On the pairing between zeros and critical points of random polynomials with independent roots

by

Noah Williams

B.A., University of Wisconsin-Eau Claire, 2013

M.A., University of Colorado Boulder, 2017

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

2019

This thesis entitled:
On the pairing between zeros and critical points of random polynomials with independent roots
written by Noah Williams
has been approved for the Department of Mathematics

Professor Sean O'Rourke

Professor Boris Hanin

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Williams, Noah (Ph.D., Mathematics)

On the pairing between zeros and critical points of random polynomials with independent roots

Thesis directed by Professor Sean O'Rourke

Consider the random, complex polynomial $p_n(z) = \prod_{j=1}^n (z - X_j)$, whose roots $X_1, \ldots, X_n$ are complex-valued random variables. It is known that for large n, when the roots are independently and identically distributed (iid), the critical points and roots of p_n are stochastically similar. In particular, Pemantle and Rivin, Kabluchko, Reddy, and others showed that when $X_1, X_2, \ldots$ are iid with distribution μ, then the empirical measure constructed from the critical points of p_n converges to μ in probability as the degree n tends to infinity.

Simulations show that, in fact, the roots and critical points of p_n "pair-up" with one another in a nearly one-to-one fashion, a phenomenon which has been initially investigated by Hanin, O'Rourke, Kabluchko and Seidel, the author, and others. This thesis seeks to quantify root-and-critical-point pairing on several scales, including macroscopic comparisons between entire collections of roots and critical points, microscopic examinations of individual critical points that lie near fixed roots, and a "mesoscopic" local law to explain the situation at scales in between.

In Chapter 2, we show that for a deterministic point ξ lying outside the support of μ, almost surely the polynomial $q_n(z) := p_n(z)(z - \xi)$ has a critical point at distance $O(1/n)$ from ξ. In other words, conditioning the random polynomials p_n to have a root at ξ almost surely forces a critical point near ξ. More generally, we prove an analogous result for the critical points of $q_n(z) := p_n(z)(z - \xi_1) \cdots (z - \xi_k)$, where $\xi_1, \ldots, \xi_k$ are deterministic. In addition, when $k = o(n)$, we show that the empirical distribution constructed from the critical points of q_n converges to μ in probability as the degree tends to infinity, extending a result of Kabluchko.

In Chapter 3, under a regularity assumption, we show that if the roots of p_n are iid, 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.

Dedication

To my grandma, Kathy Williams, who struck the first cord, to my parents, Annis and Mark Williams, who stoke the fire, and to my sister, Anneli Williams, who will keep the flame.

Acknowledgements

I thank my advisor, Sean O'Rourke, for being an outstanding mentor. He is a brilliant mathematician, teacher, and collaborator, his explanations, foresight, and patience second to none.

I am also grateful to my Dissertation Committee: to Boris Hanin for many useful conversations and comments on this work and for suggesting the line of research in Chapter 3; and to Janos Englander, Sergei Kuznetsov, and Judith Packer for their advice, support, and wonderful classes.

Many others at CU have helped, inspired, and challenged me to grow. I thank Leo Herr (B.M.) and Sarah Arpin for their loyal friendship, unconventional perspectives, and studying tactics; Robert Hines for his answers to my incessant questions about math, life, and pants; and Carlos Pinilla for his eternal positivity, endurance, and love of reckless cooking. I am grateful to Faan Tone Liu and John Martin for showing me that the only way to teach is to be human, and to Preston Cumming and the GTP for supporting my development as an educator. Thanks also to my academic sister Natalie Coston, to Barbara Wojcik and the Math Department staff, and to my countless other faculty- and grad-student colleagues for cultivating a wonderful community here.

Rosin makes tight bonds; I thank Maya Fabrikant and Monica Carroll for the spiciest music, hiking, and baking adventures; Trubee Davison, Betsy Noecker, Tyler Mixa, and the Campus Orchestra crew for the hybrid holidays and lawn games; Tim Triggs and Mike Hendricks for showing me how to count from I to V on four fingers; and Bill Rumley, who helped me through the fretting.

Special thank-yous go to Manda Riehl, Alex Smith, and the rest of my UW-Eau Claire family for opening the door, and to Roger Pope and Rollie and Pam Hicks for showing me the power of pedals, picnics, and people.

Contents

Figures

Figure

Chapter 1

Introduction

Consider Figure 1.1, which depicts the zeros (red dots) and critical points (zeros of the derivative, blue crosses) of a random, degree-100 polynomial whose roots have been chosen independently and uniformly from the unit disk in the complex plane. At a glance, one perceives that dots and

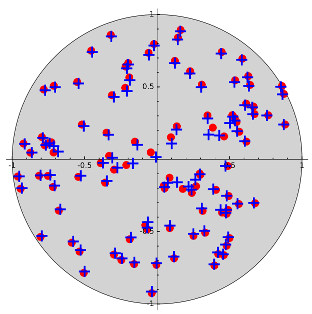


Figure 1.1: The roots (red dots) and critical points (blue crosses) of a random degree-100 polynomial, where all 100 roots have been chosen independently and uniformly from the unit disk.

crosses "pair-up," and upon re-creating the picture with n roots instead of 100, one discovers that the strength of the pairing increases with n (see Figure 1.2). Further investigations reveal that the pairing phenomenon is not unique to the unit disk but occurs when the roots are chosen from a wide range of distributions (see e.g. Figures 1.3, 1.4, and 1.5).

In the following thesis, we seek to quantitatively describe this behavior at several scales for random polynomials of the form

$$p_n(z) := \prod_{j=1}^{n}(z - X_j), \tag{1.1}$$

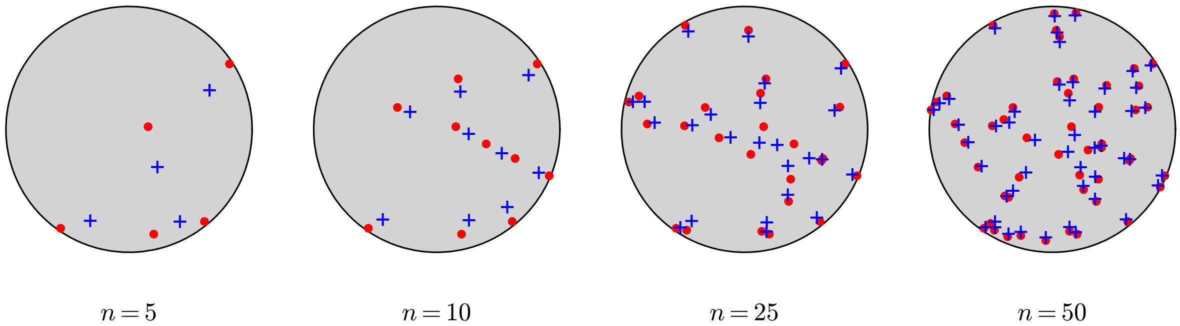


Figure 1.2: The roots (red dots) and critical points (blue crosses) of $p_n(z) = \prod_{j=1}^{n}(z - X_j)$, where the roots, $X_1, \ldots, X_{50}$ have been chosen independently and uniformly from the unit disk.

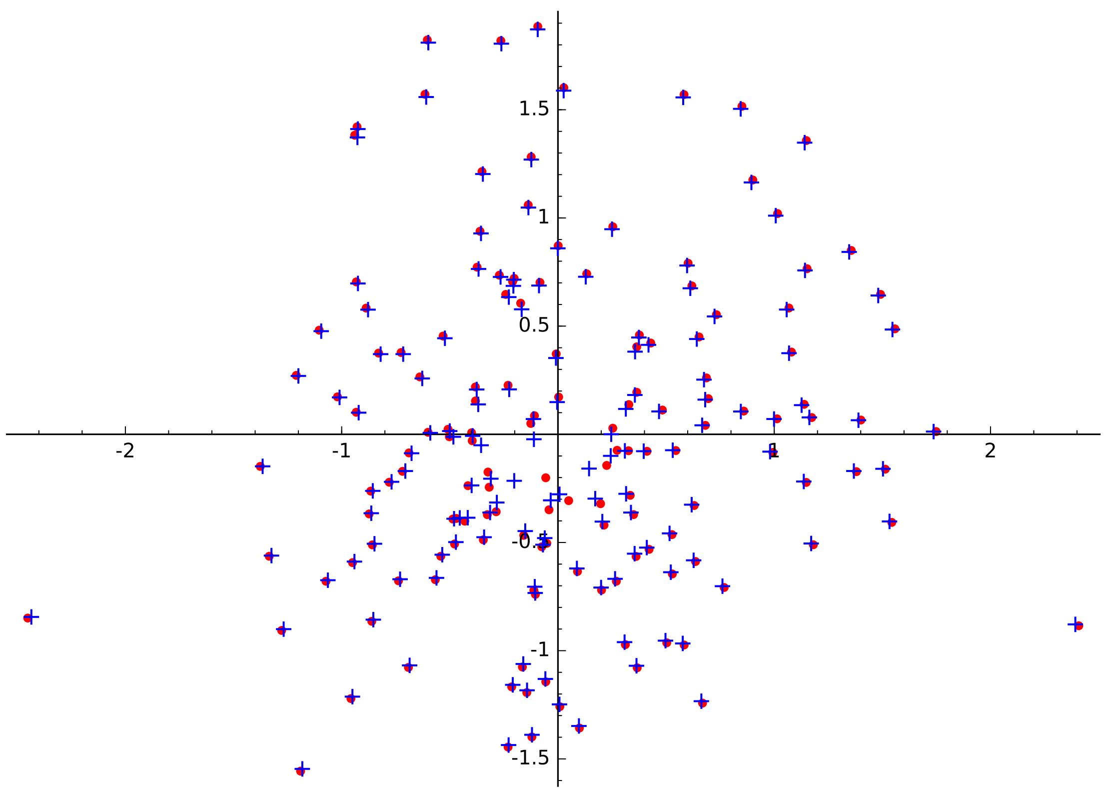


Figure 1.3: The roots (red dots) and critical points (blue crosses) of a random, degree 150 polynomial, where all 150 roots are chosen independently according to a standard complex normal distribution. See Example 3.7.

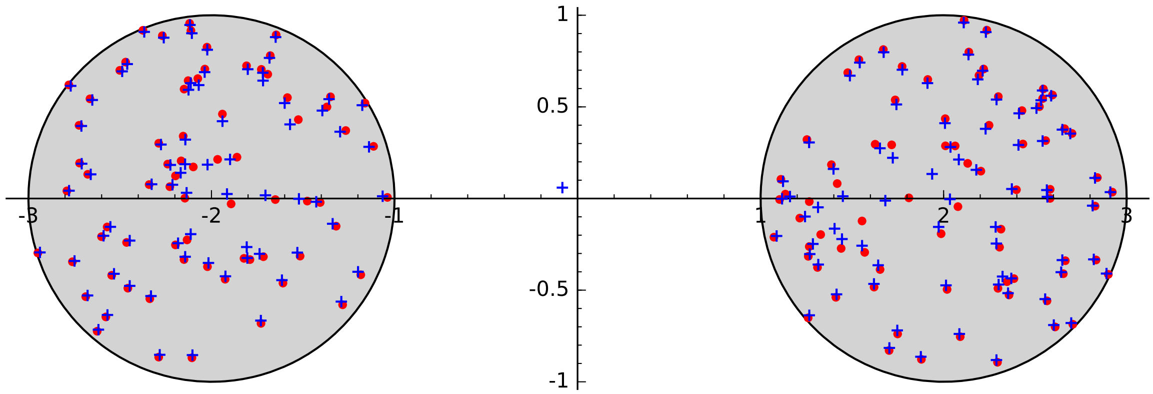


Figure 1.4: The roots (red dots) and critical points (blue crosses) of a random, degree 150 polynomial, where all 150 roots are chosen independently and uniformly from two disks. See Example 3.8.

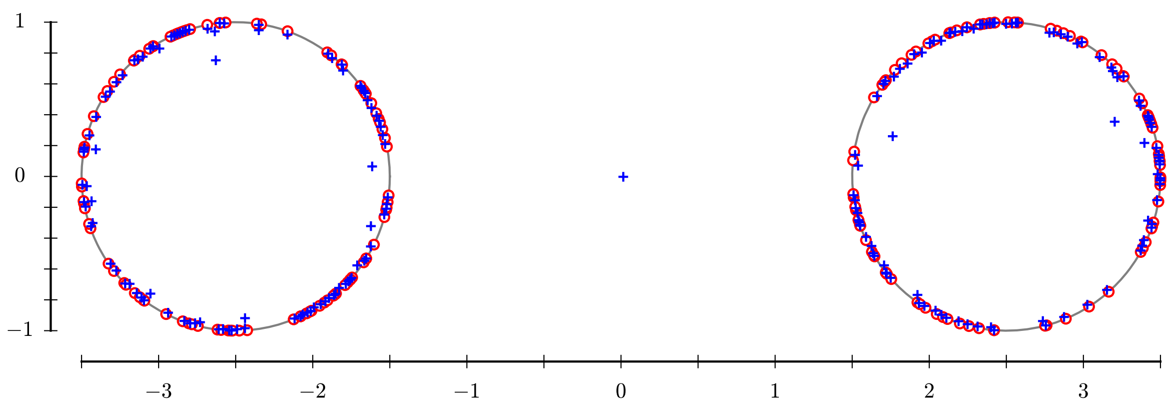


Figure 1.5: The roots (red circles) and critical points (blue crosses) of a random, degree 200 polynomial, where all 200 roots are chosen independently and uniformly from the union of the two unit circles (black curves) centered at $-5/2$ and $5/2$, respectively.

where $X_1, \ldots, X_n$ are complex-valued random variables (not necessarily independent or identically distributed). Our results include macroscopic comparisons between entire collections of roots and critical points, microscopic examinations of individual critical points that lie near fixed roots, and a "mesoscopic" local law to explain the situation at scales in between. The content throughout is joint work with Sean O'Rourke and has been adapted from the recent papers [40] and [39].

1.1 History

Understanding the critical points of polynomials with known roots has long been of interest, and there are many results that pertain to the situation where the roots are deterministic (see for example Marden's book [34], which discusses the Gauss–Lucas theorem and Walsh's two circle theorem among other things). One of the most famous examples is the Gauss–Lucas theorem, which offers a geometric connection between the roots of a polynomial and the roots of its derivative.

Theorem 1.1 (Gauss–Lucas; Theorem 6.1 from [34]). *If p is a non-constant polynomial with complex coefficients, then all zeros of p' belong to the convex hull of the set of zeros of p.*

An electrostatic interpretation of roots and critical points of complex polynomials illuminates Theorem 1.1 and its proof. Consider placing fixed electrical charges at the zeros of the polynomial p that repel a movable test charge according to a force that is inversely proportional to distance. The equilibrium points of the resulting electrical field (i.e. the places where a test charge would experience a net force of zero) are the critical points of p. Intuitively, any test charge placed outside the convex hull of the set of fixed charges is propelled to infinity, a heuristic that motivates the proof of Theorem 1.1. There are many refinements of Theorem 1.1, and we refer the reader to [2, 6, 11, 13, 14, 19, 27, 32, 33, 35, 42, 45, 46, 48, 49, 51, 59] and references therein. See also Steinerberger's recent work [50] that discusses a stability version of the Gauss–Lucas Theorem and pairing between roots and critical points of deterministic polynomials.

Motivated by discussions with Oded Schramm, Pemantle and Rivin initiated the probabilistic study of such relationships between critical points and roots [43]. They posed the following question:

for the random polynomial p_n defined in (1.1), when are the zeros of p_n' stochastically similar to the roots of p_n? In order to compare these two collections of points in their answer, Pemantle and Rivin used the language of weak convergence of random empirical measures. We introduce their notation now.

For a degree-n polynomial p, define the empirical measure constructed from the roots of p to be

$$\mu_p := \frac{1}{n} \sum_{z\in\mathbb{C}:p(z)=0} \delta_z,$$

where each root in the sum is counted with multiplicity and δ_z is the unit point mass at z. For the critical points of p, we introduce the notation

$$\mu_p' := \mu_{p'}.$$

In other words, μ_p' is the empirical measure constructed from the critical points of p. Note that when p is a random polynomial, μ_p becomes a random probability measure. One can asymptotically compare random probability measures using the following probabilistic notion of weak convergence.

Definition 1.2 (Weak convergence of random probability measures)**.** Let T be a topological space (such as $\mathbb{R}$ or $\mathbb{C}$), and let $\mathcal{B}$ be its Borel σ-field. Let $(\mu_n)_{n\geq 1}$ be a sequence of random probability measures on $(T, \mathcal{B})$, and let μ be a probability measure on $(T, \mathcal{B})$. We say μ_n **converges weakly to μ in probability** as $n \to \infty$ (and write $\mu_n \to \mu$ in probability) if for all bounded continuous $\varphi : T \to \mathbb{R}$ and any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\int \varphi d\mu_n - \int \varphi d\mu\right| > \varepsilon\right) = 0.$$

In other words, $\mu_n \to \mu$ in probability as $n \to \infty$ if and only if $\int \varphi d\mu_n \to \int \varphi d\mu$ in probability for all bounded continuous $\varphi : T \to \mathbb{R}$. Similarly, we say μ_n **converges weakly to μ almost surely** as $n \to \infty$ (and write $\mu_n \to \mu$ almost surely) if for all bounded continuous $\varphi : T \to \mathbb{R}$,

$$\lim_{n\to\infty} \int \varphi d\mu_n = \int \varphi d\mu$$

almost surely.

Pemantle and Rivin conjectured that when $X_1, X_2, \ldots$ 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 as n tends to infinity. They proved their conjecture in [43] for measures satisfying some technical assumptions, and Subramanian [52] refined their work for $X_1, X_2, \ldots$ on the unit circle. Kabluchko first proved the conjecture in full generality in [28] to obtain the following result.

Theorem 1.3 (Kabluchko; [28])**.** *Let μ be an arbitrary probability measure on $\mathbb{C}$, and let $X_1, X_2, \ldots$ be a sequence of iid random variables with distribution μ. For each $n \geq 1$, let p_n be the degree n polynomial given in* (1.1)*. Then μ'_{p_n} converges weakly to μ in probability as $n \to \infty$.*

Naturally, one may ask whether the assumptions in Theorem 1.3 (such as the roots $X_1, X_2, \ldots$ being independent) can be relaxed. O'Rourke established several versions of Theorem 1.3 for random polynomials with dependent roots that satisfy some technical conditions [37]. For example, the conclusion of Theorem 1.3 holds for characteristic polynomials of certain classes of matrices from the classical compact matrix groups (the eigenvalues of such matrices are known to not be independent). Similar results for characteristic polynomials of nearly Hermitian matrices were studied in [41, Section 2.5].

In [47], Reddy considers polynomials whose zeros are chosen randomly from two deterministic sequences of complex numbers in which the empirical measures for both sequences converge to the same limit. It is shown that the limiting empirical measures of the zeros and critical points agree for these polynomials, yielding a version of Theorem 1.3 where the randomness can be reduced and independence still remains. More recently, O'Rourke and the author [40] adapted Kabluchko's strategy to the situation where p_n is perturbed to have $o(n)$ deterministic roots; these results appear in Chapter 2 below. Byun, Lee, and Reddy [7] further generalized Kabluchko's theorem, showing that under some mild assumptions, the conclusion of Theorem 1.3 holds when p_n has mostly deterministic roots and several (potentially dependent) random ones. However, as the following example shows, the randomness in Theorem 1.3 cannot be completely eliminated (i.e., the theorem does not always hold for sequences of deterministic polynomials).

Example 1.4. Let $p_n(z) := z^n - 1$. Then the roots of p_n are the n-th roots of unity, and so μ_{p_n} converges weakly to the uniform measure on the unit circle as n tends to infinity. However, all $n-1$ critical points of p_n are located at the origin. Hence, $\mu'_{p_n} = \delta_0$ for all n.

We conclude this subsection by mentioning that in [7], Byun, Lee, and Reddy also proved several other 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 1.3 when the roots $X_1, \ldots, X_n$ are given by a $2D$ Coulomb gas density.

1.2 Local behavior

Theorem 1.3 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 1.3 with the Law of Large Numbers, one obtains that, for any bounded and continuous function $\varphi : \mathbb{C} \to \mathbb{C}$,

$$\sum_{j=1}^{n} \varphi(X_j) = \sum_{j=1}^{n-1} \varphi(w_j^{(n)}) + o(n) \tag{1.2}$$

with high probability.[1] In contrast to Theorem 1.3, this thesis primarily 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 1.3 and (1.2) describe this phenomenon at the macroscopic level by comparing the global behaviors of the critical points and roots. However, a glance at Figures 1.4 and 1.3 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 [21, 22, 23], including random polynomials with independent roots. He

[1] See Section 1.4 for a complete description of the asymptotic notation used here and in the sequel.

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 [23].

Recently, Kabluchko and Seidel determined the asymptotic fluctuations of the critical point of p_n that is nearest a given root [29]. 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 [29] and our work in Section 3.2.3 below.

1.3 Overview of Chapters 2 and 3

In the remaining chapters of this thesis, we refine the results mentioned above to obtain a more complete picture of the pairing that occurs between zeros and critical points of the polynomial p_n defined via (1.1).

Chapter 2 concerns pairing between zeros and critical points of p_n, where several roots $\xi_1, \ldots, \xi_k$ are fixed, non-random complex values, and the remaining X_j, $1 \leq j \leq n-k$, are iid with a common distribution μ. We show that when $k = 1$ and ξ_1 lies outside the support of μ, then almost surely the polynomial

$$q_n(z) := (z - \xi_1) \prod_{j=1}^{n-1} (z - X_j)$$

has a critical point at distance $O(1/n)$ from ξ_1. In other words, conditioning the random polynomial p_n to have a root at ξ_1 almost surely forces a critical point near ξ_1. More generally, we prove an analogous result for the critical points of

$$q_n(z) := (z - \xi_1) \cdots (z - \xi_k) \prod_{j=1}^{n-k} (z - X_j),$$

where $\xi_1, \ldots, \xi_k$ are deterministic. In addition, when $k = o(n)$, we show that the empirical distribution constructed from the critical points of q_n converges to μ in probability as the degree tends to infinity, extending Kabluchko's Theorem 1.3.

We begin Chapter 3 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 one to one pairing between roots and critical points in Figures 1.4 and 1.3, it does not allow one to describe the behavior near any particular root. We accomplish this feat in Section 3.2.3, 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 $X_1, \dots, X_n$ of p_n are iid, but for some of our results, we do not even require that the roots be independent (see Sections 3.2.3 and 3.2.4 for details).

1.4 Notation

Throughout the text, 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 $C_{P_1,P_2,\dots}$, to denote that the constant depends on some parameters $P_1, P_2, \dots$.

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 $\text{supp}(\mu)$ 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.

Chapter 2

Pairing between zeros and critical points of random polynomials with independent roots

2.1 Introduction to the chapter

This chapter is an adaptation of the article [40], a recently published collaboration between the author and Sean O'Rourke. There are several formatting differences between the content as it appears in [40] and in the sections below. In particular, the introductory material from [40] was assimilated into Chapter 1 above, and Appendix A from [40] appears in Section 2.4 below.

2.2 Main results

To introduce our results, we first consider the special case of the polynomial p_n, defined in (1.1), when $X_1, X_2, \ldots$ are iid with the common distribution μ that is the uniform probability measure on the unit circle centered at the origin. In this case, Theorem 1.3 implies that μ'_{p_n} converges weakly in probability to μ as $n \to \infty$. A numerical simulation of this result is shown in Figure 2.1; as can be seen, all critical points of p'_n lie very close to the unit circle. On the other hand, if we consider the polynomial $(z-\xi)p_n(z)$ for some deterministic point ξ outside the unit circle, we see in Figure 2.2 that one of the critical points leaves the unit disk and lies very close to ξ. However, the remaining critical points still lie close to the unit circle. The goal of this chapter is to describe the pairing between the root ξ and the nearby critical point. More generally, we consider the case when several deterministic zeros are appended to the random polynomial p_n and when μ is an arbitrary measure in the complex plane with compact support (not just the uniform

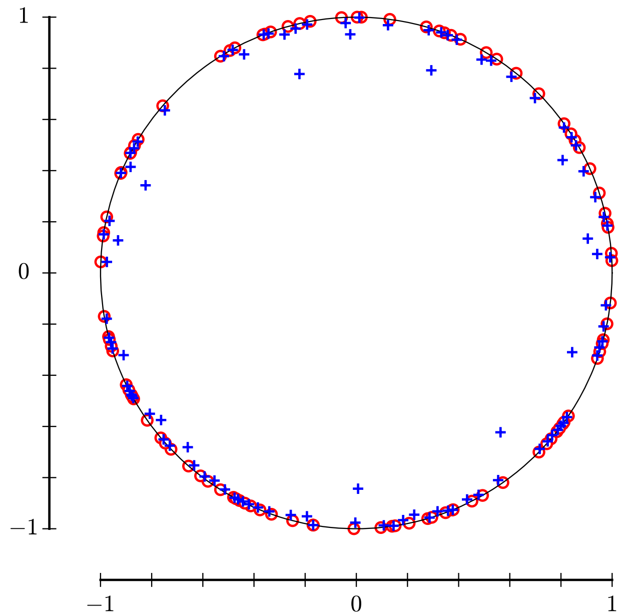


Figure 2.1: The roots (red circles) and critical points (blue crosses) of a random, degree 100 polynomial, where all 100 roots are chosen independently and uniformly from the unit circle (black curve).

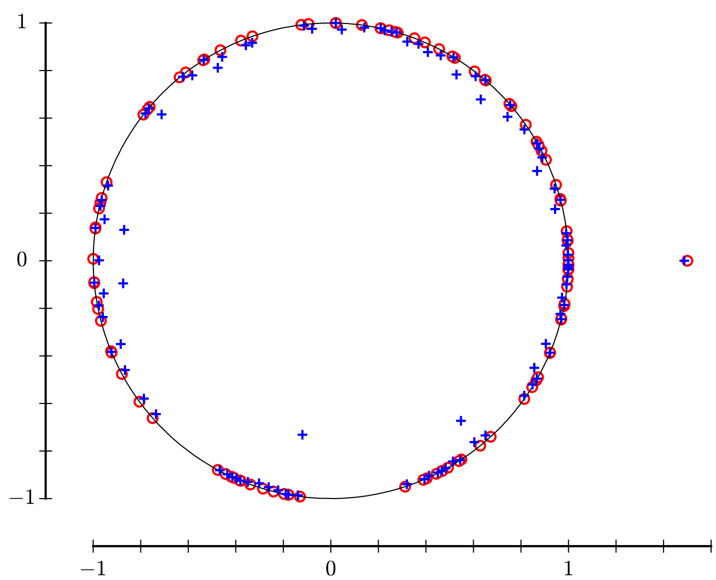


Figure 2.2: The roots (red circles) and critical points (blue crosses) of a random, degree 101 polynomial, where 100 roots are chosen independently and uniformly from the unit circle (black curve), and one root takes the deterministic value $\xi = 1.5$.

distribution on the unit circle). See, for example, Figures 2.3 and 2.4.

Let us mention that this pairing phenomenon between roots and critical points has been observed previously for random polynomials. Hanin [23] proves a similar pairing result when a number of deterministic roots are appended to a random polynomial whose roots are chosen independently from a probability measure μ supported on the Riemann sphere. Hanin's proof is guided by an intuitive electrostatic interpretation of the zeros and critical points. In contrast to many of our results, Hanin's proof works both when μ is supported on a compact subset and when μ is supported on the entire Riemann sphere. Unlike the results in [23] however, our results do not require the measure μ to have bounded density or require the deterministic roots to satisfy a separation condition. In addition, our methods are significantly different than those used in [23] and allow us to describe the exact number of critical points lying in a region outside the support of μ. In a separate paper [21], Hanin considers the joint distribution of roots and critical points for a class of Gaussian random polynomials. However, the polynomials considered in [21] are quite different than the model considered in this chapter. Finally, let us mention the work of Dennis and Hannay [12] from the physics literature, which gives an electrostatic explanation for the pairing of critical points and zeros of random polynomials and characteristic polynomials of random matrices.

2.2.1 Limiting distribution of the critical points

To begin, we first consider the analogue of Theorem 1.3 when $o(n)$ deterministic zeros are appended to the random polynomial p_n in (1.1).

Theorem 2.1 (Limiting distribution of the critical points)**.** *Let μ be an arbitrary probability measure on $\mathbb{C}$, and suppose $X_1, X_2, \dots$ are iid random variables with distribution μ. For each $n \geq 1$, let k_n be a deterministic non-negative integer no larger than n such that $k_n = o(n)$. In addition, let $\xi_1^{(n)}, \dots, \xi_{k_n}^{(n)}$ be a deterministic triangular array of complex values, and let*

$$p_n(z) := \prod_{j=1}^{n-k_n} (z - X_j) \prod_{l=1}^{k_n} (z - \xi_l^{(n)}).$$

Then μ'_{p_n} converges weakly to μ in probability as $n \to \infty$.

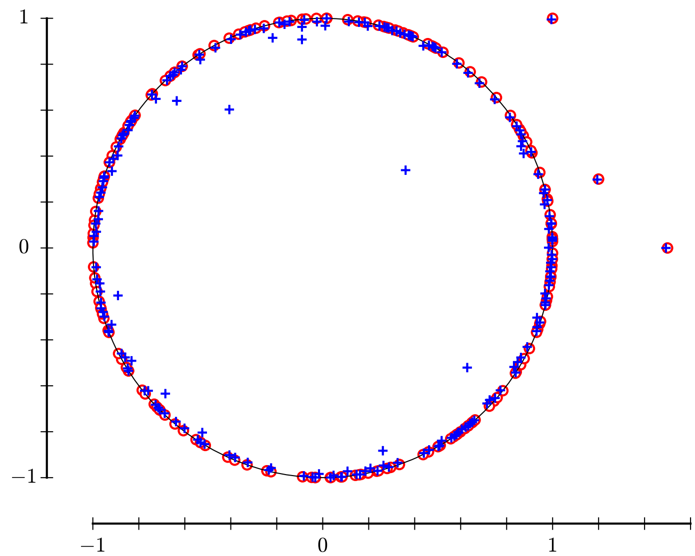


Figure 2.3: The roots (red circles) and critical points (blue crosses) of a random, degree 203 polynomial, where 200 roots are chosen independently and uniformly from the unit circle (black curve), and three roots take the deterministic values $\xi_1 = 1 + i$, $\xi_2 = 1.5$, and $\xi_3 = 1.2 + 0.3i$.

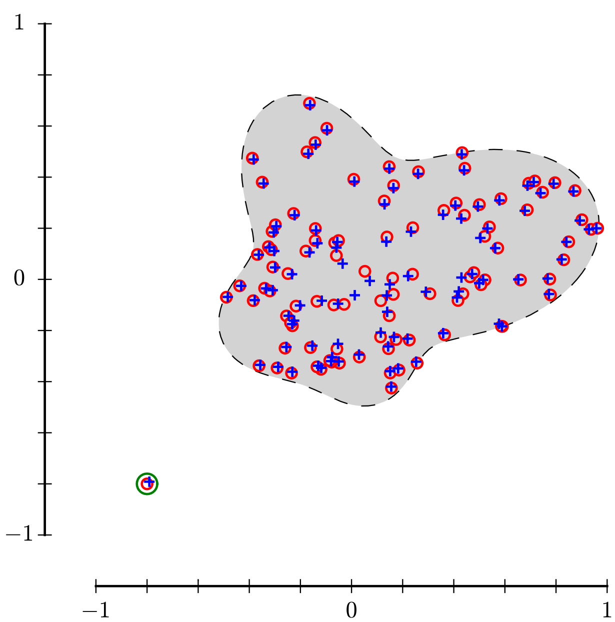


Figure 2.4: The roots (red circles) and critical points (blue crosses) of a random, degree $n = 100$ polynomial, where 99 roots are chosen independently and uniformly from the outlined region, and one root takes the deterministic value $\xi = -0.8 - 0.8i$. The small green circle centered at ξ that contains the critical point nearby has radius $4/n$.

Theorem 2.1 is a generalization of Theorem 1.3. Indeed, Theorem 1.3 can be recovered from Theorem 2.1 by taking $k_n = 0$. Unsurprisingly, we prove Theorem 2.1 in Section 2.4 by slightly generalizing the methods developed by Kabluchko in [28].

Let us discuss the intuition behind Theorem 2.1. To do so, we must begin with Theorem 1.3. Roughly speaking, Theorem 1.3 describes the phenomenon that if p_n is a degree n random polynomial, then

$$\mu_{p_n} - \mu'_{p_n} \longrightarrow 0 \tag{2.1}$$

in probability as $n \to \infty$. In other words, the limiting behavior of the critical points is the same as the limiting behavior of the roots. While Theorem 1.3 only applies to random polynomials with iid roots, the same phenomenon has been observed for other ensembles of random polynomials [37, 41], and numerical simulations show that it should be true for many other models. Stated another way, the behavior in (2.1) appears to be universal among random polynomials. Let us now consider the polynomial p_n from Theorem 2.1. It follows from the law of large numbers that $\mu_{p_n} \to \mu$ weakly almost surely as $n \to \infty$ since $k_n = o(n)$. Therefore, if the convergence in (2.1) applies to the polynomial p_n, the triangle inequality would immediately imply that μ'_{p_n} also converges weakly to μ in probability. This heuristic is the basis for our proof of Theorem 2.1.

The above heuristic also hints that the condition $k_n = o(n)$ in Theorem 2.1 is sharp. Indeed, if $\lceil \varepsilon n \rceil$ deterministic roots were to be appended, the limiting distribution is, in general, not μ as shown by the following example.

Example 2.2. Let $0 < \varepsilon < 1$ and $k_n := \lceil \varepsilon n \rceil$. Define

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

where $X_1, X_2, \ldots$ are iid random variables uniformly distributed on the unit circle centered at the origin in the complex plane. Then, by Theorem 1.3, μ'_{p_n} converges weakly to the uniform measure on the unit circle in probability as $n \to \infty$. However, the polynomial

$$q_n(z) := z^{k_n} p_n(z)$$

has at least $k_n - 1$ critical points at the origin. In particular, $\mu'_{q_n}(\{0\}) \geq \varepsilon/2$ for n sufficiently large. Among other things, this implies that μ'_{q_n} does not converge weakly to the uniform probability measure on the unit circle as $n \to \infty$.

While Theorem 2.1 shows that the global behavior of the critical points is unchanged by the addition of $o(n)$ deterministic roots, the addition of one or more deterministic roots can create a number of outlying critical points as illustrated in Figures 2.2 and 2.3. One way of viewing this phenomenon is to view the deterministic roots as a small perturbation of the original polynomial. This small perturbation is not enough to change the global distribution of the critical points; it may, however, as observed in the figures above, create a small number of outlying critical points. Our main results below describe these outliers.

2.2.2 No outlying critical points for the unperturbed model

Before we consider the perturbed model, we first consider the case when there are no deterministic roots. In this initial case, we want to determine exactly where the critical points of the random polynomial p_n, defined in (1.1), are located. This way, when we do append the small perturbation of deterministic roots, we will be able to tell exactly what effect the perturbation has had.

Let μ be a probability measure on $\mathbb{C}$, and suppose $X_1, \ldots, X_n$ are iid random variables with distribution μ. In view of the Gauss–Lucas theorem (Theorem 1.1), the roots of $p_n(z) := \prod_{j=1}^{n}(z - X_j)$, must lie in Conv(supp($\mu$)), the convex hull of the support of μ. However, as we discussed above in the case when μ is supported on the unit circle (shown in Figure 2.1), nearly all of the critical points appear near the support of μ, which is only a small subset of the convex hull. Thus, our goal is to determine the exact subset of Conv(supp(μ)) where the critical points will lie, with high probability. We do so in the theorem below. To define this set where the critical points are located, we will first need to introduce the Cauchy–Stieltjes transform.

Let μ be a probability measure on $\mathbb{C}$, and let m_μ be the Cauchy–Stieltjes transform of μ

defined by

$$m_\mu(z) := \int_{\mathbb{C}} \frac{d\mu(x)}{z-x}, \quad z \notin \operatorname{supp}(\mu).$$

Also, define

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

to be the set of zeros of m_μ. If μ has compact support, it turns out that $M_\mu \subset \operatorname{Conv}(\operatorname{supp}(\mu))$; see Proposition 2.22 for details. For $\varepsilon > 0$, we also define the set

$$\mathrm{N}_\mu(\varepsilon) := \{z \in \mathbb{C} : \operatorname{dist}(z, \operatorname{supp}(\mu) \cup M_\mu) < \varepsilon\}$$

to be the ε-neighborhood of $\operatorname{supp}(\mu) \cup M_\mu$. Here, $\operatorname{dist}(z, D) := \inf_{w \in D} |z - w|$ is the distance from $z \in \mathbb{C}$ to a set $D \subset \mathbb{C}$.

The following theorem shows that all critical points of p_n must lie inside $\mathrm{N}_\mu(\varepsilon)$ with high probability.

Theorem 2.3 (No outliers in the unperturbed model)**.** *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $X_1, \dots, X_n$ are iid random variables with distribution μ. Then, for every $\varepsilon > 0$, there exists $C, c > 0$ (depending only on μ and ε) such that, with probability at least $1 - Ce^{-cn}$, the polynomial $p_n(z) := \prod_{j=1}^n (z - X_j)$ has no critical points outside $\mathrm{N}_\mu(\varepsilon)$.*

Remark 2.4. By the Gauss–Lucas theorem (Theorem 1.1), the critical points of p_n must lie inside $\operatorname{Conv}(\operatorname{supp}(\mu))$. Thus, Theorem 2.3 actually reveals that, with high probability, p_n has no critical points outside $N_\mu(\varepsilon) \cap \operatorname{Conv}(\operatorname{supp}(\mu))$.

We now justify our choice of the set $\mathrm{N}_\mu(\varepsilon)$ as the correct location of the critical points. First, in the case that μ is degenerate, $p_n(z) = (z-a)^n$ for some $a \in \mathbb{C}$, which has critical point $z = a$ with multiplicity $n-1$. This example shows that clearly the critical points of p_n may lie in $\operatorname{supp}(\mu)$. The next example shows that the critical points can also be in a neighborhood of the zero set M_μ.

Example 2.5. Let $\mu := p\delta_a + (1-p)\delta_b$ for some $a, b \in \mathbb{C}$ with $a \neq b$ and $p \in (0,1)$, and assume $X_1, X_2, \dots$ are iid random variables with distribution μ. Then

$$p_n(z) := \prod_{j=1}^n (z - X_j) = (z-a)^\alpha (z-b)^\beta$$

for some non-negative integers α, β with $\alpha + \beta = n$. Almost surely, for n sufficiently large, $\alpha, \beta \geq 1$, and, in this case,

$$p_n'(z) = (z-a)^{\alpha-1}(z-b)^{\beta-1}\left(nz - \alpha b - \beta a\right).$$

Thus, by the law of large numbers, p_n has a critical point at

$$z = \frac{\alpha b}{n} + \frac{\beta a}{n} = pb + (1-p)a + o(1)$$

almost surely. On the other hand,

$$m_\mu(z) = \frac{p}{z-a} + \frac{1-p}{z-b}$$

has exactly one zero located at $z = pb + (1-p)a$.

By the Borel–Cantelli lemma, Theorem 2.3 immediately implies the following corollary.

Corollary 2.6. *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $X_1, X_2, \ldots$ are iid random variables with distribution μ. Fix $\varepsilon > 0$. Then, almost surely, for n sufficiently large, the polynomial $p_n(z) := \prod_{j=1}^n (z - X_j)$ has no critical points outside $N_\mu(\varepsilon)$.*

We conclude this subsection with two examples of Theorem 2.3 and Corollary 2.6.

Example 2.7. Let μ be the uniform distribution on the unit circle centered at the origin. A simple computation shows that

$$m_\mu(z) = \begin{cases} 0, & \text{if } |z| < 1, \\ \frac{1}{z}, & \text{if } |z| > 1, \end{cases}$$

and hence $M_\mu = \{z \in \mathbb{C} : |z| < 1\}$. Since $\mathrm{Conv}(\mathrm{supp}(\mu)) = \{z \in \mathbb{C} : |z| \leq 1\}$, Theorem 2.3 does not rule out the possibility of critical points in the disk $D_{1-\varepsilon} := \{z \in \mathbb{C} : |z| < 1 - \varepsilon\}$. This is not a limitation of Theorem 2.3 and is consistent with the results in [43], which imply that, with positive probability, $D_{1-\varepsilon}$ contains at least one critical point. More precisely, let $p_n(z) := \prod_{j=1}^n (z - X_j)$, where $X_1, X_2, \ldots$ are iid random variables with distribution μ. Then for any $0 < \varepsilon < 1$, there exists $\eta > 0$ (independent of n) such that p_n has a critical point in the disk $D_{1-\varepsilon}$ with probability at least η for all sufficiently large n. This follows from the determinantal structure described in [43, Theorem 2.5]. A numerical simulation of this example is shown in Figure 2.1.

Example 2.8. Let μ be the uniform distribution on the union of disjoint circles $C_1 \cup C_2$, where C_1 is the unit circle centered at $5/2$ and C_2 is the unit circle centered at $-5/2$. Then

$$m_\mu(z) = \begin{cases} \frac{4z}{4z^2-25}, & \text{if } |z-5/2| > 1 \text{ and } |z+5/2| > 1, \\ \frac{1}{2z+5}, & \text{if } |z-5/2| < 1, \\ \frac{1}{2z-5}, & \text{if } |z+5/2| < 1, \end{cases}$$

and $M_\mu = \{0\}$. Let $\varepsilon > 0$, and take $p_n(z) := \prod_{j=1}^n (z - X_j)$, where $X_1, X_2, \ldots$ are iid random variables with distribution μ. Then Corollary 2.6 guarantees that almost surely, for n sufficiently large, all critical points of p_n lie in the set

$$A_1 \cup A_2 \cup \{z \in \mathbb{C} : |z| < \varepsilon\},$$

where A_1 and A_2 are the annuli

$$A_1 := \{z \in \mathbb{C} : 1 - \varepsilon < |z - 5/2| < 1 + \varepsilon\}, \quad A_2 := \{z \in \mathbb{C} : 1 - \varepsilon < |z + 5/2| < 1 + \varepsilon\}.$$

A numerical simulation of this example is shown in Figure 1.5. In particular, the simulation depicts a single critical point near the origin, showing that critical points may lie in a neighborhood of the zero set M_μ. In fact, it follows from the law of large numbers and Walsh's two circle theorem (see, for example, [46, Theorem 4.1.1]) that, for any $0 < \varepsilon < 1/4$, almost surely, for n sufficiently large, there is exactly one critical point of p_n in the disk $\{z \in \mathbb{C} : |z| < 1 + \varepsilon\}$. Combined with Corollary 2.6, we conclude that almost surely this critical point must converge to the origin as n tends to infinity.

2.2.3 Locations of the outlying critical points in the perturbed model

We now consider the outlying critical points depicted in Figures 2.2 and 2.3. To do so, we will need the following notation. For a polynomial p of degree n, we let $w_1(p), \ldots, w_{n-1}(p)$ be the critical points of p counted with multiplicity.

Theorem 2.9 (Locations of the outlying critical points)**.** *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $X_1, X_2, \ldots$ are iid random variables with distribution μ. Let $k \geq 1$,* *and assume* $\xi_1, \ldots, \xi_k$ *are deterministic complex numbers (which do not depend on* n*); in addition, suppose there are* s *values* $\xi_1, \ldots, \xi_s$ *not in* $\operatorname{supp}(\mu) \cup M_\mu$*. Then, there exists* $\varepsilon_0 > 0$ *such that the following holds for any fixed* $0 < \varepsilon < \varepsilon_0$*. Almost surely, for* n *sufficiently large, there are exactly* s *critical points (counted with multiplicity) of the polynomial*

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

outside $N_\mu(\varepsilon)$*, and after labeling these critical points correctly,*

$$w_l(p_n) = \xi_l + o(1)$$

for each $1 \leq l \leq s$*.*

Theorem 2.9 describes exactly the phenomenon we observe in Figures 2.2 and 2.3. In particular, this theorem shows that each deterministic root outside $\operatorname{supp}(\mu) \cup M_\mu$ creates one outlying critical point, which is asymptotically close to the deterministic root.

For comparison, we provide the following example which shows that the conclusion of Theorem 2.9 fails for deterministic polynomials.

Example 2.10. Let $p_n(z) := z^{n-1} - 1$ and $q_n(z) := p_n(z)(z - 1/2)$. Then the roots of q_n are $(n-1)$-th roots of unity with an outlier at $z = 1/2$. However, we will show that q_n has no critical points near $z = 1/2$. Indeed,

$$q_n'(z) = nz^{n-1} - \frac{n-1}{2}z^{n-2} - 1,$$

and so the critical points are the solutions of

$$\frac{1}{n}q_n'(z) = z^{n-1} - \frac{1}{2}\frac{n-1}{n}z^{n-2} - \frac{1}{n} = 0.$$

For $|z| \leq 3/4$, we have

$$\left|z^{n-1} - \frac{1}{2}\frac{n-1}{n}z^{n-2}\right| \leq |z|^{n-1} + |z|^{n-2} \leq \frac{7}{4}\left(\frac{3}{4}\right)^{n-2} < \frac{1}{n}$$

for n sufficiently large. This implies that $q_n'(z) \neq 0$ for every $z \in \mathbb{C}$ with $|z| \leq 3/4$. Hence, for n sufficiently large, there are no critical points of q_n in the disk $\{z \in \mathbb{C} : |z| \leq 3/4\}$. More generally, this argument shows that for a fixed $\eta \in (0,1)$, there are no critical points of q_n in the disk $\{z \in \mathbb{C} : |z| \leq 1-\eta\}$ for sufficiently large n.

We next state two generalizations of Theorem 2.9. Both results deal with the case when the deterministic points $\xi_1, \ldots, \xi_k$ (as well as the integer k) are allowed to depend on n. Because the points can now depend on n, some additional technical assumptions are required. These technical assumptions are trivially satisfied when $\xi_1, \ldots, \xi_k$ do not depend on n. As such, Theorem 2.9 is actually a corollary of the following more general result.

Theorem 2.11 (Locations of the outlying critical points: dependence on n)**.** *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $X_1, X_2, \ldots$ are iid random variables with distribution μ. For each $n \geq 1$, let $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ be a triangular array of deterministic complex numbers with $k_n = O(1)$, and assume*

$$\max\{|\xi_1^{(n)}|, \ldots, |\xi_{k_n}^{(n)}|\} = O(1). \tag{2.2}$$

Fix $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no values of $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ in $\mathrm{N}_\mu(3\varepsilon) \setminus \mathrm{N}_\mu(\varepsilon)$ and there are s values $\xi_1^{(n)}, \ldots, \xi_s^{(n)}$ outside $\mathrm{N}_\mu(3\varepsilon)$. Then, almost surely, for n sufficiently large, there are exactly s critical points (counted with multiplicity) of the polynomial

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

outside $\mathrm{N}_\mu(2\varepsilon)$, and after labeling these critical points correctly,

$$w_l(p_n) = \xi_l^{(n)} + o(1)$$

for each $1 \leq l \leq s$.

The $O(1)$-magnitude assumption in (2.2) is required for our proof. However, we conjecture that this condition is not needed. In fact, in the case when $s = 1$, we can remove this assumption, and we obtain the following stronger result.

Theorem 2.12 (Locations of the outlying critical points: $s = 1$ case)**.** *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $X_1, X_2, \ldots$ are iid random variables with distribution μ. For* *each $n \geq 1$, let $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ be a triangular array of deterministic complex numbers with $k_n = O(1)$. Fix $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no values of $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ in $\mathrm{N}_\mu(3\varepsilon) \setminus \mathrm{N}_\mu(\varepsilon)$ and there is one value $\xi_1^{(n)}$ outside $\mathrm{N}_\mu(3\varepsilon)$. Then, almost surely, for n sufficiently large, there is exactly one critical point of the polynomial*

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

outside $\mathrm{N}_\mu(2\varepsilon)$, and after labeling the critical points correctly,

$$w_1(p_n) = \xi_1^{(n)} \left(1 + O\left(1/n\right)\right) + O(1/n). \tag{2.3}$$

Remark 2.13. If $\xi_1^{(n)} = O(1)$, then (2.3) implies that, almost surely,

$$w_1(p_n) = \xi_1^{(n)} + O(1/n).$$

More generally, if $\xi_1^{(n)} = o(n)$, Theorem 2.12 yields that, almost surely,

$$w_1(p_n) = \xi_1^{(n)} + o(1).$$

In other words, the location of the outlying critical point $w_1(p_n)$ is asymptotically close to the outlying root $\xi_1^{(n)}$.

Remark 2.14. In the case where $\xi_1^{(n)}$ lies at least a fixed distance away from the convex hull of the support of μ, the conclusion in (2.3) is a deterministic result (regardless of the asymptotic behavior of k_n). This can be deduced from Walsh's two-circle theorem (see [46, Theorem 4.1.1]).

We present a numerical simulation of Theorem 2.12 in Figure 2.4.

2.2.4 Outline

The rest of the chapter is devoted to the proof of our main results. In Section 2.3, we develop several tools we will need for the proofs. The proof of Theorem 2.1 is given in Section 2.4, and the proof of Theorem 2.3 is presented in Section 2.5. We prove Theorems 2.9, 2.11, and 2.12 in Section 2.6.

2.3 Mathematical tools

We present here some tools we will need to prove our main results.

2.3.1 Tools from probability theory

We will need the following complex-valued version of Hoeffding's inequality.

Lemma 2.15 (Hoeffding's inequality for complex-valued random variables)**.** *Let $Y_1, \ldots, Y_n$ be iid complex-valued random variables which satisfy $|Y_j| \leq K$ almost surely for some $K > 0$. Then there exist absolute constants $C, c > 0$ such that*

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

for every $t > 0$.

Proof. Let

$$S_n := \frac{1}{n}\sum_{j=1}^{n} Y_j - \frac{1}{n}\mathbb{E}\left[\sum_{j=1}^{n} Y_j\right].$$

If $|S_n| \geq t$, then $|\operatorname{Re}(S_n)| \geq t/\sqrt{2}$ or $|\operatorname{Im}(S_n)| \geq t/\sqrt{2}$. So, we have

$$\mathbb{P}(|S_n| \geq t) \leq \mathbb{P}(|\operatorname{Re}(S_n)| \geq t/\sqrt{2}) + \mathbb{P}(|\operatorname{Im}(S_n)| \geq t/\sqrt{2}).$$

The claim now follows from the classic (real-valued) version of Hoeffding's inequality (see [24]) since $|\operatorname{Re}(Y_j)| \leq K$ and $|\operatorname{Im}(Y_j)| \leq K$. $\square$

2.3.2 Nets

We introduce ε-nets as a convenient way to discretize a compact set.

Definition 2.16. Let X be a subset of $\mathbb{C}$, and $\varepsilon > 0$. A subset $\mathcal{N}$ of X is called an ε-net of X if every point $x \in X$ can be approximated within ε by some point $y \in \mathcal{N}$, i.e. so that $|x - y| \leq \varepsilon$.

For a finite set $\mathcal{N}$, we let $|\mathcal{N}|$ denote the cardinality of $\mathcal{N}$. We will need the following estimate for the size of an ε-net.

Lemma 2.17. *Let D be a compact subset of $\{z \in \mathbb{C} : |z| \leq M\}$ for some $M > 0$. Then, for every $\varepsilon > 0$, there is an ε-net $\mathcal{N}$ of D such that*

$$|\mathcal{N}| \leq \left(1 + \frac{4M}{\varepsilon}\right)^2.$$

Proof. Let $\mathcal{N}'$ be a maximal $\varepsilon/2$-separated subset of $S := \{z \in \mathbb{C} : |z| \leq M\}$. In other words, $\mathcal{N}'$ is such that $|x - y| \geq \varepsilon/2$ for all $x, y \in \mathcal{N}'$ with $x \neq y$, and no subset of S containing $\mathcal{N}'$ has this property. Such a set can always be constructed by starting with an arbitrary point in S and at each step selecting a point that is at least $\varepsilon/2$ distance away from those already selected. Since S is compact, this procedure will terminate after a finite number of steps.

The maximality property implies that $\mathcal{N}'$ is an $\varepsilon/2$-net of S. Indeed, otherwise there would exist $z \in S$ that is at least $\varepsilon/2$-far from all points in $\mathcal{N}'$. So $\mathcal{N}' \cup \{z\}$ would still be an $\varepsilon/2$-separated set, contradicting the maximality property above.

Moreover, the separation property implies that the balls of radii $\varepsilon/4$ centered at the points in $\mathcal{N}'$ are disjoint. In addition, all such balls lie in the ball of radius $M + \varepsilon/4$ centered at the origin. Comparing areas gives

$$|\mathcal{N}'| \left(\frac{\varepsilon}{4}\right)^2 \leq \left(M + \frac{\varepsilon}{4}\right)^2,$$

and hence

$$|\mathcal{N}'| \leq \left(1 + \frac{4M}{\varepsilon}\right)^2.$$

We now use $\mathcal{N}'$ to construct an ε-net of D. Indeed, we construct $\mathcal{N}$ iteratively using the following procedure. Let $(x_n)_{n=1}^N$ be an enumeration of the points in $\mathcal{N}'$, and set $\mathcal{N}_0 := \emptyset$. Given $\mathcal{N}_n$ for $0 \leq n \leq N - 1$, we construct $\mathcal{N}_{n+1}$ as follows:

(1) If the ball of radius $\varepsilon/2$ centered at x_{n+1} does not intersect D, then let $\mathcal{N}_{n+1} := \mathcal{N}_n$.

(2) If the ball of radius $\varepsilon/2$ centered at x_{n+1} does intersect D, let y_{n+1} be an element of the intersection and set $\mathcal{N}_{n+1} := \mathcal{N}_n \cup \{y_{n+1}\}$.

Now take $\mathcal{N} := \mathcal{N}_N$. By the procedure above, it follows that $|\mathcal{N}| \leq |\mathcal{N}'|$. It remains to show that $\mathcal{N}$ is an ε-net of D. Let $z \in D$. Since $D \subseteq S$, there exists $x \in \mathcal{N}'$ such that $|x - z| \leq \varepsilon/2$. This means that the ball of radius $\varepsilon/2$ centered at x intersects D. Thus, from the procedure above, there exists $y \in \mathcal{N}$ such that $|x-y| \leq \varepsilon/2$. Therefore, by the triangle inequality, $|z-y| \leq \varepsilon$. □

2.3.3 Tools from linear algebra

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 [30] (see [30, Lemma 5.7]). However, the same result was later rediscovered and significantly generalized by Cheung and Ng [10, 9].

Theorem 2.18 (Lemma 5.7 from [30]; Theorem 1.2 from [9]). *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*

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

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

Theorem 2.18 allows us to translate the problem of studying critical points to a problem involving the eigenvalues of certain matrices. For studying the eigenvalues of such matrices, we will need the following lemmata.

Lemma 2.19 (Block determinant). *Suppose $A, B, C,$ and D are matrices of dimension $n \times n$, $n \times m$, $m \times n$ and $m \times m$, respectively. If A is invertible, then*

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det(D - CA^{-1}B).$$

Proof. The conclusion follows immediately from the decomposition

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & 0 \\ C & I_m \end{pmatrix}\begin{pmatrix} I_n & A^{-1}B \\ 0 & D - CA^{-1}B \end{pmatrix},$$

where I_n and I_m are the identity matrices of dimension $n \times n$ and $m \times m$, respectively. A similar proof is given in [25, Section 0.8.5]. □

Lemma 2.20 (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 2.20 can be found in [3]; see also [25, Section 0.7.4] for a more general version of this identity known as the Sherman–Morrison–Woodbury formula. We will also require the following bound involving the difference of two determinants. For a matrix A, we let $\|A\|$ denote the spectral norm of A, i.e., $\|A\|$ is the largest singular value of A.

Lemma 2.21. *Let A and B be $k \times k$ matrices. If $\|A\|, \|B\| = O(1)$, then*

$$|\det(A) - \det(B)| \ll_k \|A - B\|.$$

Proof. By the Leibniz formula for the determinant, it follows that

$$\begin{aligned} |\det(A) - \det(B)| &= \left| \sum_{\sigma} \operatorname{sgn}(\sigma) \left(\prod_{i=1}^{k} A_{\sigma(i),i} - \prod_{i=1}^{k} B_{\sigma(i),i} \right) \right| \\ &\leq \sum_{\sigma} \left| \prod_{i=1}^{k} A_{\sigma(i),i} - \prod_{i=1}^{k} B_{\sigma(i),i} \right|, \end{aligned} \tag{2.4}$$

where the sums range over all permutations σ of $\{1, \ldots, k\}$ and $\operatorname{sgn}(\sigma)$ is the sign of the permutation σ. We now take advantage of the fact that the spectral norm of a matrix bounds the magnitude of each entry. In particular,

$$\sup_{1 \leq i,j \leq k} (|A_{ij}| + |B_{ij}|) \leq \|A\| + \|B\| = O(1)$$

and

$$\sup_{1 \leq i,j \leq k} |A_{ij} - B_{ij}| \leq \|A - B\|.$$

Thus, by multiple applications of the triangle inequality, we obtain

$$\left| \prod_{i=1}^{k} A_{\sigma(i),i} - \prod_{i=1}^{k} B_{\sigma(i),i} \right| \ll_k \|A - B\|$$

uniformly in σ. Combining this bound with (2.4) completes the proof. $\square$

2.3.4 Other tools

We collect here some additional tools and facts we will need. First, we note that if μ has compact support, then the convex hull of the support of μ is also a compact set; see [1, Corollary 5.33] for details.

The following proposition shows that the zero set of the Cauchy–Stieltjes transform of μ must lie inside the convex hull of the support of μ. It is a generalization of the Gauss–Lucas Theorem (Theorem 1.1) in the sense that Proposition 2.22 is precisely the Gauss–Lucas Theorem when μ is atomic.

Proposition 2.22. *Let μ be a probability measure on $\mathbb{C}$ with compact support. If $m_\mu(z) = 0$ for some $z \notin \operatorname{supp}(\mu)$, then $z \in \operatorname{Conv}(\operatorname{supp}(\mu))$.*

Proof. Let $S := \operatorname{Conv}(\operatorname{supp}(\mu))$, and define

$$\overline{S} := \{\overline{x} : x \in S\}.$$

Suppose $z \notin S$. Then

$$|m_\mu(z)| = |e^{i\theta} m_\mu(z)| \geq \left|\operatorname{Im}\left(e^{i\theta} m_\mu(z)\right)\right| = \left|\int_{\mathbb{C}} \frac{\operatorname{Im}\left(e^{i\theta}(\overline{z} - \overline{x})\right)}{|z-x|^2} d\mu(x)\right|$$

for any $\theta \in \mathbb{R}$. Since $\operatorname{supp}(\mu)$ is compact, it follows from [1, Corollary 5.33] that $\overline{S}$ is also compact. Thus, by the hyperplane separation theorem, there exists a pair of parallel lines, separated by a gap $\varepsilon > 0$, separating $\overline{S}$ and $\overline{z}$. Let θ be the angle these lines make with the real axis (if they do not meet the real axis take $\theta = 0$). Then $\operatorname{Im}\left(e^{i\theta}(\overline{z} - \overline{x})\right)$ is of the same sign for all $x \in S$ and

$$|\operatorname{Im}(e^{i\theta}(\overline{z} - \overline{x}))| \geq \varepsilon$$

for all $x \in S$. Thus, we obtain

$$|m_\mu(z)| \geq \varepsilon \int_{\operatorname{supp}(\mu)} \frac{d\mu(x)}{|z-x|^2}.$$

As $\operatorname{supp}(\mu)$ is compact, there exists $M > 0$ such that $|z - x| \leq M$ for all $x \in S$. Hence, we conclude that

$$|m_\mu(z)| \geq \varepsilon \frac{1}{M^2} > 0,$$

and the proof is complete. $\square$

We will also need the following observation concerning the translation of roots and critical points.

Proposition 2.23 (Translation of the critical points)**.** *Let p be a monic polynomial of degree n, and suppose $w_1, \ldots, w_{n-1}$ are the critical points of p counted with multiplicity. Then, for any $a \in \mathbb{C}$, the critical points of $q(z) := p(z-a)$ are $w_1 + a, \ldots, w_{n-1} + a$.*

Proof. Since p is a monic polynomial of degree n,

$$p'(z) = n \prod_{j=1}^{n-1} (z - w_j).$$

Thus,

$$q'(z) = p'(z-a) = n \prod_{j=1}^{n-1} (z - a - w_j),$$

and the claim follows. $\square$

2.4 Proof of Theorem 2.1

The proof of Theorem 2.1 presented here is modeled after Kabluchko's proof of [28, Theorem 1.1]. We note that Theorem 2.1 does not follow from the results in [28], and the notable difference between our proof and the one given in [28] is that we must control the additional contribution coming from the deterministic triangular array. For convenience, we use μ_n and μ_n' to mean μ_{p_n} and $\mu_{p_n'}$, respectively and define

$$\Xi := \bigcup_{n=1}^{\infty} \left\{ \xi_l^{(n)} : 1 \le l \le k_n \right\} \tag{2.5}$$

to be the collection of values present in the deterministic triangular array. We let λ represent Lebesgue measure on $\mathbb{C}$, and we denote the positive and negative parts of the real logarithm by

$$\log_- x := \begin{cases} |\log x|, & 0 \le x \le 1, \\ 0, & x \ge 1, \end{cases} \quad \text{and} \quad \log_+ x := \begin{cases} 0, & 0 \le x \le 1, \\ \log x, & x \ge 1, \end{cases}$$

for $x \in [0, \infty)$. We use the convention that $\log_-(0) := \infty$ so that $\log_-(\cdot)$ is a function taking values in the extended real line.

We prove Theorem 2.1 using the following result, which requires the deterministic array satisfy an additional assumption.

Theorem 2.24. *Under the same hypotheses as in Theorem 2.1 and with the additional assumption that there is a set E of Lebesgue measure zero for which $z \in \mathbb{C} \setminus E$ implies*

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{l=1}^{k_n} \log_- \left| z - \xi_l^{(n)} \right| = 0, \tag{2.6}$$

it follows that μ_n' converges weakly to μ in probability as $n \to \infty$.

Unfortunately, we cannot always guarantee that the deterministic array satisfies condition (2.6). To get around this issue, we will work on subsequences where the condition does hold; specifically, the proof of Theorem 2.1 will require the following corollary of Theorem 2.24.

Corollary 2.25. *Assume the same hypotheses as in Theorem 2.1 and, in addition, suppose μ_{n_m} is a subsequence of μ_n for which there is a set $E \subset \mathbb{C}$ of zero Lebesgue measure such that $z \in \mathbb{C} \setminus E$ implies*

$$\limsup_{m\to\infty} \frac{1}{n_m} \sum_{l=1}^{k_{n_m}} \log_- \left| z - \xi_l^{(n_m)} \right| = 0.$$

Then μ_{n_m}' converges weakly to μ in probability as $n \to \infty$.

Proof. We show that μ_{n_m} is a subsequence of a new sequence of random measures (modified from μ_n) for which condition (2.6) does hold. To this end, define the sequence $\tilde{k}_n$ by

$$\tilde{k}_n := \begin{cases} k_n, & \text{if } n = n_m \text{ for some } m \in \mathbb{N}, \\ 0, & \text{otherwise,} \end{cases}$$

and the random polynomial

$$\tilde{p}_n(z) := \prod_{j=1}^{n-\tilde{k}_n} (z - X_j) \prod_{l=1}^{\tilde{k}_n} (z - \xi_l^{(n)}).$$

Also let $\tilde{\mu}_n$ and $\tilde{\mu}'_n$ denote $\mu_{\tilde{p}_n}$ and $\mu_{\tilde{p}'_n}$, respectively. By construction, μ_{n_m} and μ'_{n_m} are subsequences of $\tilde{\mu}_n$ and $\tilde{\mu}'_n$, respectively. Now, $\tilde{k}_n = o(n)$, and for $z \in \mathbb{C} \setminus E$,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{l=1}^{\tilde{k}_n} \log_- \left| z - \xi_l^{(n)} \right| = \limsup_{m\to\infty} \frac{1}{n_m} \sum_{l=1}^{k_{n_m}} \log_- \left| z - \xi_l^{(n_m)} \right| = 0.$$

Thus, Theorem 2.24 implies that $\tilde{\mu}'_n$ converges weakly to μ in probability as $n \to \infty$. It follows that the subsequence μ'_{n_m} also converges to μ weakly in probability as $m \to \infty$. □

The following lemma will allow us to justify the use of Corollary 2.25.

Lemma 2.26. *Let μ_n be a sequence of random probability measures on $\mathbb{C}$, and suppose μ is a deterministic probability measure on $\mathbb{C}$. Then, μ_n converges weakly to μ in probability if and only if each subsequence of μ_n contains a further subsequence that converges weakly to μ in probability.*

Proof. Observe that, for each bounded and continuous function $f : \mathbb{C} \to \mathbb{R}$, the sequence $\int_{\mathbb{C}} f \, d\mu_n$ is a sequence of complex-valued random variables whose subsequences are of the form $\int_{\mathbb{C}} f \, d\mu_{n_m}$, where μ_{n_m} is a subsequence of μ_n. In addition, $\int f \, d\mu$ is a constant. Thus, the claim follows by applying Theorem 2.6 on page 20 of [4] to the random variables $\int_{\mathbb{C}} f \, d\mu_n$. □

We now prove Theorem 2.1 by way of Corollary 2.25 and Lemma 2.26. The proof of Theorem 2.24 is delayed until Section 2.4.1. Fix a subsequence μ'_{n_m} of μ'_n. We will show that there exists a further subsequence that converges weakly to μ in probability, which, by Lemma 2.26 would complete the proof of Theorem 2.1.

Clearly, μ_{n_m} is a subsequence of μ_n. If λ denotes Lebesgue measure on $\mathbb{C}$, then Markov's inequality implies that, for any $\varepsilon > 0$,

$$\begin{aligned}
\lambda\left(z \in \mathbb{C} : \frac{1}{n_m} \sum_{l=1}^{k_{n_m}} \log_- \left| z - \xi_l^{(n_m)} \right| \geq \varepsilon \right) &\leq \frac{1}{\varepsilon} \int_{\mathbb{C}} \frac{1}{n_m} \sum_{l=1}^{k_{n_m}} \log_- \left| z - \xi_l^{(n_m)} \right| \, d\lambda(z) \\
&= \frac{1}{\varepsilon \cdot n_m} \sum_{l=1}^{k_{n_m}} \int_{\mathbb{C}} \log_- \left| z - \xi_l^{(n_m)} \right| \, d\lambda(z) \\
&= \frac{k_{n_m}}{\varepsilon \cdot n_m} \int_{\mathbb{C}} \log_- |z| \, d\lambda(z).
\end{aligned}$$

The last expression tends to zero as $m \to \infty$ by the local integrability of the logarithm and the fact that $k_n = o(n)$. Thus, the sequence of functions

$$z \mapsto \frac{1}{n_m} \sum_{l=1}^{k_{n_m}} \log_- \left| z - \xi_l^{(n_m)} \right|$$

converges to zero in measure as $m \to \infty$. Among other things, this implies that there exists a subsequence of this sequence that converges to zero for almost every $z \in \mathbb{C}$ (see, for instance, Theorem 2.30 on page 61 of [17] for details). Let $\mu_{n_{m_j}}$ denote the corresponding subsequence of random measures. By Corollary 2.25, we have that $\mu'_{n_{m_j}}$ converges weakly to μ in probability as $j \to \infty$, completing the proof.

2.4.1 Proof of Theorem 2.24

It remains to prove Theorem 2.24. The proof presented here is modeled after the arguments given in [28]. The case where μ is degenerate is straightforward to establish by computing μ'_n explicitly and directly verifying that $\left| \int_{\mathbb{C}} f \, d\mu'_n - \int_{\mathbb{C}} f \, d\mu_n \right| \to 0$ almost surely as $n \to \infty$ for any bounded and continuous function $f : \mathbb{C} \to \mathbb{R}$. We now consider the case that μ is non-degenerate.

The proof of Theorem 2.24 will reduce to studying the logarithmic derivative L_n of p_n defined by the formula

$$L_n(z) := \frac{p'_n(z)}{p_n(z)} = \sum_{j=1}^{n-k_n} \frac{1}{z - X_j} + \sum_{l=1}^{k_n} \frac{1}{z - \xi_l^{(n)}}.$$

Specifically, Theorem 2.24 will follow from Lemma 2.28 below. We also now state a related lemma (Lemma 2.27), which we will need later. Note that these two lemmas are very similar to [28, Lemmas 2.1 and 2.2]; however, neither lemma follows directly from the results in [28] because of the deterministic contribution to L_n.

Lemma 2.27. *Under the assumptions of Theorem 2.24, there is a set $F \subset \mathbb{C}$ of Lebesgue measure zero such that if $z \in \mathbb{C} \setminus F$, then*

$$\frac{1}{n} \log |L_n(z)| \longrightarrow 0$$

in probability as $n \to \infty$.

Lemma 2.28. *Under the assumptions of Theorem 2.24, for any continuous, compactly supported function* $\varphi : \mathbb{C} \to \mathbb{R}$*, we have*

$$\frac{1}{n} \int_{\mathbb{C}} \log |L_n(z)| \, \varphi(z) \, d\lambda(z) \longrightarrow 0 \tag{2.7}$$

in probability as $n \to \infty$*. (Recall that* λ *denotes Lebesgue measure on* $\mathbb{C}$*.)*

We now prove Theorem 2.24 assuming Lemma 2.28. The key idea is the following formula (see, for instance, [26, Section 2.4.1]), which relates the integral in (2.7) to the measures μ_n and μ'_n. For any polynomial f that is not identically zero,

$$\frac{1}{2\pi} \Delta \log |f| = \sum_{z \in \mathbb{C} \,:\, f(z) = 0} \delta_z$$

in the distributional sense, where each root in the sum is counted with multiplicity. In other words, for any compactly supported, smooth function $\varphi : \mathbb{C} \to \mathbb{R}$, we have

$$\frac{1}{2\pi} \int_{\mathbb{C}} \log |f(z)| \, \Delta \varphi(z) \, d\lambda(z) = \sum_{z \in \mathbb{C} \,:\, f(z) = 0} \varphi(z).$$

From this relationship we obtain that, for any smooth, compactly supported function $\varphi : \mathbb{C} \to \mathbb{R}$,

$$\frac{1}{n} \sum_{z \in \mathbb{C} \,:\, p'_n(z) = 0} \varphi(z) - \frac{1}{n} \sum_{z \in \mathbb{C} \,:\, p_n(z) = 0} \varphi(z) = \frac{1}{2\pi n} \int_{\mathbb{C}} \log |L_n(z)| \, \Delta \varphi(z) \, d\lambda(z).$$

In view of Lemma 2.28, the integral on the right tends to zero in probability as $n \to \infty$. In addition, by the law of large numbers and the fact that $k_n = o(n)$,

$$\frac{1}{n} \sum_{z \in \mathbb{C} \,:\, p_n(z) = 0} \varphi(z) = \frac{1}{n} \sum_{j=1}^{n-k_n} \varphi(X_j) + \frac{1}{n} \sum_{l=1}^{k_n} \varphi(\xi_l^{(n)}) \longrightarrow \int_{\mathbb{C}} \varphi(z) \, d\mu(z)$$

almost surely as $n \to \infty$. Hence, for any smooth, compactly supported function $\varphi : \mathbb{C} \to \mathbb{R}$

$$\int_{\mathbb{C}} \varphi(z) \, d\mu'_n(z) = \frac{1}{n-1} \sum_{z \in \mathbb{C} \,:\, p'_n(z) = 0} \varphi(z) \longrightarrow \int_{\mathbb{C}} \varphi(z) \, d\mu(z)$$

in probability as $n \to \infty$. Since μ is a probability measure, we conclude from a simple approximation argument that μ'_n converges weakly to μ in probability. This completes the proof of Theorem 2.24.

2.4.2 Proof of Lemma 2.27

We now turn our attention to proving Lemmas 2.27 and 2.28. We begin with Lemma 2.27, which we will need to prove Lemma 2.28. First, we construct the exceptional set F described in Lemma 2.27 from several smaller subsets. The first of these, F_1, contains points where μ misbehaves, while another, F_2, includes values too close to the deterministic array. Define the set F_1 by

$$F_1 := \left\{ z \in \mathbb{C} : \int_{\mathbb{C}} \log_-^2 |z-y| \, d\mu(y) = \infty \right\}.$$

F_1 has Lebesgue measure zero since

$$\begin{aligned} \int_{\mathbb{C}} \left(\int_{\mathbb{C}} \log_-^2 |z-y| \, d\mu(y) \right) d\lambda(z) &= \int_{\mathbb{C}} \left(\int_{\mathbb{C}} \log_-^2 |z-y| \, d\lambda(z) \right) d\mu(y) \\ &= \int_{\mathbb{C}} \frac{\pi}{2} \, d\mu(y) = \frac{\pi}{2} < \infty \end{aligned}$$

by the Fubini–Tonelli theorem.

We now construct the subset F_2 by applying the Borel–Cantelli lemma. Recall that the set Ξ, defined in (2.5), is at most countable, and hence $\lambda(\Xi) = 0$. Thus, for a fixed $n \in \mathbb{N}$ and $1 \le l \le k_n$,

$$\begin{aligned} \lambda \left(z \in \mathbb{C} \setminus \Xi : \frac{1}{|z - \xi_l^{(n)}|} \ge e^{\sqrt{n}} \right) &= \lambda \left(z \in \mathbb{C} \setminus \Xi : \log_- |z - \xi_l^{(n)}| \ge \sqrt{n} \right) \\ &\le \frac{1}{n^3} \int_{\mathbb{C}} \log_-^6 |z - \xi_l^{(n)}| \, d\lambda(z) \\ &= \frac{C}{n^3} \end{aligned}$$

by Markov's inequality, where $C > 0$ is an absolute constant equal to the integral of $\log_-^6 |\cdot|$ over $\mathbb{C}$. Thus, we obtain

$$\sum_{n=1}^{\infty} \sum_{l=1}^{k_n} \lambda \left(z \in \mathbb{C} \setminus \Xi : \frac{1}{|z - \xi_l^{(n)}|} \ge e^{\sqrt{n}} \right) \le \sum_{n=1}^{\infty} \sum_{l=1}^{k_n} \frac{C}{n^3} = \sum_{n=1}^{\infty} \frac{C k_n}{n^3} < \infty$$

since $k(n) = o(n)$. It follows by the Borel–Cantelli lemma and the fact that Ξ is countable that there exists a set $F_2 \supset \Xi$ of Lebesgue measure zero such that, for every $z \in \mathbb{C} \setminus F_2$, $|z - \xi_l^{(n)}|^{-1} < e^{\sqrt{n}}$ for all but finitely many pairs (n, l). We conclude that, for $z \in \mathbb{C} \setminus F_2$,

$$\sum_{l=1}^{k_n} \frac{1}{|z - \xi_l^{(n)}|} = O_z(e^{2\sqrt{n}}), \tag{2.8}$$

where the asymptotic notation $O_z(\cdot)$ means the implicit constant is allowed to depend on z.

If we define F to be $F := E \cup F_1 \cup F_2$, then F has Lebesgue measure zero and, as we shall see, satisfies the requirements of Lemma 2.27. (Recall the definition of E from the statement of Theorem 2.24 above.) Notice that F contains the atoms of μ and the values in the deterministic triangular array.

Lemma 2.29. *For every $z \in \mathbb{C} \setminus F$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log |L_n(z)| \leq 0$$

almost surely.

Proof. Fix $z \in \mathbb{C} \setminus F$, and let $\varepsilon > 0$ be given. By Markov's inequality, for any $n \in \mathbb{N}$, we have

$$\begin{aligned} \mathbb{P}\left(\frac{1}{|z - X_n|} \geq e^{\varepsilon n} \right) &= \mathbb{P}\left(\log_- |z - X_n| \geq \varepsilon n \right) \\ &\leq \frac{\mathbb{E}\left[\log_-^2 |z - X_n| \right]}{\varepsilon^2 n^2} \\ &= \frac{1}{\varepsilon^2 n^2} \int_{\mathbb{C}} \log_-^2 |z - y| \, d\mu(y) \\ &= \frac{C_1}{\varepsilon^2 n^2}, \end{aligned}$$

for a non-negative constant C_1 since $z \notin F_1$. Hence,

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{1}{|z - X_n|} \geq e^{\varepsilon n} \right) < \infty,$$

so the Borel–Cantelli lemma applies. In particular, almost surely $\frac{1}{|z - X_n|} < e^{\varepsilon n}$ for all but finitely many n. Furthermore, z is not an atom of μ, so we have almost surely that, for all n,

$$|L_n(z)| \leq W + (n - k_n) e^{\varepsilon n} + \sum_{l=1}^{k_n} \frac{1}{|z - \xi_l^{(n)}|},$$

where W is an almost surely finite random variable. Now, since $z \in \mathbb{C} \setminus F_2$, the bound in (2.8) implies that, for n sufficiently large,

$$|L_n(z)| \leq W + n e^{\varepsilon n} + C_2 e^{2\sqrt{n}} \leq e^{2\varepsilon n}$$

for a positive constant C_2 (depending on z). It follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log |L_n(z)| \leq 2\varepsilon$$

almost surely. Since $\varepsilon > 0$ was arbitrary, the proof is complete. □

The reverse inequality in Lemma 2.27 requires an anti-concentration result that can be found, for example, in [44, Theorem 2.22 on page 76]. Before stating the lemma, we define the Lévy concentration function of a complex-valued random variable.

Definition 2.30 (Lévy concentration function)**.** Let Z be a complex-valued random variable. The **Lévy concentration function** of Z is defined as

$$\mathcal{L}(Z,t) := \sup_{u\in\mathbb{C}} \mathbb{P}\left(|Z-u| \leq t\right)$$

for all $t \geq 0$.

The Lévy concentration function bounds the **small ball probabilities** for Z, which are the probabilities that Z falls in a ball of radius t.

Lemma 2.31 (Anti-concentration estimate)**.** *Suppose that $Z_1, \ldots, Z_n$ are iid, non-degenerate, complex-valued random variables. Then, there is a positive constant C (depending only on the distribution of Z_1), so that, for any $t \geq 0$,*

$$\mathcal{L}\left(Z_1 + \cdots + Z_n, t\right) \leq C\frac{1+t}{\sqrt{n}} \tag{2.9}$$

for all $n \geq 1$.

Proof. Theorem 2.22 on page 76 in [44] implies that equation (2.9) holds when $Z_1, \ldots, Z_n$ are iid real-valued random variables and the supremum in the concentration function is taken over real numbers (see also [38, Corollary 6.8] for a more general version of this inequality). We extend this to the complex case in the following way. By assumption, $Z_1, \ldots, Z_n$ are iid and non-degenerate, so at least one of the real-valued random variables $\mathrm{Re}(Z_1)$ or $\mathrm{Im}(Z_1)$ is non-degenerate. Without loss of generality, assume $\mathrm{Re}(Z_1)$ is non-degenerate. Then

$$\begin{aligned}\mathcal{L}(Z_1 + \cdots + Z_n, t) &= \sup_{u \in \mathbb{C}} \mathbb{P}\left(|Z_1 + \cdots + Z_n - u| \leq t\right) \\ &\leq \sup_{u \in \mathbb{C}} \mathbb{P}\left(|\mathrm{Re}(Z_1) + \cdots + \mathrm{Re}(Z_n) - \mathrm{Re}(u)| \leq t\right) \\ &= \sup_{u \in \mathbb{R}} \mathbb{P}\left(|\mathrm{Re}(Z_1) + \cdots + \mathrm{Re}(Z_n) - u| \leq t\right).\end{aligned}$$

The last expression is bounded by $C\frac{1+t}{\sqrt{n}}$, for some constant C that depends only on the distribution of $\mathrm{Re}(Z_1)$ by the previously mentioned result in [44]. A nearly identical argument applies if $\mathrm{Re}(Z_1)$ is degenerate and $\mathrm{Im}(Z_1)$ is non-degenerate. □

Lemma 2.32. *For every $z \in \mathbb{C} \setminus F$ and every $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{n}\log|L_n(z)| \leq -\varepsilon\right] = 0.$$

Proof. Since $k_n = o(n)$, we assume n is sufficiently large so that $k_n < n$. Fix $z \in \mathbb{C} \setminus F$, and let $\varepsilon > 0$ be given. Since μ is non-degenerate and z is not an atom of μ, it follows that $\frac{1}{z-X_1}, \frac{1}{z-X_2}, \ldots$ are iid, non-degenerate, complex-valued random variables satisfying the hypotheses of Lemma 2.31. By absorbing the contribution of $\sum_{l=1}^{k_n}(z - \xi_l^{(n)})^{-1}$ into the complex number u in the definition of the concentration function, we conclude from Lemma 2.31 that

$$\mathbb{P}\left(|L_n(z)| \leq e^{-\varepsilon n}\right) \leq \mathcal{L}\left(\sum_{j=1}^{n-k_n} \frac{1}{z - X_j},\ e^{-\varepsilon n}\right) \leq C\frac{1 + e^{-\varepsilon n}}{\sqrt{n - k_n}}$$

for a positive constant C depending only on the distribution of $\frac{1}{z-X_1}$. As $n \to \infty$, the right-hand side goes to zero (since $k_n = o(n)$), which completes the proof. □

Together, Lemmas 2.29 and 2.32 establish Lemma 2.27.

2.4.3 Proof of Lemma 2.28

In this section, we prove Lemma 2.28 by way of the following dominated convergence result due to Tau and Vu [55].

Lemma 2.33 (Tao–Vu; Lemma 3.1 in [55])**.** *Let $(X, \mathcal{A}, \nu)$ be a finite measure space, and let $f_1, f_2, \ldots : X \to \mathbb{R}$ be random functions which are defined over a probability space $(\Omega, \mathcal{B}, \mathbb{P})$ and are jointly measurable with respect to $\mathcal{A} \otimes \mathcal{B}$. Assume that*

(i) for ν-a.e. $x \in X$ we have $f_n(x) \to 0$ in probability, as $n \to \infty$,

(ii) for some $\delta > 0$, the sequence $\int_X |f_n(x)|^{1+\delta}\, d\nu(x)$ is tight.

Then $\int_X f_n(x)\, d\nu(x)$ converges in probability to 0.

In order to prove Lemma 2.28, we will apply Lemma 2.33 to the random functions $f_n(z) := \frac{1}{n}(\log|L_n(z)|)\varphi(z)$, where φ is a continuous function with compact support. Lemma 2.27 establishes the first condition, and the tightness condition (with $\delta = 1$) follows from the next lemma. For the remainder of the paper, we let

$$\mathbb{D}_R := \{z \in \mathbb{C} : |z| < R\}$$

denote the open disk of radius $R > 0$ centered about the origin. Fix $r > 0$ such that the support of φ is contained in the open disk $\mathbb{D}_r$. We will occasionally use $\mathbb{1}_{\mathbb{D}_r}$ to denote the indicator function of the set $\mathbb{D}_r$.

Lemma 2.34. *The sequence $\frac{1}{n^2} \int_{\mathbb{D}_r} \log^2 |L_n(z)|\, d\lambda(z)$ is tight.*

In view of Lemma 2.33, the proof of Lemma 2.28 reduces to establishing Lemma 2.34. We bound the integral in Lemma 2.34 by employing the Poisson–Jensen formula as in [28]. In order to do so, we will need a uniform bound on $|L_n(z)|$ for z of certain magnitudes, which is the content of the following lemma.

Lemma 2.35. *There is an exceptional set $G \subset (0, \infty)$ of Lebesgue measure zero such that, for any $R \in (0, \infty) \setminus G$, we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log \sup_{|z|=R} |L_n(z)| \leq 0 \tag{2.10}$$

almost surely.

Proof. The proof is similar in spirit to that of Lemma 2.29. We first claim that

$$\sup_{|z|=R} \frac{1}{|z-X|} \geq K \iff ||X|-R| \leq \frac{1}{K}, \tag{2.11}$$

for any $X \in \mathbb{C}$, $R \in (0,\infty) \setminus \{|X|\}$, and $K > 0$. This equivalence will allow us to employ the method of Lemma 2.29 and control the behavior of $\log_- ||X_n| - R|$. To establish the forward direction of (2.11), observe that

$$0 < ||X| - R| = ||X| - |z|| \leq |X - z|$$

for any z satisfying $|z| = R$. Hence, $\sup_{|z|=R} |X - z|^{-1} \geq K$ implies $||X| - R| \leq K^{-1}$. On the other hand, if $||X| - R| \leq K^{-1}$, write $X = \rho e^{i\theta}$ in polar coordinates, and note that $z^* := Re^{i\theta}$ has modulus $R \neq \rho$ and satisfies

$$0 < |z^* - X| = |R - \rho| = ||X| - R| \leq \frac{1}{K}.$$

The fact that $\sup_{|z|=R} |X - z|^{-1} \geq K$ follows.

We are ready to construct G from two exceptional sets G_1 and G_2. Define

$$G_1 := \left\{ R \in (0,\infty) : \int_{\mathbb{C}} \log_-^2 ||y| - R| \; d\mu(y) = \infty \right\}.$$

It follows from the Fubini–Tonelli theorem that G_1 has Lebesgue measure zero since

$$\int_{\mathbb{R}} \int_{\mathbb{C}} \log_-^2 ||y| - R| \; d\mu(y) \, dR = \int_{\mathbb{C}} \int_{\mathbb{R}} \log_-^2 ||y| - R| \; dR \, d\mu(y) = \int_{\mathbb{C}} 2 \, d\mu(y) = 2 < \infty.$$

We now construct G_2. Let $\lambda_{\mathbb{R}}$ denote Lebesgue measure on the real line, and let

$$\Xi_{\mathbb{R}} := \bigcup_{n=1}^{\infty} \left\{ |\xi_l^{(n)}| : 1 \leq l \leq k_n \right\}.$$

Clearly, $\lambda_{\mathbb{R}}(\Xi_{\mathbb{R}}) = 0$. Equivalence (2.11) and Markov's inequality imply that for a fixed $n \in \mathbb{N}$ and

$1 \le l \le k_n$,

$$\begin{aligned}
\lambda_{\mathbb{R}} & \left(R \in (0,\infty) \setminus \Xi_{\mathbb{R}} : \sup_{|z|=R} \frac{1}{|z-\xi_l^{(n)}|} \ge e^{\sqrt{n}} \right) \\
&= \lambda_{\mathbb{R}} \left(R \in (0,\infty) \setminus \Xi_{\mathbb{R}} : \log_- ||\xi_l^{(n)}| - R| \ge \sqrt{n} \right) \\
&\le \frac{1}{n^3} \int_{[0,\infty)} \log_-^6 ||\xi_l^{(n)}| - R| \, dR \\
&\le \frac{1}{n^3} \int_{\mathbb{R}} \log_-^6 |R| \; dR \\
&= \frac{C}{n^3},
\end{aligned}$$

where $C > 0$ is an absolute constant. It follows that

$$\sum_{n=1}^{\infty} \sum_{l=1}^{k_n} \lambda_{\mathbb{R}} \left(R \in (0,\infty) \setminus \Xi_{\mathbb{R}} : \sup_{|z|=R} \frac{1}{|z-\xi_l^{(n)}|} \ge e^{\sqrt{n}} \right) \le \sum_{n=1}^{\infty} \frac{Ck_n}{n^3} < \infty,$$

so the Borel–Cantelli lemma and the countability of $\Xi_{\mathbb{R}}$ show that outside of a set $G_2 \supset \Xi_{\mathbb{R}}$ of Lebesgue measure zero,

$$\sup_{|z|=R} \frac{1}{|z-\xi_l^{(n)}|} < e^{\sqrt{n}}$$

for all but finitely many pairs (n, l). Hence, for $R \in (0,\infty) \setminus G_2$,

$$\sum_{l=1}^{k_n} \sup_{|z|=R} \frac{1}{|z-\xi_l^{(n)}|} < C_R + k_n e^{\sqrt{n}} = O_R(e^{2\sqrt{n}}), \tag{2.12}$$

where C_R is a positive constant depending on R. (Note that since $\Xi_{\mathbb{R}} \subset G_2$, $\sup_{|z|=R} |z-\xi_l^{(n)}|^{-1} < \infty$ for each pair (n, l)). If we define $G = G_1 \cup G_2$, then, $G \subset (0,\infty)$ has Lebesgue measure zero, and for $R \in (0,\infty) \setminus G$, we have that, for any $n \in \mathbb{N}$ and any $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P} \left(\sup_{|z|=R} \frac{1}{|z-X_n|} \ge e^{\varepsilon n} \right) &= \mathbb{P} \left(\log_- ||X_n| - R| \ge \varepsilon n \right) \\
&\le \frac{1}{\varepsilon^2 n^2} \mathbb{E}[\log_-^2 ||X_n| - R|] \\
&= \frac{C'_R}{n^2},
\end{aligned}$$

where we used (2.11) in the first step and Markov's inequality in the second. Here, C'_R is a positive constant depending only on R and μ. By the Borel–Cantelli lemma, it follows that almost surely,

$\sup_{|z|=R} \frac{1}{|z-X_n|} < e^{\varepsilon n}$ for all but finitely many n. This guarantees that for $R \in (0,\infty) \setminus G$, there is an almost surely bounded, real-valued random variable W_R for which

$$\sup_{|z|=R} |L_n(z)| \leq W_R + (n-k_n)e^{\varepsilon n} + \sum_{l=1}^{k_n} \sup_{|z|=R} \frac{1}{|z-\xi_l^{(n)}|} \leq e^{2\varepsilon n}$$

almost surely. (Note that $\mathbb{P}(|X_n| = R) = 0$ for all $R \in (0,\infty) \setminus G$ by the definition of the set G_1.) The last inequality holds for all sufficiently large n by (2.12). As $\varepsilon > 0$ was arbitrary, (2.10) now follows. $\square$

We now use the Poisson–Jensen formula to re-write $\log |L_n(z)|$. For any $R > r$ and $n \in \mathbb{N}$, let

$$y_1^{(n)}, \ldots, y_{s_n}^{(n)} \quad \text{and} \quad w_1^{(n)}, \ldots, w_{t_n}^{(n)}$$

be the roots and critical points, respectively, of p_n that are located in the open disk $\mathbb{D}_R$. The Poisson–Jensen formula (see, for example, [36, Chapter II.8]) implies that for any $z \in \mathbb{D}_R$ which is not a zero or pole of L_n,

$$\log |L_n(z)| = I_n(z;R) + \sum_{t=1}^{t_n} \log \left| \frac{R\left(z - w_t^{(n)}\right)}{R^2 - \overline{w_t^{(n)}}\, z} \right| - \sum_{s=1}^{s_n} \log \left| \frac{R\left(z - y_s^{(n)}\right)}{R^2 - \overline{y_s^{(n)}}\, z} \right|, \tag{2.13}$$

where

$$I_n(z;R) := \frac{1}{2\pi} \int_0^{2\pi} \log \left| L_n(Re^{i\theta}) \right| P_R(|z|, \theta - \arg z)\, d\theta,$$

and P_R denotes the Poisson kernel

$$P_R(\rho, \alpha) := \frac{R^2 - \rho^2}{R^2 + \rho^2 - 2R\rho\cos\alpha}, \quad \rho \in [0,R],\ \alpha \in [0, 2\pi]. \tag{2.14}$$

Lemma 2.36. *There exists an $R \geq \max\{1, 3r\}$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{z\in\mathbb{D}_r} I_n(z;R) \leq 0 \tag{2.15}$$

almost surely.

Proof. Fix $z \in \mathbb{D}_r$. Then, for any $\alpha \in [0, 2\pi]$ and $R \geq 3r$, we have

$$P_R(|z|, \alpha) = \frac{R^2 - |z|^2}{R^2 + |z|^2 - 2R|z|\cos\alpha} \leq \frac{(R+|z|)(R-|z|)}{R^2 + |z|^2 - 2R|z|} = \frac{R+|z|}{R-|z|} \leq 2. \tag{2.16}$$

The last inequality follows from the fact that $|z| \leq r$ and from the equivalence

$$\frac{R+r}{R-r} \leq 2 \quad \Longleftrightarrow \quad R \geq 3r,$$

which holds for all $R > r > 0$. Consequently, for any $z \in \mathbb{D}_r$ and $R \geq 3r$,

$$\begin{aligned}
\frac{1}{n} I_n(z;R) &\leq \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{n} \log \left| L_n(Re^{i\theta}) \right| \cdot 2 \, d\theta \\
&\leq \frac{1}{\pi} \int_0^{2\pi} \frac{1}{n} \log \sup_{|w|=R} |L_n(w)| \, d\theta \\
&= \frac{2}{n} \log \sup_{|w|=R} |L_n(w)| \, .
\end{aligned}$$

Therefore, we obtain

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{z\in\mathbb{D}_r} I_n(z;R) \leq \limsup_{n\to\infty} \frac{2}{n} \log \sup_{|w|=R} |L_n(w)| \, . \tag{2.17}$$

The desired result now follows by applying Lemma 2.35 to (2.17). In particular, since the exceptional set $G \subset (0,\infty)$ of Lemma 2.35 has measure zero, we can choose $R \geq \max\{1, 3r\}$ so that (2.15) holds almost surely. □

Next, we show that $I_n(z;R)$ is bounded below uniformly for $z \in \mathbb{D}_r$. We assume that $0 \notin F$, and we first consider the case when $z = 0$. There is no loss of generality in assuming $0 \notin F$, for if $0 \in F$, we can choose a different point $c \notin F$ and prove Theorem 2.24 for the random variables $\widetilde{X}_j := X_j - c$ and the deterministic array $\tilde{\xi}_l^{(n)} := \xi_l^{(n)} - c$. This follows since the translation of the roots of p_n by c simply translates the critical points by c (see Proposition 2.23).

Lemma 2.37. *Suppose $0 \notin F$. Let $R \geq \max\{1, 3r\}$ be the value from Lemma 2.36. Then there exists a non-negative constant A such that*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n} I_n(0;R) \leq -A \right) = 0.$$

Proof. Since $0 \notin F$, we have $p_n(0) \neq 0$ almost surely; in other words, 0 is almost surely not a pole of L_n. Furthermore, by Lemma 2.32, it follows that 0 is not a zero of L_n with probability $1 - o(1)$.

Consequently, on the same event, the Poisson–Jensen formula (2.13) applies to $z = 0 \in \mathbb{D}_r$, and we obtain

$$\begin{aligned}\frac{1}{n} I_n(0;R) &= \frac{1}{n}\log|L_n(0)| - \frac{1}{n}\sum_{t=1}^{t_n}\log\left|\frac{w_t^{(n)}}{R}\right| + \frac{1}{n}\sum_{s=1}^{s_n}\log\left|\frac{y_s^{(n)}}{R}\right| \\ &\geq \frac{1}{n}\log|L_n(0)| + \frac{1}{n}\sum_{s=1}^{s_n}\log\left|\frac{y_s^{(n)}}{R}\right|.\end{aligned} \tag{2.18}$$

The inequality comes from eliminating

$$\frac{1}{n}\sum_{t=1}^{t_n}\log\left|\frac{w_t^{(n)}}{R}\right| \leq 0.$$

We bound the remaining two terms in probability. A bound for the first term follows from Lemma 2.32. It remains to find a lower bound (in probability) for the last term in (2.18). Let

$$x_1^{(n)}, \dots, x_{u_n}^{(n)} \quad \text{and} \quad \zeta_1^{(n)}, \dots, \zeta_{v_n}^{(n)}$$

be the random and deterministic roots, respectively, of p_n that are contained in $\mathbb{D}_R$. (Note that $u_n + v_n = s_n$.) The law of large numbers implies that

$$\frac{1}{n-k_n}\sum_{u=1}^{u_n}\log\left|\frac{x_u^{(n)}}{R}\right| = \frac{1}{n-k_n}\sum_{j=1}^{n-k_n}\log\left|\frac{X_j}{R}\right|\mathbb{1}_{\mathbb{D}_R}(X_j) \longrightarrow -\mathbb{E}\log_-\left|\frac{X_1}{R}\right|$$

almost surely as $n \to \infty$. The expectation on the right-hand side is finite since $\mathbb{E}\log_-|X_1| < \infty$ due to the assumption $0 \notin F$ and by the bounds

$$\begin{aligned}-\mathbb{E}\left[\log_-\left|\frac{X_1}{R}\right|\right] &= -\mathbb{E}\left[\log_-\left|\frac{X_1}{R}\right| - \log_-|X_1| + \log_-|X_1|\right] \\ &\geq -\mathbb{E}\left[\log_-\left(\frac{1}{R}\right)\right] - \mathbb{E}\left[\log_-|X_1|\right] \\ &\geq -\log(R) - \mathbb{E}\left[\log_-|X_1|\right],\end{aligned}$$

which follow from the fact that $R \geq 1$. Since $k_n = o(n)$, it follows that

$$\frac{1}{n}\sum_{u=1}^{u_n}\log\left|\frac{x_u^{(n)}}{R}\right| \longrightarrow -\mathbb{E}\left[\log_-\left|\frac{X_1}{R}\right|\right] \geq -\log(R) - \mathbb{E}\log_-|X_1|$$

almost surely as $n \to \infty$, and as a consequence, we have almost surely

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{u=1}^{u_n}\log\left|\frac{x_u^{(n)}}{R}\right| \geq -A_1, \tag{2.19}$$

for some non-negative constant A_1 (depending on R). Similarly, as $R \geq 1$, we have

$$\begin{aligned}
0 \geq \frac{1}{n} \sum_{v=1}^{v_n} \log \left| \frac{\zeta_v^{(n)}}{R} \right| &= \frac{1}{n} \sum_{l=1}^{k_n} \log \left| \frac{\xi_l^{(n)}}{R} \right| \mathbb{1}_{\mathbb{D}_R} \left(\xi_l^{(n)} \right) \\
&= -\frac{1}{n} \sum_{l=1}^{k_n} \log_- \left| \frac{\xi_l^{(n)}}{R} \right| \\
&= -\frac{1}{n} \sum_{l=1}^{k_n} \left(\log_- \left| \frac{\xi_l^{(n)}}{R} \right| - \log_- \left| \xi_l^{(n)} \right| \right) - \frac{1}{n} \sum_{l=1}^{k_n} \log_- \left| \xi_l^{(n)} \right| \\
&\geq -\frac{1}{n} \sum_{l=1}^{k_n} \log_- \left| \frac{1}{R} \right| - \frac{1}{n} \sum_{l=1}^{k_n} \log_- \left| \xi_l^{(n)} \right| \\
&= -\frac{k_n}{n} \log(R) - \frac{1}{n} \sum_{l=1}^{k_n} \log_- \left| \xi_l^{(n)} \right|.
\end{aligned}$$

By condition (2.6) and the fact that $k_n = o(n)$, we obtain

$$\lim_{n \to \infty} \frac{1}{n} \sum_{v=1}^{v_n} \log \left| \frac{\zeta_v^{(n)}}{R} \right| = 0. \tag{2.20}$$

(Recall that $0 \notin F$, and hence $0 \notin E$.) Together, (2.19) and (2.20) imply the desired conclusion. $\square$

Lemma 2.38. *Suppose $0 \notin F$. Let $R \geq \max\{1, 3r\}$ be the constant from Lemma 2.36. Then there exists a non-negative constant B such that*

$$\lim_{n \to \infty} \mathbb{P} \left(\frac{1}{n} \inf_{z \in \mathbb{D}_r} I_n(z; R) \leq -B \right) = 0.$$

Proof. The proof presented here closely follows the arguments in [28]. For simplicity, define

$$q_n^+(\theta) := \frac{1}{n} \log_+ \left| L_n(Re^{i\theta}) \right| \quad \text{and} \quad q_n^-(\theta) := \frac{1}{n} \log_- \left| L_n(Re^{i\theta}) \right|$$

for $\theta \in [0, 2\pi]$. By the definition of the Poisson kernel (2.14) and reasoning similar to that used to derive the bounds in (2.16), we have

$$\frac{1}{2} \leq P_R(|z|, \theta) \leq 2$$

for all $z \in \mathbb{D}_r$ and $\theta \in [0, 2\pi]$. Notice that $P_R(0, \theta) = 1$ for all $\theta \in [0, 2\pi]$, so we have

$$\frac{2\pi}{n} I_n(0; R) = \int_0^{2\pi} q_n^+(\theta)\, d\theta - \int_0^{2\pi} q_n^-(\theta)\, d\theta.$$

It follows that, for any $n \in \mathbb{N}$ and any $z \in \mathbb{D}_r$,

$$\begin{aligned}
\frac{2\pi}{n} I_n(z;R) &= \int_0^{2\pi} q_n^+(\theta) P_R(|z|, \theta - \arg z)\, d\theta - \int_0^{2\pi} q_n^-(\theta) P_R(|z|, \theta - \arg z)\, d\theta \\
&\geq \frac{1}{2} \int_0^{2\pi} q_n^+(\theta)\, d\theta - 2 \int_0^{2\pi} q_n^-(\theta)\, d\theta \\
&= \left(\frac{1}{2} - 2\right) \int_0^{2\pi} q_n^+(\theta)\, d\theta + 2 \left(\int_0^{2\pi} q_n^+(\theta)\, d\theta - \int_0^{2\pi} q_n^-(\theta)\, d\theta \right) \\
&= -\frac{3}{2} \int_0^{2\pi} q_n^+(\theta)\, d\theta + \frac{4\pi}{n} I_n(0;R).
\end{aligned}$$

In the case where $q_n^+(\theta) = 0$ for all $\theta \in [0, 2\pi]$, we obtain the bound

$$\frac{2\pi}{n} I_n(z;R) \geq \frac{4\pi}{n} I_n(0;R).$$

Otherwise,

$$q_n^+(\theta) \leq \frac{1}{n} \log \sup_{|z|=R} |L_n(z)|$$

for all $\theta \in [0, 2\pi]$, and continuing from above,

$$\begin{aligned}
\frac{2\pi}{n} I_n(z;R) &\geq \frac{4\pi}{n} I_n(0;R) - \frac{3}{2} \int_0^{2\pi} \frac{1}{n} \log \sup_{|z|=R} |L_n(z)|\, d\theta \\
&= \frac{4\pi}{n} I_n(0;R) - \frac{3\pi}{n} \log \sup_{|z|=R} |L_n(z)|.
\end{aligned}$$

In either case, taking the infimum over all $z \in \mathbb{D}_r$ and applying the results of Lemmas 2.35 and 2.37 gives the desired conclusion. ☐

We complete the proof of Lemma 2.34 by applying Lemma 2.36 and Lemma 2.38 to (2.13). Let $R \geq \max\{1, 3r\}$ be as in Lemma 2.36. From (2.13), we apply the Cauchy–Schwarz inequality twice to obtain

$$\begin{aligned}
\frac{1}{n^2} \log^2 |L_n(z)| \leq \frac{3}{n^2} I_n^2(z;R) &+ \frac{3t_n}{n^2} \sum_{t=1}^{t_n} \log^2 \left| \frac{R\left(z - w_t^{(n)}\right)}{R^2 - \overline{w_t^{(n)}}\, z} \right| \\
&+ \frac{3s_n}{n^2} \sum_{s=1}^{s_n} \log^2 \left| \frac{R\left(z - y_s^{(n)}\right)}{R^2 - \overline{y_s^{(n)}}\, z} \right|,
\end{aligned} \tag{2.21}$$

for $z \in \mathbb{D}_R$ that is not a zero or pole of L_n. Since there are finitely many zeros and poles of L_n for a fixed n and a fixed realization of L_n, (2.21) implies

$$\begin{aligned} \frac{1}{n^2}\int_{\mathbb{D}_r} \log^2 |L_n(z)|\, d\lambda(z) \leq \int_{\mathbb{D}_r} \Bigg(\frac{3}{n^2} I_n^2(z;R) + \frac{3t_n}{n^2}\sum_{t=1}^{t_n} \log^2 \left| \frac{R\left(z - w_t^{(n)}\right)}{R^2 - \overline{w_t^{(n)}}\, z} \right| \\ + \frac{3s_n}{n^2}\sum_{s=1}^{s_n} \log^2 \left| \frac{R\left(z - y_s^{(n)}\right)}{R^2 - \overline{y_s^{(n)}}\, z} \right| \Bigg)\, d\lambda(z) \end{aligned} \tag{2.22}$$

almost surely. Lemmas 2.36 and 2.38 establish that

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{3}{n^2} \int_{\mathbb{D}_r} I_n^2(z;R)\, d\lambda(z) \right| \geq C \right) = 0$$

for some constant $C > 0$, and hence the sequence $\frac{3}{n^2}\int_{\mathbb{D}_r} I_n^2(z;R)\, d\lambda(z)$ is tight.

The remaining two terms of (2.22) are bounded almost surely. Indeed, for $z \in \mathbb{D}_r$ and $y_s^{(n)} \in \mathbb{D}_R$, we have

$$\frac{|z - y_s^{(n)}|}{2R} \leq \left| \frac{R(z - y_s^{(n)})}{R^2 - \overline{y_s^{(n)}}\, z} \right| \leq \frac{|z - y_s^{(n)}|}{R - r},$$

and hence

$$\log^2 \left| \frac{R(z - y_s^{(n)})}{R^2 - \overline{y_s^{(n)}}\, z} \right| \leq \log^2 \frac{|z - y_s^{(n)}|}{2R} + \log^2 \frac{|z - y_s^{(n)}|}{R - r}.$$

By a simple change of variables, we obtain

$$\int_{\mathbb{D}_r} \log^2 \frac{|z - y_s^{(n)}|}{2R}\, d\lambda(z) \leq \int_{\mathbb{D}_{2R}} \log^2 \frac{|z|}{2R}\, d\lambda(z),$$

and similarly

$$\int_{\mathbb{D}_r} \log^2 \frac{|z - y_s^{(n)}|}{R - r}\, d\lambda(z) \leq \int_{\mathbb{D}_{2R}} \log^2 \frac{|z|}{R - r}\, d\lambda(z).$$

Thus, by the local integrability of the squared logarithm,

$$\frac{3s_n}{n^2}\int_{\mathbb{D}_r} \sum_{s=1}^{s_n} \log^2 \left| \frac{R\left(z - y_s^{(n)}\right)}{R^2 - \overline{y_s^{(n)}}\, z} \right| d\lambda(z) \leq \frac{3s_n^2}{n^2} C' \leq 3C'$$

almost surely for all $n \in \mathbb{N}$, where $C' > 0$ is a constant that depends only on R and r, and, in the last inequality, we used the fact that $s_n \leq n$. A similar argument applies to the integral of the sum in (2.22) involving the critical points $w_t^{(n)}$; we omit the details.

We conclude that the sequence $\frac{1}{n^2}\int_{\mathbb{D}_r} \log^2 |L_n(z)|\, d\lambda(z)$ is tight, and the proof of Lemma 2.34 is complete.

2.5 Proof of Theorem 2.3

This section is devoted to the proof of Theorem 2.3. For $\varepsilon > 0$, define

$$\mathrm{S}_\mu(\varepsilon) := \{z \in \mathbb{C} : \mathrm{dist}(z, \mathrm{supp}(\mu)) < \varepsilon\}$$

to be the ε-neighborhood of the support of μ. We begin with the following concentration inequality.

Lemma 2.39. *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $X_1, \ldots, X_n$ are iid random variables with distribution μ. Then, for every $M, \varepsilon, t > 0$,*

$$\mathbb{P}\left(\sup_{z \in \mathbb{C}: |z| \leq M, z \notin \mathrm{S}_\mu(\varepsilon)} \left| \frac{1}{n} \sum_{j=1}^{n} \frac{1}{z - X_j} - m_\mu(z) \right| \geq t \right) \leq C \left(1 + \frac{40M}{\varepsilon^2 t} \right) \exp\left(-cnt^2\varepsilon^2\right)$$

for some absolute constants $C, c > 0$.

Proof. Let $M, \varepsilon, t > 0$, and define

$$D := \{z \in \mathbb{C} : |z| \leq M, z \notin \mathrm{S}_\mu(\varepsilon)\}.$$

We assume D is nonempty as the conclusion is trivial otherwise. Let $\mathcal{N}$ be an $\varepsilon^2 t/10$-net of D. By Lemma 2.17, $\mathcal{N}$ can be chosen so that

$$|\mathcal{N}| \leq \left(1 + \frac{40M}{\varepsilon^2 t} \right)^2. \tag{2.23}$$

We observe that $X_j \in \mathrm{supp}(\mu)$ almost surely for every $1 \leq j \leq n$. Thus, almost surely, for $z \in D$,

$$\left| \frac{1}{z - X_j} \right| \leq \frac{1}{\varepsilon}. \tag{2.24}$$

Hence, for $z, w \in D$,

$$\left| \frac{1}{n} \sum_{j=1}^{n} \frac{1}{z - X_j} - \frac{1}{n} \sum_{j=1}^{n} \frac{1}{w - X_j} \right| \leq \frac{|z - w|}{\varepsilon^2}.$$

In other words, the function $m_n(z) := \frac{1}{n} \sum_{j=1}^{n} \frac{1}{z - X_j}$ is almost surely Lipschitz continuous on D

with Lipschitz constant ε^{-2}. Similarly, for $z, w \in D$,

$$\begin{aligned}|m_\mu(z) - m_\mu(w)| &= \left| \int_{\mathrm{supp}(\mu)} \left(\frac{1}{z-x} - \frac{1}{w-x} \right) d\mu(x) \right| \\ &\leq \int_{\mathrm{supp}(\mu)} \frac{|z-w|}{|z-x||w-x|} d\mu(x) \\ &\leq \frac{|z-w|}{\varepsilon^2}.\end{aligned}$$

Suppose $\sup_{z\in D} |m_n(z) - m_\mu(z)| \geq t$. As m_n and m_μ are both continuous on the compact set D, there exists $z \in D$ such that $|m_n(z) - m_\mu(z)| \geq t$. Since $\mathcal{N}$ is an $\varepsilon^2 t/10$-net of D, there exists $w \in \mathcal{N}$ such that $|z-w| \leq \frac{\varepsilon^2 t}{10}$. So, by the reverse triangle inequality and the fact that the m_n and m_μ are Lipschitz continuous, we have

$$\begin{aligned}|m_n(w) - m_\mu(w)| &\geq |m_n(z) - m_\mu(z)| - |m_n(z) - m_n(w) - (m_\mu(z) - m_\mu(w))| \\ &\geq t - 2\frac{|z-w|}{\varepsilon^2} \\ &\geq \frac{4t}{5}.\end{aligned}$$

Therefore, by the union bound, we conclude that

$$\begin{aligned}\mathbb{P}\left(\sup_{z\in D} |m_n(z) - m_\mu(z)| \geq t \right) &\leq \mathbb{P}\left(\sup_{w\in\mathcal{N}} |m_n(w) - m_\mu(w)| \geq \frac{4t}{5} \right) \\ &\leq \sum_{w\in\mathcal{N}} \mathbb{P}\left(|m_n(w) - m_\mu(w)| \geq \frac{4t}{5} \right) \qquad (2.25)\end{aligned}$$

As $\mathbb{E}m_n(z) = m_\mu(z)$ for $z \in D$, Hoeffding's inequality (Lemma 2.15) and the bound in (2.24) imply that

$$\sup_{w\in\mathcal{N}} \mathbb{P}\left(|m_n(w) - m_\mu(w)| \geq \frac{4t}{5} \right) \leq C \exp(-cnt^2\varepsilon^2) \qquad (2.26)$$

for some absolute constants $C, c > 0$. Thus, combining (2.23), (2.25), and (2.26) yields

$$\mathbb{P}\left(\sup_{z\in D} |m_n(z) - m_\mu(z)| \geq t \right) \leq C\left(1 + \frac{40M}{\varepsilon^2 t}\right) \exp(-cnt^2\varepsilon^2),$$

as desired. $\square$

We now prove Theorem 2.3.

Proof of Theorem 2.3. Let $\varepsilon > 0$. With probability one, $X_j \in \operatorname{supp}(\mu)$ for each $1 \leq j \leq n$. Thus, the zeros of

$$m_n(z) := \frac{1}{n}\frac{p_n'(z)}{p_n(z)} = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{z - X_j}$$

outside of $\mathrm{N}_\mu(\varepsilon)$ are exactly the critical points of p_n outside of $\mathrm{N}_\mu(\varepsilon)$. We will show that $m_n(z)$ has no zeros in $D := \operatorname{Conv}(\operatorname{supp}(\mu)) \setminus \mathrm{N}_\mu(\varepsilon)$. The claim then follows immediately since, by the Gauss–Lucas theorem (Theorem 1.1), all the critical points of p_n lie in $\operatorname{Conv}(\operatorname{supp}(\mu))$.

Since μ has compact support, $\operatorname{Conv}(\operatorname{supp}(\mu))$ is also a compact set (see [1, Corollary 5.33]), and hence D is compact. As m_μ is a continuous function on D, $|m_\mu|$ achieves its minimum on D, which, by definition of $\mathrm{N}_\mu(\varepsilon)$ cannot be zero (since $\mathrm{N}_\mu(\varepsilon)$ contains the zero set M_μ). Thus, there exists $c' > 0$ such that

$$|m_\mu(z)| \geq c' \quad \text{for all } z \in D.$$

Since D is compact, there exists $M > 0$ (depending only on $\operatorname{supp}(\mu)$) such that $|z| \leq M$ for all $z \in D$. Thus, by Lemma 2.39 (taking $t = c'/2$), we obtain

$$\mathbb{P}\left(\sup_{z \in D} |m_n(z) - m_\mu(z)| \geq \frac{c'}{2}\right) \leq C\left(1 + \frac{80M}{\varepsilon^2 c'}\right) \exp(-cnc'\varepsilon^2)$$

for some absolute constants $C, c > 0$. Hence, on the complementary event, we have

$$|m_n(z)| \geq |m_\mu(z)| - |m_n(z) - m_\mu(z)| \geq \frac{c'}{2}$$

for all $z \in D$. Since the constants $C\left(1 + \frac{80M}{\varepsilon^2 c'}\right)$ and $cc'\varepsilon^2$ only depend on μ and ε, the proof is complete. $\square$

2.6 Proof of Theorems 2.9, 2.11, and 2.12

This section is devoted to the proof of Theorems 2.9, 2.11, and 2.12.

2.6.1 Proof of Theorem 2.9

We now prove Theorem 2.9 using Theorem 2.11. Indeed, let $\xi_1, \ldots, \xi_k$ satisfy the assumptions of Theorem 2.9. Since $\xi_1, \ldots, \xi_k$ do not depend on n, there exists $\varepsilon_0 > 0$ such that, for any $0 < \varepsilon < \varepsilon_0$,

- $\xi_1, \ldots, \xi_s$ are outside $\mathrm{N}_\mu(3\varepsilon)$,

- $\xi_{s+1}, \ldots, \xi_k$ are in $\mathrm{N}_\mu(\varepsilon)$.

In addition, condition (2.2) trivially holds because $\xi_1, \ldots, \xi_k$ do not depend on n. Thus, Theorem 2.11 is applicable for any $0 < \varepsilon < \varepsilon_0$, and hence Theorem 2.9 follows.

2.6.2 Proof of Theorems 2.11 and 2.12

We will prove Theorem 2.11 via the following result.

Theorem 2.40. *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $0 \in \operatorname{supp}(\mu)$. Let $X_1, X_2, \ldots$ be iid random variables with distribution μ. For each $n \geq 1$, let $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ be a triangular array of deterministic complex numbers with $k_n = O(1)$, and assume*

$$\max\{|\xi_1^{(n)}|, \ldots, |\xi_{k_n}^{(n)}|\} = O(1).$$

Fix $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no values of $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ in $\mathrm{N}_\mu(3\varepsilon) \setminus \mathrm{N}_\mu(\varepsilon)$ and there are s values $\xi_1^{(n)}, \ldots, \xi_s^{(n)}$ outside $\mathrm{N}_\mu(3\varepsilon)$. Then, almost surely, for n sufficiently large, there are exactly s critical points (counted with multiplicity) of the polynomial

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

outside $\mathrm{N}_\mu(2\varepsilon)$, and after labeling these critical points correctly,

$$w_l(p_n) = \xi_l^{(n)} + o(1)$$

for each $1 \leq l \leq s$.

The only difference between this theorem and Theorem 2.11 is that Theorem 2.40 assumes $0 \in \operatorname{supp}(\mu)$. Using Theorem 2.40, we prove Theorem 2.11 by applying Proposition 2.23.

Proof of Theorem 2.11. Let μ have compact support. Since $\operatorname{supp}(\mu)$ is nonempty, choose $a \in \operatorname{supp}(\mu)$. We now consider the polynomial

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

where $Y_j := X_j - a$. Let ν be the distribution of Y_1. Then ν has compact support and $0 \in \operatorname{supp}(\nu)$. In addition, the sets M_ν and $\operatorname{supp}(\nu)$ are translates by $-a$ of the sets M_μ and $\operatorname{supp}(\mu)$, respectively. Thus, by assumption, there are no values of $\xi_1^{(n)} - a, \ldots, \xi_{k_n}^{(n)} - a$ in $\mathrm{N}_\nu(3\varepsilon) \setminus \mathrm{N}_\nu(\varepsilon)$ and there are s values $\xi_1^{(n)} - a, \ldots, \xi_s^{(n)} - a$ outside $\mathrm{N}_\nu(3\varepsilon)$. Therefore, by Theorem 2.40 and Proposition 2.23, we conclude that almost surely, for n sufficiently large, there are exactly s critical points of p_n outside $\mathrm{N}_\mu(2\varepsilon)$ and after labeling correctly,

$$w_l(p_n) - a = \xi_l^{(n)} - a + o(1)$$

for $1 \leq l \leq s$. Adding a to both sides completes the proof. $\square$

Similarly, Theorem 2.12 can be proven using the following.

Theorem 2.41. *Let μ be a probability measure on $\mathbb{C}$ with compact support, and suppose $0 \in \operatorname{supp}(\mu)$. Let $X_1, X_2, \ldots$ be iid random variables with distribution μ. For each $n \geq 1$, let $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ be a triangular array of deterministic complex numbers with $k_n = O(1)$. Fix $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no values of $\xi_1^{(n)}, \ldots, \xi_{k_n}^{(n)}$ in $\mathrm{N}_\mu(3\varepsilon) \setminus \mathrm{N}_\mu(\varepsilon)$ and there is one value $\xi_1^{(n)}$ outside $\mathrm{N}_\mu(3\varepsilon)$. Then, almost surely, for n sufficiently large, there is exactly one critical point of the polynomial*

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

outside $\mathrm{N}_\mu(2\varepsilon)$, and after labeling the critical points correctly,

$$w_1(p_n) = \xi_1^{(n)} \left(1 + O\left(1/n\right)\right).$$

The proof of Theorem 2.12 using Theorem 2.41 is nearly identical to the proof of Theorem 2.11 above; we omit the details. It remains to prove Theorems 2.40 and 2.41.

2.6.3 Proof of Theorems 2.40 and 2.41

We prove Theorems 2.40 and 2.41 simultaneously. Indeed, for the first part of the proof, we continue to use the notation of Theorem 2.40. However, the same argument applies to Theorem

2.41 by simply taking $s = 1$. The conclusion of the proof will require us to consider the conditions of both theorems separately. In fact, the conclusion of the proof is the only place where we require condition (2.2). For notational convenience, throughout the proof we allow the implicit constants and rates of convergence in our asymptotic notation (such as O, o) to depend on the parameter ε without notating this dependence.

For n sufficiently large, we decompose

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

where, by assumption, $\xi_1^{(n)}, \ldots, \xi_s^{(n)}$ are outside $\mathrm{N}_\mu(3\varepsilon)$ and $\xi_{s+1}^{(n)}, \ldots, \xi_{k_n}^{(n)}$ are in $\mathrm{N}_\mu(\varepsilon)$. In addition, $X_1, \ldots, X_{n-k_n}$ are in $\operatorname{supp}(\mu) \subset \mathrm{N}_\mu(\varepsilon)$ with probability 1.

Let D be the diagonal matrix

$$D := \begin{pmatrix} D_{\text{in}} & 0 \\ 0 & D_{\text{out}} \end{pmatrix},$$

where

$$D_{\text{in}} := \operatorname{diag}(X_1, \ldots, X_{n-k_n}, \xi_{s+1}^{(n)}, \ldots, \xi_{k_n}^{(n)})$$

and

$$D_{\text{out}} := \operatorname{diag}(\xi_1^{(n)}, \ldots, \xi_s^{(n)}).$$

Here, the subscripts "in" and "out" refer to the roots inside and outside $\mathrm{N}_\mu(\varepsilon)$, respectively. Of course, D, D_{in}, and D_{out} all depend on n, but we do not denote this dependence in our notation.

By Theorem 2.18, it follows that

$$\begin{aligned}\frac{1}{n}zp_n'(z) &= \det\left(zI - D + \frac{1}{n}DJ_n\right) \\ &= \det\left[\begin{pmatrix} zI & 0 \\ 0 & zI \end{pmatrix} - \begin{pmatrix} D_{\text{in}} & 0 \\ 0 & D_{\text{out}} \end{pmatrix} + \frac{1}{n}\begin{pmatrix} D_{\text{in}} & 0 \\ 0 & D_{\text{out}} \end{pmatrix} J_n\right],\end{aligned} \tag{2.27}$$

where I is the identity matrix and J_n is the $n \times n$ all-one matrix. We decompose,

$$J_n = \begin{pmatrix} J_{n-s} & J_{n-s,s} \\ J_{s,n-s} & J_s \end{pmatrix},$$

where $J_{l,m}$ denotes the $l \times m$ all-one matrix. Thus, we conclude that

$$\frac{1}{n} z p_n'(z) = \det \begin{pmatrix} zI - D_{\text{in}} + \frac{1}{n} D_{\text{in}} J_{n-s} & \frac{1}{n} D_{\text{in}} J_{n-s,s} \\ \frac{1}{n} D_{\text{out}} J_{s,n-s} & zI - D_{\text{out}} + \frac{1}{n} D_{\text{out}} J_s \end{pmatrix}. \tag{2.28}$$

We will eventually apply Lemma 2.19 to compute this determinant, but first we will need to consider the upper-left block

$$zI - D_{\text{in}} + \frac{1}{n} D_{\text{in}} J_{n-s}.$$

Let $\mathbf{1}_n$ denote the all-one n-vector; we will often drop the subscript (and just write $\mathbf{1}$) when its size can be deduced from context. We will make use of the following lemma.

Lemma 2.42. *Under the assumptions of Theorem 2.40 (alternatively, Theorem 2.41), almost surely, for n sufficiently large, the matrix*

$$zI - D_{\text{in}} + \frac{1}{n} D_{\text{in}} J_{n-s} \tag{2.29}$$

is invertible for every $z \notin \mathrm{N}_\mu(2\varepsilon)$ and the function

$$z \mapsto \frac{1}{n} \mathbf{1}^{\mathrm{T}} \left(zI - D_{\text{in}} + \frac{1}{n} D_{\text{in}} J_{n-s} \right)^{-1} D_{\text{in}} \mathbf{1} \tag{2.30}$$

is analytic outside $\mathrm{N}_\mu(2\varepsilon)$. In addition, almost surely

$$\sup_{z \in \mathbb{C} \setminus \mathrm{N}_\mu(2\varepsilon)} \left| \frac{1}{n} \mathbf{1}^{\mathrm{T}} \left(zI - D_{\text{in}} + \frac{1}{n} D_{\text{in}} J_{n-s} \right)^{-1} D_{\text{in}} \mathbf{1} \right| = O(1).$$

Proof. Recall that the entries of the diagonal matrix D_{in} are contained in $\mathrm{N}_\mu(\varepsilon)$. Thus, for $z \notin \mathrm{N}_\mu(2\varepsilon)$, the matrix $zI - D_{\text{in}}$ is invertible. In addition, since $(zI - D_{\text{in}})^{-1}$ is a diagonal matrix, we obtain

$$\begin{aligned} \frac{1}{n} \mathbf{1}^{\mathrm{T}} (zI - D_{\text{in}})^{-1} D_{\text{in}} \mathbf{1} &= \frac{1}{n} \operatorname{tr}[(zI - D_{\text{in}})^{-1} D_{\text{in}}] \\ &= \frac{1}{n} \sum_{j=1}^{n-k_n} \frac{X_j}{z - X_j} + \frac{1}{n} \sum_{l=s+1}^{k_n} \frac{\xi_l^{(n)}}{z - \xi_l^{(n)}}. \end{aligned} \tag{2.31}$$

Among other things, this implies that the function $\frac{1}{n} \mathbf{1}^{\mathrm{T}} (zI - D_{\text{in}})^{-1} D_{\text{in}} \mathbf{1}$ is analytic outside $\mathrm{N}_\mu(2\varepsilon)$; we will use this fact later to show that the function in (2.30) is analytic on the same set. Since

$\mathrm{supp}(\mu)$ is compact, it follows from Proposition 2.22 that $\mathrm{N}_\mu(\varepsilon)$ is bounded. Let $\kappa > 0$ be such that $|z| \leq \kappa$ for all $z \in \mathrm{N}_\mu(\varepsilon)$. Let $M := 10\kappa$. Then for $|z| \geq M$, we have

$$|X_j| \leq \kappa, \quad |z - X_j| \geq M - \kappa = 9\kappa$$

for $1 \leq j \leq n - k_n$ and similarly

$$|\xi_l^{(n)}| \leq \kappa, \quad |z - \xi_l^{(n)}| \geq M - \kappa = 9\kappa$$

for each $s + 1 \leq l \leq k_n$. Thus,

$$\sup_{|z|\geq M} \left| \frac{1}{n} \mathbf{1}^{\mathrm{T}} (zI - D_{\mathrm{in}})^{-1} D_{\mathrm{in}} \mathbf{1} \right| \leq \frac{\kappa}{9\kappa} = \frac{1}{9}. \tag{2.32}$$

In particular, this bound implies that $1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}}(zI - D_{\mathrm{in}})^{-1} D_{\mathrm{in}}\mathbf{1} \neq 0$ for all $|z| \geq M$. Thus, we can apply Lemma 2.20 to conclude that the matrix in (2.29) is invertible for every $|z| \geq M$. Indeed, since $\frac{1}{n} D_{\mathrm{in}} J_{n-s} = \frac{1}{n} D_{\mathrm{in}} \mathbf{1}\mathbf{1}^{\mathrm{T}}$ is at most rank one[1] , it follows from Lemma 2.20 (taking $u = D_{\mathrm{in}}\mathbf{1}$ and $v = \frac{1}{n}\mathbf{1}$) that

$$\begin{aligned} &\frac{1}{n}\mathbf{1}^{\mathrm{T}} \left(zI - D_{\mathrm{in}} + \frac{1}{n} D_{\mathrm{in}} J_{n-s} \right)^{-1} D_{\mathrm{in}} \mathbf{1} \\ &\quad = \frac{1}{n}\mathbf{1}^{\mathrm{T}} (zI - D_{\mathrm{in}})^{-1} D_{\mathrm{in}} \mathbf{1} - \frac{\left(\frac{1}{n}\mathbf{1}^{\mathrm{T}}(zI - D_{\mathrm{in}})^{-1} D_{\mathrm{in}}\mathbf{1}\right)^2}{1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}}(zI - D_{\mathrm{in}})^{-1} D_{\mathrm{in}}\mathbf{1}}. \end{aligned} \tag{2.33}$$

Hence, by the bound in (2.32), we have, with probability one,

$$\sup_{|z|\geq M} \left| \frac{1}{n}\mathbf{1}^{\mathrm{T}} \left(zI - D_{\mathrm{in}} + \frac{1}{n} D_{\mathrm{in}} J_{n-s} \right)^{-1} D_{\mathrm{in}} \mathbf{1} \right| \leq \frac{1}{9} + \frac{\left(\frac{1}{9}\right)^2}{1 - \frac{1}{9}} = O(1).$$

In addition, the right-hand side of (2.33) is analytic in the region $|z| \geq M$, which implies that the function on the left-hand side is also analytic in the same region.

Let Ω be the compact set $\{z \in \mathbb{C} : |z| \leq M\} \setminus \mathrm{N}_\mu(2\varepsilon)$. It remains to show that, almost surely, for n sufficiently large, the matrix in (2.29) is invertible for every $z \in \Omega$, the function in (2.30) is analytic in Ω, and

$$\sup_{z\in\Omega} \left| \frac{1}{n}\mathbf{1}^{\mathrm{T}} \left(zI - D_{\mathrm{in}} + \frac{1}{n} D_{\mathrm{in}} J_{n-s} \right)^{-1} D_{\mathrm{in}} \mathbf{1} \right| = O(1).$$

[1] Here, we have used the fact that J_{n-s} is rank one, and so the product $D_{\mathrm{in}} J_{n-s}$ is either rank one or rank zero. In fact, a simple computation reveals that the product is rank zero if and only if D_{in} is the zero matrix.

To establish these results we will again apply Lemma 2.20. However, in this case, we will need more precise estimates than those established above.

Indeed, returning to (2.31), we find that

$$\frac{1}{n}\mathbf{1}^{\mathrm{T}}(zI - D_{\mathrm{in}})^{-1}D_{\mathrm{in}}\mathbf{1} = -\frac{n-s}{n} + \frac{z}{n}\sum_{j=1}^{n-k_n}\frac{1}{z - X_j} + \frac{z}{n}\sum_{l=s+1}^{k_n}\frac{1}{z - \xi_l^{(n)}}. \tag{2.34}$$

Since $\xi_{s+1}^{(n)}, \ldots, \xi_{k_n}^{(n)}$ are contained in $\mathrm{N}_\mu(\varepsilon)$, it follows from the triangle inequality that

$$\sup_{z\in\Omega}\left|\frac{z}{n}\sum_{l=s+1}^{k_n}\frac{1}{z - \xi_l^{(n)}}\right| \le \frac{k_n}{n}\frac{M}{\varepsilon} = o(1). \tag{2.35}$$

In addition, by Lemma 2.39 and the Borel–Cantelli lemma, we have, almost surely

$$\sup_{z\in\Omega}\left|\frac{z}{n}\sum_{j=1}^{n-k_n}\frac{1}{z - X_j} - zm_\mu(z)\right| = o(1). \tag{2.36}$$

As Ω is compact and m_μ cannot vanish on Ω (since $M_\mu \subset N_\mu(\varepsilon)$), there exists $C, c > 0$ such that $c \le |m_\mu(z)| \le C$ for all $z \in \Omega$. Specifically, by the assumption that $0 \in \mathrm{supp}(\mu)$, it follows that

$$\varepsilon c \le |zm_\mu(z)| \le MC, \qquad \text{for all } z \in \Omega. \tag{2.37}$$

Therefore, by (2.35), (2.36), and (2.37), we conclude from (2.34) that, almost surely, for n sufficiently large,

$$\sup_{z\in\Omega}\left|\frac{1}{n}\mathbf{1}^{\mathrm{T}}(zI - D_{\mathrm{in}})^{-1}D_{\mathrm{in}}\mathbf{1}\right| \le 2 + MC$$

and

$$\inf_{z\in\Omega}\left|1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}}(zI - D_{\mathrm{in}})^{-1}D_{\mathrm{in}}\mathbf{1}\right| \ge \frac{\varepsilon c}{2}.$$

Hence, by Lemma 2.20, we obtain (2.33) for $z \in \Omega$ which, combined with the bounds above, yields

$$\sup_{z\in\Omega}\left|\frac{1}{n}\mathbf{1}^{\mathrm{T}}\left(zI - D_{\mathrm{in}} + \frac{1}{n}D_{\mathrm{in}}J_{n-s}\right)^{-1}D_{\mathrm{in}}\mathbf{1}\right| \le 2 + MC + \frac{(2+MC)^2}{\frac{c\varepsilon}{2}} = O(1)$$

almost surely. As before, (2.33) also implies that the function in (2.30) is analytic on Ω. The proof of the lemma is complete. $\square$

Let us dispatch the simplest case of Theorem 2.40: when $s = 0$. Indeed, if $s = 0$, then $D = D_{\text{in}}$. In this case, (2.27) and the invertibility of (2.29) imply that p_n has no critical points outside $\mathrm{N}_\mu(2\varepsilon)$, completing the proof. Thus, for the remainder of the proof, we assume $s \geq 1$.

We return to the block determinant in (2.28). By Lemma 2.42, almost surely, for n sufficiently large, the upper-left block is invertible for all $z \notin \mathrm{N}_\mu(2\varepsilon)$. Thus, by Lemma 2.19, we conclude that almost surely

$$\begin{aligned}\frac{1}{n} z p_n'(z) = \det\left(zI - D_{\text{in}} + \frac{1}{n} D_{\text{in}} J_{n-s}\right)& \\ \times \det\left(zI - D_{\text{out}} + \frac{1}{n} D_{\text{out}} J_s - \frac{1}{n} D_{\text{out}} J_{s,n-s} G(z) \frac{1}{n} D_{\text{in}} J_{n-s,s}\right)&\end{aligned}$$

for all $z \notin \mathrm{N}_\mu(2\varepsilon)$, where

$$G(z) := \left(zI - D_{\text{in}} + \frac{1}{n} D_{\text{in}} J_{n-s}\right)^{-1}.$$

In other words, the zeros of p_n' outside of $\mathrm{N}_\mu(2\varepsilon)$ (counted with multiplicity) are precisely the zeros of

$$\det\left(zI - D_{\text{out}} + \frac{1}{n} D_{\text{out}} J_s - \frac{1}{n} D_{\text{out}} J_{s,n-s} G(z) \frac{1}{n} D_{\text{in}} J_{n-s,s}\right) \tag{2.38}$$

outside of $\mathrm{N}_\mu(2\varepsilon)$ (counted with multiplicity). Notice that this is the determinant of an $s \times s$ matrix, and $s \leq k_n = O(1)$. We have thus reduced the problem of studying an $n \times n$ matrix to an $s \times s$ matrix. This reduction greatly simplifies the forthcoming analysis. Before we conclude the proof, we make one final observation: since $J_{s,n-s} = \mathbf{1}_s \mathbf{1}_{n-s}^{\mathrm{T}}$ and $J_{n-s,s} = \mathbf{1}_{n-s} \mathbf{1}_s^{\mathrm{T}}$, we can rewrite the determinant in (2.38) as

$$\det\left(zI - D_{\text{out}} + \frac{1}{n} D_{\text{out}} J_s - \frac{1}{n^2}\left(\mathbf{1}_{n-s}^{\mathrm{T}} G(z) D_{\text{in}} \mathbf{1}_{n-s}\right) D_{\text{out}} J_s\right). \tag{2.39}$$

We now conclude the proof of Theorems 2.40 and 2.41 separately. Let us begin with Theorem 2.40. Indeed, under the assumptions of Theorem 2.40,

$$\|D_{\text{out}}\| = \max\{|\xi_1^{(n)}|, \ldots, |\xi_s^{(n)}|\} = O(1).$$

(Recall that $\|D_{\text{out}}\|$ denotes the spectral norm of the matrix D_{out}.) Thus, by Lemma 2.21 and

Lemma 2.42, we have, almost surely

$$\begin{aligned}
\sup_{z\notin \mathrm{N}_\mu(2\varepsilon)} &\left|\det\left(zI - D_{\text{out}} + \frac{1}{n}D_{\text{out}}J_s - \frac{1}{n^2}D_{\text{out}}J_{s,n-s}G(z)D_{\text{in}}J_{n-s,s}\right) - \det(zI - D_{\text{out}})\right| \\
&\ll \frac{1}{n}\|D_{\text{out}}\|\|J_s\| + \|D_{\text{out}}\|\|J_s\| \sup_{z\notin \mathrm{N}_\mu(2\varepsilon)}\left|\frac{1}{n^2}\mathbf{1}_{n-s}^{\mathrm{T}}G(z)D_{\text{in}}\mathbf{1}_{n-s}\right| \\
&\ll \frac{1}{n}
\end{aligned}$$

because $\|J_s\| = s \leq k_n = O(1)$. Notice that the zeros of $\det(zI - D_{\text{out}})$ are precisely the values $\xi_1^{(n)}, \ldots, \xi_s^{(n)}$. In view of Rouché's theorem (since both determinants are analytic outside $\mathrm{N}_\mu(2\varepsilon)$ due to Lemma 2.42), we conclude that, almost surely, for n sufficiently large, p_n has exactly s critical points outside $N_\mu(2\varepsilon)$, and after correctly labeling the critical points,

$$w_l(p_n) = \xi_l^{(n)} + o(1) \tag{2.40}$$

for each $1 \leq l \leq s$. This completes the proof of Theorem 2.40.

Remark 2.43. With a more careful application of Rouché's theorem, the error in (2.40) can be improved to

$$w_l(p_n) = \xi_l^{(n)} + O(n^{-\tau})$$

for each $1 \leq l \leq s$, where $\tau > 0$ depends on s. In addition, if the deterministic roots $\xi_l^{(n)}$, $1 \leq l \leq s$ satisfy some kind of separation criteria, this error term can be further improved. We do not pursue these matters here.

We now turn to the proof of Theorem 2.41. Recall that, in this case, $s = 1$. Thus, the matrix in (2.39) is just a 1×1 matrix, and hence the zeros of p_n' outside of $\mathrm{N}_\mu(2\varepsilon)$ are precisely the solutions of

$$z - \xi_1^{(n)} + \frac{1}{n}\xi_1^{(n)} - \xi_1^{(n)}\frac{1}{n^2}\mathbf{1}^{\mathrm{T}}G(z)D_{\text{in}}\mathbf{1} = 0 \tag{2.41}$$

outside $\mathrm{N}_\mu(2\varepsilon)$. By Lemma 2.42, we have, almost surely,

$$\sup_{z\notin \mathrm{N}_\mu(2\varepsilon)}\left|\left(z - \xi_1^{(n)} + \frac{1}{n}\xi_1^{(n)} - \xi_1^{(n)}\frac{1}{n^2}\mathbf{1}^{\mathrm{T}}G(z)D_{\text{in}}\mathbf{1}\right) - \left(z - \xi_1^{(n)}\right)\right| \leq \frac{C}{n}|\xi_1^{(n)}| \tag{2.42}$$

for some constant $C > 0$. Since both these terms are analytic outside $\mathrm{N}_\mu(2\varepsilon)$ due to Lemma 2.42, we can again apply Rouché's theorem. However, since $\frac{C}{n}|\xi_1^{(n)}|$ does not necessarily converge to zero, we have to be slightly more careful. Let Γ_n be any simple closed contour outside $\mathrm{N}_\mu(2\varepsilon)$ which satisfies $|z - \xi_1^{(n)}| > \frac{C}{n}|\xi_1^{(n)}|$ for all $z \in \Gamma_n$. Then, by the estimate in (2.42), Rouché's theorem implies that the number of solutions to (2.41) inside Γ_n is the same as the number of zeros of $z - \xi_1^{(n)}$ inside Γ_n. Hence, we conclude that almost surely, for n sufficiently large, there is exactly one critical point of p_n outside $\mathrm{N}_\mu(2\varepsilon)$ and that critical point takes the value $\xi_1^{(n)}(1+O(1/n))$. The proof of Theorem 2.41 is complete.

Chapter 3

On the local pairing behavior of critical points and roots of random polynomials

3.1 Introduction to the chapter

This chapter is an adaption of the paper [39], a recent collaboration with Sean O'Rourke. For clarity, much of the introductory material from [39] appears in Chapter 1, and the content of Appendix A from [39], which consisted of many supporting calculations, is included in the relevant sections below. Note that some of the results in Section 3.2.3 are similar to those in Kabluchko and Seidel's recent paper [29]. We compare the work in detail in Section 3.2.3.

3.2 Main results

We begin by introducing the Wasserstein metric in order to discuss the pairing between the roots and critical points of

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

that one sees in Figures 1.4 and 1.3. (Note: Here, p_n is defined as in (1.1).)

3.2.1 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. [61], Chapter 6).

Theorem 3.3 below gives a bound on the Wasserstein distance between the empirical measures constructed from the roots and the critical points of the polynomial p_n defined in (1.1). Before we state the theorem, we mention some notation and assumptions. For any probability measure μ on $\mathbb{C}$, let m_μ denote the Cauchy–Stieltjes transform of μ, given by

$$m_\mu(z) := \int_{\mathbb{C}} \frac{d\mu(x)}{z-x}, \tag{3.1}$$

and defined for those values of $z \in \mathbb{C}$ for which the integral exists. To denote the empirical measure constructed from the roots of p_n, we use

$$\mu_n := \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j}, \tag{3.2}$$

and our notation for the empirical measure constructed from the critical points, $w_1^{(n)}, \ldots, w_{n-1}^{(n)}$, of p_n is

$$\mu_n' := \frac{1}{n-1} \sum_{j=1}^{n-1} \delta_{w_j^{(n)}}. \tag{3.3}$$

The following assumptions describe some regularity conditions that μ must satisfy in the hypothesis of Theorem 3.3.

Assumption 3.1. Suppose there are positive constants, C_1, C_2, so that the following conditions hold when $X_1, \ldots, X_n$ are iid complex-valued random variables with common distribution μ:

(i) for any $\varepsilon > 0$,

$$\mathbb{P}(|m_\mu(X_1)| < \varepsilon) \leq C_1 \varepsilon^2;$$

(ii) the random variable $\eta_n := \max_{1 \leq j \leq n} |X_j|$ satisfies

$$\mathbb{P}\left(\eta_n \geq n^{C_2}\right) = o(1).$$

Assumption 3.2 (Alternative to Assumption 3.1 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 3.3. *Let $X_1, \dots, X_n$ be iid, complex random variables whose distribution, μ, has a bounded density and satisfies either Assumption 3.1 or Assumption 3.2. Then, there is a positive constant C, depending on μ, so that with probability $1 - o(1)$,*

$$W_1(\mu_n, \mu_n') \leq \frac{C\eta_n (\log n)^9}{n}, \tag{3.4}$$

where $\eta_n := \max_{1 \leq j \leq n} |X_j|$, and μ_n, μ_n' (defined in (3.2) and (3.3)) 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 3.3 immediately implies the following corollary.

Corollary 3.4. *Let $X_1, \dots, X_n$ be iid, complex random variables whose distribution, μ, satisfies Assumption 3.1 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 (3.2) and (3.3)) are the empirical measures constructed from the roots and critical points of

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

Theorem 3.3 and Corollary 3.4 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 1.4 and 1.3, and O'Rourke and the author believe that these bounds are optimal (up to logarithmic factors) based on the theorems of Section 3.2.3 below and the results in [29].

A couple of remarks concerning Theorem 3.3 and its corollary are in order. Due to the heuristic that motivates our proof of Theorem 3.3 (see Figure 3.3), O'Rourke and the author conjecture that Assumption 3.1 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 (3.4). 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. In this way, η_n can be viewed as quantifying how heavy-tailed the distribution μ is.

3.2.2 Examples of Theorem 3.3 and Corollary 3.4

The assumptions of Theorem 3.3 and Corollary 3.4 are rather technical, so this subsection is devoted to several specific examples worked out in detail.

The following lemma is useful for computing the Cauchy–Stieltjes transforms of radially symmetric distributions, which can expedite the verification of Assumptions 3.1 and 3.2 in a variety of situations. We note that Lemma 3.5 also appears as Proposition 3.1 in [29].

Lemma 3.5 (Computation of $m_\mu(\xi)$ for radially symmetric distributions)**.** *Suppose μ has a density $f(r,\theta) = f(r)$ that is radially symmetric about the origin. Then, $m_\mu(0) = 0$, and for $\xi \neq 0$,*

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

where $X \sim \mu$.

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

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

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

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

as is desired. Finally, observe that

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

$\square$

Example 3.6 (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

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

(Lemma 3.5 facilitates the computation of $m_\mu(z)$ when μ is radially symmetric. For this example, apply Lemma 3.5 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 3.1, and by Theorem 3.3, 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 3.7 (μ is supported on all of $\mathbb{C}$)**.** Assumption 3.2 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 3.4, with probability tending to 1, $W_1(\mu_n, \mu_n') \leq O((\log n)^{10}/n)$. Figure 1.3 illustrates this example.

Example 3.8 (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 1.4. By separately considering the cases $|z + 2| < 1$, $|z - 2| < 1$, and $|z \pm 2| \geq 1$, we can use the calculations from Example 3.6 to obtain the Cauchy–Stieltjes transform:

$$m_\mu(z) = \begin{cases} \frac{1}{2}\left(\overline{z+2} + \frac{1}{z-2}\right) & \text{if } |z+2| < 1, \\ \frac{z}{z^2-4} & \text{if } |z \pm 2| > 1, \\ \frac{1}{2}\left(\overline{z-2} + \frac{1}{z+2}\right) & \text{if } |z-2| < 1. \end{cases} \tag{3.5}$$

Since μ has compact support, Assumption 3.1 part (ii) holds trivially.

Next, we establish that μ satisfies part (i) of Assumption 3.1. Setting each of the three branches in (3.5) 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,

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

Hence, the reverse triangle inequality transforms our previous calculation into

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

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

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

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

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

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

We have verified Assumption 3.1, so by Theorem 3.3, with probability at least $1-o(1)$, $W_1(\mu_n,\mu_n') = O((\log n)^9/n)$.

3.2.3 Fluctuations of the critical points

While Theorem 3.3 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 $X_1,\dots,X_{n+1-s}$ are iid complex-valued random variables with distribution μ, and $\vec{\xi} = (\xi_1,\dots,\xi_s)$ is a deterministic vector in $\mathbb{C}^s$. Our goal is to simultaneously study the behavior of the critical points closest to ξ_l, $1 \le l \le s$.

Our first result, Theorem 3.9, covers the situation where $\xi_1,\dots,\xi_s$ are inside the support of μ. In particular, for each $1 \le l \le s$, equation (3.7) 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

$$\hat{w}_l^{(n)} := \xi_l - \frac{1}{n+1}\frac{n}{\sum_{j\neq l}\frac{1}{\xi_l-\xi_j}+\sum_{j=1}^{n+1-s}\frac{1}{\xi_l-X_j}}, \tag{3.6}$$

rather than at ξ_l, and we use this to show that the vector $(w_1^{(n)},\dots,w_s^{(n)})$ fluctuates around $(\hat{w}_1^{(n)},\dots,\hat{w}_s^{(n)})$ according to a law that converges in distribution to a multivariate normal distribution. See Figure 3.1.

In order to state Theorem 3.9 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 $B(z_0,\rho)$.

Theorem 3.9 (Locations and fluctuations of critical points when p_n has several deterministic roots)**.** *Let $X_1, X_2, \dots$ be iid complex-valued random variables with distribution μ, fix s and the distinct, deterministic values $\xi_1,\dots,\xi_s \notin \mathrm{M}_\mu$, 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*

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

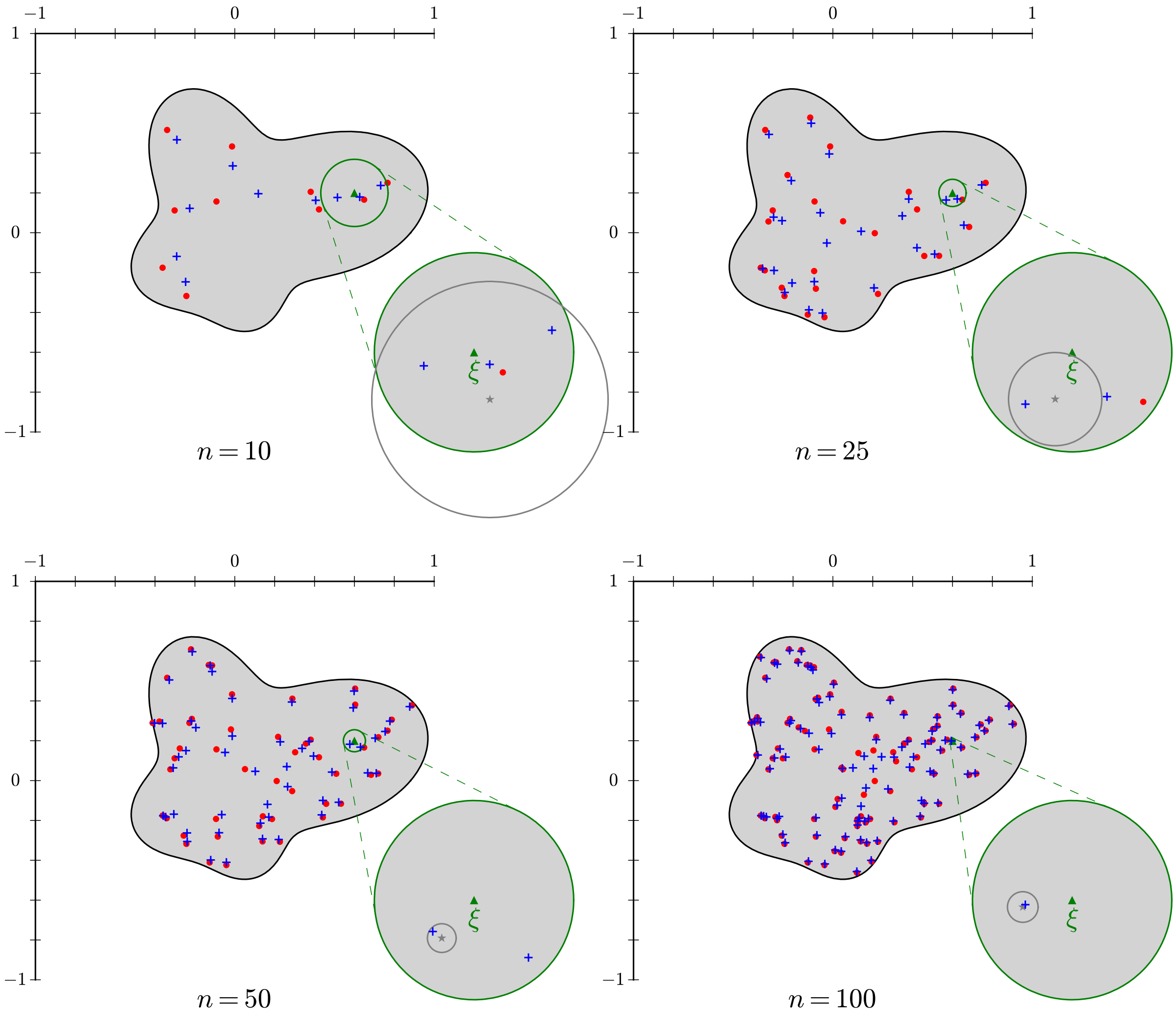


Figure 3.1: Simulation to illustrate Theorem 3.9. The roots (red dots) and critical points (blue crosses) of $p_n(z) = (z - \xi_1)\prod_{j=1}^{n}(z - X_j)$ for increasing values of n, where the roots, $X_1, \ldots, X_{100}$, are chosen independently and uniformly from the outlined region. The green circle 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 (3.6)).

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

$$\left| w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{n}{\sum_{j \neq l} \frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j}} \right| = O_{\mu, \vec{\xi}} \left(\left(\frac{\log n}{n} \right)^2 \right). \tag{3.7}$$

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

$$\left(\frac{n^{3/2}}{\sqrt{\log n}} \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right) \right)_{l=1}^{s} \longrightarrow (N_1, \dots, N_s) \tag{3.8}$$

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

$$\begin{aligned} \mathrm{Cov}(\mathrm{Re}(N_j), \mathrm{Re}(N_l)) &= \begin{cases} \frac{\pi f(\xi_j)}{2} & \text{if } l = j, \\ 0 & \text{else} \end{cases} \\ \mathrm{Cov}(\mathrm{Im}(N_j), \mathrm{Im}(N_l)) &= \begin{cases} \frac{\pi f(\xi_j)}{2} & \text{if } l = j, \\ 0 & \text{else} \end{cases} \\ \mathrm{Cov}(\mathrm{Re}(N_j), \mathrm{Im}(N_l)) &= 0. \end{aligned} \tag{3.9}$$

Remark 3.10. Theorem 3.9 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 3.9; a similar argument was used in [29].

Compare Theorem 3.9 to Theorem 2.2 of [29], 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 [29] 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 3.9 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 [29], our approach is based on a deterministic argument (see Theorem 3.23).

For values of $\xi_1, \dots, \xi_s$ outside the support of μ, (3.8) and (3.9) 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 $\operatorname{supp}(\mu)$, 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 3.2.

Theorem 3.11 (Locations and Fluctuations of critical points when p_n has several roots outside $\operatorname{supp}(\mu)$)**.** *Let $X_1, X_2, \dots$ be iid complex-valued random variables with common distribution μ, fix $s \in \mathbb{N}$, and suppose $\xi_1, \dots, \xi_s \notin \operatorname{supp}(\mu) \cup \mathrm{M}_\mu$ 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 $1 \le l \le s$, $w_l^{(n)}$ 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

$$\left| w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{n}{\sum_{j \neq l} \frac{1}{\xi_l - \xi_j} + \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j}} \right| < \frac{C_{\mu,\vec{\xi}}}{n^2}. \tag{3.10}$$

In addition, we have

$$\left(n^{3/2} \cdot m_\mu(\xi_l)^2 \cdot \left(w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{1}{m_\mu(\xi_l)} \right) \right)_{l=1}^{s} \longrightarrow (N_1, \dots, N_s) \tag{3.11}$$

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

$$\begin{aligned}
\operatorname{Cov}(\operatorname{Re}(N_j), \operatorname{Re}(N_l)) &= \operatorname{Cov}\left(\operatorname{Re}\left(\frac{1}{\xi_j - X_1} \right), \operatorname{Re}\left(\frac{1}{\xi_l - X_1} \right) \right) \\
\operatorname{Cov}(\operatorname{Im}(N_j), \operatorname{Im}(N_l)) &= \operatorname{Cov}\left(\operatorname{Im}\left(\frac{1}{\xi_j - X_1} \right), \operatorname{Im}\left(\frac{1}{\xi_l - X_1} \right) \right) \\
\operatorname{Cov}(\operatorname{Re}(N_j), \operatorname{Im}(N_l)) &= \operatorname{Cov}\left(\operatorname{Re}\left(\frac{1}{\xi_j - X_1} \right), \operatorname{Im}\left(\frac{1}{\xi_l - X_1} \right) \right).
\end{aligned} \tag{3.12}$$

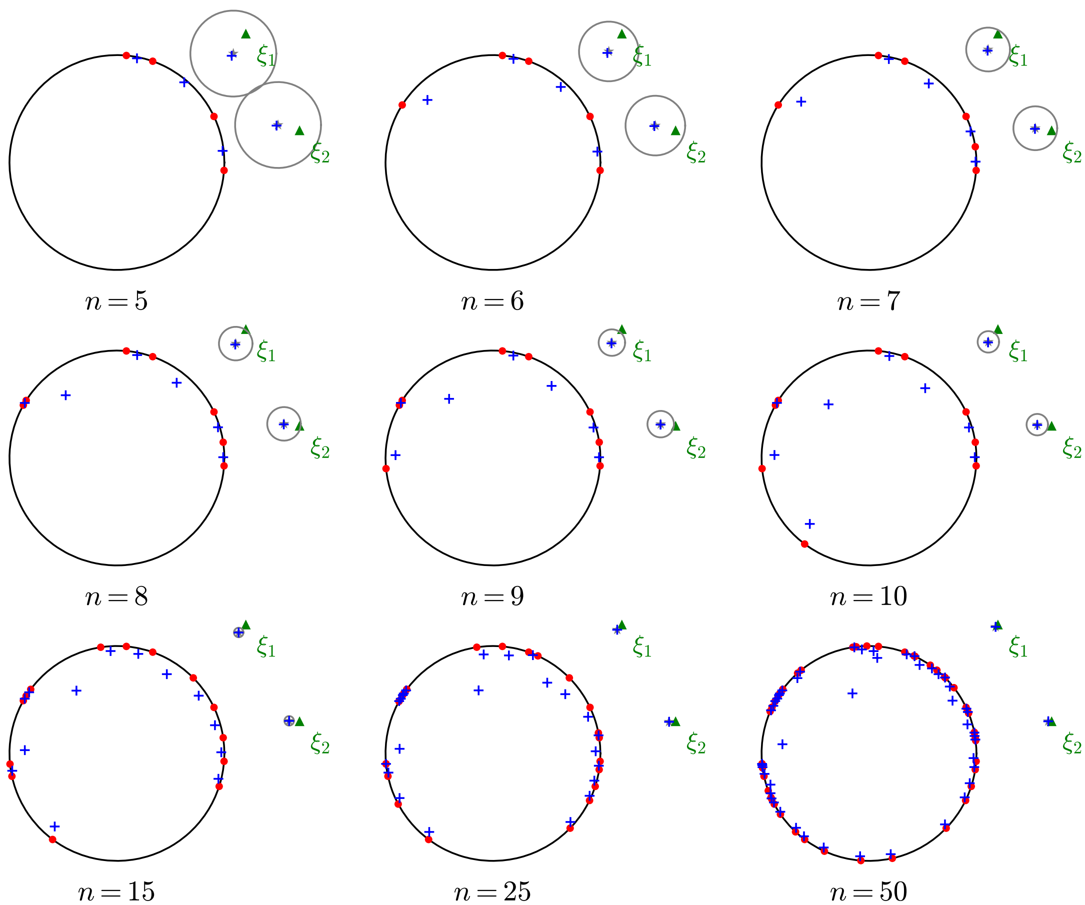


Figure 3.2: Simulation to illustrate Theorem 3.11. The roots (red circles) and critical points (blue crosses) of $p_n(z) = \prod_{l=1}^{2}(z - \xi_l)\prod_{j=1}^{n-1}(z - X_j)$ for increasing values of n, where the roots, $X_1, \ldots, X_{49}$, are chosen independently and uniformly from the unit circle. The gray circles are of radius $\frac{10}{n^2}$ and are centered at $\hat{w}_1^{(n)}$, $\hat{w}_2^{(n)}$ defined in (3.6).

Remark 3.12. After an application of the Borel–Cantelli lemma, Theorem 3.11 can be combined with Theorem 2.9 above to establish the following: when μ has compact support, almost surely, for n sufficiently large, $w_1^{(n)}, \dots, w_s^{(n)}$, characterized by (3.10), are the only critical points of p_n outside an ε-neighborhood of $\operatorname{supp}(\mu) \cup M_\mu$.

In Section 3.3, we provide a generalization of Theorem 3.11 to a situation where p_n has a number of deterministic roots that may depend on n (see Theorem 3.25 below). The proofs of Theorems 3.9 and 3.11 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 3.23). 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 3.13. *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 $O(1)$, rank $O(1)$, and whose only nonzero eigenvalue is λ. Then almost surely, for n sufficiently large, the characteristic polynomial*[1]

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

of $\frac{1}{\sqrt{n}} M + A$ satisfies the following properties:

(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 $\xi = \lambda + o(1)$.

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

$$\left| w_\xi^{(n-1)} - \xi + \frac{1}{n} \cdot \frac{1}{\frac{1}{n-1} \sum_{i=1}^{n-1} \frac{1}{\xi - X_i}} \right| = O_{\lambda,\varepsilon}\left(\frac{1}{n^2}\right). \tag{3.13}$$

and hence

$$w_\xi^{(n-1)} = \lambda + o(1) \tag{3.14}$$

[1] Here, I denotes the identity matrix.

Remark 3.14. The conclusion in (3.14) can be deduced from properties (i) and (ii) and Walsh's two circle theorem (see, for example, [46, Theorem 4.1.1]). However, the conclusion in (3.13) cannot be deduced from Walsh's two circle theorem and instead follows from Theorem 3.23. We prove Theorem 3.13 in Section 3.3.

3.2.4 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 $X_1, \ldots, X_n$ are arbitrary random variables (not assumed to be independent nor identically distributed) and then specialize our main result to several applications and examples.

Theorem 3.15 (Local law). *Fix $C > 0$, and let $X_1, \ldots, X_n$ be complex-valued random variables (not necessarily independent nor identically distributed) which satisfy the following axioms.*

(i) (Upper bound) With overwhelming probability,

$$\max_{1 \leq i \leq n} |X_i| \leq e^{n^C},$$

(ii) (Anti-concentration) For every $a > 0$, there exists $b > 0$ such that

$$\left| \sum_{i=1}^{n} \frac{1}{Z - X_i} \right| \geq n^{-b} \tag{3.15}$$

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

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

$$|\Delta\varphi(z)| \leq n^C \tag{3.16}$$

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 3.16. Condition (ii) on the random variables $X_1, \dots, X_n$ from Theorem 3.15 is implied by the following:

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

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

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 3.15 are fairly technical, and we derive some simpler conditions that guarantee when the hypotheses of Theorem 3.15 are met in Section 3.2.5. We now specialize Theorem 3.15 to the case where $X_1, \dots, X_n$ are independent random variables.

Theorem 3.17 (Local law for independent roots)**.** *Fix $C > 0$, and let $X_1, \dots, X_n$ be independent complex-valued random variables which satisfy*

$$\max_{1 \leq i \leq n} \mathbb{E}|X_i| \leq n^C.$$

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 (3.16) 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 $w_1^{(n)}, \ldots, w_{n-1}^{(n)}$ 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 3.17 can be viewed as a local version of Theorem 1.3 and (1.2). 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 3.17 is significantly better then the error term from (1.2).

Interestingly, Theorem 3.17 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 3.17 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 3.17 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 3.17 strengthens Theorem 1.6 of [7] for the empirical distribution associated with the zeros of p_n' by providing a rate of convergence. As a consequence of Theorem 3.17, we have the following central limit theorem (CLT).

Theorem 3.18 (Central limit theorem for linear statistics)**.** *Let $X_1, X_2, \ldots$ 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,*

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

in distribution as $n \to \infty$, where $w_1^{(n)}, \ldots, w_{n-1}^{(n)}$ 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 3.17 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 3.17.

Theorem 3.19 (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*

$$|m_\mu(z)| \geq c \tag{3.17}$$

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 $w_1^{(n)}, \ldots, w_{n-1}^{(n)}$ are the critical points of the polynomial

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

and $X_1, \ldots, X_n$ are iid random variables with distribution μ. Here, the implicit constants in our asymptotic notation depend on C, c, and ε.

3.2.5 Guaranteeing the assumptions in the local law

In this section, we provide some criteria for assuring the assumptions in Theorem 3.15 are met.

Lemma 3.20 (Simple criterion for an upper bound)**.** *Fix $C, \varepsilon > 0$, and suppose $X_1, \ldots, X_n$ are complex-valued random variables (not necessarily independent nor identically distributed). If*

$$\max_{1 \leq i \leq n} \mathbb{E}|X_i|^\varepsilon \leq n^C,$$

then

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

with overwhelming probability.

Proof. As

$$\mathbb{P}\left(\max_{1\leq i\leq n} |X_i| > e^{n^C}\right) \leq \sum_{i=1}^{n} \mathbb{P}(|X_i| > e^{n^C}),$$

the claim follows from a simple application of Markov's inequality. □

Lemma 3.21 (Criterion for anti-concentration)**.** *Fix $C > 0$, and let $X_1, \ldots, X_n$ be complex-valued random variables such that X_1 is independent of $X_2, \ldots, X_n$. 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 $\mathbb{E}|X_1| \leq n^C$. Then for every $a > 0$, there exists $b > 0$ such that*

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

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

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, $Z \notin \{X_1, \ldots, X_n\}$. Hence the sum

$$\sum_{i=1}^{n} \frac{1}{Z - X_i}$$

is well-defined and finite. By conditioning on the values of $X_2, \ldots, X_n$ 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 3.22 below by taking $\varepsilon := n^{-b}$ and choosing b sufficiently large in terms of C and a. □

Lemma 3.22. *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. □

3.2.6 Overview and outline

The remainder of the chapter is devoted to proving our main results. In Section 3.3, we establish Theorems 3.9, 3.11, and 3.13 of Subsection 3.2.3 by way of Theorem 3.23 for deterministic polynomials, which we also use to prove a generalization to Theorem 3.11. Section 3.4 contains the proofs of the local laws from Subsection 3.2.4 including those for Theorems 3.15, 3.17, 3.18, and 3.19. We conclude the paper with a proof of Theorem 3.3 in Section 3.5. Note that Appendix A contains some classical arguments that establish a Lindeberg CLT that we use to prove part of Theorem 3.9.

3.3 Proof of results in Section 3.2.3

The proofs of Theorems 3.9, 3.11, and 3.13 rely on the following theorem for deterministic polynomials.

Theorem 3.23. *Suppose ξ is a complex number, $\vec{X} = (X_1, X_2, \ldots, X_n)$ is a vector of complex numbers, and C_1, C_2, k_{Lip} are positive values for which the following three conditions hold:*

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

(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*

$$\left\{ z \in \mathbb{C} : |z - \xi| \leq \frac{2}{C_1 n} \right\};$$

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

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

$$C > \frac{8(1 + 2C_2^2)}{C_1^3} \quad \textit{and} \quad n > 4C_2 \max\left\{ \frac{1}{C_1}, C(k_{Lip} + 1) \right\}, \tag{3.18}$$

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*

$$\left| w_\xi^{(n)} - \xi + \frac{1}{n+1} \frac{1}{\frac{1}{n}\sum_{j=1}^n \frac{1}{\xi - X_j}} \right| < \frac{C(k_{Lip}+1)}{n^2}. \tag{3.19}$$

We remark that criteria (i) and (ii) appear relevant in view of the equality

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

which suggests that if $\frac{1}{n}\sum_{j=1}^n \frac{1}{z-X_j}$ is finite and bounded away from zero near ξ, then $p_n'(z) \approx 0$ for some z satisfying $|z - \xi| = O(1/n)$. 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 (3.19), (iii) is likely an artificial constraint related to the use of Rouché's theorem in the proof. We prove Theorem 3.23 in the next subsection.

3.3.1 Proof of Theorem 3.23

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

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

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 3.24. *Under the assumptions of Theorem 3.23:*

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

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

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

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

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

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

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

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

and the assumption $n > 4C_2C(k_{\text{Lip}}+1)$ 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 3.24 imply that under the assumptions of Theorem 3.23, $L_n(z)$ 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 3.23 and let $z \in \partial D_n^{\text{sm}} \cup \partial D_n^{\text{lg}}$. Then, the triangle inequality implies

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

where we have used hypothesis (ii) of Theorem 3.23 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

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

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 3.23 twice yields

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

Finally, we can use the reverse triangle inequality and hypothesis (i) of Theorem 3.23 to show

$$\begin{aligned}\left|L_n(z) - \widetilde{L}_n(z)\right| &\le nk_{\mathrm{Lip}}\,|z-\xi| + nC_2^2\,|z-\xi| \cdot \frac{1}{1-\left|(z-\xi)\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi-X_j}\right|} \\ &\le n\,|z-\xi|\left(k_{\mathrm{Lip}} + \frac{C_2^2}{1-|z-\xi|\,C_2}\right).\end{aligned} \tag{3.20}$$

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 3.24 guarantees that $|z-\xi| < \frac{2}{nC_1}$, and the hypotheses of Theorem 3.23 require that $\frac{1}{2} > \frac{2C_2}{nC_1}$, so we obtain

$$\left|L_n(z) - \widetilde{L}_n(z)\right| < \frac{2}{C_1}\left(k_{\mathrm{Lip}} + 2C_2^2\right) \le \frac{2}{C_1}(k_{\mathrm{Lip}}+1)(1+2C_2^2). \tag{3.21}$$

On the other hand,

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

where the last inequality follows from Lemma 3.24. One of the assumptions in Theorem 3.23 is that $C > \frac{8(1+2C_2^2)}{C_1^3}$, so

$$\left|\widetilde{L}_n(z)\right| > \frac{C_1^2}{4}(k_{\mathrm{Lip}}+1)\frac{8(1+2C_2^2)}{C_1^3} = \frac{2}{C_1}(k_{\mathrm{Lip}}+1)(1+2C_2^2). \tag{3.22}$$

Combining (3.21) and (3.22) 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 3.24 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 3.24), 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 3.24 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 (3.20), in the case where $|z-\xi| = \frac{3}{2C_1 n}$, we obtain

$$\left|L_n(z) - \widetilde{L}_n(z)\right| < \frac{3}{2C_1}(k_{\text{Lip}} + 2C_2^2) \leq \frac{3}{2C_1}(k_{\text{Lip}}+1)(1+2C_2^2), \tag{3.23}$$

where we have once again used the assumption that $\frac{1}{2} \geq \frac{2C_2}{nC_1}$. 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 3.24, (ii). From the assumptions on n and C in Theorem 3.23, it follows that

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

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

$$\left|\widetilde{L}_n(z)\right| > \frac{64}{15C_1}(k_{\text{Lip}}+1)(1+2C_2^2). \tag{3.24}$$

Combining (3.23) and (3.24) yields $\left|L_n(z) - \widetilde{L}_n(z)\right| < \left|\widetilde{L}_n(z)\right|$ for z in the boundary of D_n^{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 3.24, (ii), Rouché's theorem guarantees that $(z-\xi)L_n(z)$, $(z-\xi)\widetilde{L}_n(z)$ have the same numbers 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 3.24, (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 3.23 is complete.

In the remainder of this section, we use Theorem 3.23 to prove Theorems 3.9, 3.11 and 3.13. We also include a subsection where we sketch how the arguments could be modified to prove Theorem 3.25, which generalizes part of Theorem 3.11 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 3.11, which is more straightforward than the justification of Theorem 3.9.

3.3.2 Proof of Theorem 3.11

We begin by establishing equation (3.10) via Theorem 3.23. 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 3.23. Fix ξ_l, $1 \le l \le s$. We will show that for large n, on the complement of the "bad" event

$$E_n^l := \left\{ \left| \frac{1}{n+1-s} \sum_{j=1}^{n+1-s} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right| \ge \frac{|m_\mu(\xi_l)|}{2.1} \right\},$$

the hypotheses of Theorem 3.23 are satisfied with $\xi = \xi_l$,

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

and the positive constants

$$C_{1,l} := \frac{|m_\mu(\xi_l)|}{2},\ C_{2,l} := \frac{3\,|m_\mu(\xi_l)|}{2},\ k_{\text{Lip},l} := \frac{9}{\text{dist}(\xi_l, \text{supp}(\mu) \cup \{\xi_j : j \ne l\})^2}. \tag{3.25}$$

(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,

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

$$\leq\left|\frac{1}{n}\sum_{\substack{j=1\\j\neq l}}^{s}\frac{1}{\xi_l-\xi_j}\right|+\frac{n+1-s}{n}\left|\frac{1}{n+1-s}\sum_{j=1}^{n+1-s}\frac{1}{\xi_l-X_j}\right| \tag{3.26}$$

$$\leq o_l(1)+\left|\frac{1}{n+1-s}\sum_{j=1}^{n+1-s}\frac{1}{\xi_l-X_j}-m_\mu(\xi_l)\right|+|m_\mu(\xi_l)|\leq C_{2,l}$$

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

$$\left|\frac{1}{n}\left(\sum_{\substack{j=1\\j\neq l}}^{s}\frac{1}{\xi_l-\xi_j}+\sum_{j=1}^{n+1-s}\frac{1}{\xi_l-X_j}\right)\right|\geq C_{1,l}, \tag{3.27}$$

and condition (i) of Theorem 3.23 follows from equations (3.26) and (3.27). 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 3.23 holds, and for $|z-\xi_l|\leq\frac{2}{C_{1,l}n}$,

$$\min_{1\leq j\leq n+1-s}|z-X_j|\geq\min_{1\leq j\leq n+1-s}|\xi_l-X_j|-|z-\xi_l|\geq\varepsilon_l-\frac{2}{C_{1,l}n}>\frac{\varepsilon_l}{3},$$

$$\min_{j\neq l}|z-\xi_j|\geq\min_{j\neq l}|\xi_l-\xi_j|-|z-\xi_l|\geq\varepsilon_l-\frac{2}{C_{1,l}n}>\frac{\varepsilon_l}{3}.$$

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

$$\left|\frac{1}{n}\left(\sum_{\substack{j=1\\j\neq l}}^{s}\frac{1}{z-\xi_j}+\sum_{j=1}^{n+1-s}\frac{1}{z-X_j}\right)-\frac{1}{n}\left(\sum_{\substack{j=1\\j\neq l}}^{s}\frac{1}{z-\xi_j}+\sum_{j=1}^{n+1-s}\frac{1}{w-X_j}\right)\right|$$

$$=\frac{1}{n}\left|\sum_{\substack{j=1\\j\neq l}}^{s}\frac{w-z}{(z-\xi_j)(w-\xi_j)}+\sum_{j=1}^{n+1-s}\frac{w-z}{(z-X_j)(w-X_j)}\right|$$

$$\leq|w-z|\cdot\frac{9}{\varepsilon_l^2}=k_{\mathrm{Lip},l}\cdot|w-z|,$$

which implies condition (ii) of Theorem 3.23.

Now, fix any $C > \max_{1\leq l\leq s} \frac{8(1+2C_{2,l}^2)}{C_{1,l}^3}$. If n is a natural number large enough to guarantee inequalities (3.26) and (3.27) for $1 \leq l \leq s$ and that satisfies

$$n > \max\left\{\frac{4C_{2,l}}{C_{1,l}}, 4C_{2,l}C(k_{\mathrm{Lip},l}+1), \frac{3}{C_{1,l}\varepsilon_l} : 1 \leq l \leq s\right\}, \tag{3.28}$$

Theorem 3.23 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 $1 \leq l \leq s$, $w_l^{(n)}$ 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

$$\left|w_l^{(n)} - \xi_l + \frac{1}{n+1}\frac{n}{\sum_{j\neq l}\frac{1}{\xi_l-\xi_j} + \sum_{j=1}^{n+1-s}\frac{1}{\xi_l - X_j}}\right| < \frac{C(k_{\mathrm{Lip},l}+1)}{n^2}. \tag{3.29}$$

(Note that for large n, $w_1^{(n)}, \ldots, w_s^{(n)}$ are distinct because $\xi_1, \ldots, \xi_s$ are distinct and (3.29) implies $w_l^{(n)} \to \xi_l$ for $1 \leq l \leq s$.) We complete our justification of (3.10) from Theorem 3.11 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 $1 \leq l \leq s$, 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 Lemma 2.15 applies with $t_l := \frac{|m_\mu(\xi_l)|}{2.1}$. By Lemma 2.15, 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 (3.10). It remains to show that they satisfy the convergence in (3.11). 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 (3.10) for $1 \leq l \leq n$. It follows that with probability 1, for sufficiently large n and any l, $1 \le l \le s$,

$$\begin{aligned}
&\sqrt{n}(n+1)\cdot m_\mu(\xi_l)^2\cdot\left(w_l^{(n)}-\xi_l+\frac{1}{n+1}\frac{1}{m_\mu(\xi_l)}\right)\\
&\quad= m_\mu(\xi_l)^2\cdot\sqrt{n}\left(\frac{1}{m_\mu(\xi_l)}-\frac{n}{\sum_{j\neq l}\frac{1}{\xi_l-\xi_j}+\sum_{j=1}^{n+1-s}\frac{1}{\xi_l-X_j}}\right)+O_{\mu,\vec{\xi}}(n^{-1/2})\\
&\quad= m_\mu(\xi_l)^2\cdot\sqrt{n}\left(\frac{1}{m_\mu(\xi_l)}-\frac{1}{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi_l-X_j}+O_{\mu,\vec{\xi}}(1/n)}\right)+O_{\mu,\vec{\xi}}(n^{-1/2})\\
&\quad= \frac{m_\mu(\xi_l)}{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi_l-X_j}+O_{\mu,\vec{\xi}}(1/n)}\cdot\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi_l-X_j}-m_\mu(\xi_l)\right)+O_{\mu,\vec{\xi}}(n^{-1/2}).
\end{aligned}\tag{3.30}$$

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

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

Now, we will use the Cramér–Wold device (see e.g. Theorem 29.4 in [5]) to show the convergence (3.11). To start, let $t_1,\dots,t_s,r_1,\dots,r_s$ be arbitrary real numbers and define the random variables

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

for $1 \le l \le s$. By (3.30), we have, with probability tending to 1,

$$\begin{aligned}
Y_{n,l} &= \frac{n}{n+1}\frac{m_\mu(\xi_l)}{\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi_l-X_j}+o(1)}\cdot\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi_l-X_j}-m_\mu(\xi_l)\right)+O\left(\frac{1}{\sqrt{n}}\right)\\
&= (1+o(1))\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi_l-X_j}-m_\mu(\xi_l)\right)+O\left(\frac{1}{\sqrt{n}}\right)\\
&= \sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\xi_l-X_j}-m_\mu(\xi_l)\right)+o(1),
\end{aligned}\tag{3.31}$$

where all of the implied constants depend on $\xi_1,\dots,\xi_s$ and μ, and we have made ample use of Slutsky's theorem (see e.g. Theorem 11.4 from [20]). To obtain the last line, we also used the classical CLT (see e.g. Theorem 29.5 from [5]) 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

$$\operatorname{Var}\left(\sum_{l=1}^{s} [t_l Z_{l,1} + r_l W_{l,1}] \right).$$

This limiting distribution is also the distribution of the random variable

$$\sum_{l=1}^{s} [t_l \operatorname{Re}(N_l) + r_l \operatorname{Im}(N_l)]$$

with covariance structure given by (3.12), so by the Cramér–Wold strategy, the proof of Theorem 3.11 is complete.

The next subsection illustrates how to modify the argument above to prove a generalization of Theorem 3.25 to the case where p_n has a number of deterministic roots that may grow with n.

3.3.3 Generalization of Theorem 3.11

The following result shows how Theorem 3.23 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 [23]. 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 3.25 describes the locations of the critical points with higher precision than Theorem 2 of [23], however our theorem requires that the deterministic roots $\xi_1, \dots, \xi_{s_n}$ be outside the support of μ, while Theorem 2 in [23] doesn't make this restriction.

Theorem 3.25 (Locations of critical points when p_n has many deterministic roots.)**.** *Suppose $X_1, X_2, \dots$ are iid complex-valued random variables with distribution μ, let $\xi_1, \xi_2, \dots$ be fixed deter-*

ministic 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) $1 \leq s_n \leq l_n = o(n)$, $a_n l_n = o(n)$, $a_n = o(\sqrt{n})$;

(ii) $\min\{|m_\mu(\xi_l)| : 1 \leq l \leq s_n\} \geq \varepsilon$ *and* $\max\{|m_\mu(\xi_l)| : 1 \leq l \leq s_n\} \leq L$;

(iii) $\min\left\{|\xi_l - x| : 1 \leq l \leq s_n,\ x \in \operatorname{supp}(\mu) \cup \{\xi_j\}_{j=1, j\neq l}^{l_n}\right\} > \frac{6}{\varepsilon \cdot a_n}$.

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 $1 \leq l \leq s_n$, $w_l^{(n)}$ is the unique critical point of p_n within $\frac{3}{2\varepsilon n}$ of ξ_l and

$$\max_{1 \leq l \leq s_n} \left| w_l^{(n)} - \xi_l + \frac{1}{n+1} \frac{n}{\sum_{k=1,k\neq l}^{l_n} \frac{1}{\xi_l - \xi_k} + \sum_{j=1}^{n+1-l_n} \frac{1}{\xi_l - X_j}} \right| < \frac{C_{\mu,\varepsilon,L} \cdot a_n^2}{n^2}. \tag{3.32}$$

Theorem 3.25 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 3.3.2 for each l, $1 \leq l \leq s_n$, separately but in place of the definitions in equation (3.25) 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

$$E_n^l := \left\{ \left| \frac{1}{n+1-l_n} \sum_{j=1}^{n+1-l_n} \frac{1}{\xi_l - X_j} - m_\mu(\xi_l) \right| < \frac{|m_\mu(\xi_l)|}{2.1} \right\},\ 1 \leq l \leq s_n.$$

Notice that condition (i) from Theorem 3.23 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 (3.28) on n now become

$$n > \max\left\{\frac{4C_2}{C_1}, 4C_2C\left(\frac{\varepsilon^2 a_n^2}{4}+1\right), a_n\right\},$$

which hold uniformly for $1 \leq l \leq s_n$ by assumption (i) in the statement of Theorem 3.25. By Hoeffding's inequality (Lemma 2.15), with $Y_j^l := \frac{1}{\xi_l - X_j}$, $K_l := \frac{\varepsilon a_n}{6}$, and $t_l := \frac{|m_\mu(\xi_l)|}{2.1} > \frac{\varepsilon}{2}$, there are constants $C, c_{\mu,\varepsilon} > 0$, independent of l, ξ_l, and s_n, so that for large n

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

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

3.3.4 Proof of Theorem 3.9

We now proceed to prove Theorem 3.9. 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 3.26 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 3.26 is similar to Lemma 5.7 in [29].

Lemma 3.26 (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

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

by considering the expectations at right on 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 has probability 1. By the triangle inequality, we have

$$\begin{aligned}
&|m_\mu(x) - m_\mu(y)| \\
&\quad\le |x-y| \, \mathbb{E}\left[\frac{1}{|x-Z|\,|y-Z|} \mathbb{1}_{\mathcal{A}}\right] + \mathbb{E}\left[\frac{1}{|x-Z|} \mathbb{1}_{\mathcal{A}^c}\right] + \mathbb{E}\left[\frac{1}{|y-Z|} \mathbb{1}_{\mathcal{A}^c}\right].
\end{aligned} \tag{3.33}$$

We will bound each term separately as follows. Via Cauchy–Schwarz, we have

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

Next, observe that

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{|x-Z|}\mathbb{1}_{\mathcal{A}^c}\right] &\leq \mathbb{E}\left[\frac{1}{|x-Z|}\mathbb{1}_{\mathcal{B}}\right] + \mathbb{E}\left[\frac{1}{|x-Z|}\mathbb{1}_{\mathcal{C}\cup\mathcal{D}}\right] \\
&\leq \mathbb{E}\left[\frac{1}{|y-Z|}\mathbb{1}_{|y-Z|<|x-y|}\right] + \mathbb{E}\left[\frac{1}{|x-Z|}\mathbb{1}_{|x-Z|<|x-y|}\right] \\
&\leq 4\pi C_\mu \int_0^{|x-y|} \frac{1}{r} r\, dr \\
&= 4\pi C_\mu |x-y|.
\end{aligned}$$

For similar reasons,

$$\mathbb{E}\left[\frac{1}{|y-Z|}\mathbb{1}_{\mathcal{A}^c}\right] \leq 4\pi C_\mu |x-y|,$$

and we can combine the last few inequalities to obtain

$$|m_\mu(x) - m_\mu(y)| \leq |x-y| \left(2\pi C_\mu \log\left|\frac{1}{x-y}\right| + 1 + 8\pi C_\mu\right).$$

The proof of Lemma 3.26 is complete. □

We proceed to prove Theorem 3.9, starting with a justification of (3.7) 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 3.23 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 (3.7).

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

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

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

$$C_1 := \frac{|m_\mu(\xi)|}{2}, \quad C_2 := \frac{3\,|m_\mu(\xi)|}{2}, \quad \text{and} \quad k_{\mathrm{Lip}} := \frac{C_{\mu,\xi} \log n}{\varepsilon_n^{3/2}}, \tag{3.34}$$

for $\varepsilon_n := (\log n)^{-2/3}$ and $C_{\mu,\xi}$ defined in Lemma 3.27 below, then the conditions in Theorem 3.23 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 3.23, 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 3.27. *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

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

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 $z, w \in \overline{B}\left(\xi, \frac{2}{C_1 n}\right)$,

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

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 3.28.

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

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 3.26. Next, consider that for large n,

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

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

We now turn our attention to the events F_n^k. For $0 \le k \le d_n$ and $1 \le j \le n$, 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 $p_k \le \pi C_f \rho_\xi^2 e^{2k}/n$. 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

$$\mathbb{P}\left(N_n^k \ge \pi C_f \rho_\xi^2 e^{2k} + \frac{e^k}{\sqrt{\varepsilon_n}}\right) \le \frac{\varepsilon_n^2 \mathbb{E}\left[\left(N_n^k - \mathbb{E}[N_n^k]\right)^4\right]}{e^{4k}}. \tag{3.35}$$

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

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

Consequently, (3.35) 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 3.28 is complete. $\square$

We have established that C_1, C_2, and k_{Lip} defined in (3.34) satisfy conditions (i), (ii), and (iii) of Theorem 3.23 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 3.23 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 (3.7).

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

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

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

$$E_{l,n} = \left\{\left|\frac{1}{n+1-s}\sum_{j=1}^{n+1-s}\frac{1}{\xi_l-X_j} - m_\mu(\xi_l)\right| < \frac{|m_\mu(\xi_l)|}{2.1}\right\},\ 1\le l\le s;$$
$$F_{l,n}^k = \left\{N_{l,n}^k \ge \pi C_f\delta^2 e^{2k} + \frac{\delta e^k}{\sqrt{\varepsilon_n}}\right\},\ 0\le k\le d_n,\ 1\le l\le s;$$
$$G_{l,n} = \left\{\min_{1\le j\le n}|X_j-\xi_l| < \sqrt{\frac{\varepsilon_n}{n}}\right\},\ 1\le l\le s;$$

and the modified constants

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

(Note that $C_{\mu,\vec{\xi}}^l$, $1\le l\le s$ will be defined via lemmata similar to Lemma 3.27.) 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 3.23 hold for reasons similar to those given in the argument for $s=1$ above. (Notice that for $1\le l\le s$,

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

so computations similar to (3.26) and (3.27) establish condition (i) of Theorem 3.23.) The fact that the union of the modified "bad" events occurs with probability at most $o(1)$ follows by an updated version of Lemma 3.28 and the union bound (recall s is fixed and finite).

We now turn our attention to (3.8) which describes the joint fluctuations of $w_l^{(n)}$, $1 \le l \le s$. This is considerably more difficult than our consideration of (3.11) because in the current situation, $(\xi_l - X_j)^{-1}$, are heavy-tailed random variables. In Appendix A, we appeal to the Lindeberg exchange method with an appropriate truncation to establish Theorem A.1, a CLT that we use to prove (3.8) in a similar manner to our justification of (3.11).

To start, consider that with probability $1 - o(1)$, $w_l^{(n)}$, $1 \le l \le s$ satisfy (3.7), so with inspiration from (3.30) and (3.31), 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 A.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 A.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 (3.8) with covariance structure (3.9). Recall that studying the real parts of the linear combinations over $\mathbb{C}$ of a s-complex dimensional random vector is the same as analyzing the linear combinations over $\mathbb{R}$ of a $2s$-real dimensional random vector. Thus, we can apply the Cramér–Wold technique to conclude our justification of Theorem 3.9.

3.3.5 Proof of Theorem 3.13

We conclude this section by using Theorem 3.23 to prove Theorem 3.13.

Proof. Conclusions (i) and (ii) follow from [54, Theorem 1.7]. We now use Theorem 3.23 to establish (3.13). In particular, we will verify the three conditions of Theorem 3.23 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

$$\max_{1\leq i\leq n-1} |X_i| \leq 1+2\varepsilon, \quad \min_{1\leq i\leq n-1} |\xi - X_i| \geq \frac{\varepsilon}{2}, \quad 1+\frac{11}{4}\varepsilon \leq |\xi| \leq |\lambda|+1. \tag{3.36}$$

In fact, this event automatically guarantees the third condition from Theorem 3.23 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

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

on the same event. The upper bound in the first condition of Theorem 3.23 follows from a similar argument. The lower bound, however, is slightly more involved. Indeed, for any $\theta \in \mathbb{R}$, we have

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

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

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

Thus, on the event (3.36), we conclude that

$$\left|\frac{1}{n}\sum_{i=1}^{n-1}\frac{1}{\xi-X_i}\right| \gg_{\varepsilon,\lambda} 1, \tag{3.37}$$

which completes the proof of the lower bound. Hence, the three conditions of Theorem 3.23 are satisfied. Applying Theorem 3.23, we obtain (3.13). Lastly, (3.14) follows from (3.13) after applying conclusion (ii) and (3.37). □

3.4 Proof of results in Section 3.2.4

3.4.1 Proof of Theorem 3.15

This section is devoted to the proof of Theorem 3.15. We will need the following lemmata.

Lemma 3.29 (Monte Carlo sampling; Lemma 36 from [58])**.** *Let (X, μ) be a probability space, and let $F : X \to \mathbb{C}$ be a square-integrable function. Let $m \geq 1$, let $x_1, \ldots, x_m$ 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

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

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 3.30. *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,*

$$\max_{1 \leq i \leq n} |X_i| \leq e^{n^C}. \tag{3.38}$$

Let $\varphi : \mathbb{C} \to \mathbb{R}$ be a twice continuously differentiable function (possibly depending on n) which satisfies the pointwise bound in (3.16) for all $z \in \mathbb{C}$. Then, with overwhelming probability,

$$\int_{B(0,n^C)} |\Delta\varphi(z)|^2 \log^2 |p_n(z)| d^2z \ll n^{2C} n^{O(1)}, \tag{3.39}$$

$$\int_{B(0,n^C)} |\Delta\varphi(z)|^2 \log^2 |p_n'(z)| d^2z \ll n^{2C} n^{O(1)}, \tag{3.40}$$

and

$$\int_{B(0,n^C)} |\Delta\varphi(z)|^2 d^2z \ll n^{4C}. \tag{3.41}$$

Proof. The bound in (3.41) follows immediately from the pointwise bound in (3.16). In order to prove (3.39) it suffices, by the pointwise bound in (3.16), to prove that with overwhelming probability

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

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

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

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

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

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

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

since $\log |\cdot|$ is locally square-integrable. This completes the proof of (3.39).

For (3.40), we observe that on the event where (3.38) holds, the Gauss–Lucas theorem implies that

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

where $w_1^{(n)}, \ldots, w_{n-1}^{(n)}$ are the critical points of p_n. Working on this event, the proof follows from the same procedure as we used to prove (3.39); we omit the details. $\square$

Lemma 3.31 (Crude upper bound). *Fix $C > 0$, and let $X_1, \ldots, X_n$ be complex-valued random variables (not necessarily independent nor identically distributed). Assume Z is uniformly distributed on $B(0, n^C)$, independent of $X_1, \ldots, X_n$. Then for every $a > 0$, there exits $b > 0$ such that*

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

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

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

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

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. $\square$

We now prove Theorem 3.15.

Proof of Theorem 3.15. 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

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

(These identities can also be found in a more general form in [26, 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 3.29. Indeed, let $m := n^\beta$, and let $Z_1, \ldots, Z_m$ be iid random variables uniformly distributed on B, independent of $X_1, \ldots, X_n$. Taking β sufficiently large and applying Lemmas 3.29 and 3.30, we conclude that

$$\frac{2\pi}{|B|} \sum_{i=1}^{n} \varphi(X_i) = \frac{1}{m} \sum_{l=1}^{m} \Delta\varphi(Z_l) \log |p_n(Z_l)| + O(n^{-c-2C}), \tag{3.42}$$
$$\frac{2\pi}{|B|} \sum_{j=1}^{n-1} \varphi(w_j) = \frac{1}{m} \sum_{l=1}^{m} \Delta\varphi(Z_l) \log |p_n'(Z_l)| + O(n^{-c-2C}), \tag{3.43}$$
$$\frac{1}{|B|} \int_B |\,\Delta\varphi(z)| d^2 z = \frac{1}{m} \sum_{l=1}^{m} |\,\Delta\varphi(Z_l)| + O(n^{-c-2C-1}) \tag{3.44}$$

with probability $1 - O(n^{-\alpha})$. In addition, by (3.15), Lemma 3.31, 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

$$\sup_{1 \le l \le m} \left| \log |p_n(Z_l)| - \log |p_n'(Z_l)| \right| = O(\log n) \tag{3.45}$$

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

From (3.42) and (3.43), we find

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

with probability $1 - O(n^{-\alpha})$. Applying (3.44) and (3.45) yields

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

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

$$\left| \sum_{i=1}^{n} \varphi(X_i) - \sum_{j=1}^{n-1} \varphi(w_j) \right| \ll (\log n) \|\, \Delta\varphi\|_1 + n^{-c} \tag{3.46}$$

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

3.4.2 Proof of Theorems 3.17 and 3.18

In order to prove Theorem 3.17, it suffices to show that $X_1, \ldots, X_n$ satisfy the two axioms of Theorem 3.15. This follows from Lemmas 3.20 and 3.21.

We now turn to the proof of Theorem 3.18. By Theorem 3.17,

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

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

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

Therefore, we conclude that

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

with probability $1 - O(n^{-100})$. By the classical CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\varphi(X_i) - \mathbb{E}\varphi(X_i)) \longrightarrow N(0, v^2)$$

in distribution as $n \to \infty$, where v^2 is the variance of $\varphi(X_1)$, and the claim follows.

3.4.3 Proof of Theorem 3.19

Our proof of Theorem 3.19 involves the companion matrix result, Theorem 2.18, and the Sherman–Morrison formula, Lemma 2.20, that we used Chapter 2 above. We also require the following consequence of Lemma 2.39.

Lemma 3.32. *Under the assumptions of Theorem 3.19, 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 Lemma 2.39 and the assumption on m_μ given in (3.17). □

With Lemma 3.32 in hand, we are now prepared to present the proof of Theorem 3.19.

Proof of Theorem 3.19. Let D be the diagonal matrix $D := \operatorname{diag}(X_1, \ldots, X_n)$. Using the notation from Theorem 2.18, we observe that $zI - D$ is invertible for all $z \in \Gamma$ since $X_1, \ldots, X_n \in B(0, C)$ by supposition. In addition, by the Gauss–Lucas theorem and Theorem 2.18, 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

$$\sup_{z \in \Gamma} |\operatorname{tr} G(z) - \operatorname{tr} R(z)| = O(1) \tag{3.47}$$

with overwhelming probability.

Since $J = \mathbf{1}\mathbf{1}^{\mathrm{T}}$, where $\mathbf{1}$ is the all-ones vector, the Sherman–Morrison formula (Lemma 2.20) implies that

$$R(z) = G(z) - \frac{\frac{1}{n} G(z) D J G(z)}{1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}} G(z) D \mathbf{1}} \tag{3.48}$$

provided $1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}} G(z) D \mathbf{1} \neq 0$. In view of Lemma 3.32, there exists a constant $c' > 0$ (depending only on C, c, and ε) such that

$$\inf_{z \in \Gamma} \left| 1 + \frac{1}{n}\mathbf{1}^{\mathrm{T}} G(z) D \mathbf{1} \right| = \inf_{z \in \Gamma} \left| \frac{z}{n} \sum_{i=1}^{n} \frac{1}{z - X_i} \right| \geq c' \tag{3.49}$$

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 (3.48) and (3.49), 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 (3.47), and the proof is complete. □

3.5 Proof of Theorem 3.3

This section is devoted to proving Theorem 3.3. Our first lemma shows that Assumption 3.2 implies Assumption 3.1.

Lemma 3.33 (Sufficiency of Assumption 3.2)**.** *If μ satisfies Assumption 3.2, then μ also satisfies Assumption 3.1.*

Proof. Without loss of generality, suppose μ is radially symmetric about $z = 0$, and let $X \sim \mu$. By Lemma 3.5, we can write

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

so the hypotheses guarantee that $|m_\mu(z)|$ is continuous on $\mathbb{C} \setminus \{0\}$. (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 $|z| \leq \delta$ implies $|f(z)| \geq c > 0$. In particular, for $|z| \leq \delta$,

$$|m_\mu(z)| = \frac{1}{|z|}\int_0^{|z|} r f(r)\, dr \geq \frac{c}{|z|}\int_0^{|z|} r\, dr = \frac{c\,|z|}{2}. \tag{3.50}$$

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

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

We know that $m_{\min}$ is non-zero by (3.50) and the fact that $\mathbb{P}(|X| < r)$ is non-decreasing in r. This second fact additionally implies that for $|z| \geq r_{1/2}$,

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

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

$$\mathbb{P}(|m_\mu(X)| < \varepsilon) \leq \mathbb{P}\left(\frac{c\,|X|}{2} < \varepsilon\right) + \mathbb{P}(m_{\min} < \varepsilon) + \mathbb{P}\left(\frac{1}{2\,|X|} < \varepsilon\right) \leq C\varepsilon^2, \tag{3.51}$$

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 3.1 part (i).

To see that μ satisfies Assumption 3.1 part (ii), let $X_1, \ldots, X_n$ be iid complex-valued random variables with distribution μ, and observe that

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

By Markov's inequality,

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

which completes the argument. □

3.5.1 Introduction to and motivation for the proof of Theorem 3.3.

The following proof of Theorem 3.3 is motivated by the illustration in Figure 3.3 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 3.9, 3.11 and 3.23, 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 3.9, 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 3.3, 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 3.3. 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 3.1 and 3.2.

3.5.2 Definitions

In view of Lemma 3.33, we prove Theorem 3.3 under Assumption 3.1. Let $C_\mu > 0$ be larger than each of the constants in Assumption 3.1 and larger than the constant bounding the density associated to μ. For each $n \in \mathbb{N}$, define the following sets which partition $\mathbb{C}$ into regions based on the size of $|m_\mu(z)|$:

$$\begin{aligned}
A_n^k &:= \left\{ z \in \mathbb{C} : |m_\mu(z)| < \frac{e^k}{\sqrt{n}} \right\}, \; k = \lfloor 4\log(\log n) \rfloor, \\
A_n^k &:= \left\{ z \in \mathbb{C} : \frac{e^{k-1}}{\sqrt{n}} \leq |m_\mu(z)| < \frac{e^k}{\sqrt{n}} \right\}, \; \lfloor 4\log(\log n) \rfloor + 1 \leq k \leq \left\lfloor \log\left(\sqrt{n}\right) \right\rfloor, \\
A_n &:= \left\{ z \in \mathbb{C} : |m_\mu(z)| \geq \frac{e^{\left\lfloor \log\left(\sqrt{n}\right) \right\rfloor}}{\sqrt{n}} \right\}.
\end{aligned}$$

Additionally, define the random variables

$$\begin{aligned}
N_n^k &:= \# \left\{ 1 \leq j \leq n : X_j \in A_n^k \right\}, \; \lfloor 4\log(\log n) \rfloor \leq k \leq \left\lfloor \log\left(\sqrt{n}\right) \right\rfloor, \\
\zeta_{i,j}^{(n)} &:= \begin{cases} \dfrac{1}{X_i - X_j} \mathbb{1}_{|X_i - X_j| \geq \frac{(\log n)^2}{n|m_\mu(X_i)|}}, & |m_\mu(X_i)| \neq 0 \\ 0, & \text{otherwise} \end{cases}, \; 1 \leq i, j \leq n, \; j \neq i,
\end{aligned}$$

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 2.17. Let $\delta > 0$ be a fixed real parameter to be chosen later. We will show that the conclusion of Theorem 3.3 holds on the complement of the

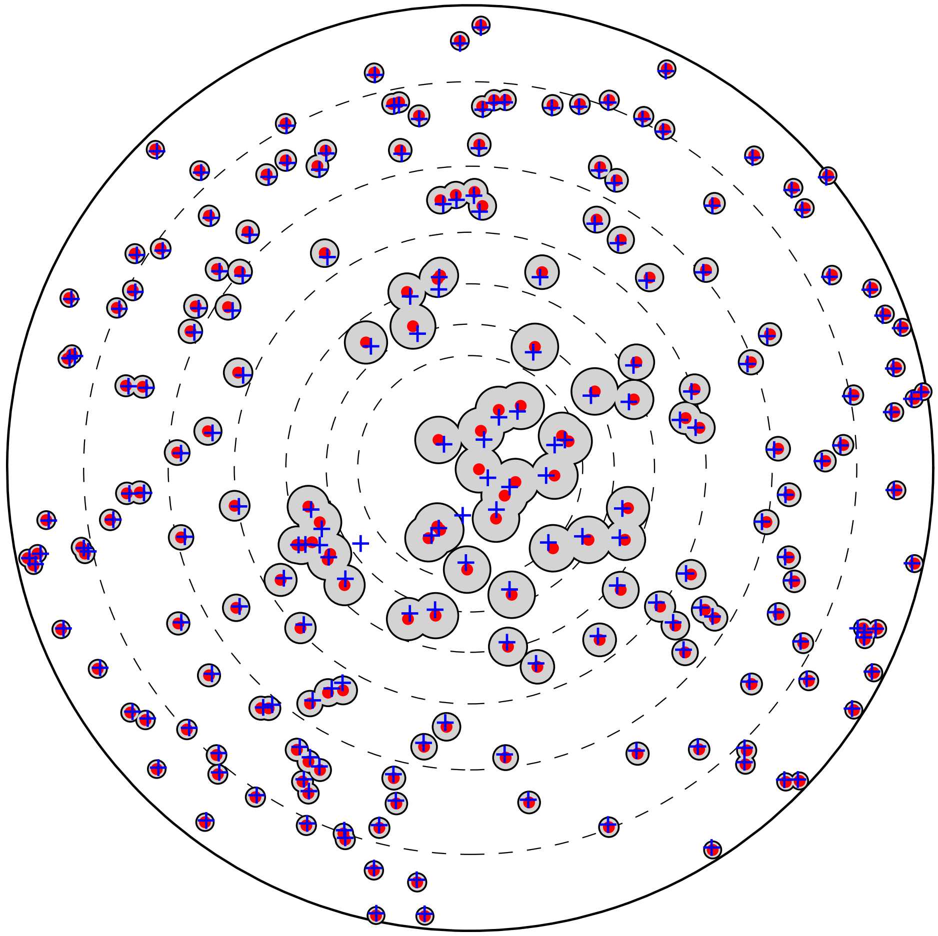


Figure 3.3: An illustration motivating the strategy we use to prove Theorem 3.3. The red circles and blue crosses represent the locations of the roots and critical points, respectively, of $p_{150}(z)$, where μ is the uniform distribution on the unit disk. Roughly speaking, the gray disks around the 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.

union of the following "bad" events:

$$
\begin{aligned}
E_n^k &:= \left\{ N_n^k \geq 2C_\mu e^{2k} \log(\log n) \right\}, \ \lfloor 4\log(\log n) \rfloor \leq k \leq \left\lfloor \log\left(\sqrt{n}\right) \right\rfloor; \\
F_n^i &:= \left\{ |m_\mu(X_i)| \geq \frac{(\log n)^4}{\sqrt{n}}, \ \left| \frac{1}{n-1} \sum_{\substack{j=1 \\ j\neq i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} | X_i] \right) \right| \geq \frac{|m_\mu(X_i)|}{2} \right\}, \\
&\qquad \text{for } 1 \leq i \leq n; \\
G_n^\delta &:= \left\{ \exists x \in \mathcal{N}_n \cup \{X_i\}_{i=1}^n \ \text{ s.t. } \# \left\{ 1 \leq j \leq n : |X_j - x| < \frac{1}{\sqrt{n}} \right\} \geq 2 + \delta \log n \right\}; \\
H_n &:= \left\{ \eta_n \geq n^{C_\mu} \right\}.
\end{aligned}
$$

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

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

3.5.3 The "bad" events are unlikely

In this subsection, we establish that

$$
\mathbb{P}\left(\mathcal{E}_n^{\text{bad}}\right) = o(1). \tag{3.52}
$$

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

Lemma 3.34.

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

Proof. Observe that for a fixed n and k, $\lfloor 4\log(\log n) \rfloor \leq k \leq \lfloor \log\left(\sqrt{n}\right) \rfloor$, 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,

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

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. $\square$

Lemma 3.35.

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

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 3.26, part (i) that $|m_\mu(X_i)|$ is almost surely bounded above by an absolute constant (that depends only on μ).

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

$$\begin{aligned}&\mathbb{E}\left[|Y - \mathbb{E}[Y \mid X]|^4 \,\middle|\, X\right]\\ &\qquad \leq \mathbb{E}\left[|Y|^4 \,\middle|\, X\right] + 6\left(\mathbb{E}\left[|Y|^2 \,\middle|\, X\right]\right)^2 + 4\mathbb{E}\left[|Y|^3 \,\middle|\, X\right]\cdot\mathbb{E}\left[|Y| \mid X\right].\end{aligned} \tag{3.53}$$

Indeed, we have

$$\begin{aligned}
&\mathbb{E}\left[|Y - \mathbb{E}[Y \mid X]|^4 \;\middle|\; X\right] \\
&\quad = \mathbb{E}\left[\left(Y - \mathbb{E}[Y \mid X]\right)^2 \left(\overline{Y} - \overline{\mathbb{E}[Y \mid X]}\right)^2 \;\middle|\; X\right] \\
&\quad = \mathbb{E}\left[|Y^4| \mid X\right] + 4\mathbb{E}\left[|Y|^2 \;\middle|\; X\right] \cdot |\mathbb{E}[Y \mid X]|^2 + 2\operatorname{Re}\left((\mathbb{E}[Y \mid X])^2 \cdot \mathbb{E}\left[\overline{Y}^2 \;\middle|\; X\right]\right) \\
&\qquad - 4\operatorname{Re}\left(\mathbb{E}\left[Y \mid X\right] \cdot \mathbb{E}\left[|Y|^2\,\overline{Y} \;\middle|\; X\right]\right) - 3\left|\mathbb{E}[Y \mid X]\right|^4 \\
&\quad \le \mathbb{E}\left[|Y|^4 \;\middle|\; X\right] + 4\left(\mathbb{E}\left[|Y|^2 \;\middle|\; X\right]\right)^2 + 2\left(\mathbb{E}\left[|Y|^2 \;\middle|\; X\right]\right)^2 + 4\mathbb{E}\left[|Y|^3 \;\middle|\; X\right] \cdot |\mathbb{E}Y|,
\end{aligned}$$

from which the desired result follows. Now, for $X = X_i$ and $Y = \zeta_{i,j}^{(n)}$, where $1 \le i, j \le n$ with $j \ne i$,

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

and similarly,

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

Consequently, via (3.53), there are positive constants C'_μ, K_μ that depend only on μ so that if $n \ge K_\mu$, on the event $|m_\mu(X_i)| \ge n^{-1/2}$,

$$\mathbb{E}\left[\left|\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} \;\middle|\; X_i\right]\right|^4 \;\middle|\; X_i\right] \le \frac{C'_\mu \left|m_\mu(X_i)\right|^2 n^2}{(\log n)^4}. \tag{3.54}$$

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

$$\mathbb{1}_{|m_\mu(X_i)| \ge \frac{1}{\sqrt{n}}} \cdot \mathbb{E}\left[\left| \sum_{\substack{j=1 \\ j \ne i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}\left[\zeta_{i,j}^{(n)} | X_i \right] \right) \right|^4 \, \middle| \, X_i \right] \le \frac{C''_\mu \left| m_\mu(X_i) \right|^2 n^3}{(\log n)^4}. \tag{3.55}$$

Write

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

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

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

where $l \ne i$ is any fixed index. From (3.54) and the bounds on $\mathbb{E}[|Y^2| \mid X_i]$ and $\mathbb{E}[|Y| \mid X_i]$ above, we can find $C''_\mu, K'_\mu > 0$ large enough so that $n \ge K'_\mu$ implies (3.55). (For the asymptotics, we are using that $n^{-1/2} \le |m_\mu(X_i)| = O_\mu(1)$, where the implied constant depends only on μ.) Via Markov's inequality, it follows that for $n \ge K'_\mu$ and a fixed i, $1 \le i \le n$, on the event $|m_\mu(X_i)| \ge n^{-1/2}$,

$$\mathbb{P}\left(\left| \frac{1}{n-1} \sum_{\substack{j=1 \\ j \ne i}}^{n} \left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} | X_i] \right) \right| \ge \frac{|m_\mu(X_i)|}{2} \, \middle| \, X_i \right) \le \frac{C''_\mu}{n \left| m_\mu(X_i) \right|^2 (\log n)^4}. \tag{3.56}$$

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

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

where we used (3.56) to bound $\mathbb{P}(F_n^1 \mid X_1)$. Assumption 3.1 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

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

Hence, for large n, our calculation from above yields

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

□

Lemma 3.36. *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. $\square$

Combining Lemmas 3.34, 3.35, and 3.36 from this subsection establishes (3.52), so for the remainder of the proof, we work on the complements of the "bad" events.

3.5.4 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 3.3 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 3.3 is one connected component, $\mathcal{U}_{n,P}$ for some $P \in \mathcal{P}_n$; a “zoomed-in” version is presented in Figure 3.5. 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 3.37. *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 [15, 60], and the discussion on pp. 13 and 67 of [62]). 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 3.38 (Theorem 1 in [60]; see also Theorem II 5.38 on p. 67 of [62])**.** *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.*

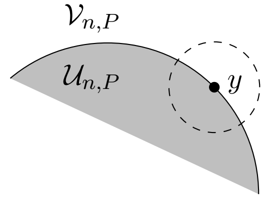
Our goal is to show that the compact set $\gamma_{n,P} = \partial\mathcal{V}_{n,P}$ has precisely two regions from which $\gamma_{n,P}$ is accessible at every point. Define $\mathcal{U}'_{n,P} := \mathbb{C} \setminus \overline{\mathcal{V}_{n,P}}$. Observe that $\mathbb{C} \setminus \gamma_{n,P} = \mathcal{V}_{n,P} \cup \mathcal{U}'_{n,P}$, 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 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 $\mathcal{U}_{n,P} \subset S$.

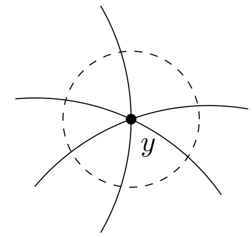
Since the (finitely many) circles $\partial B_1, \ldots, \partial B_l$ are distinct, there are only finitely many points of $\mathbb{C}$ that are contained in more than one circle. Consequently, we can choose a point $v \in \mathcal{V}_{n,P}$ such that the line segment $\overline{xv}$ does not contain 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 3.4A). 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



(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 3.4: The geometry near $y \in \gamma_{n,P}$.

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 3.4A and 3.4B, 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 3.4B). 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 3.38 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 $\mathcal{U}_{n,P} \subset \mathcal{U}'_{n,P}$. $\square$

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 3.39. 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, $1 \le j \le n$, 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 3.40. *Suppose $0 < \delta < 1/3$. There exists $C_\delta > 0$ so that for $n \ge C_\delta$, the following holds on the complement of G_n^δ: for each $P \in \mathcal{P}_n$, $|P| \le \delta \log n + 2$, 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

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

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

$B^i_k \cap B^i_{k+1} \neq \emptyset$ for $0 \leq k \leq l_i - 1$, and $B^i_0, \ldots, B^i_{l_i}$ 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^i_0 and $B^i_{l_i}$, 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^δ_n. On the other hand, if case (ii) is true, then, for large n,

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

Indeed, $\left\{B^{i^*}_k\right\}_{k=0}^{\lceil \delta \log n + 2 \rceil}$ 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 $y \in \bigcup_{k=0}^{\lceil \delta \log n + 2 \rceil} B^{i^*}_k$, 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^δ_n 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^δ_n, $|P| \leq \delta \log n + 2$ 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 3.41. *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 3.40, it suffices to show that there exist $x^*, y^* \in \Gamma$ so that

$$\sup_{x,y \in \overline{\mathcal{O}_\Gamma}} |x - y| \leq |x^* - y^*|. \tag{3.57}$$

(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 (3.57) 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 (3.57) is achieved for $x = x^*$, $y = y^*$. We conclude that $x^* \in \Gamma$. A similar argument shows that $y^* \in \Gamma$.

$\square$

3.5.5 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 3.3 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

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

The following lemma is the main result of this subsection.

Lemma 3.42. *For a fixed $\delta > 0$ chosen sufficiently small, there is a constant $C_\delta > 0$ so that for $n \geq C_\delta$, 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,*

$$|X - w| \leq \frac{(\log n)^4}{n\,|m_\mu(X)|}.$$

Proof. The proof of this lemma is similar in flavor to the proofs of Theorems 3.9 and 3.23, 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 3.5). To that end, define

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

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 $p_n'(n)$ 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 3.42.

Lemma 3.43. *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 3.26 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

$$\frac{|m_\mu(X)|}{2} \leq |m_\mu(\xi)| \leq \frac{3\,|m_\mu(X)|}{2}. \tag{3.58}$$

Indeed, via Corollary 3.41, $|\xi - X| < \frac{3\delta}{\sqrt{n}} < \frac{1}{\sqrt{n}}$ 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 3.26 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 (3.58). (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 (3.58) and the conclusion of Corollary 3.41 on the complement of G_n^δ. Since, $z, X \in \overline{\mathcal{O}_\Gamma}$, Corollary 3.41 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 $i_1, i_2 \in P^* \subset \mathcal{O}_\Gamma$, for which

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

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 $B_0, B_1, \ldots B_l \in \mathcal{C}_n$, of the form

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

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 (3.58) 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 3.40, 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

$$\begin{aligned}\left|\frac{1}{n-1}\sum_{\substack{j=1\\j\neq i}}^{n} \zeta_{i,j}^{(n)}\right| \leq &\left|\frac{1}{n-1}\sum_{\substack{j=1\\j\neq i}}^{n}\left(\zeta_{i,j}^{(n)} - \mathbb{E}[\zeta_{i,j}^{(n)} \mid X_i]\right)\right| \\ &+ \left|\mathbb{E}[\zeta_{i,l}^{(n)} \mid X_i] - m_\mu(X_i)\right| + |m_\mu(X_i)|\,,\end{aligned} \tag{3.59}$$

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 (3.59) 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, $1 \le i_X \le n$, and $X \in R_n^{\text{pair}}$, so on the complement of $\cup_{i=1}^n F_n^i$,

$$\begin{aligned}\left|\frac{1}{n-|R_{\text{near}}|}\sum_{j\notin R_{\text{near}}}\frac{1}{X-X_j}\right| &= \frac{n-1}{n-|R_{\text{near}}|}\left|\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i_X}}^{n}\zeta_{i_X,j}^{(n)}\right| \\ &\le \frac{n-1}{n-|R_{\text{near}}|}\left(\frac{3}{2}\left|m_\mu(X_{i_X})\right| + \frac{2\pi C_\mu(\log n)^2}{n\left|m_\mu(X_{i_X})\right|}\right) \\ &\le \frac{n-1}{n-|R_{\text{near}}|}\left(\frac{3}{2}\left|m_\mu(X)\right| + \frac{2\pi C_\mu}{\sqrt{n}(\log n)^2}\right).\end{aligned} \tag{3.60}$$

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 (3.60) 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, $1/4\left|m_\mu(X)\right|$ for large n.) The lower bound in (ii) is achieved similarly by using the reverse triangle inequality to obtain

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

in place of (3.59).

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

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

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 3.26, 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 3.5 which provides a visual aid to the argument.

Lemma 3.44. *Suppose $\delta < 1/3$. For large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, the polynomial*

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

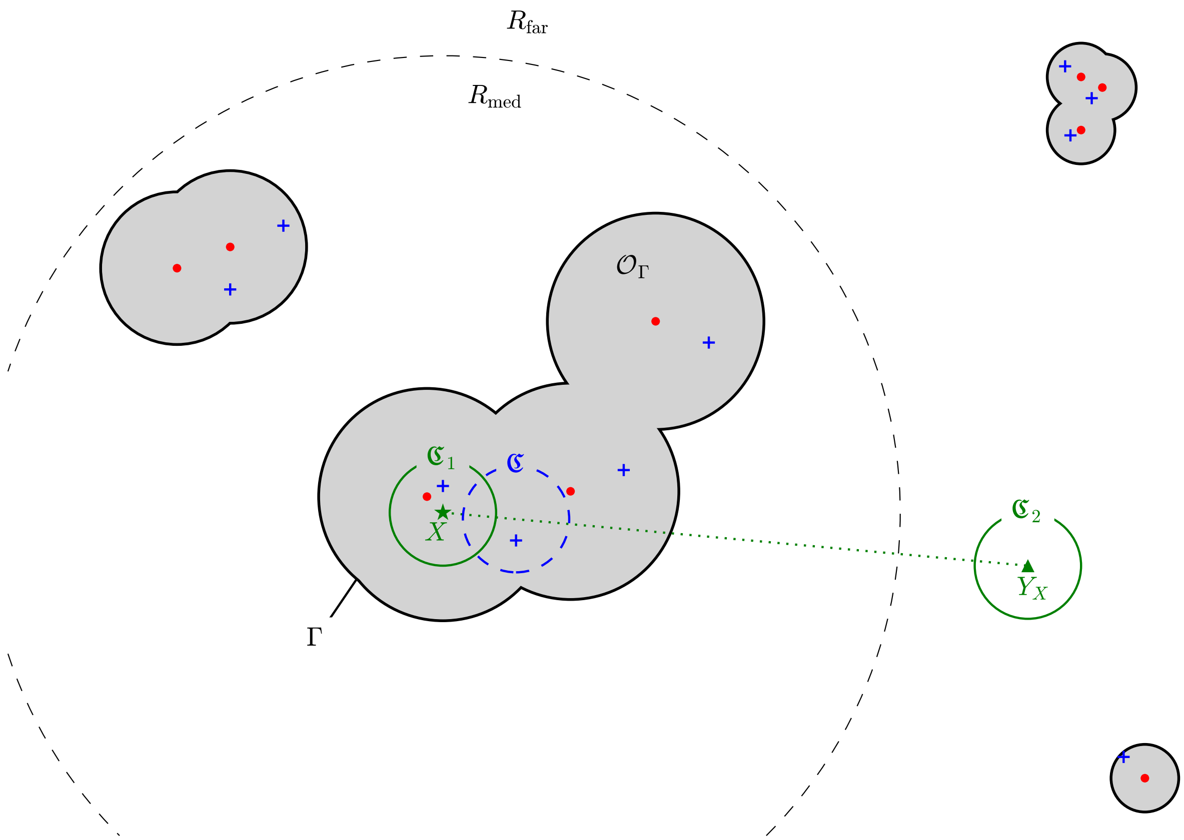


Figure 3.5: A diagram to illustrate Lemma 3.44 and its proof. The red dots and blue crosses are meant to represent roots and critical points, respectively, of p_n that lie in a region near X, which is denoted by a green star. The large dashed circle is intended to be on the order of $n^{-1/2}$. Note 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.

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 [46].) 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 3.43 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 3.43:

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

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 3.43 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 3.43, 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 [46], 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 3.44 that $p_n(z)$ and $p'_n(z)$ have the same numbers of zeros in $\mathcal{O}_\Gamma$.

Lemma 3.45. *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 $z \in \Gamma$,*

$$\left|p'_n(z) - f_X(z)\right| \leq |p_n(z)|\, \widetilde{C}_\mu \delta^2 n\, |m_\mu(X)| \tag{3.61}$$

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

Proof. For large n, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta$, Lemma 3.43 guarantees that if $z \in \Gamma$, then,

$$\begin{aligned}
&\left|p_n'(z) - f_X(z)\right| \\
&= |p_n(z)| \cdot \left| \sum_{j=1}^n \frac{1}{z - X_j} - \frac{r_X'(z)}{r_X(z)} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \\
&= |p_n(z)| \cdot \left| \sum_{j \notin R_{\text{near}}} \frac{1}{z - X_j} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \\
&\leq |p_n(z)| \cdot \left(\left| \sum_{j \notin R_{\text{near}}} \frac{1}{z - X_j} - \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right| + \left| \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} - \frac{n - |R_{\text{near}}|}{z - Y_X} \right| \right) \\
&= |p_n(z)| \cdot |z - X| \left| \sum_{j \notin R_{\text{near}}} \frac{1}{(z - X_j)(X - X_j)} \right| \\
&\quad + |p_n(z)| \left| \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right| \left| 1 - \frac{1}{(z - X)\frac{1}{n - |R_{\text{near}}|} \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} + 1} \right| \\
&\leq |p_n(z)| \, |z - X| \\
&\quad \times \left(k_n + \left| \frac{1}{n - |R_{\text{near}}|} \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right|^2 \frac{n - |R_{\text{near}}|}{1 - |z - X| \left| \frac{1}{n - |R_{\text{near}}|} \sum_{j \notin R_{\text{near}}} \frac{1}{X - X_j} \right|} \right) \\
&\leq |p_n(z)| \frac{4\delta(\log n)^4}{n \, |m_\mu(X)|} \left(k_n + 4 \, |m_\mu(X)|^2 \frac{n}{1 - \frac{4\delta(\log n)^4}{n|m_\mu(X)|} 2 \, |m_\mu(X)|} \right) \\
&\leq |p_n(z)| \left(\frac{4\delta(\log n)^4}{n \, |m_\mu(X)|} k_n + O_\delta \left((\log n)^4 \, |m_\mu(X)| \right) \right),
\end{aligned}$$

where

$$k_n := \sup_{z \in \Gamma} \left| \sum_{j \notin R_{\text{near}}} \frac{1}{(z - X_j)(X - X_j)} \right|.$$

What remains is to show that there exist positive constants $\widetilde{C}_\mu'$, $C_{\mu,\delta}'$ so that $n \geq C_{\mu,\delta}'$ implies

$$k_n = \widetilde{C}_\mu' \delta(\log(n))^{-4} n^2 \, |m_\mu(X)|^2 . \tag{3.62}$$

First observe that for any $z \in \Gamma$,

$$\begin{aligned}
&\sum_{j \notin R_{\text{near}}} \frac{1}{(z - X_j)(X - X_j)} \\
&\qquad = \sum_{j \in R_{\text{med}}} \frac{1}{(z - X_j)(X - X_j)} + \sum_{j \in R_{\text{far}}} \frac{1}{(z - X_j)(X - X_j)}.
\end{aligned} \tag{3.63}$$

We will bound each term on the right 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 $z \in \Gamma$ and $j \in R_{\text{med}}$,

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

By Lemma 3.26, 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 $\cup_{i=1}^n F_n^i \cup G_n^\delta$,

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

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 $\cup_{i=1}^n F_n^i \cup G_n^\delta$, if $z \in \Gamma$,

$$\left|\sum_{j\in R_{\text{med}}} \frac{1}{(z - X_j)(X - X_j)}\right| \leq |R_{\text{med}}|\,\frac{2n^2\,|m_\mu(X)|^2}{(\log n)^5} \leq (\delta \log n + 2)\frac{2n^2\,|m_\mu(X)|^2}{(\log n)^5}. \tag{3.64}$$

We now turn our attention to the second term on the right side of (3.63). We will split the sum into pieces based on how far away from X each X_j is. In particular, define for $1 \leq k \leq \sqrt{n} - 1$, the annuli

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

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.) Now, 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

$$\#_{k,n} \leq \left|\mathcal{N}_n \cap D'_{k,n}\right| \cdot (\delta \log n + 2). \tag{3.65}$$

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 (3.65) establishes that $\#_{k,n} \leq 16^2 k(\delta \log n + 2)$.

We can now bound the second term on the right of (3.63) as follows. For $\delta < 1/8$ and n large enough to guarantee the conclusions of Lemma 3.43, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta \cup H_n$, $j \in R_{\text{far}}$ implies

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

for $z \in \Gamma$ (note that $|m_\mu(X)| \geq \frac{(\log n)^4}{\sqrt{n}}$). Consequently, for n large and $z \in \Gamma$, on the complement of $\cup_{i=1}^n F_n^i \cup G_n^\delta \cup H_n$,

$$\begin{aligned}
&\sum_{j \in R_{\text{far}}} \frac{1}{|z - X_j|\,|X - X_j|} \\
&\quad \leq \sum_{k=1}^{\sqrt{n}-1} \sum_{j : X_j \in D_{k,n}} \frac{1}{|z - X_j|\,|X - X_j|} + \sum_{j : |X - X_j| \geq 1} \frac{1}{|z - X_j|} \\
&\quad \leq \sum_{k=1}^{\sqrt{n}-1} \sum_{j : X_j \in D_{k,n}} \frac{1}{(|X - X_j| - |z - X|)\,|X - X_j|} \\
&\qquad\quad + \sum_{j : |X - X_j| \geq 1} \frac{1}{|X - X_j| - |z - X|} \\
&\quad \leq \sum_{k=1}^{\sqrt{n}-1} \sum_{j : X_j \in D_{k,n}} \frac{2}{|X - X_j|^2} + \sum_{j : |X - X_j| \geq 1} \frac{2}{|X - X_j|} \\
&\quad \leq \sum_{k=1}^{\sqrt{n}-1} \#_{k,n} \frac{2n}{k^2} + 2n \\
&\quad \leq \sum_{k=1}^{\sqrt{n}-1} \frac{2\left(16^2 k(\delta \log n + 2)\right) n}{k^2} + 2n \\
&\quad = O\left(\delta n (\log n)^2 + 2n\right),
\end{aligned}$$

where the implied constant is independent of δ. The asymptotic comes from approximating $\sum_{k-1}^{\sqrt{n}-1} k^{-1}$ with $1+\int_1^{\sqrt{n}-1} x^{-1}\,dx$. Together with (3.64), this establishes equation (3.62) since $|m_\mu(X)| \geq \frac{(\log n)^4}{\sqrt{n}}$. $\square$

The last lemma in this subsection establishes a lower bound on $|f_X(z)|$ that will combine with (3.61) to fulfill the hypotheses of Rouché's theorem on the boundary Γ of the domain $\mathcal{O}_\Gamma$.

Lemma 3.46. *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 $z \in \Gamma$,*

$$|f_X(z)| \geq |p_n(z)|\, n\, |m_\mu(X)| \cdot e^{-9}. \tag{3.66}$$

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 3.44, 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 3.43, we have that on the complement of $\cup_{i=1}^{n} F_n^i \cup G_n^\delta$ and for $z \in \Gamma$,

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

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

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

We have used Lemma 3.43 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 (3.66) as was desired.

□

We have now established both (3.61) and (3.66), 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 3.45 and 3.46, 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

$$\left|p_n'(z) - f_X(z)\right| < |f_X(z)|\,.$$

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 3.44, 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 3.42 follows directly from this and Lemma 3.43 part (i) (note $\delta \le 1/4$).

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 3.42. □

3.5.6 Bounding the Wasserstein distance

In this subsection, we use Lemma 3.42 to prove Theorem 3.3. 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 (3.2) and (3.3)). 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 3.47. *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 3.47 and the triangle inequality to prove Theorem 3.3.

Lemma 3.48. *Let $X_1, \ldots, X_n$ be iid, complex random variables with distribution μ that has a bounded density and satisfies Assumption 3.1. Then, there is a constant C, depending only on μ, so that with probability $1 - o(1)$,*

$$W_1(\mu_n, \tilde{\mu}_n') \leq \frac{C\eta_n (\log n)^9}{n},$$

where μ_n, $\tilde{\mu}_n'$, and η_n are defined as above.

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

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

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 3.42.

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

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

For large n, on the complement of $\mathcal{E}_n^{\text{bad}}$, Lemma 3.42 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 numbers 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

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

(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 3.42 also implies that for large n, on the complement of $\mathcal{E}_n^{\text{bad}}$,

$$\left|X_j - w_{\sigma_n(j)}^{(n)}\right| \leq \frac{(\log n)^4}{n\,|m_\mu(X_j)|},$$

for each j, $1 \leq j \leq n$, such that $X_j \in R_n^{\text{pair}}$. (Indeed, $X_j \in R_n^{\text{pair}}$ implies that $X_j \in \mathcal{O}_\Gamma$ 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 $p_n(z)$. Consequently, for any $X_j \notin R_n^{\text{pair}}$, we have the trivial bound

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

It follows, for large n, on the complement of $\mathcal{E}_n^{\text{bad}}$, that

$$\begin{aligned}
&n \cdot W_1(\mu_n, \tilde{\mu}_n') \\
&\qquad \leq \sum_{j:X_j \notin R_n^{\text{pair}}} \left|X_j - w_{\sigma_n(j)}^{(n)}\right| + \sum_{j:X_j \in R_n^{\text{pair}}} \frac{(\log n)^4}{n\,|m_\mu(X_j)|} \\
&\qquad \leq \left(N_n^{\lfloor 4\log(\log n)\rfloor} + N_n^{\lfloor 4\log(\log n)\rfloor+1}\right) 2\eta_n \\
&\qquad\qquad + \sum_{k=\lfloor 4\log(\log n)\rfloor+2}^{\lfloor \log(\sqrt{n})\rfloor} \sum_{j:X_j\in A_n^k} \frac{(\log n)^4}{n\,|m_\mu(X_j)|} + n\cdot\frac{(\log n)^4\sqrt{n}}{ne^{\lfloor \log(\sqrt{n})\rfloor}} \\
&\qquad < \left(2C_\mu e^{2\lfloor 4\log(\log n)\rfloor}\log(\log n) + 2C_\mu e^{2\lfloor 4\log(\log n)\rfloor+2}\log(\log n)\right) 2\eta_n \\
&\qquad\qquad + \sum_{k=\lfloor 4\log(\log n)\rfloor+2}^{\lfloor \log(\sqrt{n})\rfloor} N_n^k \cdot \frac{(\log n)^4\sqrt{n}}{ne^{k-1}} + \frac{(\log n)^4\sqrt{n}}{e^{\lfloor \log(\sqrt{n})\rfloor}} \\
&\qquad \leq 2C_\mu \log(\log n)\left(e^{8\log(\log n)}(1+e^2)\cdot 2\eta_n + \sum_{k=\lfloor 4\log(\log n)\rfloor+2}^{\lfloor \log(\sqrt{n})\rfloor} e^{2k}\cdot\frac{(\log n)^4\sqrt{n}}{ne^{k-1}}\right) \\
&\qquad\qquad + \frac{(\log n)^4\sqrt{n}}{e^{\lfloor \log(\sqrt{n})\rfloor}} \\
&\qquad \leq 2C_\mu \log(\log n)\left(4e^2\eta_n(\log n)^8 + \frac{e(\log n)^4}{\sqrt{n}}\sum_{k=1}^{\lfloor \log(\sqrt{n})\rfloor} e^k\right) + \frac{e(\log n)^4\sqrt{n}}{\sqrt{n}} \\
&\qquad \ll_\mu (\log n)^9\eta_n + (\log n)^6.
\end{aligned}$$

To complete the proof of Lemma 3.48, recall that $\mathbb{P}(\mathcal{E}_n^{\text{bad}}) = o(1)$, and observe that with probability $1 - o(1)$, $\eta_n \log n \geq 1$. $\square$

We conclude this subsection by remarking that Theorem 3.3 follows from Lemmas 3.33, 3.47, and 3.48 and the triangle inequality for the L_1-Wasserstein metric.

Bibliography

[1] Charalambos D. Aliprantis and Kim C. Border. Infinite dimensional analysis. Springer, Berlin, third edition, 2006. A hitchhiker's guide.

[2] Abdul Aziz. On the zeros of a polynomial and its derivative. Bull. Austral. Math. Soc., 31(2):245–255, 1985.

[3] M. S. Bartlett. An inverse matrix adjustment arising in discriminant analysis. Ann. Math. Statistics, 22:107–111, 1951.

[4] Patrick Billingsley. Convergence of probability measures. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[5] Patrick 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.

[6] Hubert E. Bray. On the Zeros of a Polynomial and of Its Derivative. Amer. J. Math., 53(4):864–872, 1931.

[7] Sung-Soo Byun, Jaehun Lee, and Tulasi Ram Reddy. Zeros of random polynomials and its higher derivatives. Available at arXiv:1801.08974, 2018.

[8] Sourav Chatterjee. A generalization of the Lindeberg principle. Ann. Probab., 34(6):2061–2076, 2006.

[9] W. S. Cheung and T. W. Ng. Relationship between the zeros of two polynomials. Linear Algebra Appl., 432(1):107–115, 2010.

[10] Wai Shun Cheung and Tuen Wai 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.

[11] Branko Ćurgus and Vania Mascioni. A contraction of the Lucas polygon. Proc. Amer. Math. Soc., 132(10):2973–2981, 2004.

[12] 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.

[13] Dimitar K. Dimitrov. A refinement of the Gauss-Lucas theorem. Proc. Amer. Math. Soc., 126(7):2065–2070, 1998.

[14] Janusz Dronka. On the zeros of a polynomial and its derivative. Zeszyty Nauk. Politech. Rzeszowskiej Mat. Fiz., (9):33–36, 1989.

[15] István Fáry and Eric M. Isenberg. On a converse of the Jordan curve theorem. Amer. Math. Monthly, 81:636–639, 1974.

[16] William Feller. An introduction to probability theory and its applications. Vol. II. Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[17] Gerald B. Folland. Real analysis. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[18] Janos Galambos. Advanced probability theory, volume 10 of Probability: Pure and Applied. Marcel Dekker, Inc., New York, second edition, 1995.

[19] A. W. Goodman, Q. I. Rahman, and J. S. Ratti. On the zeros of a polynomial and its derivative. Proc. Amer. Math. Soc., 21:273–274, 1969.

[20] Allan Gut. Probability: a graduate course. Springer Texts in Statistics. Springer, New York, 2005.

[21] Boris Hanin. Correlations and pairing between zeros and critical points of Gaussian random polynomials. Int. Math. Res. Not. IMRN, 2015(2):381–421, 2015.

[22] Boris Hanin. Pairing of zeros and critical points for random meromorphic functions on Riemann surfaces. Math. Res. Lett., 22(1):111–140, 2015.

[23] Boris Hanin. Pairing of zeros and critical points for random polynomials. Ann. Inst. Henri Poincaré Probab. Stat., 53(3):1498–1511, 2017.

[24] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc., 58:13–30, 1963.

[25] Roger A. Horn and Charles R. Johnson. Matrix analysis. Cambridge University Press, Cambridge, second edition, 2013.

[26] J. Ben Hough, Manjunath Krishnapur, Yuval Peres, and Bálint Virág. Zeros of Gaussian analytic functions and determinantal point processes, volume 51 of University Lecture Series. American Mathematical Society, Providence, RI, 2009.

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

[28] Zakhar Kabluchko. Critical points of random polynomials with independent identically distributed roots. Proc. Amer. Math. Soc., 143(2):695–702, 2015.

[29] Zakhar Kabluchko and Hauke Seidel. Distances between zeroes and critical points for random polynomials with i.i.d. zeroes. Available at arXiv:1807.02140, 2018.

[30] Natalia L. Komarova and Igor Rivin. Harmonic mean, random polynomials and stochastic matrices. Adv. in Appl. Math., 31(2):501–526, 2003.

[31] J. W. Lindeberg. Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. Math. Z., 15(1):211–225, 1922.

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

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

[34] Morris Marden. The Geometry of the Zeros of a Polynomial in a Complex Variable. Mathematical Surveys, No. 3. American Mathematical Society, New York, N. Y., 1949.

[35] Morris Marden. Conjectures on the critical points of a polynomial. Amer. Math. Monthly, 90(4):267–276, 1983.

[36] A. I. Markushevich. Theory of functions of a complex variable. Vol. I, II, III. Chelsea Publishing Co., New York, english edition, 1977. Translated and edited by Richard A. Silverman.

[37] Sean O'Rourke. Critical points of random polynomials and characteristic polynomials of random matrices. Int. Math. Res. Not. IMRN, 2016(18):5616–5651, 2016.

[38] Sean O'Rourke and Behrouz Touri. On a conjecture of Godsil concerning controllable random graphs. SIAM J. Control Optim., 54(6):3347–3378, 2016.

[39] Sean O'Rourke and Noah Williams. On the local pairing behavior of critical points and roots of random polynomials. Available at arXiv:1810.06781, 2018.

[40] Sean O'Rourke and Noah Williams. Pairing between zeros and critical points of random polynomials with independent roots. Trans. Amer. Math. Soc., 371(4):2343–2381, 2019.

[41] Sean O'Rourke and Philip Matchett Wood. Spectra of nearly Hermitian random matrices. Ann. Inst. Henri Poincaré Probab. Stat., 53(3):1241–1279, 2017.

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

[43] Robin Pemantle and Igor Rivin. The distribution of zeros of the derivative of a random polynomial. In Advances in combinatorics, pages 259–273. Springer, Heidelberg, 2013.

[44] Valentin 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.

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

[46] 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.

[47] Tulasi Ram Reddy. On critical points of random polynomials and spectrum of certain products of random matrices. Available at arXiv:1602.05298, 2016.

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

[49] Blagovest Sendov. New conjectures in the Hausdorff geometry of polynomials. East J. Approx., 16(2):179–192, 2010.

[50] Stefan Steinerberger. A stability version of the gauss-lucas theorem and applications. Available at arXiv:1805.10454, 2018.

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

[52] Sneha Dey Subramanian. On the distribution of critical points of a polynomial. Electron. Commun. Probab., 17:no. 37, 9, 2012.

[53] Terence Tao. Topics in random matrix theory, volume 132 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2012.

[54] Terence Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. Probab. Theory Related Fields, 155(1-2):231–263, 2013.

[55] Terence Tao and Van Vu. Random matrices: universality of ESDs and the circular law. Ann. Probab., 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[56] Terence Tao and Van Vu. Random matrices: universality of local eigenvalue statistics. Acta Math., 206(1):127–204, 2011.

[57] Terence Tao and Van Vu. Local universality of zeroes of random polynomials. Int. Math. Res. Not. IMRN, 2015(13):5053–5139, 2015.

[58] Terence Tao and Van Vu. Random matrices: universality of local spectral statistics of non-Hermitian matrices. Ann. Probab., 43(2):782–874, 2015.

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

[60] Carsten Thomassen. The converse of the Jordan curve theorem and a characterization of planar maps. Geom. Dedicata, 32(1):53–57, 1989.

[61] Cédric Villani. Optimal transport, volume 338 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 2009. Old and new.

[62] Raymond Louis Wilder. Topology of manifolds. American Mathematical Society Colloquium Publications, Vol. XXXII. American Mathematical Society, Providence, R.I., 1963.

Appendix A

A heavy-tailed CLT

In this subsection, we prove Theorem A.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 [16], Theorem 11 in Section 6.4 of [18], and Theorem 3.10 in [44]), however, our implementation of Theorem A.1 requires specific information about the parameters of the limiting normal distribution; we include an explicit statement and proof for clarity.

Theorem A.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

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

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

Proof. We proceed by Lindeberg's exchange method [31]. (See also [8]. Similar methods have been applied to problems in random matrix theory; see e.g. [56], [57].) 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 (A.1), and let $g : \mathbb{C} \to \mathbb{R}$ be a smooth test function with compact support. We will show that

$$\left| \mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}} \sum_{j=1}^{n} \sum_{k=1}^{s} t_k \left[\frac{1}{\xi_k - X_j} - m_\mu(\xi_k) \right] \right) \right] - \mathbb{E}\left[g\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} N_j \right) \right] \right| \to 0, \tag{A.2}$$

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 [53], 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 A.2. *There is a constant $C_{\mu,s,\vec{t}} > 0$, depending only on μ, s, and $t_1, \ldots, t_k$, and there is a natural number $K_{\mu,g,\varepsilon}$ so that $n \geq K_{\mu,g,\varepsilon}$ implies*

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

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)

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

Similarly, we have

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

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

$$\begin{aligned}
&\left| \mathbb{E}\left[g \left(\frac{1}{\sqrt{n \log n}} \sum_{j=1}^{n} \widetilde{\zeta}_j \right) \right] - \mathbb{E}\left[g \left(\frac{N_1}{\sqrt{n}} + \frac{1}{\sqrt{n \log n}} \sum_{j=2}^{n} \widetilde{\zeta}_j \right) \right] \right| \\
&\quad \leq \frac{C_g}{2n} \left| \frac{1}{\log n} \mathbb{E}\left[\operatorname{Re}(\widetilde{\zeta}_1)^2 \right] - \mathbb{E}\left[\operatorname{Re}(N_1)^2 \right] \right| \\
&\quad + \frac{C_g}{2n} \left| \frac{1}{\log n} \mathbb{E}\left[\operatorname{Im}(\widetilde{\zeta}_1)^2 \right] - \mathbb{E}\left[\operatorname{Im}(N_1)^2 \right] \right| \\
&\quad + \frac{C_g}{n} \left| \frac{1}{\log n} \mathbb{E}\left[\operatorname{Re}(\widetilde{\zeta}_1) \operatorname{Im}(\widetilde{\zeta}_1) \right] - \mathbb{E}\left[\operatorname{Re}(N_1) \operatorname{Im}(N_1) \right] \right| \\
&\quad + \frac{4C_g}{(n \log n)^{3/2}} \mathbb{E}\left[\left| \widetilde{\zeta}_1 \right|^3 \right] + \frac{4C_g}{n^{3/2}} \mathbb{E}\left[\left| N_1 \right|^3 \right].
\end{aligned}$$

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

$$\begin{aligned}
&\Bigg|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} N_j + \frac{1}{\sqrt{n\log n}}\sum_{j=k}^{n} \widetilde{\zeta}_j\right)\right] \\
&\qquad\qquad - \mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{k} N_j + \frac{1}{\sqrt{n\log n}}\sum_{j=k+1}^{n} \widetilde{\zeta}_j\right)\right]\Bigg| \\
&\leq \frac{C_g}{2n}\left|\frac{1}{\log n}\mathbb{E}\left[\operatorname{Re}(\widetilde{\zeta}_k)^2\right] - \mathbb{E}\left[\operatorname{Re}(N_k)^2\right]\right| \\
&\qquad + \frac{C_g}{2n}\left|\frac{1}{\log n}\mathbb{E}\left[\operatorname{Im}(\widetilde{\zeta}_k)^2\right] - \mathbb{E}\left[\operatorname{Im}(N_k)^2\right]\right| \\
&\qquad + \frac{C_g}{n}\left|\frac{1}{\log n}\mathbb{E}\left[\operatorname{Re}(\widetilde{\zeta}_k)\operatorname{Im}(\widetilde{\zeta}_k)\right] - \mathbb{E}\left[\operatorname{Re}(N_k)\operatorname{Im}(N_k)\right]\right| \\
&\qquad + \frac{4C_g}{(n\log n)^{3/2}}\mathbb{E}\left[\left|\widetilde{\zeta}_k\right|^3\right] + \frac{4C_g}{n^{3/2}}\mathbb{E}\left[|N_k|^3\right].
\end{aligned}$$

Now, repeatedly applying the triangle inequality and using the fact that the $\widetilde{\zeta}_j$ and N_j are iid gives

$$\begin{aligned}
&\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n} \widetilde{\zeta}_j\right)\right] - \mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} N_j\right)\right]\right| \\
&\qquad \leq \frac{C_g}{2}\left|\frac{1}{\log n}\mathbb{E}\left[\operatorname{Re}(\widetilde{\zeta}_1)^2\right] - \mathbb{E}\left[\operatorname{Re}(N)^2\right]\right| \\
&\qquad\quad + \frac{C_g}{2}\left|\frac{1}{\log n}\mathbb{E}\left[\operatorname{Im}(\widetilde{\zeta}_1)^2\right] - \mathbb{E}\left[\operatorname{Im}(N)^2\right]\right| \\
&\qquad\quad + C_g\left|\frac{1}{\log n}\mathbb{E}\left[\operatorname{Re}(\widetilde{\zeta}_1)\operatorname{Im}(\widetilde{\zeta}_1)\right] - \mathbb{E}\left[\operatorname{Re}(N)\operatorname{Im}(N)\right]\right| \\
&\qquad\quad + \frac{4C_g}{\log n\sqrt{n\log n}}\mathbb{E}\left[\left|\widetilde{\zeta}_1\right|^3\right] + \frac{4C_g}{\sqrt{n}}\mathbb{E}\left[|N|^3\right].
\end{aligned} \tag{A.3}$$

In order to establish Lemma A.2, we need to show that each of the terms on the right side of (A.3) is dominated by ε as $n \to \infty$. It is in these computations that we use the fact that f is continuous at

$\xi_1, \ldots, \xi_s$. 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 3.26), there is a constant $C_{s,\vec{t},\delta}$ depending on s, $t_1, \ldots, t_s$ 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

$$\frac{1}{\log n}\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta}_1)\right] \le \sum_{k=1}^{s} |t_k|^2 \frac{\pi(f(\xi) + \eta)}{2} + o(1). \tag{A.4}$$

On the other hand, similar to above,

$$
\begin{aligned}
&\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta}_1)\right] \\
&\quad = \mathbb{E}\left[\mathrm{Re}^2(\zeta_1)\right] - \left(\mathbb{E}\left[\mathrm{Re}(\zeta_1)\right]\right)^2 \\
&\quad \geq \sum_{k=1}^{s} \mathbb{E}\left[\mathrm{Re}^2\left(\sum_{l=1}^{s} \frac{t_l}{\xi_l - X_1}\right) \mathbb{1}_{1/(\varepsilon\sqrt{n\log n} < |\xi_k - X_1| < \delta} \prod_{l \neq k} \mathbb{1}_{|\xi_l - X_1| \geq \delta}\right] - o(\log n) \\
&\quad \geq \sum_{k=1}^{s} \mathbb{E}\left[\frac{\mathrm{Re}^2((\xi_k - X_1)/t_k)}{|(\xi_k - X_1)/t_k|^4} \mathbb{1}_{1/(\varepsilon\sqrt{n\log n}) < |\xi - X_1| < \delta}\right] - o(\log n) \\
&\quad \geq \sum_{k=1}^{s} (f(\xi_k) - \eta)\,|t_k|^2 \int_0^{2\pi} \int_{1/(|t_k|\varepsilon\sqrt{n\log n})}^{\delta/|t_k|} \frac{r^2\cos^2\theta}{r^4} r\,dr\,d\theta - o(\log n) \\
&\quad \geq \sum_{k=1}^{s} \pi(f(\xi_k) - \eta)\,|t_k|^2 \log(\delta\varepsilon\sqrt{n\log n}) - o(\log n),
\end{aligned}
$$

and dividing by $\log n$ yields

$$\frac{\mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta}_1)\right]}{\log n} \geq \sum_{k=1}^{s} \frac{\pi(f(\xi_k) - \eta)\,|t_k|^2}{2} - o(1). \tag{A.5}$$

If we combine inequalities (A.4) and (A.5) and first take $\limsup_{n\to\infty}$ (respectively $\liminf_{n\to\infty}$) of both sides and then take $\eta \to 0$, we see that

$$\lim_{n\to\infty} \frac{1}{\log n} \mathbb{E}\left[\mathrm{Re}^2(\widetilde{\zeta}_1)\right] = \sum_{k=1}^{s} \frac{\pi f(\xi_k)\,|t_k|^2}{2} = \mathbb{E}[\mathrm{Re}^2(N)]. \tag{A.6}$$

(Note here that f is bounded, and by Lemma 3.26, the expectation of $|\xi - X_1|^{-1}$ is uniformly bounded, so the limit in n is uniform in ξ.) Nearly identical arguments to the one just made show that

$$\lim_{n\to\infty} \frac{1}{\log n} \mathbb{E}\left[\mathrm{Im}^2(\widetilde{\zeta}_1)\right] = \sum_{k=1}^{s} \frac{\pi f(\xi_k)\,|t_k|^2}{2} = \mathbb{E}[\mathrm{Im}^2(N)] \tag{A.7}$$

and

$$\lim_{n\to\infty} \frac{1}{\log n} \mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta}_1)\,\mathrm{Im}(\widetilde{\zeta}_1)\right] = 0 = \mathbb{E}[\mathrm{Re}(N)\,\mathrm{Im}(N)], \tag{A.8}$$

with the only modification being that to achieve upper and lower bounds for $\mathbb{E}\left[\mathrm{Re}(\widetilde{\zeta}_1)\,\mathrm{Im}(\widetilde{\zeta}_1)\right]$, one needs to consider separately the cases where the integrand is positive and negative.

In our quest to prove Lemma A.2, we next show that

$$\limsup_{n\to\infty} \frac{1}{\log n\sqrt{n\log n}} \mathbb{E}\left[\left|\widetilde{\zeta}_1\right|^3\right] \leq O_{\mu,s,\vec{t}}(\varepsilon). \tag{A.9}$$

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 $\frac{1}{\varepsilon\sqrt{n\log n}} \leq \delta_1$,

$$\begin{aligned}
\frac{\mathbb{E}\left[\left|\widetilde{\zeta}_1\right|^3\right]}{\log n\sqrt{n\log n}} &\leq \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\mathbb{E}\left[|\zeta_1|^2\right] \\
&\leq \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\left(\mathbb{E}\left[\left(\sum_{k=1}^{s}\frac{t_k}{\xi_k - X_1}\right)^2\prod_{k=1}^{s}\mathbb{1}_{|\xi_k - X_1|\geq\delta_1}\right]\right. \\
&\quad\left. + \sum_{k=1}^{s}\mathbb{E}\left[\left(\sum_{l=1}^{s}\frac{t_l}{\xi_l - X_1}\right)^2\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_k - X_1|<\delta}\prod_{l\neq k}\mathbb{1}_{|\xi_l - X_1|\geq\delta_1}\right]\right) \\
&\leq \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\left(\sum_{l=1}^{s}\frac{|t_k|}{\delta_1}\right)^2 \\
&\quad + \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\sum_{l=1}^{s}\mathbb{E}\left[\left|\frac{t_l}{\xi_l - X_1}\right|^2\mathbb{1}_{1/(\varepsilon\sqrt{n\log n}<|\xi_l - X_1|<\delta_1}\right] \\
&\quad + \sum_{k=1}^{s}\frac{8\varepsilon\,|t_k|}{\log n}\left(\sum_{l=1}^{s}2\mathbb{E}\left|\frac{t_l}{\xi_l - X_1}\right|\sum_{j\neq l}\frac{|t_j|}{\delta_1} + \sum_{l=1}^{s}\left(\sum_{j\neq l}\frac{|t_j|}{\delta_1}\right)^2\right),
\end{aligned}$$

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 (A.9). We conclude the proof of Lemma A.2 by combining equations (A.3), (A.6), (A.7), (A.8), and (A.9) in view of the facts that $|N|$ has a finite third moment and $f(z)$ and ε are bounded. $\square$

In order to establish (A.2), we still need to remove the truncation, which we will accomplish through a series of interpolations. We have

$$\begin{aligned}\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\sum_{k=1}^{s}t_k\left(\frac{1}{\xi_k-X_j}-m_\mu(\xi_k)\right)\\ &= \frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\left(\sum_{k=1}^{s}\frac{t_k}{\xi_k-X_j}-\zeta_j\right)+\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\\ &\quad+\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\left(\mathbb{E}[\zeta_j]-\sum_{k=1}^{s}t_k m_\mu(\xi_k)\right).\end{aligned}$$

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

$$\begin{aligned}\frac{n}{\sqrt{n\log n}}\sum_{k=1}^{s}|t_k|\,\mathbb{E}\left[\frac{1}{|\xi_k-X_1|}\mathbb{1}_{|\xi_k-X_1|^{-1}\geq\varepsilon\sqrt{n\log n}}\right]\\ &\leq \frac{ns\max_{1\leq k\leq s}|t_k|\cdot C_f}{\sqrt{n\log n}}\int_0^{2\pi}\int_0^{1/(\varepsilon\sqrt{n\log n})}\frac{1}{r}r\,dr\,d\theta\\ &\leq \frac{ns\max_{1\leq k\leq s}|t_k|\cdot C_f}{\sqrt{n\log n}}\frac{2\pi}{\varepsilon\sqrt{n\log n}}\\ &= o(1).\end{aligned}$$

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 A.2 now implies that for n larger than a constant depending on μ, g, ε, s, and $t_1,\dots,t_s$,

$$\begin{aligned}
&\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\sum_{k=1}^{s}t_k\left[\frac{1}{\xi-X_j}-m_\mu(\xi)\right]\right)\right]-\mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}N_j\right)\right]\right|\\
&\quad\leq\left|\mathbb{E}\left[g\left(\frac{1}{\sqrt{n\log n}}\sum_{j=1}^{n}\widetilde{\zeta}_j\right)\right]-\mathbb{E}\left[g\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}N_j\right)\right]\right|+o(1)\\
&\quad=O_{\mu,s,\vec{t},g}(\varepsilon),
\end{aligned}$$

so taking $\varepsilon\to 0$ yields equation (A.2). The conclusion of Theorem A.1 follows since our choice of g was arbitrary. $\square$