Fakultät für
Mathematik und Informatik

Michael Fleermann

Global and Local Semicircle Laws for Random Matrices with Correlated Entries

Dissertation

FernUniversität in Hagen

FernUniversität in Hagen

Global and Local Semicircle Laws for Random Matrices with Correlated Entries

by

Michael Fleermann

DISSERTATION

submitted for the degree of Doctor of Natural Sciences (Dr. rer. nat.)
at the Faculty of Mathematics and Computer Science
of the FernUniversität in Hagen

April 2019

First Referee: Prof. Dr. Werner Kirsch, FernUniversität in Hagen
Second Referee: Prof. Dr. Thomas Kriecherbauer, Universität Bayreuth
Date of Submission: April 2nd, 2019
Date of Defence: June 11th, 2019

Dedicated to my parents.

Statutory Declaration

I declare that this dissertation has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgement, the work presented is entirely my own.

Hagen, 2nd of April 2019 Michael Fleermann

Acknowledgements

This work would not have been possible without the help of many people. First and foremost, I would like to thank my supervisor, Werner Kirsch, for his excellent supervision which allowed me to grow as a researcher and mathematician. In the first part of the results (Chapter 4) he directed my research through additional research questions which I then investigated and answered using my own ideas. Here, he always encouraged me to seek for improvement of results. Owing (hopefully) to the trust in my abilities, he subsequently left the investigation and development of the second part of the results (Chapter 6) to myself. Last but not least, he allowed me to participate in the 2018 Summer School on Random Matrices in Ann Arbor, Michigan, and in two summer schools in Ghiffa, at the beautiful Lago Maggiore, for which I am very grateful. In this context, I would also like to address my sincere thanks to Thomas Kriecherbauer for helpful comments during the presentation of my results.

Second, I would like to thank Matthias Löwe and Arnoud den Boer for their continued trust in me. I have had the pleasure to be Matthias' student at the University of Münster throughout various courses, seminars and theses. He showed me that mathematics is not only about proofs, but also about people. With Arnoud I have had the pleasure to work on a project at the University of Twente, which I keep in good memory. He encouraged me to follow my passion in mathematics.

I would also like to thank the other professorial staff at the University of Münster (where I obtained a B.Sc. and an M.Sc. in mathematics) for their excellent lectures, supervision, encouragement and professionalism. Even up until now, I draw from the entrepreneurial and exploratory spirits that roam the hallways at their math faculty.

Turning to professors, colleagues and friends at the FernUniversität in Hagen, I would like to thank Hayk Asatryan, Eugen Grycko, Michael Hartz, Torsten Linß, Thomas Müller, Wolfgang Spitzer and Gabor Toth for many interesting and helpful conversations. Further, I am grateful to Dominique Andres, Helena Bergold, Winfried Hochstättler, Johanna Wiehe and Volker Winkler for being partners in crime when performing our very own interpretation of Christmas music at the Christmas party of our faculty. I thank Monika Düsterer, Anke Lüneborg, Sandra Schwarz and Jolanda van der Noll for being great company during countless visits to the canteen.

Last but not least, I would like to express my deepest gratitude to my parents and my wife Michelle, who supported me throughout my academic endeavors so far and without whom all of this would not have been possible.

Contents

1 Introduction

The theory of random matrices had its origins in the applications, namely in statistics, where John Wishart analyzed properties of multivariate normal populations (see [67]), and mathematical physics, where Eugene Wigner studied energy levels of heavy nuclei (see [64] and [65]). As quantum mechanics predicts, these energy levels are eigenvalues of self-adjoint operators. As described in [27] and [14], Wigner depicted these operators by large dimensional random matrices with independent entries. He found that asymptotically, the empirical distribution of the eigenvalues has a semicircular shape, which led to the famous *Wigner's semicircle law*.

In his experiments with heavy nuclei, Wigner also analyzed the distribution of the gaps in the set of energy levels, which he found to be independent of the underlying material, thus *universal*. Surprisingly, this gap distribution was successfully reproduced by his random matrix models. For Gaussian ensembles with independent entries and so that the distribution of the entire ensemble is invariant under conjugation by orthogonal/unitary matrices, Dyson, Gaudin and Mehta were able to analytically compute the exact gap distribution. This gave rise to the Wigner-Dyson-Mehta (WDM) universality conjecture, which states that local spectral statistics of random matrices should be independent of the exact distribution of their entries, and coincide with the Gaussian case. First breakthroughs in proving WDM universality were achieved for invariant ensembles, whose entries were not necessarily Gaussian (and then necessarily not independent) anymore. But due to a lack of analytical tools or concepts, progress was very slow for general random matrices with independent entries. Eventually in 2009, the so-called *local law* was developed, which turned out to be a powerful tool both to prove the WDM-conjecture for Wigner matrices and to give insights into the mechanisms that govern convergence of the empirical distribution of the eigenvalues to the semicircle distribution, the latter being the main focus of this text.

Ever since the historical developments just described, the reach of the theory of random matrices has grown tremendously, with fruitful interactions in the fields of information theory (e.g. wireless communication, see [62]), biology (e.g. RNA analysis, see [5]) and pure mathematics (e.g. free probability, see [47]).

But what are random matrices? In the context of the present thesis, a *random matrix* is an Hermitian $n \times n$ matrix X_n, whose entries $X_n(i,j)$ are real or complex random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then X_n possesses n real eigenvalues $\lambda_1^{X_n} \leq \ldots \leq \lambda_n^{X_n}$, all of them random. We want to analyze the following problem: Given a very large dimensional random matrix (choosing n very large) and picking uniformly at random one of the eigenvalues, where on the real line will this randomly picked random eigenvalue be located? Of course, the outcome of this experiment will follow a certain probability distribution. To answer the question, we form the empirical spectral distribution (ESD)

$$\sigma_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i^{X_n}},$$

where for each $a \in \mathbb{R}$ we denote by δ_a the Dirac measure in a. It is clear that σ_n is actually a *random probability measure*, for it depends on the realization of the eigenvalues, which in turn depend on the realization of the matrix entries. For each interval $(a, b) \subseteq \mathbb{R}$, the (random) number $\sigma_n((a, b))$ yields the proportion of the eigenvalues that fall into this interval. In other words, this is the (empirical) probability that a randomly picked eigenvalue will lie in the interval (a, b).

Given a sequence $(X_n)_n$, where each X_n is an $n \times n$ random matrix, we obtain a sequence $(\sigma_n)_n$ of random probability measures, and we can analyze its weak convergence in some probabilistic sense. For example, a common version of *Wigner's semicircle law* states that if all entries of X_n are standardized random variables which are independent (up to the symmetry constraint), identically distributed and possess moments of all orders, the sequence σ_n converges weakly almost surely to the semicircle distribution σ on the real line given by its Lebesgue density $\frac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{[-2,2]}(x)$ in x (see [6]). That is, we find a set $A \in \mathcal{A}$ with $\mathbb{P}(A) = 1$, such that for all $\omega \in A$ we have that $\sigma_n(\omega) \to \sigma$ weakly as $n \to \infty$. Wigner's semicircle law can be viewed as the central limit theorem in random matrix theory. It has been developed by Wigner in his papers [64] and [65].

In classical probability theory, the central limit theorem holds even if random variables are mildly correlated. Therefore, in the context of random matrices, a natural question to ask is whether one can relax the assumption of independence in Wigner's semicircle law and still obtain the semicircle distribution as a limit distribution of the ESDs. Such matrix ensembles with correlated entries have been studied, for example, in the papers [55], [33], [32], [37], [15], [30] or [42]. Another interesting way to relax the original assumptions is to study (periodic and non-periodic) band matrices instead of full matrices. Roughly speaking, band matrices are obtained from regular matrices by symmetrically setting certain off-diagonals to zero, thus losing randomness in the system. The ESDs of random band matrices with independent entries have been studied, for example, in [11] and [14]. Another fruitful and interesting way to deviate from the classical analysis is to study *local* convergence behavior of the ESDs to the semicircle distribution. Just as there are local versions of the central limit theorem (see [34], for example), local versions of the semicircle law have been established in random matrix theory, starting with [25] in 2009. These local laws provide very detailed insight into how exactly convergence against the limiting distribution takes place *on very small intervals.*

This dissertation will address all the extensions to the classical Wigner's semicircle law that we just mentioned. While doing so, it is our goal to provide a rather self-contained exposition that is not only aimed at the expert in the field, but should also be understandable to newcomers with only little or no background in random matrix theory. We will now describe how this dissertation is organized and which contributions it makes. Some of these contributions lie only in the detail, others might be considered folklore knowledge that just has not been written down adequately, yet, and then there are, of course, the bigger contributions that make up the main work in this text:

In Chapter 2, we will introduce and analyze in depth the concept of weak convergence for probability measures and random probability measures. Concerning random probability measures, we show in Lemma 2.18 that these are exactly stochastic kernels, the latter being a concept that is known from introductory classes in probability theory. A derivative of a random probability measure μ is its expected measure $\mathbb{E}\mu$, which plays a major role in random matrix theory. In Theorem 2.20 we point out some intricacies that arise when integrating with respect to $\mathbb{E}\mu$. We go on to define the stochastic convergence types of random probability measures, namely weak convergence almost surely, in probability and in expectation. In the literature of random matrix theory, especially the concept of weak convergence in probability is not well motivated, nor characterized. We will do so in Definition 2.24 and Theorem 2.25. A key observation that we made (which helps not only with the proof of Theorem 2.25) lies in Lemma 2.28.

In Chapter 3, we introduce the method of moments, a tool to derive weak convergence of deterministic probability measures. It postulates that probability measures converge weakly if their moments converge. Theorem 3.5 clarifies how the method of moments can be extended to random probability measures: Random probability measures converge weakly in expectation resp. in probability resp. almost surely if their random moments converge in the same sense. For the proof of this theorem, we make use of Lemma 2.28 again, and our assumptions are very mild. In particular, we do not need compactness of the target probability distribution, which is (thus unnecessarily) used in texts about random matrix theory, such as [6] or [39]. A highlight of Chapter 3 is Theorem 3.14, which can be considered the method of moments for random matrix theory.

In Chapter 4, we present our first "hard" results of the thesis, using the method of moments, Theorem 3.14 ii) and iii) with $z = 2$, thus analyzing the variance of the random moments. We strengthen the publication [37], where for "almost uncorrelated" random matrix ensembles, the semicircle law was shown in probability. We extend their results to be valid almost surely and for band matrices, where in the latter case we need to impose conditions on the bandwidth (resp. halfwidth) of the periodic (resp. non-periodic) band matrices to secure almost sure convergence. We also mildly generalize the model studied in [37] from "almost uncorrelated" to "α-almost uncorrelated" schemes. Here, $\alpha > 0$ is parameter that controls the correlation decay in the ensemble. As we will point out in Remark 4.27, α-almost uncorrelated ensembles appear quite naturally when random matrices with correlated Gaussian entries are studied. The main theorem of Chapter 4 is Theorem 4.9 for periodic random band matrices, including full matrices as a special case. Since this statement is multi-dimensional, we will draw many corollaries that exemplify its reach. Examples that fit almost uncorrelated random matrix ensembles are those matrices filled with Curie-Weiss or Gaussian random variables. Here, the Curie-Weiss distribution is a model for the behavior of ferromagnetic particles (spins) at the inverse temperature β. At low temperatures, that is, if β is large, all magnetic spins are likely to have the same alignment, resembling a strong magnetic effect. On the contrary, at high temperatures (if β is small), spins can act almost independently, resembling a weak magnetic effect. In the end of Chapter 4, we use an elegant argument in Theorem 4.46, which is also one of the main contributions, that allows us to infer asymptotic equivalence of the ESDs of periodic and non-periodic random band matrices. Therefore, it will help us to extend our results of Theorem 4.9 to non-periodic random band matrices.

In Chapter 5, we introduce the Stieltjes-transform of finite measures on $(\mathbb{R}, \mathcal{B})$. Just as the method of moments, the Stieltjes transform method is a very popular and established tool in random matrix theory. However, its relationship to the underlying ESD is much closer than the moments are. For example, the imaginary part acts as a Lebesgue density for a probability measure approximating the ESD, and the ESD can be retrieved easily and constructively from its Stieltjes transform. This relationship will be analyzed in detail in Section 5.3 and is the main reason that the Stieltjes transform can be used to greatly enhance knowledge about the convergence mechanisms in semicircle laws.

This leads directly to Chapter 6, where we will derive local laws of various types for ensembles which we call "of Curie-Weiss type." Of course, random matrices with Curie-Weiss distributed random variables will be of Curie-Weiss type, but so far only for inverse temperatures $\beta \leq 1$. The local laws now enhance precision of dynamic aspects in random matrix theory. It gives strong probability bounds on the events that $|\sigma_n(I_n) - \sigma(I_n)|$ converges to zero in probability, where $(I_n)_n$ is a sequence of intervals whose diameters do not decrease too quickly.

2 Weak Convergence

2.1 Spaces of Continuous Functions

On the set $\mathbb{R}$ of real numbers we will always consider the standard topology and the associated Borel σ-algebra $\mathcal{B}$. To study convergence of probability measures on $(\mathbb{R}, \mathcal{B})$, it is useful to get acquainted with certain spaces of functions $\mathbb{R} \to \mathbb{R}$ first. If $f : \mathbb{R} \to \mathbb{R}$ is a function, we define the *support* of f as

$$\operatorname{supp}(f) := \overline{\{x \in \mathbb{R} : f(x) \neq 0\}}.$$

Note that by definition, the support of f is always a closed subset of $\mathbb{R}$, and it is immediate that a point $x \in \mathbb{R}$ lies in the support of f if and only if for any $\epsilon > 0$ there is a $y \in B_\epsilon(x)$, such that $f(y) \neq 0$. Here and later, $B_\delta(z)$ denotes the open δ-ball around the element z in a metric space which is clear from the context.

We say that a function $f : \mathbb{R} \to \mathbb{R}$ *vanishes at infinity*, if

$$\lim_{x \to \pm\infty} f(x) = 0.$$

Denote by $\mathcal{C}(\mathbb{R})$ the vector space of continuous functions $\mathbb{R} \to \mathbb{R}$. We define the three subspaces

1. $\mathcal{C}_b(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \,|\, f \text{ is continuous and bounded}\}$,
2. $\mathcal{C}_0(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \,|\, f \text{ is continuous and vanishes at infinity}\}$ and
3. $\mathcal{C}_c(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \,|\, f \text{ is continuous with compact support}\}$.

It is clear that

$$\mathcal{C}_c(\mathbb{R}) \subsetneq \mathcal{C}_0(\mathbb{R}) \subsetneq \mathcal{C}_b(\mathbb{R}) \subsetneq \mathcal{C}(\mathbb{R}),$$

since the function $x \mapsto \min(1, 1/|x|)$ lies in $\mathcal{C}_0(\mathbb{R}) \backslash \mathcal{C}_c(\mathbb{R})$, the function $x \mapsto \mathbb{1}_{\mathbb{R}}(x)$ lies in $\mathcal{C}_b(\mathbb{R}) \backslash \mathcal{C}_0(\mathbb{R})$ and the function $x \mapsto x$ lies in $\mathcal{C}(\mathbb{R}) \backslash \mathcal{C}_b(\mathbb{R})$. Since all functions in $\mathcal{C}_c(\mathbb{R})$, $\mathcal{C}_0(\mathbb{R})$ and $\mathcal{C}_b(\mathbb{R})$ are bounded, we can equip these spaces with the supremum norm $\|\cdot\|_\infty$ defined by

$$\|f\|_\infty := \sup_{x \in \mathbb{R}} |f(x)|.$$

From now on, we will always consider the spaces $\mathcal{C}_b(\mathbb{R})$, $\mathcal{C}_0(\mathbb{R})$ and $\mathcal{C}_c(\mathbb{R})$ as vector spaces normed by the supremum norm. Convergence with respect to this norm is also called *uniform convergence*. To analyze properties of these normed spaces, we introduce continuous cutoff-functions as in [41, p. 8]:

Definition 2.1. For any real numbers $L > R \geq 0$ we define the function $\phi_{R,L} : \mathbb{R} \to [0,1]$ by

$$\phi_{R,L}(x) := \begin{cases} 1 & \text{if } |x| \leq R, \\ \frac{L-|x|}{L-R} & \text{if } R < |x| < L, \\ 0 & \text{if } |x| \geq L. \end{cases}$$

Note that for any $L > R \geq 0$, $\phi_{R,L}$ is continuous with compact support $[-L, L]$. The following theorem will summarize important properties of $\mathcal{C}_b(\mathbb{R})$, $\mathcal{C}_0(\mathbb{R})$ and $\mathcal{C}_c(\mathbb{R})$. We give a full proof for the convenience of the reader. Parts of the proof are taken from [41].

Theorem 2.2. *The following statements hold:*

i) $\mathcal{C}_b(\mathbb{R})$ *is complete, but not separable.*

ii) $\mathcal{C}_0(\mathbb{R})$ *is complete and separable.*

iii) $\mathcal{C}_c(\mathbb{R})$ *is not complete, but separable.*

iv) $\mathcal{C}_c(\mathbb{R})$ *is dense in* $\mathcal{C}_0(\mathbb{R})$*.* $\mathcal{C}_0(\mathbb{R})$ *is the completion of* $\mathcal{C}_c(\mathbb{R})$*.*

Proof. i) We first show that $\mathcal{C}_b(\mathbb{R})$ is complete: If $(f_n)_n$ is a Cauchy sequence in $\mathcal{C}_b(\mathbb{R})$, and $x \in \mathbb{R}$ is arbitrary, then $f_n(x)$ is a Cauchy sequence in $\mathbb{R}$, thus converges to a limit $f(x) \in \mathbb{R}$. We need to show that the function $x \mapsto f(x) := \lim_{n\to\infty} f_n(x)$ is continuous and bounded, and that $f_n \to f$ uniformly. To show the latter, let $\epsilon > 0$ be arbitrary, then choose $N \in \mathbb{N}$ so large that $\|f_n - f_m\|_\infty \leq \epsilon$ for all $m, n \geq N$. Then let $n \geq N$ and $x \in \mathbb{R}$ be arbitrary. Then we have for $m \geq N$ arbitrary that

$$|f(x) - f_n(x)| \leq |f(x) - f_m(x)| + |f_m(x) - f_n(x)| \leq |f(x) - f_m(x)| + \epsilon.$$

Taking the limit over m yields $|f(x) - f_n(x)| \leq \epsilon$. Therefore, since $n \geq N$ and $x \in \mathbb{R}$ were arbitrary, $\|f - f_n\|_\infty \leq \epsilon$ for all $n \geq N$, which shows $f_n \to f$ uniformly. This also implies that f is bounded, since there is an $n \in \mathbb{N}$ such that $\|f - f_n\|_\infty \leq 1$, so for $x \in \mathbb{R}$ arbitrary we have

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq 1 + \|f_n\|_\infty,$$

which yields $\|f\|_\infty \leq 1 + \|f_n\|_\infty < \infty$.

To show that f is continuous, let $x_n \to x$ in $\mathbb{R}$. Pick $\epsilon > 0$ arbitrary, then choose m large enough, such that $\|f_m - f\|_\infty \leq \frac{\epsilon}{3}$, then chose $N \in \mathbb{N}$ such that for all $n \geq N$ we find that $|f_m(x_n) - f_m(x)| \leq \frac{\epsilon}{3}$. Then it holds for all $n \geq N$:

$$|f(x_n) - f(x)| \leq |f(x_n) - f_m(x_n)| + |f_m(x_n) - f_m(x)| + |f_m(x) - f(x)| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

and thus $f(x_n) \to f(x)$ as $n \to \infty$.

Next, we show that $\mathcal{C}_b(\mathbb{R})$ is not separable, where we proceed as in [41, p. 9]. The idea is to construct an uncountable subset $\mathcal{F} \subseteq \mathcal{C}_b$, such that for all $f, g \in \mathcal{F}$ with $f \neq g$ we have $\|f-g\|_\infty = 1$. To this end, denote by Z the set of 0-1-sequences, so $Z = \{0,1\}^{\mathbb{N}}$. Note that Z is uncountable. For any sequence $z \in Z$ we define

$$\forall\, x \in \mathbb{R} : F_z(x) \coloneqq \sum_{i\in\mathbb{N}} z_i \cdot \phi_{\frac{1}{10},\frac{2}{10}}(x-i)$$

and $\mathcal{F} \coloneqq \{F_z \,|\, z \in Z\}$. Note that all F_z are $[0,1]$-valued and continuous. Since $F_z(i) = z_i$ for all $z \in Z$ and $i \in \mathbb{N}$ we find that $F_z \neq F_{z'}$ for $z \neq z' \in Z$, and even $\|F_z - F_{z'}\|_\infty = 1$ for $z \neq z' \in Z$. Therefore, if $\mathcal{G} \subseteq \mathcal{C}_b(\mathbb{R})$ is any dense subset, for all $z \in Z$ there must be a $G_z \in \mathcal{G}$ for which $\|F_z - G_z\|_\infty \leq \frac{1}{4}$ holds. But then we find for $z \neq z'$ that

$$\begin{aligned}\|G_z - G_{z'}\|_\infty &= \|F_z - F_{z'} - (F_z - G_z) - (G_{z'} - F_{z'})\|_\infty \\ &\geq \|F_z - F_{z'}\|_\infty - \|F_z - G_z\|_\infty - \|F_{z'} - G_{z'}\|_\infty \geq 1 - \frac{1}{4} - \frac{1}{4} = \frac{1}{2},\end{aligned}$$

so $G_z \neq G_{z'}$. Therefore, $\mathcal{G}$ has an uncountable subset and can thus not be countable.

iii) To show that $\mathcal{C}_c(\mathbb{R})$ is not complete, we show that it is not closed in the strict superset $\mathcal{C}_0(\mathbb{R})$. In fact, we show even more, that is, that $\mathcal{C}_c(\mathbb{R})$ is dense in $\mathcal{C}_0(\mathbb{R})$ (then since $\mathcal{C}_c(\mathbb{R}) \subsetneq \mathcal{C}_0(\mathbb{R})$, $\mathcal{C}_c(\mathbb{R})$ cannot be closed). This fact is also needed for statements ii) and iv). So let $f \in \mathcal{C}_0(\mathbb{R})$ be arbitrary. Now consider the sequence of functions $(f_n)_n$, where

$$\forall\, n \in \mathbb{N} : \forall\, x \in \mathbb{R} : \; f_n(x) \coloneqq \phi_{n,n+1}(x) f(x).$$

Then $(f_n)_n$ is a sequence in $\mathcal{C}_c(\mathbb{R})$ which converges uniformly to f. To see this, let $\epsilon > 0$ be arbitrary and $N \in \mathbb{N}$ be so large that for all $x \in \mathbb{R}$ with $|x| \geq N$ we have $|f(x)| \leq \epsilon$. Then for any $n \geq N$ we have $\|f - f_n\|_\infty \leq \epsilon$. Indeed, let $n \geq N$ be arbitrary, then since for any $x \in \mathbb{R}$ we find

$$|f(x) - f_n(x)| = |f(x) - \phi_{n,n+1}(x) f(x)| = |1 - \phi_{n,n+1}(x)| \cdot |f(x)|,$$

we have $|f(x) - f_n(x)| \leq |f(x)| \leq \epsilon$ for $|x| > n$ and $|f(x) - f_n(x)| = 0$ for $|x| \leq n$.

Next, we will show that $\mathcal{C}_c(\mathbb{R})$ is separable. To this end, denote by $\mathcal{P}$ the countable set of all polynomials with rational coefficients and set

$$\mathcal{Q} \coloneqq \{p \cdot \phi_{n,n+1} \,|\, p \in \mathcal{P}, n \in \mathbb{N}\}.$$

Then $\mathcal{Q}$ is a countable subset of $\mathcal{C}_c(\mathbb{R})$. Now let $f \in \mathcal{C}_c(\mathbb{R})$ and $\epsilon > 0$ be arbitrary. Since the support of f is compact, there is an $n \in \mathbb{N}$ such that $\operatorname{supp}(f) \subseteq [-n, n]$. It follows that $f = \phi_{n,n+1} f$. By the Weierstrass approximation theorem, we obtain a polynomial p with rational coefficients such that $|p(x) - f(x)| \leq \epsilon$ for all $x \in [-(n+1), n+1]$. Then for all $x \in [-(n+1), n+1]$ we find

$$|\phi_{n,n+1} p(x) - f(x)| = |\phi_{n,n+1}(p(x) - f(x))| \leq |\phi_{n,n+1}| \cdot |(p(x) - f(x))| \leq \epsilon$$

and for all $x \notin [-(n+1), n+1]$ we obtain $|\phi_{n,n+1} p(x) - f(x)| = 0$. As a result, $\phi_{n,n+1} \cdot p$ is ϵ-close to f.

ii) To show that $\mathcal{C}_0(\mathbb{R})$ is complete, let $(f_n)_n$ be an arbitrary Cauchy sequence in $\mathcal{C}_0(\mathbb{R})$. This is also a Cauchy sequence in $\mathcal{C}_b(\mathbb{R})$, so with i) we know that there is an $f \in \mathcal{C}_b(\mathbb{R})$ such that $f_n \to f$ uniformly. What is left to show is that f vanishes at infinity. To this end, let $\epsilon > 0$ be arbitrary and n so large that $\|f - f_n\|_\infty \leq \frac{\epsilon}{2}$. Then since f_n vanishes at infinity, we find an $R > 0$ so large that $|f_n(x)| \leq \frac{\epsilon}{2}$ whenever $|x| \geq R$. It follows for all $x \in \mathbb{R}$ with $|x| \geq R$ that

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Therefore, f vanishes at infinity. This shows that $\mathcal{C}_0(\mathbb{R})$ is complete, and to see that $\mathcal{C}_0(\mathbb{R})$ is separable, note that we have already seen that $\mathcal{C}_c(\mathbb{R})$ is separable and dense in $\mathcal{C}_0(\mathbb{R})$.
iv) The inclusion $\mathcal{C}_c(\mathbb{R}) \hookrightarrow \mathcal{C}_0(\mathbb{R})$ is an isometric embedding with dense image (as shown in the proof of iii)) and $\mathcal{C}_0(\mathbb{R})$ is complete as shown in the proof of ii), which makes $\mathcal{C}_0(\mathbb{R})$ the completion of $\mathcal{C}_c(\mathbb{R})$. $\square$

2.2 Convergence of Probability Measures

We will denote the set of measures on $(\mathbb{R}, \mathcal{B})$ by $\mathcal{M}(\mathbb{R})$, the set of finite measures by $\mathcal{M}_f(\mathbb{R})$, the set of probability measures by $\mathcal{M}_1(\mathbb{R})$, and the set of sub-probability measures by $\mathcal{M}_{\leq 1}(\mathbb{R})$. Here, a measure μ on $(\mathbb{R}, \mathcal{B})$ is called *sub-probability measure*, if $\mu(\mathbb{R}) \in [0,1]$. Note that

$$\mathcal{M}_1(\mathbb{R}) \subsetneq \mathcal{M}_{\leq 1}(\mathbb{R}) \subsetneq \mathcal{M}_f(\mathbb{R}) \subsetneq \mathcal{M}(\mathbb{R}).$$

As a shorthand notation, if $\mu \in \mathcal{M}(\mathbb{R})$ and $f : \mathbb{R} \to \mathbb{R}$ is measurable, we write

$$\langle \mu, f \rangle := \int f \, \mathrm{d}\mu$$

with the convention that when in doubt, x is the variable of integration:

$$\langle \mu, x^k \rangle = \int x^k \mu(\mathrm{d}x).$$

Definition 2.3. Let $\mathcal{F} \subseteq \mathcal{C}_b(\mathbb{R})$ be a linear subspace, then a *positive linear bounded functional* I on $\mathcal{F}$ is a bounded $\mathbb{R}$-linear map $\mathcal{F} \to \mathbb{R}$ with $I(f) \geq 0$ for all $f \in \mathcal{F}$ with $f \geq 0$.

Lemma 2.4. *Let $\mathcal{F} \subseteq \mathcal{C}_b(\mathbb{R})$ be a linear subspace with $\mathcal{C}_c(\mathbb{R}) \subseteq \mathcal{F}$. Then for any $\mu \in \mathcal{M}_f(\mathbb{R})$, the map*

$$\begin{aligned} I_\mu : \mathcal{F} &\longrightarrow \mathbb{R} \\ f &\longmapsto I_\mu(f) := \langle \mu, f \rangle \end{aligned}$$

defines a positive linear bounded functional on $\mathcal{F}$ with operator norm $\mu(\mathbb{R})$.

Proof. We only need to show that the operator norm is indeed $\mu(\mathbb{R})$. To see this, note that for any $K > 0$, we have $\phi_{K,K+1} \in \mathcal{F}$, $\phi_{K,K+1} \geq 0$ and $\|\phi_{K,K+1}\|_\infty = 1$. Further,

$$I_\mu(\phi_{K,K+1}) = \langle \mu, \phi_{K,K+1} \rangle \geq \mu([-K, K]).$$

Thus, the operator norm of I_μ is at least $\mu([-K, K])$ for all $K > 0$, hence at least $\mu(\mathbb{R})$. On the other hand, the operator norm is at most $\mu(\mathbb{R})$, since for any $f \in \mathcal{F}$ we find $|\langle \mu, f \rangle| \leq \langle \mu, |f| \rangle \leq \mu(\mathbb{R}) \cdot \|f\|_\infty$. □

The representation theorem of Riesz now states that *any* positive linear bounded functional I on $\mathcal{F} \in \{\mathcal{C}_c(\mathbb{R}), \mathcal{C}_0(\mathbb{R}), \mathcal{C}_b(\mathbb{R})\}$ has the form $I = I_\mu$ as in Lemma 2.4.

Theorem 2.5. *Let $\mathcal{F} \in \{\mathcal{C}_c(\mathbb{R}), \mathcal{C}_0(\mathbb{R}), \mathcal{C}_b(\mathbb{R})\}$ be equipped with the supremum norm. Then for any positive linear bounded functional I on $\mathcal{F}$, there exists exactly one $\mu \in \mathcal{M}_f(\mathbb{R})$ with $I = I_\mu$. It then holds $\|I\|_{\mathrm{op}} = \mu(\mathbb{R})$.*

Proof. The proof is rather lengthy. We refer the reader to [21], where the various representation theorems are discussed in detail. □

The next lemma will help us infer equality of two finite measures. Notationally, if A is a subset of a topological space, we denote its boundary by ∂A.

Lemma 2.6. *Let μ and ν be two finite measures on $(\mathbb{R}, \mathcal{B})$ and let $\mathcal{F} \subseteq \mathcal{C}_c(\mathbb{R})$ be a dense subset. Then*

i) $\mu = \nu \quad \Leftrightarrow \quad \mu(I) = \nu(I)$ *for all bounded intervals I with* $\mu(\partial I) = \nu(\partial I) = 0$,

ii) $\mu = \nu \quad \Leftrightarrow \quad \forall\, f \in \mathcal{C}_c(\mathbb{R}) : \langle \mu, f \rangle = \langle \nu, f \rangle \quad \Leftrightarrow \quad \forall\, f \in \mathcal{F} : \langle \mu, f \rangle = \langle \nu, f \rangle$.

Proof. i) "$\Rightarrow$" is clear, and for "$\Leftarrow$" we show that μ and ν agree on all finite open intervals. To this end, note that for any finite measure $\rho \in \mathcal{M}_f(\mathbb{R})$, the set of atoms $A_\rho := \{x \in \mathbb{R} \,|\, \rho(x) > 0\}$ is at most countable. As a result $\mathbb{R} \backslash (A_\mu \cup A_\nu)$ is dense in $\mathbb{R}$. For arbitrary $a < b$ in $\mathbb{R}$, we find sequences $(a_n)_n$ and $(b_n)_n$ in $\mathbb{R} \backslash (A_\mu \cup A_\nu)$ with $a_n \searrow a$ and $b_n \nearrow b$ as $n \to \infty$ and $a_n < b_n$ for all $n \in \mathbb{N}$. Then we obtain with continuity of measures from below (note that μ and ν agree on all intervals (a_n, b_n)):

$$\mu((a, b)) = \lim_{n\to\infty} \mu((a_n, b_n)) = \lim_{n\to\infty} \nu((a_n, b_n)) = \nu((a, b)).$$

ii) The two "$\Rightarrow$'s" are clear. Assume for all $f \in \mathcal{F}$ we have $\langle \mu, f \rangle = \langle \nu, f \rangle$. Now if $f \in \mathcal{C}_c(\mathbb{R})$ is arbitrary, we find a sequence $(f_n)_n$ in $\mathcal{F}$ such that $f_n \to f$ uniformly. Due to continuity of I_μ and I_ν on $\mathcal{C}_c(\mathbb{R})$ (see Theorem 2.5), we find

$$\langle \mu, f \rangle = \lim_{n\to\infty} \langle \mu, f_n \rangle = \lim_{n\to\infty} \langle \nu, f_n \rangle = \langle \nu, f \rangle\,,$$

hence $\langle \mu, f \rangle = \langle \nu, f \rangle$ for all $f \in \mathcal{C}_c(\mathbb{R})$. And if $\langle \mu, f \rangle = \langle \nu, f \rangle$ for all $f \in \mathcal{C}_c(\mathbb{R})$, we find $I_\mu = I_\nu$ on $\mathcal{C}_c(\mathbb{R})$, so $\mu = \nu$ with Theorem 2.5. □

We are especially interested in convergence behavior of sequences in $\mathcal{M}_1(\mathbb{R})$, where the limit may lie in $\mathcal{M}_{\leq 1}(\mathbb{R})$.

Definition 2.7. Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in in $\mathcal{M}_1(\mathbb{R})$.

i) The sequence $(\mu_n)_{n\in\mathbb{N}}$ is said to converge *weakly* to an element $\mu \in \mathcal{M}_1(\mathbb{R})$, if

$$\forall f \in \mathcal{C}_b(\mathbb{R}) : \lim_{n\to\infty} \langle \mu_n, f \rangle = \langle \mu, f \rangle . \tag{2.1}$$

ii) The sequence $(\mu_n)_{n\in\mathbb{N}}$ is said to converge *vaguely* to an element $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$, if

$$\forall f \in \mathcal{C}_c(\mathbb{R}) : \lim_{n\to\infty} \langle \mu_n, f \rangle = \langle \mu, f \rangle . \tag{2.2}$$

Remark 2.8. We would like to shed light on the seemingly innocent Definition 2.7:

1. Weak convergence clearly implies vague convergence. Further, due to Lemma 2.6, weak and vague limits are unique.

2. In light of Theorem 2.2, it is appropriate to say that the set of test functions for weak convergence is considerably larger than the set of test functions for vague convergence. As a result, weak limits are much more restrictive than vague limits, as clarified by the next two points.

3. The target measures $\mu \in \mathcal{M}(\mathbb{R})$, for which (2.1) can be satisfied for some sequence $(\mu_n)_n$ of probability measures are exactly all $\mu \in \mathcal{M}_1(\mathbb{R})$. To see this, if (2.1) holds for some $\mu \in \mathcal{M}(\mathbb{R})$ and a sequence $(\mu_n)_n$ in $\mathcal{M}_1(\mathbb{R})$, then we must have $\mu(\mathbb{R}) = 1$, since $\mathbb{1}_{\mathbb{R}} \in \mathcal{C}_b(\mathbb{R})$. On the other hand, if $\mu \in \mathcal{M}_1(\mathbb{R})$ is arbitrary, then (2.1) is satisfied for the sequence $(\mu_n)_n$, where $\mu_n = \mu$ for all $n \in \mathbb{N}$.

4. The measures $\mu \in \mathcal{M}(\mathbb{R})$, for which (2.2) can be satisfied for some sequence $(\mu_n)_n$ of probability measures are (somewhat surprisingly) exactly all $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$. To see this, if (2.2) holds for some $\mu \in \mathcal{M}(\mathbb{R})$ and a sequence $(\mu_n)_n$ in $\mathcal{M}_1(\mathbb{R})$, then we have for any $m \in \mathbb{N}$ that $\langle \mu_n, \phi_{m,m+1} \rangle \to_n \langle \mu, \phi_{m,m+1} \rangle$, so $\langle \mu, \phi_{m,m+1} \rangle \leq 1$, which entails $\mu([-m, m]) \leq 1$ for all $m \in \mathbb{N}$. Since measures are continous from below, we conclude that also $\mu(\mathbb{R}) \leq 1$, so μ is a sub-probability measure. On the other hand, if $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$ is arbitrary, then define $\alpha := 1 - \mu(\mathbb{R}) \in [0, 1]$ and for all $n \in \mathbb{N} : \mu_n := \mu + \alpha\delta_n$. Then $(\mu_n)_n$ is a sequence of probability measures and (2.2) is satisfied for the sequence $(\mu_n)_n$. To see this, let $f \in \mathcal{C}_c(\mathbb{R})$ be arbitrary and $N \in \mathbb{N}$ be so large that $\text{supp}(f) \subseteq [-N, N]$. Then it holds for all $n \geq N$ that $\langle \mu_n, f \rangle = \langle \mu, f \rangle + \alpha f(n) = \langle \mu, f \rangle$.

5. As a result of points 3. and 4., the limit domains for weak and vague convergence in Definition 2.7 are exact. The probability measures lie *vaguely dense* in the sub-probability measures.

Lemma 2.9. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of probability measures and μ a sub-probability measure on $(\mathbb{R},\mathcal{B})$. Then $(\mu_n)_{n\in\mathbb{N}}$ converges vaguely (resp. weakly) to μ if and only if every subsequence $(\mu_n)_{n\in J}$, $J\subseteq\mathbb{N}$, has a subsequence $(\mu_n)_{n\in I}$, $I\subseteq J$, that converges vaguely (resp. weakly) to μ.*

Proof. Of course, we only need to show "$\Leftarrow$". We assume the statement to be false, that is, that it is not true that $(\mu_n)_{n\in\mathbb{N}}$ converges vaguely (resp. weakly) to μ. Then we find a continuous function $f:\mathbb{R}\to\mathbb{R}$ which has compact support (resp. which is bounded) and an $\epsilon>0$ such that $|\langle\mu_n,f\rangle-\langle\mu,f\rangle|\geq\epsilon$ for all $n\in J$, where $J\subseteq\mathbb{N}$ is an infinite subset. But now we find a subsequence $(\mu_n)_{n\in I}$, $I\subseteq J$ that converges vaguely (resp. weakly) to μ. In particular, we find an $n\in I\subseteq J$ such that $|\langle\mu_n,f\rangle-\langle\mu,f\rangle|<\epsilon$, which leads to a contradiction to our assumption that the statement is false. □

Vague convergence of probability measures can also be characterized by convergence of the integrals $\langle\mu_n,f\rangle$ for all $f\in\mathcal{C}_0(\mathbb{R})$.

Lemma 2.10. *A sequence $(\mu_n)_n$ in $\mathcal{M}_1(\mathbb{R})$ converges vaguely to an element $\mu\in\mathcal{M}_{\leq 1}(\mathbb{R})$, if and only if*

$$\forall f\in\mathcal{C}_0(\mathbb{R}):\lim_{n\to\infty}\langle\mu_n,f\rangle=\langle\mu,f\rangle\,.$$

Proof. The condition is obviously sufficient for vague convergence. We now show necessity: Let $f\in\mathcal{C}_0(\mathbb{R})$ and then $\epsilon>0$ be arbitrary. Since $\mathcal{C}_c(\mathbb{R})\subseteq\mathcal{C}_0(\mathbb{R})$ is dense by Theorem 2.2, we find an $f_c\in\mathcal{C}_c(\mathbb{R})$ with $\|f-f_c\|_\infty\leq\epsilon/2$. Then

$$\begin{aligned}&|\langle\mu_n,f\rangle-\langle\mu,f\rangle|\\&\leq|\langle\mu_n,f\rangle-\langle\mu_n,f_c\rangle|+|\langle\mu_n,f_c\rangle-\langle\mu,f_c\rangle|+|\langle\mu,f_c\rangle-\langle\mu,f\rangle|\\&\leq\epsilon+|\langle\mu_n,f_c\rangle-\langle\mu,f_c\rangle|\end{aligned}$$

Since $\mu_n\to\mu$ vaguely, we obtain $\limsup_{n\to\infty}|\langle\mu_n,f\rangle-\langle\mu,f\rangle|\leq\epsilon$. Since $\epsilon>0$ was arbitrary, this yields $\lim_{n\to\infty}\langle\mu_n,f\rangle=\langle\mu,f\rangle$. □

If $\mu_n\to\mu$ weakly, we know that $\langle\mu_n,f\rangle\to\langle\mu,f\rangle$ for all $f\in\mathcal{C}_b(\mathbb{R})$. Often, we would like to be able to conclude $\langle\mu_n,f\rangle\to\langle\mu,f\rangle$ for more general functions f. The next lemma will be of great use in this respect, see also [20, p. 101].

Lemma 2.11. *Let $(\mu_n)_n$ and μ be probability measures such that $\mu_n\to\mu$ weakly as $n\to\infty$. Let $h:\mathbb{R}\to\mathbb{R}$ be continuous. Then to show*

$$\langle\mu_n,h\rangle\xrightarrow[n\to\infty]{}\langle\mu,h\rangle\,,$$

it is sufficient to show that there is a strictly positive continuous function $g:\mathbb{R}\to(0,\infty)$ such that h/g vanishes at infinity and $\sup_{n\in\mathbb{N}}\langle\mu_n,g\rangle<\infty$.

Proof. Let $C:=\sup_{n\in\mathbb{N}}\langle\mu_n,g\rangle\in[0,\infty)$. We first show that also $\langle\mu,g\rangle\leq C$. To this end, let $K>0$ be arbitrary, then $g\phi_{K,K+1}\in\mathcal{C}_b(\mathbb{R})$, so we know that

$$\langle\mu_n,g\phi_{K,K+1}\rangle\xrightarrow[n\to\infty]{}\langle\mu,g\phi_{K,K+1}\rangle\,.$$

Since for all $n \in \mathbb{N}$, $0 \leq \langle \mu_n, g\phi_{K,K+1} \rangle \leq \langle \mu_n, g \rangle \leq C$, also $\langle \mu, g\phi_{K,K+1} \rangle \in [0, C]$. But $K > 0$ was arbitrary. Therefore, by monotone convergence, we find

$$\langle \mu, g \rangle = \lim_{K \to \infty} \langle \mu, g\phi_{K,K+1} \rangle \in [0, C].$$

This shows that $\langle \mu, g \rangle \leq C$. Now let $\epsilon > 0$ be arbitrary, then $K > 0$ so large that $|h|/g \leq \epsilon/C$ on $[-K, K]^c$ (where if A is a set, we denote its complement by A^c, where we assume that the superset of A is clear from the context. For example, $[-K, K]^c = \mathbb{R} \backslash [-K, K]$). We conclude that for all $\nu \in \{\mu, (\mu_n)_n\}$,

$$|\langle \nu, h(1 - \phi_{K,K+1}) \rangle| \leq \left\langle \nu, \frac{|h|}{g} \cdot g(1 - \phi_{K,K+1}) \right\rangle \leq \frac{\epsilon}{C} \cdot C = \epsilon.$$

In particular, these integrals are well-defined. Since also for any $\nu \in \{\mu, (\mu_n)_n\}$, $\langle \nu, h\phi_{K,K+1} \rangle$ is well-defined, h is ν-integrable as a sum of ν-integrable functions. We find for $\epsilon > 0$ and $K > 0$ as picked above, that for all $n \in \mathbb{N}$:

$$\begin{aligned}
&|\langle \mu_n, h \rangle - \langle \mu, h \rangle| \\
&\leq |\langle \mu_n, h(1 - \phi_{K,K+1}) \rangle - \langle \mu, h(1 - \phi_{K,K+1}) \rangle| + |\langle \mu_n, h\phi_{K,K+1} \rangle - \langle \mu, h\phi_{K,K+1} \rangle| \\
&\leq \epsilon + |\langle \mu_n, h\phi_{K,K+1} \rangle - \langle \mu, h\phi_{K,K+1} \rangle|,
\end{aligned}$$

where the last summand converges to 0 as $n \to \infty$, such that

$$\limsup_{n \to \infty} |\langle \mu_n, h \rangle - \langle \mu, h \rangle| \leq \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we find $\langle \mu_n, h \rangle \to \langle \mu, h \rangle$ as $n \to \infty$. □

As we just saw in Remark 2.8, vague convergence allows the escape of probability mass. The concept of tightness prevents this from happening:

Definition 2.12. A sequence of probability measures $(\mu_n)_n$ on $(\mathbb{R}, \mathcal{B})$ is called *tight*, if for all $\epsilon > 0$ there exists a compact subset $K \subseteq \mathbb{R}$ such that

$$\forall\, n \in \mathbb{N}: \ \mu_n(K^c) \leq \epsilon.$$

A sufficient condition for tightness is given in the next Lemma, which we adopted from [20, p. 104]:

Lemma 2.13. *Let $(\mu_n)_n$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$. If there exists a measurable non-negative function $\phi : \mathbb{R} \to \mathbb{R}$ with $\phi(x) \to \infty$ for $x \to \pm\infty$ and*

$$\sup_n \langle \mu_n, \phi \rangle < \infty,$$

then $(\mu_n)_n$ is tight. In particular, this holds true if

$$\sup_n \left\langle \mu_n, x^2 \right\rangle < \infty.$$

Proof. Let $C \coloneqq \sup_n \langle \mu_n, \phi\rangle < \infty$. Then it holds for any $n \in \mathbb{N}$ and $K > 0$ that

$$C \geq \langle \mu_n, \phi\rangle \geq \left\langle \mu_n, \mathbb{1}_{[-K,K]^c} \cdot \inf_{|x|>K} \phi(x) \right\rangle = \left\langle \mu_n, \mathbb{1}_{[-K,K]^c} \right\rangle \cdot \inf_{|x|>K} \phi(x).$$

Since $\inf_{|x|>K} \phi(x) \to \infty$ as $K \to \infty$, the statement follows. □

Lemma 2.14. *Let $(\mu_n)_n$ be a sequence in $\mathcal{M}_1(\mathbb{R})$ and $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$ such that $\mu_n \to \mu$ vaguely as $n \to \infty$, then the following statements are equivalent:*

i) $(\mu_n)_n$ *is tight.*

ii) μ *is a probability measure.*

iii) μ_n *converges weakly to* μ.

Proof. *i) ⇒ iii)* Let $f \in \mathcal{C}_b(\mathbb{R})$ be arbitrary and set $s \coloneqq \max(\|f\|_\infty, 1)$. Let $\epsilon > 0$ be arbitrary, then due to tightness of $(\mu_n)_n$ and continuity from below of μ, we find a $K > 0$ such that $\mu_n([-K,K]^c) \leq \frac{\epsilon}{3s}$ and $\mu([-K,K]^c) \leq \frac{\epsilon}{3s}$. Now for $n \in \mathbb{N}$ arbitrary we find

$$\begin{aligned}
&|\langle \mu_n, f\rangle - \langle \mu, f\rangle| \\
&\leq |\langle \mu_n, f\rangle - \langle \mu_n, f\phi_{K,K+1}\rangle| + |\langle \mu_n, f\phi_{K,K+1}\rangle - \langle \mu, f\phi_{K,K+1}\rangle| + |\langle \mu, f\phi_{K,K+1}\rangle - \langle \mu, f\rangle| \\
&\leq \langle \mu_n, |f| \cdot |1 - \phi_{K,K+1}|\rangle + |\langle \mu_n, f\phi_{K,K+1}\rangle - \langle \mu, f\phi_{K,K+1}\rangle| + \langle \mu, |f| \cdot |\phi_{K,K+1} - 1|\rangle \\
&\leq s \cdot \frac{\epsilon}{3s} + |\langle \mu_n, f\phi_{K,K+1}\rangle - \langle \mu, f\phi_{K,K+1}\rangle| + s \cdot \frac{\epsilon}{3s}
\end{aligned}$$

Now choose $N \in \mathbb{N}$ so large that for all $n \geq N$ we have $|\langle \mu_n, f\phi_{K,K+1}\rangle - \langle \mu, f\phi_{K,K+1}\rangle| \leq \frac{\epsilon}{3}$. Then $|\langle \mu_n, f\rangle - \langle \mu, f\rangle| \leq \epsilon$ for all $n \geq N$. Since $\epsilon > 0$ was arbitrary, $\langle \mu_n, f\rangle \to \langle \mu, f\rangle$ as $n \to \infty$. Since $f \in \mathcal{C}_b(\mathbb{R})$ was arbitrary, $\mu_n \to \mu$ weakly as $n \to \infty$.

iii) ⇒ ii) This statement is obvious. Consider $\mathbb{1}_\mathbb{R} \in \mathcal{C}_b(\mathbb{R})$.

ii) ⇒ i). Let $\epsilon > 0$ be arbitrary. Then for $K > 0$ we find

$$\mu_n([-(K+1), K+1]) \geq \langle \mu_n, \phi_{K,K+1}\rangle \geq \langle \mu, \phi_{K,K+1}\rangle - |\langle \mu, \phi_{K,K+1}\rangle - \langle \mu_n, \phi_{K,K+1}\rangle|$$

Now first choose K large enough such that the first summand on the r.h.s. is larger than $1 - \epsilon/2$, then choose $N \in \mathbb{N}$ large enough such that for all $n > N$ the absolute value on the r.h.s. is at most $\epsilon/2$. Then we obtain for all $n > N$ that $\mu_n([-(K+1), K+1]) \geq 1 - \epsilon$. On the other hand, we find $K_1, \ldots, K_N > 0$ such that

$$\forall\, i \in \{1, \ldots, N\}: \;\; \mu_i([-K_i, K_i]) \geq 1 - \epsilon.$$

Let $K^* \coloneqq \max\{K+1, K_1, \ldots, K_N\}$, then we obtain for all $n \in \mathbb{N}$ that $\mu_n([-K^*, K^*]) \geq 1 - \epsilon$. Therefore, $(\mu_n)_n$ is tight.

□

Lemma 2.15. *Let $(\mu_n)_n$ be a sequence in $\mathcal{M}_1(\mathbb{R})$, then the following statements hold:*

i) $(\mu_n)_n$ has a vaguely convergent subsequence against some $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$.

ii) If $(\mu_n)_n$ is tight, it has a weakly convergent subsequence against some $\mu \in \mathcal{M}_1(\mathbb{R})$.

Proof. i) Let $(g_m)_m$ be a dense sequence in $\mathcal{C}_c(\mathbb{R})$, then for all $m \in \mathbb{N}$, $(\langle \mu_n, g_m \rangle)_n$ is a sequence in $\mathbb{R}$ whose absolute value is bounded by $\|g_m\|_\infty < \infty$, thus has a convergent subsequence by Bolzano-Weierstrass. By a diagonal argument, we can find a subsequence $J \subseteq \mathbb{N}$, such that for all $m \in \mathbb{N}$, $(\langle \mu_n, g_m \rangle)_{n \in J}$ converges. But since $(g_m)_m$ is dense in $\mathcal{C}_c(\mathbb{R})$, $\lim_{n \in J} \langle \mu_n, f \rangle$ exists for all $f \in \mathcal{C}_c(\mathbb{R})$ (it can be shown that $(\langle \mu_n, f \rangle)_n$ is Cauchy). The function

$$\begin{aligned} I : \mathcal{C}_c(\mathbb{R}) &\longrightarrow \mathbb{R} \\ f &\longmapsto I(f) := \lim_{n \in J} \langle \mu_n, f \rangle \end{aligned}$$

is a linear bounded positive functional on $\mathcal{C}_c(\mathbb{R})$ with operator norm at most 1, since $|\langle \mu_n, f \rangle| \leq \|f\|_\infty$ for all $n \in \mathbb{N}$ and $f \in \mathcal{C}_c(\mathbb{R})$. With Theorem 2.5, we find an element $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$ such that $I = I_\mu$, which entails $\mu_n \to \mu$ vaguely for $n \in J$.
ii) With *i)* we find a subsequence $J \subseteq \mathbb{N}$ and a $\mu \in \mathcal{M}_{\leq 1}(\mathbb{R})$ such that $(\mu_n)_{n \in J}$ converges to μ vaguely. But Lemma 2.14 yields that $\mu \in \mathcal{M}_1(\mathbb{R})$ and $\mu_n \to \mu$ weakly for $n \in J$. □

Note that statement *i)* of Lemma 2.15 is the well-known Helly's selection theorem contained in most standard books on probability theory, see [20] or [44], for example. However, we give a new proof here that differs completely from the standard proofs which utilize distribution functions.

So far we have discussed the intricacies of weak and vague convergence of probability measures. Our next goal is to better understand the topology of weak convergence on $\mathcal{M}_1(\mathbb{R})$, which will deepen our understanding of stochastic weak convergence to be discussed in the next section. Our first goal will be to reduce the number of test functions for weak convergence to a countable subset of $\mathcal{C}_b(\mathbb{R})$. However, $(\mathcal{C}_b(\mathbb{R}), \|\cdot\|_\infty)$ is large; it is not even separable. But there is no reason for despair, since the following theorem holds, which we adopted from our previous work [31].

Theorem 2.16. *Fix a sequence $(g_k)_{k \in \mathbb{N}}$ in $\mathcal{C}_c(\mathbb{R})$ which lies dense in $\mathcal{C}_c(\mathbb{R})$. Then the following statements hold:*

i) Let μ, $(\mu_n)_n \in \mathcal{M}_1(\mathbb{R})$, then the following statements are equivalent:

a) $\mu_n \to \mu$ weakly.

b) $\forall\, k \in \mathbb{N} : \langle \mu_n, g_k \rangle \xrightarrow[n \to \infty]{} \langle \mu, g_k \rangle$.

ii) Define for all $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$:

$$d_M(\mu, \nu) := \sum_{k \in \mathbb{N}} \frac{|\langle \mu, g_k \rangle - \langle \nu, g_k \rangle|}{2^k \cdot (1 + |\langle \mu, g_k \rangle - \langle \nu, g_k \rangle|)}.$$

Then d_M forms a metric on $\mathcal{M}_1(\mathbb{R})$ which metrizes weak convergence. That is, a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}_1(\mathbb{R})$ converges weakly to $\mu \in \mathcal{M}_1(\mathbb{R})$ iff $d_M(\mu_n, \mu) \to 0$ as $n \to \infty$.

iii) $(\mathcal{M}_1(\mathbb{R}), d_M)$ is a separable, but not complete, metric space.

Proof. i) Let $(\mu_n)_{n\in\mathbb{N}}$ and μ be probability measures. If $\mu_n \to \mu$ weakly, then surely we have for all $k \in \mathbb{N}$ that $\langle \mu_n, g_k \rangle \to \langle \mu, g_k \rangle$ as $n \to \infty$. If on the other hand we have for all $k \in \mathbb{N}$ that $\langle \mu_n, g_k \rangle \to \langle \mu, g_k \rangle$ as $n \to \infty$, then one can show that μ_n converges vaguely to μ, that is

$$\forall\, f \in \mathcal{C}_c(\mathbb{R}) : \langle \mu_n, f \rangle \xrightarrow[n\to\infty]{} \langle \mu, f \rangle \,.$$

To this end, let $f \in \mathcal{C}_c(\mathbb{R})$ and $\epsilon > 0$ be arbitrary, then there is an $l \in \mathbb{N}$ such that $\|f - g_l\|_\infty \leq \frac{\epsilon}{3}$. Since we know that $\langle \mu_n, g_l \rangle \to \langle \mu, g_l \rangle$ as $n \to \infty$, there is an $N \in \mathbb{N}$ so that for all $n \geq N$ we have

$$|\langle \mu_n, g_l \rangle - \langle \mu, g_l \rangle| \leq \frac{\epsilon}{3}.$$

Then it holds for all $n \geq N$:

$$\begin{aligned}
&|\langle \mu_n, f \rangle - \langle \mu, f \rangle| \\
\leq &|\langle \mu_n, f \rangle - \langle \mu_n, g_l \rangle| + |\langle \mu_n, g_l \rangle - \langle \mu, g_l \rangle| + |\langle \mu, g_l \rangle - \langle \mu, f \rangle| \\
\leq &\underbrace{\langle \mu_n, |f - g_l| \rangle}_{\leq \frac{\epsilon}{3}} + \underbrace{|\langle \mu_n, g_l \rangle - \langle \mu, g_l \rangle|}_{\leq \frac{\epsilon}{3}} + \underbrace{\langle \mu, |f - g_l| \rangle}_{\leq \frac{\epsilon}{3}} \leq \epsilon.
\end{aligned}$$

Since ϵ was arbitrary, it follows that

$$\lim_{n\to\infty} \langle \mu_n, f \rangle = \langle \mu, f \rangle \,.$$

Now since μ_n converges vaguely to μ and μ is a probability measure, we know by Lemma 2.14, that $\mu_n \to \mu$ weakly.

ii) and iii):
From Lemma 2.6, we find for any $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ that

$$\mu = \nu \;\Leftrightarrow\; \forall\, k \in \mathbb{N} : \langle \mu, g_k \rangle = \langle \nu, g_k \rangle \,.$$

Next, we will inspect the space $\mathbb{R}^{\mathbb{N}}$ endowed with the product topology. With respect to this topology, a sequence $(z_n)_n$ in $\mathbb{R}^{\mathbb{N}}$ converges to a $z \in \mathbb{R}^{\mathbb{N}}$ iff for all $i \in \mathbb{N}$ the coordinates $z_n(i)$ in $\mathbb{R}$ converge to $z(i)$ as $n \to \infty$. Further, it is well-known that the topology on $\mathbb{R}^{\mathbb{N}}$ is metrizable through the metric ρ with

$$\forall\, x, y \in \mathbb{R}^{\mathbb{N}} : \quad \rho(x, y) := \sum_{k\in\mathbb{N}} \frac{|x(k) - y(k)|}{2^k \cdot (1 + |x(k) - y(k)|)}.$$

This follows (for example) with 3.5.7 in [56, p. 121] in combination with Theorem 4.2.2 in [22, p. 259]. Further, $(\mathbb{R}^{\mathbb{N}}, \rho)$ is a *separable* metric space (Theorem 16.4 in [66, p. 109]).

We now define the following map (see [48, p. 43]):

$$\begin{aligned} T : \mathcal{M}_1(\mathbb{R}) &\longrightarrow \mathbb{R}^{\mathbb{N}} \\ \mu &\longmapsto (\langle \mu, g_1 \rangle, \langle \mu, g_2 \rangle, \ldots) \end{aligned}$$

Then surely, T is injective, since if $T(\mu) = T(\nu)$, then also for all $k \in \mathbb{N} : \langle \mu, g_k \rangle = \langle \nu, g_k \rangle$ and then $\mu = \nu$. Additionally, we have for all $\mu, \nu \in \mathcal{M}_1(\mathbb{R})$ that

$$d_M(\mu, \nu) = \sum_{k \in \mathbb{N}} \frac{|\langle \mu, g_k \rangle - \langle \nu, g_k \rangle|}{2^k \cdot (1 + |\langle \mu, g_k \rangle - \langle \nu, g_k \rangle|)} = \rho(T(\mu), T(\nu)). \tag{2.3}$$

Since T injective and ρ is a metric, d_M is a metric as well, so that $(\mathcal{M}_1(\mathbb{R}), d_M)$ is a metric space. With equation (2.3) we see that $T : (\mathcal{M}_1(X, d), d_M) \longrightarrow \mathbb{R}^{\mathbb{N}}$ is not only injective, but even isometric, especially continuous and a homeomorphism onto its image. Surely, the image is separable as a subspace of a separable metric space . Thus, $(\mathcal{M}_1(\mathbb{R}), d_M)$, being homeomorphic to a separable space, is also separable (Corollary 1.4.11 in [22, p. 31]).

With what we have shown so far we obtain for arbitrary $(\mu_n)_{n \in \mathbb{N}}, \mu \in \mathcal{M}_1(\mathbb{R})$:

$$\begin{aligned} &\mu_n \text{ converges weakly to } \mu \\ \Leftrightarrow\ &\forall\, k \in \mathbb{N} : \langle \mu_n, g_k \rangle \xrightarrow[n \to \infty]{} \langle \mu, g_k \rangle \\ \Leftrightarrow\ &T(\mu_n) \xrightarrow[n \to \infty]{} T(\mu) \text{ in } \mathbb{R}^{\mathbb{N}} \\ \Leftrightarrow\ &\rho(T(\mu_n), T(\mu)) \xrightarrow[n \to \infty]{} 0 \\ \Leftrightarrow\ &d_M(\mu_n, \mu) \xrightarrow[n \to \infty]{} 0. \end{aligned}$$

We showed the first equivalence in the first part of this proof, the second equivalence holds per definition of T and the above mentioned characterization of convergence in $\mathbb{R}^{\mathbb{N}}$, the third equivalence follows with the metrizability of $\mathbb{R}^{\mathbb{N}}$ through ρ, and the last equivalence follows from above equation (2.3). What is left to show is that $(\mathcal{M}_1(\mathbb{R}), d_M)$ is not complete. To this end, let $(\mu_n)_n$ be any sequence in $\mathcal{M}_1(\mathbb{R})$ which converges vaguely to a sub-probability measure ν with $\nu(\mathbb{R}) < 1$. Then for all $k \in \mathbb{N}$, $\langle \mu_n, g_k \rangle \to \langle \nu, g_k \rangle$ as $n \to \infty$. Thus, $d_M(\mu_n, \nu) \to 0$ as $n \to \infty$ (the function d_M makes sense even with sub-probability measures as arguments). Since for any $n, m \in \mathbb{N}$, $d_M(\mu_n, \mu_m) \leq d_M(\mu_n, \nu) + d_M(\mu_m, \nu)$, we find that $(\mu_n)_n$ is a Cauchy sequence in $(\mathcal{M}_1(\mathbb{R}), d_M)$ that does not converge weakly to an element in $\mathcal{M}_1(\mathbb{R})$. $\square$

2.3 Random Probability Measures on $(\mathbb{R}, \mathcal{B})$

As we saw in Theorem 2.16, the set $\mathcal{M}_1(\mathbb{R})$ can be metrized in such a way that the resulting convergence is exactly "weak convergence of probability measures." This shows that Definition 2.7 was adequate in the sense that it defined weak convergence for sequences of probability measures rather than for nets. The reason is that in metric spaces (or more generally, in spaces which satisfy the first axiom of countability, which means that any point has a countable neighborhood basis), the topology can be reconstructed from the knowledge of convergent sequences rather than nets. This is due to the fact that a set in such a space is closed iff any limit of a convergent *sequence* in the set is an element of the set.

From now on, we will always view $\mathcal{M}_1(\mathbb{R})$ as equipped with the topology of weak convergence and the associated Borel σ-algebra. We know that $\mathcal{M}_1(\mathbb{R})$ is separable and that d_M as in Theorem 2.16 is a metric yielding the topology of weak convergence. It is then a triviality that for any $f \in \mathcal{C}_b(\mathbb{R})$, the function

$$\begin{aligned} I_f : \mathcal{M}_1(\mathbb{R}) &\longrightarrow \mathbb{R} \\ \mu &\longmapsto I_f(\mu) := \langle \mu, f \rangle \end{aligned}$$

is continuous on $\mathcal{M}_1(\mathbb{R})$.

Since $\mathcal{M}_1(\mathbb{R})$ is a measurable space, we can study $\mathcal{M}_1(\mathbb{R})$-valued random variables, which is the subject of this section. We proceed as in our previous work [31], but streamline our argumentation and supplement our exposition with new aspects, for example, a more rigorous analysis of integrability with respect to expected measures, see Theorem 2.20.

Definition 2.17. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space.

i) A *random probability measure* on $(\mathbb{R}, \mathcal{B})$ is a measurable map $\mu : \Omega \to \mathcal{M}_1(\mathbb{R})$, $\omega \mapsto \mu(\omega, \cdot)$.

ii) A *stochastic kernel* from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B})$ is a map $\mu : \Omega \times \mathcal{B} \longrightarrow \mathbb{R}$, so that the following holds:

   a) For all $\omega \in \Omega$, $\mu(\omega, \cdot)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$.

   b) For all $B \in \mathcal{B}$, $\mu(\cdot, B)$ is $\mathcal{A}$-$\mathcal{B}$-measurable.

Lemma 2.18. *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space.*

i) A map $\mu : \Omega \times \mathcal{B} \longrightarrow \mathbb{R}$ *is a random probability measure iff it is a stochastic kernel.*

ii) If μ *is a stochastic kernel from* $(\Omega, \mathcal{A})$ *to* $(\mathbb{R}, \mathcal{B})$ *and* $f : \mathbb{R} \to \mathbb{R}$ *is measurable and bounded, then* $\omega \mapsto \langle \mu(\omega), f \rangle$ *is measurable and bounded by* $\|f\|_\infty$.

Proof. We first show *ii)*: Surely, the indicated map is bounded by $\|f\|_\infty$, since we have for all $\omega \in \Omega$:

$$|\langle \mu(\omega), f \rangle| \leq \langle \mu(\omega), |f| \rangle \leq \langle \mu(\omega), \|f\|_\infty \rangle \leq \|f\|_\infty.$$

To show measurability, we employ the monotone class argument: To start with, we know that the map $\omega \mapsto \mu(\omega, B)$ is measurable for all $B \in \mathcal{B}$. Let f be a simple function on $(\mathbb{R}, \mathcal{B})$, that is, $f = \sum_{i=1}^n \alpha_i \cdot \mathbb{1}_{B_i}$ for some $n \in \mathbb{N}$, $\alpha_i \in [0, \infty)$ and $B_i \in \mathcal{B}$, $i = 1, \ldots, n$, then also $\omega \mapsto \langle \mu(\omega), f \rangle = \sum_{i=1}^n \alpha_i \cdot \mu(\omega, B_i)$ is measurable as a linear combination of finitely many measurable functions. Now let $f \geq 0$ be measurable and bounded, then there exists sequence of simple functions $(f_n)_{n \in \mathbb{N}}$ such that $f_n \nearrow_n f$ pointwise. For

$\omega \in \Omega$ arbitrary it follows per monotone convergence that $\langle \mu(\omega), f_n \rangle \nearrow_n \langle \mu(\omega), f \rangle$, so also $\omega \mapsto \langle \mu(\omega), f \rangle$ is measurable as a pointwise limit of measurable functions. Now if $f : \mathbb{R} \longrightarrow \mathbb{R}$ is measurable and bounded, then also the positive and negative parts f_+ and f_- (then $f_+, f_- \geq 0$ with $f = f_+ - f_-$). Then $\omega \mapsto \langle \mu(\omega), f \rangle = \langle \mu(\omega), f_+ \rangle - \langle \mu(\omega), f_- \rangle$ is measurable as a difference of measurable functions.

We now show *i)*:

"$\Leftarrow$" We have just shown that for all $f \in \mathcal{C}_b(\mathbb{R})$ the map $\omega \mapsto \langle \mu(\omega), f \rangle$ is measurable. Then we obtain for all $\nu \in \mathcal{M}_1(\mathbb{R})$ that the map $\omega \mapsto d_M(\mu(\omega), \nu)$ is measurable as a limit of measurable functions, since

$$d_M(\mu(\omega), \nu) = \sum_{k \in \mathbb{N}} \frac{|\langle \mu(\omega), g_k \rangle - \langle \nu, g_k \rangle|}{2^k \cdot (1 + |\langle \mu(\omega), g_k \rangle - \langle \nu, g_k \rangle|)}.$$

To show the measurability of $\omega \mapsto \mu(\omega, \cdot)$, it suffices to show that preimages of open balls from $(\mathcal{M}_1(\mathbb{R}), d_M)$ are measurable, since the σ-algebra on $\mathcal{M}_1(\mathbb{R})$ is generated by the topology which is generated by the metric d_M, and the space $\mathcal{M}_1(\mathbb{R})$ is separable with respect to the topology of weak convergence, see [10, p. 73]. So let $\nu \in \mathcal{M}_1(\mathbb{R})$ and $\epsilon > 0$ be arbitrary, then it holds with $B_\epsilon^{\mathcal{M}_1(\mathbb{R})}(\nu) := \{\nu' \in \mathcal{M}_1(\mathbb{R}) : d_M(\nu', \nu) < \epsilon\}$:

$$\mu^{-1}\left(B_\epsilon^{\mathcal{M}_1(\mathbb{R})}(\nu)\right) = \{\omega \in \Omega : d_M(\mu(\omega), \nu) < \epsilon\} = d_M(\mu(\cdot), \nu)^{-1}([0, \epsilon)) \in \mathcal{A},$$

since above we already recognized $d_M(\mu(\cdot), \nu)$ as measurable.

"$\Rightarrow$" If μ is a random probability measure, then for all $\omega \in \Omega$, $\mu(\omega, \cdot)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$. We now argue that for any $B \in \mathcal{B}$, $\omega \mapsto \mu(\omega, B)$ is measurable. We first prove this for all open bounded intervals in $\mathbb{R}$, since these intervals generate $\mathcal{B}$. So let $a < b \in \mathbb{R}$ be arbitrary and define $\epsilon := (b - a)/4$. Then define for all $n \in \mathbb{N}$ the function $\phi_n : \mathbb{R} \to \mathbb{R}$ so that $\phi_n \equiv 1$ on $[a + \frac{1}{n}\epsilon, b - \frac{1}{n}\epsilon]$, $\phi_n \equiv 0$ on $(a, b)^c$ and ϕ_n is affine on the intervals $[a, a + \frac{1}{n}\epsilon]$ and $[b - \frac{1}{n}a, b]$ in such a way that it is continuous. Then ϕ_n is bounded, continuous and $\phi_n(x) \nearrow_n \mathbb{1}_{(a,b)}(x)$ for all $x \in \mathbb{R}$. We know that for all $n \in \mathbb{N}$, $\omega \mapsto \langle \mu(\omega), \phi_n \rangle$ is measurable as a composition of a measurable and a continuous map (see remark before Definition 2.17). Now for any $\omega \in \Omega$:

$$\lim_{n \to \infty} \langle \mu(\omega), \phi_n \rangle = \left\langle \mu(\omega), \mathbb{1}_{(a,b)} \right\rangle = \mu(\omega, (a, b)).$$

by monotone convergence. As a result, $\mu(\cdot, (a, b))$ is $\mathcal{A}$-$\mathcal{B}$-measurable as the pointwise limit of measurable functions. Now define the set

$$\mathcal{G} := \{B \in \mathcal{B} \,|\, \omega \mapsto \mu(\omega, B) \text{ is measurable}\}.$$

Surely, all open intervals lie in $\mathcal{G}$ as we have just shown. If we can show that $\mathcal{G}$ is a Dynkin system we can conclude that $\mathcal{G} = \mathcal{B}$, which is our goal. First of all, $\emptyset$, $\mathbb{R} \in \mathcal{G}$, since constant functions are always measurable. Second, since $\mu(\cdot, B^c) = 1 - \mu(\cdot, B)$, we have that $B^c \in \mathcal{G}$ whenever $B \in \mathcal{G}$. Third, if $(B_n)_n$ is a sequence of pairwise disjoint sets in $\mathcal{G}$, then $\mu(\cdot, \cup_n B_n) = \sum_n \mu(\cdot, B_n)$, so since all $\mu(\cdot, B_n)$ are measurable, then so is $\mu(\cdot, \cup_n B_n)$ as a pointwise limit of a sequence of measurable functions. This shows that $\cup_n B_n \in \mathcal{G}$ so that $\mathcal{G}$ is indeed a Dynkin system. $\square$

Random probability measures are not so uncommon in probability theory. Consider the next example:

Example 2.19. Let $Y_1, \dots, Y_n$ be real-valued random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then

$$\rho := \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i}$$

is a random probability measure on $(\mathbb{R}, \mathcal{B})$, which we call *empirical distribution* (of the Y_i). Indeed, for any $\omega \in \Omega$,

$$\rho(\omega) = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i(\omega)}$$

is a convex combination of probability measures and thus again a probability measure on $(\mathbb{R}, \mathcal{B})$. On the other hand, if $B \in \mathcal{B}$ is arbitrary, then

$$\omega \mapsto \rho(\omega, B) = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i(\omega)}(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_B(Y_i(\omega))$$

is certainly measurable. Thus, we recognize the empirical distribution ρ as a random probability measure on $(\mathbb{R}, \mathcal{B})$ via Lemma 2.18. For any measurable set B, $\rho(B)$ yields the proportion of the Y_i's that fall into the set B. Connected to the empirical distribution ρ is its empirical distribution function $F_\rho(x) := \rho((-\infty, x])$ defined for all $x \in \mathbb{R}$. This is a random distribution function and the protagonist of the famous Glivenko-Cantelli theorem and the Dvoretzky–Kiefer–Wolfowitz inequality, see [68, p. 553].

Now, let us resume our study. If μ is a random probability measure and $B \in \mathcal{B}$, then $\mu(B)$ is a bounded random variable. It is natural to consider its expectation $\mathbb{E}\mu(B)$ as the expected mass that μ prescribes to the set B. But as it turns out, $B \mapsto \mathbb{E}\mu(B)$ is yet another (deterministic) probability measure:

Theorem 2.20. *Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a probability space and μ be a random probability measure on $(\mathbb{R}, \mathcal{B})$. Then the following statements hold:*

i) The map

$$\begin{aligned} \bar{\mu} : \mathcal{B} &\longrightarrow [0,1] \\ B &\longmapsto \bar{\mu}(B) := \int_\Omega \mu(\omega, B)\, \mathbb{P}(d\omega) = \mathbb{E}\mu(B) \end{aligned}$$

is an element of $\mathcal{M}_1(\mathbb{R})$, the so called expected measure *of μ.*

ii) Any non-negative measurable function $f : \mathbb{R} \longrightarrow \mathbb{R}_+$ is $\bar{\mu}$-integrable iff $\langle \mu, f \rangle$ is $\mathbb{P}$-integrable, and in this case it holds

$$\langle \bar{\mu}, f \rangle = \int_\mathbb{R} f(x)\, \bar{\mu}(dx) = \int_\Omega \int_\mathbb{R} f(x)\, \mu(\omega, dx)\, \mathbb{P}(d\omega) = \mathbb{E}\, \langle \mu, f \rangle \,.$$

In particular, this equation is valid for any bounded measurable function $f : \mathbb{R} \to \mathbb{R}$.

iii) If $f : \mathbb{R} \to \mathbb{R}$ is $\bar{\mu}$-integrable, then $\langle \mu, f \rangle$ is $\mathbb{P}$-integrable and $\langle \bar{\mu}, f \rangle = \mathbb{E} \langle \mu, f \rangle$.

iv) Heed must be taken: If $f : \mathbb{R} \to \mathbb{R}$ is measurable and such that $\langle \mu, f \rangle$ is $\mathbb{P}$-integrable so that $\mathbb{E} \langle \mu, f \rangle$ is well-defined, f need not be $\bar{\mu}$-integrable, so that it is not true *that $\langle \bar{\mu}, f \rangle = \mathbb{E} \langle \mu, f \rangle$ whenever one of the two exists. In particular, statement ii) cannot be generalized to arbitrary measurable functions $f : \mathbb{R} \to \mathbb{R}$.*

Due to these interrelations we will also write $\mathbb{E}\mu$ instead of $\bar{\mu}$, and with what we have seen so far it holds for all $\mathbb{E}\mu$-integrable, bounded or non-negative functions f, that

$$\langle \mathbb{E}\mu, f \rangle = \langle \bar{\mu}, f \rangle = \mathbb{E} \langle \mu, f \rangle .$$

Proof. i) Clearly, $\mathbb{E}\mu(\emptyset) = 0$ and $\mathbb{E}\mu(\mathbb{R}) = 1$. Now if $(B_n)_n$ is a sequence of pairwise disjoint elements in $\mathcal{B}$, then

$$\mathbb{E}\mu(\cup_n B_n) = \mathbb{E} \sum_n \mu(B_n) = \sum_n \mathbb{E}\mu(B_n),$$

where in the last step we used dominated convergence. This shows that $\bar{\mu}$ is indeed a probability measure.
ii) Let $f : \mathbb{R} \to \mathbb{R}$ be a simple function, that is, $f = \sum_{i=1}^n \alpha_i \cdot \mathbb{1}_{B_i}$ for some $n \in \mathbb{N}$, $\alpha_i \in [0, \infty)$ and $B_i \in \mathcal{B}$, $i = 1, \dots, n$. Then

$$\langle \bar{\mu}, f \rangle = \sum_{i=1}^n \alpha_i \cdot \bar{\mu}(B_i) = \mathbb{E} \sum_{i=1}^n \alpha_i \cdot \mu(B_i) = \mathbb{E} \langle \mu, f \rangle .$$

Now let $f : \mathbb{R} \to \mathbb{R}$ be non-negative and measurable witnessed by a sequence of simple functions $(f_n)_n$ with $f_n \nearrow_n f$ pointwise, then clearly

$$\langle \bar{\mu}, f \rangle = \lim_{n\to\infty} \langle \bar{\mu}, f_n \rangle = \lim_{n\to\infty} \mathbb{E} \langle \mu, f_n \rangle = \mathbb{E} \langle \mu, f \rangle ,$$

where in the first and the last step we used monotone convergence. In particular, the non-negative f is $\bar{\mu}$-integrable iff $\langle \mu, f \rangle$ is $\mathbb{P}$-integrable and in this case it holds $\langle \bar{\mu}, f \rangle = \mathbb{E} \langle \mu, f \rangle$. Now if $f : \mathbb{R} \to \mathbb{R}$ is bounded, then there exists a $C \in \mathbb{R}$ such that $f + C$ is non-negative (and of course, it remains bounded, thus integrable). Then we immediately obtain $\langle \bar{\mu}, f \rangle = \langle \bar{\mu}, f + C \rangle - C = \mathbb{E} \langle \mu, f + C \rangle - C = \mathbb{E} \langle \mu, f \rangle$.
iii) If now $f : \mathbb{R} \to \mathbb{R}$ is $\bar{\mu}$-integrable, then $f = f_+ - f_-$ where $f_+, f_- \geq 0$ are $\bar{\mu}$-integrable. By *ii*), the non-negative random variables $\langle \mu, f_+ \rangle$ and $\langle \mu, f_- \rangle$ are both $\mathbb{P}$-integrable. Then their difference $\langle \mu, f_+ \rangle - \langle \mu, f_- \rangle = \langle \mu, f \rangle$ is also $\mathbb{P}$-integrable and we obtain with *ii*):

$$\langle \bar{\mu}, f \rangle = \langle \bar{\mu}, f_+ \rangle - \langle \bar{\mu}, f_- \rangle = \mathbb{E} \langle \bar{\mu}, f_+ \rangle - \mathbb{E} \langle \bar{\mu}, f_- \rangle = \mathbb{E} \langle \bar{\mu}, f \rangle .$$

iv) Unfortunately, this point appears to be overlooked in the literature. We need to construct a counter-example to show what we state. To this end, consider the random probability measure μ on $(\mathbb{R}, \mathcal{B})$ with

$$\forall\, n \in \mathbb{N} : \mathbb{P}(\mu = \frac{1}{2}\delta_{-n} + \frac{1}{2}\delta_n) = \frac{1}{cn^2},$$

where $c := \sum_n \frac{1}{n^2} < \infty$. Further, let f be the identity on $\mathbb{R}$, that is, $f(x) = x$ for all $x \in \mathbb{R}$. Then surely, f is measurable, and since almost all realizations of μ are symmetric measures, we have $\langle \mu, f \rangle = 0$ almost surely, which is $\mathbb{P}$-integrable with $\mathbb{E} \langle \mu, f \rangle = 0$. We now assume that f is $\bar{\mu}$-integrable and lead this to a contradiction: If f were $\bar{\mu}$-integrable, then so were $|f|$ and by *ii)* we would have $\langle \bar{\mu}, |f| \rangle = \mathbb{E} \langle \mu, |f| \rangle < \infty$. But with probability $\frac{1}{cn^2}$, μ takes the value $\frac{1}{2}\delta_{-n} + \frac{1}{2}\delta_n$, so $\langle \mu, |f| \rangle$ takes the value n, leading to the calculation

$$\mathbb{E} \langle \mu, |f| \rangle = \sum_{n \in \mathbb{N}} \frac{n}{cn^2} = \infty,$$

which is a contradiction. □

In the remainder of this section, we will derive and discuss three notions of convergence of random probability measures on $(\mathbb{R}, \mathcal{B})$, namely weak convergence in expectation, weak convergence in probability and weak convergence almost surely.

Definition 2.21. Let $(\mu_n)_{n \in \mathbb{N}}$ and μ be random probability measures on $(\mathbb{R}, \mathcal{B})$, then we say that $(\mu_n)_n$ converges weakly in expectation to μ, if the sequence of expected measures $(\mathbb{E}\mu_n)_{n \in \mathbb{N}}$ converges weakly to the expected measure $\mathbb{E}\mu$, so if:

$$\forall f \in \mathcal{C}_b(\mathbb{R}) : \langle \mathbb{E}\mu_n, f \rangle \xrightarrow[n \to \infty]{} \langle \mathbb{E}\mu, f \rangle ,$$

which is equivalent to (see Theorem 2.20)

$$\forall f \in \mathcal{C}_b(\mathbb{R}) : \mathbb{E} \langle \mu_n, f \rangle \xrightarrow[n \to \infty]{} \mathbb{E} \langle \mu, f \rangle .$$

The concept of weak convergence in expectation is extremely important for investigations in the field of random matrix theory, since it lies the foundation for stronger convergence types. This is due to the fact that weak convergence $\mathbb{P}$-almost surely or in probability will also imply weak convergence in expectation, so the latter convergence type is a necessary condition for stronger convergence types (see also Theorem 3.7). The exact interrelations between the three concepts of convergence for random probability measures are summarized in the end of this section in Theorem 2.29.

Before turning to the next convergence types, we wish to remind the reader what convergence in probability and almost surely means for random variables in metric spaces:

Definition 2.22. Let $(Y_n)_{n \in \mathbb{N}}$ and Y be random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, which take values in a metric space $(\mathcal{X}, d)$.

i) We say that $(Y_n)_{n \in \mathbb{N}}$ converges to Y in probability, if $d(Y_n, Y)$ converges to 0 in probability.

ii) We say that $(Y_n)_{n \in \mathbb{N}}$ converges to Y almost surely, if $d(Y_n, Y)$ converges to 0 almost surely.

Let us collect a quick lemma:

Lemma 2.23. *Let $(Y_n)_{n\in\mathbb{N}}$ and Y be random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, which take values in a metric space $(\mathcal{X}, d)$. If $(Y_n)_{n\in\mathbb{N}}$ converges to Y almost surely, then also in probability.*

Proof. Let $(Y_n)_{n\in\mathbb{N}}$ converge to Y almost surely. This means that the sequence of real-valued random variables $(d(Y_n, Y))_n$ converges to 0 almost surely. But this implies that $(d(Y_n, Y))_n$ converges to 0 in probability, which is precisely what it means for $(Y_n)_n$ to converge to Y in probability. □

Now let us define and analyze what it means for random probability measures to converge in probability and almost surely. Since random probability measures are nothing but random variables into the metric space $\mathcal{M}_1(\mathbb{R})$, we know what to do:

Definition 2.24. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, μ and $(\mu_n)_{n\in\mathbb{N}}$ be random probability measures on $(\mathbb{R}, \mathcal{B})$.

i) We say that $(\mu_n)_n$ converges weakly to μ in probability, if $d_M(\mu_n, \mu)$ converges to 0 in probability.

ii) We say that $(\mu_n)_n$ converges weakly to μ almost surely, if $d_M(\mu_n, \mu)$ converges to 0 almost surely.

Although stochastic types of weak convergence can be defined solidly as in Definition 2.24, this definition is not convenient to work with in practice. In addition, we would like to see that these convergence concepts do *not* depend on the choice of the metric that metrizes weak convergence on $\mathcal{M}_1(\mathbb{R})$.

Theorem 2.25. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, μ and $(\mu_n)_{n\in\mathbb{N}}$ be random probability measures on $(\mathbb{R}, \mathcal{B})$.*

i) The following statements are equivalent:

a) $(\mu_n)_n$ converges weakly to μ in probability, that is, $d_M(\mu_n, \mu) \to 0$ in probability.

b) If d is any metric on $\mathcal{M}_1(\mathbb{R})$ that metrizes weak convergence, then $d(\mu_n, \mu) \to 0$ in probability.

c) For all $f \in \mathcal{C}_b(\mathbb{R})$, the sequence of bounded real-valued random variables $(\langle \mu_n, f\rangle)_n$ converges in probability to $\langle \mu, f\rangle$, so

$$\forall\, f \in \mathcal{C}_b(\mathbb{R}) : \forall \epsilon > 0 : \mathbb{P}(|\langle \mu_n, f\rangle - \langle \mu, f\rangle| > \epsilon) \xrightarrow[n\to\infty]{} 0.$$

ii) The following statements are equivalent:

a) $(\mu_n)_n$ converges weakly to μ almost surely, that is, $d_M(\mu_n, \mu) \to 0$ almost surely.

b) For $\mathbb{P}$-almost all $\omega \in \Omega$, $\mu_n(\omega)$ converges weakly to $\mu(\omega)$.

c) If d is any metric on $\mathcal{M}_1(\mathbb{R})$ that metrizes weak convergence, then $d(\mu_n, \mu) \to 0$ almost surely.

d) For all $f \in \mathcal{C}_b(\mathbb{R})$, $\langle \mu_n, f \rangle$ converges almost surely to $\langle \mu, f \rangle$, that is,

$$\forall\, f \in \mathcal{C}_b(\mathbb{R}) : \left[\langle \mu_n, f \rangle \xrightarrow[n \to \infty]{} \langle \mu, f \rangle \;\; \textit{almost surely} \right].$$

e) Almost surely we find that for all $f \in \mathcal{C}_b(\mathbb{R})$, $\langle \mu_n, f \rangle$ converges to $\langle \mu, f \rangle$, that is,

$$\left[\forall\, f \in \mathcal{C}_b(\mathbb{R}) : \langle \mu_n, f \rangle \xrightarrow[n \to \infty]{} \langle \mu, f \rangle \right] \quad \textit{almost surely.}$$

Remark 2.26. 1. Note that in Theorem 2.25 *ii*) *d*) and *e*) we used careful bracketing [. . .] when it comes to almost sure convergence of multiple objects. This is done to avoid ambiguity. For example, questions could arise whether we find a set of measure 1 on which all objects converge (as in *e*)), or if for each object, we find a set of measure 1, *possibly depending on that object*, on which the considered object converges (as in *d*)).

2. We consider Theorem 2.25 *i*) as equivalent definitions for the concept "weak convergence in probability", and *ii*) as equivalent definitions for "weak convergence almost surely." After the proof of the theorem, we will keep on working with this characterization without always referring to Theorem 2.25.

Before we begin with the proof of Theorem 2.25, we will introduce two tools which we will make use of. For later use, we will formulate the lemmas in greater generality, that is, for complex-valued random variables.

Lemma 2.27. *Let $(X_n)_n$ and X be complex-valued random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then $(X_n)_{n\in\mathbb{N}}$ converges to X in probability iff any subsequence $J \subseteq \mathbb{N}$ has another subsequence $I \subseteq J$ so that $(X_n)_{n\in I}$ converges to X almost surely.*

Proof. The proof can be found in [44, p. 134] . □

The next extremely useful lemma generalizes the previous one by finding a *simultaneous almost surely convergent subsequence* for a countable number of sequences of random variables.

Lemma 2.28. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and for all $k \in \mathbb{N}$ let $X^{(k)}$ and $(X_n^{(k)})_{n\in\mathbb{N}}$ be complex-valued random variables. Then the following statements are equivalent:*

i) For all $k \in \mathbb{N}$, $(X_n^{(k)})_n$ converges to $X^{(k)}$ in probability.

ii) For any subsequence $J \subseteq \mathbb{N}$, we find a subsequence $I \subseteq J$ and a set $N \in \mathcal{A}$ with $\mathbb{P}(N) = 0$ such that

$$\forall\, \omega \in \Omega \backslash N : \forall k \in \mathbb{N} : X_n^{(k)}(\omega) \xrightarrow[n \in I]{} X^{(k)}(\omega).$$

Proof. The part $ii) \Rightarrow i)$ follows immediately with Lemma 2.27. So we only need to show $i) \Rightarrow ii)$: For $k = 1$ we find that $(X_n^{(1)})_{n\in J}$ converges in probability to $X^{(1)}$. Therefore, we find a subsequence $I_1 \subseteq J$ such that

$$X_n^{(1)} \xrightarrow[n\in I_1]{} X^{(1)} \quad \mathbb{P}\text{-a.s. witnessed by a set of measure zero } N_1.$$

Since $(X_n^{(2)})_{n\in I_1}$ converges to $X^{(2)}$ in probability, we find a subsequence $I_2 \subseteq I_1$ with $\min(I_2) > \min(I_1)$ such that

$$X_n^{(2)} \xrightarrow[n\in I_2]{} X^{(2)} \quad \mathbb{P}\text{-a.s. witnessed by a set of measure zero } N_2.$$

We continue this approach for all $k \in \mathbb{N}$ and obtain subsequences

$$\mathbb{N} \supseteq J \supseteq I_1 \supseteq I_2 \supseteq \ldots \supseteq I_k \supseteq \ldots$$

such that for all $k \in \mathbb{N}$ we have $\min(I_{k+1}) > \min(I_k)$ and

$$X_n^{(k)} \xrightarrow[n\in I_k]{} X^{(k)} \quad \mathbb{P}\text{-a.s. witnessed by a set of measure zero } N_k.$$

We set $N \coloneqq \cup_{k\in\mathbb{N}} N_k$ and for all $k \in \mathbb{N} : i_k \coloneqq \min(I_k)$, then we obtain that $(i_k)_{k\in\mathbb{N}}$ is strictly increasing in $\mathbb{N}$ and

$$\forall\, \omega \in \Omega \backslash N : \forall\, l \in \mathbb{N} : X_{i_k}^{(l)}(\omega) \xrightarrow[k\in\mathbb{N}]{} X^{(l)}(\omega).$$

To see this, let $\omega \in \Omega\backslash N$ and $l \in \mathbb{N}$ be arbitrary. Then we have that $\omega \in \Omega\backslash N_l$ and $i_k = \min(I_k) \in I_l$ for all $k \geq l$, so that indeed

$$X_{i_k}^{(l)}(\omega) \xrightarrow[k\in\mathbb{N}]{} X^{(l)}(\omega).$$

The proof is completed by setting $I \coloneqq \{i_k \,|\, k \in \mathbb{N}\}$. □

Now we are ready to prove Theorem 2.25:

Proof of Theorem 2.25. We show $ii)$ first.
Clearly, $a)$, $b)$ and $c)$ are equivalent, since the metrics metrize weak convergence. Also, $e)$ is just a reformulation of $b)$, thus equivalent. In addition, $d)$ follows immediately from $e)$, so we have

$$a) \;\Leftrightarrow\; b) \;\Leftrightarrow\; c) \;\Leftrightarrow\; e) \;\Rightarrow\; d)$$

We now show $\underline{d) \Rightarrow b)}$: For each $k \in \mathbb{N}$ we have that $\langle \mu_n, g_k\rangle$ converges to $\langle \mu, g_k\rangle$ almost surely on a set A_k of measure 1 (the functions $(g_k)_k$ are as in Theorem 2.16). Then the set $\Omega_1 \coloneqq \cap_k A_k$ has measure 1 and for all $\omega \in \Omega_1$ we find that

$$\forall\, k \in \mathbb{N} : \langle \mu_n(\omega), g_k\rangle \xrightarrow[n\to\infty]{} \langle \mu(\omega), g_k\rangle\,.$$

Therefore, with Theorem 2.16, we have for all $\omega \in \Omega_1$ that $\mu_n(\omega) \to \mu(\omega)$ weakly as $n \to \infty$ and hence b).
We now show i):
$a) \Leftrightarrow b)$ By exact symmetry in the argument, we will only argue $a) \Rightarrow b)$: Let $\mu_n \to \mu$ weakly in probability, that is, $(d_M(\mu_n, \mu))_{n\in\mathbb{N}}$ converges to 0 in probability. We want to show that also $(d(\mu_n, \mu))_{n\in\mathbb{N}}$ converges to 0 in probability. To use Lemma 2.27, let $J \subseteq \mathbb{N}$ be an arbitrary subsequence. Then we find a subsequence $I \subseteq J$ such that $(d_M(\mu_n, \mu))_{n\in I}$ converges to 0 almost surely. With part ii) this means that also $(d(\mu_n, \mu))_{n\in I}$ converges to 0 almost surely. But then $(d(\mu_n, \mu))_{n\in\mathbb{N}}$ converges to 0 in probability.
$a) \Rightarrow c)$ If $(\mu_n)_n$ converges weakly to μ in probability, then this means that $d_M(\mu_n, \mu)$ converges to 0 in probability. Let $f \in \mathcal{C}_b(\mathbb{R})$ be arbitrary. We must show that $\langle \mu_n, f\rangle$ converges to $\langle \mu, f\rangle$ in probability. To this end, let $J \subseteq \mathbb{N}$ be an arbitrary subsequence. Then there is a subsequence $I \subseteq J$ such that $(d_M(\mu_n, \mu))_{n\in I}$ converges to 0 almost surely on a measurable subset $\Omega_1 \subseteq \Omega$ with measure 1. Then it holds in particular for any $\omega \in \Omega_1$ that $(\langle \mu_n(\omega), f\rangle)_{n\in I}$ converges to $\langle \mu(\omega), f\rangle$, so $(\langle \mu_n, f\rangle)_{n\in I}$ converges to $\langle \mu, f\rangle$ almost surely. The statement follows with Lemma 2.27.
$c) \Rightarrow a)$ We find that for all $k \in \mathbb{N}$, $(\langle \mu_n, g_k\rangle)_{n\in\mathbb{N}}$ converges to $\langle \mu, g_k\rangle$ in probability. We must show that $d_M(\mu_n, \mu)$ converges to zero in probability. Let $J \subseteq \mathbb{N}$ be any subsequence. With Lemma 2.28, we find a subsequence $I \subseteq J$ and a measurable set $\Omega_1 \subseteq \Omega$ of measure 1, such that

$$\forall\, \omega \in \Omega_1 : \forall\, k \in \mathbb{N} : \langle \mu_n(\omega), g_k\rangle \xrightarrow[n\in I]{} \langle \mu(\omega), g_k\rangle$$

With Theorem 2.16, this entails that for all $\omega \in \Omega_1$, $(d_M(\mu_n(\omega), \mu(\omega)))_{n\in I}$ converges to 0. With Lemma 2.27, this means that $(d_M(\mu_n, \mu))_{n\in\mathbb{N}}$ converges to zero in probability. □

So, what we have seen so far is that random probability measures can converge in three different ways, namely weakly in expectation, weakly in probability and weakly almost surely. We have solidly defined and then characterized these convergence concepts. At last, we point out a hierarchy among them:

Theorem 2.29. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(\mu_n)_{n\in\mathbb{N}}$ and μ be random probability measures on $(\mathbb{R}, \mathcal{B})$.*

i) If $\mu_n \to \mu$ weakly almost surely, then also weakly in probability.

ii) If $\mu_n \to \mu$ weakly in probability, then also weakly in expectation.

Proof. i) This follows directly with Lemma 2.23.
ii) If $\mu_n \to \mu$ weakly in probability, per Theorem 2.25 this means that for all $f \in \mathcal{C}_b(\mathbb{R})$ we find $\langle \mu_n, f\rangle \to \langle \mu, f\rangle$ in probability, thus $\mathbb{E}\,\langle \mu_n, f\rangle \to \mathbb{E}\,\langle \mu, f\rangle$ by the following Lemma 2.30, since $|\langle \mu_n, f\rangle| \leq \|f\|_\infty$ and $|\langle \mu, f\rangle| \leq \|f\|_\infty$. □

Lemma 2.30. *Let $(X_n)_{n\in\mathbb{N}}$ and X be complex-valued random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $C \in \mathbb{R}$ such that $|X_n| \leq C$ for all $n \in \mathbb{N}$ and $|X| \leq C$. Then $X_n \to X$ in probability implies $\mathbb{E}X_n \to \mathbb{E}X$.*

Proof. Let $\epsilon > 0$ be arbitrary, then we calculate:

$$\begin{aligned}|\mathbb{E}X_n - \mathbb{E}X| &\leq \mathbb{E}|X_n - X| \\ &= \mathbb{E}|X_n - X|\mathbb{1}_{\{|X_n - X| \leq \epsilon\}} + \mathbb{E}|X_n - X|\mathbb{1}_{\{|X_n - X| > \epsilon\}} \\ &\leq \epsilon + \mathbb{P}(|X_n - X| > \epsilon) \cdot 2C.\end{aligned}$$

Therefore, we conclude

$$\limsup_{n\to\infty} |\mathbb{E}X_n - \mathbb{E}X| \leq \epsilon.$$

Since $\epsilon > 0$ was arbitrary, this shows the statement. □

2.4 Random Matrices and their ESDs

In this section, which is as generalization from our previous work [31], we will introduce the types of random probability measures which we would like to investigate, namely the empirical spectral distribution of random matrices. To this end, let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ and denote by $(\mathrm{Mat}_n(\mathbb{K}), \|\cdot\|_{\mathrm{op}})$ the normed $\mathbb{K}$-vector space of $n \times n$-matrices with $\mathbb{K}$-valued entries, where $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm with respect to the euclidian norm $\|\cdot\|$ on $\mathbb{K}^n$, that is,

$$\forall\, X \in \mathrm{Mat}_n(\mathbb{K}) : \|X\|_{\mathrm{op}} = \sup\left\{\|Xv\| : v \in \mathbb{K}^n, \|v\| = 1\right\}.$$

It is immediate that $(\mathrm{Mat}_n(\mathbb{K}), \|\cdot\|_{\mathrm{op}})$ is a Banach-space and a sequence of matrices $(X_m)_m$ converges to a matrix X in $\mathrm{Mat}_n(\mathbb{K})$, iff all entries $X_m(i,j)$ converge to $X(i,j)$ in $\mathbb{K}$ as $m \to \infty$. If $X \in \mathrm{Mat}_n(\mathbb{K})$ we denote its *adjoint* by X^*, which is just the transpose of X if $\mathbb{K} = \mathbb{R}$ and the conjugate transpose of X if $\mathbb{K} = \mathbb{C}$. A matrix $X \in \mathrm{Mat}_n(\mathbb{K})$ is called *self-adjoint* if $X^* = X$ (then X is also called *symmetric* if $\mathbb{K} = \mathbb{R}$ and Hermitian if $\mathbb{K} = \mathbb{C}$) and we denote the subset of all self-adjoint matrices of $\mathrm{Mat}_n(\mathbb{K})$ by $\mathrm{SMat}_n(\mathbb{K})$. Then $\mathrm{SMat}_n(\mathbb{K}) \subseteq \mathrm{Mat}_n(\mathbb{K})$ is a closed subset, since $X \mapsto X^*$ is continuous. Further, $\mathrm{SMat}_n(\mathbb{K})$ is closed under $\mathbb{R}$-linear combinations. To introduce more notation, if $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ are arbitrary, we denote by $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ the diagonal matrix $D \in \mathrm{SMat}_n(\mathbb{K})$ with entries $D(i,i) = \lambda_i$ for all $i \in \{1, \ldots, n\}$. Further, we denote by tr the trace functional $\mathrm{Mat}_n(\mathbb{K}) \longrightarrow \mathbb{K}$, that is,

$$\forall\, X \in \mathrm{Mat}_n(\mathbb{K}) : \mathrm{tr}\, X = \sum_{t=1}^{n} X(t,t).$$

The trace has some interesting properties, which are summarized in the following lemma:

Lemma 2.31. *The trace* tr *is a continuous linear functional on* $(\mathit{Mat}_n(\mathbb{K}), \|\cdot\|_{\mathrm{op}})$. *Further, if* $X, S \in \mathit{Mat}_n(\mathbb{K})$ *are arbitrary, where* S *is invertible, then* $\mathrm{tr}(X) = \mathrm{tr}(S^{-1}XS)$.

Proof. It is immediate that the trace is a continuous linear functional. The equality $\mathrm{tr}(X) = \mathrm{tr}(S^{-1}XS)$ is due to the fact that X and $S^{-1}XS$ have the same characteristic polynomial. The trace is the $(n-1)$th coefficient of the characteristic polynomial (multiplied by $(-1)^{n-1}$). For details we refer the reader to [29]. □

The next lemma clarifies the eigenvalue structure of self-adjoint matrices:

Lemma 2.32. *For any matrix $X \in SMat_n(\mathbb{K})$ we find an invertible matrix $S \in Mat_n(\mathbb{K})$ and real numbers $\lambda_1^X \leq \ldots \leq \lambda_n^X$, such that $S^{-1}XS = \operatorname{diag}(\lambda_1^X, \ldots, \lambda_n^X)$. In particular, X has exactly n real eigenvalues (counting multiplicities) and all eigenvalues are real.*

Proof. We refer the reader to [29]. □

In general, if Y is a self-adjoint $n \times n$ matrix, we will denote its n real eigenvalues by $\lambda_1^Y \leq \ldots \leq \lambda_n^Y$. The next theorem is a very versatile tool in random matrix theory. For example, it can be used to derive that eigenvalues are continuous functions of the entries of the matrix (Corollary 2.34), or it can be used to analyze asymptotic equivalence of empirical spectral distributions via the bounded Lipschitz metric, see Section 4.4.

Theorem 2.33 (Hoffman-Wielandt)**.** *For all $n \in \mathbb{N}$ and $X, Y \in SMat_n(\mathbb{K})$ it holds:*

$$\sum_{i=1}^{n} |\lambda_i^X - \lambda_i^Y|^2 \leq \operatorname{tr}(X-Y)^*(X-Y) = \operatorname{tr}(X-Y)^2.$$

Proof. See [36, p. 217]. □

We can immediately conclude that eigenvalues are continuous functions of the matrices.

Corollary 2.34. *Let $n \in \mathbb{N}$ and $l \in \{1, \ldots, n\}$ be arbitrary, then*

$$\begin{aligned} \operatorname{Eig}_l : SMat_n(\mathbb{K}) &\longrightarrow \mathbb{R} \\ X &\longmapsto \lambda_l^X \end{aligned}$$

is continuous.

Proof. Let $(X_m)_{m\in\mathbb{N}}$ and X in $\mathrm{SMat}_n(\mathbb{K})$ so that $X_m \to X$ for $m \to \infty$ (which means convergence in operator norm, or equivalently, entry-wise convergence). Then we find with Theorem 2.33 and Lemma 2.31 that

$$|\lambda_l^{X_m} - \lambda_l^X|^2 \leq \sum_{i=1}^{N} |\lambda_i^{X_m} - \lambda_i^X|^2 \leq \operatorname{tr}(X_m - X)^2 \xrightarrow[m\to\infty]{} 0.$$

□

Having studied eigenvalues of self-adjoint matrices, let us turn our attention to random matrices.

Definition 2.35. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and $n \in \mathbb{N}$ be arbitrary then a ($n \times n$ self-adjoint) random matrix is a measurable map $X : (\Omega, \mathcal{A}) \to (\mathrm{SMat}_n(\mathbb{K}), \mathcal{B}_s^{(n^2)})$, where $\mathcal{B}_s^{(n^2)}$ denotes Borel σ-algebra on $\mathrm{SMat}_n(\mathbb{K})$.

It is clear that a map $X : (\Omega, \mathcal{A}) \to (\mathrm{SMat}_n(\mathbb{K}), \mathcal{B}_s^{(n^2)})$ is measurable iff all entries $X(i,j) : (\Omega, \mathcal{A}) \to (\mathbb{K}, \mathcal{B}_{\mathbb{K}})$ are measurable, where $\mathcal{B}_{\mathbb{K}}$ denotes the Borel σ-algebra on $\mathbb{K}$. If X is an $n \times n$ random matrix on $(\Omega, \mathcal{A}, \mathbb{P})$, then for all $\omega \in \Omega$, $X(\omega) \in \mathrm{SMat}_n(\mathbb{K})$, such that $X(\omega)$ possesses eigenvalues $\lambda_1^{X(\omega)} \leq \ldots \leq \lambda_n^{X(\omega)}$. We wish to see that the maps $\omega \mapsto \lambda_l^{X(\omega)}$ for $l = 1, \ldots, n$ are measurable.

Lemma 2.36. *Let X be an $n \times n$-random matrix on $(\Omega, \mathcal{A}, \mathbb{P})$ and $l \in \{1, \ldots, n\}$ be arbitrary, then*

$$\begin{aligned} \lambda_l^X : (\Omega, \mathcal{A}) &\longrightarrow (\mathbb{R}, \mathcal{B}) \\ \omega &\longmapsto \lambda_l^{X(\omega)} \end{aligned}$$

is measurable, thus a real-valued random variable.

Proof. We know by Corollary 2.34 that

$$\begin{aligned} \mathrm{Eig}_l : \mathrm{SMat}_n(\mathbb{K}) &\longrightarrow \mathbb{R} \\ X &\longmapsto \lambda_l^X \end{aligned}$$

is continuous, in particular measurable. Further, $X : \Omega \longrightarrow \mathrm{SMat}_n(\mathbb{K})$ is measurable per definition, hence the composition $\lambda_l^X := \mathrm{Eig}_l \circ X$ is measurable as well. $\square$

Lemma 2.36 allows us to study eigenvalues of random matrices in the context of probability theory. One aspect which gains a lot of attention is the behavior of the empirical distribution of the eigenvalues (see also Example 2.19).

Definition 2.37. Let X be an $n \times n$ random matrix on $(\Omega, \mathcal{A}, \mathbb{P})$, then the *empirical spectral distribution (ESD)* σ_n of X is the random probability measure on $(\mathbb{R}, \mathcal{B})$ given by

$$\begin{aligned} \sigma_n : \Omega \times \mathcal{B} &\longrightarrow [0,1] \\ (\omega, B) &\longmapsto \sigma_n(\omega, B) := \frac{1}{n} \sum_{l=1}^{n} \delta_{\lambda_l^{X(\omega)}}(B) \end{aligned}$$

It follows from our discussion in Example 2.19 that σ_n really is a random probability measure. How is σ_n to be interpreted? For any interval $I \subseteq \mathbb{R}$, the random variable $\sigma_n(I)$ tells us the proportion of the n eigenvalues that fall into the interval I. Thus, σ_n carries information on the location of the eigenvalues, and it is of particular interest where the eigenvalues are located in the limit, that is, for $n \to \infty$.

It is a famous theorem by Wigner that allows us to conclude under fairly weak assumptions (mainly independence of matrix entries and uniformly bounded moments) that in the limit, eigenvalues will be spread according to the semicircle distribution:

Definition 2.38. The semicircle distribution σ is the probability measure on $(\mathbb{R}, \mathcal{B})$ with Lebesgue-density f_σ where

$$\begin{aligned} f_\sigma : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto f_\sigma(x) := \frac{1}{2\pi}\sqrt{4-x^2}\mathbb{1}_{[-2,2]}(x). \end{aligned}$$

Here and throughout this text, we will denote the Lebesgue measure on $(\mathbb{R}, \mathcal{B})$ by $\lambda\!\!\lambda$. With respect to Definition 2.38, we have to prove that $f_\sigma \lambda\!\!\lambda$ is actually a probability measure. We see immediately that the measure is finite, since f_σ is bounded and has compact support. We will postpone the proof that the Lebesgue integral over f_σ is 1 to Lemma 3.8. Since convergence to the semicircle distribution is an important and ubiquitous concept, we make the following definition.

Definition 2.39. If $(\sigma_n)_n$ are the ESDs of random matrices $(X_n)_n$ and $\sigma_n \to \sigma$ weakly in expectation resp. in probability resp. almost surely, then we say that *the semicircle law holds for* $(X_n)_n$ in expectation resp. in probability resp. almost surely.

We now turn to Wigner's semicircle law. Notationally, for all $n \in \mathbb{N}$ we define the index set $\square_n := \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$.

Definition 2.40. Let for all $n \in \mathbb{N}$, $a_n = (a_n(i,j))_{(i,j)\in\square_n}$ be a family of real-valued random variables, then the sequence $(a_n)_n$ is called *Wigner scheme*, if the following holds:

i) All random variables have uniformly bounded absolute moments, that is: For all $p \in \mathbb{N}$ there exists a constant $L_p \in (0, \infty)$ such that for all $n \in \mathbb{N}$ and all $(i,j) \in \square_n$: $\mathbb{E}|a_n(i,j)|^p \leq L_p$.

ii) All random variables are standardized, that is: For all $n \in \mathbb{N}$ and all $(i,j) \in \square_n$: $\mathbb{E}a_n(i,j) = 0$ and $\mathbb{V}a_n(i,j) = 1$.

iii) The families a_n are symmetric, that is: For all $n \in \mathbb{N}$ and $(i,j) \in \square_n$ we have $a_n(i,j) = a_n(j,i)$.

iv) For all $n \in \mathbb{N}$ the family $(a_n(i,j))_{1\leq i\leq j\leq n}$ is independent.

Note in particular that in Definition 2.40 we *do not* require that the whole family $((a_n(i,j))_{1\leq i\leq j\leq n})_{n\in\mathbb{N}}$ be independent. A very simple Wigner scheme is given in the following example:

Example 2.41. Let $(a(i,j))_{1\leq i\leq j}$ be an i.i.d. family of real-valued random variables such that $\mathbb{E}|a(1,1)|^p < \infty$ for all $p \in \mathbb{N}$, $\mathbb{E}a(1,1) = 0$ and $\mathbb{V}a(1,1) = 1$. Further, set $a(i,j) := a(j,i)$ for all $1 \leq j < i$. Now set for all $n \in \mathbb{N}$ and all $(i,j) \in \square_n$: $a_n(i,j) := a(i,j)$. Roughly speaking, a_n is the $n \times n$ submatrix of the infinite matrix a. Then clearly, $(a_n)_n$ is a Wigner scheme as in Definition 2.40.

The following Theorem is called "Wigner's semicircle law."

Theorem 2.42. *Let* $(a_n)_n$ *be a Wigner scheme defined on a probability space* $(\Omega, \mathcal{A}, \mathbb{P})$. *Define for all* $n \in \mathbb{N}$ *the random matrix* X_n *by*

$$\forall\, (i,j) \in \square_n : \; X_n(i,j) := \frac{1}{\sqrt{n}} a_n(i,j).$$

Then the semicircle law holds for $(X_n)_n$ *almost surely.*

Proof. This theorem is a special case of our Theorem 4.9, see Corollary 4.11. □

Of course, a valid question is how to prove Theorem 2.42. We see that certain conditions are formulated for the matrix entries. In order to use these conditions in our analysis, how can we relate σ_n back to the entries of the random matrix? And lastly, how can we conclude (stochastic) weak convergence of the ESDs? There are (at least) two standard ways to achieve this, namely the method of moments and the Stieltjes transform method. These methods will be discussed in depth in the following two sections.

3 The Method of Moments

In Chapter 2 we have studied in depth the concepts of weak convergence of probability measures and random probability measures. In this chapter, which is adopted from our previous work [31], we want to discuss a tool which helps us to infer weak convergence: The method of moments. We will carefully develop this method for both deterministic and random probability measures. To be able to use this method correctly, we also need to delve into the moment problem. But let us first define what the moments of a measure are:

Definition 3.1. Let μ be a measure on $(\mathbb{R}, \mathcal{B})$ and $k \in \mathbb{N}_0$. If $\langle \mu, |x^k| \rangle < \infty$ (where $x^0 = 1 \,\forall\, x \in \mathbb{R}$) we call the real number

$$m_k := \langle \mu, x^k \rangle$$

the k-th moment of μ. In this case, we say that μ has a finite k-th moment. On the other hand, if $\langle \mu, |x^k| \rangle = \infty$, we say the k-th moment of μ does not exist.

3.1 The Moment Problem

In numerous applications it is important to know the moments of a probability measure or at least some properties of the moments. In the Hamburger moment problem (see [51, p. 145] and [57], for example), the question is reversed. Given a sequence of real numbers $(m_k)_{k\in\mathbb{N}_0}$, what can be said about the existence and uniqueness of a measure μ on $(\mathbb{R}, \mathcal{B})$ with moments $(m_k)_{k\in\mathbb{N}_0}$? To be more precise, does there exist a measure μ on $(\mathbb{R}, \mathcal{B})$ with moments $(m_k)_{k\in\mathbb{N}_0}$, and if so, is it the only measure with those moments? Of course, if such a measure exists, it is a probability measure iff $m_0 = 1$. It is rather surprising that the existence of such a measure can be nicely characterized:

Theorem 3.2. *A sequence of real numbers $(m_k)_{k\in\mathbb{N}_0}$ constitutes the moments of at least one measure on $(\mathbb{R}, \mathcal{B})$, if and only if for all $N \in \mathbb{N}$ the corresponding Hankel matrix*

$$\begin{pmatrix} m_0 & m_1 & m_2 & \dots & m_N \\ m_1 & m_2 & m_3 & \dots & m_{N+1} \\ m_2 & m_3 & m_4 & \dots & m_{N+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_N & m_{N+1} & m_{N+2} & \dots & m_{2N} \end{pmatrix}$$

is positive semi-definite, that is, if for all $N \in \mathbb{N}_0$ and all $\beta_0, \dots, \beta_N \in \mathbb{R}$ it holds:

$$\sum_{r,s=0}^{N} \beta_r \beta_s m_{r+s} \geq 0.$$

Proof. See [51, p. 145] in combination with the fact that a real symmetric matrix is positive definite in the real sense iff it is positive definite in the complex sense. □

Oftentimes it will not be of interest if a sequence of numbers $(m_k)_{k\in\mathbb{N}_0}$ really belongs to a probability measure, since we automatically obtain this result when employing the method of moments, see Theorem 3.4. Theorem 3.2 still has two important applications: On the one hand, if the researcher is a priori assuming the target distribution to have specific moments, Theorem 3.2 can be used to check whether this is a plausible assumption and can spare the researcher from trying to prove convergence to a non-existing probability measure. On the other hand, if one has already employed the method of moments and the moments of the target distribution have been calculated, one can a posteriori evaluate the plausibility of the calculations via Theorem 3.2. Indeed, this is not uncommon practice, see [13, p. 15], for example. In any case, what will be essential for the method of moments is the knowledge about the uniqueness of a distribution with given moments, that is, the answer to the question whether there is *at most* one distribution with a given sequence of moments.

Theorem 3.3. *Let $(m_k)_{k\in\mathbb{N}}$ be a sequence of real numbers. If one of the following three conditions holds, there is at most one probability measure on $(\mathbb{R},\mathcal{B})$ with moments $(m_k)_{k\in\mathbb{N}}$:*

i) $\sum_{k=1}^{\infty} \frac{1}{\sqrt[2k]{m_{2k}}} = \infty$ *(Carleman condition),*

ii) $\limsup_{k\to\infty} \frac{\sqrt[2k]{m_{2k}}}{2k} < \infty$,

iii) $\exists\, C, D \geq 1 : \forall\, k \in \mathbb{N} : |m_k| \leq C \cdot D^k \cdot k!$.

Further, it holds that iii) ⇒ ii) ⇒ i), that is, the Carleman condition is the weakest of the three.

Proof. *i)*: See [3, p. 85].
ii): See [20, p. 122].
iii): See [51, p. 205].
Additional statement: The additional statement also proves that *ii*) and *iii*) are sufficient when knowing that *i*) is sufficient.

We assume that *ii*) holds. Let for all $k \in \mathbb{N} : \alpha_k := \sqrt[2k]{m_{2k}} \geq 0$, then we have to show $\sum_{k=1}^{\infty} \frac{1}{\alpha_k} = \infty$ under the condition that $r := \limsup_{k\to\infty} \frac{\alpha_k}{2k} < \infty$. But there exists a $K \in \mathbb{N}$ such that for all $k \geq K$ we find $\frac{\alpha_k}{2k} \leq r + 1$, thus $\alpha_k \leq 2k \cdot (r+1)$. Due to divergence of the harmonic series we obtain:

$$\sum_{k=1}^{\infty} \frac{1}{\alpha_k} \geq \sum_{k\geq K} \frac{1}{2k\cdot(r+1)} = \infty.$$

Therefore, *i*) follows from *ii*). Now if *iii*) holds, we find for all $k \in \mathbb{N}$:

$$\frac{\sqrt[2k]{m_{2k}}}{2k} \leq \frac{\sqrt[2k]{C\cdot D^{2k}\cdot(2k)!}}{2k} \leq C\cdot D\cdot\frac{\sqrt[2k]{(2k)!}}{2k} \leq C\cdot D,$$

since $(2k)^{2k} \geq (2k)!$ yields $2k \geq \sqrt[2k]{(2k)!}$ for all $k \in \mathbb{N}$. Thus, *ii*) holds. □

3.2 The Method of Moments for Probability Measures

Now we are well-prepared to introduce the method of moments, which is a means to infer weak convergence of a sequence of distributions from the convergence of their moments.

Theorem 3.4. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}_1(\mathbb{R})$, so that all moments of every μ_n exist. If there exists a sequence of real numbers $(m_k)_{k\in\mathbb{N}}$, so that*

$$\forall\, k \in \mathbb{N} : \lim_{n\to\infty} \left\langle \mu_n, x^k \right\rangle = m_k, \tag{3.1}$$

the following statements hold:

There exists a $\mu \in \mathcal{M}_1(\mathbb{R})$ and a subsequence of $(\mu_n)_{n\in\mathbb{N}}$, which converges weakly to μ. Then $\forall\, k \in \mathbb{N} : m_k = \left\langle \mu, x^k \right\rangle$. In particular, the $(m_k)_{k\in\mathbb{N}}$ are moments of a probability measure on $(\mathbb{R}, \mathcal{B})$. Further: If μ is uniquely determined by its moments, then the entire sequence $(\mu_n)_n$ converges weakly to μ.

Proof. With (3.1) it follows with $k = 2$ and Lemma 2.13 that $(\mu_n)_{n\in\mathbb{N}}$ is tight. Therefore, with Lemma 2.15 there exists a $\mu \in \mathcal{M}_1(\mathbb{R})$ and a subsequence $J \subseteq \mathbb{N}$ such that $(\mu_n)_{n\in J}$ converges weakly to μ. With Lemma 2.11, we then obtain for all $k \in \mathbb{N}$ that $(\left\langle \mu_n, x^k \right\rangle)_{n\in J}$ converges to $\left\langle \mu, x^k \right\rangle$, since the sequence $(\left\langle \mu_n, 1 + x^{2k} \right\rangle)_{n\in J}$ is bounded and the function $x \mapsto \frac{x^k}{1+x^{2k}}$ vanishes at infinity. We conclude with (3.1) that for all $k \in \mathbb{N}$ we have $\left\langle \mu, x^k \right\rangle = m_k$, so $(m_k)_k$ are indeed moments of a probability measure.

Now, if μ is uniquely determined by its moments, then the entire sequence $(\mu_n)_{n\in\mathbb{N}}$ – and not just a subsequence – converges weakly to μ. To see this, let $(\mu_n)_{n\in I}$ be an arbitrary subsequence. By Lemma 2.9, it suffices to show that this subsequence has another subsequence that converges weakly to μ. But as above (with swapped roles of I and $\mathbb{N}$) we find a probability measure ν on $(\mathbb{R}, \mathcal{B})$ and a subsequence $J' \subseteq I$, such that that $(\mu_n)_{n\in J'}$ converges weakly to ν and the numbers $(m_k)_{k\in\mathbb{N}}$ are the moments of ν. Since μ is uniquely determined by these moments, we must have $\mu = \nu$. □

3.3 The Method of Moments for Random Probability Measures

The next theorem will generalize the method of moments to the convergence types of random probability measures, namely to weak convergence in expectation, in probability and almost surely. Although this could be presented in greater generality, we will restrict our attention to convergence of random probability measures to a *deterministic* probability measure. This is the type of convergence we will encounter in our analyses ahead.

Theorem 3.5. *Let $(\mu_n)_{n\in\mathbb{N}}$ be random probability measures on $(\mathbb{R}, \mathcal{B})$ and μ be a deterministic probability measure on $(\mathbb{R}, \mathcal{B})$ which is uniquely determined by its moments. Then assuming that all following expressions (random moments, expected random moments) are well-defined and finite, we conclude:*

i) If $\forall\, k \in \mathbb{N} : \mathbb{E}\left\langle \mu_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \mu, x^k \right\rangle$*, then* $\mu_n \xrightarrow[n\to\infty]{} \mu$ *weakly in expectation.*

ii) If $\forall\, k \in \mathbb{N} : \left\langle \mu_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \mu, x^k \right\rangle$ *in probability, then* $\mu_n \xrightarrow[n\to\infty]{} \mu$ *weakly in probability.*

iii) If $\forall\, k \in \mathbb{N} : \left[\left\langle \mu_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \mu, x^k \right\rangle\ \mathbb{P}\text{-}a.s.\right]$*, then* $\mu_n \xrightarrow[n\to\infty]{} \mu$ *weakly almost surely.*

Proof. i) In the statement of the theorem, for every $n, k \in \mathbb{N}$, the random variable $\left\langle \mu_n, x^k \right\rangle$ is assumed to be well-defined, real-valued and $\mathbb{P}$-integrable. This implies:

a) $\mathbb{E}\mu_n$ has existing moments of all orders.

b) $\forall\, k \in \mathbb{N} : \left\langle \mathbb{E}\mu_n, x^k \right\rangle = \mathbb{E}\left\langle \mu_n, x^k \right\rangle$.

Of course, *b*) will follow from *a*) with Theorem 2.20. To show *a*), let $n, k \in \mathbb{N}$ be arbitrary and observe

$$\left\langle \mathbb{E}\mu_n, |x^k| \right\rangle^2 \leq \left\langle \mathbb{E}\mu_n, x^{2k} \right\rangle = \mathbb{E}\left\langle \mu_n, x^{2k} \right\rangle < \infty,$$

where we first applied Jensen's inequality and then Theorem 2.20. But now we have

$$\forall\, k \in \mathbb{N} : \left\langle \mathbb{E}\mu_n, x^k \right\rangle = \mathbb{E}\left\langle \mu_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \mu, x^k \right\rangle$$

and thus $\mathbb{E}\mu_n \to \mu$ weakly as $n \to \infty$ by Theorem 3.4.
ii) We want to show that $\mu_n \to \mu$ weakly in probability, which means that for all $f \in \mathcal{C}_b(\mathbb{R})$, $\langle \mu_n, f \rangle$ converges to $\langle \mu, f \rangle$ in probability. To this end, let $f \in \mathcal{C}_b(\mathbb{R})$ be arbitrary. To show that $(\langle \mu_n, f \rangle)_{n\in\mathbb{N}}$ converges to $\langle \mu, f \rangle$ in probability we will show that any subsequence has an almost surely convergent subsequence: Let $J \subseteq \mathbb{N}$ be a subsequence. Applying Lemma 2.28 we find a subsequence $I \subseteq J$ and a measurable set $\Omega_1 \subseteq \Omega$ of measure 1, such that

$$\forall\, \omega \in \Omega_1 : \forall\, k \in \mathbb{N} : \left\langle \mu_n(\omega), x^k \right\rangle \xrightarrow[n\in I]{} \left\langle \mu, x^k \right\rangle.$$

In particular, with Theorem 3.4 we find that for all $\omega \in \Omega_1$, $\mu_n(\omega)$ converges weakly to μ for $n \in I$, so that in particular, $\langle \mu_n(\omega), f \rangle \to \langle \mu, f \rangle$ for $n \in I$. Therefore, $\langle \mu_n, f \rangle \to \langle \mu, f \rangle$ almost surely for $n \in I$.
iii) For all $k \in \mathbb{N}$ we find a measurable set $\Omega_k \subseteq \Omega$ with measure 1 such that for all $\omega \in \Omega_k : \left\langle \mu_n(\omega), x^k \right\rangle \to \left\langle \mu, x^k \right\rangle$ as $n \to \infty$. Then $\Omega' := \cap_{k\in\mathbb{N}}\Omega_k$ has measure 1 and for all $\omega \in \Omega'$ we find that $\left\langle \mu_n(\omega), x^k \right\rangle \to \left\langle \mu, x^k \right\rangle$ for all $k \in \mathbb{N}$, so that with Theorem 3.4, for all $\omega \in \Omega'$ we have that $\mu_n(\omega)$ converges weakly to μ. Therefore, μ_n converges weakly to μ almost surely. □

We refer the reader to Remark 2.26 for an explanation on the use of brackets [. . .] in Theorem 3.5 *iii*).

Remark 3.6. The method of moments for random probability measures (Theorem 3.5) works as follows: If one wants to show weak convergence of random probability measures in expectation, in probability or almost surely, it will suffice to show that the random moments converge in expectation, in probability or almost surely. This is a very useful theorem, in particular considering we do not make any assumptions on the target measure μ except those mentioned in Theorem 3.5. In particular, we do not require the target probability measure to have a compact support. In the literature on random matrices, this condition is often used to justify the method of moments, see [6, p. 11], for example.

The next theorem will help us to determine when the conditions for Theorem 3.5 are met, to be more precise, when we are able to confirm convergence of the moments in probability or almost surely. Further, it does not assume a priori the knowledge of the target measure $\mu \in \mathcal{M}_1(\mathbb{R})$. In summary, this is the theorem that is used when applying the method of moments to random matrix theory, see also Theorems 3.12 and 3.14.

Theorem 3.7. *Let $(\mu_n)_{n\in\mathbb{N}}$ be random probability measures on $(\mathbb{R}, \mathcal{B})$ and $(m_k)_{k\in\mathbb{N}}$ be a sequence of real numbers, so that there is at most one probability measure on $(\mathbb{R}, \mathcal{B})$ with moments $(m_k)_{k\in\mathbb{N}}$. We formulate the following conditions, where we assume that all expressions (random moments, expectations and variances) are finite:*

(M1) $\forall\, k \in \mathbb{N} : \mathbb{E}\left\langle \mu_n, x^k \right\rangle \xrightarrow[n\to\infty]{} m_k,$

(M2) $\exists\, z \in \mathbb{N} : \forall\, k \in \mathbb{N} : \mathbb{E}\left(\left|\left\langle \mu_n, x^k \right\rangle - \mathbb{E}\left\langle \mu_n, x^k \right\rangle\right|^z\right) \xrightarrow[n\to\infty]{} 0,$

(M3) $\exists\, z \in \mathbb{N} : \forall\, k \in \mathbb{N} : \mathbb{E}\left(\left|\left\langle \mu_n, x^k \right\rangle - \mathbb{E}\left\langle \mu_n, x^k \right\rangle\right|^z\right) \xrightarrow[n\to\infty]{} 0$ summably fast.

Then we conclude:

i) If (M1) holds, then there is a $\mu \in \mathcal{M}_1(\mathbb{R})$ with moments $(m_k)_{k\in\mathbb{N}}$, so that $\mathbb{E}\mu_n \to \mu$ weakly (that is, $\mu_n \to \mu$ weakly in expectation). In particular, the numbers $(m_k)_{k\in\mathbb{N}}$ are the moments of a probability measure.

ii) If (M1) and (M2) hold , we conclude

$$\forall\, k \in \mathbb{N} : \left\langle \mu_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \mu, x^k \right\rangle \text{ in probability}$$

and thus $\mu_n \to \mu$ weakly in probability via Theorem 3.5.

iii) If (M1) and (M3) hold , we conclude

$$\forall\, k \in \mathbb{N} : \left[\left\langle \mu_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \mu, x^k \right\rangle \ \mathbb{P}\text{-a.s.}\right]$$

and thus $\mu_n \to \mu$ weakly almost surely via Theorem 3.5.

Proof. i) As we saw in the beginning of the proof of Theorem 3.5, we find that for all $n \in \mathbb{N}$, the expected measure $\mathbb{E}\mu_n$ has moments of all orders and that for all $k \in \mathbb{N}$:

$\langle \mathbb{E}\mu_n, x^k \rangle = \mathbb{E}\langle \mu_n, x^k \rangle$. Now given (M1), statement *i)* follows directly with Theorem 3.4.
ii)/iii) Let $k, z \in \mathbb{N}$ and $\epsilon > 0$ be arbitrary, then we first observe that

$$\begin{aligned}
&\mathbb{P}\left(\left|\langle \mu_n, x^k \rangle - \langle \mu, x^k \rangle\right| > \epsilon\right) \\
&\leq \mathbb{P}\left(\left|\langle \mu_n, x^k \rangle - \mathbb{E}\langle \mu_n, x^k \rangle + \mathbb{E}\langle \mu_n, x^k \rangle - \langle \mu, x^k \rangle\right| > \epsilon\right) \\
&\leq \mathbb{P}\left(\left|\langle \mu_n, x^k \rangle - \mathbb{E}\langle \mu_n, x^k \rangle\right| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(\left|\mathbb{E}\langle \mu_n, x^k \rangle - \langle \mu, x^k \rangle\right| > \frac{\epsilon}{2}\right) \\
&\leq \frac{\mathbb{E}|\langle \mu_n, x^k \rangle - \mathbb{E}\langle \mu_n, x^k \rangle|^z}{\frac{\epsilon^z}{2^z}} + \mathbb{P}\left(\left|\mathbb{E}\langle \mu_n, x^k \rangle - \langle \mu, x^k \rangle\right| > \frac{\epsilon}{2}\right),
\end{aligned}$$

where we applied Markov's inequality in the last step. We note that if (M1) holds, the second summand on the r.h.s. vanishes ϵ-finally (which means for all $n \geq N$, where $N \in \mathbb{N}$ is a number that depends on ϵ). In particular, if (M2) holds, we obtain that $\langle \mu_n, x^k \rangle \to_n \langle \mu, x^k \rangle$ in probability and if (M3) holds we obtain $\langle \mu_n, x^k \rangle \to_n \langle \mu, x^k \rangle$ almost surely by the Lemma of Borel-Cantelli. This shows the theorem.

□

3.4 The Moments of the Semicircle Distribution

In random matrix theory, the probability measure that appears as the limit of the empirical spectral distribution is typically the semicircle distribution as defined in Definition 2.38. What we mean by *typically* is that it appears in Wigner's semicircle law, Theorem 2.42, which is the easiest non-trivial random matrix ensemble, for it has standardized entries which are independent up to the symmetry constraint. It is safe to say that the role of the semicircle distribution in random matrix theory is as large as the role of the standard normal distribution in probability theory. To remind the reader, the semicircle distribution σ is the probability measure on $(\mathbb{R}, \mathcal{B})$ with Lebesgue-density f_σ where

$$\begin{aligned}
f_\sigma : \mathbb{R} &\longrightarrow \mathbb{R} \\
x &\longmapsto f_\sigma(x) := \frac{1}{2\pi}\sqrt{4 - x^2}\mathbb{1}_{[-2,2]}(x).
\end{aligned}$$

Since we would like to apply the method of moments to random matrix theory, we will proceed to derive the moments of the semicircle distribution. As it turns out, we will obtain that $\langle \sigma, x^0 \rangle = 1$, so that σ is identified as a probability measure, which we still owed to the reader.

Lemma 3.8. *The moments of the semicircle distribution σ are given by*

$$m_k^\sigma = \begin{cases} \frac{k!}{\frac{k}{2}!(\frac{k}{2}+1)!} & \text{for } k \in \mathbb{N}_0 \text{ even,} \\ 0 & \text{for } k \in \mathbb{N}_0 \text{ odd.} \end{cases} \tag{3.2}$$

Proof. We follow the sketch in [6, p. 7]. To this end, note that the integrand is compactly supported and bounded. Further, for odd k the integrand is odd, so the statement follows for odd k. For even k, we obtain the statement by the following calculation:

$$\begin{aligned}
m_k^\sigma &= \frac{1}{2\pi}\int_{-2}^{2} x^k\sqrt{4-x^2}\mathrm{d}x \\
&= \frac{2^{k+1}}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^k(\phi)\cos^2(\phi)\mathrm{d}\phi \\
&= \frac{2^{k+1}}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^k(\phi)(1-\cos^2(\phi))}{k+1}\mathrm{d}\phi \\
&= \frac{2^{k+1}}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^k(\phi)}{k+1}\mathrm{d}\phi - \frac{2^{k+1}}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{\sin^k(\phi)\cos^2(\phi)}{k+1}\mathrm{d}\phi,
\end{aligned}$$

where we used the substitution $x = 2\sin(\phi)$ in the second step and integration by parts in the third, where the factor to be integrated is $\sin^k \cos$. It follows

$$m_k^\sigma = \frac{2^{k+1}}{(k+2)\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^k(\phi)\mathrm{d}\phi.$$

In particular, for $k = 0$ we obtain $m_0^\sigma = 1$, so σ is indeed a probability measure. For all $k \geq 2$ even we calculate

$$\begin{aligned}
m_k^\sigma &= \frac{2^{k+1}}{(k+2)\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k-2}(\phi)(1-\cos^2(\phi))\mathrm{d}\phi \\
&= \frac{2^{k+1}}{(k+2)\pi}\left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k-2}(\phi)\mathrm{d}\phi - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^{k-2}(\phi)\cos^2(\phi)\mathrm{d}\phi\right) \\
&= \frac{2^{k+1}}{(k+2)\pi}\frac{\pi(k-1)}{2^{k-1}}m_{k-2}^\sigma \\
&= \frac{4(k-1)}{k+2}m_{k-2}^\sigma,
\end{aligned}$$

where in the third step we used the calculation from the beginning of the proof with $k-2$ instead of k. Using this recursion with $m_0^\sigma = 1$, we can prove

$$\forall\, k \in \mathbb{N}_0 \text{ even: } m_k^\sigma = \frac{k!}{\frac{k}{2}!(\frac{k}{2}+1)!}.$$

To this end, using induction, the statement is clear for $k = 0$, and if it is known for some

$k \geq 0$ even, then we calculate

$$
\begin{aligned}
m^{\sigma}_{k+2} = \frac{4(k+1)}{k+4} m^{\sigma}_k = \frac{4(k+1)}{k+4} \frac{k!}{\frac{k}{2}!(\frac{k}{2}+1)!} &\overset{!}{=} \frac{(k+2)!}{(\frac{k}{2}+1)!(\frac{k}{2}+2)!} \\
\Leftrightarrow \frac{4}{k+4}\left(\frac{k}{2}+1\right)\left(\frac{k}{2}+2\right) &= k+2 \\
\Leftrightarrow (k+2)(k+4) &= (k+2)(k+4).
\end{aligned}
$$

and the last statement is true.

□

The values of the even moments of the semicircle distribution carry a special name:

Definition 3.9. The *Catalan numbers* are elements of the sequence of natural numbers $(\mathcal{C}_k)_{k \in \mathbb{N}_0}$, where

$$\forall\, k \in \mathbb{N}_0 : \mathcal{C}_k := \frac{(2k)!}{k!(k+1)!}.$$

Combining the results of Lemma 3.8 with the definition of the Catalan numbers, we obtain for the sequence $(m^{\sigma}_k)_{k \in \mathbb{N}_0}$ of the moments of the semicircle distribution:

$$m^{\sigma}_k = \begin{cases} \mathcal{C}_{k/2} & \text{for } k \text{ even,} \\ 0 & \text{for } k \text{ odd.} \end{cases} \tag{3.3}$$

But the Catalan numbers are not only the (even) moments of the semicircle distribution. They also appear as the solution to various combinatorial problems, see [45] or [58], for example.

3.5 Application of the Method of Moments in RMT

So far we have pointed out what the method of moments is and how it works. Now we want to build the bridge to random matrix theory. To this end, we need the following observation, where as before, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$:

Lemma 3.10. *Let $n \in \mathbb{N}$ and $X \in SMat_n(\mathbb{K})$, then we obtain for all $k \in \mathbb{N}$:*

$$\sum_{i=1}^{n} (\lambda^X_i)^k = \operatorname{tr} X^k = \sum_{t_1,\dots,t_k=1}^{n} X(t_1,t_2)X(t_2,t_3)\cdots X(t_k,t_1).$$

Proof. The second equality is clear. For the first equality, note that since $X \in \mathrm{SMat}_n(\mathbb{K})$, by Lemma 2.32, there exists an invertible matrix $S \in \mathrm{Mat}_n(\mathbb{K})$ so that $X = S^{-1}DS$, where $D = \operatorname{diag}(\lambda^X_1, \dots, \lambda^X_n)$. Then

$$X^k = \underbrace{S^{-1}DS \cdot S^{-1}DS \cdot \ldots \cdot S^{-1}DS}_{k \text{ factors}} = S^{-1}D^kS = S^{-1}\operatorname{diag}\left((\lambda^X_1)^k, \dots, (\lambda^X_n)^k\right) S.$$

With Lemma 2.31, we obtain

$$\operatorname{tr}(X^k) = \operatorname{tr}\operatorname{diag}\left((\lambda_1^X)^k, \ldots, (\lambda_n^X)^k\right) = \sum_{i=1}^{n} (\lambda_i^X)^k.$$

$\square$

Corollary 3.11. *Let $(X_n)_n$ be a sequence of random matrices with corresponding ESDs $(\sigma_n)_n$. Then for all $k \in \mathbb{N}$ we find*

$$\left\langle \sigma_n, x^k \right\rangle = \frac{1}{n} \operatorname{tr} X_n^k = \frac{1}{n} \sum_{t_1,\ldots,t_k=1}^{n} X_n(t_1,t_2) X_n(t_2,t_3) \cdots X_n(t_k,t_1). \tag{3.4}$$

Proof. Using Lemma 3.10, we calculate:

$$\left\langle \sigma_n, x^k \right\rangle = \frac{1}{n} \sum_{i=1}^{n} (\lambda_i^{X_n})^k = \frac{1}{n} \operatorname{tr} X_n^k = \frac{1}{n} \sum_{t_1,\ldots,t_k=1}^{n} X_n(t_1,t_2) X_n(t_2,t_3) \cdots X_n(t_k,t_1).$$

$\square$

The next theorem will be of use in highly explorative settings, where the target distribution is not known (or assumed) yet. This is the very first step in showing that the ESDs of random matrices converge to a probability measure. To clarify terminology that we use, if Y is a $\mathbb{K}$-valued random variable, where $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, and if $p \in \mathbb{N}_0$, then we call $\mathbb{E}|Y|^p$ the p-th absolute moment of Y. Further, we say that Y has absolute moments of all orders, if $\mathbb{E}|Y|^p < \infty$ for all $p \in \mathbb{N}_0$. Note that Y is integrable iff its first absolute moment exists.

Theorem 3.12. *Let $(\sigma_n)_n$ be the empirical spectral distributions of random matrices $(X_n)_n$, whose ($\mathbb{K}$-valued) entries have absolute moments of all orders. Then if*

$$\forall\, k \in \mathbb{N} : \mathbb{E} \left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} m_k,$$

where $(m_k)_k$ is a sequence of real numbers that satisfy the Carleman condition (cf. Theorem 3.3), then $(\sigma_n)_n$ converges in expectation to a probability measure μ on $(\mathbb{R}, \mathcal{B})$ with moments $(m_k)_k$.

Proof. This follows with Theorem 3.7, since by Corollary 3.11, for each $k \in \mathbb{N}_0$, the k-th random moment is given by

$$\left\langle \sigma_n, x^k \right\rangle = \frac{1}{n} \sum_{t_1,\ldots,t_k=1}^{n} X_n(t_1,t_2) X_n(t_2,t_3) \cdots X_n(t_k,t_1),$$

which is a real-valued random variable whose expectation is finite, see the following Lemma 3.13. $\square$

Lemma 3.13. *Let $Y_1, \dots, Y_k$ be $\mathbb{K}$-valued random variables such that $\mathbb{E}|Y_i|^k < \infty$ for all $i \in \{1, \dots, k\}$, then*

$$\mathbb{E}|Y_1 Y_2 \cdots Y_k| \leq \max_{i=1,\dots,k} \mathbb{E}|Y_i|^k$$

Proof. This follows with a repeated application of Holder's inequality, see [41, p. 31]. □

We remind the reader that convergence in expectation is a necessity for stronger convergence types, see Theorem 2.29. Therefore, Theorem 3.12 is really the basis for any explorative analysis. The next Theorem will be of use either after Theorem 3.12 has been applied or if a priori, one has the target distribution of the ESDs in mind, for example if one wants to show a semicircle law.

Theorem 3.14. *Let $(\sigma_n)_n$ be the empirical spectral distributions of Hermitian random matrices $(X_n)_n$, whose entries have absolute moments of all orders. Denote by μ a probability measure which is uniquely determined by its moments (cf. Theorem 3.3). Then*

i) σ_n converges to μ weakly in expectation, if

$$\forall\, k \in \mathbb{N} : \mathbb{E}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \mu, x^k \right\rangle,$$

ii) σ_n converges to μ weakly in probability, if i) holds and for some $z \in \mathbb{N}$:

$$\forall\, k \in \mathbb{N} : \mathbb{E}\left|\left\langle \sigma_n, x^k \right\rangle - \mathbb{E}\left\langle \sigma_n, x^k \right\rangle\right|^z \xrightarrow[n\to\infty]{} 0,$$

iii) σ_n converges to μ weakly almost surely, if i) holds and for some $z \in \mathbb{N}$:

$$\forall\, k \in \mathbb{N} : \mathbb{E}\left|\left\langle \sigma_n, x^k \right\rangle - \mathbb{E}\left\langle \sigma_n, x^k \right\rangle\right|^z \xrightarrow[n\to\infty]{} 0 \quad \textit{summably fast.}$$

Proof. This is a direct consequence of Theorem 3.7, considering that since matrix entries have moments of all orders, Corollary 3.11 and Lemma 3.13 imply that expected random moments and all other expectations are well-defined and finite. □

Next, as an application, let us discuss the proof strategy behind Wigner's semicircle law, Theorem 2.42, where we restrict our attention to convergence in probability:

Example 3.15. Consider the setup of Theorem 2.42. Let $(m_k^\sigma)_{k\in\mathbb{N}}$ denote the moments of the semicircle distribution, then we can use Theorem 3.14 and show that

1. For all $k \in \mathbb{N}$:

$$\mathbb{E}\left\langle \sigma_n, x^k \right\rangle = \frac{1}{n^{1+k/2}} \sum_{t_1,\dots,t_k=1}^{n} \mathbb{E}a(t_1,t_2)a(t_2,t_3)\cdots a(t_k,t_1) \xrightarrow[n\to\infty]{} m_k^\sigma.$$

2. For all $k \in \mathbb{N}$:

$$\mathbb{E}\left(\left\langle \sigma_n, x^k \right\rangle^2\right) \xrightarrow[n\to\infty]{} (m_k^\sigma)^2.$$

This will imply statements i) and ii) from the preceding theorem with $z = 2$, thus the semicircle law in probability.

This is also exactly what is shown in [6], as can be seen from their Lemma 2.1.6 in combination with the proof of their Lemma 2.1.7. However, although Theorem 3.14 yields that above points 1 and 2 suffice for weak convergence in probability, in [6] further cumbersome calculations are carried out, utilizing the compactness of the support of the semicircle distribution, which can be observed on their pages 10 and 11.

4 Random Band Matrices with Correlated Entries

4.1 Introduction and Setup

We will start by following up on the publication [37]: In their 2015 paper, Hochstättler, Kirsch and Warzel defined a new scheme of random matrices they called *almost uncorrelated*. The entries in these matrices are allowed to exhibit a certain correlation structure. For an almost uncorrelated ensemble, they showed convergence of the empirical spectral distribution to the semicircle law weakly in probability. Naturally, it remained to investigate if convergence also holds almost surely. In this part of the thesis, we will provide a positive answer, using the method of moments. Further, we generalize this result to also be valid for band matrices. To be more precise, we study random band matrices whose entries are almost uncorrelated. We show convergence of the ESDs in probability and also almost surely, where for the latter result, a minimum growth rate of the bandwidth of the band matrices seems indispensable. In addition, for non-periodic band matrices, we will define a new parameter called "halfwidth" on which convergence statements will depend. As a special case, we will also obtain almost sure convergence of the ESDs of band matrices with independent entries to the semicircle distribution.

To achieve our results, new combinatorial ideas must be developed, and the original definition of almost uncorrelated schemes is slightly altered to be aligned with the new combinatorial arguments. The underlying model introduced in this thesis will be called *α-almost uncorrelated* to distinguish it from the 2015 ensemble and to place emphasis on the model parameter α. Here, α is a parameter which controls the correlation in our ensemble. A smaller α is associated with large correlations and vice versa.

In 2015, the driving motivating factor for the analysis of almost uncorrelated random matrices were random matrices whose entries are Curie-Weiss distributed. Curie-Weiss distributed random variables are exchangeable, whereas the general definition of almost uncorrelated schemes also admit non-exchangeable families of random variables. The updated ensemble in this thesis will still admit Curie-Weiss distributed matrix entries, but also non-exchangeable variables. In fact, we will present a new example of almost uncorrelated random matrices, which are filled with not necessarily exchangeable correlated Gaussian entries.

Chapter 4 of the thesis is organized as follows: Right after this introduction, we will develop the underlying scheme of random matrices, which we call α-almost uncorrelated. We will also introduce the structure of periodic band matrices. In Section 4.2 we will begin by stating our main result for periodic band matrices, Theorem 4.9. This theorem will be formulated in a rather general manner. Therefore, it will be important to draw simple corollaries that exemplify the reach of the theorem. The remainder of the second section will be devoted to examples of α-almost uncorrelated triangular schemes, namely those with Curie-Weiss distributed and Gaussian entries. We will also elaborate on the connection to the previous almost uncorrelated scheme developed in [37]. Section 4.3 is devoted to the proof of Theorem 4.9. In the section that follows, Section 4.4, we will extend our results to non-periodic band matrices. To do so, we will formulate a rather general theorem, which allows us to conclude asymptotic equivalence of ESDs of periodic and non-periodic random band matrices, hence might also be of general interest. Section 4.5 is devoted to auxiliary statements that we need for our proofs. A part of this chapter is already available as a preprint in [30].

Almost Uncorrelated Triangular Schemes and Band Matrices

As usual in random matrix theory, one first defines a sequence of underlying random matrices and then alters it in such a way (e.g. through rescaling), that a non-trivial limit of the empirical spectral distribution can be proved. To do this, we use the concept of triangular schemes. As before, for $n \in \mathbb{N}$ we will denote by $\square_n$ the "square of index pairs" of dimension n, the set $\{1, \dots, n\} \times \{1, \dots, n\}$.

Definition 4.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and for all $n \in \mathbb{N}$ let $\{a_n(p,q) : (p,q) \in \square_n\}$ be a family of real-valued random variables. Then we call the sequence $(a_n)_{n\in\mathbb{N}}$ a *triangular scheme*, if it is symmetric, that is, if the following holds: $\forall\, n \in \mathbb{N} : \forall\, (p,q) \in \square_n : a_n(q,p) = a_n(p,q)$.

A triangular scheme received its name due to the fact that it is completely determined by fixing the upper right triangle including the main diagonal. We have already seen an example of a triangular scheme, namely the Wigner schemes as in Definition 2.40. To identify entries in a triangular scheme which are different taking the symmetry into account, we make the following definition:

Definition 4.2. We will call two pairs of numbers $(a,b),\ (c,d) \in \mathbb{N}^2$ *fundamentally equal*, if $\{a,b\} = \{c,d\}$ and *fundamentally different*, if they are not fundamentally equal, that is, if $\{a,b\} \neq \{c,d\}$.

We will now define α-almost uncorrelated triangular schemes. These are the underlying and non-scaled random matrices that we study in this chapter. Notationally, we will denote by $\#M$ the cardinality of the set M.

Definition 4.3. Let $(a_n)_{n\in\mathbb{N}}$ be a triangular scheme and $\alpha > 0$. Then we call $(a_n)_{n\in\mathbb{N}}$ *α-almost uncorrelated*, if for all fixed $N, l \in \mathbb{N}$ and fundamentally different pairs (p_1, q_1), ..., (p_l, q_l) in $\square_N$ we have

(AAU1) Distinct decay and boundedness condition: For all $\delta_1, \dots, \delta_l \in \mathbb{N}$ we have

$$\forall\, n \geq N : \ |\mathbb{E} a_n(p_1,q_1)^{\delta_1} \cdots a_n(p_l,q_l)^{\delta_l}| \leq \frac{C_{\Phi(\delta_1,\dots,\delta_l)}}{n^{\alpha \cdot \#\{i \in \{1,\dots,l\} | \delta_i = 1\}}}, \tag{4.1}$$

(AAU2) Second moment condition:

$$\forall\, n \geq N: \; |\mathbb{E} a_n(p_1, q_1)^2 \cdots a_n(p_l, q_l)^2 - 1| \leq C_n^{(l)}, \tag{4.2}$$

and *strongly α-almost uncorrelated*, if additionally it holds:

(AAU3) Fourth moment condition:

$$\forall\, n \geq N: \; |\mathbb{E} a_n(p_1, q_1)^4 \cdot (a_n(p_2, q_2)^2 \cdots a_n(p_l, q_l)^2 - 1)| \leq D_n^{(l)}, \tag{4.3}$$

Here, Φ is a function that assigns to $(\delta_1, \ldots, \delta_l)$ a $(\delta_1 + \ldots + \delta_l)$-vector, where the m-th element is given by

$$(\Phi(\delta_1, \ldots, \delta_l))_m = \#\{i \in \{1, \ldots, l\} : \delta_i = m\}.$$

Especially, all constants of the form $C_{\Phi(\delta_1, \ldots, \delta_l)}$ do not depend on n. Further, the sequences of the form $(C_n^{(l)})_n$ and $(D_n^{(l)})_n$ shall be non-negative real sequences that converge to zero and depend on only on l. Further requirements about their convergence speed to zero will be made in the statement of the main theorem.

Note that in above definition, the conditions (4.1), (4.2) and (4.3) do not make sense for all $n \in \mathbb{N}$, but only for those n large enough so that the factors $a_n(p_i, q_i)$ exist, that is, n must be larger than all the p_i's and q_i's. This is precisely the role of the number N. That is, first an N is fixed, then pairs $(p_i, q_i) \in \square_N$ are fixed, and then the three conditions are formulated for all $n \geq N$. Note also for condition (4.3) that by convention, empty products shall have the value 1, so that the case $l = 1$ in that condition is trivial.

The properties of the entries of an almost uncorrelated triangular scheme are manifold, but let us mention a few of them: In general, entries need not be independent, and they need not have zero expectation nor unit variance, but they do so asymptotically. Further, (AAU1) quantifies the correlation decay between the variables. It is required that for each single factor in the product of the random variables, we obtain an increase of the decay rate of their expectation by n^α. A small value of α will thus lead to slowly decaying correlations whereas a large value will lead to a fast decay of correlations within the random matrices. Additionally, (AAU1) yields bounded moments of any order, uniformly over all random variables of the scheme. The second moment condition (AAU2) is to be interpreted as an extension of the requirement that asymptotically, matrix entries should have unit variance. Lastly, let us turn to condition (AAU3). One way to interpret (AAU3) is that a fourth power should not hinder the product of squares to converge to 1 in expectation, as postulated in (AAU2). Further, it is clear that (AAU3) is satisfied when we have standardized independent entries, so we require this property to carry over to the correlated case. Indeed, all examples of random matrices treated in this text will satisfy (AAU3). Further, it should be noted that in many results we obtain, condition (AAU3) is not needed, and this will be pointed out at the appropriate places by requiring either an α-almost uncorrelated or a *strongly* α-almost uncorrelated triangular scheme. A sufficient condition which helps to verify (AAU3) in practice is given in Lemma 4.18.

Although more involved examples will be discussed in the next section, we would like to give the simplest example α-almost uncorrelated schemes, namely Wigner schemes as in Definition 2.40.

Lemma 4.4. *Let $(a_n)_n$ be a Wigner scheme, then $(a_n)_n$ is strongly α-almost uncorrelated for any $\alpha > 0$.*

Proof. We need to prove that the triangular scheme $(a_n)_n$ as in the statement satisfies the conditions (AAU1), (AAU2) and (AAU3). To see this, pick $N, l \in \mathbb{N}$ and fundamentally different pairs (p_1, q_1), …, (p_l, q_l) in $\square_N$. Due to independence and unit variance, the left hand sides of (4.2) and (4.3) will vanish. Therefore, we can set $C_n^{(l)} := D_n^{(l)} := 0$ for all $n \in \mathbb{N}$. To check (4.1), denote by $(L_p)_{p\in\mathbb{N}}$ the uniform bounds on the respective p-th absolute moments of the variables $a_n(i, j)$, as given in Defintion 2.40. Let $\delta_1, \ldots, \delta_l \in \mathbb{N}$ be arbitrary. Due to independence end centrality, we can set $C_{\Phi(\delta_1,\ldots,\delta_l)} := 0$ whenever $\Phi(\delta_1, \ldots, \delta_l)_1 \geq 1$, since then the left hand side of (4.1) vanishes. On the other hand, if $\Phi(\delta_1, \ldots, \delta_l)_1 = 0$, then set $C_{\Phi(\delta_1,\ldots,\delta_l)} := L_\delta$ where $\delta := \delta_1 + \ldots + \delta_l$. Then for all $n \geq N$ we obtain

$$|\mathbb{E}a_n(p_1, q_1)^{\delta_1} \cdots a_n(p_l, q_l)^{\delta_l}| \leq \max\left\{\mathbb{E}|a_n(p_1, q_1)|^{\delta}, \ldots, \mathbb{E}|a_n(p_l, q_l)|^{\delta}\right\} \leq L_\delta = C_{\Phi(\delta_1,\ldots,\delta_l)},$$

where we used Lemma 3.13. In both cases whether there are single factors or not, we obtain validity of (4.1) with the constants we defined and for all $\alpha > 0$. □

Thus far we constructed the underlying sequence of α-almost uncorrelated triangular schemes. Since in this section we want to derive conclusions about the ESDs of periodic band matrices, let us obtain a first intuition about their structure. A 6×6 periodic band matrix M with bandwidth 3 has the structure

$$M = \begin{pmatrix} x_{1,1} & x_{1,2} & 0 & 0 & 0 & x_{1,6} \\ x_{2,1} & x_{2,2} & x_{2,3} & 0 & 0 & 0 \\ 0 & x_{3,2} & x_{3,3} & x_{3,4} & 0 & 0 \\ 0 & 0 & x_{4,3} & x_{4,4} & x_{4,5} & 0 \\ 0 & 0 & 0 & x_{5,4} & x_{5,5} & x_{5,6} \\ x_{6,1} & 0 & 0 & 0 & x_{6,5} & x_{6,6} \end{pmatrix},$$

whereas with bandwidth 5 we obtain the structure

$$M = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & 0 & x_{1,5} & x_{1,6} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & 0 & x_{2,6} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & 0 \\ 0 & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & x_{4,6} \\ x_{5,1} & 0 & x_{5,3} & x_{5,4} & x_{5,5} & x_{5,6} \\ x_{6,1} & x_{6,2} & 0 & x_{6,4} & x_{6,5} & x_{6,6} \end{pmatrix}.$$

Loosely speaking, the bandwidth is the width of the diagonal in the middle of the matrix. It is also the number of permitted non-trivial entries in each row of the matrix, and the number of non-vanishing diagonals in the upper right triangle of the matrix, including the main diagonal.

We make the observation that the concept of a "bandwidth" makes sense for each odd natural number smaller than n, or for n itself (regardless if n is odd or not), in which case we obtain the full matrix. This motivates the following definition.

Definition 4.5. Let $n \in \mathbb{N}$ be arbitrary, then a number $b_n \in \mathbb{N}$ is called *(n-)bandwidth*, if $b_n \in \{b < n \,|\, b \text{ odd}\} \cup \{n\}$.

Further, given an $n \times n$ periodic band matrix with bandwidth b_n, we want to be able to identify index pairs (i, j) which are *relevant*, that is, whose matrix entries do not vanish identically due to the band structure.

Definition 4.6. Let $n \in \mathbb{N}$ and b_n be a bandwidth, then an index pair $(i, j) \in \square_n$ is called b_n*-relevant*, if

$$|i-j| \leq \frac{b_n - 1}{2} \qquad \text{or} \qquad |i-j| \geq n - \frac{b_n - 1}{2},$$

or if $b_n = n$.

A simple observation is that if $n \in \mathbb{N}$ and b_n is a bandwidth, then there are exactly $n \cdot b_n$ relevant index pairs in $\square_n$, since there are b_n non-trivial entries per row of the matrix.

Given any $n \times n$ matrix, we can convert it into a periodic band matrix with a given bandwidth b_n by dropping the non-relevant entries:

Definition 4.7. Let $(a_n)_n$ be a sequence of arbitrary $n \times n$-matrices (for example, a triangular scheme) and $b = (b_n)_n$ be a sequence of n-bandwidths, then we define the periodic band matrices a_n^b as

$$\forall\, n \in \mathbb{N} : \forall\, (i, j) \in \square_n : \; a_n^b(i, j) := \begin{cases} a_n(i, j) & \text{if } (i, j) \text{ is } b_n\text{-relevant} \\ 0 & \text{otherwise.} \end{cases}$$

Definition 4.8. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(a_n)_{n \in \mathbb{N}}$ a triangular scheme and $b = (b_n)_n$ a sequence of n-bandwidths.

1. We say that a sequence of random matrices $(X_n)_{n \in \mathbb{N}}$ is based on the triangular scheme $(a_n)_{n \in \mathbb{N}}$, if for all $n \in \mathbb{N}$ we have:
   i) X_n has dimension $n \times n$.
   ii) $X_n(p, q) = \frac{1}{\sqrt{n}} a_n(p, q) \quad \forall\, (p, q) \in \square_n$.
2. We say that a sequence of periodic random band matrices $(X_n)_{n \in \mathbb{N}}$ is *based on the triangular scheme* $(a_n)_{n \in \mathbb{N}}$ *with bandwidth* b, if for all $n \in \mathbb{N}$ we have:
   i) X_n has dimension $n \times n$.
   ii) $X_n(p, q) = \frac{1}{\sqrt{b_n}} a_n^b(p, q) \quad \forall\, (p, q) \in \square_n$.

Please note in point 2 of the above definition that if $b_n = n$ for all n, we obtain for all n that

$$X_n(p,q) = \frac{1}{\sqrt{n}} a_n(p,q) \ \forall \, (p,q) \in \square_n,$$

since we are considering the full matrices. Thus, in case of full bandwidth, we are actually dealing with "un-banded" random matrices. In particular, the first part of Definition 4.8 is encapsulated in the second. Nevertheless, we grant the first part of the definition its own place, since it is the most common way to construct and rescale random matrices.

4.2 Results and Examples

Now, let us turn to the main theorem of this chapter and draw some corollaries. The corollaries will be proved immediately, whereas the lengthy proof of the main theorem will be postponed and carried out in Sections 4.3.1, 4.3.2 and 4.3.3. We remind the reader that Wigner schemes are defined in Definition 2.40.

Main Result and Corollaries

Theorem 4.9. *Let $(a_n)_n$ be an α-almost-uncorrelated triangular scheme, $b = (b_n)_n$ be a sequence of n-bandwidths with $b_n \to \infty$ and $(X_n)_n$ be the periodic random band matrices which are based on $(a_n)_n$ with bandwidth b. Then we obtain the following results:*

1. *If $\alpha \geq \frac{1}{2}$, then the semicircle law holds for $(X_n)_n$ in probability.*
2. *If $\alpha \geq \frac{1}{2}$, $\frac{1}{b_n^3}$ is summable over n and all entries of $(a_n)_n$ are $\{-1,1\}$-valued, then the semicircle law holds almost surely for $(X_n)_n$.*
3. *If $(a_n)_n$ is even strongly α-almost-uncorrelated with $\alpha > \frac{1}{2}$, and the sequences $\frac{1}{b_n^2}$, $\frac{1}{b_n} D_n^{(l)}$ and $C_n^{(l)}$ are summable over n, then we obtain the semicircle law almost surely for $(X_n)_n$.*
4. *If $(a_n)_n$ is a Wigner scheme and if $(\frac{1}{nb_n})_n$ is summable, then we obtain the semicircle law almost surely for $(X_n)_n$.*

Remark 4.10. Before we proceed, let us elaborate on the requirements in statement 3 of Theorem 4.9. Firstly, the requirement that $\frac{1}{b_n^2}$ be summable entails that the bandwidth has to exhibit a certain minimal growth. This condition is satisfied, for example, if $b_n \sim n^{\frac{1}{2}+\epsilon}$ for some $\epsilon > 0$, where $\sim$ denotes asymptotic equivalence of sequences, e.g. $b_n / n^{1/2+\epsilon} \to 1$.

Secondly, $\frac{1}{b_n} D_n^{(l)}$ should be summable which displays the importance of some convergence rate in the fourth moment condition. If $D_n^{(l)} \sim 1/n$, then we will *always* have that $\frac{1}{b_n} D_n^{(l)}$ is summable, since $\frac{1}{b_n^2}$ is assumed summable already. On the other hand, when considering the full bandwidth model $b_n = n$, then $D_n^{(l)} \sim 1/n^{\epsilon}$ for some $\epsilon > 0$ will suffice.

As becomes obvious in the last remark, the main theorem is a "multi-dimensional" statement. Therefore, with help of the next corollaries, we look at some important special cases of the theorem, where certain model parameters are fixed.

The first corollary will deal with the full matrix model without erased diagonals.

Corollary 4.11. *Let $(a_n)_n$ be an α-almost-uncorrelated triangular scheme and $(X_n)_n$ be the random matrices which are based on $(a_n)_n$. Then we obtain the following results:*

1. *If $\alpha \geq \frac{1}{2}$, then the semicircle law holds for $(X_n)_n$ in probability.*
2. *If $\alpha \geq \frac{1}{2}$ and all entries of $(a_n)_n$ are $\{-1, 1\}$-valued, then the semicircle law holds almost surely for $(X_n)_n$.*
3. *If $(a_n)_n$ is even strongly α-almost-uncorrelated with $\alpha > \frac{1}{2}$ and the sequences $C_n^{(l)}$ and $\frac{1}{n}D_n^{(l)}$ are summable over n, then we obtain the semicircle law almost surely for $(X_n)_n$.*
4. *If $(a_n)_n$ is a Wigner scheme, then we obtain the semicircle law almost surely for $(X_n)_n$.*

Proof. Immediate from Theorem 4.9 considering $b_n = n$. □

Note that statement 4 of the last corollary is Wigner's semicircle law. The next corollary will deal with a bandwidth increasing proportionally to n.

Corollary 4.12. *Let $(a_n)_n$ be an α-almost-uncorrelated triangular scheme, $b = (b_n)_n$ a bandwidth that grows proportionally with n, that is, there is a $p \in (0, 1)$ such that $b_n \sim p{\cdot}n$. Let $(X_n)_n$ be the periodic random band matrices which are based on $(a_n)_n$ with bandwidth b. Then we obtain the following results:*

1. *If $\alpha \geq \frac{1}{2}$, then the semicircle law holds for $(X_n)_n$ in probability.*
2. *If $\alpha \geq \frac{1}{2}$ and all entries of $(a_n)_n$ are $\{-1, 1\}$-valued, then the semicircle law holds almost surely for $(X_n)_n$.*
3. *If $(a_n)_n$ is even strongly α-almost-uncorrelated with $\alpha > \frac{1}{2}$ and the sequences $C_n^{(l)}$ and $\frac{1}{n}D_n^{(l)}$ are summable over n, then we obtain the semicircle law almost surely for $(X_n)_n$.*
4. *If $(a_n)_n$ is a Wigner scheme, then we obtain the semicircle law almost surely for $(X_n)_n$.*

Proof. This follows from Theorem 4.9: Statement 1 holds since surely, $b_n \to \infty$. Statement 2 holds since $\frac{1}{b_n^3}$ is summable, which follows from

$$\sum_{n \in \mathbb{N}} \frac{1}{b_n^3} = \sum_{n \in \mathbb{N}} \frac{(p \cdot n)^3}{b_n^3} \cdot \frac{1}{p^3 n^3}$$

and the sequence $\frac{(p\cdot n)^3}{b_n^3}$ is bounded. Statement 3 follows since $\frac{1}{b_n^2}$ is summable which follows from

$$\sum_{n\in\mathbb{N}} \frac{1}{b_n^2} = \sum_{n\in\mathbb{N}} \frac{(p\cdot n)^2}{b_n^2} \cdot \frac{1}{p^2 n^2}$$

and the sequence $\frac{(p\cdot n)^2}{b_n^2}$ is bounded. In addition, $\frac{1}{b_n} D_n^{(l)}$ is summable since

$$\sum_{n\in\mathbb{N}} \frac{1}{b_n} D_n^{(l)} = \sum_{n\in\mathbb{N}} \frac{p\cdot n}{b_n} \cdot \frac{1}{p\cdot n} D_n^{(l)}$$

and the sequence $\frac{p\cdot n}{b_n}$ is bounded. Statement 4 holds since $\frac{1}{nb_n}$ is summable due to

$$\sum_{n\in\mathbb{N}} \frac{1}{nb_n} = \sum_{n\in\mathbb{N}} \frac{p\cdot n}{b_n} \cdot \frac{1}{p\cdot n^2}$$

where the sequence $\frac{p\cdot n}{b_n}$ is bounded. □

We will now introduce some examples of α-almost uncorrelated triangular schemes. An application of Theorem 4.9 will then allow conclusions regarding the asymptotic behavior of the ESDs of the random matrices which are based on these schemes.

Curie-Weiss Ensembles and Relations to Previous Work

The previous work on almost uncorrelated ensembles in [37] (where these ensembles were invented) was motivated by random matrices with Curie-Weiss distributed entries. Let us recall their definition of almost uncorrelated triangular schemes:

Definition 4.13. A triangular scheme $(a_n)_{n\in\mathbb{N}}$ is called *almost uncorrelated*, if for all $N \in \mathbb{N}$ we have:

(AU1) Distinct decay and boundedness condition:

$$\forall\, n \geq N: \; |\mathbb{E} a_n(p_1,q_1)\cdots a_n(p_l,q_l) a_n(p_{l+1},q_{l+1})\cdots a_n(p_{l+m},q_{l+m})| \leq \frac{K_{l,m}}{n^{l/2}}$$

(AU2) Second moment condition:

$$\forall\, n \geq N: \; |\mathbb{E} a_n(p_1,q_1)^2 \cdots a_n(p_l,q_l)^2 - 1| \leq K_n^{(l)}$$

for all sequences of pairs $(p_1,q_1), \ldots, (p_{l+m},q_{l+m})$ in $\square_N$, where $l,m \in \mathbb{N}_0$, so that $(p_1,q_1),\ldots,(p_l,q_l)$ are fundamentally different from all other pairs of the entire sequence $(p_1,q_1),\ldots,(p_{l+m},q_{l+m})$.[1] Further, the constants $K_{l,m}$ are non-negative real numbers that only depend on l and m, and for all $l \in \mathbb{N}_0$ we have that $(K_n^{(l)})_{n\in\mathbb{N}}$ is a non-negative real sequence that converges to zero.

[1] To clarify, that means, for example, that (p_2,q_2) is fundamentally different from the pairs $(p_1,q_1),(p_3,q_3),(p_4,q_4),\ldots,(p_{l+m},q_{l+m})$.

How is this definition related to Definition 4.3 on page 54? On the one hand, notice that (AU2) and (AAU2) are the same. On the other hand, clearly, (AAU3) is a new condition, which we had to introduce to derive statement 3 of Theorem 4.9. But what about (AU1) and (AAU1)? In (AU1), the decay gain per single factor is set to $n^{1/2}$ whereas in (AAU1) this decay gain is n^{α} for some $\alpha > 0$. Again, this flexibility was introduced to derive statement 3 of Theorem 4.9. The following lemma sheds light on the relationship between (AU1) and (AAU1), and between almost uncorrelated and α-almost uncorrelated schemes.

Lemma 4.14. *A triangular scheme $(a_n)_n$ satisfies condition (AU1) if and only if it satisfies condition (AAU1) with $\alpha = \frac{1}{2}$. In particular, $(a_n)_n$ is almost uncorrelated iff it is $\frac{1}{2}$-almost uncorrelated.*

Proof. Due to the discussion preceding this lemma, we only need to show the first statement. We first show "$\Rightarrow$": Pick $N \in \mathbb{N}$ and an $n \geq N$ arbitrarily. Then, pick an $l \in \mathbb{N}$, $\delta_1, \ldots, \delta_l \in \mathbb{N}$ arbitrarily. Next, pick fundamentally different pairs $(p_1, q_1), \ldots, (p_l, q_l) \in \square_N$ arbitrarily. Define the two numbers

$$a := \#\{i \in \{1, \ldots, l\} | \delta_i = 1\}$$

and

$$b := \sum_{i:\delta_i \neq 1} \delta_i$$

Then we have for all $n \geq N$ by condition (AU1):

$$\begin{aligned} &|\mathbb{E} a_n(p_1, q_1)^{\delta_1} \cdots a_n(p_l, q_l)^{\delta_l}| \\ &\leq \frac{K_{a,b}}{n^{\frac{1}{2} \cdot a}} \end{aligned}$$

We observe that the numbers a and b only depend on the multiplicities of the values in the tuple $(\delta_1, \ldots, \delta_l)$. Therefore, setting

$$C_{\Phi(\delta_1, \ldots, \delta_l)} := K_{a,b}$$

will be well-defined.

Now, we show "$\Leftarrow$": Pick $N \in \mathbb{N}$ and an $n \geq N$ arbitrarily. Then, pick $l, m \in \mathbb{N}_0$ and pairs $(p_1, q_1), \ldots, (p_l, q_l), (p_{l+1}, q_{l+1}), \ldots, (p_{l+m}, q_{l+m}) \in \square_N$ arbitrarily, so that the pairs $(p_1, q_1), \ldots, (p_l, q_l)$ are fundamentally different from all other pairs in the sequence $(p_1, q_1), \ldots, (p_l, q_l), (p_{l+1}, q_{l+1}), \ldots, (p_{l+m}, q_{l+m})$. Then define the finite set

$$S(l, m) := \left\{ C_{\Phi(\delta_1, \ldots, \delta_r)} \,\middle|\, r \leq l + m,\ \#\{i : \delta_i = 1\} \geq l,\ \sum_{j=1}^{r} \delta_j = l + m \right\}$$

and the number

$$K_{l,m} := \max\{C : C \in S(l, m) \cup \{1\}\}.$$

which is well-defined, since it only depends on the numbers l and m and the set $S(l,m)$ surely is finite. Then, by (AAU1) we will have for some $C \in S(l,m)$ that

$$|\mathbb{E}a_n(p_1,q_1)\cdots a_n(p_l,q_l)a_n(p_{l+1},q_{l+1})\cdots a_n(p_{l+m},q_{l+m})| \leq \frac{C}{n^{\frac{1}{2}\cdot l}} \leq \frac{K_{l,m}}{n^{l/2}},$$

which we needed to show. □

The lemma we just proved yields the following corollary:

Corollary 4.15. *Let $(X_n)_n$ be the sequence of periodic random band matrices which is based on an almost uncorrelated triangular scheme $(a_n)_n$ with bandwidth $b = (b_n)_n$.*

1. *If $b_n \to \infty$, then the semicircle law holds for $(X_n)_n$ in probability.*
2. *If $\frac{1}{b_n^3}$ is summable over n and all entries of $(a_n)_n$ are $\{-1,1\}$-valued, then the semicircle law holds almost surely for $(X_n)_n$.*

Proof. The corollary follows directly from Lemma 4.14 and Theorem 4.9. □

We will now introduce our first example of α-almost uncorrelated triangular arrays, namely those filled with Curie-Weiss distributed random variables. We refer the reader to [41] for a rigorous treatment of the properties of the Curie-Weiss distribution.

Definition 4.16. Let $n \in \mathbb{N}$ be arbitrary and $Y_1,\ldots,Y_n$ be random variables defined on some probability space $(\Omega,\mathcal{A},\mathbb{P})$. Let $\beta > 0$, then we say that $Y_1,\ldots,Y_n$ are Curie-Weiss(β,n)-distributed, if for all $y_1,\ldots,y_n \in \{-1,1\}$ we have that

$$\mathbb{P}(Y_1 = y_1,\ldots,Y_n = y_n) = \frac{1}{Z_{\beta,n}} \cdot e^{\frac{\beta}{2n}(\sum y_i)^2}$$

where $Z_{\beta,n}$ is a normalization constant. The parameter β is called *inverse temperature*.

The Curie-Weiss(β,n) distribution is used to model the behavior of n ferromagnetic particles (spins) at the inverse temperature β. At low temperatures, that is, if β is large, all magnetic spins are likely to have the same alignment, resembling a strong magnetic effect. On the contrary, at high temperatures (if β is small), spins can act almost independently, resembling a weak magnetic effect.

Theorem 4.17. *Let $0 < \beta \leq 1$ and let for each $n \in \mathbb{N}$ the random variables $(\tilde{a}_n(i,j))_{1\leq i,j\leq n}$ be Curie-Weiss(β,n^2)-distributed. Define the triangular scheme $(a_n)_n$ by setting*

$$\forall\, n \in \mathbb{N} : \forall\, (i,j) \in \square_n : \; a_n(i,j) = \begin{cases} \tilde{a}_n(i,j) & \text{if } i \leq j \\ \tilde{a}_n(j,i) & \text{if } i > j. \end{cases}$$

Let $(X_n)_n$ be the random matrices which are based on $(a_n)_n$. Let $b = (b_n)_n$ be a sequence of n-bandwidths and $(Y_n)_n$ be the periodic random band matrices which are based on $(a_n)_n$ with bandwidth b. Then the following statements hold:

i) The triangular scheme $(a_n)_n$ is almost uncorrelated.

ii) The triangular scheme $(a_n)_n$ is $\frac{1}{2}$-almost uncorrelated.

iii) The semicircle law holds for $(X_n)_n$ almost surely.

iv) If $b_n \to \infty$ as $n \to \infty$, then the semicircle law holds for $(Y_n)_n$ in probability.

v) If $\frac{1}{b_n^3}$ is summable over n, then the semicircle law holds almost surely for $(Y_n)_n$.

Proof. In [37] it was shown that $(a_n)_n$ is almost uncorrelated, where we also refer the reader to [41] for technical details. This shows i), and ii) follows with Lemma 4.14. Statements iii), iv) and v) follow with Corollary 4.15 or Theorem 4.9. □

Correlated Gaussian Entries

We will now study random matrices filled with correlated Gaussian entries. By placing quite natural conditions on the covariance matrices of the entries with dependence on a parameter $\alpha > 0$, we will then obtain α-almost uncorrelated ensembles. In other words, α-almost uncorrelated ensembles appear quite naturally when studying correlated Gaussian ensembles. Since we are interested in semicircle laws both in probability and almost surely, we will need condition (AAU3). To validate (AAU3) in practice, we formulate the next lemma.

Lemma 4.18. *Let $(a_n)_{n\in\mathbb{N}}$ be a triangular scheme and suppose that there exists a $K \in \mathbb{R}$ such that for all $l, N \in \mathbb{N}$ and fundamentally different pairs $(p_1, q_1), \ldots, (p_l, q_l)$ in $\square_N$ we have that*

$$\forall n \geq N: \ |\mathbb{E}a_n(p_1, q_1)^4 a_n(p_2, q_2)^2 \cdots a_n(p_l, q_l)^2 - K| \leq \tilde{D}_n^{(l)}, \tag{4.4}$$

where for each $l \in \mathbb{N}$, $(\tilde{D}_n^{(l)})_n$ is a sequence converging to zero. Then (AAU3) is also satisfied with constants $D_n^{(l)} := \tilde{D}_n^{(l)} + \tilde{D}_n^{(1)}$.

Proof. We calculate

$$\begin{aligned}
&|\mathbb{E}a_n(p_1, q_1)^4 \cdot (a_n(p_2, q_2)^2 \cdots a_n(p_l, q_l)^2 - 1)| \\
&\leq |\mathbb{E}a_n(p_1, q_1)^4 a_n(p_2, q_2)^2 \cdots a_n(p_l, q_l)^2 - K| + |\mathbb{E}a_n(p_1, q_1)^4 - K| \\
&\leq \tilde{D}_n^{(l)} + \tilde{D}_n^{(1)}
\end{aligned}$$

□

What follows is a generalization of the author's work in [31]. New ideas had to be incorporated for the proceedings in the present exposition. Notationally, we define

$$\forall\, n \in \mathbb{N} : [n] := \{1, \ldots, n\}.$$

Further, for any $\alpha > 0$ we denote by $\mathrm{CovMat}(\alpha)$ the set of all sequences $(\Sigma_n)_n$, where for each $n \in \mathbb{N}$, Σ_n is a real symmetric $n \times n$ matrix with the following properties:

i) $\Sigma_n(i,i) = 1$ for all $i \in [n]$,

ii) $|\Sigma_n(i,j)| \leq 1/n^\alpha$ for all $i \neq j \in [n]$.

iii) Σ_n is positive definite.

Note that if $\alpha \geq 1$, then any symmetric $n \times n$ matrix A satisfying above conditions i) and ii) will already be positive definite, since then, A will be *strictly diagonally dominant* with strictly positive diagonal entries. We refer the reader to [49] for details.

To give examples, if $(\Sigma_n)_n \in \mathrm{CovMat}(1)$, then for $n = 4$ we could have

$$\Sigma_4 = \begin{pmatrix} 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & 1 & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 1 \end{pmatrix}$$

or

$$\Sigma_4 = \begin{pmatrix} 1 & \frac{1}{4} & -\frac{1}{5} & -\frac{1}{6} \\ \frac{1}{4} & 1 & \frac{1}{7} & \frac{1}{8} \\ -\frac{1}{5} & \frac{1}{7} & 1 & \frac{1}{9} \\ -\frac{1}{6} & \frac{1}{8} & \frac{1}{9} & 1 \end{pmatrix}.$$

In the following, we would like to define the multi-dimensional normal distribution and list some of its properties, where we followed the exposition in [44].

Definition 4.19. Let $n \in \mathbb{N}$ and Σ be a positive definite real symmetric $n \times n$ matrix, $\mu \in \mathbb{R}^n$. A random vector $Y = (Y_1, \ldots, Y_n)$ on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $\mathbb{R}^n$ is called *n-dimensional normally distributed* with expectation μ and covariance matrix Σ, if its distribution $\mathbb{P}^{(Y_1,\ldots,Y_n)}$ has Lebesgue density $f_{\mu,\Sigma}$ with

$$\forall\, y \in \mathbb{R}^n : f_{\mu,\Sigma}(y) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2} \cdot (y-\mu)^t \cdot \Sigma^{-1} \cdot (y-\mu)\right),$$

where for a (column) vector $z \in \mathbb{R}^n$ we denote by z^t the transposed vector (thus a row vector). In this case, we write $Y \sim \mathcal{N}(\mu, \Sigma)$.

Theorem 4.20. *Let $n \in \mathbb{N}$, Σ be a positive definite real symmetric $n \times n$ matrix and $\mu \in \mathbb{R}^n$. From $(Y_1, \ldots, Y_n) \sim \mathcal{N}(\mu, \Sigma)$ it follows:*

i) For all $j \in [n]$: $\mathbb{E}Y_j = \mu_j$.

ii) For all $(i,j) \in \Box_n$: $\mathrm{Cov}(Y_i, Y_j) = \Sigma(i,j)$.

iii) For all $j \in [n]$: $Y_j \sim \mathcal{N}(\mu_j, \Sigma(j,j))$.

Proof. This is part of Theorem 15.53 in [44], page 327. $\Box$

Next, we would like to be able to compute the expectation of an arbitrary product of multi-dimensional normal random variables.

Theorem 4.21. *Let $n \in \mathbb{N}$ and Σ be a positive definite, real symmetric $n \times n$ matrix. For the real-valued random variables $Y_1, \ldots, Y_n$ on $(\Omega, \mathcal{A}, \mathbb{P})$ it is assumed that $(Y_1, \ldots, Y_n) \sim \mathcal{N}(0_n, \Sigma)$ (where 0_n is the n-dimensional vector containing only zeroes). Then for all $k \in \mathbb{N}$ and $i(1), \ldots, i(k) \in [n]$, we have that*

$$\mathbb{E}Y_{i(1)} \cdots Y_{i(k)} = \sum_{\pi \in \mathcal{PP}(k)} \prod_{\{r,s\} \in \pi} \mathbb{E}Y_{i(r)}Y_{i(s)} = \sum_{\pi \in \mathcal{PP}(k)} \prod_{\{r,s\} \in \pi} \Sigma\big(i(r), i(s)\big),$$

where $\mathcal{PP}(k)$ denotes the set of all pair partitions on $\{1, \ldots, k\}$. Especially, we obtain for k odd that

$$\mathbb{E}Y_{i(1)} \cdots Y_{i(k)} = 0.$$

Proof. This theorem is known as "Wick's theorem" or "Theorem of Isserlis". It can be found in [47] or [38]. □

Example 4.22. Let $\alpha > 0$ be arbitrary and let us construct an example of a strongly α-almost uncorrelated matrix ensemble, where the sequences $(C_n^{(l)})_n$ and $(D_n^{(l)})_n$ can be chosen to be summable if $\alpha > 1/4$.

The triangular scheme will be filled with normal random variables. Due to symmetry, it suffices to specify the right upper triangle of each a_n in the triangular scheme $(a_n)_{n \in \mathbb{N}}$ now to be defined. For each $n \in \mathbb{N}$, there are at most n^2 such entries. Fix a sequence $(\Sigma_n)_n \in \text{CovMat}(\alpha)$. We will endow the right upper triangle of each a_n with variables from a random vector $(Y_1^{(n)}, \ldots, Y_{n^2}^{(n)}) \sim \mathcal{N}(0_{n^2}, \Sigma_{n^2})$, where different entries in the right upper triangle of a_n will also receive different random variables out of the vector $(Y_1^{(n)}, \ldots, Y_{n^2}^{(n)})$. To this end, for each $n \in \mathbb{N}$ we fix an injection

$$\varphi_n : \{(i,j) \in \square_n : i \leq j\} \longrightarrow \{1, \ldots, n^2\}$$

and set for all $(i,j) \in \square_n$ with $i \leq j$: $a_n(i,j) := Y_{\varphi_n(i,j)}^{(n)}$. We will proceed in this way for all $n \in \mathbb{N}$ and obtain thus a completely specified triangular scheme $(a_n)_{n \in \mathbb{N}}$. To avoid future technical difficulties, we extend for all $n \in \mathbb{N}$ the domain of φ_n onto the whole square $\square_n$, by setting for all $(i,j) \in \square_n$ with $j < i$: $\varphi_n(i,j) := \varphi_n(j,i)$.

Of course, what we have to prove next is that the triangular scheme $(a_n)_{n \in \mathbb{N}}$, which we just constructed in Example 4.22, is indeed strongly α-almost uncorrelated as in Definition 4.3 on page 54, so we need to check that the conditions (AAU1), (AAU2) and (AAU3) hold. The next three lemmas will help us in this endeavor.

Lemma 4.23. *Let $\alpha > 0$ and $(\Sigma_n)_n \in \text{CovMat}(\alpha)$ be arbitrary. Let $l \in \mathbb{N}$ and then $\delta_1, \ldots, \delta_l \in \mathbb{N}$, so that $\delta_1 + \ldots + \delta_l$ is even.*
Let $n \in \mathbb{N}$ and $i(1), i(2), \ldots, i(\delta_1 + \ldots + \delta_l) \in [n^2]$, so that

$$\begin{aligned}
&i(1) = i(2) = \ldots = i(\delta_1) \\
&i(\delta_1 + 1) = \ldots = i(\delta_1 + \delta_2) \\
&i(\delta_1 + \delta_2 + 1) = \ldots = i(\delta_1 + \delta_2 + \delta_3) \\
&\vdots \\
&i(\delta_1 + \ldots + \delta_{l-1} + 1) = \ldots = i(\delta_1 + \ldots + \delta_l)
\end{aligned}$$

but $i(1)$, $i(\delta_1+1)$, ..., $i(\delta_1+\ldots+\delta_{l-1}+1)$ *are distinct. Then we have*

$$\left|\sum_{\pi\in\mathcal{PP}(\delta_1+\ldots+\delta_l)}\prod_{\{r,s\}\in\pi}\Sigma_{n^2}\big(i(r),i(s)\big)\right|\leq\frac{\#\mathcal{PP}(\delta_1+\ldots+\delta_l)}{n^{\alpha\cdot\#\{i\,|\,\delta_i=1\}}}.$$

Proof. Each diagonal entry of Σ_{n^2} equals 1 and each off-diagonal entry lies in the interval $[-1/n^{2\alpha},1/n^{2\alpha}]$. Let $\pi\in\mathcal{PP}(\delta_1+\ldots+\delta_l)$ be arbitrary, then we have that

$$\prod_{\{r,s\}\in\pi}\Sigma_{n^2}\big(i(r),i(s)\big)$$

is a product of $(\delta_1+\ldots+\delta_l)/2$ entries of Σ_{n^2}. For each block $\{r,s\}\in\pi$ with $i(r)=i(s)$ we find $\Sigma_{n^2}\big(i(r),i(s)\big)=1$, and for each block $\{r,s\}\in\pi$ with $i(r)\neq i(s)$ we obtain $|\Sigma_{n^2}\big(i(r),i(s)\big)|\leq 1/n^{2\alpha}$. But now we have at least

$$\frac{\#\{i\,|\,\delta_i=1\}}{2}$$

blocks $\{r,s\}$ in π with $i(r)\neq i(s)$, since for any δ_i with $\delta_i=1$, the index $i(\delta_1+\ldots+\delta_i)$ is unique among all indices and must therefore share a block with a different index. Then in the worst case possible, the number of δ_i's with $\delta_i=1$ is even and their corresponding unique indices are all paired, yielding the bound above. Therefore,

$$\prod_{\{r,s\}\in\pi}|\Sigma_{n^2}\big(i(r),i(s)\big)|\leq\left(\frac{1}{n^{2\alpha}}\right)^{\frac{\#\{i|\delta_i=1\}}{2}}=\frac{1}{n^{\alpha\cdot\#\{i\,|\,\delta_i=1\}}}$$

and this bound holds for each $\pi\in\mathcal{PP}(\delta_1+\ldots+\delta_l)$, proving the lemma. □

Lemma 4.24. *Let* $\alpha>0$ *and* $(\Sigma_n)_n\in\mathrm{CovMat}(\alpha)$ *be arbitrary. Let* $n,z\in\mathbb{N}$. *Let* $i(1),\ldots,i(2z)$ *be in* $[n^2]$, *so that* $i(1)=i(2),i(3)=i(4),\ldots,i(2z-1)=i(2z)$, *but* $i(1),i(3),\ldots,i(2z-1)$ *are pairwise distinct. Then it holds:*

$$\left|\sum_{\pi\in\mathcal{PP}(2z)}\prod_{\{r,s\}\in\pi}\Sigma_{n^2}\big(i(r),i(s)\big)-1\right|\leq\frac{\#\mathcal{PP}(2z)}{n^{4\alpha}}.$$

Proof. First, let us repeat some observations that we already made in the proof of Lemma 4.23. Each diagonal entry of the matrix Σ_{n^2} equals 1 and each off-diagonal entry lies in the interval $[-1/n^{2\alpha},1/n^{2\alpha}]$. Now let $\pi\in\mathcal{PP}(2z)$ be arbitrary, then we have that

$$\prod_{\{r,s\}\in\pi}\Sigma_{n^2}\big(i(r),i(s)\big)$$

is a product of z entries of Σ_{n^2}. For each block $\{r,s\}\in\pi$ with $i(r)=i(s)$ it holds $\Sigma_{n^2}\big(i(r),i(s)\big)=1$ and for each block $\{r,s\}\in\pi$ with $i(r)\neq i(s)$ it holds $|\Sigma_{n^2}\big(i(r),i(s)\big)|\leq 1/n^{2\alpha}$.

But now consider the pair partition $\pi_0 = \{\{1,2\},\{3,4\},\dots,\{2z-1,2z\}\} \in \mathcal{PP}(2z)$. This is the only pair partition in $\mathcal{PP}(2z)$ that pairs the distinct pairs in $i(1),\dots,i(2z)$. In other words: $[\forall\,\{r,s\} \in \pi_0 : i(r) = i(s)]$. Thus

$$\prod_{\{r,s\}\in\pi_0} \Sigma_{n^2}\big(i(r),i(s)\big) = 1.$$

On the other hand, for each $\pi \in \mathcal{PP}(2z)$ with $\pi \neq \pi_0$ we find a block $\{r',s'\} \in \pi$ with $i(r') \neq i(s')$, and then necessarily at least one further block $\{r'',s''\} \in \pi$ with $i(r'') \neq i(s'')$, leading to

$$\prod_{\{r,s\}\in\pi} |\Sigma_{n^2}\big(i(r),i(s)\big)| \leq \frac{1}{n^{4\alpha}}.$$

There are at most $\#\mathcal{PP}(2z)$ partitions $\pi \in \mathcal{PP}(2z)$ with $\pi \neq \pi_0$, which concludes the proof. $\square$

Lemma 4.25. *Let $\alpha > 0$ and $(\Sigma_n)_n \in \mathrm{CovMat}(\alpha)$ be arbitrary. Let $n,z \in \mathbb{N}$ and $i(1),\dots,i(2z+2)$ be in $[n^2]$, so that $i(1) = i(2) = i(3) = i(4), i(5) = i(6), i(7) = i(8),\dots,i(2z+1) = i(2z+2)$, but $i(1), i(5), i(7),\dots,i(2z+1)$ are pairwise distinct. Then it holds:*

$$\left|\sum_{\pi\in\mathcal{PP}(2z+2)} \prod_{\{r,s\}\in\pi} \Sigma_{n^2}\big(i(r),i(s)\big) - 3\right| \leq \frac{\#\mathcal{PP}(2z+2)}{n^{4\alpha}}.$$

Proof. The proof is analogous to the proof of Lemma 4.24. Here, the pair partitions

$$\begin{aligned}
\pi_1 &= \{\{1,2\},\{3,4\},\{5,6\},\{7,8\},\dots,\{2z+1,2z+2\}\}\\
\pi_2 &= \{\{1,3\},\{2,4\},\{5,6\},\{7,8\},\dots,\{2z+1,2z+2\}\}\\
\pi_3 &= \{\{1,4\},\{2,3\},\{5,6\},\{7,8\},\dots,\{2z+1,2z+2\}\}
\end{aligned}$$

in $\mathcal{PP}(2z+2)$ are exactly the pair partitions π in $\mathcal{PP}(2z+2)$ so that each block $\{r,s\} \in \pi$ satisfies $i(r) = i(s)$. Thus for every $j \in \{1,2,3\}$, we obtain

$$\prod_{\{r,s\}\in\pi_j} \Sigma_{n^2}\big(i(r),i(s)\big) = 1.$$

On the other hand, we note that for each $\pi \in \mathcal{PP}(2z+2)$ with $\pi \neq \pi_j$ for all $j \in \{1,2,3\}$, we find a block $\{r',s'\} \in \pi$ with $i(r') \neq i(s')$, and then necessarily at least one further block $\{r'',s''\} \in \pi$ with $i(r'') \neq i(s'')$, leading to

$$\prod_{\{r,s\}\in\pi} |\Sigma_{n^2}\big(i(r),i(s)\big)| \leq \frac{1}{n^{4\alpha}}.$$

Surely, there are at most $\#\mathcal{PP}(2z+2)$ partitions $\pi \in \mathcal{PP}(2z+2)$ with $\pi \neq \pi_j$ for all $j \in \{1,2,3\}$, which concludes the proof. $\square$

Now we are ready to prove that the claims in Example 4.22 are true.

Theorem 4.26. *Let $\alpha > 0$ and $(\Sigma_n)_n \in \mathrm{CovMat}(\alpha)$ be arbitrary. Let the triangular scheme $(a_n)_{n\in\mathbb{N}}$ be constructed with respect to α and $(\Sigma_n)_n$ as in Example 4.22. Then $(a_n)_{n\in\mathbb{N}}$ is strongly α-almost uncorrelated, and if $\alpha > 1/4$, then for all $l \in \mathbb{N}$, the sequences $C_n^{(l)}$ and $D_n^{(l)}$ can be chosen summable over n. Further, this property is tight in the sense that $(a_n)_{n\in\mathbb{N}}$ need not be α'-almost uncorrelated for any $\alpha' > \alpha$.*

Proof. To prove the first statement, we first want to specify the constants $C_{\Phi(\delta_1,\ldots,\delta_l)}$ in (AAU1), which is readily done by setting $C_{\Phi(\delta_1,\ldots,\delta_l)} := \#\mathcal{PP}(\delta_1+\ldots+\delta_l)$. To prove the bound in (AAU1) with these constants, we need only consider the case where $\delta_1+\ldots+\delta_l$ is even, since if $\delta_1+\ldots+\delta_l$ is odd, then the inequality yields $0 \leq 0$ by Theorem 4.21. So let $N \in \mathbb{N}$ and $l \in \mathbb{N}$ be arbitrary, then choose arbitrary $\delta_1,\ldots,\delta_l \in \mathbb{N}$ so that $\delta_1+\ldots+\delta_l$ is even, then choose a sequence of fundamentally different pairs $(p_1,q_1),\ldots,(p_l,q_l) \in \square_N$. Then we have for all $n \geq N$ that $a_n(p_i,q_i) = Y^{(n)}_{\varphi_n(p_i,q_i)}$ for all $i \in \{1,\ldots,l\}$. Let us define the indices $i(1), i(2), \ldots, i(\delta_1+\ldots+\delta_n)$ through

$$\begin{aligned}
&i(1), i(2), \ldots, i(\delta_1) := \varphi_n(p_1,q_1)\\
&i(\delta_1+1), i(\delta_1+2), \ldots, i(\delta_1+\delta_2) := \varphi_n(p_2,q_2)\\
&\vdots\\
&i(\delta_1+\ldots+\delta_{l-1}+1), \ldots, i(\delta_1+\ldots+\delta_{l-1}+\delta_l) := \varphi_n(p_l,q_l).
\end{aligned}$$

Then the indices $i(1), i(2), \ldots, i(\delta_1+\ldots+\delta_n) \in [n^2]$ meet the conditions of Lemma 4.23. With this Lemma and Theorem 4.21 it now follows:

$$\begin{aligned}
&\left|\mathbb{E}a_n(p_1,q_1)^{\delta_1}\cdot a_n(p_2,q_2)^{\delta_2}\cdot\ldots\cdot a_n(p_l,q_l)^{\delta_l}\right|\\
&= \left|\mathbb{E}\left(Y^{(n)}_{\varphi_n(p_1,q_1)}\right)^{\delta_1}\cdot\left(Y^{(n)}_{\varphi_n(p_2,q_2)}\right)^{\delta_2}\cdot\ldots\cdot\left(Y^{(n)}_{\varphi_n(p_l,q_l)}\right)^{\delta_l}\right|\\
&= \left|\mathbb{E}Y^{(n)}_{i(1)}\cdots Y^{(n)}_{i(\delta_1+\ldots+\delta_l)}\right|\\
&= \left|\sum_{\pi\in\mathcal{PP}(\delta_1+\ldots+\delta_l)}\prod_{(r,s)\in\pi}\Sigma_{n^2}(i(r),i(s))\right| \leq \frac{\#\mathcal{PP}(\delta_1+\ldots+\delta_l)}{n^{\alpha\cdot\#\{j\,|\,\delta_j=1\}}}
\end{aligned}$$

and thus (AAU1) with $C_{\Phi(\delta_1,\ldots,\delta_l)}$ as we just defined.

Now let us turn to condition (AAU2), then we must choose appropriate sequences $(C_n^{(l)})_{n\in\mathbb{N}}$ for all $l \in \mathbb{N}$. For all $l \in \mathbb{N}$ we set $C_n^{(l)} := \frac{\#\mathcal{PP}(2l)}{n^{4\alpha}}$ for all $n \in \mathbb{N}$. Then these sequences are summable over n iff $\alpha > 1/4$.

Now let $N, l \in \mathbb{N}$ be arbitrary and $(p_1,q_1),\ldots,(p_l,q_l)$ in $\square_N$ be a sequence of fundamentally different pairs. Then we have for all $n \geq N$ that $a_n(p_i,q_i) = Y^{(n)}_{\varphi_n(p_i,q_i)}$ for all $i \in \{1,\ldots,l\}$. Further, the indices

$$\begin{aligned}
&(i(1), i(2), i(3), i(4), \ldots, i(2l-1), i(2l))\\
&\qquad := (\varphi_n(p_1,q_1), \varphi_n(p_1,q_1), \varphi_n(p_2,q_2), \varphi_n(p_2,q_2), \ldots, \varphi_n(p_l,q_l), \varphi_n(p_l,q_l))
\end{aligned}$$

in $[n^2]$ meet the conditions of Lemma 4.24. With this lemma and Theorem 4.21 it now follows:

$$\begin{aligned}|\mathbb{E}a_n(p_1,q_1)^2\cdots a_n(p_l,q_l)^2-1| &= \left|\mathbb{E}\left(Y^{(n)}_{\varphi_n(p_1,q_1)}\right)^2\cdots\left(Y^{(n)}_{\varphi_n(p_l,q_l)}\right)^2-1\right|\\ &= \left|\mathbb{E}Y^{(n)}_{i(1)}\cdots Y^{(n)}_{i(2l)}-1\right|\\ &= \left|\sum_{\pi\in\mathcal{PP}(2l)}\prod_{(r,s)\in\pi}\Sigma_{n^2}\big(i(r),i(s)\big)-1\right| \le \frac{\#\mathcal{PP}(2l)}{n^{4\alpha}}.\end{aligned}$$

and thus (AAU2) with $C_n^{(l)}=\frac{\#\mathcal{PP}(2l)}{n^{4\alpha}}$, as just defined above.

Now let us turn to condition (AAU3), then in accordance with Lemma 4.18 we choose a $K\in\mathbb{R}$ and appropriate sequences $(\tilde{D}_n^{(l)})_{n\in\mathbb{N}}$ for all $l\in\mathbb{N}$. We set $K:=3$ and for all $l\in\mathbb{N}$ we set $\tilde{D}_n^{(l)}:=\frac{\#\mathcal{PP}(2l+2)}{n^{4\alpha}}$ for all $n\in\mathbb{N}$.

Now let $N,l\in\mathbb{N}$ be arbitrary and $(p_1,q_1),\dots,(p_l,q_l)$ in $\square_N$ be a sequence of fundamentally different pairs. Then it holds for all $n\ge N$, that $a_n(p_i,q_i)=Y^{(n)}_{\varphi_n(p_i,q_i)}$ for all $i\in\{1,\dots,l\}$. Further, the indices

$$\begin{aligned}&(i(1),i(2),i(3),i(4),\dots,i(2l+1),i(2l+2))\\ &:=(\varphi_n(p_1,q_1),\varphi_n(p_1,q_1),\varphi_n(p_1,q_1),\varphi_n(p_1,q_1),\varphi_n(p_2,q_2),\varphi_n(p_2,q_2),\varphi_n(p_3,q_3),\varphi_n(p_3,q_3),\\ &\qquad\dots,\varphi_n(p_l,q_l),\varphi_n(p_l,q_l))\end{aligned}$$

in $[n^2]$ meet the conditions of Lemma 4.25. With this lemma and Theorem 4.21 it follows:

$$\begin{aligned}&|\mathbb{E}a_n(p_1,q_1)^4a_n(p_2,q_2)^2\cdots a_n(p_l,q_l)^2-3|\\ &= \left|\mathbb{E}\left(Y^{(n)}_{\varphi_n(p_1,q_1)}\right)^4\left(Y^{(n)}_{\varphi_n(p_2,q_2)}\right)^2\cdots\left(Y^{(n)}_{\varphi_n(p_l,q_l)}\right)^2-3\right|\\ &= \left|\mathbb{E}Y^{(n)}_{i(1)}\cdots Y^{(n)}_{i(2l+2)}-3\right|\\ &= \left|\sum_{\pi\in\mathcal{PP}(2l+2)}\prod_{(r,s)\in\pi}\Sigma_{n^2}\big(i(r),i(s)\big)-3\right| \le \frac{\#\mathcal{PP}(2l+2)}{n^{4\alpha}}.\end{aligned}$$

and thus (AAU3) with $D_n^{(l)}=\frac{\#\mathcal{PP}(2l+2)+\#\mathcal{PP}(4)}{n^{4\alpha}}$, where we used Lemma 4.18. Note that for all $l\in\mathbb{N}$, the sequence $D_n^{(l)}$ is summable over n iff $\alpha>1/4$.

To show the last statement, just assume that for all $n\in\mathbb{N}$ and $i\neq j\in[n]$, $\Sigma_n(i,j)=1/n^\alpha$. Then surely, for all $n\in\mathbb{N}$, Σ_n will be positive definite (since Σ_n then is a sum of a positive semi-definite matrix and a positive multiple of the identity matrix). Further for all $n\ge 2$ we obtain $|\mathbb{E}a_n(1,1)a_n(1,2)|=\mathbb{E}a_n(1,1)a_n(1,2)=1/n^\alpha$. In particular, if $\alpha'>\alpha$ is arbitrary, we will not find a constant C such that $|\mathbb{E}a_n(1,1)a_n(1,2)|\le C/n^{\alpha'}$ for all $n\ge 2$, so (AAU1) cannot be satisfied with respect to α', hence $(a_n)_n$ is not α'-almost uncorrelated. $\square$

Remark 4.27. Example 4.22 features entries which are not necessarily exchangeable. Therefore, the example shows that there are relevant models of non-exchangeable ensembles that fit the model of "α-almost uncorrelated random matrices." In addition, Theorem 4.26 shows nicely that α-almost uncorrelated ensembles appear naturally in correlated Gaussian ensembles where all correlations between different variables are absolutely bounded by $1/n^{\alpha}$.

A part of the results of the following corollary have already been established in the author's work [31], where full ensembles with specific correlation structures were considered and only weak convergence in probability was derived. The results are now strengthened to be valid (*i*) almost surely, (*ii*) for band matrices, and (*iii*) with the strongly relaxed condition that correlations in non-normalized random matrices lie in the interval $[-1/n^{2\alpha}, 1/n^{2\alpha}]$ (when the matrix is of dimension $n \times n$), without imposing an additional structure. To the best knowledge and understanding of the author, these are novel results when compared with previous work on Gaussian entries, such as [46], [8] and [16], where additional correlation structures are assumed and/or weak convergence to the semicircle distribution is not discussed.

Corollary 4.28. *Let $\alpha > 0$ and $(\Sigma_n)_n \in \mathrm{CovMat}(\alpha)$ be arbitrary, and let $(a_n)_n$ be the triangular scheme filled with correlated Gaussian entries as developed in Example 4.22, which is strongly α-almost uncorrelated by Theorem 4.26. Let $(X_n)_n$ be the random matrices which are based on $(a_n)_n$. Let $b = (b_n)_n$ be a sequence of bandwidths with $b_n \to \infty$ and $(Y_n)_n$ be the periodic random band matrices which are based on $(a_n)_n$ with bandwidth b. Then the following statements hold:*

i) If $\alpha \geq 1/2$, then the semicircle law holds in probability for $(X_n)_n$.

ii) If $\alpha > 1/2$, then the semicircle law holds almost surely for $(X_n)_n$.

iii) If $\alpha \geq 1/2$, then the semicircle law holds in probability for $(Y_n)_n$.

iv) If $\alpha > 1/2$ and $(\frac{1}{b_n^2})_n$ is summable, then the semicircle law holds almost surely for $(Y_n)_n$.

Proof. By Theorem 4.26, $(a_n)_{n\in\mathbb{N}}$ is strongly α-almost uncorrelated, and if $\alpha > 1/4$, then for all $l \in \mathbb{N}$, the sequences $C_n^{(l)}$ and $D_n^{(l)}$ can be chosen summable over n. Then all statements follow directly from 1. and 3. in Theorem 4.9, where for *i*) and *ii*), remember that full matrices are periodic band matrices with full bandwidth $\tilde{b}_n := n$. $\square$

4.3 Proof of the Main Theorem

To prove the main theorem, Theorem 4.9, we will use Theorem 3.14 with $\mu = \sigma$ and $z = 2$. Note that due to condition (AAU1), any α-almost uncorrelated triangular scheme will have absolute moments of all orders, and this remains in particular true for (band) matrices which are based on the triangular scheme.

In order to derive weak convergence in probability and almost surely, what we need to show now is:

1. The random moments converge in expectation, that is, for each $k \in \mathbb{N}$ we have
$$\mathbb{E}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \sigma, x^k \right\rangle . \tag{4.5}$$
2. The variance of the random moments decays to zero, that is, for every $k \in \mathbb{N}$ we have
$$\mathbb{V}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} 0. \tag{4.6}$$

Then (4.5) will yield weak convergence in expectation, and if in addition we show (4.6), we will have weak convergence in probability. If further, the convergence in (4.6) is summably fast, then we obtain weak convergence almost surely.

We remind the reader that if $b = (b_n)_n$ is a sequence of n-bandwidths, $(a_n)_n$ is a triangular scheme and $(X_n)_n$ the periodic random band matrices which are based on $(a_n)_n$ with bandwidth b, then we have (by Corollary 3.11) for the k-th moment of σ_n that

$$\left\langle \sigma_n, x^k \right\rangle = \frac{1}{n b_n^{k/2}} \sum_{t_1,\ldots,t_k=1}^{n} a_n^b(t_1,t_2) a_n^b(t_2,t_3) \cdots a_n^b(t_k,t_1). \tag{4.7}$$

4.3.1 Development of Combinatorics for the Method of Moments

As we observe above in (4.7), the random moment of an ESD is actually a rather elaborate sum of random variables, which means it is crucial to sort the summands in a way that makes the sum amenable for analysis. For this sorting (which amounts to subdividing indices into certain equivalence classes) we need the language of graph theory and some combinatorics, which we will introduce next. The following "definition" will incorporate all the graph theoretical notions we will need for our endeavor, where we followed the expositions as in [58] and [63], as we did in our previous work [31].

Definition 4.29. Let M be a finite set, $k \in \mathbb{N}_0$ be arbitrary, then we denote by $\binom{M}{k}$ the set of all k-element subsets of M. A *graph* G is a triple $G = (V, E, \phi)$, where the following holds:

i) V is a finite set, whose elements are called *vertices*, or *nodes*.

ii) E is a finite set, whose elements are called *edges*.

iii) $\phi : E \longrightarrow \binom{V}{1} \cup \binom{V}{2}$ is a function, which is called *incidence function*.

Given arbitrary elements $e \in E$ and $u, v \in V$, such that $\phi(e) = \{u, v\}$, then it is the underlying view that the edge e connects the vertices u and v. In this situation, if $u = v$, then e is called *loop*. If $u \neq v$, then e is called *proper edge*. Two nodes $u, v \in V$ are called *adjacent*, if they are connected by an edge, that is, if there is an $e \in E$ such that

$\phi(e) = \{u, v\}$. A vertex $v \in V$ and an edge $e \in E$ are called *incident*, if $v \in \phi(e)$, that is, if e is connected to v. Two different edges $e \neq f \in E$ are called *parallel* if they connect the same nodes, so if $\phi(e) = \phi(f)$. If there are edges $e_1, \ldots, e_k \in E$ which are all parallel to one another, but not parallel to any other edge in E, then we call each of the e_i a *k-fold edge*. For $k = 2$ we use the term *double edge*. If an edge e does not have a parallel edge, e is called a *single edge*. An edge is called *even*, if it is a k-fold edge with k even, and *odd*, if it is a k-fold edge with k odd. A *path* is a finite sequence of the form

$$v_1, e_1, v_2, e_2, v_3, e_3, \ldots, v_k, e_k, v_{k+1}$$

for a $k \in \mathbb{N}$, vertices $v_1, \ldots, v_{k+1} \in V$ and edges $e_1, \ldots, e_k \in E$, so that each two neighboring vertices are connected by the edge in between, so $\phi(e_i) = \{v_i, v_{i+1}\}$ for all $i = 1, \ldots, k$. If we also have $v_1 = v_{k+1}$, then we call the path a *cycle*. A *Eulerian cycle* in a graph $G = (V, E, \phi)$ is a cycle which traverses each edge in E exactly once. A graph in which a Eulerian cycle can be constructed is called *Eulerian graph*.

To utilize the method of moments to show the semicircle law, we have to show (4.5) and (4.6). To do this, we will analyze the sum in (4.7). We will use the language of graph theory to aid us. Recalling (4.7), we write for an arbitrary $k \in \mathbb{N}$:

$$\left\langle \sigma_n, x^k \right\rangle = \frac{1}{n b_n^{k/2}} \sum_{t_1, \ldots, t_k = 1}^{n} a_n^b(t_1, t_2) a_n^b(t_2, t_3) \cdots a_n^b(t_k, t_1) = \frac{1}{n b_n^{k/2}} \sum_{\underline{t} \in [n]^k} a_n^b(\underline{t}), \tag{4.8}$$

where $[n] := \{1, \ldots, n\}$ and for $\underline{t} \in [n]^k$ with $\underline{t} = (t_1, \ldots, t_k)$:

$$a_n^b(\underline{t}) := a_n^b(t_1, t_2) a_n^b(t_2, t_3) \cdots a_n^b(t_k, t_1).$$

Now in equation (4.8), we observe that many summands vanish due to the band structure of the matrix. To account for this, we make the following definition:

Definition 4.30. Let $b = (b_n)_n$ be a sequence of n-bandwidths. Then for fixed n and k in $\mathbb{N}$, we call a tuple $\underline{t} \in [n]^k$ b_n-relevant if each pair (t_i, t_{i+1}) for $i = 1, \ldots, k$ (where $k + 1 \equiv 1$) is b_n-relevant (cf. Definition 4.6). We further define

$$[n]_b^k := \left\{ \underline{t} \in [n]^k : \underline{t} \text{ is } b_n\text{-relevant} \right\}.$$

Then we obtain

$$\left\langle \sigma_n, x^k \right\rangle = \frac{1}{n b_n^{k/2}} \sum_{\underline{t} \in [n]_b^k} a_n^b(\underline{t}). \tag{4.9}$$

In the following, we want to sort the sum in (4.9) by classifying its index set. Generally, for a given $\underline{t} \in [n]^k$, $\underline{t} = (t_1, \ldots, t_k)$, we define the graph $G_{\underline{t}} = (V_{\underline{t}}, E_{\underline{t}}, \phi_{\underline{t}})$ with vertices $V_{\underline{t}} = \{t_1, \ldots, t_k\}$ and (abstract) edges $E_{\underline{t}} = \{e_1, \ldots, e_k\}$, where $\phi_{\underline{t}}(e_i) = \{t_i, t_{i+1}\}$ for all $i = 1, \ldots, k$ and with the convention that $k + 1 \equiv 1$. Then, through

$$t_1, e_1, t_2, e_2, \ldots t_{k-1}, e_{k-1}, t_k, e_k, t_1 \tag{4.10}$$

we obtain a Eulerian cycle which passes through the graph $G_{\underline{t}}$. We notice that each $\underline{t} \in [n]^k$ spans its own graph $G_{\underline{t}}$ and also constitutes a Eulerian cycle through that graph. Therefore, we do not just perceive $\underline{t}$ as a tuple, but also as a graph and a Eulerian cycle. When we say that $\underline{t}$ has 7 vertices and 2 different l-fold edges, or that $\underline{t}$ traverses only double edges, then those are statements concerning the graph $G_{\underline{t}}$ and the cycle through $G_{\underline{t}}$ which was induced by $\underline{t}$.

For a $\underline{t} \in [n]^k$ we will define its *profile* $\kappa(\underline{t})$ as the k-vector

$$\kappa(\underline{t}) = (\kappa_1(\underline{t}), \dots, \kappa_k(\underline{t})),$$

where for every $l \in \{1, \dots, k\}$ we define

$$\kappa_l(\underline{t}) := \#\{\text{different } l\text{-fold edges in } \underline{t}\} = \#\{\phi_{\underline{t}}(e) \,|\, e \in E_{\underline{t}} \text{ is an } l\text{-fold edge}\}.$$

To make this concept clear, the tuple $\underline{t} = (1, 1, 2, 3, 2, 6, 7, 6, 2, 6, 2)$ has only one single edge $\{1\}$ (a loop in this case), three different double edges, $\{1, 2\}$, $\{2, 3\}$ and $\{6, 7\}$, and one 4-fold edge, $\{2, 6\}$. Therefore, $\kappa(\underline{t}) = (1, 3, 0, 1, 0, 0, 0, 0, 0, 0, 0)$.

For each $\underline{t} \in [n]^k$ we immediately observe the equality

$$k = \sum_{l=1}^{k} l \cdot \kappa_l(\underline{t}).$$

Now there are certain inequalities concerning the profile of a $\underline{t} \in [n]_b^k$ which we will heavily draw upon. Denote by $\ell(\underline{t})$ the number of different loops in $\underline{t}$, so

$$\ell(\underline{t}) = \#\{\phi_{\underline{t}}(e) \,|\, e \text{ a loop in } E_{\underline{t}}\}.$$

For example, for the tuple $\underline{t} = (1, 1, 2, 2, 3, 3, 3, 3)$ we will have $\ell(\underline{t}) = 3$. The following lemma will provide upper bounds for the number of vertices of a tuple $\underline{t} \in [n]^k$ depending on the tuple's profile.

Lemma 4.31. *Let $n, k \in \mathbb{N}$ and $\underline{t} \in [n]^k$ be arbitrary, then*

i) $\#V_{\underline{t}} \leq 1 + \kappa_1(\underline{t}) + \ldots + \kappa_k(\underline{t}) - \ell(\underline{t})$.

ii) If $\underline{t}$ contains at least one odd edge, then

$$\#V_{\underline{t}} \leq \kappa_1(\underline{t}) + \ldots + \kappa_k(\underline{t}).$$

Remark 4.32. Oftentimes we only need a weaker version of statement i) in above lemma, which is that

$$\#V_{\underline{t}} \leq 1 + \kappa_1(\underline{t}) + \ldots + \kappa_k(\underline{t}).$$

Proof of Lemma 4.31. Let $\underline{t} \in [n]^k$ be arbitrary. The idea behind this proof is that we travel the cycle

$$t_1, e_1, t_2, e_2, \ldots, t_k, e_k, t_1 \tag{4.11}$$

by picking an initial node t_i and then traversing the edges in (increasing) cyclic order until reaching t_i again. On the way, we count the number of different nodes that were discovered. When traversing the path, only proper edges may discover a new vertex, whereas loops will never discover a new vertex. Further, if we pass a k-fold edge, only the first instance of that edge may discover a new vertex.

i) We write $\ell(\underline{t}) = \ell_1(\underline{t}) + \ldots + \ell_k(\underline{t})$, where $\ell_m(\underline{t})$ denotes the number of different m-fold loops in $\underline{t}$. When we start our tour along the circle $\underline{t}$ at the initial vertex t_1, we first observe this very vertex. Then, as we travel along the circle, for each $m \in \{1, \ldots, k\}$ we will pass $m \cdot \kappa_m(\underline{t})$ m-fold edges out of which only the first instance of *proper* m-fold edges can discover a new node, and there are $\kappa_m(\underline{t}) - \ell_m(\underline{t})$ of these first instances. Considering the initial node, we arrive at $\#V_{\underline{t}} \leq 1 + \kappa_1(\underline{t}) - \ell_1(\underline{t}) + \ldots + \kappa_k(\underline{t}) - \ell_k(\underline{t})$, which yields the desired inequality.

ii) The idea is that in presence of an odd edge, we can start the tour at a specific vertex such that the odd edge cannot contribute to the newly discovered vertices. To this end, fix an m-fold edge in $\underline{t}$ with m odd. Let $e_{i_1}, \ldots, e_{i_m}$, $i_1 < \ldots < i_m$, be the m-fold edges in question in the cycle (4.11). If $\underline{t}$ traverses each of these m edges in the same direction (this is in particular the case if $m = 1$), that is, $t_{i_1} = \ldots = t_{i_m}$ and $t_{i_1+1} = \ldots = t_{i_m+1}$, then start the tour at the vertex t_{i_1+1} and observe that now, none of the edges $e_{i_1}, \ldots, e_{i_m}$ can discover a new node, since if our m-fold edge is a loop, this is clear, and if it is not a loop, t_{i_2} must be discovered by an edge different from our m-fold edge. So the odd m-fold edge in question cannot discover a new node. If $\underline{t}$ does *not* traverse each of these m edges in the same direction, then, since m is odd and we are on a cycle, there still must be an index $l \in \{1, \ldots, m\}$, such that e_{i_l} and $e_{i_{l+1}}$ *are* traversed in the same direction, where i_{m+1} cycicly becomes i_1. Then, start the tour at the vertex t_{i_l+1} and we have again (with the same reasoning as before) that none of the edges $e_{i_1}, \ldots, e_{i_m}$ of our m-fold edge will discover a new node. In any of the cases, counting the initial node, our roundtrip yields at most $1 + \kappa_1(\underline{t}) + \kappa_2(\underline{t}) + \ldots + \kappa_m(\underline{t}) - 1 + \kappa_{m+1}(\underline{t}) + \ldots + \kappa_k(\underline{t})$ nodes discovered, which proves the desired inequality. □

So, what we have learned so far is an upper bound on the vertices of a $\underline{t} \in [n]^k$, depending on its profile $\kappa(\underline{t})$. The next lemma will answer the question of how many b_n-relevant tuples in $\underline{t} \in [n]_b^k$ with a maximum number of vertices we can obtain:

Lemma 4.33. *Let $b = (b_n)_n$ be a sequence of n-bandwidths. If $k, n \in \mathbb{N}$ are arbitrary and $l \in \{1, \ldots, k\}$, then*

$$\#\{\underline{t} \in [n]_b^k \mid \#V_{\underline{t}} \leq l\} \leq k^k \cdot n b_n^{l-1}$$

Proof. We bound the number of possibilities to construct a $\underline{t} \in \{\underline{t}' \in [n]_b^k \mid \#V_{\underline{t}'} \leq l\}$. Since $\underline{t}$ contains at most l different vertices, we first determine the color structure of the k-tuple $\underline{t}$, that is, we determine which places in the tuple should have equal or different vertex numbers (colors). To do this, we fix a map $f : \{1, \ldots, k\} \longrightarrow \{1, \ldots, l\}$, for which we have at most $l^k \leq k^k$ possibilities. Now, we pick a value for the first node t_1 and have n possibilities. Then if $f(2) = f(1)$, we have no choice for the node t_2, since then $t_2 \stackrel{!}{=} t_1$. But if $f(2) \neq f(1)$, we are left with at most b_n choices for t_2, since the tuple should be b_n-relevant. We proceed this way through the whole tuple $\underline{t}$. Whenever we reach a t_l with $f(l) = f(j)$ for some $j < l$, then we have no choice to make for t_l. Otherwise, we have at most b_n choices. And this happens at most $l - 1$ times after freely picking the initial node t_1. Therefore, we are left with at most $k^k \cdot nb_n^{l-1}$ possibilities, which concludes the lemma. □

Let us now develop some more combinatorics which will be used in later proofs. Fix a sequence $b = (b_n)_n$ of n-bandwidths. Now, we call two tuples $\underline{s}, \underline{t} \in [n]_b^k$ *equivalent*, if $\kappa(\underline{s}) = \kappa(\underline{t})$. Obviously, this defines an equivalence relation on $[n]_b^k$. For a tuple $\underline{s} \in [n]_b^k$, we define

$$\mathcal{T}(\underline{s}) := \{\underline{t} \in [n]_b^k : \kappa(\underline{t}) = \kappa(\underline{s})\}.$$

as the set of tuples that have the same profile as $\underline{s}$. So $\mathcal{T}(\underline{s})$ is the equivalence class that $\underline{s}$ belongs to. How many equivalence classes do we have in $[n]_b^k$, and given a $\underline{t} \in [n]_b^k$, how many elements does the equivalence class $\mathcal{T}(\underline{t})$ have?

Lemma 4.34. *Let $n, k \in \mathbb{N}$ be fixed.*

i) There are at most $(k+1)^k$ equivalence classes in $[n]_b^k$.

ii) Let $\underline{s} \in [n]_b^k$ be arbitrary, then

a) $\#\mathcal{T}(\underline{s}) \leq k^k \cdot nb_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})}$.

b) If $\underline{s}$ contains at least one odd edge, we have

$$\#\mathcal{T}(\underline{s}) \leq k^k \cdot nb_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})-1}.$$

Proof. i) There are as many different equivalence classes as there are different profiles $(\kappa_1, \ldots, \kappa_k)$ of relevant tuples. Now surely, in a profile $(\kappa_1, \ldots, \kappa_k)$, each entry is an integer in the set $\{0, \ldots, k\}$, so there are at most $(k+1)^k$ profiles which can be assumed by tuples.

ii)a) Let us fix an $\underline{s} \in [n]_b^k$. How many possibilities do we have to construct a $\underline{t} \in \mathcal{T}(\underline{s})$? Surely, each $\underline{t} \in \mathcal{T}(\underline{s})$ has the same profile as $\underline{s}$, thus has at most

$$\min(1 + \kappa_1(\underline{s}) + \ldots + \kappa_k(\underline{s}), k)$$

different vertices by Lemma 4.31 (note that it need not have the same number of loops, so we cannot obtain a generally better upper bound). Thus, Lemma 4.33 yields

$$\#\mathcal{T}(\underline{s}) \leq k^k \cdot nb_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})}$$

which is the desired inequality.

ii)b) We repeat the argument as in ii)a): Let us fix an $\underline{s} \in [n]_b^k$ with at least one odd edge. How many possibilities do we have to construct an $\underline{t} \in \mathcal{T}(\underline{s})$? Surely, each $\underline{t} \in \mathcal{T}(\underline{s})$ has the same profile as $\underline{s}$, thus has at most

$$\kappa_1(\underline{t}) + \ldots + \kappa_k(\underline{t}) \leq k$$

different vertices by Lemma 4.31. Therefore, Lemma 4.33 yields

$$\#\mathcal{T}(\underline{s}) \leq k^k \cdot nb_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})-1}.$$

□

Later on in the proofs, we will also need to deal with maximum numbers of equivalent pairs of tuples. Let us develop their combinatorics at this place.

We will introduce some mildly new notation. For $\underline{s}, \underline{s}' \in [n]_b^k$ we define

$$\mathcal{T}(\underline{s}, \underline{s}') := \{(\underline{t}, \underline{t}') \,|\, \underline{t}, \underline{t}' \in [n]_b^k, \kappa(\underline{t}) = \kappa(\underline{s}), \kappa(\underline{t}') = \kappa(\underline{s}')\}$$

and partition this set into edge disjoint tuple pairs

$$\mathcal{T}^d(\underline{s}, \underline{s}') := \{(\underline{t}, \underline{t}') \,|\, \underline{t}, \underline{t}' \in [n]_b^k, \kappa(\underline{t}) = \kappa(\underline{s}), \kappa(\underline{t}') = \kappa(\underline{s}'), \phi_{\underline{t}}(E_{\underline{t}}) \cap \phi_{\underline{t}'}(E_{\underline{t}'}) = \emptyset\}$$

and into tuples pairs with at least one common edge

$$\mathcal{T}^c(\underline{s}, \underline{s}') := \{(\underline{t}, \underline{t}') \,|\, \underline{t}, \underline{t}' \in [n]_b^k, \kappa(\underline{t}) = \kappa(\underline{s}), \kappa(\underline{t}') = \kappa(\underline{s}'), \phi_{\underline{t}}(E_{\underline{t}}) \cap \phi_{\underline{t}'}(E_{\underline{t}'}) \neq \emptyset\}.$$

We further partition the set $\mathcal{T}^c(\underline{s}, \underline{s}')$ into the subsets of equivalent tuples that have exactly l edges in common. So for each $l \in \{1, \ldots, k\}$ we define

$$\mathcal{T}_l^c(\underline{s}, \underline{s}') := \{(\underline{t}, \underline{t}') \,|\, \underline{t}, \underline{t}' \in [n]_b^k, \kappa(\underline{t}) = \kappa(\underline{s}), \kappa(\underline{t}') = \kappa(\underline{s}'), \#[\phi_{\underline{t}}(E_{\underline{t}}) \cap \phi_{\underline{t}'}(E_{\underline{t}'})] = l\}.$$

We are now interested in bounds for $\#\mathcal{T}^d(\underline{s}, \underline{s}')$, $\#\mathcal{T}^c(\underline{s}, \underline{s}')$ and $\#\mathcal{T}_l^c(\underline{s}, \underline{s}')$. The first quantity can be trivially bounded by

$$\#\mathcal{T}^d(\underline{s}, \underline{s}') \leq \#\mathcal{T}(\underline{s}) \cdot \#\mathcal{T}(\underline{s}'), \tag{4.12}$$

since the number of possibilities to pick equivalent tuples edge disjoint is bounded by the number of unrestricted possibilities. On the other hand, the latter two quantities will require some further attention. First, we realize that if the k-tuples $\underline{t}$ and $\underline{t}'$ in $[n]^k$ have a common edge, the superposition of their graphs will form another Eulerian graph, since vertex degrees remain even (see [61, p. 51]). Therefore, we can find a Eulerian cycle $\underline{u} \in [n]^{2k}$ that travels first through all the edges of $\underline{t}$ (not necessarily in the same order as $\underline{t}$ travels its edges, but in the same cyclic order) and then through all the edges of $\underline{t}'$ (again, not necessarily in the same order as $\underline{t}'$ travels its edges). The next lemma will make our statement precise. Note that by a *cyclic permutation* of a tuple we mean, for example, $(1, 2, 3, 4) \rightsquigarrow (4, 1, 2, 3)$ or $(1, 2, 3, 4) \rightsquigarrow (2, 3, 4, 1)$.

Lemma 4.35. *Let $\underline{t}$ and $\underline{t}'$ in $[n]^k$ have a common edge, then there is a $\underline{u} \in [n]^{2k}$ such that*

i) $(u_1, \dots, u_k)$ *is a cyclic permutation of* $(t_1, \dots, t_k)$ *and* $(u_{k+1}, \dots, u_{2k})$ *is a cyclic permutation of* $(t'_1, \dots, t'_k)$.

ii) $((u_1, u_2), \dots, (u_k, u_{k+1}))$ *is a cyclic permutation of* $((t_1, t_2), \dots, (t_k, t_1))$ *and* $((u_{k+1}, u_{k+2}), \dots, (u_{2k}, u_1))$ *is a cyclic permutation of* $((t'_1, t'_2), \dots, (t'_k, t'_1))$.

In particular, the Eulerian cycle $\underline{u}$ spans the graph obtained through superposition of the graphs of $\underline{t}$ and $\underline{t}'$, and it travels first through all the edges of $\underline{t}$ and then through all the edges of $\underline{t}'$.

Proof. So let $\underline{t}$ and $\underline{t}'$ be as in the statement of the lemma, then they surely have a common node. Therefore, there exist cyclic permutations of $\underline{\tilde{t}}$ of $\underline{t}$ and $\underline{\tilde{t}}'$ of $\underline{t}'$ such that $\tilde{t}_1 = \tilde{t}'_1$. Set $\underline{u} := (\tilde{t}_1, \dots, \tilde{t}_k, \tilde{t}'_1, \dots, \tilde{t}'_k)$. Then the first statement of the lemma is clear, and for the second we write

$$\begin{aligned} ((u_1, u_2), \dots, (u_k, u_{k+1})) &= ((\tilde{t}_1, \tilde{t}_2), \dots, (\tilde{t}_k, \tilde{t}'_1)) \\ &= ((\tilde{t}_1, \tilde{t}_2), \dots, (\tilde{t}_k, \tilde{t}_1)), \end{aligned}$$

and this is a cyclic permutation of $((t_1, t_2), \dots, (t_k, t_1))$, since $\underline{\tilde{t}}$ is a cyclic permutation of $\underline{t}$. Analogously, we write

$$\begin{aligned} ((u_{k+1}, u_{k+2}), \dots, (u_{2k}, u_1)) &= ((\tilde{t}'_1, \tilde{t}'_2), \dots, (\tilde{t}'_k, \tilde{t}_1)) \\ &= ((\tilde{t}'_1, \tilde{t}'_2), \dots, (\tilde{t}'_k, \tilde{t}'_1)), \end{aligned}$$

and this is a cyclic permutation of $((t'_1, t'_2), \dots, (t'_k, t'_1))$, since $\underline{\tilde{t}}'$ is a cyclic permutation of $\underline{t}'$. □

Next, we formulate a lemma which is in the spirit of Lemma 4.34, but for overlapping edges.

Lemma 4.36. *Let $b = (b_n)_n$ be a sequence of n-bandwidths and $n, k \in \mathbb{N}$ fixed. Let $\underline{s}, \underline{s}' \in [n]_b^k$. Then the following statements hold:*

1. *If $\underline{s}$ and $\underline{s}'$ have only even edges, we have for each $(\underline{t}, \underline{t}') \in \mathcal{T}^c(\underline{s}, \underline{s}')$ that*

$$\#(V_{\underline{t}} \cup V_{\underline{t}'}) \leq k.$$

2. *Let at least one of the tuples $\underline{s}$ or $\underline{s}'$ contain at least one odd edge and let $l \in \{1, \dots, k\}$. Then it holds for each $(\underline{t}, \underline{t}') \in \mathcal{T}^c(\underline{s}, \underline{s}')$ so that $\underline{t}$ and $\underline{t}'$ have at least l different common edges, that*

$$\#(V_{\underline{t}} \cup V_{\underline{t}'}) \leq \sum_{i=1}^{k} \kappa_i(\underline{s}) + \sum_{j=1}^{k} \kappa_j(\underline{s}') - l.$$

Proof. To prove the first statement, we need to show that for each $(\underline{t}, \underline{t}') \in \mathcal{T}^c(\underline{s}, \underline{s}')$ we have that - taken together - $\underline{t}$ and $\underline{t}'$ can have at most k different nodes. To argue this, we have by Lemma 4.31 that $\underline{t}$ spans at most $\frac{k}{2}+1$ nodes, and in this case $\underline{t}$ consists of only double edges, all of them *proper*. In particular, since having a proper edge with $\underline{t}$ in common, $\underline{t}'$ can span at most $\frac{k}{2}+1-2$ additional nodes, leading to a total of at most $\frac{k}{2}+1+\frac{k}{2}+1-2 = k$ different nodes. Now on the other hand, if initially, we assume $\underline{t}$ to contain less than $\frac{k}{2}+1$ nodes, thus at most $\frac{k}{2}$ nodes, $\underline{t}$ might have loops, especially $\underline{t}'$ could have a loop in common with $\underline{t}$, thus contributing up to $\frac{k}{2}+1-1$ additional nodes, again leading to a total number of different nodes of at most $\frac{k}{2}+\frac{k}{2}+1-1 = k$.

To prove the second statement, we need to dive a little deeper. To start, we assume w.l.o.g. that $\underline{s}$ contains an odd edge.

Now, since $\underline{t}$ and $\underline{t}'$ have at least one edge in common, we will find a $\underline{u} \in [n]^{2k}$ as in Lemma 4.35. Surely, we have that $\#(V_{\underline{t}} \cup V_{\underline{t}'}) = \#V_{\underline{u}}$. Thus, we will travel through the Eulerian cycle

$$u_1, e_1^u, u_2, e_2^u, \ldots, u_{2k-1}, e_{2k-1}^u, u_{2k}, e_{2k}^u, u_1$$

and observe how many nodes we can discover. By traveling the first k edges of $\underline{u}$, we actually travelled the edges of $\underline{t}$ (by Lemma 4.35) and can thus discover at most

$$\kappa_1(\underline{t}) + \ldots + \kappa_k(\underline{t})$$

nodes by Lemma 4.31. Now, while traveling the last k edges of $\underline{u}$, how many additional nodes can we discover? By Lemma 4.35, we will travel all the edges of $\underline{t}'$. Then at most all the single edges and first instances of m-fold edges with $m \geq 2$ may discover a new node, but only if that edge has not been travelled before during the crossing of the first k edges. Since we have at least l common edges, while crossing the last k edges we can see at most

$$\kappa_1(\underline{t}') + \ldots + \kappa_k(\underline{t}') - l$$

new nodes. Thus, in total we can observe at most

$$\kappa_1(\underline{t}) + \ldots + \kappa_k(\underline{t}) + \kappa_1(\underline{t}') + \ldots + \kappa_k(\underline{t}') - l$$

nodes while traveling the cycle $\underline{u}$. □

Lemma 4.37. *Let $b = (b_n)_n$ be a sequence of n-bandwidths. Fix $n, k \in \mathbb{N}$ and let $\underline{s}$ and $\underline{s}'$ in $[n]_b^k$ be arbitrary.*

a) If both $\underline{s}$ and $\underline{s}'$ contain only even edges, then $\#\mathcal{T}^c(\underline{s}, \underline{s}') \leq k^2 \cdot (2k)^{2k} \cdot n b_n^{k-1}$.

b) If $\underline{s}$ or $\underline{s}'$ contains at least one odd edge, we have

$$\#\mathcal{T}^c(\underline{s}, \underline{s}') \leq k^2 \cdot (2k)^{2k} \cdot n b_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})+\kappa_1(\underline{s}')+\ldots+\kappa_k(\underline{s}')-2}.$$

c) If $\underline{s}$ or $\underline{s}'$ contains at least one odd edge, we have for all $l \in \{1, \ldots, k\}$, that

$$\#\mathcal{T}_l^c(\underline{s}, \underline{s}') \leq k^2 \cdot (2k)^{2k} \cdot n b_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})+\kappa_1(\underline{s}')+\ldots+\kappa_k(\underline{s}')-l-1}.$$

Proof. a) Let us pick $\underline{s}$ and $\underline{s}'$ in $[n]_b^k$ with only even edges. How many possibilities do we have to construct a tuple pair $(\underline{t}, \underline{t}') \in \mathcal{T}^c(\underline{s}, \underline{s}')$? By Lemma 4.35, $\underline{t}$ and $\underline{t}'$ are cyclic permutations of the first and second half of a tuple $\underline{u} \in [n]^{2k}$, which serves as a Eulerian cycle through the superposition of the graphs of $\underline{t}$ and $\underline{t}'$. Further, by Lemma 4.36, $\underline{u}$ has at most k different nodes. Therefore, the number of possible $(\underline{t}, \underline{t}') \in \mathcal{T}^c(\underline{s}, \underline{s}')$ is surely bounded by the number of all $\underline{u} \in [n]^{2k}$ with at most k different nodes multiplied by the number of possibilities to cyclicly permute the first and second half of those tuples. Since the latter admits at most k^2 choices, we obtain by Lemma 4.33:

$$\#\mathcal{T}^c(\underline{s}, \underline{s}') \le k^2 \cdot (2k)^{2k} \cdot nb_n^{k-1}.$$

b) We can imitate the proof of part a) almost word by word: Let us pick $\underline{s}$ and $\underline{s}'$ in $[n]_b^k$, so that at least one of the two tuples contains an odd edge. How many possibilities do we have to construct a tuple pair $(\underline{t}, \underline{t}') \in \mathcal{T}^c(\underline{s}, \underline{s}')$? By Lemma 4.35, $\underline{t}$ and $\underline{t}'$ are cyclic permutations of the first and second half of a tuple $\underline{u} \in [n]^{2k}$, which serves as a Eulerian cycle through the superposition of the graphs of $\underline{t}$ and $\underline{t}'$. Further, since we have at least one odd edge, we obtain by Lemma 4.36 that $\underline{u}$ has at most $\sum_j \kappa_j(\underline{s}) + \sum_j \kappa_j(\underline{s}') - 1$ different nodes. Therefore, the number of possible $(\underline{t}, \underline{t}') \in \mathcal{T}^c(\underline{s}, \underline{s}')$ is surely bounded by the number of all $\underline{u} \in [n]^{2k}$ with at most $\sum_j \kappa_j(\underline{s}) + \sum_j \kappa_j(\underline{s}') - 1$ different nodes multiplied by the number of possibilities to cyclicly permute the first and second half of those tuples. Since the latter admits at most k^2 choices, we obtain by Lemma 4.33:

$$\#\mathcal{T}^c(\underline{s}, \underline{s}') \le k^2 \cdot (2k)^{2k} \cdot nb_n^{\sum_i \kappa_i(\underline{s}) + \sum_j \kappa_j(\underline{s}') - 2}.$$

c) We proceed as in the previous parts of this proof: Let us pick $\underline{s}$ and $\underline{s}'$ in $[n]_b^k$, so that at least one of the two tuples contains an odd edge. How many possibilities do we have to construct a tuple pair $(\underline{t}, \underline{t}') \in \mathcal{T}_l^c(\underline{s}, \underline{s}')$? By Lemma 4.35, $\underline{t}$ and $\underline{t}'$ are cyclic permutations of the first and second half of a tuple $\underline{u} \in [n]^{2k}$ which serves as a Eulerian cycle through the superposition of the graphs of $\underline{t}$ and $\underline{t}'$. Further, by Lemma 4.36, $\underline{u}$ has at most $\sum_j \kappa_j(\underline{s}) + \sum_j \kappa_j(\underline{s}') - l$ different nodes. Therefore, the number of possible $(\underline{t}, \underline{t}') \in \mathcal{T}_l^c(\underline{s}, \underline{s}')$ is surely bounded by the number of all $\underline{u} \in [n]^{2k}$ with at most $\sum_j \kappa_j(\underline{s}) + \sum_j \kappa_j(\underline{s}') - l$ different nodes multiplied by the number of possibilities to cyclicly permute the first and second half of those tuples. Since the latter admits at most k^2 choices, we obtain by Lemma 4.33:

$$\#\mathcal{T}^c(\underline{s}, \underline{s}') \le k^2 \cdot (2k)^{2k} \cdot nb_n^{\sum_i \kappa_i(\underline{s}) + \sum_j \kappa_j(\underline{s}') - l - 1}.$$

□

The last lemma concludes our combinatorial endeavors.

4.3.2 Convergence of Expected Moments

So now, let us show (4.5) on page 71, convergence in expectation.

Theorem 4.38. *Let $(X_n)_n$ be a sequence of periodic random band matrices which is based on an α-almost uncorrelated triangular array $(a_n)_n$ with $\alpha \geq \frac{1}{2}$ and bandwidth $b = (b_n)_n$. Denote by $(\sigma_n)_n$ the ESDs of $(X_n)_n$. Then if $b_n \to \infty$, we have for all $k \in \mathbb{N}$ that*

$$\mathbb{E}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} \left\langle \sigma, x^k \right\rangle .$$

Proof. By Lemma 3.8, if k is even, then the k-th moment of the semicircle distribution σ is given by the Catalan number $\mathcal{C}_{\frac{k}{2}}$. The odd moments of σ vanish.
Step 1: Let $k \in \mathbb{N}$ be even.
We need to show that

$$\mathbb{E}\left\langle \sigma_n, x^k \right\rangle = \frac{1}{nb_n^{k/2}} \sum_{\underline{t}\in[n]_b^k} \mathbb{E}a_n^b(\underline{t}) \xrightarrow[n\to\infty]{!} \mathcal{C}_{\frac{k}{2}}.$$

Case 1:
At first we consider an $\underline{s} \in [n]_b^k$ which consists of only double edges. This means it has the profile $(0, k/2, 0, \ldots, 0)$. We partition the set $\mathcal{T}(\underline{s})$ into the sets

$$\mathcal{T}_{\frac{k}{2}+1}(\underline{s}) := \left\{\underline{t} \in \mathcal{T}(\underline{s}) : \#V_{\underline{t}} = \frac{k}{2} + 1\right\}$$

and

$$\mathcal{T}_{\leq\frac{k}{2}}(\underline{s}) := \left\{\underline{t} \in \mathcal{T}(\underline{s}) : \#V_{\underline{t}} \leq \frac{k}{2}\right\}.$$

To count the possibilities to construct a $\underline{t} \in \mathcal{T}_{\frac{k}{2}+1}(\underline{s})$, we first pick an appropriate surjective coloring $f : \{1, \ldots, k\} \twoheadrightarrow \{1, \ldots, k/2 + 1\}$ in *standard form*. This means $f(1) = 1$ and for $l > 1$ we have that if $f(l) \neq f(j)$ for all $j < l$, then $f(l) = \max\{f(j) : j < l\} + 1$. Intuitively, a standard coloring always uses the lowest color number possible. Now the possible standard colorings for k-tuples with only proper double edges and $k/2+1$ different vertices are in bijective correspondence to Dyck paths of length k, and there are exactly $\mathcal{C}_{\frac{k}{2}}$ of them. For example, the tuple $(8, 5, 6, 9, 6, 2, 6, 5)$ has the standard coloring scheme $(1, 2, 3, 4, 3, 5, 3, 2)$, which is associated with the difference sequence of the Dyck path $(1, 1, 1, -1, 1, -1, -1, -1)$. For a formal proof of this we refer the reader to [6, p. 15]: There, note that given a coloring f as above, we obtain the associated Wigner word representative $(f(1), f(2), \ldots, f(k), f(1))$ (and vice versa) as in the proof of their Lemma 2.1.6.

Now given such a standard coloring f, to construct a b_n-relevant tuple $(t_1, \ldots, t_k) \in [n]^k$ matching this coloring, we first have n choices for t_1, then for subsequent choices of t_{i+1}, $i = 1, \ldots, k-1$, whenever $f(i + 1) \in \{f(1), \ldots, f(i)\}$ we have no choice for t_{i+1} because then $t_{i+1} \stackrel{!}{=} t_j$ where $j \in \{1, \ldots, i\}$ with $f(j) = f(i + 1)$. On the other hand, if $f(i + 1) \notin \{f(1), \ldots, f(i)\}$, then t_{i+1} must be different from $t_1, \ldots, t_i$. At the same time, (t_i, t_{i+1}) needs to be b_n-relevant. Accounting for both, we sill have at least $b_n - \#\{f(1), \dots, f(i)\} \geq b_n - i \geq 1$ possibilities for n large enough (so that $b_n \geq k$), but at most b_n possibilities. Therefore, given a standard coloring f, to construct a b_n-relevant tuple $\underline{t} \in [n]^k$, we have at least

$$n \cdot (b_n - 1) \cdots (b_n - \frac{k}{2}) = \frac{n(b_n - 1)!}{(b_n - \frac{k}{2} - 1)!}$$

possibilities (for all n large enough), but at most $n \cdot b_n^{\frac{k}{2}}$ possibilities. Since we had $\mathcal{C}_{\frac{k}{2}}$ choices for the standard coloring, we obtain

$$\mathcal{C}_{\frac{k}{2}} \frac{n(b_n - 1)!}{(b_n - \frac{k}{2} - 1)!} \leq \#\mathcal{T}_{\frac{k}{2}+1}(\underline{s}) \leq \mathcal{C}_{\frac{k}{2}} \cdot n \cdot b_n^{\frac{k}{2}}. \tag{4.13}$$

Further, we have

$$\#\mathcal{T}_{\leq \frac{k}{2}}(\underline{s}) \leq k^k n b_n^{\frac{k}{2}-1},$$

which follows from Lemma 4.33.

In addition, it holds in conjunction with the second moment property of our almost uncorrelated scheme, that for all $\underline{t} \in \mathcal{T}(\underline{s})$ we find

$$|\mathbb{E}a_n^b(\underline{t}) - 1| \leq C_n^{(k/2)},$$

where $C_n^{(k/2)}$ converges to 0 as $n \to \infty$.

Therefore, we obtain

$$\begin{aligned}
\left| \frac{1}{nb_n^{k/2}} \sum_{\underline{t} \in \mathcal{T}(\underline{s})} (\mathbb{E}a_n^b(\underline{t}) - 1) \right| &\leq \frac{1}{nb_n^{k/2}} \sum_{\underline{t} \in \mathcal{T}(\underline{s})} |\mathbb{E}a_n^b(\underline{t}) - 1| \\
&= \frac{1}{nb_n^{k/2}} \sum_{\underline{t} \in \mathcal{T}_{\frac{k}{2}+1}(\underline{s})} |\mathbb{E}a_n^b(\underline{t}) - 1| + \frac{1}{nb_n^{k/2}} \sum_{\underline{t} \in \mathcal{T}_{\leq \frac{k}{2}}(\underline{s})} |\mathbb{E}a_n^b(\underline{t}) - 1| \\
&\leq \frac{1}{nb_n^{k/2}} \#\mathcal{T}_{\frac{k}{2}+1}(\underline{s}) \cdot C_n^{(k/2)} + \frac{1}{nb_n^{k/2}} \#\mathcal{T}_{\leq \frac{k}{2}}(\underline{s}) \cdot C_n^{(k/2)} \\
&\leq \frac{nb_n^{k/2}}{nb_n^{k/2}} \cdot \mathcal{C}_{\frac{k}{2}} \cdot \underbrace{C_n^{(k/2)}}_{\xrightarrow[n]{} 0} + \frac{k^k n b_n^{k/2-1}}{nb_n^{k/2}} \cdot C_n^{(k/2)} \xrightarrow[n \to \infty]{} 0
\end{aligned}$$

and since with (4.13) we have

$$\frac{1}{nb_n^{k/2}} \mathcal{C}_{\frac{k}{2}} \frac{n(b_n - 1)!}{(b_n - \frac{k}{2} - 1)!} \leq \frac{1}{nb_n^{k/2}} \#\mathcal{T}_{\frac{k}{2}+1}(\underline{s}) \leq \frac{1}{nb_n^{k/2}} \mathcal{C}_{\frac{k}{2}} \cdot n \cdot b_n^{\frac{k}{2}},$$

thus

$$\frac{1}{nb_n^{k/2}} \#\mathcal{T}_{\frac{k}{2}+1}(\underline{s}) \xrightarrow[n \to \infty]{} \mathcal{C}_{\frac{k}{2}},$$

it follows

$$\frac{1}{nb_n^{k/2}}\#\mathcal{T}(\underline{s}) = \frac{1}{nb_n^{k/2}}\#\mathcal{T}_{\frac{k}{2}+1}(\underline{s}) + \underbrace{\frac{1}{nb_n^{k/2}}\#\mathcal{T}_{\leq\frac{k}{2}}(\underline{s})}_{\to 0 \text{ as } n\to\infty} \xrightarrow[n\to\infty]{} \mathcal{C}_{\frac{k}{2}}.$$

Therefore,

$$\frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}\mathbb{E}a_n^b(\underline{t}) = \frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}(\mathbb{E}a_n^b(\underline{t})-1) + \frac{1}{nb_n^{k/2}}\#\mathcal{T}(\underline{s}) \xrightarrow[n\to\infty]{} \mathcal{C}_{\frac{k}{2}}.$$

<ins>Case 2:</ins>
Let $\underline{s}\in[n]^k$ have only even edges, but at least one m-fold edge with $m\geq 4$ even. Then it holds by Lemma 4.34, that

$$\#\mathcal{T}(\underline{s}) \leq k^k\cdot nb_n^{\kappa_2(\underline{s})+\kappa_4(\underline{s})\ldots+\kappa_k(\underline{s})}.$$

The exponent is maximized when $\underline{s}$ has one 4-fold edge and just double edges otherwise, and then we obtain the exponent

$$\kappa_2(\underline{s})+\kappa_4(\underline{s}) = \frac{k-4}{2}+1 = \frac{k}{2}-1.$$

Therefore, we surely have

$$\#\mathcal{T}(\underline{s}) \leq k^k\cdot nb_n^{\frac{k}{2}-1}.$$

Further, due to the boundedness property in (AAU1) we have for each $\underline{t}\in\mathcal{T}(\underline{s})$ that (note that $\kappa(\underline{t})=\kappa(\underline{s})$)

$$|\mathbb{E}a_n^b(\underline{t})| \leq C_{\kappa(\underline{s})}.$$

Therefore, we obtain

$$\begin{aligned}\left|\frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}\mathbb{E}a_n^b(\underline{t})\right| &\leq \frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}|\mathbb{E}a_n^b(\underline{t})| \\ &\leq \frac{1}{nb_n^{k/2}}\#\mathcal{T}(\underline{s})\cdot C_{\kappa(\underline{s})} \\ &\leq \frac{1}{nb_n^{k/2}}\cdot k^k\cdot nb_n^{\frac{k}{2}-1}\cdot C_{\kappa(\underline{s})} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

<ins>Case 3:</ins> Let $\underline{s}\in[n]_b^k$, so that $\underline{s}$ contains an odd edge. Since k is even, $\underline{s}$ must contain a second odd edge. Further, it holds by Lemma 4.34, that

$$\#\mathcal{T}(\underline{s}) \leq k^k\cdot nb_n^{\kappa_1(\underline{s})+\kappa_2(\underline{s})\ldots+\kappa_k(\underline{s})-1}.$$

Due to the distinct decay property in (AAU1), we have for all $\underline{t}\in\mathcal{T}(\underline{s})$, that

$$|\mathbb{E}a_n^b(\underline{t})| \leq \frac{C_{\kappa(\underline{s})}}{n^{\alpha\cdot\kappa_1(\underline{s})}} \leq \frac{C_{\kappa(\underline{s})}}{b_n^{\frac{1}{2}\cdot\kappa_1(\underline{s})}}.$$

This yields

$$\left|\frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}\mathbb{E}a_n^b(\underline{t})\right| \leq \frac{1}{nb_n^{k/2}}\cdot\#\mathcal{T}(\underline{s})\cdot\frac{C_{\kappa(\underline{s})}}{b_n^{\frac{1}{2}\cdot\kappa_1(\underline{s})}}$$
$$\leq \frac{C_{\kappa(\underline{s})}}{nb_n^{k/2}}\cdot k^k\cdot nb_n^{\frac{1}{2}\cdot\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{s})-1}.$$

The last exponent is maximized, for example, when the two odd edges of $\underline{s}$ are simple and all other edges of $\underline{s}$ are simple as well, yielding the bound

$$\frac{1}{2}\cdot\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{t})-1\leq\frac{k}{2}-1,$$

leading to

$$\left|\frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}\mathbb{E}a_n^b(\underline{t})\right| \leq \frac{C_{\kappa(\underline{s})}}{b_n^{k/2}}\cdot k^k\cdot b_n^{k/2-1}\xrightarrow[n\to\infty]{}0.$$

Conclusion of Step 1: We observe that for even k, the sum over all tuples in only one equivalence class does not vanish at infinity, namely the class of tuples which contains only double edges. This sum converges to the desired quantity $\mathcal{C}_{\frac{k}{2}}$. All sums over each of the finitely many other classes (Lemma 4.34 i)) converge to zero, thus proving the theorem for k even.

Step 2: Let $k\in\mathbb{N}$ be odd.
Then we know that for each $\underline{s}\in[n]_b^k$ we must have at least one edge which is odd. Also, due to the distinct decay property (AAU1) of our ensemble, we have for every $\underline{t}\in\mathcal{T}(\underline{s})$ (thus $\kappa(\underline{t})=\kappa(\underline{s})$), that

$$|\mathbb{E}a_n^b(\underline{t})|\leq\frac{C_{\kappa(\underline{s})}}{n^{\alpha\cdot\kappa_1(\underline{s})}}\leq\frac{C_{\kappa(\underline{s})}}{b_n^{\alpha\cdot\kappa_1(\underline{s})}}.$$

Therefore, we obtain with Lemma 4.34, that

$$\left|\frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}\mathbb{E}a_n^b(\underline{t})\right| \leq \frac{1}{nb_n^{k/2}}\cdot\#\mathcal{T}(\underline{s})\cdot\frac{C_{\kappa(\underline{s})}}{b_n^{\alpha\cdot\kappa_1(\underline{s})}}$$
$$\leq \frac{C_{\kappa(\underline{s})}}{nb_n^{k/2}}\cdot k^k\cdot nb_n^{(1-\alpha)\cdot\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{s})-1}$$

Since $\alpha\geq\frac{1}{2}$, the last exponent is maximal if $\underline{s}$ consists of only one single edge and double edges otherwise, giving an exponent of $(1-\alpha)+\frac{k-1}{2}-1\leq\frac{k}{2}-1$. Therefore,

$$\left|\frac{1}{nb_n^{k/2}}\sum_{\underline{t}\in\mathcal{T}(\underline{s})}\mathbb{E}a_n^b(\underline{t})\right| \leq \frac{C_{\kappa(\underline{s})}}{b_n^{k/2}}\cdot k^k\cdot b_n^{k/2-1}\xrightarrow[n\to\infty]{}0.$$

Conclusion of Step 2: We have shown that for odd k, the sum over all the tuples in any equivalence class converges to zero. Since there are only finitely many equivalence classes (Lemma 4.34 i)), the entire sum converges to zero for odd k, completing the proof. □

Remark 4.39. In Step 2 of the last proof it becomes apparent that $\alpha \geq \frac{1}{2}$ is a necessity for our proof to work for all bandwidths. Suppose $\alpha < \frac{1}{2}$. Then the last exponent in question is maximal iff $\underline{s}$ consists of single edges only and becomes $(1-\alpha)\cdot k - 1$. To ensure convergence to zero, we must have

$$\begin{aligned}(1-\alpha)\cdot k - 1 &< \frac{k}{2}\\ \Leftrightarrow \frac{1}{2} - \frac{1}{k} &< \alpha.\end{aligned}$$

But this must hold for all odd $k \in \mathbb{N}$, which cannot be true if $\alpha < \frac{1}{2}$.

4.3.3 Decay of Variance of Moments

Now, let us show (4.6) on page 71, a decay of the variance, which is somewhat more involved.

Theorem 4.40. *Let $(X_n)_n$ be a sequence of periodic random band matrices which is based on an α-almost uncorrelated triangular array $(a_n)_n$ with $\alpha \geq \frac{1}{2}$ and bandwidth $b = (b_n)_n$. Denote by $(\sigma_n)_n$ the ESDs of $(X_n)_n$. Then we obtain the following results:*

i) If $b_n \to \infty$, we have for all $k \in \mathbb{N}$ that

$$\mathbb{V}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} 0.$$

ii) If all random variables of $(a_n)_n$ are $\{+1,-1\}$-valued and $\frac{1}{b_n^3}$ is summable over n, then we have for all $k \in \mathbb{N}$ that

$$\mathbb{V}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} 0 \quad \textit{summably fast.}$$

iii) If $(a_n)_n$ is even strongly α-almost uncorrelated with $\alpha > \frac{1}{2}$, and the sequences $\frac{1}{b_n^2}$, $\frac{1}{b_n} D_n^{(l)}$ and $C_n^{(l)}$ are summable over n for all $l \in \mathbb{N}$, then we have for all $k \in \mathbb{N}$ that

$$\mathbb{V}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} 0 \quad \textit{summably fast.}$$

iv) If $(a_n)_{n\in\mathbb{N}}$ is a Wigner scheme and if $\frac{1}{nb_n}$ is summable over n, we have that

$$\mathbb{V}\left\langle \sigma_n, x^k \right\rangle \xrightarrow[n\to\infty]{} 0 \quad \textit{summably fast.}$$

The proof is rather long and will be subdivided into several steps, each containing several cases and sometimes even subcases. The case-by-case analysis allows for a fine analysis, since only certain combinations of cases are relevant for each part of Theorem 4.40. An overview over the cases to be analyzed is given. Before we begin with the proof of Theorem 4.40, let us formulate a lemma which will facilitate the use of condition (AAU3).

Lemma 4.41. *Let $(a_n)_n$ be a strongly α-almost uncorrelated triangular scheme. Then for all $l, N \in \mathbb{N}$, $l \geq 3$ odd, and fundamentally different pairs $P_1, \dots, P_l$ in $\square_N$ we have for all $n \geq N$:*

$$\begin{aligned}&\left|\mathbb{E}\left[a_n(P_1)^4 a_n(P_2)^2 \cdots a_n(P_l)^2\right] - \mathbb{E}\left[a_n(P_1)^4 a_n(P_2)^2 \cdots a_n(P_{l_1})^2\right] \cdot \mathbb{E}\left[a_n(P_{l_2})^2 \cdots a_n(P_l)^2\right]\right| \\ &\leq D_n^{(l)} + C_{(0,0,0,1)} \cdot C_n^{(l_2)} + D_n^{(l_1)} \cdot C_{(0,l_2,0,\dots,0)_{2l_2}},\end{aligned}$$

where we set $l_1 := \frac{l-1}{2}$ and $l_2 := \frac{l+1}{2}$ (thus $l - l_2 = l_1$) and the subscript $2l_2$ of the vector $(0, l_2, 0, \dots, 0)$ indicates its dimension.

Proof. We calculate

$$\begin{aligned}&\left|\mathbb{E}\left[a_n(P_1)^4 a_n(P_2)^2 \cdots a_n(P_l)^2\right] - \mathbb{E}\left[a_n(P_1)^4 a_n(P_2)^2 \cdots a_n(P_{l_1})^2\right] \cdot \mathbb{E}\left[a_n(P_{l_2})^2 \cdots a_n(P_l)^2\right]\right| \\ &= |\mathbb{E}\left[a_n(P_1)^4 [a_n(P_2)^2 \cdots a_n(P_l)^2 - 1]\right] + \mathbb{E}\left[a_n(P_1)^4\right] \\ &\quad - \mathbb{E}\left[a_n(P_1)^4 [a_n(P_2)^2 \cdots a_n(P_{l_1})^2 - 1]\right] \cdot \mathbb{E}\left[a_n(P_{l_2})^2 \cdots a_n(P_l)^2\right] \\ &\quad - \mathbb{E}\left[a_n(P_1)^4\right] \cdot \mathbb{E}\left[a_n(P_{l_2})^2 \cdots a_n(P_l)^2\right]| \\ &\leq D_n^{(l)} + |\mathbb{E}\left[a_n(P_1)^4\right]||\mathbb{E}\left[a_n(P_{l_2})^2 \cdots a_n(P_l)^2 - 1\right]| + D_n^{(l_1)} \cdot C_{(0,l_2,0,\dots,0)_{2l_2}} \\ &\leq D_n^{(l)} + C_{(0,0,0,1)} \cdot C_n^{(l_2)} + D_n^{(l_1)} \cdot C_{(0,l_2,0,\dots,0)_{2l_2}}.\end{aligned}$$

□

Proof of Theorem 4.40. To begin with, we note that

$$\mathbb{V}\left\langle \sigma_n, x^k \right\rangle = \mathbb{E}\left(\left\langle \sigma_n, x^k \right\rangle\right)^2 - \left(\mathbb{E}\left\langle \sigma_n, x^k \right\rangle\right)^2.$$

Therefore, considering (4.9) on page 72, we need to show that

$$\frac{1}{n^2 b_n^k} \sum_{\underline{t}, \underline{t}' \in [n]_b^k} \left(\mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}') - \mathbb{E} a_n^b(\underline{t}) \mathbb{E} a_n^b(\underline{t}')\right) \xrightarrow[n \to \infty]{} 0 \tag{4.14}$$

and determine when this convergence is summably fast.

To show (4.14), we will subdivide the sum into finitely many subsums and determine the convergence for each of these subsums.

We remind the reader of the following notation (from page 76): For $\underline{s}, \underline{s}' \in [n]_b^k$ we have

$$\mathcal{T}(\underline{s}, \underline{s}') = \{(\underline{t}, \underline{t}') \,|\, \underline{t}, \underline{t}' \in [n]_b^k, \kappa(\underline{t}) = \kappa(\underline{s}), \kappa(\underline{t}') = \kappa(\underline{s}')\}$$

and a partitioning of this set into

$$\mathcal{T}^d(\underline{s}, \underline{s}') = \{(\underline{t}, \underline{t}') \,|\, \underline{t}, \underline{t}' \in [n]_b^k, \kappa(\underline{t}) = \kappa(\underline{s}), \kappa(\underline{t}') = \kappa(\underline{s}'),\ \underline{t} \text{ and } \underline{t}' \text{ are edge-disjoint}\}$$

and

$$\mathcal{T}^c(\underline{s}, \underline{s}') := \{(\underline{t}, \underline{t}') \,|\, \underline{t}, \underline{t}' \in [n]_b^k, \kappa(\underline{t}) = \kappa(\underline{s}), \kappa(\underline{t}') = \kappa(\underline{s}'),\ \underline{t} \text{ and } \underline{t}' \text{ have a common edge}\}.$$

In the following, we analyze the convergence in (4.14) by partitioning the sum into subsums over different (subsets of) equivalence classes of tuple pairs $(\underline{s}, \underline{s}')$. For example, we might consider the subsum

$$\frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}') \in \mathcal{T}^c(\underline{s},\underline{s}')} \left(\mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}') - \mathbb{E} a_n^b(\underline{t}) \mathbb{E} a_n^b(\underline{t}') \right),$$

where we suppose the tuple pair $(\underline{s}, \underline{s}')$ to belong to a specific equivalence class of tuple pairs, that is, we assume the profiles of $\underline{s}$ and $\underline{s}'$ to have certain properties. Since there are only finitely many equivalence classes of tuples pairs (this number is bounded by $(k+1)^k \cdot (k+1)^k$ by Lemma 4.34), this line of argumentation is valid.

For a better overview, we provide an outline of the upcoming case-by-case analysis:

1. Step: The tuples in the tuple pairs have disjoint edge sets.
   1. Case: Both $\underline{s}$ and $\underline{s}'$ have only even edges.
      1. Subcase: $\kappa_2(\underline{s}) = \frac{k}{2} = \kappa_2(\underline{s}')$.
      2. Subcase: $\kappa_l(\underline{s}) \geq 1$ or $\kappa_l(\underline{s}') \geq 1$ for some $l \in \{4, 6, 8, \ldots\}$.
   2. Case: $\underline{s}$ has at least one odd edge and $\underline{s}'$ has only even edges, or vice versa.
      1. Subcase: $\underline{s}$ has an m-fold edge with $m \geq 3$.
      2. Subcase: $\underline{s}$ has no m-fold edge with $m \geq 3$, but $\underline{s}'$ does.
      3. Subcase: Both $\underline{s}$ and $\underline{s}'$ have no m-fold edge with $m \geq 3$.
   3. Case: Both $\underline{s}$ and $\underline{s}'$ have at least one odd edge.
2. Step: The tuples in the tuple pairs have non-disjoint edge sets.
   1. Case: $\underline{s}$ and $\underline{s}'$ have only even edges.
   2. Case: $\underline{s}$ or $\underline{s}'$ contains at least one odd edge.

Here, if in "Step 1, Case 2" the vice versa case is treated, then $\underline{s}$ and $\underline{s}'$ will also swap their roles in the subcases.

In each case, we will determine which conditions are needed for regular convergence (i.e. convergence per se) and summable convergence in (4.14), and we will summarize our findings at the beginning of each case. After the case-by-case analysis we will argue for the statements i) through iv) of Theorem 4.40 by combining the findings of the relevant cases.

1. Step: Disjoint edge sets.

We analyze convergence to zero for subsums of

$$\frac{1}{n^2 b_n^k} \sum_{\substack{\underline{t},\underline{t}' \in [n]_b^k \\ \phi_{\underline{t}}(E_{\underline{t}}) \cap \phi_{\underline{t}'}(E_{\underline{t}'}) = \emptyset}} \left(\mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}') - \mathbb{E} a_n^b(\underline{t}) \mathbb{E} a_n^b(\underline{t}') \right).$$

1. Case: Both $\underline{s}$ and $\underline{s}'$ have only even edges.

1. Subcase: $\kappa_2(\underline{s}) = \frac{k}{2} = \kappa_2(\underline{s}')$.

[Outcome: We achieve regular convergence if the sequences $(C_n^{(l)})_n$ converge to zero and summably fast convergence if the sequences $(C_n^{(l)})_n$ are summable, where for both versions, we use conditions (AAU1) and (AAU2).]

So this is the subcase in which both $\underline{s}$ and $\underline{s}'$ consist only of double edges. Then we have due to the second moment property (AAU2) for all $\underline{t} \in \mathcal{T}(\underline{s})$ and $\underline{t}' \in \mathcal{T}(\underline{s}')$ with disjoint edge sets, that

$$\begin{aligned}
&|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - 1| \leq C_n^{(k)},\\
&|\mathbb{E}a_n^b(\underline{t}) - 1| \leq C_n^{(k/2)},\\
&|\mathbb{E}a_n^b(\underline{t}') - 1| \leq C_n^{(k/2)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')|\\
\leq &|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - 1| + |1 - (\mathbb{E}a_n^b(\underline{t}) - 1)\mathbb{E}a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t}')|\\
\leq &|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - 1| + |\mathbb{E}a_n^b(\underline{t}) - 1| \cdot \underbrace{|\mathbb{E}a_n^b(\underline{t}')|}_{\leq C_{\kappa(\underline{s}')}} + |\mathbb{E}a_n^b(\underline{t}') - 1| \leq D_n
\end{aligned}$$

for the real sequence $D_n := C_n^{(k)} + C_{\kappa(\underline{s}')}C_n^{(k/2)} + C_n^{(k/2)}$ that converges to 0.

Further, by Lemma 4.34, we have at most

$$\#\mathcal{T}^d(\underline{s}, \underline{s}') \leq \#\mathcal{T}(\underline{s}) \cdot \#\mathcal{T}(\underline{s}') \leq k^k \cdot nb_n^{\frac{k}{2}} \cdot k^k \cdot nb_n^{\frac{k}{2}} = k^{2k} \cdot n^2 b_n^k$$

pairs of tuples $\underline{t} \in \mathcal{T}(\underline{s})$ and $\underline{t}' \in \mathcal{T}(\underline{s}')$ with disjoint edge sets. Therefore,

$$\left| \frac{1}{n^2 b_n^k} \sum_{(\underline{t}, \underline{t}') \in \mathcal{T}^d(\underline{s}, \underline{s}')} \left(\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}') \right) \right| \leq k^{2k} D_n \xrightarrow[n\to\infty]{} 0$$

and this convergence is summably fast if the sequences $(C_n^{(l)})_n$ converge to zero summably fast.

2. Subcase: We have $\kappa_l(\underline{s}) \geq 1$ or $\kappa_l(\underline{s}') \geq 1$ for some $l \in \{4, 6, 8, \ldots\}$.
[Outcome: We achieve regular convergence if $b_n \to \infty$, and summably fast convergence if $\frac{1}{b_n^2}$, $\frac{1}{b_n} D_n^{(l)}$ and $C_n^{(l)}$ are summable for all l. For the regular version we use (AAU1), for the summable version (AAU1) and (AAU3).]

We first argue for regular convergence without using condition (AAU3): For all $\underline{t} \in \mathcal{T}(\underline{s})$ and $\underline{t}' \in \mathcal{T}(\underline{s}')$ with disjoint edge sets we have that $|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')|$ remains uniformly (in $n \in \mathbb{N}$, $(\underline{t}, \underline{t}') \in \mathcal{T}^d(\underline{s}, \underline{s}')$) bounded by some real number B, which follows from the boundedness property (AAU1). To make this more precise, note that if $n, z \in \mathbb{N}$ are arbitrary and $P_1, \ldots, P_z \in \square_n$ are arbitrary, then using Lemma 3.13 and (AAU1) we obtain

$$\mathbb{E}|a_n(P_1) \cdots a_n(P_z)| \leq \max_{i=1,\ldots,z} \mathbb{E}|a_n(P_i)|^z \leq \max_{i=1,\ldots,z} \sqrt{\mathbb{E}a_n(P_i)^{2z}} \leq \sqrt{C_{\Phi(2z)}}. \tag{4.15}$$

and thus

$$|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')| + |\mathbb{E}a_n^b(\underline{t})| \cdot |\mathbb{E}a_n^b(\underline{t}')| \leq \sqrt{C_{\Phi(4k)}} + C_{\Phi(2k)} =: B.$$

We assume w.l.o.g. that $\kappa_l(\underline{s}) \geq 1$ with $l \in \{4,6,8,\ldots\}$, then by Lemma 4.34 we obtain

$$\#\mathcal{T}(\underline{s}) \leq k^k \cdot nb_n^{\kappa_2(\underline{s})+\kappa_4(\underline{s})+\ldots+\kappa_k(\underline{s})} \leq k^k \cdot nb_n^{\frac{k}{2}-1},$$

where the bound follows from the fact that the exponent in the term before is maximized when $\kappa_l(\underline{s}) = 1$ and the rest of the edges of $\underline{s}$ are double edges, thus giving the exponent

$$1 + \frac{k - l\kappa_l(\underline{s})}{2} \leq 1 + \frac{k}{2} - \frac{l}{2} \leq 1 + \frac{k}{2} - \frac{4}{2} = \frac{k}{2} - 1.$$

To choose the other tuple $\underline{t}' \in \mathcal{T}(\underline{s}')$ edge-disjoint from $\underline{t}$, we have at most $k^k \cdot nb_n^{\frac{k}{2}}$ possibilities, since this is the unrestricted upper bound. Therefore, in total, we are considering at most

$$k^k \cdot nb_n^{\frac{k}{2}-1} \cdot k^k \cdot nb_n^{\frac{k}{2}} = k^{2k} \cdot n^2 b_n^{k-1}$$

tuple pairs in this subcase, in formulas

$$\#\mathcal{T}^d(\underline{s}, \underline{s}') \leq k^{2k} \cdot n^2 b_n^{k-1}.$$

Therefore, we calculate

$$\left| \frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}') \in \mathcal{T}(\underline{s},\underline{s}')} \mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}') - \mathbb{E} a_n^b(\underline{t}) \mathbb{E} a_n^b(\underline{t}') \right| \leq \frac{1}{n^2 b_n^k} \cdot k^{2k} \cdot n^2 b_n^{k-1} \cdot B \xrightarrow[n\to\infty]{} 0,$$

if $b_n \to \infty$. Note that since $b_n \leq n$, $\frac{1}{b_n}$ is not summable.

Now we will obtain faster decays: First note that in above calculation, if l can be chosen in the set $\{6,8,\ldots\}$ or if $l = 4$ and $\kappa_l(\underline{s}) \geq 2$, we achieve an exponent of b_n of at most $\frac{k}{2} - 2$ in the upper bound of $\#\mathcal{T}(\underline{s})$ and thus summably fast convergence to zero if $\frac{1}{b_n^2}$ is summable. Also note in above calculation, that if also $\kappa_{l'}(\underline{s}') \geq 1$ for some $l' \in \{4,6,8,\ldots\}$, then also $\#\mathcal{T}(\underline{s}') \leq k^k n b_n^{k/2-1}$, thus $\#\mathcal{T}^d(\underline{s},\underline{s}') \leq k^{2k} n^2 b_n^{k-2}$ and thus summably fast convergence if $1/b_n^2$ is summable.

Now if $l = 4$, $\kappa_l(\underline{s}) = 1$, $\kappa_2(\underline{s}) = (k-4)/2$ and $\kappa_2(\underline{s}') = k/2$ (that is, we have one 4-fold edge and double edges otherwise), then we must resort to (AAU3) and Lemma 4.41 (note that $a_n^b(\underline{t}) = a_n(\underline{t})$ for b_n-relevant tuples $\underline{t}$), since then there are natural numbers z_1 and z_2 and constants C_1 and C_2 such that

$$\begin{aligned}&\left| \frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}') \in \mathcal{T}(\underline{s},\underline{s}')} \mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}') - \mathbb{E} a_n^b(\underline{t}) \mathbb{E} a_n^b(\underline{t}') \right| \\ &\leq \frac{1}{n^2 b_n^k} \cdot k^{2k} \cdot (D_n^{(z_1)} + C_1 \cdot C_n^{(z_2+1)} + D_n^{(z_2)} \cdot C_2) \cdot n^2 b_n^{k-1} \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

where the convergence is summably fast if $\frac{1}{b_n} D_n^{(l)}$ and $C_n^{(l)}$ are summable for all l.

This completes the 1. Case of the 1. Step, in which $\underline{s}$ and $\underline{s}'$ had only even edges.

2. Case: $\underline{s}$ has at least one odd edge and $\underline{s}'$ has only even edges, or vice versa.
Throughout this case, we assume w.l.o.g. that $\underline{s}$ has at least one odd edge and $\underline{s}'$ has only even edges (in the vice versa case, $\underline{s}$ and $\underline{s}'$ will also swap their roles in the subcases). Then, actually, $\underline{s}$ has at least two odd edges, since the total number of edges is even.
1. Subcase: $\underline{s}$ has an m-fold edge, $m \geq 3$.
[Outcome: We achieve regular convergence if $b_n \to \infty$ and summably fast convergence if $\frac{1}{b_n^2}$ is summable and $\alpha > \frac{1}{2}$. For both versions we use (AAU1).]

For each $(\underline{t}, \underline{t}') \in \mathcal{T}^d(\underline{s}, \underline{s}')$ we now have due to the distinct decay property (AAU1):

$$|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')| \leq \frac{C_{[\kappa(\underline{s})+\kappa(\underline{s}')]}}{n^{\alpha\cdot\kappa_1(\underline{s})}}$$

and

$$|\mathbb{E}a_n^b(\underline{t})| \cdot \underbrace{|\mathbb{E}a_n^b(\underline{t}')|}_{\leq C_{\kappa(\underline{s}')}} \leq \frac{C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}}{n^{\alpha\cdot\kappa_1(\underline{s})}}$$

Now, we determine an upper bound for $\#\mathcal{T}^d(\underline{s}, \underline{s}')$. Clearly, the number of possibilities to choose $\underline{t} \in \mathcal{T}(\underline{s})$ and $\underline{t}' \in \mathcal{T}(\underline{s}')$ edge-disjoint is bounded by the number of unrestricted possibilities, where by Lemma 4.34 we have at most

$$k^k \cdot nb_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})-1}$$

possibilities for $\underline{t}$ and at most

$$k^k \cdot nb_n^{\kappa_1(\underline{s}')+\ldots+\kappa_k(\underline{s}')} \leq k^k \cdot nb_n^{k/2}$$

possibilities for $\underline{t}'$, yielding the bound

$$\#\mathcal{T}^d(\underline{s}, \underline{s}') \leq k^{2k} \cdot n^2 b_n^{\frac{k}{2}+\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})-1}.$$

Therefore, we obtain

$$\begin{aligned}
&\left|\frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}')\in\mathcal{T}^d(\underline{s},\underline{s}')} \mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')\right| \\
&\leq \frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}')\in\mathcal{T}^d(\underline{s},\underline{s}')} |\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')| + \frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}')\in\mathcal{T}^d(\underline{s},\underline{s}')} |\mathbb{E}a_n^b(\underline{t})| \cdot |\mathbb{E}a_n^b(\underline{t}')| \\
&\leq \frac{1}{n^2 b_n^k} \cdot k^{2k} \cdot C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} \cdot n^2 \cdot \frac{b_n^{\frac{k}{2}+\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{s})-1}}{n^{\alpha\cdot\kappa_1(\underline{s})}} \\
&\quad + \frac{1}{n^2 b_n^k} \cdot k^{2k} \cdot C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')} \cdot n^2 \cdot \frac{b_n^{\frac{k}{2}+\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{s})-1}}{n^{\alpha\cdot\kappa_1(\underline{s})}} \\
&= \frac{k^{2k}}{b_n^{\frac{k}{2}}} \cdot (C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}) \cdot \frac{b_n^{\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{s})-1}}{n^{\alpha\cdot\kappa_1(\underline{s})}} \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

if $b_n \to \infty$, where the convergence is summably fast if $\frac{1}{b_n^2}$ is summable over n and $\alpha > \frac{1}{2}$. Of course, we will have to prove the last statements, so we will analyze the decay of

$$\frac{1}{b_n^{\frac{k}{2}}} \cdot \frac{b_n^{\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{s})-1}}{n^{\alpha\cdot\kappa_1(\underline{s})}}.$$

To this end, we first assume that $\kappa_1(\underline{s}) = 0$. Then the exponent

$$\kappa_1(\underline{s}) + \kappa_2(\underline{s}) + \ldots + \kappa_k(\underline{s}) - 1$$

is maximal if $m = 3$, $\kappa_m(\underline{s}) = 2$ and $\kappa_2(\underline{s}) = \frac{k-6}{2} = \frac{k}{2} - 3$ (note that since k is even, we need to have at least two odd edges) and then assumes the value

$$\frac{k}{2} - 3 + 2 - 1 = \frac{k}{2} - 2.$$

Therefore, we obtain regular convergence if $b_n \to \infty$ and a summable convergence if $\frac{1}{b_n^2}$ is summable.

Now, we assume that $\kappa_1(\underline{s}) > 0$. Then the exponent in the numerator can surely be bounded in the following way, in which we assume we have $\kappa_1(\underline{s})$ single edges, one m-fold edge, and that all remaining edges (besides the single edges and the m-fold edge) can be allocated to double edges (if this cannot be done, i.e. $k - m\kappa_m(\underline{s}) - \kappa_1(\underline{s})$ is not even, then our bound gets even better, i.e. tighter, since then we have to allocate the remaining edges to edges of higher multiplicity, of which there can be fewer):

$$\kappa_1(\underline{s}) + \kappa_2(\underline{s}) + \ldots + \kappa_k(\underline{s}) - 1 \leq \kappa_1(\underline{s}) + \frac{k - m - \kappa_1(\underline{s})}{2} + 1 - 1 = \frac{\kappa_1(\underline{s})}{2} + \frac{k}{2} - \frac{m}{2}$$

Remembering that $|ab| \leq \frac{1}{2}(a^2 + b^2)$, we calculate

$$\begin{aligned}
\frac{1}{b_n^{\frac{k}{2}}} \cdot \frac{b_n^{\kappa_1(\underline{s})+\kappa_2(\underline{s})+\ldots+\kappa_k(\underline{s})-1}}{n^{\alpha\cdot\kappa_1(\underline{s})}} &\leq \frac{1}{b_n^{\frac{k}{2}}} \cdot \frac{b_n^{\frac{\kappa_1(\underline{s})}{2}+\frac{k}{2}-\frac{m}{2}}}{n^{\alpha\cdot\kappa_1(\underline{s})}} \\
&= \frac{b_n^{\frac{\kappa_1(\underline{s})}{2}-\frac{m}{2}}}{n^{\alpha\cdot\kappa_1(\underline{s})}} \\
&= \frac{1}{b_n^{\frac{m}{2}-\frac{\kappa_1(\underline{s})}{2}}} \cdot \frac{1}{n^{\alpha\cdot(\kappa_1(\underline{s})-1)}} \cdot \frac{1}{n^{\alpha}} \\
&\leq \frac{1}{2}\frac{1}{b_n^{m-\kappa_1(\underline{s})}}\frac{1}{n^{2\alpha(\kappa_1(\underline{s})-1)}} + \frac{1}{2}\frac{1}{n^{2\alpha}} \\
&\leq \frac{1}{2}\frac{1}{b_n^{m-\kappa_1(\underline{s})}}\frac{1}{b_n^{\kappa_1(\underline{s})-1}} + \frac{1}{2}\frac{1}{n^{2\alpha}} \\
&= \frac{1}{2}\frac{1}{b_n^{m-1}} + \frac{1}{2}\frac{1}{n^{2\alpha}} \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

if $b_n \to \infty$ since $m \geq 3$, where the convergence is summably fast if $\frac{1}{b_n^2}$ is summable and $\alpha > \frac{1}{2}$.

2. Subcase: $\underline{s}$ has no m-fold edge with $m \geq 3$, but $\underline{s}'$ has an m-fold edge with $m \geq 3$.
[Outcome: We achieve regular convergence if $b_n \to \infty$ and summably fast convergence if $\frac{1}{b_n^2}$ is summable. For both versions we use condition (AAU1).]

First, considering that $\underline{s}'$ has only even edges, we have by Lemma 4.34, that

$$\#\mathcal{T}(\underline{s}') \leq k^k \cdot n b_n^{\kappa_1(\underline{s}')+\ldots+\kappa_k(\underline{s}')} \leq k^k \cdot n b_n^{\frac{k}{2}-1}.$$

We need to justify the last inequality: Since $\underline{s}'$ has only even edges and an m-fold edge with $m \geq 3$, it actually has an m-fold edge with $m \geq 4$. Then the exponent is maximized if $\underline{s}'$ has just one such m-fold edge, thus $\kappa_m(\underline{s}') = 1$, and all other edges are double, thus $\kappa_2(\underline{s}') = \frac{k-m}{2}$. This yields an exponent of at most

$$\frac{k-m}{2} + 1 = \frac{k}{2} - \frac{m}{2} + 1 \leq \frac{k}{2} - 1,$$

since $m \geq 4$.

Now the number of possibilities to choose a $\underline{t} \in \mathcal{T}(\underline{s})$ edge-disjoint from a chosen $\underline{t}' \in \mathcal{T}(\underline{s}')$ is bounded by the number of unrestricted possibilities. There are at most $k^k n b_n^{\kappa_1(\underline{s})+\kappa_2(\underline{s})-1}$ such possibilities by Lemma 4.34, yielding

$$\#\mathcal{T}^d(\underline{s}, \underline{s}') \leq k^{2k} \cdot n^2 b_n^{\frac{k}{2}+\kappa_1(\underline{s})+\kappa_2(\underline{s})-2}.$$

Fortunately, in this subcase we can bound more generously than in the previous subcase: For each $(\underline{t}, \underline{t}') \in \mathcal{T}^d(\underline{s}, \underline{s}')$ we now have due to the distinct decay property (AAU1):

$$|\mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}')| \leq \frac{C_{[\kappa(\underline{s})+\kappa(\underline{s}')]}}{n^{\alpha \cdot \kappa_1(\underline{s})}} \leq \frac{C_{[\kappa(\underline{s})+\kappa(\underline{s}')]}}{b_n^{\frac{1}{2}\cdot\kappa_1(\underline{s})}}$$

and

$$|\mathbb{E} a_n^b(\underline{t})| \cdot \underbrace{|\mathbb{E} a_n^b(\underline{t}')|}_{\leq C_{\kappa(\underline{s}')}} \leq \frac{C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}}{n^{\alpha \cdot \kappa_1(\underline{s})}} \leq \frac{C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}}{b_n^{\frac{1}{2}\cdot\kappa_1(\underline{s})}}.$$

We thus arrive at

$$\begin{aligned}
&\left| \frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}') \in \mathcal{T}^d(\underline{s},\underline{s}')} \mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}') - \mathbb{E} a_n^b(\underline{t}) \mathbb{E} a_n^b(\underline{t}') \right| \\
&\leq \frac{k^{2k}}{n^2 b_n^k} \cdot \left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}\right) \cdot n^2 b_n^{\frac{k}{2}+\frac{1}{2}\kappa_1(\underline{s})+\kappa_2(\underline{s})-2} \\
&= \frac{k^{2k}}{b_n^k} \cdot \left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}\right) \cdot b_n^{k-2} \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

if $b_n \to \infty$ and where the convergence is summably fast if $\frac{1}{b_n^2}$ is summable over n. Note that the last equality follows due to the fact that $\underline{s}$ has k edges, all of them single or double, so $1/2\kappa_1(\underline{s}) + \kappa_2(\underline{s}) = k/2$.

3. Subcase: Both $\underline{s}$ and $\underline{s}'$ have no m-fold edge with $m \geq 3$.
[Outcome: We achieve regular convergence for $\alpha \geq \frac{1}{2}$ and summably fast convergence if $\alpha > \frac{1}{2}$. For both versions we use (AAU1).]

In this subcase, $\underline{s}$ has an even number of single edges, but at least 2 of them, and the remaining edges are double edges, whereas $\underline{s}'$ contains only double edges.

By above standard argumentation and Lemma 4.34, we have

$$\#\mathcal{T}^d(\underline{s}, \underline{s}') \leq k^{2k} \cdot n^2 b_n^{\frac{k}{2}+\kappa_1(\underline{s})+\kappa_2(\underline{s})-1},$$

and for each $(\underline{t}, \underline{t}') \in \mathcal{T}^d(\underline{s}, \underline{s}')$ we will bound $|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')|$ and $|\mathbb{E}a_n^b(\underline{t})||\mathbb{E}a_n^b(\underline{t}')|$ as follows:

$$|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')| \leq \frac{C_{[\kappa(\underline{s})+\kappa(\underline{s}')]}}{n^{\alpha \cdot \kappa_1(\underline{s})}}$$

and

$$|\mathbb{E}a_n^b(\underline{t})| \cdot \underbrace{|\mathbb{E}a_n^b(\underline{t}')|}_{\leq C_{\kappa(\underline{s}')}} \leq \frac{C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}}{n^{\alpha \cdot \kappa_1(\underline{s})}}.$$

Using $\kappa_2(\underline{s}) = \frac{k-\kappa_1(\underline{s})}{2}$, we arrive at

$$\begin{aligned}
&\left| \frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}') \in \mathcal{T}^d(\underline{s},\underline{s}')} \mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}') \right| \\
&\leq \frac{k^{2k}}{b_n^{k/2}} \cdot \left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}\right) \cdot \frac{b_n^{\kappa_1(\underline{s})+\kappa_2(\underline{s})-1}}{n^{\alpha \cdot \kappa_1(\underline{s})}} \\
&= \frac{k^{2k}}{b_n^{k/2}} \cdot \left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}\right) \cdot \frac{b_n^{\kappa_1(\underline{s})+\frac{k-\kappa_1(\underline{s})}{2}-1}}{n^{\alpha \cdot \kappa_1(\underline{s})}} \\
&= k^{2k} \cdot \left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}\right) \cdot \frac{b_n^{\frac{\kappa_1(\underline{s})}{2}-1}}{n^{\alpha \cdot \kappa_1(\underline{s})}} \\
&\leq k^{2k} \cdot \left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}\right) \cdot \frac{n^{\frac{\kappa_1(\underline{s})}{2}-1}}{n^{\alpha \cdot \kappa_1(\underline{s})}} \\
&= k^{2k} \cdot \left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}\right) \cdot \frac{1}{n^{(2\alpha-1)\frac{\kappa_1(\underline{s})}{2}+1}} \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

where the convergence is regular for $\alpha \geq \frac{1}{2}$ and summably fast if $\alpha > \frac{1}{2}$, since $\kappa_1(\underline{s}) \geq 2$.

3. Case: Both $\underline{s}$ and $\underline{s}'$ have at least one odd edge.
[Outcome: We achieve regular convergence if $b_n \to \infty$ and summably fast convergence if $\frac{1}{b_n^3}$ is summable. For both versions we use (AAU1).]
To begin, by Lemma 4.34, we have $\#\mathcal{T}(\underline{s}) \leq k^k \cdot n b_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})-1}$ and $\#\mathcal{T}(\underline{s}') \leq k^k \cdot n b_n^{\kappa_1(\underline{s}')+\ldots+\kappa_k(\underline{s}')-1}$ and therefore surely

$$\#\mathcal{T}^d(\underline{s}, \underline{s}') \leq k^{2k} \cdot n^2 b_n^{\kappa_1(\underline{s})+\ldots+\kappa_k(\underline{s})+\kappa_1(\underline{s}')+\ldots+\kappa_k(\underline{s}')-2}.$$

Further, by (AAU1) we have for each $(\underline{t}, \underline{t}') \in \mathcal{T}^d(\underline{s}, \underline{s}')$, that

$$|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')| \leq \frac{C_{[\kappa(\underline{s})+\kappa(\underline{s}')]}}{n^{\alpha\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}} \leq \frac{C_{[\kappa(\underline{s})+\kappa(\underline{s}')]}}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}}$$

and

$$|\mathbb{E}a_n^b(\underline{t})||\mathbb{E}a_n^b(\underline{t}')| \leq \frac{C_{\kappa(\underline{s})}\cdot C_{\kappa(\underline{s}')}}{n^{\alpha\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}} \leq \frac{C_{\kappa(\underline{s})}\cdot C_{\kappa(\underline{s}')}}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}}.$$

By the triangle inequality and the bounds above we obtain

$$\begin{aligned}
&\left|\frac{1}{n^2 b_n^k}\sum_{(\underline{t},\underline{t}')\in\mathcal{T}^d(\underline{s},\underline{s}')}\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')\right| \\
&\leq \frac{1}{n^2 b_n^k}\cdot\left(C_{[\kappa(\underline{s})+\kappa(\underline{s}')]} + C_{\kappa(\underline{s})}C_{\kappa(\underline{s}')}\right)\cdot k^{2k}n^2\cdot b_n^{\sum_{i\leq k}\kappa_i(\underline{s})+\sum_{j\leq k}\kappa_j(\underline{s}')-2}\cdot\frac{1}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}} \\
&= \frac{1}{b_n^k}\cdot C\cdot b_n^{\sum_{i\leq k}\kappa_i(\underline{s})+\sum_{j\leq k}\kappa_j(\underline{s}')-2}\cdot\frac{1}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}} \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

if $b_n \to \infty$, where the convergence is summably fast if $\frac{1}{b_n^3}$ is summable (and where C absorbs all constants).

Of course, we have to argue the last statement, for which we will distinguish three subcases, but we will not declare these as subcases formally.
First, we assume that $\kappa_1(\underline{s}) = 0 = \kappa_1(\underline{s}')$. Then the exponent of b_n is maximized if both $\underline{s}$ and $\underline{s}'$ have only one triple edge and just double edges otherwise (if this is possible, i.e., k is odd. If k is even, we will get an even tighter bound, since we will have more edges of higher multiplicity, of which there can only be fewer. This reasoning will be used in the second and third point as well). Then the exponent of b_n in the numerator is bounded by

$$\frac{k-3}{2} + 1 + \frac{k-3}{2} + 1 - 2 = k - 3,$$

showing the above statement.
Second, we assume that $\kappa_1(\underline{s}) \geq 1$ but $\kappa_1(\underline{s}') = 0$ (or vice versa). In this case the term

$$\frac{1}{b_n^k}\cdot b_n^{\sum_{i\leq k}\kappa_i(\underline{s})+\sum_{j\leq k}\kappa_j(\underline{s}')-2}\cdot\frac{1}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}}$$

is maximized if $\underline{s}$ consists of only one single edge and double edges otherwise, whereas $\underline{s}'$ consists of only one triple edge and double edges otherwise, and we get

$$\begin{aligned}
&\frac{1}{b_n^k}\cdot b_n^{\sum_{i\leq k}\kappa_i(\underline{s})+\sum_{j\leq k}\kappa_j(\underline{s}')-2}\cdot\frac{1}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}} \\
&\leq \frac{1}{b_n^k}b_n^{1+\frac{k-1}{2}+\frac{k-3}{2}+1-2}\cdot\frac{1}{n^{\frac{1}{2}}} \\
&= \frac{1}{b_n^2}\cdot\frac{1}{n^{\frac{1}{2}}} \leq \frac{3}{5}\cdot\frac{1}{b_n^{\frac{10}{3}}} + \frac{2}{5}\cdot\frac{1}{n^{\frac{5}{4}}},
\end{aligned}$$

which converges to zero as $b_n \to \infty$ and this summably fast if $\frac{1}{b_n^3}$ is summable. In the last step in above calculation we used Young's inequality

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

which holds for all non-negative real numbers a, b, p and q such that $\frac{1}{p} + \frac{1}{q} = 1$.
Third, we assume that $\kappa_1(\underline{s}) \geq 1$ and $\kappa_1(\underline{s}') \geq 1$. In this case the term

$$\frac{1}{b_n^k} \cdot b_n^{\sum_{j\leq k}\kappa_j(\underline{s})+\sum_{j\leq k}\kappa_j(\underline{s}')-2} \cdot \frac{1}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}}$$

is maximized if $\underline{s}$ and $\underline{s}'$ both consist of only one single edge and double edges otherwise, and we get

$$\begin{aligned}
&\frac{1}{b_n^k} \cdot b_n^{\sum_{i\leq k}\kappa_i(\underline{s})+\sum_{j\leq k}\kappa_j(\underline{s}')-2} \cdot \frac{1}{n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}} \\
&\leq \frac{1}{b_n^k} \cdot b_n^{1+\frac{k-1}{2}+1+\frac{k-1}{2}-2} \cdot \frac{1}{n} \\
&= \frac{1}{b_n} \cdot \frac{1}{n} \leq \frac{1}{3} \cdot \frac{1}{b_n^3} + \frac{2}{3} \cdot \frac{1}{n^{\frac{3}{2}}},
\end{aligned}$$

which converges to zero if $b_n \to \infty$ and this summably fast if $\frac{1}{b_n^3}$ is summable.

Now, the first step of the proof, when we sum over edge-wise disjoint tuples, is completed.

2. Step: Non-disjoint edge sets.

We analyze convergence to zero for subsums of

$$\frac{1}{n^2 b_n^k} \sum_{\substack{\underline{t},\underline{t}'\in[n]_b^k \\ \phi_{\underline{t}}(E_{\underline{t}})\cap\phi_{\underline{t}'}(E_{\underline{t}'})\neq\emptyset}} \left(\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')\right).$$

1. Case: $\underline{s}$, $\underline{s}'$ have only even edges.

[Outcome: We achieve regular convergence as is and summably fast convergence if $\frac{1}{nb_n}$ is summable. For both versions we use (AAU1).]

By Lemma 4.37, we obtain the bound

$$\#\mathcal{T}^c(\underline{s},\underline{s}') \leq k^2 \cdot (2k)^{2k} n b_n^{k-1}.$$

Next, recall the constant B which we constructed in "Step 1, Case 1, Subcase 2" (see (4.15) and beyond) independent of n, which clearly satisfies for all $n \in \mathbb{N}$ and all $(\underline{t},\underline{t}') \in \mathcal{T}^c(\underline{s},\underline{s}')$:

$$|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')| \;\leq\; \mathbb{E}|a_n^b(\underline{t})a_n^b(\underline{t}')| + \mathbb{E}|a_n^b(\underline{t})|\mathbb{E}|a_n^b(\underline{t}')| \;\leq\; B.$$

With these bounds, we achieve

$$\left|\frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}')\in\mathcal{T}^c(\underline{s},\underline{s}')} \mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') - \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')\right| \leq \frac{1}{n^2 b_n^k} \cdot k^2 \cdot (2k)^{2k} \cdot n b_n^{k-1} \cdot B \xrightarrow[n\to\infty]{} 0,$$

where the convergence is summably fast if $\frac{1}{nb_n}$ is summable over n.

We now turn to the final case of this proof.

2. Case: $\underline{s}$ or $\underline{s}'$ contains at least one odd edge.

[Outcome: We achieve regular convergence as is and summably fast convergence if $(\frac{1}{nb_n})_n$ is summable. For both versions we use (AAU1).]

Of course, we have

$$\begin{aligned}&\left|\frac{1}{n^2b_n^k}\sum_{(\underline{t},\underline{t}')\in\mathcal{T}^c(\underline{s},\underline{s}')}\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')-\mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')\right|\\&\leq\frac{1}{n^2b_n^k}\sum_{(\underline{t},\underline{t}')\in\mathcal{T}^c(\underline{s},\underline{s}')}(|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')|+|\mathbb{E}a_n^b(\underline{t})|\cdot|\mathbb{E}a_n^b(\underline{t}')|)\\&=\sum_{l=1}^k\frac{1}{n^2b_n^k}\sum_{(\underline{t},\underline{t}')\in\mathcal{T}_l^c(\underline{s},\underline{s}')}(|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')|+|\mathbb{E}a_n^b(\underline{t})|\cdot|\mathbb{E}a_n^b(\underline{t}')|).\end{aligned}$$

Thus, it suffices to show that each of the k summands converges to zero if $b_n\to\infty$, and that this convergence is summably fast as long as $(\frac{1}{nb_n})_n$ is summable. To this end, pick an $l\in\{1,\ldots,k\}$ and a $(\underline{t},\underline{t}')\in\mathcal{T}_l^c(\underline{s},\underline{s}')$. How can we bound $|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')|$ and $|\mathbb{E}a_n^b(\underline{t})|\cdot|\mathbb{E}a_n^b(\underline{t}')|$?

Surely, by the distinct decay property (AAU1), we obtain

$$|\mathbb{E}a_n^b(\underline{t})|\cdot|\mathbb{E}a_n^b(\underline{t}')|\leq\frac{C_{\kappa(\underline{s})}\cdot C_{\kappa(\underline{s}')}}{n^{\alpha\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}}\leq\frac{C_{\kappa(\underline{s})}\cdot C_{\kappa(\underline{s}')}}{b_n^{\frac{1}{2}\cdot(\kappa_1(\underline{s})+\kappa_1(\underline{s}'))}}.$$

Now to treat $|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')|$, note that with Lemma 4.35, we find a Eulerian cycle $\underline{u}\in[n]^{2k}$ which passes through the graph obtained through superposition of the graphs of $\underline{t}$ and $\underline{t}'$. But how many single edges does $\underline{u}$ have, i.e., what can we say about $\kappa_1(\underline{u})$? If $\underline{t}$ and $\underline{t}'$ were edge-disjoint, we would have $\kappa_1(\underline{u})=\kappa_1(\underline{t})+\kappa_1(\underline{t}')$. But now, for each common edge of $\underline{t}$ and $\underline{t}'$, the number of single edges can be reduced by at most 2 after superposition of the graphs, which happens if the common edge is a single edge in both $\underline{t}$ and $\underline{t}'$. Therefore, we obtain

$$\kappa_1(\underline{u})\geq\max(\kappa_1(\underline{t})+\kappa_1(\underline{t}')-2l,0)=\max(\kappa_1(\underline{s})+\kappa_1(\underline{s}')-2l,0)$$

and thus

$$|\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}')|=|\mathbb{E}a_n^b(\underline{u})|\leq\frac{C_{\kappa(\underline{u})}}{n^{\alpha\cdot\kappa_1(\underline{u})}}\leq\frac{C_{\kappa(\underline{u})}}{b_n^{\frac{1}{2}\cdot\kappa_1(\underline{u})}}\leq\frac{B}{b_n^{\frac{1}{2}\cdot\max(\kappa_1(\underline{s})+\kappa_1(\underline{s}')-2l,0)}},$$

where the constant B is taken from the cases above. Since Lemma 4.37 yields

$$\#\mathcal{T}_l^c(\underline{s},\underline{s}')\leq k^2\cdot(2k)^{2k}\cdot nb_n^{\sum_i\kappa_i(\underline{s})+\sum_j\kappa_j(\underline{s}')-l-1},$$

we obtain

$$\begin{aligned}
&\frac{1}{n^2 b_n^k} \sum_{(\underline{t},\underline{t}') \in \mathcal{T}_l^c(\underline{s},\underline{s}')} \left(|\mathbb{E} a_n^b(\underline{t}) a_n^b(\underline{t}')| + |\mathbb{E} a_n^b(\underline{t})| \cdot |\mathbb{E} a_n^b(\underline{t}')| \right) \\
&\leq \frac{1}{n^2 b_n^k} \cdot k^2 \cdot (2k)^{2k} \cdot n b_n^{\sum_i \kappa_i(\underline{s}) + \sum_j \kappa_j(\underline{s}') - l - 1} \cdot \left(\frac{B}{b_n^{\frac{1}{2} \cdot \max(\kappa_1(\underline{s}) + \kappa_1(\underline{s}') - 2l, 0)}} + \frac{C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}}{b_n^{\frac{1}{2} \cdot (\kappa_1(\underline{s}) + \kappa_1(\underline{s}'))}} \right) \\
&\leq \frac{k^2 (2k)^{2k}}{n b_n^k} \cdot (B + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}) b_n^{\frac{1}{2}\kappa_1(\underline{s}) + \sum_{i \geq 2} \kappa_i(\underline{s}) + \frac{1}{2}\kappa_1(\underline{s}') + \sum_{j \geq 2} \kappa_j(\underline{s}') - 1} \\
&\leq \frac{k^2 (2k)^{2k}}{n b_n^k} \cdot (B + C_{\kappa(\underline{s})} \cdot C_{\kappa(\underline{s}')}) b_n^{k-1} \xrightarrow[n \to \infty]{} 0,
\end{aligned}$$

where the convergence is summably fast if $\frac{1}{nb_n}$ is summable. We have to justify the second and third inequality in above calculation. The second inequality follows by a case-by-case analysis whether $\kappa_1(\underline{s}) + \kappa_1(\underline{s}') \geq 2l$ or $\kappa_1(\underline{s}) + \kappa_1(\underline{s}') < 2l$. In the first case, the inequality is clear, while in the second case we obtain $l > \frac{\kappa_1(\underline{s})}{2} + \frac{\kappa_1(\underline{s}')}{2}$, which also yields the inequality.

The third inequality follows since the exponent of b_n is maximized, for example, when both tuples $\underline{s}$ and $\underline{s}'$ contain only single edges and thus assumes the value $k-1$.

This concludes the lengthy case-by-case analysis. For each of the different cases we pointed out which conditions we need for regular and summable convergence of the variance to zero. Now we will argue for the statements i) through iv) of Theorem 4.40:

i) As can be seen from the outcome of each of the cases, the condition that $b_n \to \infty$ suffices for regular convergence of the variance to zero.

ii) Assuming $\{+1,-1\}$-valued entries, we observe that the term in (4.14) on page 85 vanishes for each subsum in our case-by-case analysis except in "Step 1, Case 3" and "Step 2, Case 2" (since if $\underline{t}$, say, consists of only even edges, we have $a_n(\underline{t}) \equiv 1$). In those cases, for a summably fast convergence we need that $\frac{1}{b_n^3}$ is summable and that $\frac{1}{nb_n}$ is summable. Since the former implies the latter through

$$\frac{1}{nb_n} \leq \frac{2}{3} \cdot \frac{1}{n^{\frac{3}{2}}} + \frac{1}{3} \cdot \frac{1}{b_n^3},$$

where we used Young's inequality, it is enough to assume that $\frac{1}{b_n^3}$ is summable.

iii) Without extra assumptions on the entries of $(a_n)_n$, all of the above subcases are relevant. Therefore, for a summably fast convergence of the variance we need $\alpha > \frac{1}{2}$ and the sequences $(\frac{1}{b_n^2})_n$, $(\frac{1}{b_n} D_n^{(l)})_n$ and $(C_n^{(l)})_n$ for all $l \in \mathbb{N}$ to be summable over n. In particular, we used condition (AAU3).

iv) Assuming independent entries in $(a_n)_n$ with existing moments, zero expectation and unit variance, we see that the term in (4.14) vanishes for each subsum in our case-by-case analysis except for "Step 2, Case 1" and "Step 2, Case 2", since if the edge sets of tuples $\underline{t}$ and $\underline{t}'$ are disjoint, we obtain $\mathbb{E}a_n^b(\underline{t})a_n^b(\underline{t}') = \mathbb{E}a_n^b(\underline{t})\mathbb{E}a_n^b(\underline{t}')$. For those two subcases just mentioned, we achieve a summable convergence to zero if $(\frac{1}{nb_n})_n$ is summable over n.

□

We have now proved Theorem 4.9:

Proof of Theorem 4.9. This is a direct consequence of Theorem 3.14 with $z = 2$, Theorem 4.38 and Theorem 4.40. □

4.4 Extension of Results to Non-Periodic Band Matrices

In this section, we will extend our results to non-periodic band matrices. In order to analyze these matrices in a sensible manner, the concept of a bandwidth should be replaced by the concept called *halfwidth*, which we adopted from [42]. Intuitively, the halfwidth $h = (h_n)_n$ should be interpreted as half of a bandwidth $b = (b_n)_n$, hence the name. A 6×6 non-periodic band matrix M with halfwidth 2 has the structure

$$M = \begin{pmatrix} x_{1,1} & x_{1,2} & 0 & 0 & 0 & 0 \\ x_{2,1} & x_{2,2} & x_{2,3} & 0 & 0 & 0 \\ 0 & x_{3,2} & x_{3,3} & x_{3,4} & 0 & 0 \\ 0 & 0 & x_{4,3} & x_{4,4} & x_{4,5} & 0 \\ 0 & 0 & 0 & x_{5,4} & x_{5,5} & x_{5,6} \\ 0 & 0 & 0 & 0 & x_{6,5} & x_{6,6} \end{pmatrix},$$

whereas with halfwidth 4 we obtain the structure

$$M = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & 0 & 0 \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & 0 \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & x_{3,6} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & x_{4,6} \\ 0 & x_{5,2} & x_{5,3} & x_{5,4} & x_{5,5} & x_{5,6} \\ 0 & 0 & x_{6,3} & x_{6,4} & x_{6,5} & x_{6,6} \end{pmatrix}.$$

The halfwidth denotes the number of possible non-trivial entries in the first row of the matrix. For an $n \times n$ matrix, a valid halfwidth is thus a number in the set $\{1, 2, \ldots, n\}$. For a visual comparison of the structure of periodic and non-periodic band matrices, the reader is encouraged to turn to page 56. We notice the difference between the periodic and the non-periodic case is that in the latter case the triangular areas in the upper right and lower left corner of the matrices are missing, leading to the possibility that the *inner band* is so wide that it reaches the top right and lower left corners of the matrix. Also, the rows of the matrix do not possess the same number of non-trivial entries any longer. This makes analysis of such matrices a little less straight-forward when applying the method of moments. We will use a different route to extend our results to the non-periodic case:

We will show that under reasonable conditions, the ESDs of periodic and non-periodic matrices are asymptotically equivalent, so that we can directly make use of our previous results. In order to compare these two types of band matrices, we need to relate the concepts halfwidth and bandwidth:

Definition 4.42. Let $n \in \mathbb{N}$ be arbitrary, then an $h_n \in \mathbb{N}$ is called *(n-)halfwidth*, if $h_n \in \{1 \ldots n\}$. By slight abuse of language, a sequence $h = (h_n)_n$ of halfwidths will also be called *halfwidth*. Given a halfwidth $h = (h_n)_n$, we set

$$\forall\, n \in \mathbb{N}: \; b_n := \min(2h_n - 1, n)$$

and call b_n (resp. $b = (b_n)_n$) the *bandwidth associated with the halfwidth* h_n (resp. $h = (h_n)_n$).

It is clear that for any $n \in \mathbb{N}$, the bandwidth b_n that is associated with a halfwidth h_n is either n itself or an odd number in the set $\{1, \ldots, n\}$, thus coincides with the concept of a bandwidth in previous sections. As before, b_n is the number of possible non-trivial entries in "the middle row" of the matrix, whereas h_n is the number of possible non-trivial entries in the first row.

Definition 4.43. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(a_n)_{n\in\mathbb{N}}$ a triangular scheme, $h = (h_n)_n$ be a sequence of n-halfwidths with associated bandwidths $b = (b_n)_n$.

1. We define the non-periodic random matrices which are based on the triangular scheme $(a_n)_{n\in\mathbb{N}}$ with halfwidth h as

$$\forall\, n \in \mathbb{N}: \forall\, (i,j) \in \square_n: \; X_n^{NP}(i,j) := \begin{cases} \frac{1}{\sqrt{b_n}} a_n(i,j) & \text{if } |i-j| \leq h_n - 1 \\ 0 & \text{otherwise.} \end{cases}$$

2. We define the periodic random matrices which are based on the triangular scheme $(a_n)_{n\in\mathbb{N}}$ with associated bandwidth b as

$$\forall\, n \in \mathbb{N}: \forall\, (i,j) \in \square_n: \; X_n^{P}(i,j) := \begin{cases} \frac{1}{\sqrt{b_n}} a_n(i,j) & \text{if } |i-j| \leq h_n - 1 \\ \frac{1}{\sqrt{b_n}} a_n(i,j) & \text{if } |i-j| \geq \max(n - h_n + 1, h_n) \\ 0 & \text{otherwise.} \end{cases}$$

Note that in the previous definition, we always normalize with the square root of the bandwidth. In particular, the normalization does not increase for $h_n \geq (n+1)/2$, despite the fact that (in the non-periodic case) more non-trivial entries are allowed. Note also that the definition of periodic random matrices has not changed in comparison to previous sections, see Definitions 4.6, 4.7 and 4.8 and the following lemma:

Lemma 4.44. *Let for some $n \in \mathbb{N}$, h_n be an n-halfwidth with associated bandwidth b_n. Then an index pair $(i,j) \in \square_n$ is b_n-relevant iff $|i-j| \leq h_n - 1$ or $|i-j| \geq \max(n - h_n + 1, h_n)$.*

Proof. By definition, $b_n = \min(2h_n - 1, n)$. We show "$\Rightarrow$" first. If (i,j) is b_n-relevant, then this means that

$$|i-j| \leq \frac{b_n - 1}{2} \quad \text{or} \quad |i-j| \geq n - \frac{b_n - 1}{2} \quad \text{or} \quad b_n = n. \tag{4.16}$$

Now if the first condition in (4.16) is satisfied, we calculate

$$|i-j| \leq \frac{b_n - 1}{2} = \frac{\min(2h_n - 1, n) - 1}{2} \leq h_n - 1.$$

If the second condition in (4.16) is satisfied (but not $b_n = n$, which we consider further below, in particular, we assume $b_n = 2h_n - 1$), the first one cannot be satisfied by transitivity of $\leq$ (since $b_n \leq n$) and hence

$$|i-j| > \frac{b_n - 1}{2} = \frac{2h_n - 1 - 1}{2} \geq h_n - 1,$$

which yields $|i-j| \geq h_n$. In addition, the second condition itself yields:

$$|i-j| \geq \frac{2n - b_n + 1}{2} = \frac{2n - (2h_n - 1) + 1}{2} = n - h_n + 1,$$

so in combination with $|i-j| \geq h_n$ we obtain $|i-j| \geq \max(n - h_n + 1, h_n)$. Now let the third condition in (4.16) be satisfied, that is, $b_n = n$ which implies $n \leq 2h_n - 1$ by definition of b_n, so that $h_n \geq n - h_n + 1$. Then if $\neg(|i-j| \leq h_n - 1)$, this entails $|i-j| \geq h_n \geq n - h_n + 1$, so indeed $|i-j| \geq \max(n - h_n + 1, h_n)$. This completes the direction "$\Rightarrow$." For "$\Leftarrow$" we assume that $b_n \neq n$, which entails that $b_n < n$, so $b_n = 2h_n - 1 < n$, hence $n - h_n + 1 > h_n$ and we must show that

$$|i-j| \leq \frac{b_n - 1}{2} \quad \text{or} \quad |i-j| \geq n - \frac{b_n - 1}{2}$$

given that $|i-j| \leq h_n - 1$ or $|i-j| \geq n - h_n + 1$. But if $|i-j| \leq h_n - 1$, we find

$$|i-j| \leq h_n - 1 = \frac{b_n - 1}{2},$$

and if $|i-j| \geq n - h_n + 1$, we obtain

$$|i-j| \geq n - h_n + 1 = n - (h_n - 1) = n - \frac{b_n - 1}{2}.$$

$\square$

We have secured that Theorem 4.9 and its corollaries are applicable to periodic band matrices X_n^P which are based on bandwidths that are associated with halfwidths. We want to study next in which way those results can be extended to the non-periodic versions X_n^{NP}. In addition, since the significant parameter in the non-periodic case is the halfwidth of which the bandwidth is just a derivate, we should also formulate the results for non-periodic band matrices in terms of the halfwidth.

The question we should ask ourselves is what kind of results we can hope for. Bogachev et al. have shown in [11] that for the i.i.d. case, the semicircle law holds in probability for $(X_n^{NP})_n$ if

$$\lim_{n\to\infty} h_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{h_n}{n} \in \{0,1\}, \tag{4.17}$$

whereas the semicircle law does not hold if $\lim_n h_n/n = p$ for some $p \in (0,1)$.

Therefore, our analysis which follows will always impose condition (4.17), since we are interested in convergence to the semicircle distribution (and the i.i.d. case is a special case of ours). Now instead of applying the method of moments once again, we realize that if (4.17) holds, X_n^{NP} and X_n^P will only differ in small triangles in the upper right and lower left corners of the matrix. To visualize this effect for small fractions h_n/n, assume that $n = 10$, $h_n = 3$ and thus $b_n = 5$, yielding the following structures for X_{10}^{NP} and X_{10}^P:

$$X_{10}^{NP} = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & 0 & 0 & 0 & 0 & 0 & 0 \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & 0 & 0 & 0 & 0 & 0 \\ 0 & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & x_{4,6} & 0 & 0 & 0 & 0 \\ 0 & 0 & x_{5,3} & x_{5,4} & x_{5,5} & x_{5,6} & x_{5,7} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{6,4} & x_{6,5} & x_{6,6} & x_{6,7} & x_{6,8} & 0 & 0 \\ 0 & 0 & 0 & 0 & x_{7,5} & x_{7,6} & x_{7,7} & x_{7,8} & x_{7,9} & 0 \\ 0 & 0 & 0 & 0 & 0 & x_{8,6} & x_{8,7} & x_{8,8} & x_{8,9} & x_{8,10} \\ 0 & 0 & 0 & 0 & 0 & 0 & x_{9,7} & x_{9,8} & x_{9,9} & x_{9,10} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{10,8} & x_{10,9} & x_{10,10} \end{pmatrix}$$

$$X_{10}^{P} = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & 0 & 0 & 0 & 0 & 0 & x_{1,9} & x_{1,10} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & 0 & 0 & 0 & 0 & 0 & x_{2,10} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & 0 & 0 & 0 & 0 & 0 \\ 0 & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & x_{4,6} & 0 & 0 & 0 & 0 \\ 0 & 0 & x_{5,3} & x_{5,4} & x_{5,5} & x_{5,6} & x_{5,7} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{6,4} & x_{6,5} & x_{6,6} & x_{6,7} & x_{6,8} & 0 & 0 \\ 0 & 0 & 0 & 0 & x_{7,5} & x_{7,6} & x_{7,7} & x_{7,8} & x_{7,9} & 0 \\ 0 & 0 & 0 & 0 & 0 & x_{8,6} & x_{8,7} & x_{8,8} & x_{8,9} & x_{8,10} \\ x_{9,1} & 0 & 0 & 0 & 0 & 0 & x_{9,7} & x_{9,8} & x_{9,9} & x_{9,10} \\ x_{10,1} & x_{10,2} & 0 & 0 & 0 & 0 & 0 & x_{10,8} & x_{10,9} & x_{10,10} \end{pmatrix}$$

On the other hand, for large fractions h_n/n, we assume $n = 10$, $h_n = 8$ and thus $b_n = 10$, and we obtain the following structures for X_{10}^{NP} and X_{10}^P:

$$X_{10}^{NP} = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & x_{1,6} & x_{1,7} & x_{1,8} & 0 & 0 \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & x_{2,6} & x_{2,7} & x_{2,8} & x_{2,9} & 0 \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & x_{3,6} & x_{3,7} & x_{3,8} & x_{3,9} & x_{3,10} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & x_{4,6} & x_{4,7} & x_{4,8} & x_{4,9} & x_{4,10} \\ x_{5,1} & x_{5,2} & x_{5,3} & x_{5,4} & x_{5,5} & x_{5,6} & x_{5,7} & x_{5,8} & x_{5,9} & x_{5,10} \\ x_{6,1} & x_{6,2} & x_{6,3} & x_{6,4} & x_{6,5} & x_{6,6} & x_{6,7} & x_{6,8} & x_{6,9} & x_{6,10} \\ x_{7,1} & x_{7,2} & x_{7,3} & x_{7,4} & x_{7,5} & x_{7,6} & x_{7,7} & x_{7,8} & x_{7,9} & x_{7,10} \\ x_{8,1} & x_{8,2} & x_{8,3} & x_{8,4} & x_{8,5} & x_{8,6} & x_{8,7} & x_{8,8} & x_{8,9} & x_{8,10} \\ 0 & x_{9,2} & x_{9,3} & x_{9,4} & x_{9,5} & x_{9,6} & x_{9,7} & x_{9,8} & x_{9,9} & x_{9,10} \\ 0 & 0 & x_{10,3} & x_{10,4} & x_{10,5} & x_{10,6} & x_{10,7} & x_{10,8} & x_{10,9} & x_{10,10} \end{pmatrix}$$

$$X_{10}^{P} = \begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} & x_{1,5} & x_{1,6} & x_{1,7} & x_{1,8} & x_{1,9} & x_{1,10} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} & x_{2,5} & x_{2,6} & x_{2,7} & x_{2,8} & x_{2,9} & x_{2,10} \\ x_{3,1} & x_{3,2} & x_{3,3} & x_{3,4} & x_{3,5} & x_{3,6} & x_{3,7} & x_{3,8} & x_{3,9} & x_{3,10} \\ x_{4,1} & x_{4,2} & x_{4,3} & x_{4,4} & x_{4,5} & x_{4,6} & x_{4,7} & x_{4,8} & x_{4,9} & x_{4,10} \\ x_{5,1} & x_{5,2} & x_{5,3} & x_{5,4} & x_{5,5} & x_{5,6} & x_{5,7} & x_{5,8} & x_{5,9} & x_{5,10} \\ x_{6,1} & x_{6,2} & x_{6,3} & x_{6,4} & x_{6,5} & x_{6,6} & x_{6,7} & x_{6,8} & x_{6,9} & x_{6,10} \\ x_{7,1} & x_{7,2} & x_{7,3} & x_{7,4} & x_{7,5} & x_{7,6} & x_{7,7} & x_{7,8} & x_{7,9} & x_{7,10} \\ x_{8,1} & x_{8,2} & x_{8,3} & x_{8,4} & x_{8,5} & x_{8,6} & x_{8,7} & x_{8,8} & x_{8,9} & x_{8,10} \\ x_{9,1} & x_{9,2} & x_{9,3} & x_{9,4} & x_{9,5} & x_{9,6} & x_{9,7} & x_{9,8} & x_{9,9} & x_{9,10} \\ x_{10,1} & x_{10,2} & x_{10,3} & x_{10,4} & x_{10,5} & x_{10,6} & x_{10,7} & x_{10,8} & x_{10,9} & x_{10,10} \end{pmatrix}$$

To measure the difference between spectral distributions, we employ the bounded Lipschitz metric $\mathrm{d_{BL}}$, which can be used particularly well to analyze the effect of small perturbations of random matrices. Please turn to Section 4.5 for details and notational conventions, which we will use in passing in what follows. The main reason this metric is so convenient is that the distance between the ESDs of two random matrices M_1 and M_2 is expressed through the entries of the difference matrix $M_1 - M_2$ (see Lemma 4.52). Comparing our non-periodic and periodic band matrices given some halfwidth h and associated bandwidth b (as we just visualized, but see also Definition 4.43), we realize that both matrices contain a non-trivial area with indices $|i-j| \leq h_n - 1$, which is the band in the middle of the matrix, and additionally, periodic matrices contain non-trivial triangular areas with indices $|i-j| \geq \max(n-h_n+1, h_n)$.

To employ the bounded Lipschitz metric, it is useful to count the index pairs that either point to the common band in the middle of the matrix or to the entries in the triangular areas that make up the difference. Since we only consider real symmetric random matrices and the main diagonal is trivial to deal with (and will never be part of any of the mentioned triangular areas), for a given halfwidth $(h_n)_n$ we set for all $n \in \mathbb{N}$:

$$B(h_n) := \{(i,j) \,|\, 1 \leq i < j \leq n,\, j-i \leq h_n - 1\}$$

and

$$T(h_n) := \{(i,j) \,|\, 1 \leq i < j \leq n,\, j-i \geq \max(n-h_n+1, h_n)\}\,,$$

where "B" stands for "band" and "T" for "triangular area." These sets are obviously disjoint and we will denote their union by $S(h_n)$, so $S(h_n) := B(h_n) \cup T(h_n)$.

Lemma 4.45. *Let $n \in \mathbb{N}$ and $h_n \in \{1, \dots, n\}$, then*

i) $\#B(h_n) = \frac{(h_n-1)(2n-h_n)}{2}$.

ii) $\#T(h_n) = \min\left(\frac{(h_n-1)h_n}{2}, \frac{(n-h_n)(n-h_n+1)}{2}\right)$.

iii) $\#S(h_n) = \min\left(n(h_n-1), \frac{n(n-1)}{2}\right)$.

Proof. i) $B(h_n)$ contains all index pairs (i,j) corresponding the diagonals where $j - i \in \{1, \dots, h_n - 1\}$. Their count is

$$\sum_{k=1}^{h_n-1}(n-k) = n(h_n-1) - \sum_{k=1}^{h_n-1} k = n(h_n-1) - \frac{(h_n-1)h_n}{2} = \frac{(h_n-1)(2n-h_n)}{2}.$$

ii) To count the index pairs in $T(h_n)$, let $k \in \{1, \dots, n\}$ be fixed and let us count the index pairs in the set

$$\{1 \le i < j \le n \,|\, j - i \ge k\}.$$

To this end, if $j = n$, then we can pick $i \in \{1, \dots, n-k\}$, if $j = n-1$ we can pick $i \in \{1, \dots, n-k-1\}$ and so on until finally, if $j = k+1$, we can pick $i \in \{1\}$. Therefore, with $k = \max(n - h_n + 1, h_n)$, hence $n - k = \min(h_n - 1, n - h_n)$, we calculate

$$\sum_{l=1}^{n-k} l = \sum_{l=1}^{\min(h_n-1,n-h_n)} l = \min\left(\sum_{l=1}^{h_n-1} l, \sum_{l=1}^{n-h_n} l\right) = \min\left(\frac{(h_n-1)h_n}{2}, \frac{(n-h_n)(n-h_n+1)}{2}\right)$$

index pairs in $T(h_n)$.
iii) We calculate, using that $B(h_n)$ and $T(h_n)$ are disjoint,

$$\begin{aligned}
\#S(h_n) &= \#(B(h_n) \cup T(h_n)) \\
&= \#B(h_n) + \#T(h_n) \\
&= \frac{(h_n-1)(2n-h_n)}{2} + \min\left(\frac{(h_n-1)h_n}{2}, \frac{(n-h_n)(n-h_n+1)}{2}\right) \\
&= \min\left(\frac{2n(h_n-1)}{2}, \frac{n(h_n-1) + (h_n-1)(n-h_n) + (n-h_n)(n-(h_n-1))}{2}\right) \\
&= \min\left(n(h_n-1), \frac{n(h_n-1) + n(n-h_n)}{2}\right) \\
&= \min\left(n(h_n-1), \frac{n(n-1)}{2}\right).
\end{aligned}$$

$\square$

Having studied the areas of non-trivial entries away from the main diagonal of periodic and non-periodic band matrices, we now turn to the statement that the conditions $h_n \to \infty$ and $\lim_n h_n/n \in \{0,1\}$ render the ESDs of these matrices asymptotically equivalent. The theorem that follows is the main tool of this section and might also be of independent interest, since it relates the asymptotic ESDs of period and non-periodic random matrices in quite general settings. One of the statements in the theorem concerns sub-Gaussian entries. We refer the reader to Section 4.5 for details on both the metric d_{BL} and sub-Gaussian random variables.

Theorem 4.46. *Let $(a_n)_n$ be a triangular scheme and let $(h_n)_n$ be a halfwidth with $h_n \to \infty$ and $\lim_n h_n/n \in \{0,1\}$. Denote by X_n^P the periodic and by X_n^{NP} the non-periodic random matrices which are based on a_n with halfwidth h and associated bandwidth b. Then the following statements hold:*

i) If the entries of $(a_n)_n$ possess uniformly bounded second moments, then

$$\mathrm{d_{BL}}(X_n^P, X_n^{NP}) \xrightarrow[n\to\infty]{} 0 \qquad \text{in probability.}$$

ii) If the entries of $(a_n)_n$ are uniformly bounded, then

$$\mathrm{d_{BL}}(X_n^P, X_n^{NP}) \xrightarrow[n\to\infty]{} 0 \qquad \text{surely.}$$

iii) If for all $n \in \mathbb{N}$, the family of random variables $(a_n(i,j))_{1\leq i\leq j\leq n}$ is independent, and if all variables in $(a_n)_n$ possess uniformly bounded eighth moments, then

$$\mathrm{d_{BL}}(X_n^P, X_n^{NP}) \xrightarrow[n\to\infty]{} 0 \qquad \text{almost surely.}$$

iv) If the entries of $(a_n)_n$ are β-sub-Gaussian for some $\beta > 0$, and if there exist $C, d > 1$ such that

$$\forall\, n \in \mathbb{N}:\ \min\left(1 - \frac{h_n}{n}, \frac{h_n}{n}\right) \leq \frac{C}{\log^d(n)},$$

where $\log^d(n) = (\log(n))^d$, then

$$\mathrm{d_{BL}}(X_n^P, X_n^{NP}) \xrightarrow[n\to\infty]{} 0 \qquad \text{almost surely.}$$

Before we start with the proof, let us analyze the condition on the halfwidth in statement *iv*). Knowing already that $\lim_n h_n/n \in \{0,1\}$, this condition lets us conclude that the convergence takes place with some polylogarithmic convergence speed. An important special case is that $h_n/n^\rho \to 1$ for some $\rho \in (0,1)$. Then we find

$$\frac{h_n}{n} = \frac{h_n}{n^\rho} \cdot \frac{1}{n^{1-\rho}} \leq \frac{C'}{n^\gamma} \leq \frac{C}{\log^d(n)}$$

where $\gamma = 1 - \rho > 0$, $C' > 0$ is a bound for the convergent sequence h_n/n^ρ, $d > 1$ can actually be chosen arbitrarily, and C is a bound for the bounded sequence $(C' \log^d(n)/n^\gamma)_n$. Therefore, the case $h_n/n^\rho \to 1$ for some $\rho \in (0,1)$ is covered by the theorem.

Proof of Theorem 4.46. We start by making the following calculation, using Lemma 4.52:

$$\begin{aligned}\mathrm{d_{BL}}(X_n^P, X_n^{NP})^2 &\leq \frac{1}{n}\operatorname{tr}\left[(X_n^P - X_n^{NP})^*(X_n^P - X_n^{NP})\right]\\ &= \frac{1}{n}\sum_{(i,j)\in\square_n}\left|(X_n^{NP} - X_n^P)(i,j)\right|^2 \qquad (4.18)\\ &= \frac{2}{nb_n}\sum_{(i,j)\in T(h_n)}|a_n(i,j)|^2.\end{aligned}$$

We also note the following: If $h_n/n \to 0$, we will have $b_n = \min(2h_n - 1, n) = 2h_n - 1$ finally, since

$$2h_n - 1 \leq n \Leftrightarrow \frac{2h_n}{n} - \frac{1}{n} \leq 1,$$

and the latter is finally true. (Here and throughout the whole proof, the word "finally" is used exclusively to mean "for all $n \in \mathbb{N}$ with $n \geq N$, where $N \in \mathbb{N}$ depends only on the specific choice of the halfwidth h".) On the other hand, if $h_n/n \to 1$, we will have $b_n = \min(2h_n - 1, n) = n$ finally, since

$$2h_n - 1 \geq n \Leftrightarrow \frac{2h_n}{n} - \frac{1}{n} \geq 1,$$

and the latter is finally true. Therefore, in our calculations below, whenever we consider the cases $h_n/n \to 0$ or $h_n/n \to 1$, we may and will replace b_n by $2h_n - 1$ or n, respectively, without further notice.
Statement $i)$ Let $m_2 \geq 0$ denote a uniform bound of the second moment of the entries of $(a_n)_n$. Then for any $\epsilon > 0$, we find using (4.18) and Lemma 4.45 $ii)$, that

$$\begin{aligned}\mathbb{P}(\mathrm{d_{BL}}(X_n^P, X_n^{NP}) > \epsilon) &\leq \frac{\mathbb{E}\,\mathrm{d_{BL}}(X_n^P, X_n^{NP})^2}{\epsilon^2}\\ &\leq \frac{2m_2}{\epsilon^2 nb_n}\cdot\min\left(\frac{(h_n-1)h_n}{2}, \frac{(n-h_n)(n-h_n+1)}{2}\right),\end{aligned}$$

Now if $h_n/n \to 0$, we will finally have

$$\mathbb{P}(\mathrm{d_{BL}}(X_n^P, X_n^{NP}) > \epsilon) \leq \frac{2m_2}{\epsilon^2 n(2h_n - 1)}\frac{(h_n-1)h_n}{2} = \frac{m_2}{\epsilon^2}\underbrace{\frac{h_n - 1}{2h_n - 1}}_{\leq 1/2}\frac{h_n}{n} \xrightarrow[n\to\infty]{} 0,$$

whereas if $h_n/n \to 1$ we will finally have

$$\begin{aligned}\mathbb{P}(\mathrm{d_{BL}}(X_n^P, X_n^{NP}) > \epsilon) &\leq \frac{2m_2}{\epsilon^2 n^2}\frac{(n-h_n)(n-h_n+1)}{2}\\ &= \frac{m_2}{\epsilon^2}\left(1 - \frac{h_n}{n}\right)\left(1 - \frac{h_n}{n} + \frac{1}{n}\right) \xrightarrow[n\to\infty]{} 0.\end{aligned}$$

So indeed, $\mathrm{d}_{\mathrm{BL}}(X_n^P, X_n^{NP}) \to 0$ in probability, which is exactly statement i).
Statement ii) Denote by $C \geq 0$ a uniform bound of the entries of $(a_n)_n$. Then, using Lemma 4.45 ii) and (4.18), we obtain

$$\begin{aligned} \mathrm{d}_{\mathrm{BL}}(X_n^P, X_n^{NP})^2 &\leq \frac{2}{nb_n} \sum_{(i,j)\in T(h_n)} |a_n(i,j)|^2 \\ &\leq \frac{2C^2}{nb_n} \min\left(\frac{(h_n-1)h_n}{2}, \frac{(n-h_n)(n-h_n+1)}{2}\right). \end{aligned}$$

Now if $h_n/n \to 0$, we conclude that finally

$$\mathrm{d}_{\mathrm{BL}}(X_n^P, X_n^{NP})^2 \leq \frac{2C^2}{n(2h_n-1)} \frac{(h_n-1)h_n}{2} = C^2 \frac{h_n-1}{2h_n-1} \frac{h_n}{n} \xrightarrow[n\to\infty]{} 0,$$

whereas if $h_n/n \to 1$, we conclude that finally

$$\mathrm{d}_{\mathrm{BL}}(X_n^P, X_n^{NP})^2 \leq \frac{2C^2}{n^2} \frac{(n-h_n)(n-h_n+1)}{2} = C^2 \left(1 - \frac{h_n}{n}\right)\left(1 - \frac{h_n}{n} + \frac{1}{n}\right) \xrightarrow[n\to\infty]{} 0,$$

so in particular, $\mathrm{d}_{\mathrm{BL}}(X_n^P, X_n^{NP}) \to 0$ surely.
Statement iii) For a truncation level $T > 0$, denote by $X_n^{P,T}$ and $X_n^{NP,T}$ the truncated versions of X_n^P and X_n^{NP}, that is, for all $n \in \mathbb{N}$ and $(i,j) \in \square_n$, set

$$X_n^{P,T}(i,j) := X_n^P(i,j)\mathbb{1}_{\{|\sqrt{b_n}X_n^P(i,j)| \leq T\}}$$

and

$$X_n^{NP,T}(i,j) := X_n^{NP}(i,j)\mathbb{1}_{\{|\sqrt{b_n}X_n^{NP}(i,j)| \leq T\}}.$$

Since the entries of $(a_n)_n$ have uniformly bounded eighth moments, the family $((|a_n(i,j)|^2)_{(i,j)\in\square_n})_{n\in\mathbb{N}}$ has uniformly bounded fourth moments and is thus uniformly integrable. In particular, for any $\epsilon > 0$ we can find a truncation level $T > 0$ such that for all $n \in \mathbb{N}$ and $(i,j) \in \square_n$ we have $\mathbb{E}|a_n(i,j)|^2\mathbb{1}_{\{|a_n(i,j)|>T\}} \leq \epsilon^2$. We show that then the following holds:

1. $\limsup_{n\to\infty} d_{BL}(X_n^P, X_n^{P,T}) \leq \epsilon$ almost surely,
2. $\limsup_{n\to\infty} d_{BL}(X_n^{NP}, X_n^{NP,T}) \leq \epsilon$ almost surely.

To this end, we calculate using Lemma 4.52

$$\begin{aligned} d_{BL}(X_n^P, X_n^{P,T})^2 &\leq \frac{1}{nb_n} \sum_{(i,j)\in\square_n} |\sqrt{b_n}X_n^P(i,j)|^2\mathbb{1}_{\{|\sqrt{b_n}X_n^P(i,j)|>T\}} \\ &= \frac{1}{nb_n}\sum_{i=1}^n |a_n(i,i)|^2\mathbb{1}_{\{|a_n(i,i)|>T\}} + \frac{2}{nb_n}\sum_{(i,j)\in S(h_n)} |a_n(i,j)|^2\mathbb{1}_{\{|a_n(i,j)|>T\}} \qquad (4.19) \\ &=: A_n^{(1)} + A_n^{(2)}. \end{aligned}$$

We will handle these two series separately and invoke the SLLN as in Theorem 4.51 each time. First of all, since $h_n \to \infty$, also $b_n \to \infty$ and with

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |a_n(i,i)|^2 \mathbb{1}_{\{|a_n(i,i)|>T\}} \leq \epsilon^2$$

almost surely by the SLLN, we obtain

$$A_n^{(1)} \xrightarrow[n\to\infty]{} 0 \qquad \text{almost surely.} \tag{4.20}$$

For $A_n^{(2)}$ we obtain via Lemma 4.45 *iii)* that

$$\frac{2\cdot \#S(h_n)}{nb_n} = \frac{2}{nb_n} \min\left(n(h_n-1), \frac{n(n-1)}{2}\right) = \frac{1}{b_n} \min\left(2(h_n-1), n-1\right).$$

Now if $h_n/n \to 0$, we finally have $b_n = 2h_n - 1$ and $\min(2(h_n-1), n-1) = 2(h_n-1)$, so that finally

$$\frac{2\cdot \#S(h_n)}{nb_n} = \frac{2(h_n-1)}{2h_n-1} \xrightarrow[n\to\infty]{} 1.$$

On the other hand, if $h_n/n \to 1$, we finally have $b_n = n$ and $\min(2(h_n-1), n-1) = n-1$, so that finally

$$\frac{2\cdot \#S(h_n)}{nb_n} = \frac{n-1}{n} \xrightarrow[n\to\infty]{} 1.$$

Therefore, in either case we have $2\cdot \#S(h_n)/(nb_n) \to 1$ and since finally

$$\#S(h_n) = \min\left(n(h_n-1), \frac{n(n-1)}{2}\right) \geq n, \quad \text{thus} \quad \frac{1}{(\#S(h_n))^2} \leq \frac{1}{n^2},$$

$1/(\#S(h_n))^2$ is finally summable (cf. Definition 4.50). Therefore, using the SLLN once again we obtain

$$\begin{aligned}\limsup_{n\to\infty} A_n^{(2)} &= \limsup_{n\to\infty} \frac{2}{nb_n} \sum_{(i,j)\in S(h_n)} |a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T\}} \\ &= \limsup_{n\to\infty} \frac{2\cdot \#S(h_n)}{nb_n} \frac{1}{\#S(h_n)} \sum_{(i,j)\in S(h_n)} |a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T\}} \leq \epsilon^2\end{aligned}$$

almost surely. Therefore,

$$\limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^P, X_n^{P,T}) \leq \epsilon \qquad \text{almost surely}$$

for all T so large that $\mathbb{E}|a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T\}} \leq \epsilon^2$ for all $n \in \mathbb{N}$ and $(i,j) \in \square_n$.

To treat the non-periodic case, we calculate

$$\begin{aligned}
\mathrm{d_{BL}}(X_n^{NP}, X_n^{NP,T})^2 &\leq \frac{1}{nb_n}\sum_{i,j=1}^{n} |\sqrt{b_n}X_n^{NP}(i,j)|^2 \mathbb{1}_{\{|\sqrt{b_n}X_n^{NP}(i,j)|>T\}} \\
&= \frac{1}{nb_n}\sum_{i=1}^{n} |a_n(i,i)|^2 \mathbb{1}_{\{|a_n(i,i)|>T\}} + \frac{2}{nb_n}\sum_{(i,j)\in B(b_n)} |a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T\}} \\
&\leq \frac{1}{nb_n}\sum_{i=1}^{n} |a_n(i,i)|^2 \mathbb{1}_{\{|a_n(i,i)|>T\}} + \frac{2}{nb_n}\sum_{(i,j)\in S(b_n)} |a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T\}} \\
&= A_n^{(1)} + A_n^{(2)}.
\end{aligned}$$

As we just argued above, it holds

$$A_n^{(1)} \xrightarrow[n\to\infty]{} 0 \qquad \text{almost surely.}$$

and

$$\limsup_{n\to\infty} A_n^{(2)} \leq \epsilon^2 \qquad \text{almost surely,}$$

so that

$$\limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^{NP}, X_n^{NP,T}) \leq \epsilon \qquad \text{almost surely.}$$

In total, we achieved that for all $\epsilon > 0$ we find a $T > 0$ such that

$$\begin{aligned}
&\limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^{NP}, X_n^{P}) \\
&\leq \limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^{NP}, X_n^{NP,T}) + \limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^{NP,T}, X_n^{P,T}) + \limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^{P,T}, X_n^{P}) \\
&\leq \epsilon + 0 + \epsilon = 2\epsilon \qquad \text{almost surely.}
\end{aligned}$$

where we used statement *ii)* for the summand in the middle. Since $\epsilon > 0$ was arbitrary, for each $k \in \mathbb{N}$ we can find a set $A_k \in \mathcal{A}$ of measure 1 where

$$\limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^{P}, X_n^{NP}) \leq \frac{1}{k}.$$

Then $A_\infty := \bigcap_{k\in\mathbb{N}} A_k$ has measure one and

$$\limsup_{n\to\infty} \mathrm{d_{BL}}(X_n^{P}, X_n^{NP}) = 0 \qquad \text{on } A_\infty.$$

Thus,

$$\lim_{n\to\infty} \mathrm{d_{BL}}(X_n^{P}, X_n^{NP}) = 0 \qquad \text{almost surely.}$$

Statement *iv)* We proceed similarly as in statement *iii)*, but must be careful: We do not have a powerful SLLN at our disposal for sub-Gaussian random variables with arbitrary correlation structures. Therefore, we must follow a different idea, which is comprised of an increasing truncation level $T_n \nearrow \infty$ and the use of tail probability bounds witnessed by the sub-Gaussian property. Given $\beta > 0$ as in statement *iv*), we set $T_n := 4\beta\sqrt{\log(n)}$ as our dynamic truncation parameter. Using (4.18) as above and Lemma 4.45 *ii*), we obtain finally

$$\begin{aligned}
\mathrm{d}_{\mathrm{BL}}(X_n^{P,T_n}, X_n^{NP,T_n})^2 &\leq \frac{2}{nb_n} \sum_{(i,j)\in T(h_n)} |a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|\leq T_n\}}. \\
&\leq \frac{2}{nb_n} \cdot \min\left(\frac{(h_n-1)h_n}{2}, \frac{(n-h_n)(n-h_n+1)}{2}\right) T_n^2.
\end{aligned}$$

We note that by choice of T_n, $T_n^2/\log^d(n) \to 0$ as $n \to \infty$, since $d > 1$. Now if $h_n/n \to 0$, using the convergence speed as in the statement of *iv*), we conclude that finally,

$$\begin{aligned}
\mathrm{d}_{\mathrm{BL}}(X_n^{P,T_n}, X_n^{NP,T_n})^2 &\leq \frac{2}{n(2h_n-1)} \frac{(h_n-1)h_n}{2} \cdot T_n^2 \\
&= \frac{h_n-1}{2h_n-1} \cdot \frac{h_n}{n} \cdot T_n^2 \\
&\leq \frac{h_n-1}{2h_n-1} \frac{C}{\log^d(n)} \cdot T_n^2 \xrightarrow[n\to\infty]{} 0 \quad \text{surely.}
\end{aligned}$$

On the other hand, if $h_n/n \to 1$ we conclude that finally

$$\begin{aligned}
\mathrm{d}_{\mathrm{BL}}(X_n^{P,T_n}, X_n^{NP,T_n})^2 &\leq \frac{2}{n^2} \frac{(n-h_n)(n-h_n+1)}{2} \cdot T_n^2 \\
&= \left(1 - \frac{h_n}{n}\right)\left(1 - \frac{h_n}{n} + \frac{1}{n}\right) \cdot T_n^2 \\
&\leq \frac{C}{\log^d(n)} \left(\frac{C}{\log^d(n)} + \frac{1}{n}\right) \cdot T_n^2 \xrightarrow[n\to\infty]{} 0 \quad \text{surely.}
\end{aligned}$$

Therefore, it suffices to show that

1. $\limsup_{n\to\infty} d_{BL}(X_n^P, X_n^{P,T_n}) \xrightarrow[n\to\infty]{} 0$ almost surely,
2. $\limsup_{n\to\infty} d_{BL}(X_n^{NP}, X_n^{NP,T_n}) \xrightarrow[n\to\infty]{} 0$ almost surely.

By Markov's inequality and the Borel-Cantelli lemma, it suffices to show

I. $\mathbb{E}\,\mathrm{d}_{\mathrm{BL}}(X^P, X^{P,T_n})^2 \to 0$ summably fast.

II. $\mathbb{E}\,\mathrm{d}_{\mathrm{BL}}(X^{NP}, X^{NP,T_n})^2 \to 0$ summably fast.

To show I. we obtain finally, using (4.19), the Cauchy-Schwarz inequality, Lemma 4.45 *iii*) and Theorem 4.53:

$$\begin{aligned}
&\mathbb{E}\,\mathrm{d}_{\mathrm{BL}}(X_n^P, X_n^{P,T_n})^2 \\
&\leq \frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}|a_n(i,i)|^2 \mathbb{1}_{\{|a_n(i,i)|>T_n\}} + \frac{2}{nb_n}\sum_{(i,j)\in S(h_n)} \mathbb{E}|a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T_n\}} \\
&\leq \frac{1}{nb_n}\sum_{i=1}^n \sqrt{\mathbb{E}|a_n(i,i)|^4}\cdot\sqrt{\mathbb{P}\left(|a_n(i,i)|>T_n\right)} \\
&\quad + \frac{2}{nb_n}\sum_{(i,j)\in S(h_n)} \sqrt{\mathbb{E}|a_n(i,i)|^4}\cdot\sqrt{\mathbb{P}\left(|a_n(i,j)|>T_n\right)} \\
&\leq \frac{\sqrt{2}\sqrt{m_4}}{b_n}\cdot e^{-\frac{T_n^2}{4\beta^2}} + \frac{2\sqrt{2}\sqrt{m_4}}{nb_n}\min\left(n(h_n-1), \frac{n(n-1)}{2}\right)\cdot e^{-\frac{T_n^2}{4\beta^2}},
\end{aligned} \tag{4.21}$$

where m_4 denotes a fourth moment bound of the entries. To be more precise, if Y is β-sub-Gaussian, then by Theorem 4.53,

$$\mathbb{E}Y^4 = \int_0^\infty \mathbb{P}(Y^4 > t)\mathrm{d}t = \int_0^\infty \mathbb{P}(|Y| > t^{\frac{1}{4}})\mathrm{d}t \leq 2\int_0^\infty e^{-\frac{\sqrt{t}}{2\beta^2}}\mathrm{d}t =: m_4 < \infty,$$

where the finiteness of the integral is easily seen since for all t large enough we obtain $e^{-\frac{\sqrt{t}}{2\beta^2}} \leq \frac{1}{t^2}$. Now to argue the summability of (4.21), note that $e^{-T_n^2/(4\beta^2)}$ is summable in n since by choice of $T_n = 4\beta\sqrt{\log(n)}$,

$$e^{-\frac{T_n^2}{4\beta^2}} = e^{-4\log(n)} = \frac{1}{n^4}.$$

Therefore, only summability of the second summand on the r.h.s. of (4.21) must be argued. But this is straight-forward with arguments we have used before: Assuming $h_n/n \to 0$, we may finally substitute b_n by $2h_n - 1$ and then consider the first term in the minimum. On the other hand, if $h_n/n \to 1$, we may finally substitute b_n by n and then consider the second term on the minimum. In both cases, the summability of $e^{-T_n^2/(4\beta^2)}$ will then ensure summability of the summand in question.

Analogously, for II. we calculate

$$\begin{aligned}
&\mathbb{E}d_{BL}(X_n^{NP}, X_n^{NP,T_n})^2 \\
&\leq \frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}|a_n(i,i)|^2 \mathbb{1}_{\{|a_n(i,i)|>T_n\}} + \frac{2}{nb_n}\sum_{(i,j)\in B(h_n)} \mathbb{E}|a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T_n\}} \\
&\leq \frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}|a_n(i,i)|^2 \mathbb{1}_{\{|a_n(i,i)|>T_n\}} + \frac{2}{nb_n}\sum_{(i,j)\in S(h_n)} \mathbb{E}|a_n(i,j)|^2 \mathbb{1}_{\{|a_n(i,j)|>T_n\}}
\end{aligned}$$

and we have just seen that this expression converges to zero summably fast. □

We will derive the following theorem for the non-periodic case:

Theorem 4.47. *Let $(a_n)_n$ be an α-almost-uncorrelated triangular scheme, $h = (h_n)_n$ a sequence of n-halfwidths and $(X_n^{NP})_n$ the non-periodic random matrices which are based on $(a_n)_n$ with halfwidth h. We assume that*

$$h_n \to \infty \qquad \textit{but} \qquad \lim_{n\to\infty} \frac{h_n}{n} \in \{0,1\}.$$

Then we obtain the following results:

1. *If $\alpha \geq \frac{1}{2}$, then the semicircle law holds for $(X_n^{NP})_n$ in probability.*
2. *If $\alpha \geq \frac{1}{2}$, $\frac{1}{h_n^3}$ is summable over n and all entries of $(a_n)_n$ are $\{-1,1\}$-valued, then the semicircle law holds almost surely for $(X_n^{NP})_n$.*
3. *If $(a_n)_{n\in\mathbb{N}}$ is a Wigner scheme (Def. 2.40) and if $(\frac{1}{nh_n})_n$ is summable, then we obtain the semicircle law almost surely for $(X_n^{NP})_n$.*

Proof. First of all, note that if $h_n \to \infty$, also $b_n \to \infty$. The theorem is then a direct implication of Theorem 4.46 and Theorem 4.9. To give details, 1. follows with the fact that entries in α-almost uncorrelated triangular schemes have uniformly bounded second moments, 2. follows with the fact that

$$\frac{1}{b_n^3} \text{ is summable} \quad \Leftrightarrow \quad \frac{1}{h_n^3} \text{ is summable} \tag{4.22}$$

and 3. follows with the fact that

$$\frac{1}{nb_n} \text{ is summable} \quad \Leftrightarrow \quad \frac{1}{nh_n} \text{ is summable,} \tag{4.23}$$

where (4.22) and (4.23) follow easily since there exists a $c \geq 1$ such that finally,

$$\frac{1}{c}h_n \leq b_n \leq ch_n. \tag{4.24}$$

To see this, we know from the proof of Theorem 4.46 that if $h_n/n \to 0$, then $b_n = 2h_n - 1$ finally and then (4.24) holds with $c = 2$. If $h_n/n \to 1$, then $b_n = n$ finally and we obtain again that (4.24) holds with $c = 2$, since $h_n \in \{1, \ldots, n\}$ and

$$n \leq 2h_n \quad \Leftrightarrow \quad \frac{1}{2} \leq \frac{h_n}{n},$$

which is finally true since $h_n/n \to 1$.

□

The next corollary will deal with non-periodic random band matrices with Curie-Weiss entries:

Corollary 4.48. *Let $0 < \beta \leq 1$ and let for each $n \in \mathbb{N}$ the random variables $\tilde{a}_n(i,j)_{1\leq i,j\leq n}$ be Curie-Weiss(β, n^2)-distributed. Define the triangular scheme $(a_n)_n$ by setting*

$$\forall\, n \in \mathbb{N} : \forall\, (i,j) \in \square_n : \ a_n(i,j) = \begin{cases} \tilde{a}_n(i,j) & \text{if } i \leq j \\ \tilde{a}_n(j,i) & \text{if } i > j. \end{cases}$$

Let $h = (h_n)_n$ be a sequence of n-halfwidths with $h_n \to \infty$ and $\lim_n h_n/n \in \{0,1\}$. Let $(X_n^{NP})_n$ be the non-periodic random band matrices which are based on $(a_n)_n$ with halfwidth h. Then the following statements hold:

i) The semicircle law holds for $(X_n^{NP})_n$ in probability.

ii) If $\frac{1}{h_n^3}$ is summable over n, then the semicircle law holds almost surely for $(X_n^{NP})_n$.

Proof. This is a direct consequence of Theorem 4.17 and Theorem 4.47 *i*) and *ii*). □

Obviously, a statement analogous to the third statement of Theorem 4.9 is missing in Theorem 4.47. The statement in question was formulated mainly to be able to treat correlated Gaussian ensembles, as in Example 4.22, which is 1-almost uncorrelated with summable sequences $(D_n^{(l)})_n$ and $(C_n^{(l)})_n$ by Theorem 4.26. This motivated the formulation of Theorem 4.46 *iv*). Inspecting the third statement of Theorem 4.9 and the fourth statement of Theorem 4.46 (and keeping in mind that finally, $1/2h_n \leq b_n \leq 2h_n$ as pointed out in the proof of Theorem 4.47), the halfwidth h_n should satisfy $h_n \to \infty$, $\lim_n h_n/n \in \{0,1\}$ and both

$$\frac{1}{h_n^2} \xrightarrow[n\to\infty]{} 0 \quad \text{summably fast} \qquad \text{and} \qquad \min\left(1 - \frac{h_n}{n}, \frac{h_n}{n}\right) \leq \frac{C}{\log^d(n)} \tag{4.25}$$

for some $C, d > 1$ and all $n \in \mathbb{N}$. Let us analyze these conditions. The first condition in (4.25) requires $h_n \to \infty$ with some minimal speed. For example, $h_n \sim \sqrt{n \log^a(n)}$ grows too slowly for $a = 1$ but fast enough for all $a > 1$. The second condition in (4.25) states how h_n shall grow in comparison to n. If $h_n/n \to 0$, then h_n should not grow too quickly. The quickest growth rate allowed is of the form $h_n \sim C'n/\log^d(n)$ for some $C' > 0$ and $d > 1$. If $h_n/n \to 1$, then h_n should not grow too slowly. A minimal speed of

$$h_n \geq n\left(1 - \frac{C'}{\log^d(n)}\right)$$

for some $C' > 0$, $d > 1$ suffices. This will render the first condition to be satisfied, since then

$$\frac{1}{h_n^2} = \underbrace{\frac{1}{h_n^2} n^2 \left(1 - \frac{C'}{\log^d(n)}\right)^2}_{\to 1} \cdot \frac{1}{n^2 \underbrace{\left(1 - \frac{C'}{\log^d(n)}\right)^2}_{\geq \frac{1}{2} \text{ finally}}}.$$

Let us discuss some classic examples that satisfy the conditions in (4.25). For halfwidths $(h_n)_n$ with $h_n/n \to 0$, a very slow growth of h_n would be $h_n \sim \sqrt{n}\log(n)$, whereas a rather fast growth rate would be of the type $h_n \sim n/\log^2(n)$. Rates of the type $h_n \sim n^\rho$ for $\rho \in (1/2, 1)$ lie in between and are thus also admissible. Now if $h_n/n \to 1$, a rate of

$$h_n \geq n\left(\frac{\log(n)-1}{\log(n)}\right)$$

does not satisfy the convergence condition, for then, h_n does not grow quickly enough. On the other hand, a rate of

$$h_n \geq n\left(\frac{\log^2(n)-1}{\log^2(n)}\right)$$

would be admissible.

Theorem 4.49. *Let $(a_n)_n$ be the triangular scheme with correlated Gaussian entries as in Example 4.22 with respect to arbitrarily fixed $\alpha > 0$ and $(\Sigma_n)_n \in \mathrm{CovMat}(\alpha)$.*

Let $(h_n)_n$ be n-halfwidths which satisfy $h_n \to \infty$ and $\lim_n h_n/n \in \{0,1\}$, and let $(X_n^{NP})_n$ be the non-periodic random matrices which are based on $(a_n)_n$ with halfwidth $(h_n)_n$. Then the following statements hold:

i) If $\alpha \geq 1/2$, then the semicircle law holds for $(X_n^{NP})_n$ in probability.

ii) If $\alpha > 1/2$ and the conditions in (4.25) *are satisfied, then the semicircle law holds almost surely for $(X_n^{NP})_n$.*

Proof. By Theorem 4.26, $(a_n)_{n\in\mathbb{N}}$ is strongly α-almost uncorrelated. Thus, the first statement follows immediately from Theorem 4.47 *i*). The second statement of almost sure convergence follows from Corollary 4.28 *iv*) (note that summability of $1/h_n^2$ is equivalent to summability of $1/b_n^2$ by (4.24)) in combination with Theorem 4.46 *iv*), since entries of $(a_n)_n$ are 1-sub-Gaussian. □

4.5 Auxillary Statements

The SLLN for $\mathcal{L}_4$-Schemes with Independent Variables

We would like to make statements about terms involving a factor $1/\#I(n)$, where $I(n)$ is an n-dependent and possibly empty index set. In this context, we use the convention $1/0 = \infty$ and make the following definition:

Definition 4.50. A sequence $(x_n)_n$ in $\mathbb{R}_+ \cup \{\infty\}$ is called *finally summable*, if there exists an $N \in \mathbb{N}$ such that $x_n \neq \infty$ for all $n \geq N$ and

$$\sum_{n\geq N} x_n < \infty.$$

Theorem 4.51. *Let* $\left((Y_i^{(n)})_{i\in I(n)}\right)_{n\in\mathbb{N}}$ *be a sequence of real-valued random variables, where* $I(n)$ *is an* n*-dependent finite index set, such that the following holds:*

a) The random variables have a uniformly bounded fourth moment:

$$\exists C \geq 0 : \forall n \in \mathbb{N} : \forall i \in I(n) : \ \mathbb{E}\left(Y_i^{(n)}\right)^4 \leq C.$$

b) $(1/(\#I(n))^2)_{n\in\mathbb{N}}$ *is finally summable.*

Then the following statements hold:

i) If for all $n \in \mathbb{N}$ *and* $i \in I(n)$, $\mathbb{E}Y_i^{(n)} = 0$, *then*

$$\frac{1}{\#I(n)} \sum_{i\in I(n)} Y_i^{(n)} \xrightarrow[n\to\infty]{} 0 \qquad \textit{almost surely.}$$

ii) If there are real numbers $l \leq u$ *such that for all* $n \in \mathbb{N}$ *and* $i \in I(n)$, $\mathbb{E}Y_i^{(n)} \in [l, u]$, *then almost surely*

$$l \ \leq \ \liminf_{n\to\infty} \frac{1}{\#I(n)} \sum_{i\in I(n)} Y_i^{(n)} \ \leq \ \limsup_{n\to\infty} \frac{1}{\#I(n)} \sum_{i\in I(n)} Y_i^{(n)} \ \leq \ u.$$

Proof. It is clear that *i*) follows from *ii*). However, in the proof it is convenient to show *i*) first. To this end, let $\epsilon > 0$ be arbitrary, then we calculate:

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{\#I(n)} \sum_{i\in I(n)} Y_i^{(n)}\right| > \epsilon\right) \\
&\leq \frac{1}{\epsilon^4(\#I(n))^4} \sum_{i,j,k,l\in I(n)} \mathbb{E}Y_i^{(n)}Y_j^{(n)}Y_k^{(n)}Y_l^{(n)} \\
&= \frac{1}{\epsilon^4(\#I(n))^4} \sum_{i=j\neq k=l} \mathbb{E}\left(Y_i^{(n)}\right)^2\left(Y_k^{(n)}\right)^2 + \frac{1}{\epsilon^4(\#I(n))^4} \sum_{i=k\neq j=l} \mathbb{E}\left(Y_i^{(n)}\right)^2\left(Y_j^{(n)}\right)^2 \\
&\quad + \frac{1}{\epsilon^4(\#I(n))^4} \sum_{i=l\neq j=k} \mathbb{E}\left(Y_i^{(n)}\right)^2\left(Y_j^{(n)}\right)^2 + \frac{1}{\epsilon^4(\#I(n))^4} \sum_{i=j=k=l} \mathbb{E}\left(Y_i^{(n)}\right)^4 \\
&\leq 3 \cdot \frac{1}{\epsilon^4(\#I(n))^4} \cdot (\#I(n))^2 \cdot C + \frac{1}{\epsilon^4(\#I(n))^4} \cdot \#I(n) \cdot C \xrightarrow[n\to\infty]{} 0 \qquad \text{summably fast,}
\end{aligned}$$

where we used Markov's inequality in the first step, independence and centrality in the second and the Cauchy-Schwarz inequality in the third step. The statement follows by Borel-Cantelli.

Now to show *ii*), we write

$$\frac{1}{\#I(n)} \sum_{i\in I(n)} Y_i^{(n)} = \frac{1}{\#I(n)} \sum_{i\in I(n)} \left(Y_i^{(n)} - \mathbb{E}Y_i^{(n)}\right) + \frac{1}{\#I(n)} \sum_{i\in I(n)} \mathbb{E}Y_i^{(n)} =: S_n^{(1)} + S_n^{(2)}.$$

We will analyze the two subsums separately. For $S_n^{(1)}$ note that for all $n \in \mathbb{N}$ and $i \in I(n)$,

$$\|Y_i^{(n)} - \mathbb{E}Y_i^{(n)}\|_4 \leq \|Y_i^{(n)}\|_4 + \|\mathbb{E}Y_i^{(n)}\|_4 \leq \sqrt[4]{C} + |\mathbb{E}Y_i^{(n)}| \leq \sqrt[4]{C} + \max\{|l|, |u|\}.$$

Therefore, for all $n \in \mathbb{N}$ and $i \in I(n)$ we find

$$\mathbb{E}\left(Y_i^{(n)} - \mathbb{E}Y_i^{(n)}\right)^4 \leq (\sqrt[4]{C} + \max\{|l|, |u|\})^4,$$

so we can apply Statement $i)$ to the scheme $\left((Y_i^{(n)} - \mathbb{E}Y_i^{(n)})_{i\in I(n)}\right)_{n\in\mathbb{N}}$ and obtain

$$S_n^{(1)} \xrightarrow[n\to\infty]{} 0 \quad \text{almost surely,}$$

that is, this convergence takes place on some measurable set A with measure 1. For $S_n^{(2)}$, note that for finally all $n \in \mathbb{N}$, $I(n)$ will be nonempty (due to condition $b)$) and then for all such n,

$$S_n^{(2)} = \frac{1}{\#I(n)} \sum_{i\in I(n)} \mathbb{E}Y_i^{(n)} \in [l, u].$$

Therefore, in total, on the set A we find

$$\limsup_{n\to\infty} \frac{1}{\#I(n)} \sum_{i\in I(n)} Y_i^{(n)} \leq \limsup_{n\to\infty} S_n^{(1)} + \limsup_{n\to\infty} S_n^{(2)} \leq 0 + u.$$

and

$$\liminf_{n\to\infty} \frac{1}{\#I(n)} \sum_{i\in I(n)} Y_i^{(n)} \geq \liminf_{n\to\infty} S_n^{(1)} + \liminf_{n\to\infty} S_n^{(2)} \geq 0 + l.$$

□

The Bounded Lipschitz Metric

We denote by $\mathcal{M}_1(\mathbb{R})$ the space of probability measures on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra over $\mathbb{R}$. For all $\nu, \mu \in \mathcal{M}_1(\mathbb{R})$ we define

$$\mathrm{d}_{\mathrm{BL}}(\nu, \mu) := \sup\left\{\left|\int_{\mathbb{R}} f\,\mathrm{d}\mu - \int_{\mathbb{R}} f\,\mathrm{d}\nu\right| : \|f\|_\infty \leq 1, f \in \mathrm{Lip}_1\right\},$$

where Lip_1 denotes the set of functions $\mathbb{R} \to \mathbb{R}$, which are Lipschitz-continuous with a Lipschitz constant of at most 1. Then d_{BL} is a metric on $\mathcal{M}_1(\mathbb{R})$ which metrizes weak convergence (see [12, p. 191] and [50, p. 74]). That is, a sequence of probability measures $(\mu_n)_n$ in $\mathcal{M}_1(\mathbb{R})$ converges weakly to an element $\mu \in \mathcal{M}_1(\mathbb{R})$ if and only if $\mathrm{d}_{\mathrm{BL}}(\mu_n, \mu) \to 0$ as $n \to \infty$. Since we are mainly interested in weak convergence of empirical spectral distributions of random matrices, we use the following notation: Whenever d_{BL} receives an Hermitian matrix X as an argument, we will interpret this argument as the ESD of X. For example, if X and Y are Hermitian matrices with ESDs μ_X and μ_Y, then

$$\mathrm{d_{BL}}(X,Y) := \mathrm{d_{BL}}\left(\mu_X, \mu_Y\right)$$

and, keeping in mind that σ will always denote the semicircle distribution,

$$\mathrm{d_{BL}}(X,\sigma) := \mathrm{d_{BL}}\left(\mu_X, \sigma\right).$$

Now there is a fruitful inequality concerning the $\mathrm{d_{BL}}$-distance of two ESDs.

Lemma 4.52. *Let X and Y be two Hermitian $n \times n$ matrices, then we obtain*

$$\mathrm{d_{BL}}(X,Y)^2 \leq \frac{1}{n}\operatorname{tr}\left[(X-Y)^*(X-Y)\right] = \frac{1}{n}\sum_{i,j=1}^{n} |(X-Y)(i,j)|^2,$$

where tr *denotes the trace functional.*

Proof. We only need to show the first inequality. Let $f \in \mathrm{Lip}_1$ be arbitrary, λ_i^Z denote the eigenvalues and μ_Z the ESD of Z, where $Z \in \{X,Y\}$ and $i \in \{1,\ldots,n\}$, then

$$\begin{aligned}
\left|\int_{\mathbb{R}} f\,\mu_X - \int_{\mathbb{R}} f\,\mu_Y\right|^2 &= \left|\frac{1}{n}\sum_{i=1}^{n} f\left(\lambda_i^X\right) - \frac{1}{n}\sum_{i=1}^{n} f\left(\lambda_i^Y\right)\right|^2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|f\left(\lambda_i^X\right) - f\left(\lambda_i^Y\right)\right|^2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|\lambda_i^X - \lambda_i^Y\right|^2 \\
&\leq \frac{1}{n}\operatorname{tr}\left[(X-Y)^*(X-Y)\right],
\end{aligned}$$

where we used convexity in the second step and the Hoffman-Wielandt inequality (Theorem 2.33) in the fourth step. □

Sub-Gaussian Random Variables

If $\beta > 0$ is arbitrary, then a real-valued random variable X is said to be β-sub-Gaussian, if its moment generating function is dominated by the one of an $\mathcal{N}(0,\beta^2)$-distributed random variable, thus if

$$\forall\, t \in \mathbb{R} : \mathbb{E}e^{tX} \leq e^{\frac{\beta^2 t^2}{2}}.$$

In particular, a standard normal random variable is 1-sub-Gaussian. Sub-Gaussian random variables have rich properties, which can be studied in [52]. For example, if X is β-sub-Gaussian, then $\mathbb{E}X = 0$ and $\mathbb{V}X \leq \beta^2$. The following theorem characterizes sub-Gaussian random variables:

Theorem 4.53. *For a real-valued random variable X, the following statements are equivalent:*

1. *X is β-sub-Gaussian for some $\beta > 0$, i.e.* $\forall\, t \in \mathbb{R} : \mathbb{E}e^{tX} \le e^{\frac{\beta^2 t^2}{2}}$.
2. *There is a* $c > 0$ *such that* $\forall\, \lambda > 0 : \mathbb{P}(|X| \ge \lambda) \le 2e^{-c\lambda^2}$.
3. *There is an* $a > 0$ *such that* $\mathbb{E}e^{aX^2} \le 2$.

Further, if statement 1 holds with constant $\beta > 0$, *statement 2 holds with constant* $c = \frac{1}{2\beta^2}$.

Proof. See [52]. □

5 The Stieltjes Transform Method

5.1 Motivation and Basic Properties

In order to analyze properties of random variables and their distributions, it is a common technique to use transforms of these distributions which make analysis more accessible due to their favorable algebraic structure. For example, a common and short proof of the central limit theorem is conducted by using the Fourier transform of the random variables involved, owing to the property that Fourier transforms handle convolutions particularly well (and the central limit theorem is about a sum of random variables).

In random matrix theory, however, when analyzing empirical spectral distributions of diverse matrix ensembles, it is desirable to use a tool for analysis that relates the behavior of the empirical spectral distribution back to the level of the entries of the matrices. For example, using the method of moments, one sees in equation (3.4) that the moments of the ESD σ_n of a random matrix X_n can be calculated through:

$$\forall\, k \in \mathbb{N}: \ \left\langle \sigma_n, x^k \right\rangle = \frac{1}{n}\operatorname{tr}(X_n^k) = \frac{1}{n}\sum_{i_1,\dots,i_k=1}^{n} X_n(i_1,i_2)X_n(i_2,i_3)\cdots X_n(i_k,i_1).$$

In other words, instead trying to work with an ESD directly, we can analyze its moments which allows us to work on the level of the matrix entries.

A tool that combines both worlds, that is, that provides the structure of a transform with favorable algebraic properties and that allows us to work on the level of the matrix entries is the so called Stieltjes transform:

Definition 5.1. Let μ be a finite measure on $(\mathbb{R},\mathcal{B})$. Then we define the Stieltjes transform S_μ of μ as the map

$$\begin{aligned} S_\mu : \mathbb{C}\backslash\mathbb{R} &\longrightarrow \mathbb{C} \\ z &\longmapsto \int_{\mathbb{R}} \frac{1}{x-z}\mu(\mathrm{d}x) \end{aligned}$$

.

We note that the Stieltjes transform is defined via a measure-theoretical integral over a complex-valued function. We refer the reader to Section 5.6 for elementary but important properties concerning these integrals. The following lemma studies the Stieltjes transform $S_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z}\mu(\mathrm{d}x)$. Note that we do not have to consider the trivial case where $\mu \equiv 0$, since in this case, $S_\mu \equiv 0$. Notationally, we set $\mathbb{C}_+ := \{z \in \mathbb{C} \,|\, \operatorname{Im}(z) > 0\}$.

Lemma 5.2. *Let μ be a finite measure on $(\mathbb{R},\mathcal{B})$ with $\mu(\mathbb{R}) > 0$ and S_μ be its Stieltjes transform. Further, let $E \in \mathbb{R}$, $\eta \in \mathbb{R}\backslash\{0\}$ and $z := E + i\eta$, then we obtain:*

i) *For any $x \in \mathbb{R}$ we find:* $\frac{1}{x-z} = \frac{x-E}{(x-E)^2+\eta^2} + i\frac{\eta}{(x-E)^2+\eta^2}$.

ii) $\operatorname{Re} S_\mu(z) = \int \frac{x-E}{(x-E)^2+\eta^2}\mu(dx)$ *and* $\operatorname{Im} S_\mu(z) = \int \frac{\eta}{(x-E)^2+\eta^2}\mu(dx)$.

iii) $\operatorname{Im}(z) \gtrless 0 \Leftrightarrow \operatorname{Im} S_\mu(z) \gtrless 0$.

iv) $S_\mu(\overline{z}) = \overline{S_\mu(z)}$.

v) *S_μ is uniquely determined by its restriction $S_\mu : \mathbb{C}_+ \to \mathbb{C}_+$.*

vi) $|S_\mu(z)| \le \frac{\mu(\mathbb{R})}{|\operatorname{Im}(z)|}$

vii) *S_μ is holomorphic.*

viii) *In particular, S_μ is continuous, can be represented by a power series around any $z_0 \in \mathbb{C}\backslash\mathbb{R}$, and is infinitely often differentiable.*

Proof. Statement i) is obvious, ii) follows from i) by construction of the integral, iii) follows directly from ii) and so does iv) in combination with the construction of the integral. Statement v) follows directly from iii) and iv), and vi) follows from

$$\left|\frac{1}{x-z}\right| = \frac{1}{|x-z|} \le \frac{1}{|\operatorname{Im}(x-z)|} = \frac{1}{|\operatorname{Im}(z)|}.$$

To show statement vii), let $(z_n)_n$ and $z \in \mathbb{C}\backslash\mathbb{R}$ with $z_n \to z$, but $z_n \neq z$ be arbitrary, then:

$$\begin{aligned}\frac{S_\mu(z_n) - S_\mu(z)}{z_n - z} &= \frac{1}{z_n - z}\int \frac{1}{x-z_n} - \frac{1}{x-z}\mu(\mathrm{d}x)\\ &= \frac{1}{z_n - z}\int \frac{z_n - z}{(x-z_n)(x-z)}\mu(\mathrm{d}x) \xrightarrow[n\to\infty]{} \int \frac{1}{(x-z)^2}\mu(\mathrm{d}x)\end{aligned}$$

by dominated convergence, since for some $C > 0$ and all $n \in \mathbb{N}$,

$$\left|\frac{1}{(x-z_n)(x-z)}\right| \le \frac{1}{|\operatorname{Im}(z_n)||\operatorname{Im}(z)|} \le C,$$

for convergent sequences are bounded. □

It was mentioned in [59, p. 143] that the Stieltjes transform is holomorphic; we just gave a proof for the convenience of the reader. Also in [59, p. 144], boundedness properties are stated for the derivatives of the Stieltjes transform. They follow directly from our calculation of the derivatives in the following lemma.

Lemma 5.3 (Derivatives of S_μ). *For all $k \in \mathbb{N}_0$ we denote by $S_\mu^{(k)}$ the k-th derivative of S_μ. We find*

$$\forall\, k \in \mathbb{N}_0 : \ \forall\, z \in \mathbb{C}\backslash\mathbb{R} : S_\mu^{(k)}(z) = \int \frac{k!}{(x-z)^{k+1}} \mu(dx).$$

Proof. For $k = 0$ the statement is trivial, and for $k = 1$ we have shown the statement in the proof of Lemma 5.2. We proceed by induction. We assume the statement to be true for some $k \in \mathbb{N}$ and show that it is valid for $k+1$. To this end, let $z, (z_n)_n \in \mathbb{C}\backslash\mathbb{R}$ be arbitrary with $z_n \to z$ but $z_n \neq z$. Using that for any $a, b \in \mathbb{C}$ and $k \in \mathbb{N}_0$, we have

$$a^{k+1} - b^{k+1} = \left(\sum_{l=0}^{k} a^{k-l} b^l \right) (a-b),$$

we calculate, using the induction hypothesis in the first step,

$$\begin{aligned} \frac{S_\mu^{(k)}(z_n) - S_\mu^{(k)}(z)}{z_n - z} &= \frac{k!}{z_n - z} \int \frac{1}{(x-z_n)^{k+1}} - \frac{1}{(x-z)^{k+1}} \mu(\mathrm{d}x) \\ &= \frac{k!}{z_n - z} \int \left(\sum_{l=0}^{k} \frac{1}{(x-z_n)^{k-l}} \frac{1}{(x-z)^l} \right) \left(\frac{1}{x-z_n} - \frac{1}{x-z} \right) \mu(\mathrm{d}x) \\ &= k! \int \sum_{l=0}^{k} \frac{1}{(x-z_n)^{k-l+1}} \frac{1}{(x-z)^{l+1}} \mu(\mathrm{d}x). \end{aligned}$$

Realizing that for all $n \in \mathbb{N}$,

$$\left| \sum_{l=0}^{k} \frac{1}{(x-z_n)^{k-l+1}} \frac{1}{(x-z)^{l+1}} \right| \leq \sum_{l=0}^{k} \frac{1}{|\mathrm{Im}(z_n)|^{k-l+1} |\mathrm{Im}(z)|^{l+1}} \leq C,$$

for some positive constant C (since z_n converges), we obtain via dominated convergence:

$$S_\mu^{(k+1)}(z) = \lim_{n\to\infty} \frac{S_\mu^{(k)}(z_n) - S_\mu^{(k)}(z)}{z_n - z} = k! \int \sum_{l=0}^{k} \frac{1}{(x-z)^{k+2}} \mu(\mathrm{d}x) = \int \frac{(k+1)!}{(x-z)^{k+2}} \mu(\mathrm{d}x)$$

□

Theorem 5.4 (Retrieval of Measure). *For any bounded interval $I \subseteq \mathbb{R}$ with end points $\alpha < \beta$, we obtain the following:*

$$\mu((\alpha,\beta)) + \frac{1}{2}(\mu(\{\alpha\}) + \mu(\{\beta\})) = \lim_{\eta \searrow 0} \frac{1}{\pi} \int_I \mathrm{Im}\, S_\mu(E + i\eta) \lambda\!\!\lambda(dE).$$

Proof. Let I be an interval with end points $\alpha < \beta$ and $\eta > 0$. Then we obtain via Fubini:

$$\begin{aligned} \frac{1}{\pi} \int_I \mathrm{Im}\, S_\mu(E+i\eta) \lambda\!\!\lambda(\mathrm{d}E) &= \frac{1}{\pi} \int_I \int_{\mathbb{R}} \frac{\eta}{(x-E)^2 + \eta^2} \mu(\mathrm{d}x) \lambda\!\!\lambda(\mathrm{d}E) \\ &= \frac{1}{\pi} \int_{\mathbb{R}} \int_I \frac{\eta}{(x-E)^2 + \eta^2} \lambda\!\!\lambda(\mathrm{d}E) \mu(\mathrm{d}x) \\ &= \frac{1}{\pi} \int_{\mathbb{R}} \int_\alpha^\beta \frac{\eta}{(x-E)^2 + \eta^2} \mathrm{d}E \mu(\mathrm{d}x). \end{aligned}$$

Now since

$$\begin{aligned}
\int_\alpha^\beta \frac{\eta}{(x-E)^2+\eta^2}\mathrm{d}E &= \frac{1}{\eta}\int_\alpha^\beta \frac{1}{(\frac{E-x}{\eta})^2+1}\mathrm{d}E \\
&= \int_{\frac{\alpha-x}{\eta}}^{\frac{\beta-x}{\eta}} \frac{1}{E^2+1}\mathrm{d}E \\
&= \arctan\left(\frac{\beta-x}{\eta}\right) - \arctan\left(\frac{\alpha-x}{\eta}\right),
\end{aligned}$$

and $\arctan : \mathbb{R} \to (-\frac{\pi}{2}, +\frac{\pi}{2})$ is strictly increasing with $\lim_{x\to\pm\infty}\arctan(x) = \pm\frac{\pi}{2}$, we obtain

$$\lim_{\eta\searrow 0}\left[\arctan\left(\frac{\beta-x}{\eta}\right) - \arctan\left(\frac{\alpha-x}{\eta}\right)\right] = \begin{cases} \pi & \text{if } x\in(\alpha,\beta) \\ 0 & \text{if } x\notin[\alpha,\beta] \\ \frac{\pi}{2} & \text{if } x=\alpha \vee x=\beta. \end{cases}$$

Thus, by dominated convergence we find

$$\begin{aligned}
\lim_{\eta\searrow 0}\frac{1}{\pi}\int_I \operatorname{Im} S_\mu(E+i\eta)\lambda\!\!\lambda(\mathrm{d}E) &= \lim_{\eta\searrow 0}\frac{1}{\pi}\int_{\mathbb{R}} \arctan\left(\frac{\beta-x}{\eta}\right) - \arctan\left(\frac{\alpha-x}{\eta}\right)\mu(\mathrm{d}x) \\
&= \int_{\mathbb{R}} \mathbb{1}_{(\alpha,\beta)}(x) + \frac{1}{2}\mathbb{1}_{\{\alpha,\beta\}}(x)\mu(\mathrm{d}x) \\
&= \mu((\alpha,\beta)) + \frac{1}{2}(\mu(\{\alpha\})+\mu(\{\beta\}))
\end{aligned}$$

□

The previous theorem and the following corollary are similar to Theorem 2.4.3 in [6]. As usual, for a subset I of a topological space, we denote by ∂I its boundary, which is a concept we assume to be known to the reader.

Corollary 5.5. *For any bounded interval $I \subseteq \mathbb{R}$ with $\mu(\partial I) = 0$, we find:*

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

Thus, any finite measure μ on $(\mathbb{R},\mathcal{B})$ is uniquely determined by S_μ. In other words, $\mu \mapsto S_\mu$ is injective.

Proof. The convergence statement follows from Theorem 5.4. Now if μ and ν are finite measures on $(\mathbb{R},\mathcal{B})$ with $S_\mu = S_\nu$, denote their sets of atoms by $A_\rho := \{x\in\mathbb{R}, \rho(\{x\})>0\}$ for $\rho\in\{\mu,\nu\}$. Then A_μ and A_ν are at most countable. Let $a<b$ be arbitrary real numbers, then there are sequences $(a_n)_n$ and $(b_n)_n$ in $\mathbb{R}\backslash(A_\mu\cup A_\nu)$ with $a<a_n<b_n<b$ for all $n\in\mathbb{N}$ such that $a_n \searrow a$ and $b_n \nearrow b$. It follows

$$\mu((a,b)) = \lim_{n\to\infty}\mu((a_n,b_n)) = \lim_{n\to\infty}\nu((a_n,b_n)) = \nu((a,b)),$$

where we used continuity of measures in the first and last step and Theorem 5.4 in the middle step. Since μ and ν agree on all open bounded intervals, we conclude $\mu=\nu$. □

The last theorem and its corollary suggest that for any finite measure μ on $(\mathbb{R}, \mathcal{B})$ and $\eta > 0$ small, $E \mapsto \frac{1}{\pi} \operatorname{Im} S_\mu(E + i\eta)$ acts as a Lebesgue density for (a measure approximating) μ. In particular, even measures that do not possess a Lebesgue density (for example, all empirical measures) can be approximated in this way by using the Stieltjes transform. In Section 5.3 we will see how this can be made precise.

5.2 The Stieltjes Transform and Weak Convergence

For any finite measure, S_μ carries all the information of μ (cf. Corollary 5.5). Therefore, it is not surprising that this tool can be used particularly well to analyze weak convergence of probability measures. The following theorem generalizes Theorem 2.4.4 in [6].

Theorem 5.6 (Convergence Theorem)**.** *Let $Z \subseteq \mathbb{C} \backslash \mathbb{R}$ be a subset that has an accumulation point in $\mathbb{C} \backslash \mathbb{R}$ (which is not necessarily an element of Z itself). Then the following statements hold:*

1. *Let $(\mu_n)_n$ in $\mathcal{M}_1(\mathbb{R})$, such that for all $z \in Z$ we find that $S(z) := \lim_{n\to\infty} S_{\mu_n}(z)$ exists. Then there is a sub-probability measure μ with $\mu_n \to \mu$ vaguely and $S_\mu = S$.*
2. *Let $(\mu_n)_n$ and μ in $\mathcal{M}_1(\mathbb{R})$, then we find:*
$$\mu_n \to \mu \text{ weakly } \Leftrightarrow \ S_{\mu_n}(z) \to S_\mu(z) \text{ for all } z \in Z.$$
3. *Let $(\mu_n)_n$ be random probability measures and μ be a deterministic probability measure, then:*
   a) *$\mu_n \to \mu$ weakly in expectation $\Leftrightarrow \mathbb{E}S_{\mu_n}(z) \to S_\mu(z)$ for all $z \in Z$.*
   b) *$\mu_n \to \mu$ weakly in probability $\Leftrightarrow S_{\mu_n}(z) \to S_\mu(z)$ in probability for all $z \in Z$.*
   c) *$\mu_n \to \mu$ weakly almost surely $\Leftrightarrow [S_{\mu_n}(z) \to S_\mu(z)$ almost surely$]$ for all $z \in Z$.*

Proof. 1. Let $(\mu_n)_{n\in J}$ be an arbitrary subsequence of $(\mu_n)_{n\in\mathbb{N}}$. Due to Lemma 2.15, there exists a subsequence $(\mu_n)_{n\in I}$, $I \subseteq J$, such that $\mu_n \to \mu$ vaguely for $n \in I$ and a sub-probability measure μ. Since $x \mapsto \frac{1}{x-z}$ vanishes at $\pm\infty$, it follows $S_{\mu_n}(z) \to S_\mu(z)$ for $n \in I$ for all $z \in Z$ (cf. Lemma 2.10). Therefore, $S(z) = S_\mu(z)$ for all $z \in Z$. If ν is another subsequential limit of $(\mu_n)_{n\in J}$, we find by the same argument that $S_\mu(z) = S(z) = S_\nu(z)$ for all $z \in Z$. This implies $S_\mu = S_\nu$, since Stieltjes transforms are holomorphic. Therefore, $\mu = \nu$ by Theorem 5.4. By Lemma 2.9, we find $\mu_n \to \mu$ vaguely for $n \in \mathbb{N}$.
2. Since $x \mapsto \frac{1}{x-z}$ is continuous, "$\Rightarrow$" is obvious. To show "$\Leftarrow$", statement 1 yields that $\mu_n \to \mu$ vaguely, thus $\mu_n \to \mu$ weakly, since all measures involved are probability measures (cf. Lemma 2.14).
3.a) This follows directly from statement 2, considering

$$\mathbb{E}S_{\mu_n}(z) = \mathbb{E} \int \frac{1}{x-z} \mu_n(\mathrm{d}x) = \int \frac{1}{x-z} \mathbb{E}\mu_n(\mathrm{d}x) = S_{\mathbb{E}\mu_n}(z),$$

see also Theorem 2.20.
3.c) If $\mu_n \to \mu$ weakly on a measurable set A with $\mathbb{P}(A) = 1$, then we have on A that for all $z \in Z$ we find $S_{\mu_n}(z) \to S_\mu(z)$ (by statement 2). This shows "$\Rightarrow$", and to show "$\Leftarrow$", fix a sequence $(z_k)_k$ in Z that converges to some $z \in \mathbb{C}\backslash\mathbb{R}$. For all $k \in \mathbb{N}$ we find a measurable set A_k with $\mathbb{P}(A_k) = 1$ on which $S_{\mu_n}(z_k) \to S_\mu(z_k)$ as $n \to \infty$. Then $A := \cap_{k\in\mathbb{N}} A_k$ is measurable with $\mathbb{P}(A) = 1$, and on A we find that for all $z \in Z' := \{z_k | k \in \mathbb{N}\}$ we have $S_{\mu_n}(z) \to S_\mu(z)$. Since Z' has an accumulation point in $\mathbb{C}\backslash\mathbb{R}$, we find on the set A that $\mu_n \to \mu$ weakly by statement 2.
3.b) The direction "$\Rightarrow$" is trivial since $x \mapsto \frac{x-E}{(x-E)^2+\eta^2}$ and $x \mapsto \frac{\eta}{(x-E)^2+\eta^2}$ are bounded and continuous (cf. Theorem 2.25). For "$\Leftarrow$" we let $f \in \mathcal{C}_b(\mathbb{R})$ be arbitrary. Then we need to show that $\langle \mu_n, f\rangle \to \langle \mu, f\rangle$ in probability. Let $J \subseteq \mathbb{N}$ be a subsequence, then by Lemma 2.28, we find a subsequence $I \subseteq J$ and a measurable set N with $\mathbb{P}(N) = 0$, such that for $(z_k)_k$ fixed as in the proof of 3.c):

$$\forall \omega \in \Omega\backslash N : \forall k \in \mathbb{N} : S_{\mu_n(\omega)}(z_k) \xrightarrow[n\in I]{} S_{\mu(\omega)}(z_k).$$

Therefore, it follows with statement 3.c) that $\mu_n \xrightarrow[n\in I]{} \mu$ almost surely, so in particular $\langle \mu_n, f\rangle \xrightarrow[n\in I]{} \langle \mu, f\rangle$ almost surely. Then $\langle \mu_n, f\rangle \xrightarrow[n\in \mathbb{N}]{} \langle \mu, f\rangle$ in probability by Lemma 2.27.
□

We refer the reader to Remark 2.26 for an explanation on the use of brackets [...] in Theorem 5.6 3. *c*).

5.3 The Imaginary Part of the Stieltjes Transform

In Corollary 5.5 we saw that if $\mu \in \mathcal{M}_1(\mathbb{R})$, then for a small $\eta > 0$, the function $E \mapsto \frac{1}{\pi} \operatorname{Im} S_\mu(E + i\eta)$ should be the Lebesgue density of a probability measure on $(\mathbb{R}, \mathcal{B})$ that approximates μ well. But so far, we do not even know whether $E \mapsto \frac{1}{\pi} \operatorname{Im} S_\mu(E + i\eta)$ yields a density of a *probability measure* at all. How can this intuition be portrayed in the right context, and is there a connection to the weak convergence results of Section 5.2? This section aims to shed light onto these aspects. First, we will rigorously delve into convolution of probability measures, which will be based on [4]. Second, we will introduce kernel density estimators, which motivate further the use of the Stieltjes transform when analyzing ESDs of random matrices. We begin by making the following definition:

Definition 5.7. Let μ and ν be probability measures on $(\mathbb{R}, \mathcal{B})$ and $f, g : \mathbb{R} \to \mathbb{R}$ Lebesgue-density functions (i.e. $h \geq 0$ and $\int h \mathrm{d}\lambda = 1$, $h \in \{f, g\}$).

i) The convolution of the probability measures μ and ν is defined as $\mu * \nu := (\mu \otimes \nu)^+$. Here, $\mu \otimes \nu$ is the product measure on $(\mathbb{R}^2, \mathcal{B}^2)$, $+ : \mathbb{R}^2 \to \mathbb{R}$ is the addition map, and $(\mu \otimes \nu)^+$ is the push-forward of the product measure under the addition map.

ii) The convolution of the density f and the probability measure ν is defined as the function $f * \nu : \mathbb{R} \to \mathbb{R}$ with

$$\forall\, x \in \mathbb{R}: \ (f * \nu)(x) := \int_{\mathbb{R}} f(x-y)\nu(\mathrm{d}y).$$

iii) The convolution of the densities f and g is the function $f * g : \mathbb{R} \to \mathbb{R}$ with

$$\forall\, x \in \mathbb{R}: \ (f * g)(x) := \int_{\mathbb{R}} f(x-y)g(y)\lambda\!\!\lambda(\mathrm{d}y).$$

Note that in ii) and iii) above, the definitions of the convolution are to be understood for $\lambda\!\!\lambda$-almost all $x \in \mathbb{R}$, since the respective integrals are well-defined only for $\lambda\!\!\lambda$-almost all $x \in \mathbb{R}$, which can be observed via Fubini/Tonelli. The convolutions are understood to equal zero on the respective sets of measure zero.

We will now casually discuss Definition 5.7 and summarize the most important aspects of our findings in the next lemma. So let us assume we are in the situation of said definition.

Let us first discuss point *i)* of Definition 5.7: Per construction, the convolution of two probability measures yields another probability measure on the real line, and if $B \in \mathcal{B}$ is arbitrary, we find

$$(\mu * \nu)(B) = (\mu \otimes \nu)\left(\left\{(x,y) \in \mathbb{R}^2 : x + y \in B\right\}\right).$$

Further, if $f : \mathbb{R} \to \mathbb{R}$ is $\mu * \nu$-integrable, then we obtain by transformation:

$$\int_{\mathbb{R}} f \mathrm{d}(\mu{*}\nu) = \int_{\mathbb{R}^2} (f \circ +)\mathrm{d}(\mu{\otimes}\nu) = \int_{\mathbb{R}^2} f(x{+}y)(\mu{\otimes}\nu)(\mathrm{d}(x,y)) = \int_{\mathbb{R}}\int_{\mathbb{R}} f(x{+}y)\mu(\mathrm{d}x)\nu(\mathrm{d}y),$$

so that in particular for an indicator function $f = \mathbb{1}_B$ for some $B \in \mathcal{B}$:

$$(\mu * \nu)(B) = \int_{\mathbb{R}} \mathbb{1}_B \mathrm{d}(\mu * \nu) = \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_B(x+y)\mu(\mathrm{d}x)\nu(\mathrm{d}y) = \int_{\mathbb{R}} \mu(B-y)\nu(\mathrm{d}y),$$

where the fact that the first term is equal to the third shows in a particularly nice way that the convolution is commutative (via Fubini). Let us introduce a quick but enlightening example:

Example 5.8 (Convolution with Dirac measures)**.** For all $a \in \mathbb{R}$, denote by δ_a the Dirac measure in a and by T_a the translation by a, that is, $T_a : \mathbb{R} \to \mathbb{R}$, $T_a(x) = x + a$ for all $x \in \mathbb{R}$. Then we find for any probability measure $\mu \in \mathcal{M}_1(\mathbb{R})$:

$$\mu * \delta_a = \mu^{T_a}, \quad \text{in particular:} \quad \mu * \delta_0 = \mu,$$

since $T_0 = \mathrm{id}_{\mathbb{R}}$. We conclude that δ_0 is the neutral element of convolution (there is no other neutral element, since $*$ is commutative). To prove our claim, we calculate for an arbitrary $B \in \mathcal{B}$:

$$(\mu * \delta_a)(B) = \int_{\mathbb{R}} \mu(B-y)\delta_a(\mathrm{d}y) = \mu(B-a) = \mu^{T_a}(B),$$

where we used that $T_a^{-1} = T_{-a}$.

Now, let us discuss point *ii)* of Definition 5.7: First of all, we point out that $f * \nu$ is a Lebesgue-density function, for it is non-negative, and via Fubini we obtain immediately that $\int f * \nu \mathrm{d}\lambda\!\!\lambda = 1$. But even more holds: $f * \nu$ is the Lebesgue-density of the convolution $(f\lambda\!\!\lambda) * \nu$, so that the equality $(f\lambda\!\!\lambda) * \nu = (f * \mu)\lambda\!\!\lambda$ holds. In particular, this convolution is Lebesgue-continuous. To verify our statement, we calculate for an arbitrary $B \in \mathcal{B}$:

$$\begin{aligned} [(f\lambda\!\!\lambda) * \nu](B) &= \int_{\mathbb{R}} (f\lambda\!\!\lambda)(B - y)\nu(\mathrm{d}y) \\ &= \int_{\mathbb{R}} \int_{B-y} f(x)\lambda\!\!\lambda(\mathrm{d}x)\nu(\mathrm{d}y) \\ &= \int_{\mathbb{R}} \int_{B} f(x-y)\lambda\!\!\lambda(\mathrm{d}x)\nu(\mathrm{d}y) \\ &= \int_{B} \int_{\mathbb{R}} f(x-y)\nu(\mathrm{d}y)\lambda\!\!\lambda(\mathrm{d}x) \\ &= \int_{B} [f * \nu](x)\lambda\!\!\lambda(\mathrm{d}x), \end{aligned}$$

where the third step follows from

$$\int_{B-y} f(x)\lambda\!\!\lambda(\mathrm{d}x) = \int_{T_y^{-1}(B)} (f \circ T_y^{-1} \circ T_y)(x)\lambda\!\!\lambda(\mathrm{d}x) = \int_{B} (f \circ T_y^{-1})(x)\lambda\!\!\lambda^{T_y}(\mathrm{d}x),$$

and the Lebesgue measure is translation invariant, thus $\lambda\!\!\lambda^{T_y} = \lambda\!\!\lambda$.

Lastly, let us discuss point *iii)* of Definition 5.7: Again by Fubini, we find immediately that $f * g$ is a Lebesgue-density function. Now since from the definition we have for all $x \in \mathbb{R}$ that $(f * g)(x) = (f * (g\lambda\!\!\lambda))(x)$, we find through our discussion of point *ii)* that $f * g$ is the Lebesgue-density of the convolution $(f\lambda\!\!\lambda) * (g\lambda\!\!\lambda)$, so $(f * g)\lambda\!\!\lambda = (f\lambda\!\!\lambda) * (g\lambda\!\!\lambda)$. Let us summarize our findings in the following lemma:

Lemma 5.9. *In the situation of Definition 5.7, we make the following observations (point x here is with respect to point x in Definition 5.7, $x \in \{i), ii), iii)\}$):*

i) The convolution is a commutative binary operation on the space of probability measures. The neutral element is given by δ_0. Further, the following formula holds:

$$\forall\, B \in \mathcal{B} : (\mu * \nu)(B) = \int_{\mathbb{R}} \mu(B - y)\nu(dy).$$

ii) $f * \nu$ is a Lebesgue-density for the convolution $(f\lambda\!\!\lambda) * \nu$, that is, $(f\lambda\!\!\lambda) * \nu = (f * \nu)\lambda\!\!\lambda$.

iii) $f * g$ is a Lebesgue-density for the convolution $(f\lambda\!\!\lambda) * (g\lambda\!\!\lambda)$, that is, $(f\lambda\!\!\lambda) * (g\lambda\!\!\lambda) = (f * g)\lambda\!\!\lambda$.

Proof. This follows from the discussion preceding the lemma. □

The following lemma will capture a very important property of the convolution:

Lemma 5.10. *The convolution of probability measures on* $(\mathbb{R}, \mathcal{B})$ *is continous with respect to weak convergence. That is, if* $(\mu_n)_n$, $(\nu_n)_n$, μ *and* ν *are probability measures on* $(\mathbb{R}, \mathcal{B})$ *with* $\mu_n \to \mu$ *and* $\nu_n \to \nu$ *weakly, then* $\mu_n * \nu_n \to \mu * \nu$ *weakly.*

Proof. With [10, p. 23] it follows that $\mu_n \otimes \nu_n \to \mu \otimes \nu$. Now if $f \in \mathcal{C}_b(\mathbb{R})$ is arbitrary, then we also have that $(x, y) \mapsto f(x + y)$ is a continuous and bounded function on $\mathbb{R}^2$, so

$$\int_{\mathbb{R}} f \mathrm{d}(\mu_n * \nu_n) = \int_{\mathbb{R}^2} (f \circ +) \mathrm{d}(\mu_n \otimes \nu_n) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}^2} (f \circ +) \mathrm{d}(\mu \otimes \nu) = \int_{\mathbb{R}} f \mathrm{d}(\mu * \nu).$$

□

Now, we will bring the Stieltjes transform into play:

Definition 5.11. For all $\eta > 0$, we define the Cauchy kernel $P_\eta : \mathbb{R} \to \mathbb{R}$ as the function with

$$\forall\, x \in \mathbb{R} : P_\eta(x) := \frac{1}{\pi} \frac{\eta}{x^2 + \eta^2},$$

which is the $\lambda\!\!\lambda$-density function of the Cauchy distribution with scale parameter η.

We will collect a quick lemma before proceeding:

Lemma 5.12. *As* $\eta \searrow 0$, *we find* $(P_\eta \lambda\!\!\lambda) \to \delta_0$ *weakly.*

Proof. The characteristic function of the measure $P_\eta \lambda\!\!\lambda$ is given by $t \mapsto e^{-\eta|t|}$, see [44, p. 330] and [54, p. 208]. Fixing $t \in \mathbb{R}$ and letting $\eta \to 0$ will yield the statement, since e^0 is the characteristic function of δ_0. □

Now, as we see, for any probability measure μ on $(\mathbb{R}, B)$, we have

$$\frac{1}{\pi} \operatorname{Im} S_\mu(E + i\eta) = \int_{\mathbb{R}} \frac{1}{\pi} \frac{\eta}{(E - x)^2 + \eta^2} \mu(\mathrm{d}x) = (P_\eta * \mu)(E)$$

Therefore, $1/\pi \operatorname{Im} S_\mu(\cdot + i\eta)$ is the convolution of the density P_η with μ and thus a Lebesgue-density for the measure $(P_\eta \lambda\!\!\lambda) * \mu$. In particular, as $\eta \searrow 0$ we have that

$$\frac{1}{\pi} \operatorname{Im} S_\mu(\cdot + i\eta) \lambda\!\!\lambda = (P_\eta \lambda\!\!\lambda) * \mu \longrightarrow \delta_0 * \mu = \mu \quad \text{weakly.}$$

This immediately proves Corollary 5.5 again (using the Portmanteau theorem). But due to continuity of the convolution, we can say much more:

Assume that $(\sigma_n)_n$ is a sequence of ESDs of Hermitian random matrices, so that σ_n converges almost surely to the semicircle distribution σ. We assume this convergence takes place on a measurable set A with $\mathbb{P}(A) = 1$. Then we find on A that the following commutative diagram holds, where all arrows indicate weak convergence:

$$\begin{tikzcd}
(P_\eta * \sigma_n)\lambda\!\!\lambda \arrow[r, "n \to \infty"'] \arrow[d, "\eta \searrow 0"'] \arrow[dr, "\substack{n\to\infty \\ \eta\searrow 0}"'] & (P_\eta * \sigma)\lambda\!\!\lambda \arrow[d, "\eta \searrow 0"] \\
\delta_0 * \sigma_n = \sigma_n \arrow[r, "n \to \infty"'] & \sigma
\end{tikzcd}$$

In particular, the diagonal arrow says that we obtain weak convergence $(P_{\eta_n} * \sigma_n)\lambda\!\!\lambda \to \sigma$ as $n \to \infty$ for any sequence $\eta_n \searrow 0$. This is an interesting result, but it does not tell us if also densities align. More concretely, write $\sigma = f_\sigma \lambda\!\!\lambda$, then from $(P_\eta * \sigma_n)\lambda\!\!\lambda \to f_\sigma \lambda\!\!\lambda$ weakly we cannot infer that also $P_\eta * \sigma_n \to f_\sigma$ in some sense, for example in $\|\cdot\|_\infty$ over a specified compact interval. This is desirable since it allows conclusion about *local* estimation of σ_n by σ. If $\eta = \eta_n$ drops too quickly to zero as $n \to \infty$, then $(P_{\eta_n} * \sigma_n)$ will have steep peaks at each eigenvalue, thus will not approximate the density of the semicircle distribution uniformly. This "problem" is typical for kernel density estimators in general (see [54], especially their Remark 11.2.10), which we will introduce next.

Definition 5.13. A kernel K is a Lebesgue-probability-density function $\mathbb{R} \to \mathbb{R}$, that is, K is non-negative and

$$\int_{\mathbb{R}} K(y)\lambda\!\!\lambda(dy) = 1.$$

Further, if K is a kernel and $h > 0$, we define K_h as the kernel with $K_h(x) = \frac{1}{h}K(\frac{x}{h})$ for all $x \in \mathbb{R}$ and call K_h the kernel K at bandwidth h. In particular, $K = K_1$.

In above definition, it is clear that K_h is a kernel if K is a kernel and $h > 0$. An example of a kernel is the Cauchy kernel P_1 from Definition 5.11, which yields the standard Cauchy distribution. We have for all $x \in \mathbb{R}$ and $\eta > 0$:

$$P_1(x) = \frac{1}{\pi}\frac{1}{x^2+1} \quad \text{and} \quad P_\eta(x) = \frac{1}{\pi\eta}\frac{1}{\left(\frac{x}{\eta}\right)^2+1} = \frac{1}{\pi}\frac{\eta}{x^2+\eta^2}.$$

Now given a vector $v = (v_1, \ldots, v_N)$ of real-valued observations, we are interested in constructing a Lebesgue-density that describes the experiment of drawing uniformly at random from these observations, in other words that approximates the empirical probability measure

$$\nu_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{v_i}. \tag{5.1}$$

This can be done with help of a kernel K, which is oftentimes chosen to be unimodal and symmetric around 0, just as the Cauchy kernel P_1.

Definition 5.14. The *kernel density estimator with kernel K and bandwidth $h > 0$* for an empirical measure ν as in (5.1) is the Lebesgue-density given by the convolution $K_h * \nu$, thus

$$\begin{aligned} K_h * \nu : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto (K_h * \nu)(x) = \frac{1}{N}\sum_{i=1}^{N} K_h(x - v_i) = \frac{1}{Nh}\sum_{i=1}^{N} K_1\left(\frac{x - v_i}{h}\right) \end{aligned}$$

Heuristically speaking, the concept works in the following way: The center of the kernel is placed upon each observation, whose influence (i.e. probability mass of $1/N$) is smoothed over its neighborhood. The size of this neighborhood is governed by the bandwidth h: A small h will restrain the probability mass of $1/N$ to be closer to its observation, whereas a larger h will result in a wider spread of probability mass. Therefore, a smaller h will result in a peaky density function (with steep peaks at the observation), whereas a larger h will result in a smoother density function.

We now assume we are given an empirical spectral distribution σ_N from a real symmetric $N \times N$ matrix X_N. The kernel density estimator at location $E \in \mathbb{R}$ for σ_N with kernel P_1 at bandwidth $\eta > 0$ is then given by

$$(P_\eta * \sigma_N)(E) = \frac{1}{N\eta}\sum_{i=1}^{N}\frac{1}{\pi}\frac{1}{\left(\frac{E-\lambda_i^{X_N}}{\eta}\right)^2 + 1} = \frac{1}{\pi N}\sum_{i=1}^{N}\frac{\eta}{(E - \lambda_i^{X_N})^2 + \eta^2} = \frac{1}{\pi}\operatorname{Im} S_{\sigma_N}(E + i\eta).$$

This gives the imaginary part of the Stieltjes transform the new role of a kernel density estimator for the empirical spectral distribution. Let us conduct a simulation study for $N = 100$. Let A_{100} be a symmetric 100×100 random matrix with independent Rademacher distributed variables in the upper half triangle, including the main diagonal. Let $X_{100} := \frac{1}{\sqrt{100}} A_{100}$. Denote by σ_{100} the empirical spectral distribution of X_{100}. Further, we define the bandwidths $\eta_1 := 1/N^{1/2} = 1/10$ and $\eta_2 := 1/N^1 = 1/100$. With respect to the commutative diagram after Lemma 5.12 and the discussion below it, let us analyze how well $P_{\eta_1} * \sigma_{100}$ and $P_{\eta_2} * \sigma_{100}$ can be approximated by the density of the semicircle distribution, f_σ, in Figures 5.1 and 5.2, which are based on the same simulation outcome.

As we see, considering that we are in the case of a very low $N = 100$, we already obtain a decent approximation by the semicircle density in Figure 5.1. Reducing the scale from η_1 to η_2 we obtain the result in Figure 5.2. There we observe that for the smaller bandwidth parameter η_2, we do not obtain a useful approximation by the semicircle density anymore. Indeed, as we will elaborate in the next chapter, the scale $1/N^1$ is too fast to obtain uniform convergence of the estimated density to the target density, whereas a scale of $1/N^{1-\gamma}$ for any $\gamma \in (0, 1)$ would be sufficient. Nevertheless, Figure 5.2 displays nicely how the kernel density estimator works: A closer look – in particular to the edges of the bulk – shows how the probability mass of each individual eigenvalue is spread around its neighborhood.

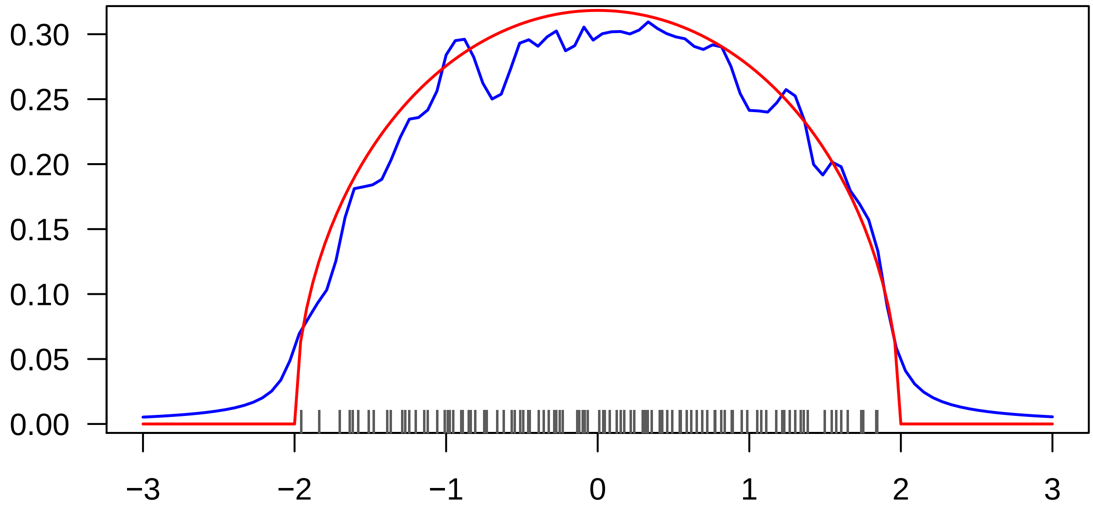

Figure 5.1: Red line: f_σ. Blue line: $\frac{1}{\pi}\operatorname{Im} S_{\sigma_{100}}(\cdot + i\eta_1) = P_{\eta_1} * \sigma_{100}$. Grey bars: eigenvalue locations.

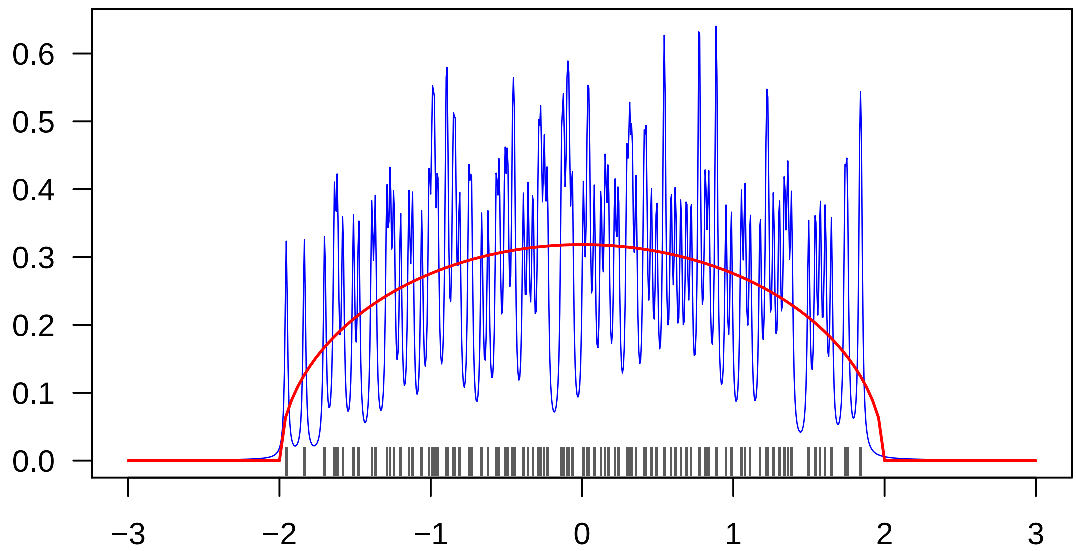

Figure 5.2: Red line: f_σ. Blue line: $\frac{1}{\pi}\operatorname{Im} S_{\sigma_{100}}(\cdot + i\eta_2) = P_{\eta_2} * \sigma_{100}$. Grey bars: Eigenvalue locations.

5.4 The Stieltjes Transform of the Semicircle Distribution

This section is devoted to the Stieltjes transform of a very specific probability measure, namely the semicircle distribution σ on $(\mathbb{R}, \mathcal{B})$ as in Definition 2.38. We will simply write $s := S_\sigma$ for brevity. Here and later, for any $z \in \mathbb{C}\backslash\mathbb{R}_+$, we will denote by $\sqrt{z}$ the square root with positive imaginary part.

Lemma 5.15. *Denote by s the Stieltjes transform of the semicircle distribution, then we obtain*

$$\forall\, z \in \mathbb{C}_+ : s(z) = \frac{-z + \sqrt{z^2 - 4}}{2}.$$

Proof. See [7, p. 32]. □

The Stieltjes transform s of the semicircle law has important properties, which are summarized in the following theorem (compare with Lemma 6.2 in [27], which was not proved):

Theorem 5.16. *For the Stieltjes transform s of the semicircle distribution, we find:*

1. *For $z \in \mathbb{C}_+$, the equation in m*
$$m^2 + zm + 1 = 0 \tag{5.2}$$
has the solutions $\frac{-z \pm \sqrt{z^2-4}}{2}$ in $\mathbb{C}$. Further, $s(z)$ is the positive and $1/s(z)$ is the negative branch of these solutions.
2. *For $z \in \mathbb{C}_+$, we find $s(z) = -\frac{1}{z+s(z)}$.*
3. *There exists a constant $C_s > 0$ such that for all $z = E + i\eta \in [-10, 10] + i(0, 10]$ we find*
   a) $|s(z)| \leq C_s$.
   b) $\left|\frac{1}{s(z)}\right| \leq C_s$.
   c) $\frac{1}{C_s}\sqrt{\kappa + \eta} \leq \left|s(z) - \frac{1}{s(z)}\right| \leq C_s\sqrt{\kappa + \eta}$,

   where $\kappa = ||E| - 2|$. C_s can be chosen to be 16.

Proof. Clearly, the solutions of (5.2) are given by

$$m_{1,2} = -\frac{z}{2} \pm \sqrt{\frac{z^2}{4} - 1} = \frac{-z \pm \sqrt{z^2 - 4}}{2}.$$

In particular, $s(z)$ is the positive branch of the solutions to (5.2), that is, $s(z)^2 + zs(z) + 1 = 0$. Especially, $s(z) + z \neq 0$, hence statement 2 holds, and from this we obtain (since $s(z) \neq 0$ by Lemma 5.2)

$$\frac{1}{s(z)} = -z - s(z) = -z - \frac{-z + \sqrt{z^2 - 4}}{2} = \frac{-z - \sqrt{z^2 - 4}}{2},$$

which completes 1. To see 3. a) and b), note that

$$\left|\frac{-z \pm \sqrt{z^2-4}}{2}\right| \leq \frac{|z| + \sqrt{|z|^2+4}}{2} \leq \frac{2|z|+2}{2} \leq 16,$$

since $\sqrt{\cdot}$ is sub-additive on $[0,\infty)$ and $|z| \leq 10\sqrt{2} \leq 15$. So a) and b) hold with $C_s := 16$. For c) we must conduct some light calculations: First of all, with $z = E + i\eta$ we notice that

$$\begin{aligned}\left|s(z) - \frac{1}{s(z)}\right| &= \left|\frac{-z+\sqrt{z^2-4}}{2} - \frac{-z-\sqrt{z^2-4}}{2}\right| \\ &= |\sqrt{z^2-4}| \\ &= \sqrt{|z+2||z-2|} \\ &= \sqrt{|E+2+i\eta| \cdot |E-2+i\eta|}.\end{aligned}$$

Second, since for all real numbers a and b we easily find

$$\frac{1}{\sqrt{2}}(|a|+|b|) \leq \sqrt{a^2+b^2} \leq |a|+|b|,$$

it follows (since then for example, $\frac{1}{\sqrt{2}}(|E+2|+\eta) \leq |E+2+i\eta| \leq |E+2|+|\eta|$)

$$\begin{aligned}\frac{1}{\sqrt{2}}\sqrt{(|E+2|+\eta)\cdot(|E-2|+\eta)} &\leq \sqrt{|E+2+i\eta| \cdot |E-2+i\eta|} \\ &\leq \sqrt{(|E+2|+\eta)\cdot(|E-2|+\eta)}.\end{aligned}$$

For both cases $E \geq 0$ and $E < 0$ it follows immediately that

$$\sqrt{(|E+2|+\eta)\cdot(|E-2|+\eta)} \leq \sqrt{22}\sqrt{||E|-2|+\eta},$$

which shows that

$$\left|s(z) - \frac{1}{s(z)}\right| \leq \sqrt{22}\sqrt{\kappa+\eta} \leq 16\sqrt{\kappa+\eta}.$$

On the other hand, for $E \geq 0$ arbitrary, we find

$$\begin{aligned}\sqrt{||E|-2|+\eta} &= \sqrt{|E-2|+\eta} \\ &\leq \sqrt{\frac{|E+2|+\eta}{2}} \cdot \sqrt{|E-2|+\eta} = \frac{1}{\sqrt{2}}\sqrt{|E+2|+\eta}\sqrt{|E-2|+\eta}\end{aligned}$$

and similarly for $E < 0$ that

$$\begin{aligned}\sqrt{||E|-2|+\eta} &= \sqrt{|E+2|+\eta} \\ &\leq \sqrt{\frac{|E-2|+\eta}{2}} \cdot \sqrt{|E+2|+\eta} = \frac{1}{\sqrt{2}}\sqrt{|E+2|+\eta}\sqrt{|E-2|+\eta},\end{aligned}$$

which implies

$$\sqrt{\kappa+\eta} \leq \left| s(z) - \frac{1}{s(z)} \right|.$$

In total, we obtain

$$\frac{1}{C_s}\sqrt{\kappa+\eta} \leq \left| s(z) - \frac{1}{s(z)} \right| \leq C_s\sqrt{\kappa+\eta}.$$

□

In the following we would like to unveil that as $\eta \searrow 0$, the function $E \mapsto \frac{1}{\pi} \operatorname{Im} s(E+i\eta)$, that is $P_\eta * f_\sigma$, converges uniformly to f_σ over any compact interval and with a speed of $O(\sqrt{\eta})$. We know that

$$\frac{1}{\pi} \operatorname{Im} s(z) = -\frac{\operatorname{Im}(z)}{2\pi} + \frac{\operatorname{Im}(\sqrt{z^2-4})}{2\pi}.$$

How do we gain access to $\operatorname{Im}(\sqrt{z^2-4})$? We need an auxiliary lemma:

Lemma 5.17. *Let $a, b, c, d \in \mathbb{R}$, where $b > 0$. If $(a+ib)^2 = c + id$, then*

$$b = \sqrt{\frac{-c + \sqrt{c^2+d^2}}{2}}.$$

Proof. We find:

$$\begin{aligned}
&c + id = (a+ib)^2 = a^2 + 2abi - b^2 \\
&\Rightarrow c = a^2 - b^2 \ \wedge \ d = 2ab \\
&\Rightarrow b^2 + c = a^2 = \frac{d^2}{4b^2} \\
&\Rightarrow (b^2)^2 + cb^2 - \frac{d^2}{4} = 0 \\
&\Rightarrow (b^2)_{1,2} = -\frac{c}{2} \pm \sqrt{\frac{c^2}{4} + \frac{d^2}{4}} \\
&\Rightarrow b^2 = -\frac{c}{2} + \sqrt{\frac{c^2}{4} + \frac{d^2}{4}} \\
&\Rightarrow b = \sqrt{\frac{-c + \sqrt{c^2+d^2}}{2}},
\end{aligned}$$

where we used in the fifth step that $b^2 > 0$ and in the sixth step that $b > 0$. □

We can now prove the following lemma, which is also one of the contributions of this thesis.

Lemma 5.18. *Let $C \geq 2$ be arbitrary, then we obtain for any $\eta \in (0, C]$:*

$$\sup_{E \in [-C,C]} \left| \frac{1}{\pi} \operatorname{Im}(s(E+i\eta)) - f_\sigma(E) \right| \leq \sqrt{C\eta}.$$

Proof. Throughout the proof we use that for $x, y \geq 0$ we have $\sqrt{x+y} \leq \sqrt{x}+\sqrt{y}$, without further mention. With $C \geq 2$ and $z = E + i\eta$, where $E \in [-C, C]$ and $\eta > 0$, we find that

$$z^2 - 4 = \underbrace{E^2 - \eta^2 - 4}_{=:c} + i \underbrace{2E\eta}_{=:d}.$$

Applying Lemma 5.17 (with c and d as just defined) yields:

$$\operatorname{Im}(\sqrt{z^2-4}) = \sqrt{\frac{4+\eta^2-E^2+\sqrt{(E^2-\eta^2-4)^2+4E^2\eta^2}}{2}},$$

so

$$\frac{1}{\pi}\operatorname{Im}(s(z)) = \frac{1}{2\pi}\left(-\eta + \sqrt{\frac{4+\eta^2-E^2+\sqrt{(E^2-\eta^2-4)^2+4E^2\eta^2}}{2}}\right).$$

Assuming at first that $E \in [-2, 2]$, we find

$$\begin{aligned}
&\left|\frac{1}{\pi}\operatorname{Im}(s(z)) - f_\sigma(E)\right| \\
&= \left|\frac{1}{2\pi}\left(-\eta + \sqrt{\frac{4+\eta^2-E^2+\sqrt{(E^2-\eta^2-4)^2+4E^2\eta^2}}{2}}\right) - \frac{1}{2\pi}\sqrt{4-E^2}\right| \\
&\leq \frac{\eta}{2\pi} + \frac{1}{2\pi}\left(\sqrt{\frac{4+\eta^2-E^2+\sqrt{(E^2-\eta^2-4)^2+4E^2\eta^2}}{2}} - \sqrt{4-E^2}\right).
\end{aligned}$$

Since $\sqrt{\cdot}$ is uniformly continuous, we analyze the difference of the arguments:

$$\begin{aligned}
&\left|\frac{4+\eta^2-E^2+\sqrt{(E^2-\eta^2-4)^2+4E^2\eta^2}-8+2E^2}{2}\right| \\
&= \frac{E^2-4+\eta^2+\sqrt{(E^2-\eta^2-4)^2+4E^2\eta^2}}{2} \\
&\leq \frac{E^2-4+\eta^2+4-E^2+\eta^2+2|E|\eta}{2} \\
&\leq \frac{2\eta^2+4\eta}{2} = \eta^2 + 2\eta
\end{aligned}$$

Since the modulus of continuity of $\sqrt{\cdot}$ is given by $\sqrt{\cdot}$ itself, we obtain

$$\left|\frac{1}{\pi}\operatorname{Im}(s(z)) - f_\sigma(E)\right| \leq \frac{\eta}{2\pi} + \frac{\sqrt{\eta^2+2\eta}}{2\pi} \leq \frac{\eta+\sqrt{\eta}}{\pi} \leq \frac{\sqrt{\eta}\sqrt{C}+\sqrt{\eta}}{\pi} \leq \frac{2\sqrt{C\eta}}{\pi} \leq \sqrt{C\eta}.$$

Now assuming that $E \in [-C, C] \backslash [-2, 2]$ we find

$$\begin{aligned}
&\left| \frac{1}{\pi} \operatorname{Im}(s(z)) - f_\sigma(E) \right| \\
&= \left| \frac{1}{2\pi} \left(-\eta + \sqrt{\frac{4 + \eta^2 - E^2 + \sqrt{(E^2 - \eta^2 - 4)^2 + 4E^2\eta^2}}{2}} \right) \right|
\end{aligned}$$

Then if $\eta^2 \leq E^2 - 4$, we have

$$\begin{aligned}
&4 + \eta^2 - E^2 + \sqrt{(E^2 - \eta^2 - 4)^2 + 4E^2\eta^2} \\
&\leq 4 + \eta^2 - E^2 + E^2 - \eta^2 - 4 + 2|E|\eta \\
&= 2|E|\eta \leq 2C\eta,
\end{aligned}$$

whereas if $\eta^2 > E^2 - 4$, we find

$$\begin{aligned}
&4 + \eta^2 - E^2 + \sqrt{(E^2 - \eta^2 - 4)^2 + 4E^2\eta^2} \\
&\leq 4 + \eta^2 - E^2 + 4 - E^2 + \eta^2 + 2|E|\eta \\
&= 2(\eta^2 - (E^2 - 4)) + 2|E|\eta \\
&\leq 2\eta^2 + 2C\eta,
\end{aligned}$$

where we used that $|E| > 2$ in the last step. We conclude that if $E \in [-C, C] \backslash [-2, 2]$, we obtain

$$\left| \frac{1}{\pi} \operatorname{Im}(s(z)) - f_\sigma(E) \right| \leq \frac{\eta}{2\pi} + \frac{1}{2\pi} \sqrt{\eta^2 + C\eta} \leq \frac{2\eta + \sqrt{C\eta}}{2\pi} \leq \frac{2\sqrt{C\eta} + \sqrt{C\eta}}{2\pi} \leq \sqrt{C\eta}.$$

So in total, we obtain

$$\sup_{E \in [-C, C]} \left| \frac{1}{\pi} \operatorname{Im}(s(E + i\eta)) - f_\sigma(E) \right| \leq \sqrt{C\eta}.$$

□

5.5 The Stieltjes Transform of ESDs of Hermitian Matrices

As we motivated the Stieltjes transform in the beginning of this chapter, it is possible to relate the Stieltjes transform of an ESD of a random matrix to the entries of the random matrix. We will now see how this is done. Notationally, as the Stieltjes transform of the semicircle distribution received the special letter $s := S_\sigma$, the Stieltjes transform of an ESD σ_n of an Hermitian $n \times n$ matrix X_n is denoted by $s_n := S_{\sigma_n}$. The following theorem summarizes the findings of this section (see also [7, pp. 470-472]):

Theorem 5.19. *Let X_n be an Hermitian $n \times n$ matrix with ESD σ_n.*

i) For all $z \in \mathbb{C}\backslash\mathbb{R}$ we find:

$$s_n(z) = S_{\sigma_n}(z) = \frac{1}{n}\operatorname{tr}(X_n - z)^{-1} = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{X_n(k,k) - z - x_k^*(X_n^{(k)} - z)^{-1}x_k}.$$

ii) For $z = E + i\eta$, where $E \in \mathbb{R}$ and $\eta > 0$, we obtain for all $k \in \{1, \ldots, n\}$:

$$\left|\operatorname{tr}(X_n - z)^{-1} - \operatorname{tr}(X_n^{(k)} - z)^{-1}\right| \leq \frac{1}{\eta}.$$

Here, $X_n^{(k)}$ denotes the k-th principal minor of X_n (thus an $(n-1)\times(n-1)$ matrix) and x_k the k-th column of X_n without the k-th entry (thus an $(n-1)$-vector).

Proof. *i)* The first equality is just a notational convention and the last equality is the statement of Corollary 5.23 below. For the second equality, let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of X_n, then by the spectral theorem for normal operators, $\frac{1}{\lambda_1 - z}, \ldots, \frac{1}{\lambda_n - z}$ are the eigenvalues of $(X_n - z)^{-1}$. Since for normal matrices, the trace yields the sum of the eigenvalues, we conclude

$$S_{\sigma_n}(z) = \int_{\mathbb{R}} \frac{1}{x - z}\sigma_n(\mathrm{d}x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z} = \frac{1}{n}\operatorname{tr}(X_n - z)^{-1}.$$

ii) This is the statement of Corollary 5.25 below. □

Note that Theorem 5.19 *i)* also allows us to work with the Stieltjes transform $S_{\mathbb{E}\sigma_n}$ of the expected ESD $\mathbb{E}\sigma_n$, since as in the proof of Theorem 5.6 we have $S_{\mathbb{E}\sigma_n} = \mathbb{E}S_{\sigma_n} = \mathbb{E}s_n$.

The remainder of this section will be devoted to the proof of Theorem 5.19, for which we follow the roadmap as in [7]. We begin by noting:

Lemma 5.20. *Let A be an $n \times n$ matrix with $\det(A) \neq 0$, then it holds for all $k \in \{1, \ldots, n\}$:*

$$A^{-1}(k,k) = \frac{\det A^{(k)}}{\det A},$$

where as before, $A^{(k)}$ denotes the k-th principle minor of A.

Proof. For all $i, j \in [n]$ we define $A^{(i,j)}$ to be the $(n-1)\times(n-1)$ matrix obtained from A through elimination of the i-th row and j-th column. Further, for all such (i,j) we let $C(i,j) := (-1)^{i+j}\det A^{(i,j)}$ be the (i,j)-th cofactor of A. We obtain an $n \times n$ matrix C of cofactors of A, and invoke the well-known identity $AC^T = \det(A)\cdot I_n$ (see [29, p. 204]). Multiplying by A^{-1} on the left and dividing by $\det(A)$ yields $A^{-1} = \det(A)^{-1}C^T$, from which the statement follows. □

In the following Lemma, the Schur complement is defined and studied (see also [69]).

Lemma 5.21. *Let*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

be a quadratic block matrix with A_{11} invertible. Then the Schur complement of A_{11} in A is defined as

$$B := A_{22} - A_{21} A_{11}^{-1} A_{12}$$

and has the following properties, where I resp. 0 are identity matrices resp. 0-matrices of appropriate dimension:

i) We obtain the Schur complement formula

$$\begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ 0 & B \end{pmatrix}.$$

ii) We find the Schur complement determinant formula

$$\det(A) = \det(A_{11}) \det(B) = \det(A_{11}) \det(A_{22} - A_{21} A_{11}^{-1} A_{12})$$

iii) If A is invertible, so is $B = A_{22} - A_{21} A_{11}^{-1} A_{12}$.

iv) In case A is invertible, we find the Schur complement inversion formula

$$\begin{aligned} A^{-1} &= \begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{pmatrix} \\ &= \begin{pmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}B^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}B^{-1} \\ -B^{-1}A_{21}A_{11}^{-1} & B^{-1} \end{pmatrix}. \end{aligned}$$

Proof. Statement *i*) requires mere verification through multiplication of the matrices, *ii*) follows directly from *i*) and *iii*) follows directly from *ii*). The first equality of statement *iv*) follows directly by inverting the Schur complement formula and multiplying from the left and right with the appropriate matrices. The second equality is again verified through simple multiplication of the matrices. □

Lemma 5.22. *Let A be an invertible $n \times n$ matrix. If $A^{(k)}$ is invertible for some $k \in \{1, \ldots, n\}$, then*

$$A^{-1}(k,k) = \frac{1}{A(k,k) - r_k A^{(k)-1} c_k},$$

where r_k is the k-th row of A without the k-th entry and c_k is the k-th column of A without the k-th entry.

Proof. We first prove the statement for $k = n$. We obtain

$$A = \begin{pmatrix} A^{(n)} & c_n \\ r_n & A(n,n) \end{pmatrix}$$

By the Schur complement determinant formula, we obtain $\det(A) = \det(A^{(n)})\det(A(n,n) - r_n A^{(n)-1} c_n)$. Therefore, with Lemma 5.20, we obtain

$$\frac{1}{A(n,n) - r_n A^{(n)-1} c_n} = \frac{\det(A^{(n)})}{\det(A)} = A^{-1}(n,n).$$

Next, we assume $k < n$. Then define a permutation matrix column-wise as

$$V := (e_1|e_2|\dots|\widehat{e_k}|\dots|e_n|e_k)$$

where the e_i are the standard n-dimensional basis vectors, and the hat over e_k indicates that this vector is left out. In other words, V is obtained through the identity matrix by erasing its k-th column e_k and appending it at the end of the matrix. We obtain immediately that $V^T = V = V^{-1}$. Then AV is the matrix A with erased and then appended k-th column and VA is the matrix A with erased and then appended k-th row. Therefore, $(VAV)^{(n)} = A^{(k)}$ and by the case above

$$A^{-1}(k,k) = (VA^{-1}V)(n,n) = (VAV)^{-1}(n,n) = \frac{1}{(VAV)(n,n) - r_n'(VAV)^{(n)-1}c_n'},$$

where r_n' denotes the n-th row of VAV and c_n' denotes the n-th column of VAV, both without their n-th entry. But $r_n' = r_k$, $c_n' = c_k$ and $(VAV)(n,n) = A(k,k)$. □

Corollary 5.23. *Let X_n be an Hermitian $n \times n$ matrix, then it holds for $z \in \mathbb{C}\backslash\mathbb{R}$:*

$$\operatorname{tr}(X_n - z)^{-1} = \sum_{k=1}^{n} \frac{1}{X_n(k,k) - z - x_k^*(X_n^{(k)} - z)^{-1} x_k}.$$

Proof. X_n and all $X_n^{(k)}$ are Hermitian, thus $X_n - z$ and $X_n^{(k)} - z = (X_n - z)^{(k)}$ are invertible for all k. We also know that the k-th column (resp. row) of X_n without the k-th entry is also the k-th column (resp. row) of $(X_n - z)$ without the k-th entry. Therefore, the statement follows directly with Lemma 5.22. □

Lemma 5.24. *Let A be an invertible $n \times n$ matrix and $k \in \{1,\dots,n\}$, such that $A^{(k)}$ is invertible. Then we obtain:*

$$\operatorname{tr} A^{-1} - \operatorname{tr} A^{(k)-1} = \frac{1 + r_k A^{(k)-2} c_k}{A(k,k) - r_k A^{(k)-1} c_k},$$

where r_k denotes the k-th row of A without the k-th entry and c_k denotes the k-th column of A without the k-th entry.

Proof. We first prove the statement vor $k = n$. The Schur complement inversion formula for

$$A = \begin{pmatrix} A^{(n)} & c_n \\ r_n & A(n,n) \end{pmatrix}$$

yields with $B := A(n,n) - r_n A^{(n)-1} c_n \in \mathbb{C}$, that

$$A^{-1} = \begin{pmatrix} A^{(n)-1} + A^{(n)-1} c_n B^{-1} r_n A^{(n)-1} & -A^{(n)-1} c_n B^{-1} \\ -B^{-1} r_n A^{(n)-1} & B^{-1} \end{pmatrix}.$$

Therefore, since the trace is linear and only depends on the diagonal block matrices, we find

$$\begin{aligned} \operatorname{tr} A^{-1} - \operatorname{tr} A^{(n)-1} &= \operatorname{tr} \begin{pmatrix} A^{(n)-1} c_n B^{-1} r_n A^{(n)-1} & 0 \\ 0 & B^{-1} \end{pmatrix} \\ &= \frac{1}{B} \operatorname{tr} \begin{pmatrix} A^{(n)-1} c_n r_n A^{(n)-1} & 0 \\ 0 & 1 \end{pmatrix} \\ &= \frac{1}{B} \left(1 + \sum_{k,l,m=1}^{n-1} A^{(n)-1}(k,l) c_n(l) r_n(m) A^{(n)-1}(m,k) \right) \\ &= \frac{1}{B} \left(1 + \sum_{k,l,m=1}^{n-1} r_n(m) A^{(n)-1}(m,k) A^{(n)-1}(k,l) c_n(l) \right) \\ &= \frac{1}{B} \left(1 + r_n A^{(n)-2} c_n \right), \end{aligned}$$

which concludes the statement for $k = n$. Now if $k < n$, let V be the permutation matrix as in the proof of Lemma 5.22, then since $A^{(k)} = (VAV)^{(n)}$, we obtain with first part that

$$\begin{aligned} \operatorname{tr} A^{-1} - \operatorname{tr} A^{(k)-1} &= \operatorname{tr} V A^{-1} V - \operatorname{tr} (VAV)^{(n)-1} \\ &= \operatorname{tr} (VAV)^{-1} - \operatorname{tr} (VAV)^{(n)-1} \\ &= \frac{1 + r_n'(VAV)^{(n)-2} c_n'}{(VAV)(n,n) - r_n'(VAV)^{(n)-2} c_n'} \end{aligned}$$

where r_n' (resp. c_n') is the n-th row (resp. column) of VAV without the n-th entry. This concludes the statement, since $r_n' = r_k$, $c_n' = c_k$, and $(VAV)(n,n) = A(k,k)$. □

Corollary 5.25. *Let X_n be an Hermitian $n \times n$ matrix, $z = E + i\eta$ where $E \in \mathbb{R}$ and $\eta > 0$, then we find for any $k \in \{1, \dots, n\}$:*

$$\left| \operatorname{tr}(X_n - z)^{-1} - \operatorname{tr}(X_n^{(k)} - z)^{-1} \right| \le \frac{1}{\eta},$$

where for all $k \in \{1, \dots, n\}$, $X_n^{(k)}$ denotes the k-th principal minor of X_n and x_k denotes the k-th column of X_n without the k-th entry.

Proof. By Lemma 5.24, we know that

$$\begin{aligned} \left| \operatorname{tr}(X_n - z)^{-1} - \operatorname{tr}(X_n^{(k)} - z)^{-1} \right| &= \left| \frac{1 + x_k^*(X_n^{(k)} - z)^{-2} x_k}{X_n(k,k) - z - x_k^*(X_n^{(k)} - z)^{-1} x_k} \right| \\ &\le \frac{|1 + x_k^*(X_n^{(k)} - z)^{-2} x_k|}{|-\eta - \operatorname{Im}(x_k^*(X_n^{(k)} - z)^{-1} x_k)|} \end{aligned}$$

where x_k denotes the k-th column of X_n without the k-th entry. We also used that $X_n(k,k) \in \mathbb{R}$, since X_n is Hermitian. We proceed by analyzing the numerator and denominator separately. Before we begin, consider the following: If T is a bounded normal operator on some Hilbert space and x is a vector, then $\|Tx\| = \|T^*x\|$, so

$$|\langle T^2x \mid x\rangle| = |\langle Tx \mid T^*x\rangle| \leq \|Tx\|\|T^*x\| = \|T^*x\|^2 = \langle T^*x \mid T^*x\rangle = \langle TT^*x \mid x\rangle \,.$$

Since $(X_n^{(k)} - z)^{-1}$ is normal, we obtain

$$|1 + x_k^*(X_n^{(k)} - z)^{-2}x_k| \leq 1 + x_k^*(X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}x_k.$$

Now let us turn to the denominator. To do so, we write $P := (X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}$, then P is a positive operator, and so is $\sqrt{P}$. Both P and $\sqrt{P}$ commute with $X_n^{(k)}$ as images of the continuous functional calculus for normal operators. Now we calculate:

$$\begin{aligned}
\left\langle (X_n^{(k)} - z)^{-1}x_k \mid x_k\right\rangle &= \left\langle (X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}(X_n^{(k)} - \bar{z})x_k \mid x_k\right\rangle \\
&= \left\langle P(X_n^{(k)} - \bar{z})x_k \mid x_k\right\rangle \\
&= \left\langle \sqrt{P}(X_n^{(k)} - \bar{z})x_k \mid \sqrt{P}x_k\right\rangle \\
&= \underbrace{\left\langle X_n^{(k)}\sqrt{P}x_k \mid \sqrt{P}x_k\right\rangle}_{\in\mathbb{R}} - \bar{z}\underbrace{\langle Px_k \mid x_k\rangle}_{\geq 0},
\end{aligned}$$

Therefore, $\operatorname{Im}(x_k^*(X_n^{(k)} - z)^{-1}x_k) = \eta x_k^*(X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}x_k$. It follows with what we have shown above that

$$\begin{aligned}
\left|\operatorname{tr}(X_n - z)^{-1} - \operatorname{tr}(X_n^{(k)} - z)^{-1}\right| &\leq \frac{1 + x_k^*(X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}x_k}{|-\eta - \eta x_k^*(X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}x_k|} \\
&= \frac{1 + x_k^*(X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}x_k}{\eta(1 + x_k^*(X_n^{(k)} - z)^{-1}(X_n^{(k)} - \bar{z})^{-1}x_k)},
\end{aligned}$$

from which the statement follows. □

5.6 Auxillary Statements

Integration of Complex-Valued Functions

We assume the reader to be acquainted with measure-theoretical integration of real-valued functions on measure spaces. Since we will also need to integrate complex-valued functions, we give a very short introduction in the form of one definition and two lemmata.

Definition 5.26. Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $f : (\Omega, \mathcal{A}) \to \mathbb{C}$ measurable, then f is called *μ-integrable*, if the real-valued functions $\operatorname{Re} f$ and $\operatorname{Im} f$ both are μ- integrable. In this case, we define

$$\int_\Omega f \mathrm{d}\mu := \int_\Omega \operatorname{Re} f \mathrm{d}\mu + i \int_\Omega \operatorname{Im} f \mathrm{d}\mu.$$

We will denote the space of $\mathbb{C}$-valued integrable functions as $\mathcal{L}_1(\mu, \mathbb{C})$.

It is worth noting the following lemma about the properties of the integral:

Lemma 5.27. *Let* $(\Omega, \mathcal{A}, \mu)$ *be a measure space, then the following statements hold:*

1. *The map* $\mathcal{L}_1(\mu, \mathbb{C}) \to \mathbb{C},\ f \mapsto \int f\, d\mu$ *is* $\mathbb{C}$*-linear.*
2. $\forall\, f \in \mathcal{L}_1(\mu, \mathbb{C}) : \overline{\int f\, d\mu} = \int \overline{f}\, d\mu.$
3. $\forall\, f \in \mathcal{L}_1(\mu, \mathbb{C}) : \left|\int f\, d\mu\right| \leq \int |f|\, d\mu.$

Proof. 1) follows by elementary calculations and 2) holds by the definition of the integral. To see 3), let $z \in \mathbb{C}$ with $|z| = 1$, such that $z \int f \mathrm{d}\mu = \left|\int f \mathrm{d}\mu\right|$, then it follows

$$\left|\int f \mathrm{d}\mu\right| = z \int f \mathrm{d}\mu = \int \mathrm{Re}(zf) \mathrm{d}\mu + i \underbrace{\int \mathrm{Im}(zf) \mathrm{d}\mu}_{=0} \leq \int |zf| \mathrm{d}\mu = \int |f| \mathrm{d}\mu.$$

□

Lemma 5.28 (Lebesgue's Dominated Convergence Theorem)**.** *Let* $(\Omega, \mathcal{A}, \mu)$ *be a measure space,* $(f_n)_n, f : \Omega \to \mathbb{C}$ *be measurable with* $f_n \to f$ μ*-almost everywhere. If there exists a* μ*-integrable* $g : \Omega \to \mathbb{R}$ *with* $|f_n| \leq g$ μ*-almost everywhere for all* n*, then* f *is* μ*-integrable and it holds*

$$\lim_{n\to\infty} \int |f - f_n|\, d\mu = 0,$$

so that in particular

$$\lim_{n\to\infty} \int f_n\, d\mu = \int f\, d\mu.$$

Proof. Certainly, $|\mathrm{Re}\, f_n|, |\mathrm{Im}\, f_n| \leq |f_n| \leq |g|$ and $\mathrm{Re}\, f_n \to \mathrm{Re}\, f$, $\mathrm{Im}\, f_n \to \mathrm{Im}\, f$ μ-almost everywhere. Also, $|f - f_n| \leq |\mathrm{Re}\, f - \mathrm{Re}\, f_n| + |\mathrm{Im}\, f - \mathrm{Im}\, f_n|$. Now for real-valued measurable functions, the theorem is assumed to be known. See [44, p. 142] for a reference. □

6 The Local Law for Curie-Weiss Type Ensembles

The local semicircle law is a rather recent result that was derived to gain a more detailed understanding of the convergence of the ESDs of random matrices to the semicircle distribution (and also to other deterministic and not necessarily known limiting distributions). Further, as mentioned in the introduction of this thesis, it was also used to establish universality results for Wigner matrices. In the literature, the local law has many shapes and forms, so it is hard to speak of "the local law." A common formulation of this type of theorem is a uniform alignment of the Stieltjes transforms of the ESDs and the semicircle distribution, see [9], for example. Reducing this statement to the imaginary parts and keeping in mind our exposition in Section 5.3, this will directly imply alignment of Lebesgue densities of the probability measures approximating the ESDs and the semicircle distribution. Another formulation of the local law is as follows, see [60]: For any sequence of intervals $(I_N)_N$, whose diameter is not decaying to zero too quickly, $\sigma_N(I_N)$ can be well approximated by $\sigma(I_N)$ for large N. In fact, the second formulation of the local law will follow from the first, as we will show further below. But there are even more versions of the local law that also allow conclusions about the asymptotic behaviour of eigenvectors of random matrices. These types will not be treated in this thesis, but is rather ongoing work as this thesis is handed in.

Although there were some previous results into the direction of a local law in [40] and [26], it is safe to say that on the level of strength available today, it first appeared in [25] by Erdős, Schlein and Yau. Ever since, the results were strengthened (see [35], for example) and proof layouts were refined to make the theory more accessible to a broader audience. Indeed, although there are areas of lesser gravity in probability theory - both in power and complexity of concepts and proofs - the local laws are displayed in an accessible manner in the text [9] by Benaych-Georges and Knowles and the book [27] by Erdős and Yau. Both of these texts are in turn based on their joint publication [28].

As the semicircle law itself, the local semicircle law was initially considered for matrices with independent and identically distributed entries in [25]. Further generalizations can be found in [28], where entries are still assumed to be independent, but not identically distributed anymore.

Of course, the next question is if and how local laws can also be proved for random matrices with correlated entries. Even up to today, these ensembles are not well understood in terms of the local law. In [1], the local law was proved for random matrices with correlated Gaussian entries, where the covariance matrix is assumed to posses a certain translation invariant structure. In [2], ensembles with correlated entries were considered, where the correlation decays arbitrarily polynomially fast in the distance of the entries. This result has been improved by [24] (who reference an older preprint version of [2]), where fast polynomial decay is assumed only for entries outside of neighborhoods of a size growing slower than $\sqrt{N}$, and a slower correlation decay between entries within these neighborhoods. Meanwhile, another correlation structure was analyzed in [17], where correlation was only allowed for entries close to each other and independence was assumed otherwise. What all four mentioned publications have in common is that their results are formulated in such a general manner that the limiting distribution of the local law need not be the semicircle distribution. They all require additional analysis to derive a local *semicircle* law.

In this chapter, we will derive a semicircle law of the weak type for a random matrix ensemble with very slow correlation decay for entries that are arbitrarily close or far apart. In fact, our correlation decay between any two different matrix entries (regardless of their distance) in the upper right half of the matrix will be of order N^{-1}. In particular, our model is not covered by the previous work on correlated entries that was mentioned above (for example, in [24], Assumption (D) is violated), and new proof techniques must be developed.

The ensemble we study will be called "of Curie-Weiss type", and not surprisingly, Curie-Weiss(β)-distributed entries will be admissible (as long as $\beta \leq 1$). For our proof we need to establish new sets of so called large-deviation inequalities. We state and prove them in such a general manner that they can also be used to derive stronger local laws.

The second goal of this chapter is to turn the proofs of the local law that are available in the literature so far into a proof that is both complete and structured in a way that is easy to follow by a broader audience. To this end, we have molded the expositions in [9] and [27] into one comprehensible and complete proof.

Notationally, we would like to point out that throughout this chapter, the dimension parameter n will be capitalized (written as N), which has some notational advantages.

6.1 De-Finetti Type Random Variables

In this section we introduce random variables of de-Finetti type. It should be noted that the expectation operator $\mathbb{E}$ will always denote the expectation with respect to the generic probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In addition, probability spaces with finite sample space will always be equipped with the power set as σ-algebra. Further, if I is an index set and for all $i \in I$, Z_i is a mathematical object, then we write $Z_I \equiv (Z_i)_{i\in I}$ for better readability. Now, we proceed as in [41].

Definition 6.1. Let I be a finite index set and Y_I be a family of $\{\pm 1\}$-valued random variables on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Further, let μ be a probability measure on the interval $[-1, 1]$ equipped with its Borel σ-algebra. Then the random vector Y_I is called *of de-Finetti type with mixture μ,* if for all configurations $y_I \in \{\pm 1\}^I$ we have

$$\mathbb{P}(Y_I = y_I) = \int_{[-1,1]} P_t^{\otimes I}(y_I) \mathrm{d}\mu(t), \tag{6.1}$$

where $P_t^{\otimes I} := \otimes_{i \in I} P_t$ for all $t \in [-1,1]$ and P_t is the probability measure on $\{\pm 1\}$ with

$$P_t(1) = \frac{1+t}{2} \qquad \text{and} \qquad P_t(-1) = \frac{1-t}{2}.$$

In particular, $P_t^{\otimes I}$ is the I-fold product measure of P_t on $\{\pm 1\}^I$.

Remark 6.2. We would like to make several observations concerning Definition 6.1:

1. As becomes clear from the definition, for $\{\pm 1\}$ -valued random variables to be of de-Finetti type is solely a property of their distribution and *not* a property concerning the specific construction of the variables or the probability space on which they are defined.

2. If Y_I is of de-Finetti type with mixture μ, so is Y_J for any subset $J \subseteq I$.

3. It is interesting to point out that if the variables Y_I are of de-Finetti type, then their distribution is a mixture of product distributions with identically distributed coordinates. Further, the expectation of the measure P_t is t. Thus, coordinates of the identity map on the probability space $(\{\pm 1\}^I, P_t^{\otimes I})$ are i.i.d. P_t-distributed with expectation t.

4. From Definition 6.1 it is clear that a family Y_I of de-Finetti-type is *exchangeable*, that is, if $\pi : I \to I$ is a bijection, then $(Y_i)_{i \in I}$ and $(Y_{\pi(i)})_{i \in I}$ have the same distribution.

Lemma 6.3. *Let Y_I be of de-Finetti type with mixture μ as in Definition 6.1. Then we obtain for any function $F : \{\pm 1\}^I \to \mathbb{C}$ that*

$$\mathbb{E} F(Y_I) = \int_{[-1,1]} \int_{\{\pm 1\}^I} F(y_I)\, dP_t^{\otimes I}(y_I)\, d\mu(t).$$

Further, if $J, K \subseteq I$ are nonempty with $J \cap K = \emptyset$ and $J \cup K = I$, then

$$\mathbb{E} F(Y_{J \cup K}) = \int_{[-1,1]} \int_{\{\pm 1\}^K} \int_{\{\pm 1\}^J} F(y_J, y_K)\, dP_t^{\otimes J}(y_J)\, dP_t^{\otimes K}(y_K)\, d\mu(t).$$

Proof. This is straightforward considering

$$\mathbb{E} F(Y_I) = \sum_{y_I \in \{\pm 1\}^I} F(y_I) \mathbb{P}(Y_I = y_I).$$

and applying Fubini. $\square$

It is clear that the second statement of Lemma 6.3 can be generalized to an arbitrary number of pairwise disjoint nonempty subsets $J_1, J_2, \ldots, J_l$ whose union is I. However, the case $l = 2$ already captures the essence of the technique.

We now study a prominent example for random variables of de-Finetti type, see Definition 4.16.

Theorem 6.4. *If $Y_1, \dots, Y_N$ are Curie-Weiss(β,N)-distributed, then they are of de-Finetti-type with mixture μ_N^β, which is Lebesgue-continuous with density on $(-1,1)$ given by*

$$t \mapsto f_N(t) := \frac{1}{\int_{(-1,1)} \frac{e^{-\frac{N}{2}F_\beta(s)}}{1-s^2} \lambda\!\!\lambda(ds)} \frac{e^{-\frac{N}{2}F_\beta(t)}}{1-t^2} \mathbb{1}_{(-1,1)}(t),$$

where for all $t \in (-1,1)$ we define

$$F_\beta(t) := \frac{1}{\beta}\left(\frac{1}{2}\ln\left(\frac{1+t}{1-t}\right)\right)^2 + \ln(1-t^2).$$

Further, if $\beta \leq 1$, the mixtures $(\mu_N^\beta)_{N\in\mathbb{N}}$ satisfy the following moment decay:

$$\forall\ p \in 2\mathbb{N} : \int_{[-1,+1]} t^p \mu_N^\beta(dt) \leq \frac{K_{\beta,p}}{N^{\frac{p}{4}}},$$

where $K_{\beta,p} \in \mathbb{R}_+$ is a constant that depends on β and p only.

Proof. This was shown rigorously in [41], see Theorem 5.6, Remark 5.7, Proposition 5.9 and Theorem 5.17 in their text. □

6.2 Stochastic Domination

For the statement of the weak local law and its proof we need the concept of stochastic domination. The following exposition is based on [27]. The first time that this concept was used was in [23].

For the remainder of this thesis, we will say that a statement which depends on $N \in \mathbb{N}$ holds v-finally, where v is a parameter or a parameter-vector, if the statement holds for all $N \geq N^*$, where $N^* \in \mathbb{N}$ depends on v. We will also write $N^* = N(v)$ and say the statement holds for all $N \geq N(v)$ in this case.

Definition 6.5. Let $X = X^{(N)}$ be a sequence of complex-valued and $Y = Y^{(N)}$ be a sequence of non-negative random variables, then we say that X is *stochastically dominated* by Y, if for all $\epsilon, D > 0$ there is a constant $C_{\epsilon,D} \geq 0$ such that

$$\forall\ N \in \mathbb{N} : \mathbb{P}\left(|X^{(N)}| > N^\epsilon Y^{(N)}\right) \leq \frac{C_{\epsilon,D}}{N^D}.$$

In this case, we write

$$X \prec Y \quad \text{or} \quad X^{(N)} \prec Y^{(N)}.$$

If both X and Y depend on a possibly N-dependent index set $U = U^{(N)}$, such that

$$X = \left(X^{(N)}(u), N \in \mathbb{N}, u \in U^{(N)}\right), \quad Y = \left(Y^{(N)}(u), N \in \mathbb{N}, u \in U^{(N)}\right),$$

then we say that X is stochastically dominated by Y *uniformly in* $u \in U$, if for all $\epsilon, D > 0$ we can find a $C_{\epsilon,D} \geq 0$ such that

$$\forall N \in \mathbb{N} : \sup_{u \in U^{(N)}} \mathbb{P}\left(|X^{(N)}(u)| > N^{\epsilon} Y^{(N)}(u)\right) \leq \frac{C_{\epsilon,D}}{N^D}. \tag{6.2}$$

In this case, we write

$$X \prec Y \quad \text{or} \quad X(u) \prec Y(u), u \in U \quad \text{or} \quad X^{(N)}(u) \prec Y^{(N)}(u), u \in U^{(N)},$$

where the first version is used if U is clear from the context. If in above situation, all $Y(u)$ are strictly positive, then we say that X is stochastically dominated by Y, *simultaneously in* $u \in U$, if for all $\epsilon, D > 0$ we can find a $C_{\epsilon,D} \geq 0$, such that

$$\forall N \in \mathbb{N} : \mathbb{P}\left(\sup_{u \in U^{(N)}} \frac{|X^{(N)}(u)|}{Y^{(N)}(u)} > N^{\epsilon}\right) \leq \frac{C_{\epsilon,D}}{N^D},$$

and then we write

$$\sup_{u \in U} \frac{|X(u)|}{Y(u)} \prec 1 \quad \text{or} \quad \sup_{u \in U^{(N)}} \frac{|X^{(N)}(u)|}{Y^{(N)}(u)} \prec 1.$$

Remark 6.6. We make the following important observations with regards to Definition 6.5:

1. Simultaneous stochastic domination implies uniform stochastic domination.

2. The intuition of stochastic domination is that if $X \prec Y$ then X does not grow faster in N than any N^{ϵ}-multiple of Y.

3. If (6.2) holds for all $N \geq N(\epsilon, D)$, then also for all $N \in \mathbb{N}$ after rescaling $C_{\epsilon,D}$. Therefore, it suffices to show (6.2) (ϵ, D)-finally. To validate our claim, suppose we have shown (6.2) to hold for all $N \geq N(\epsilon, D)$, then replace $C_{\epsilon,D}$ by $\max(C_{\epsilon,D}, N(\epsilon, D)^D)$ which is also a constant depending only on ϵ and D, making (6.2) valid for all $N \in \mathbb{N}$, since the left hand side is bounded by 1.

4. In order to show $X \prec Y$, it suffices to show that (6.2) holds for all ϵ small enough. To be more precise, if there is an $\epsilon_0 > 0$, such that for all $\epsilon \in (0, \epsilon_0]$ and $D > 0$ we find a constant $C_{\epsilon,D} \geq 0$ such that (6.2) holds, then clearly for all $\epsilon > \epsilon_0$ and $N \in \mathbb{N}$:

   $$\sup_{u \in U^{(N)}} \mathbb{P}\left(|X^{(N)}(u)| > N^{\epsilon} Y^{(N)}(u)\right) \leq \sup_{u \in U^{(N)}} \mathbb{P}\left(|X^{(N)}(u)| > N^{\epsilon_0} Y^{(N)}(u)\right) \leq \frac{C_{\epsilon_0,D}}{N^D}.$$

   Therefore, for all $\epsilon \geq \epsilon_0$ and $D > 0$ the constant $C_{\epsilon_0,D}$ can be utilized.

5. Another characterization of $\prec$ is often used in the literature (see [9] or [27]), that is, $X \prec Y$ holds if and only if for any $\epsilon, D > 0$ there exists an $N(\epsilon, D) \in \mathbb{N}$ such that

   $$\forall N \geq N(\epsilon, D) : \mathbb{P}\left(|X^{(N)}| > N^{\epsilon} Y^{(N)}\right) \leq \frac{1}{N^D}.$$

Let us validate the equivalence: If the condition holds then $X \prec Y$ is satisfied with constants $C_{\epsilon,D} := N(\epsilon, D)^D$, and now assume that we have $X \prec Y$, then let $\epsilon, D > 0$ be arbitrary, then there exists a constant $C_{\epsilon,D+1} \geq 0$ such that

$$\forall\ N \in \mathbb{N} : \mathbb{P}\left(|X^{(N)}| > N^{\epsilon} Y^{(N)}\right)\right) \leq \frac{C_{\epsilon,D+1}}{N^{D+1}} = \frac{1}{N^D} \frac{C_{\epsilon,D+1}}{N}.$$

so that the alternative condition holds for all $N \geq N(\epsilon, D) := \lceil C_{\epsilon,D+1} \rceil$, where $\lceil \cdot \rceil$ is the ceiling function.

Stochastic domination admits several important and intuitive rules of calculation, which we collect in the following lemma. It supplements the findings in the literature, see [9] and [27]:

Lemma 6.7. *Let X, X_1, X_2 be $\mathbb{C}$-valued and $(W_i)_{i\in I}, Y, Y_1, Y_2, Z$ be $\mathbb{R}_+$-valued random variables, all depending on $N \in \mathbb{N}$ and $u \in U^{(N)}$ as in Definition 6.5. Further, the index set I shall depend on N with $|I| \leq C \cdot N^k$ for some fixed $C \geq 0$ and $k \in \mathbb{N}$. Then the following holds:*

i) If $X \prec Y$ and $Y \prec Z$, then $X \prec Z$.

ii) If $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then $X_1 + X_2 \prec Y_1 + Y_2$.

iii) If $X_1 \prec Y_1$ and $X_2 \prec Y_2$, then $X_1 \cdot X_2 \prec Y_1 \cdot Y_2$.

iv) If $W_i \prec Z\ \forall\, i \in I$, and if the constants $C_{\epsilon,D}$ for $\prec$ can be chosen independently of $i \in I$, then $\max_{i\in I} W_i \prec Z$.

v) If $Y \leq Z$, then $Y \prec Z$. In particular, $Y \prec Y$.

vi) If $Y \prec Z$ and $p > 0$, then $Y^p \prec Z^p$.

vii) If $X \prec Y$ and $c > 0$, then $X \prec cY$.

viii) If for all $N \in \mathbb{N}$, $U_1^{(N)}$ and $U_2^{(N)}$ are subsets of $U^{(N)}$, then if $X(u) \prec Y(u), u \in U_1^{(N)}$ and $X(u) \prec Y(u), u \in U_2^{(N)}$, then also $X(u) \prec Y(u), u \in (U_1^{(N)} \cup U_2^{(N)})$.

Proof. Notationally, we drop the N from all N-dependent quantities (except for N itself, which would be disastrous). Let $\epsilon, D > 0$ be arbitrary.

i) We know the following:

$$X \prec Y \Rightarrow \forall\, N \in \mathbb{N} : \sup_{u\in U} \mathbb{P}(|X(u)| > N^{\epsilon/2} Y(u)) \leq C^{(1)}_{\epsilon/2,D} N^{-D}$$

$$Y \prec Z \Rightarrow \forall\, N \in \mathbb{N} : \sup_{u\in U} \mathbb{P}(Y(u) > N^{\epsilon/2} Z(u)) \leq C^{(2)}_{\epsilon/2,D} N^{-D}.$$

We conclude that for all $N \in \mathbb{N}$ and with $C^*_{\epsilon,D} \coloneqq C^{(1)}_{\epsilon/2,D} + C^{(2)}_{\epsilon/2,D}$, we obtain for any $u \in U$ (which we drop from the notation here and later in this proof):

$$\begin{aligned}\mathbb{P}(|X| > N^\epsilon Z) &\le \mathbb{P}(|X| > N^{\epsilon/2} Y \text{ or } Y > N^{\epsilon/2} Z)\\ &\le \mathbb{P}(|X| > N^{\epsilon/2} Y) + \mathbb{P}(Y > N^{\epsilon/2} Z)\\ &\le C^{(1)}_{\epsilon/2,D} N^{-D} + C^{(2)}_{\epsilon/2,D} N^{-D} = C^*_{\epsilon,D} N^{-D}\end{aligned}$$

ii) We know the following:

$$\begin{aligned}X_1 \prec Y_1 &\Rightarrow \forall\, N \in \mathbb{N} : \sup_{u \in U} \mathbb{P}(|X_1(u)| > N^\epsilon Y_1(u)) \le C^{(1)}_{\epsilon,D} N^{-D}\\ X_2 \prec Y_2 &\Rightarrow \forall\, N \in \mathbb{N} : \sup_{u \in U} \mathbb{P}(|X_2(u)| > N^\epsilon Y_2(u)) \le C^{(2)}_{\epsilon,D} N^{-D}\end{aligned}$$

We conclude that for all $N \in \mathbb{N}$ and with $C^*_{\epsilon,D} \coloneqq C^{(1)}_{\epsilon,D} + C^{(2)}_{\epsilon,D}$, we obtain for any $u \in U$:

$$\begin{aligned}\mathbb{P}(|X_1 + X_2| > N^\epsilon (Y_1 + Y_2)) &\le \mathbb{P}(|X_1| + |X_2| > N^\epsilon Y_1 + N^\epsilon Y_2)\\ &\le \mathbb{P}(|X_1| > N^\epsilon Y_1 \text{ or } |X_2| > N^\epsilon Y_2)\\ &\le \mathbb{P}(|X_1| > N^\epsilon Y_1) + \mathbb{P}(|X_2| > N^\epsilon Y_2)\\ &\le C^{(1)}_{\epsilon,D} N^{-D} + C^{(2)}_{\epsilon,D} N^{-D} = C^*_{\epsilon,D} N^{-D}.\end{aligned}$$

iii) We know the following:

$$\begin{aligned}X_1 \prec Y_1 &\Rightarrow \forall\, N \in \mathbb{N} : \sup_{u \in U} \mathbb{P}(|X_1(u)| > N^{\epsilon/2} Y_1(u)) \le C^{(1)}_{\epsilon/2,D} N^{-D}\\ X_2 \prec Y_2 &\Rightarrow \forall\, N \in \mathbb{N} : \sup_{u \in U} \mathbb{P}(|X_2(u)| > N^{\epsilon/2} Y_2(u)) \le C^{(2)}_{\epsilon/2,D} N^{-D}.\end{aligned}$$

We conclude that for all $N \in \mathbb{N}$ and with $C^*_{\epsilon,D} \coloneqq C^{(1)}_{\epsilon/2,D} + C^{(2)}_{\epsilon/2,D}$, we obtain for any $u \in U$:

$$\begin{aligned}\mathbb{P}(|X_1||X_2| > N^\epsilon Y_1 Y_2) &\le \mathbb{P}(|X_1| > N^{\epsilon/2} Y_1 \text{ or } |X_2| > N^{\epsilon/2} Y_2)\\ &\le C^{(1)}_{\epsilon/2,D} N^{-D} + C^{(2)}_{\epsilon/2,D} N^{-D} = C^*_{\epsilon,D} N^{-D}.\end{aligned}$$

iv) We know the following, considering all W_i are $\mathbb{R}_+$-valued:

$$[\forall\, i \in I : W_i \prec Z] \Rightarrow \forall\, N \in \mathbb{N} : \forall\, i \in I : \sup_{u \in U} \mathbb{P}(W_i(u) > N^\epsilon Z(u)) \le C_{\epsilon,D+k} N^{-D-k}$$

We conclude that for all $N \in \mathbb{N}$ and with $C^*_{\epsilon,D} \coloneqq C \cdot C_{\epsilon,D+k}$, we obtain for any $u \in U$:

$$\begin{aligned}\mathbb{P}(\max_{i \in I} W_i > N^\epsilon Z) &= \mathbb{P}(\exists i \in I : W_i > N^\epsilon Z)\\ &\le \sum_{i \in I} \mathbb{P}(W_i > N^\epsilon Z)\\ &\le C \cdot N^k \cdot C_{\epsilon,D+k} \cdot N^{-D-k} = C^*_{\epsilon,D} N^{-D}.\end{aligned}$$

v) This is immediate, since $\mathbb{P}(Y(u) > N^\epsilon Z(u)) = 0$ for all $N \in \mathbb{N}, u \in U$.

vi) We find for $N \in \mathbb{N}$ and $u \in U$:

$$\mathbb{P}(Y(u)^p > N^\epsilon Z(u)^p) = \mathbb{P}(Y(u) > N^{\epsilon/p} Z(u)) \leq \frac{C_{\epsilon/p,D}}{N^D},$$

thus the statement holds with constants $(\epsilon, D) \mapsto C_{\epsilon/p,D}$.

vii) For $c \geq 1$, the statement is clear, since then $Y \leq cY$, hence $Y \prec cY$ by v), and now i). If $c < 1$, then there is an $N(\epsilon) \in \mathbb{N}$, such that $cN^{\epsilon/2} \geq 1$ for all $N \geq N(\epsilon)$. This entails that for all $N \geq N(\epsilon)$ and $u \in U$:

$$\mathbb{P}(|X| > N^\epsilon cY) \leq \mathbb{P}(|X| > N^{\epsilon/2} Y) \leq \frac{C_{\epsilon/2,D}}{N^D}.$$

Therefore, the constants $(\epsilon, D) \mapsto C_{\epsilon/2,D}$ can be used ϵ-finally. Now consider Remark 6.6.

viii) Denote by $C^{(1)}_{\epsilon,D}$ the constants for $\prec$ over $U_1^{(N)}$ and by $C^{(2)}_{\epsilon,D}$ the constants for $\prec$ over $U_2^{(N)}$, then $C_{\epsilon,D} := \max(C^{(1)}_{\epsilon,D}, C^{(2)}_{\epsilon,D})$ will clearly yield valid constants for $\prec$ over $U_1^{(N)} \cup U_2^{(N)}$. □

6.3 The Weak Local Law and its Consequences

Definition 6.8. An ensemble of real symmetric random matrices $(X_N)_N$ is called *of Curie-Weiss type*, if the following holds:

1. For all $N \in \mathbb{N}$, the random variables $(\sqrt{N} X_N(i,j))_{1 \leq i \leq j \leq N}$ are of de-Finetti type with mixture μ_N.
2. The sequence of mixtures $(\mu_N)_N$ satisfies the moment decay condition

$$\forall\, p \in 2\mathbb{N} : \exists\, K_p \in \mathbb{R}_+ : \forall\, N \in \mathbb{N} : \int_{[-1,1]} t^p \mathrm{d}\mu_N(t) \leq \frac{K_p}{N^{p/2}}. \tag{6.3}$$

Example 6.9. Let $0 < \beta \leq 1$ and let for each $N \in \mathbb{N}$ the random variables $(\tilde{a}_N(i,j))_{i,j \in [N]}$ be Curie-Weiss(β, N^2)-distributed. Define the ensemble $(X_N)_N$ by setting

$$\forall\, N \in \mathbb{N} : \forall\, (i,j) \in \square_N : \; X_N(i,j) = \begin{cases} \frac{1}{\sqrt{N}} \tilde{a}_N(i,j) & \text{if } i \leq j \\ \frac{1}{\sqrt{N}} \tilde{a}_N(j,i) & \text{if } i > j. \end{cases}.$$

Then by Theorem 6.4, $(X_N)_N$ is an ensemble of Curie-Weiss type with mixtures $(\mu_N)_N := (\mu^\beta_{N^2})_N$ and constants $K_p := K_{\beta,p}$.

The local law, which we will formulate in a moment, is about a locally uniform approximation of the ESDs of the Curie-Weiss type ensemble by the semicircle distribution in terms of their Stieltjes transforms. Since the semicircle distribution has compact support on $[-2, 2]$, it suffices to consider a region around this interval, say, $[-10, 10]$ for the real part of the domain and a moderately falling imaginary part, so to be able to obtain better and better Stieltjes transform kernel density estimates, see Section 5.3.

We will proceed to define certain quantities and regions which will be needed in the sequel. Whenever a $z \in \mathbb{C}$ is considered, we will denote by E its real part and by η its imaginary part and denote by κ the minimal distance of E to -2 or 2. To be more precise, if a $z \in \mathbb{C}$ is considered,

$$\begin{aligned} E &= E(z) = \mathrm{Re}(z) \\ \eta &= \eta(z) = \mathrm{Im}(z) \\ \kappa &= \kappa(z) = ||E| - 2|. \end{aligned}$$

Further, for all $N \in \mathbb{N}$ and $\gamma \in (0,1)$ we define the domains

$$\begin{aligned} \mathcal{D}_I &:= \{z \in \mathbb{C} \mid -10 \leq E \leq 10,\ 1 \leq \eta \leq 10\}, \\ \mathcal{D}_N(\gamma) &:= \left\{z \in \mathbb{C} \mid -10 \leq E \leq 10,\ \frac{1}{N^{1-\gamma}} \leq \eta \leq 10\right\}, \\ \mathcal{D}_N &:= \left\{z \in \mathbb{C} \mid -10 \leq E \leq 10,\ \frac{1}{N} \leq \eta \leq 10\right\}. \end{aligned}$$

For all $N \in \mathbb{N}$ we find $\mathcal{D}_I \subseteq \mathcal{D}_N(\gamma) \subseteq \mathcal{D}_N$. The region $\mathcal{D}_I$ will be used for an initial estimate, and quantities to be analyzed will behave nicely here. The region $\mathcal{D}_N(\gamma)$ will be used in the formulation of the local law and is thus the main region of interest. The region $\mathcal{D}_N$ covers all regions $\mathcal{D}_N(\gamma)$ with $\gamma \in (0,1)$ and will serve as a domain on which continuity properties of certain functions of interest will be proven (so that we know they hold on all regions $\mathcal{D}_N(\gamma)$).

We will now turn to our main theorem of this chapter. In formulation and proof, our weak local law is closer to [27] than to [9].

Theorem 6.10 (Weak Local Law for Curie-Weiss Type Ensembles). *Let $\gamma \in (0,1)$ be fixed and $(X_N)_N$ be an ensemble of Curie-Weiss type. Further, denote by s_N the Stieltjes transform of the empirical spectral distribution σ_N of X_N and by s the Stieltjes transform of the semicircle distribution σ. Then we obtain*

$$|s_N(z) - s(z)| \prec \min\left\{\frac{1}{\sqrt{N\eta\kappa}}, \frac{1}{(N\eta)^{\frac{1}{4}}}\right\}, \ z \in \mathcal{D}_N(\gamma).$$

To interpret the weak local law very roughly, it ensures that on $\mathcal{D}_N(\gamma)$ and up to a factor of N^ϵ, $|s_N(z) - s(z)|$ is bounded by $(N\eta)^{-1/4}$, and a minimal distance away from the edges of the bulk -2 and 2, the bound sharpens to $(N\eta)^{-1/2}$.

Before turning to the proof of the theorem, we will discuss some of its consequences. The simplest corollary is perhaps:

Corollary 6.11. *In the setting of Theorem 6.10 we obtain that $\sigma_N \to \sigma$ weakly in probability and weakly almost surely.*

Proof. Fix a $z \in \mathcal{D}_I$, then any $\eta \geq 1$, Theorem 6.10 implies that for $\epsilon = 1/8$ and $D = 2$ we obtain a constant $C_{\epsilon,D} \geq 0$ such that

$$\mathbb{P}\left(|s_N(z) - s(z)| > \frac{N^{\frac{1}{8}}}{N^{\frac{1}{4}}}\right) \leq \frac{C_{\epsilon,D}}{N^2}.$$

This inequality certainly implies $s_N(z) \to s(z)$ in probability, but by Borel-Cantelli also almost surely. Since $z \in \mathcal{D}_I$ was arbitrary, Theorem 5.6 yields that $\sigma_N \to \sigma$ in probability and almost surely. □

Now, let us strengthen Theorem 6.10, which is a statement about the supremum of certain probabilities. As it turns out, this supremum can be taken inside the probability, which is possible due to the Lipschitz continuity of all quantities involved in the statement. This will imply that $\prec$ does not only hold uniformly for $z \in \mathcal{D}_N(\gamma)$, but simultaneously for these z (cf. Definition 6.5).

The following theorem is far-reaching and can even be used to prove uniformity in the statement of stronger local laws (Theorem 2.6 in [9], for example). To state it in a general manner (in which we need it), we define for any sequence of regions $\mathcal{G}_N \subseteq \mathcal{D}_N$ the subsets

$$\mathcal{G}_N^L := \mathcal{G}_N \cap \frac{1}{N^L}(\mathbb{Z} + i\mathbb{Z}).$$

For example, we will set $\mathcal{G}_N := \mathcal{D}_N(\gamma)$ for all $N \in \mathbb{N}$ and consider the sets $\mathcal{G}_N^4$. At another point, we will set $\mathcal{G}_N := \mathcal{D}_I$ for all $N \in \mathbb{N}$ and consider the sets $\mathcal{G}_N^4$. We notice that in both cases, $\mathcal{G}_N^4$ forms a $\frac{2}{N^4}$-net in $\mathcal{G}_N$ (if $A \subseteq B \subseteq \mathbb{C}$ are subsets and $\tau > 0$, then A is called τ-net in B, if for any $b \in B$ there is an $a \in A$ such that $|b - a| \leq \tau$). In formulation and proof, the following theorem strongly generalizes Remark 2.7 in [9].

Theorem 6.12. *Suppose we are given stochastic domination of the form*

$$F_i^{(N)}(z) \prec \Psi^{(N)}(z), \qquad i \in I_N, z \in \mathcal{G}_N^L,$$

where for all $N \in \mathbb{N}$:

- *$\mathcal{G}_N \subseteq \mathcal{D}_N$ is a non-empty subset with a geometry such that $\mathcal{G}_N^L$ forms a $\frac{2}{N^L}$-net in $\mathcal{G}_N$.*
- *$(F_i^{(N)})_{i \in I_N}$ is a family of complex-valued functions on $\mathcal{D}_N$, where $\#I_N \leq C_1 N^{d_1}$ and for all $i \in I_N$, $F_i^{(N)}$ is $C_2 N^{d_2}$-Lipschitz-continuous on $\mathcal{D}_N$,*
- *$\Psi^{(N)}$ is an $\mathbb{R}_+$-valued function on $\mathcal{D}_N$, which is $C_3 N^{d_3}$-Lipschitz-continuous and bounded from below by $\frac{1}{C_4 N^{d_4}}$,*

where $C_1, \ldots, C_4 > 0$ and $d_1, \ldots, d_4 > 0$ are N-independent constants and $L > \max(d_2 + d_4,\ d_3 + d_4)$.

Then we obtain the simultaneous statement:

$$\sup_{z \in \mathcal{G}_N} \max_{i \in I_N} \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} \prec 1. \tag{6.4}$$

Proof. The following statements hold trivially for all $N \in \mathbb{N}$:

i) $\#\mathcal{G}_N^L \leq \#\mathcal{D}_N^L \leq 21N^L \cdot 11N^L = C_5 N^{2L}$,

ii) $\forall\, z \in \mathcal{G}_N : \exists\, z' \in \mathcal{G}_N^L : |z - z'| \leq \frac{2}{N^L}$,

where we set $C_5 \coloneqq 231$.
Step 1: (6.4) holds if $\mathcal{G}_N$ is replaced by $\mathcal{G}_N^L$.
This is easily done by the following calculation for $\epsilon, D > 0$ arbitrary:

$$\mathbb{P}\left(\sup_{z\in\mathcal{G}_N^L} \max_{i\in I_N} \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^\epsilon\right) \leq \sum_{z\in\mathcal{G}_N^L} \sum_{i\in I_N} \mathbb{P}\left(\frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^\epsilon\right) \leq C_5 N^{2L} C_1 N^{d_1} \frac{C_{\epsilon,D}}{N^D}$$

This concludes the first step by shifting $D \rightsquigarrow D + 2L + d_1$ and absorbing $C_1 \cdot C_5$ into $C_{\epsilon,D+2L+d_1}$. Note that we did not use any Lipschitz-continuity yet, but rather the union bound and the polynomial growth of the index sets.
Step 2: Extension from $\mathcal{G}_N^L$ to $\mathcal{G}_N$.
Now, Lipschitz-continuity comes into play: For an arbitrary $\epsilon > 0$, suppose

$$\exists z \in \mathcal{G}_N,\, \exists i \in I_N : |F_i^{(N)}(z)| > \Psi^{(N)}(z) N^\epsilon.$$

Then there exists a $z' \in \mathcal{G}_N^L$ with $|z - z'| \leq \frac{2}{N^L}$, and then due to Lipschitz-continuity of $F_i^{(N)}$:

$$|F_i^{(N)}(z')| > \Psi^{(N)}(z) N^\epsilon - \frac{2}{N^L} \cdot C_2 N^{d_2}.$$

Further, due to Lipschitz-continuity of $\Psi^{(N)}$:

$$\frac{2}{N^L} \cdot C_3 N^{d_3+\epsilon} + |F_i^{(N)}(z')| > \Psi^{(N)}(z') N^\epsilon - \frac{2}{N^L} \cdot C_2 N^{d_2}.$$

It follows, using the lower bound on $\Psi^{(N)}$:

$$\frac{|F_i^{(N)}(z')|}{\Psi^{(N)}(z')} > N^\epsilon - 2\frac{C_2 N^{d_2} + C_3 N^{d_3+\epsilon}}{N^L \Psi^{(N)}(z')} \geq N^\epsilon - 2C_4 N^{d_4} \frac{C_2 N^{d_2} + C_3 N^{d_3+\epsilon}}{N^L}.$$

We may assume w.l.o.g. that ϵ is small, for example, $\epsilon \in (0, L - d_3 - d_4)$ (see Remark 6.6). Then there exists an $N(\epsilon) \in \mathbb{N}$, such that for all $N \geq N(\epsilon)$:

$$N^\epsilon - 2C_4 N^{d_4} \frac{C_2 N^{d_2} + C_3 N^{d_3+\epsilon}}{N^L} > N^{\frac{\epsilon}{2}}.$$

We have shown that for all $N \geq N(\epsilon)$:

$$\left[\exists\, z \in \mathcal{G}_N,\, \exists i \in I_N : \frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^\epsilon\right] \Rightarrow \left[\exists\, z' \in \mathcal{G}_N^L,\, \exists i \in I_N : \frac{|F_i^{(N)}(z')|}{\Psi^{(N)}(z')} > N^{\frac{\epsilon}{2}}\right].$$

Therefore, if $D > 0$ is arbitrary, we obtain for all $N \geq N(\epsilon)$:

$$\mathbb{P}\left(\sup_{z\in\mathcal{G}_N}\max_{i\in I_N}\frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^{\epsilon}\right) \leq \mathbb{P}\left(\sup_{z\in\mathcal{G}_N^L}\max_{i\in I_N}\frac{|F_i^{(N)}(z)|}{\Psi^{(N)}(z)} > N^{\frac{\epsilon}{2}}\right) \leq \frac{C_{\frac{\epsilon}{2},D}}{N^D},$$

where we used Step 1 for the last inequality. This concludes the proof by choosing constants as $(\epsilon, D) \mapsto C_{\frac{\epsilon}{2},D}$ and with Remark 6.6. $\square$

In order to apply Theorem 6.12 to Theorem 6.10 and to other local laws, we must analyze the Lipschitz-continuity of the quantities involved. This is the task of the following two lemmas. The first lemma will study general properties of Lipschitz continuity, the second lemma will then analyze Lipschitz-continuity of the relevant quantities involved.

Lemma 6.13. *Let V, W and X be normed $\mathbb{K}$-vector spaces, where $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$, and let $f, g : U \to W$, $h : W \to X$ be maps, where $U \subseteq V$ is a subset, then the following statements hold:*

i) If f is L-Lipschitz, so is $\|f\|$.

ii) If f is L-Lipschitz and $\lambda \in \mathbb{K}$, then λf is $|\lambda|L$-Lipschitz.

iii) If f is L-Lipschitz and g is K-Lipschitz, then $f + g$ is $(L + K)$-Lipschitz.

iv) If f resp. g are $\mathbb{R}$-valued and L- resp. K-Lipschitz, then $\min(f, g)$ and $\max(f, g)$ are both $\max(L, K)$-Lipschitz.

v) If f is L-Lipschitz and h is K-Lipschitz, then $h \circ f$ is $L \cdot K$-Lipschitz.

Proof. To show i), note that

$$|\|f(v_1)\| - \|f(v_2)\|| \leq \|f(v_1) - f(v_2)\| \leq L\|v_1 - v_2\|.$$

For ii), we see that

$$\|\lambda f(v_1) - \lambda f(v_2)\| = |\lambda|\|f(v_1) - f(v_2)\| \leq |\lambda|L\|v_1 - v_2\|.$$

To see iii), we observe

$$\|f(v_1)+g(v_1)-f(v_2)-g(v_2)\| \leq \|f(v_1)-f(v_2)\|+\|g(v_1)-g(v_2)\| \leq L\|v_1-v_2\|+K\|v_1-v_2\|.$$

In iv), we will only show the statement for the minimum (the statement for the maximum then follows, since $\max(f, g) = -\min(-f, -g)$ and *ii)* was already shown). Setting $M := \max(L, K)$, we need to show for all $v_1, v_2 \in U$ that

$$|\min(f(v_1), g(v_1)) - \min(f(v_2), g(v_2))| \leq M\|v_1 - v_2\|.$$

We assume that there exist v_1, v_2 in U such that ">" holds. Then the two minima cannot be assumed by the same function, since both f and g are M-Lipschitz. Thus, w.l.o.g. we may assume that $\min(f(v_1), g(v_1)) = f(v_1)$ and $\min(f(v_2), g(v_2)) = g(v_2)$. The statement is clear if $f(v_1) = g(v_2)$. Now on the one hand, if $f(v_1) < g(v_2)$, then we actually obtain $f(v_1) < g(v_2) \leq f(v_2)$. But then

$$|f(v_2) - f(v_1)| = f(v_2) - f(v_1) \geq g(v_2) - f(v_1) > M\|v_2 - v_1\|,$$

which is a contradiction, since f is M-Lipschitz. On the other hand, if $f(v_1) > g(v_2)$, then we actually obtain $g(v_1) \geq f(v_1) > g(v_2)$. But then

$$|g(v_2) - g(v_1)| = g(v_1) - g(v_2) \geq f(v_1) - g(v_2) > M\|v_2 - v_1\|,$$

which is a contradiction, since g is M-Lipschitz. Finally, for v) we note

$$\|h(f(v_1)) - h(f(v_2))\| \leq K\|f(v_1) - f(v_2)\| \leq K \cdot L\|v_1 - v_2\|.$$

□

Lemma 6.14. *Let $N \in \mathbb{N}$ be arbitrary and X_N be an Hermitian $N \times N$ matrix.*

i) The Stieltjes transform s of the semicircle distribution σ is N^2-Lipschitz and its reciprocal $1/s$ is $2N^2$-Lipschitz on $\mathcal{D}_N$.

ii) The resolvent $G = G^{(N)}$ of X_N, that is, $z \mapsto G(z) = (X_N - z)^{-1}$ is N^2-Lipschitz on $\mathcal{D}_N$.

iii) For any $i, j \in [N]$, the resolvent entry G_{ij} is N^2-Lipschitz on $\mathcal{D}_N$.

iv) The Stieltjes transform s_N of the ESD of X_N, that is, $s_N(z) = \frac{1}{N}\sum_{i=1}^N G_{ii}(z)$, is N^2-Lipschitz on $\mathcal{D}_N$.

v) The absolute difference $|s_N(z) - s(z)|$ is $2N^2$-Lipschitz and $|s_N(z) - 1/s(z)|$ is $3N^2$-Lipschitz on $\mathcal{D}_N$.

vi) The minimum $S_N(z) := \min\left\{|s_N(z) - s(z)|, \left|s_N(z) - \frac{1}{s(z)}\right|\right\}$ is $3N^2$-Lipschitz on $\mathcal{D}_N$.

vii) The error term $R_N(z) := \min\left\{\frac{1}{\sqrt{N\eta\kappa}}, \frac{1}{(N\eta)^{\frac{1}{4}}}\right\}$ is $10N$-Lipschitz on $\mathcal{D}_N$.

Proof. Throughout the proof, we will tacitly use Lemma 6.13. Let $N \in \mathbb{N}$ and $z, y \in \mathcal{D}_N$ be arbitrary. It should be very clear, when and which part of Lemma 6.13 is used.

i) We calculate

$$\begin{aligned}
|s(z) - s(y)| &= \left|\int_{\mathbb{R}} \frac{1}{x - z}\sigma(\mathrm{d}x) - \int_{\mathbb{R}} \frac{1}{x - y}\sigma(\mathrm{d}x)\right| \\
&\leq \int_{\mathbb{R}} \left|\frac{z - y}{(x - z)(x - y)}\right| \sigma(\mathrm{d}x) \\
&\leq |z - y| \int_{\mathbb{R}} \frac{1}{|\mathrm{Im}(z)||\mathrm{Im}(y)|}\sigma(\mathrm{d}x) \\
&\leq |z - y| \cdot N^2,
\end{aligned}$$

since $\mathrm{Im}(z), \mathrm{Im}(y) \geq 1/N^2$. Now with

$$\frac{1}{s(z)} = -z - s(z),$$

we see that the reciprocal $1/s$ is (N^2+1)-Lipschitz, hence $2N^2$ - Lipschitz.

ii) Since the resolvent identity yields $G(z) - G(y) = (z-y)G(z)G(y)$, we conclude

$$\|G(z) - G(y)\| \leq |z-y| \|G(z)\| \|G(y)\| \leq |z-y| \cdot \frac{1}{\mathrm{Im}(z)} \cdot \frac{1}{\mathrm{Im}(y)} \leq |z-y| N^2,$$

where we used that for any $z' \in \mathbb{C} \backslash \mathbb{R}$:

$$\|G(z')\| = \sup \left\{ \left| \frac{1}{\lambda_i - z'} \right| \mid i = 1, \ldots, n, \lambda_i \text{ eigenvalue of } X_N \right\} \leq \frac{1}{|\mathrm{Im}(z')|},$$

by the spectral theorem.

iii) We calculate

$$|G_{ij}(z) - G_{ij}(y)| = |(G(z) - G(y))_{ij}| \leq \|e_i\| \cdot \|G(z) - G(y)\| \cdot \|e_j\| \leq |z-y| N^2.$$

iv) We calculate

$$\begin{aligned} |s_N(z) - s_N(y)| &= \left| \frac{1}{N} \sum_{i=1}^{N} G_{ii}(z) - \frac{1}{N} \sum_{i=1}^{N} G_{ii}(y) \right| \\ &\leq \frac{1}{N} \sum_{i=1}^{N} |G_{ii}(z) - G_{ii}(y)| \leq \frac{1}{N} \cdot N \cdot N^2 |z-y|. \end{aligned}$$

v) This follows immediately from *i)* and *iv)*.

vi) This follows immediately from *v)*.

vii) For this result, we actually need to work. First we write

$$A_N(z) = \frac{1}{\sqrt{N\eta ||E| - 2|}} \qquad \text{and} \qquad B_N(z) = \frac{1}{(N\eta)^{\frac{1}{4}}},$$

so that $R_N(z) = \min(A_N(z), B_N(z))$. We note that

$$A_N(z) \leq B_N(z) \quad \Leftrightarrow \quad \frac{1}{\sqrt{N\eta ||E| - 2|}} \leq \frac{1}{(N\eta)^{\frac{1}{4}}} \quad \Leftrightarrow \quad \frac{1}{\sqrt{N\eta}} \leq ||E| - 2|$$

The question is: For which constant $L > 0$ do we find

$$\forall\, z_1, z_2 \in \mathcal{D}_N : \; |R_N(z_1) - R_N(z_2)| \leq L|z_1 - z_2|?$$

Case 1: $R_N(z_1) = A_N(z_1)$ and $R_N(z_2) = A_N(z_2)$. Then this entails that

$$\frac{1}{\sqrt{N\eta_1}} \leq ||E_1| - 2| \qquad \text{and} \qquad \frac{1}{\sqrt{N\eta_2}} \leq ||E_2| - 2| \tag{6.5}$$

Therefore, with explanations right after the calculation,

$$\begin{aligned}
|R_N(z_1) - R_N(z_2)| &= \left| \frac{1}{\sqrt{N\eta_1||E_1| - 2|}} - \frac{1}{\sqrt{N\eta_2||E_2| - 2|}} \right| \\
&= \left| \frac{\sqrt{N\eta_2||E_2| - 2|} - \sqrt{N\eta_1||E_1| - 2|}}{N\sqrt{\eta_1||E_1| - 2|}\sqrt{\eta_2||E_2| - 2|}} \right| \\
&\leq \left| \sqrt{N\eta_2||E_2| - 2|} - \sqrt{N\eta_1||E_1| - 2|} \right| \\
&\leq \frac{1}{2} \Big| N\eta_2||E_2| - 2| - N\eta_1||E_1| - 2| \Big| \\
&= \frac{N}{2} \Big| \eta_2||E_2| - 2| - \eta_1||E_1| - 2| \Big| \\
&\leq 10N \left\| \begin{pmatrix} ||E_2| - 2| \\ \eta_2 \end{pmatrix} - \begin{pmatrix} ||E_1| - 2| \\ \eta_1 \end{pmatrix} \right\|_2 \\
&\leq 10N \left\| \begin{pmatrix} E_2 \\ \eta_2 \end{pmatrix} - \begin{pmatrix} E_1 \\ \eta_1 \end{pmatrix} \right\|_2 \\
&= 10N|z_1 - z_2|
\end{aligned}$$

where the third step follows since with (6.5), the denominator is lower bounded as follows:

$$\begin{aligned}
N\sqrt{\eta_1||E_1| - 2|}\sqrt{\eta_2||E_2| - 2|} &\geq N\sqrt{\eta_1 \frac{1}{\sqrt{N\eta_1}} \eta_2 \frac{1}{\sqrt{N\eta_2}}} \\
&= N\sqrt{\frac{1}{N}\sqrt{\eta_1\eta_2}} \geq \frac{N}{\sqrt{N}}\sqrt{\frac{1}{N}} = 1.
\end{aligned}$$

The fourth step follows from the fact that $\sqrt{\cdot}$ is $\frac{1}{2}$-Lipschitz on $[1, \infty)$ and this is the domain of the arguments of $\sqrt{\cdot}$, as seen with (6.5), for example

$$N\eta_2||E_2| - 2| \geq N\eta_2 \frac{1}{\sqrt{N\eta_2}} = \sqrt{N\eta_2} \geq \sqrt{N\frac{1}{N}} = 1.$$

Further, the sixth step follows since $f : [0, 10]^2 \to \mathbb{R}$, $f(x, y) = xy$ is $10\sqrt{2}$-Lipschitz, since $\|\nabla f(x, y)\|_2 = \sqrt{x^2 + y^2} \leq 10\sqrt{2}$. The seventh step follows from the direct calculation

$$|E_2 - E_1| \geq ||E_2| - |E_1|| = |(|E_2| - 2) - (|E_1| - 2)| \geq \Big| ||E_2| - 2| - ||E_1| - 2| \Big|,$$

where we applied the reverse triangle inequality two times.

As we see, in Case 1 the Lipschitz constant of $10N$ suffices.
Case 2: $R_N(z_1) = B_N(z_1)$ and $R_N(z_2) = B_N(z_2)$. We calculate

$$\begin{aligned}|R_N(z_1) - R_N(z_2)| &= \left|\frac{1}{(N\eta_1)^{\frac{1}{4}}} - \frac{1}{(N\eta_2)^{\frac{1}{4}}}\right| \\ &= \left|\frac{(N\eta_2)^{\frac{1}{4}} - (N\eta_1)^{\frac{1}{4}}}{\sqrt{N}(\eta_1\eta_2)^{\frac{1}{4}}}\right| \\ &\leq |(N\eta_2)^{\frac{1}{4}} - (N\eta_1)^{\frac{1}{4}}| \\ &\leq \frac{N}{4}|\eta_1 - \eta_2|,\end{aligned}$$

where in the third step we used that $\sqrt{N}(\eta_1\eta_2)^{\frac{1}{4}} \geq \sqrt{N}\left(\frac{1}{N}\frac{1}{N}\right)^{\frac{1}{4}} = 1$ and in the fourth step we used that $(\cdot)^{\frac{1}{4}}$ is $\frac{1}{4}$-Lipschitz on $[1,\infty)$ and the arguments are at least 1, for example, $N\eta_1 \geq N\frac{1}{N} = 1$. Again, the Lipschitz-constant of $10N$ suffices.
Case 3: $[R_N(z_1) = A_N(z_1)$ and $R_N(z_2) = B_N(z_2)]$
or $[R_N(z_1) = B_N(z_1)$ and $R_N(z_2) = A_N(z_2)]$.
Per symmetry we may assume that $R_N(z_1) = A_N(z_1)$ and $R_N(z_2) = B_N(z_2)$. In particular, we have

$$||E_1| - 2| \geq \frac{1}{\sqrt{N\eta_1}} \qquad \text{and} \qquad ||E_2| - 2| \leq \frac{1}{\sqrt{N\eta_2}}.$$

We conclude

$$\begin{aligned}|R_N(z_1) - R_N(z_2)| &= \left|\frac{1}{\sqrt{N\eta_1||E_1| - 2|}} - \frac{1}{(N\eta_2)^{\frac{1}{4}}}\right| \\ &= \left|\frac{(N\eta_2)^{\frac{1}{4}} - \sqrt{N\eta_1||E_1| - 2|}}{\sqrt{N\eta_1||E_1| - 2|}(N\eta_2)^{\frac{1}{4}}}\right| \\ &\leq \left|\sqrt{\sqrt{N\eta_2}} - \sqrt{N\eta_1||E_1| - 2|}\right| \\ &\leq \frac{1}{2}\left|\sqrt{N\eta_2} - N\eta_1||E_1| - 2|\right|,\end{aligned} \tag{6.6}$$

where the third step follows since the denominator can be lower bounded by 1:

$$\sqrt{N\eta_1||E_1| - 2|}(N\eta_2)^{\frac{1}{4}} \geq \sqrt{N\eta_1\frac{1}{\sqrt{N\eta_1}}}\sqrt{\sqrt{1}} = \sqrt{\sqrt{N\eta_1}} \geq 1.$$

The fourth step follows since $\sqrt{\cdot}$ is $\frac{1}{2}$-Lipschitz on $[1,\infty)$ and this is clearly the domain of the arguments.

Subcase 1: $\sqrt{N\eta_2} > N\eta_1||E_1|-2|$. Then

$$\begin{aligned}\left|\sqrt{N\eta_2} - N\eta_1||E_1|-2|\right| &= \sqrt{N\eta_2} - N\eta_1||E_1|-2| \\ &\leq \sqrt{N\eta_2} - \sqrt{N\eta_1} \\ &\leq \frac{1}{2}|N\eta_2 - N\eta_1| \\ &\leq \frac{N}{2}\sqrt{(\eta_2-\eta_1)^2 + (E_2-E_1)^2} = \frac{N}{2}|z_1 - z_2|\end{aligned}$$

and thus again, the constant of $10N$ suffices with (6.6).
Subcase 2: $\sqrt{N\eta_2} \leq N\eta_1||E_1|-2|$. Then

$$\begin{aligned}\left|\sqrt{N\eta_2} - N\eta_1||E_1|-2|\right| &= N\eta_1||E_1|-2| - \sqrt{N\eta_2} \\ &\leq N\eta_1||E_1|-2| - N\eta_2||E_2|-2| \\ &= N\Big|\eta_1||E_1|-2| - \eta_2||E_2|-2|\Big| \\ &\leq 20N|z_1 - z_2|,\end{aligned}$$

where the second step follows with our findings at the very beginning of Case 3 and the last step follows from our calculation in Case 1. As we see, again a Lipschitz constant of $10N$ suffices (note the factor 1/2 in the end of (6.6)). □

We will now show that Theorem 6.10 actually holds simultaneously. It should be noted that later in Section 6.4, where we prove Theorem 6.10, we actually already prove it simultaneously. But this is merely due to the nature of our proof. It is still nice (and important) to see here that the simultaneous version actually *follows from* the seemingly weaker uniform version, employing Theorem 6.12.

Theorem 6.15 (Simultaneous Weak Local Law for Curie-Weiss-Type Ensembles)**.** *In the setting of the weak local law for Curie-Weiss type ensembles (Theorem 6.10) we obtain*

$$\sup_{z\in\mathcal{D}_N(\gamma)} \frac{|s_N(z) - s(z)|}{\min\left\{\frac{1}{\sqrt{N\eta\kappa}}, \frac{1}{(N\eta)^{\frac{1}{4}}}\right\}} \prec 1$$

Proof. We know by Lemma 6.14 that $F^{(N)}(z) := |s_N(z) - s(z)|$ is $2N^2$-Lipschitz and $\Psi^{(N)}(z) := \min\left\{\frac{1}{\sqrt{N\eta\kappa}}, \frac{1}{(N\eta)^{\frac{1}{4}}}\right\}$ is $10N$-Lipschitz on $\mathcal{D}_N$. Since on $\mathcal{D}_N$ we find $\eta, \kappa \leq 10$, we obtain that $\frac{1}{\sqrt{N\eta\kappa}} \geq \frac{1}{10\sqrt{N}}$ and $\frac{1}{(N\eta)^{\frac{1}{4}}} \geq \frac{1}{10N^{\frac{1}{4}}}$, such that $\Psi^{(N)}(z) \geq \frac{1}{10\sqrt{N}}$.

Further, we surely obtain by Theorem 6.10 that

$$F^{(N)}(z) \prec \Psi^{(N)}(z),\ z \in D_N^4(\gamma).$$

Therefore, the statement follows with Theorem 6.12 by choosing constants $C_2 = 2$, $d_2 = 2$, $C_3 = 10$, $d_3 = 1$, $C_4 = 10$, $d_4 = 1/2$ and $L = 4$. □

We draw two further immediate but important corollaries from Theorem 6.15.

Corollary 6.16. *In the situation of Theorem 6.10, we find that for any $\gamma \in (0,1)$:*

$$\sup_{z\in\mathcal{D}_N(\gamma)} |s_N(z) - s(z)| \prec \frac{1}{N^{\frac{\gamma}{4}}}.$$

Proof. Since for any $z \in \mathcal{D}_N(\gamma)$ we find

$$\frac{1}{(N\eta)^{\frac{1}{4}}} \leq \frac{1}{\left(N\frac{1}{N^{1-\gamma}}\right)^{\frac{1}{4}}} = \frac{1}{N^{\frac{\gamma}{4}}},$$

it follows

$$\sup_{z\in\mathcal{D}_N(\gamma)} \frac{|s_N(z)-s(z)|}{\frac{1}{N^{\frac{\gamma}{4}}}} \leq \sup_{z\in\mathcal{D}_N(\gamma)} \frac{|s_N(z)-s(z)|}{\min\left\{\frac{1}{\sqrt{N\eta\kappa}}, \frac{1}{(N\eta)^{\frac{1}{4}}}\right\}} \prec 1$$

by Theorem 6.15. Multiplying both sides by $1/N^{\gamma/4}$ concludes the proof. □

Corollary 6.16 allows us to conclude that on sets with high probability, s_N converges uniformly to s on a growing domain $D_N(\gamma)$ that approaches the real axis.

Corollary 6.17. *In the situation of Theorem 6.10 let $\gamma \in (0,1)$ be arbitrary and define the scale $\eta_N := 1/N^{1-\gamma}$ for all $N \in \mathbb{N}$. Then for all $\epsilon, D > 0$ there exists a constant $C_{\epsilon,D} \geq 0$ such that for all $N \in \mathbb{N}$ we have*

$$\mathbb{P}\left(\sup_{E\in[-10,10]} \left|\frac{1}{\pi}\operatorname{Im}(s_N(E+i\eta_N)) - f_\sigma(E)\right| > \frac{N^\epsilon}{\pi N^{\frac{\gamma}{4}}} + \frac{\sqrt{10}}{N^{\frac{1}{2}-\frac{\gamma}{2}}}\right) \leq \frac{C_{\epsilon,D}}{N^D}.$$

Proof. We find for all $N \in \mathbb{N}$ that

$$\begin{aligned}
&\sup_{E\in[-10,10]} \left|\frac{1}{\pi}\operatorname{Im}(s_N(E+i\eta_N)) - f_\sigma(E)\right| \\
&\leq \sup_{E\in[-10,10]} \left|\frac{1}{\pi}\operatorname{Im}(s_N(E+i\eta_N)) - \frac{1}{\pi}\operatorname{Im}(s(E+i\eta_N))\right| \\
&\quad + \sup_{E\in[-10,10]} \left|\frac{1}{\pi}\operatorname{Im}(s(E+i\eta_N)) - f_\sigma(E)\right| \\
&\leq \frac{1}{\pi} \sup_{z\in\mathcal{D}_N(\gamma)} |s_N(z)-s(z)| + \sqrt{10\eta_N} \\
&\leq \frac{N^\epsilon}{\pi N^{\frac{\gamma}{4}}} + \sqrt{10\eta_N}.
\end{aligned}$$

on a set A_N with $\mathbb{P}(A_N) > 1 - C_{\epsilon,D}/N^D$ for an N-independent constant $C_{\epsilon,D}$, witnessed by Corollary 6.16. We also used Lemma 5.18 for the second summand. This proves the statement. □

Corollary 6.17 states in particular that, at the scale $\eta_N = 1/N^{1-\gamma}$ ($\gamma \in (0,1)$ fixed), we find uniform convergence in probability of the kernel density estimator $P_{\eta_N} * \sigma_N$ to f_σ on the interval $[-10, 10]$, where we have strong control on the probability estimates. In his publication [40], Khorunzhy showed for the Wigner case that for arbitrary but fixed $E \in (-2, 2)$ and for slow scales $\eta_N = 1/N^{1-\gamma}$ ($\gamma \in (3/4, 1)$ fixed), $P_{\eta_N} * \sigma_N(E) \to f_\sigma(E)$ in probability. But very importantly, he showed that this does *not* hold in general for scales that decay too quickly, such as the scale $\eta_N = 1/N$, see his Remark 4 on page 149 in above mentioned publication. Therefore, the scale that is used in our weak local law, Theorem 6.10, cannot be improved *while still implying convergence in probability* of $P_{\eta_N} * \sigma_N$ to f_σ pointwise or uniformly. See also Figures 5.1 and 5.2 on page 128 for a visulization of these findings. We need to keep this in mind when interpreting the Theorem 10.1 in [9, p. 45] which implies that (in particular), Theorem 6.10 remains true even if every $\mathcal{D}_N(\gamma)$ is replaced by $[-10, 10] + i(0, 10]$. For note that for faster decays such as $\eta_N = 1/N^5$ and for $E \in [-10, 10]$ fixed, we would obtain the statement that for each $\epsilon, D > 0$ we find a constant $C_{\epsilon,D} \geq 0$ such that

$$\forall\, N \in \mathbb{N} : \mathbb{P}\left(|s_N(E + i\eta_N) - s(E + i\eta)| > N^\epsilon \min\left\{ \frac{N^2}{\sqrt{\kappa}}, N \right\} \right) \leq \frac{C_{\epsilon,D}}{N^D},$$

which is hardly a statement from which we could infer convergence in probability or almost surely of $|s_N(E + i\eta_N) - s(E + i\eta_N)|$ to zero. Rather, this statement has the structure of a tail probability bound, which is a whole different matter.

Theorem 6.10 and Theorem 6.15 guarantee closeness of the Stieltjes transforms of the ESDs and of the semicircle distribution. But how can we conclude that for certain classes of functions f, $\langle \sigma_N, f \rangle$ is close to $\langle \sigma, f \rangle$? The following lemma is a key ingredient. Indeed, in the equality in the following lemma, if we integrate both sides with respect to (say) $\sigma_N(\mathrm{d}\lambda)$, we obtain a triple integral on the right hand side. Applying Fubini, we retrieve a double integral over a term that includes the Stieltjes transform of σ_N at z.

Lemma 6.18 (Non-Holomorphic Cauchy Integral Formula)**.** *Let $f : \mathbb{C} \to \mathbb{C}$ be a function that is continuously differentiable in the real sense. Further, we assume f to be compactly supported. Then it holds for all $\lambda \in \mathbb{C}$:*

$$f(\lambda) = \frac{1}{2\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(\partial_x + i\partial_y) f(z)}{\lambda - z}\, dx\, dy \qquad (z = x + iy).$$

In particular, the integral on the right hand side exists.

Proof. This lemma is proved in [53, p. 388] using elementary arguments. The idea is to use polar coordinates with epicenter λ. □

Next, we will formulate and prove a semicircle law on small scales, where we will follow the sketch in [9] (see their Theorem 2.8), but implement own ideas such as the use of the Cauchy-Riemann equations. What we are after is a probabilistic evaluation of how well the semicircle distribution predicts the fraction of eigenvalues in given intervals $I \subseteq \mathbb{R}$. This is a very important theorem and exemplifies well what the local law is capable of.

In fact, a variant of the following theorem (see Theorem 6.20 below) even constitutes the local law *per se* in [60]. Notationally, if $A \subseteq \mathbb{R}$ is a subset, denote by $\mathcal{I}(A)$ the set of all intervals $I \subseteq A$.

Theorem 6.19 (Semicircle Law on Small Scales)**.** *In the setting of the weak local law for Curie-Weiss type ensembles (Theorem 6.10) we obtain*

$$\sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \prec \frac{1}{N^{\frac{1}{4}}}.$$

Further, for any fixed $\theta \in (0,1)$,

$$\sup_{I \in \mathcal{I}([-2+\theta, 2-\theta])} |\sigma_N(I) - \sigma(I)| \prec \frac{1}{N^{\frac{1}{2}}}.$$

Proof. We will show the first statement and afterwards discuss the minor changes to be made for the proof of the second statement.

First of all, since $\sigma_N(I)$ is a random variable (i.e. measurable) for all $I \in \mathcal{I}(\mathbb{R})$, so is $\sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)|$, since this value coincides with the supremum over all intervals $I \subseteq \mathbb{R}$ with rational end points, which is then a countable supremum of measurable functions, hence measurable. Therefore, it is valid to analyze events as in the statement of Theorem 6.19 in a probabilistic manner.

Step 1: Initialization of smooth indicator functions

We will start by introducing certain quantities that we will employ in the proof. For any interval $I \in \mathcal{I}([-3,3])$ and $\eta \in (0,1]$ denote by $f = f_{I,\eta} \in \mathcal{C}_c^\infty(\mathbb{R},[0,1])$ a smoothed indicator function with

- $f(x) = 1$ for all $x \in I$,
- $f(x) = 0$ for all $x \in \mathbb{R}$ with $\operatorname{dist}(x, I) \geq \eta$,
- $\|f'\|_\infty \leq \frac{C_1}{\eta}$,
- $\|f''\|_\infty \leq \frac{C_2}{\eta^2}$,

where C_1 and C_2 are suitable constants independent of η and I, and $\operatorname{dist}(x, I) := \inf_{y \in I} |x - y|$. It follows that the supports of f, f' and f'' are contained in $[-4,4]$. Further, the supports of f' and f'' have Lebesgue measure of at most 2η. In particular, Lebesgue integrals over $|f'|$ are bounded by $2C_1$ and Lebesgue integrals over $|f''|$ are bounded by $2C_2/\eta$. These facts will be used later on without always mentioning them again.

Now, we pick a smooth *even* cutoff function $\chi \in \mathcal{C}_c^\infty(\mathbb{R},[0,1])$ with

- $\chi(y) = 1$ for all $y \in [-1,1]$.
- $\chi(y) = 0$ for all $y \in \mathbb{R} \backslash [-2,2]$.
- $\|\chi'\|_\infty \leq C_3$ for some $C_3 \geq 0$.

Note that the support of χ' lies within the set $[-2,2]\backslash(-1,1)$. The purpose of χ is to serve as a cutoff function for the imaginary part in the calculations below. Instead of the smooth approximations of indicators $f_{I,\eta}$ and χ, that is, infinitely often differentiable ones, for our analysis we only need two times continuously differentiable approximations, as seen in the calculations that follow. However, even the existence of smooth approximations is well-known, see for example [19, p. 20].

Step 2: Applying the Non-Holomorphic Cauchy Integral Formula
If f and χ are as in Step 1, note that the function $g:\mathbb{C}\to\mathbb{C}$ with $g(x+iy):=(f(x)+iyf'(x))\chi(y)$ vanishes outside of a compact set, and is continuously differentiable in (x,y) when regarded as a function $\mathbb{R}^2\to\mathbb{R}^2$. Note also that per construction, for real arguments λ it holds that $g(\lambda)=f(\lambda)$. Therefore, if ν is an arbitrary probability measure on $(\mathbb{R},\mathcal{B})$ with Stieltjes transform S_ν, we obtain with Lemma 6.18:

$$\begin{aligned}\int_{\mathbb{R}} f(\lambda)\nu(\mathrm{d}\lambda) &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}}\frac{(\partial_x+i\partial_y)[(f(x)+iyf'(x))\chi(y)]}{\lambda-(x+iy)}\mathrm{d}y\mathrm{d}x\nu(\mathrm{d}\lambda)\\ &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}}(\partial_x+i\partial_y)[(f(x)+iyf'(x)\chi(y))]S_\nu(x+iy)\mathrm{d}y\mathrm{d}x\end{aligned}$$

Note that the partial derivatives ∂_x and ∂_y are only applied to the term in the brackets $[\ldots]$. This derivative can be evaluated as

$$\begin{aligned}&(\partial_x+i\partial_y)[(f(x)+iyf'(x))\chi(y)]\\ &= f'(x)\chi(y)+iyf''(x)\chi(y)+\partial_y(if(x)\chi(y)-yf'(x)\chi(y))\\ &= f'(x)\chi(y)+iyf''(x)\chi(y)+if(x)\chi'(y)-f'(x)\chi(y)-yf'(x)\chi'(y)\\ &= iyf''(x)\chi(y)+if(x)\chi'(y)-yf'(x)\chi'(y).\end{aligned}$$

With our calculations so far, and writing S_ν instead of $S_\nu(x+iy)$ in the following calculation for better readability, we obtain for any $\eta\in(0,1]$ (note also that f depends on η):

$$\begin{aligned}&\int f(\lambda)\nu(\mathrm{d}\lambda) = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}}\left[iyf''(x)\chi(y)+if(x)\chi'(y)-yf'(x)\chi'(y)\right]S_\nu\mathrm{d}y\mathrm{d}x\\ &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} iyf''(x)\chi(y)(\operatorname{Re}S_\nu+i\operatorname{Im}S_\nu)\mathrm{d}y\mathrm{d}x+\frac{i}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}}\left[f(x)+iyf'(x)\right]\chi'(y)S_\nu\mathrm{d}y\mathrm{d}x\\ &= -\frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f''(x)\chi(y)y\operatorname{Im}S_\nu\mathrm{d}y\mathrm{d}x\\ &\quad+\frac{i}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f''(x)\chi(y)y\operatorname{Re}S_\nu\mathrm{d}y\mathrm{d}x+\frac{i}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}}\left[f(x)+iyf'(x)\right]\chi'(y)S_\nu\mathrm{d}y\mathrm{d}x\\ &= -\frac{1}{2\pi}\int_{\mathbb{R}}\int_{|y|\leq\eta} f''(x)\chi(y)y\operatorname{Im}S_\nu\mathrm{d}y\mathrm{d}x-\frac{1}{2\pi}\int_{\mathbb{R}}\int_{|y|>\eta} f''(x)\chi(y)y\operatorname{Im}S_\nu\mathrm{d}y\mathrm{d}x\\ &\quad+\frac{i}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f''(x)\chi(y)y\operatorname{Re}S_\nu\mathrm{d}y\mathrm{d}x+\frac{i}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}}\left[f(x)+iyf'(x)\right]\chi'(y)S_\nu\mathrm{d}y\mathrm{d}x.\end{aligned}$$

Step 3: Bounding the difference of smoothed indicators
In particular, the last calculation yields for $\hat{s}_N := s_N - s$ (where as before, $s_N = S_{\sigma_N}$ and $s = S_\sigma$):

$$\begin{aligned}
&\int f(\lambda)\sigma_N(\mathrm{d}\lambda) - \int f(\lambda)\sigma(\mathrm{d}\lambda) \\
&= -\frac{1}{2\pi}\int_{\mathbb{R}}\int_{|y|\le\eta} f''(x)\chi(y)y\,\mathrm{Im}\,\hat{s}_N(x+iy)\mathrm{d}y\mathrm{d}x && \text{(T1)} \\
&\quad -\frac{1}{2\pi}\int_{\mathbb{R}}\int_{|y|>\eta} f''(x)\chi(y)y\,\mathrm{Im}\,\hat{s}_N(x+iy)\mathrm{d}y\mathrm{d}x && \text{(T2)} \\
&\quad +\frac{i}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} f''(x)\chi(y)y\,\mathrm{Re}\,\hat{s}_N(x+iy)\mathrm{d}y\mathrm{d}x. && \text{(T3)} \\
&\quad +\frac{i}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} (f(x)+iyf'(x))\chi'(y)\hat{s}_N(x+iy)\mathrm{d}y\mathrm{d}x. && \text{(T4)}
\end{aligned}$$

Now, we pick $\epsilon \in (0, 1/4)$ and $D > 0$ arbitrarily and let η decay at a rate compatible with our weak local law: For all $N \in \mathbb{N}$, let $\eta = \eta(N) := \frac{1}{N^{\frac{1}{4}-\epsilon}}$. We then know from the simultaneous weak local law (Theorem 6.15 with $\gamma = 1 - (1/4 - \epsilon)$) that there is a constant $C_{\epsilon,D} \ge 0$, which is independent of N, such that the set

$$A_N := \left\{ |s_N(x+iy) - s(x+iy)| \le \frac{N^\epsilon}{(N|y|)^{\frac{1}{4}}}, |x| \le 10, |y| \in [\eta, 10] \right\}$$

has high probability, namely $\mathbb{P}(A_N) > 1 - \frac{C_{\epsilon,D}}{N^D}$. Here, we used that for any Stieltjes transform S_ν of a probability measure ν we have $S_\nu(x - iy) = \overline{S_\nu(x+iy)}$ by Lemma 5.2. Next, we want to bound the terms (T1), (T2), (T3) and (T4) on A_N.

We begin with the term (T4). Considering that χ' has support in $J := [-2, 2]\backslash(-1, 1)$, $|f|$ and $|f'|$ and are bounded by 1 and C_1/η (respectively), both have support in $[-4, 4]$, where the support of f' has Lebesgue measure of at most 2η, we obtain on A_N:

$$\begin{aligned}
|\text{(T4)}| &\le \frac{1}{2\pi}\int_{[-4,4]}\int_J (|f(x)| + |y||f'(x)|)|\chi'(y)||\hat{s}_N(x+iy)|\mathrm{d}y\mathrm{d}x \\
&\le \frac{1}{2\pi}\int_{[-4,4]}\int_J (1 + 2|f'(x)|)C_3 \underbrace{\frac{N^\epsilon}{(N|y|)^{\frac{1}{4}}}}_{\le N^\epsilon/N^{1/4}=\eta} \mathrm{d}y\mathrm{d}x \\
&\le \frac{\eta C_3}{2\pi}\left(\int_{[-4,4]}\int_J 1\mathrm{d}y\mathrm{d}x + \int_{[-4,4]}|f'(x)|\int_J 2\mathrm{d}y\mathrm{d}x\right) \\
&\le \frac{\eta C_3}{2\pi}\left(16 + 4\int_{[-4,4]}|f'(x)|\mathrm{d}x\right) \\
&\le \frac{\eta C_3}{2\pi}(16 + 8C_1) \le \eta C_3(4 + 2C_1).
\end{aligned}$$

Next, we will bound (T3). To this end, notice that the left hand side of the equation in the beginning of Step 3 (in which (T1), (T2), (T3) and (T4) were defined) is a real number. On the right hand side of this equation, (T1) and (T2) are both real, so we must have $\operatorname{Im}(\mathrm{T3}) = -\operatorname{Im}(\mathrm{T4})$. Since $\operatorname{Re}(\mathrm{T3}) = 0$, we have $(\mathrm{T3}) = i\operatorname{Im}(\mathrm{T3})$ and conclude

$$|(\mathrm{T3})| = |i\operatorname{Im}(\mathrm{T3})| = |\operatorname{Im}(\mathrm{T4})| \leq |(\mathrm{T4})| \leq \eta C_3(4+2C_1).$$

Let us turn to (T1). If ν is any probability measure on $(\mathbb{R},\mathcal{B})$, then the expression $|y||\operatorname{Im} S_\nu(x+iy)|$ is non-decreasing in $|y|$ for any $x \in \mathbb{R}$. To see this, we calculate:

$$\begin{aligned}
|y||\operatorname{Im} S_\nu(x+iy)| &= |y|\left|\int_{\mathbb{R}} \frac{y}{(a-x)^2+y^2}\nu(\mathrm{d}a)\right| \\
&= \int_{\mathbb{R}} \frac{|y|^2}{(a-x)^2+|y|^2}\nu(\mathrm{d}a),
\end{aligned}$$

which is clearly non-decreasing in $|y|$ for any $x \in \mathbb{R}$. Therefore, on A_N we have for all $y \in \mathbb{R}$ with $|y| \leq \eta$ and all $x \in [-4,4]$:

$$\begin{aligned}
|y||\operatorname{Im} \hat{s}_N(x+iy)| &\leq |y|\left(|\operatorname{Im} s_N(x+iy)| + |\operatorname{Im} s(x+iy)|\right) \\
&\leq \eta\left(|\operatorname{Im} s_N(x+i\eta)| + |\operatorname{Im} s(x+i\eta)|\right) \\
&\leq \eta\left(|s_N(x+i\eta)| + |s(x+i\eta)|\right) \\
&\leq \eta\left(|s_N(x+i\eta) - s(x+i\eta)| + 2|s(x+i\eta)|\right) \\
&\leq \eta\left(\frac{N^\epsilon}{(N\eta)^{\frac{1}{4}}} + 32\right) \\
&\leq 33\eta,
\end{aligned}$$

where in the fifth step we used Theorem 5.16 and in the last step that

$$\frac{N^\epsilon}{(N\eta)^{\frac{1}{4}}} = \frac{N^\epsilon}{N^{(\frac{3}{4}+\epsilon)\frac{1}{4}}} = \frac{N^{\frac{3}{4}\epsilon}}{N^{\frac{3}{4}\cdot\frac{1}{4}}} \leq 1, \tag{6.7}$$

recalling that $\epsilon \in (0,1/4)$. Now to bound (T1) we calculate

$$\begin{aligned}
|(\mathrm{T1})| &\leq \frac{1}{2\pi}\int_{[-4,4]}\int_{|y|\leq\eta} |f''(x)||y||\operatorname{Im} \hat{s}_N(x+iy)|\mathrm{d}x\mathrm{d}y \\
&\leq \frac{1}{2\pi}\int_{[-4,4]} |f''(x)| \int_{|y|\leq\eta} 33\eta\mathrm{d}y\mathrm{d}x \\
&\leq \frac{66}{2\pi}\eta^2\int_{[-4,4]} |f''(x)|\mathrm{d}x \\
&\leq \frac{66}{2\pi}\eta^2\frac{2C_2}{\eta} \\
&\leq 22\eta C_2.
\end{aligned}$$

Next, we will bound (T2). With s_N and s, also $\hat{s}_N = s_N - s$ is holomorphic on $\mathbb{C}\backslash\mathbb{R}$. We will write

$$\hat{s}_N(x+iy) = u(x,y) + iv(x,y)$$

where $u, v : \mathbb{R}^2 \to \mathbb{R}$. Further, denote by $u_x(x,y)$, $u_y(x,y)$, $v_x(x,y)$ and $v_y(x,y)$ the corresponding partial derivatives. We now obtain that

$$\hat{s}_N(x-iy) = \overline{\hat{s}_N(x+iy)} = u(x,y) - iv(x,y)$$

where we applied Lemma 5.2 *iv)* to both summands of $s - s_N$. In particular, we obtain

$$u(x,y) = u(x,-y), \quad \text{thus} \quad u_y(x,y) = -u_y(x,-y).$$

Since $\hat{s}_N$ is holomorphic, we know that $v_x(x,y) = -u_y(x,y)$. In the following calculation, in order to get rid of the second derivative of f, we integrate by parts with respect to x and then with respect to y (also, keeping in mind that χ is an even function):

$$\begin{aligned}
-2\pi(\mathrm{T2}) &= \int_{\mathbb{R}} \int_{|y|>\eta} f''(x)\chi(y)y \operatorname{Im} \hat{s}_N(x+iy)\mathrm{d}y\mathrm{d}x \\
&= \int_{|y|>\eta} \chi(y)y \int_{\mathbb{R}} f''(x)v(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{|y|>\eta} \chi(y)y \left([f'(x)v(x,y)]_{x=-\infty}^{+\infty} - \int_{\mathbb{R}} f'(x)v_x(x,y)\mathrm{d}x \right) \mathrm{d}y \\
&= \int_{|y|>\eta} \int_{\mathbb{R}} \chi(y)yf'(x)u_y(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{\mathbb{R}} f'(x) \int_{|y|>\eta} \chi(y)yu_y(x,y)\mathrm{d}y\mathrm{d}x \\
&= 2\int_{\mathbb{R}} f'(x) \int_{\eta}^{\infty} \chi(y)yu_y(x,y)\mathrm{d}y\mathrm{d}x \\
&= 2\int_{\mathbb{R}} f'(x) \left([u(x,y)\chi(y)y]_{y=\eta}^{\infty} - \int_{\eta}^{\infty} u(x,y)\chi(y)\mathrm{d}y - \int_{\eta}^{\infty} u(x,y)\chi'(y)y\mathrm{d}y \right) \mathrm{d}x \\
&= -2\int_{\mathbb{R}} f'(x)u(x,\eta)\chi(\eta)\eta\mathrm{d}x - 2\int_{\mathbb{R}}\int_{\eta}^{\infty} f'(x)u(x,y)\chi(y)\mathrm{d}y\mathrm{d}x \\
&\quad - 2\int_{\mathbb{R}}\int_{\eta}^{\infty} f'(x)u(x,y)\chi'(y)y\mathrm{d}y\mathrm{d}x,
\end{aligned}$$

and therefore with $|u(x,y)| = |\operatorname{Re} \hat{s}_N(x+iy)| \leq |\hat{s}_N(x+iy)|$

$$\begin{aligned}
|(\mathrm{T2})| \leq \frac{1}{\pi} &\int_{[-4,4]} |f'(x)\hat{s}_N(x+i\eta)\chi(\eta)\eta|\mathrm{d}x && \text{(T5)} \\
&+ \frac{1}{\pi}\int_{[-4,4]}\int_{\eta}^{\infty} |f'(x)\hat{s}_N(x+iy)\chi(y)|\mathrm{d}y\mathrm{d}x && \text{(T6)} \\
&+ \frac{1}{\pi}\int_{[-4,4]}\int_{\eta}^{\infty} |f'(x)\hat{s}_N(x+iy)\chi'(y)y|\mathrm{d}y\mathrm{d}x. && \text{(T7)}
\end{aligned}$$

Now to bound (T5), (T6) and (T7) on A_N is easy for once: For (T5) we find immediately

$$(\mathrm{T5}) \leq \frac{2\eta}{\pi} C_1 \leq \eta C_1,$$

where we used that $|\hat{s}_N(x+i\eta)| \leq 1$ on A_N for all relevant x, which follows from (6.7). For (T6) we calculate

$$(\mathrm{T6}) \leq \frac{1}{\pi} \int_{[-4,4]} |f'(x)| \int_{\eta}^{2} \frac{N^{\epsilon}}{(Ny)^{\frac{1}{4}}} \mathrm{d}y \mathrm{d}x,$$

and since

$$\int_{\eta}^{2} \frac{N^{\epsilon}}{N^{\frac{1}{4}}} \frac{1}{y^{\frac{1}{4}}} \mathrm{d}y = \eta \left[\frac{4}{3} y^{\frac{3}{4}} \right]_{\eta}^{2} \leq 4\eta,$$

we conclude

$$(\mathrm{T6}) \leq \frac{4\eta}{\pi} \int_{[-4,4]} |f'(x)| \mathrm{d}x \leq \frac{8\eta C_1}{\pi} \leq 3\eta C_1.$$

Lastly, considering that the support of χ' lies in $[-2,2] \backslash (-1,1)$, we obtain

$$(\mathrm{T7}) \leq \frac{C_3}{\pi} \int_{[-4,4]} |f'(x)| \int_{1}^{2} \frac{N^{\epsilon}}{(Ny)^{\frac{1}{4}}} y \mathrm{d}y \mathrm{d}x \leq \frac{2\eta C_3}{\pi} \int_{[-4,4]} |f'(x)| \mathrm{d}x \leq \frac{4\eta C_1 C_3}{\pi} \leq 2\eta C_1 C_3,$$

where we used

$$\int_{1}^{2} \frac{N^{\epsilon}}{(Ny)^{\frac{1}{4}}} y \mathrm{d}y = \eta \int_{1}^{2} y^{\frac{3}{4}} \mathrm{d}y = \eta \left[\frac{4}{7} y^{\frac{7}{4}} \right]_{1}^{2} \leq 2\eta.$$

Putting things together, on the high-probability set A_N we find that for an arbitrary interval $I \subseteq [-3,3]$ (note that f depends on I and η, and η depends on N):

$$\begin{aligned}
&\left| \int f(\lambda) \sigma_N(\mathrm{d}\lambda) - \int f(\lambda) \sigma(\mathrm{d}\lambda) \right| \\
&\leq |(\mathrm{T1})| + |(\mathrm{T2})| + |(\mathrm{T3})| + |(\mathrm{T4})| \\
&\leq |(\mathrm{T1})| + |(\mathrm{T3})| + |(\mathrm{T4})| + (\mathrm{T5}) + (\mathrm{T6}) + (\mathrm{T7}) \\
&\leq 22\eta C_2 + \eta C_3(4 + 2C_1) + \eta C_3(4 + 2C_1) + \eta C_1 + 3\eta C_1 + 2\eta C_1 C_3 \\
&= K\eta
\end{aligned}$$

for a constant $K \geq 0$ that does not depend on N, I or η.

Step 4: "Unsmoothing" of the indicators

We need to translate the integration over smoothed indicator functions back to integration over "regular" indicator functions. We fix an $I \in \mathcal{I}([-3,3])$. Then we have on A_N:

$$\sigma_N(I) \leq \int f_{I,\eta}(\lambda) \sigma_N(\mathrm{d}\lambda) \leq \int f_{I,\eta}(\lambda) \sigma(\mathrm{d}\lambda) + K\eta \leq \sigma(I) + \frac{2\eta}{\pi} + K\eta \leq \sigma(I) + (K+1)\eta,$$

where the third inequality is due to the fact that the density of the semicircle distribution is bounded by $1/\pi$. On the other hand, define the (possibly empty) intervals $I_\eta^* \subseteq [-3,3]$ as $I_\eta^* := \{x \in \mathbb{R} : \operatorname{dist}(x, I^c) \geq \eta\}$, then we obtain

$$\sigma_N(I) \geq \int f_{I_\eta^*,\eta}(\lambda)\sigma_N(\mathrm{d}\lambda) \geq \int f_{I_\eta^*,\eta}(\lambda)\sigma(\mathrm{d}\lambda) - K\eta \geq \sigma(I) - \frac{2\eta}{\pi} - K\eta \geq \sigma(I) - (K+1)\eta.$$

Therefore, on A_N we obtain for all intervals $I \subseteq [-3,3]$ that

$$|\sigma_N(I) - \sigma(I)| \leq (K+1)\eta.$$

In particular, we obtain the useful information that on A_N,

$$\begin{aligned}\sigma_N([-2,2]^c) &= 1 + \sigma([-2,2]) - \sigma_N([-2,2]) - \sigma([-2,2]) \\ &\leq |\sigma([-2,2]) - \sigma_N([-2,2])| \leq (K+1)\eta,\end{aligned}$$

where we used that σ is a probability measure with $\sigma([-2,2]) = 1$. This is helpful: Let $I \subseteq \mathbb{R}$ now be an *arbitrary* interval, then we obtain on A_N:

$$\begin{aligned}&|\sigma_N(I) - \sigma(I)| \\ &= |\sigma_N(I \cap [-2,2]) + \sigma_N(I \cap [-2,2]^c) - \sigma(I \cap [-2,2]) - \sigma(I \cap [-2,2]^c)| \\ &\leq |\sigma_N(I \cap [-2,2]) - \sigma(I \cap [-2,2])| + \sigma_N(I \cap [-2,2]^c) \\ &\leq (K+1)\eta + \sigma_N([-2,2]^c) \\ &\leq 2(K+1)\eta\end{aligned}$$

We have seen that on A_N, for all intervals $I \subseteq \mathbb{R}$ it holds

$$|\sigma_N(I) - \sigma(I)| \leq 2(K+1)\eta = 2(K+1)\frac{N^\epsilon}{N^{\frac{1}{4}}},$$

so in particular

$$\mathbb{P}\left(\sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \leq 2(K+1)\frac{N^\epsilon}{N^{\frac{1}{4}}}\right) > 1 - \frac{C_{\epsilon,D}}{N^D}.$$

But $D > 0$ and $\epsilon \in (0, 1/4)$ were arbitrary, yielding

$$\sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \prec \frac{2(K+1)}{N^{\frac{1}{4}}},$$

where we used Remark 6.6, thus with Lemma 6.7 *vii*):

$$\sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \prec \frac{1}{N^{\frac{1}{4}}}.$$

This is the first statement of the theorem, which we wanted to show rigorously.

For the second statement, we will discuss the changes to be made to the proof we just conducted: Let $\theta \in (0,1)$ be fixed. We notice that the intervals I for which we consider the smoothed indicators $f = f_{I,\eta}$ are elements of $\mathcal{I}([-2+\theta, 2-\theta])$. Let such an I be fixed. In the third step as above, we then pick an $\epsilon \in (0, 1/2)$ and $D > 0$ arbitrarily and set $\eta = \eta(N) := 1/N^{1/2-\epsilon}$. Note that due to the smoothing of the indicators by f, the support of f may come closer to the edge than θ. To account for this, we need to allow ourselves a little room of $\theta/2$: Theorem 6.15 implies with Lemma 6.7 that for $\gamma := 1 - (1/2 - \epsilon)$ we have

$$\sup_{\substack{z \in \mathcal{D}_N(\gamma) \\ |\mathrm{Re}(z)| \leq 2-\theta/2}} \frac{|s_N(z) - s(z)|}{\frac{1}{\sqrt{N \,\mathrm{Im}(z) ||\mathrm{Re}(z)|-2|}}} \prec \sqrt{\theta/2},$$

where we just restricted the domain $\mathcal{D}_N(\gamma)$ on the left hand side of $\prec$ and multiplied the right hand side of $\prec$ by a positive constant. But since in the supremum above, $||\mathrm{Re}(z)| - 2|$ is lower bounded by $\theta/2$, we immediately obtain

$$\sup_{\substack{z \in \mathcal{D}_N(\gamma) \\ |\mathrm{Re}(z)| \leq 2-\theta/2}} \frac{|s_N(z) - s(z)|}{\frac{1}{\sqrt{N \,\mathrm{Im}(z)}}} \prec 1.$$

With γ and $\eta = \eta(N)$ as we just defined, we obtain a constant $C_{\epsilon,D} \geq 0$ independent of N such that for all $N \in \mathbb{N}$ the set

$$A_N := \left\{ |s_N(x+iy) - s(x+iy)| \leq \frac{N^\epsilon}{\sqrt{N|y|}}, |x| \leq 2 - \frac{\theta}{2}, |y| \in [\eta, 10] \right\}$$

has high probability, namely $\mathbb{P}(A_N) > 1 - \frac{C_{\epsilon,D}}{N^D}$. This is the *new* set A_N on which we will operate, but we will not give it a new name to preserve familiarity to the first part of the proof (just as we are considering a newly defined scale parameter η here). Since $I \in \mathcal{I}([-2+\theta, 2-\theta])$ and $\eta = \eta(N) = \frac{1}{N^{1/2-\epsilon}}$ where $\epsilon \in (0, 1/2)$, we obtain that f, f' and f'' all have support in $[-2+\theta/2, 2-\theta/2]$ for all $N \geq N(\epsilon)$, where $N(\epsilon)$ is so large that $\eta(N(\epsilon)) = 1/(N(\epsilon))^{1/2-\epsilon} \leq \theta/2$. Note that $N(\epsilon)$ also depends on θ, but θ is a super-parameter in the statement of the theorem. Note also that $I \subseteq [-2+\theta, 2-\theta] \subseteq [-4, 4]$, which we mention so that we can follow the notation as in the first part of the proof. In what follows, we will operate on A_N for all $N \geq N(\epsilon)$ (which suffices by Remark 6.6).

The bounds of (T1), (T2), (T3) and (T4) will take place on A_N for all $N \geq N(\epsilon)$. Also notice that integrals for the real part over $[-4, 4]$ are actually just integrals over $[-2+\theta/2, 2-\theta/2]$, since f and its derivatives are not supported elsewhere. In addition, we can use bounds on $\hat{s}$ given by operating on A_N. When bounding $|(T4)|$, we use the following upper bound (note that as before, integration with respect to y is over $[-2,2] \backslash (-1,1)$):

$$|\hat{s}_N(x+iy)| \leq \frac{N^\epsilon}{\sqrt{N|y|}} \leq \frac{N^\epsilon}{\sqrt{N}} = \frac{1}{N^{\frac{1}{2}-\epsilon}} = \eta.$$

This leaves the final bound of $|(T4)|$ unchanged. When bounding $|(T3)|$, nothing changes.

In the analysis of $|(T1)|$, we obtain that on A_N we have for all $y \in \mathbb{R}$ with $|y| \leq \eta$ and all $x \in [-2+\theta/2, 2-\theta/2]$:

$$|y||\operatorname{Im} \hat{s}_N(x+iy)| \leq \eta \left(\frac{N^\epsilon}{\sqrt{N\eta}} + 32 \right) \leq 33\eta,$$

where we used that

$$\frac{N^\epsilon}{\sqrt{N\eta}} = \frac{N^\epsilon}{N^{(\frac{1}{2}+\epsilon)\frac{1}{2}}} = \frac{N^{\frac{1}{2}\epsilon}}{N^{\frac{1}{2}\cdot\frac{1}{2}}} \leq 1,$$

recalling that $\epsilon \in (0, 1/2)$. Then in the bound of $|(T1)|$, we again use that f'' has support in $[-2+\theta/2, 2-\theta/2]$ for all $N \geq N(\epsilon)$, so we can use above bound of $|y||\operatorname{Im} \hat{s}(x+iy)|$. The final bound of $|(T1)|$ will remain unchanged.

Now bounding $|(T2)|$ will as before lead to bounding (T5), (T6) and (T7). In all these three terms, we use that for $N \geq N(\epsilon)$, f' has support in $[-2+\theta/2, 2-\theta/2]$, so we can use that we operate on A_N. Now the bound for (T5) will remain unchanged, considering that for all $x \in [-2+\theta/2, 2-\theta/2]$, we have on A_N that

$$|\hat{s}(x+i\eta)| \leq \frac{N^\epsilon}{\sqrt{N\eta}} = \frac{N^\epsilon}{N^{(\frac{1}{2}+\epsilon)\frac{1}{2}}} = \frac{N^{\frac{1}{2}\epsilon}}{N^{\frac{1}{2}\cdot\frac{1}{2}}} \leq 1.$$

For (T6) we obtain

$$(T6) \leq \frac{1}{\pi} \int_{[-4,4]} |f'(x)| \int_\eta^2 \frac{N^\epsilon}{(Ny)^{1/2}} \mathrm{d}y \mathrm{d}x,$$

and since

$$\int_\eta^2 \frac{N^\epsilon}{N^{\frac{1}{2}}} \frac{1}{y^{\frac{1}{2}}} \mathrm{d}y = \eta \left[2y^{\frac{1}{2}} \right]_\eta^2 \leq 4\eta,$$

we may leave the bound for (T6) unchanged. The bound for (T7) can be left unchanged as well using

$$\int_1^2 \frac{N^\epsilon}{(Ny)^{\frac{1}{2}}} y \mathrm{d}y = \eta \int_1^2 y^{\frac{1}{2}} \mathrm{d}y = \eta \left[\frac{2}{3} y^{\frac{3}{2}} \right]_1^2 \leq 2\eta.$$

Finally, putting things together, we obtain that for any interval $I \subseteq [-2+\theta, 2-\theta]$ and for all $N \geq N(\epsilon)$, on A_N it holds

$$\left| \int f(\lambda)\sigma_N(\mathrm{d}\lambda) - \int f(\lambda)\sigma(\mathrm{d}\lambda) \right| \leq K\eta$$

for a constant $K \geq 0$ independent of N, I and η, which can be chosen as in the proof of the first statement of the theorem. Proceeding as before, this will entail that for all $N \geq N(\epsilon)$ and all intervals $I \in \mathcal{I}([-2+\theta, 2-\theta])$ we have on A_N that

$$|\sigma_N(I) - \sigma(I)| \leq (K+1)\eta = (K+1)\frac{N^\epsilon}{N^{\frac{1}{2}}},$$

so in particular

$$\mathbb{P}\left(\sup_{I\in\mathcal{I}([-2+\theta,2-\theta])}|\sigma_N(I)-\sigma(I)|\leq(K+1)\frac{N^\epsilon}{N^{\frac{1}{2}}}\right)>1-\frac{C_{\epsilon,D}}{N^D},$$

But $D>0$ and $\epsilon\in(0,1/2)$ were arbitrary, yielding

$$\sup_{I\in\mathcal{I}([-2+\theta,2-\theta])}|\sigma_N(I)-\sigma(I)|\prec\frac{(K+1)}{N^{\frac{1}{2}}},$$

where we used Remark 6.6, thus with Lemma 6.7 *vii*):

$$\sup_{I\in\mathcal{I}([-2+\theta,2-\theta])}|\sigma_N(I)-\sigma(I)|\prec\frac{1}{N^{\frac{1}{2}}},$$

which is what we wanted to show. □

Having just proved Theorem 6.19, let us see how we may interpret it: It says in particular that for any $\epsilon\in(0,1/4)$ and $D>0$ we find a constant $C_{\epsilon,D}\geq 0$ such that

$$\forall\,N\in\mathbb{N}:\ \mathbb{P}\left(\sup_{I\in\mathcal{I}(\mathbb{R})}|\sigma_N(I)-\sigma(I)|\leq\frac{N^{\frac{3}{4}+\epsilon}}{N}\right)>1-\frac{C_{\epsilon,D}}{N^D},\tag{6.8}$$

This tells us that when predicting interval probabilities of σ_N by those of σ, the absolute error will be bounded by $1/N^{1/4-\epsilon}$ with a probability that grows arbitrarily polynomially fast to 1. We just used $\sigma(I)$ as a predictor for $\sigma_N(I)$ (note that this viewpoint is exactly opposite from statistics). Alternatively, this can be translated directly to estimating the number of eigenvalues in a given interval I, $N\sigma_N(I)$, by $N\sigma(I)$. Then (6.8) allows to control an absolute error of $N^{3/4+\epsilon}$ out of N eigenvalues, where we just multiplied the inequality inside the probability by N. We will switch back and fourth between these views, depending on what is needed for the argument we would like to make.

Inspecting (6.8), we ask: For which kind of intervals is this a good statement? Imagine I to be very small in comparison to $1/N$ (for example, with diameter of order e^{-N}), so that there is only very little chance that an eigenvalue falls into this interval (the average distance of eigenvalues of a well-behaved ensemble is of order $1/N$, so it is likely that eigenvalues miss an interval with diameter of order e^{-N}). Then still, (6.8) only allows us to predict the number of eigenvalues in I by $N\sigma(I)$ up to $N^{3/4+\epsilon}$ out of N eigenvalues, which is then not a useful estimate. In other words, (6.8) holds uniformly over all intervals, but it does not take into account the size of the interval. The natural way to account for the size of the interval is to divide both sides of the inequality inside the probability by it (since in well-behaved ensembles, the expected number of eigenvalues should be roughly proportional to the diameter of the interval). Since we want the right hand side to be non-increasing (eventually, we want to keep it slightly decreasing so we can underbid any given positive number), for any fixed $\epsilon\in(0,1/4)$ we can afford intervals of length $|I|\geq 1/N^{1/4-\epsilon}$ (by $|I|$ we will denote the diameter of an interval $I\subseteq\mathbb{R}$). This yields the following theorem, which for Tao and Vu actually constitutes "The Local Semicircle Law" (instead of a statement as our Theorem 6.10 involving Stieltjes transforms), see their Theorem 7 in [60, p. 7] and our Remark 6.22.

Theorem 6.20 (Interval-Type Weak Semicircle Laws). *Let $\theta \in (0,1)$ be fixed.*

i) For all $\epsilon \in (0,1/4)$ and $D > 0$, there is a constant $C_{\epsilon,D} \geq 0$, such that for all $N \in \mathbb{N}$:

$$\mathbb{P}\left(\sup_{\substack{I \in \mathcal{I}(\mathbb{R}) \\ |I| \geq \frac{1}{N^{1/4-\epsilon}}}} \frac{|\sigma_N(I) - \sigma(I)|}{|I|} > \frac{1}{N^{\frac{\epsilon}{2}}} \right) \leq \frac{C_{\epsilon,D}}{N^D}.$$

ii) For all $\epsilon \in (0,1/2)$ and $D > 0$ there is a constant $C'_{\epsilon,D}$ such that for all $N \in \mathbb{N}$:

$$\mathbb{P}\left(\sup_{\substack{I \in \mathcal{I}([-2+\theta, 2-\theta]) \\ |I| \geq \frac{1}{N^{1/2-\epsilon}}}} \frac{|\sigma_N(I) - \sigma(I)|}{|I|} > \frac{1}{N^{\frac{\epsilon}{2}}} \right) \leq \frac{C'_{\epsilon,D}}{N^D}.$$

iii) For all $\epsilon \in (0,1/2)$ and $D > 0$ there is a constant $C''_{\epsilon,D}$ such that for all $N \in \mathbb{N}$:

$$\mathbb{P}\left(\sup_{\substack{I \in \mathcal{I}([-2+\theta, 2-\theta]) \\ |I| \geq \frac{1}{N^{1/2-\epsilon}}}} \left| \frac{\sigma_N(I)}{\sigma(I)} - 1 \right| > \frac{1}{N^{\frac{\epsilon}{2}}} \right) \leq \frac{C''_{\epsilon,D}}{N^D}.$$

Proof. To prove the first statement, let $\epsilon \in (0,1/4)$ and $D > 0$ be given. From the first statement of Theorem 6.19 we find that for the events $(A_N)_N$ with

$$\forall\, N \in \mathbb{N}: \; A_N := \left\{ \sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \leq \frac{1}{N^{\frac{1}{4}-\frac{\epsilon}{2}}} \right\},$$

there is a constant $C_{\epsilon/2,D} \geq 0$ such that

$$\forall\, N \in \mathbb{N}: \; \mathbb{P}(A_N) > 1 - \frac{C_{\epsilon/2,D}}{N^D}.$$

Now let $N \in \mathbb{N}$ be arbitrary and $I \subseteq \mathbb{R}$ be an interval with $|I| \geq 1/N^{1/4-\epsilon}$, then on A_N:

$$\frac{|\sigma_N(I) - \sigma(I)|}{|I|} \leq \frac{|\sigma_N(I) - \sigma(I)|}{1/N^{1/4-\epsilon}} \leq \frac{N^{\frac{1}{4}-\epsilon}}{N^{\frac{1}{4}-\frac{\epsilon}{2}}} = \frac{1}{N^{\frac{\epsilon}{2}}}.$$

Therefore:

$$\forall\, N \in \mathbb{N}: \mathbb{P}\left(\sup_{\substack{I \in \mathcal{I}(\mathbb{R}) \\ |I| \geq \frac{1}{N^{1/4-\epsilon}}}} \frac{|\sigma_N(I) - \sigma(I)|}{|I|} \leq \frac{1}{N^{\frac{\epsilon}{2}}} \right) \geq \mathbb{P}(A_N) > 1 - \frac{C_{\epsilon/2,D}}{N^D},$$

which concludes the proof of the first statement by using the constants $(\epsilon, D) \mapsto C_{\epsilon/2,D}$. The proof of the second statement can be carried out analogously, using the second statement of Theorem 6.19 instead of the first. We will now proceed to rigorously prove the third statement, which is very similar to the second. So to start, we assume that $\theta \in (0,1)$ is fixed and choose $\epsilon \in (0,1/2)$ and $D > 0$ arbitrarily. Then the density f_σ of the semicircle distribution σ is lower bounded on $[-2+\theta, 2-\theta]$ by some constant $\beta_\theta > 0$. Therefore, for any interval $I \subseteq [-2+\theta, 2-\theta]$ we find $\sigma(I) \geq \beta_\theta |I|$. By the second statement of Theorem 6.19 and Lemma 6.7 *vii*), we have

$$\sup_{I \in \mathcal{I}([-2+\theta,2-\theta])} |\sigma_N(I) - \sigma(I)| \prec \frac{\beta_\theta}{N^{\frac{1}{2}}}.$$

In particular, we find that for the sets A'_N with

$$\forall\, N \in \mathbb{N} : A'_N := \left\{ \sup_{I \in \mathcal{I}([-2+\theta,2-\theta])} |\sigma_N(I) - \sigma(I)| \leq \frac{\beta_\theta}{N^{\frac{1}{2}-\frac{\epsilon}{2}}} \right\},$$

there exists are a constant $C_{\epsilon/2,D} \geq 0$ such that

$$\forall\, N \in \mathbb{N} : \mathbb{P}(A'_N) > 1 - \frac{C_{\epsilon/2,D}}{N^D}.$$

Now let $N \in \mathbb{N}$ be arbitrary and $I \subseteq [-2+\theta, 2-\theta]$ be an interval with $|I| \geq 1/N^{\frac{1}{2}-\epsilon}$. Then on A'_N:

$$\left| \frac{\sigma_N(I)}{\sigma(I)} - 1 \right| = |\sigma_N(I) - \sigma(I)| \cdot \frac{1}{\sigma(I)} \leq |\sigma_N(I) - \sigma(I)| \frac{1}{\beta_\theta |I|} \leq \frac{\beta_\theta}{N^{\frac{1}{2}-\frac{\epsilon}{2}}} \cdot \frac{N^{\frac{1}{2}-\epsilon}}{\beta_\theta} = \frac{1}{N^{\frac{\epsilon}{2}}}.$$

Therefore:

$$\forall\, N \in \mathbb{N} : \mathbb{P}\left(\sup_{\substack{I \in \mathcal{I}(\mathbb{R}) \\ |I| \geq \frac{1}{N^{1/2-\epsilon}}}} \left| \frac{\sigma_N(I)}{\sigma(I)} - 1 \right| \leq \frac{1}{N^{\frac{\epsilon}{2}}} \right) \geq \mathbb{P}(A'_N) > 1 - \frac{C_{\epsilon/2,D}}{N^D},$$

which concludes the proof of the third statement by using the constants $(\epsilon, D) \mapsto C_{\epsilon/2,D}$. $\square$

Statement *iii*) of Theorem 6.20 is very interesting. It allows conclusions about the *relative* deviation $\sigma_N(I)/\sigma(I)$ for all intervals I in the bulk with a minimal length and thus resolves the problems discussed before Theorem 6.20. In particular, we now know exactly for which intervals I, $\sigma_N(I)$ can be sensibly predicted by $\sigma(I)$, and which relative error to expect.

Remark 6.21. Let us now investigate how far we would get by squeezing as much out of the *global law* as possible. We remember that the global law merely states that σ_N converges to σ weakly in probability (or almost surely). Denote by F_{σ_N} and F_σ the corresponding distribution functions, then since F_σ is continuous, we obtain that $\|F_{\sigma_N} - F_\sigma\|_\infty \to 0$ in probability. To see this, let $I \subseteq \mathbb{N}$ be an arbitrary subsequence and d be a metric on the space of probability measures that metrizes weak convergence, then $d(\sigma_N, \sigma) \to 0$ in probability for $N \in I$, so for some subsequence $J \subseteq I$, $d(\sigma_N, \sigma) \to 0$ almost surely for $N \in J$, that is, this happens on a measurable set A with $\mathbb{P}(A) = 1$. Then for all $\omega \in A$ we find $\sigma_N(\omega) \to \sigma$ weakly for $N \in J$, but this entails that $\|F_{\sigma_N(\omega)} - F_\sigma\|_\infty \to 0$ for $N \in J$, since F_σ is continuous (see [34, p. 141]). So indeed, the global semicircle law in probability yields $\|F_{\sigma_N} - F_\sigma\|_\infty \to 0$ in probability (since any subsequence has an almost surely convergent subsequence). However, we know nothing about the rate of convergence. We will be pragmatic and assume a C_ϵ/N-rate, which means that for all $\epsilon > 0$ there is a $C_\epsilon \geq 0$ such that

$$\forall\, N \in \mathbb{N}: \; \mathbb{P}\left(\|F_{\sigma_N} - F_\sigma\|_\infty \leq \epsilon\right) \geq 1 - \frac{C_\epsilon}{N}. \tag{6.9}$$

This is a valid probabilistic statement, that is, $\|F_{\sigma_N} - F_\sigma\|_\infty$ is measurable, since

$$\begin{aligned}\sup_{x\in\mathbb{R}} |F_{\sigma_N}(x) - F_\sigma(x)| &= \sup_{x\in\mathbb{R}} |\sigma_N((-\infty, x]) - \sigma((-\infty, x])| \\ &= \sup_{x\in\mathbb{Q}} |\sigma_N((-\infty, x]) - \sigma((-\infty, x])|,\end{aligned}$$

where the last equation follows easily from continuity of probability measures. Thus, measurability of $\|F_{\sigma_N} - F_\sigma\|_\infty$ is due to the measurability of $|\sigma_N((-\infty, x]) - \sigma((-\infty, x])|$ for each $x \in \mathbb{Q}$.

Now if $\|F_{\sigma_N} - F_\sigma\|_\infty \leq \epsilon$, this entails that for any interval $I \subseteq \mathbb{R}$, $|\sigma_N(I) - \sigma(I)| \leq 2\epsilon$. To show this, it is easily verified that the bound is even ϵ instead of 2ϵ for intervals of the type $(-\infty, x]$, $(-\infty, x)$, $[x, \infty)$ and (x, ∞), where $x \in \mathbb{R}$. This is then used to derive the 2ϵ-bound for all other interval types. Therefore, from (6.9) it follows that

$$\forall\, N \in \mathbb{N}: \; \mathbb{P}\left(\sup_{I\in\mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \leq 2\epsilon\right) \geq 1 - \frac{C_\epsilon}{N}.$$

Renaming our constants C_ϵ, we find in particular that for all $\epsilon \in (0, 1)$,

$$\forall\, N \in \mathbb{N}: \; \mathbb{P}\left(\sup_{I\in\mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \leq \frac{\epsilon N}{N}\right) \geq 1 - \frac{C_\epsilon}{N}. \tag{6.10}$$

What that means now is that when estimating $\sigma_N(I)$ by $\sigma(I)$ under knowledge of the global law, we can merely control an error of ϵ independent of N. In contrast to the small scale results from the local law, a growing N will not yield more accurate predictability of $\sigma_N(I)$ by $\sigma(I)$. Likewise, when it comes to predicting the number of eigenvalues in a given interval I, we can only do so up to a number of eigenvalues proportional to N, namely

ϵN, and the probability that our forecast does not fall within these bounds decays only at a rate of C_ϵ/N (by assumption) instead of arbitrarily polynomially fast. Likewise, for any fixed $\theta \in (0,1)$ and $\epsilon, c > 0$, we may conclude that there is a constant $K_{\epsilon,c} > 0$ such that

$$\mathbb{P}\left(\sup_{\substack{I \in \mathcal{I}([-2+\theta, 2-\theta]) \\ |I| \geq c}} \left| \frac{\sigma_N(I)}{\sigma(I)} - 1 \right| > \epsilon \right) \leq \frac{K_{\epsilon,c}}{N}. \tag{6.11}$$

To see this, denote by $\beta_\theta > 0$ the minimum of f_σ on $[-2+\theta, 2-\theta]$ and set

$$\forall\, N \in \mathbb{N}: \; A_N := \left\{ \sup_{I \in \mathcal{I}(\mathbb{R})} |\sigma_N(I) - \sigma(I)| \leq \epsilon c \beta_\theta \right\}.$$

Then on A_N, for any $I \subseteq [-2+\theta, 2-\theta]$ with $|I| > c$, we obtain

$$\left| \frac{\sigma_N(I)}{\sigma(I)} - 1 \right| = |\sigma_N(I) - \sigma(I)| \frac{1}{\sigma(I)} \leq \epsilon c \beta_\theta \frac{1}{\beta_\theta |I|} \leq \epsilon.$$

Since $\mathbb{P}(A_N) \geq 1 - C_{\epsilon c \beta_\theta}/N$, setting $K_{\epsilon,c} := C_{\epsilon c \beta_\theta}$, we obtain (6.11). Note the slow convergence speed of $O(1/N)$, which stems from our assumption in (6.9). Note also that the size of the intervals is not allowed to decrease as $N \to \infty$, but that $|I| \geq c$ is required. In other words, the global law truly is not a local law.

Remark 6.22. In this remark, we would like to mention what kind of improvements can be made to all of the above theorems if the stronger local law as in [9] were known. In the weak local law, Theorem 6.10, the error term on the right hand side of $\prec$ could be replaced by $1/N\eta$. Accordingly, in the simultaneous weak local law, the denominator on the left hand side of $\prec$ can be replaced by the term $1/N\eta$. In Corollary 6.16, the denominator on the right hand side of $\prec$ can be replaced by $1/N^\gamma$. Further, in Corollary 6.17, the term $N^{\gamma/4}$ inside the probability can be replaced by N^γ. In the semicircle law on small scales, Theorem 6.19, the first statement of $\prec$ can be improved in such a way that the right hand side is replaced by $1/N$. The second $\prec$ statement will then be redundant. In the interval-type weak semicircle law, Theorem 6.20, in all these statements we may consider $\epsilon \in (0,1)$ instead of $\epsilon \in (0,1/4)$ or $\epsilon \in (0,1/2)$. Further, in all statements, intervals of length $|I| \geq N^{\epsilon-1}$ may be considered, rendering statement *ii*) redundant. As mentioned before, these statements are in the spirit of "The Local Semicircle Law" as formulated by Tao and Vu in [60, p. 7].

Having discussed the weak local law, Theorem 6.10, and its consequences in this chapter, we will now move on to prove it.

6.4 Proof of The Weak Local Law

The proof of Theorem 6.10 is obtained through five steps, thus it is carried out in the following five subsections (actually, we choose an argumentation that already yields the stronger simultaneous version, Theorem 6.15). We follow the line of the sketched proof in Chapter 7 of [27] as inspiration. Our proof accommodates the setting of Curie-Weiss type ensembles. Further, we incorporate some ideas as in [9]. For example, we prefer the initial estimate as in [9], since it is simplified and sufficient.

At first, a purely deterministic (but very thorough) stability analysis will analyze the equation in $m \in \mathbb{C}$

$$m + \frac{1}{z+m} = 0$$

for fixed $z \in [-10, 10] \times i(0, 10]$. We know from Theorem 5.16 that if an m satisfies this equation, then $m \in \{s(z), 1/s(z)\}$. The question is, if m *almost* satisfies this equation (which means that $|m + 1/(z + m)|$ is small), then how far is m from $s(z)$ or $1/s(z)$? In the second step, an initial estimate for the local law will be derived. To be more precise, it will be proved to be valid on the smaller (and N-independent) domain $\mathcal{D}_I$ instead of $\mathcal{D}_N(\gamma)$. In the third step, with help of bootstrap and continuity arguments, the validity of the local law will be extended to the domain $\mathcal{D}_N(\gamma)$.

In the analysis that follows we will use the notational convention that exponents bind stronger than other operations, for example, if Y is an invertible $N \times N$ matrix and tr is the trace operator, then $\operatorname{tr} Y^{-1} = \operatorname{tr}[Y^{-1}]$.

6.4.1 Step 1: Deterministic Stability Analysis

The Landau symbol O is well-known: For functions f and g we write $f = O(g)$, if $|f| \leq C \cdot g$ for some constant $C > 0$. To be more precise in the following, we specifically write $f = O_C(g)$ with $C > 0$, if $|f| \leq C \cdot g$.

Now to start with our analysis, we need the following lemma:

Lemma 6.23. *Let $(X_N)_N$ be a Curie-Weiss type ensemble and $(s_N)_N$ the Stieltjes transforms of the ESDs of $(X_N)_N$. Then if $z = E + i\eta$ where $E \in \mathbb{R}$ and $\eta > 0$, we obtain:*

$$\forall\, N \in \mathbb{N} : s_N(z) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{-z - s_N(z) + \Omega_k},$$

where

$$\Omega_k = X_N(k,k) - Z_k + O_1\left(\frac{1}{N\eta}\right) \quad \text{and} \quad Z_k = \sum_{i \neq j} x_k(i)(X_N^{(k)} - z)^{-1}(i,j)x_k(j).$$

In particular, the terms Ω_k, Z_k and x_k all depend on N, which is dropped from the notation. Here, $X_N^{(k)}$ denotes the k-th principle minor of X_N and x_k denotes the k-th column of X_N without the k-th entry.

Proof. For $N \in \mathbb{N}$, Theorem 5.19 yields

$$s_N(z) = S_{\sigma_N}(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{X_N(k,k) - z - x_k^*(X_N^{(k)} - z)^{-1}x_k}.$$

Now for $k \in \{1, \dots, N\}$ arbitrary we find, considering that entries of X_N are $\pm 1/\sqrt{N}$-valued,

$$\begin{aligned}
&X_N(k,k) - z - x_k^*(X_N^{(k)} - z)^{-1}x_k \\
&= X_N(k,k) - z - \underbrace{\sum_{i \neq j} x_k(i)(X_N^{(k)} - z)^{-1}(i,j)x_k(j)}_{=Z_k} - \frac{1}{N}\operatorname{tr}(X_N^{(k)} - z)^{-1} \\
&= X_N(k,k) - z - Z_k - s_N(z) + \underbrace{s_N(z) - \frac{1}{N}\operatorname{tr}(X_N^{(k)} - z)^{-1}}_{|\dots| \leq \frac{1}{N\eta} \text{by Theorem 5.19}} \\
&= -z - s_N(z) + \underbrace{X_N(k,k) - Z_k + O_1\left(\frac{1}{N\eta}\right)}_{=\Omega_k},
\end{aligned}$$

□

We should obtain

$$s_N(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{-z - s_N(z) + \Omega_k} \approx \frac{1}{N}\sum_{k=1}^{N} \frac{1}{-z - s_N(z)} = \frac{1}{-z - s_N(z)}$$

if we can show that all Ω_k are small, which then should entail

$$s_N(z) \approx s(z).$$

The following lemma and theorem show us how we can make this rigorous, see also pages 41 through 43 in [27]:

Theorem 6.24 (Geometric Series Expansion)**.** *In the situation above, if*

$$s_N(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{-z - s_N(z) + \Omega_k}$$

and

$$\frac{\max_k |\Omega_k|}{|z + s_N(z)|} \leq \frac{1}{2},$$

then

$$\left| s_N(z) + \frac{1}{z + s_N(z)} \right| \leq \frac{2\max_k |\Omega_k|}{|z + s_N(z)|^2}.$$

Proof. First, we make a general observation: If $x \in \mathbb{C}$ with $|x| < 1$, then

$$\frac{1}{1-x} = \sum_{n\in\mathbb{N}_0} x^n = 1 + x + \sum_{n\geq 2} x^n = 1 + x + x^2 \sum_{n\in\mathbb{N}_0} x^n = 1 + x + \frac{x^2}{1-x},$$

thus

$$\frac{1}{1-x} = 1 + x + R(x), \qquad |R(x)| = \frac{|x|^2}{|1-x|} \leq \frac{|x|^2}{1-|x|}.$$

Now,

$$\begin{aligned}
s_N(z) &= \frac{1}{N}\sum_{k=1}^{N} \frac{1}{-z - s_N(z) + \Omega_k} \\
&= \frac{1}{N}\sum_{k=1}^{N} \frac{1}{-z - s_N(z)} \frac{1}{1 - \frac{\Omega_k}{z+s_N(z)}} \\
&= \frac{1}{N}\sum_{k=1}^{N} \frac{1}{-z - s_N(z)} \left(1 + \frac{\Omega_k}{z+s_N(z)} + R\left(\frac{\Omega_k}{z+s_N(z)}\right)\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left| s_N(z) + \frac{1}{z+s_N(z)} \right| &\leq \frac{1}{N}\sum_{k=1}^{N} \left(\frac{|\Omega_k|}{|z+s_N(z)|^2} + \frac{1}{|z+s_N(z)|} \left| R\left(\frac{\Omega_k}{z+s_N(z)}\right)\right| \right) \\
&\leq \frac{\max_k |\Omega_k|}{|z+s_N(z)|^2} + \frac{1}{N}\sum_{k=1}^{N} \frac{1}{|z+s_N(z)|} \frac{\frac{|\Omega_k|^2}{|z+s_N(z)|^2}}{1 - \frac{|\Omega_k|}{|z+s_N(z)|}} \\
&\leq \frac{\max_k |\Omega_k|}{|z+s_N(z)|^2} + \frac{\max_k |\Omega_k|}{|z+s_N(z)|^2} \cdot \frac{\frac{\max_k |\Omega_k|}{|z+s_N(z)|}}{1 - \frac{\max_k |\Omega_k|}{|z+s_N(z)|}} \\
&\leq 2\frac{\max_k |\Omega_k|}{|z+s_N(z)|^2},
\end{aligned}$$

where we used that $a \mapsto \frac{a^2}{1-a}$ and $a \mapsto \frac{a}{1-a}$ are isotonic on $[0,1)$ and $\frac{\frac{1}{2}}{1-\frac{1}{2}} = 1$. □

Next, if $z \in \mathbb{C}_+$ is fixed, we want to answer the following question: If $|m + \frac{1}{z+m}| \leq \delta \leq 1$ for some $\delta \in [0,1]$, then how close is m to one of the roots of the equation

$$m^2 + zm + 1 = 0$$

in dependence of δ? By Theorem 5.16, we know the roots to be $s(z)$ and $\frac{1}{s(z)}$, so the question is to find an upper bound on

$$\min\left(|m - s(z)|, \left|m - \frac{1}{s(z)}\right|\right)$$

in dependence of δ. If $\delta = 0$, above equation is satisfied by m, so m is exactly $s(z)$ or $\frac{1}{s(z)}$ and thus

$$\min\left(|m - s(z)|, \left|m - \frac{1}{s(z)}\right|\right) = 0.$$

For general $\delta \in [0,1]$, the following theorem helps. Although deterministic in nature, it constitutes one of the key ingredients for the proof of the local law.

Theorem 6.25 (Proximity Theorem)**.** *There is a $C_P > 0$, so that the following holds: If $z \in [-10, 10] + i(0, 10]$ and $m \in \mathbb{C}$ are arbitrary with*

$$\left|m + \frac{1}{z+m}\right| \leq \delta$$

for a $\delta \in [0,1]$, then it follows

$$\min\left\{|m - s(z)|, \left|m - \frac{1}{s(z)}\right|\right\} \leq \frac{C_P \delta}{\sqrt{\kappa + \eta + \delta}} \leq C_P \sqrt{\delta}.$$

where $E = E(z) = \mathrm{Re}(z)$, $\eta = \eta(z) = \mathrm{Im}(z)$ and $\kappa = ||E| - 2|$.

Proof. Clearly, we only need to show the first inequality. The proof will be subdivided into three steps:
Step 1: It holds that $|m| \leq 17$.
We find

$$\left|m + \frac{1}{z+m}\right| \leq \delta \leq 1, \quad \text{thus} \quad m + \frac{1}{z+m} = x \quad \text{for some } x \in \mathbb{C} \text{ with } \quad |x| \leq 1.$$

It follows with $|z| \leq 15$ that

$$\begin{aligned}
&m^2 + zm + 1 = zx + mx \\
\Rightarrow\ &m^2 + (z - x)m + 1 - zx = 0 \\
\Rightarrow\ &m_{1,2} = -\frac{z-x}{2} \pm \sqrt{\frac{(z-x)^2}{4} - 1 + zx} \\
\Rightarrow\ &|m| \leq \frac{|z| + |x|}{2} + \sqrt{\frac{(|z| + |x|)^2}{4} + 1 + |z|} \\
&\quad\ \leq \frac{16}{2} + \sqrt{64 + 16} \\
&\quad\ \leq 17.
\end{aligned}$$

Step 2: Bounding the Product of the Terms in the Minimum
Let again $x = m + \frac{1}{z+m}$, then $|x| \leq \delta \leq 1$. Since

$$s(z) + \frac{1}{z + s(z)} = 0,$$

it follows

$$\begin{aligned}
& s(z) + \frac{1}{z+s(z)} - x = -x \\
\Rightarrow\ & s(z) + \frac{1}{z+s(z)} - m - \frac{1}{z+m} = -x \\
\Rightarrow\ & m - s(z) - \frac{m-s(z)}{(z+m)(z+s(z))} = x \\
\Rightarrow\ & (m-s(z)) \left[1 - \frac{1}{(z+m)(z+s(z))}\right] = x \\
\Rightarrow\ & (m-s(z)) \left[m + z - \frac{1}{z+s(z)}\right] = x(m+z).
\end{aligned}$$

Now

$$z - \frac{1}{z+s(z)} = z + s(z) = -\frac{1}{s(z)},$$

where we used for both equalities that $s(z) + 1/(z+s(z)) = 0$. We obtain

$$|m-s(z)| \left| m - \frac{1}{s(z)} \right| = |x(m+z)| \leq 32\delta = D\delta, \tag{6.12}$$

where we set $D \coloneqq 32$. From this we obtain through $s(z) = s(z)^2/s(z)$:

$$|m-s(z)| \left| m - s(z) - \frac{1-s(z)^2}{s(z)} \right| \leq D\delta \tag{6.13}$$

Step 3: Proof of Statement by Case-by-Case Analysis

Throughout this step, we will use Theorem 5.16 on page 129 and the constant C_s therein. We set

$$C' \coloneqq C_s \sqrt{\frac{D}{2}}, \qquad \text{then} \qquad 2\left(\frac{C'}{|s(z)|}\right)^2 \geq D, \qquad \text{since} \qquad |s(z)| \leq C_s. \tag{6.14}$$

1. Case: $|1 - s(z)^2| \leq C'\sqrt{\delta}$

Then (6.13) implies

$$|m-s(z)| \leq \frac{2C'\sqrt{\delta}}{|s(z)|}.$$

To see this, if $|m-s(z)| > \frac{2C'\sqrt{\delta}}{|s(z)|}$ were true, it would follow for the second factor in (6.13) that

$$\left| m - s(z) - \frac{1-s(z)^2}{s(z)} \right| \geq |m-s(z)| - \left| \frac{1-s(z)^2}{s(z)} \right| > \frac{C'\sqrt{\delta}}{|s(z)|}.$$

Therefore, the l.h.s. of (6.13) would be strictly larger than $2\left(\frac{C'}{|s(z)|}\right)^2 \delta$, which is a contradiction due to (6.14). We conclude

$$|m-s(z)| \leq \frac{2C'\sqrt{\delta}}{|s(z)|} \leq 2C'C_s\sqrt{\delta}.$$

From Theorem 5.16 and the case assumption it also follows

$$\sqrt{\kappa+\eta+\delta} \leq \sqrt{\kappa+\eta} + \sqrt{\delta} \leq C_s^2|1-s(z)^2| + \sqrt{\delta} \leq C_s^2 C'\sqrt{\delta} + \sqrt{\delta} = (C_s^2 C' + 1)\sqrt{\delta}.$$

Therefore,

$$(\sqrt{\kappa+\eta+\delta})|m-s(z)| \leq 2C'C_s(C_s^2 C'+1)\cdot\delta,$$

which shows the statement with constant $C_p^{(1)} := 2C'C_s(C_s^2 C'+1)$.
2. Case: $|1-s(z)^2| > C'\sqrt{\delta}$
Then with Theorem 5.16 we obtain

$$\sqrt{\kappa+\eta} \geq \frac{1}{C_s^2}|1-s(z)^2| > \frac{C'}{C_s^2}\sqrt{\delta},$$

and using this to bound $\sqrt{\delta}$ from above we obtain

$$\sqrt{\kappa+\eta+\delta} \leq \sqrt{\kappa+\eta} + \sqrt{\delta} \leq \left(\frac{C_s^2}{C'}+1\right)\sqrt{\kappa+\eta},$$

which we will use in the following two subcases:
1. Subcase: $|m-s(z)| \leq \frac{1}{2}\frac{|1-s(z)^2|}{|s(z)|}$.
Then it holds for the second factor in (6.13), that

$$\begin{aligned}
\left|m-s(z)-\frac{1-s(z)^2}{s(z)}\right| &\geq \left|\frac{1-s(z)^2}{s(z)}\right| - |m-s(z)| \\
&\geq \frac{1}{2}\left|\frac{1-s(z)^2}{s(z)}\right| \\
&\geq \frac{1}{2}\frac{1}{C_s}\sqrt{\kappa+\eta} \\
&\geq \frac{1}{\frac{C_s^2}{C'}+1}\cdot\frac{1}{2}\frac{1}{C_s}\sqrt{\kappa+\eta+\delta}.
\end{aligned}$$

We deduce in combination with (6.13) that

$$|m-s(z)| \leq 2DC_s\left(\frac{C_s^2}{C'}+1\right)\frac{\delta}{\sqrt{\kappa+\eta+\delta}},$$

which shows the statement with constant $C_p^{(2)} := 2DC_s(C_s^2/C'+1)$.
2. Subcase: $|m-s(z)| > \frac{1}{2}\frac{|1-s(z)^2|}{|s(z)|}$
Then as above,

$$|m-s(z)| \geq \frac{1}{2}\frac{|1-s(z)^2|}{|s(z)|} \geq \frac{1}{\frac{C_s^2}{C'}+1}\cdot\frac{1}{2}\frac{1}{C_s}\sqrt{\kappa+\eta+\delta}$$

and it follows with (6.12) that

$$\left|m-\frac{1}{s(z)}\right| \leq 2DC_s\left(\frac{C_s^2}{C'}+1\right)\frac{\delta}{\sqrt{\kappa+\eta+\delta}},$$

which again shows the statement with constant $C_p^{(2)}$ as defined above. Setting $C_P := \max\left\{C_p^{(1)}, C_p^{(2)}\right\}$ concludes the proof. □

The results we obtained in this section so far allow us to prove a very important lemma, which we will heavily draw upon:

Lemma 6.26 (Deterministic Root Approximation)**.** *In the situation of Lemma 6.23, assume $z \in [-10, 10] + i(0, 10]$ is arbitrary so that in particular,*

$$s_N(z) = \frac{1}{N} \sum_{k\in[N]} \frac{1}{-z - s_N(z) + \Omega_k}.$$

If we have

$$\frac{\max_k |\Omega_k|}{|z + s_N(z)|}, \; \frac{\max_k |\Omega_k|}{|z + s_N(z)|^2} \leq \frac{1}{2},$$

then it follows that

$$\min\left\{|s_N(z) - s(z)|, \left|s_N(z) - \frac{1}{s(z)}\right|\right\} \leq C_{Det} \min\left\{\frac{\max_k |\Omega_k|}{\sqrt{\kappa}|z + s_N(z)|^2}, \sqrt{\frac{\max_k |\Omega_k|}{|z + s_N(z)|^2}}\right\},$$

where $C_{Det} = 2C_P$, and C_P is the constant from Theorem 6.25.

Proof. Due to Theorem 6.24 we have

$$\left|s_N(z) + \frac{1}{z + s_N(z)}\right| \leq \frac{2\max_k |\Omega_k|}{|z + s_N(z)|^2} =: \delta \leq 1.$$

With Theorem 6.25 we obtain

$$\begin{aligned}
\min\left\{|s_N(z) - s(z)|, \left|s_N(z) - \frac{1}{s(z)}\right|\right\} &\leq C_P \min\left\{\frac{\delta}{\sqrt{\kappa}}, \sqrt{\delta}\right\} \\
&= C_P \min\left\{\frac{2\max_k |\Omega_k|}{\sqrt{\kappa}|z + s_N(z)|^2}, \sqrt{\frac{2\max_k |\Omega_k|}{|z + s_N(z)|^2}}\right\} \\
&\leq \underbrace{2C_P}_{=:C_{Det}} \min\left\{\frac{\max_k |\Omega_k|}{\sqrt{\kappa}|z + s_N(z)|^2}, \sqrt{\frac{\max_k |\Omega_k|}{|z + s_N(z)|^2}}\right\}
\end{aligned}$$

□

6.4.2 Step 2: Large Deviations Estimates

In the setting of Lemma 6.23, we would like to show the smallness of

$$\max_k |\Omega_k| = \max_k \left|X_N(k,k) - Z_k + O_1\left(\frac{1}{N\eta}\right)\right|.$$

Since $|X_N(k,k)| = \frac{1}{\sqrt{N}}$, the only component left to analyze is

$$\max_k |Z_k| = \max_k \left| \sum_{i \neq j} x_k(i)(X_N^{(k)} - z)^{-1}(i,j)x_k(j) \right|.$$

To this end, we need a small collection of large deviation inequalities. We would like to emphasize that the term "large deviation inequality" in the local law literature is not used in the sense of the standard probabilistic theory of large deviations. Rather, these inequalities are meant to establish bounds on the deviations of random variables in the sense of stochastic domination $\prec$. In the following, for $p \geq 1$ the norm $\|\cdot\|_p$ shall denote the $\mathcal{L}_p(\mathbb{P})$-seminorm, so for any random variable $Y : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow \mathbb{C}$, $\|Y\|_p = (\mathbb{E}|Y|^p)^{1/p}$.

Theorem 6.27 (Marcinkiewicz-Zygmund Inequality)**.** *If $Y_1, \ldots, Y_N$ are independent, centered and complex-valued random variables with existing absolute moments, then for every $p \geq 1$ there exists a positive constant A_p which depends only on p, such that*

$$\left\| \sum_{i=1}^N Y_i \right\|_p \leq A_p \left\| \left(\sum_{i=1}^N |Y_i|^2 \right)^{\frac{1}{2}} \right\|_p$$

Proof. In [18, p. 386], the statement is proved for independent real-valued random variables. As was also shown in [7, p. 33], the statement easily carries over to the complex case by the following calculation, where we assume A'_p to be the constants in the real case:

$$\begin{aligned}
\left\| \sum_{i=1}^N Y_i \right\|_p &\leq \left\| \sum_{i=1}^N \operatorname{Re} Y_i \right\|_p + \left\| \sum_{i=1}^N \operatorname{Im} Y_i \right\|_p \\
&\leq A'_p \left\| \left(\sum_{i=1}^N |\operatorname{Re} Y_i|^2 \right)^{\frac{1}{2}} \right\|_p + A'_p \left\| \left(\sum_{i=1}^N |\operatorname{Im} Y_i|^2 \right)^{\frac{1}{2}} \right\|_p \\
&\leq 2A'_p \left\| \left(\sum_{i=1}^N |Y_i|^2 \right)^{\frac{1}{2}} \right\|_p .
\end{aligned}$$

Therefore, the statement is true for the complex case with constants $A_p = 2A'_p$. □

The following three lemmas (and their proofs) are taken from [9] and are included for completeness and convenience.

Lemma 6.28. *Let $Y_1, \ldots, Y_N$ be independent, complex-valued random variables which are centered and uniformly $\|\cdot\|_p$-bounded by constants μ_p for all $p \geq 2$. Then it holds for any complex numbers $b_1, \ldots, b_N$ that*

$$\forall p \geq 2 : \left\| \sum_{i=1}^N b_i Y_i \right\|_p \leq A_p \mu_p \left(\sum_{i=1}^N |b_i|^2 \right)^{\frac{1}{2}},$$

where A_p is a constant depending only on p, which can be chosen as in Theorem 6.27.

Proof. We may assume that not all b_i vanish. Setting

$$\beta^2 := \sum_{i=1}^{N} |b_i|^2 > 0,$$

we calculate, using Theorem 6.27 (and the constant A_p therein) and convexity of $t \mapsto |t|^{p/2}$ for $p \geq 2$, that

$$\begin{aligned}
\mathbb{E}\left|\sum_{i=1}^{N} b_i X_i\right|^p &\leq A_p^p \cdot \mathbb{E}\left|\sum_{i=1}^{N} |b_i|^2 |X_i|^2\right|^{\frac{p}{2}} \\
&= A_p^p \cdot \beta^p \cdot \mathbb{E}\left|\sum_{i=1}^{N} \frac{|b_i|^2}{\beta^2} |X_i|^2\right|^{\frac{p}{2}} \\
&\leq A_p^p \cdot \beta^p \cdot \sum_{i=1}^{N} \frac{|b_i|^2}{\beta^2} \mathbb{E}|X_i|^p \\
&\leq A_p^p \cdot \beta^p \cdot \mu_p^p.
\end{aligned}$$

□

Lemma 6.29. *Let $Y_1, \ldots Y_N, Z_1, \ldots Z_N$ be independent, complex-valued random variables which are centered and uniformly $\|\cdot\|_p$-bounded by constants μ_p for all $p \geq 2$. Then it holds for any complex numbers $(a_{i,j})_{i,j\in[N]}$ that*

$$\forall\, p \geq 2 : \left\|\sum_{i,j=1}^{N} a_{i,j} Y_i Z_j\right\|_p \leq A_p^2 \mu_p^2 \left(\sum_{i,j=1}^{N} |a_{i,j}|^2\right)^{\frac{1}{2}},$$

where A_p is a constant depending only on p, which can be chosen as in Theorem 6.27.

Proof. We set for all $j = 1, \ldots, N$:

$$B_j := \sum_{i=1}^{N} a_{i,j} Y_i,$$

so that

$$\sum_{i,j=1}^{N} a_{i,j} Y_i Z_j = \sum_{j=1}^{N} B_j Z_j.$$

Now define $B := (B_1, \ldots, B_N)$ and $Z := (Z_1, \ldots, Z_N)$ as vector-valued random variables, then B and Z are independent, which allows us to conclude with Fubini and Lemma 6.28:

$$\begin{aligned}\mathbb{E}\left|\sum_{i,j=1}^{N} a_{i,j}Y_iZ_j\right|^p &= \mathbb{E}\left|\sum_{j=1}^{N} B_jZ_j\right|^p = \int\left|\sum_{j=1}^{N} b_jz_j\right|^p \mathrm{d}\mathbb{P}^{(Z,B)}(z,b)\\ &= \int\int\left|\sum_{j=1}^{N} b_jz_j\right|^p \mathrm{d}\mathbb{P}^{Z}(z)\mathrm{d}\mathbb{P}^{B}(b) = \int\mathbb{E}\left|\sum_{j=1}^{N} b_jZ_j\right|^p \mathrm{d}\mathbb{P}^{B}(b)\\ &\le \int A_p^p\cdot\mu_p^p\cdot\left(\sum_{j=1}^{N}|b_j|^2\right)^{\frac{p}{2}}\mathrm{d}\mathbb{P}^{B}(b) = A_p^p\cdot\mu_p^p\cdot\mathbb{E}\left(\sum_{j=1}^{N}|B_j|^2\right)^{\frac{p}{2}}.\end{aligned}$$

We conclude

$$\begin{aligned}\left\|\sum_{i,j=1}^{N} a_{i,j}Y_iZ_j\right\|_p &\le A_p\cdot\mu_p\cdot\left(\mathbb{E}\left(\sum_{j=1}^{N}|B_j|^2\right)^{\frac{p}{2}}\right)^{\frac{1}{p}} = A_p\cdot\mu_p\cdot\left(\left\|\sum_{j=1}^{N}|B_j|^2\right\|_{\frac{p}{2}}\right)^{\frac{1}{2}}\\ &\le A_p\cdot\mu_p\cdot\left(\sum_{j=1}^{N}\|B_j\|_p^2\right)^{\frac{1}{2}} \le A_p\cdot\mu_p\cdot\left(\sum_{j=1}^{N}A_p^2\cdot\mu_p^2\cdot\sum_{i=1}^{N}|a_{i,j}|^2\right)^{\frac{1}{2}}\\ &= A_p^2\cdot\mu_p^2\cdot\left(\sum_{i,j=1}^{N}|a_{i,j}|^2\right)^{\frac{1}{2}},\end{aligned}$$

where in the fourth step we used Lemma 6.28 again. □

Lemma 6.30. *Let $Y_1,\dots,Y_N$ be independent, complex-valued random variables which are centered and uniformly $\|\cdot\|_p$-bounded by constants μ_p for all $p\ge 2$. Then it holds for any complex numbers $(a_{i,j})_{i,j\in[N]}$ that*

$$\forall p\ge 2:\left\|\sum_{i\ne j=1}^{N} a_{i,j}Y_iY_j\right\|_p \le 4A_p^2\mu_p^2\left(\sum_{i\ne j=1}^{N}|a_{i,j}|^2\right)^{\frac{1}{2}},$$

where A_p is a constant depending only on p, which can be chosen as in Theorem 6.27.

Proof. We begin with noting that for all $i,j\in[N]$ with $i\ne j$, we obtain

$$1=\frac{1}{Z_N}\sum_{I\dot\cup J=[N]}\mathbb{1}_I(i)\mathbb{1}_J(j),$$

where the sum ranges over all partitions of $[N]=\{1,\dots,N\}$ into two sets, and $Z_N := 2^{N-2}$. To see that this is true, we just have to count all possible partitions of $[N]$ into sets I and J such that $i\in I$ and $j\in J$, where $i\ne j$ are fixed elements in $[N]$. To this end, we must count all subsets of $[N]$ containing i but not j. Finally, these are exactly all sets of the form $A \cup \{i\}$ for subsets $A \subseteq [N]\backslash\{i,j\}$, and these are 2^{N-2}. Now, we calculate for $p \geq 2$:

$$\begin{aligned}
\left\| \sum_{i\neq j=1}^{N} a_{i,j} Y_i Y_j \right\|_p &= \frac{1}{Z_N} \left\| \sum_{i\neq j\in[N]} \sum_{I\dot{\cup} J=[N]} \mathbb{1}_I(i)\mathbb{1}_J(j) a_{i,j} Y_i Y_j \right\|_p \\
&= \frac{1}{Z_N} \left\| \sum_{I\dot{\cup} J=[N]} \sum_{i\in I, j\in J} a_{i,j} Y_i Y_j \right\|_p \\
&\leq \frac{1}{Z_N} \sum_{I\dot{\cup} J=[N]} \left\| \sum_{i\in I, j\in J} a_{i,j} Y_i Y_j \right\|_p \\
&\leq \frac{2^N}{2^{N-2}} A_p^2 \mu_p^2 \left(\sum_{i\neq j=1}^{N} |a_{i,j}|^2 \right)^{\frac{1}{2}},
\end{aligned}$$

where in the last step we used Lemma 6.29. To see why the lemma is applicable, note that I and J are disjoint, and that complementing families of independent random variables with constant zero-variables will again yield independent families. □

The next theorem generalizes Lemma 6.28, Lemma 6.29 and Lemma 6.30 to independent random variables with a common expectation $t \in \mathbb{C}$, which is different from zero. Note that if t=0, we obtain exactly the statements of the Lemmas mentioned, so that the next theorem is a true generalization. Theorem 6.31 is a key step to obtaining the local law for Curie-Weiss type ensembles and thus a key contribution of this part of the dissertation.

Theorem 6.31. *Let $N \in \mathbb{N}$ be arbitrary, $(a_{i,j})_{i,j\in[N]}$ and $(b_i)_{i\in[N]}$ be deterministic complex numbers, $(Y_i)_{i\in[N]}$ and $(Z_i)_{i\in[N]}$ be complex-valued random variables with common expectation $t \in \mathbb{C}$, so that the whole family $\mathcal{W} := \{Y_i \mid i \in [N]\} \cup \{Z_i \mid i \in [N]\}$ is independent. Further, we assume that for all $p \geq 2$ there exists a $\mu_p \in \mathbb{R}_+$ such that $\|W - t\|_p \leq \mu_p$ for all $W \in \mathcal{W}$. Then we obtain for all $p \geq 2$:*

i) $$\left\| \sum_{i\in[N]} b_i Y_i \right\|_p \leq \left(A_p\mu_p + \sqrt{N}|t| \right) \sqrt{\sum_{i\in[N]} |b_i|^2},$$

ii) $$\left\| \sum_{i,j\in[N]} a_{i,j} Y_i Z_j \right\|_p \leq \left(A_p^2\mu_p^2 + 2A_p\mu_p\sqrt{N}|t| + N|t|^2 \right) \sqrt{\sum_{i,j\in[N]} |a_{i,j}|^2},$$

iii) $$\left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j} Y_i Y_j \right\|_p \leq \left(4A_p^2\mu_p^2 + 2A_p\mu_p\sqrt{N}|t| + N|t|^2 \right) \sqrt{\sum_{\substack{i,j\in[N]\\ i\neq j}} |a_{i,j}|^2}.$$

where $A_p \in \mathbb{R}_+$ is the constant from Theorem 6.27, which depends only on p.

Proof. Surely, $(Y_i - t)_i$ and $(Z_i - t)_i$ are centered and uniformly $\|\cdot\|_p$-bounded by μ_p for all $p \geq 2$.
We first show that *iii)* holds, which is the most important statement for our purposes. For $p \geq 2$ we find:

$$\begin{aligned}
&\left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j} Y_i Y_j \right\|_p \\
&= \left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j}[(Y_i - t) + t][(Y_j - t) + t] \right\|_p \\
&= \left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j}[(Y_i - t)(Y_j - t) + t(Y_j - t) + t(Y_i - t) + t^2] \right\|_p \\
&\leq \left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j}(Y_i - t)(Y_j - t) \right\|_p + \left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j} t (Y_j - t) \right\|_p \\
&\quad + \left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j} t (Y_i - t) \right\|_p + \left\| \sum_{\substack{i,j\in[N]\\ i\neq j}} a_{i,j} t^2 \right\|_p \\
&=: T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

We will now proceed to analyze the four terms separately. Note that in general, $T_2 \neq T_3$. Their bounds can be derived in the same manner, though. To bound T_1, we have by Lemma 6.30, that

$$T_1 \leq 4A_p^2\mu_p^2 \sqrt{\sum_{\substack{i,j\in[N]\\ i\neq j}} |a_{i,j}|^2}.$$

For T_2 we obtain through Lemma 6.28 that

$$\begin{aligned}
T_2 &= |t| \left\| \sum_{j\in[N]} \left(\sum_{i\in[N]\setminus\{j\}} a_{i,j} \right) (Y_j - t) \right\|_p \\
&\leq |t| A_p \mu_p \sqrt{\sum_{j\in[N]} \left| \sum_{i\in[N]\setminus\{j\}} a_{i,j} \right|^2} \leq \sqrt{N} |t| A_p \mu_p \sqrt{\sum_{\substack{i,j\in[N]\\ i\neq j}} |a_{i,j}|^2},
\end{aligned}$$

where we used that for any $j \in [N]$ we find

$$\left| \sum_{i \in [N] \setminus \{j\}} a_{i,j} \right|^2 \leq N \cdot \sum_{i \in [N] \setminus \{j\}} |a_{i,j}|^2$$

by the Cauchy-Schwarz inequality (and $N - 1 \leq N$). The bound we just derived for T_2 analogously holds for T_3, thus

$$T_2 + T_3 \leq 2\sqrt{N}|t|A_p\mu_p \sqrt{\sum_{\substack{i,j \in [N] \\ i \neq j}} |a_{i,j}|^2}.$$

Lastly, using the Cauchy-Schwarz inequality again, we obtain

$$T_4 = \left| \sum_{\substack{i,j \in [N] \\ i \neq j}} a_{i,j} t^2 \right| = |t|^2 \left| \sum_{\substack{i,j \in [N] \\ i \neq j}} a_{i,j} \right| \leq |t|^2 \sqrt{\sum_{\substack{i,j \in [N] \\ i \neq j}} |a_{i,j}|^2} \cdot \sqrt{N^2} = N|t|^2 \sqrt{\sum_{\substack{i,j \in [N] \\ i \neq j}} |a_{i,j}|^2}.$$

This shows that *iii*) holds. Now *ii*) is shown analogously to *iii*), with the difference that sums over i and j are always over $[N]$ without further restrictions such as $i \neq j$. In addition, instead of using Lemma 6.30 to bound T_1, we then use Lemma 6.29 (where constants are smaller, thus we can replace $4A_p^2\mu_p^2$ by $A_p^2\mu_p^2$).

To show that *i*) holds, we calculate for $p \geq 2$:

$$\begin{aligned}
\left\| \sum_{i \in [N]} b_i Y_i \right\|_p &= \left\| \sum_{i \in [N]} b_i((Y_i - t) + t) \right\|_p \\
&\leq \left\| \sum_{i \in [N]} b_i(Y_i - t) \right\|_p + \left\| \sum_{i \in [N]} b_i t \right\|_p \\
&\leq A_p\mu_p \sqrt{\sum_{i \in [N]} |b_i|^2} + |t| \left| \sum_{i \in [N]} b_i \right| \\
&\leq (A_p\mu_p + |t|\sqrt{N}) \sqrt{\sum_{i \in [N]} |b_i|^2},
\end{aligned}$$

where in the third step we used Lemma 6.28, and in the fourth step we used the Cauchy-Schwarz inequality. $\square$

We proceed to show the main large deviations result in relation to the stochastic order relation $\prec$. Together with Theorem 6.31, this is the key contribution in order to obtain the local law for Curie-Weiss type ensembles. We show more than we need for this thesis. In particular, statement *ii*) of the following theorem is needed to obtain the stronger local law as in [9]. Work on this generalization is ongoing as this thesis is handed in.

Theorem 6.32. *Let for all $N \in \mathbb{N}$, Y and W be N-dependent objects ($Y = Y^{(N)}, W = W^{(N)}$) that satisfy the following for all $N \in \mathbb{N}$:*

- *$W = W^{(N)}$ is a finite index set.*
- *$Y_W = (Y_i)_{i \in W} = (Y_i^{(N)})_{i \in W^{(N)}} = Y_{W^{(N)}}^{(N)}$ is a tuple of random variables of de-Finetti type with mixture μ_N satisfying the moment decay condition* (6.3) *on page 148.*

Further, denote for all subsets $K \subseteq W$ by $\mathcal{F}_W(\{\pm 1\}^K)$ the set of tuples $C = (C_i)_{i \in W}$, where for each $i \in W$, $C_i : \{\pm 1\}^K \to \mathbb{C}$ is a complex-valued function. Analogously, define for all subsets $K \subseteq W$ by $\mathcal{F}_{W \times W}(\{\pm 1\}^K)$ the set of tuples $C = (C_{i,j})_{i,j \in W}$, where for all $i, j \in W$, $C_{i,j} : \{\pm 1\}^K \to \mathbb{C}$ is a complex-valued function. Then we obtain the following large deviation bounds:

i) $\sum\limits_{i \in I} B_i[Y_K] Y_i \prec \sqrt{\sum\limits_{i \in I} |B_i[Y_K]|^2}$*, uniformly over all pairwise disjoint subsets $I, K \subseteq W$ with $\# I \leq N$, and $B \in \mathcal{F}_W(\{\pm 1\}^K)$.*

ii) $\sum\limits_{i \in I, j \in J} Y_i A_{i,j}[Y_K] Y_j \prec \sqrt{\sum\limits_{i \in I, j \in J} |A_{i,j}[Y_K]|^2}$*, uniformly over all pairwise disjoint subsets $I, J, K \subseteq W$ with $\# I = \# J \leq N$, and $A \in \mathcal{F}_{W \times W}(\{\pm 1\}^K)$.*

iii) $\sum\limits_{i,j \in I, i \neq j} Y_i A_{i,j}[Y_K] Y_j \prec \sqrt{\sum\limits_{i,j \in I, i \neq j} |A_{i,j}[Y_K]|^2}$*, uniformly over all pairwise disjoint subsets $I, K \subseteq W$ with $\# I \leq N$, and $A \in \mathcal{F}_{W \times W}(\{\pm 1\}^K)$.*

Proof. We prove *iii*) first: Let $\epsilon, D > 0$ be arbitrary and choose $p \in \mathbb{N}$ with $p \geq 2$ so large that $p\epsilon > D$. Now, we pick an $N \in \mathbb{N}$, then choose pairwise disjoint subsets $I, K \subseteq W^{(N)}$ with $\# I \leq N$ and $A \in \mathcal{F}_{W \times W}(\{\pm 1\}^K)$ arbitrarily. To avoid division by zero, we define the set:

$$\mathcal{A}_3 := \left\{ y_K \in \{\pm 1\}^K \mid \sum_{i,j \in I, i \neq j} |A_{i,j}[y_K]|^2 > 0 \right\}.$$

Then we conduct the following calculation (explanations are found below the calculation; the sums over "$i \neq j$" are over all $i, j \in I$ with $i \neq j$):

$$
\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i\neq j} Y_i A_{i,j}[Y_K]Y_j\right| > N^{\epsilon}\left(\sum_{i\neq j}|A_{i,j}[Y_K]|^2\right)^{\frac{1}{2}}\right)\\
&= \mathbb{P}\left(\left|\frac{\sum_{i\neq j} Y_i A_{i,j}[Y_K]Y_j}{\left(\sum_{i\neq j}|A_{i,j}[Y_K]|^2\right)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_3}(Y_K) > N^{p\epsilon}\right)\\
&\leq \frac{1}{N^{p\epsilon}}\mathbb{E}\left|\frac{\sum_{i\neq j} Y_i A_{i,j}[Y_K]Y_j}{\left(\sum_{i\neq j}|A_{i,j}[Y_K]|^2\right)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_3}(Y_K)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\int_{\{\pm 1\}^{I\cup K}}\left|\frac{\sum_{i\neq j} y_i A_{i,j}[y_K]y_j}{\left(\sum_{i\neq j}|A_{i,j}[y_K]|^2\right)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_3}(y_K)\mathrm{d}P_t^{\otimes I\cup K}(y_{I\cup K})\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\int_{\{\pm 1\}^{K}}\int_{\{\pm 1\}^{I}}\left|\frac{\sum_{i\neq j} y_i A_{i,j}[y_K]y_j}{\left(\sum_{i\neq j}|A_{i,j}[y_K]|^2\right)^{\frac{1}{2}}}\right|^p \mathrm{d}P_t^{\otimes I}(y_I)\mathbb{1}_{\mathcal{A}_3}(y_K)\mathrm{d}P_t^{\otimes K}(y_K)\mathrm{d}\mu_N(t)\\
&\leq \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\int_{\{\pm 1\}^{K}}\left[4A_p^2\mu_p^2 + 2A_p\mu_p\sqrt{N}|t| + N|t|^2\right]^p \mathrm{d}P_t^{\otimes K}(y_K)\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\left[C_1 + (C_2\sqrt{N} + N|t|)|t|\right]^p \mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\sum_{l=0}^{p}\binom{p}{l}C_1^{p-l}(C_2\sqrt{N} + N|t|)^l|t|^l\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\sum_{l=0}^{p}\binom{p}{l}C_1^{p-l}\left(\sum_{k=0}^{l}\binom{l}{k}(C_2\sqrt{N})^{l-k}(N|t|)^k\right)|t|^l\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\sum_{l=0}^{p}\sum_{k=0}^{l}\binom{p}{l}\binom{l}{k}C_1^{p-l}C_2^{l-k}N^{\frac{l+k}{2}}|t|^{l+k}\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\sum_{k=0}^{l}\binom{p}{l}\binom{l}{k}C_1^{p-l}C_2^{l-k}N^{\frac{l+k}{2}}\int_{[-1,1]}|t|^{l+k}\mathrm{d}\mu_N(t)\\
&\leq \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\sum_{k=0}^{l}\binom{p}{l}\binom{l}{k}C_1^{p-l}C_2^{l-k}N^{\frac{l+k}{2}}\sqrt{\int_{[-1,1]}|t|^{2(l+k)}\mathrm{d}\mu_N(t)}\\
&\leq \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\sum_{k=0}^{l}\binom{p}{l}\binom{l}{k}C_1^{p-l}C_2^{l-k}N^{\frac{l+k}{2}}\frac{\sqrt{K_{2(l+k)}}}{N^{\frac{l+k}{2}}}\\
&= \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\sum_{k=0}^{l}\binom{p}{l}\binom{l}{k}(4A_p^2\mu_p^2)^{p-l}(2A_p\mu_p)^{l-k}\sqrt{K_{2(l+k)}}\\
&\leq \frac{1}{N^{D}}\cdot\mathrm{const}(p(\epsilon, D)),
\end{aligned}
$$

where the first step follows from the fact that for

$$\left|\sum_{i\neq j} Y_i A_{i,j}[Y_K] Y_j\right| > N^{\epsilon}\left(\sum_{i\neq j} |A_{i,j}[Y_K]|^2\right)^{\frac{1}{2}}$$

to hold, not all $A_{i,j}[Y_K]$ may vanish, in the third step we used Lemma 6.3, in the fourth step Fubini, in the fifth step we used part *iii*) of Theorem 6.31 (notice that the ± 1-valued coordinates $(y_i)_{i\in I}$ are independent under $P_t^{\otimes I}$ and have expectation $t \in [-1,1]$, thus $(\int_{\{\pm 1\}} |y_i - t|^p \mathrm{d}P_t(y_i))^{1/p} \leq 2$, which makes Theorem 6.31 applicable with $\mu_p = 2$ for any $t \in [-1,1]$. Further, $\#I \leq N$), in the sixth step we set temporarily for the duration of above calculation $C_1 := 4A_p^2\mu_p^2$ and $C_2 := 2A_p\mu_p$, in the seventh and eighth step we used the binomial theorem, in the eleventh step Cauchy-Schwarz, and in the twelfth step we used the moment decay property (6.3). Lastly,

$$\mathrm{const}(p(\epsilon, D)) := \sum_{l=0}^{p}\sum_{k=0}^{l}\binom{p}{l}\binom{l}{k}(4A_p^2\mu_p^2)^{p-l}(2A_p\mu_p)^{l-k}\sqrt{K_{2(l+k)}}$$

denotes a constant which depends only on p, which in turn depends only on the choices of ϵ and D, as is obvious in the beginning of the proof. In particular, this constant does not depend on the choice of $N \in \mathbb{N}$, the sets I and K or the function tuple A. This shows *iii*).

To show *ii*), we can proceed analogously to the proof of part *iii*), with the following minor modifications: Instead of $\mathcal{A}_3$, we use the set

$$\mathcal{A}_2 := \left\{ y_K \in \{\pm 1\}^K \mid \sum_{i\in I, j\in J} |A_{i,j}[y_K]|^2 > 0 \right\}.$$

The summation in the beginning is over $i \in I$, $j \in J$ instead of $i, j \in I$, $i \neq j$. This leads to the term $P_t^{\otimes I\cup J\cup K}(y_{I\cup J\cup K})$ in the third step and the term $P_t^{\otimes I\cup J}(y_{I\cup J})P_t^{\otimes K}(y_K)$ in the fourth step. In the fifth step, we use part *ii*) of Theorem 6.31 instead of part *iii*), giving slightly different constants.

It is left to show part *i*). Let again $\epsilon, D > 0$ be arbitrary and choose $p \in \mathbb{N}$ with $p \geq 2$ so large that $p\epsilon > D$. Now pick $N \in \mathbb{N}$ arbitrarily, then pairwise disjoint subsets $I, K \subseteq W^{(N)}$ with $\#I \leq N$ and $B \in \mathcal{F}_W(\{\pm 1\}^K)$ arbitrarily. We define the set

$$\mathcal{A}_1 := \left\{ y_K \in \{\pm 1\}^K \mid \sum_{i\in I} |B_i[y_K]|^2 > 0 \right\}.$$

Then we calculate (where again, explanations are found below the calculation, and the summation over i means summing over $i \in I$):

Now

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_i B_i[Y_K]Y_i\right| > N^{\epsilon}\left(\sum_i |B_i[Y_K]|^2\right)^{\frac{1}{2}}\right)\\
&= \mathbb{P}\left(\left|\frac{\sum_i B_i[Y_K]Y_i}{(\sum_i |B_i[Y_K]|^2)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_1}(Y_K) > N^{p\epsilon}\right)\\
&\leq \frac{1}{N^{p\epsilon}}\mathbb{E}\left|\frac{\sum_i B_i[Y_K]Y_i}{(\sum_i |B_i[Y_K]|^2)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_1}(Y_K)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\int_{\{\pm 1\}^{I\cup K}}\left|\frac{\sum_i B_i[y_K]y_i}{(\sum_i |B_i[y_K]|^2)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_1}(y_K)\mathrm{d}P_t^{\otimes I\cup K}(y_{I\cup K})\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\int_{\{\pm 1\}^{K}}\int_{\{\pm 1\}^{I}}\left|\frac{\sum_i B_i[y_K]y_i}{(\sum_i |B_i[y_K]|^2)^{\frac{1}{2}}}\right|^p \mathrm{d}P_t^{\otimes I}(y_I)\mathbb{1}_{\mathcal{A}_1}(y_K)\mathrm{d}P_t^{\otimes K}(y_K)\mathrm{d}\mu_N(t)\\
&\leq \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\int_{\{\pm 1\}^{K}}\left[A_p\mu_p + \sqrt{N}|t|\right]^p \mathrm{d}P_t^{\otimes K}(y_K)\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\int_{[-1,1]}\sum_{l=0}^{p}\binom{p}{l}(A_p\mu_p)^{p-l}(\sqrt{N}|t|)^l\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\binom{p}{l}(A_p\mu_p)^{p-l}N^{\frac{l}{2}}\int_{[-1,1]}|t|^l\mathrm{d}\mu_N(t)\\
&= \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\binom{p}{l}(A_p\mu_p)^{p-l}N^{\frac{l}{2}}\sqrt{\int_{[-1,1]}|t|^{2l}\mathrm{d}\mu_N(t)}\\
&\leq \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\binom{p}{l}(A_p\mu_p)^{p-l}N^{\frac{l}{2}}\frac{\sqrt{K_{2l}}}{N^{\frac{l}{2}}}\\
&\leq \frac{1}{N^{p\epsilon}}\sum_{l=0}^{p}\binom{p}{l}(A_p\mu_p)^{p-l}\sqrt{K_{2l}}\\
&\leq \frac{1}{N^{D}}\mathrm{const}'(p(\epsilon, D)),
\end{aligned}$$

where in the third step we used Lemma 6.3, in the fourth step Fubini, in the fifth step we used part i) of Theorem 6.31 (again, we notice that the ± 1-valued coordinates $(y_i)_{i\in I}$ are independent under $P_t^{\otimes I}$ and have expectation $t \in [-1,1]$, thus for any such t, $(\int_{\{\pm 1\}} |y_i - t|^p \mathrm{d}P_t(y_i))^{1/p} \leq 2$, which makes Theorem 6.31 applicable with $\mu_p = 2$ and any $t \in [-1,1]$. Further, $\#I \leq N$), in the sixth step we used the binomial theorem, in the eighth step Cauchy-Schwarz, and in the ninth step we used the moment decay property (6.3). Lastly,

$$\mathrm{const}'(p(\epsilon, D)) := \sum_{l=0}^{p}\binom{p}{l}(A_p\mu_p)^{p-l}\sqrt{K_{2l}}$$

denotes a constant which depends only on p, which in turn depends only on the choices of ϵ and D, as is obvious in the beginning of the proof of i). In particular, this constant does not depend on the choice of $N \in \mathbb{N}$, the sets I and K or the function tuple B. This shows i). $\square$

Remark 6.33. We would like to portray the calculations in Theorem 6.32 in a different light, which might also be used to expand our ideas to other situations, but at least will make these calculations very intuitive. First of all, it is clear that $\prec$-statements pertain only to the *distributions* of the random variables involved and have nothing to do with the underlying probability space on which the random variables are defined. Therefore, one can construct the probability space in a favourable way that reveals probabilistic structure. For example, let μ be a probability measure on $[-1,1]$, W be a finite non-empty index set. Let $P_t^{\otimes W} := \otimes_{i\in W} P_t$ be the stochastic kernel from $([-1,1], \mathcal{B}([-1,1]))$ to $\left(\{\pm 1\}^W, \mathcal{P}\left(\{\pm 1\}^W\right)\right)$ as in Definition 6.1, where the notation for the σ-algebras should be clear. Then define the probability space $\left([-1,1] \times \{\pm 1\}^W, \mu(\mathrm{d}t) \otimes P_t^{\otimes W}\right)$, where $\mu(\mathrm{d}t) \otimes P_t^{\otimes W} =: \mathbb{P}$ is the product the probability measure μ and the kernel $P_t^{\otimes W}$. On this space, define the random variable M, which is just the projection of the first coordinate to $[-1,1]$, thus μ-distributed, and $Y_W = (Y_i)_{i\in W}$, which is the projection of the second coordinate to $\{\pm 1\}^I$ and thus $\mu(\mathrm{d}t) \circ P_t^{\otimes W}$-distributed (composition of probability measure and kernel), which exactly means (6.1). Therefore, we have effectively constructed random variables of de-Finetti type with mixture μ. But in addition, we have also generated a mixing variable M. The clou is that Y_I are conditionally independent given M. And this was actually the starting point in the beginning of our investigations for Theorem 6.32. In the setting of the theorem, given pairwise disjoint subsets $K, I \subseteq W$, we repeat the calculation with only slightly different – yet more intuitive – notation:

$$
\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i\neq j} Y_i A_{i,j}[Y_K] Y_j\right| > N^{\epsilon}\left(\sum_{i\neq j}|A_{i,j}[Y_K]|^2\right)^{\frac{1}{2}}\right) \\
&\leq \frac{1}{N^{p\epsilon}} \mathbb{E}\left|\frac{\sum_{i\neq j} Y_i A_{i,j}[Y_K] Y_j}{\left(\sum_{i\neq j}|A_{i,j}[Y_K]|^2\right)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_3}(Y_K) \\
&= \frac{1}{N^{p\epsilon}} \int_{[-1,1]} \int_{\{\pm 1\}^{I\cup K}} \left|\frac{\sum_{i\neq j} y_i A_{i,j}[y_K] y_j}{\left(\sum_{i\neq j}|A_{i,j}[y_K]|^2\right)^{\frac{1}{2}}}\right|^p \mathbb{1}_{\mathcal{A}_3}(y_K) \mathrm{d}\mathbb{P}^{(Y_I,Y_K)|M=t}(y_I, y_K) \mathrm{d}\mathbb{P}^M(t) \\
&= \frac{1}{N^{p\epsilon}} \int_{[-1,1]} \int_{\{\pm 1\}^K} \int_{\{\pm 1\}^I} \left|\frac{\sum_{i\neq j} y_i A_{i,j}[y_K] y_j}{\left(\sum_{i\neq j}|A_{i,j}[y_K]|^2\right)^{\frac{1}{2}}}\right|^p \mathrm{d}\mathbb{P}^{Y_I|M=t}(y_I) \mathbb{1}_{\mathcal{A}_3}(y_K) \mathrm{d}\mathbb{P}^{Y_K|M=t}(y_K) \mathrm{d}\mathbb{P}^M(t) \\
&\leq \dots
\end{aligned}
$$

In the third step of above calculation we used conditional independence.

Theorem 6.32 is very powerful. It is an analog to Theorem 3.6 in [9] which is a key ingredient to the full local law. Statement *iii*) of our Theorem 6.32 is all that is needed for the proof of the weak local law, whereas for the proof of the stronger local law as in [9] (which is ongoing work), we will also need statement *ii*). For our purposes, statement *iii*) immediately yields the following corollary:

Corollary 6.34. *In the setting of Lemma 6.23, we find*

$$\sum_{i\neq j\in[N-1]}(\sqrt{N}x_k(i))(X_N^{(k)}-z)^{-1}(i,j)(\sqrt{N}x_k(j))\prec\left(\sum_{i\neq j\in[N-1]}|(X_N^{(k)}-z)^{-1}(i,j)|^2\right)^{\frac{1}{2}},\quad {\scriptstyle z\in\mathbb{C}_+,\; k\in[N]}$$

Proof. Note that for all $N\in\mathbb{N}$ and $k\in[N]$, the vector $\sqrt{N}x_k$ is $(N-1)$-dimensional with distinct entries out of the family $(\sqrt{N}X_N(i,j))_{1\leq i\leq j\leq N}$, which is of de-Finetti type with mixture μ_N satisfying the moment decay condition (6.3) on page 148. Further, for any $z\in\mathbb{C}_+$ and $i\neq j\in[N-1]$ we have that $(X_N^{(k)}-z)^{-1}(i,j)$ is a complex function of variables in $(\sqrt{N}X_N(i,j))_{1\leq i\leq j\leq N}$ disjoint from those in $\sqrt{N}x_k$. Therefore, the statement follows with Theorem 6.32. □

We would like to remind the reader that the eventual goal is to be able to analyze the magnitude of the error terms Z_k as laid out in the beginning of this section. We have already come very far, but need a last ingredient before we can finally turn to Theorem 6.36, which is the main stochastic large deviations estimate we are after. The missing ingredient is the Ward identity:

Lemma 6.35 (Ward Identity)**.** *Let $N\in\mathbb{N}$, H be an Hermitian $N\times N$- matrix, $z=E+i\eta$ with $E\in\mathbb{R}$ and $\eta>0$ and G be the resolvent of H at z, that is, $G=(H-z)^{-1}$. Then for any $i\in[N]$:*

$$\sum_{j\in[N]}|G_{ij}|^2=\frac{\operatorname{Im}G_{ii}}{\eta}.$$

Proof. Since H is Hermitian, it has real eigenvalues $\lambda_1,\dots,\lambda_N$ and we find corresponding orthonormal eigenvectors $u_1,\dots,u_N$ so that

$$H=\sum_{j\in[N]}\lambda_j u_j u_j^*.$$

By simple calculation, we find

$$(GG^*)_{ii}=\sum_{j\in[N]}G_{ij}G^*_{ji}=\sum_{j\in[N]}G_{ij}\overline{G_{ij}}=\sum_{j\in[N]}|G_{ij}|^2.$$

By the spectral theorem, we have

$$(GG^*)_{ii}=\left(\sum_{j\in[N]}\frac{1}{|\lambda_j-z|^2}u_ju_j^*\right)_{ii}=\sum_{j\in[N]}\frac{1}{|\lambda_j-z|^2}|u_j(i)|^2.$$

This clearly shows

$$\sum_{j\in[N]} |G_{ij}|^2 = \sum_{j\in[N]} \frac{1}{|\lambda_j - z|^2}|u_j(i)|^2.$$

It remains to see that this coincides with the right hand side of the Ward identity. We calculate

$$\begin{aligned}
\frac{1}{\eta}\operatorname{Im} G_{ii} &= \frac{1}{\eta}\operatorname{Im}\left(\sum_{j\in[N]} \frac{1}{\lambda_j - z} u_j u_j^*\right)_{ii} \\
&= \frac{1}{\eta}\operatorname{Im}\sum_{j\in[N]} \frac{1}{\lambda_j - z}|u_j(i)|^2 \\
&= \frac{1}{\eta}\sum_{j\in[N]} \frac{\eta}{(\lambda_j - E)^2 + \eta^2}|u_j(i)|^2 \\
&= \sum_{j\in[N]} \frac{1}{(\lambda_j - E)^2 + \eta^2}|u_j(i)|^2 \\
&= \sum_{j\in[N]} \frac{1}{|\lambda_j - z|^2}|u_j(i)|^2.
\end{aligned}$$

□

Now, we are equipped to show the main stochastic large deviations estimate of this section:

Theorem 6.36. *We find*

$$\max_{k\in[N]} |Z_k(z)| \prec \frac{1}{\sqrt{N\eta}}\sqrt{\operatorname{Im} s_N(z)} + \frac{1}{N\eta}, \quad z \in \mathbb{C}_+.$$

Proof. By Lemma 6.7, it suffices to show the statement for a fixed k and reveal that the constants $C_{\epsilon,D}$ do not depend on k. We know that

$$Z_k(z) = \sum_{i\neq j} x_k(i)(X_N^{(k)} - z)^{-1}(i,j)x_k(j) = \frac{1}{N}\sum_{i\neq j}(\sqrt{N}x_k(i))(X_N^{(k)} - z)^{-1}(i,j)(\sqrt{N}x_k(j)).$$

Therefore, by Corollary 6.34 and Lemma 6.7 it follows for all $k \in [N]$, that

$$|Z_k(z)|^2 \prec \frac{1}{N^2}\sum_{i\neq j}\left|(X_N^{(k)} - z)^{-1}(i,j)\right|^2, \qquad z \in \mathbb{C}_+,$$

where the constants $C_{\epsilon,D}$ do not depend on $k \in [N]$. But now

$$\begin{aligned}
&\frac{1}{N^2}\sum_{i\neq j}\left|(X_N^{(k)}-z)^{-1}(i,j)\right|^2\\
&\leq \frac{1}{N^2}\sum_{i,j}\left|(X_N^{(k)}-z)^{-1}(i,j)\right|^2\\
&= \frac{1}{N^2\eta}\sum_{i}\operatorname{Im}(X_N^{(k)}-z)^{-1}(i,i)\\
&= \frac{1}{N^2\eta}[\operatorname{Im}\operatorname{tr}(X_N-z)^{-1}+\operatorname{Im}\operatorname{tr}(X_N^{(k)}-z)^{-1}-\operatorname{Im}\operatorname{tr}(X_N-z)^{-1})]\\
&\leq \frac{1}{N\eta}\operatorname{Im}s_N(z)+\frac{1}{N^2\eta}\left|\operatorname{tr}(X_N^{(k)}-z)^{-1}-\operatorname{tr}(X_N-z)^{-1}\right|\\
&\leq \frac{1}{N\eta}\operatorname{Im}s_N(z)+\frac{1}{N^2\eta^2},
\end{aligned}$$

where in the second step we used the Ward identity (Lemma 6.35) and in the last step we used Corollary 5.25. We conclude

$$|Z_k(z)|\prec\frac{1}{\sqrt{N\eta}}\sqrt{\operatorname{Im}s_N(z)}+\frac{1}{N\eta},\qquad z\in\mathbb{C}_+,$$

by taking the square root on both sides, and where the constants $C_{\epsilon,D}$ are independent of $k\in[N]$. This is what we wanted to show. □

6.4.3 Step 3: The Initial Estimate

In this section we prove a preliminary version of the weak local law, namely:

Theorem 6.37. *In the situation of Theorem 6.10, we find*

$$|s_N(z)-s(z)|\prec\min\left(\frac{1}{(N\eta)^{\frac14}},\frac{1}{\sqrt{N\eta\kappa}}\right),\quad z\in\mathcal{D}_I.$$

Proof. <u>Step 1</u> We show that uniformly on $\mathcal{D}_I$,

$$\frac{\max_k|\Omega_k|}{|z+s_N(z)|}\prec\frac{1}{\sqrt{N}}\qquad\text{and}\qquad\frac{\max_k|\Omega_k|}{|z+s_N(z)|^2}\prec\frac{1}{\sqrt{N}}.$$

To see this, note that since $\eta\geq 1$ on $\mathcal{D}_I$, we have on $\mathcal{D}_I$:

$$0<\operatorname{Im}s_N(z)\leq\frac{1}{\eta}\leq 1\quad\text{and}\quad\frac{1}{N\eta}\leq\frac{1}{N}.$$

Hence, it follows with Theorem 6.36 uniformly over $z\in\mathcal{D}_I$:

$$\max_{k\in[N]}|Z_k|\prec\frac{1}{\sqrt{N\eta}}\sqrt{\operatorname{Im}s_N(z)}+\frac{1}{N\eta}\leq\frac{1}{\sqrt{N}}+\frac{1}{N}\prec\frac{2}{\sqrt{N}}\prec\frac{1}{\sqrt{N}},$$

where we used Lemma 6.7 here and will do so tacitly for the remainder of the proof. Since $|X_N(i,i)| = \frac{1}{\sqrt{N}}$, we find

$$\max_{k\in[N]} |\Omega_k| \leq \max_{k\in[N]} \left(|Z_k| + |X_N(k,k)| + O_1\left(\frac{1}{N\eta}\right)\right) \prec \frac{1}{\sqrt{N}} + \frac{1}{\sqrt{N}} + \frac{1}{N} \prec \frac{1}{\sqrt{N}}.$$

Since $\operatorname{Im}(z + s_N(z)) \geq \eta \geq 1$, we have $\frac{1}{|z+s_N(z)|} \leq 1$ and $\frac{1}{|z+s_N(z)|^2} \leq 1$, so that on $\mathcal{D}_I$:

$$\frac{\max_k |\Omega_k|}{|z + s_N(z)|} \prec \frac{1}{\sqrt{N}} \qquad \text{and} \qquad \frac{\max_k |\Omega_k|}{|z + s_N(z)|^2} \prec \frac{1}{\sqrt{N}}.$$

<u>Step 2</u> The second step consists of utilizing the deterministic root approximation lemma, Lemma 6.26, to show

$$\min\left\{|s_N(z) - s(z)|, \left|s_N(z) - \frac{1}{s(z)}\right|\right\} \prec \min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right), \quad z \in \mathcal{D}_I.$$

We know that uniformly over $z \in \mathcal{D}_I$,

$$\frac{\max_k |\Omega_k|}{|z + s_N(z)|} \prec \frac{1}{\sqrt{N}} \qquad \text{and} \qquad \frac{\max_k |\Omega_k|}{|z + s_N(z)|^2} \prec \frac{1}{\sqrt{N}}.$$

so that also

$$\max\left(\frac{\max_k |\Omega_k|}{|z + s_N(z)|}, \frac{\max_k |\Omega_k|}{|z + s_N(z)|^2}\right) \prec \frac{1}{\sqrt{N}}, \quad z \in \mathcal{D}_I.$$

Now let $\epsilon, D > 0$ be arbitrary, then w.l.o.g. $\epsilon \leq 1/4$ with Remark 6.6. Denote by A_N the set

$$A_N := \left\{\frac{\max_k |\Omega_k|}{|z + s_N(z)|} \leq \frac{N^\epsilon}{\sqrt{N}} \quad \text{and} \quad \frac{\max_k |\Omega_k|}{|z + s_N(z)|^2} \leq \frac{N^\epsilon}{\sqrt{N}}\right\}.$$

Then there is a $C_{\epsilon,D} > 0$ such that for all $N \in \mathbb{N} : \mathbb{P}(A_N) \geq 1 - \frac{C_{\epsilon,D}}{N^D}$.

For all $N \geq 16$, we find that $\frac{N^\epsilon}{\sqrt{N}} \leq \frac{1}{N^{\frac{1}{4}}} \leq \frac{1}{2}$, so that on A_N we have

$$\frac{\max_k |\Omega_k|}{|z + s_N(z)|} \leq \frac{N^\epsilon}{\sqrt{N}} \leq \frac{1}{2} \quad \text{and} \quad \frac{\max_k |\Omega_k|}{|z + s_N(z)|^2} \leq \frac{N^\epsilon}{\sqrt{N}} \leq \frac{1}{2}.$$

for all $N \geq 16$. Hence, we conclude via Lemma 6.26 that for all $N \geq 16$, we have on the set A_N that

$$\begin{aligned}\min\left(|s_N(z) - s(z)|, \left|s_N(z) - \frac{1}{s(z)}\right|\right\} &\leq C_{Det} \min\left(\frac{\max_k |\Omega_k|}{\sqrt{\kappa}|z + s_N(z)|^2}, \sqrt{\frac{\max_k |\Omega_k|}{|z + s_N(z)|^2}}\right) \\ &\leq C_{Det} \min\left(\frac{N^\epsilon}{\sqrt{\kappa N}}, \sqrt{\frac{N^\epsilon}{\sqrt{N}}}\right) \\ &\leq N^\epsilon C_{Det} \min\left(\frac{1}{\sqrt{\kappa N}}, \frac{1}{N^{\frac{1}{4}}}\right).\end{aligned}$$

It follows that

$$\min\left(|s_N(z)-s(z)|, \left|s_N(z)-\frac{1}{s(z)}\right|\right) \prec C_{Det}\min\left(\frac{1}{\sqrt{\kappa N}}, \frac{1}{N^{\frac{1}{4}}}\right), \quad z\in\mathcal{D}_I.$$

Therefore, with Lemma 6.7, we conclude that uniformly over $z\in\mathcal{D}_I$:

$$\min\left(|s_N(z)-s(z)|, \left|s_N(z)-\frac{1}{s(z)}\right|\right) \prec \frac{1}{10}\min\left(\frac{1}{\sqrt{\kappa N}}, \frac{1}{N^{\frac{1}{4}}}\right) \leq \min\left(\frac{1}{\sqrt{\kappa N\eta}}, \frac{1}{(N\eta)^{\frac{1}{4}}}\right),$$

since $\eta \leq 10$ over $\mathcal{D}_I$.
Step 3 In this last step, we wish to conclude from

$$\min\left(|s_N(z)-s(z)|, \left|s_N(z)-\frac{1}{s(z)}\right|\right) \prec \min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right), \quad z\in\mathcal{D}_I, \tag{6.15}$$

that actually

$$|s_N(z)-s(z)| \prec \min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right), \quad z\in\mathcal{D}_I.$$

To this end, note that for all $N\in\mathbb{N}$ and $z\in\mathcal{D}_I$,

$$\left|s_N(z)-\frac{1}{s(z)}\right| = |s_N(z)+s(z)+z| > \eta \geq 1 \geq N^{\frac{1}{4}}\min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right).$$

As a result, for $\epsilon>0$ with $\epsilon\leq\frac{1}{4}$ and $D>0$ arbitrary, we find for all $N\in\mathbb{N}$:

$$\begin{aligned}
&\mathbb{P}\left(|s_N(z)-s(z)| > N^{\epsilon}\min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right)\right)\\
&= \mathbb{P}\Bigg(|s_N(z)-s(z)| > N^{\epsilon}\min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right)\\
&\qquad\qquad \text{and} \quad \left|s_N(z)-\frac{1}{s(z)}\right| > N^{\epsilon}\min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right)\Bigg)\\
&= \mathbb{P}\left(\min\left(|s_N(z)-s(z)|, \left|s_N(z)-\frac{1}{s(z)}\right|\right) > N^{\epsilon}\min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right)\right) \leq \frac{C_{\epsilon,D}}{N^D}
\end{aligned}$$

for a suitable constant $C_{\epsilon,D}$ due to (6.15). This concludes the Initial Estimate. □

Before we continue, we would like to apply Theorem 6.12 and Lemma 6.14 to Theorem 6.37 to increase uniformity in the statement, which we will use in a later step. To this end we use the abbreviation

$$R_N(z) := \min\left(\frac{1}{(N\eta)^{\frac{1}{4}}}, \frac{1}{\sqrt{N\eta\kappa}}\right).$$

Theorem 6.38. *In the situation of Theorem 6.10, we find*

$$\sup_{z\in\mathcal{D}_I}\frac{|s_N(z)-s(z)|}{R_N(z)}\prec 1.$$

Proof. We know from Theorem 6.37 that

$$|s_N(z)-s(z)|\prec R_N(z),\quad z\in\mathcal{D}_I.$$

Lemma 6.14 yields that $|s_N(z)-s(z)|$ is $2N^2$-Lipschitz on $\mathcal{D}_I$, whereas $R_N(z)$ is $10N$-Lipschitz and lower bounded by $1/(10\sqrt{N})$ on D_I as in the proof of Theorem 6.15. The statement now follows with Theorem 6.12 with $L=4$, $C_2=2$, $d_2=2$, $C_3=10$, $d_3=1$, $C_4=10$ and $d_4=1/2$, $\mathcal{G}_N^4=\mathcal{D}_I^4$. □

6.4.4 Step 4: The Bootstrap Argument

As in Lemma 6.14, we set for all $z\in[-10,10]+i(0,10]$:

$$S_N(z):=\min\left\{|s_N(z)-s(z)|,\left|s_N(z)-\frac{1}{s(z)}\right|\right\}\quad\text{and}\quad R_N(z):=\min\left\{\frac{1}{\sqrt{N\eta\kappa}},\frac{1}{(N\eta)^{\frac{1}{4}}}\right\}.$$

From the initial estimate, we know that

$$S_N(z)\prec R_N(z),\quad z\in\mathcal{D}_I \tag{6.16}$$

and even that

$$|s_N(z)-s(z)|\prec R_N(z),\quad z\in\mathcal{D}_I. \tag{6.17}$$

In this part of the proof we wish to see that (6.16) actually holds uniformly over all $z\in\mathcal{D}_N(\gamma)$, even simultaneously. The strategy we will follow is to fix a real part $E\in[-10,10]$ and to show that

$$S_N(z(N,k))\prec R_N(z(N,k)),\quad k=0,1,\dots,m(N), \tag{6.18}$$

where for all $N\in\mathbb{N}$,

$$m(N):=\lfloor N^4-N^{3+\gamma}\rfloor$$

and

$$z(N,k):=z_E(N,k):=E+i(1-kN^{-4}),\quad k=0,1,\dots,m(N).$$

We will show that the choice of the constants $C_{\epsilon,D}$ in (6.18) does not depend on $E\in[-10,10]$, so that actually, we obtain

$$S_N(z)\prec R_N(z),\quad z\in\mathcal{D}_N^4(\gamma)\cap([-10,10]+i(0,1]),$$

and since the initial estimate will give us

$$S_N(z)\prec R_N(z),\quad z\in\mathcal{D}_N^4(\gamma)\cap([-10,10]+i[1,10]),$$

we will have

$$S_N(z) \prec R_N(z), z \in \mathcal{D}_N^4(\gamma),$$

thus by Theorem 6.12

$$\sup_{z\in\mathcal{D}_N(\gamma)} \frac{S_N(z)}{R_N(z)} \prec 1, \tag{6.19}$$

so in particular

$$S_N(z) \prec R_N(z), \quad z \in \mathcal{D}_N(\gamma).$$

Then later, in Step 5 of the proof, we will use (6.19) and a continuity argument to improve (6.17) to the statement

$$|s_N(z) - s(z)| \prec R_N(z), \quad z \in \mathcal{D}_N(\gamma).$$

But let us get started with the bootstrapping argument. We fix an $E \in [-10, 10]$ and define as above:

$$z(N,k) := E + i(1 - kN^{-4}), \quad k = 0, 1, \ldots, m(N).$$

The initial estimate tells us that

$$S_N(z(N,0)) \prec R_N(z(N,0)). \tag{6.20}$$

By reducing the imaginary part by very small increments, we wish to see that for N large enough (i.e. $N \geq N(\epsilon,\gamma)$), we will have

$$S_N(z(N,k)) \prec R_N(z(N,k)), \quad k = 0, 1, \ldots, m(N),$$

where the constants do not depend on the previously fixed E.

This is done by showing that each small decrease of the imaginary part will only forfeit a negligible amount of probability. Here, the large deviations estimate from Step 2 will play a big role. The next two theorems make up the heart of the bootstrapping argument. Theorem 6.40 will analyze how target quantities change when the imaginary part is decreased by one step, that is, by N^{-4}. Then, Theorem 6.41 will analyze how the high probability (certainty) which we initially obtained through (6.20), decays with each step. In order for these two theorems to work, we will need to fix certain constants a priori, which is done in the following remark.

Remark 6.39. For the next three theorems we fix a parameter $\gamma \in (0, 1)$ and define the following constants:

- We fix $c \geq 3$ such that $|s|, |\frac{1}{s}|, |\frac{1}{s^2}|$ are bounded by c on $[-10, 10] + i(0, 10]$. This choice is possible due to Theorem 5.16. In particular, S_N is then cN^2-Lipschitz on $\mathcal{D}_N$ by Lemma 6.14.
- Let $C := C_P$ denote the constant from the Proximity Theorem, Theorem 6.25.
- Let $C_1 := 8C(1 + c)^{5/2}$.

Further, for any $\epsilon > 0$ (and the previously fixed $\gamma \in (0,1)$) we choose $N(\epsilon,\gamma)$ so large, that for all $N \geq N(\epsilon,\gamma)$ the following statements hold:

1. $c\left(\frac{C_1}{N^{\frac{\epsilon}{4}}} + \frac{c}{N^2}\right) < 1$,
2. $0 < \frac{c}{1-c(\frac{C_1}{N^{\frac{\epsilon}{4}}}+\frac{c}{N^2})} \leq c+1$.
3. $4N^{-\frac{3\gamma}{20}}(1+c)^{\frac{5}{2}} \leq \frac{1}{2}$.

Note that instead of $N(\epsilon,\gamma)$ we could just write $N(\epsilon)$, since γ is a super-parameter in our model. However, we write $N(\epsilon,\gamma)$ to clarify to the reader that the choice also depends on γ. We will now conduct a one-step deviation analysis.

Theorem 6.40. *Let $E \in [-10,10]$ and $N \geq N(\epsilon,\gamma)$ be fixed. Let $\epsilon \in (0,\gamma/10)$ and $\eta \geq \frac{1}{N^{1-5\epsilon}}$ (for example, this holds if $\eta \geq \frac{1}{N^{1-\gamma}}$) be arbitrary so that $\eta - N^{-4} \geq N^{-1}$. Define $z := E + i\eta$ and $z_1 := z - iN^{-4}$ (then $z, z_1 \in \mathcal{D}_N$). Suppose it holds that*

$$S_N(z) \leq C_1 N^{\epsilon} R_N(z).$$

Then it follows:

i) $S_N(z_1) \leq C_1 N^{\epsilon} R_N(z) + \frac{c}{N^2} \leq C_1 \frac{1}{N^{\frac{\epsilon}{4}}} + \frac{c}{N^2}$.

ii) $|\mathrm{Im}\, s_N(z_1)| \leq 1 + c$.

iii) $\frac{1}{|z_1 + s_N(z_1)|} \leq 1 + c$.

Proof. We use that s_N and s are N^2-Lipschitz and S_N is cN^2-Lipschitz on $\mathcal{D}_N$, see Lemma 6.14. i)

$$\begin{aligned} S_N(z_1) &\leq |S_N(z_1) - S_N(z)| + S_N(z) \leq \frac{c}{N^2} + C_1 N^{\epsilon} R_N(z) \\ &\leq \frac{c}{N^2} + C_1 N^{\epsilon} \frac{1}{(N\eta)^{\frac{1}{4}}} \leq \frac{c}{N^2} + C_1 N^{\epsilon} \frac{1}{(N^{5\epsilon})^{\frac{1}{4}}} \\ &= \frac{c}{N^2} + C_1 \frac{1}{N^{\frac{\epsilon}{4}}}. \end{aligned}$$

ii) We have

$$|s_N(z_1)| \leq \underbrace{|s(z_1)|}_{\leq c} + |s_N(z_1) - s(z_1)|$$

and

$$|s_N(z_1)| \leq \underbrace{\left|\frac{1}{s(z_1)}\right|}_{\leq c} + \left|s_N(z_1) - \frac{1}{s(z_1)}\right|,$$

which entails with i) that

$$|\operatorname{Im} s_N(z_1)| \leq |s_N(z_1)| \leq c + S_N(z_1) \leq c + \frac{c}{N^2} + \frac{C_1}{N^{\frac{\epsilon}{4}}} \leq c+1,$$

since $c \geq 3$ and already $c\left(\frac{c}{N^2} + \frac{C_1}{N^{\frac{\epsilon}{4}}}\right) < 1$ per choice of $N(\epsilon, \gamma)$.

iii) We have by choice of c (and Theorem 5.16) that

$$\left|\frac{1}{z_1 + s(z_1)}\right| = |s(z_1)| \leq c \qquad \text{and} \qquad \frac{1}{\left|z_1 + \frac{1}{s(z_1)}\right|} = \frac{1}{|s(z_1)|} \leq c,$$

so

$$\begin{aligned} \left|\frac{1}{z_1 + s_N(z_1)}\right| &\leq \left|\frac{1}{z_1 + s(z_1)}\right| + \left|\frac{1}{z_1 + s_N(z_1)} - \frac{1}{z_1 + s(z_1)}\right| \\ &\leq c + \frac{|s(z_1) - s_N(z_1)|}{|z_1 + s_N(z_1)||z_1 + s(z_1)|} \\ &\leq c + c\frac{|s(z_1) - s_N(z_1)|}{|z_1 + s_N(z_1)|}. \end{aligned}$$

Analogously,

$$\begin{aligned} \left|\frac{1}{z_1 + s_N(z_1)}\right| &\leq \left|\frac{1}{z_1 + \frac{1}{s(z_1)}}\right| + \left|\frac{1}{z_1 + s_N(z_1)} - \frac{1}{z_1 + \frac{1}{s(z_1)}}\right| \\ &\leq c + c\frac{\left|\frac{1}{s(z_1)} - s_N(z_1)\right|}{|z_1 + s_N(z_1)|}. \end{aligned}$$

Therefore, combining our results:

$$\left|\frac{1}{z_1 + s_N(z_1)}\right| \leq c + c\frac{S_N(z_1)}{|z_1 + s_N(z_1)|}.$$

We arrive at

$$\frac{1}{|z_1 + s_N(z_1)|}(1 - cS_N(z_1)) \leq c.$$

Per choice of $N(\epsilon, \gamma)$ and by using the bound on $S_N(z_1)$ given by i), this allows the conclusion

$$\frac{1}{|z_1 + s_N(z_1)|} \leq \frac{c}{1 - cS_N(z_1)} \leq \frac{c}{1 - c(\frac{C_1}{N^{\frac{\epsilon}{4}}} + \frac{c}{N^2})} \leq c+1.$$

□

We now arrive at the heart of the bootstrap argument:

Theorem 6.41. *Let $E \in [-10, 10]$ be fixed and for all $N \in \mathbb{N}$ and $k = 0, 1, \ldots, m(N)$ let $z(N, k) := E + i(1 - kN^{-4})$ and*

$$A(N,k) := \{S_N(z(N,k)) \le C_1 N^\epsilon R_N(z(N,k))\}.$$

Then for any $\epsilon \in (0, \gamma/10)$ and $D > 0$, there exists a constant $C_{\epsilon,D} \ge 0$ (independent of E), such that for all $N \ge N(\gamma, \epsilon)$ and $k = 0, 1, \ldots, m(N)$:

$$\mathbb{P}(A(N,k)^c) \le \frac{(k+1)C_{\epsilon,D}}{N^D}.$$

Proof. Let $N \ge N(\epsilon, \gamma)$, $\epsilon \in (0, \gamma/10)$ and $D > 0$ be fixed throughout the proof. For better readability in formulas below, we define $z_k := z(N,k)$ and $\eta_k := \operatorname{Im} z(N,k)$ for $k = 0, 1, \ldots, m(N)$. For all $k \in \{1, \ldots, m(N)\}$, we set

$$D(N,k) := \left\{\forall\, i \in [N] : |Z_i(z_k)| \le N^\epsilon \frac{1}{\sqrt{N\eta_k}}\sqrt{\operatorname{Im} s_N(z_k)} + \frac{N^\epsilon}{N\eta_k}\right\}$$

and

$$\tilde{A}(N,k) = A(N,0) \cap \bigcap_{j=1}^{k} D(N,j),$$

where an empty intersection shall yield the ground set Ω, in particular, $\tilde{A}(N,0) = A(N,0)$. We proceed to show that it holds for all $k = 0, \ldots, m(N)$, that

$$\text{On } \tilde{A}(N,k): \; S_N(z_k) \le C_1 N^\epsilon R_N(z_k),$$

so that in particular,

$$\tilde{A}(N,k) \subseteq A(N,k).$$

We use induction in k and Theorem 6.40.

Induction basis: The statement is true for $k = 0$.
The statement is clear since $\tilde{A}(N,0) = A(N,0)$.

Induction hypothesis:
We assume the statement to be valid for all $k' \le k$, where $k \in \{0, 1, \ldots, m(N) - 1\}$ is fixed.

Induction step: $k \to k + 1$
We find $\eta_k, \eta_{k+1} \ge \frac{1}{N^{1-\gamma}} \ge \frac{1}{N^{1-5\epsilon}}$. On $\tilde{A}(N, k+1)$ it holds (since $\tilde{A}(N,k+1) \subseteq \tilde{A}(N,k)$):

$$S_N(z_k) \le C_1 N^\epsilon R_N(z_k),$$

where we used the induction hypothesis. Therefore, it follows with Theorem 6.40 that on $\tilde{A}(N, k+1)$,

- $|\operatorname{Im} s_N(z_{k+1})| \le 1 + c$
- $\frac{1}{|z_{k+1} + s_N(z_{k+1})|} \le 1 + c$

- $|\Omega_i(z_{k+1})| \leq |Z_i(z_{k+1})| + \frac{1}{\sqrt{N}} + \frac{1}{N\eta_{k+1}} \leq \frac{N^\epsilon}{\sqrt{N\eta_{k+1}}}\sqrt{\operatorname{Im} s_N(z_{k+1})} + \frac{N^\epsilon}{N\eta_{k+1}} + \frac{1}{\sqrt{N}} + \frac{1}{N\eta_{k+1}} \leq \frac{4N^\epsilon}{\sqrt{N\eta_{k+1}}}\sqrt{1+c}$
- $\frac{\max_i |\Omega_i(z_{k+1})|}{|z_{k+1}+s_N(z_{k+1})|}, \frac{\max_i |\Omega_i(z_{k+1})|}{|z_{k+1}+s_N(z_{k+1})|^2} \leq \frac{4N^\epsilon}{\sqrt{N\eta_{k+1}}}(1+c)^{\frac{5}{2}} \leq \frac{1}{2}$,

where the last inequality holds since $\eta_{k+1} \geq 1/N^{1-\gamma} \geq 1/N^{1-\gamma/2}$ and by choice of $N(\epsilon,\gamma)$ right before Theorem 6.40, we find

$$\frac{N^\epsilon}{\sqrt{N\eta_{k+1}}} \leq \frac{N^{\frac{\gamma}{10}}}{\sqrt{N^{\frac{\gamma}{2}}}} = N^{\frac{\gamma}{10}-\frac{\gamma}{4}} = N^{-\frac{3\gamma}{20}}.$$

Using the Deterministic Root Approximation (Lemma 6.26), we find on $\tilde{A}(N,k+1)$:

$$\begin{aligned} S_N(z_{k+1}) &\leq C_{Det} \min\left\{\frac{\max_i |\Omega_i|}{\sqrt{\kappa}|z_{k+1}+s_N(z_{k+1})|^2}, \sqrt{\frac{\max_i |\Omega_i|}{|z_{k+1}+s_N(z_{k+1})|^2}}\right\} \\ &\leq 2C \min\left\{\frac{4N^\epsilon(1+c)^{\frac{5}{2}}}{\sqrt{\kappa}\sqrt{N\eta_{k+1}}}, \sqrt{\frac{4N^\epsilon(1+c)^{\frac{5}{2}}}{\sqrt{N\eta_{k+1}}}}\right\} \\ &\leq 2 \cdot C \cdot 4 \cdot N^\epsilon \cdot (1+c)^{\frac{5}{2}} \min\left\{\frac{1}{\sqrt{\kappa N\eta_{k+1}}}, \frac{1}{(N\eta_{k+1})^{\frac{1}{4}}}\right\} \\ &\leq C_1 N^\epsilon R_N(z_{k+1}). \end{aligned}$$

In particular, we have $\tilde{A}(N,k+1) \subseteq A(N,k+1)$. This concludes the induction. We have shown:

$$\forall\, k = 0,\dots,m(N) : \tilde{A}(N,k) \subseteq A(N,k).$$

Since $N \geq N(\epsilon,\gamma)$ was arbitrary, we obtain

$$\forall\, N \geq N(\epsilon,\gamma) : \forall\, k = 0,\dots,m(N) : \tilde{A}(N,k) \subseteq A(N,k).$$

We still want to show that for all such N and k we have

$$\mathbb{P}\left(A(N,k)^c\right) \leq \frac{(k+1)C_{\epsilon,D}}{N^D},$$

where the constant $C_{\epsilon,D}$ does not depend on E, which we fixed in the statement of the theorem. It suffices to show the inequality for the sets $\tilde{A}(N,k)$.

But our large deviations estimate Theorem 6.36 yields a constant $C'_{\epsilon,D}$, such that *in particular* for all $N \geq N(\epsilon,D)$ and $k = 0,1,\dots,m(N)$:

$$\mathbb{P}(D(N,k)^c) \leq \frac{C'_{\epsilon,D}}{N^D},$$

whereas the initial estimate yields a constant $C^*_{\epsilon,D}$ such that *in particular* for these N and k we have

$$\mathbb{P}(A(N,0)^c) \leq \frac{C^*_{\epsilon,D}}{N^D}.$$

and the constants $C^*_{\epsilon,D}$ and $C'_{\epsilon,D}$ are independent of E. Since we also have for all such N and k that

$$\tilde{A}(N,k) = A(N,0) \cap \bigcap_{i=1}^{k} D(N,i),$$

we find, setting $C_{\epsilon,D} \coloneqq \max(C^*_{\epsilon,D},\ C'_{\epsilon,D})$ that

$$\begin{aligned}\mathbb{P}\left(A(N,k)^c\right) \leq \mathbb{P}(\tilde{A}(N,k)^c) &= \mathbb{P}\left(A(N,0)^c \cup \bigcup_{i=1}^{k} D(N,i)^c\right)\\ &\leq \mathbb{P}(A(N,0)^c) + \sum_{i=1}^{k} \mathbb{P}(D(N,i)^c)\\ &\leq \frac{C^*_{\epsilon,D}}{N^D} + \frac{kC'_{\epsilon,D}}{N^D}\\ &\leq (k+1)\frac{C_{\epsilon,D}}{N^D}.\end{aligned}$$

□

Theorem 6.41 now allows us to conclude the main theorem of this part of the proof:

Theorem 6.42. *In the setting above (in particular in the setting of Theorem 6.10), we find*

$$\sup_{z\in\mathcal{D}_N(\gamma)} \frac{S_N(z)}{R_N(z)} \prec 1,$$

thus in particular,

$$S_N(z) \prec R_N(z), \quad z \in \mathcal{D}_N(\gamma).$$

Proof. Since the constants $C_{\epsilon,D}$ in Theorem 6.41 did not depend on the choice of $E \in [-10,10]$, we find with $z_E(N,k) \coloneqq E + i(1 - kN^{-4})$ that for all $\epsilon \in (0, \gamma/10)$, $D > 0$ and $N \geq N(\epsilon, \gamma)$:

$$\sup_{E\in[-10,10]}\ \sup_{k\in\{0,1,\dots,m(N)\}} \mathbb{P}(S_N(z_E(N,k)) > N^\epsilon C_1 R_N(z_E(N,k))) \leq \frac{N^4 C_{\epsilon,D+4}}{N^{D+4}} \leq \frac{C_{\epsilon,D+4}}{N^D},$$

where we used that $m(N) \leq N^4 - 1$. We conclude with Remark 6.6 that

$$S_N(z) \prec R_N(z), \quad z \in D^4_N(\gamma) \cap ([-10,10] + i(0,1]).$$

Since the intial estimate especially yields

$$S_N(z) \prec R_N(z), \quad z \in D^4_N(\gamma) \cap ([-10,10] + i[1,10]),$$

we conclude (again with Remark 6.6) that

$$S_N(z) \prec R_N(z), \quad z \in D^4_N(\gamma).$$

With Theorem 6.12 we conclude

$$\sup_{z\in\mathcal{D}_N(\gamma)} \frac{S_N(z)}{R_N(z)} \prec 1,$$

since it was shown in the proof of Theorem 6.15 that R_N is lower bounded by $\frac{1}{10\sqrt{N}}$ on $\mathcal{D}_N$, and S_N and R_N are suitably Lipschitz on $\mathcal{D}_N$ by Lemma 6.14. □

6.4.5 Step 5: The Continuity Argument

By Theorem 6.42, we know that

$$\sup_{z\in\mathcal{D}_N(\gamma)} \frac{S_N(z)}{R_N(z)} \prec 1. \tag{6.21}$$

In this step, we wish to conclude that

$$\sup_{z\in\mathcal{D}_N(\gamma)} \frac{|s_N(z)-s(z)|}{R_N(z)} \prec 1, \tag{6.22}$$

which immediately yields the weak local law (in fact, this is even the statement of the Simultaneous Weak Local Law, Theorem 6.15).

By the uniform initial estimate, Theorem 6.38, we know

$$\sup_{z\in\mathcal{D}_I} \frac{|s_N(z)-s(z)|}{R_N(z)} \prec 1. \tag{6.23}$$

Due to knowledge of (6.23), it suffices to show

$$\sup_{z\in\mathcal{G}_N} \frac{|s_N(z)-s(z)|}{R_N(z)} \prec 1, \tag{6.24}$$

where

$$\mathcal{G}_N \coloneqq \left\{ z = E+i\eta \in \mathbb{C} \,|\, |E| \leq 10, \frac{1}{N^{1-\gamma}} \leq \eta \leq 1 \right\}.$$

Pick $\epsilon > 0$ and $D > 0$ arbitrarily, and define for all $N \in \mathbb{N}$ the set

$$\begin{aligned} H(N) \coloneqq \{\forall z \in \mathcal{G}_N : S_N(z) \leq N^\epsilon R_N(z)\} \\ \cap \{\forall\, E \in [-10, 10] : |s(E+i) - s_N(E+i)| \leq N^\epsilon R_N(E+i)\}\,. \end{aligned}$$

By (6.21) and (6.23), $H(N)$ has high probability, that is, there exists a $C_{\epsilon,D} > 0$, such that

$$\forall N \in \mathbb{N} : \mathbb{P}(H(N)^c) \leq \frac{C_{\epsilon,D}}{N^D}.$$

We will show:

$$\forall\, N \in \mathbb{N} : \text{On } H(N){:} \quad \forall\, z \in \mathcal{G}_N : |s_N(z) - s(z)| \leq (2C_s^2+1)N^\epsilon R_N(z), \tag{6.25}$$

where C_s is the constant from Theorem 5.16, such that for all $z \in [-10, 10] + i(0, 10]$ we have

$$\frac{1}{C_s}\sqrt{\kappa+\eta} \le \left| s(z) - \frac{1}{s(z)} \right| \le C_s\sqrt{\kappa+\eta}.$$

To show (6.25), we fix an $N \in \mathbb{N}$ and an $E \in [-10, 10]$ arbitrarily. Then we set

$$\forall\, \eta > 0 : z(\eta) := E + i\eta.$$

We know (6.25) already for $z(1) = E + i$ and we would like to see it be true for all $\eta \in [1/N^{1-\gamma}, 1]$. As we drop η from 1 down to $1/N^{1-\gamma}$, we analyze what happens to the validity of the inequality

$$2N^{\epsilon} R_N(z(\eta)) < \frac{1}{C_s}\sqrt{\kappa+\eta}. \tag{6.26}$$

We notice that both sides of (6.26) are continuous in η, and that through decreasing η, the l.h.s. increases and the r.h.s. decreases. Therefore, if (6.26) is violated for some η, then also for all $\eta' \le \eta$. It follows that we can partition the interval $[1/N^{1-\gamma}, 1]$ into two intervals, the lower part L and the upper part U, where on U, (6.26) holds, and on L, (6.26) is violated. L or U are allowed to be empty, indicating that (6.26) holds or is violated on the entire interval $[1/N^{1-\gamma}, 1]$.

Now let $\eta' \in [1/N^{1-\gamma}, 1]$ be arbitrary. Since $[1/N^{1-\gamma}, 1] = L \dot\cup U$, we consider two cases: <u>Case 1: $\eta' \in U$.</u> This indicates that U is not empty and that, in particular, $[\eta', 1] \subseteq U$. This means that for all $\eta \in [\eta', 1]$, on $H(N)$:

$$S_N(z(\eta)) \le N^{\epsilon} R_N(z(\eta)) \le 2N^{\epsilon} R_N(z(\eta)) < \frac{1}{C_s}\sqrt{\kappa+\eta} \le \left| s(z(\eta)) - \frac{1}{s(z(\eta))} \right|. \tag{6.27}$$

This is now a very important inequality. It tells us that over $H(N)$ and $[\eta', 1]$, $s_N(z(\cdot))$ remains closer either to $s(z(\cdot))$ or to $1/s(z(\cdot))$ and cannot change this alignment. In mathematical terms, we find on $H(N)$:

$$\begin{aligned} \text{Either:} \quad & \forall\, \eta \in [\eta', 1] : |s_N(z(\eta)) - s(z(\eta))| < \left| s_N(z(\eta)) - \frac{1}{s(z(\eta))} \right|, \\ \text{or:} \quad & \forall\, \eta \in [\eta', 1] : \left| s_N(z(\eta)) - \frac{1}{s(z(\eta))} \right| < |s_N(z(\eta)) - s(z(\eta))|. \end{aligned}$$

This is the case, since otherwise, due to continuity, there would be an $\eta_0 \in [\eta', 1]$, such that

$$|s_N(z(\eta_0)) - s(z(\eta_0))| = \left| s_N(z(\eta_0)) - \frac{1}{s(z(\eta_0))} \right|,$$

and then

$$\begin{aligned} \left| s(z(\eta_0)) - \frac{1}{s(z(\eta_0))} \right| &\le |s(z(\eta_0)) - s_N(z(\eta_0))| + \left| s_N(z(\eta_0)) - \frac{1}{s(z(\eta_0))} \right| \\ &= S_N(z(\eta_0)) + S_N(z(\eta_0)) \\ &\le 2N^{\epsilon} R_N(z(\eta_0)), \end{aligned}$$

which is a contradiction to (6.27). Now since we are on $H(N)$ and with (6.27), we know that

$$\begin{aligned}\left|s_N(z(1)) - \frac{1}{s(z(1))}\right| &= \left|s_N(z(1)) - s(z(1)) + s(z(1)) - \frac{1}{s(z(1))}\right| \\ &\geq \left|s(z(1)) - \frac{1}{s(z(1))}\right| - |s_N(z(1)) - s(z(1))| \\ &> 2N^\epsilon R_N(z(1)) - N^\epsilon R_N(z(1)) \\ &= N^\epsilon R_N(z(1)) \geq |s_N(z(1)) - s(z(1))|,\end{aligned}$$

and therefore, on $H(N)$, for all $\eta \in [\eta', 1]$,

$$|s_N(z(\eta)) - s(z(\eta))| = S_N(z(\eta)) \leq N^\epsilon R_N(z(\eta)) \leq (2C_s^2 + 1)N^\epsilon R_N(z(\eta)),$$

and especially, this inequality holds for η', what we wanted to show.
Case 2: $\eta' \in L$ This implies that η' violates (6.26), such that

$$2N^\epsilon R_N(z(\eta')) \geq \frac{1}{C_s}\sqrt{\kappa + \eta'}.$$

But this implies

$$2C_s^2 N^\epsilon R_N(z(\eta')) \geq C_s\sqrt{\kappa + \eta'} \geq \left|s(z(\eta')) - \frac{1}{s(z(\eta'))}\right|.$$

Then if $S_N(z(\eta')) = |s(z(\eta')) - s_N(z(\eta'))|$, we find on $H(N)$:

$$|s(z(\eta')) - s_N(z(\eta'))| \leq N^\epsilon R(z(\eta')) \leq (2C_s^2 + 1)N^\epsilon R_N(z(\eta')),$$

and if $S_N(z(\eta')) = |s_N(z(\eta')) - 1/s(z(\eta'))|$, we find on $H(N)$:

$$\begin{aligned}|s_N(z(\eta')) - s(z(\eta'))| &\leq \left|s_N(z(\eta')) - \frac{1}{s(z(\eta'))}\right| + \left|\frac{1}{s(z(\eta'))} - s(z(\eta'))\right| \\ &\leq N^\epsilon R_N(z(\eta')) + 2C_s^2 N^\epsilon R_N(z(\eta')) \\ &= (2C_s^2 + 1)N^\epsilon R_N(z(\eta')).\end{aligned}$$

Thus, we have finally shown (6.25) (since $E \in [-10, 10]$ was arbitrary), and with the inequality preceding (6.25), we find that

$$\sup_{z \in \mathcal{G}_N} \frac{|s_N(z) - s(z)|}{R_N(z)} \prec 2C_s^2 + 1,$$

from which (6.24) follows immediately with Lemma 6.7.

6.5 Ongoing and Future Research

The results obtained in this chapter can be greatly expanded. First of all, with help of our large deviation results from Theorem 6.32 and methods outlined in [9], it should be possible to expand our weak local law firstly to the matrix-valued weak local law as in Proposition 5.1 in [9], and secondly to the stronger local law as in Theorem 2.6 in [9]. The first expansion will allow conclusions about the behavior of eigenvectors of Curie-Weiss type ensembles, while the second expansion will derive optimal error bounds (up to factors of N^{ϵ} in the $\prec$-formalism) in the statement of the local law.

Further, combining our techniques with techniques outlined in [43], it is likely that one can broaden the concept of Curie-Weiss type ensembles to allow for Curie-Weiss entries with an inverse temperature $\beta > 1$ and still obtain local semicircle laws.

Lastly, one turn to other limit laws in random matrix theory, such as the Marchenko-Pastur law for covariance matrices, and analyze how the techniques developed in this thesis can be applied to those settings.

All of the points mentioned so far are ongoing research as this dissertation is handed in. On the other hand, future research endeavors that are planned lead into two directions: Firstly, we would like to investigate how to obtain results about local gap statistics of eigenvalues using the local laws that we derived. Secondly, by methods similar to our Theorem 6.31, we will investigate if more general types of random matrices with exchangeable entries can be analyzed towards local law results, and not just those with values in the set $\{\pm 1\}$.

List of Symbols

The following list contains not all, but most of the mathematical symbols which are used in this thesis. The page number references the location in the text where the symbol is defined and/or used in context. Note that some symbols have multiple meanings. However, in their relative context their use is unambiguous.

Symbol	Description	Page
$\emptyset$	empty set	28
$\#M$	number of elements in the set M	54
M^c	complement of M in contextual superset	22
$\mathbb{N}$	set of natural numbers, $\{1,2,3,\ldots\}$	16
$\mathbb{N}_0$	extended set of natural numbers, $\{0,1,2,\ldots\}$	41
$\mathbb{Z}$	set of integers	150
$\mathbb{Q}$	set of rational numbers	172
$\mathbb{R}$	set of real numbers	12
$\mathbb{R}_+$	set of non-negative real numbers	29
$\mathbb{C}$	set of complex numbers	117
$\mathbb{C}_+$	set $\{z \in \mathbb{C},\ \operatorname{Im}(z) > 0\}$	117
$\mathbb{K}$	field, $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$	36
$\operatorname{Mat}_n(\mathbb{K})$	set of $n \times n$ matrices over $\mathbb{K}$	36
$\operatorname{SMat}_n(\mathbb{K})$	set of self-adjoint $n \times n$ matrices over $\mathbb{K}$	36
λ_i^X	i-th smallest eigenvalue of self-adjoint matrix X	37
tr	trace functional on sets of matrices	36
$\operatorname{diag}(x_1,\ldots,x_n)$	diagonal matrix with entries $x_1,\ldots,x_n$	36
$\square_n$	alternative notation for $\{1,\ldots,n\} \times \{1,\ldots,n\}$	39
$[n]$	aternative notation for $\{1,\ldots,n\}$	63
$[n]^k$	k-fold cartesian product of $[n]$	72
$[n]_b^k$	b-relevant tuples in $[n]^k$	72
$\mathbb{1}_A$	indicator function of set A	12
$\mathcal{C}(\mathbb{R})$	the set of real-valued continuous functions on $\mathbb{R}$	15
$\mathcal{C}_b(\mathbb{R})$	set of all bounded $f \in \mathcal{C}(\mathbb{R})$	15
$\mathcal{C}_0(\mathbb{R})$	set of all $f \in \mathcal{C}(\mathbb{R})$ that vanish at ∞	15
$\mathcal{C}_c(\mathbb{R})$	set of all compactly supported $f \in \mathcal{C}(\mathbb{R})$	15
$\operatorname{supp}(f)$	support of the function f	15
$(\mathcal{X}, d)$	metric space with ground set X and metric d	31
$B_\delta(z)$	open δ-ball around z in contextual metric space	15
$\phi_{R,L}$	continuous cutoff function	16

Symbol	Description	Page
$\\|\cdot\\|_\infty$	supremum norm for real or complex valued functions	15
$\\|\cdot\\|_{\text{op}}$	operator norm for linear maps, including matrices	19
$\\|\cdot\\|$	euclidian norm for vectors in $\mathbb{K}^n$	36
$\\|\cdot\\|_p$	$\mathcal{L}_p(\mathbb{P})$ norm, $\\|Y\\|_p = (\mathbb{E}\|Y\|^p)^{1/p}$	181
$\mathcal{M}(\mathbb{R})$	space of measures on $(\mathbb{R}, \mathcal{B})$	18
$\mathcal{M}_f(\mathbb{R})$	space of finite measures on $(\mathbb{R}, \mathcal{B})$	18
$\mathcal{M}_{\leq 1}(\mathbb{R})$	space of sub-probability measures on $(\mathbb{R}, \mathcal{B})$	18
$\mathcal{M}_1(\mathbb{R})$	space of probability measures on $(\mathbb{R}, \mathcal{B})$	18
$\mathcal{B}$	Borel σ-algebra over $\mathbb{R}$	15
$\mathcal{B}_{\mathbb{K}}$	Borel σ-algebra over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$	15
$\mathcal{B}_s^{(n^2)}$	Borel σ-algebra over $\text{SMat}_n(\mathbb{K})$	37
$\langle \mu, f \rangle$	alternative notation for $\int f \, \mathrm{d}\mu$	18
$\left\langle \mu, x^k \right\rangle$	alternative notation for $\int x^k \, \mu(\mathrm{d}x)$	18
δ_x	Dirac measure in x	20
d_M	metric on the space of probability measures	24
$\mathbb{E}X$	expectation of the random variable X	29
$\mathbb{E}\mu$	expected measure of random measure μ	30
$\mathbb{V}X$	variance of the random variable X	71
$\text{Cov}(X, Y)$	covariance between the random variables X and Y	64
$\mathbb{P}$	underlying probability measure	27
$\mathbb{P}^X$	distribution of (or push-forward of $\mathbb{P}$ under) X	64
σ_n	empirical spectral distribution of $n \times n$ (random) matrix	38
σ	semicircle distribution on $(\mathbb{R}, \mathcal{B})$	39
f_σ	Lebesgue density function of σ	39
m_k^σ	k-th moment of σ	46
$\mathcal{C}_n$	n-th Catalan number	48
$(a_n)_n$	triangular scheme	54
$(b_n)_n$	bandwidth sequence	57
$(h_n)_n$	halfwidth sequence	98
$(a_n^b)_n$	periodic band matrices based on bandwidth b	57
X_n	$n \times n$ random matrix	40
X_n^P	periodic $n \times n$ random band matrix	98
X_n^{NP}	non-periodic $n \times n$ random band matrix	98
$X_n^{(k)}$	k-th principal minor of X_n	134
x_k	k-th column of X_n without k-th entry	134
$\sim$	asymptotic equivalence of sequences	58
Curie-Weiss(β, n)	Curie-Weiss distribution with n spins at temperature $1/\beta$	62
$\mathcal{PP}(k)$	set of all pair partitions of $\{1, \ldots, k\}$	65
$\mathcal{N}(\mu, \Sigma)$	multi-dimensional normal distribution	64
$f_{\mu,\Sigma}$	Lebesgue density of $\mathcal{N}(\mu, \Sigma)$	64
$\text{CovMat}(\alpha)$	Sequences of specific covariance matrices	63
$\underline{t}$	tuple in $[n]^k$	72

Symbol	Description	Page
$(V_{\underline{t}}, E_{\underline{t}}, \phi_{\underline{t}})$	Eulerian graph based on tuple $\underline{t}$	72
$\kappa(\underline{t})$	profile of tuple $\underline{t}$	73
$\kappa_l(\underline{t})$	l-th element of profile $\kappa(\underline{t})$	73
$\ell(\underline{t})$	number of loops in $\underline{t}$	73
$\mathcal{T}(\underline{s})$	equivalence class of tuple $\underline{s}$	75
$\mathcal{T}_{\frac{k}{2}+1}(\underline{s})$	equivalence class of $\underline{s}$ with $k/2+1$ nodes	80
$\mathcal{T}_{\leq\frac{k}{2}}(\underline{s})$	equivalence class of $\underline{s}$ with at most $k/2$ nodes	80
$\mathcal{T}(\underline{s}, \underline{s}')$	equivalence class of tuple pairs	76
$\mathcal{T}^d(\underline{s}, \underline{s}')$	equivalence class of disjoint tuple pairs	76
$\mathcal{T}^c(\underline{s}, \underline{s}')$	equivalence class of overlapping tuple pairs	76
$\mathcal{T}^c_l(\underline{s}, \underline{s}')$	equivalence class of tuple pairs with l overlaps	76
$B(h_n)$	band area depending on halfwidth	101
$T(h_n)$	triangular area depending on halfwidth	101
$S(h_n)$	nontrivial entries depending on halfwidth	102
d_{BL}	bounded Lipschitz metric	114
S_μ	Stieltjes transform of μ	117
$\mathrm{Re}(z)$	real part of complex number z	118
$\mathrm{Im}(z)$	imaginary part of complex number z	117
E	real part of contextual complex number z	149
η	imaginary part of contextual complex number z	149
κ	$\lvert\lvert E\rvert - 2\rvert$ for contextual complex number z	149
$\lambda\!\!\lambda$	Lebesgue measure on $(\mathbb{R}, \mathcal{B})$	39
$f\lambda\!\!\lambda$	measure with Lebesgue density f	39
$\mu * \nu$	convolution of μ and ν	122
$\mu \otimes \nu$	product measure of μ and ν	122
$f * \nu$	convolution of f and ν	123
$f * g$	convolution of f and g	123
P_η	Cauchy kernel with bandwidth η	125
$\sqrt{z}$	complex root with non-negative imaginary part	129
$s(z)$	Stieltjes transform of semicircle distribution	129
$s_n(z)$	Stieltjes transfrom of contextual ESD	133
P_t	probability measure on $\{-1, 1\}$ with expectation t	143
$P_t^{\otimes I}$	I-fold product measure of the P_t	143
$\prec$	stochastic domination	144
$\mathcal{D}_I$	the set $[-10, 10] \times i[1, 10]$	149
$\mathcal{D}_N(\gamma)$	the set $[-10, 10] \times i[N^{\gamma-1}, 10]$	149
$\mathcal{D}_N$	the set $[-10, 10] \times i[N^{-1}, 10]$	149
$\mathcal{I}(A)$	set of intervals $I \subseteq A \subseteq \mathbb{R}$	160
$\lvert I\rvert$	length of interval I	169
Ω_k	error term in local law analysis	174
Z_k	error term in local law analysis	174
$C_{\Phi(\delta_1,\ldots,\delta_l)}$	constants for condition (AAU1)	55

Symbol	Description	Page
$C_n^{(l)}$	constants for condition (AAU2)	55
$D_n^{(l)}$	constants for condition (AAU3)	55
$K_{l,m}$	constants for (AU1)	60
$K_n^{(l)}$	constants for (AU2)	60
$Z_{\beta,n}$	Curie-Weiss normalization constant	62
$C_{\kappa(\underline{s})}$	constants for condition (AAU1)	82
C_s	constant for bounds on $s = S_\sigma$	129
$K_{\beta,p}$	constants for Curie-Weiss moment decay	144
$C_{\epsilon,D}$	constants for $\prec$	144
K_p	constant for Curie-Weiss type moment decay	148
C_P	constant for proximity analysis	177
C_{Det}	constant for root approximation	180
$\neg$	mathematical negation	99

Bibliography

[1] Oskari H. Ajanki, László Erdős, and Torben Krüger. "Local Spectral Statistics of Gaussian Matrices with Correlated Entries". In: *Journal of Statistical Physics* 163.2 (Apr. 2016), pp. 280–302.

[2] Oskari H. Ajanki, László Erdős, and Torben Krüger. "Stability of the Matrix Dyson Equation and Random Matrices with Correlations". In: *Probability Theory and Related Fields* (2018). URL: `https://doi.org/10.1007/s00440-018-0835-z`.

[3] N. I. Akhiezer. *The classical moment problem*. Oliver and Boyd, 1965.

[4] Gerold Alsmeyer. *Wahrscheinlichkeitstheorie*. 5th ed. Skripten zur Mathematischen Statistik 30. (Münster). 2007.

[5] Jörgen Andersen et al. "Enumeration of RNA complexes via random matrix theory". In: *Biochemical Society Transactions* (2013).

[6] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2010.

[7] Zhidong Bai and Jack W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, 2010.

[8] Marwa Banna, Florence Merlevède, and Magda Peligrad. "On the limiting spectral distribution for a large class of symmetric random matrices with correlated entries". In: *Stochastic Processes and their Applications* 125.7 (July 2015), pp. 2700–2726.

[9] Florent Benaych-Georges and Antti Knowles. "Lectures on the Local Semicircle Law for Wigner Matrices". Sept. 2018. URL: `http://www.unige.ch/~knowles/SCL.pdf`.

[10] Patrick Billingsley. *Convergence of Probability Measures*. 2nd ed. John Wiley and Sons, 1999.

[11] Leonid Bogachev, Stanislav Molchanov, and Leonid Pastur. "On the level density of random band matrices". In: *Mathematical Notes* (1991).

[12] Vladimir Bogachev. *Measure Theory*. Vol. 2. Springer, 2006.

[13] Wlodzimierz Bryc, Amir Dembo, and Tiefeng Jiang. "Spectral Measure of Large Random Hankel, Markov and Toeplitz Matrices". In: *The Annals of Probability* 34.1 (2006), pp. 1–38.

[14] G. Casati and V. Girko. "Wigner's semicircle law for band random matrices". In: *Random Operators and Stochastic Equations* 1.1 (1993), pp. 15–21.

[15] Riccardo Catalano. "On Weighted Random Band-Matrices with Dependencies". PhD thesis. FernUniversität in Hagen, Germany, 2016.

[16] Arijit Chakrabarty, Rajat Subhra Hazra, and Deepayan Sarkar. “From random matrices to long range dependence”. In: *Random Matrices: Theory and Applications* 5.2 (2016).

[17] Ziliang Che. “Universality of random matrices with correlated entries”. In: *Electronic Journal of Probability* 22.30 (2017), pp. 1–38.

[18] Yuan Shih Chow and Henry Teicher. *Probability Theory*. 3rd ed. Springer, 1997.

[19] Leszek F. Demkowicz. “Lecture Notes on Energy Spaces”. May 2018. URL: `http://users.ices.utexas.edu/~leszek/classes/EM394H/book2.pdf`.

[20] Rick Durrett. *Probability*. 4th ed. Cambridge University Press, 2010.

[21] Jürgen Elstrodt. *Maß- und Integrationstheorie*. 6th ed. Springer, 2009.

[22] Ryszard Engelking. *General Topology*. Heldermann, 1989.

[23] László Erdős, Antti Knowles, and Horng-Tzer Yau. “Averaging Fluctuations in Resolvents of Random Band Matrices”. In: *Annales Henri Poincaré* 14.8 (Dec. 2013), pp. 1837–1926.

[24] László Erdős, Torben Krüger, and Dominik Schröder. “Random Matrices with Slow Correlation Decay”. Dec. 2017. URL: `https://arxiv.org/pdf/1705.10661.pdf`.

[25] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. “Local Semicircle Law and Complete Delocalization for Wigner Random Matrices”. In: *Communications in Mathematical Physics* 287 (2009), pp. 641–655.

[26] László Erdős, Benjamin Schlein, and Horng-Tzer Yau. “Semicircle Law on Short Scales and Delocalization of Eigenvectors for Wigner Random Matrices”. In: *The Annals of Probability* 37.3 (2009), pp. 815–852.

[27] László Erdős and Horng-Tzer Yau. *A Dynamical Approach to Random Matrix Theory*. American Mathematical Society, 2017.

[28] László Erdős et al. “The local semicircle law for a general class of random matrices”. In: *Electronic Journal of Probability* 18 (Jan. 2013), pp. 1–58.

[29] Gerd Fischer. *Lineare Algebra*. SpringerSpektrum, 2014.

[30] Michael Fleermann. “The Almost Sure Semicircle Law for Random Band Matrices with Dependent Entries”. Nov. 2016. URL: `https://arxiv.org/pdf/1711.10196.pdf`.

[31] Michael Fleermann. “The empirical spectral distribution of symmetric random matrices with correlated entries. An asymptotic analysis employing the method of moments.” MA thesis. University of Münster, Germany, 2015.

[32] Olga Friesen and Matthias Löwe. “A phase transition for the limiting spectral density of random matrices”. In: *Electronic Journal of Probability* 18.17 (2013), pp. 1–17.

[33] Olga Friesen and Matthias Löwe. “The semicircle law for matrices with independent diagonals”. In: *Journal of Theoretical Probability* 26.4 (2013).

[34] Hans-Otto Georgii. *Stochastik*. 4th ed. Walter de Gruyter, 2009.

[35] Friedrich Götze et al. "On the Local Semicircular Law for Wigner Ensembles". In: *Bernoulli* 24.3 (2018), pp. 2358–2400.

[36] Martin Hanke-Bourgeois. *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens*. 3rd ed. Vieweg+Teubner, 2009.

[37] Winfried Hochstättler, Werner Kirsch, and Simone Warzel. "Semicircle Law for a Matrix Ensemble with Dependent Entries". In: *Journal of Theoretical Probability* (2015). URL: `http://dx.doi.org/10.1007/s10959-015-0602-3`.

[38] Leon Isserlis. "On a formula for the product-moment coefficient of each order of a normal frequency distribution in every number of variables". In: *Biometrika* 12 (1918), pp. 134–139.

[39] Todd Kemp. "Math 247A: Introduction to Random Matrix Theory". Dec. 2016. URL: `http://www.math.ucsd.edu/~tkemp/247A.Notes.pdf`.

[40] Alexei Khorunzhy. "On smoothed density of states for Wigner random matrices". In: *Random Operators and Stochastic Equations* 5.2 (Jan. 1997), pp. 147–162.

[41] Werner Kirsch. "A Survey on the Method of Moments". Oct. 2015. URL: `https://www.fernuni-hagen.de/stochastik/downloads/momente.pdf`.

[42] Werner Kirsch and Thomas Kriecherbauer. "Random matrices with exchangeable entries". July 2018. URL: `https://arxiv.org/pdf/1807.05159.pdf`.

[43] Werner Kirsch and Thomas Kriecherbauer. "Semicircle Law for Generalized Curie-Weiss Matrix Ensembles at Subcritical Temperature". In: *Journal of Theoretical Probability* 31.4 (2018), pp. 2446–2458.

[44] Achim Klenke. *Wahrscheinlichkeitstheorie*. 2nd ed. Springer, 2008.

[45] Thomas Koshy. *Catalan Numbers with Applications*. Oxford University Press, 2009.

[46] Anne Boutet de Monvel and Alexei Khorunzhy. "On the Norm and Eigenvalue Distribution of Large Random Matrices". In: *The Annals of Probability* 27.2 (1999), pp. 913–944.

[47] Alexandru Nica and Roland Speicher. *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, 2006.

[48] K.R. Parthasarathy. *Probability measures on metric spaces*. Academic Press, 1967.

[49] Robert Plato. *Numerische Mathematik Kompakt*. 4th ed. Vieweg und Teubner, 2010.

[50] David Pollard. *Convergence of Stochastic Processes*. Springer, 1984.

[51] Michael Reed and Barry Simon. *Fourier Analysis, Self-Adjointness*. Academic Press, 1975.

[52] Omar Rivasplata. "Subgaussian random variables: An expository note". Nov. 2012. URL: `https://sites.ualberta.ca/~omarr/publications/subgaussians.pdf`.

[53] Walter Rudin. *Real and Complex Analysis*. 3rd ed. McGraw-Hill, 1987.

[54] Ludger Rüschendorf. *Mathematische Statistik*. 1st ed. Springer Spektrum, 2014.

[55] Jeffrey H. Schenker and Hermann Schulz-Baldes. "Semicircle law and freeness for random matrices with symmetries or correlations". In: *Mathematical Research Letters* 12 (2005), pp. 531–542.

[56] Satish Shirali and Harkrishan L. Vasudeva. *Metric Spaces*. Springer, 2006.

[57] J. A. Shohat and J. D. Tamarkin. *The problem of moments*. American Mathematical Society, 1943.

[58] Richard P. Stanley. *Enumerative Combinatorics*. Vol. 1. Cambridge University Press, 2012.

[59] Terence Tao. *Topics in Random Matrix Theory*. American Mathematical Society, 2012.

[60] Terence Tao and Van Vu. "The Universality Phenomenon for Wigner Ensembles". Feb. 2012. URL: `https://arxiv.org/pdf/1202.0068`.

[61] Anusch Taraz. *Diskrete Mathematik*. Birkhäuser, 2012.

[62] Emre Telatar. "Capacity of Multi-antenna Gaussian Channels". In: *Transactions on Emerging Telecommunications Technologies* (1999).

[63] Peter Tittmann. *Einführung in die Kombinatorik*. 2nd ed. Springer, 2014.

[64] Eugene P. Wigner. "Characteristic Vectors of Bordered Matrices With Infinite Dimensions". In: *The Annals of Mathematics* 62.3 (1955), pp. 548–564.

[65] Eugene P. Wigner. "On the Distribution of Roots of Certain Symmetric Matrices". In: *The Annals of Mathematics* 67.2 (1958), pp. 225–327.

[66] Stephen Willard. *General Topology*. Addison-Wesley, 1970.

[67] John Wishart. "The generalised product moment distribution in samples from a normal multivariate population". In: *Biometrika* 20 (1928), pp. 32–52.

[68] Hermann Witting and Ulrich Müller-Funk. *Mathematische Statistik II*. 1st ed. Teubner, 1995.

[69] Fuzhen Zhang. *Matrix Theory*. 2nd ed. Springer, 2011.