y-Z| \ge |x-y|\}, \\ \mathcal{B} &:= \{|x-Z| \ge |x-y| \text{ and } |y-Z| < |x-y|\}, \\ \mathcal{C} &:= \{|x-Z| < |x-y| \text{ and } |y-Z| \ge |x-y|\}, \\ \mathcal{D} &:= \{|x-Z| < |x-y| \text{ and } |y-Z| < |x-y|\},\end{aligned}$$

whose union occurs almost surely. The Cauchy–Schwarz inequality implies

$$\begin{aligned}\mathbb{E}\left[\frac{\mathbb{1}_{\mathcal{A}}}{|x-Z|\,|y-Z|}\right] &\le \sqrt{\mathbb{E}\left[\frac{\mathbb{1}_{|x-Z|\ge|x-y|}}{|x-Z|^2}\right] \mathbb{E}\left[\frac{\mathbb{1}_{|y-Z|\ge|x-y|}}{|y-Z|^2}\right]} \\ &\le 2\pi C_\mu \int_{|x-y|}^1 \frac{1}{r^2} r\, dr + \mathbb{E}[1] = 2\pi C_\mu \log\left|\frac{1}{x-y}\right| + 1.\end{aligned}$$

If $\mathcal{B}$ occurs, then $|x-Z| \ge |x-y|$, so the expectation on the right of (35) is bounded by

$$\mathbb{E}\left[\frac{\mathbb{1}_{\mathcal{B}}}{|x-Z|\,|y-Z|}\right] \le \frac{1}{|x-y|}\mathbb{E}\left[\frac{\mathbb{1}_{|y-Z|<|x-y|}}{|y-Z|}\right] \le \frac{4\pi C_\mu}{|x-y|}\int_0^{|x-y|} \frac{1}{r} r\, dr = 4\pi C_\mu.$$

We can bound $\mathbb{E}[\mathbb{1}_{\mathcal{C}}(|x-Z|\,|x-Y|)^{-1}]$ in similar fashion. For the expectations over the event $\mathcal{D}$, we have the following bound on the middle expression of (35):

$$\mathbb{E}\left[\frac{\mathbb{1}_{\mathcal{D}}}{|x-Z|}\right] + \mathbb{E}\left[\frac{\mathbb{1}_{\mathcal{D}}}{|y-Z|}\right] \le 8\pi C_\mu \int_0^{|x-y|} \frac{1}{r} r\, dr \le 8\pi C_\mu\, |x-y|.$$

In view of (35), these last few inequalities establish (ii). □

We proceed to prove Theorem 2.8, starting with a justification of (9) in the case $s = 1$ and $\xi_1 = \xi$. Choose $\rho_\xi > 0$ so that in the disk $B(\xi, 3\rho_\xi)$, μ has a density f that is bounded by C_f. Our plan of attack will be to show that the hypotheses of Theorem 3.1 are satisfied on the complement of a "bad" event whose probability tends to 0 as n grows. To optimize our control over this event, we allow it to depend on the parameter $\varepsilon_n = o(1)$ that we will choose appropriately to achieve the asymptotic bound in (9).

To that end, suppose $\varepsilon_n \in (0,1)$, let $d_n := \lceil \log(\sqrt{n}) \rceil$, and for each $n \geq 1$ define the annuli

$$A_n^0 := \left\{ z \in \mathbb{C} : |z - \xi| < \frac{\rho_\xi}{\sqrt{n}} \right\},$$
$$A_n^k := \left\{ z \in \mathbb{C} : \frac{\rho_\xi e^{k-1}}{\sqrt{n}} \leq |z - \xi| < \frac{\rho_\xi e^k}{\sqrt{n}} \right\},\ 1 \leq k \leq d_n,$$

and the binomial random variables

$$N_n^k := \#\left\{ 1 \leq j \leq n : X_j \in A_n^k \right\},\ 0 \leq k \leq d_n.$$

Consider the "bad" events

$$E_n = \left\{ \left| \frac{1}{n} \sum_{j=1}^n \frac{1}{\xi - X_j} - m_\mu(\xi) \right| \geq \frac{|m_\mu(\xi)|}{2} \right\},$$
$$F_n^k = \left\{ N_n^k \geq \pi C_f \rho_\xi^2 e^{2k} + \frac{e^k}{\sqrt{\varepsilon_n}} \right\},\ 0 \leq k \leq d_n,$$
$$G_n = \left\{ \min_{1 \leq j \leq n} |X_j - \xi| < \sqrt{\frac{\varepsilon_n}{n}} \right\}.$$

We will demonstrate that if

$$C_1 := \frac{|m_\mu(\xi)|}{2}, \quad C_2 := \frac{3\,|m_\mu(\xi)|}{2}, \quad \text{and} \quad k_{\text{Lip}} := \frac{C_{\mu,\xi} \log n}{\varepsilon_n^{3/2}}, \tag{36}$$

for $\varepsilon_n := (\log n)^{-2/3}$ and $C_{\mu,\xi}$ defined in Lemma 3.6 below, then the conditions in Theorem 3.1 hold on the complement of $E_n \cup G_n \cup \bigcup_k F_n^k$ for large enough n. Furthermore, we will show that the union of these events occurs with probability tending to 0. Notice that events E_n, F_n^k, and G_n are related to conditions (i), (ii), and (iii) of Theorem 3.1, respectively.

It is clear that condition (i) holds on the complement of E_n because $m_\mu(\xi) \neq 0$. For $n > \frac{9}{C_1^2 \varepsilon_n}$, (iii) is true, on the complement of G_n, because in this case, $\sqrt{\frac{\varepsilon_n}{n}} > \frac{3}{C_1 n}$. The following lemma establishes condition (ii).

Lemma 3.6. *There exists a constant $C_{\mu,\xi} > 0$, depending only on μ and ξ, so that if $\varepsilon_n \in (0,1)$, and*

$$n > \max\left\{ \left(\frac{8\rho_\xi}{C_1 \varepsilon_n} \right)^2, \left(\frac{8e^2}{C_1 \rho_\xi} \right)^2, \frac{8}{C_1 \rho_\xi} \right\},$$

then, on the complement of $\bigcup_{k=0}^{d_n} F_n^k \cup G_n$, any complex numbers $z, w \in \overline{B}\left(\xi, \frac{2}{C_1 n}\right)$ satisfy

$$\left| \sum_{j=1}^n \frac{1}{(z - X_j)(w - X_j)} \right| \leq C_{\mu,\xi} \cdot \frac{n \log n}{\varepsilon_n^{3/2}}.$$

Proof. Fix $z, w \in \overline{B}\left(\xi, \frac{2}{C_1 n}\right)$ and $1 \le j \le n$. By applying the triangle inequality several times, we obtain

$$\begin{aligned}
|z - X_j|\,|w - X_j| &\ge \Big|\big(\,|\xi - X_j| - |z - \xi|\,\big)\big(\,|\xi - X_j| - |w - \xi|\,\big)\Big| \\
&\ge |\xi - X_j|^2 - |\xi - X_j|\,(|z - \xi| + |w - \xi|) \\
&\ge |\xi - X_j|^2 - |\xi - X_j|\,\frac{4}{C_1 n}.
\end{aligned}$$

Consequently, on the complement of $\bigcup_{k=0}^{d_n} F_n^k \cup G_n$,

$$\begin{aligned}
\left|\sum_{j=1}^{n} \frac{1}{(z - X_j)(w - X_j)}\right| &\le \sum_{j=1}^{n} \frac{1}{|z - X_j| \cdot |w - X_j|} \\
&\le \sum_{j=1}^{n} \frac{1}{|\xi - X_j|^2 - |\xi - X_j|\,\frac{4}{C_1 n}} \\
&\le \sum_{\substack{1 \le j \le n \text{ s.t.} \\ \frac{\sqrt{\varepsilon_n}}{\sqrt{n}} \le |X_j - \xi| < \frac{\rho_\xi}{\sqrt{n}}}} \frac{1}{\frac{\varepsilon_n}{n} - \frac{\rho_\xi}{\sqrt{n}}\frac{4}{C_1 n}} \\
&\quad + \sum_{k=1}^{d_n} \sum_{\substack{1 \le j \le n \text{ s.t.} \\ X_j \in A_n^k}} \frac{1}{\frac{\rho_\xi^2 e^{2k-2}}{n} - \frac{\rho_\xi e^k}{\sqrt{n}}\frac{4}{C_1 n}} \\
&\quad + \sum_{\substack{1 \le j \le n \text{ s.t.} \\ |X_j - \xi| \ge \rho_\xi}} \frac{1}{|\xi - X_j|^2 - |\xi - X_j|\,\frac{4}{C_1 n}}.
\end{aligned}$$

We have split the sum over $1 \le j \le n$ into $d_n + 2$ pieces. Notice that for $n > \left(\frac{8\rho_\xi}{C_1 \varepsilon_n}\right)^2$,

$$0 < \frac{1}{\frac{\varepsilon_n}{n} - \frac{\rho_\xi}{\sqrt{n}}\frac{4}{C_1 n}} \le \frac{2n}{\varepsilon_n}$$

and for $n > \left(\frac{8e^2}{C_1 \rho_\xi}\right)^2$,

$$0 < \frac{1}{\frac{\rho_\xi^2 e^{2k-2}}{n} - \frac{\rho_\xi e^k}{\sqrt{n}}\frac{4}{C_1 n}} \le \frac{2n}{\rho_\xi^2 e^{2k-2}}, \text{ for } 1 \le k \le d_n.$$

Additionally, if $n > \frac{8}{C_1 \rho_\xi}$ and $|X_j - \xi| \ge \rho_\xi$, then,

$$0 < \frac{1}{|\xi - X_j|^2 - |\xi - X_j|\,\frac{4}{C_1 n}} \le \frac{1}{|\xi - X_j|\left(\rho_\xi - \frac{4}{C_1 n}\right)} \le \frac{2}{\rho_\xi^2}.$$

It follows that if

$$n > \max\left\{\left(\frac{8\rho_\xi}{C_1 \varepsilon_n}\right)^2, \left(\frac{8e^2}{C_1 \rho_\xi}\right)^2, \frac{8}{C_1 \rho_\xi}\right\},$$

on the complement of $\bigcup_{k=0}^{d_n} F_n^k \cup G_n$, for all $z, w \in \overline{B}\left(\xi, \frac{2}{C_1 n}\right)$,

$$\begin{aligned}
&\left|\sum_{j=1}^{n} \frac{1}{(z-X_j)(w-X_j)}\right| \\
&\qquad \leq N_n^0 \cdot \frac{2n}{\varepsilon_n} + \sum_{k=1}^{d_n} N_n^k \cdot \frac{2n}{\rho_\xi^2 e^{2k-2}} + \frac{2n}{\rho_\xi^2} \\
&\qquad \leq \frac{(\pi C_f \rho_\xi^2 + 1)}{\sqrt{\varepsilon_n}} \cdot \frac{2n}{\varepsilon_n} + \sum_{k=1}^{d_n} \left(\pi C_f \rho_\xi^2 e^{2k} + \frac{e^k}{\sqrt{\varepsilon_n}}\right) \frac{2n}{\rho_\xi^2 e^{2k-2}} + \frac{2n}{\rho_\xi^2} \\
&\qquad = O_{\mu,\xi}\left(\frac{n}{\varepsilon_n^{3/2}}\right) + \sum_{k=1}^{d_n} O_{\mu,\xi}\left(\frac{n}{\sqrt{\varepsilon_n}}\right) = O_{\mu,\xi}\left(\frac{n \log n}{\varepsilon_n^{3/2}}\right),
\end{aligned}$$

which completes the proof. □

It remains to find an upper bound on the probability of $E_n \cup \bigcup_{k=1}^{d} F_n^k \cup G_n$, which we accomplish in the next lemma.

Lemma 3.7.

$$\mathbb{P}\left(E_n \cup \bigcup_{k=0}^{d_n} F_n^k \cup G_n\right) = o_{\mu,\xi}(1) + O_{\mu,\xi}\left(\log n \cdot \varepsilon_n^2 + \varepsilon_n\right) = o_{\mu,\xi}(1)$$

Proof. To control $\mathbb{P}(E_n)$, apply the Weak Law of Large Numbers to the random variables $\frac{1}{\xi - X_j}$, which have finite expectation by Lemma 3.5. Next, consider that for large n,

$$\mathbb{P}(G_n) \leq n \cdot \mathbb{P}\left(|X_1 - \xi| \leq \sqrt{\frac{\varepsilon_n}{n}}\right) \leq n \cdot \pi C_f \cdot \frac{\varepsilon_n}{n} = \pi C_f \varepsilon_n,$$

which establishes $\mathbb{P}(G_n) = O_{\mu,\xi}(\varepsilon_n)$.

We now turn our attention to the events F_n^k. For $0 \leq k \leq d_n$ and $1 \leq j \leq n$, define the random variables

$$\chi_{j,k} := \mathbb{1}_{\{X_j \in A_n^k\}},$$

which, for a fixed k, are independent and identically distributed according to a Bernoulli distribution with parameter $p_k \leq \pi C_f \rho_\xi^2 e^{2k}/n$. Since $N_n^k = \sum_{j=1}^{n} \chi_{j,k}$ has expectation at most $\pi C_f \rho_\xi^2 e^{2k}$, Markov's inequality yields

$$\mathbb{P}\left(N_n^k \geq \pi C_f \rho_\xi^2 e^{2k} + \frac{e^k}{\sqrt{\varepsilon_n}}\right) \leq \frac{\varepsilon_n^2 \mathbb{E}\left[\left(N_n^k - \mathbb{E}[N_n^k]\right)^4\right]}{e^{4k}}. \tag{37}$$

In order to control the fourth central moment of N_n^k, recall that for two independent, real-valued random variables X and Y,

$$\begin{aligned}
&\mathbb{E}\left[(X + Y - \mathbb{E}[X] - \mathbb{E}[Y])^4\right] \\
&\qquad = \mathbb{E}\left[(X - \mathbb{E}[X])^4\right] + \mathbb{E}\left[(Y - \mathbb{E}[Y])^4\right] + 6 \operatorname{Var}(X) \operatorname{Var}(Y).
\end{aligned}$$

Since $\chi_{j,k}$ are iid, it follows by inductively applying the previous identity that

$$\begin{aligned}
\mathbb{E}\left[\left(N_n^k - \mathbb{E}[N_n^k]\right)^4\right] &= n\mathbb{E}\left[\left(\chi_{1,k} - \mathbb{E}[\chi_{1,k}]\right)^4\right] + 6\frac{n(n-1)}{2}\operatorname{Var}(\chi_{1,k})^2 \\
&\le n\left(\mathbb{E}[\chi_{1,k}^4] + 6\operatorname{Var}(\chi_{1,k})\left(\mathbb{E}[\chi_{1,k}]\right)^2\right) + 3n^2\operatorname{Var}(\chi_{1,k})^2 \\
&\le n\left(\mathbb{E}[\chi_{1,k}] + 6\left(\mathbb{E}[\chi_{1,k}]\right)^2\right) + 3n^2\left(\mathbb{E}[\chi_{1,k}]\right)^2 \\
&\le n\left(\frac{\pi C_f \rho_\xi^2 e^{2k}}{n} + 6\frac{\pi^2 C_f^2 \rho_\xi^4 e^{4k}}{n^2}\right) + 3n^2\frac{\pi^2 C_f^2 \rho_\xi^4 e^{4k}}{n^2} \\
&= O_{\mu,\xi}\left(e^{4k}\right).
\end{aligned}$$

Consequently, (37) becomes

$$\mathbb{P}\left(N_n^k \ge \pi C_f \rho_\xi^2 e^{2k} + \frac{e^k}{\sqrt{\varepsilon_n}}\right) = O_{\mu,\xi}(\varepsilon_n^2),$$

and by the union bound

$$\mathbb{P}\left(\bigcup_{k=0}^{d_n} F_n^k\right) = O_{\mu,\xi}\left(\log n \cdot \varepsilon_n^2\right).$$

The proof of Lemma 3.7 is complete. □

We have established that C_1, C_2, and k_{Lip} defined in (36) satisfy conditions (i), (ii), and (iii) of Theorem 3.1 for large n, on the complement of $E_n \cup \bigcup_{k=0}^{d_n} F_n^k \cup G_n$, a "bad" event whose probability tends to zero. Consequently, the conclusion of Theorem 3.1 guarantees that with probability at least $1 - o_{\mu,\xi}(1)$, the polynomial p_n has a unique critical point $w_\xi^{(n)}$ that fulfills (9).

We now consider the case $s > 1$. The argument in this more general situation is much the same as the one just presented for $s = 1$, so we sketch the proof and point out the major differences. Consider each of the roots ξ_l, $1 \le l \le s$ separately and modify the argument above in the obvious ways. In particular, we replace the annuli A_n^k with

$$\begin{aligned}
A_{l,n}^0 &:= \left\{z \in \mathbb{C} : |z - \xi_l| < \frac{\delta}{\sqrt{n}}\right\},\ 1 \le l \le s; \\
A_{l,n}^k &:= \left\{z \in \mathbb{C} : \frac{\delta e^{k-1}}{\sqrt{n}} \le |z - \xi_l| < \frac{\delta e^k}{\sqrt{n}}\right\},\ 1 \le k \le d_n,\ 1 \le l \le s;
\end{aligned}$$

where $\delta > 0$ is any real number such that f is a density for μ in the balls $B(\xi_l, \delta)$ and so that $2\delta < \min_{1 \le j < l \le s} |\xi_j - \xi_l|$. Define the random variables $N_{l,n}^k$ accordingly, in addition to the modified "bad" events

$$\begin{aligned}
E_{l,n} &= \left\{\left|\frac{1}{n+1-s}\sum_{j=1}^{n+1-s}\frac{1}{\xi_l - X_j} - m_\mu(\xi_l)\right| \ge \frac{|m_\mu(\xi_l)|}{3}\right\},\ 1 \le l \le s; \\
F_{l,n}^k &= \left\{N_{l,n}^k \ge \pi C_f \delta^2 e^{2k} + \frac{\delta e^k}{\sqrt{\varepsilon_n}}\right\},\ 0 \le k \le d_n,\ 1 \le l \le s; \\
G_{l,n} &= \left\{\min_{1 \le j \le n} |X_j - \xi_l| < \sqrt{\frac{\varepsilon_n}{n}}\right\},\ 1 \le l \le s;
\end{aligned}$$

and the modified constants

$$C_1^l := \frac{|m_\mu(\xi_l)|}{2}, \quad C_2^l := \frac{3\,|m_\mu(\xi_l)|}{2}, \quad \text{and} \quad k_{\text{Lip}}^l := C^l_{\mu,\vec{\xi}} \cdot \frac{\log n}{\varepsilon_n^{3/2}},\ 1 \le l \le s.$$

(Note that $C^l_{\mu,\vec{\xi}}$, $1 \le l \le s$, will be defined via lemmata similar to Lemma 3.6.) On the complement of the union of the modified "bad" events, for each l, $1 \le l \le s$, conditions (i), (ii), and (iii) of Theorem 3.1 hold for reasons similar to those given in the argument for $s = 1$ above. (Notice that for $1 \le l \le s$,

$$\left| \frac{1}{n} \sum_{k=1, k\neq l}^{s} \frac{1}{\xi_l - \xi_k} \right| = o(1),$$

so computations similar to (28) and (29) establish condition (i) of Theorem 3.1.) The fact that the union of the modified "bad" events occurs with probability at most $o(1)$ follows by an updated version of Lemma 3.7 and the union bound (recall s is fixed and finite).

We now turn our attention to (10) which describes the joint fluctuations of $w_l^{(n)}$, $1 \le l \le s$. This is considerably more difficult than our consideration of (13) because in the current situation, $(\xi_l - X_j)^{-1}$, are heavy-tailed random variables. In Appendix B, we appeal to the Lindeberg exchange method with an appropriate truncation to establish Theorem B.1, a CLT that we use to prove (10) in a similar manner to our justification of (13).

To start, consider that with probability $1 - o(1)$, $w_l^{(n)}$, $1 \le l \le s$ satisfy (9), so with inspiration from (32) and (33), we obtain with probability at least $1 - o(1)$ that for $1 \le l \le s$,

$$\begin{aligned} \frac{n^{3/2}}{\sqrt{\log n}} \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right) \\ = \sqrt{n} \left(\frac{1}{n} \sum_{j=1}^{n} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right) + o(1), \end{aligned}$$

where all of the implied constants depend on $\xi_1, \ldots, \xi_s$ and μ, and we have used Slutsky's theorem several times. (We also used the heavy-tailed CLT, Theorem B.1 once.) For the arbitrary constants $t_1, \ldots, t_s \in \mathbb{C}$, we have with probability at least $1 - o(1)$,

$$\begin{aligned} \operatorname{Re} \left(\sum_{l=1}^{s} t_l \frac{n^{3/2}}{\sqrt{\log n}} \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right) \right) \\ = \operatorname{Re} \left(\sqrt{\frac{n}{\log n}} \left(\frac{1}{n} \sum_{j=1}^{n} \sum_{l=1}^{s} t_l \left[\frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right] \right) \right) + o(1), \end{aligned}$$

which converges in distribution by Slutsky's theorem and Theorem B.1 to a normal distribution with with mean zero and variance $\sum_{l=1}^{s} \frac{\pi |t_l|^2 f(\xi_l)}{2}$. This is exactly the same distribution as the sum $\operatorname{Re}\left(\sum_{l=1}^{s} t_l N_l\right)$, where N_l are defined as in (10) with covariance structure (11). By the Cramér–Wold technique, this completes the proof of Theorem 2.8.

3.5. **Proof of Theorem 2.12.** We conclude this section by using Theorem 3.1 to prove Theorem 2.12.

Proof. Conclusions (i) and (ii) follow from [32, Theorem 1.7]. We now use Theorem 3.1 to establish (15). In particular, we will verify the three conditions of Theorem 3.1 hold for some constants $C_1, C_2, k_{\mathrm{Lip}} > 0$ which depend only on ε and λ. In view of parts (i) and (ii), it suffices to work on the event where

$$\max_{1\le i\le n-1} |X_i| \le 1+2\varepsilon, \quad \min_{1\le i\le n-1} |\xi - X_i| \ge \frac{\varepsilon}{2}, \quad 1+\frac{11}{4}\varepsilon \le |\xi| \le |\lambda|+1. \tag{38}$$

In fact, this event automatically guarantees the third condition from Theorem 3.1 for all values of n sufficiently large. The second condition also follows for large n since, for $z, w \in \mathbb{C}$ with $|z|, |w| > 1 + 5/2\varepsilon$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{z-X_i} - \frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{w-X_i}\right| \le \frac{|z-w|}{n}\sum_{i=1}^{n-1}\frac{1}{|z-X_i||w-X_i|} \ll_\varepsilon |z-w|$$

on the same event. The upper bound in the first condition of Theorem 3.1 follows from a similar argument. The lower bound, however, is slightly more involved. Indeed, for any $\theta \in \mathbb{R}$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\xi-X_i}\right| = \left|\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\xi e^{\sqrt{-1}\theta} - X_i e^{\sqrt{-1}\theta}}\right| \ge \frac{1}{n}\sum_{i=1}^{n-1}\frac{\mathrm{Re}(\xi e^{\sqrt{-1}\theta}) - \mathrm{Re}(X_i e^{\sqrt{-1}\theta})}{|\xi - X_i|^2}.$$

Choose $\theta \in \mathbb{R}$ so that $\xi e^{\sqrt{-1}\theta}$ is real-valued and positive. This gives

$$\left|\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\xi-X_i}\right| \ge \frac{1}{n}\sum_{i=1}^{n-1}\frac{\xi e^{\sqrt{-1}\theta} - \mathrm{Re}(X_i e^{\sqrt{-1}\theta})}{|\xi-X_i|^2} \ge \frac{1}{n}\sum_{i=1}^{n-1}\frac{|\xi|-|X_i|}{(|\xi|+|X_i|)^2}.$$

Thus, on the event (38), we conclude that

$$\left|\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\xi-X_i}\right| \gg_{\varepsilon,\lambda} 1, \tag{39}$$

which completes the proof of the lower bound. Hence, the three conditions of Theorem 3.1 are satisfied. Applying Theorem 3.1, we obtain (15). Lastly, (16) follows from (15) after applying conclusion (ii) and (39). □

4. Proof of results in Section 2.4

4.1. **Proof of Theorem 2.14.** This section is devoted to the proof of Theorem 2.14. We will need the following lemmata.

Lemma 4.1 (Monte Carlo sampling; Lemma 36 from [35])**.** *Let (X, μ) be a probability space, and let $F : X \to \mathbb{C}$ be a square-integrable function. Let $m \ge 1$, let $x_1, \dots, x_m$ be drawn independently at random from X with distribution μ, and let S be the empirical average*

$$S := \frac{1}{m}(F(x_1) + \cdots + F(x_m)).$$

Then S has mean $\int_X F d\mu$ and variance $\frac{1}{m}\int_X |F - \int_X F d\mu|^2 d\mu$. In particular, by Chebyshev's inequality, one has

$$\mathbb{P}\left(\left|S - \int_X F d\mu\right| \ge t\right) \le \frac{1}{mt^2}\int_X \left|F - \int_X F d\mu\right|^2 d\mu$$

for any $t > 0$, or equivalently, for any $\delta > 0$ one has with probability at least $1 - \delta$ that

$$\left| S - \int_X F d\mu \right| \leq \frac{1}{\sqrt{m\delta}} \left(\int_X \left| F - \int_X F d\mu \right|^2 d\mu \right)^{1/2}.$$

Lemma 4.2. *Fix $C > 0$, and let $X_1, \ldots, X_n$ be complex-valued random variables (not necessarily independent nor identically distributed) such that, with overwhelming probability,*

$$\max_{1 \leq i \leq n} |X_i| \leq e^{n^C}. \tag{40}$$

Let $\varphi : \mathbb{C} \to \mathbb{R}$ be a twice continuously differentiable function (possibly depending on n) which satisfies the pointwise bound in (18) *for all $z \in \mathbb{C}$. Then, with overwhelming probability,*

$$\int_{B(0,n^C)} |\Delta \varphi(z)|^2 \log^2 |p_n(z)| d^2 z \ll n^{2C} n^{O(1)}, \tag{41}$$

$$\int_{B(0,n^C)} |\Delta \varphi(z)|^2 \log^2 |p_n'(z)| d^2 z \ll n^{2C} n^{O(1)}, \tag{42}$$

and

$$\int_{B(0,n^C)} |\Delta \varphi(z)|^2 d^2 z \ll n^{4C}. \tag{43}$$

Proof. The bound in (43) follows immediately from the pointwise bound in (18). In order to prove (41) it suffices, by the pointwise bound in (18), to prove that with overwhelming probability

$$\int_{B(0,n^C)} \log^2 |p_n(z)| d^2 z \ll n^{O(1)}.$$

By supposition, we now work on the event where $X_1, \ldots, X_n \in B(0, e^{n^C})$. As

$$\log^2 |p_n(z)| \ll n \sum_{i=1}^n \log^2 |z - X_i|,$$

it suffices to prove that

$$\max_{1 \leq i \leq n} \int_B \log^2 |z - X_i| d^2 z \ll n^{O(1)},$$

where $B := B(0, n^C)$. Since $X_1, \ldots, X_n \in B(0, e^{n^C})$, it follows that

$$\max_{1 \leq i \leq n} \int_{B \setminus B(X_i,1)} \log^2 |z - X_i| d^2 z \ll n^{2C} |B| \ll n^{O(1)},$$

where $|B|$ is the Lebesgue measure of B, and $|B| = O(n^{2C})$. Near each root, we have

$$\max_{1 \leq i \leq n} \int_{B \cap B(X_i,1)} \log^2 |z - X_i| d^2 z \leq \max_{1 \leq i \leq n} \int_{B(X_i,1)} \log^2 |z - X_i| d^2 z \ll 1$$

since $\log |\cdot|$ is locally square-integrable. This completes the proof of (41).

For (42), we observe that on the event where (40) holds, the Gauss–Lucas theorem implies that

$$\max_{1 \leq j \leq n-1} |w_j| \leq e^{n^C},$$

where $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ are the critical points of p_n. Working on this event, the proof follows from the same procedure as we used to prove (41); we omit the details. □

Lemma 4.3 (Crude upper bound)**.** *Fix $C > 0$, and let $X_1, \dots, X_n$ be complex-valued random variables (not necessarily independent nor identically distributed). Assume Z is uniformly distributed on $B(0, n^C)$, independent of $X_1, \dots, X_n$. Then for every $a > 0$, there exits $b > 0$ such that*

$$\left| \sum_{i=1}^{n} \frac{1}{Z - X_i} \right| \leq n^b$$

with probability $1 - O_a(n^{-a})$.

Proof. Conditioning on $X_1, \dots, X_n$, we find that

$$\mathbb{P}\left(\min_{1 \leq i \leq n} |Z - X_i| \leq \varepsilon \right) \leq \sum_{i=1}^{n} \mathbb{P}(Z \in B(X_i, \varepsilon)) \ll n \frac{\varepsilon^2}{n^{2C}}$$

for all $\varepsilon > 0$. In addition, on the event where $\min_{1 \leq i \leq n} |Z - X_i| > \varepsilon$, we have

$$\left| \sum_{i=1}^{n} \frac{1}{Z - X_i} \right| \leq \frac{n}{\varepsilon}.$$

In order to prove the claim, it suffices to assume $a > 2C$. In this case, by taking $\varepsilon := \sqrt{\frac{n^{2C}}{n^{a+1}}}$, the result follows from the estimates above. □

We now prove Theorem 2.14.

Proof of Theorem 2.14. Let $B := B(0, n^C)$, and let $|B|$ denote its Lebesgue measure. Fix $\alpha > 0$, and let $\beta \in \mathbb{N}$ be a large constant (depending on C, c, α) to be chosen later.

Using the log-transform of the empirical measures constructed from the roots and critical points of p, we obtain

$$\sum_{i=1}^{n} \varphi(X_i) = \frac{1}{2\pi} \int_B \Delta\varphi(z) \log |p_n(z)| d^2 z, \quad \sum_{j=1}^{n-1} \varphi(w_j) \quad = \frac{1}{2\pi} \int_B \Delta\varphi(z) \log |p_n'(z))| d^2 z.$$

(These identities can also be found in a more general form in [16, Section 2.4.1].) Instead of working with the integrals on the right-hand sides, we will work with large empirical averages by applying Lemma 4.1. Indeed, let $m := n^\beta$, and let $Z_1, \dots, Z_m$ be iid random variables uniformly distributed on B, independent of $X_1, \dots, X_n$. Taking β sufficiently large and applying Lemmas 4.1 and 4.2, we conclude that

$$\frac{2\pi}{|B|} \sum_{i=1}^{n} \varphi(X_i) = \frac{1}{m} \sum_{l=1}^{m} \Delta\varphi(Z_l) \log |p_n(Z_l)| + O(n^{-c-2C}), \tag{44}$$

$$\frac{2\pi}{|B|} \sum_{j=1}^{n-1} \varphi(w_j) = \frac{1}{m} \sum_{l=1}^{m} \Delta\varphi(Z_l) \log |p_n'(Z_l)| + O(n^{-c-2C}), \tag{45}$$

$$\frac{1}{|B|} \int_B |\Delta\varphi(z)| d^2 z = \frac{1}{m} \sum_{l=1}^{m} |\Delta\varphi(Z_l)| + O(n^{-c-2C-1}) \tag{46}$$

with probability $1-O(n^{-\alpha})$. In addition, by (17), Lemma 4.3, and the union bound it follows that there exists $b > 0$ such that

$$n^{-b} \leq \min_{1\leq l\leq m} \left| \sum_{i=1}^{n} \frac{1}{Z_l - X_i} \right| \leq \max_{1\leq l\leq m} \left| \sum_{i=1}^{n} \frac{1}{Z_l - X_i} \right| \leq n^b$$

with probability $1 - O(n^{-\alpha})$. Thus, since $\frac{p_n'(z)}{p_n(z)} = \sum_{i=1}^{n} \frac{1}{z-X_i}$, we obtain

$$\sup_{1\leq l\leq m} |\log|p_n(Z_l)| - \log|p_n'(Z_l)|| = O(\log n) \tag{47}$$

with probability $1 - O(n^{-\alpha})$.

From (44) and (45), we find

$$\left| \frac{2\pi}{|B|} \sum_{i=1}^{n} \varphi(X_i) - \frac{2\pi}{|B|} \sum_{j=1}^{n-1} \varphi(w_j) \right|$$
$$\leq \frac{1}{m} \sum_{l=1}^{m} |\,\Delta\varphi(Z_l)|\,|\log|p_n(Z_l)| - \log|p_n'(Z_l)|| + O(n^{-c-2C})$$

with probability $1 - O(n^{-\alpha})$. Applying (46) and (47) yields

$$\left| \frac{2\pi}{|B|} \sum_{i=1}^{n} \varphi(X_i) - \frac{2\pi}{|B|} \sum_{j=1}^{n-1} \varphi(w_j) \right|$$
$$\ll (\log n) \frac{1}{m} \sum_{l=1}^{m} |\,\Delta\varphi(Z_l)| + n^{-c-2C}$$
$$\ll (\log n) \frac{1}{|B|} \int_B |\,\Delta\varphi(z)| d^2z + n^{-c-2C}$$

with probability $1 - O(n^{-\alpha})$. Since $|B| = \Theta(n^{2C})$, we rearrange to obtain

$$\left| \sum_{i=1}^{n} \varphi(X_i) - \sum_{j=1}^{n-1} \varphi(w_j) \right| \ll (\log n) \|\,\Delta\varphi\|_1 + n^{-c} \tag{48}$$

with probability $1 - O(n^{-\alpha})$. The proof of the theorem is complete. □

4.2. **Proof of Theorems 2.16 and 2.17.** In order to prove Theorem 2.16, it suffices to show that $X_1, \ldots, X_n$ satisfy the two axioms of Theorem 2.14. This follows from Lemmas 2.19 and 2.20.

We now turn to the proof of Theorem 2.17. By Theorem 2.16,

$$\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) = \sum_{i=1}^{n} \varphi(X_i) + O(\log n)$$

with probability $1 - O(n^{-100})$. Since φ is bounded, we obtain

$$\mathbb{E}\left[\sum_{j=1}^{n-1} \varphi(w_j^{(n)}) \right] = \sum_{i=1}^{n} \mathbb{E}\varphi(X_i) + O(\log n).$$

Therefore, we conclude that

$$\frac{1}{\sqrt{n}}\left(\sum_{j=1}^{n-1}\varphi(w_j^{(n)}) - \mathbb{E}\left[\sum_{j=1}^{n-1}\varphi(w_j^{(n)})\right]\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi(X_i) - \mathbb{E}\varphi(X_i)\right) + o(1)$$

with probability $1 - O(n^{-100})$. The claim now follows by applying the classical CLT to the right-hand side.

4.3. **Proof of Theorem 2.18.** We will need the following companion matrix result, which describes a matrix whose eigenvalues are the critical points of a given polynomial. This result appears to have originally been developed in [19] (see [19, Lemma 5.7]). However, the same result was later rediscovered and significantly generalized by Cheung and Ng [5, 6].

Theorem 4.4 (Lemma 5.7 from [19]; Theorem 1.2 from [6])**.** *Let $p(z) := \prod_{j=1}^{n}(z - z_j)$ for some complex numbers $z_1, \ldots, z_n$, and let D be the diagonal matrix $D := \operatorname{diag}(z_1, \ldots, z_n)$. Then*

$$\frac{1}{n}zp'(z) = \det\left(zI - D\left(I - \frac{1}{n}J\right)\right),$$

where I is the $n \times n$ identity matrix and J is the $n \times n$ all-one matrix.

We will also need the Sherman–Morrison formula for computing the inverse of a rank one update to a matrix.

Lemma 4.5 (Sherman–Morrison formula)**.** *Suppose A is an invertible matrix and u, v are column vectors. If $1 + v^{\mathrm{T}}A^{-1}u \neq 0$, then*

$$(A + uv^{\mathrm{T}})^{-1} = A^{-1} - \frac{A^{-1}uv^{\mathrm{T}}A^{-1}}{1 + v^{\mathrm{T}}A^{-1}u}.$$

Lemma 4.5 can be found in [1]; see also [15, Section 0.7.4] for a more general version of this identity known as the Sherman–Morrison–Woodbury formula. We also require the following consequence of [23, Lemma 4.1].

Lemma 4.6. *Under the assumptions of Theorem 2.18, there exists a constant $c' > 0$ (depending only on C, c, and ε) such that*

$$\inf_{z\in\Gamma}\left|\frac{z}{n}\sum_{i=1}^{n}\frac{1}{z - X_i}\right| \geq c'$$

with overwhelming probability.

Proof. Clearly $|z| = C + \varepsilon$ for all $z \in \Gamma$. Thus, it suffices to prove that

$$\inf_{z\in\Gamma}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{z - X_i}\right| \geq c'$$

with overwhelming probability. The claim now follows from the uniform bound in [23, Lemma 4.1] and the assumption on m_μ given in (19). □

With Lemma 4.6 in hand, we are now prepared to present the proof of Theorem 2.18.

Proof of Theorem 2.18. Let D be the diagonal matrix $D := \operatorname{diag}(X_1, \dots, X_n)$. Using the notation from Theorem 4.4, we observe that $zI - D$ is invertible for all $z \in \Gamma$ since $X_1, \dots, X_n \in B(0, C)$ by supposition. In addition, by the Gauss–Lucas theorem and Theorem 4.4, it must be the case that the eigenvalues of $D\left(I - \frac{1}{n}J\right)$ are also contained in $B(0, C)$. This implies that $zI - D\left(I - \frac{1}{n}J\right)$ is also invertible for every $z \in \Gamma$. In view of these observations, we define the resolvents

$$G(z) := (zI - D)^{-1}, \qquad R(z) := \left(zI - D\left(I - \frac{1}{n}J\right)\right)^{-1}$$

for $z \in \Gamma$.

Thus, by Cauchy's integral formula

$$\sum_{i=1}^{n} \varphi(X_i) = \frac{1}{2\pi\sqrt{-1}} \oint_{\Gamma} \varphi(z) \operatorname{tr} G(z) dz$$

and

$$\sum_{j=1}^{n-1} \varphi(w_j) + \varphi(0) = \frac{1}{2\pi\sqrt{-1}} \oint_{\Gamma} \varphi(z) \operatorname{tr} R(z) dz.$$

We now take the difference of these two equalities. Since $|\varphi(0)| \ll \int_\Gamma |\varphi(z)||dz|$, it suffices by the triangle inequality to show

$$\sup_{z \in \Gamma} |\operatorname{tr} G(z) - \operatorname{tr} R(z)| = O(1) \tag{49}$$

with overwhelming probability.

Since $J = \mathbf{1}\mathbf{1}^{\mathrm{T}}$, where $\mathbf{1}$ is the all-ones vector, the Sherman–Morrison formula (Lemma 4.5) implies that

$$R(z) = G(z) - \frac{\frac{1}{n}G(z)DJG(z)}{1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}}G(z)D\mathbf{1}} \tag{50}$$

provided $1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}}G(z)D\mathbf{1} \neq 0$. In view of Lemma 4.6, there exists a constant $c' > 0$ (depending only on C, c, and ε) such that

$$\inf_{z \in \Gamma} \left|1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}}G(z)D\mathbf{1}\right| = \inf_{z \in \Gamma} \left|\frac{z}{n}\sum_{i=1}^{n} \frac{1}{z - X_i}\right| \geq c' \tag{51}$$

with overwhelming probability. Here, we have exploited the fact that D and $G(z)$ are diagonal matrices, which implies that $\mathbf{1}^{\mathrm{T}}G(z)D\mathbf{1} = \sum_{i=1}^{n} \frac{X_i}{z - X_i}$. Using (50) and (51), we conclude that with overwhelming probability

$$\sup_{z \in \Gamma} |\operatorname{tr} G(z) - \operatorname{tr} R(z)| \leq \frac{1}{nc'} \sup_{z \in \Gamma} |\operatorname{tr}[G(z)DJG(z)]|.$$

To bound this last remaining term, we again exploit the fact that $J = \mathbf{1}\mathbf{1}^{\mathrm{T}}$. Indeed, from the cyclic property of the trace, we have the deterministic bound

$$|\operatorname{tr}[G(z)DJG(z)]| = \left|\mathbf{1}^{\mathrm{T}}G^2(z)D\mathbf{1}\right| = \left|\sum_{i=1}^{n} \frac{X_i}{(z - X_i)^2}\right| \leq \sum_{i=1}^{n} \frac{|X_i|}{|z - X_i|^2} \leq n\frac{C}{\varepsilon^2}$$

for all $z \in \Gamma$. Combining the bounds above, we obtain (49), and the proof is complete. □

5. Proof of Theorem 2.3

This section is devoted to proving Theorem 2.3. Our first lemma shows that Assumption 2.2 implies Assumption 2.1.

Lemma 5.1 (Sufficiency of Assumption 2.2)**.** *If μ satisfies Assumption 2.2, then μ also satisfies Assumption 2.1.*

Proof. Without loss of generality, suppose μ is radially symmetric about $z = 0$, and let $X \sim \mu$. By Lemma A.1, we can write

$$|m_\mu(z)| = \frac{\mathbb{P}(|X| < |z|)}{|z|},$$

so the hypotheses guarantee that $|m_\mu(z)|$ is continuous on $\mathbb{C}\backslash\{0\}$. (Indeed, $\mathbb{P}(|X| < r)$ is the cumulative distribution function associated to the radial part of μ, which has a continuous density.) Since $f(0) > 0$, there are $\delta, c > 0$ so that $|z| \le \delta$ implies $|f(z)| \ge c > 0$. In particular, for $|z| \le \delta$,

$$|m_\mu(z)| = \frac{1}{|z|}\int_0^{|z|} r f(r)\, dr \ge \frac{c}{|z|}\int_0^{|z|} r\, dr = \frac{c\,|z|}{2}. \tag{52}$$

Let $r_{1/2}$ be any value for which $\mathbb{P}(|X| < r_{1/2}) = 1/2$. By the extreme value theorem, $|m_\mu(z)|$ achieves its minimum, $m_{\min}$, on the closed, bounded annulus

$$A := \left\{z \in \mathbb{C} : \delta \le |z| \le r_{1/2}\right\}.$$

We know that $m_{\min}$ is non-zero by (52) and the fact that $\mathbb{P}(|X| < r)$ is non-decreasing in r. This second fact additionally implies that for $|z| \ge r_{1/2}$,

$$|m_\mu(z)| = \frac{\mathbb{P}(|X| < |z|)}{|z|} \ge \frac{1}{2\,|z|}.$$

We conclude that for any $\varepsilon \in (0, m_{\min})$,

$$\mathbb{P}(|m_\mu(X)| < \varepsilon) \le \mathbb{P}\left(\frac{c\,|X|}{2} < \varepsilon\right) + \mathbb{P}(m_{\min} < \varepsilon) + \mathbb{P}\left(\frac{1}{2\,|X|} < \varepsilon\right) \le C\varepsilon^2, \tag{53}$$

for some $C > 0$. (We have used the fact that μ has two finite absolute moments to bound the last probability.) It follows that μ satisfies Assumption 2.1 part (i).

To see that μ satisfies Assumption 2.1 part (ii), let $X_1, \dots, X_n$ be iid complex-valued random variables with distribution μ, and observe that

$$\mathbb{P}\left(\max_j |X_j| > \sqrt{n \log n}\right) = 1 - \mathbb{P}\left(|X_1| \le \sqrt{n \log n}\right)^n.$$

By Markov's inequality,

$$\mathbb{P}\left(|X_1| \le \sqrt{n \log n}\right)^n \ge \left(1 - \frac{\mathbb{E}\,|X_1|^2}{n \log n}\right)^n \xrightarrow{n\to\infty} 1,$$

which completes the argument. □

5.1. **Introduction to and motivation for the proof of Theorem 2.3.** The following proof of Theorem 2.3 is motivated by the illustration in Figure 5 that depicts the roots (red dots) and critical points (blue crosses) of $p_n(z)$ when the roots, $X_1, \ldots, X_{150}$ are chosen independently and uniformly in the unit disk centered at the origin. The observer will notice two things:

1) since the X_j are chosen uniformly at random, they tend to "clump together," and
2) the roots further from the origin tend to "pair" more closely with nearby critical points than the roots near the origin.

The first of these makes it difficult to use our strategy from Theorems 2.8, 2.10 and 3.1, where it was a simple matter to "zoom in" on a fixed root and ensure that no other roots were nearby. We address this concern by grouping the critical points that lie near each "clump" of roots and simultaneously considering all of the critical points that lie in the same group. We will show that each "clump" of roots (and its corresponding group of critical points) is far away from other "clumps," for large n.

The second observation can be explained by Theorem 2.8, which suggests that the closest critical point, $w_j^{(n)}$, to a given root X_j is at a distance $\frac{1}{n|m_\mu(X_j)|}$ from X_j. For example, in the case where μ is uniform on the unit disk, $|m_\mu(z)| = |z|$ for $|z| \leq 1$, so near the origin, it makes sense that the "pairing" phenomenon gets worse. We tackle this problem by counting the "clumps" of roots and critical points in exponentially widening, nested regions that avoid the zeros of m_μ. (In Figure 5, these are the annuli delimited by concentric dashed circles.) Using this method, we can take advantage of the fact that the number of "clumps" that are a given distance from the zero set of m_μ is roughly proportional to the strength of the "pairing" within those "clumps." The "pairing" phenomenon is quite unreliable near the zeros of m_μ, so for any "clumps" that are sufficiently close to the zeros of m_μ, we bound the distances between the roots and critical points using the Gauss–Lucas theorem. (In fact, this is where we expect to find the "extra," un-paired root that results because p_n has a higher degree than p_n').

In order to synthesize these two ideas, we will form random, disjoint, simple closed curves to encircle each "clump" of roots and critical points. We will build the curves from the arcs of circles centered at the roots of p_n and will use smaller circles for roots that are farther away from the zeros of m_μ. See, for example, the boundaries of the gray domains depicted in Figure 5. We will conclude with an argument involving Rouché's theorem to count the number of critical points interior to each curve by comparing p_n' to a simpler polynomial whose critical points can be located with Walsh's two circle theorem. Near the zeros of m_μ, our method breaks down, and we use the Gauss–Lucas theorem for a bound on the distances between the critical points and roots of p_n. Luckily, there are few critical points near the zeros of m_μ, a fact which follows in part from Assumptions 2.1 and 2.2.

5.2. **Definitions.** In view of Lemma 5.1, we prove Theorem 2.3 under Assumption 2.1. Let $C_\mu > 0$ be larger than each of the constants in Assumption 2.1 and larger than the constant bounding the density associated to μ. For each $n \in \mathbb{N}$, define the

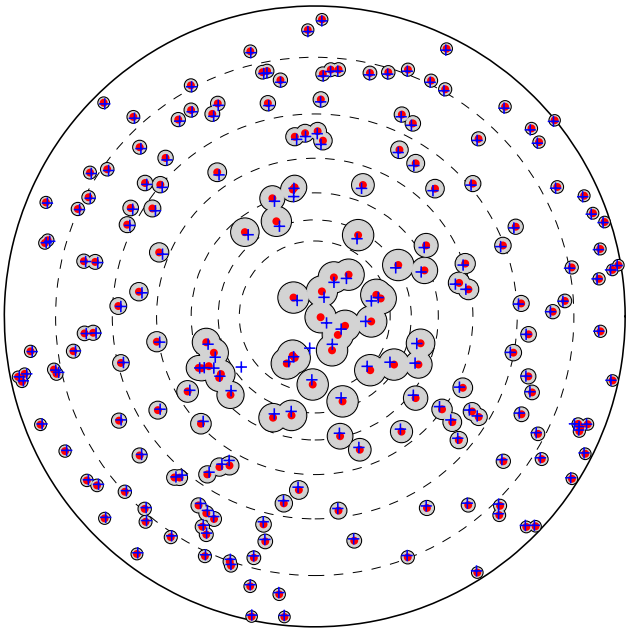

FIGURE 5. An illustration motivating the strategy we use to prove Theorem 2.3. The red circles and blue crosses represent the locations of the roots and critical points, respectively, of $p_{150}(z)$, where μ is the uniform distribution on the unit disk. Roughly speaking, the gray disks around the X_j are of radius $\max\{1/(n\,|m_\mu(X_j)|), 1/\sqrt{n}\}$. The dashed concentric circles are meant to divide the unit disk into exponentially widening annuli.

following sets which partition $\mathbb{C}$ into regions based on the size of $|m_\mu(z)|$:

$$A_n^k := \left\{ z \in \mathbb{C} : |m_\mu(z)| < \frac{e^k}{\sqrt{n}} \right\}, \ k = \lfloor 4\log(\log n) \rfloor,$$

$$A_n^k := \left\{ z \in \mathbb{C} : \frac{e^{k-1}}{\sqrt{n}} \le |m_\mu(z)| < \frac{e^k}{\sqrt{n}} \right\}, \ \lfloor 4\log(\log n) \rfloor + 1 \le k \le \left\lfloor \log\left(\sqrt{n}\right) \right\rfloor,$$

$$A_n := \left\{ z \in \mathbb{C} : |m_\mu(z)| \ge \frac{e^{\lfloor \log(\sqrt{n}) \rfloor}}{\sqrt{n}} \right\}.$$

Additionally, define the random variables

$$N_n^k := \#\left\{ 1 \le j \le n : X_j \in A_n^k \right\}, \ \lfloor 4\log(\log n) \rfloor \le k \le \left\lfloor \log\left(\sqrt{n}\right) \right\rfloor,$$

$$\zeta_{i,j}^{(n)} := \begin{cases} \dfrac{1}{X_i - X_j} \mathbb{1}_{|X_i - X_j| \ge \frac{(\log n)^2}{n|m_\mu(X_i)|}}, & |m_\mu(X_i)| \ne 0 \\ 0, & \text{otherwise} \end{cases}, \ 1 \le i, j \le n, \ j \ne i,$$

and let $\mathcal{N}_n$ be a $n^{-1/2}$-net of the closed disk $\overline{B}(0, n^{C_\mu})$ that satisfies:

(i) $\overline{B}(0, n^{C_\mu}) \subseteq \bigcup_{x \in \mathcal{N}_n} B(x, n^{-1/2})$,

(ii) if $x, y \in \mathcal{N}_n$, and $x \neq y$, then $|x - y| \geq \frac{1}{2\sqrt{n}}$,
(iii) $\#\mathcal{N}_n = O_\mu(n^{1+2C_\mu})$.

Such a collection of points exists by e.g. Lemma 3.3 in [23]. Let $\delta > 0$ be a fixed real parameter to be chosen later. We will show that the conclusion of Theorem 2.3 holds on the complement of the union of the following "bad" events:

$$E_n^k := \left\{ N_n^k \geq 2C_\mu e^{2k} \log(\log n) \right\}, \ \lfloor 4\log(\log n) \rfloor \leq k \leq \left\lfloor \log\left(\sqrt{n}\right) \right\rfloor;$$

$$F_n^i := \left\{ |m_\mu(X_i)| \geq \frac{(\log n)^4}{\sqrt{n}}, \left| \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} | X_i] \right) \right| \geq \frac{|m_\mu(X_i)|}{2} \right\},$$
for $1 \leq i \leq n$;

$$G_n^\delta := \left\{ \exists x \in \mathcal{N}_n \cup \{X_i\}_{i=1}^n \text{ s.t. } \# \left\{ 1 \leq j \leq n : |X_j - x| < \frac{1}{\sqrt{n}} \right\} \geq 2 + \delta \log n \right\};$$

$$H_n := \left\{ \eta_n \geq n^{C_\mu} \right\}.$$

For convenience, we use $\mathcal{E}_n^{\text{bad}}$ to denote the union of all of the "bad" events:

$$\mathcal{E}_n^{\text{bad}} := \bigcup_{k=\lfloor 4\log(\log n) \rfloor}^{\lfloor \log(c_\mu \sqrt{n}) \rfloor} E_n^k \cup \bigcup_{i=1}^{n} F_n^i \cup G_n^\delta \cup H_n.$$

5.3. **The "bad" events are unlikely.** In this subsection, we establish that

$$\mathbb{P}\left(\mathcal{E}_n^{\text{bad}}\right) = o(1). \tag{54}$$

By assumption, $\mathbb{P}(H_n) = o(1)$, so it remains to bound the probabilities of the remaining events.

Lemma 5.2.

$$\mathbb{P}\left(\bigcup_{k=\lfloor 4\log(\log n) \rfloor}^{\lfloor \log(\sqrt{n}) \rfloor} E_n^k \right) \leq \frac{1}{C_\mu [\log(\log n)]^2} = o(1).$$

Proof. Observe that for a fixed n and k, $\lfloor 4\log(\log n) \rfloor \leq k \leq \lfloor \log(\sqrt{n}) \rfloor$, N_n^k is a binomial random variable with parameters n and $p_k \leq C_\mu e^{2k}/n$. By Markov's inequality, we have,

$$\begin{aligned}
\mathbb{P}\left(N_n^k \geq 2C_\mu e^{2k} \log(\log n) \right) &\leq \mathbb{P}\left(\left| N_n^k - \mathbb{E}\left[N_n^k \right] \right| \geq C_\mu e^{2k} \log(\log n) \right) \\
&\leq \frac{\text{Var}\left(N_n^k \right)}{C_\mu^2 e^{4k} [\log(\log n)]^2} \\
&= \frac{np_k(1-p_k)}{C_\mu^2 e^{4k} [\log(\log n)]^2} \\
&\leq \frac{1}{C_\mu e^{2k} [\log(\log n)]^2}.
\end{aligned}$$

If we take the union over k, we obtain

$$\mathbb{P}\left(\bigcup_{k=\lfloor 4\log(\log n)\rfloor}^{\lfloor \log(\sqrt{n})\rfloor} E_n^k\right) \leq \sum_{k=1}^{\infty} \frac{1}{C_\mu e^{2k}[\log(\log n)]^2} = \frac{1}{C_\mu(e^2-1)[\log(\log n)]^2},$$

which implies the desired result. □

Lemma 5.3. $\mathbb{P}\left(\bigcup_{i=1}^n F_n^i\right) = o(1)$.

Proof. We will use the method of moments to control the probability of each F_n^i, $1 \leq i \leq n$. Since $F_n^i \subset \left\{|m_\mu(X_i)| \geq n^{-1/2}\right\}$, we will often assume that $|m_\mu(X_i)| \geq n^{-1/2}$ in our calculations. Recall from Lemma 3.5, part (i) that $|m_\mu(X_i)|$ is almost surely bounded above by an absolute constant (that depends only on μ).

First, consider that for complex-valued random variables X and Y, where Y has a finite fourth absolute moment,

$$\begin{aligned}\mathbb{E}\left[|Y - \mathbb{E}[Y \mid X]|^4 \;\middle|\; X\right] \\ \leq \mathbb{E}\left[|Y|^4 \;\middle|\; X\right] + 6\left(\mathbb{E}\left[|Y|^2 \;\middle|\; X\right]\right)^2 + 4\mathbb{E}\left[|Y|^3 \;\middle|\; X\right] \cdot \mathbb{E}\left[|Y| \mid X\right].\end{aligned} \tag{55}$$

(This inequality could be derived by writing

$$|Y - \mathbb{E}[Y \mid X]|^4 = \left(Y - \mathbb{E}[Y \mid X]\right)^2 \left(\overline{Y} - \overline{\mathbb{E}[Y \mid X]}\right)^2,$$

expanding the expression at right, and bounding each of the resulting terms with an appropriate term from the right side of (55).)

Now, for $X = X_i$ and $Y = \zeta_{i,j}^{(n)}$, where $1 \leq i, j \leq n$ with $j \neq i$,

$$\begin{aligned}\mathbb{E}\left[|Y|^4 \;\middle|\; X_i\right] &\leq \mathbb{E}\left[\frac{1}{|X_i - X_j|^4} \mathbb{1}_{\frac{(\log n)^2}{n|m_\mu(X_i)|} \leq |X_i - X_j| \leq 1} \;\middle|\; X_i\right] \\ &\qquad + \mathbb{E}\left[\frac{1}{|X_i - X_j|^4} \mathbb{1}_{|X_i - X_j| > 1} \;\middle|\; X_i\right] \\ &\leq 2\pi C_\mu \int_{\frac{(\log n)^2}{n|m_\mu(X_i)|}}^1 \frac{r}{r^4}\, dr + 1 \\ &= \frac{\pi C_\mu n^2 |m_\mu(X_i)|^2}{(\log n)^4} - \pi C_\mu + 1,\end{aligned}$$

and similarly,

$$\begin{aligned}\mathbb{E}\left[|Y|^3 \;\middle|\; X_i\right] &\leq \frac{2\pi C_\mu n |m_\mu(X_i)|}{(\log n)^2} - 2\pi C_\mu + 1, \\ \mathbb{E}\left[|Y|^2 \;\middle|\; X_i\right] &\leq 2\pi C_\mu \log\left(\frac{n|m_\mu(X_i)|}{(\log n)^2}\right) + 1, \\ \mathbb{E}\left[|Y| \mid X_i\right] &\leq 2\pi C_\mu + 1.\end{aligned}$$

Consequently, via (55), there are positive constants C'_μ, K_μ that depend only on μ so that if $n \geq K_\mu$, on the event $|m_\mu(X_i)| \geq n^{-1/2}$,

$$\mathbb{E}\left[\left|\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} \;\middle|\; X_i\right]\right|^4 \;\middle|\; X_i\right] \leq \frac{C'_\mu |m_\mu(X_i)|^2 n^2}{(\log n)^4}. \tag{56}$$

Next, we show that there are constants $C''_\mu, K'_\mu > 0$ that depend only on μ, so that for $n \geq K'_\mu$ and any fixed i, $1 \leq i \leq n$,

$$\mathbb{1}_{|m_\mu(X_i)| \geq \frac{1}{\sqrt{n}}} \cdot \mathbb{E}\left[\left| \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} | X_i \right] \right) \right|^4 \,\middle|\, X_i \right] \leq \frac{C''_\mu \left| m_\mu(X_i) \right|^2 n^3}{(\log n)^4}. \tag{57}$$

Write

$$\begin{aligned} &\mathbb{E}\left[\left| \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} | X_i \right] \right) \right|^4 \,\middle|\, X_i \right] \\ &\qquad = \mathbb{E}\left[\left(\sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} | X_i \right] \right) \right)^2 \left(\overline{\sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} | X_i \right] \right)} \right)^2 \,\middle|\, X_i \right], \end{aligned}$$

and observe that if we distribute the factors inside the expectation, the independence of $\{X_j\}_{j=1}^n$ implies that the only terms which contribute to a nonzero expectation are bounded by expectations of the form

$$\mathbb{E}\left[\left| \zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} \;\middle|\; X_i \right] \right|^2 \cdot \left| \zeta_{i,k}^{(n)} - \mathbb{E}\left[\zeta_{i,k}^{(n)} \;\middle|\; X_i \right] \right|^2 \,\middle|\, X_i \right],$$

where $1 \leq j, k \leq n$ and $j, k \neq i$. By a routine counting argument and the fact that $\zeta_{i,j}^{(n)}$, $j \neq i$ are identically distributed, it follows that

$$\begin{aligned} &\mathbb{E}\left[\left| \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} \;\middle|\; X_i \right] \right) \right|^4 \,\middle|\, X_i \right] \\ &\qquad \leq (n-1)\mathbb{E}\left[\left| \zeta_{i,l}^{(n)} - \mathbb{E}\left[\zeta_{i,l}^{(n)} \;\middle|\; X_i \right] \right|^4 \,\middle|\, X_i \right] \\ &\qquad\qquad + \binom{n-1}{2}\binom{4}{2}\left(\mathbb{E}\left[\left| \zeta_{i,l}^{(n)} - \mathbb{E}\left[\zeta_{i,l}^{(n)} \;\middle|\; X_i \right] \right|^2 \,\middle|\, X_i \right] \right)^2, \end{aligned}$$

where $l \neq i$ is any fixed index. From (56) and the bounds on $\mathbb{E}[\left|Y^2\right| \mid X_i]$ and $\mathbb{E}[|Y| \mid X_i]$ above, we can find $C''_\mu, K'_\mu > 0$ large enough so that $n \geq K'_\mu$ implies (57). (For the asymptotics, we are using that $n^{-1/2} \leq |m_\mu(X_i)| = O_\mu(1)$, where the implied constant depends only on μ.) Via Markov's inequality, it follows that for $n \geq K'_\mu$ and a fixed i, $1 \leq i \leq n$, on the event $|m_\mu(X_i)| \geq n^{-1/2}$,

$$\mathbb{P}\left(\left| \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} | X_i] \right) \right| \geq \frac{|m_\mu(X_i)|}{2} \,\middle|\, X_i \right) \leq \frac{C''_\mu}{n \left| m_\mu(X_i) \right|^2 (\log n)^4}. \tag{58}$$

We conclude the proof by demonstrating that $\mathbb{P}(\cup_{i=1}^n F_n^i) = o(1)$. Indeed, for $n \geq K'_\mu$,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^n F_n^i\right) &\leq n\mathbb{P}\left(F_n^1\right) \\
&= n\mathbb{P}(\emptyset) + n \sum_{k=\lfloor 4\log(\log n)\rfloor+1}^{\lfloor \log(\sqrt{n})\rfloor} \mathbb{P}\left(\left\{X_1 \in A_n^k\right\} \cap F_n^1\right) + n \cdot \mathbb{P}\left(\left\{X_1 \in A_n\right\} \cap F_n^1\right) \\
&= n \sum_{k=\lfloor 4\log(\log n)\rfloor+1}^{\lfloor \log(\sqrt{n})\rfloor} \mathbb{E}\left(\mathbb{1}_{\{X_1 \in A_n^k\}} \cdot \mathbb{P}(F_n^1 \mid X_1)\right) + n \cdot \mathbb{E}\left(\mathbb{1}_{\{X_1 \in A_n\}} \cdot \mathbb{P}(F_n^1 \mid X_1)\right) \\
&\leq n \sum_{k=\lfloor 4\log(\log n)\rfloor+1}^{\lfloor \log(\sqrt{n})\rfloor} \mathbb{E}\left(\frac{C''_\mu \cdot \mathbb{1}_{\{X_1 \in A_n^k\}}}{n\left|m_\mu(X_1)\right|^2 (\log n)^4}\right) + n \cdot \mathbb{E}\left(\frac{C''_\mu \cdot \mathbb{1}_{\{X_1 \in A_n\}}}{n\left|m_\mu(X_1)\right|^2 (\log n)^4}\right) \\
&\leq \sum_{k=\lfloor 4\log(\log n)\rfloor+1}^{\lfloor \log(\sqrt{n})\rfloor} \frac{C''_\mu \cdot n^2 \cdot \mathbb{P}(X_1 \in A_n^k)}{n(\log n)^4 e^{2k-2}} + \frac{C''_\mu \cdot n^2 \cdot \mathbb{P}(X_1 \in A_n)}{n(\log n)^4 e^{2\lfloor \log(\sqrt{n})\rfloor}},
\end{aligned}$$

where we used (58) to bound $\mathbb{P}(F_n^1 \mid X_1)$. Assumption 2.1 guarantees that

$$\mathbb{P}(X_1 \in A_n^k) \leq C_\mu \cdot \frac{e^{2k}}{n}, \ \lfloor 4\log(\log n)\rfloor \leq k \leq \left\lfloor \log\left(\sqrt{n}\right)\right\rfloor.$$

We also have

$$e^{2\lfloor \log(\sqrt{n})\rfloor} \geq e^{2\log(\sqrt{n})-2} = ne^{-2}.$$

Hence, for large n, our calculation from above yields

$$\mathbb{P}\left(\bigcup_{i=1}^n F_n^i\right) \leq \sum_{k=1}^{\lfloor \log(\sqrt{n})\rfloor} \frac{C''_\mu C_\mu e^2}{(\log n)^4} + \frac{C''_\mu e^2}{(\log n)^4} \cdot 1 = o(1).$$

□

Lemma 5.4. *For a fixed* $\delta \in \left(0, \frac{1}{2\pi C_\mu}\right)$,

$$\mathbb{P}(G_n^\delta) = O_\mu\left(\frac{n^{2+2C_\mu}}{(1+\delta\log n)^{(2+\delta\log n)}}\right) = o_\delta(1).$$

Proof. This is a straight-forward application of the Chernoff bound for binomial random variables. In particular, for each $x \in \mathcal{N}_n$, define the random variable

$$N_x := \sum_{j=1}^n \mathbb{1}_{|X_j - x| \leq \frac{1}{\sqrt{n}}},$$

which has a binomial distribution with parameters n and $p \leq \pi C_\mu / n$. The moment generating function for N_x is

$$\mathbb{E}[e^{tN_x}] = (1 + p(e^t - 1))^n \leq e^{np(e^t-1)} \leq e^{\pi C_\mu (e^t - 1)}.$$

Choosing $t = \log(1 + 1/(\pi C_\mu)\log n)$ establishes

$$\mathbb{E}\left[(1 + 1/(\pi C_\mu)\log n)^{N_x}\right] \leq n,$$

and by Markov's inequality, we obtain

$$\mathbb{P}\left(N_x \geq 2 + \delta \log n\right) \leq \frac{\mathbb{E}\left[(1 + 1/(\pi C_\mu) \log n)^{N_x}\right]}{(1 + 1/(\pi C_\mu) \log n)^{(2+\delta \log n)}} \leq \frac{n}{(1 + 1/(\pi C_\mu) \log n)^{(2+\delta \log n)}}.$$

Note that the bound is independent of x, and that the argument can be easily modified (by conditioning on X_i) to show that for a fixed $1 \leq i \leq n$,

$$\mathbb{P}\left(\sum_{\substack{j=1 \\ j \neq i}} \mathbb{1}_{|X_j - X_i| \leq \frac{1}{\sqrt{n}}} \geq 2 + \delta \log n\right) \leq \frac{n}{(1 + 1/(2\pi C_\mu) \log n)^{(2+\delta \log n)}}.$$

Hence, we can apply the the union bound over all $x \in \mathcal{N}_n$ and $X_1, \ldots, X_n$ to obtain the desired result. □

Combining Lemmas 5.2, 5.3, and 5.4 from this subsection establishes (54), so for the remainder of the proof, we work on the complements of the "bad" events.

5.4. **Constructing disjoint domains that partition the roots.** We will create disjoint domains which contain clusters of roots of $p_n(z)$ that are close to one another and show that inside each domain, the numbers of roots and critical points of $p_n(z)$ are the same. The domains will be disjoint to ensure that no roots or critical points are counted more than once (see Figure 5 for reference). For technical reasons involving Rouché's theorem, we will require that the boundaries of the regions be simple, closed curves.

Our strategy will be to make an open ball around each X_i, $1 \leq i \leq n$, and to consider the path-connected components of the union of these balls. Some of the resulting regions may not be simply connected, so we need to "fill in the holes." To start, define the random collection of open balls

$$\mathcal{C}_n := \left\{ B\left(x, \frac{(\log n)^3}{n \cdot \max\left\{|m_\mu(x)|, \frac{(\log n)^4}{\sqrt{n}}\right\}}\right) : x \in \{X_j\}_{j=1}^n \right\},$$

and define on $\{1, 2, \ldots, n\}$ the equivalence relation given by the following rule: $i \sim j$ if and only if there is a collection

$$\{B_0, B_1, \ldots, B_l\} \subset \mathcal{C}_n,$$

with

$$B_0 = B\left(X_i, \frac{(\log n)^3}{n \cdot \max\left\{|m_\mu(X_i)|, \frac{(\log n)^4}{\sqrt{n}}\right\}}\right),$$

and

$$B_l = B\left(X_j, \frac{(\log n)^3}{n \cdot \max\left\{|m_\mu(X_j)|, \frac{(\log n)^4}{\sqrt{n}}\right\}}\right),$$

such that $B_k \cap B_{k+1} \neq \emptyset$ for $0 \leq k \leq l-1$. Let $\mathcal{P}_n$ be the set of equivalence classes induced by $\sim$. The idea is that for a fixed $P \in \mathcal{P}_n$,

$$\mathcal{U}_{n,P} := \bigcup_{i \in P} B\left(X_i, \frac{(\log n)^3}{n \cdot \max\left\{|m_\mu(X_i)|, \frac{(\log n)^4}{\sqrt{n}}\right\}}\right)$$

forms a connected component of $\cup_{B \in \mathcal{C}_n} B$. Each light gray region in Figure 5 is one connected component, $\mathcal{U}_{n,P}$ for some $P \in \mathcal{P}_n$; a "zoomed-in" version is presented in Figure 7. Notice that some of the $\mathcal{U}_{n,P}$, $P \in \mathcal{P}_n$ may not have simple, closed boundaries, and some could be "nested" inside "holes" formed by others. We address these concerns in the following discussion, where we demonstrate how to select a simple, closed component of the boundary of each $\mathcal{U}_{n,P}$, $P \in \mathcal{P}_n$, whose interior contains $\mathcal{U}_{n,P}$.

More specifically, for each equivalence class $P \in \mathcal{P}_n$, we will create a simple closed curve, $\gamma_{n,P} \subset \partial\mathcal{U}_{n,P}$, such that each X_j, $j \in P$ is contained interior to the bounded component of $\mathbb{C} \setminus \gamma_{n,P}$. Furthermore, we will show that the interiors of the bounded regions defined by the curves $\{\gamma_{n,P}\}_{P \in \mathcal{P}_n}$ are partially ordered with respect to set inclusion. This will allow us to combine "nested" regions.

To that end, fix an equivalence class $P \in \mathcal{P}_n$, and recall the definition of the open set $\mathcal{U}_{n,P}$ from above. For simplicity, write $\mathcal{U}_{n,P} = \bigcup_{i=1}^{l} B_i$, where $B_1, \ldots, B_l$ are *distinct* open balls (in the definition of $\mathcal{U}_{n,P}$, some of the open balls could coincide if, for example $X_i = X_j$ for $i, j \in P$, $i \neq j$). We use $\mathcal{V}_{n,P}$ to denote the unique unbounded, path-connected component of the complement of $\overline{\mathcal{U}_{n,P}}$. (The complement of $\overline{\mathcal{U}_{n,P}}$ has a unique unbounded, path-connected component because $\overline{\mathcal{U}_{n,P}}$, a union of finitely many closed disks, is compact.) By construction, the boundaries $\partial\mathcal{U}_{n,P} \supseteq \partial\mathcal{V}_{n,P}$ consist of arcs of the finitely many circles $\partial B_1, \ldots, \partial B_l$.

Lemma 5.5. *The curve $\gamma_{n,P} := \partial\mathcal{V}_{n,P}$ is a simple, closed curve (i.e. a Jordan curve), and $\mathcal{U}_{n,P}$ is contained in the bounded component of $\mathbb{C} \setminus \gamma_{n,P}$.*

Proof. There are several ways that one could proceed. One method is to construct a simple path starting on the boundary $\partial\mathcal{V}_{n,P}$ that follows circle arcs until it returns to the start. A second approach is to consider the genus of the region $\mathcal{U}_{n,P}$, find generators for its fundamental group, and "close-off" any "holes." We present, in detail, a third method that relies on the following converse of the Jordan curve theorem due to Schönflies (see [8, 36], and the discussion on pp. 13 and 67 of [38]). The theorem statement requires two definitions.

A *region* of the closed set $F \subset \mathbb{C}$ is defined as a path-connected component of $\mathbb{C} \setminus F$. A point x in F is *accessible* from a region $\mathcal{R}$ if there is a point $y \in \mathcal{R}$ and a simple path from y to x, whose intersection with F is $\{x\}$.

Theorem 5.6 (Theorem 1 in [36]; see also Theorem II 5.38 on p. 67 of [38])**.** *If F is a compact set in $\mathbb{C}$ with precisely two regions such that every point of F is accessible from each of those regions, then F is a simple closed curve.*

Our goal is to show that the compact set $\gamma_{n,P} = \partial\mathcal{V}_{n,P}$ has precisely two regions from which $\gamma_{n,P}$ is accessible at every point. Define $\mathcal{U}'_{n,P} := \mathbb{C} \setminus \overline{\mathcal{V}_{n,P}}$. Observe that $\mathbb{C} \setminus \gamma_{n,P} = \mathcal{V}_{n,P} \cup \mathcal{U}'_{n,P}$, where the union is disjoint. It is clear that $\mathcal{V}_{n,P}$ is a region of $\gamma_{n,P}$; next, we argue that $\mathcal{U}'_{n,P}$ is also a region of $\gamma_{n,P}$.

Since $\mathcal{U}'_{n,P} \subset \mathbb{C}$ is open, it suffices to show that $\mathcal{U}'_{n,p}$ is connected. Suppose, for a contradiction, that this is not the case. Then, there are disjoint, non-empty open

(A) Case 1: y is on precisely one circle among $\{\partial B_i\}_{i=1}^l$.

(B) Case 2: y is on more than one of the circles $\{\partial B_i\}_{i=1}^l$.

FIGURE 6. The geometry near $y \in \gamma_{n,P}$.

sets $S, T \subset \mathbb{C}$ such that $S \cup T = \mathcal{U}'_{n,P}$. By construction, the open set $\mathcal{U}_{n,P} \subset \mathcal{U}'_{n,P}$ is path-connected, and hence connected, so $\mathcal{U}_{n,P}$ must be completely contained in either S or T. Suppose, without loss of generality, that $\mathcal{U}_{n,P} \subset S$. Since T is non-empty, there is some $x \in T$. We will demonstrate that a path whose image is contained entirely in $\mathcal{U}'_{n,P}$ connects x to a point of $\mathcal{U}_{n,P} \subset S$, which results in a contradiction. We may assume that $x \notin \partial\mathcal{U}_{n,P}$ because otherwise x lies on a one of the circles ∂B_i, $1 \leq i \leq l$, and there is a path in $\mathcal{U}'_{n,P}$ between x and a point of $\mathcal{U}_{n,P} \subset S$.

Since the (finitely many) circles $\partial B_1, \ldots, \partial B_l$ are distinct, there are only finitely many points of $\mathbb{C}$ that are contained in more than one circle. Consequently, we can choose a point $v \in \mathcal{V}_{n,P}$ such that the line segment $\overline{xv}$ does not contain any points of $\mathbb{C}$ that lie in the intersection of two or more distinct B_i, $1 \leq i \leq l$. (Indeed, choose a circle $\mathfrak{C}_x \subset \mathcal{V}_{n,P}$, centered at x, whose interior contains the compact set $\overline{\mathcal{U}'_{n,P}}$. Then, the collection $\{\overline{xz} : z \in \mathfrak{C}_x\}$ of line segments connecting x to points of $\mathfrak{C}_x$ is infinite in number. Also, $x \notin \partial U_{n,P}$ by assumption.) Define the path $\ell : [0,1] \to \mathbb{C}$ via $t \mapsto tx + (1-t)v$, whose image is the line segment $\overline{xv}$. Since $\overline{xv}$ is connected, it cannot be the case that $\overline{xv} \in \mathbb{C} \setminus \gamma_{n,P}$ (indeed, $\mathcal{U}'_{n,P} \cup \mathcal{V}_{n,P} = \mathcal{C} \setminus \gamma_{n,P}$ is a disjoint union of non-empty open sets). Consequently, $\overline{xv}$ contains a point of $\gamma_{n,P}$. Let $t^* := \min\{t : \ell(t) \in \gamma_{n,P}\}$ and set $y := \ell(t^*)$. Note that $t^* > 0$ since $x \notin \mathcal{U}_{n,P}$.

By construction, y lies on precisely one of the circles $\{\partial B_i\}_{i=1}^l$; suppose, without loss of generality, that $y \in \partial B_1$. Hence, we can choose an open ball $\mathcal{B}_y \ni y$ small enough that $\mathcal{B}_y \setminus \partial B_1$ consists of exactly two disjoint, path-connected open regions (See Figure 6A). One of these regions must be a subset of $B_1 \subset \mathcal{U}_{n,P}$, and the other must be a subset of $\mathcal{V}_{n,P}$. (The second region is connected and open, contains no points of $\partial\mathcal{V}_{n,P}$, and must contain a point of $\mathcal{V}_{n,P}$ because $y \in \partial\mathcal{V}_{n,P}$.)

Choose $\eta > 0$ small enough so that $t^* - \eta > 0$ and $\ell(t^* - \eta) \in \mathcal{B}_y$. It follows that the line segment

$$L := \{\ell(t) : 0 \leq t \leq t^* - \eta\}$$

is connected and disjoint from $\gamma_{n,P}$. We conclude that L is contained entirely in T, for it contains $x \in T$. This means L does not contain any points of $\mathcal{V}_{n,P}$, so $\ell(t^* - \eta) \in \mathcal{B}_y \cap B_1 \subset \mathcal{U}_{n,P} \subset S$. We have reached a contradiction since S and T are disjoint, so $\mathcal{U}'_{n,P}$ must be connected.

We have shown that $\gamma_{n,P}$ has precisely two regions, $\mathcal{V}_{n,P}$ and $\mathcal{U}'_{n,P}$. It remains to show that every point of $\gamma_{n,P}$ is accessible from both of these regions. Suppose

$y \in \gamma_{n,P}$. There are two cases: y is contained in precisely one of ∂B_i, $1 \leq i \leq l$, or y is contained in more than one of these circles. (See Figures 6A and 6B, respectively.)

If the first case is true, just as we did above, we can choose an open ball $\mathcal{B}_y \ni y$ small enough that $\mathcal{B}_y \setminus \partial B_1$ consists of the two disjoint, path-connected open regions $\mathcal{B}_y \cap \mathcal{U}_{n,P}$ and $\mathcal{B}_y \cap \mathcal{V}_{n,P}$. It is now clear that y is accessible from both $\mathcal{V}_{n,P}$ and $\mathcal{U}'_{n,P} \supset \mathcal{U}_{n,P}$.

On the other hand, suppose, without loss of generality, that y is contained in the circles $\partial B_1, \partial B_2, \ldots, \partial B_j$. Then, we can choose an open ball $\mathcal{B}_y \ni y$ small enough that $\mathcal{B}_y \setminus \bigcup_{i=1}^{j} \partial B_i$ consists of $2j$ disjoint path-connected, open regions that do not contain points from $\gamma_{n,P}$ (see Figure 6B). Consequently, each of these regions must be entirely contained in one of the disjoint open sets $\mathcal{U}'_{n,P}$ or $\mathcal{V}_{n,P}$. Since $y \in \partial \mathcal{U}'_{n,P} = \partial \mathcal{V}_{n,P}$, at least one of the $2j$ regions must be contained in $\mathcal{U}'_{n,P}$ and at least one must be contained in $\mathcal{V}_{n,P}$. It follows that y is accessible from both $\mathcal{V}_{n,P}$ and $\mathcal{U}'_{n,P}$.

We conclude via Theorem 5.6 that $\gamma_{n,P}$ is a simple closed curve whose interior contains $\mathcal{U}_{n,P}$ because $\mathcal{U}'_{n,P}$ is the bounded component of $\mathbb{C} \setminus \gamma_{n,P}$, and $\mathcal{U}_{n,P} \subset \mathcal{U}'_{n,P}$. □

We have shown that there are simple, closed curves $\{\gamma_{n,P}\}_{P \in \mathcal{P}_n}$ so that for each $P \in \mathcal{P}_n$, $\gamma_{n,P} \subseteq \partial \mathcal{U}_{n,P}$ and $\mathcal{U}_{n,P}$ is contained in the interior of the bounded region defined by $\gamma_{n,P}$. Furthermore, the path-connected, open regions $\{\mathcal{U}_{n,P}\}_{P \in \mathcal{P}_n}$ are disjoint by the definition of the equivalence relation $\sim$. This means that no curve $\gamma_{n,P}$ can pass through the interior of any region $\mathcal{U}_{n,P}$, and as a result, we can identify "maximal" curves which we will use in the remainder of the proof.

Definition 5.7. We say that a simple, closed curve γ_{n,P^*} among $\{\gamma_{n,P}\}_{P \in \mathcal{P}_n}$ is *maximal* if whenever $\mathcal{U}_{n,P^*}$ is in the bounded component of $\mathbb{C} \setminus \gamma_{n,P}$ for some $P \in \mathcal{P}_n$, we have $P = P^*$. We use $\mathcal{M}_n$ to denote the collection of maximal curves. For each $\Gamma \in \mathcal{M}_n$, let $\mathcal{O}_\Gamma$ denote the bounded component of $\mathbb{C} \setminus \Gamma$, so that $\partial \mathcal{O}_\Gamma = \Gamma$.

Notice that the domains $\mathcal{O}_\Gamma$, $\Gamma \in \mathcal{M}_n$ are disjoint by construction and that each X_j, $1 \leq j \leq n$, is contained in precisely one $\mathcal{O}_\Gamma$. We conclude this subsection with two important lemmas that restrict the sizes of the equivalence classes P, $P \in \mathcal{P}_n$ and domains $\mathcal{O}_\Gamma$, $\Gamma \in \mathcal{M}_n$.

Lemma 5.8. *Suppose $0 < \delta < 1/3$. There exists $C_\delta > 0$ so that for $n \geq C_\delta$, the following holds on the complement of G_n^δ: for each $P \in \mathcal{P}_n$, $|P| \leq \delta \log n + 2$, and if $x, y \in \overline{\mathcal{U}_{n,P}}$, then, $|x - y| < \frac{3\delta}{\sqrt{n}}$.*

Proof. Assume, for a contradiction, that there is a $P \in \mathcal{P}_n$ for which $|P| > \delta \log n + 2$, and suppose, without loss of generality, that $1 \in P$. By the definition of $\mathcal{P}_n$, for each $i \in P \setminus \{1\}$, there are elements $B_0^i, B_1^i, \ldots B_{l_i}^i \in \mathcal{C}_n$, where

$$B_0^i = B\left(X_1, \frac{(\log n)^3}{n \cdot \max\left\{|m_\mu(X_1)|, \frac{(\log n)^4}{\sqrt{n}}\right\}}\right),$$

$$B_{l_i}^i = B\left(X_i, \frac{(\log n)^3}{n \cdot \max\left\{|m_\mu(X_i)|, \frac{(\log n)^4}{\sqrt{n}}\right\}}\right),$$

$B_k^i \cap B_{k+1}^i \neq \emptyset$ for $0 \leq k \leq l_i - 1$, and $B_0^i, \ldots, B_{l_i}^i$ are balls with radius at most $(\log n)^{-1} n^{-1/2}$. Notice that the distance between X_1 and any X_i, $i \in P \setminus \{1\}$ is bounded by $2 + 2(l_i - 1)$ times this maximum radius (recall that X_1 and X_i, $i \in P \setminus \{1\}$ are the centers of B_0^i and $B_{l_i}^i$, respectively). We consider two cases:

(i) for every $i \in P \setminus \{1\}$, $l_i < \delta \log n + 2$
(ii) there is an $i^* \in P \setminus \{1\}$ for which $l_{i^*} \geq \delta \log n + 2$.

If case (i) is true, then, for n large enough to guarantee $\delta \log n \geq 3$,

$$\max_{i \in P \setminus \{1\}} |X_1 - X_i| < \max_{i \in P \setminus \{1\}} \frac{2 + 2(l_i - 1)}{\log n \sqrt{n}} < \frac{2 + 2(\delta \log n + 1)}{\log n \sqrt{n}} \leq \frac{3\delta}{\sqrt{n}} < \frac{1}{\sqrt{n}},$$

so every X_i, $i \in P$ is in the ball of radius $n^{-1/2}$ centered at X_1, which is impossible on the complement of G_n^δ. On the other hand, if case (ii) is true, then, for large n,

$$\bigcup_{k=0}^{\lceil \delta \log n + 2 \rceil} B_k^{i^*} \subset B\left(X_1, \frac{1}{\sqrt{n}}\right).$$

Indeed, $\left\{B_k^{i^*}\right\}_{k=0}^{\lceil \delta \log n + 2 \rceil}$ are overlapping balls with radius at most $(\log n)^{-1} n^{-1/2}$, so if n is large enough that $\delta \log n \geq 7$ and $y \in \bigcup_{k=0}^{\lceil \delta \log n + 2 \rceil} B_k^{i^*}$, then,

$$|y - X_1| \leq \frac{1 + 2\lceil \delta \log n + 2 \rceil}{\log n \sqrt{n}} < \frac{2\delta \log n + 7}{\log n \sqrt{n}} \leq \frac{3\delta}{\sqrt{n}} < \frac{1}{\sqrt{n}}.$$

This is impossible on the complement of G_n^δ because it would imply too many roots among $\{X_j\}_{j=1}^n$ in the ball of radius $n^{-1/2}$ centered at X_1.

Now, suppose $x, y \in \overline{\mathcal{U}_{n,P}}$ and n is large enough to guarantee that, on the complement of G_n^δ, $|P| \leq \delta \log n + 2$ and $\delta \log n > 4$. Since the path-connected set $\overline{\mathcal{U}_{n,P}}$ consists of $|P|$ overlapping closed disks of radius at most $(\log n)^{-1} n^{-1/2}$, we have

$$|x - y| \leq |P| \frac{2}{\log n \sqrt{n}} \leq \frac{2(\delta \log n + 2)}{\log n \sqrt{n}} < \frac{3\delta}{\sqrt{n}}.$$

□

Corollary 5.9. *Suppose $0 < \delta < 1/3$. There exists $C_\delta > 0$ such that for $n \geq C_\delta$, on the complement of G_n^δ, each $\Gamma \in \mathcal{M}_n$ satisfies the following. There exist $x^*, y^* \in \Gamma$ so that if $x, y \in \overline{\mathcal{O}_\Gamma}$, then $|x - y| \leq |x^* - y^*| < \frac{3\delta}{\sqrt{n}}$.*

Proof. In view of Lemma 5.8, it suffices to show that there exist $x^*, y^* \in \Gamma$ so that

$$\sup_{x, y \in \overline{\mathcal{O}_\Gamma}} |x - y| \leq |x^* - y^*|. \tag{59}$$

(Recall that there exists $P^* \in \mathcal{P}_n$ so that $\Gamma \subset \partial \overline{\mathcal{U}_{n,P^*}}$.) Since $\overline{\mathcal{O}_\Gamma}$ is compact and $(x, y) \mapsto |x - y|$ is continuous, the extreme value theorem guarantees the existence of $x^*, y^* \in \overline{\mathcal{O}_\Gamma}$ so that the supremum in (59) is achieved when $x = x^*$ and $y = y^*$. Suppose, for a contradiction, that $x^* \notin \Gamma$. Then, x^* is in the open set $\mathcal{O}_\Gamma$, and there is a $\rho > 0$ so that $x^* \in B(x^*, \rho) \subset \mathcal{O}_\Gamma$. Consequently, the line segment $\overline{x^* y^*}$ can be extended along the line connecting x^* and y^* by length $\rho/2$ without leaving $\overline{\mathcal{O}_\Gamma}$. This contradicts the assumption that the supremum in (59) is achieved for $x = x^*$, $y = y^*$. We conclude that $x^* \in \Gamma$. A similar argument shows that $y^* \in \Gamma$.

□

5.5. **Pairing of roots and critical points inside each domain.** We now show that on the complement of the "bad" events, the roots and critical points within most of the domains $\mathcal{O}_\Gamma$, $\Gamma \in \mathcal{M}_n$ are "paired." The only domains for which this does not occur are those that contain roots of $p_n(z)$ that are "too close" to the zeros of m_μ. (See Figure 5 for reference; recall that $m_\mu(z) = 0$ precisely when $z = 0$ in the case where μ is the uniform measure on the unit disk.) To make "too close" rigorous, we define the random collection of roots

$$\begin{aligned} R_n^{\text{pair}} &:= \left\{ X_j : 1 \le j \le n \text{ and } X_j \in \mathbb{C} \setminus \left(A_n^{\lfloor 4\log(\log n)\rfloor} \cup A_n^{\lfloor 4\log(\log n)\rfloor+1} \right) \right\} \\ &\subseteq \left\{ X_j : 1 \le j \le n \text{ and } |m_\mu(X_j)| > \frac{(\log n)^4}{\sqrt{n}} \right\}. \end{aligned}$$

The following lemma is the main result of this subsection.

Lemma 5.10. *For a fixed $\delta > 0$ chosen sufficiently small, there is a constant $C_\delta > 0$ so that for $n \ge C_\delta$, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta \cup H_n$, the following conclusion holds. For each $\mathcal{O}_\Gamma$, $\Gamma \in \mathcal{M}_n$, such that $\mathcal{O}_\Gamma \cap R_n^{pair} \neq \emptyset$, the number of critical points of $p_n(z)$ that lie inside $\mathcal{O}_\Gamma$ is equal to the number of roots of $p_n(z)$ that lie inside $\mathcal{O}_\Gamma$ (where both counts include multiplicity). Furthermore, if $X \in \mathcal{O}_\Gamma \cap R_n^{pair}$ and $w \in \mathcal{O}_\Gamma$ is a critical point of $p_n(z)$, then, $|X - w| \le \frac{(\log n)^4}{n|m_\mu(X)|}$.*

Proof. The proof of this lemma is similar in flavor to the proofs of Theorems 2.8 and 3.1, although the argument presented here is much more technical. Fix $n \in \mathbb{N}$, suppose $\mathcal{O}_\Gamma$, $\Gamma \in \mathcal{M}_n$ is such that $\mathcal{O}_\Gamma \cap R_n^{\text{pair}} \neq \emptyset$, and choose an $X \in \mathcal{O}_\Gamma \cap R_n^{\text{pair}}$ to be a distinguished root that will be a reference point in our calculations. We classify the roots $\{X_j\}_{j=1}^n$ into three groups based on their proximity to X (see Figure 7). To that end, define

$$\begin{aligned} R_{\text{near}} &:= \left\{ j : 1 \le j \le n,\ |X_j - X| < \frac{(\log n)^2}{n\,|m_\mu(X)|} \right\} \\ R_{\text{med}} &:= \left\{ j : 1 \le j \le n,\ |X_j - X| < \frac{1}{\sqrt{n}} \right\} \setminus R_{\text{near}} \\ R_{\text{far}} &:= \left\{ j : 1 \le j \le n,\ |X_j - X| \ge \frac{1}{\sqrt{n}} \right\}, \end{aligned}$$

and let

$$q_X(z) := \prod_{j \notin R_{\text{near}}} (z - X_j) \quad \text{and} \quad r_X(z) := \prod_{j \in R_{\text{near}}} (z - X_j),$$

so that $p_n(z) = q_X(z) r_X(z)$. Note that $|R_{\text{med}}|$ and $|R_{\text{near}}|$ are of size at most $\delta \log n + 2$ on the complement of G_n^δ. We will compare the zeros of $p_n'(z)$ inside $\mathcal{O}_\Gamma$ to the zeros of the function

$$f_X(z) := q_X(z) \left(r_X'(z) + r_X(z) \frac{n - |R_{\text{near}}|}{z - Y_X} \right)$$

that are inside $\mathcal{O}_\Gamma$, where Y_X is defined by

$$Y_X := X - \frac{n - |R_{\text{near}}|}{\sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j}}.$$

The idea is that

$$\frac{f_X(z)}{p_n(z)} = \frac{r_X'(z)}{r_X(z)} + \frac{n - |R_{\text{near}}|}{z - Y_X}$$

is similar to the logarithmic derivative of $p_n(z)$ for z near X. Furthermore, the number of roots of the equation

$$0 = r'_X(z) + r_X(z)\frac{n - |R_{\text{near}}|}{z - Y_X}$$

that are inside $\mathcal{O}_\Gamma$ will be easy to calculate since these are the same as the critical points of

$$\widetilde{p}_X(z) := r_X(z) \cdot (z - Y_X)^{n-|R_{\text{near}}|}$$

that lie inside $\mathcal{O}_\Gamma$ (we will show that $Y_X \notin \mathcal{O}_\Gamma$), and these can be located with Walsh's two circle theorem.

The following lemma contains a few facts that we will frequently reference for the remainder of the proof of Lemma 5.10.

Lemma 5.11. *Suppose $\delta < 1/3$. There is a constant $K_{\mu,\delta} \in \mathbb{N}$, depending only on μ and δ (and not on X, P, Γ, etc...), so that $n \geq K_{\mu,\delta}$ implies the following. On the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, if $X \in \mathcal{O}_\Gamma \cap R_n^{pair}$ and $z \in \overline{\mathcal{O}_\Gamma}$, then*

(i) $|z - X| \leq \dfrac{4\delta(\log n)^4}{n\,|m_\mu(X)|}$, *and* $|z - X| \geq \frac{(\log n)^3}{n|m_\mu(X)|}$ *if* $z \in \Gamma$;

(ii) $\dfrac{|m_\mu(X)|}{4} \leq \left| \dfrac{1}{n - |R_{near}|} \displaystyle\sum_{j \notin R_{near}} \dfrac{1}{X - X_j} \right| \leq 2\,|m_\mu(X)|$;

(iii) $\dfrac{1}{4\,|m_\mu(X)|} \leq |z - Y_X| \leq \dfrac{5}{|m_\mu(X)|}$, *so in particular, $f_X(z)$ is analytic in $\mathcal{O}_\Gamma$.*

Proof. Much of this proof relies on the fact that $m_\mu(\cdot)$ is nearly Lipschitz (see Lemma 3.5 part (ii)). To establish (i), we first observe that for large n, on the complement of G_n^δ, if $\xi \in \overline{\mathcal{O}_\Gamma}$, then

$$\frac{|m_\mu(X)|}{2} \leq |m_\mu(\xi)| \leq \frac{3\,|m_\mu(X)|}{2}. \tag{60}$$

Indeed, via Corollary 5.9, $|\xi - X| < \frac{3\delta}{\sqrt{n}} < \frac{1}{\sqrt{n}}$ for large n, on the complement of G_n^δ, so as long as we also have $\frac{1}{\sqrt{n}} < \min\left\{\varepsilon_\mu, e^{-1}\right\}$, Lemma 3.5 guarantees that

$$|m_\mu(\xi) - m_\mu(X)| \leq \kappa_\mu \frac{3\delta}{\sqrt{n}} \log\left(\frac{\sqrt{n}}{3\delta}\right).$$

(We have used the fact that on the interval $[0, e^{-1}]$, the function $-x \log x$ is increasing.) It follows that for $n \geq 5$ and larger than some constant depending on μ and δ, on the complement of G_n^δ,

$$|m_\mu(\xi) - m_\mu(X)| \leq \frac{(\log n)^2}{\sqrt{n}} \leq \frac{(\log n)^4}{2\sqrt{n}} \leq \frac{|m_\mu(X)|}{2},$$

which implies equation (60). (The last inequality follows since $X \in R_n^{\text{pair}}$.) We will use this inequality to compute $|z - X|$, for $z \in \overline{\mathcal{O}_\Gamma}$, in a way that references the balls that we started with when we constructed Γ.

Let n be large enough to establish (60) and the conclusion of Corollary 5.9 on the complement of G_n^δ. Since, $z, X \in \overline{\mathcal{O}_\Gamma}$, Corollary 5.9 guarantees the existence

of $w_1, w_2 \in \Gamma$ for which $|z - X| \leq |w_1 - w_2|$. Recall that $\Gamma \subseteq \partial \mathcal{U}_{n,P^*}$ for some $P^* \in \mathcal{P}_n$, so there are $i_1, i_2 \in P^*$ for which, $X_{i_1}, X_{i_2} \in \mathcal{O}_\Gamma$, and

$$w_1 \in \partial B \left(X_{i_1}, \frac{(\log n)^3}{n \max\left\{ |m_\mu(X_{i_1})|, \frac{(\log n)^4}{\sqrt{n}} \right\}} \right)$$

and

$$w_2 \in \partial B \left(X_{i_2}, \frac{(\log n)^3}{n \max\left\{ |m_\mu(X_{i_2})|, \frac{(\log n)^4}{\sqrt{n}} \right\}} \right).$$

Furthermore, since i_1 and i_2 are related by the equivalence that defines $\mathcal{P}_n$, there are open balls $B_0, B_1, \dots B_l \in \mathcal{C}_n$, of the form

$$B \left(X_j, \frac{(\log n)^3}{n \cdot \max\left\{ |m_\mu(X_j)|, \frac{(\log n)^4}{\sqrt{n}} \right\}} \right), \ j \in P^*, \ X_j \in \mathcal{O}_\Gamma,$$

where

$$B_0 = B \left(X_{i_1}, \frac{(\log n)^3}{n \cdot \max\left\{ |m_\mu(X_{i_1})|, \frac{(\log n)^4}{\sqrt{n}} \right\}} \right),$$
$$B_l = B \left(X_{i_2}, \frac{(\log n)^3}{n \cdot \max\left\{ |m_\mu(X_{i_2})|, \frac{(\log n)^4}{\sqrt{n}} \right\}} \right),$$

and $B_k \cap B_{k+1} \neq \emptyset$ for $0 \leq k \leq l-1$. Notice that on the complement of G_n^δ, equation (60) guarantees that the radii of these balls are bounded by $\frac{2(\log n)^3}{n|m_\mu(X)|}$ (recall that $X \in R_n^{\text{pair}}$), and if n is large enough to guarantee the conclusion of Lemma 5.8, the number of balls, l, is less than $|P^*| \leq \delta \log n + 2$. It follows that for n larger than a constant depending on δ, on the complement of G_n^δ,

$$|z - X| \leq |w_1 - w_2| \leq |P^*| \cdot 2 \frac{2(\log n)^3}{n\,|m_\mu(X)|} \leq \frac{4(\delta \log n + 2)(\log n)^3}{n\,|m_\mu(X)|} \leq \frac{4\delta(\log n)^4}{n\,|m_\mu(X)|}.$$

We have established the first half of (i). To see the second inequality, simply recall that Γ does not pass through $\mathcal{U}_{n,P}$ for any $P \in \mathcal{P}_n$, so if $z \in \Gamma$, then

$$|z - X_j| \geq \frac{(\log n)^3}{n \cdot \max\left\{ |m_\mu(X_j)|, \frac{(\log n)^4}{\sqrt{n}} \right\}}$$

for any root X_j, $1 \leq j \leq n$. In particular, this is true for $X \in R_n^{\text{pair}}$, which satisfies $|m_\mu(X)| \geq \frac{(\log n)^4}{\sqrt{n}}$, so we obtain the second part of (i).

Inequality (ii) holds for large n on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$ after several interpolations. For each i, $1 \leq i \leq n$, the random variables $\mathbb{E}[\zeta_{i,j}^{(n)} \mid X_i]$, $1 \leq j \leq n$,

$j \neq i$ are identically distributed, so

$$\left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\zeta_{i,j}^{(n)}\right| \leq \left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} \mid X_i]\right)\right| + \left|\mathbb{E}[\zeta_{i,l}^{(n)} \mid X_i] - m_\mu(X_i)\right| + |m_\mu(X_i)|\,, \tag{61}$$

where l is any index different from i. Since the X_j are iid, we have

$$\begin{aligned}\left|\mathbb{E}[\zeta_{i,l}^{(n)} \mid X_i] - m_\mu(X_i)\right| &= \left|\mathbb{E}\left[\frac{1}{X_i - X_l}\mathbb{1}_{|X_i - X_l| < \frac{(\log n)^2}{n|m_\mu(X_i)|}} \;\middle|\; X_i\right]\right| \\ &\leq 2\pi C_\mu \int_0^{\frac{(\log n)^2}{n|m_\mu(X_i)|}} \frac{1}{r}\cdot r\, dr \\ &= 2\pi C_\mu \frac{(\log n)^2}{n\,|m_\mu(X_i)|},\end{aligned}$$

so equation (61) implies that for any i, $1 \leq i \leq n$,

$$\left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\zeta_{i,j}^{(n)}\right| \leq \left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} \mid X_i]\right)\right| + \frac{2\pi C_\mu (\log n)^2}{n\,|m_\mu(X_i)|} + |m_\mu(X_i)|\,.$$

Now, $X = X_{i_X}$ for some i_X, $1 \leq i_X \leq n$, and $X \in R_n^{\text{pair}}$, so on the complement of $\cup_{i=1}^n F_n^i$,

$$\begin{aligned}\left|\frac{1}{n-|R_{\text{near}}|}\sum_{j\notin R_{\text{near}}}\frac{1}{X - X_j}\right| &= \frac{n-1}{n-|R_{\text{near}}|}\left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i_X}}^{n}\zeta_{i_X,j}^{(n)}\right| \\ &\leq \frac{n-1}{n-|R_{\text{near}}|}\left(\frac{3}{2}\,|m_\mu(X_{i_X})| + \frac{2\pi C_\mu(\log n)^2}{n\,|m_\mu(X_{i_X})|}\right) \\ &\leq \frac{n-1}{n-|R_{\text{near}}|}\left(\frac{3}{2}\,|m_\mu(X)| + \frac{2\pi C_\mu}{\sqrt{n}(\log n)^2}\right).\end{aligned} \tag{62}$$

On the complement of G_n^δ, $|R_{\text{near}}|$ is at most $\delta \log n + 2$, so for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$ inequality (62) establishes the upper bound in (ii). (We have used that $X \in R_n^{\text{pair}}$ to bound $\frac{2\pi C_\mu}{\sqrt{n}(\log n)^2}$ above by, say, $1/4\,|m_\mu(X)|$ for large n.) The lower bound in (ii) is achieved similarly by using the reverse triangle inequality to obtain

$$\left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\zeta_{i,j}^{(n)}\right| \geq |m_\mu(X_i)| - \left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} \mid X_i]\right)\right| - \left|\mathbb{E}[\zeta_{i,l}^{(n)} \mid X_i] - m_\mu(X_i)\right|,$$

in place of (61).

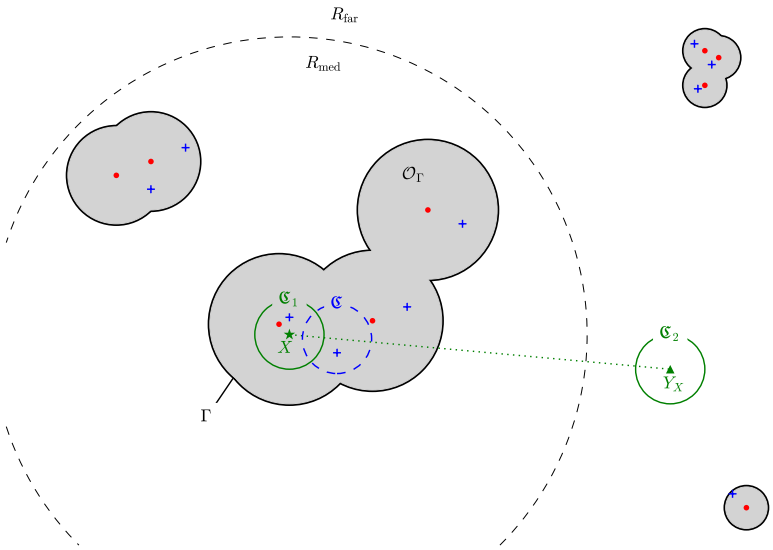

FIGURE 7. A diagram to illustrate Lemma 5.12 and its proof. The red dots and blue crosses are meant to represent roots and critical points, respectively, of p_n that lie in a region near X, which is denoted by a green star. The large dashed circle is intended to be on the order of $n^{-1/2}$. Note that indices $1 \leq j \leq n$ in R_{near} correspond to roots X_j that lie interior to $\mathfrak{C}_1$. This figure is neither to scale nor the result of a simulation.

We conclude by establishing (iii) as a consequence of (i) and (ii). Indeed, via the triangle inequality, we have for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, that

$$|z - Y_X| \leq |z - X| + \left| \frac{n - |R_{\text{near}}|}{\sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j}} \right| \leq \frac{4\delta (\log n)^4}{n\,|m_\mu(X)|} + \frac{4}{|m_\mu(X)|} \leq \frac{5}{|m_\mu(X)|},$$

where the rightmost inequality holds for large n. The lower bound in (iii) follows for similar reasons, and f_X is analytic because $|m_\mu(X)|$ is almost surely bounded above by an constant that depends only on μ (apply Lemma 3.5, part (i) with $\xi = 0$ and $\rho = +\infty$). □

The next Lemma justifies our choice of $f_X(z)$ as an intermediate comparison between $p_n(z)$ and $p_n'(z)$ because it establishes that under the right conditions, $f_X(z)$ and $p_n(z)$ have the same number of roots in the domain $\mathcal{O}_\Gamma$. Consider Figure 7 which provides a visual aid to the argument.

Lemma 5.12. *Suppose $\delta < 1/3$. For large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, the polynomial $\widetilde{p}_X(z) = r_X(z)(z - Y_X)^{n - |R_{near}|}$ has $|R_{near}|$ critical points inside $B\left(X, \frac{5(\log n)^2}{n|m_\mu(X)|}\right) \subset \mathcal{O}_\Gamma$, and none of these is $Y_X \notin \mathcal{O}_\Gamma$. In particular, under these conditions, $f_X(z)$ has the same number of roots inside $\mathcal{O}_\Gamma$ as $p_n(z)$ does.*

Proof. This follows from Walsh's two circle theorem (see e.g. Theorem 4.1.1 in [27].) First, we will show that $r_X(z)$ and $\widetilde{p}'_X(z)$ have the same number of roots, $|R_{\text{near}}|$, inside $\mathcal{O}_\Gamma$ by using Walsh's two circle theorem, and then, we will use this fact to compare the roots of $p_n(z)$ and $f_X(z)$ inside $\mathcal{O}_\Gamma$.

To that end, choose n large enough so that the statements in Lemma 5.11 hold on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, and define the circular domains

$$\mathfrak{C}_1 := B\left(X, \frac{(\log n)^2}{n\,|m_\mu(X)|}\right) \quad \text{and} \quad \mathfrak{C}_2 := B\left(Y_X, \frac{(\log n)^2}{n\,|m_\mu(X)|}\right).$$

Note that $\mathfrak{C}_1$ and $\mathfrak{C}_2$ are disjoint for large n on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$ by inequality (iii) of Lemma 5.11:

$$|X - Y_X| \geq \frac{1}{4\,|m_\mu(X)|} > \frac{(\log n)^2}{n\,|m_\mu(X)|}.$$

In fact, for n large enough,

$$\frac{1}{4\,|m_\mu(X)|} > \frac{4\delta(\log n)^4}{n\,|m_\mu(X)|} + \frac{(\log n)^2}{n\,|m_\mu(X)|},$$

so on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, Lemma 5.11 part (i) guarantees that $\mathfrak{C}_2$ is disjoint from $\mathcal{O}_\Gamma$.

Next, observe that all of the roots of $\widetilde{p}_X(z)$ lie in $\mathfrak{C}_1 \cup \mathfrak{C}_2$, so by Walsh's two circle theorem, the critical points of $\widetilde{p}_X$ lie in $\mathfrak{C}_1 \cup \mathfrak{C}_2 \cup \mathfrak{C}$, where $\mathfrak{C}$ is the open ball

$$\mathfrak{C} := B\left(\frac{n - |R_{\text{near}}|}{n} X + \frac{|R_{\text{near}}|}{n} Y_X, \frac{(\log n)^2}{n\,|m_\mu(X)|}\right).$$

By Lemma 5.11, for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, $x \in \mathfrak{C}$ implies

$$\begin{aligned}
|x - X| &\leq \left|\frac{n - |R_{\text{near}}|}{n} X + \frac{|R_{\text{near}}|}{n} Y_X - X\right| + \frac{(\log n)^2}{n\,|m_\mu(X)|} \\
&= \frac{|R_{\text{near}}|}{n} \left|\frac{n - |R_{\text{near}}|}{\sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j}}\right| + \frac{(\log n)^2}{n\,|m_\mu(X)|} \\
&\leq \frac{|R_{\text{near}}|}{n} \frac{4}{|m_\mu(X)|} + \frac{(\log n)^2}{n\,|m_\mu(X)|} \\
&\leq \frac{4(\delta \log n + 2) + (\log n)^2}{n\,|m_\mu(X)|} \\
&< \frac{5(\log n)^2}{n\,|m_\mu(X)|},
\end{aligned}$$

where the last inequality holds for large n. It follows that for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$,

$$\mathfrak{C} \subseteq B\left(X, \frac{5(\log n)^2}{n\,|m_\mu(X)|}\right) \subseteq B\left(X, \frac{(\log n)^3}{n \cdot \max\left\{|m_\mu(X)|\,, \frac{(\log n)^4}{\sqrt{n}}\right\}}\right)$$

(recall $X \in R_n^{\text{pair}}$), so in particular, $\mathfrak{C} \cup \mathfrak{C}_1$ is contained in $\mathcal{O}_\Gamma$, and this union is disjoint from $\mathfrak{C}_2$. Consequently, by the Supplement Theorem 4.1.1 in [27], for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, $\widetilde{p}'_X(z)$ has $|R_{\text{near}}|$ roots inside $\mathcal{O}_\Gamma$, just like $r_X(z)$ does. Under these conditions, $f_X(z)$ has the same roots as $q_X(z)\widetilde{p}'_X(z)$ inside

$\mathcal{O}_\Gamma$ because $Y_X \notin \mathcal{O}_\Gamma$, so it follows that $f_X(z)$ and $p_n(z) = q_X(z)r_X(z)$ have the same number of roots inside $\mathcal{O}_\Gamma$. □

We conclude this subsection with two lemmas and an application of Rouché's theorem to establish that $f_X(z)$ and $p_n'(z)$ have the same numbers of zeros in $\mathcal{O}_\Gamma$. This will imply via Lemma 5.12 that $p_n(z)$ and $p_n'(z)$ have the same numbers of zeros in $\mathcal{O}_\Gamma$.

Lemma 5.13. *Suppose $\delta < 1/8$. There exist positive constants $\widetilde{C}_\mu$, dependent only on μ, and $C_{\mu,\delta}$, dependent only on μ and δ (and not on X, Γ, etc...), so that for $n \geq C_{\mu,\delta}$, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta \cup H_n$, if $z \in \Gamma$,*

$$|p_n'(z) - f_X(z)| \leq |p_n(z)|\, \widetilde{C}_\mu \delta^2 n\, |m_\mu(X)| \tag{63}$$

(here, $\widetilde{C}_\mu$ is independent of δ).

Proof. Fix $z \in \Gamma$. By the definition of $f_X(z)$ and the triangle inequality, we have

$$\begin{aligned}
&|p_n'(z) - f_X(z)| \\
&\quad = |p_n(z)| \cdot \left| \sum_{j=1}^n \frac{1}{z - X_j} - \frac{r_X'(z)}{r_X(z)} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \\
&\quad = |p_n(z)| \cdot \left| \sum_{j \notin R_{\text{near}}} \frac{1}{z - X_j} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \\
&\quad \leq |p_n(z)| \left(\left| \sum_{j \notin R_{\text{near}}} \frac{1}{z - X_j} - \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right| + \left| \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \right).
\end{aligned} \tag{64}$$

We find upper bounds for the two terms at right separately. First, factor $\left|\sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j}\right|$ from the second term, and combine the resulting fractions to obtain

$$\begin{aligned}
&\left| \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \\
&\quad = \left| \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right| \left| \frac{(z - X) \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j}}{(z - X) \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} + n - |R_{\text{near}}|} \right| \\
&\quad = |z - X| \left| \frac{1}{n - |R_{\text{near}}|} \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right|^2 \left| \frac{n - |R_{\text{near}}|}{(z - X) \frac{1}{n - |R_{\text{near}}|} \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} + 1} \right|.
\end{aligned}$$

Lemma 5.11 and an application of the reverse triangle inequality to the denominator of the rightmost factor yield for large n, on the complement of $\cup_{i=1}^{n} F_n^i \cup G_n^\delta$,

$$\begin{aligned}
&\left| \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \\
&\quad \leq \frac{4\delta (\log n)^4}{n \left| m_\mu(X) \right|} \cdot 4 \left| m_\mu(X) \right|^2 \cdot \frac{n - |R_{\text{near}}|}{1 - |z - X| \left| \frac{1}{n - |R_{\text{near}}|} \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right|} \\
&\quad = O_\delta \left((\log n)^4 \left| m_\mu(X) \right| \right).
\end{aligned} \tag{65}$$

We now find a bound on the first term in (64). Combining the summands gives

$$\begin{aligned}
&\left| \sum_{j \notin R_{\text{near}}} \frac{1}{z - X_j} - \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right| \\
&\quad = |z - X| \left| \sum_{j \notin R_{\text{near}}} \frac{1}{(z - X_j)(X - X_j)} \right| \\
&\quad \leq |z - X| \left(\left| \sum_{j \in R_{\text{med}}} \frac{1}{(z - X_j)(X - X_j)} \right| + \left| \sum_{j \in R_{\text{far}}} \frac{1}{(z - X_j)(X - X_j)} \right| \right),
\end{aligned}$$

so in view of (65) and the upper bound Lemma 5.11 gives for $|z - X|$, we can establish (63) by showing there exist positive constants $\widetilde{C}'_\mu$, $C'_{\mu,\delta}$ satisfying the following: on the complement of $\cup_{i=1}^{n} F_n^i \cup G_n^\delta$, $n \geq C'_{\mu,\delta}$ implies

$$\left| \sum_{j \in R_{\text{med}}} \frac{1}{(z - X_j)(X - X_j)} \right| + \left| \sum_{j \in R_{\text{far}}} \frac{1}{(z - X_j)(X - X_j)} \right| \leq \frac{\widetilde{C}'_\mu \delta n^2 \left| m_\mu(X) \right|^2}{(\log(n))^4}. \tag{66}$$

We will bound each term on the left separately. By construction of the sets $\{\mathcal{O}_\Gamma\}_{\Gamma \in \mathcal{M}_n}$, recall that the curves $\Gamma \in \mathcal{M}_n$ do not intersect the interiors of the open balls forming $\mathcal{U}_{n,P}$, $P \in \mathcal{P}_n$. Hence, for $j \in R_{\text{med}}$,

$$|z - X_j| \geq \frac{(\log n)^3}{n \cdot \max \left\{ \left| m_\mu(X_j) \right|, \frac{(\log n)^4}{\sqrt{n}} \right\}}.$$

By Lemma 3.5, it follows that for large n, $|m_\mu(X_j)| \leq 2 |m_\mu(X)|$ (Recall that for $j \in R_{\text{med}}$, $|X - X_j| < \frac{1}{\sqrt{n}}$ and $X \in R_n^{\text{pair}}$). Consequently, for large n, on the complement of $\cup_{i=1}^{n} F_n^i \cup G_n^\delta$,

$$|z - X_j| \geq \frac{(\log n)^3}{n \cdot \max \left\{ 2 \left| m_\mu(X) \right|, \frac{(\log n)^4}{\sqrt{n}} \right\}} \geq \frac{(\log n)^3}{2n \left| m_\mu(X) \right|}.$$

In addition, for $j \in R_{\text{med}}$, $|X - X_j| \geq \frac{(\log n)^2}{n |m_\mu(X)|}$. Hence, for n large, on the complement of $\cup_{i=1}^{n} F_n^i \cup G_n^\delta$,

$$\left| \sum_{j \in R_{\text{med}}} \frac{1}{(z - X_j)(X - X_j)} \right| \leq |R_{\text{med}}| \frac{2n^2 \left| m_\mu(X) \right|^2}{(\log n)^5} \leq (\delta \log n + 2) \frac{2n^2 \left| m_\mu(X) \right|^2}{(\log n)^5}. \tag{67}$$

We now turn our attention to the second term on the left side of (66). Since μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{C}$, we expect that the number of X_j, $j \in R_{\text{far}}$ within a given distance of X is roughly proportional to the square of that distance, and hence, the sum over $j \in R_{\text{far}}$ in (66) should be roughly on the order of n. To take advantage of this intuition, we will split the sum into pieces by grouping terms $[(z - X_j)(X - X_j)]^{-1}$ according to the distance between X_j and X. To that end, define for $1 \leq k \leq \sqrt{n} - 1$, the annuli

$$D_{k,n} := \left\{ w \in \mathbb{C} : \frac{k}{\sqrt{n}} \leq |w - X| \leq \frac{k+1}{\sqrt{n}} \right\}$$
$$D'_{k,n} := \left\{ w \in \mathbb{C} : \frac{k-1}{\sqrt{n}} \leq |w - X| \leq \frac{k+2}{\sqrt{n}} \right\}$$
$$D''_{k,n} := \left\{ w \in \mathbb{C} : \frac{k-2}{\sqrt{n}} \leq |w - X| \leq \frac{k+3}{\sqrt{n}} \right\}$$

and the random variables

$$\#_{k,n} := \#\left\{ j : 1 \leq j \leq n,\ X_j \in D_{k,n} \right\}.$$

(Note that $D'_{1,n}, D''_{1,n}, D''_{2,n}$ are disks.) On the complement of H_n, each X_j, $1 \leq j \leq n$ is within $n^{-1/2}$ of some $x_j \in \mathcal{N}_n$, and on the complement of G_n^δ, there are at most $2 + \delta \log n$ roots X_l, $1 \leq l \leq n$ within $n^{-1/2}$ of x_j. It follows that

$$\#_{k,n} \leq \left| \mathcal{N}_n \cap D'_{k,n} \right| \cdot (\delta \log n + 2). \tag{68}$$

We will argue that due to the fact that any distinct $x, y \in \mathcal{N}_n$ are separated by at least $\frac{1}{2\sqrt{n}}$, the size of $\mathcal{N}_n \cap D'_{k,n}$ is bounded by $16^2 k$. Indeed, for any distinct $x, y \in \mathcal{N}_n$, the balls $B(x, n^{-1/2}/4)$ and $B(y, n^{-1/2}/4)$ are disjoint, and if $x \in D'_{k,n}$, then $B(x, n^{-1/2}/4) \subset D''_{k,n}$. The area of $D''_{k,n}$ for $k \geq 2$ is $\frac{\pi}{n}(10k + 5)$, so at most $16(10k + 5)$ disjoint balls of radius $n^{-1/2}/4$ can fit in $D''_{k,n}$. Similarly, at most 16^2 balls of radius $n^{-1/2}/4$ can fit in $D''_{1,n}$. Combining this with equation (68) establishes that, on the complement of $G_n^\delta \cup H_n$, we have $\#_{k,n} \leq 16^2 k(\delta \log n + 2)$.

We can now find an upper bound for the second term on the left of (66) by breaking this sum into pieces that correspond to the annuli $D_{k,n}$, $1 \leq k \leq \sqrt{n} - 1$. To start, observe that

$$\begin{aligned}
&\sum_{j \in R_{\text{far}}} \frac{1}{|z - X_j|\,|X - X_j|} \\
&\leq \sum_{k=1}^{\sqrt{n}-1} \sum_{j : X_j \in D_{k,n}} \frac{1}{|z - X_j|\,|X - X_j|} + \sum_{j : |X - X_j| \geq 1} \frac{1}{|z - X_j|} \\
&\leq \sum_{k=1}^{\sqrt{n}-1} \sum_{j : X_j \in D_{k,n}} \frac{1}{(|X - X_j| - |z - X|)\,|X - X_j|} + \sum_{j : |X - X_j| \geq 1} \frac{1}{|X - X_j| - |z - X|},
\end{aligned} \tag{69}$$

where the last line follows from the reverse triangle inequality (the next equation justifies why $|z - X|$ is smaller than $|X - X_j|$). If $\delta < 1/8$ and n is large enough to guarantee the conclusions of Lemma 5.11, then on the complement of $\cup_{i=1}^n F_n^i \cup$

$G_n^\delta \cup H_n$, for $j \in R_{\text{far}}$, we have

$$|X - X_j| \geq \frac{1}{\sqrt{n}} > \frac{8\delta}{\sqrt{n}} \geq \frac{8\delta(\log n)^4}{n\,|m_\mu(X)|} \geq 2\,|z - X|$$

(note that $|m_\mu(X)| \geq \frac{(\log n)^4}{\sqrt{n}}$). Substituting $|X - X_j|/2$ for $|z - X|$ into the last line of (69) establishes that for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta \cup H_n$,

$$\sum_{j \in R_{\text{far}}} \frac{1}{|z - X_j|\,|X - X_j|} \leq \sum_{k=1}^{\sqrt{n}-1} \sum_{j : X_j \in D_{k,n}} \frac{2}{|X - X_j|^2} + \sum_{j : |X - X_j| \geq 1} \frac{2}{|X - X_j|},$$

at which point we can use the fact that $|X_j - X| \geq \frac{k}{\sqrt{n}}$ for $X_j \in D_{k,n}$, $1 \leq k \leq \sqrt{n} - 1$ to obtain

$$\sum_{j \in R_{\text{far}}} \frac{1}{|z - X_j|\,|X - X_j|} \leq \sum_{k=1}^{\sqrt{n}-1} \#_{k,n} \frac{2n}{k^2} + 2n \leq \sum_{k=1}^{\sqrt{n}-1} \frac{2\left(16^2 k(\delta \log n + 2)\right) n}{k^2} + 2n.$$

By approximating $\sum_{k-1}^{\sqrt{n}-1} k^{-1}$ with $1 + \int_1^{\sqrt{n}-1} x^{-1}\,dx$, we see that this last expression is on the order of $O\left(\delta n (\log n)^2\right)$, where the implied constant is independent of δ. Together with (67), this establishes equation (66) since $|m_\mu(X)| \geq \frac{(\log n)^4}{\sqrt{n}}$. □

The last lemma in this subsection establishes a lower bound on $|f_X(z)|$ that will combine with (63) to fulfill the hypotheses of Rouché's theorem on the boundary Γ of the domain $\mathcal{O}_\Gamma$.

Lemma 5.14. *For fixed $\delta > 0$, there is a constant $\check{C}_{\mu,\delta}$ depending only on μ and δ so that when $n \geq \check{C}_{\mu,\delta}$, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, if $z \in \Gamma$,*

$$|f_X(z)| \geq |p_n(z)|\, n\, |m_\mu(X)| \cdot e^{-9}. \tag{70}$$

Proof. We have

$$|r_X(z)| = \prod_{X_j \in R_{\text{near}}} |z - X_j| \leq \left(|z - X| + \frac{(\log n)^2}{n\,|m_\mu(X)|}\right)^{|R_{\text{near}}|}$$

and

$$\left|\frac{f_X(z)}{q_X(z)}\right| = \frac{1}{|z - Y_X|}\left|(z - Y_X) r_X'(z) + r_X(z)\,(n - |R_{\text{near}}|)\right|.$$

By Lemma 5.12, for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, the polynomial expression

$$(z - Y_X) r_X'(z) + r_X(z)\,(n - |R_{\text{near}}|) = \frac{\widetilde{p}_X'(z)}{(z - Y_X)^{n - |R_{\text{near}}| - 1}}$$

has degree $|R_{\text{near}}|$, leading coefficient n, and $|R_{\text{near}}|$ roots in $B\left(X, \frac{5(\log n)^2}{n|m_\mu(X)|}\right) \subset \mathcal{O}_\Gamma$. It follows that under these conditions,

$$\begin{aligned}\left|\frac{f_X(z)}{q_X(z)}\right| &= \frac{n}{|z-Y_X|} \prod_{\substack{w\in\mathcal{O}_\Gamma \\ \widetilde{p}'_X(w)=0}} |z-w| \\ &\geq \frac{n}{|z-Y_X|}\left(|z-X| - \frac{5(\log n)^2}{n\,|m_\mu(X)|}\right)^{|R_{\text{near}}|},\end{aligned}$$

where the critical points of $\widetilde{p}_X(z)$ that index the product are considered with multiplicity.

If additionally, $\delta < 1$ and n is large enough to guarantee the bounds on $|z-X|$ in Lemma 5.11, we have that on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$ and for $z \in \Gamma$,

$$\frac{(\log n)^2}{n\,|m_\mu(X)|} \leq \frac{|z-X|}{\delta \log n}.$$

Hence, if n is large enough, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, for $z \in \Gamma$,

$$\begin{aligned}|f_X(z)| &= |p_n(z)| \cdot \frac{1}{|r_X(z)|} \cdot \left|\frac{f_X(z)}{q_X(z)}\right| \\ &\geq |p_n(z)| \cdot \frac{n}{|z-Y_X|} \cdot \left(\frac{|z-X|\left(1-\frac{5}{\delta\log n}\right)}{|z-X|\left(1+\frac{1}{\delta\log n}\right)}\right)^{|R_{\text{near}}|} \\ &\geq |p_n(z)| \cdot \frac{n\,|m_\mu(X)|}{5} \cdot \left(\frac{1-\frac{5}{\delta\log n}}{1+\frac{1}{\delta\log n}}\right)^{\delta\log n+2} \\ &\geq |p_n(z)|\, n\,|m_\mu(X)| \cdot e^{-9}.\end{aligned}$$

We have used Lemma 5.11 to bound $|z-Y_X|$, and the last inequality holds for large n and comes from the fact that

$$\left(1+\frac{x}{\delta\log n}\right)^{\delta\log n+2} \xrightarrow{n\to\infty} e^x.$$

(Note that the rate of convergence possibly depends on δ.) We have achieved (70) as was desired. □

We have now established both (63) and (70), where the inequalities are independent of X, Γ, and $z \in \Gamma$. Since $\widetilde{C}_\mu$ is independent of δ, we can choose $\delta \in (0, 1/8)$ small enough that $\widetilde{C}_\mu \delta^2 < e^{-9}$. For such a δ, by Lemmas 5.13 and 5.14, for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta \cup H_n$, any $z \in \Gamma$ satisfies

$$|p'_n(z) - f_X(z)| < |f_X(z)|.$$

It follows by Rouché's theorem that for large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta \cup H_n$, $p'_n(z)$ and $f_X(z)$ have the same number of zeros inside $\mathcal{O}_\Gamma$, and by Lemma 5.12, we conclude that $p'_n(z)$ and $p_n(z)$ have the same number of zeros inside $\mathcal{O}_\Gamma$. The inequality in the conclusion of Lemma 5.10 follows directly from this and Lemma 5.11 part (i) (note $\delta \leq 1/4$).

In the argument above, the particular curve $\Gamma \in \mathcal{M}_n$ and the root $X \in \mathcal{O}_\Gamma \cap R_n^{\text{pair}}$ were arbitrary, and all of the constants involved were independent of Γ, so we have proved Lemma 5.10. □

5.6. **Bounding the Wasserstein distance.** In this subsection, we use Lemma 5.10 to prove Theorem 2.3. Let $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ denote the (not necessarily distinct) critical points of $p_n(z)$, and recall the definitions of the empirical measures, μ_n and μ_n' (see (3)) Since the numbers of roots and critical points of a polynomial differ by one, we first compare the measure μ_n' to the intermediate measure

$$\tilde{\mu}_n' := \frac{1}{n}\left(\delta_{\overline{X}} + \sum_{j=1}^{n-1} \delta_{w_j^{(n)}}\right), \text{ where } \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

The following lemma justifies our choice of $\tilde{\mu}_n'$.

Lemma 5.15. *Let μ_n', $\tilde{\mu}_n'$, and $\eta_n := \max_{1\le j\le n} |X_j|$ be defined as above. Then, with probability 1, $W_1(\mu_n', \tilde{\mu}_n') \le \frac{2\eta_n}{n}$.*

Proof. Let $\widetilde{\pi}$ be the measure on $\mathbb{C}\times\mathbb{C}$ given by

$$\widetilde{\pi} := \frac{1}{n}\sum_{j=1}^{n-1} \delta_{(w_j^{(n)}, w_j^{(n)})} + \frac{1}{n(n-1)}\sum_{j=1}^{n-1} \delta_{(w_j^{(n)}, \overline{X})},$$

whose marginal distributions are easily seen to be μ_n' and $\tilde{\mu}_n'$. It follows from the definition of the L_1-Wasserstein metric that, almost surely,

$$W_1(\mu_n', \tilde{\mu}_n') \le \frac{1}{n}\sum_{j=1}^{n-1} \left|w_j^{(n)} - w_j^{(n)}\right| + \frac{1}{n(n-1)}\sum_{j=1}^{n-1} \left|w_j^{(n)} - \overline{X}\right| \le 0 + \frac{1}{n}\cdot 2\eta_n,$$

where the last inequality follows from the Gauss–Lucas theorem. □

The next result is an L_1-Wasserstein comparison between μ_n and $\tilde{\mu}_n'$ that we will use in conjunction with Lemma 5.15 and the triangle inequality to prove Theorem 2.3.

Lemma 5.16. *Let $X_1, \dots, X_n$ be iid, complex random variables with distribution μ that has a bounded density and satisfies Assumption 2.1. Then, there is a constant C, depending only on μ, so that with probability $1 - o(1)$, $W_1(\mu_n, \tilde{\mu}_n') \le \frac{C\eta_n (\log n)^9}{n}$, where μ_n, $\tilde{\mu}_n'$, and η_n are defined as above.*

Proof. Suppose $w_1^{(n)}, \dots, w_{n-1}^{(n)}$ are critical points of $p_n(z)$ defined above, and define $w_n^{(n)} := \overline{X}$. Then, for any permutation σ_n of $\{1, 2, \dots, n\}$, the measure

$$\pi_{\sigma_n} := \frac{1}{n}\sum_{j=1}^{n} \delta_{(X_j, w_{\sigma_n(j)}^{(n)})}$$

has marginal distributions μ_n and $\tilde{\mu}_n'$, so

$$W_1(\mu_n, \tilde{\mu}_n') \le \int |x-y| \, d\pi_{\sigma_n}(x,y) = \frac{1}{n}\sum_{j=1}^{n} \left|X_j - w_{\sigma_n(j)}^{(n)}\right|. \tag{71}$$

We will now make a judicious choice of σ_n in order to take advantage of the "clumping" behavior of the roots and critical points of $p_n(z)$ proclaimed in the conclusion of Lemma 5.10.

To start, define the index sets S_Γ, $\Gamma \in \mathcal{M}_n$ by

$$S_\Gamma := \{1 \le j \le n : X_j \in \mathcal{O}_\Gamma\}.$$

For large n, on the complement of $\mathcal{E}_n^{\text{bad}}$, Lemma 5.10 guarantees that each $\mathcal{O}_\Gamma$, $\Gamma \in \mathcal{M}_n$ satisfying $\mathcal{O}_\Gamma \cap R_n^{\text{pair}} \neq \emptyset$ contains the same number of critical points and roots of $p_n(z)$. Consequently, we can choose σ_n so that for each $\Gamma \in \mathcal{M}_\Gamma$ satisfying $\mathcal{O}_\Gamma \cap R_n^{\text{pair}} \neq \emptyset$, we have

$$\sigma_n(S_\Gamma) = \left\{1 \leq j \leq n-1 : w_j^{(n)} \in \mathcal{O}_\Gamma\right\}$$

(recall that $\mathcal{O}_\Gamma$, $\Gamma \in \mathcal{M}_n$ are pairwise disjoint). For the remaining indices whose images under σ_n we haven't specified, arbitrarily assign them from among the remaining choices. (There is at least one index $1 \leq i \leq n$ for which $\sigma_n(i)$ is still undefined because the number of roots and critical points of $p_n(z)$ differs by 1. Recall that we have added $w_n^{(n)} = \overline{X}$ to account for this fact.)

Based on our construction of σ_n, Lemma 5.10 also implies that for large n, on the complement of $\mathcal{E}_n^{\text{bad}}$,

$$\left|X_j - w_{\sigma_n(j)}^{(n)}\right| \leq \frac{(\log n)^4}{n\,|m_\mu(X_j)|} \text{ when } X_j \in R_n^{\text{pair}},\ 1 \leq j \leq n \tag{72}$$

(Indeed, $X_j \in R_n^{\text{pair}}$ implies that $X_j \in \mathcal{O}_\Gamma$, with $\mathcal{O}_\Gamma \cap R_n^{\text{pair}} = \emptyset$ for some $\Gamma \in \mathcal{M}_n$.) By the Gauss–Lucas theorem, each critical point (and each root) of $p_n(z)$ is in the convex hull of the set $\{X_j\}_{j=1}^n$ of roots of $p_n(z)$, so for any $X_j \notin R_n^{\text{pair}}$, we have the trivial bound

$$\left|X_j - w_{\sigma_n(j)}^{(n)}\right| \leq 2\eta_n. \tag{73}$$

We now find an upper bound for $W_1(\mu_n, \tilde{\mu}_n')$ by splitting the sum on the right side of (71) into many pieces. To start, we classify the terms based on whether or not $X_j \in R_n^{\text{pair}}$ and apply (72) and (73) to obtain

$$W_1(\mu_n, \tilde{\mu}_n') \leq \frac{1}{n} \sum_{j:X_j \notin R_n^{\text{pair}}} 2\eta_n + \frac{1}{n} \sum_{j:X_j \in R_n^{\text{pair}}} \frac{(\log n)^4}{n\,|m_\mu(X_j)|}, \tag{74}$$

for large n, on the complement of $\mathcal{E}_n^{\text{bad}}$. According to the definition of R_n^{pair}, any $X_j \notin R_n^{\text{pair}}$ is an element of $A^{\lfloor 4\log(\log n)\rfloor} \cup A^{\lfloor 4\log(\log n)\rfloor+1}$, so there are $N_n^{\lfloor 4\log(\log n)\rfloor} + N_n^{\lfloor 4\log(\log n)\rfloor+1}$ terms in the leftmost sum. It follows that for large n, on the complement of $\mathcal{E}_n^{\text{bad}}$,

$$\sum_{j:X_j \notin R_n^{\text{pair}}} 2\eta_n \leq 4\eta_n C_\mu \log(\log n)\left(e^{2\lfloor 4\log(\log n)\rfloor} + e^{2\lfloor 4\log(\log n)\rfloor+2}\right) \ll_\mu (\log n)^9 \eta_n. \tag{75}$$

It remains to control the rightmost sum in (74), which we accomplish by grouping indices j, $1 \leq j \leq n$ that correspond to X_j that fall in the same region among A_n and A_n^k, $\lfloor 4\log(\log n)\rfloor + 2 \leq k \leq \lfloor \log(\sqrt{n})\rfloor$. We have,

$$\begin{aligned}\sum_{j:X_j \in R_n^{\text{pair}}} \frac{(\log n)^4}{n\,|m_\mu(X_j)|} &= \sum_{k=\lfloor 4\log(\log n)\rfloor+2}^{\lfloor \log(\sqrt{n})\rfloor} \sum_{j:X_j \in A_n^k} \frac{(\log n)^4}{n\,|m_\mu(X_j)|} + \sum_{j:X_j \in A_n} \frac{(\log n)^4}{n\,|m_\mu(X_j)|} \\ &\leq \sum_{k=\lfloor 4\log(\log n)\rfloor+2}^{\lfloor \log(\sqrt{n})\rfloor} N_n^k \cdot \frac{(\log n)^4\sqrt{n}}{ne^{k-1}} + n \cdot \frac{(\log n)^4\sqrt{n}}{ne^{\lfloor \log(\sqrt{n})\rfloor}},\end{aligned}$$

where the last inequality follows after applying the lower bounds on $|m_\mu(X_j)|$ that are given in the definitions of A_n and A_n^k (note that there are N_n^k terms in each inner

sum, and the very last sum has at most n terms). For large n, on the complement of $\mathcal{E}_n^{\text{bad}}$, we have $N_n^k < 2C_\mu e^{2k}\log(\log n)$, for $\lfloor 4\log(\log n)\rfloor \le k \le \lfloor \log(\sqrt{n})\rfloor$, so, continuing from above,

$$\begin{aligned}\sum_{j:X_j\in R_n^{\text{pair}}} \frac{(\log n)^4}{n\,|m_\mu(X_j)|} &\le 2C_\mu \log(\log n)\frac{e(\log n)^4}{\sqrt{n}} \sum_{k=\lfloor 4\log(\log n)\rfloor+2}^{\lfloor \log(\sqrt{n})\rfloor} e^k + \frac{(\log n)^4\sqrt{n}}{e^{\lfloor \log(\sqrt{n})\rfloor}} \\ &\le 2C_\mu \log(\log n)\frac{e(\log n)^4}{\sqrt{n}} \cdot \log(\sqrt{n})e^{\log(\sqrt{n})} + \frac{e(\log n)^4\sqrt{n}}{\sqrt{n}} \\ &\ll_\mu (\log n)^6.\end{aligned}$$

In view of the fact that $\mathbb{P}(\mathcal{E}_n^{\text{bad}}) = o(1)$, we complete the proof of Lemma 5.16 by combining the last asymptotic with (74) and (75). (Note that with probability $1 - o(1)$, $\eta_n \log n \ge 1$.) □

We conclude this subsection by remarking that Theorem 2.3 follows from Lemmas 5.1, 5.15, and 5.16 and the triangle inequality for the L_1-Wasserstein metric.

Appendix A. Proof of assorted results from Section 2

A.1. **Computation of m_μ for radially symmetric distributions.** The following lemma is useful for computing the Cauchy–Stieltjes transforms of radially symmetric distributions, which can expedite the verification of Assumptions 2.1 and 2.2 in a variety of situations. We note that Lemma A.1 also appears as Proposition 3.1 in [18].

Lemma A.1. *Suppose μ has a density $f(r,\theta) = f(r)$ that is radially symmetric about the origin. Then, $m_\mu(0) = 0$, and for $\xi \neq 0$,*

$$m_\mu(\xi) = \frac{2\pi}{\xi}\int_0^{|\xi|} r f(r)\,dr = \frac{1}{\xi}\mathbb{P}(|X| < |\xi|),$$

where $X \sim \mu$.

Proof. For $\xi \neq 0$, we can use polar coordinates and Laurent series to obtain

$$\begin{aligned} m_\mu(\xi) &= \int_0^{2\pi}\int_0^\infty \frac{f(r)}{\xi - re^{i\theta}}\cdot r\,dr\,d\theta \\ &= \frac{1}{\xi}\int_0^{|\xi|} r f(r)\int_0^{2\pi}\frac{1}{1-\frac{r}{\xi}e^{i\theta}}\,d\theta\,dr - \frac{1}{\xi}\int_{|\xi|}^\infty r f(r)\int_0^{2\pi}\frac{\frac{\xi}{r}e^{-i\theta}}{1-\frac{\xi}{r}e^{-i\theta}}\,d\theta\,dr \\ &= \frac{1}{\xi}\int_0^{|\xi|} r f(r)\sum_{j=0}^\infty \int_0^{2\pi}\left(\frac{r}{\xi}e^{i\theta}\right)^j d\theta\,dr \\ &\qquad - \frac{1}{\xi}\int_{|\xi|}^\infty r f(r)\sum_{j=0}^\infty \frac{\xi}{r}\int_0^{2\pi} e^{-i\theta}\left(\frac{\xi}{r}e^{-i\theta}\right)^j d\theta\,dr. \end{aligned}$$

The only nonzero integral occurs when the power on the exponential is 0, so we obtain

$$m_\mu(\xi) = \frac{2\pi}{\xi}\int_0^{|\xi|} r f(r)\,dr$$

as is desired. Finally, observe that

$$m_\mu(0) = \int_0^\infty \int_0^{2\pi} \frac{-f(r)}{re^{i\theta}} \cdot r \, d\theta \, dr = -\int_0^\infty f(r) \int_0^{2\pi} e^{-i\theta} \, d\theta \, dr = 0.$$

□

A.2. **Calculations for Example 2.7.** In this subsection, we establish that μ from Example 2.7 satisfies part (i) of Assumption 2.1. Setting each of the three branches in (7) to zero shows that the only zeros of $m_\mu(z)$ are when $z = 0, \pm\sqrt{3}$. We claim that there is a $C > 0$ such that if $X \sim \mu$, and $\varepsilon > 0$ is small, then, $\mathbb{P}(|m_\mu(X)| < \varepsilon) \leq C\varepsilon^2$. To start, consider that for $|z+2| < 1$,

$$\begin{aligned}
|m_\mu(z)| &= \frac{1}{2(z-2)} \left|(z-2)(\overline{z+2}) + 1\right| \\
&= \frac{1}{2\,|z-2|} \left|\left(z + \sqrt{3} - \sqrt{3} - 2\right)\left(\overline{z+\sqrt{3}} - \sqrt{3} + 2\right) + 1\right| \\
&= \frac{1}{2\,|z-2|} \left|\left|z+\sqrt{3}\right|^2 + (2-\sqrt{3})(z+\sqrt{3}) - (2+\sqrt{3})(\overline{z+\sqrt{3}})\right| \\
&= \frac{|z+\sqrt{3}|}{2\,|z-2|} \left|\overline{(z+\sqrt{3})} + 2 - \sqrt{3} - (2+\sqrt{3})\frac{\overline{z+\sqrt{3}}}{z+\sqrt{3}}\right|.
\end{aligned}$$

Since $|z+2| < 1$, it follows that $|z-2| < 6$ and also, by the triangle inequality,

$$\left|z+\sqrt{3}\right| \leq |z+2| + \left|\sqrt{3}-2\right| \leq 1 + 2 - \sqrt{3} = 3 - \sqrt{3}.$$

Hence, the reverse triangle inequality transforms our previous calculation into

$$\begin{aligned}
|m_\mu(z)| &\geq \frac{|z+\sqrt{3}|}{12} \left(\left|(2+\sqrt{3})\frac{\overline{z+\sqrt{3}}}{z+\sqrt{3}}\right| - \left|z+\sqrt{3}\right| - \left|2-\sqrt{3}\right|\right) \\
&\geq \frac{|z+\sqrt{3}|}{12} \left((2+\sqrt{3}) - (3-\sqrt{3}) - (2-\sqrt{3})\right) \\
&= \frac{\sqrt{3}-1}{4} \left|z+\sqrt{3}\right|,
\end{aligned}$$

for $|z+2| < 1$. Similarly, $|z-2| < 1$ implies that

$$|m_\mu(z)| \geq \frac{\sqrt{3}-1}{4} \left|z-\sqrt{3}\right|.$$

Since the random variable X can only take values z for which $|z \pm 2| < 1$, it follows that

$$\mathbb{P}\left(|m_\mu(X)| < \varepsilon\right) \leq \mathbb{P}\left(\left|X+\sqrt{3}\right| < c\varepsilon\right) + \mathbb{P}\left(\left|X-\sqrt{3}\right| < c\varepsilon\right) \leq \frac{c^2\varepsilon^2}{2},$$

where $c = 4/(\sqrt{3}-1)$ and $\varepsilon > 0$ is small enough that $B(\sqrt{3}, c\varepsilon) \subset B(2,1)$.

A.3. **Proof of Lemma 2.20.**

Proof. Fix $a > 0$, and let $b > 0$ be a large constant (depending on C and a) to be chosen later. Since Z is independent of $X_1, \ldots, X_n$ it follows that, with probability

1, $Z \notin \{X_1, \dots, X_n\}$. Hence the sum $\sum_{i=1}^n \frac{1}{Z-X_i}$ is well-defined and finite. By conditioning on the values of $X_2, \dots, X_n$ and Z, it suffices to prove that

$$\sup_{w \in \mathbb{C}} \sup_{z \in B(0,n^C)} \mathbb{P}\left(\left| \frac{1}{z - X_1} - w \right| \leq n^{-b} \right) \ll_a n^{-a}.$$

The claim now follows from Lemma A.2 below by taking $\varepsilon := n^{-b}$ and choosing b sufficiently large in terms of C and a. □

Lemma A.2. *Fix $C > 0$, and let X be a complex-valued random variable that is absolutely continuous (with respect to Lebesgue measure on $\mathbb{C}$) and which has density bounded by n^C. If $\mathbb{E}|X| \leq n^C$, then for every $a > 0$ and $0 < \varepsilon < 1$,*

$$\sup_{w \in \mathbb{C}} \sup_{z \in B(0,n^C)} \mathbb{P}\left(\left| \frac{1}{z - X} - w \right| \leq \varepsilon \right) \leq 4\sqrt{\varepsilon} n^C + 4\pi\varepsilon n^{3C+2a} + n^{-a}.$$

Proof. Fix $w \in \mathbb{C}$ and $z \in B(0, n^C)$. We consider two cases. If $|w| \leq \sqrt{\varepsilon}$, then

$$\begin{aligned}
\mathbb{P}\left(\left| \frac{1}{z - X} - w \right| \leq \varepsilon \right) &\leq \mathbb{P}\left(\left| \frac{1}{z - X} \right| \leq 2\sqrt{\varepsilon} \right) \\
&\leq \mathbb{P}\left(|X - z| \geq \frac{1}{2\sqrt{\varepsilon}} \right) \\
&\leq 2\sqrt{\varepsilon} \left(\mathbb{E}|X| + n^C \right) \\
&\leq 4\sqrt{\varepsilon} n^C
\end{aligned}$$

by Markov's inequality.

We now consider the case where $|w| > \sqrt{\varepsilon}$. Define the event $\mathcal{E} := \{|X| \leq n^{C+a}\}$. By Markov's inequality, it follows that $\mathbb{P}(\mathcal{E}^c) \leq n^{-a}$. Thus, we obtain

$$\begin{aligned}
\mathbb{P}\left(\left| \frac{1}{z - X} - w \right| \leq \varepsilon \right) &\leq \mathbb{P}\left(\left| \frac{1}{w} - (z - X) \right| \leq \sqrt{\varepsilon} |z - X| \right) \\
&\leq \mathbb{P}\left(\left| \frac{1}{w} - (z - X) \right| \leq \sqrt{\varepsilon} |z - X| \;\middle|\; \mathcal{E} \right) \mathbb{P}(\mathcal{E}) + \mathbb{P}(\mathcal{E}^c) \\
&\leq \mathbb{P}\left(\left| \frac{1}{w} - (z - X) \right| \leq 2\sqrt{\varepsilon} n^{C+a} \right) + n^{-a} \\
&\leq \mathbb{P}\left(X \in B(w^{-1} - z, 2\sqrt{\varepsilon} n^{C+a}) \right) + n^{-a} \\
&\leq \pi n^C \left(2\sqrt{\varepsilon} n^{C+a} \right)^2 + n^{-a}.
\end{aligned}$$

Combining the bounds above yields

$$\mathbb{P}\left(\left| \frac{1}{z - X} - w \right| \leq \varepsilon \right) \leq 4\sqrt{\varepsilon} n^C + 4\pi\varepsilon n^{3C+2a} + n^{-a}$$

for any $w \in \mathbb{C}$ and $z \in B(0, n^C)$. The proof of the lemma is complete. □

Appendix B. A heavy-tailed CLT

In this subsection, we prove Theorem B.1, a CLT for "heavy-tailed" random variables that have the same distribution as $Y := \frac{1}{\xi - X}$, where $X \sim \mu$ and μ has a continuous density f in a neighborhood of ξ. Notice that $\mathbb{E}|Y|^p < \infty$ for $p \in [0, 2)$, but $\mathbb{E}|Y|^2 = \infty$. Many results demonstrate that Y is in the domain of attraction of a normal random variable (see e.g. Section XVII.5 in [9], Theorem

11 in Section 6.4 of [10], and Theorem 3.10 in [26]), however, our implementation of Theorem B.1 requires specific information about the parameters of the limiting normal distribution; we include an explicit statement and proof for clarity.

Theorem B.1. *Let $X_1, X_2, \dots$ be iid, complex-valued random variables with common distribution μ, fix $s, k \in \mathbb{N}$, and suppose $\xi_1, \dots, \xi_s, t_1, \dots, t_k \in \mathbb{C}$ are deterministic values with $\xi_1, \dots, \xi_s$ distinct. In addition, assume that μ has a bounded density f in a neighborhood of each ξ_l, $1 \le l \le s$, that is continuous at these points. Then,*

$$\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \sum_{k=1}^{s} t_k \left[\frac{1}{\xi_k - X_j} - m_\mu(\xi) \right] \longrightarrow N$$

in distribution as $n \to \infty$, where N is a complex random variable with mean zero whose real and imaginary parts have a joint Gaussian distribution that has covariance matrix

$$\Sigma := \sum_{k=1}^{s} \frac{\pi |t_k|^2 f(\xi_k)}{2} I. \tag{76}$$

(Here, I denotes the 2×2 identity matrix.)

Proof. We proceed by Lindeberg's exchange method [20]. (See also [4]. Similar methods have been applied to problems in random matrix theory; see e.g. [33], [34].) To that end, let $N, N_1, N_2, \dots$ be a sequence of iid complex random variables independent of $\{X_j\}$, whose components have a joint Gaussian distribution with mean zero and covariance matrix Σ, defined in (76), and let $g : \mathbb{C} \to \mathbb{R}$ be a smooth test function with compact support. We will show that

$$\left| \mathbb{E}\left[g\left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \sum_{k=1}^{s} t_k \left[\frac{1}{\xi_k - X_j} - m_\mu(\xi_k) \right] \right) \right] - \mathbb{E}\left[g\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} N_j \right) \right] \right| \to 0, \tag{77}$$

as $n \to \infty$, which implies convergence of the corresponding measures in the vague topology. Convergence in distribution follows because for each n, $n^{-1/2} \sum_{j=1}^{n} N_j$ has the same distribution as the random variable N. (See e.g. Exercise 1.1.25 of [31], pages 23-33.)

Since the random variables $\sum_{k=1}^{s} \frac{t_k}{\xi_k - X_j}$ are heavy-tailed, we initially need to truncate them. Let $\varepsilon \in (0,1)$ be fixed, and define

$$\zeta_j := \sum_{k=1}^{s} \frac{t_k}{\xi_k - X_j} \mathbb{1}_{\{|\xi_k - X_j|^{-1} < \varepsilon \sqrt{n \log n}\}},$$
$$\widetilde{\zeta}_j := \zeta_j - \mathbb{E}[\zeta_j].$$

(Be aware that this notation suppresses the dependence of ζ_j and $\widetilde{\zeta}_j$ on ε and n.)

Lemma B.2. *There is a constant $C_{\mu,s,\vec{t}} > 0$, depending only on μ, s, and $t_1, \dots, t_k$, and there is a natural number $K_{\mu,g,\varepsilon}$ so that $n \ge K_{\mu,g,\varepsilon}$ implies*

$$\left| \mathbb{E}\left[g\left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \widetilde{\zeta}_j \right) \right] - \mathbb{E}\left[g\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} N_j \right) \right] \right| \le C_{\mu,s,\vec{t}} \varepsilon.$$

Proof. By Taylor's theorem applied to the Taylor series for g centered at

$$A_{1,n} := \frac{1}{\sqrt{n \log n}} \sum_{j=2}^{n} \widetilde{\zeta}_j,$$

we have

$$\begin{aligned} g\left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \widetilde{\zeta}_j \right) &= g\left(A_{1,n}\right) + \frac{g_x\left(A_{1,n}\right)}{\sqrt{n \log n}} \operatorname{Re}\left(\widetilde{\zeta}_1\right) + \frac{g_y\left(A_{1,n}\right)}{\sqrt{n \log n}} \operatorname{Im}\left(\widetilde{\zeta}_1\right) \\ &\quad + \frac{g_{xx}\left(A_{1,n}\right)}{2n \log n} \operatorname{Re}\left(\widetilde{\zeta}_1\right)^2 + \frac{g_{yy}\left(A_{1,n}\right)}{2n \log n} \operatorname{Im}\left(\widetilde{\zeta}_1\right)^2 \\ &\quad + \frac{g_{xy}\left(A_{1,n}\right)}{n \log n} \operatorname{Re}\left(\widetilde{\zeta}_1\right) \operatorname{Im}\left(\widetilde{\zeta}_1\right) + R_3\left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \widetilde{\zeta}_j \right), \end{aligned}$$

where

$$\left| R_3\left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \widetilde{\zeta}_j \right) \right| \leq \frac{8 \cdot C_g}{2! \cdot (n \log n)^{3/2}} \left| \widetilde{\zeta}_1 \right|^3,$$

and C_g is any constant that is an upper bound for the mixed partial derivatives of g up to and including order three (which are compactly supported and thus bounded). Taking the expectation of both sides yields (by independence and the fact that $\widetilde{\zeta}_j$ are centered)

$$\begin{aligned} &\mathbb{E}\left[g\left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \widetilde{\zeta}_j \right) \right] \\ &\quad = \mathbb{E}[g(A_{1,n})] + \frac{\mathbb{E}[g_{xx}\left(A_{1,n}\right)]}{2n \log n} \mathbb{E}\left[\operatorname{Re}\left(\widetilde{\zeta}_1\right)^2 \right] + \frac{\mathbb{E}[g_{yy}\left(A_{1,n}\right)]}{2n \log n} \mathbb{E}\left[\operatorname{Im}\left(\widetilde{\zeta}_1\right)^2 \right] \\ &\quad + \frac{\mathbb{E}[g_{xy}\left(A_{1,n}\right)]}{n \log n} \mathbb{E}\left[\operatorname{Re}\left(\widetilde{\zeta}_1\right) \operatorname{Im}\left(\widetilde{\zeta}_1\right) \right] + \mathbb{E}\left[R_3\left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \widetilde{\zeta}_j \right) \right]. \end{aligned}$$

Similarly, we have

$$\begin{aligned} &\mathbb{E}\left[g\left(\frac{N_1}{\sqrt{n}} + \frac{1}{\sqrt{n \log n}} \sum_{j=2}^{n} \widetilde{\zeta}_j \right) \right] \\ &\quad = \mathbb{E}[g(A_{1,n})] + \frac{\mathbb{E}[g_{xx}\left(A_{1,n}\right)]}{2n} \mathbb{E}\left[\operatorname{Re}\left(N_1\right)^2 \right] + \frac{\mathbb{E}[g_{yy}\left(A_{1,n}\right)]}{2n} \mathbb{E}\left[\operatorname{Im}\left(N_1\right)^2 \right] \\ &\quad + \frac{\mathbb{E}[g_{xy}\left(A_{1,n}\right)]}{n} \mathbb{E}\left[\operatorname{Re}\left(N_1\right) \operatorname{Im}\left(N_1\right) \right] + \mathbb{E}\left[R_3\left(\frac{N_1}{\sqrt{n}} + \frac{1}{\sqrt{n \log n}} \sum_{j=2}^{n} \widetilde{\zeta}_j \right) \right], \end{aligned}$$

where

$$\left| R_3\left(\frac{N_1}{\sqrt{n}} + \frac{1}{\sqrt{n \log n}} \sum_{j=2}^{n} \widetilde{\zeta}_j \right) \right| \leq \frac{8 \cdot C_g}{2! \cdot n^{3/2}} \left| N_1 \right|^3.$$

The difference between these two equations is bounded by

$$\begin{aligned}\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\right)\right]-\mathbb{E}\left[g\left(\frac{N_1}{\sqrt{n}}+\frac{1}{\sqrt{n\log n}}\sum_{j=2}^{n}\widetilde{\zeta}_j\right)\right]\right| \\ \leq \frac{C_g}{2n}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta}_1)^2\right]-\mathbb{E}\left[\mathrm{Re}\,(N_1)^2\right]\right| \\ +\frac{C_g}{2n}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Im}(\widetilde{\zeta}_1)^2\right]-\mathbb{E}\left[\mathrm{Im}\,(N_1)^2\right]\right| \\ +\frac{C_g}{n}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta}_1)\,\mathrm{Im}(\widetilde{\zeta}_1)\right]-\mathbb{E}\left[\mathrm{Re}\,(N_1)\,\mathrm{Im}\,(N_1)\right]\right| \\ +\frac{4C_g}{(n\log n)^{3/2}}\mathbb{E}\left[\left|\widetilde{\zeta}_1\right|^3\right]+\frac{4C_g}{n^{3/2}}\mathbb{E}\left[|N_1|^3\right].\end{aligned}$$

If we continue, for $2 \leq k \leq n$, the process of computing the second order Taylor polynomials of g centered at

$$A_{k,n} := \frac{1}{\sqrt{n}}\sum_{j=1}^{k-1}N_j+\frac{1}{\sqrt{n\log n}}\sum_{j=k+1}^{n}\widetilde{\zeta}_j$$

and evaluating them at both

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1}N_j+\frac{1}{\sqrt{n\log n}}\sum_{j=k}^{n}\widetilde{\zeta}_j \quad\text{and}\quad \frac{1}{\sqrt{n}}\sum_{j=1}^{k}N_j+\frac{1}{\sqrt{n\log n}}\sum_{j=k+1}^{n}\widetilde{\zeta}_j,$$

we find that

$$\begin{aligned}\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1}N_j+\frac{1}{\sqrt{n\log n}}\sum_{j=k}^{n}\widetilde{\zeta}_j\right)\right]\right. \\ \left.-\mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{k}N_j+\frac{1}{\sqrt{n\log n}}\sum_{j=k+1}^{n}\widetilde{\zeta}_j\right)\right]\right| \\ \leq \frac{C_g}{2n}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta}_k)^2\right]-\mathbb{E}\left[\mathrm{Re}\,(N_k)^2\right]\right| \\ +\frac{C_g}{2n}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Im}(\widetilde{\zeta}_k)^2\right]-\mathbb{E}\left[\mathrm{Im}\,(N_k)^2\right]\right| \\ +\frac{C_g}{n}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta}_k)\,\mathrm{Im}(\widetilde{\zeta}_k)\right]-\mathbb{E}\left[\mathrm{Re}\,(N_k)\,\mathrm{Im}\,(N_k)\right]\right| \\ +\frac{4C_g}{(n\log n)^{3/2}}\mathbb{E}\left[\left|\widetilde{\zeta}_k\right|^3\right]+\frac{4C_g}{n^{3/2}}\mathbb{E}\left[|N_k|^3\right].\end{aligned}$$

Now, repeatedly applying the triangle inequality and using the fact that the $\widetilde{\zeta_j}$ and N_j are iid gives

$$\begin{aligned}
&\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\right)\right]-\mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}N_j\right)\right]\right| \\
&\qquad\le \frac{C_g}{2}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta_1})^2\right]-\mathbb{E}\left[\mathrm{Re}\,(N)^2\right]\right| \\
&\qquad\quad+\frac{C_g}{2}\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Im}(\widetilde{\zeta_1})^2\right]-\mathbb{E}\left[\mathrm{Im}\,(N)^2\right]\right| \\
&\qquad\quad+C_g\left|\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta_1})\,\mathrm{Im}(\widetilde{\zeta_1})\right]-\mathbb{E}\left[\mathrm{Re}\,(N)\,\mathrm{Im}\,(N)\right]\right| \\
&\qquad\quad+\frac{4C_g}{\log n\sqrt{n\log n}}\mathbb{E}\left[\left|\widetilde{\zeta_1}\right|^3\right]+\frac{4C_g}{\sqrt{n}}\mathbb{E}\left[|N|^3\right].
\end{aligned}\tag{78}$$

In order to establish Lemma B.2, we need to show that each of the terms on the right side of (78) is dominated by ε as $n\to\infty$. It is in these computations that we use the fact that f is continuous at $\xi_1,\dots,\xi_s$. To take advantage of this hypothesis, fix $\eta>0$, and note that there is a $\delta=\delta(\eta)>0$ such that $\delta<\frac{1}{2}\min_{1\le k<l\le s}|\xi_k-\xi_l|$ and for which $|z-\xi_k|<\delta$ implies $|f(z)-f(\xi_k)|<\eta$ for $1\le k\le s$. We have

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta_1})\right]&=\mathbb{E}\left[\mathrm{Re}^2(\zeta_1)\right]-\left(\mathbb{E}\left[\mathrm{Re}(\zeta_1)\right]\right)^2\le\mathbb{E}\left[\mathrm{Re}^2(\zeta_1)\right] \\
&\le\mathbb{E}\left[\mathrm{Re}^2\left(\sum_{k=1}^{s}\frac{t_k}{\xi_k-X_1}\right)\prod_{k=1}^{s}\mathbb{1}_{|\xi_k-X_1|\ge\delta}\right] \\
&\quad+\sum_{k=1}^{s}\mathbb{E}\left[\mathrm{Re}^2\left(\sum_{l=1}^{s}\frac{t_l}{\xi_l-X_1}\right)\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_k-X_1|<\delta}\prod_{l\ne k}\mathbb{1}_{|\xi_l-X_1|\ge\delta}\right] \\
&\le\left(\sum_{k=1}^{s}\frac{|t_k|}{\delta}\right)^2+\sum_{k=1}^{s}\mathbb{E}\left[\mathrm{Re}^2\left(\frac{t_k}{\xi_k-X_1}\right)\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_k-X_1|<\delta}\right] \\
&\quad+2\sum_{k=1}^{s}\sum_{l\ne k}\frac{|t_l|}{\delta}\mathbb{E}\left|\frac{t_k}{\xi_k-X_1}\right|+\sum_{k=1}^{s}\left(\sum_{l\ne k}\frac{|t_l|}{\delta}\right)^2,
\end{aligned}$$

where the last inequality follows from the fact that

$$\mathrm{Re}^2(z+w)=\mathrm{Re}^2(z)+2\,\mathrm{Re}(z)\,\mathrm{Re}(w)+\mathrm{Re}^2(w)\le\mathrm{Re}^2(z)+2\,|z|\,|w|+|w|^2.$$

Since $\mathbb{E}\left|\frac{1}{\xi_k-X_1}\right|$ is bounded by a constant that depends only on μ (see Lemma 3.5), there is a constant $C_{s,\vec{t},\delta}$ depending on s, $t_1,\dots,t_s$ and δ so that, continuing from

above,

$$\begin{aligned}\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta_1})\right] &\le C_{s,\vec{t},\delta} + \sum_{k=1}^{s}\mathbb{E}\left[\frac{\mathrm{Re}^2((\xi_k - X_1)/t_k)}{|(\xi_k - X_1)/t_k|^4}\mathbb{1}_{1/(\varepsilon\sqrt{n\log n})<|\xi_k - X_1|<\delta}\right]\\ &\le C_{s,\vec{t},\delta} + \sum_{k=1}^{s}(f(\xi_k)+\eta)\,|t_k|^2\int_0^{2\pi}\int_{1/(|t_k|\varepsilon\sqrt{n\log n})}^{\delta/|t_k|}\frac{r^2\cos^2\theta}{r^4}r\,dr\,d\theta\\ &\le C_{s,\vec{t},\delta} + \sum_{k=1}^{s}\pi(f(\xi_k)+\eta)\,|t_k|^2\log(\delta\varepsilon\sqrt{n\log n}).\end{aligned}$$

Dividing both sides by $\log n$ yields

$$\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta_1})\right] \le \sum_{k=1}^{s}|t_k|^2\,\frac{\pi(f(\xi)+\eta)}{2} + o(1). \tag{79}$$

On the other hand, similar to above,

$$\begin{aligned}&\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta_1})\right]\\ &\quad= \mathbb{E}\left[\mathrm{Re}^2(\zeta_1)\right] - \left(\mathbb{E}\left[\mathrm{Re}(\zeta_1)\right]\right)^2\\ &\quad\ge \sum_{k=1}^{s}\mathbb{E}\left[\mathrm{Re}^2\left(\sum_{l=1}^{s}\frac{t_l}{\xi_l - X_1}\right)\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_k - X_1|<\delta}\prod_{l\ne k}\mathbb{1}_{|\xi_l - X_1|\ge\delta}\right] - o(\log n)\\ &\quad\ge \sum_{k=1}^{s}\mathbb{E}\left[\frac{\mathrm{Re}^2((\xi_k - X_1)/t_k)}{|(\xi_k - X_1)/t_k|^4}\mathbb{1}_{1/(\varepsilon\sqrt{n\log n})<|\xi - X_1|<\delta}\right] - o(\log n)\\ &\quad\ge \sum_{k=1}^{s}(f(\xi_k)-\eta)\,|t_k|^2\int_0^{2\pi}\int_{1/(|t_k|\varepsilon\sqrt{n\log n})}^{\delta/|t_k|}\frac{r^2\cos^2\theta}{r^4}r\,dr\,d\theta - o(\log n)\\ &\quad\ge \sum_{k=1}^{s}\pi(f(\xi_k)-\eta)\,|t_k|^2\log(\delta\varepsilon\sqrt{n\log n}) - o(\log n),\end{aligned}$$

and dividing by $\log n$ yields

$$\frac{\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta_1})\right]}{\log n} \ge \sum_{k=1}^{s}\frac{\pi(f(\xi_k)-\eta)\,|t_k|^2}{2} - o(1). \tag{80}$$

If we combine inequalities (79) and (80) and first take $\limsup_{n\to\infty}$ (respectively $\liminf_{n\to\infty}$) of both sides and then take $\eta\to 0$, we see that

$$\lim_{n\to\infty}\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta_1})\right] = \sum_{k=1}^{s}\frac{\pi f(\xi_k)\,|t_k|^2}{2} = \mathbb{E}[\mathrm{Re}^2(N)]. \tag{81}$$

(Note: f is bounded, and by Lemma 3.5, the expectation of $|\xi - X_1|^{-1}$ is uniformly bounded, so the limit in n is uniform in ξ.) Nearly identical arguments show that

$$\lim_{n\to\infty}\frac{1}{\log n}\mathbb{E}\left[\mathrm{Im}^2(\widetilde{\zeta_1})\right] = \sum_{k=1}^{s}\frac{\pi f(\xi_k)\,|t_k|^2}{2} = \mathbb{E}[\mathrm{Im}^2(N)] \tag{82}$$

and

$$\lim_{n\to\infty}\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta_1})\,\mathrm{Im}(\widetilde{\zeta_1})\right] = 0 = \mathbb{E}[\mathrm{Re}(N)\,\mathrm{Im}(N)], \tag{83}$$

with the only difference being that when we find bounds for $\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta_1})\,\mathrm{Im}(\widetilde{\zeta_1})\right]$, we consider separately the cases where the integrand is positive and negative.

In our quest to prove Lemma B.2, we next show that

$$\limsup_{n\to\infty} \frac{1}{\log n\sqrt{n\log n}}\mathbb{E}\left[\left|\widetilde{\zeta_1}\right|^3\right] \leq O_{\mu,s,\vec{t}}(\varepsilon). \tag{84}$$

Note that

$$\mathbb{E}\left[\left|\widetilde{\zeta_1}\right|^3\right] \leq 2\cdot\mathbb{E}\left[|\zeta_1|^3\right] + 6\cdot\mathbb{E}\left[|\zeta_1|^2\right]\cdot\mathbb{E}\,|\zeta_1| \leq 8\varepsilon\sqrt{n\log n}\sum_{k=1}^{s}|t_k|\,\mathbb{E}\left[|\zeta_1|^2\right],$$

where the last inequality comes from using the fact that, almost surely, $|\zeta_1| \leq \varepsilon\sqrt{n\log n}\sum_{k=1}^{s}|t_k|$. Choose $\delta_1 > 0$ so that $\delta_1 < \frac{1}{2}\min_{1\leq k<l\leq s}|\xi_k - \xi_l|$ and that for $|z-\xi_k|<\delta_1$, $1\leq k\leq s$, we have $|f(z)-f(\xi_k)|<1$. Then, it follows that for n large enough to ensure $\frac{1}{\varepsilon\sqrt{n\log n}} \leq \delta_1$,

$$\begin{aligned}
\frac{\mathbb{E}\left[\left|\widetilde{\zeta_1}\right|^3\right]}{\log n\sqrt{n\log n}} &\leq \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\mathbb{E}\left[|\zeta_1|^2\right] \\
&\leq \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\left(\mathbb{E}\left[\left(\sum_{k=1}^{s}\frac{t_k}{\xi_k - X_1}\right)^2\prod_{k=1}^{s}\mathbb{1}_{|\xi_k - X_1|\geq\delta_1}\right]\right. \\
&\quad \left. + \sum_{k=1}^{s}\mathbb{E}\left[\left(\sum_{l=1}^{s}\frac{t_l}{\xi_l - X_1}\right)^2\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_k - X_1|<\delta_1}\prod_{l\neq k}\mathbb{1}_{|\xi_l - X_1|\geq\delta_1}\right]\right) \\
&\leq \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\left(\sum_{l=1}^{s}\frac{|t_k|}{\delta_1}\right)^2 \\
&\quad + \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\sum_{l=1}^{s}\mathbb{E}\left[\left|\frac{t_l}{\xi_l - X_1}\right|^2\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_l - X_1|<\delta_1}\right] \\
&\quad + \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\left(\sum_{l=1}^{s}2\mathbb{E}\left|\frac{t_l}{\xi_l - X_1}\right|\sum_{j\neq l}\frac{|t_j|}{\delta_1} + \sum_{l=1}^{s}\left(\sum_{j\neq l}\frac{|t_j|}{\delta_1}\right)^2\right),
\end{aligned}$$

where we have used the fact that $|z+w|^2 \leq |z|^2 + 2\,|z|\,|w| + |w|^2$. Continuing from above, where the second sum is the only one of non-negligible order, we have

$$\begin{aligned}
&\frac{1}{\log n\sqrt{n\log n}}\mathbb{E}\left[\left|\widetilde{\zeta_1}\right|^3\right] \\
&\quad\leq \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\sum_{l=1}^{s}\mathbb{E}\left[\left|\frac{t_l}{\xi_l - X_1}\right|^2\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_l - X_1|<\delta_1}\right] + o(1) \\
&\quad\leq \sum_{k,l=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}8\varepsilon(f(\xi_l)+1)\,|t_l|^2\int_0^{2\pi}\int_{1/(\varepsilon\sqrt{n\log n})}^{\delta_1}\frac{1}{r^2}r\,dr\,d\theta + o(1) \\
&\quad= \sum_{k,l=1}^{s}\frac{16\pi\varepsilon(f(\xi_l)+1)\,|t_k|\,|t_l|^2\log(\delta_1\varepsilon\sqrt{n\log n})}{\log n} + o(1)
\end{aligned}$$

and taking $\limsup_{n\to\infty}$ establishes (84). We conclude the proof of Lemma B.2 by combining equations (78), (81), (82), (83), and (84) in view of the facts that $|N|$ has a finite third moment and $f(z)$ and ε are bounded. □

In order to establish (77), we still need to remove the truncation, which we will accomplish through a series of interpolations. We have

$$\begin{aligned}\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\sum_{k=1}^{s} t_k\left(\frac{1}{\xi_k - X_j} - m_\mu(\xi_k)\right)\\ = \frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\left(\sum_{k=1}^{s}\frac{t_k}{\xi_k - X_j} - \zeta_j\right) + \frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\\ + \frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\left(\mathbb{E}[\zeta_j] - \sum_{k=1}^{s} t_k m_\mu(\xi_k)\right).\end{aligned}$$

For n large enough to guarantee that the density, f, is well-defined and bounded by a constant, C_f, on

$$\bigcup_{k=1}^{s} B\left(\xi_k, \frac{1}{\varepsilon\sqrt{n\log n}}\right),$$

it follows that

$$\mathbb{E}\left|\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\left(\sum_{k=1}^{s}\frac{t_k}{\xi_k - X_j} - \zeta_j\right)\right| \quad \text{and} \quad \left|\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\left(\mathbb{E}[\zeta_j] - \sum_{k=1}^{s} t_k m_\mu(\xi_k)\right)\right|$$

are both less than

$$\begin{aligned}&\frac{n}{\sqrt{n\log n}}\sum_{k=1}^{s}|t_k|\,\mathbb{E}\left[\frac{1}{|\xi_k - X_1|}\mathbb{1}_{|\xi_k - X_1|^{-1}\geq\varepsilon\sqrt{n\log n}}\right]\\ &\qquad\leq \frac{ns\max_{1\leq k\leq s}|t_k|\cdot C_f}{\sqrt{n\log n}}\int_0^{2\pi}\int_0^{1/(\varepsilon\sqrt{n\log n})}\frac{1}{r}r\,dr\,d\theta\\ &\qquad\leq \frac{ns\max_{1\leq k\leq s}|t_k|\cdot C_f}{\sqrt{n\log n}}\frac{2\pi}{\varepsilon\sqrt{n\log n}}\\ &\qquad= o(1).\end{aligned}$$

Consequently,

$$\mathbb{E}\left|\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\sum_{k=1}^{s} t_k\left(\frac{1}{\xi - X_j} - m_\mu(\xi)\right) - \frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\right| = o(1).$$

We can take advantage of the fact that g is Lipshitz (indeed, g is smooth with compact support, so it has bounded partial derivatives), to obtain

$$\mathbb{E}\left|g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\sum_{k=1}^{s} t_k\left(\frac{1}{\xi - X_j} - m_\mu(\xi)\right)\right) - g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\right)\right| = o(1).$$

Lemma B.2 now implies that for n larger than a constant depending on μ, g, ε, s, and $t_1, \ldots, t_s$,

$$\begin{aligned}\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\sum_{k=1}^{s}t_k\left[\frac{1}{\xi - X_j} - m_\mu(\xi)\right]\right)\right] - \mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}N_j\right)\right]\right| \\ \leq \left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\right)\right] - \mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}N_j\right)\right]\right| + o(1) \\ = O_{\mu,s,\vec{t},g}(\varepsilon),\end{aligned}$$

so taking $\varepsilon \to 0$ yields equation (77). The conclusion of Theorem B.1 follows since our choice of g was arbitrary. □

References

[1] M. S. Bartlett. An inverse matrix adjustment arising in discriminant analysis. *Ann. Math. Statistics*, 22:107–111, 1951.

[2] P. Billingsley. *Probability and measure*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, 2012. Anniversary edition [of MR1324786], With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.

[3] S.-S. Byun, J. Lee, and T. R. Reddy. Zeros of random polynomials and its higher derivatives. Available at arXiv:1801.08974, 2018.

[4] S. Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.*, 34(6):2061–2076, 2006.

[5] W. S. Cheung and T. W. Ng. A companion matrix approach to the study of zeros and critical points of a polynomial. *J. Math. Anal. Appl.*, 319(2):690–707, 2006.

[6] W. S. Cheung and T. W. Ng. Relationship between the zeros of two polynomials. *Linear Algebra Appl.*, 432(1):107–115, 2010.

[7] M. R. Dennis and J. H. Hannay. Saddle points in the chaotic analytic function and Ginibre characteristic polynomial. *J. Phys. A*, 36(12):3379–3383, 2003. Random matrix theory.

[8] I. Fáry and E. M. Isenberg. On a converse of the Jordan curve theorem. *Amer. Math. Monthly*, 81:636–639, 1974.

[9] W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[10] J. Galambos. *Advanced probability theory*, volume 10 of *Probability: Pure and Applied*. Marcel Dekker, Inc., New York, second edition, 1995.

[11] A. Gut. *Probability: a graduate course*. Springer Texts in Statistics. Springer, New York, 2005.

[12] B. Hanin. Correlations and pairing between zeros and critical points of Gaussian random polynomials. *Int. Math. Res. Not. IMRN*, 2015(2):381–421, 2015.

[13] B. Hanin. Pairing of zeros and critical points for random meromorphic functions on Riemann surfaces. *Math. Res. Lett.*, 22(1):111–140, 2015.

[14] B. Hanin. Pairing of zeros and critical points for random polynomials. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(3):1498–1511, 2017.

[15] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, second edition, 2013.

[16] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág. *Zeros of Gaussian analytic functions and determinantal point processes*, volume 51 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2009.

[17] Z. Kabluchko. Critical points of random polynomials with independent identically distributed roots. *Proc. Amer. Math. Soc.*, 143(2):695–702, 2015.

[18] Z. Kabluchko and H. Seidel. Distances between zeroes and critical points for random polynomials with i.i.d. zeroes. Available at arXiv:1807.02140, 2018.

[19] N. L. Komarova and I. Rivin. Harmonic mean, random polynomials and stochastic matrices. *Adv. in Appl. Math.*, 31(2):501–526, 2003.

[20] J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.*, 15(1):211–225, 1922.
[21] M. Marden. *Geometry of polynomials*. Second edition. Mathematical Surveys, No. 3. American Mathematical Society, Providence, R.I., 1966.
[22] S. O'Rourke. Critical points of random polynomials and characteristic polynomials of random matrices. *Int. Math. Res. Not. IMRN*, 2016(18):5616–5651, 2016.
[23] S. O'Rourke and N. Williams. Pairing between zeros and critical points of random polynomials with independent roots. *Trans. Amer. Math. Soc.*, 371(4):2343–2381, 2019.
[24] S. O'Rourke and P. M. Wood. Spectra of nearly Hermitian random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 53(3):1241–1279, 2017.
[25] R. Pemantle and I. Rivin. The distribution of zeros of the derivative of a random polynomial. In *Advances in combinatorics*, pages 259–273. Springer, Heidelberg, 2013.
[26] V. V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.
[27] Q. I. Rahman and G. Schmeisser. *Analytic theory of polynomials*, volume 26 of *London Mathematical Society Monographs. New Series*. The Clarendon Press, Oxford University Press, Oxford, 2002.
[28] T. R. Reddy. On critical points of random polynomials and spectrum of certain products of random matrices. Available at arXiv:1602.05298, 2016.
[29] S. Steinerberger. A stability version of the gauss-lucas theorem and applications. Available at arXiv:1805.10454, 2018.
[30] S. D. Subramanian. On the distribution of critical points of a polynomial. *Electron. Commun. Probab.*, 17:no. 37, 9, 2012.
[31] T. Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.
[32] T. Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields*, 155(1-2):231–263, 2013.
[33] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011.
[34] T. Tao and V. Vu. Local universality of zeroes of random polynomials. *Int. Math. Res. Not. IMRN*, 2015(13):5053–5139, 2015.
[35] T. Tao and V. Vu. Random matrices: universality of local spectral statistics of non-Hermitian matrices. *Ann. Probab.*, 43(2):782–874, 2015.
[36] C. Thomassen. The converse of the Jordan curve theorem and a characterization of planar maps. *Geom. Dedicata*, 32(1):53–57, 1989.
[37] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.
[38] R. L. Wilder. *Topology of manifolds*. American Mathematical Society Colloquium Publications, Vol. XXXII. American Mathematical Society, Providence, R.I., 1963.

Department of Mathematics, University of Colorado, Campus Box 395, Boulder, CO 80309-0395, USA
E-mail address: sean.d.orourke@colorado.edu

Department of Mathematical Sciences, Appalachian State University, 342 Walker Hall, 121 Bodenheimer Dr., Boone, NC 28608, USA
E-mail address: williamsnn@appstate.edu