Rate of convergence for non-Hermitian random matrices and their products

Dissertation

zur Erlangung des Doktorgrades (Dr.math.)

vorgelegt von

Jonas Jalowy

16. Februar 2020

Fakultät für Mathematik
Universität Bielefeld

Dissertation

Rate of convergence for non-Hermitian random matrices and their products

Betreuer: Prof. Dr. Friedrich Götze
Fakultät für Mathematik
Universität Bielefeld

Nicht immer war ich schon so alt -
das machten erst die Jahre.
Die Stirne wuchs mit dem Verstand
im Laufe meiner Haare.
Nun wünsch ich mir, dass, was ich schrieb,
auch frohe Leser findet,
dann möge dieser Band das Band
sein, welches uns verbindet.

– Heinz Erhardt

Acknowledgements

First of all, I would like to express my sincere gratitude to my advisor Friedrich Götze for shaping my whole academic path that eventually led me to study this fascinating, yet challenging, topic. Without his stimulating ideas in insightful discussions and his steady support in several ways, this thesis would not be the same.

I would also like to thank Gernot Akemann, Mario Kieburg, Holger Kösters, Alexey Naumov, Thorsten Neuschel and especially Alexander Tikhomirov for their interest in my work and for sharing their expertise in plenty interesting discussions from which I learned a lot. Furthermore, I am grateful to Ji Oon Lee for his hospitality and collaboration during my stay in Korea.

Many thanks go to Anita Cole for her kind aid in organisatorial issues, as well as to the administration of the IRTG 2235, Nadine Brehme and Claudia Köhler. Financial support by the German Research Foundation (DFG) through the IRTG 2235 is gratefully acknowledged. I am absolutely thankful for the community of IRTG students that we built together - It was a wonderful time in Korea with all of you. And to all my Korean friends; 감사합니다.

Spending my daily life at the offices at Bielefeld University with Anna Gusakova, Holger Sambale, Arthur Sinulis and Sophiane Yahiatene was a pleasure - Thank you for all the enlightening mathematical chats as well as non-mathematical distractions. In particular, I would like to thank Arthur Sinulis and Anna Gusakova for proofreading this thesis and providing valuable feedback.

Nothing would have been possible without all my friends who have made me what I am today and who still accompany me. For the longest time I know my hometown guys Alex, Carsten, Fabian, Markus and Mathias - Thank you that I can always count on you. In Bielefeld, I thank my “Dokogruppe” for all the fun we had and all the amazing memories that we share from inside as well as outside of the University. A big chunk of thanks goes to Thoren, Basti and Arthur for a helping hand whenever needed and the fantastic activities which left no room for boredom. And to Anna for her way of cooling me down with her heartwarming character; Спасибо.

Finally, I thank my whole family for their cohesion in stressful times and encouraging belief in me. Thanks to Finn and Laurenz for our “Bruderschaft” and for never letting math look uncool. Most importantly, I am grateful to my parents, Markus and Petra, for their never-ending support - I owe you everything and I hope to be able to give you back what you gave me.

Danke euch allen.

ABSTRACT: The aim of this thesis is to investigate the rate of convergence of empirical spectral distributions of non-Hermitian random matrices with independent entries and their products. The distance to the deterministic limiting distribution will be measured in terms of a uniform Kolmogorov-like metric.

We will show that the optimal rate of convergence to the Circular Law is determined by Ginibre matrices and is given by $1/\sqrt{n}$. For products of Ginibre matrices, the optimal rate of convergence to powers of the Circular Law is shown to be $1/\sqrt{n}$ as well. Interestingly, the rate of convergence of the mean empirical spectral distribution is even faster in the bulk of the spectrum.

Furthermore, we develop an approach to study the rate of convergence for matrices with independent entries, which are not necessarily Gaussian. A smoothing inequality for complex measures that quantitatively relates the uniform Kolmogorov-like distance to the concentration of logarithmic potentials is shown. Combining it with results from Local Circular Laws, we apply it to prove nearly optimal rate of convergence to the Circular Law. Moreover, we show that also products of matrices with independent entries attain the optimal rate in the bulk up to a logarithmic factor.

The robustness of this approach enables us to similarly obtain the same rate of convergence in terms of the classical two-dimensional Kolmogorov distance as well as for the empirical measure of the roots of random Weyl polynomials. Finally, we shall relate our result to the spectral radius of non-Hermitian random matrices and investigate its rate of convergence.

Contents

Chapter 1

Introduction

Motivated by the statistics of energy levels of heavy atomic nuclei, Eugene Wigner discovered the *Semicircle Law* in his seminal work [Wig55] in 1955. He found that the empirical eigenvalue distribution of Hermitian random matrices tends to a semicircle distribution on the real line as the size n of the matrices tends to infinity. First, Wigner only considered symmetric random sign matrices, but shortly after, in [Wig58], he "point[s] out that the distribution law obtained before for a very special set of matrices is valid for more general sets" – A phenomenon that is nowadays called *Universality*: Many limiting eigenvalue statistics do not depend on the distribution of the entries, but only on the symmetry class of the random matrix.

This Universality Phenomenon makes Random Matrix Theory not only practically more applicable but also serves as the following guiding principle for mathematical research. If a statement holds true for a certain, say Gaussian, distribution of the entries, then it is expected to be still valid in wider generality. In particular, Gaussian matrices give access to explicit formulas and simpler methods, hence they play a central role in the theory as well as in this thesis.

Random Matrix Theory (RMT) was initially driven by physical (and statistical) applications, but soon more and more methods have been developed, links have been revealed and new problems arose. As the attention of many mathematicians increased over the years, it became a widespread and very active area of probability theory. Apart from their applications, random matrices are particularly interesting for various connections to other branches of mathematics, like number theory, non-commutative algebra, combinatorics and stochastics as well as classical, discrete and harmonic analysis. We postpone the discussion of some of these connections and applications to Chapter 2 and some will be encountered in relevant passages throughout this work. Instead, we will keep this introduction focused on our subject:

> "The theory of *non-Hermitian random matrices*, though not applicable to any physical problems, is a fascinating subject and must be studied for its own sake." – Madan Lal Mehta [Meh67]

1.1. Motivation: The Circular Law

We start with a natural mathematical question that was posed already by the physicist Jean Ginibre in 1965 without having any applications in mind.

> What can we say about the complex empirical eigenvalue distribution if we drop the symmetry constraint on the random matrices?

Obviously, the eigenvalues may lie anywhere in the complex plane, which introduces several new difficulties in comparison to Hermitian ensembles, since most of its techniques that will be presented in Section 2.1 are not directly applicable here. Yet, as Ginibre already answered himself, the eigenvalues tend to be uniformly distributed in the unit disk, see Figure 1.1. In order to formulate this *Circular Law*, let us define the following two classes of non-Hermitian random matrices that we shall be interested in throughout this work.

Definition 1.1. (i) A (complex) *Ginibre matrix* X is a non-Hermitian random $n \times n$-matrix with independent standard complex Gaussian entries X_{ij}, i.e. $\mathrm{Re} X_{ij}$ and $\mathrm{Im} X_{ij}$ are independent Gaussian random variables with mean 0 and variance $\frac{1}{2}$.

(ii) A non-Hermitian random $n \times n$-matrix X is said to have *independent entries* if X_{ij} are independent complex or real random variables, and in the complex case we additionally assume $\mathrm{Re} X_{ij}$ and $\mathrm{Im} X_{ij}$ to be independent.

Furthermore we define the central object of study, *the empirical spectral distribution* (ESD), by

$$\mu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j (X/\sqrt{n})}, \tag{1.1}$$

where δ_{λ_j} are Dirac measures in the eigenvalues λ_j of the scaled matrix $X/\sqrt{n}$. Moreover let $\bar{\mu}_n = \mathbb{E}\, \mu_n$ be the *mean empirical spectral distribution*. Denote the Lebesgue measure on $\mathbb{C}$ by $\lambda\!\!\lambda$, weak convergence of measures by $\Rightarrow$ and the underlying probability space by $(\Omega, \mathcal{A}, \mathbb{P})$.

Theorem 1.2 (The Circular Law, [Gin65; Bai97; GT07; GT10c; PZ10; TV10b])**.** *Let X be a matrix with independent entries. If $\mathbb{E}\, X_{ij} = 0$ and $\mathbb{E}\, |X_{ij}|^2 = 1$, then $\mathbb{P}$-a.s. we have*

$$\mu_n \Rightarrow \mu_\infty, \text{ where } d\mu_\infty = \frac{1}{\pi} \mathbb{1}_{B_1(0)} d\lambda\!\!\lambda \tag{1.2}$$

is the uniform distribution on the complex unit disc.

Conjectured since the early 1950's, the proof of the Circular Law has a long and interesting history with contributions of many different researchers that we shall recap in Section 2.2.

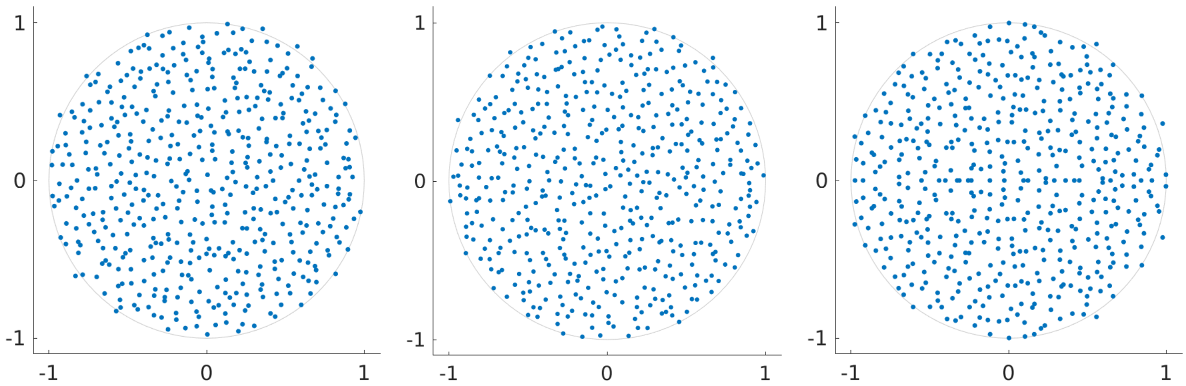

Figure 1.1: Samples of the spectrum of different non-Hermitian random matrices of size $n = 500$. The independent entries are chosen to be complex standard Gaussian (left), uniform on the discrete corners of a square (center) and real standard Gaussian (right). All three entry distributions lead to the Circular Law, the only noticeable difference being the natural symmetry around the real axis in the case of real entries (right).

This Circular Law is an instance of what is called a *Global Law*. The term “global” means that the entire spectrum contributes and individual eigenvalues do not play a role. It is complemented by a *Local Law* stating that the eigenvalue distribution is well approximated by a limiting distribution down to microscopic scale containing only a small portion (or even a finite amount) of the eigenvalues. Recently, there has been made significant progress for Local Laws [BYY14a; BYY14b; GNT19a; TV15; AEK19; Yin14] among others, and for Universality of the correlation functions, see [TV15; CES19b]. Additionally, modified models have been studied, for instance non-homogeneous variances [AEK18; AEK19], an Elliptic Law as interpolation between the Circular Law and the Semicircle Law [Nau12], sparse matrices [RT19] and products of independent matrices [GT10b; OS11]. The latter is most important for this thesis and we will present the result in detail in the sequel. However, for any given distance between distributions the rate of convergence to the Circular Law has barely been treated in the literature so far.

1.2. Rate of convergence

In this thesis, we address Universality of the *rate of convergence*, containing local as well as global Universality in a uniform and quantitative manner. We will provide an explicit optimal rate of convergence for Ginibre Matrices, nearly optimal rate of convergence for certain matrices with independent entries and generalize the results to products of independent matrices. The study is completed by a rate of convergence for the spectral radius and applications to random polynomials.

Rates of convergence in Hermitian RMT has been studied intensively, see [GT03a; CB04; GT05; BS10; BHPZ11; GT16; GNTT18; CFLW19]. Here, the distance between the empirical spectral distribution and the Semicircle Law is measured in terms of the usual Kolmogorov distance.

We are interested in the rate of convergence to the Circular Law, more precisely in the *Kolmogorov distances* over balls

$$D(\mu_n,\mu_\infty) := \sup_{z_0\in\mathbb{C},R>0} |\mu_n(B_R(z_0)) - \mu_\infty(B_R(z_0))| \tag{1.3}$$

as $n \to \infty$. The study of Kolmogorov-like metrics of complex measures is widely uncommon in the literature of non-Hermitian Random Matrix Theory so far. Therefore, let us provide some additional information about advantages of studying D.

Most importantly, convergence in this distance coincides with weak convergence in the case of an absolutely continuous limiting distribution as we will prove in Lemma 2.25. Hence D is a reasonable object to study the rate of convergence to the Circular Law. The choice to uniformly consider all balls reflects the essential structure of the rotational symmetry of μ_∞ and of the mean empirical spectral distribution $\bar{\mu}_n$ of the Ginibre ensemble. Moreover, in the latter case some explicit computations of $D(\mu_n,\mu_\infty)$ are possible such as the following.

Lemma 3.1. *The mean ESD $\bar{\mu}_n = \mathbb{E}\,\mu_n$ of the Ginibre ensemble satisfies*

$$D(\bar{\mu}_n,\mu_\infty) \sim \frac{1}{\sqrt{2\pi n}} \tag{1.4}$$

and

$$\sup_{\substack{B_R(z_0)\subseteq\mathbb{C}\setminus B_{1+\varepsilon}(0)\\ \text{or } B_R(z_0)\subseteq B_{1-\varepsilon}(0)}} |\bar{\mu}_n(B_R(z_0)) - \mu_\infty(B_R(z_0))| \lesssim e^{-n\varepsilon^2}. \tag{1.5}$$

Here and in the sequel $\sim$ denotes asymptotic equivalence, $\lesssim$ will denote an inequality that holds up to a parameter-independent constant $c > 0$ that may differ in each occurrence. Moreover we write $A \asymp B$ if $c\,|B| \le |A| \le C\,|B|$ for some constants $0 < c < C$.

According to (1.4), the optimal rate of convergence of μ_n to the Circular Law turns out to be $\mathcal{O}(1/\sqrt{n})$. This follows directly from $D(\bar{\mu}_n,\mu_\infty) \le \mathbb{E}\,D(\mu_n,\mu_\infty)$. Interestingly, if one stays away from the edge $\partial B_1(0)$ at a fixed distance $\varepsilon > 0$, then (1.5) implies that the rate of convergence is exponentially fast. This is a striking difference to the Hermitian case, where the optimal rate of convergence to the Semicircle Law is $\mathcal{O}(1/n)$ even inside the bulk, see [GFF05; GT05; KB02].

Nevertheless we cannot expect an exponentially fast rate of convergence for the *non-averaged* empirical spectral distribution μ_n, since it is still sensitive to individual eigenvalue fluctuations. In particular, for each fixed set of eigenvalues $\{\lambda_i\}_{i\le n}$ we may select a ball of radius $(10\sqrt{n})^{-1}$ contained in $B_1(0)$ such that it does not cover any eigenvalue and obtain $D(\mu_n,\mu_\infty) \gtrsim 1/n$.

Heuristically, the typical distance between n uniformly distributed eigenvalues is $n^{-1/2}$. Therefore one may vary $B_R(z_0)$ up to a magnitude of $n^{-1/2}$ without $B_{R+n^{-1/2}}(z_0)$ covering a new eigenvalue. This leads to a deviation in μ_∞ of order $n^{-1/2}$ and hence we expect $D(\mu_n,\mu_\infty)$ to be of order $n^{-1/2}$ as well. We confirm this rate of convergence in Theorem 3.7 up to a logarithmic factor. In accordance with the Universality phenomenon, one of our main results states that a nearly optimal rate of convergence still holds for non-Gaussian entry distributions of the underlying matrix X. Under certain conditions that we specify in Definition 2.14, we will show the following theorem in Chapter 4.

Theorem 4.5. *If X satisfies the conditions of Theorem 1.2 and in addition its entries have bounded moments, then for every (small) $\varepsilon > 0$ and (large) $Q > 0$*

$$\mathbb{P}\left(D(\mu_n,\mu_\infty) \leq n^{-1/2+\varepsilon}\right) \geq 1 - n^{-Q} \tag{1.6}$$

holds for sufficiently large n.

Thus, with overwhelming probability the distance between the ESD and the Circular Law does not considerably exceed the optimal rate $1/\sqrt{n}$. In view of Lemma 3.1, the obtained rate of convergence does not depend on the distribution of the entries, it is *universal*. Most importantly, such a bound on D allows to choose the (worst) ball $B_R(z_0)$ depending on the random sample of the eigenvalues $(\lambda_j(X(\omega)/\sqrt{n}))_j$, cf. Figure 1.2. In other words eigenvalues do not cluster anywhere and do not have gaps or areas that are too sparse. In contrast to that, a Local Law can be interpreted as a non-uniform rate at fixed position z_0 of shrinking radius $R = n^{-a}$, $a \in (0,1/2)$.

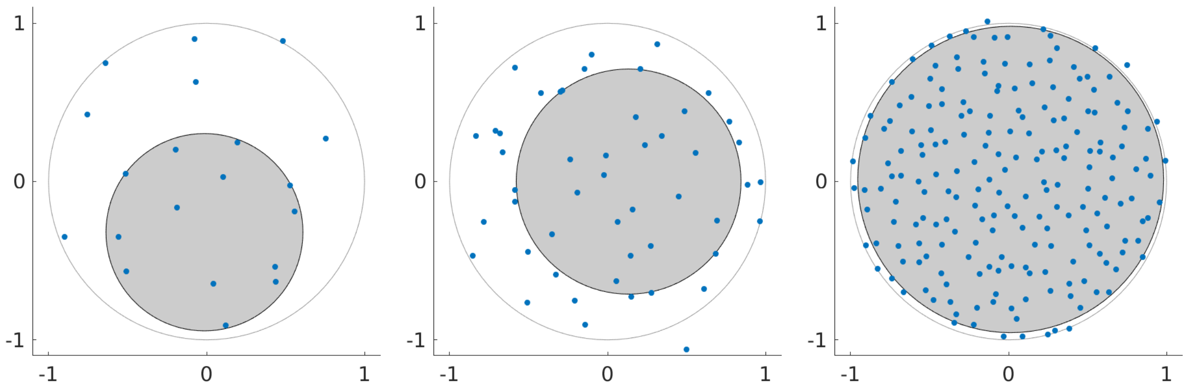

Figure 1.2: Samples of the spectrum of X for $n = 20$, 50, 200 and the gray Ball $B_R(z_0)$, which attains the supremum in D. Theorem 4.5 above shows that clusters (like in the left sample) and sparse areas (like in the middle sample) do not significantly differ from the uniform distribution. Here, we chose the entries to be uniformly distributed over a centered square in $\mathbb{C}$. Even for these non-Gaussian entries, we expect the maximizer $B_R(z_0)$ to be close to $B_1(0)$. This statement is exact for Ginibre matrices, as we see in the proof of Lemma 3.1.

The proof of Theorem 4.5 relies on a new *Smoothing Inequality*, Theorem 4.2, that quantitatively relates the uniform Kolmogorov distance D to a concentration of the logarithmic potentials. Albeit the Smoothing Inequality has an interesting interpretation and might be of independent interest, we will postpone its discussion to Section 4.1. In particular we will introduce and discuss the importance of logarithmic potentials in Section 2.2 together with other necessary technicalities. In the second step after the Smoothing Inequality, parts of the proof of a Local Circular Law from [AEK19] are used in order to control the distance of logarithmic potentials.

This approach is quite robust in the sense that the Smoothing Inequality for logarithmic potentials applies in many other settings. For instance in Theorem 4.7, the assumptions are weakened at the cost of constraining to the interior $B_{1-\tau}(0)$, the so called *bulk of the spectrum*. Moreover we provide an analogue for the classical two dimensional Kolmogorov distance

$$d_K(\mu_n,\mu_\infty) = \sup_{s,t\in\mathbb{R}} |(\mu_n - \mu_\infty)((-\infty,s] \times (-\infty,t])| \lesssim n^{-1/2+\varepsilon} \tag{1.7}$$

with overwhelming probability in Theorem 4.6. Our results on the rate of convergence to the Circular Law have been obtained in [GJ18].

Let us compare our results to related works. Tao and Vu [TV08] showed that $d_K(\mu_n,\mu_\infty) \lesssim n^{-\eta}$ holds $\mathbb{P}$-a.s. for some unknown $\eta > 0$, if the entries have finite $2+\varepsilon$-moments. Comparing this to our result (1.7), we see that a nearly optimal rate of convergence is obtained which holds with overwhelming probability. On the other hand a stronger moment assumption on the entries is needed. In particular, the explicit rate of convergence (1.7) gives a partial answer to an open problem mentioned in [TV09a].

As already discussed above, non-uniform rates can be read off from Local Circular Laws [BYY14a; BYY14b; TV15; GNT19a; AEK19] and fluctuations of linear spectral statistics [RS06; KOV18]. Note that these results deal with certain classes of smooth functions, whereas our metric uniformly covers a class of non-smooth indicator functions.

In the special case of Gaussian entries, i.e. for the Ginibre ensemble, pointwise convergence of the density of $\bar{\mu}_n$ has also been discussed in [AC18; TV15], similar to the integrated version (1.5). Furthermore $\mathbb{P}$-a.s. convergence rates of the same order $\sqrt{\log n}/n^{1/4}$ in p-Wasserstein distance d_{W_p} for $1 \leq p \leq 2$ have been proven in [MM15]. Very recently this has been extended by O'Rourke and Williams [OW19] to matrices satisfying a stronger condition (A) (see Definition 2.14), where a non-optimal rate of $\mathcal{O}(n^{-1/4+o(1)})$ in 1-Wasserstein distance has been shown. Though the distances d_{W_1} and D are not directly comparable, see Section 2.3, both optimal rates are expected to be $n^{-1/2}$ up to logarithmic factors.

More generally, Chafaï, Hardy and Maïda studied invariant β-ensembles with external potential V instead of independent-entry matrices in [CHM18]. Their result implies a rate of convergence to the limiting measure with density $c\Delta V$ of order $\sqrt{\log n/n}$ with respect to the bounded Lipschitz metric and the 1-Wasserstein distance. The paper [CHM18] is also based on an inequality between distances of measures to their energy, an integrated logarithmic potential, similar to our Smoothing Inequality. However it relies critically on the existence of a confining potential, hence a joint probability density function for the eigenvalues. Note that their result is given for a Coulomb gas point process in arbitrary dimension $d \in \mathbb{N}$, yielding a bound of order $n^{-1/d}$ up to logarithmic factors. This coincides with the rate of order $1/n$ for the Semicircle Law for $d = 1$ as well as the optimal order $1/\sqrt{n}$ in the Circular Law. Similar questions in this context of log-gases, but for the non-uniform variant of D (the discrepancy) have been addressed in [Ser17].

1.3. Products of random matrices

The Circular Law states that the empirical spectral distribution of a single non-Hermitian random matrix with i.i.d. entries converges to the uniform distribution on the complex disk as the size of the matrix tends to infinity. Interestingly, for the product of m independent matrices of such type the limit coincides with the distribution of the m-th power of a random variable uniformly distributed in the unit disk. Let us consider the product

$$\mathbf{X} = \frac{1}{\sqrt{n^m}} \prod_{q=1}^{m} X^{(q)} \tag{1.8}$$

of m independent random matrices $X^{(1)}, \ldots, X^{(m)}$, each of size $n \times n$. For fixed $m \in \mathbb{N}$, the asymptotic in $n \to \infty$ will be of interest. Its empirical spectral distribution is given by

$$\mu_n^m = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j(\mathbf{X})}. \tag{1.9}$$

Note that $\mu_n^1 = \mu_n$ from the previous section. The empirical spectral distribution of $\mathbf{X}$ also converges weakly to a deterministic probability measure on the complex plane that generalizes the Circular Law.

Theorem 1.3 ([GT10b]). *If $X^{(1)}, \ldots, X^{(m)}$ have independent entries satisfying $\mathbb{E}\, X_{ij}^{(q)} = 0$ and $\mathbb{E}|X_{ij}^{(q)}|^2 = 1$ for all $q = 1, \ldots, m$, then $\mathbb{P}$-a.s. we have*

$$\mu_n^m \Rightarrow \mu_\infty^m, \text{ where } d\mu_\infty^m(z) = \frac{|z|^{2/m-2}}{\pi m} \mathbb{1}_{B_1}(z) d\lambda(z) \tag{1.10}$$

is the m-th power of the uniform distribution $\mu_\infty = \mu_\infty^1$ on the complex unit disc.

The Gaussian case has been treated in [BJW10; AB12], more general models can be found in [KT15; GKT15; Bor11; AI15; IK14], for the convergence of the singular values see [AGT10] and furthermore for local results we refer to [Nem17; KOV18; Nem18; GNT19a; CO19].

Recalling the approach of the previous section, one may ask for generalizations of Lemma 3.1 and Theorem 4.5. In particular it is interesting whether the optimal rate of convergence for products of independent Ginibre matrices depends on m. Surprisingly, the answer is negative.

Theorem 3.2. *The mean empirical spectral distribution $\bar{\mu}_n^m = \mathbb{E}\,\mu_n^m$ satisfies*

$$\sup_{R>0} |\bar{\mu}_n^m(B_R) - \mu_\infty^m(B_R)| \asymp \frac{1}{\sqrt{n}}.$$

The following more detailed estimates hold as long as the boundary of the complex disk is avoided

$$\sup_{R<1-\frac{m}{2}\sqrt{\log n/n}} |\bar{\mu}_n^m(B_R) - \mu_\infty^m(B_R)| \lesssim \frac{\log^{3/2} n}{n}$$

and uniformly in $R > 1 + \sqrt{\log n/n}$

$$|\bar{\mu}_n^m(B_R) - \mu_\infty^m(B_R)| \lesssim e^{-n(R-1)^2}.$$

While the proof of Lemma 3.1 is an elementary calculation, the proof of Theorem 3.2 is more technical and relies on a saddle-point method of a double contour integral representation for the density of μ_n^m. Figure 1.3 illustrates the statements of Theorem 3.2.

The exact constants of the upper and lower bound can be chosen to be $C = \sqrt{\pi/2m}$ and $c = 1/(\sqrt{2\pi m})$, coinciding with Lemma 3.1. We will see that the maximal distance is attained at $R = 1$. The rate of convergence is much faster inside and outside of the bulk.

Theorem 3.2 provides the optimal rate of convergence for centered balls. Using adapted techniques, we will also prove a nearly optimal rate of convergence result for products of matrices with independent entries. To make the statement more comprehensible when comparing with Theorem 3.2, we state the following result. These statements on the rate of convergence of products of random matrices have been obtained in [Jal19].

Corollary 4.10. *Let $X^{(1)}, \dots, X^{(m)}$ be independent matrices with independent entries. If $\mathbb{E}\, X_{ij}^{(q)} = 0$ and $\mathbb{E}\,\big|X_{ij}^{(q)}\big|^2 = 1$ for all $q = 1, \dots, m$ and $\max_{i,j,q,n} \mathbb{E}\,\big|X_{ij}^{(q)}\big|^{4+\delta} < \infty$ for*

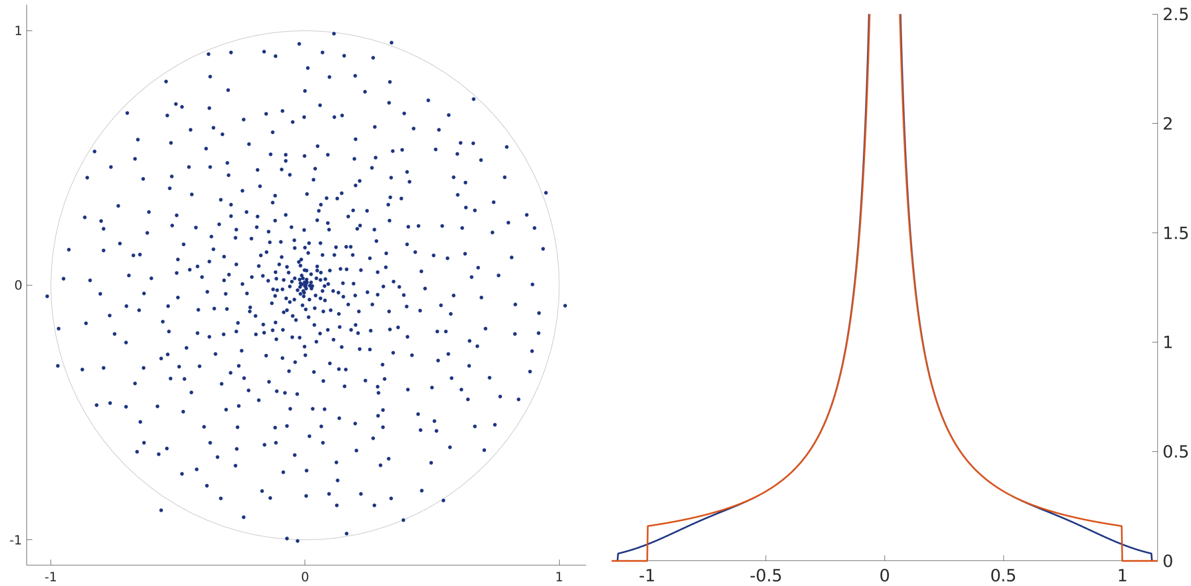

Figure 1.3: The empirical spectral distribution of the product $\mathbf{X}$ of $m = 2$ Ginibre matrices
Left: The eigenvalues of a sample for $n = 500$ and the unit ball $B_1(0)$ as reference. Corollary 4.10 shows that gaps (like one can see at the top) and clusters do not significantly differ from the limiting distribution.
Right: The radial part of the densities of $\bar{\mu}_n^m$ for $n = 15$ in blue and of the limiting distribution μ_∞^m in orange. Clearly the rate of convergence in the bulk is faster than close to the edge, illustrating the statement of Theorem 3.2.

some $\delta > 0$, then for every τ,$Q > 0$ we have

$$\mathbb{P}\left(\sup_B |(\mu_n^m - \mu_\infty^m)(B)| \lesssim \frac{\log^2 n}{\sqrt{n}}\right) \geq 1 - n^{-Q},$$

where the supremum runs over all balls B such that $\partial B_R(z_0) \subseteq B_{1+\tau}^c \cup B_{1-\tau} \setminus B_\tau$ avoids the edge and the origin.

We already know from Theorem 3.2 that the optimal rate is given by $\mathcal{O}(1/\sqrt{n})$, hence Corollary 4.10 shows that this rate is also satisfied for matrices with independent entries, as long as edge and origin are avoided. Comparing the previous Theorem 4.5 for $m = 1$ to Corollary 4.10, we remark that the latter holds for products of m matrices under weaker assumptions at the cost of being restricted to balls in the bulk. The rate of convergence however is much more precise and close to the optimal rate.

1.4. Random polynomials

Eigenvalues of a (random) matrix are zeros of the characteristic polynomial f_n, e.g. for $n = 2$, we have $f_2(z) = z^2 - \mathrm{trace}(X)z + \det(X)$. If the entries are independent random variables, we notice a certain dependency of the coefficients of the corresponding characteristic polynomial. Instead, we may consider a random polynomial with *independent identically distributed coefficients* ξ_k that are centered and normalized. We call

$$f_n(z) = \sum_{k=0}^{n} \sqrt{\frac{n^k}{k!}} \xi_k z^k,$$

a Weyl polynomial. In the same spirit as in RMT, we associate to a random polynomial f_n its multiset of zeros $\Lambda := \{\lambda \in \mathbb{C} : f_n(\lambda) = 0\}$ and its empirical measure $\mu_n^W = \frac{1}{n} \sum_{\lambda \in \Lambda} \delta_\lambda$. Interestingly, Kabluchko and Zaporozhets [KZ14a] showed that the *Circular Law* also holds for μ_n^W, if $\mathbb{E} \log(1 + |\xi_0|) < \infty$, i.e. $\mu_n^W \Rightarrow \mu_\infty$ holds $\mathbb{P}$-a.s..

As an application of our Smoothing Inequality that we used in RMT, we obtain the following rate of convergence for random polynomials with independent coefficients by using the same versatile approach.

Theorem 5.2. *If* $\mathbb{E}\,\xi_0 = 0$*,* $\mathbb{E}\,|\xi_0|^2 = 1$*,* $\mathbb{E}\,|\xi_0|^{2+\delta} < \infty$ *and* $\mathbb{E}\,|1/\xi_0|^\delta < \infty$ *for some* $\delta > 0$*, then for every* ε*,*$Q > 0$ *and sufficiently large* n *we have*

$$\mathbb{P}(D(\mu_n^W, \mu_\infty) \leq n^{-1/2+\varepsilon}) \geq 1 - n^{-Q}.$$

1.5. The spectral radius

So far, we considered the rate of weak convergence of the empirical measures, or, in other words we uniformly looked at at all eigenvalues simultaneously. In the end of this thesis, we will briefly change the viewpoint and look at one specific eigenvalue that is important for the large n behavior in the Circular Law. The largest absolute value of the eigenvalues, given by the the *spectral radius*, converges

$$|\lambda|_{\max} := \max\left\{ \left|\lambda_j(X/\sqrt{n})\right| : 1 \leq j \leq n \right\} \to 1 \tag{1.11}$$

$\mathbb{P}$-a.s. as $n \to \infty$, see [BS10; BCCT18].

In Hermitian RMT, the largest (real) eigenvalue is fairly well understood, see [LY14; TW94; Sos99; EKYY13; EKYY12; BY88; Joh07] just to name a few. Its fluctuation around the edge is of order $n^{-2/3}$ and given by the ubiquitous Tracy-Widom distribution.

For Ginibre matrices Kostlan [Kos92] discovered and Rider [Rid03] reestablished that the correct fluctuation of $|\lambda|_{\max}$ is of order $n^{-1/2}$ up to logarithmic factors and is given by a Gumbel distribution. For non-Gaussian entries the question about the fluctuation of $|\lambda|_{\max}$ remains open.

An application of the rate of convergence (1.6) readily implies a lower bound on the spectral radius $|\lambda|_{\max}$ of order $1 - n^{-1/2+\varepsilon}$. Recently, Alt, Erdős and Krüger [AEK19] managed to show a rate of convergence

$$\mathbb{P}(||\lambda|_{\max} - 1| \leq cn^{-b}) \geq 1 - o(n^{-b}).$$

for any $0 < b < 1/2, Q > 0$ in a very general setting. At the time of this thesis being written, we used very different techniques, namely matching moments up to order 6, and proved a first non-optimal rate $\mathbb{P}(||\lambda|_{\max} - 1| \leq cn^{-b}) \geq 1 - o(n^{-b})$ for some unknown $b > 0$. Though this result is outdated by now, we will state it in case our method of proof may provide new insight into similar problems.

1.6. Structure of the thesis

The structure of this work is essentially already indicated by the preceding introduction.

Chapter 2 introduces basic methods that will be used in the sequel. We provide historical remarks along with necessary concepts of Random Matrix Theory. We will see why tools from Hermitian RMT do not apply in the non-Hermitian setup and how Girko's Hermitization Trick builds the bridge between both areas. Experts in the area of RMT may skip these sections. After presenting the logarithmic potential, important known results will be collected. Furthermore, we will clarify in which sense the problem of rate of convergence is to be understood and motivate our investigations with numerical simulations.

Chapter 3 is dedicated to the mean ESD of Gaussian matrices. The results here will show exact upper and lower bounds, so that later results can be understood to be optimal and universal. First, we consider a single random matrix and then turn to products of matrices. The study of Gaussian matrices will be completed with an upper bound for the rate of convergence of the non-averaged ESD.

Chapter 4 contains estimates on the rate of convergence for matrices with independent entries, in particular non-Gaussian entries. We begin with a Smoothing Inequality for logarithmic potentials that provides a unified approach for rate of convergence results. Upper bounds on the rate of convergence for the ESD with overwhelming probability will follow. As before, we first discuss single matrices and turn to products afterwards.

Chapter 5 will be about how to apply our ideas to related models. As an example, we will show a similar rate of convergence result for the zeros of random polynomials with independent coefficients. Moreover, we will compare previously discussed results to the spectral radius. Using moment matching techniques, we shall prove a rate of convergence for the spectral radius.

Chapter 2
Preliminaries

2.1. (Hermitian) Random Matrix Theory

Random Matrix Theory is the study of matrix-valued random variables. Although this statement is a natural starting point for many mathematical textbooks, it does not help to grasp the importance of the subject. Instead, we will give some historical background information first and turn to the development of the theory afterwards. Here and there, we will provide the necessary definitions whenever necessary and emphasize only the most important ones for this thesis.

2.1.1. Basics

The first notable appearance of a random matrix was in 1928, when John Wishart [Wis28] studied large sample covariance matrices. We call X a *random matrix*, if it is a random variable taking values in a space of matrices, say $\mathbb{C}^{n\times n}$. Wishart's covariance matrices take the form $X = (Y - \mathbb{E}Y)(Y - \mathbb{E}Y)^*$ for some random (not necessarily square) matrix Y. These random matrices do not only play an important role in multivariate statistics and computer science but similarly structured matrices will also appear in this thesis, e.g. (2.32).

In fact, the starting point of Random Matrix theory is motivated by nuclear physics. In order to understand the energy levels of heavy nuclei, in principle one needs to solve the eigenvalue problem

$$H\psi_k = E_k\psi_k,$$

where H is a Hermitian operator, called Hamiltonian, with (real) eigenvalues E_k which describe the energy of the quantum system in state ψ_k, the corresponding eigenfunctions. Systems for nuclei consisting of a hundred or more nucleons are non-integrable, i.e. it is too complicated to be solved exactly. Instead Wigner and Dyson suggested to study the statistics of the energy distribution, see e.g. [Wig55; Wig58; Dys62b]. A reasonable approximation of H could be a random self adjoint matrix X of large dimension n - Random Matrix Theory was born.

> "Fifty years ago, the world-wide community of experts in the theory of random matrices consisted of about ten people. There was our leader Eugene Wigner who first invented the subject." – Freeman Dyson [ABD11]

As has been mentioned in the introduction, Eugene Wigner discovered the Semicircle Law in 1955. He found that the empirical eigenvalue distribution (ESD) of correctly scaled Hermitian random matrices of growing size $n \to \infty$ tends almost surely to a semicircle distribution with density $\rho_{sc}(x) = \frac{1}{2\pi}\sqrt{(4-x^2)_+}$. More generally this holds for *Wigner Matrices* X, which have independent entries X_{ij}, $1 \le i \le j \le n$, up to the symmetry constraint with $\mathbb{E}\, X_{ij} = 0$ and $\mathbb{V}\mathrm{ar}\, X_{ij} = 1$.

In 1967, Vladimir Marčenko and Leonid Pastur obtained a limiting distribution for the ESD of *covariance matrices* X^*X depending on the limit of the ratio between row size and column size of the matrix X. Its eigenvalues $s_1^2 \ge \ldots \ge s_n^2 \ge 0$ are squared singular values of X, hence we define the *empirical singular value distribution*

$$\nu_n = \sum_{j=1}^{n} \delta_{s_j(X/\sqrt{n})}$$

of X to be the ESD of $\sqrt{X^*X}/\sqrt{n}$. Let us state one special case of the Marčenko-Pastur Law that is mentioned for it's relation to non-Hermitian RMT as we will see later.

Theorem 2.1 (Quarter Circular Law, [MP67])**.** *Let X be an $n \times n$ random matrix with independent entries. If $\mathbb{E}\, X_{ij} = 0$ and $\mathbb{E}\,|X_{ij}|^2 = 1$, then $\mathbb{P}$-a.s. we have*

$$\nu_n \Rightarrow \nu_\infty, \text{ where } d\nu_\infty(x) = \frac{1}{\pi}\sqrt{4-x^2}\,\mathbb{1}_{[0,2]}dx \tag{2.1}$$

is the quarter Circular Law.

The shape of the density of ν_∞ is a quarter circle in contrast to the linear limiting distribution of radii of eigenvalues of X. This indicates asymptotical non-normality[1] of the random matrices X.

An obvious link between the limiting law's that we mentioned so far is as follows. Given the Circular Law μ_∞, its rescaled real marginal distribution is the Semicircle Law that is also the symmetrization of ν_∞. Hence for later purposes we will write $\widetilde{\nu}_\infty$ for the Semicircle Law. On the level of eigenvalues one may introduce correlations $\mathbb{E}\, \bar{X}_{ij}X_{ji} = \tau$ between the off diagonal entries of X. Then, $\tau = 0$ corresponds to independent entries leading to the Circular Law, whereas $\tau = 1$ forces X to be a (selfadjoint) Wigner Matrix and hence its real eigenvalues converge to the Semicircle Law. The transition for arbitrary $\tau \in [0,1]$ is called *Elliptic Law* with its obvious limiting shape, see [Nau12].

On the other hand, the (chiral) block matrix

$$V = \begin{bmatrix} 0 & X/\sqrt{n} \\ X^*/\sqrt{n} & 0 \end{bmatrix}. \tag{2.2}$$

1 A matrix A is normal iff $A^*A = AA^*$ and that is the case iff A is unitarily diagonalizable

has eigenvalues $\pm s_j$, where s_j are the singular values of X. The Wigner-type block structure suggest that its ESD converges to the semicircle distribution, which is indeed the case. An important link between μ_∞ and ν_∞ will be presented in Subsection 2.2.2.

Since Wigner discovered his result, the Semicircle Law and its local versions have also been proven for various other structured matrices, e.g. adjacency matrices of Erdös Rényi random graphs [EKYY13; EKYY12] and some d-regular graphs [TVW13], band matrices [EYY12a], matrices with comparable variances [Erd11] and stochastic matrices [GNT15d; Erd11]. Many more detailed limit theorems, e.g. on the fluctuations of linear statistics, edge and gap distribution, were obtained in the past, but in order to avoid an extensive list, we refer to the monographs [PS11; AGZ10].

2.1.2. Methodologies

In this subsection we would like to give a short overview on common techniques used in Hermitian RMT without going into the details, but rather pointing out why most of them are not applicable for non-Hermitian random matrices. Parts of it we mention in order to understand the difference to their non-Hermitian counterparts and parts because we will need them in subsequent sections.

Truncation and perturbation. A common first simplifying step for the proofs of statements in Hermitian RMT is *truncation*. Truncating large entries of a Hermitian matrix (and centering again) usually does not change the limiting distribution, see [GNT15a, Appendix D]. This is due to the fact that the spectrum of normal matrices, in particular of Hermitian matrices, is stable under small perturbations. One may view this as a consequence of the property that the ε-pseudospectrum of a normal matrix coincides with the ε-neigborhood of its eigenvalues, see [TE05]. Changing only a few entries or slightly modifying all entries will only cause minor perturbation in the ESD. More precisely, one may use rank or trace inequalities, see for instance [BS10, Appendix A.5]. Non-normal matrices however fail to be stable under perturbations as the following counterexample shows.

Counterexample 2.2. Define the non-normal $n \times n$-matrices

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ 0 & 0 & \cdots & \cdots & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ n^{-10} & 0 & \cdots & \cdots & 0 \end{pmatrix}. \tag{2.3}$$

Albeit their difference is a perturbation of rank 1 and has small operator norm $\|A - B\| = n^{-10}$, the spectra differ significantly. Since A is nilpotent, its only eigenvalue is 0, but B has all eigenvalues lying on the circle of radius $n^{-10/n}$. In particular the large n limiting distribution δ_0 of the ESD of A is strikingly different from the one of B, which is the uniform distribution on $S^1 = \partial B_1(0)$.

This counterexample and the non-validity of rank and trace inequalities suggest that direct truncation of the entries of non-Hermitian random matrices does not lead to a simplification of the problem. Furthermore, methods that use the perturbation of an arbitrary Wigner matrix with a small Gaussian noise, e.g. in [ESY11; Joh01; BP05; EPR10], also can't be generalized in order to apply them to non-Hermitian random matrices.[1]

Method of Moments. If all moments of a sequence of probability measures converge to the moments of a limiting distribution, which is determined by its moments, then the measures also converge weakly. Since limiting measures of ESD's of (most) random matrices have compact support, their moments fully characterize the measure. In general, this determinancy of the *Moment problem* is a delicate question that holds true if the so called Carleman condition is satisfied and counterexamples are given by Krein's condition, see [Sch]. The k-th moment of the mean empirical spectral distribution $\bar{\mu}_X = \mathbb{E}\,\mu_X$ of any $n \times n$ random matrix X is fairly simple to study because of the following identity[2]

$$\int x^k d\bar{\mu}_X(x) = \frac{1}{n}\,\mathbb{E}\,\mathrm{trace}X^k = \frac{1}{n}\sum_{i_1,\ldots,i_k=1}^{n} \mathbb{E}\left(X_{i_1,i_2}X_{i_2,i_3}\cdots X_{i_k,i_1}\right). \tag{2.4}$$

Excluding certain combinations of the product and counting the combinatorial factor, already Wigner used this method to prove the Semicircle Law, cf. [Wig55; BS10; AGZ10].

For probability measures on $\mathbb{C}$, however, even a bounded support is not sufficient for the measure being characterized by its moments.

Counterexample 2.3. The moments of the Circular Law are vanishing $\int x^k d\mu_\infty(x) = 0$ as well as the moments of the complex Gaussian distribution $\mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ and any other sufficiently light tailed rotationally invariant distribution.

Naively one may hope that passing over to mixed moments of the form $\int x^k \bar{x}^l d\mu$ would be a solution, but as explained in [MS17, Section 11] non-normality prevents us from passing over to mixed normalized traces.[3]

Occasionally, the method of moments is still used in current research of normal

1 Note that the eigenvalues of such matrices have the same distribution as those of Dyson's Brownian motion, which evolve according to a coupled stochastic differential equation, see [Dys62a].

2 This identity serves as a definition of the distribution of non-commutative random variables in free probability theory. Without going into the details, we refer to [AGZ10; NS06] and only mention selected connections in the sequel.

3 Ultimately, free probability theory provides a way to circumvent the issue by introducing the so called $\star$-distribution. Replacing X^k in(2.4) by $X^{\varepsilon_1}\cdots X^{\varepsilon_k}$ for $\varepsilon_i \in \{1,*\}$, it is possible to show that the ESD of Ginibre matrices converge in $\star$-moments to the Circular Law, see [Sni02] or [MS17, p. 11.6.3].

random matrices, e.g. [Sos99; DS94], though it has been outdated by other more versatile techniques, such as the following.

Stieltjes transform. In classical probability theory arguably the most important tool used in the study of sums of independent random variables is the Fourier transform. The *Stieltjes transform* takes on the same significance in RMT.

Definition 2.4. For any probability measure μ its Stieltjes transform is defined via

$$m_\mu : \mathbb{C} \setminus \mathbb{R} \to \mathbb{C}\,, \quad w \mapsto \int \frac{1}{t-w} d\mu(t).$$

Analogously to the role of the Fourier transform, the Stieltjes transform captures all the information of the measure itself. It can be inverted and its pointwise convergence coincides with weak convergence of the measures.

Proposition 2.5. *Let μ_n, $n \in \mathbb{N}$, and μ be probability measures on $\mathbb{R}$.*

1. *m_μ is analytic, more precisely it is a Nevanlinna function, with $|m_\mu(E+i\eta)| \le \frac{1}{|\eta|}$.*
2. *For any open interval I with no boundary point being an atom of μ*

$$\mu(I) = \lim_{\eta \searrow 0} \frac{1}{\pi} \int_I \operatorname{Im}(m_\mu(E+i\eta))dE. \tag{2.5}$$

3. *Weak convergence $\mu_n \Rightarrow \mu$ holds if and only if $m_{\mu_n}(w) \to m_\mu(w)$ for all $w \in \mathbb{C} \setminus \mathbb{R}$.*

It is important to mention that all properties mentioned above rely only on the knowledge of $m_\mu(w)$ for values of w that lie outside of $\operatorname{supp}\mu \subseteq \mathbb{R}$. We refer to [AGZ10] for more general statements, its link to the generating function of moments and the proof of this proposition. It is also easy to see that $\frac{1}{\pi}\operatorname{Im}(m_\mu(E+i\eta))$ is nothing but the density of μ convoluted with a centered Cauchy distribution of scale parameter η.

It is well known that the Stieltjes transform of the semicircle distribution $\widetilde{\nu}_\infty$ is given by $s(w) = -\frac{1}{2}(w - \sqrt{w^2-4})$[1] and satisfies the equation

$$s(w) = -\frac{1}{w+s(w)}. \tag{2.6}$$

The Stieltjes transform was first used in RMT in 1967 by Marčenko and Pastur [MP67; Pas73; BS10]. Its usefulness is due the fact that the Stieltjes transform of an ESD μ_X of any Hermitian random matrix X is given in terms of its resolvent $(X-w)^{-1}$ (or Green's

1 The branch of the root is chosen such that $\operatorname{Im} s \ge 0$ in accordance with (2.5).

function)

$$m_{\mu_X}(w) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j(X) - w} = \operatorname{trace}(X - w)^{-1}. \tag{2.7}$$

Here and in the sequel we abbreviate the diagonal matrix Iw by w. By applying Schur's inversion formula (see [HJ12]) to the resolvent and denoting $X^{(k)}$ to be the sub matrix of removing the k-th column and row of X, one may see that $m_{\mu_X}(w)$ satisfies

$$m_{\mu_X}(w) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{X_{kk} - w - \sum_{i,j\neq k}(X^{(k)} - w)^{-1}_{ij} X_{kj}X_{ki}}. \tag{2.8}$$

Up to an error ε this is close to a recurrence relation (or self-consistent equation), since the last term in the denominator is expected to be close to $\mathbb{E}\, m_{X^{(k)}}(w)$ and hence it can be seen as a perturbed version of (2.6). This is the standard approach to prove the Semicircle Law or Theorem 2.1 for scaled Wigner matrices $X/\sqrt{n}$, see for instance [AGZ10, §2.4.2.] or [BS10, §2.3 and §3.3].

Notably, this idea is still at the heart of the proof of many results in RMT. For instance, Erdös, Schlein and Yau, in [ESY11], gained control over the error ε that leads to the so called *Local Semicircle Law* which has been improved and generalized extensively during the last years, hence we may refer to the surveys [BK16; Erd11; TV12]. Roughly speaking, the Local Law states that with overwhelming probability[1] for any spectral parameter E in the bulk we have

$$\left| m_{X/\sqrt{n}}(E + i\eta) - s(E + i\eta) \right| \lesssim \frac{\log n}{n\eta}, \tag{2.9}$$

see for instance the most recent result in [GNT19b]. Taking into account the Stieltjes inversion formula (2.5), the Local Semicircle Law states that as the scaling η becomes smaller, the ESD is getting closer to the Semicircle Law on sets containing much less than $\mathcal{O}(n)$ eigenvalues. The Local Semicircle Law will only hold up to the optimal scaling $\eta = 1/n$, when finitely many eigenvalues are considered, since then individual eigenvalue fluctuations appear that are studied in terms of correlation functions.

The Local Semicircle Law is the key tool that provides access to many other local problems such as eigenvector delocalization, eigenvalue rigidity [EKYY13; EYY12b; TV10a; GNT15a], Universality of the correlation functions [EPR10; LY14; TV11] and the rate of convergence to the Semicircle Law [GNTT18; GT03a]. Generally speaking, there is no doubt that the Stieltjes transform method is the most powerful tool in RMT.

1 A sequence of events Ω_n is said to hold with overwhelming probability (in short w.o.p.) if $\mathbb{P}(\Omega_n^c) \lesssim n^{-Q}$ for any $Q > 0$ and it holds with high probability if $\mathbb{P}(\Omega_n^c) = o(1)$.

In order to show *rate of convergence* results for Wigner matrices, one needs a quantitative direct relation between the Stieltjes transform and the distance of measures. A prototype of such so called Smoothing Inequalities has been proven by Bai [Bai93a; Bai93b; BS10]. The term “smoothing” is due to $\frac{1}{\pi}\mathrm{Im}(m_\mu(E+i\eta))$ being the distribution μ mollified by a Cauchy kernel, as mentioned above.

Theorem 2.6 (Bai’s inequality)**.** *Let μ,ν be probability distributions on $\mathbb{R}$ with Stieltjes transforms m_μ,m_ν and distribution functions F_μ, F_ν satisfying $\int |F_\mu(x) - F_\nu(x)|\, dx < \infty$.*[1]*. Then, for any $\alpha > 0$ there exists a constant $\gamma > 1/2$ such that the Kolmogorov distance can be estimated by*

$$d^*(\mu,\nu) = \sup_{x\in\mathbb{R}} |F_\mu(x) - F_\nu(x)| \tag{2.10}$$

$$\leq \frac{1}{\pi(2\gamma-1)} \int_{-\infty}^{\infty} |m_\mu - m_\nu|\,(u+iv)du + \frac{1}{v}\sup_{x\in\mathbb{R}} \int_{-2v\alpha}^{2v\alpha} \nu((x,x+y])dy. \tag{2.11}$$

The proof of one of our main results in Section 4.1, the rate of convergence to the Circular Law, will rely on a similar idea.

Unfortunately, a direct generalization of the Stieltjes transform to complex measures is not useful at all. If X is a non-Hermitian random matrix satisfying the conditions of the Circular Law, Theorem 1.2, then $\mathrm{trace}(X/\sqrt{n} - z)^{-1}$ is unbounded inside the support of μ_∞.[2] This causes serious problems for its analysis and therefore convergence seems unlikely, though convergence for $|z| > 1$ is doable, see [BS10]. The reason for the uselessness is that, unlike for measures supported on the real line, it is not enough to know the values of the Stieltjes transform outside the distribution’s support in order to recover the distribution itself.

Counterexample 2.7. By Cauchy’s integral formula we obtain the Stieltjes transform of the Circular Law

$$\hat{m}_{\mu_\infty}(z) = \int_{\mathbb{C}} \frac{1}{t-z} d\mu_\infty(t) = \frac{i}{\pi z}\int_0^1 \oint_{\partial B_1(0)} \frac{r}{\xi - r/z} d\xi dr = \begin{cases} -1/z & \text{, if } |z| > 1, \\ \bar{z} & \text{, if } |z| \leq 1. \end{cases} \tag{2.12}$$

But μ_∞ is not uniquely determined by the values of $\hat{m}_{\mu_\infty}(z)$ outside of its support $B_1(0)$. In particular any uniform distribution on a disk $B_R(0)$ with $R < 1$ leads to the same value $1/z$ in (2.12), hence the direct generalization of Proposition 2.5 to distributions

1 Note that this equals the Wasserstein metric $d_{W_1}(\mu,\nu)$.

2 The careful reader might have noticed that we changed the notation from the spectral parameter w for real eigenvalues to z for complex eigenvalues. This is intended and will become more clear in later considerations.

on the complex plane does not hold.

As we will see in Section 2.2 the key to studying complex spectra is to connect the non-Hermitian world with the Hermitian one for which the method of Stieltjes transform is applicable again.[1]

The GUE, orthogonal polynomials and the log-gas picture. A very special role among all Wigner matrices is taken by the so called Gaussian Unitary ensemble, in short *GUE*. These matrices have standardized complex Gaussian entries and, just like the Gaussian distribution, their distribution on the space of $n \times n$ Hermitian matrices is invariant under unitary conjugation. Denoting by dH the Lebesgue measure on the space of Hermitian matrices, we may write the (matrix-) distribution of a GUE matrix as

$$d\,\mathbb{P}(H) = Z_n^{-1} \cdot \exp\left[-\tfrac{1}{2}\mathrm{trace}(H^2)\right] dH \tag{2.13}$$

for some normalizing constant Z_n. Using the unitary invariance, it is possible to perform the integration over the unitary group, corresponding to the eigenvectors, so that an explicit formula for the joint probability density function of the eigenvalues can be derived

$$P(\lambda_1, \ldots, \lambda_n) = \widetilde{Z}_n^{-1} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_j|^2 \exp\Big[-\frac{1}{2}\sum_{j=1}^{n} \lambda_j^2\Big]. \tag{2.14}$$

The term $\prod_{1 \le j < k \le n}(\lambda_j - \lambda_j)$ is called Vandermonde determinant and comes from the Jacobian of the unitary diagonalization in the group integral, see [Meh67; AGZ10]. These group integrals do not decouple for non-Gaussian Wigner matrices and hence these matrices do not possess an explicit representation for the eigenvalue density. This is the main reason why Gaussian matrices have been studied earlier and why much more details are known. Basically, Random Matrix Theory splits into two areas, one that considers (invariant) ensembles defined via a density on some symmetric matrix space and one that considers matrices with independent entries. Mostly, both theories only intersect in Gaussian matrices and the question of Universality can always be asked in both directions.

Formula (2.14) gives rise to an interpretation from statistical mechanics of the eigenvalues being viewed as particles on the real line, confined by a potential $V = |\cdot|^2$ and

1 It also tempting to define a generalized Stieltjes transform by using a small quaternionic parameter that directs perpendicular out of the complex plane. However this naive idea only covers resolvent formulas as (2.7), if the non-Hermitian matrix is normal, see [FZ97]. Useful generalizations of a quaternionic resolvent, as described in [BC12] and [MS17, Section 11.9], will then be 2×2-matrix-structured paraphrase of the Hermitization method in Section 2.2.

having a logarithmic repulsion[1] from other particles. Nothing changes in the previously described derivation if we replace the potential V by any other admissible function. The particle system is then called *log-gas*[2] and we refer to the monograph [For10] for a detailed description. Moreover, the exponent $\beta = 2$ of the Vandermonde determinant may be replaced by any Dyson index $\beta > 0$ interpreted as the inverse temperature. Apart from $\beta = 2$, only $\beta = 1$ and $\beta = 4$ have interpretation in RMT, counting the number of real components in a self-adjoint matrix (real symmetric for $\beta = 1$ and quaternionic self dual matrices for $\beta = 4$). Already Dyson's threefold way [Dys62c] classified random matrices according to these three selfadjoint symmetry classes.

As it is done in [ABD11, §4] and [For10, §5], by introducing polynomials ψ_k that are orthogonal in L^2 with respect to the weight $w = e^{-V}$, one may rewrite the Vandermonde determinant, and hence the whole density given in (2.14) as a determinant over the Kernel

$$K_n(x,y) = \sqrt{w(x)w(y)} \sum_{k=0}^{n-1} c_k \psi_k(x)\psi_k(y).$$

Since the polynomials have been chosen to be orthogonal, also the marginal distributions of $k \leq n$ many eigenvalues can be calculated by using Dyson's theorem [Meh04, Theorem 5.1.4]. The so called *k-point correlation functions*

$$\varrho_k(x_1,\dots,x_k) := \frac{n!}{(n-k)!} \int P(x_1,\dots,x_n) dx_{k+1} \dots dx_n = \det\left[K_n(x_i,x_j)\right]_{1 \leq i,j \leq k}$$

have a determinantal structure, hence the eigenvalues form a *determinantal point processes*, cf. [Joh05; AGZ10; For10; Ser15] for more information. In the case of the GUE we have $w(x) = e^{-x^2/2}$ and ψ_k are Hermite polynomials. The normalized 1-point correlation function ϱ_1/n is the density of a randomly picked eigenvalue, and hence the density of the mean empirical spectral distribution.

In the next section, we shall see that this viewpoint carries over to particles on the complex plane, which are eigenvalues of Gaussian non-Hermitian random matrices.

2.2. Non-Hermitian Random Matrix Theory

As was mentioned in the introduction, non-Hermitian RMT emerged out of pure curiosity without any applications in mind. The first noteworthy appearance of non-Hermitian

1 This is visible if we write (2.14) as a Gibbs measure, so that the Vandermonde determinant appears as $\sum_{j<k} \log|\lambda_j - \lambda_k|$ in the energy term of exponent.

2 One may also see it as a Coulomb gas restricted to the real line, since the logarithm corresponds to the Coulomb interaction in dimension 2.

random matrices was in 1965, when Ginibre expressed his viewpoint on the topic as follows.

> "Apart from the intrinsic interest of the problem, one may hope that the methods and results will provide further insight in the cases of physical interest or suggest as yet lacking applications." – Jean Ginibre [Gin65]

Ginibre showed the first instance of the Circular Law, Theorem 1.2, stating that the mean empirical spectral distribution of Ginibre matrices converges to μ_∞, see also [Meh67; Ede97]. According to [Hwa86], the $\mathbb{P}$-a.s. convergence of $\mu_n \Rightarrow \mu_\infty$ is due to Silverstein in an unpublished note from 1984. Having the Universality phenomenon in mind, one should expect the Circular Law to hold for non-Gaussian matrices as well. Twenty years later, Girko [Gir85] presented a more general strategy for the proof of the Circular Law, however his proof was lacking arguments regarding the control of smallest singular values. It took until 1997 when Bai [Bai97] used Girko's Hermitization Trick (see (2.32) below) and the rate of convergence for the empirical singular value distribution in order to prove the Circular Law under the assumptions of an existing density and higher moments for the entries.

The density assumption was removed by Götze and Tikhomirov [GT07] and several weakenings of the moment conditions appeared in [GT10c; PZ10; TV08]. Significant progress was possible due to the control of the smallest singular value in [Rud08; TV09b; RV08; TV08]. Ultimately, the Circular Law was proven under optimal second moment assumption by Tao and Vu (with an appendix by Krishnapur) [TV10b] in the form stated in Theorem 1.2. The survey [BC12] presents a great overview of the Circular Law.

As the history of the Circular Law already indicates, non-Hermitian RMT has become a fast-growing field especially actively developing in the last years. Since then, the Circular Law has also been shown for random d-regular digraphs [Coo19], sparse matrices [RT19; GT10c] and matrices with exchangeable entries [ACW16]. Moreover, the model has been modified to obtain an Elliptic Law [Nau12], to products and matrices with inhomogeneous variances [AEK18; AEK19] or infinite variances [BCC11]. Dyson's desire for applications is variously fulfilled and include dynamics of (neural) networks [RA06; SCSS88], gaps between Buzzard nests [ABC20], scattering in chaotic quantum systems [FKS97; ABD11], Coulomb plasma [ABD11; For10] and see (2.22) below.

Products of Hermitian matrices lose their hermiticity and hence they naturally belong to non-Hermitian RMT. The pioneering works by Bellman [Bel54] and Furstenberg and Kesten [FK60] are devoted to the study of Lyapunov exponents of a product of random matrices. Above all, this means that they considered a large product of fixed size matrices, contrary to the investigation of a fixed number of large sized matrices. In this way, this community is separated from random matrix theoretical researchers by their interest in different questions.

Applications of products of random matrices include disordered and chaotic systems [CPV12], products of scattering matrices in wireless telecommunication [Mül02; TV04], the Dirac operator in quantum chromodynamics with chemical potential [Ake17] and stability of (ecological) dynamical systems [Cas06]. In order to motivate products of random matrices, let us look at the last application in more detail.

In 1972 in his celebrated paper [May72], Robert May posed the question "Will a large complex system be stable?". May considered the difference equation $x_m = Ax_{m-1}$ for some $n \times n$ 'interaction' matrix $A = B - I$ and the vector x_m of the number of species of different types after m time steps. The system is stable if all eigenvalues of A have non-positive real part. The Circular Law reveals that this should be the case iff the random entries of B have variance ≤ 1, which essentially is May's claim. Cohen and Newman [CN84] added some missing rigorous restrictions and generalized it to the time dependent case in which products of random matrices appear as

$$x_m = A(m)x_{m-1} = A(m) \cdots A(1)x_0.$$

Of course, stability questions are answered in terms of Lyapunov exponents and hence belong to the category of large products, but May's simple question also attracts researcher from classical Random Matrix Theory, e.g. [CM16].

However, the study of the case $n \to \infty$ and $m \in \mathbb{N}$ fixed is a much younger research area. Even the global limit for products of Ginibre matrices has only been found in 2010 (45 years after the case $m = 1$) by Burda, Janik and Waclaw [BJW10] and microscopic correlation functions have been computed by Akemann and Burda [AB12]. For random matrices with independent entries, Theorem 1.3 has been proven by Götze and Tikhomirov [GT10b], see also [OS11]. Since then, this result have been generalized to products of rectangular matrices [GKT15; IK14; AI15; BJL10], products of elliptic matrices [GNT15c; ORSV15], powers of Ginibre matrices [BNS12], products of two coupled matrices [AB10], sums of products [KT15; Bor11], quaternionic Ginibre matrices [Ips13] and products involving inverse matrices [ARRS16; Bor11]. For the convergence of the empirical singular value distribution of products of random matrices see [AGT10]. Of interest for our considerations are in particular local results [Nem17; KOV18; Nem18; GNT19a; CO19] that we shall discuss in comparison with our results later on.

In the following subsection, we will introduce relevant techniques that are used in previously mentioned articles and will also be used in this work. Chronologically we begin with Ginibre matrices.

2.2.1. Gaussian ensembles

The Ginibre ensemble. Recall that a Ginibre matrix X is a non-Hermitian random $n \times n$ matrix with independent standard complex Gaussian entries X_{ij}. Equivalently to Definition 1.1 and in accordance with (2.13), Ginibre [Gin65] defined the density of X as

$$d\,\mathbb{P}(X) = Z_n^{-1} \exp\left[-\mathrm{trace}(XX^*)\right] d\lambda^{n^2}(X), \tag{2.15}$$

where $\lambda\!\!\!\lambda^{n^2}$ is the Lebesgue measure on the space of complex matrices $\mathbb{C}^{n^2}$. In the same manner as for Hermitian matrices, by unitary invariance, Schur decomposition (instead of diagonalization) and transformation of the group integral, Ginibre [Gin65] showed that the joint probability density of the eigenvalues is given by

$$P(\lambda_1, \ldots, \lambda_n) = \frac{1}{\pi^n \prod_{j=1}^{n} j!} \prod_{1 \leq j < k \leq n} |\lambda_j - \lambda_k|^2 \exp\Big[- \sum_{j=1}^{n} |\lambda_j|^2 \Big]. \tag{2.16}$$

Apart from normalizing factor, the only change from (2.14) to (2.16) is the respective measure, which here is the 2-dimensional Lebesgue measure $\lambda\!\!\!\lambda$ on the underlying space $\mathbb{C}$. The factor 1/2 in the exponent (2.14) is due to the Hermiticity constraint. Following [Meh04], a determinantal structure can be unveiled, i.e. the correlation functions are given by

$$\varrho_k(\lambda_1, \ldots, \lambda_k) = \exp\Big(- \sum_{j=1}^{k} |\lambda_j|^2 \Big) \det\left[K_n(\lambda_i, \lambda_j)\right]_{1 \leq i,j \leq k}, \tag{2.17}$$

where the kernel, without the weight function $w(z) = e^{-|z|^2}$, is given by

$$K_n(\lambda_i, \lambda_j) = \sum_{k=0}^{n-1} \frac{(\lambda_i \bar{\lambda}_j)^k}{\pi k!}. \tag{2.18}$$

Note that the orthogonal polynomials on the complex plane with respect to the exponential weight function are the monomials.

Lemma 2.8. *The mean empirical spectral distribution $\bar{\mu}_n = \mathbb{E}\,\mu_n$ of the Ginibre ensemble has a (Lebesgue) density given by*

$$p_n(z) = \frac{1}{\pi} e^{-n|z|^2} \sum_{k=0}^{n-1} \frac{n^k\, |z|^{2k}}{k!}. \tag{2.19}$$

Proof. A randomly picked eigenvalue of a Ginibre matrix X has a density given by the 1-point correlation function (2.17) normalized by $1/n$. According to the definition (1.1) of the ESD, we apply a scaling of order $1/\sqrt{n}$ to the matrix and hence to the eigenvalues $\lambda_1(X) = \sqrt{n}\lambda_1(X/\sqrt{n}) = \sqrt{n}z$. The differential of this change of variables gives a factor of n, cancelling the factor of the normalization, i.e.

$$p_n(z) = \frac{1}{n} \varrho_1(\sqrt{n}z) \frac{d\lambda_1(X)}{dz} = \frac{1}{\pi} e^{-n|z|^2} K_1(z,z).$$

□

Elementary calculations show that this lemma implies a version of the Circular Law. In the proof of Lemma 3.1, we will perform these steps in order to show a first result on the rate of convergence to the Circular Law.

Remark 2.9. Ginibre [Gin65] also described *real and quaternionic Ginibre ensembles*, which are described by a Pfaffian structure. For real or quaternionic Gaussian entries, the eigenvalues will come in conjugate pairs and Figure 1.1 already showed a symmetry and repulsion at the real axis.

This becomes particularly visible in the joint probability density of the eigenvalues of the quaternionic Ginibre, which (compared to (2.16)) has extra factors of the form $\left|\bar{\lambda}_j - \lambda_j\right|^2$, see [AKMP19; Meh04]. The density of the mean ESD of quaternionic Ginibre matrices reads

$$d\bar{\mu}_n(z) = \frac{4}{\pi}e^{-2n|z|^2} \sum_{0\leq j\leq k\leq n-1} \frac{2^k k!}{2^j j!(2k+1)!}(2n)^{k+j+1/2} \left|z\right|^{4j} \operatorname{Im}(z)\operatorname{Im}(z^{2(k-j)+1})d\lambda\!\!\lambda(z), \tag{2.20}$$

where we also see how eigenvalues avoid the real line.

For real Ginibre matrices with entries $X_{ij} \sim \mathcal{N}_{\mathbb{R}}(0,1)$, the joint probability density is much more involved, since it contains interaction terms between real/real, complex/complex and real/complex eigenvalues. The one-point correlation functions $\varrho_1^{\mathbb{R}}$, $\varrho_1^{\mathbb{C}}$ consider real or complex eigenvalues separately, see [Som07; FN07]. Thus, the density of the mean ESD splits into two parts

$$\begin{aligned} d\bar{\mu}_n(z) &= \frac{1}{\sqrt{n}}\varrho_1^{\mathbb{R}}(\sqrt{n}x)d\lambda\!\!\lambda^{\mathbb{R}}(x) + \varrho_1^{\mathbb{C}}(\sqrt{n}z)d\lambda\!\!\lambda(z) \qquad (2.21)\\ &= \frac{\sqrt{n}}{2\sqrt{2\pi}}\Big(\int_{-\infty}^{\infty} |x-t|\, e^{-n(x^2+t^2)/2} \sum_{k=0}^{n-2} \frac{n^k (xt)^k}{k!} dt\Big) d\lambda\!\!\lambda^{\mathbb{R}}(x) \\ &+ \frac{2\sqrt{2n}}{\pi} e^{-n(x^2-y^2)} \int_{|y|\sqrt{2n}}^{\infty} e^{-t^2} dt\, |y| \sum_{k=0}^{n-2} \frac{n^k \left|z\right|^{2k}}{n!} d\lambda\!\!\lambda(z) \end{aligned}$$

where $z = x+iy$ and $d\lambda\!\!\lambda^{\mathbb{R}}$ is the Lebesgue measure on $\mathbb{R}$.[1] Once the correlation functions are given, the derivation of the density is similar to Lemma 2.8 and we skip it.

We should also point out that, similar to the log-gas picture of real eigenvalues and based on (2.16), the classical *Coulomb gas* (also called one-component plasma) in

1 Edelman, Kostlan and Shub [EKS94] derived the real density in order to show that the expected number of the real eigenvalues equals $\int \varrho_1^{\mathbb{R}}(x)dx \sim \sqrt{2n/\pi}$

dimension 2 is defined by

$$P(z_1,\dots,z_n) = Z_n^{-1} \exp\Big[-\frac{\beta}{2}\Big(\sum_{1\le j<k\le n} -\log|z_j-z_k|^2 + \sum_{j=1}^n V(z_j)\Big)\Big] \tag{2.22}$$

for some Dyson index (the inverse temperature) $\beta > 0$ and admissible confining potential $V : \mathbb{C} \to \mathbb{R}$, see [For10; Ser15]. The equilibrium measure, and hence the large n limiting distribution, is given by the Laplacian $\frac{\Delta V}{c}$ on its support, coinciding with the Circular Law in the case of $V(z) = |z|^2$.

Products of Ginibre matrices. We would like to study the eigenvalue distribution of the product of m independent Ginibre matrices $X^{(1)},\dots,X^{(m)}$, each of size $n \times n$. By performing a generalized Schur decomposition as change of variables, it can be shown that the Jacobian is again the Vandermonde determinant, see [ARRS16]. The product structure leads to a recurrence relation of the corresponding weight functions, which can be solved using the Mellin inversion formula for powers of Gamma functions. As it was proven in [AB12], the joint probability density function then reads

$$P(\lambda_1,\dots,\lambda_n) = \pi^{m-1} \prod_{1\le j<k\le n} |\lambda_j-\lambda_k|^2 \cdot \prod_{j=1}^n G^{m,0}_{0,m}\left(\begin{matrix}-\\0\end{matrix}\,\middle|\,|\lambda_j|^2\right), \tag{2.23}$$

where the definition of the Meijer-G function $G^{m,0}_{0,m}$ is given as follows. The *Meijer G-function* for $\xi \in \mathbb{C}\backslash\{0\}$ is defined as the Mellin inverse of products of Gamma functions

$$G^{m,n}_{p,q}\left(\begin{matrix}a_1,\dots,a_p\\b_1,\dots,b_q\end{matrix}\,\middle|\,\xi\right) = \frac{1}{2\pi i}\int_L \frac{\prod_{j=1}^m \Gamma(b_j-t)\prod_{j=1}^n \Gamma(1-a_j+t)}{\prod_{j=m+1}^q \Gamma(1-b_j+t)\prod_{j=n+1}^p \Gamma(a_j-t)}\xi^t dt, \tag{2.24}$$

where $0 \le m \le q$, $0 \le n \le p$, $a_k - b_j \notin \mathbb{N}$ for $k = 1,\dots,n$ and $j = 1,\dots,m$. The contour L goes from $-i\infty$ to $i\infty$, but can be chosen arbitrarily as long as the poles of $\Gamma(b_j - t)$ are on the right hand side of the path and the poles of $\Gamma(1 - a_j - t)$ are on the left hand side. Since the Gamma function is the Mellin transform of the exponential function, we have for instance $G^{1,0}_{0,1}\big(\begin{smallmatrix}-\\0\end{smallmatrix}\big|\xi^2\big) = \exp(-\xi^2)$ and hence it is indeed a generalization of the previous discussion for Ginibre matrices.

The appearance of the Meijer-G function is not surprising by the following reasons. First, the importance of the Mellin transform for products is equivalent to the role of the Fourier transform for sums, in particular due to its factorization for independent random variables. More specifically, the statement of (2.23) for the scalar case $n = 1$ goes back to a theorem by Springer and Thompson [ST70]: The product of m independent Gaussian random variables has a density that is proportional to $G^{m,0}_{0,m}\big(\begin{smallmatrix}-\\0\end{smallmatrix}\big|c\xi^2\big)$.

The resulting density of the eigenvalues of a product of Ginibre matrices (2.23) is also a determinantal point process with correlation functions similar to (2.17),

$$\varrho_k(\lambda_1,\ldots,\lambda_k) = \prod_{j=1}^{k}\left(\pi^{m-1}G^{m,0}_{0,m}\left(\begin{smallmatrix}-\\0\end{smallmatrix}\,\middle|\,|\lambda_j|^2\right)\right)\det\left[K_n(\lambda_i,\lambda_j)\right]_{1\le i,j\le k}, \tag{2.25}$$

where the kernel is given by

$$K_n(\lambda_i,\lambda_j) = \sum_{k=0}^{n-1}\frac{(\lambda_i\bar{\lambda}_j)^k}{(\pi k!)^m}. \tag{2.26}$$

Analogously to Lemma 2.8, we obtain the following lemma for the scaled product.

Lemma 2.10. *The mean empirical spectral distribution* $\bar{\mu}_n^m = \mathbb{E}\,\mu_n^m$ *of*

$$\mathbf{X} = \frac{1}{\sqrt{n^m}}\prod_{q=1}^{m}X^{(q)},$$

where $X^{(q)}$ *are independent Ginibre matrices, has a Lebesgue density given by*

$$\rho_n^m(z) = n^{m-1}\sum_{k=0}^{n-1}\frac{n^{mk}\,|z|^{2k}}{\pi(k!)^m}G^{m,0}_{0,m}\left(\begin{smallmatrix}-\\0\end{smallmatrix}\,\middle|\,n^m\,|z|^2\right).$$

We would like to point out that the Coulomb gas picture partially breaks down here, since $V = -\log G^{m,0}_{0,m}\left(\begin{smallmatrix}-\\0\end{smallmatrix}\,\middle|\,|\cdot|^2\right)$ is not admissible (according to [Ser17]). In particular, as V is unbounded from below in its singularity and hence $\Delta V \sim \mu_\infty^m$ fails for finite n.

2.2.2. The logarithmic potential

Similar to the role of the Stieltjes transform in the theory of Hermitian random matrices, the weak topology of measures μ on $\mathbb{C}$ can be expressed in terms of the so called logarithmic potential U. Before we turn to its importance for non-Hermitian RMT, let us collect the most important properties from

Potential Theory and Harmonic Analysis. The logarithmic potential of a measure μ is a solution of the distributional Poisson equation. More precisely for every finite Radon measure μ on $\mathbb{C}$ the *logarithmic potential* defined by

$$U_\mu(z) := -\int_{\mathbb{C}}\log|t-z|\,d\mu(t) = (-\log|\cdot| * \mu)(z) \quad \text{satisfies} \quad \Delta U_\mu = -2\pi\mu \tag{2.27}$$

in the sense of distributions. This follows directly from the definition of convolutions and derivatives in the sense of distributions and the fundamental solution to the Laplace equation $\Delta\frac{1}{2\pi}\log|\cdot| = \delta_0$. Most importantly, (2.27) provides *Unicity*: If the logarithmic potentials of two Radon measure coincide almost everywhere, then the measures are equal. Obviously the logarithmic potential of a measure is superharmonic in $\mathbb{C}$, harmonic outside the support of μ and the solution to (2.27) is unique up to addition of harmonic function. It follows from Fubini's theorem that U_μ is locally Lebesgue integrable on $\mathbb{C}$. We refer to [ST13] for a complete discussion of the (logarithmic) potential theory.

From the viewpoint of electrostatics, U_μ is the electrostatic potential of charged particles in $\mathbb{C}$, which repel each other with logarithmic (Coulomb) interaction. The system reaches equilibrium when the energy functional

$$\mathcal{E}(\mu) = \int U_\mu(z) d\mu(z)$$

is minimal. The existence and uniqueness of the minimalizing measure goes back to Frostman [Fro35]. Note the apparent link to the log-gas picture, in which ΔV is the density of a measure μ_V that minimizes a weighted energy functional $\mathcal{E}_V = \mathcal{E}(\mu) + \int V d\mu$ with some external field given by $V = -U_{\mu_V}$. In this setting we may put $V(z) = |z|^2$ again, and restrict ourselves to measures of fixed variance $\int |z|^2 \, d\mu(z) = \frac{1}{2}$. It follows that the Circular Law μ_∞ is the constrained minimizer of $\mathcal{E}(\mu)$ among all measures with fixed variance, see [ST13]. Let us explicitly calculate its logarithmic potential.

Lemma 2.11. *The logarithmic potential U_∞ of the Circular Law μ_∞ is given by*

$$U_\infty(z) = \begin{cases} -\log|z| & , \text{ if } |z| > 1, \\ \frac{1}{2}(1 - |z|^2) & , \text{ if } |z| \leq 1. \end{cases} \tag{2.28}$$

This is also a neat example of yet another characterization of the logarithmic potential that is studied in the calculus of variations, the so called *obstacle problem*: The logarithmic potential of the equilibrium measure μ_V is the smallest superharmonic function bounded from below by $c - V/2$ and harmonic on $\{U_\mu \neq c - V/2\}$, see [KS80] for more information.

Proof. We will use the mean-value property of harmonic functions, see for instance [Eva98]. Since $\log|\cdot|$ is harmonic in $\mathbb{C} \setminus \{0\}$, we obtain

$$⨍_{B_r(0)} \log|t - z| \, dt = \log|z| = ⨍_{\partial B_r(0)} \log|t - z| \, dt \tag{2.29}$$

for $|z| > r$. We denoted by $⨍$ the normalized integral, i.e. the mean value. Setting $r = 1$, we directly get $U_\infty(z) = -\log|z|$ for $|z| > 1$. On the other hand for $|z| < r$ the mean value property also yields

$$\frac{1}{2\pi} \int_0^{2\pi} \log\left|re^{i\theta} - z\right| d\theta = \frac{1}{2\pi} \int_0^{2\pi} \log\left|r - \bar{z}e^{i\theta}\right| d\theta = \log r.$$

Let now $|z| \leq 1$. We apply both previous equations to obtain

$$
\begin{aligned}
U_\infty(z) &= \frac{-1}{\pi} \int_0^1 \int_0^{2\pi} \log \left| re^{i\theta} - z \right| d\theta \; rdr \\
&= \int_{|z|}^1 -2r \log r dr + \int_0^{|z|} -2r \log |z| \, dr = \frac{1}{2}(1 - |z|^2).
\end{aligned}
$$

□

The values of U_∞ are strikingly similar to the Stieltjes transform computed in (2.12). In order to clarify this connection, let us introduce the *Wirtinger derivatives*

$$
\partial = \frac{1}{2}(\partial_x - i\partial_y) \text{ and } \bar{\partial} = \frac{1}{2}(\partial_x + i\partial_y), \tag{2.30}
$$

cf. [FL12]. In complex analysis the Wirtinger derivatives are frequently used as the partial derivatives in the direction of the independent variables z and $\bar{z}$. In particular, we see that $2\partial \log |z| = 1/z$, hence $2\partial U_\infty(z) = m_\infty(z)$. This is not an instance of the Circular Law, but a very general relation between the Stieltjes transform and the logarithmic potential. Since the Laplacian is given by $\Delta = 4\partial\bar{\partial} = 4\bar{\partial}\partial$, we have for any probability measure μ on $\mathbb{C}$

$$
-2\pi\mu = 2\bar{\partial} m_\mu = \Delta U_\mu \text{ and } m_\mu = 2\partial U_\mu \tag{2.31}
$$

in the sense of distributions.

Now that we have summarized the theory of logarithmic potentials, let us turn to its usefulness in non-Hermitian RMT.

Girko's Hermitization Trick. The advantage of the logarithmic potentials $U_n = U_{\mu_n}$ of μ_n in non-Hermitian RMT is the following identity known as *Girko's Hermitization trick*

$$
\begin{aligned}
U_n(z) &= -\frac{1}{n} \sum_{j=1}^n \log |\lambda_j - z| = -\frac{1}{n} \log \left| \det \left(\frac{1}{\sqrt{n}} X - z \right) \right| \\
&= -\frac{1}{n} \log \det \sqrt{\left(\frac{1}{\sqrt{n}} X - z \right) \left(\frac{1}{\sqrt{n}} X - z \right)^*} = -\int_0^\infty \log(x) d\nu_n^z(x),
\end{aligned} \tag{2.32}
$$

where ν_n^z is the empirical singular value distribution of the shifted matrix $X/\sqrt{n} - z$.[1] Due to this fact, the entire information on the complex spectrum of $X/\sqrt{n}$ is stored in

1 This identity gives rise to the definition of the Brown measure of non-self-adjoint elements in free probability, see [HL99] and [MS17, Equation (11.13)].

the real and positive spectra of $(X/\sqrt{n}-z)(X/\sqrt{n}-z)^*$ for *all* shifts z. Note that the symmetrized version $\widetilde{\nu}_n^z$ of ν_n^z is the empirical eigenvalue distribution of the Hermitian matrix

$$V(z)=\begin{bmatrix} 0 & (X/\sqrt{n}-z) \\ (X/\sqrt{n}-z)^* & 0 \end{bmatrix}. \tag{2.33}$$

Therefore, Girko's Hermitization Trick builds a very useful bridge between Hermitian and non-Hermitian Random Matrix Theory.

We emphasize that the same notation as in Section 2.1 has been used on purpose. Particularly, we already discussed the case $z=0$, corresponding to the the matrix $V=V(0)$ in (2.2), where $\nu_n=\nu_n^0$ converges to the Quarter Circular Law (2.1) and $\widetilde{\nu}_n=\widetilde{\nu}_n^0$ converges to the Semicircle Law $\widetilde{\nu}_\infty$. For non-zero $z\in\mathbb{C}$, the limiting distribution exists as well, but differ as follows.[1]

Let $\widetilde{\nu}_n^z$ be the symmetrized empirical singular value distribution of the shifted matrices $X/\sqrt{n}-z$ defined above and

$$m_n(z,\cdot):\mathbb{C}\setminus\mathbb{R}\to\mathbb{C}, w\mapsto\int_{\mathbb{R}}\frac{1}{w-t}d\widetilde{\nu}_n^z(t) \tag{2.34}$$

be its Stieltjes transform. Under the conditions of the Circular Law it is known that $m_n(z,\cdot)$ converges a.s. to the solution of

$$s(z,w)=-\frac{s(z,w)+w}{(w+s(z,w))^2-|z|^2}, \tag{2.35}$$

see for instance [GT10c]. Note again the analogy to the case $z=0$ and the generalization of (2.6). Furthermore $s(z,\cdot)$ corresponds to a limiting measure $\widetilde{\nu}^z$ which has a symmetric bounded density ρ^z (the bound holds uniformly in z) and has compact support

$$\mathbb{J}^z:=\begin{cases}[-\lambda_+,-\lambda_-]\cup[\lambda_-,\lambda_+], & \text{if } |z|>1\\ [-\lambda_+,\lambda_+], & \text{if } |z|\le 1\end{cases}, \tag{2.36}$$

where the endpoints are given by

$$\lambda_\pm^2:=\frac{(\alpha\pm 3)^3}{8(\alpha\pm 1)}\wedge 0,\quad \alpha:=\sqrt{1+8|z|^2}.$$

1 From the viewpoint of free probability we may interpret it as the free convolution $\widetilde{\nu}_\infty^z=\widetilde{\nu}_\infty\boxplus(\frac{1}{2}\delta_{|z|}+\frac{1}{2}\delta_{-|z|})$, see [GKT15, Theorem 6.1]. This corresponds to (2.33) being the sum of two asymptotically free matrices, if X is Ginibre.

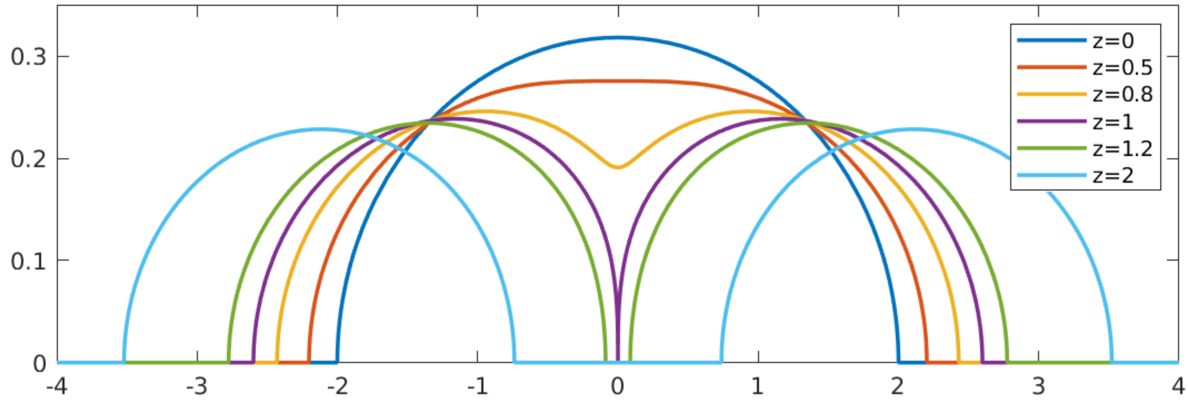

Figure 2.1: The densities of $\widetilde{\nu}^z_\infty$ for different values of z including the Semicircle Law $\widetilde{\nu}^0_\infty$. As the spectral parameter gets closer to the edge $|z| \approx 1$, the support separates which causes several difficulties in the analysis.

As z approaches the spectral edge, the following problem is encountered. Since $\lambda_- \sim (1-|z|)^{3/2}$ as $|z| \to 1$, a new gap in the support emerges at 0. This leads to the unboundedness of s and equation (2.35) becomes unstable. This is the reason for the bulk constraint in many results, for instance also in the following Proposition 2.18 and our main result Theorem 4.9. Very recently, in [AEK19], this restriction has been removed by separating stable from instable directions and the usage of a cusp fluctuation averaging method.

Basically, one may analyze the Stieltjes transform $m_n(z,w)$ uniformly in the shift parameter z and follow the same route as the Semicircle Law has been obtained from the self recurrence relation (2.8). Proposition 2.5 implies weak convergence $\nu^z_n \Rightarrow \nu^z_\infty$, i.e. for all bounded and continuous test functions φ we have $\int \varphi d\nu^z_n \to \varphi d\nu^z_\infty$. The nuisance is that this does not yield convergence of the logarithmic potentials, since the logarithm is unbounded. Therefore, we need an additional improved tightness constraint. If in addition the logarithm is *uniformly integrable* with respect to the family $\{\nu^z_n\}_n$, then there exists a weak limit μ of the ESD's μ_n such that $U_\mu(z) = \int \log(t) d\nu^z_\infty(t)$ for $\lambda\!\!\lambda$ almost all z, see [BC12; TV10b]. In order to complete this subsection with the claim we began with, let us emphasize the following simplification of the preceding discussion. We state it in the same spirit as Proposition 2.5 was given for pointwise convergence of Stieltjes transforms.

Lemma 2.12. *Let μ_n, $n \in \mathbb{N}$, and μ be probability measures on $\mathbb{C}$. If $U_{\mu_n}(z) \to U_\mu(z)$ for almost all z and if U_n is locally uniformly Lebesgue integrable, then $\mu_n \Rightarrow \mu$.*

Proof. From the locally uniformly integrability, it follows that $U_n \to U$ in the sense of distributions, i.e. $\int \varphi U_n d\lambda\!\!\lambda \to \int \varphi U d\lambda\!\!\lambda$ for all $\varphi \in \mathcal{C}^\infty_c$. By Fatou's lemma, U is also locally Lebesgue integrable. Moreover, by continuity of Δ on the space of distributions, we have $\mu_n = -\frac{1}{2\pi}\Delta U_n \to \frac{1}{2\pi}\Delta U \geq 0$. Now since μ is already assumed to be a probability measure, this convergence also holds weakly. □

As already mentioned, we will only need this simplified version. More generally, if U_{μ_n} converges to some function U and the logarithm function is uniformly integrable with respect to μ_n, then weak convergence $\mu_n \Rightarrow \mu = \frac{1}{2\pi}\Delta U$ holds, see [BC12, Remark 4.8].

Although Girko's ingenious idea of Hermitization can be already found in his original paper [Gir85], his proof of the Circular Law lacked the argument of uniformly integrability of the logarithm with respect to $\widetilde{\nu}_n^z$, as can be seen in his formula (31). Not the integrability of large singular values is a problem, but the singularity of the logarithm at 0 is problematic: One has to control the small singular values of the matrix $V(z)$. Though Girko revisits the Circular Law every decade, see [Gir94; Gir04a; Gir04b; Gir05; Gir12], it should have taken 26 years to close the gap in his proof.

Ultimately it was the idea of Rudelson and Vershynin's [Rud08; RV08], further developed by Tao and Vu [TV08; TV09b; TV09a], that allowed to gain optimal control over the smallest singular value of random matrices by applying geometric arguments and the (inverse) Littlewood-Offord problem. As we already presented at the beginning of this section, this led to the final proof of the Circular Law. Still, Girko was right when he wrote

> "The crucial step in the proof of the C-Law is my VICTORIA transform."[1]
> – Vyacheslav L. Girko [Gir12]

2.2.3. State of the art

In order to approach our goal of obtaining rates of convergences, we essentially need two ingredients: A rate of convergence of the logarithmic potentials and direct relations between measures and their logarithmic potentials. In this section, we will recall what is known so far in the theory of non-Hermitian matrices with independent entries and transfer certain statements into our setting.

Concentration of logarithmic potentials. Under certain conditions on the matrix entries, the logarithmic potential U_n concentrates around that of the Circular Law U_∞. Let us fix some notation and these conditions.

Definition 2.13. Two non-Hermitian random matrices X,$\widetilde{X}$ are said to *match moments up to the d-th order* for some $d \in \mathbb{N}$ if

$$\mathbb{E}\,\mathrm{Re}(X_{ij})^p\mathrm{Im}(X_{ij})^q = \mathbb{E}\,\mathrm{Re}(\widetilde{X}_{ij})^p\mathrm{Im}(\widetilde{X}_{ij})^q$$

for all $1 \le i,j \le n$ and $p,q \in \mathbb{N}_0$ with $p + q \le d$.

1 Victoria stands for *Very Important Computational Transformation Of Randomly Independent Arrays* and is equal to the Stieltjes transform $m_n(z,w)$ in our setting.

Starting with the *Four Moment Theorem* by Tao and Vu [TV11], many Universality results in RMT are based on moment matching conditions for the entries of the underlying matrices. If the first few (mostly 4) moments of two matrix ensembles coincide, then local spectral statistics remain unaffected. In this way Universality results can be deduced from properties of Gaussian matrices. Hence, the approach is to firstly prove a statement for Gaussian matrices for which the analysis is much easier due to the special determinantal structure we discussed in the previous sections. Afterwards one compares general matrices with independent entries to the Gaussian case via a Lindeberg replacement of the entries.

For instance, this method of proof has been applied in [TV15; TV14; EYY12a; KOV18; TV10a] just to name a few. We will use this approach in Chapter 5.2 for a rate of convergence result of the spectral radius. In particular, the first result about Universality of correlation functions of non-Hermitian random matrices was proven by Tao and Vu using this technique. The Four Moment Theorem [TV15] for non-Hermitian random matrices states that if the first four moments match the Ginibre ensemble, then the weak limit of the correlation functions is identical to the corresponding limit of Ginibre matrices, which can be read off from (2.18).

Throughout this thesis, we frequently use the following conditions for our random matrices.

Definition 2.14. (A) We say X satisfies *condition (A)* if it has independent entries X_{ij} with mean zero, variance $\mathbb{E}\,|X_{ij}|^2 = 1$, subexponential tails

$$\mathbb{P}(|X_{ij}| \geq t) \leq C\exp(-t^c) \tag{2.37}$$

for some fixed $c,C > 0$ and matches either the real or complex Gaussian moments up to third order, i.e.

$$\mathbb{E}\,X_{ij} = \mathbb{E}\,\mathrm{Re}(X_{ij})^3 = \mathbb{E}\,\mathrm{Im}(X_{ij})^3 = 0$$

and either $\mathbb{E}\,|\mathrm{Re}X_{ij}|^2 = \mathbb{E}\,|\mathrm{Im}X_{ij}|^2 = 1/2$ or $\mathbb{E}\,|\mathrm{Re}X_{ij}|^2 = 1, \mathbb{E}\,|\mathrm{Im}X_{ij}|^2 = 0$.

(B) We say X satisfies *condition (B)* if it has independent entries X_{ij} with mean zero, variance $\mathbb{E}\,|X_{ij}|^2 = 1$ and if for all $p \in \mathbb{N}$ it holds $\max_{i,j}\mathbb{E}\,|X_{i,j}|^p < \infty$.

(C) We say X satisfies *condition (C)* if it has independent entries, with

$$\max_{i,j}|\mathbb{E}\,X_{ij}| \leq n^{-1-\varepsilon} \text{ and } \max_{i,j}\left|1-\mathbb{E}\,|X_{ij}|^2\right| \leq n^{-1-\varepsilon}$$

for some $\varepsilon > 0$ and furthermore

$$\max_{i,j,n}\mathbb{E}\,|X_{ij}|^{4+\delta} < \infty$$

for some $\delta > 0$.

Note that in contrast to Wigner matrices, the distributions of the entries may be different and clearly, condition (A) implies condition (B), which in turn implies condition (C).

The following concentration of the logarithmic potentials was proven in [TV15, Theorem 25] and is one of the main ingredients for the proof of the Four Moment Theorem for non-Hermitian random matrices.

Proposition 2.15 ([TV15])**.** *If X satisfies (A), then for every $\varepsilon,\tau,Q > 0$ there exists a constant $c > 0$ such that*

$$\mathbb{P}\left(|U_n(z) - U_\infty(z)| \leq cn^{-1+\varepsilon}\right) \geq 1 - n^{-Q} \tag{2.38}$$

holds for any $z \in B_{1+\tau}(0)$.

In other words, (2.38) states that $|U_n(z) - U_\infty(z)| \lesssim n^{-1+\varepsilon}$ with overwhelming probability uniformly in z. However such shortened formulations should be treated with caution.

Remark 2.16. The statement of Proposition 2.15 should not be confused with an assertion that holds uniformly in z, since the uniform term $\sup_{z\in B_{1+\tau}(0)} |U_n(z) - U_\infty(z)|$ is unbounded whenever an eigenvalue lies in $B_{1+\tau}(0)$. Due to the same reasons, it is important to carefully distinguish between events holding with overwhelming probability uniformly in z and uniform events that hold w.o.p.. The former leads to Local Circular Laws, see Theorem 2.19 below, and hence do not imply the latter, which is an estimate on $D(\mu_n,\mu_\infty)$.

First of all, we clarify that only three matching moments are required for this theorem, instead of four, because the above error has a divergent quantity n^ε, whereas for the convergence of correlation functions a negative power of n is necessary. Up to certain extent the agreement of the logarithmic and Stieltjes transforms gets better the more moments are matching. We will make this precise in Lemma 5.8.

The bound is nearly optimal, apart from the additional power ε. For the Ginibre ensemble, Rider and Virag [RV07] showed that $nU_n - n\,\mathbb{E}\,U_n$ converges weakly in distribution to the planar Gaussian free field. It follows that for any test function φ, the standard deviation of $\int \varphi(U_n - U_\infty)d\lambda\!\!\lambda$ is of order n^{-1} in the limit. Pointwise, one may look at the special case $z = 0$ and see that an additional logarithmic factor appears. Nguyen and Vu [NV14] showed that, for matrices with indepenedent subexponential entries, the log determinant $nU_n(0)/\sqrt{1/2\log n}$ after centering converges to a standard Gaussian distribution, i.e. pointwise U_n fluctuates at order $\sqrt{\log n}/n$.

A very recent update[1] of [AEK19] allows to extract the following concentration of logarithmic potentials from their proof, which weakens the assumptions of the previous Proposition.

1 Published on arXiv on February 10th, 2020

Proposition 2.17 ([AEK19]). *If X obeys (B), then for every $\varepsilon,\tau,Q > 0$ there exists a constant $c > 0$ such that*

$$\mathbb{P}\left(|U_n(z) - U_\infty(z)| \leq cn^{-1+\varepsilon}\right) \geq 1 - n^{-Q} \tag{2.39}$$

holds for any $z \in B_{1+\tau}(0)$.

Note that the result in [AEK19] holds in a more general setting of inhomogeneous variances under some additional assumptions. Therefore, it is possible to show a rate of convergence in Kolmogorov distance D for the inhomogeneous Circular Law as well. However, we stick to normalized variances in this work in order to avoid exhaustive notation and technicalities.

In many different situations Götze, Naumov and Tikhomirov weakened conditions of various Local Laws to the existence of a small number of moments, see for instance [GNT19b; GNT19a; GT16; GNT15a; GNT15b]. In [GNT19a], the assumptions of Proposition 2.15 have been weakened, the rate has been improved and the result has been generalized to products of independent matrices, but at the cost of restricting the region to the bulk $||z| - 1| \geq \tau$.

Proposition 2.18 ([GNT19a]). *If X obeys (C), then for every $\tau,Q > 0$ there exists a constant $c > 0$ such that*

$$\mathbb{P}\left(|U_n(z) - U_\infty(z)| \leq c\frac{\log^4 n}{n}\right) \geq 1 - n^{-Q} \tag{2.40}$$

holds for any $\{z \in B_{1+\tau^{-1}}(0) : |1 - |z|| \geq \tau\}$.

Since both results Proposition 2.17 and Proposition 2.18 are not explicitly worked out in [GNT19a], or [AEK19] respectively, we will derive them in the end of this section, based on the results proved in these papers. Generally speaking, in order to prove the above statements one needs to show a Local Law for $\widetilde{\nu}_n^z$ like (2.9).

Local Circular Laws. Results on the concentration of the logarithmic potentials are used to derive *Local Circular Laws.* In the same spirit as Local Semicircle Laws, these type of results assert that the Circular Law holds down to nearly microscopic scale, i.e. considering much less than $\mathcal{O}(n)$ eigenvalues. For any center $z_0 \in \mathbb{C}$, testfunction $f : \mathbb{C} \to \mathbb{R}$, $f \in \mathcal{C}_c^\infty$, and scale $s \in [0,1/2]$, define the function $f_{z_0}(z) := n^{2s} f((z - z_0)n^s)$ which zooms into z_0 at speed n^s. Heuristically, the support of f_{z_0} should only cover $n^{2s} \in \mathcal{O}(n)$ many eigenvalues. For comparison, we state an earlier result by Bourgade, Yau and Yin.

Theorem 2.19 ([BYY14a; BYY14b; Yin14]). *Let X be a random matrix with independent entries having subexponential decay, i.e.* (2.37) *holds. For any $Q,\varepsilon > 0$ there exists* *a constant $c > 0$ such that*

$$\mathbb{P}\left(\left|\int f_{z_0} d\mu_n - \int_{\mathbb{C}} f_{z_0} d\mu_\infty\right| \leq cn^{-1+2s+\varepsilon} \left\|\Delta f\right\|_{L^1(\mathbb{C})}\right) \geq 1 - n^{-Q}$$

holds for any function $f \in \mathcal{C}_c^\infty$.

The Local Circular Law in [TV15] states this theorem for indicator functions: If Condition 2.14 (A) holds, then for all fixed centers $z_0 \in B_C(0)$ and radii $R \geq 1/\sqrt{n}$, we have

$$\mathbb{P}\left(|\mu_n(B_R(z_0)) - \mu_\infty(B_R(z_0))| \leq cRn^{-1/2+\varepsilon}\right) \geq 1 - n^{-Q}. \tag{2.41}$$

The core of the proof of the Local Law for non-Hermitian matrices is the following identity

$$\int f_{z_0} d(\mu_n - \mu_\infty) = -\frac{n^{2s}}{2\pi} \int \Delta f \left(U_n - U_\infty\right) d\lambda\!\!\lambda, \tag{2.42}$$

which follows directly from integration by parts or, in other words, the distributional Poisson equation (2.27).

Theorem 2.19 already gives a hint on a rate of convergence: For fixed smooth functions with a fixed center z_0 the rate of convergence is roughly n^{-1}. Lambert [Lam19] proved a rate of convergence of order $\log^{1+\varepsilon} n/n$ for Ginibre matrices which holds *uniformly* over smooth functions with controlled derivative and with vanishing Laplacian outside $B_{1-\tau}(0)$, see also [CHM18; MM15] for classes of Lipschitz functions. In Chapter 4, we will present a versatile approach to prove that uniformly over any non-smooth indicator function $f = \mathbb{1}_{B_R(z_0)}$ the rate is roughly $n^{-1/2}$. Contrary to Local Circular Laws, a bound on D with overwhelming probability allows to choose functions f depending on the random sample of eigenvalues $(\lambda_j(X(\omega)/\sqrt{n}))_j$, see Remark 2.16 and also Figure 1.2.

Some identities between μ_n, m_n and U_n. One may see formula (2.42) as a direct relation between the complex distributions and the corresponding logarithmic potentials. Another suggestion how to restrict to a local region was provided by Tao and Vu [TV15, Formula (2.2)] in order to prove their version of the Local Law. Jensen's formula [Kra12, §9.1.2.] applied to the function $z \mapsto \det(X/\sqrt{n} - z)$ yields

$$U_n(z) = \frac{1}{n}\sum_{j=1}^{n}\left(\log\frac{r}{|\lambda_j(X/\sqrt{n}) - z|}\right)_+ + \frac{1}{2\pi}\int_0^{2\pi} U_n(z + re^{i\theta})d\theta.$$

Formal differentiation with respect to r leads to the following formula that seems not being used in RMT so far and will not be applied in this thesis. Anyhow, it provides a new insight into the measure of balls and its link to the local behavior of the logarithmic potential. Also, it may be seen as an analogue of the Stieltjes inversion formula (2.5).

Lemma 2.20. *It holds*

$$\mu_n(B_r(z_0)) = -\frac{r}{2\pi}\partial_r \int_{-\pi}^{\pi} U_n(z_0 + re^{i\theta})d\theta.$$

Proof. Let $r > 0$, since the case $r = 0$ is trivial. For each eigenvalue λ_j we have

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|\lambda_j - z_0 - re^{i\theta}\right| d\theta = \begin{cases} \log r & \text{, if } \lambda_j \in B_r(z_0), \\ \log|\lambda_j - z_0| & \text{, if } \lambda_j \notin B_r(z_0), \end{cases}$$

as we computed in the proof of Lemma 2.11 for $z = \lambda_j - z_0$. We rewrite this as $\log(|\lambda_j - z_0| \vee r)$, where $\vee$ denotes the maximum of both numbers. For any $a \in \mathbb{R}$ it holds $\partial_r(a \vee r) = \mathbb{1}_{(a,\infty)}(r)$ in the sense of distributions, i.e. for any $\varphi \in \mathcal{C}_c^\infty((0,\infty))$ we have

$$\int \varphi(r)\partial_r(a \vee r)dr = -\int_a^\infty \varphi(r)dr$$

where integration by parts was used. The chain rule implies

$$\frac{1}{2\pi}\partial_r \int_{-\pi}^{\pi} \log\left|\lambda_j - z_0 - re^{i\theta}\right| d\theta = \frac{1}{r}\mathbb{1}_{B_r(z_0)}(\lambda_j).$$

The claim follows after summing over all $j = 1, \ldots, n$. □

Note that we did not use the fact that μ_n is a discrete measure and the method of proof carries over to any complex measure μ.

Another quantitative relation between the deviation of logarithmic potentials and the Kolmogorov-like distance of the corresponding measures will be presented in Theorem 4.2. This new Smoothing Inequality will imply many of our rate of convergence results.

We will now describe a direct connection between the Stieltjes transform of $\widetilde{\nu}_n^z$ and the logarithmic potential $U_n(z)$. We already know an indirect connection; if $m_n(z,\cdot)$ converges to $s(z,\cdot)$, then $\widetilde{\nu}_n^z \Rightarrow \widetilde{\nu}_\infty^z$ by Proposition 2.5 and under additional integrability conditions U_n converges to U_∞, see Lemma 2.12. In order to prove the results from the previous subsection, more precise and direct relations are necessary.

The following identity (2.43) goes back to Tao and Vu [TV15] and is nowadays used by others as well, e.g. [AEK18; AEK19; CES19a].

Lemma 2.21. *For any $T > 0$ it holds*

$$U_n(z) = \int_0^T \operatorname{Im}(m_n(z,i\eta))d\eta - \frac{1}{2n}\log|\det(V(z) - iT)| \tag{2.43}$$

and in the limit $T \to \infty$ we also have

$$U_\infty(z) = \int_0^\infty \operatorname{Im}(s(z,i\eta)) - \frac{1}{1+\eta} d\eta. \tag{2.44}$$

Proof. The distributional equation for Stieltjes transforms (2.31) in the case of $\mu = \widetilde{\nu}_n^z$ yields $\operatorname{Im}(m_n(z,i\eta)) = -\partial_\eta U_{\widetilde{\nu}_n^z}(i\eta)$. We integrate with respect to η and readily obtain

$$\int_0^T \operatorname{Im}(m_n(z,i\eta)) d\eta = -U_{\widetilde{\nu}_n^z}(iT) + U_{\widetilde{\nu}_n^z}(0).$$

The first term is as in the claim and the second term follows from rephrasing Girko's Hermitization Trick (2.32) as

$$U_n(z) = U_{\mu_n}(z) = U_{\widetilde{\nu}_n^z}(0) = U_{\nu_n^z}(0).$$

The same arguments yield

$$\begin{aligned} U_\infty(z) &= \int_0^T \operatorname{Im}(s(z,i\eta)) d\eta - \int_{-\lambda_+}^{\lambda_+} \log|x - iT| \, d\widetilde{\nu}_\infty^z(x) \\ &= \int_0^T \operatorname{Im}(s(z,i\eta)) d\eta - \int_0^T \frac{1}{1+\eta} d\eta + \mathcal{O}(T^{-1}), \end{aligned}$$

since the second integral is asymptotically equivalent to $\log T$ as $T \to \infty$. For large η we see that

$$\operatorname{Im}(s(z,i\eta)) - \frac{1}{1+\eta} = \int \frac{1 - x^2/\eta}{(x^2/\eta + \eta)(1+\eta)} d\widetilde{\nu}_\infty^z(x) \sim \frac{1}{\eta^2}, \tag{2.45}$$

hence it is integrable and we may pass to the limit $T \to \infty$ to obtain the claim. □

A more involved, but plausible method of relating U_n to m_n is applying a Smoothing Inequality to $\widetilde{\nu}_n^z$, similar to Theorem 2.6. This is the method that is used by Götze, Naumov and Tikhomirov, e.g. in order to proof Proposition 2.18. We will explain these ideas in more detail below, when we sketch this proof. Most importantly, it will be sufficient to control the Stieltjes transforms in the interior of the support of the corresponding measures - in contrast to Lemma 2.21 where always the imaginary axis is considered. Before we turn to this method, let us provide some details on how the proof of Proposition 2.17 follows from the previous lemma.

Proof of Proposition 2.17. We will show all steps until a result from [AEK19] can be directly applied. Using Lemma 2.21, we need to estimate

$$
\begin{aligned}
U_\infty(z) - U_n(z) &= \int_0^T \operatorname{Im}(s(z,i\eta) - m_n(z,i\eta))d\eta \\
&\quad + \int_T^\infty \operatorname{Im}(s(z,i\eta)) - \frac{1}{1+\eta}d\eta \\
&\quad + \frac{1}{2n}\log|\det(V(z) - iT)| - \int_0^T \frac{1}{1+\eta}d\eta, \qquad (2.46)
\end{aligned}
$$

which corresponds to a pointwise estimate of the integral in [AEK19, Equation (6.1)]. By (2.45), the second term is of order $\mathcal{O}(T^{-1})$. Regarding the third term it holds

$$
\begin{aligned}
\frac{1}{2n}\log|\det(V(z) - iT)| &= \log T + \log\left|\det(i - T^{-1}V(z))\right| \\
&= \log T + \sum_{j=1}^n \log\left(1 + \frac{s_j(z)^2}{T^2}\right), \qquad (2.47)
\end{aligned}
$$

where $s_j(z)$ are the non-negative eigenvalues of $V(z)$, or equivalently the singular values of $X/\sqrt{n} - z$. For any $Q > 0$ it holds

$$
\mathbb{P}(s_{\max}(z) \geq n^{(Q+1)/2}) \leq \frac{\mathbb{E}\left\|(X/\sqrt{n} - z)\right\|^2}{n^{Q+1}} \leq \frac{1}{n^{Q+2}}\sum_{i,j=1}^n \mathbb{E}\left|X_{ij}\right|^2 + \frac{|z|^2}{n^Q} \lesssim n^{-Q}, \quad (2.48)
$$

where the operator norm $\|\cdot\|$ has been estimated by the Hilbert Schmidt norm. Thus, the last term in (2.47) is neglegible if we choose $T = n^C$ for C large enough. The remaining $\log T$ cancels the last term of (2.46). Furthermore, [TV08, Theorem 2.1] states that for any $Q > 0$ there exists a constant $B > 0$ such that $\mathbb{P}(s_{\min}(z) < n^{-B}) \lesssim n^{-Q}$. Consequently, Equation (2.46) becomes

$$
U_\infty(z) - U_n(z) = \int_{n^{-B}}^{n^C} \operatorname{Im}(s(z,i\eta) - m_n(z,i\eta))d\eta + \mathcal{O}(n^{-1})
$$

with overwhelming probability. At this stage, according to [AEK19, Remark 6.2], it holds [AEK19, Lemma 6.1] stating

$$
\mathbb{E}\left|\int_{n^{-B}}^{n^C} \operatorname{Im}(s(z,i\eta) - m_n(z,i\eta))\right|^p \lesssim \frac{n^{\delta p}}{n^p}
$$

for any $\delta > 0$, $p \in \mathbb{N}$. Choosing $\delta = \varepsilon/2$ and p sufficiently large, an application of Markov's inequality finishes the proof. We shall point out that the conditions [AEK19, (6.4) Remark 2.5] are satisfied in our case, while [AEK19, (A1), (A2)] coincide with condition (B) in our claim. $\square$

Products of independent matrices. An issue in studying a product

$$\mathbf{X} = \frac{1}{\sqrt{n^m}} \prod_{q=1}^{m} X^{(q)},$$

is the non-linearity in the entries, when it comes to applying Schur's inversion formula, cf. (2.8). The problem can be 'linearized', using the method due to [GJJN03], which consists of replacing the $n \times n$ matrix by a bigger $nm \times nm$ matrix that is linear and has a comparable spectrum. We define the *linearization matrix* as the block matrix

$$\mathbf{W} := \frac{1}{\sqrt{n}} \begin{pmatrix} 0 & X^{(1)} & 0 & \cdots & 0 \\ \vdots & 0 & X^{(2)} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & X^{(m-1)} \\ X^{(m)} & 0 & \cdots & \cdots & 0 \end{pmatrix} \tag{2.49}$$

and note that $\mathbf{W}^m$ is a block diagonal matrix with cyclic products of $X^{(1)}, \ldots, X^{(m)}$. Consequently, its spectrum consists of the eigenvalues $\lambda_j(\mathbf{X})$ of $\mathbf{X}$ with multiplicity m. Furthermore define its shifted Hermitization, similar to (2.33), by

$$\mathbf{V}(z) := \begin{pmatrix} 0 & \mathbf{W} - z \\ (\mathbf{W} - z)^* & 0 \end{pmatrix} \tag{2.50}$$

for $z \in \mathbb{C}$. The eigenvalues of $\mathbf{V}(z)$ are again given by $\pm s_j(\mathbf{W} - z)$, where $s_{\max} = s_1 \geq \ldots \geq s_{mn} = s_{\min}$ are the singular values of $\mathbf{W} - z$. We will also write $\widetilde{\nu}_n^z$ for ESD of $\mathbf{V}(z)$ and $U_n^{(m)} = U_n$ for the logarithmic potential of the ESD of $\mathbf{W}$, skipping the dependency on m for readability. It should be clear from the corresponding chapter whether $m = 1$ or $m \geq 1$.

The idea of the linearization matrix $\mathbf{W}$ is also used in [GNT19a; AB12; BJW10; BJL10; Nem17; Nem18; KOV18] among others.

Definition 2.22. We say $\mathbf{X} = X^{(1)} \cdots X^{(m)} / \sqrt{n^m}$ satisfies *condition (D)* if the matrices $X^{(q)}$, $q = 1, \ldots, m$, uniformly satisfy condition (C), i.e. have jointly independent entries and satisfy

$$\left| \mathbb{E}\, X_{ij}^{(q)} \right| \leq n^{-1-\varepsilon} \text{ and } \left| 1 - \mathbb{E} \left| X_{ij}^{(q)} \right|^2 \right| \leq n^{-1-\varepsilon}$$

for some $\varepsilon > 0$ independent of n and furthermore

$$\max_{i,j,q,n} \mathbb{E} \left| X_{ij}^{(q)} \right|^{4+\delta} < \infty$$

for some $\delta > 0$.

It is neither useful nor easy to show concentration of the logarithmic potentials of μ_n^m and μ_∞^m due to non-linearity addressed above. Instead, Götze, Naumov and Tikhomirov use the following generalization of Proposition 2.18.

Proposition 2.23 ([GNT19a]). *If* $\mathbf{X}$ *obeys (D), then for every* $\tau, Q > 0$ *there exists a constant* $c > 0$ *such that*

$$\mathbb{P}\left(|U_n(z) - U_\infty(z)| \le c\frac{\log^4 n}{n}\right) \ge 1 - n^{-Q} \tag{2.51}$$

holds uniformly in $\{z \in B_{1+\tau^{-1}} : |1 - |z|| \ge \tau\}$.

Since this is not explicitly worked out in [GNT19a], we will derive it here, based on the results proved in this paper. Note that Proposition 2.23 shows that the ESD of $\mathbf{W}$ converges to the Circular Law. We will directly follow the approach of [GNT19a], making use of Girko's Hermitization trick (2.32) to convert the non-Hermitian problem into a Hermitian one, apply the local Stieltjes transform estimate from [GNT19a] and the Smoothing Inequality from [GT03a]. Let $\widetilde{\nu}_n^z$ be the symmetrized empirical singular value distribution of the shifted linearized matrices $\mathbf{W} - z$, defined in (2.32) and $m_n(z,\cdot)$ be its Stieltjes transform. The discussion of the limit $s(z,\cdot)$ is identical to (2.35).

Proof of Proposition 2.18 and Proposition 2.23. Fix some arbitrary $Q,\tau > 0$ and $z \in B_{1+\tau^{-1}}$ satisfying $|1 - |z|| \ge \tau$. As is explained in Girko's Hermitization trick (2.32),

$$|U_n(z) - U_\infty(z)| = \left|\int_{\mathbb{R}} \log|x|\, d(\widetilde{\nu}_n^z - \widetilde{\nu}^z)(x)\right|$$

and therefore it is necessary to estimate the extremal singular values as well as the rate of convergence of $\widetilde{\nu}_n^z$ to $\widetilde{\nu}^z$ in Kolmogorov distance

$$d_n^*(z) = \sup_{x\in\mathbb{R}} |(\widetilde{\nu}_n^z - \widetilde{\nu}^z)(-\infty,x]|\,.$$

Introduce the events

$$\Omega_0 := \{s_{\min} \ge n^{-B}\},\quad \Omega_1 := \{s_{\max} \le n^{B'}\},\quad \Omega_2 := \{d_n^*(z) \le c\log^3 n/n\}$$

for some constants $B,B',c > 0$ yet to be chosen. From [OS11, Theorem 31] it follows that there exists a constant $B > 0$ such that $\mathbb{P}(\Omega_0^c) \lesssim n^{-Q}$. Moreover for any $Q > 0$ it holds

$$\mathbb{P}(s_{\max} \ge n^{(Q+1)/2}) \le \frac{\mathbb{E}\,\|(\mathbf{W}-z)\|^2}{n^{Q+1}} \le \frac{1}{n^{Q+1}}\sum_{ij}^{nm} \mathbb{E}\,|\mathbf{W}_{ij}|^2 + \frac{|z|^2}{n^Q} \le \big(m + \tau^{-1}\big)n^{-Q},$$

similar to (2.48). Thus, there exists a constant $B' > 0$ with $\mathbb{P}(\Omega_1^c) \lesssim n^{-Q}$. Since $\widetilde{\nu}^z$ has a bounded density, we get

$$\left| \int_{-n^{-B}}^{n^{-B}} \log|x|\, d\widetilde{\nu}^z(x) \right| \lesssim \frac{\log n}{n^B}$$

and furthermore on Ω_2 it holds that

$$\left| \int_{n^{-B} \le |x| \le n^{B'}} \log|x|\, d(\widetilde{\nu}_n^z - \widetilde{\nu}^z)(x) \right| \lesssim d_n^*(z) \log n \lesssim \frac{\log^4 n}{n}.$$

Hence, the claimed concentration of U_n holds on $\Omega_0 \cap \Omega_1 \cap \Omega_2$, implying

$$\mathbb{P}\left(|U_n(z) - U_\infty(z)| \ge c \frac{\log^4 n}{n} \right) \le \mathbb{P}(\Omega_0^c) + \mathbb{P}(\Omega_1^c) + \mathbb{P}(\Omega_2^c).$$

It remains to check $\mathbb{P}(\Omega_2^c) \le n^{-Q}$, which has been done explicitly in [GNT19a, (4.14)-(4.16)], using the Smoothing Inequality [GNT19a, Corollary B.3] (originally obtained in [GT03a]) and the Local Law for $\mathbf{V}(z)$ in terms of its Stieltjes transform. Let us provide both, for completeness. In our setting, the Smoothing Inequality reads as follows.

Proposition 2.24 ([GT03a; GNT19a])**.** *Let $0 < v_0^{2/3} \lesssim \varepsilon < 1/2$ and $z \in B_{1+\tau^{-1}}$ with $|1 - |z|| \ge \tau$. There exist absolute constants c,$V > 0$ such that for any $v_0 < V$ we have*

$$d_n^*(z) \lesssim \int_{-\infty}^{\infty} |m_n - s|\,(z,u+iV)du + \sup_{\substack{x \in \mathbb{J}^z \\ \gamma(x) \ge \frac{\varepsilon}{2}}} \int_{v_0/\sqrt{\gamma(x)}}^{V} |m_n - s|\,(z, x+iv)dv + v_0 + \varepsilon^{3/2},$$

where $\gamma(x) = (\lambda_+ - x)$ if $|z| < 1$ and $\gamma(x) = (x - \lambda_-) \wedge (\lambda_+ - x)$ if $|z| > 1$.

This is a much more precise and useful improvement of Theorem 2.6. In particular, it solely needs information on the Stieltjes transform from the bulk of the spectrum and it makes use of the squareroot-behavior of the limit measure $\widetilde{\nu}^z$ at the edges $\pm\lambda_\pm$, which fails for $|z| = 1$. The Local Law (comparable to (2.9)) for $\mathbf{V}(z)$ states that there exists $c > 0$ such that for any $Q > 0$ and any z given as above, we have

$$\mathbb{P}\Big(\bigcap_{w \in \mathcal{D}} \Big\{ |m_n - s|\,(z,w) \le \frac{c \log^2 n}{nv} \Big\} \Big) \ge 1 - n^{-Q}, \tag{2.52}$$

where $\mathcal{D} = \{w = x+iv \in \mathbb{C} : x \in \mathbb{J}^z, \gamma(x) \ge \varepsilon/2, v_0/\sqrt{\gamma(x)} \le v \le V\}$ coincides with the arguments of the second integral in the Proposition 2.24. Using this, for $v_0 \sim \frac{\log^2 n}{n} = \varepsilon^{3/2}$, it remains to estimate the difference of Stieltjes transforms at arguments lying far away from the real line, which is explicitly done in [GNT19a] using [GNT19a, Lemma 4.4]. □

2.3. Statement of the Problem

As soon as there is a limit theorem in probability theory, the natural question of rate of convergence arises immediately. A uniform rate of convergence provides an explicit error in applications and numerical simulations, if finite n distributions are replaced by the limiting approximation.

In Hermitian RMT, the investigation of the rate of convergence began with the works of Bai [Bai93a; Bai93b; BHPZ11] and was further developed by Götze and Tikhomirov [GT03a; GT10a; GT04; GT16; GNTT18], see also [BB19; CTX19]. We shall discuss these results in direct comparison to our results in Chapter 4. More detailed rates are available for GUE matrices, see Chapter 3.

The rate of convergence in the Central Limit Theorem goes back to Berry [Ber41] and Esseen [Ess45], see also [BR10; Cra38]. More comparable to our study would be the rate of convergence in the multivariate Central Limit Theorem that was and is studied extensively, e.g. in [Saz68; Göt91; Bha68; FK20]. Therein, the considered metrics on the space of multivariate distributions are similar to our choice of D. Furthermore, the rate of convergence in Kolmogorov distance of empirical processes of i.i.d. vectors to their common distribution is studied in [KW58]. Note that in both cases, the rate of convergence is $n^{-1/2}$ as well, independent of the dimension.

2.3.1. Distance of probability measures

Let μ,ν be probability distributions on $\mathbb{C}$.[1] In order to study rates of convergence between such measures, we need to fix a metric on the space of probability measures. We define the (spherical) *Kolmogorov distance* over balls

$$D(\mu,\nu) := \sup_{z_0\in\mathbb{C}\,,R>0} \left|\mu(B_R(z_0)) - \nu(B_R(z_0))\right|.$$

In [GS02], D is called discrepancy metric[2]. Since the name discrepancy is rather vacuous, we prefer the term Kolmogorov distance due to its uniform character that is comparable to Kolmogorov's original metric, see [Kol33; Ste92]. Obviously, D defines a metric on the space of probability measures on $(\mathbb{C}\,,\mathcal{B}(\mathbb{C}))$. Moreover we justify the name by formally retrieving the one-dimensional Kolmogorov distance $d^*(\mu_j,\nu_j)$ of the marginals $j = 1,2$

1 Not to be confused with ν^z from the previous section, which is a particular distribution on $\mathbb{R}$.
2 Note that [GS02] defines D with closed balls in the supremum instead of open balls, which does not make any difference in our considerations.

in limits such as

$$(\mu_1 - \nu_1)((-\infty,t]) = \lim_{K\to\infty} (\nu - \mu)(B_K(t + K,0)).$$

The most important property of this distance of distributions is that it captures the weak topology of measures.

Lemma 2.25. *Convergence of distributions on $\mathbb{C}$ with respect to the spherical Kolmogorov distance D implies weak convergence.*

In general, convergence in the Kolmogorov metric D is stronger than weak convergence, similar to the usual Kolmogorov metric. For absolutely continuous limiting distributions however, the converse statement is also true, see for instance [TDH76]. Hence D is a reasonable object for studying the rate of convergence to the Circular Law.

Proof. We will prove vague convergence and tightness. Let μ,ν be distributions on $\mathbb{C}$, $f \in \mathcal{C}_c(\mathbb{C})$ be a continuous function with compact support and $f_r = \frac{1}{\pi r^2} f * \mathbb{1}_{B_r(0)}$ be its ball mean function. Furthermore denote by λ the Lebesgue measure on $\mathbb{C}$ and set $\eta = \mu - \nu + \lambda$. It holds

$$\begin{aligned}
\int f d(\mu - \nu) - \int f - f_r d\eta &= \int f_r d\eta - \int f d\lambda \\
&= \int \int \frac{1}{\pi r^2} \mathbb{1}_{B_r(0)}(y - x) d\lambda(y) d\eta(x) - \int f d\lambda \\
&= \int f(y) \left(\int \frac{1}{\pi r^2} \mathbb{1}_{B_r(y)}(x) d\eta(x) - 1 \right) d\lambda(y) \\
&= \frac{1}{\pi r^2} \int f(y) \left(\mu(B_r(y)) - \nu(B_r(y)) \right) d\lambda(y).
\end{aligned}$$

Now choosing a sequence $\mu = \mu_n$ converging to ν with respect to D implies for all $r > 0$

$$\begin{aligned}
\left| \int f d(\mu_n - \nu) \right| &\leq \left| \int f - f_r d\eta_n \right| + \frac{1}{\pi r^2} \int f(y) \left| \mu(B_r(y)) - \nu(B_r(y)) \right| d\lambda(y) \\
&\leq \int |f - f_r| \, d(\mu_n + \nu + \lambda) + \frac{D(\mu_n,\nu)}{\pi r^2} \int |f(y)| \, d\lambda(y) \\
&\leq 2 \|f - f_r\|_{L^\infty(\lambda)} + \|f - f_r\|_{L^1(\lambda)} + \frac{D(\mu_n,\nu)}{\pi r^2} \int |f(y)| \, d\lambda(y).
\end{aligned}$$

First as $n \to \infty$, the last term converges to 0, then as $r \to 0$, the first term vanishes due to the continuity of f and the second due to Lebesgue Differentiation Theorem. This implies that μ_n converges to ν in weak$*$ convergence.

Tightness follows easily from the convergence in D. For any ε let $N \in \mathbb{N}$ be sufficiently large for $D(\mu_n,\mu) \leq \varepsilon/2$ for all $n > N$. Then choose $K_\varepsilon > 0$ sufficiently large for

$\mu(B^c_{K_\varepsilon}(0))$,$\mu_n(B^c_{K_\varepsilon}(0)) \leq \varepsilon/2$ for all $n \leq N$, then $\mu_n(B^c_{K_\varepsilon}(0)) \leq \varepsilon$ for all n. □

Let us briefly compare D to other distances between probability measures. The metric D belongs to the class of *integral probability metrics*, which can written as $\sup_{f\in\mathcal{F}} \left|\int f d(\mu-\nu)\right|$ for some family $\mathcal{F}$ of functions, see [Mül97]. Other relevant metrics are the

- the total variation distance d_{TV} for $\mathcal{F} = \{\mathbb{1}_A : A \in \mathcal{B}(\mathbb{C})\}$,
- the two dimensional Kolmogorov distance d_K for $\mathcal{F} = \{\mathbb{1}_{(-\infty,s]\times(-\infty,t]} : s,t \in \mathbb{R}\}$,
- the 1-Wasserstein metric d_{W_1} for 1-Lipschitz functions $\mathcal{F}$,
- the bounded Lipschitz metric d_{BL} for $\mathcal{F}$ being 1-Lipschitz functions that are bounded by 1.

The Prokhorov metric is given by

$$d_P(\mu,\nu) = \inf\{\varepsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon \text{ for all } A \in \mathcal{B}(\mathbb{C})\},$$

where A^ε is the ε-neighborhood of A. It is not an integral probability metric, but it metrizes weak convergence.

We recommend [GS02] for a good overview *on choosing and bounding probability metrics* including the proofs of the following bounds, see also [Hub04]. The metrics are related as follows

$$\begin{aligned} d_P &\leq \sqrt{d_{BL}} \leq \sqrt{d_{W_1}}, \\ \tfrac{1}{2}d_{BL} &\leq d_P \leq d_{TV}, \\ \text{and } D &\leq d_{TV}. \end{aligned}$$

The converse $d_{TV} \lesssim d_{W_1}$ only holds true on discrete spaces. In our case, the limiting measure will be absolutely continuous, hence ε-environments A^ε of Borel-sets A are directly related to the measure of A and from [GS02, Theorem 1] we obtain

$$D(\mu_n,\mu_\infty) \leq 3d_P(\mu_n,\mu_\infty).^1$$

Thus, rate of convergence in D is weaker compared to the other distances. Not much is stated in the literature about d, but in our particular case it can be classified as D above, see also [Rac91].

Furthermore, for n sufficiently large we can (with high probability) restrict ourselves from $\mathbb{C}$ to a bounded region $B_K(0)$ for some constant $K > 1$. In this case, the Wasserstein distance is not as strong anymore, i.e.

$$d_{W_1} \leq (K+1)d_P \leq (K+1)d_{TV}.$$

The biggest advantage of D over other metrics for our considerations is its computability. It is possible to explicitly calculate D, for instance in Lemma 3.1.

Let us give a short note on the difficulty of extending our results about the rate of convergence to other integral probability metrics. As we have seen in (2.42), Δf needs to be estimated uniformly in $f \in \mathcal{F}$. In the proof of our Smoothing Inequality 4.2, we will consider smoothened $f \approx \mathbb{1}_{B_R(z_0)}$ for which the Laplacian has small support with controllable bounds. This is not true for arbitrary Lipschitz functions, that may have unbounded Laplacian on a large area.

2.3.2. Numerical Simulations

Before we turn to our results, let us have a look at a few numerical simulations. Some of which confirm and illustrate our results, some lead to new conjectures. Since there is no visible difference between Ginibre and non-Gaussian matrices in any of the following simulations, we chose to focus on the latter in order to underline that these simulations are not an instance of the special case of Gaussian matrices. In the sequel, let X be a random non-Hermitian matrix with independent entries, which are distributed on a normalized discrete cube in $\mathbb{C}$, i.e.

$$\mathbb{P}(X_{ij} = \pm 1/\sqrt{2} \pm i/\sqrt{2}) = 1/4. \tag{2.53}$$

For products $\mathbf{X}$, the simulations do not seem to differ, thus we skip them.

In Figure 1.2, we saw already how the eigenvalues distribute in the complex plane and how the maximizing ball $B_R(z_0)$ of $D(\mu_n, \mu_\infty)$ is attained. Figure 2.2 shows a different viewpoint that underlines why $B_R(z_0)$ is conjectured to be close to $B_1(0)$.

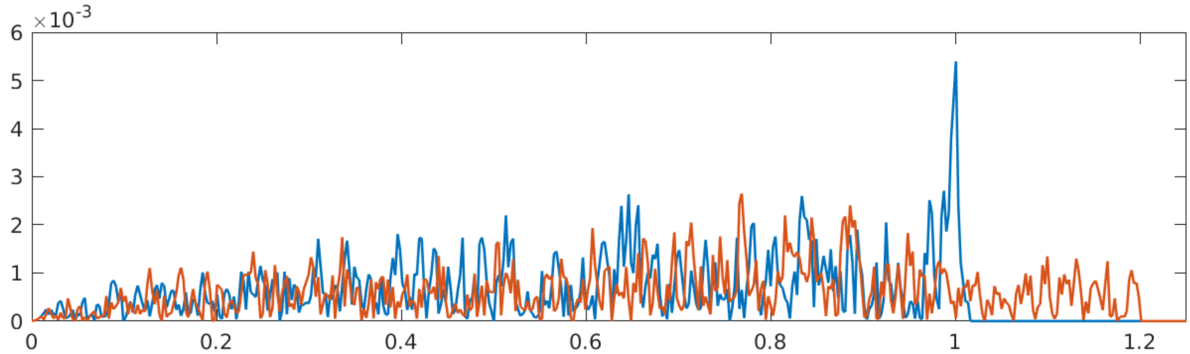

Figure 2.2: The plot shows the values of $|\mu_n(B_R(z_0)) - \mu_\infty(B_R(z_0))|$ as a function of R for $z_0 = 0$ (blue) and $z_0 = 0.2$ (orange) for a single realization of X of size $n = 5000$.

The maximal value $|\mu_n(B_1(0)) - \mu_\infty(B_1(0))| \approx 0.0054$ of the matrix X given in Figure 2.2 is already extraordinarily close to the value $D(\bar{\mu}_n, \mu_\infty) \sim \frac{1}{\sqrt{2\pi n}} \approx 0.0056$, which we compute for the maximal average distance of Ginibre matrices in Lemma 3.1. Furthermore, the plot suggests that for the non-averaged empirical spectral distributions the rate of convergence inside the interior of the bulk seems to be faster than closer to the edge. As already mentioned in the introduction, we cannot expect it to be faster than $\mathcal{O}(1/n)$ due to individual eigenvalues. An average over multiple matrices smoothens the randomness, but keeps the maximal value at $B_1(0)$ as can be seen in Figure 2.3.

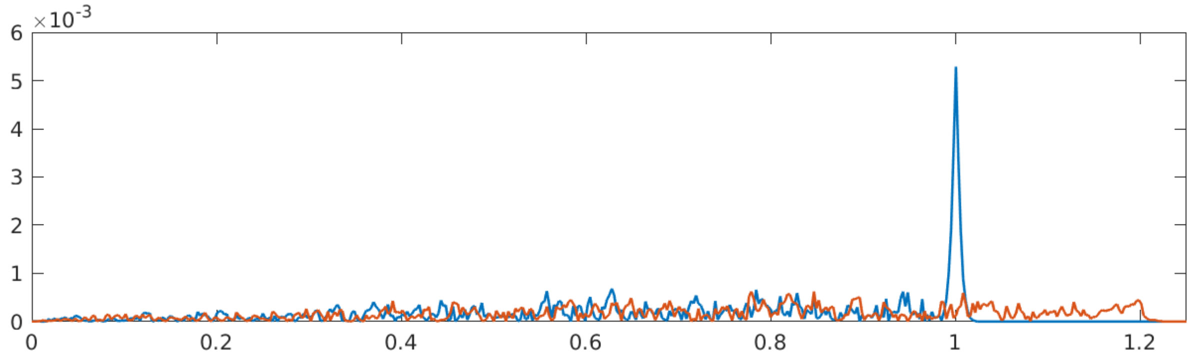

Figure 2.3: The same plot as in Figure 2.2 for the sample mean ESD's averaged over $N = 20$ independent matrices X of size $n = 5000$.

Obviously, the rate of convergence inside the interior of the bulk seems to be significantly faster after averaging. For the Ginibre matrices, where the density of the mean ESD $\mathbb{E}\,\mu_n = \bar{\mu}_n$ is explicitly given by Lemma 2.8, we verify this phenomenon in Lemma 3.1 in the next chapter. In particular, the peak in Figure 2.3 corresponds to the error-function behavior of $\bar{\mu}_n$ at the edge. To avoid repetition, the density of $\bar{\mu}_n$ is plotted in Figure 3.1.

In order to visualize the large n asymptotic of $D(\mu_n,\mu_\infty)$, we consider the supremum over centered balls only, which we abbreviate by

$$D_n = \sup_{R>0} |\mu_n(B_R(0)) - \mu_\infty(B_R(0))|$$

and for centered balls avoiding the edge, we denote

$$D_n^{\text{Bulk}} = \sup_{0<R<0.9} |\mu_n(B_R(0)) - \mu_\infty(B_R(0))|\,.$$

Figure 2.4 shows that we expect the rate of convergence of $D(\mu_n,\mu_\infty)$ to be of order $\mathcal{O}(n^{-1/2})$, apart from a hidden logarithmic factor. We confirm this asymptotic in Chapter 4.

Interestingly, the non-averaged ESD's seem to rather converge at a rate of order $\mathcal{O}(n^{-3/4})$ as long as the edge is avoided. Here, the Coulomb gas picture provides a possible explanation (at least for Gaussian matrices): Let $\mathcal{M}(A)$ be the number of particles in a set $A \in \mathcal{B}(\mathbb{C})$ of a Coulomb gas defined as in (2.22). For instance, the eigenvalues of an unscaled Ginibre matrix X, which are mostly contained in $B_{\sqrt{n}}(0)$. The variance $\mathbb{V}\mathrm{ar}\,\mathcal{M}(B_{R\sqrt{n}}(0))$ of the number of points in a ball is proportional to the volume of the boundary of the ball $\left|\partial B_{R\sqrt{n}}(0)\right| \sim R\sqrt{n}$, see [JLM93; Rid04]. This means

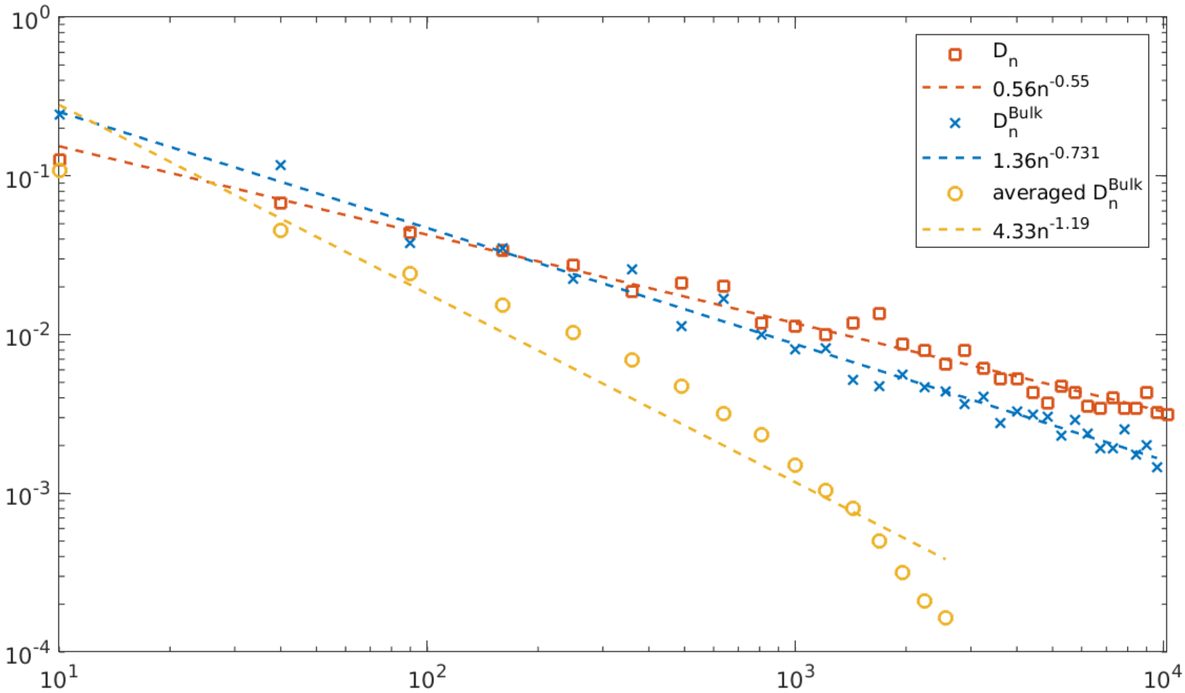

Figure 2.4: Let X be matrices as in (2.53) of different sizes $10 \leq n \leq 10000$. The double logarithmic plot shows D_n (orange) and D_n^{Bulk} (blue) as well as D_n^{Bulk} after averaging over $N = 2000$ matrices (yellow) a function of n. The linear regressions (dashed lines) express rates of convergence as given in the legend.

that the typical fluctuations of the random variable $\mathcal{M}(B_{R\sqrt{n}}(0))$ is $n^{1/4}$, hence after rescaling we expect $\mathcal{M}(B_{R\sqrt{n}}(0))/n = \mu_n(B_R(0))$ to have standard deviation of order $n^{-3/4}$.

After sample averaging, we expect the ESD's to be close to $\mathbb{E}\,\mu_n$. In this case, Figure 2.4 suggests that the rate in the interior of the bulk is faster than any power of n. In particular the linear regression becomes less meaningful. In Lemma 3.1 we prove that for Ginibre matrices the rate of convergence is in fact of exponential order. It should be mentioned that in a plot of the averaged D_n^{Bulk} for larger matrix size n or lower samples averages N, the curve becomes close to a rate n^{-1} again. This observation coincides with the optimal rate $1/n$ mentioned above, in the case where the averaging did not rule out individual eigenvalue fluctuations.

Chapter 3

Rate of convergence for Gaussian matrices

We begin our investigation with the special case of random matrices having Gaussian entries. As explained in Subsections 2.1.2 and 2.2.1, the mean ESD $\bar{\mu}_n = \mathbb{E}\,\mu_n$ is absolutely continuous with respect to the Lebesgue measure and has an explicit density due to the determinantal structure of μ_n. In this chapter we will derive upper and lower bounds for the rate of convergence for the mean ESD. Generally, the upper bounds show that we expect slightly better rates of convergence for the averaged distributions and lower bounds provide optimality. In particular if $D(\mu_n,\mu_\infty) \lesssim \varepsilon_n$ with overwhelming probability for some rate ε_n, then $\varepsilon_n \gtrsim D(\bar{\mu}_n,\mu_\infty) + n^{-Q}$ follows from $D(\bar{\mu}_n,\mu_\infty) \leq \mathbb{E}\,D(\mu_n,\mu_\infty)$.

3.1. Ginibre matrices

For Ginibre matrices X (see Definition 1.1) we obtain the most detailed and explicit rate of convergence, compared to any other model that we are going to study.

Lemma 3.1. *The mean ESD $\bar{\mu}_n$ of the Ginibre ensemble satisfies*

$$D(\bar{\mu}_n,\mu_\infty) \sim \frac{1}{\sqrt{2\pi n}} \tag{3.1}$$

and

$$\sup_{\substack{B_R(z_0)\subseteq \mathbb{C}\setminus B_{1+\varepsilon}(0) \\ or\ B_R(z_0)\subseteq B_{1-\varepsilon}(0)}} |\bar{\mu}_n(B_R(z_0)) - \mu_\infty(B_R(z_0))| \lesssim e^{-n\varepsilon^2}. \tag{3.2}$$

In the introduction, we already gave a few explanations about why (3.1) is a reasonable optimal rate. Meckes and Meckes [MM15, Proposition 5.1] showed a rate of convergence in total variation distance

$$\frac{e^{-1}}{\sqrt{n}} \leq d_{TV}(\bar{\mu}_n,\mu_\infty) = \sup_{A\in\mathcal{B}(\mathbb{C})} |\bar{\mu}_n(A) - \mu_\infty(A)| \leq \frac{e^{1}}{\sqrt{n}} \tag{3.3}$$

and claim this to be the first rate of convergence of that type. Note that our constant is exact in the limit and fits between the constants in (3.3).

Nevertheless, the exponential rate of convergence inside the bulk (3.2) seems surprising at first sight.

In the case of the Gaussian unitary ensemble, i.e. for Hermitian Gaussian matrices, the rate of the mean ESD has been studied in [GT03b; GT05]. In this case, Götze and Tikhomirov showed that the optimal rate of convergence to the Semicircle Law in terms of the Kolmogorov distance is given by $\mathcal{O}(n^{-1})$. Moreover, this rate does not improve if we restrict ourselves to the bulk $[-2+\varepsilon, 2-\varepsilon]$ of the semicircle law. We illustrate this phenomenon in Figure 3.1.

Very recently Claeys, Fahs, Lambert and Webb [CFLW19, Remark 1.1] showed that the optimal rate of convergence for the non-averaged ESD μ_n of GUE matrices in Kolmogorov distance is asymptotically equivalent to $\log(n)/\pi n$. Notably, this shows that almost sure convergence rates in general have additional logarithmic factors that might belong to the optimal rate.

Proof of Lemma 3.1. As we have seen in Lemma 2.8, the density p_n of $\bar{\mu}_n$ is given by

$$p_n(z) = \frac{1}{\pi} e^{-n|z|^2} \sum_{k=0}^{n-1} \frac{n^k |z|^{2k}}{k!},$$

which converges to $p_\infty(z) = \frac{1}{\pi}\mathbb{1}_{B_1(0)}(z)$. In the case of $z_0 = 0$, we can explicitly calculate

$$\begin{aligned}
\bar{\mu}_n(B_R(0)) &= \frac{1}{\pi} \int_{B_R(0)} e^{-n|z|^2} \sum_{k=0}^{n-1} \frac{n^k |z|^{2k}}{k!} dz \\
&= \frac{1}{n} \sum_{k=0}^{n-1} \int_0^{nR^2} e^{-r} \frac{r^k}{k!} dr \\
&= \frac{1}{n} \sum_{k=0}^{n-1} \left(1 - e^{-nR^2} \sum_{j=0}^{k} \frac{(nR^2)^j}{j!}\right) \\
&= 1 - e^{-nR^2} \sum_{k=0}^{n-1} \frac{(n-k)(nR^2)^k}{nk!} \\
&= 1 - e^{-nR^2} \left(\frac{(nR^2)^n}{n!} + (1-R^2) \sum_{k=0}^{n-1} \frac{(nR^2)^k}{k!} \right),
\end{aligned}$$

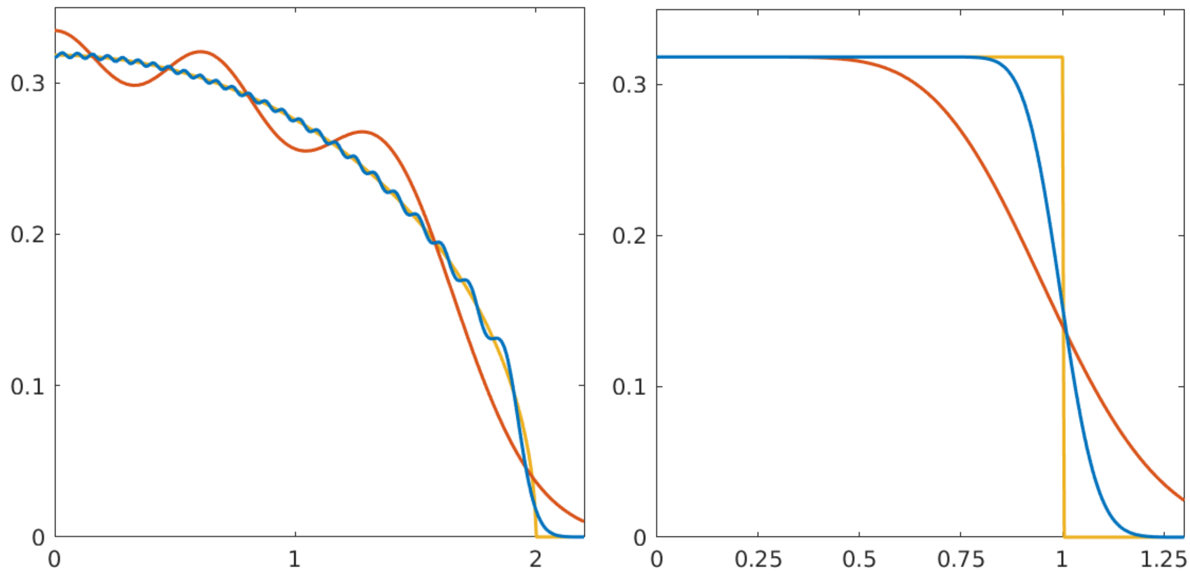

Figure 3.1: Comparison of the densities of the mean ESD of GUE matrices (left) and Ginibre matrices (right) on the positive real line for $n = 5$ (red), $n = 50$ (blue) and $n = \infty$ (yellow). The real eigenvalues tend to be in their predicted locations, but the radial part of complex eigenvalues do not have rigid locations. Therefore we can see much less oscillation of the complex density and the faster rate of convergence in the interior of the bulk does not appear in the case of GUE matrices. A very close look at the edge even reveals the different rates $1/n$ and $1/\sqrt{n}$.

where we used the substitution $r = n\left|z\right|^2$ and integration by parts. The function

$$\begin{aligned}\bar{D}_n(R) &= \mu_\infty(B_R(0)) - \bar{\mu}_n(B_R(0)) \\ &= 1 \wedge R^2 - 1 + e^{-nR^2}\left(\frac{(nR^2)^n}{n!} + (1-R^2)\sum_{k=0}^{n-1}\frac{(nR^2)^k}{k!}\right)\end{aligned}$$

is continuous in R and differentiable for $R \neq 1$ with radial derivative as above

$$2R\left(\mathbb{1}_{[0,1)}(R) - e^{-nR^2}\sum_{k=0}^{n-1}\frac{(nR^2)^k}{k!}\right)\begin{cases} > 0 & \text{, if } R < 1 \\ < 0 & \text{, if } R > 1 \end{cases}.$$

Hence the maximum is attained at $R = 1$ and Stirling's formula yields

$$\sup_{R>0}\left|\bar{\mu}_n(B_R(0)) - \mu_\infty(B_R(0))\right| = \mu_\infty(B_1(0)) - \bar{\mu}_n(B_1(0)) = \frac{n^n}{e^n n!} \sim \frac{1}{\sqrt{2\pi n}}. \tag{3.4}$$

The distances of arbitrary balls are likewise bounded by

$$\mu_\infty(B_R(z_0)) - \bar{\mu}_n(B_R(z_0)) \leq \mu_\infty(B_1(0)) - \bar{\mu}_n(B_1(0))$$
$$\bar{\mu}_n(B_R(z_0)) - \mu_\infty(B_R(z_0)) \leq \bar{\mu}_n(B_1(0)^c) = \mu_\infty(B_1(0)) - \bar{\mu}_n(B_1(0)),$$

hence the first part of the statement is proven. For $R \leq 1$ we have

$$\bar{D}_n(R) = e^{-nR^2}\left(\frac{(nR^2)^n}{n!} - (1-R^2)\sum_{k=n}^{\infty}\frac{(nR^2)^k}{k!}\right)$$

and

$$\begin{aligned} e^{-nR^2}\sum_{k=n}^{\infty}\frac{n^k R^{2k}}{k!} &\leq e^{-nR^2}\frac{(nR^2)^n}{n!}\sum_{k=0}^{\infty}\left(\frac{nR^2}{(n+1)}\right)^k \\ &= e^{-nR^2}\frac{(nR^2)^n}{n!}\frac{n+1}{n(1-R^2)+1} \\ &\sim \frac{1}{\sqrt{2\pi n}}e^{-n(R^2-1-\log(R^2))}\frac{n+1}{n(1-R^2)+1}, \end{aligned}$$

where we applied Stirling's formula again. Consequently

$$\begin{aligned} \left|\bar{D}_n(R)\right| &\lesssim \frac{1}{\sqrt{n}}e^{-n(R^2-1-\log(R^2))}\left(1+(1-R^2)\frac{n+1}{n(1-R^2)+1}\right) \\ &\lesssim \frac{1}{\sqrt{n}}e^{-n(R^2-1-\log(R^2))} \end{aligned}$$

for $R \leq 1$. On the other hand if $R \geq 1$, then

$$\bar{D}_n(R) = e^{-nR^2}\left(\frac{(nR^2)^n}{n!} - (R^2-1)\sum_{k=0}^{n-1}\frac{(nR^2)^k}{k!}\right),$$

where analogously we have

$$\sum_{k=0}^{n-1}\frac{(nR^2)^k}{k!} \leq \frac{(nR^2)^{n-1}}{(n-1)!}\sum_{k=0}^{n-1}\left(\frac{n-1}{nR^2}\right)^k \leq \frac{(nR^2)^n}{(n)!}\frac{1}{(R^2-1)+1}$$

and hence

$$\left|\bar{D}_n(R)\right| \lesssim \frac{1}{\sqrt{n}}e^{-n(R^2-1-\log(R^2))}.$$

Finally choose $R = 1 - \varepsilon$ (or $R = 1 + \varepsilon$, respectively) and note that $R^2 - 1 - \log R^2 \geq 2\varepsilon^2 + \mathcal{O}(\varepsilon^3)$, we conclude

$$\left|\bar{D}_n(1-\varepsilon)\right| \lesssim e^{-n\varepsilon^2}$$

and the second part of the Lemma follows. □

3.2. Products of Ginibre matrices

In this section, we will consider normalized products $\mathbf{X}$ of $m \geq 1$ independent Ginibre matrices. The density of $\bar{\mu}_n^m$ is given in terms of non-elementary functions having a singularity at the origin, as we saw in Lemma 2.10. The main result of this section is the following analogue of Lemma 3.1 showing that the rate of convergence is independent of a fixed $m \in \mathbb{N}$ and that the singularity does not affect this behavior. Let us set $B_R = B_R(0)$ for brevity.

Theorem 3.2. *The mean empirical spectral distribution $\bar{\mu}_n^m = \mathbb{E}\,\mu_n^m$ satisfies*

$$\sup_{R>0} |\bar{\mu}_n^m(B_R) - \mu_\infty^m(B_R)| \asymp \frac{1}{\sqrt{nm}}. \tag{3.5}$$

The following more detailed estimates hold as long as the boundary of the complex disk is avoided

$$\sup_{R<1-\frac{m}{2}\sqrt{\log n/n}} |\bar{\mu}_n^m(B_R) - \mu_\infty^m(B_R)| \lesssim \frac{\log^{3/2} n}{n} \tag{3.6}$$

and uniformly in $R > 1 + \sqrt{\log n/n}$

$$|\bar{\mu}_n^m(B_R) - \mu_\infty^m(B_R)| \lesssim e^{-n(R-1)^2}. \tag{3.7}$$

The precise constants of the upper and lower bound of (3.5) can be chosen to be $C = \sqrt{\pi}/\sqrt{2}$ and $c = 1/(\sqrt{2\pi})$, coinciding with Lemma 3.1. We will see that the maximal distance is, just as in Lemma 3.1, attained at $R = 1$. The rate of convergence is faster inside and much faster outside of the bulk. However, it might be an artifact of the method of proof that we do not obtain an exponential rate of convergence inside the bulk in the case of products of Gaussian random matrices. Only the rate $\mathcal{O}(1/n)$ seems to be achievable due to the discrete nature of the residue calculus, as we will see in (3.14) below. While the proof of Lemma 3.1 is an elementary calculation, the proof of Theorem 3.2 is more involved and relies on a saddle-point method of a double *contour integral representation* for the density of μ_n^m, see Lemma 3.4. An idea of the proof is given after definition of the contours, see Figure 3.2. For an illustration of the statement of Theorem 3.2 we also refer to Figure 1.3.

Remark 3.3. Since the constants and errors in Theorem 3.2 are explicit in m, it is possible to consider the double scaling limit and let $m = m(n) \to \infty$. In this case, the rate will be faster, depending on m, and in particular by setting $m = n$ we obtain a rate

$$\sup_{R>0} |\bar{\mu}_n^n(B_R) - \delta_0(B_R)| \asymp \frac{1}{n}.$$

This follows from the fact that μ_∞^n converges to its weak limit $\mu_\infty^\infty = \delta_0$ at rate $\mathcal{O}(1/n)$. Note that $m \sim n$ is also the critical scaling of Lyapunov exponents between deterministic and GUE statistics, see [ABK19; LW19].

We start with the following double contour integral representation for the density of μ_n^m that is essential for Theorem 3.2.

Lemma 3.4. *The density of $\bar{\mu}_n^m$ satisfies*

$$\rho_n^m(z) = \frac{1}{n(2\pi i)^2} \oint_\gamma \int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} \left(\frac{\Gamma(s)}{\Gamma(t)}\right)^m n^{m(t-s)} |z|^{2(t-s-1)} \cot(\pi t) ds dt,$$

where γ is any closed contour that encircles the numbers $1, \ldots, n$ counter clockwise and no natural number greater than n.

In [KZ14b], a similar double contour integral representation for the correlation kernel of the singular values of $\mathbf{X}$ was derived. This was used in [LWZ16] to prove bulk universality for singular values of products of independent Ginibre matrices. In general, double contour integrals appear regularly in the theory of products of random matrices, e.g. [KZ14b; FW17; KKS15].

Recall that we showed in Lemma 2.10 that the density of $\bar{\mu}_n^m$ has the representation

$$\rho_n^m(z) = n^{m-1} \sum_{k=0}^{n-1} \frac{n^{mk} |z|^{2k}}{\pi (k!)^m} G_{0,m}^{m,0}\left({- \atop 0} \middle| n^m |z|^2 \right), \tag{3.8}$$

see [AB12] and compare to the case $m = 1$, where $G_{0,1}^{1,0}\left({- \atop 0} \middle| n|z|^2 \right) = e^{-n|z|^2}$. According to Theorem 1.3, ρ_n^m converges to the density $\rho_\infty^m(z) = \frac{|z|^{2/m-2}}{\pi m} \mathbb{1}_{B_1}(z)$ of $\bar{\mu}_\infty^m$.

Remark 3.5. The viewpoint of studying products of m matrices and Definition (3.8) of ρ_n^m makes sense for $m \in \mathbb{N}$ only. However the representation of Lemma 3.4 makes sense for arbitrary $m \in \mathbb{R}_+$. Furthermore, as we can see from the proof of Theorem 3.2, its statements (3.5) and (3.7) remain true for real $m > 1$, as well as (3.6) for real $m \geq 2$.

Proof of Lemma 3.4. For the contour L of the Meijer G-function in (3.8), we choose the straight vertical line $L = [-1/2 - i\infty, -1/2 + i\infty]$ that after a simple change of variables $-t = s$ leads to

$$G^{m,0}_{0,m}\left(\begin{matrix} - \\ 0 \end{matrix}\middle| n^m |z|^2\right) = \frac{1}{2\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} \Gamma(s)^m (n^m |z|^2)^{-s} ds \tag{3.9}$$

The remaining part of (3.8) is the kernel (2.26) of the (monic) orthogonal polynomials with respect to the Meijer-G-weight. It can be rewritten with the help of the residue theorem. For any closed curve γ encircling the numbers $1, \ldots, n$ but no natural number greater than n, we have

$$\frac{1}{2\pi i}\oint_\gamma \frac{(n^m |z|^2)^{t-1}}{\Gamma(t)^m}\cot(\pi t)dt = \sum_{k=0}^{n-1}\frac{n^{mk}|z|^{2k}}{\pi (k!)^m},$$

since the integrand is holomorphic except for its simple poles in $\mathbb{N}$ with residues $1/\pi$ each. Combining both contour integrals proves the claim. $\square$

Asymptotic expansions of $G^{m,0}_{0,m}$, like [Luk14, §5.9.1.], together with heuristics for the hypergeometric kernel give rise to pointwise limits in [AB12]. A rigorous estimation of the error bound uniformly in z seems to be absent in the literature so far. Hence it is reasonable to study the problem by a direct analysis.

Proof of Theorem 3.2. By Lemma 3.1, it is sufficient to consider $m \geq 2$. For $R > 1$ we have $|(\bar{\mu}^m_n - \mu^m_\infty)(B_R)| < |(\bar{\mu}^m_n - \mu^m_\infty)(B_1)|$, since $\operatorname{supp}(\mu^m_\infty) = B_1(0)$. Throughout the proof we assume

$$\log^{3m/4} n / n^{m/2} \leq R \leq 1 \tag{3.10}$$

since for smaller values of R it holds

$$|(\bar{\mu}^m_n - \mu^m_\infty)(B_R(0))| \leq \left|(\bar{\mu}^m_n - \mu^m_\infty)(B_{\log^{3m/4} n/n^{m/2}}(0))\right| + \mathcal{O}(\log^{3/2} n/n),$$

due to $\mu^m_\infty(B_R) = R^{2/m}$. We first use spherical symmetry of ρ^m_n and Lemma 3.4 in order to calculate

$$\begin{aligned}\bar{\mu}^m_n(B_R) &= \int_0^{R^2} \pi\rho^m_n(\sqrt{r})dr \\ &= \frac{\pi}{n(2\pi i)^2}\oint_\gamma\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}\left(\frac{\Gamma(s)}{\Gamma(t)}\right)^m \frac{(n^m R^2)^{t-s}}{t-s}\cot(\pi t)dsdt. \end{aligned}\tag{3.11}$$

This holds in the case where s and t have distance bounded from below, which is what we will choose in the following. We will now show that shifting the vertical contour in Lemma 3.4 to $L = [\eta - i\infty, \eta + i\infty]$ for another real part $\eta \geq 1/2$, $\eta \neq 1, \ldots, n$, produces an additional term. Cauchy's integral formula implies

$$\frac{\pi}{(2\pi i)^2}\left(\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}-\int_L\right)\left(\frac{\Gamma(s)}{\Gamma(t)}\right)^m\frac{(n^mR^2)^{t-s}}{t-s}ds=\frac{\pi}{2\pi i}\mathbb{1}_{(1/2,\eta)}(\mathrm{Re}(t)).$$

We temporarily split γ into two parts γ_l and γ_r such that γ_l encircles $\{1,\dots,\lfloor\eta\rfloor\wedge n\}$ and γ_r encircles $\{\lceil\eta\rceil,\dots,n\}$. Soon we will make the path of γ more explicit. Continuing the integration of the right hand side of the last equation over $\gamma_l\cup\gamma_r$ as in (3.11) yields

$$\frac{\pi}{2\pi i}\oint_{\gamma_l}\cot(\pi t)dt=\lfloor\eta\rfloor\wedge n, \tag{3.12}$$

hence we conclude

$$\bar{\mu}_n^m(B_R)=\frac{\pi}{n(2\pi i)^2}\oint_\gamma\int_L\left(\frac{\Gamma(s)}{\Gamma(t)}\right)^m\frac{(n^mR^2)^{t-s}}{t-s}\cot(\pi t)dsdt+\frac{\lceil\eta\rceil}{n}\wedge 1.$$

Choosing $\eta=\lfloor nR^{2/m}\rfloor+1/2$ we see that the second term is $\mathcal{O}(1/n)$ close to $\mu_\infty^m(B_R)=R^{2/m}\wedge 1$. Moreover, by Cauchy's integral formula, we may artificially add the removed part $\gamma-\gamma_l-\gamma_r$ again as long as γ is symmetric around the x-axis. Let γ be the rectangular contour connecting the vertices $3/4-i$, $n+1/4-i$, $n+1/4+i$ and $3/4+i$. This ensures a constant distance to the singularities. The scaled version $\widetilde{\gamma}=\gamma/(R^{2/m}n)$ is illustrated below in Figure 3.2. Furthermore note that the integral exists as we will explicitly show below, see (3.24). Recall Stirling's formula for the Gamma function

$$\log\Gamma(z)=(z-1/2)\log z-z+\frac{1}{2}\log 2\pi+\mathcal{O}(1/\mathrm{Re}z),$$

which holds uniformly for $\mathrm{Re}z\geq 1/2$, cf. for instance [WW96, p.249]. Thus, we have

$$\log\left(\frac{\Gamma(s)^m}{(n^mR^2)^s}\right)=m\left(s\left(\log\left(\frac{s}{nR^{2/m}}\right)-1\right)+\frac{1}{2}\log\left(\frac{2\pi}{s}\right)\right)+\mathcal{O}(1/\mathrm{Re}(s)), \tag{3.13}$$

where for all $s\in L$ (and analogously for $t\in\gamma$) the error term is at most $\mathcal{O}(1)$. We rescale the integration by $nR^{2/m}$ and denote $\widetilde{\gamma}=\gamma/nR^{2/m}$, $\widetilde{L}=L/nR^{2/m}$ as well as

$$F(z)=z\log z-z$$

to obtain

$$\begin{aligned}
&\bar{\mu}_n^m(B_R) - \frac{\lceil nR^{2/m} \rceil}{n} \\
&\quad = \frac{\pi R^{2/m}}{(2\pi i)^2} \oint_{\widetilde{\gamma}} \int_{\widetilde{L}} \left(\frac{\Gamma(nR^{2/m}s)}{\Gamma(nR^{2/m}t)} \right)^m \frac{(n^m R^2)^{(nR^{2/m})(t-s)}}{t-s} \cot(\pi n R^{2/m} t) ds dt \\
&\quad = \frac{\pi}{(2\pi i)^2} \oint_{\widetilde{\gamma}} \int_{\widetilde{L}} \exp\left[nmR^{2/m} \left(F(s) - F(t) \right) \right] \left(\frac{t}{s} \right)^{m/2} \frac{\cot(\pi n R^{2/m} t)}{t-s} \qquad (3.14) \\
&\qquad \cdot \left(R^{2/m} + \mathcal{O}\left(\tfrac{1}{n\mathrm{Re}(s)} \right) + \mathcal{O}\left(\tfrac{1}{n\mathrm{Re}(t)} \right) \right) ds dt.
\end{aligned}$$

Observe that $\mathcal{O}(1/(n\mathrm{Re}(t))) = \mathcal{O}(R^{2/m})$ and $\mathcal{O}(1/(n\mathrm{Re}(s))) = \mathcal{O}(1/n)$. We will analyze this main formula using the method of steepest descent, hence we are interested in the saddle points of F. The saddle point equation simply reads

$$F'(z) = \log z = 0$$

and is obviously satisfied only for $z = 1$ only, with $F''(1) = 1 > 0$. Denoting $z = x + iy$, the Cauchy-Riemann equations for F imply

$$\partial_y \mathrm{Re} F(z) = -\mathrm{Im} F'(z) = -\arg(z) > 0 \Leftrightarrow y < 0, \qquad (3.15)$$

$$\partial_x \mathrm{Re} F(z) = \mathrm{Re} F'(z) = \log |z| > 0 \Leftrightarrow |z| > 1,. \qquad (3.16)$$

Hence $\mathrm{Re}F$ attains its local maximum $F(1) = -1$ in y-direction and its minimum in x-direction. Define the box $Q_{\delta_n}(1) = [1-\delta_n, 1+\delta_n] \times [-\delta_n, \delta_n]$ around $z = 1$ of range

$$\delta_n = \sqrt{\frac{\log n}{nR^{2/m}}} \leq \log^{-1/4} n \to 0$$

by our assumption (3.10). Note that $\widetilde{\gamma}$ is $1/nR^{2/m} \in \mathcal{O}(\delta_n^2)$-close to the real axis and the vertical path $\widetilde{L}$ is equally close to the saddle point $z = 1$. The local part of the paths are given by $L_{loc} = \widetilde{L} \cap Q_{\delta_n}(1)$ and $\gamma_{loc} = \widetilde{\gamma} \cap Q_{\delta_n}(1)$ as well as L_{loc}^c and γ_{loc}^c denotes the remaining part of the path (under slight abuse of notation).

Let us collect the necessary bounds for each part of the contour by applying a Taylor approximation around $z = 1$. We have $(s-1) = i\mathrm{Im}(s) + \mathcal{O}(\delta_n^2)$, hence for $s \in L_{loc}$

$$F(s) = -1 - \mathrm{Im}(s)^2/2 + \mathcal{O}(\delta_n^3), \qquad (3.17)$$

and similarly for $t \in \gamma_{loc}$

$$F(t) = -1 + (1 - \mathrm{Re}(t))^2/2 + \mathcal{O}(\delta_n^3), \qquad (3.18)$$

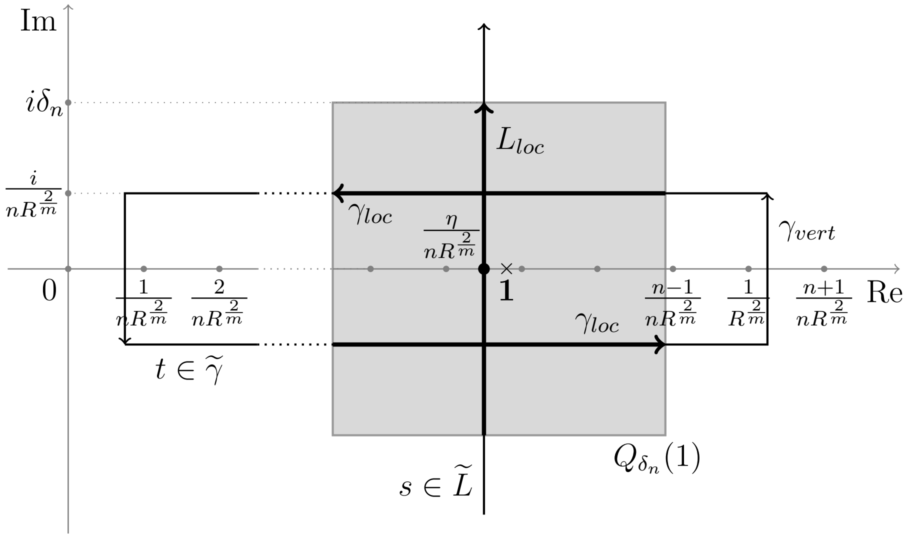

Figure 3.2: The scaled contours of integration and their local parts in thicker lines. As we will see later, the main contribution comes from γ_{loc} that is in a box δ_n-close to the saddle point at $z = 1$. If $R > 1$, there is no γ_{loc} and the integral vanishes exponentially fast (depending on the distance $|R - 1|$). If $R < 1$, then both horizontal contours of γ_{loc} will cancel, because of their symmetry. In this case we will obtain a rate of convergence of microscopic order $1/n$ due to the discrete nature of the residues. The maximal rate will be attained for $R = 1$, where the integrals do not cancel, yet the vertical part γ_{vert} is small enough.

since $|\mathrm{Im}(t)| \lesssim \delta_n$. On the other hand for $s \in L_{loc}^c$ by using (3.15) it holds

$$\mathrm{Re}F(s) < \mathrm{Re}F(\eta/nR^{2/m} + i\delta_n) = -1 - \delta_n^2/2 + \mathcal{O}(\delta_n^3) \tag{3.19}$$

and for $t \in \gamma_{loc}^c$ we see from (3.16)

$$\mathrm{Re}F(t) > \mathrm{Re}F(1 \pm \delta_n) = -1 + \delta_n^2/2 + \mathcal{O}(\delta_n^3). \tag{3.20}$$

The nonlocal terms are negligible, e.g. we apply (3.18) and (3.19) to obtain

$$\begin{aligned}
&R^{2/m} \int_{\gamma_{loc}} \int_{L_{loc}^c} \exp\left[nmR^{2/m}\left(\mathrm{Re}F(s) - \mathrm{Re}F(t)\right)\right] \left|\frac{t}{s}\right|^{m/2} \frac{\left|\cot(\pi n R^{2/m} t)\right|}{|t-s|} ds dt \\
&\lesssim R^{2/m} \int_{\gamma_{loc}} \int_{L_{loc}^c} \exp\left[-\frac{m}{2}\log n + \mathcal{O}(\delta_n)\right] \frac{1}{|s|^{m/2}\,|\mathrm{Im}(s)|} ds dt \\
&\lesssim R^{2/m} n^{-m/2} \lesssim n^{-1},
\end{aligned}$$

where we used $|\mathrm{Im}(s)| \gtrsim \delta_n$, $|\gamma_{loc}| \in \mathcal{O}(\delta_n)$, $t \in \mathcal{O}(1)$, $\left|\cot(\pi n R^{2/m} t)\right| \asymp 1$ and $m \geq 2$.

Moreover from (3.19), (3.20) and $t \in \mathcal{O}(R^{-2/m})$ it follows

$$
\begin{aligned}
&R^{2/m} \int_{\gamma_{loc}^c} \int_{L_{loc}^c} \exp\left[nmR^{2/m}\left(\mathrm{Re}F(s) - \mathrm{Re}F(t)\right)\right] \left|\frac{t}{s}\right|^{m/2} \frac{\left|\cot(\pi n R^{2/m} t)\right|}{|t-s|} dsdt \\
&\lesssim R^{2/m} \int_{\gamma_{loc}^c} \int_{L_{loc}^c} \exp\left[-m\log n + \mathcal{O}(\delta_n)\right] \frac{R^{-1}}{|s|^{m/2}\,|\mathrm{Im}(s)|} dsdt \\
&\lesssim R^{-1} n^{-m} \lesssim n^{-1},
\end{aligned}
$$

where the last step once more follows from the assumption (3.10). Analogously we obtain from (3.17), (3.20)

$$
\begin{aligned}
&R^{2/m} \int_{\gamma_{loc}^c} \int_{L_{loc}} \exp\left[nmR^{2/m}\left(\mathrm{Re}F(s) - \mathrm{Re}F(t)\right)\right] \left|\frac{t}{s}\right|^{m/2} \frac{1}{|t-s|} dsdt \\
&\lesssim R^{2/m} \int_{\gamma_{loc}^c} \int_{L_{loc}} \exp\left[-\frac{m}{2}\log n + \mathcal{O}(\delta_n)\right] \frac{|t|^{m/2}}{|\mathrm{Re}(t) - 1|} dsdt \\
&\lesssim \delta_n R^{-1+2/m} n^{-m/2} \lesssim n^{-1}.
\end{aligned} \tag{3.21}
$$

Locally close to $z = 1$, the error term of Stirling's formula (3.14) is $\mathcal{O}(n^{-1})$. Thus, it remains to control

$$
\begin{aligned}
&\int_{\gamma_{loc}} \int_{L_{loc}} \exp\left[nmR^{2/m}\left(F(s) - F(t)\right)\right] \left(\frac{t}{s}\right)^{\frac{m}{2}} \frac{\cot(\pi n R^{2/m} t)}{t-s} \left(R^{2/m} + \mathcal{O}(n^{-1})\right) dsdt \\
&= \int_{\gamma_{loc}} \int_{L_{loc}} \exp\left[-\frac{nmR^{2/m}}{2}\left(\mathrm{Im}(s)^2 + (1-\mathrm{Re}(t))^2\right)\right] \frac{\cot(\pi n R^{2/m} t)}{t-s} \\
&\qquad \cdot \left(R^{2/m} + \mathcal{O}\left(R^{1/m}\sqrt{\frac{\log^3 n}{n}}\right)\right) dsdt,
\end{aligned} \tag{3.22}
$$

where we used (3.17), (3.18) and $t/s = 1 + \mathcal{O}(\delta_n)$. We parameterize L_{loc} as the straight line

$$
s = \frac{\eta}{nR^{2/m}} + \frac{i}{\sqrt{nmR^{2/m}}} u, \quad u \in I = (-\sqrt{m\log n}, +\sqrt{m\log n}).
$$

The vertical microscopic part

$$
\gamma_{vert} = [3/4 + i, 3/4 - i]/nR^{2/m} \cup ([n + 1/4 - i, n + 1/4 + i]/nR^{2/m})
$$

receives the same scaling, e.g. for the right part we choose

$$t = R^{-2/m} + \frac{1}{4nR^{2/m}} + \frac{i}{\sqrt{nmR^{2/m}}}v, \quad v \in (-\sqrt{m/nR^{2/m}}, \sqrt{m/nR^{2/m}}).$$

This part of the integral (3.22) on γ_{vert} is visible if and only if R is close to 1. The exponential function becomes $e^{u^2/2}$ after dropping the negligible part in t. Using $R \sim 1$ and $|\cot(\pi/4 + ix)| = 1$ for $x \in \mathbb{R}$, the integration over γ_{vert} can then be bounded by

$$\begin{aligned}
&\left| \int_{-\sqrt{m/nR^{2/m}}}^{\sqrt{m/nR^{2/m}}} \int_I \frac{e^{-u^2/2} \cot\left(\pi/4 + ivR^{1/m}\sqrt{n/m}\right)}{m(n-\eta+1/4)/R^{1/m} + i(v-u)\sqrt{nm}} R^{1/m} dudv \right| \\
&\lesssim \int_{-\sqrt{m/nR^{2/m}}}^{\sqrt{m/nR^{2/m}}} \int_I \frac{e^{-u^2/2}}{\sqrt{m(n-\eta+1/4)^2/R^{2/m} + n(v-u)^2}} dudv \\
&\lesssim \frac{1}{n} \int_{-\infty}^{\infty} \frac{e^{-u^2/2}}{\sqrt{1/n + u^2}} du \lesssim \frac{\log n}{n},
\end{aligned} \tag{3.23}$$

where in the second step we shifted u by $v = \mathcal{O}(1/\sqrt{n})$ and used $m(n-\eta+1/4)^2/R^{2/m} \gtrsim 1$. The last step follows from the asymptotics of the modified Bessel function $K_0(1/4n)$ or more elementary by splitting the integration into $|u| \lesseqgtr 1$. From $\delta_n \to 0$ and $1/nR^{2/m} \to 0$ it follows that the left vertical path is not contained in $Q_{\delta_n}(1)$.

We parameterize the remaining path of $\gamma_{loc} \setminus \gamma_{vert}$ as horizontal lines

$$t = 1 \mp \frac{1}{\sqrt{nmR^{2/m}}} v \pm \frac{i}{nR^{2/m}}, \quad v \in \widetilde{I} \subseteq I,$$

where $\widetilde{I}$ is the part of I such that the corresponding contour overlaps $\widetilde{\gamma}$. The integral (3.22) becomes the sum of

$$\begin{aligned}
&\int_{\widetilde{I}} \int_I e^{-\frac{u^2+v^2}{2}} \frac{\cot(\pi n R^{2/m} \pm i\pi \mp \pi\sqrt{nR^{2/m}/m}v)}{\mp v + (nR^{2/m} - \eta)\sqrt{m/nR^{2/m}} + i\left(\pm\sqrt{m/nR^{2/m}} - u\right)} \\
&\qquad \frac{iR^{1/m} + \mathcal{O}(\sqrt{\frac{\log^3 n}{n}})}{\mp\sqrt{nm}} dudv \\
&\leq \frac{1}{\sqrt{nm}} \int_{\mathbb{R}^2} e^{-|z|^2/2} \frac{1}{|z|} \Big(1 + \mathcal{O}\big(\sqrt{\log^3 n/n}\big)\Big) d\lambda(z) \sim \sqrt{\frac{2\pi^3}{nm}},
\end{aligned} \tag{3.24}$$

where we shifted $u,v \in I$ by $\mathcal{O}(\sqrt{m/nR^{2/m}}) = \mathcal{O}(1/\log n)$ and extended the area of integration. Recalling the correct prefactor $c = -1/4\pi$ from (3.14), we conclude

$$|\bar{\mu}_\infty^m(B_R) - \bar{\mu}_n^m(B_R)| \leq \sqrt{\frac{\pi}{2nm}} + o(n^{-1/2}).$$

In order to obtain the lower bound of the claim, it suffices to consider $R = 1$. Moreover we will only study the sum of (3.24) keeping the phase factor of the integrand, since all the other parts of the double contour integral are proven to be of strictly lower order than $o(n^{-1/2})$. We have the approximation

$$\begin{aligned}
&\bar{\mu}_\infty^m(B_1) - \bar{\mu}_n^m(B_1) \\
&= \frac{1}{4\pi}\Bigg(\int_{-\sqrt{m}/4\sqrt{n}}^{\sqrt{m}\log n}\int_{-\sqrt{m}\log n}^{+\sqrt{m}\log n} e^{-\frac{u^2+v^2}{2}} \frac{\cot(i\pi - \pi\sqrt{n/m}v)}{v + \frac{\sqrt{m}}{2\sqrt{n}} + i\left(-\sqrt{\frac{m}{n}} + u\right)} \frac{i + o(1)}{\sqrt{nm}} dudv \\
&+ \int_{-\sqrt{m}\log n}^{+\sqrt{m}/4\sqrt{n}}\int_{-\sqrt{m}\log n}^{+\sqrt{m}\log n} e^{-\frac{u^2+v^2}{2}} \frac{\cot(-i\pi + \pi\sqrt{n/m}v)}{v - \frac{\sqrt{m}}{2\sqrt{n}} + i\left(-\sqrt{\frac{m}{n}} - u\right)} \frac{i + o(1)}{\sqrt{nm}} dudv \\
&= \frac{1}{4\pi\sqrt{nm}}\int_{\sqrt{m}/4\sqrt{n}}^{\sqrt{m}\log n}\int_{-\sqrt{m}\log n}^{+\sqrt{m}\log n} e^{-\frac{u^2+v^2}{2}} \frac{-i}{v + ui} \\
&\quad \cdot (\tan(i\pi - \pi\sqrt{n/m}v) + \tan(i\pi + \pi\sqrt{n/m}v))dudv + o(n^{-1/2}).
\end{aligned}$$

In the second step we shifted v by a negligible value $\pm\sqrt{m}/2\sqrt{n}$ and u by $\pm\sqrt{m/n}$, inverted $v \to -v$ in the second integral in order to merge both and put $\cot(\pi/2 - z) = -\tan(-z) = \tan(z)$. Note that $\sup_{x\in\mathbb{R}} |\tan(i\pi - x) + \tan(i\pi + x) - 2i| =: \varepsilon < 0.01$ and that the left hand side of the previous equation is real, hence

$$\begin{aligned}
&\bar{\mu}_\infty^m(B_1) - \bar{\mu}_n^m(B_1) \\
&\geq \frac{1-\varepsilon}{2\pi\sqrt{nm}}\int_{\sqrt{m}/4\sqrt{n}}^{\sqrt{m}\log n}\int_{-\sqrt{m}\log n}^{+\sqrt{m}\log n} e^{-\frac{u^2+v^2}{2}} \frac{v}{u^2+v^2} dudv + o(n^{-1/2}) \\
&\sim \frac{1-\varepsilon}{\sqrt{2\pi mn}}.
\end{aligned}$$

The same upper bound holds with $1 + \varepsilon$ instead. For a better control of the constant one may vary the distance of γ to the real axis from the start. The above asymptotic yields the first claim and coincides with $m = 1$ from Lemma 3.1.

If we avoid the edge by some distance $\left|1 - R^{1/m}\right| \gtrsim \sqrt{\log n/n} > 0$, then $\delta_n < \left|1 - 1/R^{2/m}\right|$. Hence γ_{vert} is not a part of γ_{loc} and (3.23) drops out. This is the case for $R < 1 - \frac{m}{2}\sqrt{\log n/n}$, for which we have $I = \widetilde{I}$. As before, we shift u and v by

$\mathcal{O}(1/\sqrt{nR^{2/m}})$ and obtain

$$\int_I \int_I e^{-\frac{u^2+v^2}{2}} \left(\frac{-\cot(\pi n R^{2/m} + i\pi - \pi\sqrt{nR^{2/m}/m}v)}{-v + (nR^{2/m} - \eta)\sqrt{m/nR^{2/m}} + i\left(\sqrt{m/nR^{2/m}} - u\right)} \right.$$
$$\left. + \frac{\cot(\pi n R^{2/m} - i\pi + \pi\sqrt{nR^{2/m}/m}v)}{v + (nR^{2/m} - \eta)\sqrt{m/nR^{2/m}} - i\left(\sqrt{m/nR^{2/m}} + u\right)} \right) \frac{iR^{1/m} + \mathcal{O}(\sqrt{\frac{\log^3 n}{n}})}{\sqrt{nm}} dudv$$
$$= \int_I \int_I e^{-\frac{u^2+v^2}{2}} \frac{iR^{1/m} + \mathcal{O}(\sqrt{\frac{\log^3 n}{n}})}{\sqrt{nm}(v+ui)}$$
$$\cdot \tan(i\pi - \pi\sqrt{nR^{2/m}/m}v) - \tan(-i\pi + \pi\sqrt{nR^{2/m}/m}v) dudv.$$

From the last line it is obvious that the horizontal contour integrals (3.24) cancel, hence

$$\begin{aligned} &\bar{\mu}_\infty^m(B_R) - \bar{\mu}_n^m(B_R) \\ &\lesssim \frac{\sqrt{\log^3 n}}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{u^2+v^2}{2}} \frac{1}{|v+ui|} dudv + \frac{\log n}{n} \\ &\lesssim \frac{\log^{3/2} n}{n} \end{aligned}$$

due to symmetry.

Lastly we turn to the statement about the exponential decay for $R > 1 + \sqrt{\log n/n}$. As before it is sufficient to consider $R \le 2$, because of $\mathrm{supp}(\mu_\infty^m) = B_1$. The position of the minimum in x-direction in (3.16) and $\mathrm{Re}(t) \le (n+1)/nR^{2/m}$ for $t \in \widetilde{\gamma}$ yield

$$\begin{aligned} \mathrm{Re}F(t) &= \mathrm{Re}F(\mathrm{Re}(t)) + \mathcal{O}(1/n) \\ &\ge F\left(\frac{n+1}{nR^{2/m}}\right) + \mathcal{O}(1/n) \\ &= -R^{-2/m}\left(\log(R^{2/m}) + 1\right) + \mathcal{O}(1/n) \end{aligned} \tag{3.25}$$

for n sufficiently large. Since $\gamma_{loc} = \emptyset$, we estimate (3.14) similar to (3.21), hence apply

(3.17), (3.19), (3.25) to obtain

$$
\begin{aligned}
&\int_{\widetilde{\gamma}}\int_{\widetilde{L}} \exp\left[nmR^{2/m}\left(\mathrm{Re}F(s)-\mathrm{Re}F(t)\right)\right]\left|\frac{t}{s}\right|^{m/2}\frac{\left|\cot(\pi n R^{2/m}t)\right|}{|t-s|}dsdt\\
&\lesssim \int_{\widetilde{\gamma}}\int_{\widetilde{L}} \exp\left[-nm(R^{2/m}-1-\log(R^{2/m}))+\mathcal{O}(1)\right]\frac{1}{|s|^{m/2}\left|\mathrm{Im}(s)\right|}dsdt\\
&\lesssim \exp\left[-2n(R-1-\log R)\right]\\
&\leq \exp\left[-n(R-1)^2\right],
\end{aligned}
$$

where again Bernoulli's inequality was used and the last inequality holds for $R > 1$. Ultimately all claims are proven. $\square$

In Lemma 3.1, we were able to consider non-centered balls by a simple monotonicity argument of the density. Here, the same questions turns into the problem of $\rho_n^m/\rho_\infty^m\,|\cdot|$ being monotone[1]. We conjecture that $\rho_n^m/\rho_\infty^m\,|\cdot|$ is unimodal, hence it has value $= 1$ on at most two spheres, but the answer seems unknown. If the conjecture is true, it would imply the rate of convergence to hold on arbitrary balls.

3.3. A slight modification of the model

Having Stirling's formula (3.13) in mind, we notice that $\Gamma(s)^m$ is comparable to $\Gamma(ms)$. Its error (3.14) makes an exponential rate of convergence in the interior of the bulk as in Lemma 3.1 inaccessible. However one may hope for a adaption of the model $\bar{\mu}_n^m$ by replacing $\Gamma(s)^m$ by an identity, the Gauss's multiplication formula

$$
\prod_{j=0}^{m-1}\Gamma(s+\frac{j+\alpha}{m}) = (2\pi)^{(m-1)/2}m^{1/2-ms-\alpha}\Gamma(ms+\alpha) \tag{3.26}
$$

for any $\alpha \in \mathbb{Z}$. In analogy to the mESD of products of Ginibre matrices and its density given in Lemma 3.4, we define a density

$$
\hat{\rho}_n^{m,\alpha}(z) = \frac{1}{n(2\pi i)^2}\oint_\gamma\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}\left(\prod_{j=0}^{m-1}\frac{\Gamma(s+\frac{j+\alpha}{m})}{\Gamma(t+\frac{j+\alpha}{m})}\right)n^{m(t-s)}\left|z\right|^{2(t-s-1)}\cot(\pi t)dsdt,
$$

where γ is any closed contour that encircles the numbers $1,\dots,n$ counter clockwise and no natural number greater than n. For the measure $\hat{\mu}_n^m$ defined via the density $\hat{\rho}_n^m$ of

1 This quotient can also be rewritten in terms of Meijer-G functions.

the average

$$\hat{\rho}_n^m(z) = \frac{1}{m} \sum_{\alpha=-m+1}^{0} \hat{\rho}_n^{m,\alpha}(z)$$

we find the following analogue to Lemma 3.1.

Lemma 3.6. *For $\hat{\mu}_n^m$ defined above, we have*

$$D(\hat{\mu}_n^m, \mu_\infty) \sim \frac{1}{\sqrt{2\pi n m}}$$

and for any fixed $\varepsilon > 0$

$$\sup_{\substack{B_R(z_0) \subseteq \mathbb{C} \setminus B_{1+\varepsilon} \\ \text{or } B_R(z_0) \subseteq B_{1-\varepsilon}}} |\hat{\mu}_n^m(B_R(z_0)) - \mu_\infty(B_R(z_0))| \lesssim e^{-n\varepsilon^2}.$$

Proof. Applying Gauss's multiplication formula (3.26), we obtain

$$\prod_{j=0}^{m-1} \frac{\Gamma(s + \frac{j+\alpha}{m})}{\Gamma(t + \frac{j+\alpha}{m})} = \frac{m^{mt}\Gamma(ms+\alpha)}{m^{ms}\Gamma(mt+\alpha)}.$$

By the residue theorem, we get

$$\frac{1}{2\pi i} \oint_\gamma \frac{\left(nm\,|z|^{2/m}\right)^{mt}}{\Gamma(mt+\alpha)} \cot(\pi t) dt = \sum_{k=0}^{n-1} \frac{\left(nm\,|z|^{2/m}\right)^{m(k+1)}}{\pi\Gamma(m(k+1)+\alpha)}.$$

Furthermore, the term of the Meijer G-function can now be explicitly calculated as the Γ function is nothing but the Mellin transform of $\exp[-\cdot]$, i.e.

$$\frac{1}{2\pi i} \int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} \Gamma(ms+\alpha)\left(nm\,|z|^{2/m}\right)^{-ms} ds = \frac{1}{m}\left(nm\,|z|^{2/m}\right)^{\alpha} \exp\left[-nm\,|z|^{2/m}\right].$$

Combining both parts, we obtain

$$\hat{\rho}_n^{m,\alpha}(z) = \frac{|z|^{-2}}{\pi m n} \exp\left[-nm\,|z|^{2/m}\right] \sum_{k=0}^{n-1} \frac{\left(nm\,|z|^{2/m}\right)^{m(k+1)+\alpha}}{\Gamma(m(k+1)+\alpha)},$$

and consequently after averaging over $\alpha = -m+1,\ldots,0$

$$\hat{\rho}_n^m(z) = \frac{|z|^{2/m-2}}{\pi m} \exp\big[-nm\,|z|^{2/m}\big] \sum_{k=0}^{nm-1} \frac{\big(nm\,|z|^{2/m}\big)^k}{k!},$$

where we recognize the limit density (1.10) as well as the same structure of the density as in the case $m=1$. Consequently we can apply the same ideas as for Lemma 3.1. In particular polar coordinates yield

$$\hat{\mu}_n^m(B_R) = \frac{1}{nm}\int_0^{nmR^{2/m}} e^{-r} \sum_{k=0}^{nm-1} \frac{r^k}{k!} dr.$$

From here, we may follow the lines of the proof of Lemma 3.1 and the claim follows. □

3.4. Non-averaged rate of convergence

So far, we considered mean empirical spectral distributions only. In this section we prove a rate of convergence result for the non-averaged ESD of products of Ginibre matrices, which holds with overwhelming probability.

Theorem 3.7. *For products of Ginibre matrices and any $\varepsilon, Q > 0$ there exists a constant $c > 0$ such that*

$$\mathbb{P}\left(\sup_{R>0} |\mu_n^m(B_R) - \mu_\infty^m(B_R)| \le c\sqrt{\frac{\log n}{n}}\right) \ge 1 - n^{-Q}.$$

This rate coincides exactly with the Wasserstein rate obtained in [CHM18]. The logarithmic factor is expected to be non-optimal and might be removed by controlling the variance term in (3.28) below, but this will not be pursued here. The proof makes use of the determinantal structure of the eigenvalues of Gaussian matrices in order to apply Bernstein's inequality.

Proof. Without loss of generality, we consider $0 < R < 1$, since for $R \ge 1$, we have $|\mu_n^m(B_R) - \mu_\infty(B_R)| \le |\mu_n^m(B_1) - \mu_\infty(B_1)|$. Suppose we had shown already that for any $Q > 0$ there exist a $c > 0$ such that

$$\mathbb{P}\left(|\mu_n^m(B_R) - \mu_\infty^m(B_R)| > t_n/5\right) \le n^{-Q-\lceil m/4\rceil} \tag{3.27}$$

for $t_n = c\sqrt{\log n/n}$ and fixed $0 < R < 1$. Set $p = \lceil m/4\rceil$ and consider the equidistant points $r_k = k/n^p$ for $k = 1,\dots,n$. We have

$$\begin{aligned}|\mu_n^m(B_R) - \mu_\infty^m(B_R)| &\le \mu_n^m(B_{r_{k+1}} \setminus B_{r_k}) + \mu_\infty^m(B_{r_{k+1}} \setminus B_{r_k}) + |\mu_n^m(B_{r_k}) - \mu_\infty^m(B_{r_k})| \\ &\le \left|\mu_n^m(B_{r_{k+1}}) - \mu_\infty^m(B_{r_{k+1}})\right| + 2\mu_\infty^m(B_{r_{k+1}} \setminus B_{r_k}) + 2\left|\mu_n^m(B_{r_k}) - \mu_\infty^m(B_{r_k})\right|\end{aligned}$$

where we chose $k = \lfloor Rn^p \rfloor$. Taking the supremum over R is equivalent to taking the maximum in k, hence the union bound implies

$$\begin{aligned}\mathbb{P}\Big(\sup_{R>0}|\mu_n^m(B_R) - \mu_\infty^m(B_R)| > t_n\Big) \le \sum_{k=1}^{n^p}\Big(&\mathbb{P}\left(\left|\mu_n^m(B_{r_{k+1}}) - \mu_\infty^m(B_{r_{k+1}})\right| > t_n/5\right) \\ &+ \mathbb{P}\left(\mu_\infty^m(B_{r_{k+1}} \setminus B_{r_k}) > t_n/5\right) \\ &+ \mathbb{P}\left(|\mu_n^m(B_{r_k}) - \mu_\infty^m(B_{r_k})| > t_n/5\right)\Big).\end{aligned}$$

The first and last term are covered by (3.27) and the second term vanishes because of

$$\mu_\infty^m(B_{r_{k+1}} \setminus B_{r_k}) = \frac{(k+1)^{2/m} - k^{2/m}}{n^{2p/m}} \le n^{-p(1\wedge 2/m)} \le n^{-1/2}.$$

Thus, we have shown that for all $Q > 0$ there exists a constant $c > 0$ such that $\sup_{R>0}|\mu_n^m(B_R) - \mu_\infty^m(B_R)| \le c\sqrt{\log n/n}$ holds with probability at least $1 - n^{-Q}$.

It remains to show (3.27) for which we follow the ideas of [MM15, Proposition 4] for the case $m = 1$. Fix $R < 1$ and let $\xi_k \sim Ber(\eta_j) \in \{0,1\}$ be independent Bernoulli variables with parameter $\eta_k \in [0,1]$. According to [AGZ10, Corollary 4.2.24], the determinantal point process

$$\#\{\lambda_j(\sqrt{n^m}\mathbf{X}) \in B_{\sqrt{n^m}R}\} = n\mu_n^m(B_R) \stackrel{\mathcal{D}}{=} \sum_{k=1}^{n} \xi_k$$

has the same distribution as the sum of Bernoulli variables, where the parameter η_k are given by the eigenvalues of the trace class operator

$$\mathcal{K}_R f(z) = \int_{B_{\sqrt{n^m}R}} \sqrt{G_{0,m}^{m,0}\left(\begin{smallmatrix}-\\0\end{smallmatrix}\,\middle|\,|z|^2\right) G_{0,m}^{m,0}\left(\begin{smallmatrix}-\\0\end{smallmatrix}\,\middle|\,|w|^2\right)} \sum_{k=0}^{n-1} \frac{(z\bar{w})^k}{\pi k!^m} f(w)\, d\lambda\!\!\lambda(w)$$

for $f \in L^2(B_{\sqrt{n^m}R})$, cf. (2.25). Due to rotational symmetry, like we argued for the orthogonality of the monomials, $\mathcal{K}_R$ has eigenfunctions $\varphi_k(w) = \sqrt{G_{0,m}^{m,0}\left(\begin{smallmatrix}-\\0\end{smallmatrix}\,\middle|\,|w|^2\right)}\, w^k$

with eigenvalues

$$\eta_k = \int_{B_{\sqrt{n^m}R}} G^{m,0}_{0,m}\Big(\begin{smallmatrix} - \\ 0 \end{smallmatrix}\Big| \, |w|^2\Big) \frac{|w|^{2k}}{\pi k!^m} d\lambda\!\!\lambda(w) \le 1.$$

It follows from Theorem 3.2 that

$$\begin{aligned}\mathbb{P}\Big(|\mu_n^m(B_R) - \mu_\infty^m(B_R)| \ge \frac{t}{\sqrt{n}}\Big) &\le \mathbb{P}\Big(|\mu_n^m(B_R) - \bar{\mu}_n^m(B_R)| \ge \frac{t-c}{\sqrt{n}}\Big) \\ &= \mathbb{P}\Big(\Big|\sum_{k=1}^n \xi_k - \mathbb{E}\Big(\sum_{k=1}^n \xi_k\Big)\Big| \ge (t-c)\sqrt{n}\Big).\end{aligned}$$

Applying Bernstein's inequality, see [BLM13, Equation (2.10)] yields

$$\begin{aligned}\mathbb{P}\Big(|\mu_n^m(B_R) - \mu_\infty^m(B_R)| > \frac{t}{\sqrt{n}}\Big) &\le 2\exp\Big(- \frac{n(t-c)^2}{2\sum_{k=1}^n \mathbb{E}\,|\xi_k|^2 + \frac{2}{3}(t-c)\sqrt{n}}\Big) \\ &\le e^{-t^2/3} \end{aligned} \tag{3.28}$$

for $t \gtrsim 1$. In particular for $t = c\sqrt{\log n}/5$ with $c = 5\sqrt{3(Q + \lceil m/4 \rceil)}$, we obtain (3.27) as claimed. $\square$

Chapter 4

Rate of convergence for matrices with independent entries

Our approach to prove rate of convergence results in the Kolmogorov-like Distance D goes back to the ideas of Zhidong Bai [Bai93a] that were used to prove the earliest rate of convergence results for Wigner matrices. He shows a Smoothing Inequality, Theorem 2.6, that quantitatively relates estimates on the difference of Stieltjes transforms to the Kolmogorov distance. Hereafter, a concentration of the Stieltjes transforms directly yields a rate of convergence. This method is very robust in the sense that it simultaneously applies to different models [Bai93a; Bai93b; BHPZ11], and improved Stieltjes transform estimates from Local Laws immediately lead to improved rates of convergence. Bai even used the rate of convergence of $\widetilde{\nu}_n^z$ for his proof of the Circular Law [Bai97].

> "For one of these papers I worked for 13 years from 1984 to 1997 [...]. It was the hardest problem I have ever worked on. The problem is: [The Circular Law]" – Zhidong Bai [BCZH08]

Götze and Tikhomirov [GT03a; GT10a; GT04; GT16; GNTT18] further developed Bai's methods with their Smoothing Inequality, Proposition 2.24, in order to make several improvements on the rate of convergence.

In this chapter, we transfer this method to the non-Hermitian setting. First, we prove Smoothing Inequalities for logarithmic potentials. Then, we apply the results of Subsection 2.2.3 about the concentration of logarithmic potentials in order to obtain rates of convergence results in Kolmogorov distance.

In the case of products of random matrices, the smoothing inequality does not directly apply, but the method of proof will implicitly use the same ideas.

4.1. The Smoothing Inequality for logarithmic potentials

Consider a sequence of probability measures μ_n on $\mathbb{C}$ with logarithmic potentials U_n. In Lemma 2.12 we have seen that if U_n converges pointwise to some function U_μ and if U_n is locally uniformly Lebesgue integrable, then $\mu_n \Rightarrow \mu$. The following smoothing inequality quantifies this statement by relating $D(\mu_n,\mu)$ to the concentration of logarithmic potentials.

Theorem 4.1. *Let μ,ν be probability measures on $\mathbb{C}$ with $\operatorname{supp}\nu \subseteq B_K(0)$ for some $K > 0$, let U_μ, U_ν be their logarithmic potentials and fix some $1 \leq p \leq \infty$. For any $a \geq 1/2$ we have*

$$D(\mu,\nu) \lesssim a^{1+1/p} \left\| U_\mu - U_\nu \right\|_{L^p(B_{K+1/a}(0))} + \sup_{R\geq 0, z_0 \in \mathbb{C}} \nu\left(R \leq |\cdot - z_0| \leq R + 1/a\right).$$

In the same manner it is possible to show an analogue for the classical Kolmogorov distance between 2-dimensional distribution functions, see Corollary 4.4. For measures μ, ν on $\mathbb{R}$, where ν has a bounded density, Dinh and Vu showed in [DV17] another direct relation of similar type

$$|\mu(I) - \nu(I)| \lesssim \left\| U_\mu - U_\nu \right\|_{L^\infty(\operatorname{supp}\nu)}^{1/2}$$

for all intervals $I \subseteq \mathbb{R}$ and it was used to show a rate of convergence in Wigner's Semicircular Law and the Marchenko-Pastur Law.

Theorem 4.1 may be of independent interest, since it provides a complex counterpart of other Smoothing Inequalities of distributions μ,ν on the real line, some of which we have encountered already. Let us take a look at Theorem 4.1 in direct comparison to known Smoothing Inequalities.

For instance in the case of Fourier transforms $\varphi_\mu(t) = \int e^{itx} d\mu(x)$, the well known *Berry-Essen inequality* is a Smoothing Inequality of the type

$$d^*(\mu,\nu) = \sup_{x\in\mathbb{R}} |(\mu - \nu)((-\infty,x])| \lesssim \int_{-a}^{a} \left| \frac{\varphi_\mu(t) - \varphi_\nu(t)}{t} \right| dt + \sup_{x\in\mathbb{R}} \nu((x,x+c/a]), \quad (4.1)$$

see e.g. [Pet12, V§1]. This leads to a rate of convergence of order $1/\sqrt{n}$ in the Central Limit Theorem, when choosing $\nu = \mathcal{N}(0,1)$ and $\mu = \mathbb{P}_{S_n}$ for the normalized sum $S_n = n^{-1/2} \sum_{k=1}^n X_k$ of i.i.d. random variables X_k with $\mathbb{E}\, X_1 = 0, \mathbb{E}\, X_1^2 = 1$ and finite third moment $\mathbb{E}\, X_1^3 < \infty$, see e.g. [Ess45; Ber41]. Berry-Esseen bounds also occur in RMT, see for instance [XGL19; BB19].

More important for Random Matrix Theory is *Bai's inequality*, Theorem 2.6. It is a handy tool to profit from the control of the Stieltjes' transforms m_μ that can be simplified[1] to

$$d^*(\mu,\nu) \lesssim \int |m_\mu - m_\nu| (t + i/a) dt + \sup_{x\in\mathbb{R}} \nu((x,x+c/a]). \quad (4.2)$$

Roughly speaking, [BS10, Chapter 8] uses $a \simeq \sqrt{n}$ to show a rate of convergence of

1 More precisely we set α as a constant, $v = 1/a$ and roughly estimate the second integral. Note that a in [BS10] corresponds to our α and not to our a that is $1/v$ in [BS10].

order $1/\sqrt{n}$ for the Kolmogorov distance in Wigner's Semicircle Law under finite sixth moment condition. Using the improved, but more involved, Smoothing Inequality from Proposition 2.24, it is shown in [GT16] that the optimal expected rate of convergence to the semicircle distribution is given by $\mathcal{O}(1/n)$. To our knowledge, the best rate of convergence of the non-averaged ESD to the Semicircle Law is given by $\mathcal{O}(\log^2(n)/n)$ obtained in [GNTT18, Equation (1.10)].

All smoothing inequalities (4.1), (4.2) and Theorem 4.1 are used to derive convergence rates under moment conditions. Furthermore, they share the essential structure of bounding the Kolmogorov distance by the distance of certain integral-transforms and an additional maximal annulus probability of width $\mathcal{O}(1/a)$ with respect to the limiting distribution. In the next section, we shall consider the ESD's $\mu = \mu_n, \nu = \mu_\infty$ of non-Hermitian random matrices in Theorem 4.1. Heuristically, if we choose $a = \sqrt{n}$ and $K = 1$, then the remainder term is of order $n^{-1/2}$, which suggests a rate of convergence for $D(\mu_n, \mu_\infty)$.

We prove the following slightly more general statement that covers all variants we need. This version is comparable to the generalized Bai's inequality [BS10, Theorem B.14].

Theorem 4.2. *Let μ,ν be probability measures on $\mathbb{C}$ with logarithmic potentials U_μ, U_ν respectively, fix $1 \leq p \leq \infty$ and for some $z^* \in \mathbb{C}$, $K > 0$, $\eta \geq 0$ define the annuli $V = B_K(z^*) \setminus B_{2\eta/a}(z^*)$ and $V' = B_{K+2/a}(z^*) \setminus B_{\eta/a}(z^*)$. For any $a > 1$ we have*

$$\begin{aligned} D(\mu,\nu) \lesssim & a^{1+1/p} \left\| U_\mu - U_\nu \right\|_{L^p(V')} + \mu(V^c) + \nu(V^c) \\ & + \sup_{R \geq 0, z_0 \in \mathbb{C}} \nu\left(z \in V' : R \leq |z - z_0| \leq R + (2 \vee \eta)/a\right). \end{aligned}$$

Here, $\eta \neq 0$ is only needed for the applications to random polynomials, where the logarithmic potential near the origin cannot be controlled. From the proof it follows that the implicit constant hidden in $\lesssim$ can be chosen to be $c = 3 \vee 4(K \vee \eta)^{1-1/p}$.

We retrieve Theorem 4.1 by taking $\eta = 0$, $z^* = 0$, $\nu(V^c) = 0$, replacing a by $2a$ for simplicity and noting that for probability distributions we estimate

$$\mu(V^c) = (\nu - \mu)(V) \leq \sup_{R \geq 0, z_0 \in \mathbb{C}} |(\mu - \nu)(B_R(z_0) \cap V)|$$

in the following proof.

Proof of Theorem 4.2. First, note that

$$\sup_{R \geq 0, z_0 \in \mathbb{C}} |(\mu - \nu)(B_R(z_0))| \leq \sup_{R \geq 0, z_0 \in \mathbb{C}} |(\mu - \nu)(B_R(z_0) \cap V)| + \mu(V^c) + \nu(V^c),$$

hence we have to estimate the first term. Fix some $a > 1$, let $\varphi \in \mathcal{C}^\infty(\mathbb{R})$ be nonnegative with $\operatorname{supp} \varphi \subseteq [-1,1]$ and $\int \varphi = 1$, and define $\varphi_a(\rho) = a\varphi(a\rho)$. For arbitrary $R > 0$

and $z_0 \in \mathbb{C}$ we mollify the indicator function appearing in $D(\mu,\nu)$ via the rotationally invariant approximation

$$\begin{aligned} f_1(z) :&= \left(\mathbb{1}_{(-\infty,R-1/a]} * \varphi_a\right)(|z-z_0|) \\ &\leq \mathbb{1}_{B_R(z_0)}(z) \\ &\leq \left(\mathbb{1}_{(-\infty,R+1/a]} * \varphi_a\right)(|z-z_0|) =: f_2(z), \end{aligned}$$

where we choose $f_1 \equiv 0$ if $R \leq 2/a$ for smoothness reasons. Furthermore we will approximate $\mathbb{1}_V$ by smooth functions h_1 from inside and by h_2 from outside, more precisely define

$$h_1(z) := \begin{cases} \left((\mathbb{1}_{[5\eta/2a,\infty)} * \varphi_{2a/\eta}) \cdot (\mathbb{1}_{(-\infty,K-1/a]} * \varphi_a)\right)(|z-z^*|) & \text{, if } \eta > 0, \\ \mathbb{1}_{(-\infty,K-1/a]} * \varphi_a(|z-z^*|) & \text{, if } \eta = 0, \end{cases}$$

$$h_2(z) := \begin{cases} \left((\mathbb{1}_{[3\eta/2a,\infty)} * \varphi_{2a/\eta}) \cdot (\mathbb{1}_{(-\infty,K+1/a]} * \varphi_a)\right)(|z-z^*|) & \text{, if } \eta > 0, \\ \mathbb{1}_{(-\infty,K+1/a]} * \varphi_a(|z-z^*|) & \text{, if } \eta = 0. \end{cases}$$

We apply $h_1 f_1 \leq \mathbb{1}_{B_R(z_0)\cap V}$ and integration by parts (in other words we use the definition of the distributional Poisson equation (2.27)) back and forth to obtain

$$\begin{aligned} \mu(B_R(z_0)\cap V) &\geq \int h_1 f_1 d\mu = -\frac{1}{2\pi}\int \Delta(h_1 f_1) U_\mu d\lambda\!\!\lambda \\ &= -\frac{1}{2\pi}\int \Delta(h_1 f_1)(U_\mu - U_\nu) d\lambda\!\!\lambda - \int (\mathbb{1}_{B_R(z_0)\cap V} - h_1 f_1) d\nu + \int \mathbb{1}_{B_R(z_0)\cap V} d\nu. \end{aligned}$$

A rough estimate of the error of approximation yields for the second term

$$\begin{aligned} \int (\mathbb{1}_{B_R(z_0)\cap V} - h_1 f_1) d\nu &\leq \nu\left(z \in V' : R - 2/a \leq |z-z_0| \leq R\right) + \nu(V' \setminus V) \\ &\leq 3 \sup_{R\geq 0, z_0 \in \mathbb{C}} \nu\left(z \in V' : R \leq |z-z_0| \leq R + (2 \vee \eta)/a\right) \\ &=: 3M_\nu(a). \end{aligned}$$

We use Hölder's inequality to estimate the first term, implying

$$(\mu-\nu)(B_R(z_0)\cap V) \geq -\frac{1}{2\pi}\left\|\Delta(h_1 f_1)\right\|_{L^q}\left\|U_\mu - U_\nu\right\|_{L^p} - 3M_\nu(a), \tag{4.3}$$

where $L^p = L^p(V')$, $L^q = L^q(V')$ (we omit V' in the sequel), $1/p + 1/q = 1$. Noting $\mu(B_R(z_0)\cap V) \leq \int h_2 f_2 d\mu$ and taking the same route for $h_2 f_2$ as for $h_1 f_1$, we obtain the same upper bound, i.e.

$$\begin{aligned}&-\frac{1}{2\pi}\left\|\Delta(h_1 f_1)\right\|_{L^q}\left\|U_\mu - U_\nu\right\|_{L^p} - 3M_\nu(a)\\ \leq&(\mu-\nu)(B_R(z_0)\cap V)\\ \leq&\frac{1}{2\pi}\left\|\Delta(h_2 f_2)\right\|_{L^q}\left\|U_\mu - U_\nu\right\|_{L^p} + 3M_\nu(a).\end{aligned} \tag{4.4}$$

Therefore it remains to control

$$\left\|\Delta(h_j f_j)\right\|_{L^q} \leq \left\|h_j \Delta f_j\right\|_{L^q} + 2\left\|\nabla h_j \cdot \nabla f_j\right\|_{L^q} + \left\|f_j \Delta h_j\right\|_{L^q}.$$

We see that the supports of all three functions are (unions of) annulus-segments, e.g. $V' \cap (B_{R+2/a}(z_0) \setminus B_R(z_0))$ for $h_2 \Delta f_2$, with length at most $2\pi(K+2/a)$ and the width equal to $(2 \vee \eta)/a$. Hence uniformly in $R > 0$ and $z_0 \in \mathbb{C}$, the size of the area of integration is bounded by $cK(2\vee\eta)/a$ and we arrive at

$$\left\|\Delta(h_j f_j)\right\|_{L^q} \leq (cK(2\vee\eta)/a)^{1/q}\left(\left\|\Delta f_j\right\|_{L^\infty} + 2\left\|\nabla h_j \cdot \nabla f_j\right\|_{L^\infty} + \left\|\Delta h_j\right\|_{L^\infty}\right).$$

With our choice of f_j and h_j, the radial derivatives become fairly simple, e.g.

$$\partial_r(f_2(z+z_0)) = \partial_r \int_{|z|-R-1/a}^{\infty} \varphi_a(\rho)d\rho = -a\varphi(a|z| - aR - 1).$$

Due to the rotational symmetry of f_2, we have $\left\|\nabla f_2\right\|_{L^\infty} \leq \left\|\varphi_a\right\|_{L^\infty} \lesssim a$ and again exploiting rotational symmetry it follows that the maximal curvature is attained in radial direction, i.e.

$$\left\|\Delta f_2\right\|_{L^\infty} = \sup_{r>0}\left|\partial_r^2 f_2(z_0+r)\right| = a^2\left\|\varphi'\right\|_{L^\infty}.$$

The same bounds hold for $j=1$ and h_j, where for h_j we replace a by $a \vee 2a/\eta$ if $\eta > 0$. Finally we conclude

$$\left\|\Delta h_j f_j\right\|_{L^q} \lesssim a^2(K/a)^{1/q} \lesssim K^{1-1/p}a^{1+1/p}, \tag{4.5}$$

where the implicit constant in the last $\lesssim$ depends on p,η and φ only. The claim now follows from taking the supremum over $R > 0$ and $z_0 \in \mathbb{C}$ in (4.4). □

In fact if we restrict ourselves to a certain region, we obtain a local smoothing inequality that makes it possible to invoke Proposition 2.18.

Corollary 4.3. *Let μ,ν be probability measures on $\mathbb{C}$ with logarithmic potentials U_μ, U_ν respectively, and fix some $z^* \in \mathbb{C}$, K,$\tau > 0$ and $1 \leq p \leq \infty$. There exists a constant*

$c > 0$ *such that for any* $a > 1 \wedge \tau^{-1}$

$$\sup_{B_R(z_0)\subseteq B_{K-\tau}(z^*)} |(\mu-\nu)(B_R(z_0))| \le ca^{1+1/p}\,\|U_\mu - U_\nu\|_{L^p(B_K(z^*))} + \sup_{R\ge 0, z_0\in\mathbb{C}} \nu\left(z\in B_K(z^*) : R \le |z-z_0| \le R+2/a\right).$$

Proof. Replace K by $K-\tau$, set $\eta = 0$ and note that the cutoff h in the previous proof is not necessary anymore. □

Although we only use this inequality for $K = 1, z^* = 0$, it reveals that the local distance of the measures only depends on the local distance of the logarithmic potentials. Girko's Hermitization Trick however transforms it into a highly nonlocal problem, taking the whole support (or spectrum of $\mu = \mu_n$) into account.

Moreover, the method of proof extends to the case of the classical Kolmogorov distance between 2-dimensional distribution functions.

Corollary 4.4. *Let* μ,ν *be probability measures on* $\mathbb{C}$ *with* $\operatorname{supp}\nu \subseteq [-K,K]^2$ *for some* $K > 0$*, let* U_μ, U_ν *be their logarithmic potentials and fix some* $\tau > 0$ *and* $1 \le p \le \infty$*. There exists a constant* $c > 0$ *such that for any* $a > 1$

$$\sup_{s,t\in\mathbb{R}} |(\mu-\nu)((-\infty,s]\times(-\infty,t])| \le ca^{1+1/p}\,\|U_\mu-U_\nu\|_{L^p([-K-\tau,K+\tau]^2)} + 3\sup_{s,t\in\mathbb{R}} \nu(([s,s+2/a]\times\mathbb{R})\cup(\mathbb{R}\times[t,t+2/a])).$$

Proof. We continue with the same notation as in the last proof and exploit the same ideas. Define now

$$\begin{aligned} f_1(z) := {} & \mathbb{1}_{(-\infty,s-1/a]} * \varphi_a(\mathrm{Re} z)\cdot \mathbb{1}_{(-\infty,t-1/a]} * \varphi_a(\mathrm{Im} z) \\ & \le \mathbb{1}_{(-\infty,s]\times(-\infty,t]}(z) \\ & \le \mathbb{1}_{(-\infty,s+1/a]} * \varphi_a(\mathrm{Re} z)\cdot \mathbb{1}_{(-\infty,t+1/a]} * \varphi_a(\mathrm{Im} z) =: f_2(z), \end{aligned}$$

and $h(z) = \mathbb{1}_{[-K-\tau/2,K+\tau/2]} * \varphi_{\tau/2}(\mathrm{Re} z)\cdot \mathbb{1}_{[-K-\tau/2,K+\tau/2]} * \varphi_{\tau/2}(\mathrm{Im} z)$. Here, if ν has compact support, we do not need h_1 in order to restrict ourselves to V. By similar arguments as above, e.g. $hf_1 \le \mathbb{1}_{(-\infty,s]\times(-\infty,t]}$, we obtain

$$(\mu-\nu)((-\infty,s]\times(-\infty,t]) \ge -\frac{1}{2\pi}\,\|\Delta(hf_1)\|_{L^q}\,\|U_\mu-U_\nu\|_{L^p} - M_\nu(a),$$

where now $M_\nu(a) = \sup_{s,t\in\mathbb{R}} \nu(([s,s+2/a]\times\mathbb{R})\cup(\mathbb{R}\times[t,t+2/a]))$ and we abbreviated $L^p = L^p([-K-\tau,K+\tau]^2)$, $L^q = L^q([-K-\tau,K+\tau]^2)$. For a short moment, consider

$$f_1^0(z) = \mathbb{1}_{[-K+1/a,K-1/a]} * \varphi_a(\mathrm{Re} z)\cdot \mathbb{1}_{[-K+1/a,K-1/a]} * \varphi_a(\mathrm{Im} z),$$

which analogously to the idea mentioned before Corollary 4.3 yields

$$1 - \mu([-K,K]^2) = (\nu - \mu)([-K,K]^2) \leq \frac{1}{2\pi} \left\|\Delta(f_1^0)\right\|_{L^q} \|U_\mu - U_\nu\|_{L^p} + 2M_\nu(a).$$

We conclude

$$\begin{aligned}
&-\frac{1}{2\pi} \|\Delta(hf_1)\|_{L^q} \|U_\mu - U_\nu\|_{L^p} - M_\nu(a) \\
\leq&(\mu - \nu)((-\infty,s] \times (-\infty,t]) \\
\leq&\frac{1}{2\pi} \left(\|\Delta(hf_2)\|_{L^q} + \left\|\Delta(f_1^0)\right\|_{L^q}\right) \|U_\mu - U_\nu\|_{L^p} + 3M_\nu(a).
\end{aligned}$$

Consequently it remains to derive similar estimates $\|\Delta(hf_j)\|_{L^q} \lesssim a^{1+1/p}$ using the same arguments as before. We omit the details here. □

4.2. The Circular Law

In this section, we will state different results on the rate of convergence to the Circular Law in Kolmogorov distance $D(\mu_n,\mu_\infty)$. Naively, we would like to apply the Smoothing inequality for $p = \infty$ and use the pointwise convergence of the logarithmic potentials from Subsection 2.2.3. Unfortunately, the inequality becomes meaningless, since the uniform term $\sup_{z\in B_K(0)} |U_n(z) - U_\infty(z)|$ explodes whenever an eigenvalue lies in $B_K(0)$, cf. Remark 2.16.

Thus, one cannot simply take $p = \infty$ in Theorem 4.1. Instead, we will choose a sufficiently large p and approximate the L^p norm by a random sum, which avoids the logarithmic singularities, in order to obtain the following result.

Theorem 4.5. *If Condition 2.14 (B) holds, then for every (small) $\varepsilon > 0$ and (large) $Q > 0$*

$$\mathbb{P}\left(D(\mu_n,\mu_\infty) \leq n^{-1/2+\varepsilon}\right) \geq 1 - n^{-Q}$$

holds for sufficiently large n.

Originally, we applied Proposition 2.15 instead of Proposition 2.17 in [GJ18], thus one may still find Condition 2.14 (A) therein. By virtue of Corollary 4.4, the following analogue for the Kolmogorov distance $d_K(\mu_n,\mu_\infty) = \sup_{s,t\in\mathbb{R}} |(\mu_n - \mu_\infty)((-\infty,s] \times (-\infty,t])|$ holds.

Theorem 4.6. *If Condition 2.14 (B) holds, then for every $\varepsilon,Q > 0$*

$$\mathbb{P}\left(d_K(\mu_n,\mu_\infty) \leq n^{-1/2+\varepsilon}\right) \geq 1 - n^{-Q}$$

holds for sufficiently large n.

Invoking Proposition 2.18, we prove a rate of convergence result weakening the conditions of the last statements at the cost of restricting to sets from the bulk.

Theorem 4.7. *If Condition 2.14 (C) holds, then for every* $\varepsilon,\tau,Q > 0$

$$\mathbb{P}\Big(\sup_{B_R(z_0)\subseteq B_{1-\tau}(0)} |(\mu_n - \mu_\infty)(B_R(z_0))| \leq n^{-1/2+\varepsilon}\Big) \geq 1 - n^{-Q}$$

holds for sufficiently large n.

Proof of Theorem 4.5. Without loss of generality $\varepsilon < 4$, we choose $p > 4/\varepsilon$ and apply Theorem 4.1 to $\mu = \mu_n$, $\nu = \mu_\infty$, $K = 1$ and $a = \sqrt{n}$,

$$D(\mu_n,\mu_\infty) \lesssim n^{1/2+\varepsilon/2} \left\| U_n - U_\infty \right\|_{L^p(B_{1+\tau}(0))} + \sup_{R\geq 0, z_0\in\mathbb{C}} \mu_\infty \left(R \leq |\cdot - z_0| \leq R + n^{-1/2}\right).$$

Since μ_∞ has bounded support and bounded density it is clear that the second term is of order $\mathcal{O}(n^{-1/2})$. In order to obtain a bound of the $L^p(B_{1+\tau}(0))$-norm of the log potentials from the pointwise estimate in Proposition 2.17, we adapt the Monte Carlo sampling method which was used in [TV15] (in a different form); we approximate

$$⨏ I(z)^p dz := \frac{1}{\pi(1+\tau)^2} \int_{B_{1+\tau}(0)} |U_n(z) - U_\infty(z)|^p \, dz \approx \frac{1}{N}\sum_{j=1}^N I(z_j)^p =: S_N,$$

where $(z_j)_{j=1,\dots,N}$ are independent random variables (also independent of X_{ij}) uniformly distributed on $B_{1+\tau}(0)$. More precisely we will show that for every $Q > 0$

$$\left| ⨏ I(z)^p dz - S_N \right|^{1/p} \lesssim n^{-1} \tag{4.6}$$

as well as

$$|S_N|^{1/p} \lesssim n^{-1+\varepsilon/2} \tag{4.7}$$

holds with probability at least $1 - n^{-Q}$ for some large n-dependent N. Assuming (4.6) and (4.7) are true, it holds

$$\begin{aligned}
&\mathbb{P}(D(\mu_n,\mu_\infty) \geq cn^{-1/2+\varepsilon}) \\
&\leq \mathbb{P}\left(cn^{1/2+\varepsilon/2}\Big(\Big| ⨏ I(z)^p dz - S_N \Big|^{1/p} + |S_N|^{1/p} \Big) + cn^{-1/2} \geq cn^{-1/2+\varepsilon} \right) \\
&\leq \mathbb{P}\left(\Big| ⨏ I(z)^p dz - S_N \Big|^{1/p} \geq cn^{-1+\varepsilon/2} \right) + \mathbb{P}\left(|S_N|^{1/p} \geq cn^{-1+\varepsilon/2} \right) \\
&\leq n^{-Q}
\end{aligned}$$

proving the claim.
Let's turn to the proof of (4.6). First, we restrict ourselves to the set of polynomially bounded eigenvalues. On the one hand the largest absolute value of eigenvalues $|\lambda|_{\max}$ is bounded by the largest singular value $s_{\max}$ and on the other hand for every $Q > 0$ we have

$$\mathbb{P}(s_{\max} \geq n^{(Q+1)/2}) \leq \frac{1}{n^{Q+1}} \mathbb{E} \left\| X/\sqrt{n} \right\|^2 \leq \frac{1}{n^{Q+2}} \sum_{ij}^{n} \mathbb{E} \left| X_{ij} \right|^2 \leq n^{-Q}, \tag{4.8}$$

similar to (2.48). We freeze the coefficients X_{ij} and use Chebyshev's inequality for the probability measure conditioned on X

$$\mathbb{P}\left(\left| S_N - ⨍ I(z)^p dz \right|^{1/p} \geq \frac{c}{n} \middle| X \right) \leq \frac{n^{2p}}{c^{2p}} \mathbb{V}\mathrm{ar}(S_N|X) \leq \frac{n^{2p}}{Nc^{2p}} \mathbb{V}\mathrm{ar}(I^p|X).$$

The variance of I^p given X can be estimated from above by

$$\mathbb{V}\mathrm{ar}(I^p|X) \leq \mathbb{E}(I^{2p}|X) \leq ⨍ \left| U_n(z) \right|^{2p} + \left| U_\infty(z) \right|^{2p} dz.$$

If we assume the eigenvalues $\lambda_1, \ldots, \lambda_n$ to be fixed and use Jensen's inequality, we may estimate

$$\begin{aligned} ⨍ \left| U_n(z) \right|^{2p} dz &\leq \frac{1}{n} \sum_{j=1}^{n} ⨍ \left| \log |\lambda_j - z| \right|^{2p} dz \\ &\leq \frac{c}{n} \sum_{j=1}^{n} \iint_{B_{1+\tau}(-\lambda_j)} r \left| \log r \right|^{2p} dr d\varphi \\ &\leq c_p(1 + \tau + |\lambda|_{\max}) \log^{2p}(1 + \tau + |\lambda|_{\max}) \\ &\leq c_p n^{(Q+1)/2} \log^{2p} n \end{aligned}$$

for some p-dependent constant c_p. Similarly we get $⨍ \left| U_\infty(z) \right|^{2p} dz = c_p$. Now choosing $N := n^{2p+3Q/2+1}$ and putting the estimates together we have shown

$$\begin{aligned} &\mathbb{P}\Big(\Big| ⨍ I(z)^p dz - S_N \Big|^{1/p} \geq cn^{-1} \Big) \\ &\leq \mathbb{E}\left(\mathbb{P}\left(\left\{ \Big| ⨍ I(z)^p dz - S_N \Big|^{1/p} \geq \frac{c}{n} \right\} \cap \left\{ |\lambda|_{\max} \leq n^{\frac{Q+1}{2}} \right\} \middle| X \right) \right) + n^{-Q} \\ &\leq cn^{-Q}. \end{aligned}$$

It remains to show (4.7). To this end we use Proposition 2.17 with an adjusted error probability stating

$$\mathbb{P}(I(z) \geq cn^{-1+\varepsilon/2}) \leq n^{-2p-5Q/2-1} \tag{4.9}$$

uniformly in $B_{1+\tau}(0)$. If $I(z_j) \leq n^{-1+\varepsilon/2}$ for all $j = 1, \dots, n$ then $|S_N|^{1/p} \leq n^{-1+\varepsilon/2}$ which implies

$$\begin{aligned}\mathbb{P}(|S_N|^{1/p} \geq n^{-1+\varepsilon/2}) &\leq \sum_{j=1}^{N} \mathbb{P}(I(z_j) \geq n^{-1+\varepsilon/2}) \\ &\leq cNn^{-2p-5Q/2-1} = cn^{-Q}.\end{aligned}$$

The proof is now complete, since these constants may be absorbed by the n^{-Q} (respectively n^{ε}) term for some slightly larger Q (respectively smaller ε). □

Analogously, Theorem 4.6 follows from Corollary 4.4 and Theorem 4.7 follows from Corollary 4.3. The details are exactly the same as above and we skip them. Moreover using the same techniques it is possible to show the following version of a Local Circular Law. Compared to [GNT19a] it improves the statement to hold with overwhelming probability but replaces the constant $\|\Delta f\|_{L^1}$ by $\|\Delta f\|_{L^q}$ and is stated for a single matrix, instead for a product of m many.

Corollary 4.8 (Local Circular Law). *Let $q > 1$, $z_0 \in B_{1+\tau^{-1}}(0)$ with $|1 - |z_0|| \geq \tau$, $f : \mathbb{C} \to \mathbb{R}_+$ be a bounded smooth function, which is compactly supported with $\|f'\|_{L^\infty} \leq n^{\widetilde{c}}$ for some constant $\widetilde{c} > 0$. Define the function $f_{z_0}(z) := n^{2s} f((z - z_0)n^s)$ which zooms into z_0 at speed $s \in (0,1/2)$. For any $Q > 0$ there exists a constant $c > 0$ such that*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=1}^{n} f_{z_0}(\lambda_j) - \int_{\mathbb{C}} f_{z_0}(z)d\mu_\infty(z)\right| \leq \frac{c\log^4 n}{n^{1-2s}} \|\Delta f\|_{L^q}\right) \geq 1 - n^{-Q}.$$

Recalling the discussion at the beginning of this section, z_0 and f are not allowed to depend on ω here.

Proof. As in the proof of Theorem 4.1, integration by parts yields

$$\frac{1}{n}\sum_{j=1}^{n} f_{z_0}(\lambda_j) - \int_{\mathbb{C}} f_{z_0}(z)d\mu_\infty(z) = -\frac{n^{2s}}{2\pi}\int_{\mathbb{C}} \Delta f(z)\left(U_n(z) - U_\infty(z)\right)dz.$$

After applying Hölder's inequality as was done in (4.3), it remains to show the estimate $\|U_n - U_\infty\|_{L^p} \lesssim \log^4 n/n$ which we already showed in the proof of Theorem 4.5 via Monte Carlo sampling and Proposition 2.18. □

4.3. Products of matrices with independent entries

Based on the ideas of the two previous sections, we shall generalize the results to products of matrices with independent entries. Recall that necessary notation and definitions have been fixed in Section 2.2 and we shall use them without repeating them.

Theorem 4.9. *If Condition 2.22 (D) holds, then for every $\tau,Q>0$ there exist a constant $c>0$ such that*

$$\mathbb{P}\Big(\sup_{B\subseteq B_{1-\tau}\cup B_{1+\tau}^c}|(\mu_n^m-\mu_\infty^m)(B)|\leq ch_m(n)\Big)\geq 1-n^{-Q},$$

where the asymptotic error is given by

$$h_m(n)=\begin{cases} n^{-1/2}\log^2 n & \text{for } m=1,\\ n^{-1/2}\log^3 n & \text{for } m=2,\\ n^{-2/(m+2)}\log^{8/(m+2)} n & \text{for } m\geq 3.\end{cases}$$

Theorem 3.2 provides the optimal rate of convergence, which is determined by Ginibre matrices. In the proof of Theorem 4.9 we will see that the m-dependent term is only visible for balls touching the origin. To make the statement more comprehensible when comparing with Theorem 3.2, we also state the following result.

Corollary 4.10. *If Condition 2.22 (D) holds, then for every $\tau,Q>0$ we have*

$$\mathbb{P}\Big(\sup_B|(\mu_n^m-\mu_\infty^m)(B)|\lesssim\frac{\log^2 n}{\sqrt{n}}\Big)\geq 1-n^{-Q},$$

where the supremum runs over all balls B such that $\partial B_R(z_0)\subseteq B_{1+\tau}^c\cup B_{1-\tau}\setminus B_\tau$ avoids the edge and the origin.

We already know from Theorem 3.2 that the optimal rate is given by $\mathcal{O}(1/\sqrt{n})$, hence Corollary 4.10 shows that this rate is also satisfied for matrices with independent entries, if edge and origin are avoided. Note that centered balls are allowed here. For $m=1$, Theorem 4.9 improves the rate of convergence from Theorem 4.7. Naively, we would like to choose $p\sim\log n$ in the Smoothing Inequality in order to remove the appearing factor n^ε. Unfortunately this is impossible, because the number N of points approximating the L^p norm depends on p, see (4.9).

We will circumvent the problem by replacing the Monte Carlo sampling by the following random grid approximation.

Lemma 4.11. *Let $\alpha,\beta>0$, S be a random variable uniformly distributed on $[0,1]^2$, and $\lambda_1,\ldots,\lambda_n\in\mathbb{C}$ be fixed with corresponding logarithmic potential U of the corresponding* *empirical distribution. Define the random grid* $A = 2\beta n^{-\alpha/2}(\mathbb{Z}^2 + S) \cap [-\beta,\beta]^2$ *enumerated by* $z_1, \dots, z_{\lceil n^\alpha \rceil}$. *For any function* $f \in \mathcal{C}_c^3(\mathbb{C})$ *with* $\operatorname{supp} f \subseteq (-\beta,\beta)^2$ *it holds*

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j) - \frac{-2\beta^2}{n^\alpha \pi}\sum_{i=1}^{\lceil n^\alpha \rceil} \Delta f(z_i) U(z_i) \qquad\qquad\qquad\qquad (4.10)\\ = \mathcal{O}(\|\nabla \Delta f\|_\infty n^{-\alpha/2}) + \mathcal{O}(\|\Delta f\|_\infty \log(n)^2 n^{-\alpha/4}).\end{aligned}$$

with overwhelming probability. More precisely if S *is chosen independently of the random matrix elements, then* (4.10) *and*

$$\sup_i |U(z_i)| = \mathcal{O}(\log^2 n)$$

hold on an event Ω_* *of probability* $1 - \mathcal{O}(n^{-\log n})$, *which does not depend on* f.

Note that (4.10) holds uniformly in $f \in \mathcal{C}_c^3(\mathbb{C})$, hence we could choose a function depending on the positions of λ_j. In order to make the statement more intuitive, suppose we replace the logarithmic potential U by a more regular function $U \in \mathcal{C}^1$. Then (4.10) is nothing but Riemann approximation of the integral

$$\int \Delta f(z) U(z) d\lambda(z) - \frac{(2\beta)^2}{n^\alpha}\sum_{i=1}^{\lceil n^\alpha \rceil} \Delta f(z_i) U(z_i) \lesssim (\|\nabla \Delta f\|_\infty + \|\Delta f\|_\infty)\, n^{-\alpha/2}. \qquad (4.11)$$

This follows directly from the mean value theorem, very similar to what we will do in (4.13) below.

In the Monte Carlo approximation used in [TV15] and [KOV18], the random points z_i are not ordered in a grid but drawn independently, thus variance bounds are of importance for improving bounds as (4.10). By using reference points or eigenvalue rigidity, the error estimates in [TV15] and [KOV18] are stronger by a factor of $1/n$ for the same number of points z_i. On the other hand, in order to control the singularities of U_n, one has to handle many random affects of all z_i, whereas in (4.10) only a single random shift affects all points z_i. Heuristically speaking, this leads to a higher probability than in previous approaches, so that the weaker error bound is negligible.

Proof of Lemma 4.11. Using the definition (2.27), in other words integration by parts, we find

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j) = \int f d\mu_n = -\frac{1}{2\pi}\int \Delta f(z) U(z) dz.$$

It suffices to show that with probability at least $1 - n^{-\log n - 1}$ we have

$$\begin{aligned}\int \Delta f(z) \log|\lambda - z|\, dz - \frac{4\beta^2}{n^\alpha} \sum_{i=1}^{\lceil n^\alpha \rceil} \Delta f(z_i) \log|\lambda - z_i| \\ = \mathcal{O}(\|\nabla \Delta f\|_\infty n^{-\alpha/2}) + \mathcal{O}(\|\Delta f\|_\infty \log(n)^2 n^{-\alpha/4}) \end{aligned} \tag{4.12}$$

for fixed $\lambda \in \mathbb{C}$, since the claim then follows from freezing the eigenvalues, i.e. conditioning on X, summation and the union bound. The event, where (4.12) holds, will be $\bigcap_{z_i \in A}\{|z_i - \lambda| > 2\beta n^{-(\log n + 1 + \alpha)/2}\}$ which fails if S is $n^{(-\log n - 1)/2}$ close to λ shifted by the grid. More precisely let $z^* \in 2\beta n^{-\alpha/2}\,\mathbb{Z}^2 \cap [-\beta,\beta]^2$ be the corner of this box with $\lambda \in z^* + [0,2\beta n^{-\alpha/2}]^2$, then

$$\begin{aligned}&\mathbb{P}\left(\exists i = 1, \ldots, \lceil n^\alpha \rceil : |z_i - \lambda| \leq 2\beta n^{-(\log n + 1 + \alpha)/2}\right) \\ &= \mathbb{P}\left(\operatorname{dist}\left(S, \frac{\lambda - z^*}{2\beta n^{-\alpha/2}}\right) \leq n^{(-\log n - 1)/2}\right) = \mathcal{O}(n^{-\log n - n}),\end{aligned}$$

where the distance in $[0,1]^2$ is measured according to the metric of the quotient space $\mathbb{T}^2$. From now on we will restrict ourselves to this event. Rewrite (4.12) as

$$\sum_{i=1}^{\lceil n^\alpha \rceil} \int_{K_i} \Delta f(z) \log|\lambda - z| - \Delta f(z_i) \log|\lambda - z_i|\, dz,$$

where we denoted the boxes with corner z_i by $K_i = z_i + [0,2\beta n^{-\alpha/2}]^2$. Adding and removing $\Delta f(z_i)\log|\lambda - z|$, we obtain one error of order

$$\sum_{i=1}^{\lceil n^\alpha \rceil} \int_{K_i} (\Delta f(z) - \Delta f(z_i)) \log|\lambda - z|\, dz = \mathcal{O}(\|\nabla \Delta f\|_\infty n^{-\alpha/2}), \tag{4.13}$$

where we used the mean value theorem and local integrability of log in $\mathbb{C}$. The second term can be bounded by

$$\begin{aligned}&\sum_{i=1}^{\lceil n^\alpha \rceil} \int_{K_i} \Delta f(z_i) \left(\log|\lambda - z| - \log|\lambda - z_i|\right) dz \\ &\quad \leq \|\Delta f\|_\infty \left(\sum_{i:|z_i - \lambda| \geq n^{-\frac{\alpha}{4}}} + \sum_{i:|z_i - \lambda| < n^{-\frac{\alpha}{4}}} \right) \int_{K_i} \left(\log|\lambda - z| - \log|\lambda - z_i|\right) dz.\end{aligned}$$

Applying the mean value theorem for log, yields a bound of order $\mathcal{O}(n^{-\alpha/4})$ for the first sum. The second sum can be bounded by performing the integration

$$\int_0^{2n^{-\alpha/4}} r\log r + \sup_{z_i\in A}\log|\lambda - z_i|\,dr = \mathcal{O}(n^{-\alpha/2}\log n) + \mathcal{O}(n^{-\alpha/4}\log(n)^2),$$

where we finally used the prescribed event. Putting all estimates together proves the first claim.

The bound for U follows from the choice of A and a trivial upper bound on $|\lambda|_{\max}$. On the one hand $|\lambda|_{\max}$ is bounded by the largest singular value $s_{\max}$ and on the other hand we have

$$\mathbb{P}(s_{\max}\geq n^{\log n}) \leq \frac{1}{n^{2\log n}}\,\mathbb{E}\left\|X/\sqrt{n}\right\|^2 \leq \frac{1}{n^{2\log n+1}}\sum_{ij}^{n}\mathbb{E}\,|X_{ij}|^2 \leq n^{-2\log n+1},$$

as before. Therefore on an event Ω_* with probability at least $1 - O(n^{-\log n})$, we have

$$\begin{aligned}\sup_{z_i\in A}|U(z_i)| &\leq \frac{1}{n}\sum_{j=1}^{n}\sup_{z_i\in A}\big|\log|\lambda_j - z_i|\,\big| \\ &\lesssim (\log n + 1 + \alpha)\log n + \log\big|\,|\lambda|_{\max} + 5\big| = \mathcal{O}(\log^2 n).\end{aligned}$$

□

The core of the proof of the local law for products of non-Hermitian matrices in [GNT19a] is the following identity. First, for any function $f\in\mathcal{C}_c^2(\mathbb{C})$ define $\widetilde{f}$ by $\widetilde{f}(z) = f(z^m)$ and note that $\int f d\mu_\infty^m = \int \widetilde{f} d\mu_\infty^1$, which follows from definition of μ_∞^m in (1.10). Using the distributional Poisson equation (2.27) and the representation of the eigenvalues of $\mathbf{W}$, we get

$$\int f d(\mu_n^m - \mu_\infty^m) = \frac{1}{nm}\sum_{j=1}^{nm}\widetilde{f}(\lambda_j(\mathbf{W})) - \int \widetilde{f} d\mu_\infty^1 = -\frac{1}{2\pi}\int \Delta\widetilde{f}\,(U_n - U_\infty)\,d\boldsymbol{\lambda}. \quad (4.14)$$

In [Nem17; GNT19a], locally shrinking functions f_{z_0} (cf. Theorem 2.19) have been considered and for fixed global f, Gaussian fluctuation has been proven in [KOV18]. As in the previous sections, we would like to uniformly approximate all indicator functions $\mathbb{1}_{B_R(z_0)}$ by smooth functions, replace the right hand side of (4.14) by a discrete random sum and use the pointwise estimate from Proposition 2.23. In contrast to what was done in the previous section, we cannot use the smoothing inequality, since there is no control of the difference of logarithmic potentials of μ_n^m, but of the matrix $\mathbf{W}$. Therefore we will use a direct approach, where the ideas of the Smoothing Inequality will appear again.

Proof of Theorem 4.9. First, note that we only need to consider $\tau < 1$. In order to restrict ourselves to a bounded region, say $V = B_7(0)$, we separate

$$\begin{aligned} D^\circ(\mu_n^m,\mu_\infty^m) &= \sup_{\substack{B_R(z_0)\subseteq\mathbb{C}\setminus B_{1+\tau} \\ \text{or } B_R(z_0)\subseteq B_{1-\tau}}} |(\mu_n^m - \mu_\infty^m)(B_R(z_0))| \\ &\leq \sup_{\substack{B_R(z_0)\subseteq\mathbb{C}\setminus B_{1+\tau} \\ \text{or } B_R(z_0)\subseteq B_{1-\tau}}} |(\mu_n^m - \mu_\infty^m)(B_R(z_0)\cap V)| + \mu_n^m(V^c). \end{aligned} \tag{4.15}$$

Fix some $\tau,R > 0$, $z_0 \in \mathbb{C}$ such that $B_R(z_0) \subseteq B_{1-\tau} \cup B_{1+\tau}^c$.

Let $\varphi \in \mathcal{C}^\infty(\mathbb{R})$ be nonnegative with $\operatorname{supp}\varphi \subseteq [-1,1]$ and $\int \varphi = 1$, and define $\varphi_a(\rho) = a\varphi(a\rho)$ for some $a > 1$ to be determined later. We define the cutoff to V^c by

$$f_0(z) = \left(\mathbb{1}_{(7-1/a,\infty)} * \varphi_a\right)(|z|) \geq \mathbb{1}_{V^c}(z).$$

Moreover, we mollify the indicator function appearing in (4.15) via the approximation

$$\begin{aligned} f_1(z) :&= \left(\mathbb{1}_{(-\infty,R-1/a]} * \varphi_a\right)(|z-z_0|)\cdot\left(\mathbb{1}_{(-\infty,7-1/a]} * \varphi_a\right)(|z|) \\ &\leq \mathbb{1}_{B_R(z_0)\cap V}(z) \\ &\leq \left(\mathbb{1}_{(-\infty,R+1/a]} * \varphi_a\right)(|z-z_0|)\cdot\left(\mathbb{1}_{(-\infty,7+1/a]} * \varphi_a\right)(|z|) =: f_2(z), \end{aligned}$$

where we choose $f_1 \equiv 0$ if $R \leq 2/a$ for smoothness reasons.

We apply $f_1 \leq \mathbb{1}_{B_R(z_0)\cap V}$ and integration by parts to $\widetilde{f}_1 : z \mapsto f_1(z^m)$ as was forecast in (4.14) to obtain

$$\begin{aligned} &\mu_n^m(B_R(z_0)\cap V) \geq \int f_1 d\mu_n^m = -\frac{1}{2\pi}\int(\Delta\widetilde{f}_1)U_n d\lambda \\ &= -\frac{1}{2\pi}\int \Delta\widetilde{f}_1(U_n - U_\infty)d\lambda - \int(\mathbb{1}_{B_R(z_0)\cap V} - f_1)d\mu_\infty^m + \int \mathbb{1}_{B_R(z_0)\cap V}d\mu_\infty^m. \end{aligned} \tag{4.16}$$

Analogous upper bounds hold for f_0 and f_2. A rough estimate of the error of approximation yields for the second term

$$\int(\mathbb{1}_{B_R(z_0)\cap V} - f_1)d\mu_\infty^m \leq \mu_\infty^m\left(z\in\mathbb{C} : R - 2/a \leq |z-z_0| \leq R\right). \tag{4.17}$$

Due to the radial monotonicity of μ_∞^m's density, this value increases by bending two halves of the given annulus of width $2/a$ into two straight rectangles $[-1,1]\times[-4/a,4/a]$. The density is bounded in the case of $m = 1$ or for Corollary 4.10, where the origin is avoided, and hence the term in (4.17) is of order $\mathcal{O}(1/a)$. In general we can bound it by

$$\begin{aligned}
\mu_\infty^m([-1,1]\times[-4/a,4/a]) &\leq \frac{1}{\pi m}\int_{-1}^{1}\int_{-4/a}^{4/a}(x^2+y^2)^{1/m-1}dxdy \\
&\leq \frac{4}{\pi m a}\int_{2/a}^{1} x^{2/m-2}dx + \frac{2}{m}\int_0^{4/a} r^{2/m-1}dr \\
&= \frac{2}{\pi(m-2)}\left((2/a)^{2/m}-2/a\right)+(4/a)^{2/m} \lesssim a^{-2/m},
\end{aligned}$$

where the equation only holds for $m>2$. For $m=2$ we get

$$\begin{aligned}
&\mu_\infty^2([-1,1]\times[-4/a,4/a]) \leq \frac{1}{2\pi}\int_{-1}^{1}\int_{-4/a}^{4/a}(x^2+y^2)^{1/2}dxdy \\
&= \frac{8}{\pi a}\log\left(\sqrt{1+16/a^2}+1\right)+\frac{2}{\pi}\log\left(\sqrt{1+16/a^2}+4/a\right)+\frac{8}{\pi a}\log(a/4) \\
&\sim a^{-1}\log a
\end{aligned}$$

and we see that the log-term appears naturally. Define the error function

$$\widetilde{h}_m(a) = \begin{cases} \mathcal{O}(a^{-1}) & \text{for } m=1 \text{ or Corollary 4.10,} \\ \mathcal{O}(a^{-1}\log a) & \text{for } m=2, \\ \mathcal{O}(a^{-2/m}) & \text{for } m\geq 3. \end{cases} \tag{4.18}$$

Let us continue to estimate the first term of (4.16) by using our random grid approximation Lemma 4.11. Let $\beta=7$ and S be a random variable, independent of $\mathbf{X}$ and uniformly distributed on $[0,1]^2$. Conditioned on $\mathbf{X}$, we have with overwhelming probability

$$\begin{aligned}
&\int \Delta\widetilde{f}_1(z)U_n(z)d\lambda(z) - \frac{(2\beta)^2}{n^\alpha}\sum_{i=1}^{\lceil n^\alpha\rceil}\Delta\widetilde{f}(z_i)U_n(z_i) \\
&= \mathcal{O}(\|\nabla\Delta\widetilde{f}_1\|_\infty n^{-\alpha/2}) + \mathcal{O}(\|\Delta\widetilde{f}_1\|_\infty \log(n)^2 n^{-\alpha/4}).
\end{aligned}$$

Due to our explicit choice of functions f_1 and f_2 as product of shifted radial symmetric functions, the partial derivatives become fairly simple. Each derivative that hits one of the φ_a produces a factor of a, more precisely any k-th directional derivative satisfies $\|\partial^{(k)}f_1(z)\|_\infty \lesssim a^k$. This estimate, again, is independent on the choice of the ball $B_R(z_0)$.

Together with the Riemann approximation (4.11), we conclude that for any matrix

$\mathbf{X}$ we have with overwhelming probability

$$\sup_B \left| \int \Delta \widetilde{f}_1(z)(U_n(z) - U_\infty(z))d\lambda\!\!\lambda(z) - \frac{(2\beta)^2}{n^\alpha} \sum_{i=1}^{\lceil n^\alpha \rceil} \Delta \widetilde{f}_1(z_i) \left(U_n(z_i) - U_\infty(z_i)\right) \right|$$
$$= \mathcal{O}(a^3 n^{-\alpha/2}) + \mathcal{O}(a^2 \log(n)^2 n^{-\alpha/4}),$$

where the supremum runs over all choices of $B \subseteq B_{1-\tau} \cup B_{1+\tau}^c$. Since we will always choose $a \lesssim n$ (actually we will make it even smaller, cf. (4.19)), it is possible to freely choose $\alpha > 0$ sufficiently big such that the error is arbitrarily small. For instance $\alpha = 13$ is more than enough to ensure that the error is of order $\mathcal{O}(n^{-1})$. It should be emphasized that still no randomness of $\mathbf{X}$ has been used and the only randomness is the shifted grid. Combining the previous steps yield

$$(\mu_n^m - \mu_\infty^m)(B \cap V) \geq -\frac{(2\beta)^2}{2\pi n^\alpha} \sum_{i=1}^{\lceil n^\alpha \rceil} \Delta \widetilde{f}_1(z_i)\Big(U_n(z_i) - U_\infty(z_i)\Big) - \widetilde{h}_m(a) - \mathcal{O}(n^{-1})$$

uniformly in $B \subseteq B_{1-\tau} \cup B_{1+\tau}^c$ with overwhelming probability. Noting $\mu_n^m(B_R(z_0) \cap V) \leq \int f_2 d\mu_n^m$ and taking the same route for f_2 as for f_1, we obtain the same upper bound

$$(\mu_n^m - \mu_\infty^m)(B \cap V) \leq -\frac{(2\beta)^2}{2\pi n^\alpha} \sum_{i=1}^{\lceil n^\alpha \rceil} \Delta \widetilde{f}_2(z_i)\Big(U_n(z_i) - U_\infty(z_i)\Big) + \widetilde{h}_m(a) + \mathcal{O}(n^{-1}).$$

In the same way, the bound holds for $\mu_n^m(V^c)$ as well. Finally we use the randomness of $\mathbf{X}$ by applying Proposition 2.23. Conditioning on S, i.e. freezing the lattice points z_i, we obtain for any $Q > 0$

$$\mathbb{P}\left(|U_n(z_i) - U_\infty(z_i)| \geq c\frac{\log^4(n)}{n} \Big| S \right) \leq n^{-Q-\alpha}$$

for each $i = 1, \dots, \lceil n^\alpha \rceil$. By the union bound this implies that with probability at least $1 - n^{-Q}$ the logarithmic potentials concentrate like $U_n(z_i) - U_\infty(z_i) = \mathcal{O}(\log^4 n/n)$ simultaneously at all lattice points. Therefore, for $k = 0,1,2$,

$$\frac{(2\beta)^2}{2\pi n^\alpha} \sum_{i=1}^{\lceil n^\alpha \rceil} \Delta \widetilde{f}_k(z_i)\Big(U_n(z_i) - U_\infty(z_i)\Big) \lesssim \frac{(2\beta)^2 \log^4 n}{n^{1+\alpha}} \sum_{i=1}^{\lceil n^\alpha \rceil} |\Delta \widetilde{f}_k(z_i)|$$
$$= \frac{\log^4 n}{n} \big\|\Delta \widetilde{f}_k\big\|_{L^1} + \mathcal{O}\Big(a^3 \frac{\log^4}{n^{\alpha/2+1}}\Big)$$

where the integral of the a^3-Lipschitz function $|\Delta \widetilde{f}_k|$ has been approximated by its

Riemann sum. Write $\Delta = 4\bar{\partial}\partial$ in terms of the Wirtinger derivatives $\partial = \frac{1}{2}(\partial_x - i\partial_y)$ and $\bar{\partial} = \frac{1}{2}(\partial_x + i\partial_y)$. Since $g(z) = z^m$ is holomorphic, i.e. $\bar{\partial}g = 0$, we obtain by applying the chain rule and changing variables from z to $g(z)$

$$\|\Delta \widetilde{f}_k\|_{L^1} = \|4\bar{\partial}\partial(f_k \circ g)\|_{L^1} = 4\|(\bar{\partial}\partial f_k) \circ g \cdot \bar{\partial}\bar{g} \cdot \partial g)\|_{L^1} = \|\Delta f_k\|_{L^1}.$$

Since $\Delta f_k \lesssim a^2$ and has support on an area of order a^{-1}, we have

$$\sup_B \|\Delta f_k\|_{L^1} \lesssim a.$$

So overall we have proven that for all Q there exists a constant $c > 0$ such that with probability $1 - n^{-Q}$ we have

$$\sup_{B \subseteq B_{1-\tau} \cup B_{1+\tau}^c} |(\mu_n^m - \mu_\infty^m)(B)| \leq ca \frac{\log^4 n}{n} + \widetilde{h}_m(a) + \mathcal{O}(n^{-1}). \tag{4.19}$$

Optimizing in a yields $a = \sqrt{n}/\log^2 n$ for Corollary 4.10 and $m = 1$, as well as $h_2(n) = \log^3 n/\sqrt{n}$. The asymptotic $h_m(n)$ for higher m follows from choosing $a = n^{m/m+2} \log^{-4m/(m+2)} n$.

□

In the proof we have seen that the maximal error for the limiting distribution μ_∞^m is by balls, which touch the origin (technically these balls of growing size are not even admissible here). This yields the non-optimal rate in Theorem 4.9 if we do not exclude the origin. Having Theorem 3.2 in mind however, we expect the maximizing ball to appear roughly at $B_1(0)$, where the error would be optimal again.

Chapter 5

Applications and related models

In the previous section, we developed an approach to derive rate of convergence results in Kolmogorov distance d_K and D from the concentration of logarithmic potentials. We may use the same ideas in order to obtain rates of convergence for other models, for instance the empirical measure of i.i.d. variables or the inhomogeneous Circular Law. Instead of repeating the same proofs with more cumbersome definitions, we shall stick to the following example.

5.1. Random polynomials

The Smoothing Inequality can also be applied to the empirical distribution of roots of random polynomials in order to obtain the same rate of convergence to the Circular Law as for random matrices. In the previous sections we considered the roots of the characteristic polynomial of a random matrix, where the coefficients of the polynomial exhibit specific dependencies. We begin by replacing the independence condition on the matrix entries by independent coefficients in the polynomial.

Definition 5.1. Given $n \in \mathbb{N}$ many complex numbers $c_0, \dots, c_n$ and i.i.d. centered complex random variables $\xi_0, \dots, \xi_n$ with $\mathbb{E}\,|\xi_0|^2 = 1$ and $\mathbb{E}\,|\xi_0|^{2+\delta} < \infty$ for some $\delta > 0$, we define the random polynomial $f_n : \mathbb{C} \to \mathbb{C}$ by

$$f_n(z) = \sum_{k=0}^{n} c_k \xi_k z^k.$$

In particular we will work with so called *Weyl (or Flat) polynomials* f_n^W corresponding to $c_k = \sqrt{n^k/k!}$. By analogy to the Introduction, we associate to a random polynomial f_n its multiset of zeros $\Lambda := \{\lambda \in \mathbb{C} : f_n(\lambda) = 0\}$ taking their multiplicities into account and its empirical measure given by

$$\mu_{f_n} = \frac{1}{n} \sum_{\lambda \in \Lambda} \delta_\lambda.$$

It should be remarked that μ_{f_n} is not necessarily normalized, since a random polynomial may have degree $\deg(f_n) < n$. Unsurprisingly this does not affect the large n limit, since $n - \deg(f_n) \in \mathcal{O}(1)$ $\mathbb{P}$-a.s. and as in [IZ13], we may always assume $\mathbb{P}(\xi_0 = 0) = 0$, since otherwise we may restrict ourselves to $\{\deg(f_n) = k, \min\{j \leq n : \xi_j \neq 0\} = l\}$.

The Circular Law for the empirical measure μ_n^W of the roots of Weyl polynomials has been established in [KZ14a] by Kabluchko and Zaporozhets, see also [FH99] for the Gaussian case, stating

$$\mu_n^W \Rightarrow \mu_\infty \quad \mathbb{P}\text{-a.s..}$$

Note that their result holds for much more general random analytic functions and under the much weaker condition of the coefficients having finite logarithmic moments $\mathbb{E}\log(1 + |\xi_0|) < \infty$.

We aim to quantify this result by showing a rate of convergence of order $n^{-1/2+\varepsilon}$ by using results about logarithmic potentials. Since local Universality for certain random polynomials has been proven in by Tao and Vu [TV14] using concentration of logarithmic magnitudes $\log|f_n|$, we can apply the same methods as before. We denote $U_n^W = -\frac{1}{n}\log|f_n|$ and rephrase [TV14, Lemma 12.1]: Under the conditions mentioned above and for every ε,δ,τ,$Q > 0$ there exists a constant $c > 0$ such that

$$\mathbb{P}\left(\left|U_n^W(z) - U_\infty(z) + 1/2\right| \leq cn^{-(1-\varepsilon)}\right) \geq 1 - n^{-Q} \tag{5.1}$$

holds for any $n^{-1/2+\delta} \leq |z| \leq 1 + \tau$. The origin has to be avoided, since the distribution of $U_n^W(0) = -\frac{1}{n}\log|\xi_0|$ around 0 stays arbitrary. In particular, the bound (5.1) will not hold in $z = 0$ if $\mathbb{P}(\xi_0 = 0) > 0$. Due a rough bound on the largest root, we still need a technical assumption on the concentration of ξ_0 near $z = 0$ in the following rate of convergence result which we deduce from Theorem 4.2.

Theorem 5.2. *If $\mathbb{E}|1/\xi_0|^\delta < \infty$ for some $\delta > 0$, then for every ε,$Q > 0$ and sufficiently large n we have*

$$\mathbb{P}(D(\mu_n^W,\mu_\infty) \leq n^{-1/2+\varepsilon}) \geq 1 - n^{-Q}.$$

It seems likely that other polynomials, like elliptic polynomials, admit the same asymptotics to their corresponding limit root distributions, but we focus on Circular Laws here.

Note that the analogue of products of random matrices in the world of random polynomials is given by the polynomial with weights $c_k = (n^k/k!)^{m/2}$, which are the m-th powers of the weights of a Weyl polynomial. For the corresponding random polynomial $f_n^{(m)}$, Kabluchko and Zaporozhets [KZ14a] showed that $\mu_{f_n^{(m)}} \Rightarrow \mu_\infty^m$. However, $f_n^{(m)}$ violates the non-clustering property of [TV14], hence there is no analogue of (5.1) available.

The proof of Theorem 5.2 does not differ much from those in the previous chapter.

Proof. The claim is trivial for $\varepsilon \geq 1/2$, hence we fix $\varepsilon < 1/2$. As above, we choose

$p > (1-\varepsilon)/\varepsilon$ large enough and apply Theorem 4.2 to $\mu = \mu_n^W, \nu = \mu_\infty, K = 2, \eta = 1$, $z^* = 0$ and $a = n^{1/2-\varepsilon}$, and obtain

$$\begin{aligned} D(\mu_n^W,\mu_\infty) \lesssim{}& n^{1/2} \left\| U_n^W - U_\infty + 1/2 \right\|_{L^p(B_3(0)\setminus B_{n^{-1/2+\varepsilon}}(0))} \\ &+ \mu_n^W(B_{2n^{-1/2+\varepsilon}}(0)) + \mu_n^W(B_2(0)^c) \\ &+ \mu_\infty(B_{2n^{-1/2+\varepsilon}}(0)) + \mu_\infty(B_2(0)^c) \\ &+ \sup_{R\geq 0, z_0\in\mathbb{C}} \mu_\infty\left(R \leq |\cdot - z_0| \leq R + 2n^{-1/2+\varepsilon}\right). \end{aligned}$$

Let us consider each term starting with the last one. Obviously the last term is of order $n^{-1/2+\varepsilon}$ and the third line equals $4\pi n^{-1+2\varepsilon}$. From an already existing (non-uniform) Local Circular Law for random polynomials, see [TV14, Formula (87)], it follows that with overwhelming probability the second line of our estimation can also be bounded by $cn^{-1+2\varepsilon}$. Therefore it remains to control the L^p distance of the logarithmic potentials. The application of Monte Carlo sampling and the pointwise control of the logarithmic potentials from (5.1) remains unchanged. The only notable difference to the proof of Theorem 4.5 is the restriction to polynomially bounded moduli of the zeros. From Rouché's Theorem, we deduce an upper bound for the largest root

$$|\lambda|_{\max} \leq 1 + \frac{\max\{c_0\,|\xi_0|\,, \ldots, c_{n-1}\,|\xi_{n-1}|\}}{c_n\,|\xi_n|}$$

of any polynomial. Hence for any $Q > 0$ we have

$$\begin{aligned} \mathbb{P}(|\lambda|_{\max} \geq n^{(Q+1)/\delta}) &\leq \mathbb{P}\left(\frac{\max\{|\xi_0|\,, \ldots, |\xi_{n-1}|\}}{|\xi_n|} \gtrsim n^{(Q+1)/\delta}\right) \\ &\leq (n-1)\,\mathbb{P}(|\xi_0| \gtrsim n^{(Q+1)/\delta}\,|\xi_n|) \\ &\lesssim \frac{n-1}{n^{Q+1}}\,\mathbb{E}\,|\xi_0|^\delta\,\mathbb{E}\,|1/\xi_0|^\delta \lesssim n^{-Q}, \end{aligned}$$

which replaces (4.8) and the proof is finished. $\square$

5.2. Rate of convergence of the spectral radius

The spectral radius

$$|\lambda|_{\max} := \max\left\{\left|\lambda_j(X/\sqrt{n})\right| : 1 \leq j \leq n\right\}$$

converges $\mathbb{P}$-a.s. to 1 as $n \to \infty$, see [BS10]. Here, we would like to study the rate of convergence. If $\widetilde{X}$ is drawn from the complex Ginibre ensemble, then the distribution of the properly rescaled spectral radius converges to a Gumbel distribution, because the radii form an independent family of random variables with light tails. More precisely, it was initiated by Kostlan [Kos92] and reestablished by Rider [Rid03] that

$$\lim_{n\to\infty} \mathbb{P}\left(\sqrt{4n\gamma_n}(|\lambda|_{\max} - 1 - \sqrt{\gamma_n/4n}) \le x\right) = \exp(-e^{-x}), \tag{5.2}$$

where $\gamma_n = \log\left(\frac{n}{2\pi\log^2 n}\right)$.[1] Thus, we expect the spectral radius $|\lambda|_{\max}$ to converge to 1 at a rate $\sqrt{\log(n)/n}$. From the rate of convergence of the ESD's given by Theorem 4.5, we immediately obtain a lower bound on the spectral radius.

Lemma 5.3. *If Condition 2.14 (B) holds, then for any $b < 1/2$, $Q > 0$ we have*

$$\mathbb{P}\left(|\lambda|_{\max} \ge 1 - n^{-b}\right) \ge 1 - n^{-Q}$$

for n sufficiently large.

Proof. From Theorem 4.5 with $b = 1/2 - \varepsilon$ it follows

$$\begin{aligned}\mathbb{P}(|\lambda|_{\max} < 1 - n^{-b}) &= \mathbb{P}(\mu_n(B_{1-n^{-b}}(0)) = 1)\\ &\le \mathbb{P}(|\mu_n(B_{1-n^{-b}}(0)) - \mu_\infty(B_{1-n^{-b}}(0))| \ge n^{-b})\\ &\le \mathbb{P}(D(\mu_n,\mu_\infty) \ge n^{-b}) \le n^{-Q}.\end{aligned}$$

□

Naturally, a lower bound on $|\lambda|_{\max}$ should be easier to obtain than an upper bound. For the spectral radius to be too small, *all* eigenvalues have to be too small, which is an extraordinarily improbable event. For the spectral radius to be too big, it is enough that a single eigenvalue is an outlier. We will prove the following two-sided estimate on $|\lambda|_{\max}$ assuming an extra moment matching condition.

Theorem 5.4. *If X is a non-Hermitian complex random matrix satisfying Condition 2.14 (A) and additionally matches Gaussian moments up to the sixth order, then there exist constants b, $c > 0$ such that*

$$\mathbb{P}(||\lambda|_{\max} - 1| \le cn^{-b}) \ge 1 - o(n^{-b}). \tag{5.3}$$

Under the sole assumption of bounded $2 + \varepsilon$ moments of the matrix entries, in [BCCT18] it has been shown that for any $\delta > 0$

$$\mathbb{P}(|\lambda|_{\max} \ge 1 + \delta) \lesssim 1/(\log^2 n),$$

see also [Nem18] for products of non-Hermitian random matrices. From the above perspective one may think of (5.3) as a rate of convergence of the spectral radius valid

1 Gumbel fluctuations were also obtained for more general Coulomb gases by Chafaï and Péché [CP14].

with high probability. Having (5.2) in mind, one would expect b to be close to $1/2$. In fact, this has been proven very recently by Alt, Erdős and Krüger [AEK19]. Though, the authors considered a more general setting of inhomogeneous variances of the entries, we will only state the homogeneous case for comparison.

Theorem 5.5 ([AEK19])**.** *Let X be a non-Hermitian random matrix having independent entries X_{ij} with mean zero, variance $\mathbb{E}|X_{ij}|^2 = 1$ and existing moments $\mathbb{E}|X_{ij}|^p < \infty$ of all orders $p \in \mathbb{N}$. For all $b < 1/2$, $Q > 0$, we have for n sufficiently large*

$$\mathbb{P}\left(|\lambda|_{\max} \leq 1 + n^{-b}\right) \geq 1 - n^{-Q}$$

and if additionally the entries have bounded densities, then also the lower bound holds

$$\mathbb{P}\left(|\lambda|_{\max} \geq 1 - n^{-b}\right) \geq 1 - n^{-Q}.$$

Compared to Theorem 5.4, the previous Theorem 5.5 holds in a more general setting, without any moment matching hypothesis and provides nearly optimal rate of convergence $\mathcal{O}(n^{-1/2+\varepsilon})$. For certain discrete matrices, where the density assumption of [AEK19] is violated, our result still holds. Since we will use entirely different methods, some ideas may still provide new insight for further problems.

We will deduce Theorem 5.4 as a direct corollary from the following result. In the sequel, we will denote by $\mu_n = \frac{1}{n}\sum_{j=1}^n \delta_{\lambda_j(X/\sqrt{n})}$ and $\widetilde{\mu}_n$ the ESD's of two non-Hermitian complex random matrices X,$\widetilde{X}$, respectively and by $U_n(z) = -\frac{1}{n}\log|\det(X/\sqrt{n} - z)|$ and $\widetilde{U}_n$ the corresponding logarithmic potentials.

Theorem 5.6. *Let X,$\widetilde{X}$ be two independent non-Hermitian complex random matrices X,$\widetilde{X}$ satisfying Condition 2.14 (A). If X and $\widetilde{X}$ match moments up to sixth order, then there exists a constant $b > 0$ such that*

$$\mathbb{P}(|\lambda|_{\max} > 1 + 2n^{-b}) \leq n\,\mathbb{E}\left(\widetilde{\mu}_n(B^c_{1+n^{-b}}(0))\right) + \mathcal{O}(n^{-b}).$$

The naive idea of the proof is as follows. First, we bound the left hand side by the expectation of the linear statistic $n\int f d\mu_n$ of a smooth cutoff f on $B^c_{1+2n^{-b}}(0)$, then we rewrite the linear statistic as a complex integral of logarithmic potentials using (2.27). The integral involving the logarithmic potentials is going to be approximated by a sum with n^α many points. Each summand of the approximation can then be replaced by the one corresponding to $\widetilde{X}$. Similar to Lindeberg exchange method we swap the distribution of the entries one by one. This swapping mechanism is done in [TV15], from which we cite the following Theorem.

Theorem 5.7 ([TV15])**.** *Let $c_0 > 0$ be some sufficiently small absolute constant. Fix some $C > 0, 1 \leq k \leq n^{c_0}$ and points $z_1, \ldots, z_k \in B_C(0)$. Let X,$\widetilde{X}$ be two independent random matrices satisfying Condition 2.14 (A) and which match each others moments* *to fourth order. There exists a constant $b > 0$ such that for any smooth bounded function $G : \mathbb{R}^k \to \mathbb{C}$ obeying the derivative bounds $|\nabla^j G(x_1, \dots ,x_k)| = \mathcal{O}(n^{c_0})$ for all $j = 0, \dots ,5$ we have*

$$\mathbb{E}\, G(nU_n(z_1), \dots ,nU_n(z_k)) = \mathbb{E}\, G(n\widetilde{U}_n(z_1), \dots ,n\widetilde{U}_n(z_k)) + \mathcal{O}(n^{-b}).$$

Since we would like to compare $\mathbb{E}\, n \int f d\mu_n$ to $\mathbb{E}\, n \int f d\widetilde{\mu}_n$ (including the factor n), we will need the swapping mechanism to have an improved error bound $\mathcal{O}(n^{-1-b})$. As already indicated in [TV15, Remark 47], the exponent of the error in Theorem 5.7 may be slightly improved by imposing additional matching moment hypothesis. Unfortunately, only limited improvements are possible, since the event, where the resolvent swapping fails (see Proposition 5.9), has probability n^{-b} which is linked to smallest singular value problems that cannot be improved without introducing additional restrictions.[1]

Restricted to the correct event, we are able to prove the following variant that makes it possible to detect a better rate of approximation depending directly on higher imposed matching moment hypothesis.

Lemma 5.8. *Let $X,\widetilde{X}$ be two independent random matrices satisfying Condition 2.14 (A) and which match each others moments up to order $d \in \mathbb{N}$. For any $C > 0$ there exists a constant $b > 0$ such that for all $z \in B_C(0)$ the following holds. For any smooth bounded function $G : \mathbb{R} \to \mathbb{C}$ obeying the derivative bounds $G^{(j)}(x) = \mathcal{O}(1)$ for all $j = 1, \dots ,d+1$ there exist $[0,1]$-valued random variables $\chi, \widetilde{\chi}$ only dependent on z and $X,\widetilde{X}$ such that we have*

$$\mathbb{E}\left(G(nU_n(z))\chi\right) = \mathbb{E}\left(G(n\widetilde{U}_n(z))\widetilde{\chi}\right) + \|G\|_\infty\, \mathcal{O}(n^{-(d-4)/2-b}).$$

Moreover we have

$$\mathbb{P}(\chi = 1) = 1 - \mathcal{O}(n^{-b}) = \mathbb{P}(\widetilde{\chi} = 1). \tag{5.4}$$

By the random variable χ we denote (under slight abuse of notation) a smooth cutoff function χ to the region $|x| \le n^{3c_0}$ that equals 1 for $|x| \le n^{3c_0}/2$ evaluated at $\mathrm{Im}(m_n(z,in^{-1-4c_0}))$, see [TV15].

Let us provide some details on how Lemma 5.8 is used to prove Theorem 5.4. As mentioned above, we need to approximate $\int n\Delta f U_n d\lambda\!\!\lambda$ by a sum of n^α points. For local functions f zooming into a fixed point $z_0 \in \mathbb{C}$ at rate $n^{-1/2}$ this investigation coincides with the Local Law, cf. Theorem 2.19. On the other hand for a fixed (global) function f

1 For $|z| > 1$, the limiting singular value distribution has support away from the origin, therefore one may expect a better lower bound on the smallest singular value for z_i in our given region. Here, [AEK19] obtained an improved estimate that might improve the constant b to $1/2 - \varepsilon$ in our approach as well.

similar methods have been used in [KOV18] to prove that the limiting distributions of linear statistics are Gaussian. In both cases $\|\Delta f\|_{L^1}$ are of order $\mathcal{O}(1)$, while in our case we have to deal with a growing term $\|n\Delta f\|_{L^1} \sim n^{1+b}$. Basically, this will be bounded by (5.4), Proposition 2.18 and the fact that $\mu_\infty(\operatorname{supp}\Delta f) = 0$, see (5.10).

For local or fixed functions f and Monte Carlo approximation[1] the amount n^α of approximating points z_i is small ($\alpha \ll 1/2$) and hence the approach of [TV15; KOV18] works fine in their cases. However if one considers a global *and* approximating function f, as we wish, much more than $n^{1/2}$ points might be needed. Thus, we will use the approximation by a randomly shifted grid, see Lemma 4.11, where we can control the randomness much better than for Monte Carlo approximation.

In order to avoid the logarithmic singularities of U_n, [TV15] introduces a smooth bounded cutoff function $G : \mathbb{R}^{n^\alpha} \to \mathbb{C}$ depending on all approximating points z_i. If the amount $n^\alpha = n^{c_0}$ of approximating points z_i is small ($\alpha \ll 1/2$), then the approach of [TV15] works without problems: Resolvent swapping with Proposition 5.9 below and Taylor approximation of G into all directions of z_i. But the Taylor approximation of G becomes unfavorable if α is close to $1/2$, cf. (5.14) below, hence we cannot allow G to depend on all points. The random grid approximation exploits the linear structure of G by pulling the approximating sum out of the expectation, hence making it only depend on a single logarithmic potential's value, see equation (5.9). Thus, the Taylor approximation is not a problem anymore and we are allowed to use as many approximating points n^α as we wish.

Proof of Theorem 5.6. First, we restrict ourselves to the event of bounded eigenvalues of X and $\widetilde{X}$. In our situation we have

$$\mathbb{P}(|\lambda|_{\max} > C) \leq \mathbb{P}(\|X\| > C\sqrt{n}) = \mathcal{O}(n^{-b})$$

for some $b,C > 0$, see [RV08, §2.1]. Note that the explicit value of b might change in the sequel. Let $\varphi_{\varepsilon/2} \in \mathcal{C}_c^\infty(\mathbb{R})$ be a mollifier on scale $\varepsilon/2$ and set

$$f(z) = \left(\mathbb{1}_{[1+3\varepsilon/2, C+\varepsilon/2]} * \varphi_{\varepsilon/2}\right)(|z|),$$

Note that f is smooth with

$$\left\|\partial^j f\right\|_\infty \lesssim \varepsilon^{-j} \tag{5.5}$$

1 We use Monte Carlo in the proof of Theorem 4.5.

for all $j \geq 0$.[1] We now have

$$\mathbb{P}(|\lambda|_{\max} > 1 + 2\varepsilon) \leq \mathbb{E}\Big(\sum_{j=1}^{n} f(\lambda_j)\Big) + \mathcal{O}(n^{-b}) \tag{5.6}$$

by roughly bounding the indicator variables with counting function. We aim to apply the grid approximation Lemma 4.11. Let S be random variable uniformly distributed on $[0,1]^2$, independent of X. If Ω_* is the event where (4.10) holds for X, then

$$\mathbb{E}\Big(\sum_{j=1}^{n} f(\lambda_j)\Big) = \mathbb{E}\Big(\mathbb{1}_{\Omega_*}\sum_{j=1}^{n} f(\lambda_j)\Big) + \mathcal{O}(n^{1-\log n}). \tag{5.7}$$

Freezing X first, we apply Lemma 4.11 for fixed $\lambda_1,\ldots,\lambda_n$, $\beta > C+1$ and α sufficiently large in order to dominate the derivatives (5.5) and obtain

$$\mathbb{E}\Big(\mathbb{1}_{\Omega_*}\sum_{j=1}^{n} f(\lambda_j)\Big) = \mathbb{E}\Big(\mathbb{1}_{\Omega_*}\frac{-2\beta^2 n}{n^\alpha \pi}\sum_{i=1}^{\lceil n^\alpha \rceil} \Delta f(z_i) U_n(z_i)\Big) + \mathcal{O}(n^{-b}) \tag{5.8}$$

for some $b > 0$. On the other hand freezing S, we may write

$$\mathbb{E}\Big(\mathbb{1}_{\Omega_*}\frac{n}{n^\alpha}\sum_{z_i\in A} \Delta f(z_i) U_n(z_i)\Big) = \frac{1}{n^\alpha}\,\mathbb{E}\Big(\sum_{z_i \in A} \Delta f(z_i)\,\mathbb{E}\big(\mathbb{1}_{\Omega_*} n U_n(z_i)\big|S\big)\Big), \tag{5.9}$$

where we used the fact that the grid $A = 2\beta n^{-\alpha/2}(\mathbb{Z}^2 + S) \cap [-\beta,\beta]^2$ depends on S only. For each summand $0 \leq i \leq n^\alpha$ let χ_i be the $[0,1]$-valued random variable from Lemma 5.8, for (fixed) $z_i \in \mathbb{C}$. Moreover on Ω_* we have bound $\sup_i |U_n(z_i)| = \mathcal{O}(\log^2 n)$ from Lemma 4.11 and hence introduce a smooth cutoff $G : \mathbb{R} \to \mathbb{R}$, $G(x) = x$ on $|x| \leq n\log^2 n$ and vanishing on $|x| \geq 2n\log^2 n$. We split the expectation into a part of χ_i and $1 - \chi_i$, yielding

$$\begin{aligned}
&\frac{1}{n^\alpha}\,\mathbb{E}\Big(\sum_{z_i\in A} \Delta f(z_i)\,\mathbb{E}\big(\mathbb{1}_{\Omega_*} n U_n(z_i)\big|S\big)\Big) \\
&= \frac{1}{n^\alpha}\,\mathbb{E}\Big(\sum_{z_i\in A} \Delta f(z_i)\Big(\mathbb{E}\big(G(nU_n(z_i))\chi_i\big|S\big) + \mathbb{E}\big(\mathbb{1}_{\Omega_*} n U_n(z_i)(1-\chi_i)\big|S\big)\Big)\Big) + \mathcal{O}(n^{-Q}).
\end{aligned}$$

First, let us estimate the error term that is the sum of expectations on $(1 - \chi_i)$. Proposition 2.15 states that for any fixed z_i, uniformly in $z_i \in [-\beta,\beta]^2$ with overwhelming

1 Here, ε plays the role of $1/a$ of Chapter 4.

probability it holds $U_n(z_i) = U_\infty(z_i) + O(n^{-1+\widetilde{\varepsilon}})$ for any $\widetilde{\varepsilon} > 0$ small enough. On the opposite event with probability $\mathcal{O}(n^{-Q})$ for all $Q > 0$ we use the rough bound $\sup_i |U_n(z_i)| = \mathcal{O}(\log^2 n)$ from Lemma 4.11. Therefore, we have

$$\begin{aligned}
&\frac{n}{n^\alpha}\mathbb{E}\Big(\sum_{z_i\in A}\mathbb{E}\left(\mathbb{1}_{\Omega_*}\Delta f(z_i)U_n(z_i)(1-\chi_i)\middle|S\right)\Big)\\
&\qquad\qquad = \frac{n}{n^\alpha}\sum_{i=1}^{\lceil n^\alpha\rceil}\mathbb{E}\Big(\mathbb{1}_{\Omega_*}\big(U_\infty(z_i)+O(n^{-1+\widetilde{\varepsilon}})\big)(1-\chi_i)\Delta f(z_i)\Big)+\mathcal{O}(n^{-Q}).
\end{aligned}$$

Let us remove the randomness of z_i by observing that the non-shifted grid $A^* = 2\beta n^{-\alpha/2}\,\mathbb{Z}^2\cap[-\beta,\beta]^2$ (enumerated by z_i^*) satisfies $|z_i - z_i^*| \lesssim n^{-\alpha/2}$ for all i, hence

$$\Delta f(z_i) = \Delta f(z_i^*) + \mathcal{O}(n^{-\alpha/2}\varepsilon^{-3}).$$

as well as $U_\infty(z_i) = U_\infty(z_i^*) + \mathcal{O}(n^{-\alpha/2})$. Together with $\mathbb{E}(1-\chi_i) = \mathcal{O}(n^{-b})$ from (5.4) we obtain

$$\begin{aligned}
&\frac{n}{n^\alpha}\mathbb{E}\Big(\sum_{z_i\in A}\mathbb{E}\left(\mathbb{1}_{\Omega_*}U_n(z_i)\Delta f(z_i)(1-\chi_i)\middle|S\right)\Big) \qquad (5.10)\\
&= \frac{n}{n^\alpha}\sum_{i=1}^{\lceil n^\alpha\rceil}\mathbb{E}\left(\mathbb{1}_{\Omega_*}(1-\chi_i)\right)\Big(U_\infty(z_i^*)+\mathcal{O}(n^{-1+\widetilde{\varepsilon}})\big)\Delta f(z_i^*)\Big)+\mathcal{O}(n^{1-\alpha/2}\varepsilon^{-3}).\\
&= \frac{\mathcal{O}(n^{1-b})}{n^\alpha}\sum_{i=1}^{\lceil n^\alpha\rceil}U_\infty(z_i^*)\Delta f(z_i^*)+\frac{\mathcal{O}(n^{-b+\widetilde{\varepsilon}})}{n^\alpha}\sum_{i=1}^{\lceil n^\alpha\rceil}\Delta f(z_i^*)+\mathcal{O}(n^{1-\alpha/2}\varepsilon^{-3}).
\end{aligned}$$

Both terms are Riemann sums that we estimate as in (4.11). Using $\Delta U_\infty = -2\pi\mu_\infty$, the first one approximates $\int f d\mu_\infty = 0$ at rate $\mathcal{O}(n^{-\alpha/2}\varepsilon^{-3})$, thus it vanishes at order $O(n^{1-b-\alpha/2}\varepsilon^{-3})$. The second term converges to $\|\Delta f\|_{L^1}$, more precisely we have $\sum_i |\Delta f(z_i^*)| = \mathcal{O}(n^\alpha/\varepsilon)$ using (5.5) and the size of $\operatorname{supp}\Delta f$. Choosing $\varepsilon = n^{-b/3}$ and $\widetilde{\varepsilon} = b/3$, we conclude that the total error term (5.10) is of order $\mathcal{O}(n^{-b})$ for some new $b > 0$.

The resolvent swapping applies whenever χ_i is nonzero, so that Lemma 5.8 with $\|G\|_\infty = \mathcal{O}(n\log^2 n)$ and for $d = 6$ matching moments yields

$$\begin{aligned}
&\frac{1}{n^\alpha}\mathbb{E}\Big(\sum_{z_i\in A}\Delta f(z_i)\Big(\mathbb{E}\left(G(nU_n(z_i))\chi_i\middle|S\right)\Big)\Big)\\
=&\frac{1}{n^\alpha}\mathbb{E}\Big(\sum_{z_i\in A}\Delta f(z_i)\Big(\mathbb{E}\left(G(n\widetilde{U}_n(z_i))\widetilde{\chi}_i\middle|S\right)\Big)\Big)+\mathcal{O}(n^{-b})\,\mathbb{E}\Big(\frac{1}{n^\alpha}\sum_{z_i\in A}\Delta f(z_i)\Big). \qquad (5.11)
\end{aligned}$$

The second term is again a Riemann sum converging to $\|\Delta f\|_{L^1} = \mathcal{O}(n^{b/3})$. Taking the same steps (5.10), (5.9), (5.8) and (5.7) backwards for $\widetilde{X}$, we end up with

$$\begin{aligned} \mathbb{E}\sum_{j=1}^{n} f(\lambda_j) &\leq \mathbb{E}\sum_{j=1}^{n} f(\widetilde{\lambda}_j) + \mathcal{O}(n^{-b}), \\ \frac{1}{n}\,\mathbb{E}\sum_{j=1}^{n} g(\lambda_j) &\leq \frac{1}{n}\,\mathbb{E}\sum_{j=1}^{n} g(\widetilde{\lambda}_j) + \mathcal{O}(n^{-b}) \end{aligned}$$

for some (possibly smaller) $b > 0$. Roughly bounding $f \leq \mathbb{1}_{B^c_{1+\varepsilon}(0)}$, we conclude

$$\mathbb{P}(|\lambda|_{\max} > 1 + 2\varepsilon) \leq \mathbb{E}\left(n\widetilde{\mu}_n(B^c_{1+\varepsilon}(0))\right) + \mathcal{O}(n^{-b})$$

for $\varepsilon = n^{-b/3}$. □

Proof of Theorem 5.4. The lower bound on $|\lambda|_{\max}$ is given in Lemma 5.3. It remains to estimate the right hand side in Theorem 5.6 for a Ginibre matrix $\widetilde{X}$. As we have shown in the proof of Lemma 3.1, it holds

$$\mathbb{E}\,\widetilde{\mu}_n(B^c_R(0)) = e^{-nR^2}\sum_{k=0}^{n-1}\frac{(n-k)(nR^2)^k}{nk!} \leq e^{-nR^2}\frac{(nR^2)^n}{(n)!}\frac{1}{(R^2-1)+1}$$

for $R > 0$. Applying Stirling's Formula yields

$$n\,\mathbb{E}\,\widetilde{\mu}_n(B_R(0)^c) \lesssim \sqrt{n}e^{-n(R^2-1-\log(R^2))}.$$

Finally choose $R = 1+n^{-b}$ and note that $(1+n^{-b})^2 - 1 - \log(1+n^{-b})^2 \geq n^{-2b} + \mathcal{O}(n^{-3b})$, we conclude

$$n\,\mathbb{E}\,\widetilde{\mu}_n(B_R(0)^c) \lesssim \sqrt{n}e^{-n^{1-2b}/4}$$

and the claim follows from Theorem 5.6 after a possible change to $b < 1/2$. □

Proof of Lemma 5.8. We follow the proof of the Four Moment Theorem for log determinants in [TV15], where essentially we only pay attention to the very last lines. Lemma 2.21 for $T = n^c$ states

$$U_n(z) = -\frac{1}{2n}\log|\det(V(z) - in^c)| + \operatorname{Im}\int_0^{n^c} m_n(z,i\eta)d\eta.$$

Here $V(z)$ is the Hermitization matrix (2.33), with empirical spectral distribution $\widetilde{\nu}^z_n$

and $m_n(z,\cdot)$ is its Stieltjes transform. It holds

$$\frac{1}{2n}\log|\det(V(z)-in^c)| = c\log n + \log\left|\det(i-n^{-c}V(z))\right| = c\log n + o(n^{-c/2}),$$

thus we may shift the argument of G by $c\log n$ and the error remains negligible for c sufficiently large by an argument similar to what we saw already in (2.48). Therefore, it remains to show that $\mathbb{E}(G(\operatorname{Im}\int_0^{n^c} s^z(i\eta)d\eta)\chi)$ is insensitive to exchanging the matrix ensembles with errors given in the claim. By applying resolvent bounds from smallest singular value estimates, Tao and Vu give an event of high probability where the following resolvent swapping statement applies. Let us first fix notation.

Enumerate the entries of a $n\times n$ non-Hermitian matrix by $l=0,\ldots,n^2$ and let $V_l(z)$ be the matrix $V(z)$ corresponding to the first l entries of X being swapped to $\widetilde{X}$. In particular we have $V_0(z)=V(z)$ and $V_{n^2}(z)=\widetilde{V}(z)$. Denote by $V_l^0(z)$ the Hermitization matrix corresponding to the first $l-1$ entries of the underlying matrix being swapped to $\widetilde{X}$, the l-th entry being 0 and the last entries being taken from X. Furthermore, let s_l^z, $s_l^{0,z}$ be the Stieltjes transforms of the empirical spectral distributions of $V_l(z)$, $V_l^0(z)$ respectively, hence the normalized traces of their resolvents R_l^z, $R_l^{0,z}$. Lastly we will need the matrix norm $\|M\|_{(\infty,1)}=\sup_{1\le i,j\le n}|M_{i,j}|$. Fixing l and all the matrices mentioned before, we quote the resolvent swapping [TV15, Proposition 45] in the form we will use it and suppress the appearances of z for readability.

Proposition 5.9 ([TV15]). *For $w=E+i\eta$ suppose that*

$$|X_l|\cdot\left\|R_l^0(w)\right\|_{(\infty,1)} = o(\sqrt{n}). \tag{5.12}$$

Then for fixed $d\in\mathbb{N}$ one has

$$\begin{aligned} s_l(w) &= s_l^0(w) + \sum_{i=1}^{d} n^{-i/2}c_i(w)X_l^i \\ &\quad + \mathcal{O}\left(n^{-\frac{d+1}{2}}|X_l|^{d+1}\left\|R_l^0(w)\right\|_{(\infty,1)}^{d+1}\min\left(\left\|R_l^0(w)\right\|_{(\infty,1)},\tfrac{1}{n\eta}\right)\right), \end{aligned}$$

where the coefficients c_i are independent of X_l and satisfy

$$|c_i(w)| = \mathcal{O}\left(\left\|R_l^0(w)\right\|_{(\infty,1)}^{i}\min\left(\left\|R_l^0(w)\right\|_{(\infty,1)},\tfrac{1}{n\eta}\right)\right)$$

for all $1\le i\le d$.

For some sufficiently small constant c_0 (cf. Theorem 5.7), let $\chi:\mathbb{R}\to\mathbb{R}$ be a smooth cutoff function to the region $|x|\le n^{3c_0}$ that equals 1 for $|x|\le n^{3c_0}/2$. We denote the random variable χ (under slight abuse of notation) as $\chi(\operatorname{Im}(m_n(z,in^{-1-4c_0})))$, see [TV15], which satisfies $\chi>0$ (and also $\chi=1$) with $1-\mathcal{O}(n^{-b})$-high probability. This is explicitly shown in the proof of [TV15, Theorem 23], by means of smallest singular value estimates, and we do not repeat the proof here. On the event $\chi > 0$, we have

$$\sup_{\eta>0} \left\| R_l^0(i\eta) \right\|_{(\infty,1)} = n^{\mathcal{O}(c_0)} \tag{5.13}$$

with overwhelming probability thus (5.12) holds (after truncation of the entries) for c_0 sufficiently small, cf. [TV15, Lemma 46]. Integrating in η shows that

$$n\mathrm{Im} \int_0^{n^c} s_l(i\eta) d\eta = n\mathrm{Im} \int_0^{n^c} s_l^0(i\eta) d\eta + \mathcal{P}_d(X_l) + \mathcal{O}(n^{-\frac{d+1}{2}+\mathcal{O}(c_0)})$$

for some polynomial $\mathcal{P}_d$ of order d with coefficients $a_j = \mathcal{O}(n^{-j/2+\mathcal{O}(c_0)})$, independent of X_l and $\mathcal{P}_d(0) = a_0 = 0$. A Taylor approximation of G up to order d around $n\mathrm{Im} \int_0^{n^c} s_l^0(i\eta) d\eta$ yields

$$G\Big(n\mathrm{Im} \int_0^{n^c} s_l(i\eta) d\eta\Big) = G\Big(n\mathrm{Im} \int_0^{n^c} s_l^0(i\eta) d\eta\Big) + \mathcal{Q}_d(X_l) + \mathcal{O}(n^{-\frac{d+1}{2}+\mathcal{O}(c_0)}) \tag{5.14}$$

for some other polynomial $\mathcal{Q}_d$. Its coefficients satisfy the same bounds as before, because of $\left\| G^{(j)} \right\|_\infty = \mathcal{O}(1)$ for $j \geq 1$. In a similar way by a Taylor approximation of χ up to order d we find that $\chi(\mathrm{Im}(s_l(in^{-1-4c_0})))$ is equal to a polynomial of degree d in X_l plus an error $\mathcal{O}(n^{-\frac{d+1}{2}+o(1)})$. By the matching moment hypothesis it follows

$$\begin{aligned}
&\mathbb{E}\left(G\Big(n\mathrm{Im} \int_0^{n^c} s_l(i\eta) d\eta\Big) \chi(\mathrm{Im}(s_l(in^{-1-4c_0})))\right) \\
&= \mathbb{E}\left(G\Big(n\mathrm{Im} \int_0^{n^c} s_{l+1}(i\eta) d\eta\Big) \chi(\mathrm{Im}(s_{l+1}(in^{-1-4c_0})))\right) + \|G\|_\infty \mathcal{O}(n^{-\frac{d+1}{2}+\mathcal{O}(c_0)}).
\end{aligned}$$

Repeating this swapping procedure for all $1 \leq l \leq n^2$ and choosing c_0 small enough s.t. $\mathcal{O}(c_0) < 1/2 - b$, we conclude

$$\begin{aligned}
&\mathbb{E}\left(G\Big(n\mathrm{Im} \int_0^{n^c} m_n(z,i\eta) d\eta\Big) \chi(\mathrm{Im}(m_n(z,in^{-1-4c_0})))\right) \\
&= \mathbb{E}\left(G\Big(n\mathrm{Im} \int_0^{n^c} \widetilde{s}^z(i\eta) d\eta\Big) \chi(\mathrm{Im}(\widetilde{s}^z(in^{-1-4c_0})))\right) + \|G\|_\infty \mathcal{O}(n^{-\frac{d-4}{2}-b})
\end{aligned}$$

for some $b > 0$. □

In the resolvent bound (5.13), the part $\sup_{\eta>1/n} \|R_l^0(i\eta)\|_{(\infty,1)} = n^{o(1)}$ holds with overwhelming probability and only the part $\eta < 1/n$ causes the swapping mechanism to work for high, instead of overwhelming, probability.

List of Symbols

Notation	Description
$(\Omega,\mathcal{A},\mathbb{P})$	The underlying probability space
$\mathbb{E}$	Expectation with respect to $\mathbb{P}$
$\mathbb{V}\mathrm{ar}$	Variance $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}\,\lvert X - \mathbb{E}\,X\rvert^2$ of a random variable X
$\mathbb{1}_A$	Indicator function of a measurable set A, i.e. $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise
δ_x	Dirac delta distribution in a point x, i.e. $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise
$\lambda\!\!\!\lambda$	Lebesgue measure on $\mathbb{C}$
$\mathbf{X}$	Product of m independent random matrices, see (1.8)
$\mathbf{W}$	Linearization matrix of $\mathbf{X}$, see (2.49)
$V(z)$	Hermitization of a shifted random matrix, see (2.32)
$\mathbf{V}(z)$	Hermitization of the shifted matrix $\mathbf{W}$, see (2.33)
$\mathrm{trace}(\cdot)$	Trace of a matrix
λ_j	An eigenvalue
s_j	Singular values, ordered decreasingly
μ_n	Empirical spectral distribution (ESD) of a random matrix, see (1.1)
μ_n^m	The ESD of $\mathbf{X}$, see (1.9)
$\bar{\mu}_n, \bar{\mu}_n^m$	The mean ESD's, i.e. $\bar{\mu}_n(A) = \mathbb{E}\,\mu_n(A)$ or $\bar{\mu}_n^m(A) = \mathbb{E}\,\mu_n^m(A)$
μ_∞	The Circular Law $d\mu_\infty(z) = \frac{1}{\pi}\mathbb{1}_{B_1(0)}(z)dz$, see (1.2)
ν_n^z	Empirical singular value distribution of matrices shifted by $z \in \mathbb{C}$, see (2.32) and of the matrix $\mathbf{W}$, see (2.49)

Notation	Description
$\widetilde{\nu}_n^z$	The ESD of $V(z)$ or $\mathbf{V}(z)$, which is the symmetrized version of $\widetilde{\nu}_n^z$, see (2.33) and (2.50)
U_n	Logarithmic Potential of μ_n, see (2.27) and (2.32), and for $m > 1$ see (2.50)
U_∞	Logarithmic Potential of μ_∞, see (2.27) and (2.28)
$m_n(z,\cdot)$	Stieltjes Transform of $\widetilde{\nu}_n^z$, see Definition 2.4 and (2.34)
$\widetilde{\nu}_\infty^z$	The limit distribution of $\widetilde{\nu}_n^z$, see (2.35) and Figure 2.1
$s(z,\cdot)$	Stieltjes Transform of the limit distribution $\widetilde{\nu}_\infty^z$, see (2.35)
D	The Kolmogorov-like metric (over balls) for probability measures on $\mathbb{C}$, defined in (1.3)
d_K	The classical two dimensional Kolmogorov-like metric for probability measures on $\mathbb{C}$, see (1.7). For this and other metrics see Section 2.3.1
d^*	The classical one dimensional Kolmogorov-like metric for probability measures on $\mathbb{R}$, see (2.10) or (4.1).
$\oint$	A line integral over a closed curve, e.g. $\oint_\gamma f dz = \int f(\gamma(t))\dot{\gamma}(t)dt$
$⨍$	Mean integral, i.e. $⨍_A f d\lambda^2 = \frac{1}{\lambda^2(A)} \int f d\lambda^2$
$*$	Convolution of functions and probability measures
$\Rightarrow$	Weak convergence / convergence in distribution
$\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$	Rounding a real number down or up to the closest natural number
$\wedge$ and $\vee$	Minimum and maximum of two real numbers
$(\cdot)_+$	Positive part of a real number, i.e. $x_+ = x \vee 0$
$\mathcal{O}$	Asymptotically bounded, i.e. $f = \mathcal{O}(g)$ if $\limsup_{n\to\infty} \left\lvert\frac{f(n)}{g(n)}\right\rvert < \infty$
o	Asymptotically negligible, i.e. $f = o(g)$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$
$\sim$	Asymptotic equivalence, i.e. $f \sim g$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$

Notation	Description
$\lesssim$	Bounded up to a parameter independent constant, i.e. $f \lesssim g$ if $f \leq cg$ for some $c > 0$
$\asymp$	$A \asymp B$ if $c\,\lvert B\rvert \leq \lvert A\rvert \leq C\,\lvert B\rvert$ for some constants $0 < c < C$
∂_x, ∂_y	Partial derivatives with respect to the variable x or y, denoting real and imaginary direction
$\partial, \bar{\partial}$	Wirtinger derivatives, i.e. $\partial = \frac{1}{2}(\partial_x - i\partial_y)$ and $\bar{\partial} = \frac{1}{2}(\partial_x + i\partial_y)$
Δ	Laplace operator in the complex plane, i.e. $\Delta f = (\partial_x^2 + \partial_y^2)(f)$ or $\Delta = 4\partial\bar{\partial}$ in terms of Wirtinger derivatives
w.o.p.	With overwhelming probability; Ω_n holds w.o.p. if $\mathbb{P}(\Omega_n^c) \lesssim n^{-Q}$ for any $Q > 0$
$\log^b n$	Shorthand for $\log(n)^b$
$\mathcal{C}_c^\infty$	Infinitely often differentiable functions $f : \mathbb{C} \to \mathbb{R}$ with compact support
Γ	The Gamma function $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ for $\mathrm{Re}(z) > 0$
$G_{p,q}^{m,n}$	The Meijer-G function, defined in (2.24)

Bibliography

[ACW16] Adamczak, Radosław, Djalil Chafaï, and Paweł Wolff: 'Circular law for random matrices with exchangeable entries'. *Random Structures & Algorithms* (2016), vol. 48(3): pp. 454–479.

[ARRS16] Adhikari, Kartick, Nanda Kishore Reddy, Tulasi Ram Reddy, and Koushik Saha: 'Determinantal point processes in the plane from products of random matrices'. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques.* Vol. 52. 1. Institut Henri Poincaré. 2016: pp. 16–46.

[AI15] Akemann, G. and J.R. Ipsen: 'Recent Exact and Asymptotic Results for Products of Independent Random Matrices'. *Acta Physica Polonica B* (2015), vol. 46(9): p. 1747.

[Ake17] Akemann, Gernot: 'Random matrix theory and quantum chromodynamics'. *Stochastic Processes and Random Matrices: Lecture Notes of the Les Houches Summer School: Volume 104, July 2015* (2017), vol. 104: p. 228.

[ABC20] Akemann, Gernot, Michael Baake, Nayden Chakarov, Oliver Krüger, Adam Mielke, Meinolf Ottensmann, and Rebecca Werdehausen: *Territorial Behaviour of Buzzards Versus Random Matrix Spacing Distributions.* In Preparation. 2020.

[ABD11] Akemann, Gernot, Jinho Baik, and Philippe Di Francesco: *The Oxford handbook of random matrix theory.* Oxford University Press, 2011.

[AB10] Akemann, Gernot and Martin Bender: 'Interpolation between Airy and Poisson statistics for unitary chiral non-Hermitian random matrix ensembles'. *Journal of Mathematical Physics* (2010), vol. 51(10): p. 103524.

[AB12] Akemann, Gernot and Zdzislaw Burda: 'Universal microscopic correlation functions for products of independent Ginibre matrices'. *Journal of Physics A: Mathematical and Theoretical* (2012), vol. 45(46): p. 465201.

[ABK19] Akemann, Gernot, Zdzislaw Burda, and Mario Kieburg: 'From integrable to chaotic systems: Universal local statistics of Lyapunov exponents'. *EPL (Europhysics Letters)* (2019), vol. 126(4): p. 40001.

[AC18] Akemann, Gernot and Milan Cikovic: 'Products of random matrices from fixed trace and induced Ginibre ensembles'. *Journal of Physics A: Mathematical and Theoretical* (2018), vol. 51(18): p. 184002.

[AKMP19] Akemann, Gernot, Mario Kieburg, Adam Mielke, and Tomaž Prosen: 'Universal signature from integrability to chaos in dissipative open quantum systems'. *Physical review letters* (2019), vol. 123(25).

[AGT10] Alexeev, Nikita V., Friedrich Götze, and Alexander Tikhomirov: 'On the singular spectrum of powers and products of random matrices'. *Doklady mathematics*. Vol. 82. 1. Springer. 2010: pp. 505–507.

[AEK18] Alt, Johannes, Laszlo Erdős, and Torben Krüger: 'Local inhomogeneous circular law'. *Ann. Appl. Probab.* (Feb. 2018), vol. 28(1): pp. 148–203.

[AEK19] Alt, Johannes, Laszlo Erdős, and Torben Krüger: *Spectral radius of random matrices with independent entries*. 2019. arXiv: `1907.13631 [math.PR]`.

[AGZ10] Anderson, Greg W., Alice Guionnet, and Ofer Zeitouni: *An introduction to random matrices*. Vol. 118. Cambridge university press, 2010.

[Bai93a] Bai, Zhi Dong: 'Convergence Rate of Expected Spectral Distributions of Large Random Matrices. Part I. Wigner Matrices'. *The Annals of Probability* (1993), vol. 21(2): pp. 625–648.

[Bai93b] Bai, Zhi Dong: 'Convergence rate of expected spectral distributions of large random matrices. Part II. Sample covariance matrices'. *The Annals of Probability* (1993), vol. 21(2): pp. 649–672.

[Bai97] Bai, Zhi Dong: 'Circular Law'. *The Annals of Probability* (1997), vol.: pp. 494–529.

[BY88] Bai, Zhi Dong and Yong-Qua Yin: 'Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix'. *The Annals of Probability* (1988), vol.: pp. 1729–1741.

[BCZH08] Bai, Zhidong, Zehua Chen, Jin-Ting Zhang, and Feifang Hu: *Advances in Statistics: Proceedings of the Conference in Honor of Professor Zhidong Bai on His 65th Birthday, National University of Singapore, 20 July 2008*. World Scientific, 2008.

[BHPZ11] Bai, Zhidong, Jiang Hu, Guangming Pan, and Wang Zhou: 'A note on rate of convergence in probability to semicircular law'. *Electronic Journal of Probability* (2011), vol. 16: pp. 2439–2451.

[BS10] Bai, Zhidong and Jack W. Silverstein: *Spectral analysis of large dimensional random matrices*. Vol. 20. Springer, 2010.

[Bel54] Bellman, Richard: 'Limit theorems for non-commutative operations. I.' *Duke Mathematical Journal* (1954), vol. 21(3): pp. 491–500.

[BP05] Ben Arous, Gérard and Sandrine Péché: 'Universality of local eigenvalue statistics for some sample covariance matrices'. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* (2005), vol. 58(10): pp. 1316–1357.

[BK16] Benaych-Georges, Florent and Antti Knowles: *Lectures on the local semicircle law for Wigner matrices*. 2016. arXiv: `1601.04055 [math.PR]`.

[BB19] Berezin, Sergey and Alexander I. Bufetov: *On the Rate of Convergence in the Central Limit Theorem for Linear Statistics of Gaussian, Laguerre, and Jacobi Ensembles.* 2019. arXiv: `1904.09685 [math.PR]`.

[Ber41] Berry, Andrew C.: 'The accuracy of the Gaussian approximation to the sum of independent variates'. *Transactions of the american mathematical society* (1941), vol. 49(1): pp. 122–136.

[Bha68] Bhattacharya, Rabindra Nath: 'Berry-Esseen bounds for the multi-dimensional central limit theorem'. *Bulletin of the American Mathematical Society* (1968), vol. 74(2): pp. 285–287.

[BR10] Bhattacharya, Rabindra Nath and R. Ranga Rao: *Normal approximation and asymptotic expansions.* SIAM, 2010.

[Bor11] Bordenave, Charles: 'On the spectrum of sum and product of non-hermitian random matrices'. *Electron. Commun. Probab.* (2011), vol. 16: pp. 104–113.

[BCC11] Bordenave, Charles, Pietro Caputo, and Djalil Chafaï: 'Spectrum of non-Hermitian heavy tailed random matrices'. *Communications in mathematical physics* (2011), vol. 307(2): p. 513.

[BCCT18] Bordenave, Charles, Pietro Caputo, Djalil Chafaï, and Konstantin Tikhomirov: 'On the spectral radius of a random matrix: An upper bound without fourth moment'. *The Annals of Probability* (2018), vol. 46(4): pp. 2268–2286.

[BC12] Bordenave, Charles and Djalil Chafaï: 'Around the circular law'. *Probability surveys* (2012), vol. 9.

[BLM13] Boucheron, Stéphane, Gábor Lugosi, and Pascal Massart: *Concentration inequalities: A nonasymptotic theory of independence.* Oxford university press, 2013.

[BYY14a] Bourgade, Paul, Horng-Tzer Yau, and Jun Yin: 'Local circular law for random matrices'. *Probability Theory and Related Fields* (2014), vol. 159(3-4): pp. 545–595.

[BYY14b] Bourgade, Paul, Horng-Tzer Yau, and Jun Yin: 'The local circular law II: the edge case'. *Probability Theory and Related Fields* (2014), vol. 159(3-4): pp. 619–660.

[BJW10] Burda, Zdzisław, Romuald A. Janik, and Bartek Waclaw: 'Spectrum of the product of independent random Gaussian matrices'. *Physical Review E* (2010), vol. 81(4): p. 041132.

[BJL10] Burda, Zdzisław, Andrzej Jarosz, Giacomo Livan, Maciej A. Nowak, and Artur Swiech: 'Eigenvalues and singular values of products of rectangular Gaussian random matrices'. *Physical Review E* (2010), vol. 82(6): p. 061114.

[BNS12] BURDA, ZDZISŁAW, MACIEJ A. NOWAK, and ARTUR SWIECH: 'Spectral relations between products and powers of isotropic random matrices'. *Physical Review E* (2012), vol. 86(6): p. 061137.

[Cas06] CASWELL, HAL: 'Matrix population models'. *Encyclopedia of Environmetrics* (2006), vol. 3.

[CHM18] CHAFAÏ, DJALIL, ADRIEN HARDY, and MYLÈNE MAÏDA: 'Concentration for Coulomb gases and Coulomb transport inequalities'. *J. Funct. Anal.* (2018), vol. 275(6): pp. 1447–1483.

[CP14] CHAFAÏ, DJALIL and SANDRINE PÉCHÉ: 'A note on the second order universality at the edge of Coulomb gases on the plane'. *Journal of Statistical Physics* (2014), vol. 156(2): pp. 368–383.

[CTX19] CHAN, CHIN HEI, VAHID TAROKH, and MAOSHENG XIONG: *Convergence Rate of Empirical Spectral Distribution of Random Matrices from Linear Codes.* 2019. arXiv: `1902.08428 [math.PR]`.

[CB04] CHATTERJEE, SOURAV and ARUP BOSE: 'A new method for bounding rates of convergence of empirical spectral distributions'. *Journal of Theoretical Probability* (2004), vol. 17(4): pp. 1003–1019.

[CM16] CICUTA, GIOVANNI M. and LUCA GUIDO MOLINARI: 'Random antagonistic matrices'. *Journal of Physics A: Mathematical and Theoretical* (2016), vol. 49(37): p. 375601.

[CES19a] CIPOLLONI, GIORGIO, LÁSZLÓ ERDŐS, and DOMINIK SCHRÖDER: *Central Limit Theorem for Linear Eigenvalue Statistics of non-Hermitian Random Matrices.* 2019. arXiv: `1912.04100 [math.PR]`.

[CES19b] CIPOLLONI, GIORGIO, LÁSZLÓ ERDŐS, and DOMINIK SCHRÖDER: *Edge Universality for non-Hermitian Random Matrices.* 2019. arXiv: `1908.00969 [math.PR]`.

[CFLW19] CLAEYS, TOM, BENJAMIN FAHS, GAULTIER LAMBERT, and CHRISTIAN WEBB: *How much can the eigenvalues of a random Hermitian matrix fluctuate?* 2019. arXiv: `1906.01561 [math.PR]`.

[CN84] COHEN, JOEL E. and CHARLES M. NEWMAN: 'The stability of large random matrices and their products'. *The Annals of Probability* (1984), vol.: pp. 283–310.

[Coo19] COOK, NICHOLAS: 'The circular law for random regular digraphs'. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques.* Vol. 55. 4. Institut Henri Poincaré. 2019: pp. 2111–2167.

[CO19] COSTON, NATALIE and SEAN O'ROURKE: 'Gaussian fluctuations for linear eigenvalue statistics of products of independent iid random matrices'. *Journal of Theoretical Probability* (2019), vol.: pp. 1–72.

[Cra38] CRAMÉR, HARALD: 'On a new limit theorem in probability theory'. *Actual. Sci. Ind.* (1938), vol. 736.

[CPV12] Crisanti, Andrea, Giovanni Paladin, and Angelo Vulpiani: *Products of random matrices: in Statistical Physics*. Vol. 104. Springer Science & Business Media, 2012.

[DS94] Diaconis, Persi and Mehrdad Shahshahani: 'On the eigenvalues of random matrices'. *Journal of Applied Probability* (1994), vol. 31(A): pp. 49–62.

[DV17] Dinh, Tien-Cuong and Duc-Viet Vu: *Large deviation theorem for random covariance matrices*. 2017. arXiv: `1707.07174 [math.CV]`.

[Dys62a] Dyson, Freeman J.: 'A Brownian-motion model for the eigenvalues of a random matrix'. *Journal of Mathematical Physics* (1962), vol. 3(6): pp. 1191–1198.

[Dys62b] Dyson, Freeman J.: 'Statistical theory of the energy levels of complex systems. I'. *Journal of Mathematical Physics* (1962), vol. 3(1): pp. 140–156.

[Dys62c] Dyson, Freeman J.: 'The threefold way. Algebraic structure of symmetry groups and ensembles in quantum mechanics'. *Journal of Mathematical Physics* (1962), vol. 3(6): pp. 1199–1215.

[Ede97] Edelman, Alan: 'The Probability that a Random Real Gaussian Matrix has k Real Eigenvalues, Related Distributions, and the Circular Law'. *Journal of Multivariate Analysis* (1997), vol. 60(2): pp. 203 –232.

[EKS94] Edelman, Alan, Eric Kostlan, and Michael Shub: 'How many eigenvalues of a random matrix are real?' *Journal of the American Mathematical Society* (1994), vol. 7(1): pp. 247–267.

[Erd11] Erdős, László: 'Universality of Wigner random matrices: a survey of recent results'. *Russian Mathematical Surveys* (2011), vol. 66(3): p. 507.

[EKYY12] Erdős, László, Antti Knowles, Horng-Tzer Yau, and Jun Yin: 'Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues'. *Communications in Mathematical Physics* (2012), vol. 314(3): pp. 587–640.

[EPR10] Erdős, László, Sandrine Péché, José A Ramírez, Benjamin Schlein, and Horng-Tzer Yau: 'Bulk universality for Wigner matrices'. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* (2010), vol. 63(7): pp. 895–925.

[ESY11] Erdős, László, Benjamin Schlein, and Horng-Tzer Yau: 'Universality of random matrices and local relaxation flow'. *Inventiones mathematicae* (2011), vol. 185(1): pp. 75–119.

[EYY12a] Erdős, László, Horng-Tzer Yau, and Jun Yin: 'Bulk universality for generalized Wigner matrices'. *Probability Theory and Related Fields* (Oct. 2012), vol. 154(1): pp. 341–407.

[EYY12b] Erdős, László, Horng-Tzer Yau, and Jun Yin: 'Rigidity of eigenvalues of generalized Wigner matrices'. *Advances in Mathematics* (2012), vol. 229(3): pp. 1435–1515.

[EKYY13] Erdős, László, Antti Knowles, Horng-Tzer Yau, and Jun Yin: 'Spectral statistics of Erdős–Rényi graphs I: Local semicircle law'. *Ann. Probab.* (May 2013), vol. 41(3B): pp. 2279–2375.

[Ess45] Esseen, Carl-Gustav: 'Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law'. *Acta Mathematica* (1945), vol. 77: pp. 1–125.

[Eva98] Evans, Lawrence C.: *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 1998.

[FK20] Fang, Xiao and Yuta Koike: *High-dimensional Central Limit Theorems by Stein's Method*. 2020. arXiv: `2001.10917 [math.PR]`.

[FZ97] Feinberg, Joshua and Anthony Zee: 'Non-hermitian random matrix theory: Method of hermitian reduction'. *Nuclear Physics B* (1997), vol. 504(3): pp. 579–608.

[FL12] Fischer, Wolfgang and Ingo Lieb: 'A course in complex analysis'. *Vieweg+ Teubner Verlag, Berlin* (2012), vol.

[For10] Forrester, Peter J.: *Log-gases and random matrices (LMS-34)*. Princeton University Press, 2010.

[FH99] Forrester, Peter J. and Graeme Honner: 'Exact statistical properties of the zeros of complex random polynomials'. *Journal of Physics A: Mathematical and General* (1999), vol. 32(16): p. 2961.

[FN07] Forrester, Peter J. and Taro Nagao: 'Eigenvalue statistics of the real Ginibre ensemble'. *Physical review letters* (2007), vol. 99(5): p. 050603.

[FW17] Forrester, Peter J. and Dong Wang: 'Muttalib-Borodin ensembles in random matrix theory - realisations and correlation functions'. *Electron. J. Probab.* (2017), vol. 22: 43 pp.

[Fro35] Frostman, Otto: 'Potentiel d'équilibre et capacité des ensembles'. PhD thesis. 1935.

[FK60] Furstenberg, Harry and Harry Kesten: 'Products of random matrices'. *The Annals of Mathematical Statistics* (1960), vol. 31(2): pp. 457–469.

[FKS97] Fyodorov, Yan V, Boris A Khoruzhenko, and Hans-Juergen Sommers: 'Almost-Hermitian random matrices: Eigenvalue density in the complex plane'. *Physics Letters A* (1997), vol. 226(1-2): pp. 46–52.

[GFF05] Garoni, TM, Peter J. Forrester, and NE Frankel: 'Asymptotic corrections to the eigenvalue density of the GUE and LUE'. *Journal of mathematical physics* (2005), vol. 46(10): p. 103301.

[GS02] Gibbs, Alison L. and Francis Edward Su: 'On choosing and bounding probability metrics'. *International statistical review* (2002), vol. 70(3): pp. 419–435.

[Gin65] GINIBRE, JEAN: 'Statistical ensembles of complex, quaternion, and real matrices'. *J. Mathematical Phys.* (1965), vol. 6: pp. 440–449.

[Gir85] GIRKO, VYACHESLAV L.: 'Circular law'. *Theory of Probability & Its Applications* (1985), vol. 29(4): pp. 694–706.

[Gir94] GIRKO, VYACHESLAV L.: 'The circular law: ten years later'. *Random Operators and Stochastic Equations* (1994), vol. 2(3): pp. 235–276.

[Gir04a] GIRKO, VYACHESLAV L.: 'The strong circular law. Twenty years later. Part I'. *Random Operators and Stochastic Equations rose* (2004), vol. 12(1): pp. 49–104.

[Gir04b] GIRKO, VYACHESLAV L.: 'The strong circular law. Twenty years later. Part II'. *Random Operators and Stochastic Equations rose* (2004), vol. 12(3): pp. 255–312.

[Gir05] GIRKO, VYACHESLAV L.: 'The Circular Law. Twenty years later. Part III'. *Random Operators and Stochastic Equations rose* (2005), vol. 13(1): pp. 53–109.

[Gir12] GIRKO, VYACHESLAV L.: 'The circular law. Thirty years later'. *Random Operators and Stochastic Equations* (2012), vol. 20(2): pp. 143–187.

[Göt91] GÖTZE, FRIEDRICH: 'On the rate of convergence in the multivariate CLT'. *The Annals of Probability* (1991), vol.: pp. 724–739.

[GJ18] GÖTZE, FRIEDRICH and JONAS JALOWY: *Rate of Convergence to the Circular Law via Smoothing Inequalities for Log-Potentials.* 2018. arXiv: `1807.00489 [math.PR]`.

[GKT15] GÖTZE, FRIEDRICH, HOLGER KÖSTERS, and ALEXANDER TIKHOMIROV: 'Asymptotic spectra of matrix-valued functions of independent random matrices and free probability'. *Random Matrices: Theory and Applications* (2015), vol. 4(02): p. 1550005.

[GNT15a] GÖTZE, FRIEDRICH, ALEXEY NAUMOV, and ALEXANDER TIKHOMIROV: *Local semicircle law under moment conditions. Part I: The Stieltjes transform.* 2015. arXiv: `1510.07350 [math.PR]`.

[GNT15b] GÖTZE, FRIEDRICH, ALEXEY NAUMOV, and ALEXANDER TIKHOMIROV: *Local semicircle law under moment conditions. Part II: Localization and delocalization.* 2015. arXiv: `1511.00862 [math.PR]`.

[GNT15c] GÖTZE, FRIEDRICH, ALEXEY NAUMOV, and ALEXANDER TIKHOMIROV: 'On a Generalization of the Elliptic Law for Random Matrices'. *Acta Physica Polonica B* (2015), vol. 46(9): p. 1737.

[GNT19a] GÖTZE, FRIEDRICH, ALEXEY NAUMOV, and ALEXANDER TIKHOMIROV: 'Local laws for non-Hermitian random matrices and their products'. *Random Matrices: Theory and Applications* (2019), vol.: p. 2150004.

[GNT19b] GÖTZE, FRIEDRICH, ALEXEY NAUMOV, and ALEXANDER TIKHOMIROV: 'Local semicircle law under fourth moment condition'. *Journal of Theoretical Probability* (2019), vol.: pp. 1–36.

[GNTT18] Götze, Friedrich, Alexey Naumov, Alexander Tikhomirov, and Dmitry Timushev: 'On the local semicircular law for Wigner ensembles'. *Bernoulli* (2018), vol. 24(3): pp. 2358–2400.

[GT10a] Götze, Friedrich and Aleksander Tikhomirov: 'The rate of convergence of spectra of sample covariance matrices'. *Theory of Probability & Its Applications* (2010), vol. 54(1): pp. 129–140.

[GT03a] Götze, Friedrich and Alexander Tikhomirov: 'Rate of convergence to the semi-circular law'. *Probability Theory and Related Fields* (2003), vol. 127(2): pp. 228–276.

[GT04] Götze, Friedrich and Alexander Tikhomirov: 'Rate of convergence in probability to the Marchenko-Pastur law'. *Bernoulli* (2004), vol. 10(3): pp. 503–548.

[GT05] Götze, Friedrich and Alexander Tikhomirov: 'The rate of convergence for spectra of GUE and LUE matrix ensembles'. *Open Mathematics* (2005), vol. 3(4): pp. 666–704.

[GT07] Götze, Friedrich and Alexander Tikhomirov: *On the Circular Law.* 2007. arXiv: `math/0702386 [math.PR]`.

[GT10b] Götze, Friedrich and Alexander Tikhomirov: *On the asymptotic spectrum of products of independent random matrices.* 2010. arXiv: `1012.2710 [math.PR]`.

[GT10c] Götze, Friedrich and Alexander Tikhomirov: 'The circular law for random matrices'. *The Annals of Probability* (2010), vol. 38(4): pp. 1444–1491.

[GT16] Götze, Friedrich and Alexander Tikhomirov: 'Optimal bounds for convergence of expected spectral distributions to the semi-circular law'. *Probability Theory and Related Fields* (2016), vol. 165(1-2): pp. 163–233.

[GJJN03] Gudowska-Nowak, Ewa, Romuald A Janik, Jerzy Jurkiewicz, and Maciej A Nowak: 'Infinite products of large random matrices and matrix-valued diffusion'. *Nuclear Physics B* (2003), vol. 670(3): pp. 479–507.

[GNT15d] Götze, Friedrich, Alexey Naumov, and Alexander Tikhomirov: 'Limit Theorems for Two Classes of Random Matrices with Dependent Entries'. *Theory of Probability & Its Applications* (2015), vol. 59(1): pp. 23–39.

[GT03b] Götze, Friedrich and Alexander Tikhomirov: 'Rate of convergence to the semicircular law for the Gaussian unitary ensemble'. *Theory of Probability & Its Applications* (2003), vol. 47(2): pp. 323–330.

[HL99] Haagerup, Uffe and Flemming Larsen: *Brown's spectral distribution measure for R-diagonal elements in finite von Neumann algebras.* Centre for Mathematical Physics and Stochastics, University of Aarhus, 1999.

[HJ12] Horn, Roger A. and Charles R. Johnson: *Matrix analysis.* Cambridge university press, 2012.

[Hub04] Huber, Peter J: *Robust statistics*. Vol. 523. John Wiley & Sons, 2004.

[Hwa86] Hwang, Chii-Ruey: 'A brief survey on the spectral radius and the spectral distribution of large random matrices with iid entries'. *Random Matrices and Their Applications (Contemporary Mathematics 50)* (1986), vol.: pp. 145–152.

[IZ13] Ibragimov, Ildar and Dmitry Zaporozhets: 'On distribution of zeros of random polynomials in complex plane'. *Prokhorov and contemporary probability theory*. Springer, 2013: pp. 303–323.

[Ips13] Ipsen, Jesper R.: 'Products of independent quaternion Ginibre matrices and their correlation functions'. *Journal of Physics A: Mathematical and Theoretical* (2013), vol. 46(26): p. 265201.

[IK14] Ipsen, Jesper R. and Mario Kieburg: 'Weak commutation relations and eigenvalue statistics for products of rectangular random matrices'. *Phys. Rev. E* (3 Mar. 2014), vol. 89: p. 032106.

[Jal19] Jalowy, Jonas: *Rate of Convergence for products of independent non-Hermitian random matrices*. 2019. arXiv: `1912.09300 [math.PR]`.

[JLM93] Jancovici, B, Joel L Lebowitz, and G Manificat: 'Large charge fluctuations in classical Coulomb systems'. *Journal of statistical physics* (1993), vol. 72(3-4): pp. 773–787.

[Joh01] Johansson, Kurt: 'Universality of the Local Spacing Distribution in Certain Ensembles of Hermitian Wigner Matrices'. *Communications in Mathematical Physics* (2001), vol. 215(3): pp. 683–705.

[Joh05] Johansson, Kurt: *Random matrices and determinantal processes*. 2005. arXiv: `math-ph/0510038 [math-ph]`.

[Joh07] Johansson, Kurt: 'From Gumbel to Tracy-Widom'. *Probability theory and related fields* (2007), vol. 138(1-2): pp. 75–112.

[KZ14a] Kabluchko, Zakhar and Dmitry Zaporozhets: 'Asymptotic distribution of complex zeros of random analytic functions'. *The Annals of Probability* (2014), vol. 42(4): pp. 1374–1395.

[KB02] Kalisch, Frieder and Daniel Braak: 'Exact density of states for finite Gaussian random matrix ensembles via supersymmetry'. *Journal of Physics A: Mathematical and General* (2002), vol. 35(47): p. 9957.

[KKS15] Kieburg, Mario, Arno B.J. Kuijlaars, and Dries Stivigny: 'Singular Value Statistics of Matrix Products with Truncated Unitary Matrices'. *International Mathematics Research Notices* (Aug. 2015), vol. 2016(11): pp. 3392–3424.

[KW58] Kiefer, J and J Wolfowitz: 'On the deviations of the empiric distribution function of vector chance variables'. *Transactions of the American Mathematical Society* (1958), vol. 87(1): pp. 173–186.

[KS80] KINDERLEHRER, DAVID and GUIDO STAMPACCHIA: *An introduction to variational inequalities and their applications.* Vol. 31. Siam, 1980.

[Kol33] KOLMOGOROV, AN: 'Sulla determinazione empirica di una legge di distribuzione'. *Giornale dell'Istituto Italiano degli Attuari* (1933), vol. 4: pp. 83–91.

[KOV18] KOPEL, PHIL, SEAN O'ROURKE, and VAN VU: *Random matrix products: Universality and least singular values.* 2018. arXiv: `1802.03004 [math.PR]`.

[KT15] KÖSTERS, HOLGER and ALEXANDER TIKHOMIROV: *Limiting spectral distributions of sums of products of non-Hermitian random matrices.* 2015. arXiv: `1506.04436 [math.PR]`.

[Kos92] KOSTLAN, ERIC: 'On the spectra of Gaussian matrices'. *Linear algebra and its applications* (1992), vol. 162: pp. 385–388.

[Kra12] KRANTZ, STEVEN G: *Handbook of complex variables.* Springer Science & Business Media, 2012.

[KZ14b] KUIJLAARS, ARNO B.J. and LUN ZHANG: 'Singular values of products of Ginibre random matrices, multiple orthogonal polynomials and hard edge scaling limits'. *Communications in Mathematical Physics* (2014), vol. 332(2): pp. 759–781.

[Lam19] LAMBERT, GAULTIER: *The law of large numbers for the maximum of the characteristic polynomial of the Ginibre ensemble.* 2019. arXiv: `1902.01983 [math.PR]`.

[LY14] LEE, JI OON and JUN YIN: 'A necessary and sufficient condition for edge universality of Wigner matrices'. *Duke Mathematical Journal* (2014), vol. 163(1): pp. 117–173.

[LWZ16] LIU, DANG-ZHENG, DONG WANG, and LUN ZHANG: 'Bulk and soft-edge universality for singular values of products of Ginibre random matrices'. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques.* Vol. 52. 4. Institut Henri Poincaré. 2016: pp. 1734–1762.

[LW19] LIU, DANG-ZHENG and YANHUI WANG: *Phase transitions for infinite products of large non-Hermitian random matrices.* 2019. arXiv: `1912.11910 [math.PR]`.

[Luk14] LUKE, YUDELL L: *Mathematical functions and their approximations.* Academic Press, 2014.

[MP67] MARČENKO, VLADIMIR A and LEONID ANDREEVICH PASTUR: 'Distribution of eigenvalues for some sets of random matrices'. *Mathematics of the USSR-Sbornik* (1967), vol. 1(4): p. 457.

[May72] MAY, ROBERT M: 'Will a large complex system be stable?' *Nature* (1972), vol. 238(5364): pp. 413–414.

[MM15] MECKES, ELIZABETH S. and MARK W. MECKES: 'A rate of convergence for the circular law for the complex Ginibre ensemble'. *Ann. Fac. Sci. Toulouse Math. (6)* (2015), vol. 24(1): pp. 93–117.

[Meh67] Mehta, Madan Lal: *Random matrices and the statistical theory of energy levels*. Academic Press, 1967.

[Meh04] Mehta, Madan Lal: *Random matrices*. Vol. 142. Elsevier, 2004.

[MS17] Mingo, James A and Roland Speicher: *Free probability and random matrices*. Vol. 35. Springer, 2017.

[Mül97] Müller, Alfred: 'Integral probability metrics and their generating classes of functions'. *Advances in Applied Probability* (1997), vol. 29(2): pp. 429–443.

[Mül02] Müller, Ralf Reiner: 'On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels'. *IEEE Transactions on Information Theory* (2002), vol. 48(7): pp. 2086–2091.

[Nau12] Naumov, Alexey: *Elliptic law for real random matrices*. 2012. arXiv: `1201.1639 [math.PR]`.

[Nem17] Nemish, Yuriy: 'Local law for the product of independent non-Hermitian random matrices with independent entries'. *Electron. J. Probab.* (2017), vol. 22.

[Nem18] Nemish, Yuriy: 'No Outliers in the Spectrum of the Product of Independent Non-Hermitian Random Matrices with Independent Entries'. *Journal of Theoretical Probability* (Mar. 2018), vol. 31(1): pp. 402–444.

[NV14] Nguyen, Hoi H and Van Vu: 'Random matrices: Law of the determinant'. *The Annals of Probability* (2014), vol. 42(1): pp. 146–167.

[NS06] Nica, Alexandru and Roland Speicher: *Lectures on the combinatorics of free probability*. Vol. 13. Cambridge University Press, 2006.

[ORSV15] O'Rourke, Sean, David Renfrew, Alexander Soshnikov, and Van Vu: 'Products of independent elliptic random matrices'. *Journal of Statistical Physics* (2015), vol. 160(1): pp. 89–119.

[OS11] O'Rourke, Sean and Alexander Soshnikov: 'Products of independent non-Hermitian random matrices'. *Electronic Journal of Probability* (2011), vol. 16: pp. 2219–2245.

[OW19] O'Rourke, Sean and Noah Williams: *Partial linear eigenvalue statistics for non-Hermitian random matrices*. 2019. arXiv: `1912.08856 [math.PR]`.

[PZ10] Pan, Guangming and Wang Zhou: 'Circular law, extreme singular values and potential theory'. *Journal of Multivariate Analysis* (2010), vol. 101(3): pp. 645–656.

[Pas73] Pastur, Leonid Andreevich: 'Spectra of random self adjoint operators'. *Russian mathematical surveys* (1973), vol. 28(1): p. 1.

[PS11] Pastur, Leonid Andreevich and Mariya Shcherbina: *Eigenvalue distribution of large random matrices*. 171. American Mathematical Soc., 2011.

[Pet12] Petrov, Valentin Vladimirovich: *Sums of independent random variables*. Vol. 82. Springer Science & Business Media, 2012.

[Rac91] Rachev, Svetlozar Todorov: *Probability metrics and the stability of stochastic models*. Vol. 269. John Wiley & Son Ltd, 1991.

[RA06] Rajan, Kanaka and Larry F. Abbott: 'Eigenvalue spectra of random matrices for neural networks'. *Physical review letters* (2006), vol. 97(18): p. 188104.

[Rid03] Rider, Brian: 'A limit theorem at the edge of a non-Hermitian random matrix ensemble'. *Journal of Physics A: Mathematical and General* (2003), vol. 36(12): p. 3401.

[Rid04] Rider, Brian: 'Deviations from the circular law'. *Probability theory and related fields* (2004), vol. 130(3): pp. 337–367.

[RS06] Rider, Brian C. and Jack W. Silverstein: 'Gaussian fluctuations for non-Hermitian random matrix ensembles'. *The Annals of Probability* (2006), vol. 34(6): pp. 2118–2143.

[RV07] Rider, Brian C. and Bálint Virág: 'The noise in the circular law and the Gaussian free field'. *International Mathematics Research Notices* (2007), vol. 2007.

[Rud08] Rudelson, Mark: 'Invertibility of random matrices: norm of the inverse'. *Annals of Mathematics* (2008), vol.: pp. 575–600.

[RT19] Rudelson, Mark and Konstantin Tikhomirov: 'The sparse circular law under minimal assumptions'. *Geometric and Functional Analysis* (Apr. 2019), vol. 29(2): pp. 561–637.

[RV08] Rudelson, Mark and Roman Vershynin: 'The Littlewood–Offord problem and invertibility of random matrices'. *Advances in Mathematics* (2008), vol. 218(2): pp. 600–633.

[ST13] Saff, Edward B and Vilmos Totik: *Logarithmic potentials with external fields*. Vol. 316. Springer Science & Business Media, 2013.

[Saz68] Sazonov, Vasiliy V.: 'On the multi-dimensional central limit theorem'. *Sankhyā: The Indian Journal of Statistics, Series A* (1968), vol.: pp. 181–204.

[Sch] Schmüdgen, Konrad: *The moment problem*. Vol. 9. Springer.

[Ser15] Serfaty, Sylvia: *Coulomb gases and Ginzburg–Landau vortices*. 2015.

[Ser17] Serfaty, Sylvia: *Microscopic description of log and Coulomb gases*. 2017. arXiv: `1709.04089 [math-ph]`.

[Sni02] Sniady, Piotr: 'Random Regularization of Brown Spectral Measure'. *Journal of Functional Analysis* (2002), vol. 193: pp. 291–313.

[Som07] Sommers, Hans-Jürgen: 'Symplectic structure of the real Ginibre ensemble'. *Journal of Physics A: Mathematical and Theoretical* (2007), vol. 40(29): F671.

[SCSS88] Sommers, H.J., A. Crisanti, H. Sompolinsky, and Y. Stein: 'Spectrum of large random asymmetric matrices'. *Physical review letters* (1988), vol. 60(19): p. 1895.

[Sos99] Soshnikov, Alexander: 'Universality at the Edge of the Spectrum in Wigner Random Matrices'. *Communications in mathematical physics* (1999), vol. 207(3): pp. 697–733.

[ST70] Springer, Melvin Dale and William E. Thompson: 'The distribution of products of beta, gamma and Gaussian random variables'. *SIAM Journal on Applied Mathematics* (1970), vol. 18(4): pp. 721–737.

[Ste92] Stephens, MA: 'Introduction to Kolmogorov (1933) on the empirical determination of a distribution'. *Breakthroughs in statistics*. Springer, 1992: pp. 93–105.

[TV08] Tao, Terence and Van Vu: 'Random matrices: the circular law'. *Communications in Contemporary Mathematics* (2008), vol. 10(02): pp. 261–307.

[TV09a] Tao, Terence and Van Vu: 'From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices'. *Bulletin of the American Mathematical Society* (2009), vol. 46(3): pp. 377–396.

[TV10a] Tao, Terence and Van Vu: *Random Matrices: Localization of the eigenvalues and the necessity of four moments*. 2010. arXiv: `1005.2901 [math.PR]`.

[TV10b] Tao, Terence and Van Vu: 'Random matrices: Universality of ESDs and the circular law'. *The Annals of Probability* (2010), vol. with an appendix by Manjunath Krishnapur: pp. 2023–2065.

[TV11] Tao, Terence and Van Vu: 'Random matrices: universality of local eigenvalue statistics'. *Acta mathematica* (2011), vol. 206(1): pp. 127–204.

[TV12] Tao, Terence and Van Vu: 'Random matrices: the universality phenomenon for Wigner ensembles'. *Modern aspects of random matrix theory* (2012), vol. 72: pp. 121–172.

[TV14] Tao, Terence and Van Vu: 'Local universality of zeroes of random polynomials'. *International Mathematics Research Notices* (2014), vol. 2015(13): pp. 5053–5139.

[TV15] Tao, Terence and Van Vu: 'Random matrices: universality of local spectral statistics of non-Hermitian matrices'. *The Annals of Probability* (2015), vol. 43(2): pp. 782–874.

[TV09b] Tao, Terence and Van H Vu: 'Inverse Littlewood-Offord theorems and the condition number of random discrete matrices'. *Annals of Mathematics* (2009), vol.: pp. 595–632.

[TDH76] Topsøe, Flemming, Richard M Dudley, and Jørgen Hoffmann-Jørgensen: 'Two examples concerning uniform convergence of measures wrt balls in Banach spaces'. *Empirical Distributions and Processes*. Springer, 1976: pp. 141–146.

[TW94] TRACY, CRAIG A and HAROLD WIDOM: 'Level-spacing distributions and the Airy kernel'. *Communications in Mathematical Physics* (1994), vol. 159(1): pp. 151–174.

[TVW13] TRAN, LINH V, VAN H VU, and KE WANG: 'Sparse random graphs: Eigenvalues and eigenvectors'. *Random Structures & Algorithms* (2013), vol. 42(1): pp. 110–134.

[TE05] TREFETHEN, LLOYD N and MARK EMBREE: *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*. Princeton University Press, 2005.

[TV04] TULINO, ANTONIA M and SERGIO VERDÚ: 'Random matrix theory and wireless communications'. *Foundations and Trends® in Communications and Information Theory* (2004), vol. 1(1): pp. 1–182.

[WW96] WHITTAKER, EDMUND TAYLOR and GEORGE NEVILLE WATSON: *A course of modern analysis*. Cambridge university press, 1996.

[Wig55] WIGNER, EUGENE P.: 'Characteristic Vectors of Bordered Matrices With Infinite Dimensions'. *Annals of Mathematics* (1955), vol.: pp. 548–564.

[Wig58] WIGNER, EUGENE P.: 'On the Distribution of the Roots of Certain Symmetric Matrices'. *Annals of Mathematics* (1958), vol. 67(2): pp. 325–327.

[Wis28] WISHART, JOHN: 'The generalized product moment distribution in samples from a normal multivariate population'. *Biometrika* (Dec. 1928), vol. 20A(1-2): pp. 32–52.

[XGL19] XIAO, HUI, ION GRAMA, and QUANSHENG LIU: *Berry-Esseen bound and precise moderate deviations for products of random matrices*. 2019. arXiv: `1907.02438 [math.PR]`.

[Yin14] YIN, JUN: 'The local circular law III: general case'. *Probability Theory and Related Fields* (2014), vol. 160(3-4): pp. 679–732.