TOPICS IN LINEAR SPECTRAL STATISTICS
OF RANDOM MATRICES

by

Asad Lodhia

Submitted to the Department of Mathematics on May 3, 2017
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Mathematics

Abstract:

The behavior of the spectrum of a large random matrix is a topic of great interest in probability theory and statistics. At a global level, the limiting spectra of certain random matrix models have been known for some time. For example, the limiting spectral measure of a Wigner matrix is a semicircle law and the limiting spectral measure of a sample covariance matrix under certain conditions is a Marčenko-Pastur law.

The local behavior of eigenvalues for specific random matrix ensembles (GUE and GOE) have been known for some time as well and until recently, were conjectured to be universal. There have been many recent breakthroughs in the universality of this local behavior of eigenvalues for Wigner Matrices. Furthermore, these universality results laws have been proven for other probabilistic models of particle systems, such as Beta Ensembles.

In this thesis we investigate the fluctuations of linear statistics of eigenvalues of Wigner Matrices and Beta Ensembles in regimes intermediate to the global regime and the microscopic regime (called the mesoscopic regime). We verify that these fluctuations are Gaussian and derive the covariance for a range of test functions and scales.

On a separate line of investigation, we study the global spectral behavior of a random matrix arising in statistics, called Kendall’s Tau and verify that it satisfies an analogue of the Marčenko-Pastur Law.

Thesis Supervisor: Alice Guionnet
Title: Professor of Mathematics
INTRODUCTION

§0.1 Limit Theorems in Probability

In classical probability theory, the Law of Large Numbers (LLN) predicts that if \( \{X_i\}_{i=1}^N \) is a sequence of \( \mathbb{R} \)-valued independent identically distributed random variables with distribution \( P \), then for any \( f \in C_b(\mathbb{R}) \) we have

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(X_i) = \mathbb{E}[f(X)] \quad \text{where} \quad X \sim P, \quad \text{(LLN)}
\]

where the convergence holds almost surely. Further, the Central Limit Theorem (CLT) implies

\[
\lim_{N \to \infty} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \{ f(X_i) - \mathbb{E}[f(X_i)] \} = \mathcal{N}(0, \sigma^2(f)) \quad \text{where} \quad \sigma^2(f) := \text{Var}(f(X)), \quad \text{(CLT)}
\]

where the convergence holds in distribution. A key requirement in the LLN and the CLT is the independence of the sequence \( X_i \) along with the observation that each \( f(X_i) \) is bounded (and therefore has a mean and variance). If we conceive of the sequence \( \{X_i\}_{i=1}^N \) as a collection of particles randomly scattered on the line, then the LLN can be interpreted as describing the limiting shape of the configuration they form. Further, the CLT tells us the typical deviation from this shape.

If we were to go further and ask what the local behavior of these particle positions are, say, in the simpler setting where the \( X_i \) have a continuous distribution \( \rho(x) \) with compact support \([a; b]\), then we obtain another interesting limit theorem from probability theory. Specifically, let \( E \in [a; b] \) be a point such that \( \rho > 0 \) in a neighborhood of \( E \) (this is called a point in the bulk of \( \rho \)) and let \( s > 0 \). We have that

\[
\# \left\{ X_i : X_i \in \left[ E, E + \frac{s}{N \rho(E)} \right] \right\} \Rightarrow \mathcal{P}(s), \quad \text{(LRE)}
\]

that is, the number of points landing in an interval of the form \([E, E + \frac{s}{N \rho(E)}]\) converges to a Poisson random variable with parameter \( s \). We refer to this as the Law of Rare Events (LRE).

We may also ask what the behavior of the maximum (or minimum) of the sequence \( X_i \) is. The Borel-Cantelli lemma easily shows that almost surely, \( \max_{i \leq N} X_i \rightarrow b \) in the limit as \( N \to \infty \). The deviation from this limit can have an explicit distribution in certain cases. Suppose that \( \rho(x) = A|x-b|^\gamma(1+o(1)) \) as \( x \to b \) for some constants \( A > 0, \gamma > -1, \gamma \neq 0 \) and \( s > 0 \) then

\[
\lim_{N \to \infty} \mathbb{P} \left( \left\{ \max_{i \leq N} X_i - b \leq -\frac{s}{BN^{\frac{1}{\gamma+1}}} \right\} \right) = e^{-s^{(1+\gamma)}}, \quad \text{(LEV)}
\]
for some scaling constant $B$. The distribution on the right hand side above is called the \textit{Gumbel distribution}; we refer to the above as the Law of Extreme Values (LEV).

In the study of random matrix eigenvalues, there are analogues to the above Theorems that differ greatly from the classical setting. For the (LLN), there is the famous \textit{Wigner Semicircle Law} which states [AGZ09] that if $\lambda_1 \leq \cdots \leq \lambda_N$ are the $N$ real eigenvalues of a Wigner matrix (see Definition 0.1 below) then

$$\frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) \to \int_{-2}^{2} f(x) \frac{\sqrt{4-x^2}}{2\pi} \, dx$$

(0.1)

almost surely for all bounded continuous functions $f$. The fluctuations of the above are also known

$$\sum_{i=1}^{N} f(\lambda_i) - E[f(\lambda_i)] \Rightarrow \mathcal{N}(0, \sigma^2(f)),$$

and scale differently from the classic (CLT) which has a normalization of $\frac{1}{\sqrt{N}}$. Here $\sigma^2(f)$ has been computed explicitly [Joh98, BY05]. Furthermore, the local (bulk) behavior of the eigenvalues does not converge to a classical Poisson point process as in the (LRE). Rather for the case of GUE matrices, it converges to a \textit{determinantal point process} [AGZ09, Lemma 3.5.6] with two-point correlation function given by the \textit{sine-kernel}:

$$K(x, y) = \frac{1}{\pi} \frac{\sin(x - y)}{x - y},$$

this indicates that the dependency structure of the eigenvalues create a novel interaction. Finally, the maximal eigenvalue follows a behavior that is distinct from the usual (LEV). As it turns out the limiting behavior of the largest eigenvalue of the GUE follows the Tracy-Widom distribution [AGZ09, Theorem 3.1.4]

$$\lim_{N \to \infty} P\left( \left\{ N^{\frac{2}{3}} (\lambda_N - 2) \leq t \right\} \right) = F_2(t),$$

which is distinct from the Gumbel distribution (although the $N^{\frac{2}{3}}$ scaling is predicted if we use the $\gamma = \frac{1}{2}$ value of the semicircle density) and is related to the solution to the Painlevé equation (see [AGZ09, Theorem 3.1.5] for the formula for $F_2(t)$). One question that has been heavily investigated in recent years is the universality of these results for a broad class of random matrix models, see [Erd11] for a survey. In this line of recent investigation it was found that in the intermediate scale (between the local sine-kernel behavior and the global semicircle behavior) the semicircle law continues to hold for a broad class of Wigner matrices. Our first line of work in this thesis is to investigate the fluctuations about the semicircle law at this intermediate scale.

These limit Theorems for random matrix eigenvalues have been observed for random processes arising in other contexts. For instance, the length of the longest increasing subsequence of a random permutation follows the Tracy-Widom law, also, the limiting behaviour of certain statistics of random tilings follow the Tracy-Widom Law (see [BDS16] for a recent survey).
\textbf{§0.2 Summary of Part 1}

In Part 1, we study linear spectral statistics of $N \times N$ Wigner random matrices $\mathcal{H}$ on mesoscopic scales. This particular section is based on joint work with N. J. Simm. The definition of our matrix model is:

\textbf{Definition 0.1.} A Wigner matrix is an $N \times N$ Hermitian random matrix $W$ whose entries $W_{ij} = \overline{W_{ji}}$ are centered, independent identically distributed complex random variables satisfying $\mathbb{E}|W_{ij}|^2 = 1$ and $\mathbb{E}W_{ij}^2 = 0$ for all $i$ and $j$. We assume that the common distribution $\mu$ of $W_{ij}$ satisfies the sub-Gaussian decay $\int_{\mathbb{C}} e^{c|z|^2} d\mu(z) < \infty$ for some $c > 0$. This implies that the higher moments are finite, in fact we have $\mathbb{E}|W_{ij}|^q < (Cq)^q$ for some $C > 0$. We denote by $\mathcal{H} = N^{-1/2}W$ the normalized Wigner matrix. Finally, in case all the entries $W_{ij}$ are Gaussian distributed, the ensemble is known as the Gaussian Unitary Ensemble (GUE).

As stated above, a mesoscopic linear statistic is a functional of the form:

$$X_N^{\text{meso}}(f) := \sum_{j=1}^{N} f(d_N(E - \lambda_j))$$

and we define the centered version of the above as

$$\tilde{X}_N^{\text{meso}}(f) = X_N^{\text{meso}}(f) - \mathbb{E}[X_N^{\text{meso}}(f)],$$

where $f \in C^{1,\alpha,\beta}(\mathbb{R})$ and $C^{1,\alpha,\beta}(\mathbb{R})$ denotes the space of all functions with $\alpha$-Hölder continuous first derivative such that $f(x)$ and $f'(x)$ decay faster than $O(|x|^{-1-\beta})$ as $|x| \to \infty$.

Suppose that $d_N = N^\gamma$ where $\gamma$ satisfies the condition $0 < \gamma < 1/3$ and consider test functions $f_1, \ldots, f_M \in C^{1,\alpha,\beta}(\mathbb{R})$ for some $\alpha > 0$ and $\beta > 0$. Then for a fixed $E \in (-2, 2)$ in (0.2) we prove the convergence in distribution

$$(\tilde{X}_N^{\text{meso}}(f_1), \ldots, \tilde{X}_N^{\text{meso}}(f_M)) \Rightarrow (X(f_1), \ldots, X(f_M))$$

where $(X(f_1), \ldots, X(f_M))$ is an $M$-dimensional Gaussian vector with zero mean and covariance matrix

$$\mathbb{E}(X(f_p)X(f_q)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk |k| \hat{f}_p(k)\hat{f}_q(k), \quad 1 \leq p, q \leq M$$

and $\hat{f}(k) := (2\pi)^{-1/2} \int_{\mathbb{R}} f(x) e^{-ikx} dx$. 

\begin{equation}
(0.3)
\end{equation}
§0.3 Summary of Part 2.

The results in this section are from a joint work with Florent Bekerman. In a similar vein to Part 1, we studied linear statistics of particles $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ distributed according a $\beta$-ensemble with potential $V$ whose density is defined by:

$$
P_N^\beta(d\lambda_1, \ldots, d\lambda_N) = \frac{\exp \left( - \frac{N}{\beta} \sum_{i=1}^{N} V(\lambda_i) \right) \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^\beta}{Z_{N,\beta}} \prod_{i=1}^{N} d\lambda_i.
$$

When $V$ is continuous and satisfies

$$
\lim \inf_{x \to \infty} \frac{V(x)}{\beta \log |x|} > 1,
$$

the empirical measure of the $\lambda_i$, defined $L_N := \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$, converges almost surely towards a unique probability measure $\mu_V$. For the sake of this paper, in addition to the above condition on $V$ we made the following additional assumptions on $V$:

- $V \in C^6(\mathbb{R})$.
- The support of $\mu_V$ is a connected interval $[a, b]$ and

$$
\frac{d\mu_V}{dx} =: \rho_V(x) = S(x) \sqrt{(b - x)(x - a)} 	ext{ with } S > 0 \text{ on } [a, b].
$$

- The function $V(\cdot) - \beta \int \log |\cdot - y| d\mu_V(y)$ achieves its minimum on the support only.

With the above assumptions (the second and third are referred to as one-cut assumption and off-criticality respectively) we analyze the statistic $M_N(f)$ given in (1) where now $f \in C^5(\mathbb{R})$ (compactly supported test functions with 5 derivatives) and $0 < \alpha < 1$. We prove that the limiting distribution of $M_N(f)$ is the same as in the Wigner setting above, that is $M_N(f)$ tends towards a gaussian with mean 0 and variance:

$$
\frac{1}{2\beta \pi^2} \int \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 \, dx \, dy,
$$

which up to a scale factor is the same as the variance (0.3).
§0.4 Summary of Part 3.

The results in this section are from a joint work with Afonso Bandeira and Philippe Rigollet. This section differs from the above two in that it studies the limiting distribution of eigenvalues of a random matrix model rather than its fluctuations about a limit, further the results were done at a macroscopic scale rather than the mesoscopic results in the above two cases. The model in the case consisted of a random matrix generated by a sequence of i.i.d random vectors $X_i \in \mathbb{R}^p$ with $1 \leq i \leq n$. The distribution of each component of $\mathbb{R}^p$ is unspecified, we only assume each $X_i(j)$ is independent of any other $X_i(k)$ and each $X_i(j)$ has a distribution that is absolutely continuous with respect to the Lebesgue Measure. From this, we construct the matrix
\[
\tau = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign}(X_i - X_j) \otimes \text{sign}(X_i - X_j),
\]
referred to as Kendall’s rank correlation matrix, or Kendall’s $\tau$. This is a $p \times p$ symmetric matrix with integer entries. Letting
\[
L_N(\tau) := \frac{1}{p} \sum_{i=1}^{p} \delta_{\lambda_i},
\]
where $\lambda_i$ are the ordered eigenvalues of $\tau$, we show under these assumptions that in the limit as $n \to \infty$ and
\[
\lim_{n \to \infty} \frac{p}{n} := \gamma \in (0, \infty),
\]
we have that $L_N$ converges weakly in probability to the distribution of the random variable
\[
\frac{2}{3} Y + \frac{1}{3},
\]
where $Y$ is distributed according to the standard Marčenko-Pastur law with parameter $\gamma$, whose density is
\[
p_\gamma(x) := \begin{cases} \frac{1}{2\pi \gamma x} \sqrt{(b-x)(x-a)} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise}, \end{cases}
\]
and has a point mass $1 - \frac{1}{\gamma}$ at the origin if $\gamma > 1$. Here $a := (1 - \sqrt{\gamma})^2$ and $b := (1 + \sqrt{\gamma})^2$.

We prove that Kendall’s Rank correlation matrix converges to the Marčenko Pastur law, under the assumption that observations are i.i.d random vectors $X_1, \ldots, X_n$ with components that are independent and absolutely continuous with respect to the Lebesgue measure. This is the first result on the empirical spectral distribution of a multivariate $U$-statistic.
Acknowledgements: The authors express thanks to Alice Guionnet for suggesting the main techniques used in the paper. The author (A. L.) in particular expresses his gratitude to Alice Guionnet, who provided helpful advice and support through the NSF grant 6927980 “Random Matrices, Free Probability and the enumeration of maps”. The author (N. J. Simm) expresses his gratitude to Yan Fyodorov, Anna Maltsev and Jérémie Unterberger for stimulating discussions. N. J. Simm was supported on EPSRC grant EP/J002763/1 “Insights into Disordered Landscapes via Random Matrix Theory and Statistical Mechanics".
§1.1 Overview

As stated in the Introduction, the goal of this chapter is to study the limiting fluctuations as \( N \to \infty \) of the linear spectral statistic

\[
X_N^{\text{meso}}(f) := \sum_{j=1}^{N} f(d_N(E - \lambda_j)) \tag{1.1}
\]

where \( \lambda_1, \ldots, \lambda_N \) are the eigenvalues of an \( N \times N \) Wigner random matrix \( \mathcal{H} \) as defined in Definition 0.1. Recall that the mesoscopic or intermediate scale is defined by the assumption that \( d_N \to \infty \) as \( N \to \infty \), but \( d_N/N \to 0 \) as \( N \to \infty \). Therefore, if \( f \) is decaying suitably at \( \infty \), only a fraction \( N/d_N \) of the total number of eigenvalues will contribute in the sum (1.1).

In recent years, there has been growing interest in understanding the limiting distribution of (1.1) on such mesoscopic scales. This interest has stemmed from, e.g., the appearance of novel stochastic processes in probability theory [FKS13], conductance fluctuations in disordered systems [EK14a, EK14b] and linear statistics of the zeros of Riemann’s zeta function [BK14], among others [BD14, BEYY14, DJ13, JL15].

Previously, the majority of studies concentrated exclusively on the macroscopic scale where \( d_N = 1 \) and \( E = 0 \) in (1.1), denoted \( X_N^{\text{macro}}(f) \). In this case it was proved for many different types of random matrix ensembles that, provided \( f \) has at least one derivative, the centered random variable

\[
\tilde{X}_N^{\text{macro}}(f) := X_N^{\text{macro}}(f) - \mathbb{E}X_N^{\text{macro}}(f) \tag{1.2}
\]

converges in distribution to the normal law \( \mathcal{N}(0, \sigma^2) \) as \( N \to \infty \). Furthermore, an explicit formula for the limiting variance \( \sigma^2 \) was obtained, see e.g. [Joh98, BY05]. In analogy with classical probability, we refer to such results as central limit theorems (CLTs).

Going to finer scales, the mesoscopic fluctuations of (1.1) are known to be highly sensitive when compared to the macroscopic scale: in fact the CLT must break down if \( d_N \) grows too quickly [Pas06]. In particular if \( d_N = N \), only a finite number of terms contribute in the sum (1.1) and we cannot expect a Gaussian limit. The latter case \( d_N = N \) is known as microscopic and will not be considered in this article, though see [ErdlI] for an extensive review.

This ensemble was introduced by Wigner who proved that in the limit \( N \to \infty \), the mean eigenvalue distribution of the normalized Wigner matrix \( \mathcal{H} \) converges to the semicircle law. A modern version of this result (e.g. Theorem 2.9 in [AGZ09]) states that this convergence holds weakly almost surely, i.e.

\[
\frac{1}{N} \sum_{j=1}^{N} f(\lambda_j) \to \frac{1}{2\pi} \int_{-2}^{2} f(x) \sqrt{4 - x^2} \, dx, \quad N \to \infty, \quad \text{almost surely.}
\]

for all bounded and continuous functions \( f \). In order to state our main Theorem, we need a condition on the regularity and decay of the test functions \( f \) entering in (1.1). For \( \alpha, \beta > 0 \), let \( C^{1,\alpha,\beta}(\mathbb{R}) \) denote the space of all functions with \( \alpha \)-Hölder continuous first derivative such that \( f(x) \) and \( f'(x) \) decay faster than \( O(|x|^{-1-\beta}) \) as \( |x| \to \infty \). Finally, recall the notation \( \tilde{X}_N^{\text{meso}}(f) := X_N^{\text{meso}}(f) - \mathbb{E}X_N^{\text{meso}}(f) \).
Theorem 1.1. Let $\mathcal{H}$ be a normalized Wigner matrix as in Definition 0.1. Suppose that $d_N = N^\gamma$ where $\gamma$ satisfies the condition $0 < \gamma < 1/3$ and consider test functions $f_1, \ldots, f_M \in C^{1,\alpha,\beta}(\mathbb{R})$ for some $\alpha > 0$ and $\beta > 0$. Then for a fixed $E \in (-2, 2)$ in (1.1) we have the convergence in distribution

$$(\tilde{X}_N^{\text{meso}}(f_1), \ldots, \tilde{X}_N^{\text{meso}}(f_M)) \Rightarrow (X(f_1), \ldots, X(f_M))$$

(1.3)

where $(X(f_1), \ldots, X(f_M))$ is an $M$-dimensional Gaussian vector with zero mean and covariance matrix

$$E(X(f_p)X(f_q)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk |k| \hat{f}_p(k) \hat{f}_q(k), \quad 1 \leq p, q \leq M$$

(1.4)

and $\hat{f}(k) := (2\pi)^{-1/2} \int_{\mathbb{R}} f(x) e^{-ikx} dx$.

This result improves and extends earlier work of Boutet de Monvel and Khorunzhy [BdMK99a] who proved Theorem 1.1 when $0 < \gamma < 1/8$, $M = 1$ and $f(x) = (x - z)^{-1}$ (see also Theorem 1.4 below). Using very different methods, Erdős and Knowles [EK14a, EK14b] proved an analogue of Theorem 1.1 for random band matrices at the same scale $1 \ll d_N \ll N^{1/3}$, which includes the Wigner matrices as a special case, but includes a heavier assumptions on the allowed test functions $f$. One year after the present work originally appeared on arXiv, He and Knowles [HK16] extended our Theorem 1.1 to all scales $1 \ll d_N \ll N$. Due to exploiting a similar Helffer-Sjöstrand functional calculus approach to ours (crucially, in the present work this part of the calculation actually works on all scales $1 \ll d_N \ll N$), the regularity conditions in [HK16] remain identical to our Theorem 1.1. Apart from these works, CLTs for (1.1) were also obtained in several other ensembles [BdMK99b, Sos00, FKS13, DJ13, BD14, BEYY14]. Although these works extend to scales $0 < \gamma < 1$, the proofs rely on exact formulas for the distribution of the eigenvalues, which are unavailable in the Wigner setting.

Let us now make some general remarks about Theorem 1.1. On macroscopic scales $\gamma = E = 0$ the results are different for Wigner matrices since the limiting covariance depends on the fourth moment of the matrix entries [BS04, Shc11]. On mesoscopic scales with fixed $E \in (-2, 2)$, we show that this difference vanishes, indicating a particularly strong form of universality for formula (1.4) (see also [KKP96]). As with the local regime, the limiting distribution of (1.1) is universal in the choice $E \in (-2, 2)$ around which ones samples the eigenvalues. Apparently unique to the mesoscopic regime, however, is the scale invariance of the limiting Gaussian process: formula (1.4) is unchanged after rescaling the arguments of the test functions by any parameter (see also Section 1.3 ). Optimal conditions on the test functions given in Theorem 1.1 remains a significant issue ever since the seminal work of Johansson [Joh98]. The latter article suggests that in the macroscopic regime, only finiteness of the limiting variance should suffice to conclude asymptotic Gaussianity, see [SW13] for recent progress in this direction. In the mesoscopic regime we believe analogously that optimal conditions for asymptotic Gaussianity of (1.1) should be that $\int_{\mathbb{R}} |k| ||\hat{f}(k)||^2 dk < \infty$. It is historically interesting to remark that (1.4) already appeared in a famous 1963 paper of Dyson and Mehta [DM63].
From a probabilistic viewpoint, the semi-circle law (0.1) from the Introduction may be interpreted as the law of large numbers for the eigenvalues of a Wigner matrix; this may be considered the first natural step for the probabilist. The second natural step is to prove the CLT for the macroscopic fluctuations around (0.1). Going now to mesoscopic scales, only the first step has been investigated in detail for Wigner matrices, with the corresponding results known as local semi-circle laws, so called because they track the convergence to the semi-circle closer to the scale of individual eigenvalues. The local semi-circle laws turned out to be a very important tool in proving the long-standing universality conjectures for eigenvalue statistics at the microscopic scale [Meh04, Erd11, TV11]. Consequently, a number of refinements of Wigner’s semi-circle law of increasing optimality were obtained in recent years [ESY09a, EYY12b, EYY12a, EKYY13]. Such results will play a crucial role in our proof of the CLT at mesoscopic scales.

In order to state the local semi-circle law, it is convenient to work with the resolvent $G(z) = (\mathcal{H} - z)^{-1}$, $\Im(z) > 0$. Then according to (0.1), the Stieltjes transform of $\mathcal{H}$

$$s_N(z) := \frac{1}{N} \text{Tr} G(z)$$

should be close to the Stieltjes transform of the semi-circle:

$$s(z) := \frac{1}{2\pi} \int_{-2}^{2} (x - z)^{-1} \sqrt{4 - x^2} \, dx$$

The local semi-circle law shows that this convergence remains valid at mesoscopic scales $\Im(z) = O(d_N^{-1})$ for $1 \ll d_N \ll N$. The following is a version (in our notation) of this result that we will use.

**Theorem 1.2** (Cacciapuoti, Maltsev, Schlein, 2014). [CMS14, Theorem 1 (i)] Fix $\tilde{\eta} > 0$ and let $z = t + i\eta$ with $t \in \mathbb{R}$ and $\eta > 0$ fixed. Then there are constants $M_0, N_0, C, c, c_0 > 0$ such that

$$\mathbb{P}\left( |s_N(E + z/d_N) - s(E + z/d_N)| > \frac{K d_N}{N \tilde{\eta}} \right) \leq (Cq)^c q^{2K^q}$$

for all $\frac{\eta}{d_N} \leq \tilde{\eta}$, $|E + t/d_N| \leq 2 + \eta/d_N$, $N > N_0$ such that $\frac{N \eta}{d_N} \geq M_0$, and $q \leq c_0 \left( \frac{N \eta}{d_N} \right)^{1/8}$.

To prove Theorem 1.1, we will start by proving it for the special case of the resolvent $f(x) = (x - z)^{-1}$. Indeed, one can interpret $s_N(E + z/d_N)$ as a random process on the upper-half plane $\mathbb{H}$ and ask whether, after appropriate centering and normalization, a universal limiting process exists. We will show that the function $(N/d_N)(s_N(E + z/d_N) - \mathbb{E}s_N(E + z/d_N))$ converges to the $\Gamma^+-$processes. These are certain analytic-pathed Gaussian processes defined on $\mathbb{H}$. In order to define these processes, we first review a well-known Gaussian process—the Fractional Brownian motion.

Fractional Brownian motion is a continuous time Gaussian process $B_H(t)$ indexed by a number $H \in (0, 1)$ and having covariance

$$\mathbb{E}(B_H(t)B_H(s)) = c_H(|t|^{2H} + |s|^{2H} - |t - s|^{2H})$$

(1.5)
where $c_H$ is a normalization constant. A generalization of the usual Brownian motion ($H = 1/2$), these processes are characterized by their fundamental properties of stationary increments, scale invariance (i.e. $B_H(at) = a^H B_H(t)$) and Gaussianity. The parameter $H$ is known as the Hurst index and describes the raggedness of the resulting stochastic motion, with the limit of vanishingly small $H$ to be considered the most irregular (see e.g. Proposition 2.5 in [DOT03]). Although the fBm processes were invented by Kolmogorov, they were very widely popularized due to a famous work of Mandelbrot and van Ness [MvN68] and since have appeared prominently across mathematics, engineering and finance, among other fields, see [LSSW14] for a survey of fractional Gaussian fields. Until recently however, no relation between the fBm processes and random matrix theory was known. The latter relation (discovered in [FKS13]) goes via the limit $H \to 0$ and the following regularization

$$B_H^{(n)}(t) := \frac{1}{2\sqrt{2}} \int_0^\infty \frac{e^{-\eta s}}{s^{1/2+H}} \left( [e^{-ists} - 1]B_C(ds) + [e^{ists} - 1]B_C(ds) \right)$$

(1.6)

where $B_C(s) := B_1(s) + iB_2(s)$ and $B_1, B_2$ are independent copies of standard Brownian motion. One can verify that in the limit $\eta \to 0$, one recovers precisely the fractional Brownian motion, i.e. $B_H^{(0)}(t) = B_H(t)$. On the other hand, taking instead the limit $H \to 0$ in (1.6), one obtains a process $B_0^{(n)}(t)$, about which the following was proved:

**Theorem 1.3** ([FKS13] Fyodorov, Khoruzhenko and Simm). For a GUE random matrix $\mathcal{H}_{GUE}$, consider the sequence of stochastic processes

$$W_N^{(n)}(t) := \log \left| \det \left( \mathcal{H}_{GUE} - E - \tau + i\eta \frac{\tau}{d_N} \right) \right| - \log \left| \det \left( \mathcal{H}_{GUE} - E - i\eta \frac{\tau}{d_N} \right) \right|$$

(1.7)

and $\bar{W}_N^{(n)}(t) := W_N^{(n)}(t) - E(W_N^{(n)}(t))$. On any mesoscopic scales of the form $d_N \to \infty$ with $d_N = o(N/\log(N))$ and with fixed $\tau \in \mathbb{R}$, $\eta > 0$ and $E \in (-2, 2)$, the process $\bar{W}_N^{(n)}$ converges weakly in $L^2[a,b]$ to $B_0^{(n)}$ as $N \to \infty$.

In particular, this gives a functional (in $L^2$) version of the CLT of Theorem 1.1 for GUE random matrices with $f_k(x) = \log |x - \tau_k - i\eta| - \log |x - i\eta|$. Either by computing the resulting $H^{1/2}$ norm (1.4) or by computing the covariance of $B_0^{(n)}$ as defined in (1.6) one finds the logarithmic correlations

$$E((B_0^{(n)}(t) - B_0^{(n)}(s))^2) = \frac{1}{2} \log \left( \frac{(t-s)^2}{\eta^2} + 1 \right).$$

(1.8)

Thus $B_0^{(n)}$ inherits many of the fundamental properties of fBm, including Gaussianity, stationary increments, although now one has the ‘regularized self-similarity’ $B_0^{(an\eta)}(at) \sim B_0^{(n)}(t)$ (the latter following from the scale invariance of the inner product (1.4). More generally, Gaussian fields with logarithmic correlations have received a great deal of recent attention across mathematics and physics, see [FLDR12, FK14] and references therein. The
most famous example of such a field is undoubtedly the 2D Gaussian Free Field (GFF) [She07], which has important applications in areas such as quantum gravity [DRSV14], Gaussian multiplicative chaos and Stochastic Loewner Evolution [SS13]. The GFF is also believed to play a central role in random matrix theory. For example, similarly to (1.7), it has appeared in relation to the characteristic polynomial, either explicitly [RV07, AHM11] or in what appear to be its various one-dimensional slices [HKO01, FKS13, Web14]. More recently it has appeared as the height function for the minor processes of random matrices [BG, Bor14].

Going now to the Stieltjes transform \( s_N(z) \) of Theorem 1.2, the trivial relation

\[
\frac{N}{d_N} \Re \{ s_N(\tau + i\eta) \} = \frac{\partial}{\partial \tau} W_N^{(\eta)}(\tau)
\]

suggests that the appropriate limiting object should be related to the derivative of \( B_0^{(\eta)}(\tau) \). Although such a derivative could obviously be represented by differentiating inside the Fourier integral in (1.6), it can also be conveniently represented by a random series. More generally, for \( z \in \mathbb{H} \) and Hurst index \( H < 1 \), define the following ‘Cayley’ series

\[
\Gamma^+_H(z) := \frac{1}{\sqrt{2}} \left( \frac{z + i}{2} \right)^{2H-2} \sum_{k=0}^{\infty} \frac{\Gamma(2-2H+k)}{\Gamma(2-2H)k!} \left( \frac{z - i}{z + i} \right)^k (\xi_k^{(1)} + i\xi_k^{(2)})
\] (1.9)

where \( \{\xi_k^{(1)}, \xi_k^{(2)}\}_{k=0}^{\infty} \) are a family of real i.i.d. standard Gaussians. A quick computation with the series (1.9) shows that it has zero mean and covariance structure

\[
\mathbb{E}(\Gamma^+_H(z_1)\Gamma^+_H(z_2)) = \frac{1}{(i(z_1 - z_2))^{2-2H}}.
\] (1.10)

with \( \mathbb{E}(\Gamma^+_H(z_1)\Gamma^+_H(z_2)) = 0 \). It follows that the processes \( \Gamma^+_H \) are stationary on horizontal line segments of the complex plane (c.f. the stationary increments (1.8) for the integrated version). The \( \Gamma^+_H \)-processes were originally introduced by Unterberger [Unt09] in the context of geometric rough path theory and stochastic partial differential equations, but since then the relation to random matrix theory has apparently gone unnoticed. We will show that \( \Gamma^+_0(z) \) is directly related to a fundamental object of random matrix theory: the normalized trace of the resolvent.

**Theorem 1.4.** Consider the resolvent \( G(z) = (\mathcal{H} - z)^{-1} \). Under the same assumptions as Theorem 1.1, the centered and normalized trace

\[
V_N(z) := \frac{1}{d_N} \left( \text{Tr} G(E + z/d_N) - \mathbb{E} \text{Tr} G(E + z/d_N) \right), \quad \Re(z) > 0
\]

converges in the sense of finite dimensional distributions to \( \Gamma^+_0(z) \) as \( N \to \infty \). That is, for any finite set of points \( z_1, \ldots, z_M \) in the upper half-plane \( \mathbb{H} \), we have

\[
(V_N(z_1), \ldots, V_N(z_M)) \Rightarrow (\Gamma^+_0(z_1), \ldots, \Gamma^+_0(z_M)), \quad N \to \infty.
\] (1.11)
Furthermore, the process \( V_N \) is tight in the space \( \mathcal{U}(D) \) of continuous functions defined on a bounded \( N \)-independent rectangle \( D \subset \mathbb{H} \) and \( V_N \) converges weakly to \( \Gamma_0^+ \) in \( \mathcal{U}(D) \).

**Proof.** For the finite dimensional convergence in (1.11), see Section 1.2. The tightness condition in \( \mathcal{U}(D) \) follows from Corollary 1.30.

Intuitively, the underlying reason for the covariance structure (1.10) (with \( H = 0 \)) appearing in random matrix theory can be traced back to the fundamental relation with the sine-kernel

\[
\lim_{\eta_1, \eta_2 \to 0} \lim_{N \to \infty} \mathbb{E}(V_N(t_1 + i\eta_1)V_N(t_2 + i\eta_2)) \bigg|_{d_N = N} = \left( \frac{1}{\pi} \frac{\sin(\pi(t_1 - t_2))}{(t_1 - t_2)} \right)^2
\]

where (heuristically) going to slightly larger scales \( d_N = N^\gamma \) with \( 0 < \gamma < 1 \) has the effect of a large time separation \([t_1 - t_2]\) smoothening out the oscillations in the numerator, thus reproducing (1.10) with \( H = 0 \) (see e.g. [BZ93, Pas06] for additional heuristics).

Theorem 1.3 can now be easily extended to Wigner matrices, starting with the identity

\[
W_{N}^{(n)}(\tau) = \int_0^\tau \Re(V_N(t + i\eta)) \, dt. \tag{1.12}
\]

Next, by the rigidity of Theorem 1.2, we have \( \mathbb{E}|V_N(t+i\eta)|^2 \) bounded uniformly on compact subsets of \( t \) and \( \eta \in [\delta, \infty) \) for fixed \( \delta > 0 \) (see Proposition B.4). Then a standard tightness argument (see e.g. [Gri76]) combined with (1.11) allows us to conclude the convergence in distribution as \( N \to \infty \),

\[
\int_0^\tau \Re(V_N(t + i\eta)) \, dt \longrightarrow \int_0^\tau \Re(\Gamma_0^+(t + i\eta)) \, dt \sim B_0^{(n)}(\tau).
\]

This implies that \( W_{N}^{(n)} \to B_0^{(n)} \), though now in the Wigner case, subject to a more restricted growth of the parameter \( d_N \) than in Theorem 1.3. In all cases considered here, optimal conditions on the growth of \( d_N \) should be anything asymptotically slower than the microscopic scale, i.e. we expect our main results to hold provided only that \( d_N = o(N) \).

Our proof of Theorem 1.4 will follow closely the approach popularized by Bai and Silverstein [BS10a, BS04]. The technique begins by exploiting the independence of the matrix entries of \( \mathcal{H} \) to write \( \operatorname{Tr} G(z) \) as a sum of martingale differences. Then a classical version of the martingale CLT implies that only 2 estimates are required in order to conclude asymptotic Gaussianity. For the macroscopic regime, this technique was applied successfully to conclude CLTs for many random matrix ensembles, though not without significant computations [BS04, RS06, BWZ09, TV12, BGGM13, OR14]. The mesoscopic regime is characterized by the situation that \( \Re(z) = O(d_N^{-1}) \) as \( N \to \infty \), which is further problematic in that the majority of bounds for resolvents involve powers of \( \Re(z)^{-1} \). To overcome this we use the rigid control provided by Theorem 1.2 many times, but for technical reasons we were not able to avoid obtaining estimates of order \( N^{-1} \Re(z)^{-3} \). Such estimates are the source of the restriction on \( d_N \) in Theorem 1.1.
To pass Theorem 1.4 onto the general linear statistic (1.1) of Theorem 1.1, we use an exact formula (see Lemma C.1):

$$\tilde{X}_N^{\text{meso}}(f) = \frac{1}{\pi} \Re \int_0^\infty \int_{-\infty}^\infty V_N(\tau + i\eta) \overline{\partial} \Psi_f(\tau, \eta) \, d\tau \, d\eta$$

(1.13)

where $\overline{\partial} := \frac{\partial}{\partial \tau} + i\frac{\partial}{\partial \eta}$ and $\Psi_f$ is a certain 2-dimensional extension of $f$, known as an almost-analytic extension [Dav95]. Since $\overline{\partial} \Psi_f$ is deterministic, we can use our CLT for $V_N(\tau + i\eta)$ to conclude a CLT for $\tilde{X}_N^{\text{meso}}(f)$. The main problem there is to interchange the distributional convergence for $V_N$ with the integrals appearing in (1.13). To perform such an interchange it will suffice to prove a certain tightness condition which will boil down to having sharp control on $E|V_N(\tau + i\eta)|^2$ in the various regimes of $\tau$ and $\eta$. In the bulk of the Wigner semi-circle with $\eta/d_N \gg N^{-1}$, the optimal bound of Theorem 1.2 plays a key role, since earlier estimates involving $\log(N)$ and $N^\epsilon$ factors would lead to a divergent estimate in the mesoscopic regime. In the regions outside the bulk, or with very small imaginary part $\eta/d_N \ll N^{-1}$, we employ the recent variance estimates of [SW13] (see Proposition 1.23) which have the advantage of holding uniformly in $\eta > 0$, but the disadvantage of an additional factor $d_N^\epsilon$ appearing in the bound. In this way we are able to remove the assumption of very rapid decay, which appears in most studies on the mesoscopic regime [Sos00, BD14, BEYY14]. In contrast, there is no decay requirement in the macroscopic regime and the main important characteristic is the regularity of $f$ [SW13], while here the decay adds an additional complexity to the problem.

The structure of this chapter is as follows. In Section 1.2 we prove the finite dimensional convergence in Theorem 1.4 on scales $1 \ll d_N \ll N^{1/3}$. In Section 1.3 we extend the obtained results to compactly supported functions $f \in C_{c,1}^{1,\alpha}(R)$ and show how to replace the assumption of compact support with a suitable decay condition on $f$. 

15
§1.2 Convergence in law of the Stieltjes transform

The goal of this section is to prove the following:

**Proposition 1.5.** Let \( z_1, \ldots, z_M \) be \( M \) fixed numbers in the upper half of the complex plane \( \mathbb{H} \). Under the same assumptions as Theorem 1.4, the function \( V_N \) converges in the sense of finite dimensional distributions to \( \Gamma_0^{+} \), i.e. we have the convergence in law

\[
(V_N(z_1), \ldots, V_N(z_M)) \to (\Gamma_0^{+}(z_1), \ldots, \Gamma_0^{+}(z_M)), \quad N \to \infty,
\]

where \( \Gamma_0^{+}(z) \) is a Gaussian process on \( \mathbb{H} \) with covariance \( C(z_1, z_2) \) defined by

\[
E(\Gamma_0^{+}(z_1)\overline{\Gamma_0^{+}(z_2)}) = \frac{1}{(i(z_1 - z_2))^2}
\]

and

\[
E(\Gamma_0^{+}(z_1)\Gamma_0^{+}(z_2)) = 0.
\]

To prove Proposition 1.5, it is enough to fix a linear combination

\[
Z_M := \sum_{p=1}^{M} c_p V_N(z_p) = \sum_{p=1}^{M} c_p \frac{1}{d_N} (\text{Tr}(G(E + z_p/d_N)) - E \text{Tr}(G(E + z_p/d_N)))
\]

and to prove that \( Z_M \) converges in distribution to a Gaussian random variable with the appropriate variance. Our starting point is that \( Z_M \) can be expressed as a sum of martingale differences, to which a classical version of the martingale CLT can be applied, see Theorem A.1 in the appendix. To satisfy the conditions of the martingale CLT we shall follow the technique outlined in Chapter 9 of the book [BS10a] of Bai and Silverstein, which is equivalent to the work [BY05]. Our approach is also valid at the macroscopic scales considered in [BY05] and we feel gives a somewhat more accessible proof in this case.

We first outline the martingale method and provide the notation used in the remainder of this Section. Let \( E_k \) denote the conditional expectation with respect to the \( \sigma \)-algebra generated by the upper-left \( k \times k \) corner of the Wigner matrix \( W \). Then we have the martingale decomposition

\[
Z_M = \sum_{k=1}^{N} X_{k,N}
\]

where

\[
X_{k,N} := (E_k - E_{k-1}) \sum_{p=1}^{M} c_p \frac{1}{d_N} \text{Tr}(G(E + z_p/d_N))
\]

Therefore, to prove Proposition 1.5, it will suffice to check the following two conditions:

1. **The Lindeberg condition:** for all \( \epsilon > 0 \), we have

\[
\sum_{k=1}^{N} E(|X_{k,N}|^2 1_{|X_{k,N}| > \epsilon}) \to 0, \quad N \to \infty.
\]  

16
2. **Conditional variance**: we have the convergence in probability

\[
\sum_{k=1}^{N} E_{k-1} |X_{k,N}|^2 \to \sum_{l,m=1}^{M} c_l \overline{c}_m C(z_l, \overline{z}_m), \quad N \to \infty, \quad (1.15)
\]

\[
\sum_{k=1}^{N} E_{k-1} X_{k,N}^2 \to \sum_{l,m=1}^{M} c_l c_m C(z_l, z_m), \quad N \to \infty, \quad (1.16)
\]

where \(C(z_l, z_m) = (i(\tau_l - \tau_m + i(\eta_l + \eta_m)))^{-2}\) denotes the covariance in Proposition 1.5.

Before we proceed with the proof of these conditions, we provide some of the relevant notation.

**Important notation**: Until now the complex numbers \(z_p\) were independent of \(N\). For notational convenience and in the remainder of this Section only, we will now allow the implicit \(N\)-dependence

\[z_p := E + \frac{\tau_p + i\eta_p}{d_N}\]

As before, the sequence \(d_N \to \infty\) as \(N \to \infty\) with \(d_N/N \to 0\) and \(\tau_p, \eta_p\) are fixed real numbers with \(\eta_p \neq 0\). We fix \(E \in (-2+\delta, 2-\delta)\) strictly inside the support of the limiting semi-circle for some small \(\delta > 0\). Let \(H_k\) be the \((N-1) \times (N-1)\) Wigner matrix obtained by erasing the \(k\)-th row and column from \(H\). We denote by \(G_k(z) = (H_k - z)^{-1}\) the corresponding resolvent. The following formula, a consequence of the *Schur complement formula* from Linear Algebra, will play an important role:

\[
\text{Tr}(G(z)) - \text{Tr}(G_k(z)) = \frac{1 + h_k^\dagger G_k(z)^2 h_k}{H_{kk} - z - h_k^\dagger G_k(z) h_k}
\]

where \(h_k\) is the \(k\)-th column of \(H\) with the \(k\)-th entry removed [BS10a, Theorem A.5].

Recall the following standard notation for convergence of random variables in \(L^p\). For a sequence of random variables \(\{X_N\}_{N=1}^{\infty}\), we write \(X_N = O_{L^p}(u(N))\) to mean there exists a constant \(c\) such that \(E|X_N|^p \leq cu(N)\) for all \(N\) large enough. We will repeatedly use the standard fact that if \(X_N\) converges to \(X\) in probability and \(Y_N\) converges to zero in \(L^p\), \(p \geq 1\), then \(X_N + Y_N\) converges to \(X\) in probability.

We start with the proof of the Lindeberg condition (1.14) which follows from the following stronger result (due to the trivial inequality \(|X_{k,N}|^2 1_{|X_{k,N}| > \epsilon} \leq \epsilon^{2-\delta}|X_{k,N}|^\delta\) with \(\delta > 2\)):

**Lemma 1.6** (Lyapunov condition). For all mesoscopic scales \(1 \ll d_N \ll N^{1-\epsilon}\) with \(\epsilon > 0\), there is an integer \(\delta > 2\) such that

\[
\sum_{k=1}^{N} E \left| (E_k - E_{k-1}) \sum_{p=1}^{M} c_p \frac{1}{d_N} \text{Tr} G(z) \right|^\delta \to 0, \quad N \to \infty.
\]
Proof. By the triangle inequality it suffices to verify the claim when \( M = 1, c_1 = 1 \). By definition of \( E_k \) we have \((E_k - E_{k-1})d_N^{-1}\text{Tr}G(z_1) = (E_k - E_{k-1})Z_{k,N} \) where \( Z_{k,N} := d_N^{-1}(\text{Tr}G(z_1) - \text{Tr}G_k(z_1)) \). Then Schur's complement formula implies

\[
Z_{k,N} = \frac{1}{d_N} \frac{1 + h_k^1 G_k(z_1)^2 h_k}{\mathcal{H}_{kk} - z_1 - h_k^1 G_k(z_1) h_k}
\]

\[
= \frac{1}{d_N} (1 + \delta_N^{k,2}(z_1) + N^{-1}\text{Tr}(G_k(z_1)^2)G_{kk}(z_1))
\]

where we made use of the identity for the diagonal elements of the resolvent

\[
G_{kk}(z_1) = \frac{1}{\mathcal{H}_{kk} - z_1 - h_k^1 G_k(z_1) h_k}
\]

and defined

\[
\delta_N^{k,n}(z_1) := h_k^1 G_k(z_1)^n h_k - N^{-1}\text{Tr}(G_k(z_1)^n)
\] (1.17)

By the conditional Jensen inequality, we have \(|E_k Z_k|^{\delta} \leq E_k |Z_k|^\delta\). Hence it is sufficient to prove

\[
\sum_{k=1}^N E[Z_{k,N}]^{\delta} \to 0, \quad N \to \infty
\] (1.18)

The limit (1.18) follows from standard concentration inequalities applied to the variables \( d_N^{-1} \delta_N^{k,2}(z) \), \( G_{kk}(z) \) and \( d_N^{-1}N^{-1}\text{Tr}G(z)^2 \). In particular, Lemmas B.1, B.2 and 1.11 show that for any fixed \( q > 0 \), we have the estimates

\[
d_N^{-1} \delta_N^{k,2}(z) = O_{L^q}((d_N/N)^q/2),
\]

\[
G_{kk}(z) = O_{L^q}(1),
\]

\[
d_N^{-1}N^{-1}\text{Tr}G(z)^2 \leq O_{L^q} \left( \max\{d_N^{-q}, (d_N/N)^q\} \right),
\]

Then applying Cauchy-Schwarz and choosing \( \delta > 0 \) large enough, we obtain (1.18). \( \blacksquare \)

Remark 1.7. In the macroscopic regime \( d_N = 1 \), one can argue similarly that \( Z_{k,N} = (1 + s'(z)(-z - s(z))^{-1} + O(N^{-1/2}) \) with high probability. The leading term in this asymptotic is deterministic and does not contribute to \((E_k - E_{k-1})Z_{k,N} \), while the error term is small enough to imply (1.18).

We now proceed to the remaining and most challenging part of the proof of Proposition 1.5, which is to verify condition (1.5). Before we proceed, it’s worth noting that both \( X_k^2 \) and \(|X_k|^2 \) are finite linear combinations of terms of the form

\[
(E_k - E_{k-1}) \frac{1}{d_N} \text{Tr}G(z_1) \times (E_k - E_{k-1}) \frac{1}{d_N} \text{Tr}G(z_2)
\]

and so it suffices to prove the convergence for a single mixed term in the linear combination. Setting

\[
Y_k(z) := (E_k - E_{k-1})d_N^{-1}\text{Tr}G(z),
\]

18
our essential goal in the remainder of this section will be to prove that we have the convergence in probability

\[ C_N(z_1, z_2) := \sum_{k=1}^{N} \mathbf{E}_{k-1}[Y_k(z_1)Y_k(z_2)] \to \frac{1}{(i(\tau_1 - \tau_2 + i(\eta_1 + \eta_2)))^2}, \quad N \to \infty \]  

(1.19)

In what follows, the proof of (1.19) is divided into 3 main parts: first we rewrite \( C_N(z_1, z_2) \) in a form suitable for the computation of asymptotics, then the main asymptotic results are obtained and finally they are used to prove (1.19). **Simplifying the covariance kernel:** Our first Proposition shows that \( C_N(z_1, z_2) \) can be approximated in the following way

**Proposition 1.8.** In terms of the variables (1.17), define the covariance kernel

\[ \hat{C}_N(z_1, z_2) := \frac{1}{d_N} \frac{\partial^2}{\partial z_1 \partial z_2} \left[ s(z_1)s(z_2) \sum_{k=1}^{N} \mathbf{E}_{k-1}[E_k \delta_N^{k,1}(z_1)E_k \delta_N^{k,1}(z_2)] \right] \]

(1.20)

Then we have

\[ C_N(z_1, z_2) = \hat{C}_N(z_1, z_2) + O_L \left( \sqrt{d_N^2/N} \right) \]

(1.21)

**Proof.** As in the proof of the Lindeberg condition, we start with Schur's complement formula which implies that

\[ Y_k(z) = (E_k - E_{k-1}) \frac{1}{d_N} \frac{1 + h_k^I G_k(z)^2 h_k}{H_{kk} - z - h_k^I G_k(z) h_k} \]

Rewriting \( Y_k(z) \) via the small terms (1.17) and expanding, we obtain the exact identity

\[ Y_k(z) = (E_k - E_{k-1}) \frac{\partial}{\partial z} \frac{H_{kk} - \delta_N^{k,1}(z)}{d_N \left( z + N^{-1} \text{Tr}(G_k(z)) \right)} + \epsilon_{k,N}(z) \]

where

\[ \epsilon_{k,N}(z) := (E_k - E_{k-1}) \left( \frac{1}{d_N} \frac{(H_{kk} - \delta_N^{k,1}(z))^2 G_k(z)}{z + N^{-1} \text{Tr}(G_k(z))} - \frac{1}{d_N} \frac{\delta_N^{k,2}(H_{kk} - \delta_N^{k,1}(z))}{z + N^{-1} \text{Tr}(G_k(z))} \right). \]

This identity is implicit in the work [BY05] (see Section 4.1 in [BY05]), but we provide the derivation in the Appendix, Lemma B.3. Then as in the proof of (1.18), Lemma B.1 implies the bound \( \epsilon_{k,N}(z) = O_L \left( (d_N/N)^q \right) \) uniformly in \( k \). Now we replace \( (z + N^{-1} \text{Tr}(G_k(z)))^{-1} \) with \( -s(z) \), causing an error to \( Y_k(z) \) of the form

\[ (E_k - E_{k-1}) \frac{1}{2\pi i d_N} \int_{S_z} d\omega \frac{H_{kk} - \delta_N^{k,1}(\omega)}{(\omega - z)^2} \left( \frac{1}{\omega + N^{-1} \text{Tr}(G_k(\omega)) + s(\omega)} \right) \]
where we employed analyticity to write the \( z \)-derivative as a contour integral over a small circle \( S_z \) centered at \( z \) with radius of order \( d^{-1}_N \) (avoiding intersection with \( \mathbb{R} \)). By Lemma B.1, \( \delta^{k,1}_N(\omega) \) is \( O_{L^\delta}((d_N/N)^{q/2}) \) and \( (\omega + N^{-1} \text{Tr} G_k(\omega))^{-1} + \delta(\omega) \) is \( O_{L^\delta}((d_N/N)^{q-\epsilon'}) \) for some small \( \epsilon' \), so the error caused by this replacement is at most \( O_{L^\delta}((d_N/N)^q) \).

Therefore, we have \( Y_k(z) = \tilde{Y}_k(z) + O_{L^\delta}((d_N/N)^q) \) where

\[
\tilde{Y}_k(z) = -(E_k - E_{k-1}) \frac{\partial}{\partial z} \frac{1}{d_N} s(z)(H_{kk} - \delta^{k,1}_N(z))
\]

Using properties of the conditional expectation, we compute that

\[
\sum_{k=1}^N E_{k-1}[\tilde{Y}_k(z_1)\tilde{Y}_k(z_2)] = \frac{1}{d_N^2} \frac{\partial^2}{\partial z_1 \partial z_2} s(z_1)s(z_2) + \tilde{C}_N(z_1, z_2)
\]

Then the covariance \( C_N(z_1, z_2) \) can be estimated as

\[
C_N(z_1, z_2) \approx \tilde{C}_N(z_1, z_2) + \sum_{k=1}^N E_{k-1}(Y_k(z_1) - \tilde{Y}_k(z_1))\tilde{Y}_k(z_2)
\]

\[
+ \sum_{k=1}^N E_{k-1}(Y_k(z_2) - \tilde{Y}_k(z_2))\tilde{Y}_k(z_1)
\]

\[
+ \sum_{k=1}^N E_{k-1}(Y_k(z_1) - \tilde{Y}_k(z_1))(Y_k(z_2) - \tilde{Y}_k(z_2)) + O_{L^1}(1/d_N^2)
\]

where we used that \( s'(z_1)s'(z_2) \) are uniformly bounded for any fixed \( E \in (-2, 2) \). By our estimates for \( Y_k - \tilde{Y}_k \), the second last term above is \( O_{L^1}((d^2_N/N)) \). For the middle terms, we apply Cauchy-Schwarz (twice) to obtain

\[
E \left[ \sum_{k=1}^N E_{k-1}[\tilde{Y}_k(z_1)(Y_k(z_2) - \tilde{Y}_k(z_2))] \right] \leq \sqrt{\sum_{k=1}^N E[Y_k(z_1) - \tilde{Y}_k(z_1)]^2 \sum_{k=1}^N E[\tilde{Y}_k(z_2)]^2}
\]

\[
= \sqrt{\sum_{k=1}^N E[Y_k(z_1) - \tilde{Y}_k(z_1)]^2 \sqrt{E[C_N(z_2, z_2) + d^{-2}_N |s'(z_2)|^2}}}
\]

The first term in the product above is \( O(\sqrt{d^2_N/N}) \). In the remainder of this section, it will become clear that \( \tilde{C}_N(z_1, z_2) \) is bounded in \( L^1 \), see Proposition 1.22. This completes the proof of Proposition 1.8.

**Remark 1.9.** Before we proceed further, note that the approximate covariance kernel \( \tilde{C}_N(z_1, z_2) \) in (1.20) is naturally expressed in terms of the auxiliary kernel

\[
K_N(z_1, z_2) := \sum_{k=1}^N E_{k-1}[E_k \delta^{k,1}_N(z_1) E_k \delta^{k,1}_N(z_2)]
\]
Then by Cauchy’s integral formula and analyticity, we can write the derivatives in (1.20) as

\[ C_N(z_1, z_2) := \frac{1}{d_N^2 (2\pi i)^2} \oint_{S_{z_1}} d\omega_1 \oint_{S_{z_2}} d\omega_2 \frac{s(\omega_1)s(\omega_2)}{(z_1 - \omega_1)^2(z_2 - \omega_2)^2} K_N(\omega_1, \omega_2) \]

where \( S_z \) is a small circle with center \( z \) and radius \( 1/(2d_N)\sqrt{(\tau_1 - \tau_2)^2 + (\eta_1 - \eta_2)^2} \). This ensures that for fixed \( \tau_1, \tau_2, \eta_1, \eta_2 \) and \( N \) large enough, \( S_{z_1} \) and \( S_{z_2} \) are disjoint sets. In the degenerate case that \( z_1 = z_2 \), it is enough to use the Cauchy integral formula with a single circle \( S_{z_1} \). If we obtain uniform estimates on \( K_N \) of the form \( K_N = \tilde{K}_N + O_{L^1}(u(N)) \), then the error in approximating \( C_N(z_1, z_2) \) is of the same order in \( N \):

\[
\mathbb{E} \left| \frac{1}{d_N^2 (2\pi i)^2} \oint_{S_{z_1}} d\omega_1 \oint_{S_{z_2}} d\omega_2 \frac{s(\omega_1)s(\omega_2)}{(z_1 - \omega_1)^2(z_2 - \omega_2)^2} |K_N - \tilde{K}_N| \right| \leq \frac{1}{(\tau_1 - \tau_2)^2 + (\eta_1 - \eta_2)^2} |u(N)|
\]

The conclusion of this remark is that it will be sufficient just to understand the convergence in \( L^1 \) of the kernel \( K_N(z_1, z_2) \).

Our first Lemma in this direction rewrites \( K_N(z_1, z_2) \) in terms of the matrix elements of the resolvent \( G_k(z) := (\mathcal{H}_k - z)^{-1} \). We will frequently make use of the shorthand notation \( G_k^{(p)} := G_k(z_p), p = 1, 2 \) to emphasize the dependence on the variables \( z_1 \) and \( z_2 \).

**Lemma 1.10.** The covariance kernel \( K_N(z_1, z_2) \) satisfies the exact identity

\[ K_N(z_1, z_2) = N^{-2} \sum_{k=1}^{N} \sum_{i<k,j<k} \mathbb{E}_k(G_k^{(1)}_{ij}) \mathbb{E}_k(G_k^{(2)}_{ji}) \]

\[ + N^{-2} \sum_{k=1}^{N} \sum_{i<k} \mathbb{E}_k(G_k^{(1)}_{ii}) \mathbb{E}_k(G_k^{(2)}_{ii}) \beta_{ik} \]

where \( \beta_{ik} \) is expressed in terms of the fourth moments

\[ \beta_{ik} := \mathbb{E}(|W_{ik}|^2 - 1)^2 \]

**Proof.** By definition we have

\[ \mathbb{E}_k(\Delta_N^{(1)}(z_1)) = N^{-1} \sum_{i<k, j<k} \mathbb{E}_k(G_k^{(1)}_{ij}) \overline{W_{ik}} W_{jk} - N^{-1} \sum_{j<k} \mathbb{E}_k(G_k^{(1)}_{jj}) = N^{-1} \sum_{i<k, j<k} \mathbb{E}_k(G_k^{(1)}_{ij}) \overline{W_{ik}} W_{jk} + N^{-1} \sum_{j<k} \mathbb{E}_k(G_k^{(1)}_{jj})(|W_{jk}|^2 - 1) \]
Multiplying out the resulting terms, we obtain
\[
E_{k-1}[E_k(\delta_{N}^{1,k}(z_1))E_k(\delta_{N}^{1,k}(z_2))]
\]
\[
= N^{-2}E_{k-1} \sum_{i<k,j<k, p<k, q<k, i\neq j, p \neq q} E_k(G_k^{(1)})_{ij} E_k(G_k^{(2)})_{pq} \overline{W}_{ik} W_{jk} \overline{W}_{pk} W_{qk}
\]
(1.24)
\[
+ N^{-2}E_{k-1} \sum_{i<k,j<k} E_k(G_k^{(1)})_{ij} \overline{W}_{ik} W_{jk} \sum_{j<k} E_k(G_k^{(2)})_{jj} (|W_{jk}|^2 - 1)
\]
(1.25)
\[
+ N^{-2}E_{k-1} \sum_{i<k,j<k} E_k(G_k^{(2)})_{ij} \overline{W}_{ik} W_{jk} \sum_{j<k} E_k(G_k^{(1)})_{jj} (|W_{jk}|^2 - 1)
\]
(1.26)
\[
+ N^{-2}E_{k-1} \sum_{i<k,j<k} E_k(G_k^{(1)})_{ii} E_k(G_k^{(2)})_{jj} (|W_{jk}|^2 - 1)(|W_{ik}|^2 - 1)
\]
(1.27)

It is clear that the sums (1.25) and (1.26) are identically zero, since the vector \(\{W_{ik}\}_{i \neq k}\) consists of centered independent random variables satisfying \(\mathbb{E}|W_{ik}|^2 = 1\). Similarly, the first summation (1.24) will be zero unless \(i = q\) and \(p = j\), this gives the first sum on the right-hand side of (1.22). The last term (1.27) will be zero unless \(i = j\), which gives the second sum in (1.23).

We now proceed with the estimation of the two sums in (1.22) and (1.23). To do this we need precise estimates on the resolvent matrix elements appearing in the sums.

**Lemma 1.11 (Bound on the resolvent).** Let \(z = E + \frac{\tau + in}{d_N}\) as above and consider an off-diagonal resolvent matrix element \((G_k(z))_{pq}\) with \(p \neq q\). Then for any positive integer \(s\) we have positive constants \(c, C\) such that
\[
\mathbb{E}|(G_k(z))_{pq}|^s \leq (Cs)^s \left(\frac{d_N}{\eta N}\right)^{s/2}
\]  
(1.28)

for all \(k, p, q, E, \tau \in \mathbb{R}, \eta > 0\) and \(d_N > 0\). For the diagonal matrix elements, the same result holds but with \((G_k(z))_{pp} = s(z)\) in place of \((G_k(z))_{pq}\).

**Proof.** This follows from Lemma 5.3 in [CMS14], see also previous works on the local semi-circle laws [ESY09b, ESY09a, EYY12a, EYY12b].

The second sum (1.23) over diagonal elements of the resolvent can now be dispensed with immediately:

**Lemma 1.12.** For all points \(z_1\) and \(z_2\) with non-zero imaginary part and on all mesoscopic scales \(1 \ll d_N \ll N\), we have the convergence in \(L^1\):
\[
\frac{1}{d_N^2} \frac{\partial^2}{\partial z_1 \partial z_2} \left[ s(z_1)s(z_2)N^{-2} \sum_{k=1}^{N} \sum_{i<k} E_k(G_k^{(1)})_{ii} E_k(G_k^{(2)})_{ii} \beta_{ik} \right] \to 0, \quad N \to \infty
\]
(1.29)

**Proof.** By Lemma 1.11 we have
\[
N^{-2} \sum_{k=1}^{N} \sum_{i<k} E_k(G_k^{(1)})_{ii} E_k(G_k^{(2)})_{ii} \beta_{ik} = N^{-2} s(z_1)s(z_2) \sum_{k=1}^{N} \sum_{i<k} \beta_{ik} + O_L(\sqrt{d_N/N})
\]
(1.30)
where we assumed that the fourth moment of each matrix entry \( W_{ik} \) is finite. Since the left-hand side of (1.29) is analytic, we can use the strategy outlined in Remark 1.9. Hence after inserting (1.30) into the left-hand side of (1.29), the \( O_L(\sqrt{d_N/N}) \) bound remains of the same order while the leading term is of order \( d_N^2 \) since \( s \) and \( s' \) are analytic and uniformly bounded provided \( E \in (-2+\delta, 2-\delta) \). This completes the proof of the Lemma.

\[ \square \]

**Remark 1.13.** This shows that in the mesoscopic regime, the process \( V_N(z) \) is insensitive to the value of the fourth cumulants of the matrix elements, provided they are finite. This is in contrast to the regime of global fluctuations where the fourth cumulant is known to appear explicitly in the limiting covariance formula, see e.g. [LP09a, BY05, Shc11]. This suggests that the Gaussian fluctuations obtained in the mesoscopic regime are more universal than in the global regime.

The remaining challenge now is to compute the first sum in (1.22). For that we shall show that to leading order on the scales \( 1 \ll d_N \ll N^{1/3} \), the sum

\[ S_1 := \frac{1}{N} \sum_{p<k,q<k} E_k(G_k(z_1))_{pq} E_k(G_k(z_2))_{qp} \tag{1.31} \]

satisfies a self-consistent equation. The proof relies on heavy computations.

**Calculation of the sum \( S_1 \):** Our goal is to prove the following

**Proposition 1.14.** Consider the quantity \( S_1 \) in (1.31). We have the estimate

\[ z_1 S_1 = -s(z_1) S_1 - \frac{k-1}{N} s(z_2) S_1 + O_L((d_N^3 N^{-1})^{1/2}) \tag{1.32} \]

where the \( O_L \) bound is uniform in \( k = 1, \ldots, N \).

In what follows, the proof of this Proposition will be divided into a number of Lemmas which eventually culminate in Lemma 1.21. We start with the Green function perturbation identity

\[ z_1(G_k(z_1))_{pq} = -\delta_{pq} + N^{-1/2} \sum_j (W_k)_{pj}(G_k)_{jq} \tag{1.33} \]

It turns out to be helpful to separate out the correlations between \( W_{ij} \) and \( G_k \) by introducing the matrix \( G_{kj} \) defined as the resolvent of the matrix \( \mathcal{H}_k \) with entries \((i,j)\) and \((j,i)\) replaced with 0. Simple algebra shows this is a perturbation of the original resolvent:

\[ G_k^{(1)} - G_{kj}^{(1)} = -N^{-1/2} G_{kj}^{(1)} c_{ij} ((W_k)_{ij} E_{ij} + (W_k)_{ji} E_{ji}) G_k^{(1)}, \tag{1.34} \]

where \( E_{ij} \) is the elementary matrix which is entirely zero, except entry \((i,j)\) which is equal to 1, while the coefficients \( c_{ij} = 1 \) if \( i \neq j \) and \( c_{ii} = 1/2 \). We now insert (1.34) directly into (1.33) leading to the following expansion.
Lemma 1.15. The matrix elements of the resolvent \((G^1_k)_{pq}\) satisfy
\[
z_1(G^1_k)_{pq} = -\delta_{pq} + N^{-1/2} \sum_j (W_k)_{pj}(G^1_{kp_j})_{jq} - s(z_1)(1 - 3/(2N))(G^1_k)_{pq} - \frac{1}{2} \sum_j c_{pj}((W_k)_{pj})^2 - 1)s(z_1)(G^1_k)_{pq}
\]
(1.35)

Now inserting (1.35) into the definition (1.31), we obtain the decomposition
\[
z_1S_1 = S_{1,1} + S_{1,2} + S_{1,3} + S_{1,4} + S_{1,5} + S_{1,6}
\]
where
\[
S_{1,1} = -N^{-1} \sum_{p < k} E_k (G^{(2)}_k)_{pp}
\]
(1.36)
\[
S_{1,2} = N^{-3/2} \sum_{j < k} \sum_{p < k, q < k} (W_k)_{pj} E_k (G^{(1)}_{kp})_{jq} E_k (G^{(2)}_k)_{qp}
\]
(1.37)
\[
S_{1,3} = -s(z_1)(1 - 3/(2N))S_1
\]
(1.38)
\[
S_{1,4} = -N^{-2} \sum_{p < k, q < k} c_{pj} E_k (W_k)_{pj}^2 ((G^{(1)}_{kp})_{jq} - s(z_1)(G^1_k)_{pq} E_k (G^{(2)}_k)_{qp}
\]
(1.39)
\[
S_{1,5} = -N^{-2} \sum_{p < k, q < k} c_{pj} E_k ((W_k)_{pj})^2 - 1)s(z_1)(G^1_k)_{pq} E_k (G^{(2)}_k)_{qp}
\]
(1.40)
\[
S_{1,6} = -N^{-2} \sum_{p < k, q < k} c_{pj} E_k (W_k)_{pj}^2 (G^{(1)}_{kp})_{jq} E_k (G^{(1)}_k)_{jq} E_k (G^{(2)}_k)_{qp}
\]
(1.41)

To estimate these sums we will apply Lemma 1.11 repeatedly, together with the following Lemma which allows us to bound the matrix elements of the perturbation \((G_{kij}(z))_{pq}\).

Lemma 1.16 (Bound on the perturbation). The same bound (1.28) of Lemma 1.11 holds with \((G_{kij}(z))_{pq}\) in place of \((G_k(z))_{pq}\).

Proof. We will use that higher moments of the entries \(W_{ij}\) are finite. Using the perturbation identity (1.34) we get
\[
(G_{kij})_{pq} = (G_k)_{pq} + c_{ij}(W_k)_{ij}N^{-1/2}(G_{kij})_{pq} + c_{ij}(W_k)_{ji}N^{-1/2}(G_{kij})_{pq} (G_k)_{ij}
\]
(1.42)

Iterating (1.42) once more and applying Minkowski's inequality shows that
\[
|(G_{kij})_{pq}|^s \leq c_s |(G_k)_{pq}|^s + c_s |(W_k)_{ij}|^s N^{-s/2} (|(G_{kij})_{pq}|^s + |(G_{kij})_{pq} (G_k)_{ij}|^s
\]
+ c_s |(W_k)_{ij}|^s N^{-s} (|(G_{kij})_{pq} (G_k)_{ji}|^s + |(G_{kij})_{pq} (G_k)_{ij}|^s
\]
+(G_{kij})_{pq} |(G_k)_{jq}|^s + |(G_{kij})_{pq} (G_k)_{ij}|^s)
\]

24
for some constant $c_s$ depending only on $s$. From the deterministic bound $|(G_{kij})_{pq}| \leq \|G_{kij}\| \leq d_N/\eta$ and Lemma 1.11, we find by Cauchy-Schwarz that

$$E|(G_{kij})_{pq}|^s \leq c_s E|(G_k)_{pq}|^s + O(N^{-s/2}) + O((d_N/\eta N)^s)$$

Note that the obtained error term is smaller than $(d_N/\eta N)^{s/2}$ found in (1.28). This completes the proof of the Lemma.

As is suggested by the structure of the terms (1.36)-(1.41), in what follows we will encounter many sums of the following generic form

$$\Omega := \sum_{\zeta_1, \zeta_2, \ldots, \zeta_r} \phi(W) \prod_{m \in I} E_k(G_k^{(m)})_{\zeta_m, \zeta_{m+1}}$$

for some index set $I$ and a complex valued function $\phi$. By simply counting the number of occurrences of $(G_k^{(m)})_{\zeta_m, \zeta_{m+1}}$ we obtain

**Lemma 1.17 (Trivial Bound).** Suppose that $E|\phi(W)|^2$ is uniformly bounded. Then

$$E|\Omega| \leq c N^r (d_N/N)^{1/2}$$

with the same result holding if any of the factors in (1.43) are replaced with the perturbation $G_k^{(m)}$ for any $a, b \in I$.

**Remark 1.18.** In practice we will apply a slightly more general result, including cases where $\phi$ or $G_k$ may or may not appear inside a conditional expectation. By conditional Jensen and Cauchy-Schwarz inequalities, the bound (1.44) continues to hold.

**Proof.** This follows by repeatedly applying the Cauchy-Schwarz inequality in conjunction with Lemmas 1.11 and 1.16.

We start by giving some first estimates on the terms $S_{1,1}$ to $S_{1,6}$ defined in (1.36)-(1.41).

**Lemma 1.19 (Estimates of $S_{1,1}$, $S_{1,3}$, $S_{1,4}$ and $S_{1,6}$).** We have the following bounds, holding uniformly in $k = 1, \ldots, N$:

$$S_{1,1} = -\frac{k-1}{N} s(z_2) + O_L((d_N N^{-1})^{1/2})$$

$$S_{1,3} = -s(z_1) S_1 + O_L(d_N/N)$$

$$E|S_{1,4}| \leq c \sqrt{d_N^3 N^{-1}}$$

$$E|S_{1,6}| \leq c \sqrt{d_N^3 N^{-1}}$$

**Proof.** These estimates are straightforward consequences of the trivial bound of Lemma 1.17.

With $S_{1,5}$ we have to take a bit more care because the trivial bound actually produces a bound of order $d_N$ which is divergent. To fix this we have to exploit independence and the fact that $E|(W_k)_{pq}|^2 = 1$. 

25
Lemma 1.20 (Estimate for $S_{1,5}$). We have the following bound, holding uniformly in $k = 1, \ldots, N$:

$$
E|S_{1,5}| \leq c\sqrt{d^2_N/N}
$$

Proof. First denote

$$
R_p := \sum_{q<k} (G^{(1)}_k)_{pq} E_k (G^{(2)}_k)_{qp}
$$

and note that the usual estimates imply that $E|R_p|^2 = O(d^2_N)$. Now by Cauchy-Schwarz we can estimate $S_{1,5}$ as

$$
E|S_{1,5}| \leq N^{-2} \sum_{p<k} \left( E|R_p|^2 \sum_{j_1,j_2} c_{p_{j_1}c_{p_{j_2}}} (|(W_k)_{p_{j_1}}|^2 - 1)(|(W_k)_{p_{j_2}}|^2 - 1) \right)^{1/2} 
$$

$$
= N^{-2} \sum_{p<k} \left( E|R_p|^2 \sum_j c_{p_j}^2 E(|(W_k)_{p_j}|^2 - 1)^2 \right)^{1/2} \leq c\sqrt{d^2_N/N}
$$

which is the desired result.

To complete the proof of Proposition 1.14 the main task is to control $S_{1,2}$. The problem is that the matrix $G^{(2)}_k$ still depends on $(W_k)_{p_{j_1}}$, so in the following we will replace $G^{(2)}_k$ with $G^{(2)}_{kp_{j_2}}$.

Lemma 1.21. We have the following bound, holding uniformly in $k = 1, \ldots, N$:

$$
S_{1,2} = -(k-1)N^{-1}s(z_2)S_1 + O_L1((d^3_N N^{-1})^{1/2})
$$

(1.49)

Proof. We start by replacing $(G^{(2)}_k)_{qp}$ with $(G^{(2)}_{kp_{j_2}})_{qp}$ in the definition (1.37) of $S_{1,2}$ and denote the modified sum by $\hat{S}_{1,2}$. We will show that $\hat{S}_{1,2}$ converges to 0 in $L^2$. We have

$$
E|\hat{S}_{1,2}|^2 = N^{-3} \sum_{j_1,j_2,p_{j_1},p_{j_2}} (W_k)_{p_{j_1}j_1}(W_k)_{p_{j_2}j_2} E_k(G^{(1)}_{kp_{j_1}})_{j_1q_1} E_k(G^{(2)}_{kp_{j_2}})_{q_1p_{j_1}}
$$

$$
\times E_k(G^{(1)}_{kp_{j_2}})_{j_2q_2} E_k(G^{(2)}_{kp_{j_2}})_{q_2p_{j_2}}
$$

(1.50)

where for the proof of this Lemma and unless otherwise stated, the summation indices run from 1 to $k-1$. To make $(W_k)_{p_{j_1}j_1}$ and $G_{kp_{j_2}}$ independent we use a similar perturbation formula to remove the matrix element $(W_k)_{p_{j_1}j_1}$ from $G_{kp_{j_2}}$. Therefore, we define $G^{(2),p_{j_1}j_1}_{kp_{j_2}}$ as the resolvent of the matrix $W$ with the $k$-th column and row erased and with entries $(W_k)_{p_{j_1}j_1}, (W_k)_{j_1p_{j_1}}, (W_k)_{p_{j_2}j_2}$ and $(W_k)_{j_2p_{j_2}}$ replaced with 0. It is easy to show that we have a similar identity

$$
(G^{(2)}_{kp_{j_2}})_{ab} = (G^{(2),p_{j_1}j_1}_{kp_{j_2}})_{ab} - c_{p_{j_1}j_1} (W_k)_{p_{j_1}j_1} N^{-1/2} (G^{(2),p_{j_1}j_1}_{kp_{j_2}})_{ap_{j_1}} (G^{(2)}_{kp_{j_2}})_{j_1b} 
$$

$$
- c_{p_{j_1}j_1} (W_k)_{j_1p_{j_1}} N^{-1/2} (G^{(2),p_{j_1}j_1}_{kp_{j_2}})_{aj_1} (G^{(2)}_{kp_{j_2}})_{p_{j_1}b}
$$

(1.51)

26
Then if we replace the last two factors in (1.50) with (1.51) the main term has identically zero expectation unless both \( p_1 = j_1 \) and \( p_2 = j_2 \), which we assume for the moment does not hold. The higher order terms in (1.51) give rise to an error of the form

\[
A = N^{-7/2} \mathbb{E} \sum_{j_1,j_2,p_1,p_2,q_1,q_2} (W_k)_{p_1j_1} (W_k)_{p_2j_2} \mathbb{E}_k(G^{(1)}_{k_{p_1j_1}})_{j_1q_1} \mathbb{E}_k(G^{(2)}_{k_{p_1j_1}})_{q_1p_1} \\
\times \mathbb{E}_k(G^{(1)}_{k_{p_2j_2}})_{j_2q_2} \mathbb{E}_k((W_k)_{p_1j_1}(G^{(2)}_{k_{p_2j_2}})_{q_2p_1}(G^{(2)}_{k_{p_2j_2}})_{j_1p_2})
\]

and four further error terms which do not differ in any important way from \( A \) above. To estimate \( A \), we now force independence in \((W_k)_{p_2,j_2}\) by replacing \( G^{(2)}_{k_{p_1j_1}}\) with \( G^{(2)}_{k_{p_2j_2}}\). Again the main term obtained by this replacement has expectation 0 because \( \mathbb{E}(W_k)_{p_2j_2} = 0 \). The error terms in (1.51) give rise to sums of the form

\[
A' = N^{-4} \mathbb{E} \sum_{j_1,j_2,p_1,p_2,q_1,q_2} (W_k)_{p_1j_1} (W_k)_{p_2j_2} \mathbb{E}_k(G^{(1)}_{k_{p_1j_1}})_{j_1q_1} \mathbb{E}_k((W_k)_{p_2j_2}(G^{(2)}_{k_{p_1j_1}})_{q_1p_1}(G^{(2)}_{k_{p_1j_1}})_{j_2p_1}) \\
\times \mathbb{E}_k(G^{(1)}_{k_{p_2j_2}})_{j_2q_2} \mathbb{E}_k((W_k)_{p_1j_1}(G^{(2)}_{k_{p_2j_2}})_{q_2p_1}(G^{(2)}_{k_{p_2j_2}})_{j_1p_2})
\]

We call such a term maximally expanded because we can no longer exploit independence of the different factors to reduce the size of the sum (none of the factors have zero expectation at this point). See also [EK13] for related methods. By the prescription (1.44), we have

\[
A' \leq cN^{-4}N^6(d_NN^{-1})^3 \leq cd^3_N/N
\]

It is clear that all error terms resulting from the replacement (1.51) give the same bounds after employing this procedure.

Now consider the diagonal terms \( p_1 = p_2 \) and \( j_1 = j_2 \) contributing to \( \tilde{\mathcal{E}}_{1,2} \):

\[
N^{-3} \mathbb{E} \sum_{j,p,q_1,q_2} |(W_k)_{pj}|^2 \mathbb{E}_k(G^{(1)}_{kpj})_{jq_1} \mathbb{E}_k(G^{(1)}_{kpj})_{q_1p} \mathbb{E}_k(G^{(2)}_{kpj})_{j_2q_2} \mathbb{E}_k(G^{(2)}_{kpj})_{q_2p}
\leq cN^{-3}N^4(d_NN^{-1})^2 \leq cd^2_N/N
\]

We conclude that \( \mathbb{E}|\tilde{\mathcal{E}}_{1,2}|^2 \leq cd^3_N/N \). Now again using (1.34) we have

\[
S_{1,2} = N^{-3/2} \sum_{j,p,q} (W_k)_{pj} \mathbb{E}_k(G^{(1)}_{kpj})_{jq} \mathbb{E}_k((G^{(2)}_{k})_{jp} - (G^{(2)}_{kpj})_{qp}) + O_L \left( \sqrt{d^3_N/N} \right)
\]

\[
= -N^{-2} \sum_{j,p,q} c_{pj} (W_k)_{pj}^2 \mathbb{E}_k(G^{(1)}_{kpj})_{jq} \mathbb{E}_k((G^{(2)}_{kpj})_{qp}(G^{(2)}_{k})_{jp})
\]  \hfill (1.52)

\[
= N^{-2} \sum_{j,p,q} c_{pq} (W_k)_{pj}^2 \mathbb{E}_k(G^{(1)}_{kpj})_{jq} \mathbb{E}_k((G^{(2)}_{kpj})_{pj}(G^{(2)}_{k})_{pp}) + O_L \left( \sqrt{d^3_N/N} \right) \]

\hfill (1.53)
By (1.44), we see that (1.52) is $O_{L^1}(\sqrt{d_N^3 N^{-1}})$. The final term (1.53) (let us denote it $\tilde{S}_{1,2}$) can be written

\begin{equation}
\tilde{S}_{1,2} = - N^{-2}s(z_2) \sum_{j,p,q} c_{pj} E_k(G_{k_{pj}}^{(1)})_{jq} E_k(G_{k_{pj}}^{(2)})_{qj} \tag{1.54}
\end{equation}

\begin{equation}
- N^{-2}s(z_2) \sum_{j,p} (|(W_k)_{pj}|^2 - 1)c_{pj} \sum_{q} E_k(G_{k_{pj}}^{(1)})_{jq} E_k(G_{kj}^{(2)})_{qj} \tag{1.55}
\end{equation}

\begin{equation}
- N^{-2} \sum_{j,p,q} c_{pj} (|(W_k)_{pj}|^2 E_k(G_{k_{pj}}^{(1)})_{jq} E_k(G_{kj}^{(2)})_{qj} (G_{kj}^{(2)})_{pp} - s(z_2)) \tag{1.56}
\end{equation}

To estimate the term (1.55), first replace $G^{(2)}$ with $G^{(2)}$ and note that by (1.44) the error in making this replacement is $O_{L^1}(\sqrt{d_N^3 N^{-1/2}})$. We denote the remaining sum $\tilde{S}_{1,2}$. We have

\begin{equation}
E[\tilde{S}_{1,2}]^2 = N^{-4}|s(z_2)|^2 \sum_{j_1,p_1,q_1} (|(W_k)_{p_1j_1}|^2 - 1) (|(W_k)_{p_2j_2}|^2 - 1) E_k(G_{k_{p_1j_1}}^{(1)})_{j_1q_1} E_k(G_{k_{p_2j_2}}^{(2)})_{q_1j_1} \tag{1.57}
\end{equation}

\begin{equation}
\times E_k(G_{k_{p_2j_2}}^{(1)})_{j_2q_2} E_k(G_{k_{p_2j_2}}^{(2)})_{q_2j_2} \tag{1.58}
\end{equation}

This sum has a similar structure to that seen in the computation of $E[\tilde{S}_{1,2}]^2$. Applying exactly the same procedure shows that $E[\tilde{S}_{1,2}]^2 \leq cN^3 / N^2$. The term (1.56) is $O_{L^1}(\sqrt{d_N^3 N^{-1}})$ as follows directly from the generic bound (1.44). In (1.54) all terms in the sum where $p = j$ will only contribute $O_{L^1}(d_N N^{-1})$ so can be neglected, which justifies us setting $c_{pj} = 1$ in (1.54).

Finally, note that (1.54) has a similar form to $S_1$ except with the perturbed resolvent elements $(G_{k_{pj}}^{(1)})_{jq}$. By the trivial bound (1.44) they can be exchanged with the original ones $(G_{kj}^{(1)})_{jq}$ at a cost $O_{L^1}(d_N N^{-1})).$ Then the summation over $p$ just gives a prefactor $(k - 1)$, leading to (1.49).

**Proof of Proposition 1.5.** With the most significant challenges dealt with above, our aim now is to complete the proof of Proposition 1.5 by solving relation (1.32) and computing the limit $N \to \infty$. In other words, we finally prove the conditional variance formula (1.15) which is enough to verify Proposition 1.5.

**Proposition 1.22.** Consider the approximate covariance kernel $\tilde{C}_N(z_1, z_2)$ given by equation (1.20), where $z_k = E + T_k + \eta_k$, we have the estimate

\begin{equation}
\tilde{C}_N(z_1, z_2) = \frac{1}{(i(\tau_1 - \tau_2 + i(\eta_1 + \eta_2)))^2} + O_{L^1}(1/d_N) + O_{L^1}\left(\log(d_N)\sqrt{d_N^3/N}\right) \tag{1.57}
\end{equation}

and

\begin{equation}
\tilde{C}_N(z_1, z_2) = O_{L^1}(1/d_N) + O_{L^1}\left(\log(d_N)\sqrt{d_N^3/N}\right) \tag{1.58}
\end{equation}

28
Proof. Recall that \( \tilde{C}_N(z_1, z_2) \) can be expressed in terms of the auxiliary kernel \( K_N(z_1, \overline{z}_2) \) in (1.22):

\[
\tilde{C}_N(z_1, z_2) = \frac{1}{d_N^2} \frac{\partial^2}{\partial z_1 \partial \overline{z}_2} s(z_1) s(\overline{z}_2) K_N(z_1, z_2)
\]

In turn, \( K_N(z_1, z_2) \) was expressed in terms of the sum \( S_1 \) defined in (1.31) and computed in Proposition 1.14. Let us denote the error term in that Proposition by \( E_{N,k} \). Then we have

\[
S_1 = \frac{k - 1}{N} s(z_1) s(z_2) + \frac{E_{N,k}}{1 - k^{-1} s(z_1) s(z_2)}
\]

where \( E_{N,k} = O_L(\sqrt{d_N^3/N}) \) uniformly in \( k \). Then by definition (1.31) the kernel \( K_N \) can be written as

\[
K_N(z_1, z_2) = \frac{1}{d_N^2} \sum_{k=1}^{N} \frac{k - 1}{N} s(z_1) s(z_2) + \frac{1}{N} \sum_{k=1}^{N} \frac{E_{N,k}}{1 - k^{-1} s(z_1) s(z_2)} + O_L(1/N)
\]

These sums are close to Riemann integrals, but before we calculate them we must take into account a subtle feature of the mesoscopic regime: when the points \( z_1 \) and \( z_2 \) have opposing signs in their imaginary parts, there is a singularity in the denominators of (1.59), due to the asymptotic formula

\[
1 - s(z_1)i \overline{s}(z_2) = \frac{\eta_1 + \eta_2 + (\tau_2 - \tau_1)i}{\sqrt{4 - r^2}} + O(d_N^{-2})
\]

If the signs of the imaginary part are the same, there is no singularity and one finds that the limiting covariance is identically zero, as in (1.58). Now let us control the errors in (1.59), assuming there is a singularity. For \( 0 < u < 1 \), let \( \psi(u) := |1 - us(z_1)i \overline{s}(z_2)|^{-1} \).

Then standard results about the Riemann integral show that the error term in (1.59) is bounded in \( L^1 \) by

\[
\sqrt{d_N^3/N} \sum_{k=1}^{N} \frac{1}{|1 - k^{-1} s(z_1) s(z_2)|} \leq \sqrt{d_N^3/N} \left( \int_0^1 \psi(u) du + \frac{\|\psi\|_{TV}}{N} \right)
\]

where \( \|\cdot\|_{TV} \) is the total variational norm of the function \( \psi \). Since \( us(z_1)i \overline{s}(z_2) \) has non-zero imaginary part, \( \psi(u) \) exists and is Riemann integrable. Then the total variational norm can be written \( \|\psi\|_{TV} = \int_0^1 |\psi'(u)| du \) and a simple calculation taking into account the asymptotics (1.60) shows that \( \|\psi\|_{TV} = O(d_N) \) as \( N \to \infty \), while \( \int_0^1 \psi(u) du = O(\log(d_N)) \) as \( N \to \infty \).

The error from approximating the first term in (1.59) can be estimated in the same way, and we get

\[
K_N(z_1, \overline{z}_2) = \int_0^1 \frac{us(z_1)i \overline{s}(z_2)}{1 - us(z_1)i \overline{s}(z_2)} du + O_L(\log(d_N) \sqrt{d_N^3/N})
\]

\[
= -1 - \log(1 - s(z_1)i \overline{s}(z_2)) + O_L(\log(d_N) \sqrt{d_N^3/N})
\]
Inserting (1.61) into the definition of $\tilde{C}_N(z_1, z_2)$ and bearing in mind Remark 1.9, we have

$$
\tilde{C}_N(z_1, z_2) := \frac{1}{d_N} \frac{\partial^2}{\partial z_1 \partial z_2} s(z_1) s(z_2) K_N(z_1, z_2)
$$

$$
= \frac{1}{d_N^2} \frac{s'(z_1) s'(z_2)}{1 - s(z_1) s(z_2)} - \frac{1}{d_N^2} s'(z_1) s'(z_2) + O_L \left( \log(d_N) \sqrt{\frac{d_N^3}{N}} \right)
$$

$$
= \frac{1}{(i(\tau_1 - \tau_2 + i(\eta_1 + \eta_2)))^2} + O(1/d_N) + O_L \left( \log(d_N) \sqrt{\frac{d_N^3}{N}} \right)
$$

where in the last line we applied formula (1.60) and used that

$$
s'(z_1) s'(z_2) = \frac{1}{4 - E^2} + O(d_N^{-1}).
$$

This completes the proof of Proposition 1.22. Therefore we have finally verified the conditions (1.15) and (1.16), hence completing the proof of Proposition 1.5.

§1.3 Tightness and linear statistics

The goal of this section is to extend the CLT obtained in the previous section to a wider range of test functions, focusing on the distributional convergence of (1.1). Our approach will be to combine the Helffer-Sjöstrand formula (1.13) with the main result of the last Section, telling us that the finite dimensional distributions of $V_N$ converge to $\Gamma_0^+$. If we can interchange the integrals in (1.13) with this distributional convergence then the CLT for mesoscopic linear statistics (1.1) would be proved.

In practice we will need to prove a certain tightness criteria in order to justify this interchange, involving uniform estimates on the second moment $E|V_N(\tau + i\eta)|^2$. We will handle those estimates mainly with Theorem 1.2, but this Theorem only applies in the bulk region with $\eta \geq d_N/N$. For smaller $\eta$ we need the following variance bound, due to Sosoe and Wong (in our notation).

**Proposition 1.23.** Let $\epsilon > 0$. Then for $0 < \epsilon < 1$, and $|E + \tau/d_N| \leq 5$ we have that there is a universal constant $C > 0$ such that:

$$
E|V_N(\tau + i\eta)|^2 \leq C d_N \eta^{-2-\epsilon},
$$

(1.62)

**Proof.** See Proposition 4.1 in [SW13].

The reader should compare this result with that obtained from Theorem 1.2 which gives an improved bound $E|V_N(t + i\eta)|^2 \leq C \eta^{-2}$ (Proposition B.4) but in a more restricted region.

To prove the CLT for (1.1), we consider two situations. Firstly we consider the case that $f \in C_c^{1, \alpha}(\mathbb{R})$, the Hölder space of compactly supported functions $f$ such that $f'$ is Hölder continuous with exponent $\alpha > 0$. The hypothesis of compact support will allow us to avoid complications coming from the edges of the spectrum and will serve as a warm up, illustrating our general approach.
In the last subsection we consider the more challenging case where the hypothesis of compact support is replaced with a more general decay condition \( f(x) \) and \( f'(x) \). In particular we will prove our main result, Theorem 1.1.

**Compactly supported functions with Hölder continuous first derivative:** To obtain the CLT we will apply the Helffer-Sjöstrand formula (1.13) with

\[
\Psi_f(t, \eta) = (f(t) + i(f(t + \eta) - f(t)))J(\eta)
\]

where \( J(\eta) \) is a smooth function of compact support, equal to 1 in a neighborhood of \( \eta = 0 \) and equal to 0 if \( \eta > 1 \), see Lemma C.1. It follows that

\[
\frac{\partial \Psi_f(t, \eta)}{\partial \eta} = (f'(t) - f'(t + \eta) + i(f'(t + \eta) - f'(t)))J(\eta) + (if(t) - (f(t + \eta) - f(t)))J'(\eta)
\]

and if we assume that \( f' \) is Hölder continuous with exponent \( 0 < \alpha < 1 \) then we also have the bound \( |\frac{\partial \Psi_f(t, \eta)}{\partial \eta}| \leq c\eta^\alpha \) for some constant \( c > 0 \) for \( \eta \) small. Also \( \Psi_f(t, \eta) \) satisfies \( \Psi_f(t, 0) = f(t) \), and that \( \frac{\partial \Psi_f(t, \eta)}{\partial \eta} \) is compactly supported whenever \( f \) is.

**Theorem 1.24.** Suppose that \( d_N \) satisfies the condition \( 1 \ll d_N \ll N^{1/3} \) and consider compactly supported test functions \( f_1, \ldots, f_M \) whose first derivatives are Hölder continuous for some exponent \( \alpha > 0 \). Then for any fixed \( E \in (-2, 2) \) in (1.1) we have the convergence in distribution

\[
(X^{\text{meso}}_{N}(f_1), \ldots, X^{\text{meso}}_{N}(f_M)) \to (X_1, \ldots, X_M)
\]

where \( (X_1, \ldots, X_M) \) is an \( M \)-dimensional Gaussian vector with zero mean and covariance matrix

\[
E(X_pX_q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |k| \hat{f}_p(k)\bar{\hat{f}}_q(k) \, dk, \quad 1 \leq p, q \leq M \quad (1.63)
\]

**Proof.** Since \( f \) has compact support, we may write

\[
\hat{X}^{\text{meso}}_{N}(f) = \frac{1}{\pi} \Re \int_0^1 \int_{\tau-}^{\tau+} V_N(\tau + i\eta)\bar{\frac{\partial \Psi_f(\tau, \eta)}{\partial \eta}} \, d\tau \, d\eta
\]

for some fixed \( \tau_- \) and \( \tau_+ \). First note that the region \( 0 \leq \eta \leq d_N/N \) can be neglected, since

\[
\int_0^{d_N/N} \int_{\tau-}^{\tau+} E(\frac{|V_N(\tau + i\eta)|\frac{\partial \Psi_f(\tau, \eta)}{\partial \eta}}{\eta^\alpha}) \, d\tau \, d\eta \leq c\eta^\alpha \int_0^{d_N/N} \int_{\tau-}^{\tau+} \eta^{\alpha-1-\epsilon} \, d\tau \, d\eta (1.64)
\]

The latter goes to zero after choosing \( 0 < \epsilon < (1-\gamma)\alpha \), where \( d_N = N^\gamma \) with \( 0 < \gamma < 1 \). Therefore, it suffices to study the convergence in distribution of

\[
X(V_N, f) = \frac{1}{\pi} \Re \int_0^1 \int_{\tau-}^{\tau+} V_N(\tau + i\eta)\chi_N(\eta)\bar{\frac{\partial \Psi_f(\tau, \eta)}{\partial \eta}} \, d\tau \, d\eta
\]

where \( \chi_N(\eta) \) is an indicator function, equal to 1 on the region \( d_N/N \leq \eta \leq 1 \) and 0 otherwise. We wish to show that \( X(V_N, f) \) converges in distribution to \( X(\Gamma_0^+, f) \). Since
we proved in the previous section that the finite dimensional distributions of $V_N(\tau + i\eta)$ converge pointwise to $\Gamma_0^+(\tau + i\eta)$, we can appeal to Theorem A.2. Let $\Phi$ denote the set of functions $\phi(z, w) : H \times C \to C$ of the form $\phi(z, w) = \overline{\partial \Psi}_g(z) w$ where $g$ is a function in the class stated by Theorem 1.24 and $M$ is as in the Theorem A.2.

Let $D$ be the domain $[0, 1] \times [-\tau_+, \tau_+]$. Then Theorem A.2 guarantees the convergence in distribution $(X(V_N, f_1), \ldots, X(V_N, f_M)) \Rightarrow (X(\Gamma_0^+, f_1), \ldots, X(\Gamma_0^+, f_M))$ provided we check the following tightness conditions:

\[
\inf_{B \subset H, \lambda(B) < \infty} \limsup_{N \to \infty} \mathbb{P}\left( \frac{1}{\pi} \int_{D \setminus B} |V_N(\tau + i\eta)\chi_N(\eta)\overline{\partial \Psi}_f(\tau, \eta)| \, d\tau \, d\eta \geq \epsilon \right) = 0 \quad (1.65)
\]

\[
\lim_{K \to \infty} \limsup_{N \to \infty} \mathbb{P}\left( \frac{1}{\pi} \int_D |(V_N(\tau + i\eta)\chi_N(\eta)\overline{\partial \Psi}_f(\tau, \eta) - K)^+| \, d\tau \, d\eta \geq \epsilon \right) = 0, \quad (1.66)
\]

First we prove (1.65). By Markov’s inequality it suffices to check that

\[
\inf_{B \subset D, \lambda(B) < \infty} \limsup_{N \to \infty} \int_{D \setminus B} \mathbb{E}[|V_N(\tau + i\eta)|\chi_N(\eta)|\overline{\partial \Psi}_f(\tau, \eta)|] \, d\tau \, d\eta = 0 \quad (1.67)
\]

Now since $\tau$ is fixed, the real part appearing in the denominator of the resolvent is bounded away from the edges (due to the compact support of $f$). Also the imaginary part appearing in the denominator is no less than $1/N$. Hence we can apply Proposition B.4 and obtain $\mathbb{E}[|V_N(\tau + i\eta)|] \leq c\eta^{-1}$. By the assumptions on $f$, $\eta^{-1}||\overline{\partial \Psi}_f(\tau, \eta)||$ is integrable on $D \setminus B$, and so (1.67) is bounded by

\[
\inf_{B \subset D, \lambda(B) < \infty} \int_{D \setminus B} c\eta^{-1}|\overline{\partial \Psi}_f(\tau, \eta)| \, d\tau \, d\eta = 0
\]

where the last equality follows from the dominated convergence theorem. To check (1.66) we proceed similarly, noting that for any $\delta > 0$, we have

\[
(|V_N(\tau + i\eta)\chi_N(\eta)\overline{\partial \Psi}_f(\tau, \eta)| - K)^+ \leq \frac{1}{K^\delta} |V_N(\tau + i\eta)\chi_N(\eta)\overline{\partial \Psi}_f(\tau, \eta)|^{1+\delta}
\]

Then (1.66) is bounded by

\[
\lim_{K \to \infty} K^{-\delta} \limsup_{N \to \infty} \int_D \mathbb{E}[|V_N(\tau + i\eta)|^{1+\delta}\chi_N(\eta)|\overline{\partial \Psi}_f(\tau, \eta)|^{1+\delta}] \, d\tau \, d\eta
\]

\[
\leq \lim_{K \to \infty} K^{-\delta} \int_D c_1 \eta^{-1-\delta}||\overline{\partial \Psi}_f(\tau, \eta)||^{1+\delta} \, d\tau \, d\eta = 0 \quad (1.68)
\]

where in (1.68), we choose $\delta$ so that $0 < \delta < \alpha/(1 - \alpha)$ so that the above integral is finite by the behavior $||\overline{\partial \Psi}_f(\tau, \eta)|| \leq c\eta^{-\alpha}$ for small $\eta$. This completes the proof that $X(V_N, f)$ converges in distribution to the random variable

\[
X(\Gamma_0^+, f) = \frac{1}{\pi} \Re \int_D \Gamma_0^+(\tau + i\eta)\overline{\partial \Psi}_f(\tau + i\eta) \, d\tau \, d\eta,
\]

Since the integral of a Gaussian process is Gaussian, it just remains to compute the co-variance and verify it is well-defined, see the next Lemma.  

32
Lemma 1.25. Consider the random functional defined by:

\[ X(\Gamma_0^{t+}, f) = \frac{1}{\pi} \mathcal{R} \int_{t}^{T} \Gamma_0^{t+} (\tau + i\eta) \overline{\partial \Psi}_f (\tau + i\eta) \, d\tau \, d\eta \]  

(1.69)

Then the covariance \( \mathbb{E} X(\Gamma_0^{t+}, f_1) X(\Gamma_0^{t+}, f_2) \) is given by (1.4), provided that \( f_1 \) and \( f_2 \) are in \( C^{1, \alpha}(R) \) for some \( \alpha > 0 \), and \( |f_i|, |f'_i| \) are \( O(|x|^{-(1+\beta)}) \) for some \( \beta > 0 \).

Proof. Recall that the process \( \Gamma_0^{t+}(\tau + i\eta) \) for \( \eta > 0 \) appearing in (1.69) has the covariance structure

\[ \mathbb{E}(\Gamma_0^{t+}(\tau_1 + i\eta_1) \overline{\Gamma_0^{t+}}(\tau_2 + i\eta_2)) = \frac{1}{(i(\tau_1 - \tau_2 + i(\eta_1 + \eta_2)))^2} \]  

(1.70)

while \( \mathbb{E}(\Gamma_0^{t+}(\tau_1 + i\eta_1) \overline{\Gamma_0^{t+}}(\tau_2 + i\eta_2)) = 0 \). We compute

\[ \mathbb{E}(X(\Gamma_0^{t+}, f_1), X(\Gamma_0^{t+}, f_2)) = \frac{1}{2\pi^2} \mathbb{E} \left[ \int_{\mathbb{H} \times \mathbb{H}} \, dz_1 \, dz_2 \left\{ \mathcal{R} \left[ \Gamma_0^{t+}(z_1) \Gamma_0^{t+}(z_2) \overline{\partial \Psi}_{f_1}(z_1) \overline{\partial \Psi}_{f_2}(z_2) \right] \right\} \right] \]  

(1.71)

\[ + \mathbb{R} \left[ \Gamma_0^{t+}(z_1) \overline{\partial \Psi}_{f_1}(z_1) \Gamma_0^{t+}(z_2) \overline{\partial \Psi}_{f_2}(z_2) \right] \]  

(1.72)

we proceed to check that the expectation is finite and that we can exchange expectation with integrals by utilizing Fubini’s theorem. It suffices to verify that

\[ \int_{\mathbb{H} \times \mathbb{H}} \, dz_1 \, dz_2 \mathbb{E}|\Gamma_0^{t+}(z_1) \overline{\Gamma_0^{t+}}(z_2)| |\overline{\partial \Psi}_{f_1}(z_1)| |\overline{\partial \Psi}_{f_2}(z_2)| < \infty, \]

by using Cauchy-Schwarz, the above integral is bounded by

\[ \int_{\mathbb{H} \times \mathbb{H}} \, dz_1 \, dz_2 \eta_1^{-1} \eta_2^{-1} |\overline{\partial \Psi}_{f_1}(z_1)| |\overline{\partial \Psi}_{f_2}(z_2)| < \infty, \]

since by our assumptions on \( f_i \), we have

\[ |f'_i(t + \eta) - f'_i(t)| \leq \min(C_1 \eta^{\alpha}, C_2 |t|^{-(1+\beta)}) \leq c \eta^{\alpha \sigma} |t|^{-(1+\beta)(1-\sigma)}, \]

for any \( \sigma \in (0, 1) \) and \( c > 0 \) a constant independent of \( t \) and \( \eta \), giving us the bound

\[ |\overline{\partial \Psi}_{f_i}(t, \eta)| \leq c \eta^{\alpha \sigma} |t|^{-(1+\beta)(1-\sigma)}, \]

taking \( \sigma < \beta/(1 + \beta) \) gives a bound that is integrable (recall that \( \overline{\partial \Psi}_{f_i}(t, \eta) \) is supported on \( \eta \in [0, 1] \)). Hence, the covariance of \( X(\Gamma_0^{t+}, f) \) is finite, which implies, in particular that \( X(\Gamma_0^{t+}, f) \) is finite almost-surely — yielding that the process is well-defined as required.

After taking the expectation inside the integral, the first term (1.71) vanishes identically. In the second term (1.72), we write the covariance in the Sobolev form

\[ \mathbb{E} \Gamma_0^{t+}(\tau + i\eta) \overline{\Gamma_0^{t+}}(\tau + i\eta) = \frac{1}{2\pi} \int_{\mathbb{R}} \, dk |k| r_{\tau_1, \eta_1}(k) r_{\tau_2, \eta_2}(k) \]

33
where \( r_{\tau_1,\eta_1}(x) = (x - \tau_1 - i\eta_1)^{-1} \). We now interchange the integration over \( k \) with the integration over \( H \times H \). To justify this, note that \( r_{\tau_1,\eta_1}(k) = -2\pi ie^{-ik\tau_1}e^{-|k|\eta_1} \) and due to the conditions on \( f_1, f_2 \), we have the bound:

\[
\int_{H \times H} dz_1 dz_2 \int_{\mathbb{R}} dk |\overline{\partial \Psi_{f_1}(z_1)}||\overline{\partial \Psi_{f_2}(z_2)}||k|e^{-|k|\eta_1}e^{-|k|\eta_2} \leq C \int_0^1 \int_0^1 d\eta_1 d\eta_2 \frac{(\eta_1 \eta_2)^{\alpha\sigma}}{(\eta_1 + \eta_2)^2} < \infty.
\]

After interchanging these integrals, the integration over \( H \times H \) factorizes as a product. Now we would like to interchange the Fourier transform with the integral over \( H \), for which it suffices to bound the following

\[
\int_0^1 d\eta_1 \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} d\tau_1 \left| \frac{\overline{\partial \Psi_f(\tau_1, \eta_1)}}{x - \tau_1 - i\eta_1} \right| = \int_0^1 d\eta_1 \|g_{\eta_1} * h_{\eta_1}\|_1
\]

where \( g_{\eta_1}(\tau_1) = |\tau_1 + i\eta_1|^{-1} \) and \( h_{\eta_1}(\tau_1) = |\overline{\partial \Psi_f(\tau_1, \eta_1)}| \). To bound the \( L^1 \) norm of the convolution, we apply Young’s inequality \( \|g_{\eta_1} * h_{\eta_1}\|_1 \leq \|g_{\eta_1}\|_p \|h_{\eta_1}\|_q \) with \( q = 1 - \delta \) and \( p = 1 + \delta/(1 - 2\delta) \) with \( \delta > 0 \). A simple computation shows that \( \|g_{\eta_1}\|_p \leq c\eta_1^{-\delta/(1-\delta)} \) while for sufficiently small \( \delta \), \( \|h_{\eta_1}\|_q \) is bounded uniformly in \( \eta_1 \) due to the integrability assumptions on \( f \) and its derivatives. This shows that \( \|g_{\eta_1} * h_{\eta_1}\|_1 \leq c\eta_1^{-\delta/(1-\delta)} \) so that (1.73) is finite.

After performing all such interchanges of integration, we finally obtain

\[
E(X(\Gamma_0^+, f_1)X(\Gamma_0^+, f_2)) =
\frac{1}{8\pi} \int_{-\infty}^{\infty} |k| dk \left\{ \int_{-\infty}^{\infty} dx e^{-ikx} \frac{1}{\pi} \int_{\mathbb{H}} dz_1 \frac{1}{x - z_1} \overline{\partial \Psi_{f_1}(z_1)} \right.
\times \left. \int_{-\infty}^{\infty} dx e^{-ikx} \frac{1}{\pi} \int_{\mathbb{H}} dz_2 \frac{1}{x - z_2} \overline{\partial \Psi_{f_2}(z_2)} \right\} + \text{c.c.},
\]

where c.c. refers to the complex conjugate of the term preceding it.

Now the inner integrals over \( \mathbb{H} \) can be evaluated by Lemma C.1. There is a caveat however, Lemma C.1 requires the function \( f \) to be compactly supported. We remedy this by taking our function \( f \) and multiplying it by a cutoff function \( \phi_n = \phi(x/n) \) where \( \phi(x) \) is 1 on \([-1, 1]\) and vanishes outside \([-2, 2]\), we let \( f_n = \phi_n f \). By Lemma C.1 we have the identity

\[
\frac{1}{\pi} \int_0^\infty d\eta_1 \int_{-\infty}^\infty d\tau_1 \frac{1}{x - \tau_1 - i\eta_1} \overline{\partial \Psi_{f_n}(\tau_1, \eta_1)} = f_n(x) + iH[f_n](x).
\]

It is well known that \( H \) is a bounded operator from \( L^2(\mathbb{R}) \) to itself, therefore if we take the limit as \( n \to \infty \) on both sides of (1.75) (and note that \( f_n \to f \) pointwise everywhere and in \( L^2(\mathbb{R}) \)) we have

\[
\lim_{n \to \infty} \frac{1}{\pi} \int_0^\infty d\eta_1 \int_{-\infty}^\infty d\tau_1 \frac{1}{x - \tau_1 - i\eta_1} \overline{\partial \Psi_{f_n}(\tau_1, \eta_1)} = f(x) + iH[f](x),
\]

34
we check that the limit interchanges with integral by noting that
\[
|\tilde{\phi}_f(t, \eta)| \leq 2|f'_n(t) - f'_n(t + \eta)||J(\eta)| + 2(|f(t) - f(t + \eta)||J(\eta)|,
\]
\[
|f'(t) - f'_n(t + \eta)| \leq |\phi'_n(t + \eta)||f(t + \eta) - f(t)| + |\phi'_n(t) - \phi'_n(t)||f(t)|
\]
\[
+ |\phi'_n(t + \eta)||f'(t + \eta) - f'(t)| + |f'(t)||\phi'_n(t + \eta) - \phi'_n(t)|,
\]
and that \( \phi_n(t) = \phi(t/n) \) is infinitely differentiable and therefore in \( C^{k, \alpha}(\mathbb{R}) \) for any \( k \in \mathbb{N}, \alpha \in (0, 1) \). So we may bound for large \( t \), \( |\tilde{\phi}_f(t, \eta)| \) by \( C|\eta^{\alpha}\sigma|^{-1/2} \) for some constant \( C \) and any \( \sigma \in (0, 1) \), it follows by the dominated convergence theorem that we may interchange limit with integral to obtain
\[
\frac{1}{\pi} \int_0^\infty d\eta_1 \int_{-\infty}^\infty d\tau_1 \frac{1}{x - \tau_1 - i\eta_1} \tilde{\phi}_f(\tau_1, \eta_1) = f(x) + iH[f](x).
\]
(1.76)

The Fourier transform of the Hilbert transform is given by
\[
\hat{H}[f](k) = -i \text{sgn}(k) \hat{f}(k).
\]
(1.77)

Hence, inserting (1.76) into (1.74) and applying (1.77) yields the limiting covariance structure
\[
\mathbb{E}(X(f_1)X(f_2)) = \frac{1}{8\pi} \int_{-\infty}^\infty dk |k| \hat{f}_1(k) \hat{f}_2(k) |1 - i \text{sgn}(k)|^2 dk + \text{c.c.}
\]
\[
= \frac{1}{2\pi} \int_{-\infty}^\infty dk |k| \hat{f}_1(k) \hat{f}_2(k)
\]
as required.

**Functions supported on the real line** The main goal of this subsection is to remove the assumption of compact support from the functions \( f \) in Theorem 1.24 subject to the following decay condition on \( f \) and \( f' \): for some \( \beta > 0 \) and \( |x| \) large enough, \( f(x) \) and \( f'(x) \) are \( O(|x|^{-1-\beta}) \). This will complete the proof of the corresponding statement in our main Theorem 1.1. We begin by approximating \( f \) by a compactly supported function whose support grows at a rate \( O(d_N) \) as \( N \to \infty \). The support of the test function can now extend over the edges of the spectrum, so that the resolvent bounds of Proposition B.4 cannot be applied. The goal of this subsection is to prove the following

**Theorem 1.26.** Let \( d_N = N^\gamma \) with \( 0 < \gamma < 1/3 \). If for some \( \alpha > 0 \) and \( \beta > 0 \), we have \( f \in C^{1, \alpha}(\mathbb{R}) \) where \( f(x) \) and \( f'(x) \) decay faster than \( |x|^{-1-\beta} \) for large \( |x| \), then the random variable \( X_N^{\text{meso}}(f) \) converges in distribution to a Gaussian random variable with variance given by (1.63) of Theorem 1.24. Moreover, the multidimensional version stated in Theorem 1.24 continues to hold for functions of this class.

We will prove this Theorem by means of the following Lemma and two Propositions.

**Lemma 1.27.** Let \( \phi_N(x) \) denote a smooth cutoff function equal to 1 in when \( |E + x/d_N| \leq 2 \) and equal to 0 when \( |E + x/d_N| \geq 4 \). Let \( f_N(x) := f(x)\phi_N(x) \). Then \( X_N^{\text{meso}}(f) = X_N^{\text{meso}}(f_N) + o_L^2(1) \).
Proof. We follow the same technique here as in Section 4 of [SW13]. Note that $\hat{X}_N^{\text{meso}}((1 - \phi_N)f)$ is only non-zero when $\lambda_1 < -4$ or $\lambda_N > 4$, so that

$$P \left( \left| \sum_i ((1 - \phi_N)f)(d_N \lambda_i) \right| > 0 \right) \leq P (\lambda_1 > 4) + P (\lambda_N < -4),$$

these quantities are bounded by $e^{-N^c}$ by [EYY12a, Lemma 7.2]. Thus

$$E|X_N^{\text{meso}}((1 - \phi_N)f)|^2 = 2 \int_{-N^c}^{N^c} \int_{\mathbb{R}} x P \left( \left| \sum_i (1 - \phi_N)f(d_N \lambda_i) \right| \geq x \right) dx,$$

$$\leq N^2 \|f\|_L^2(R)P \left( \left| \sum_i (1 - \phi_N)f(d_N \lambda_i) \right| > 0 \right),$$

$$\leq N^2 e^{-N^c} \|f\|_L^2(R) \to 0.$$

Now we apply the Helffer-Sjöstrand formula (1.13) to the function $f_N$, obtaining

$$\hat{X}_N^{\text{meso}}(f_N) = \frac{1}{\pi} \Re \int_0^1 \int_{\mathbb{R}} V_N(\tau + i\eta)\tilde{\chi}_N(\tau)\bar{\partial}\psi_{f_N}(\tau, \eta) d\tau d\eta$$

where $\tilde{\chi}_N(\tau)$ is the indicator function of the region $|E + \tau/d_N| \leq 4$. Repeating the derivation of (1.64), we see that

$$\int_0^{d_N/N} \int_{\mathbb{R}} E(|V_N(\tau + i\eta)|\tilde{\chi}_N(\tau))|\bar{\partial}\psi_{f_N}(\tau, \eta)| d\tau d\eta \to 0$$

on all scales of the form $d_N = N^\gamma$ with $0 < \gamma < 1$. To apply Theorem A.2, we replace the $N$-dependent $f_N$ with $f$, noting that $\bar{\partial}\psi_{f_N} - \bar{\partial}\psi_f = \bar{\partial}\psi_{f_N - f}$, where $f_N(\tau) - f(\tau) = f(\tau)(\phi_N(\tau) - 1)$ is supported on the region $|\tau/d_N - E| \geq 4$.

**Proposition 1.28.** We have

$$\frac{1}{\pi} \Re \int_0^1 \int_{\mathbb{R}} V_N(\tau + i\eta)\bar{\partial}\psi_{f_N - f}(\tau, \eta) d\tau d\eta = o_{L^1}(1). \quad (1.78)$$

We postpone the proof of this Proposition until the end of this subsection, where it will follow from a more general argument. Thus $X_N^{\text{meso}}(f) = I(f, V_N) + o_{L^1}(1)$ where

$$I(f, V_N) = \frac{1}{\pi} \Re \int_0^1 \int_{-\infty}^{\infty} V_N(\tau + i\eta)\chi_N(\eta)\tilde{\chi}_N(\tau)\bar{\partial}\psi_f(\tau, \eta) d\tau d\eta.$$

To show that $I(f, V_N)$ converges to $I(f, \Gamma^{'\dagger}_0)$, thus completing the proof of Theorem 1.26, it remains to check the following tightness result.
Proposition 1.29. Consider the domain $D = [0, 1] \times \mathbb{R}$. We have the following estimates:

\begin{equation}
\inf_{B \subset H, \lambda(B) < \infty} \limsup_{N \to \infty} \int_{D \setminus B} \mathbb{E}(|V_N(t + i\eta)|\chi_N(\eta)) \varphi_N(\tau))|\overline{\partial} \Psi_f(t, \eta)| \, d\eta \, dt = 0 \tag{1.79}
\end{equation}

\begin{equation}
\lim_{K \to \infty} \lim_{N \to \infty} \int_D \mathbb{E}(|V_N(t + i\eta)|^{1+\delta}) \chi_N(\eta) \varphi_N(\tau))|\overline{\partial} \Psi_f(t, \eta)|^{1+\delta} \, d\eta \, dt = 0. \tag{1.80}
\end{equation}

Proof. First we give some bounds on $|\overline{\partial} \Psi_f(t, \eta)|$. By our assumptions on $f$, we have

$$|f'(t + \eta) - f'(t)| \leq \min(C_1\eta^{\alpha}, C_2|t|^{-(1+\beta)}) \leq c\eta^{\alpha}|t|^{-(1+\beta)(1-\sigma)},$$

for any $\sigma \in (0, 1)$ and $c > 0$ a constant independent of $t$ and $\eta$. Hence by construction we have the bound

$$|\overline{\partial} \Psi_f(t, \eta)| \leq c\eta^{\alpha}|t|^{-(1+\beta)(1-\sigma)} \tag{1.81}$$

for large $|t|$ and small $\eta$. We proceed by splitting the integration in (1.79) and (1.80) into the regions $D_{\text{bulk}} = [0, 1] \times \{\tau : |E + \tau/d_N| \leq 2\}$ and $D_{\text{out}} = [0, 1] \times \{\tau : 2 \leq |E + \tau/d_N| \leq 4\}$. Starting with region $D_{\text{bulk}}$, the variance bound of Proposition B.4 is applicable and we see that $\mathbb{E}|V_N(t + i\eta)|^{1+\delta} \leq c_1\eta^{-(1+\delta)}$ uniformly in $t$. Combined with (1.81) we see that the integrands of (1.79) and (1.80) restricted to $D_{\text{bulk}}$ are dominated by an integrable function and the proof proceeds as in the compactly supported case of Theorem 1.24.

It remains to bound the contribution to the integrals on the domain $D_{\text{out}}$. Here we can exploit the decay of the test function $f$ and show that the inner lim sup over $N$ will already be zero in (1.79) and (1.80). Therefore it suffices to take $B = \emptyset$ and $K = 1$. Then the variance bound of Proposition 1.23 applies, yielding

\begin{align*}
\int_{D_{\text{out}}} & \mathbb{E}(|V_N(z)|^{1+\delta}) |\overline{\partial} \Psi_f(t, \eta)|^{1+\delta} \, dt \, d\eta \\
& \leq cd_N^{(1+\delta)} \int_{D_{\text{out}}} \eta^{-(1-\epsilon)(1+\delta) + \sigma\alpha}|t|^{-(1-\epsilon)(1-\sigma)(1+\delta)} \, dt \, d\eta \\
& \leq cd_N^{(1+\delta)-(1+\beta)(1-\sigma)(1+\delta)+1} \tag{1.82}
\end{align*}

where we choose $\epsilon$ and $\delta$ small so that the integral over $\eta$ is finite while the integral over $t$ goes to zero as $N \to \infty$. Indeed, if $\epsilon < \sigma\alpha/3$ and $\delta < \sigma\alpha/3$ then $-(1-\epsilon)(1+\delta) + \sigma\alpha > -1$ and the $\eta$ integral is finite. In the integral over $t$, we choose $\sigma < \delta/(1+\delta)$, $\epsilon < \beta/(1+\delta)$ and deduce that $\epsilon(1+\delta) -(1+\beta)(1-\sigma)(1+\delta)+1 < 0$. To make the bounds work simultaneously we take $\epsilon < \min\{\sigma\alpha/3, \beta/(1+\delta)\}$. We conclude that the limit of (1.82) is zero. To prove Proposition 1.28 notice that the integrand is supported on $D_{\text{out}}$ with the same regularity conditions on $f$. Hence an identical calculation to that given in (1.82) shows that (1.78) converges to zero in $L^1$ as $N \to \infty$. This completes the proof of Propositions 1.29 and 1.28. Consequently, by means of Theorem A.1, this also completes the proof of Theorem 1.26.
Corollary 1.30. The sequence of stochastic processes $V_N(z)$ with $z \in \mathbb{H}$ is tight in the space of continuous functions on any $N$-independent rectangle in the upper half-plane $\mathbb{H}$.

Proof. It suffices to verify the Arzela-Ascoli criterion:

$$E|V_N(z_1) - V_N(z_2)|^2 \leq C|z_1 - z_2|^2,$$

for $z_1 = u_1 + iv_1$ and $z_2 = u_2 + iv_2$ in some $N$-independent bounded rectangle in $\mathbb{R} \subset \mathbb{H}$ and $C$ a constant depending only on the vertices of the rectangle (i.e. not on $N$). To prove this, note that

$$\frac{1}{x - u_1 - iv_1} - \frac{1}{x - u_2 - iv_2} = \frac{(u_1 - u_2) + i(v_1 - v_2)}{(x - u_1 - iv_1)(x - u_2 - iv_2)}$$

implies

$$E|V_N(z_1) - V_N(z_2)|^2 = |z_1 - z_2|^2 E|\tilde{X}_N^{meso}(h)|^2$$

where $h(x) = ((x - u_1 - iv_1)(x - u_2 - iv_2))^{-1}$ is a smooth function with decay $h(x) = O(|x|^{-2})$ as $|x| \to \infty$. Hence the techniques of the present subsection are applicable with $\alpha = \beta = 1$. Indeed, after replacing $h$ with a smooth cut-off $h_N$ as in Lemma 1.27, an application of formula (1.13) followed by Cauchy-Schwarz leads to $E|\tilde{X}_N^{meso}(v)|^2 \leq I_v^2$ where

$$I_v := \frac{1}{\pi} \int_0^1 \int_{-\infty}^{\infty} \sqrt{E|V_N(\tau + i\eta)|^2} |\partial \Psi_{h_N}(\tau, \eta)| d\tau d\eta$$

Following the proof of Proposition 1.29 we easily deduce that $I_v$ is uniformly bounded in $N$ and $z_1, z_2 \in \mathbb{R}$. \qed
PART 2

MESOSCOPIC CENTRAL LIMIT THEOREM FOR GENERAL BETA ENSEMBLES

Based on joint work with F. Bekerman

Acknowledgements: The authors thank Alice Guionnet for the fruitful discussions and the very helpful comments and her support through the NSF grant MS-1307704. The authors also thank the anonymous referees who helped us improve the manuscript.
 §2.1 Introduction

We consider a system of \( N \) particles on the real line distributed according to a density proportional to

\[
\prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^0 e^{-N \sum \lambda_i} \prod \lambda_i,
\]

where \( V \) is a continuous potential and \( \beta > 0 \). This system is called the \( \beta \)-log gas, or general \( \beta \)-ensemble and for classical values of \( \beta \in \{1, 2, 4\} \), this distribution corresponds to the joint law of the eigenvalues of symmetric, hermitian or quaternionic random matrices with density proportional to \( e^{-N \text{Tr} V(M)} dM \) where \( N \) is the size of the random matrix \( M \).

Recently, great progress has been made to understand the behaviour of \( \beta \)-log gases. At the microscopic scale, the eigenvalues exhibit a universal behaviour (see [BFG13], [BEY14b], [BEY14a], [Bek15], [Shc14]) and the local statistics of the eigenvalues are described by the Sine\( \beta \) process in the bulk and the Stochastic Airy Operator at the edge (see [VV09] and [RRV11] for definitions). In this article, we study the linear fluctuations of the eigenvalues of general \( \beta \)-ensembles at the mesoscopic scale; we prove that for \( \alpha \in (0; 1) \) fixed, \( f \) a smooth function (whose regularity and decay at infinity will be specified later), and \( E \) a fixed point in the bulk of the spectrum

\[
\sum_{i=1}^{N} f\left(N^\alpha (\lambda_i - E)\right) - N \int f\left(N^\alpha (x - E)\right) d\mu_V(x)
\]

converges towards a Gaussian random variable. At the macroscopic level (i.e when \( \alpha = 0 \)), it is known that the eigenvalues satisfy a central limit theorem and the re-centered linear statistics of the eigenvalues converge towards a Gaussian random variable. This was first proved in [Joh98] for polynomial potentials satisfying the one-cut assumption. In [BG13a], the authors derived a full expansion of the free energy in the one-cut regime from which they deduce the central limit theorem for analytic potentials. The multi-cut regime is more complicated and in this setting, the central limit theorem does not hold anymore for all test functions (see [BG13b], [Shc13]). Similar results have also been obtained for the eigenvalues of Random Matrices from different ensembles (see [AZ06], [LP*t09b], [Shc11]).

Interest in mesoscopic linear statistics has surged in recent years. Results in this field of study were obtained in a variety of settings, for Gaussian random matrices [BdMK99b, FKS13], and for invariant ensembles [BD14, Lam16]. In many cases the results were shown at all scales \( \alpha \in (0; 1) \), often with the use of distribution specific properties. In more general settings, the absence of such properties necessitates other approaches to obtain the limiting behaviour at the mesoscopic regime. For example, an early paper studying mesoscopic statistics for Wigner Matrices was [BdMK99a], here the regime studied was \( \alpha \in (0; \frac{1}{2}) \). This was pushed to \( \alpha \in (0; \frac{1}{3}) \) [Lod15] using improved local law results , and recent work has pushed this to all scales [HK16]. The central limit theorem at the mesoscopic scale has also been obtained recently for the two dimensional \( \beta \)-log gas (or Coulomb gas) in [LS16, BBNY16].

Extending these results to one dimensional \( \beta \)-ensembles is a natural step. We also prove convergence at all mesoscopic scales. The proof of the main Theorem relies on the analysis of the loop equations (see Section 2.2) from which we can deduce a recurrence
relationship between moments, and the rigidity results from [BEY14b], [BEY14a] to control the linear statistics. Similar results have been obtained before in [BEYY15, Theorem 5.4]. There, the authors showed the mesoscopic CLT in the case of a quadratic potential, for small $\alpha$ (see Remark 5.5).

In Section 2.2, we introduce the model and recall some background results and Section 2.4 will be dedicated to the proof of Theorem 2.5.

§2.2 Definitions and Background

We consider the general $\beta$-matrix model. For a potential $V : \mathbb{R} \to \mathbb{R}$ and $\beta > 0$, we denote the measure on $\mathbb{R}^N$

$$P^N_V (d\lambda_1, \cdots, d\lambda_N) := \frac{1}{Z^N_V} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^\beta e^{-N \sum V(\lambda_i)} \prod_i d\lambda_i,$$

with

$$Z^N_V = \int \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^\beta e^{-N \sum V(\lambda_i)} \prod_i d\lambda_i.$$

It is well known (see [MPS95] for the Hölder case, and [J] Theorem 2.1 for the continuous case) that under $P^N_V$ the empirical measure of the eigenvalues converge towards an equilibrium measure:

**Theorem 2.1.** Assume that $V : \mathbb{R} \to \mathbb{R}$ is continuous and that

$$\liminf_{x \to \infty} \frac{V(x)}{\beta \log |x|} > 1.$$

Then the energy defined by

$$E(\mu) = \int \int \left( \frac{V(x_1) + V(x_2)}{2} - \frac{\beta}{2} \log |x_1 - x_2| \right) d\mu(x_1) d\mu(x_2) \quad (2.1)$$

has a unique global minimum on the space $\mathcal{M}_1(\mathbb{R})$ of probability measures on $\mathbb{R}$. Moreover, under $P^N_V$ the normalized empirical measure $L_N = N^{-1} \sum_{i=1}^N \delta_{\lambda_i}$ converges almost surely and in expectation towards the unique probability measure $\mu_V$ which minimizes the energy. Furthermore, $\mu_V$ has compact support $A$ and is uniquely determined by the existence of a constant $C$ such that:

$$\beta \int \log |x - y| d\mu_V(y) - V(x) \leq C,$$

with equality almost everywhere on the support.

**Hypothesis 2.2.** For what proceeds, we assume the following

- $V$ is continuous and goes to infinity faster than $\beta \log |x|$.
- The support of $\mu_V$ is a connected interval $A = [a; b]$ and

$$\frac{d\mu_V}{dx} = \rho_V(x) = S(x) \sqrt{b - x(x - a)} \quad \text{with} \quad S > 0 \quad \text{on}[a; b].$$
The function $V(\cdot) - \beta \int \log | - y | d\mu_V(y)$ achieves its minimum on the support only.

**Remark 2.3.** The second and third assumptions are typically known as the one-cut and off-criticality assumptions. In the case where the support of the equilibrium measure is no longer connected, the macroscopic central limit theorem does not hold anymore in generality (see [BG13b], [Shc13]). Whether the theorem holds for critical potentials is still an open question.

**Remark 2.4.** If the previous assumptions are fulfilled, and $V \in C^p(\mathbb{R})$ then $S \in C^{p-3}(\mathbb{R})$ (see for instance [BFG13], Lemma 3.2).

**Theorem 2.5.** Let $0 < \alpha < 1$, $E$ a point in the bulk $(a; b)$, $V \in C^6(\mathbb{R})$ and $f \in C^6(\mathbb{R})$ with compact support. Then, under $P_N^V$

$$\sum_{i=1}^{N} f(N^\alpha(\lambda_i - E)) - N \int f(N^\alpha(x - E)) d\mu_V(x) \Rightarrow \mathcal{N}(0, \sigma_f^2),$$

where the convergence holds in moments (and thus, in distribution), and

$$\sigma_f^2 = \frac{1}{2\beta^2} \int \int \left( \frac{f(x) - f(y)}{x - y} \right)^2 dx dy.$$

Note that, as in the macroscopic central limit theorem, the variance is universal in the potential with a multiplicative factor proportional to $\beta$. Interestingly and in contrast with the macroscopic scale, the limit is always centered.

The proof relies on an explicit computation of the moments of the linear statistics. We will use two tools: optimal rigidity for the eigenvalues of beta-ensembles to provide a bound on the linear statistics (as in [BEY14b], [BEY14a]) and the loop equations at all orders to derive a recurrence relationship between the moments.

For what follows, set

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}, \quad M_N = \sum_{i=1}^{N} \delta_{\lambda_i} - N \mu_V.$$

and for a measure $\nu$ and an integrable function $h$ set

$$\nu(h) = \int h d\nu \quad \text{and} \quad \tilde{\nu}(h) = \int h d\nu - E_N^V \left( \int h d\nu \right), \quad (2.2)$$

when $\nu$ is random and where $E_N^V$ is expectation with respect to $P_N^V$. Further $f$ will be any function as in Theorem 2.5, and

$$f_N(x) := f(N^\alpha(x - E)).$$

Finally, for any function $g \in C^p(\mathbb{R})$, let

$$\|g\|_{C^p(\mathbb{R})} := \sum_{l=0}^{p} \sup_{x \in \mathbb{R}} |g^{(l)}(x)|,$$

42
when it exists.

**Loop Equations:** To prove the convergence, we use the loop equations at all orders. Loop equations have been used previously to derive recurrence relationships between correlators and derive a full expansion of the free energy for \(\beta\)-ensembles in [Shc13], [BG13b], and [BG13a] (from which the authors also derive a macroscopic central limit theorem). The first loop equation was used to prove the central limit theorem at the macroscopic scale in [J] and used subsequently in [BEYY15]. Here, rather than using the first loop equation to control the Stieltjes transform as in [J] and [BEYY15], we rely on the analysis of the loop equations at all orders to compute directly the moments.

**Proposition 2.6.** Let \(h, h_1, h_2, \ldots\) be a sequence of bounded functions in \(C^1(\mathbb{R})\). Define

\[
F_k^N(h) := \frac{N\beta}{2} \int \int \frac{h(x) - h(y)}{x-y} dL_N(x)dL_N(y) - NL_N(hV') + \left(1 - \frac{\beta}{2}\right) L_N(h')
\]  
and for all \(k \geq 1\)

\[
F_{k+1}^N(h, h_1, \ldots, h_k) := F_k^N(h, h_1, \ldots, h_{k-1}) \tilde{M}_N(h_k) + \left(\prod_{l=1}^{k-1} \tilde{M}_N(h_l)\right) L_N(hh'_k)
\]

where the product is equal to 1 when \(k = 1\) and \(\tilde{M}_N\) was defined by the convention eq. (2.2). Then we have for all \(k \geq 1\)

\[
E_V^N(F_k^N(h, h_1, \ldots, h_{k-1})) = 0,
\]

which is called the loop equation of order \(k\).

**Proof.** The first loop equation (2.3) is derived by integration by parts (see also [J] eq. (2.18) for a proof using a change of variables). More precisely, for a fixed index \(l\), integration by parts with respect to \(\lambda_l\) yields the equality:

\[
E_V^N(h'(\lambda_l)) = -E_V^N\left(h(\lambda_l) \left(\beta \sum_{1 \leq i \leq N, i \neq l} \frac{1}{\lambda_i - \lambda_l} - NV'(\lambda_l)\right)\right).
\]

Summing over \(l\) we get by symmetry

\[
E_V^N\left(\frac{\beta}{2} \sum_{l=1}^{N} \sum_{1 \leq i \leq N, i \neq l} h(\lambda_l) - h(\lambda_i)}{\lambda_l - \lambda_i} - N \sum_{l=1}^{N} V'(\lambda_l)h(\lambda_l) + \sum_{l=1}^{N} h'(\lambda_l)\right) = 0
\]

Writing the sums in term of \(L_N\) gives eq. (2.5) for \(k = 1\).

To derive the loop equation at order \(k + 1\) from the one at order \(k\), replace \(V\) by \(V + \delta h_k\) and notice that for any functional \(F\) that is independent of \(\delta\),

\[
\left.\frac{\partial E_V^N(F)}{\partial \delta}\right|_{\delta=0} = -N E_V^N\left(F \tilde{M}_N(h_k)\right).
\]

43
Also observe that the loop equation eq. (2.5) is now
\[ E_{V + \delta h_k}^N (F_k^N (h, h_1, \cdots, h_{k-1})) - \delta N E_{V + \delta h_k}^N \left( \prod_{i=1}^{k-1} \tilde{M}_N (h_i) \right) L_N (hh'_k) = 0, \]
by induction and the definitions given in eqns. (2.3) and (2.4). Differentiating both sides with respect to \( \delta \) and setting \( \delta = 0 \) yields the loop equation at order \( k + 1 \).

It will be easier to compute the moments of \( M_N (\tilde{f}) \) by re-centering the first loop equation — that is, we wish to replace \( L_N \) by \( L_N - \mu_V \). To that end, define the operator \( \Xi \) acting on smooth functions \( h : \mathbb{R} \to \mathbb{R} \) by
\[ \Xi h (x) := \beta \int \frac{h(x) - h(y)}{x - y} d\mu_V (y) - V'(x) h(x). \]
We now prove the equilibrium relation (2.6) below, to recenter \( L_N \) by \( \mu_V \). Consider for a smooth function \( h : \mathbb{R} \to \mathbb{R} \) and \( \delta \) in a neighbourhood of 0, \( \mu_{V, \delta} = (x + \delta h(x)) \# \mu_V \), where for a map \( T \) and measure \( \mu, T \# \mu \) refers to the push-forward measure of \( \mu \) by \( T \). Then by (2.1) we have \( E(\mu_{V, \delta}) \geq E(\mu V) \), which writes
\[ \int \int \left\{ \frac{V(x_1 + \delta h(x_1)) + V(x_2 + \delta h(x_2))}{2} - \frac{\beta}{2} \log |x_1 - x_2 + \delta(h(x_1) - h(x_2))| \right\} d\mu_V (x_1) d\mu_V (x_2) \geq \int \int \left( \frac{V(x_1) + V(x_2)}{2} - \frac{\beta}{2} \log |x_1 - x_2| \right) d\mu_V (x_1) d\mu_V (x_2). \]
As \( \delta \) approaches 0 we get
\[ \delta \left( \frac{1}{N} M_N (\Xi h) + \frac{1}{N} \int \int \frac{h(x) - h(y)}{x - y} dL_N (x) dL_N (y) - L_N (h') \right) + O(\delta^2) \geq 0, \]
So that
\[ \frac{\beta}{2} \int \int \frac{h(x) - h(y)}{x - y} d\mu_V (x) d\mu_V (y) = \int V'(x) h(x) d\mu_V (x), \]
and thus
\[ \frac{\beta}{2} \int \int \frac{h(x) - h(y)}{x - y} dL_N (x) dL_N (y) = \frac{1}{N} M_N (\Xi h) + \frac{\beta}{2 N^2} \int \int \frac{h(x) - h(y)}{x - y} dM_N (x) dM_N (y). \]
Consequently, we can write
\[ F_1^N (h) = M_N (\Xi h) + \left( 1 - \frac{\beta}{2} \right) L_N (h') + \frac{1}{N} \left[ \frac{\beta}{2} \int \int \frac{h(x) - h(y)}{x - y} dM_N (x) dM_N (y) \right]. \]
One of the key features of the operator \( \Xi \) is that it is invertible (modulo constants) in the space of smooth functions. More precisely, we have the following Lemma (see [BFG13, Lemma 3.2] for the proof):
Lemma 2.7 (Inversion of Ξ). Assume that \( V \in C^p(\mathbb{R}) \) and satisfies Hypothesis 2.2. Let \([a; b]\) denote the support of \( \mu_V \) and set
\[
\frac{d\mu_V}{dx} = S(x)\sqrt{(b - x)(x - a)} = S(x)\sigma(x),
\]
where \( S > 0 \) on \([a; b]\). Then for any \( k \in C^r(\mathbb{R}) \) there exists a unique constant \( c_k \) and \( h \in C^{(r-2)\wedge(p-3)}(\mathbb{R}) \) such that
\[
\Xi(h) = k + c_k.
\]
Moreover the inverse is given by the following formulas:

\( \forall x \in \text{supp}(\mu_V) \)
\[
h(x) = -\frac{1}{\beta \pi^2 S(x)} \left( \int_a^b k(y) - k(x) \frac{dy}{\sigma(y)(y - x)} \right)
\]

\( \forall x \notin \text{supp}(\mu_V) \)
\[
h(x) = \beta \int \frac{h(y)}{x - y} d\mu_V(y) + k(x) + c_k \beta \int \frac{1}{x - y} d\mu_V(y) - V'(x).
\]

And \( c_k = -\beta \int \frac{h(y)}{a - y} d\mu_V(y) - k(a) \). Note that the definition (2.9) is proper since \( h \) has been defined on the support.

We shall denote this inverse by \( \Xi^{-1}k \).

Remark 2.8. For \( f \) and \( V \) as in Theorem 2.5, \( p = 6 \) and \( r = 6 \) so \( \Xi^{-1}f_N \in C^4(\mathbb{R}) \).

Remark 2.9. The denominator \( \beta \int \frac{1}{x - y} d\mu_V(y) - V'(x) \) is identically null on \( \text{supp} \mu_V \) and behaves like a square root at the edges. As we can modify freely the potential outside any neighbourhood of the support (see for instance the large deviation estimates Section 2.1 of [BG13a]), we may assume that it doesn’t vanish outside \( \mu_V \).

In order to bound the linear statistics we use the following lemma to bound \( \Xi^{-1}(f_N) \) and its derivatives.

Lemma 2.10. Let \( \text{supp} f \subset [-M, M] \) for some constant \( M > 0 \). For each \( p \in \{0, 1, 2, 3\} \), there is a constant \( C > 0 \) such that
\[
\|\Xi^{-1}(f_N)\|_{C^p(\mathbb{R})} \leq C N^{p\alpha} \log N,
\]
Moreover, there is a constant \( C \) such that whenever \( x \in \text{supp} \mu_V \) and \( N^\alpha|x - E| \geq M + 1 \)
\[
|\Xi^{-1}(f_N)^{(p)}(x)| \leq \frac{C}{N^\alpha (x - E)^{p+1}},
\]
and when \( x \notin \text{supp} \mu_V \)
\[
|\Xi^{-1}(f_N)^{(p)}(x)| \leq \frac{C \log N}{N^\alpha}.
\]
Proof. We start by proving (2.10) on the support. For $x \in \text{supp} \mu_V$ we use

$$
\Xi^{-1}(f_N)(x) = -\frac{N^\alpha}{\beta \pi^2 S(x)} \int_a^b \frac{1}{\sigma(y)} \int_0^1 f'(N^\alpha t(x - E) + N^\alpha(1 - t)(y - E)) dt dy
$$

so that

$$
\Xi^{-1}(f_N)^{(p)}(x) = -\frac{1}{\beta \pi^2} \sum_{l=0}^p \left\{ \left( \begin{array}{c} p \\ l \end{array} \right) \left( \frac{1}{S} \right)^{(p-l)}(x) \right\} \times \int_a^b \frac{N^{(l+1)}\alpha}{\sigma(y)} \int_0^1 t^l f^{(l+1)}(N^\alpha t(x - E) + N^\alpha(1 - t)(y - E)) dt dy.
$$

Let $A(x) = \{(t, y) \in [0; 1] \times [a; b] : N^\alpha|t(x - E) + (1 - t)(y - E)| \leq M\}$. We have

$$
\int_0^1 1_{A(x)}(t, y) dt \leq \frac{2M}{N^\alpha|x - y|} \land 1
$$

and thus

$$
\int_a^b \frac{N^{(l+1)}\alpha}{\sigma(y)} \int_0^1 |f^{(l+1)}(N^\alpha t(x - E) + N^\alpha(1 - t)(y - E))| dt dy \leq C \log NN^\alpha,
$$

and this proves (2.10).

We now proceed with the proof of (2.11). First, let $x \in \text{supp} \mu_V$ such that $N^\alpha|x - E| \geq M + 1$. The inversion formula (2.8) writes

$$
\Xi^{-1}(f_N)(x) = -\frac{1}{\beta \pi^2 S(x)} \int_a^b \frac{f(N^\alpha(y - E))}{\sigma(y)(y - x)} dy

= -\frac{1}{\beta \pi^2 S(x)} \int_{-M}^M \frac{f(u)}{\sigma(E + N^\alpha u - N^\alpha(x - E))} du.
$$

and we can conclude in this setting by differentiating under the integral. Moreover we see that $\Xi^{-1}(f_N)$ is in fact of class $C^5$ on $\text{supp} \mu_V$ and similar bounds holds for $p \in \{4, 5\}$.

We now prove the bounds for $x \notin \text{supp} \mu_V$. Let $\psi_N$ be an arbitrary extension of $\Xi^{-1}(f_N)|_{\text{supp} \mu_V}$ in $C^6(\mathbb{R})$, bounded by $C/N^\alpha$ outside the support (and its five first derivatives as well). This is possible by what we just proved and a Taylor expansion. Using (2.9) we notice that

$$
\Xi^{-1}(f_N)(x) = \frac{\beta \int \frac{\psi_N(y)}{x - y} d\mu_V(y) + c_{f_N}}{\beta \int \frac{1}{x - y} d\mu_V(y) - V'(x)}

= \frac{\beta \int \frac{\psi_N(x) - \psi_N(y)}{x - y} d\mu_V(y) + \beta \psi_N(x) \int \frac{d\mu_V(y)}{x - y} + c_{f_N}}{\beta \int \frac{1}{x - y} d\mu_V(y) - V'(x)}

= \psi_N(x) - \frac{\Xi(\psi_N)(x) - c_{f_N}}{\beta \int \frac{1}{x - y} d\mu_V(y) - V'(x)}.
$$
f has compact support we may write \( \Xi(\psi_N) - c_{f_N} = \Xi(\psi_N) - c_{f_N} - f_N \) on \([a; a + \varepsilon]\) and \([b - \varepsilon; b]\) for \(\varepsilon\) small enough. Furthermore this quantity vanishes identically on these intervals by definition of \(\psi_N\). Consequently, \(\Xi(\psi_N) - c_{f_N}\) and its four first derivatives vanish at the edges. By definition we also get that

\[
|c_{f_N}| = \left| \beta \int \frac{\psi_N(y)}{a - y} d\mu_V(y) \right| \\
\leq C \log N \int_{|y - E| \leq 2M/N} \frac{d\mu_V(y)}{y - a} + \frac{C}{N^\alpha} \int_{|y - E| \geq 2M/N} \frac{d\mu_V(y)}{(y - a)|y - E|} \\
\leq C \frac{\log N}{N^\alpha}
\]

On the other hand, for \(p \in [4]\) and \(x \notin \text{supp } \mu_V\),

\[
\Xi(\psi_N)^{(p)}(x) = \beta d! \left( \psi_N(y) - \psi_N(x) - \cdots - \psi_N^{(p)}(x)(y - x)^p/p! \right) d\mu_V(y) - (V'\psi_N)^{(p)}(x) .
\]

By doing a similar splitting, and bounding the fifth derivative of \(\psi_N\) uniformly away from \(E\), we obtain the same bound \(C \log N/N^\alpha\) on \(\Xi(\psi_N)^{(p)}\) outside the support. By Remark 2.9 and (2.14), we conclude that we can bound the \(C^3\) norm of \(\Xi^{-1}(f_N)\) by \(C \log N/N^\alpha\) outside the support.

**Sketch of the Proof:** We have developed the tools we need to prove Theorem 2.5. In order to motivate the technical estimates in the following section, we now sketch the proof by computing the first moments. Consider a function \(f\) satisfying the hypothesis of Theorem 2.5. Applying (2.7) to \(\Xi^{-1}(f_N)\) yields

\[
F_1^N (\Xi^{-1}(f_N)) = M_N(f_N) + \left(1 - \frac{\beta}{2}\right) L_N ((\Xi^{-1}f_N)'') \\
+ \frac{1}{N} \left[ \frac{\beta}{2} \int \int \frac{\Xi^{-1}f_N(y) - \Xi^{-1}f_N(x)}{x - y} dM_N(x) dM_N(y) \right] ,
\]

If the central limit theorem holds, we expect terms of the type \(M_N(h)\) where \(h\) is fixed to be almost of constant order, and this an easy consequence of the rigidity estimates from [BEY14a] (stated as Theorem 2.11 below). Due to the dependency in \(N\) of \(f_N\) (and its inverse under \(\Xi\)), a little care must be taken for these estimates to yield a bound on the last term in the right handside, and this is precisely the point of Lemma 2.14, eq. (2.24). Similarly, we have

\[
L_N ((\Xi^{-1}f_N)') = \mu_V ((\Xi^{-1}f_N)') + \frac{1}{N} M_N ((\Xi^{-1}f_N)') ,
\]

and Lemma 2.13 shows the term in the right handside is a small error term. Thus admitting the results of the next section, we would get with high probability and for \(\varepsilon_N\) small

\[
F_1^N (\Xi^{-1}(f_N)) = M_N(f_N) + \left(1 - \frac{\beta}{2}\right) \mu_V ((\Xi^{-1}f_N)') + \varepsilon_N .
\]
By the first loop equation from Proposition 2.6, the expectation of $F_1^N$ is zero and this shows that the first moment

$$E_V^N \left( M_N(f_N) \right) = \left( 1 - \frac{\beta}{2} \right) \mu_V \left( (\Xi^{-1} f_N)' \right) + o(1)$$

The term on the right handside is deterministic and is shown to decrease towards zero in Lemma *INSERT*. Thus the first moment converges to 0.

In order to exhibit all the terms we will need to control, we proceed with the computation of the second moment. By definition

$$F_2^N (\Xi^{-1}(f_N), f_N) = F_1^N (\Xi^{-1}(f_N)) \tilde{M}_N(f_N) + L_N (\Xi^{-1}(f_N)f_N') ,$$

which we can write (with now an $\epsilon_N$ incorporating the deterministic mean converging to zero)

$$F_2^N (\Xi^{-1}(f_N), f_N) = M_N(f_N)\tilde{M}_N(f_N) + \epsilon_N \tilde{M}_N(f_N) + L_N (\Xi^{-1}(f_N)f_N')$$

Lemma 2.13 ensures that $\epsilon_N \tilde{M}_N(f_N)$ remains small, and that the term in the right handside of the decomposition

$$L_N (\Xi^{-1}(f_N)f_N') = \mu_V (\Xi^{-1}(f_N)f_N') + \frac{1}{N} M_N (\Xi^{-1}(f_N)f_N') ,$$

is also controlled. Consequently, using the second loop equation we see that

$$E_V^N \left( M_N(f_N)^2 \right) = -\mu_V (\Xi^{-1}(f_N)f_N') + o(1)$$

The limit of the term appearing on the right handside is then computed in Lemma 2.14 equation (2.23). The following moments are computed similarly (see Section 2.4).

In the following section, we establish all the bounds we need for the proof of Theorem 2.5. The previous steps will then be made rigorous in the last section.
§2.3 Control of the linear statistics.

We now make use of the strong rigidity estimates proved in [BEY14a, Theorem 2.4] to control the linear statistics. We recall the result here.

**Theorem 2.11.** Let $\gamma_i$ the quantile defined by

$$\int_{a}^{\gamma_i} d\mu_V(x) = \frac{i}{N}.$$ 

Then, under Hypothesis 2.2 and for all $\xi > 0$ there exists constants $c > 0$ such that for $N$ large enough

$$P^N_{\nu}(\{|\lambda_i - \gamma_i| \geq N^{-2/3} + \xi \gamma_i^{-1/3}\}) \leq e^{-N^c},$$

where $\gamma = i \wedge (N + 1 - i)$.

We will use the following lemma quite heavily in what proceeds.

**Lemma 2.12.** Let $\gamma_i$ and $i$ be as in Theorem 2.11, let $t \in [0; 1]$, and let $\lambda_i, i \in [N]$, be a configuration of points such that $|\lambda_i - \gamma_i| \leq N^{-2/3} + \xi \gamma_i^{-1/3}$ for $0 < \xi < (1 - \alpha) \wedge \frac{2}{3}$, and let $M > 1$ be a constant. Define the pairwise disjoint sets:

- $J_1 := \{i \in [N], |N^\alpha(\gamma_i - E)| \leq 2M\}$,
- $J_2 := \left\{i \in J_1^c, |(\gamma_i - E)| \leq \frac{1}{2}(E - a) \wedge (b - E)\right\}$,
- $J_3 := J_1^c \cap J_2^c$.

The following statements hold:

(a) For all $i \in J_1 \cup J_2$, $i \geq CN$, for some $C > 0$ that depend only on $\mu_V$ in a neighborhood of $E$, also for all such $i$, $|\gamma_i - \gamma_{i+1}| \leq \frac{C}{N}$ for a constant $C > 0$ depending only on $\mu_V$ in a neighborhood of $E$.

(b) Uniformly in all $i \in J_1^c = J_2 \cup J_3$, $x \in [\gamma_i, \gamma_{i+1}]$ and all $t \in [0; 1]$,

$$|N^\alpha t(\lambda_i - x) + N^\alpha(x - E)| > M + 1,$$

for $N$ large enough.

(c) The cardinality of $J_1$ is of order $CN^{1-\alpha}$, where again, $C > 0$ depends only on $\mu_V$ in a neighborhood of $E$.

**Proof.** The first part of statement (a) holds by the observation that for $i \in J_1 \cup J_2$, $\gamma_i$ is in the bulk, so

$$0 < c \leq \int_{a}^{\gamma_i} d\mu_V(x) = \frac{i}{N} \leq C < 1$$

for constants $C, c > 0$ depending only on $\mu_V$. For the second part of statement (a), the density of $\mu_V$ is bounded below uniformly in $i \in J_1 \cup J_2$, so

$$c|\gamma_i - \gamma_{i+1}| \leq \int_{\gamma_i}^{\gamma_{i+1}} d\mu_V(x) = \frac{1}{N}.$$
Statement (b) can be seen as follows: let $i \in J_2$ and consider first $x = \gamma_i$. On this set $i \geq CN$ by (a), so uniformly in such $i$, $N^\alpha|\lambda_i - \gamma_i| \leq CN^\alpha_{-1+\xi}$, which goes to zero, while $N^\alpha|\gamma_i - E| > 2M$. On the other hand, for $i \in J_3$, we have $N^\alpha|\gamma_i - E| > \frac{1}{2}N^\alpha(E-a)\wedge(b-E)$, which goes to infinity faster than $N^\alpha|\lambda_i - \gamma_i| \leq N^{\alpha-\frac{3}{2}+\xi}$, by our choice of $\xi$. When we substitute $\gamma_i$ by $x$, the same argument holds because $N^\alpha|x - \gamma_i| \leq N^\alpha|\gamma_i - \gamma_{i+1}|$, which is of order $N^\alpha_{-1}$ on $J_2$ (as we showed in statement (a)) and bounded by $CN^\alpha_{-\frac{4}{3}}$ on $J_3$.

Statement (c) follows by the observation that on the set $x \in [a, b]$ such that $|x - E| \leq \frac{2M}{N^\alpha}$ the density of $\mu_V$ is bounded uniformly above and below, so
\[
\frac{c}{N^\alpha} \leq \int_{|x - E| \leq \frac{2M}{N^\alpha}} d\mu_V(x) = \sum_{i \in J_1} \int_{\gamma_i}^{\gamma_{i+1}} d\mu_V(x) + O\left(\frac{1}{N}\right) \leq \frac{C}{N^\alpha},
\]
giving the required result.

The rigidity of eigenvalues, Theorem 2.11, along with the previous Lemma leads to the following estimates.

**Lemma 2.13.** For all $0 < \xi < (1 - \alpha) \wedge \frac{2}{3}$ there exists constants $C$, $c > 0$ such that for $N$ large enough we have the concentration bounds
\[
P_N^V (|M_N(f_N)| \geq CN^\xi\|f\|_{C^1(R)}) \leq e^{-N^c},
\]
\[
P_N^V (|M_N(\Xi^{-1}(f_N'))| \geq CN^{\alpha+\xi}\|f\|_{C^1(R)}) \leq e^{-N^c},
\]
\[
P_N^V (|M_N(\Xi^{-1}(f_N)f_N')| \geq CN^{\alpha+\xi}\|f\|_{C^1(R)}) \leq e^{-N^c}
\]

**Proof.** Let $M > 1$ such that supp $f \subset [-M, M]$ and fix $0 < \xi < (1 - \alpha) \wedge \frac{2}{3}$. For the remainder of the proof, we may assume that we are on the event $\Omega := \{\forall i \, |\lambda_i - \gamma_i| \leq N^{-2/3+\xi} i^{-1/3}\}$. This follows from the fact that, for example,
\[
P_N^V (|M_N(f_N)| \geq CN^\xi\|f\|_{C^1(R)}) \leq P_N^V (\{ |M_N(f_N)| \geq CN^\xi\|f\|_{C^1(R)} \} \cap \Omega) + P_N^V (\Omega^c),
\]
and by Theorem 2.11, we may bound $P_N^V (\Omega^c)$ by $e^{-N^c}$ for some constant $c > 0$, and $N$ large enough. On $\Omega$, the $\lambda_i$ satisfy the conditions of Lemma 2.12, we will utilize the sets $J_1$, $J_2$, and $J_3$ as defined there.

We begin by controlling (2.18). We have that
\[
|M_N(f_N)| = \left| \sum_{i=1}^{N} f(N^\alpha(\lambda_i - E)) - N\mu_V(\tilde{f}) \right| \\
\leq \sum_{i=1}^{N} f(N^\alpha(\lambda_i - E)) - \sum_{i=1}^{N} f(N^\alpha(\gamma_i - E)) \right| + \left| \sum_{i=1}^{N} f(N^\alpha(\gamma_i - E)) - N\mu_V(\tilde{f}) \right|,
\]
(2.21)
the first term in (2.21) may be bounded (on $\Omega$) by

$$\left| \sum_{i=1}^{N} f(N^\alpha(\lambda_i - E)) - \sum_{i=1}^{N} f(N^\alpha(\gamma_i - E)) \right| =$$

$$\left| \sum_{i=1}^{N} N^\alpha(\lambda_i - \gamma_i) \int_{0}^{1} f'(tN^\alpha(\lambda_i - \gamma_i) + N^\alpha(\gamma_i - E)) dt \right|$$

$$\leq \sum_{i=1}^{N} N^{\alpha-2/3+\xi} \int_{0}^{1} \left| f'(tN^\alpha(\lambda_i - \gamma_i) + N^\alpha(\gamma_i - E)) \right| dt,$$

By Lemma 2.12 (b), for $N$ large enough, we have

$$\int_{0}^{1} \sum_{i=1}^{N} N^{\alpha-2/3+\xi} \int_{0}^{1/3} \left| f'(tN^\alpha(\lambda_i - \gamma_i) + N^\alpha(\gamma_i - E)) \right| dt$$

$$= \int_{0}^{1} \sum_{i \in J_1} N^{\alpha-2/3+\xi} \int_{0}^{1/3} \left| f'(tN^\alpha(\lambda_i - \gamma_i) + N^\alpha(\gamma_i - E)) \right| dt,$$

$$\leq \sum_{i \in J_1} N^{\alpha-1+\xi} \| f \|_{C^1(R)} \leq CN^\xi \| f \|_{C^1(R)},$$

where, in the third line we used Lemma 2.12 (a) and (c) in order. Thus

$$\left| \sum_{i=1}^{N} f(N^\alpha(\lambda_i - E)) - \sum_{i=1}^{N} f(N^\alpha(\gamma_i - E)) \right| \leq CN^\xi \| f \|_{C^1(R)}.$$

For the second term in (2.21),

$$\left| \sum_{i=1}^{N} f(N^\alpha(\gamma_i - E)) - N \int_{a}^{b} f(N^\alpha(x - E)) d\mu_V(x) \right|$$

$$\leq N \sum_{i \in J_1} \int_{\gamma_i}^{\gamma_{i+1}} \left| f(N^\alpha(\gamma_i - E)) - f(N^\alpha(x - E)) \right| d\mu_V(x)$$

$$\leq N^{1+\alpha} \| f \|_{C^1(R)} |J_1| \sup_{i \in J_1} (\gamma_{i+1} - \gamma_i) \int_{\gamma_i}^{\gamma_{i+1}} d\mu_V(x) \leq C \| f \|_{C^1(R)}$$

since the spacing of the quantiles in $J_1$ is bounded by $\frac{C}{N}$. This proves (2.18).

We now proceed with the proof of (2.19).

$$\left| M_N(\Xi^{-1}(f_N))' \right| = \left| \sum_{i=1}^{N} \left( \Xi^{-1}(f_N)'(\lambda_i) - N \int_{\gamma_i}^{\gamma_{i+1}} \Xi^{-1}(f_N)'(x) d\mu_V(x) \right) \right|$$

$$\leq N \sum_{i=1}^{N} \int_{\gamma_i}^{\gamma_{i+1}} \left| \Xi^{-1}(f_N)'(\lambda_i) - \Xi^{-1}(f_N)'(x) \right| d\mu_V(x)$$

$$\leq N \sum_{i=1}^{N} \int_{\gamma_i}^{\gamma_{i+1}} \int_{0}^{1} |\lambda_i - x| \left| \Xi^{-1}(f_N)(t(\lambda_i - x) + x) \right| dt d\mu_V(x),$$

51
Recall from the proof of Lemma 2.12 that uniformly in $i \in J_2$ and $x \in [\gamma_i, \gamma_{i+1}]$, $|\gamma_i - E| \geq \frac{2M}{N^\alpha}$ while $|x - \gamma_i| \leq \frac{C}{N}$; further, $|\lambda_i - x| \leq CN^{-1+\xi}$ so for $N$ large enough we can replace $|t(\lambda_i - x) + (\gamma_i - E)|$ by $|\gamma_i - E|$ uniformly in $t \in [0; 1]$. Likewise, uniformly in $i \in J_3$, $x \in [\gamma_i, \gamma_{i+1}]$ and $t \in [0; 1]$ we can bound below $|t(\lambda_i - x) + (x - E)|$ by a constant for what follows.

For $i \in J_2$, by the observations in the previous paragraph, along with Lemma 2.12 (b), Lemma 2.10 (2.11), and Lemma 2.12 (a),

$$N \sum_{i \in J_2} \int_{\gamma_i}^{\gamma_{i+1}} \int_0^1 |\lambda_i - x| \left| \Xi^{-1}(f_N)^{(2)}(t(\lambda_i - x) + x) \right| dt \, d\mu_V(x)$$

$$\leq N \sum_{i \in J_2} \int_{\gamma_i}^{\gamma_{i+1}} \int_0^1 \frac{C|\lambda_i - x|}{N^\alpha(|t(\lambda_i - x) + x - E|)^3} \, dt \, d\mu_V(x) \leq \sum_{i \in J_2} \frac{CN^{\xi-1/3}}{(\gamma_i - E)^3},$$

The same reasoning for $i \in J_3$ using also (2.12) yields

$$N \sum_{i \in J_3} \int_{\gamma_i}^{\gamma_{i+1}} \int_0^1 |\lambda_i - x| \left| \Xi^{-1}(f_N)^{(2)}(t(\lambda_i - x) + x) \right| dt \, d\mu_V(x) \leq \log N \sum_{i \in J_3} CN^{\xi-\frac{2}{3} - \frac{1}{3}} i^{-\frac{1}{3}}.$$

For $i \in J_1$, by Lemma 2.10 (2.10) and Lemma 2.12 (a),

$$N \sum_{i \in J_1} \int_{\gamma_i}^{\gamma_{i+1}} \int_0^1 |\lambda_i - x| \left| \Xi^{-1}(f_N)^{(2)}(t(\lambda_i - x) + x) \right| dt \, d\mu_V(x)$$

$$\leq N \sum_{i \in J_1} \int_{\gamma_i}^{\gamma_{i+1}} CN^{2\alpha} \log N |\lambda_i - x| \, d\mu_V(x) \leq \sum_{i \in J_1} CN^{2\alpha + \xi - 1} \log N.$$

It follows that

$$|M_N(\Xi^{-1}(f_N))| \leq \sum_{i \in J_1} CN^{2\alpha + \xi - 1} \log N + \sum_{i \in J_2} \frac{CN^{\xi-1/3}}{(\gamma_i - E)^3} + \log N \sum_{i \in J_3} CN^{\xi-\frac{2}{3} - \frac{1}{3}} i^{-\frac{1}{3}}$$

$$\leq CN^{\alpha + \xi} \log N + C \log NN^{-\xi} \leq CN^{\alpha + \xi} \log N,$$

where we have used $|J_1| \leq CN^{1-\alpha}$ from Lemma 2.12, and the following estimates:

$$\sum_{i \in J_2} \frac{N^{\xi-1}}{(\gamma_i - E)^3} \leq CN^{\xi-1} \left( \int_a^{E-\frac{2M}{N^\alpha}} \frac{dx}{(x - E)^3} + \int_{E+\frac{2M}{N^\alpha}}^b \frac{dx}{(x - E)^3} \right) \leq CN^{\xi+\alpha},$$

$$CN^{\xi-\frac{2}{3}} \sum_{i \in J_3} i^{-\frac{1}{3}} \leq CN^{\xi-1} \frac{1}{N} \sum_{i=1}^N \left( \frac{i}{N} \right)^{-\frac{1}{3}} \leq CN^{\xi-1}.$$

This proves (2.19). The bound (2.20) is obtained in a similar way and we omit the details.

For convenience we introduce the following notation: for a sequence of random variable $(X_N)_{N \in \mathbb{N}}$ we write $X_N = \omega(1)$ if there exists constants $c$, $C$ and $\delta > 0$ such that the bound $|X_N| \leq \frac{C}{N^\delta}$ holds with probability greater than $1 - e^{-N^\delta}$.

Using Lemma 2.13 we prove the following bounds:
Lemma 2.14. The following estimates hold:

\[ L_N (\Xi^{-1}(f_N)') = \omega(1), \]

(2.22)

\[ L_N (\Xi^{-1}(f_N)f_N') + \sigma_f^2 = \omega(1), \]

(2.23)

\[ \frac{1}{N} \int \int \frac{\Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y)}{x - y} dM_N(x) dM_N(y) = \omega(1). \]

(2.24)

Proof. For both (2.22) and (2.23), we use

\[ L_N (\Xi^{-1}(f_N)') = \frac{M_N(\Xi^{-1}(f_N)')}{N} + \mu_V (\Xi^{-1}(f_N)'), \]

\[ L_N (\Xi^{-1}(f_N)f_N') = \frac{M_N (\Xi^{-1}(f_N)f_N')}{N} + \mu_V (\Xi^{-1}(f_N)f_N'), \]

Lemma 2.13 implies that the first term in both equations are \( \omega(1) \) so (2.22) and (2.23) simplify to deterministic statements about the speed of convergence of the integrals against \( \mu_V \) above.

To show (2.22), integration by parts yields:

\[ \int (\Xi^{-1}f_N)'(x) d\mu_V(x) = - \int_a^b (\Xi^{-1}f_N)(x)(S'(x)\sigma(x) + S(x)\sigma'(x)) dx, \]

inserting the formula for \( \Xi^{-1}f_N \) we obtain

\[ \left| \int (\Xi^{-1}f_N)'(x) d\mu_V(x) \right| \leq \frac{1}{\beta \pi^2} \int_a^b \int_a^b \left| f_N(x) - f_N(y) \right| \left( \left| \frac{S'(x)\sigma(x)}{S(x)\sigma(y)} \right| + \left| \frac{\sigma'(x)}{\sigma(y)} \right| \right) dx dy. \]

Recall that \( S \) is bounded below on \([a, b]\), \( S' \) is bounded above on \([a, b]\), further, up to a constant, \( \frac{\sigma'(x)}{\sigma(y)} \) can be bounded above by \((\sigma(x)\sigma(y))^{-1}\). We define the sets

\[ A_N := [N^\alpha(a - E); N^\alpha(b - E)], \]

\[ B_N := \left[ \frac{1}{2} N^\alpha(a - E); \frac{1}{2} N^\alpha(b - E) \right]. \]

By the observations above, and the change of variable \( u = N^\alpha(x - E) \) and \( v = N^\alpha(y - E) \) we get

\[ \left| \int (\Xi^{-1}f_N)'(x) d\mu_V(x) \right| \leq \frac{C}{N^\alpha} \int_{A_N} \int_{A_N} \left| f(u) - f(v) \right| \left( \frac{\sigma(E + \frac{u}{N^\alpha})}{\sigma(E + \frac{v}{N^\alpha})} + \frac{1}{\sigma(E + \frac{u}{N^\alpha})\sigma(E + \frac{v}{N^\alpha})} \right) dudv. \]

(2.25)
For large enough \( N \), on the set \( (u, v) \in (A_N \setminus B_N)^2 \), the function \(|f(u) - f(v)|\) is always zero, thus the integral on the right above can be divided into integrals over the sets:

\[
(A_N \times A_N) \cap (A_N \setminus B_N \times A_N \setminus B_N)^c = B_N \times B_N \cup B_N \times (A_N \setminus B_N) \cup (A_N \setminus B_N) \times B_N. \tag{2.26}
\]

We bound the integral in (2.25) over each set in (2.26). We begin with the first set in (2.26). For \((u, v) \in B_N \times B_N\), \(\sigma(E + \frac{v}{N^\alpha})\) and \(\sigma(E + \frac{u}{N^\alpha})\) are uniformly bounded above and below. Therefore, the integral in (2.25) can be bounded in this region by

\[
\int \int_{B_N^2} \left| \frac{f(u) - f(v)}{u - v} \right| \, du \, dv = 
\int \int_{[-M; M]^2} \left| \frac{f(u) - f(v)}{u - v} \right| \, du \, dv + 2 \int_{-M}^M \int_{B_N \cap \{|u| \geq M\}} \left| \frac{f(v)}{u - v} \right| \, du \, dv,
\]

the integral over \([-M; M]^2\) exists by the differentiability of \( f \), while:

\[
\int_{-M}^M \int_{B_N \cap \{|u| \geq M\}} \left| \frac{f(v)}{u - v} \right| \, du \, dv \leq C \int_{-M}^M |f(v)| \log[N|v + M||v - M|] \, dv \leq C \log N,
\]

for \( N \) large enough.

For the second set in (2.25), observe that for \((u, v) \in B_N \times (A_N \setminus B_N)\), \(f(v) = 0\) for \( N \) sufficiently large, and \(\sigma(E + \frac{v}{N^\alpha})\) is bounded uniformly above and below while \(f(u)\) is 0 outside \([-M; M]\). This implies that the integral in (2.25) can be bounded in this region by

\[
\int_{A_N \setminus B_N} \int_{-M}^M \left| \frac{f(u)}{u - v} \right| \left( \frac{\sigma(E + \frac{v}{N^\alpha})}{\sigma(E + \frac{u}{N^\alpha})} + \frac{1}{\sigma(E + \frac{u}{N^\alpha})\sigma(E + \frac{v}{N^\alpha})} \right) \, du \, dv 
\leq C \frac{\|f\|_{C(R)}}{N^\alpha} \int_{A_N \setminus B_N} \frac{1}{\sigma(E + \frac{u}{N^\alpha})} \, dv \leq C,
\]

where in the final line we used \(|u - v| \geq cN^{\alpha}\) for \( u \in [-M; M] \), \( v \in A_N \setminus B_N\).

We can do similarly for the third set in (2.25) and putting together these bounds on the right hand side of (2.25) gives

\[
\left| \int (\Xi^{-1}f_N)'(x) d\mu_V(x) \right| \leq \frac{C \log N}{N^\alpha},
\]

which is \( \omega(1) \) as claimed.

We continue with (2.23). Recall that we reduced this problem to computing the limit of \(\mu_V(\Xi^{-1}(f_N)f_N')\). Using the inversion formula we see that

\[
\int \Xi^{-1}f_N(x)f_N'(x) d\mu_V(x) = -\frac{1}{\beta \pi^2} \int_a^b \int_a^b \frac{\sigma(x) f_N'(x)(f_N(x) - f_N(y))}{\sigma(y)(x - y)} \, dx \, dy
\]

54
Observe that
\[
\frac{1}{2} \sigma(x) \frac{\partial_x(f_N(x) - f_N(y))^2}{x - y} = f'_N(x)(f_N(x) - f_N(y)) \cdot \frac{\partial_x}{x - y} \left( \frac{\sigma(x)}{x - y} \right) = -\frac{1}{2} (a + b)(x + y) + ab + xy.
\]

Therefore, integration by parts yields
\[
\int \Xi^{-1} f_N(x) f'_N(x) d\mu_V(x) = -\frac{1}{2\beta \pi^2} \int_a^b \int_a^b \frac{\sigma(x) \partial_x(f_N(x) - f_N(y))^2}{\sigma(y)(x - y)} dxdy
\]
\[
= \frac{1}{2\beta \pi^2} \int_a^b \int_a^b \left( \frac{f_N(x) - f_N(y)}{x - y} \right)^2 \left( ab + xy - \frac{1}{2} (a + b)(x + y) \right) dxdy,
\]
By changing variables again to \((u, v) = (N^\alpha(x - E), N^\alpha(y - E))\) and observing that
\[
ab + xy - \frac{1}{2} (a + b)(x + y) = -\sigma(E)^2 + \frac{u + v}{N^\alpha} \left( a + b \right) + \frac{uv}{N^{2\alpha}}.
\]
we obtain
\[
\int \Xi^{-1} f_N(x) f'_N(x) d\mu_V(x) = -\frac{1}{2\beta \pi^2} \int \int A_N^2 \left( \frac{f(u) - f(v)}{u - v} \right)^2 \left( \frac{\sigma(E)^2 - \frac{u + v}{N^\alpha} \left( a + b \right) + E - \frac{uv}{N^{2\alpha}}}{\sigma(E + \frac{u + v}{N^\alpha})} \right) dudv.
\]
As before, \((f(u) - f(v))^2\) is zero for all \((u, v) \in (A_N \setminus B_N)^2\) for large enough \(N\), therefore we split the above integral into the regions defined in (2.26).

Notice that uniformly in \(u \in B_N\)
\[
\frac{1}{\sigma(E + \frac{u + v}{N^\alpha})} = \frac{1}{\sigma(E)} + O \left( \frac{|u|}{N^\alpha} \right),
\]
and further notice \((u + v)/N^\alpha\) and \(uv/N^{2\alpha}\) are bounded uniformly by constants in the entire region \(A_N \times A_N\) and converge pointwise to 0 for each \((u, v)\).

Consequently the integral (2.27) over the region \(B_N \times B_N\) is:
\[
\int \int_{B_N^2} \left( \frac{f(u) - f(v)}{u - v} \right)^2 \left( \frac{\sigma(E)^2 - \frac{u + v}{N^\alpha} \left( a + b \right) + E - \frac{uv}{N^{2\alpha}}}{\sigma(E + \frac{u + v}{N^\alpha})} \right) dudv
\]
\[
= \int \int_{B_N^2} \left( \frac{f(u) - f(v)}{u - v} \right)^2 \left( 1 - \frac{u + v}{N^\alpha \sigma(E)^2} \left( a + b \right) + E - \frac{uv}{N^{2\alpha} \sigma(E)^2} \right) dudv
\]
\[
+ O \left( \frac{1}{N^\alpha} \int \int_{B_N^2} \left( \frac{f(u) - f(v)}{u - v} \right)^2 (|u| + |v|) dudv \right),
\]
(2.28)
the first term of (2.28) is equal to,
\[ \frac{1}{2\beta \pi^2} \int \int \left( \frac{f(u) - f(v)}{u-v} \right)^2 \, dudv + O\left( \frac{1}{N^\alpha} \right) \]
while the second term in (2.28) can be written as
\[
\int \int_{B_N^2} \left( \frac{f(u) - f(v)}{u-v} \right)^2 (|u| + |v|)dudv = \int \int_{[-M;M]^2} \left( \frac{f(u) - f(v)}{u-v} \right)^2 (|u| + |v|)dudv \\
\quad + 2 \int_{-M}^M \int_{B_N \cap \{|u| \geq M\}} \left( \frac{f(v)}{u-v} \right)^2 (|u| + |v|)dudv,
\]
the integral over \([-M;M]^2\) is finite by differentiability of \(f\) while the second is bounded by
\[
\int_{-M}^M \int_{B_N \cap \{|u| \geq M\}} |f(v)|^2 \left( \frac{1}{|u-v|} + \frac{2|v|}{|u-v|^2} \right) \, dudv \\
\leq C \int_{-M}^M |f(v)|^2 \left( \frac{1}{|v-M|} + \frac{1}{|M+v|} + \log |v-M||v+M| \right) \leq C \log N
\]
since supp \(f \subset [-M, M]\).

In the region \((u,v) \in B_N \times (A_N \setminus B_N)\), \(\sigma(E + \frac{v}{N^\alpha})\) is bounded above and below while, for \(N\) large enough \(f(v) = 0\), thus the integral over \(B_N \times (A_N \setminus B_N)\) is bounded above by
\[
\int_{A_N \setminus B_N} \int_{B_N} \left( \frac{f(u) - f(v)}{u-v} \right)^2 \left( \frac{1}{\sigma(E + \frac{u}{N^\alpha})\sigma(E + \frac{v}{N^\alpha})} \right) \, dudv \\
\leq \int_{A_N \setminus B_N} \int_{-M}^M \left( \frac{f(u)}{u-v} \right)^2 \frac{1}{\sigma(E + \frac{v}{N^\alpha})} \, dudv \\
\leq \frac{C}{N^{2\alpha}} \int_{A_N \setminus B_N} \frac{1}{\sigma(E + \frac{v}{N^\alpha})} \, dv \leq \frac{C}{N^\alpha},
\]
where in the second line we used \(|u-v| \geq cN^\alpha\) for \(u \in [-M;M]\) and \(v \in A_N \setminus B\). By symmetry of the integrand in (2.27) this argument extends to the region \((u,v) \in (A_N \setminus B_N) \times B_N\).

Altogether, our bounds show
\[
\int \Xi^{-1} f_N(x) f'_N(x) d\mu_N(x) = - \frac{1}{2\beta \pi^2} \int \int \left( \frac{f(x) - f(y)}{x-y} \right)^2 \, dxdy + O\left( \frac{\log N}{N^\alpha} \right),
\]
which shows (2.23).

We conclude by proving (2.24). The proof will be similar to the proof of Lemma 2.13. As in Lemma 2.13 we may restrict our attention to the event \(\Omega = \{\forall i : |\lambda_i - \gamma_i| \leq \)
\( N^{-\frac{\xi}{2}+\varepsilon} \) by applying Theorem 2.11. Further, we use again the sets \( J_1, J_2, \) and \( J_3 \) defined in Lemma 2.12.

The general idea will be that we can use the uniform bounds (2.10) for particles close to the bulk point \( E \) (corresponding to the indices in \( J_1 \)), and control the number of such particles. In the intermediary regime we will use the bounds (2.12) or the explicit formula (2.13). On the other hand, for the particles far away from \( E \) (corresponding to \( J_3 \)) we can use the uniform decay of \( \Xi^{-1}f_N \) and its derivative by (2.11) and (2.12).

Define for \( j \in \{1, 2, 3\} \):

\[
M^{(j)}_N = \sum_{i \in J_j} \left( \delta_{x_i} - N1_{[\gamma_i, \gamma_{i+1}]} \mu \right)
\]

so that \( M_N = M^{(1)}_N + M^{(2)}_N + M^{(3)}_N \). We can write

\[
\int \int \frac{\Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y)}{x - y} dM_N(x) dM_N(y) = \sum_{1 \leq j_1, j_2 \leq 3} \int \int \frac{\Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y)}{x - y} dM_N^{(j_1)}(x) dM_N^{(j_2)}(y)
\]

Integrating repeatedly for each \((j_1, j_2)\), and using that \( N\mu(x) = 1 \) for all indices \( i \) yields:

\[
\int \int \frac{\Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y)}{x - y} dM_N^{(j_1)}(x) dM_N^{(j_2)}(y) = \sum_{i_1 \in J_{j_1}, i_2 \in J_{j_2}} \int_{\gamma_i} d\mu(x_1) \int_{\gamma_i} d\mu(x_2) \int_T dudv dt \left\{ (\lambda_{i_1} - x_1)(\lambda_{i_2} - x_2)t(1-t) \right. \\
\left. \times \Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y) + ut(x_2 - \lambda_{i_2}) + u(\lambda_{i_2} - x_2) + t(x_1 - x_2) + x_2 \right\}
\]

where \( T = [0, 1]^3 \). We will bound (2.29) for each pair \((j_1, j_2)\).

For \((j_1, j_2) = (1, 1)\). Recall by Lemma 2.12 (c) that \( |J_i| \leq CN^{1-\alpha} \), and further from the proof of Lemma 2.12 uniformly in \( i \in J_1, |\lambda_i - x| \leq CN^{\xi-1} \) whenever \( x \in [\gamma_i, \gamma_{i+1}] \). We use (2.29), Lemma 2.10 (2.10) to obtain the upper bound

\[
\int \int \frac{\Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y)}{x - y} dM_N^{(1)}(x) dM_N^{(1)}(y) \leq \sum_{i_1 \in J_{i_1}, i_2 \in J_{i_2}} \int_{\gamma_{i_1}} \int_{\gamma_{i_2}} N^{3\alpha} \log N |\lambda_{i_1} - x_1||\lambda_{i_2} - x_2| d\mu(x_1) d\mu(x_2) \leq CN^{2\xi+\alpha} \log N,
\]

which is \( o(1) \) when divided by \( N \).
For \((j_1, j_2) = (3,3)\). We bound as in the previous case, using \((2.12)\) and that uniformly uniformly in \(i \in J_3\), \(|\lambda_i - x| \leq C \leq N^{-\frac{3}{2}} + \epsilon_i^{-\frac{1}{2}}\)

\[
\int \int \frac{\Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y)}{x-y} dM_N^{(3)}(x) dM_N^{(3)}(y) \leq \\
N^2 \sum_{i_1 \in J_3, i_2 \in J_3} \int_{\gamma_{i_1}}^{\gamma_{i_1}+1} \int_{\gamma_{i_2}}^{\gamma_{i_2}+1} \log N \frac{N^\alpha}{N^\alpha} |\lambda_{i_1} - x_1||\lambda_{i_2} - x_2| d\mu_N(x_1) d\mu_N(x_2) \leq CN^{2\xi - \alpha} \log N,
\]

which is \(\omega(1)\) when divided by \(N\).

For \((j_1, j_2) = (2,2)\). We remark that the strategy is not as straightforward as the case \(i \in J_2\) in the proof of Lemma 2.13 \((2.19)\), this is because the term \(t(x_1 - x_2) + x_2\) appearing as an argument in \((2.29)\) may enter a neighborhood of \(E\) depending on the indices \(i_1, i_2 \in J_2\); so we may not use the bound Lemma 2.10 \((2.11)\) uniformly in \(i_1, i_2 \in J_2\). Some care is needed also because \(M_N\) is a signed measure so \(|M_N(g)|\) need not be bounded by \(M_N(|g|)\).

It will be convenient to use directly \((2.13)\) from the proof of Lemma 2.10 \((\text{this can be done as } J_2 \text{ is located outside the support of } f)\). We can write for \(x, y \in \{z \in \text{supp } \mu, N^\alpha |z - E| > M + 1\}\)

\[
\Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y) \leq \\
\frac{1}{\beta \pi^2} \int_{-M}^{M} \frac{f(u)}{\sigma(E + \frac{u}{N^\alpha})(x-y)} \left( \frac{1}{S(y)(u-N^\alpha(y-E))} - \frac{1}{S(x)(u-N^\alpha(x-E))} \right) du \\
= \frac{1}{\beta \pi^2} \int_{-M}^{M} \frac{f(u)}{\sigma(E + \frac{u}{N^\alpha})} \left\{ \frac{S(x)-S(y)}{(x-y)} \frac{1}{S(x)S(y)(u-N^\alpha(y-E))} \\
+ \frac{N^\alpha}{S(x)(u-N^\alpha(x-E))(u-N^\alpha(y-E))} \right\} du.
\]

(2.30)

When we integrate the term on the third line of \((2.30)\) against \(M_N^{(2)} \otimes M_N^{(2)}\), we obtain

\[
\int_{-M}^{M} \frac{f(u)}{\sigma(E + \frac{u}{N^\alpha})} \left\{ \int M_N^{(2)} \left( \int_{0}^{1} \frac{S'(t(\cdot - y) + y)}{S(\cdot)S(y)} dt \right) \frac{1}{(u-N^\alpha(y-E))} dM_N^{(2)}(y) \right\} du,
\]

(2.31)

define the function

\[
g(y) := M_N^{(2)} \left( \int_{0}^{1} \frac{S'(t(\cdot - y) + y)}{S(\cdot)S(y)} dt \right),
\]

58
first, \( g(y) \) is bounded for any \( y \in [a; b] \):

\[
\left| M_N^{(2)} \left( \int_0^1 \frac{S'(t(-y) + y)}{S(t)S(y)} \, dt \right) \right| = \frac{N}{S(y)} \sum_{i \in J_2} \int_{\gamma_i} \int_0^1 \frac{S'(t(\lambda_i - y) + y) - S'(t(x - y) + y)}{S(\lambda_i)} \, dt \, d\mu_V(x) \leq \frac{N}{S(y)} \sum_{i \in J_2} \int_{\gamma_i} \int_0^1 \frac{S'(t(\lambda_i - y) + y) - S'(t(x - y) + y)}{S(\lambda_i)} \, dt \, d\mu_V(x) \leq CN^\xi,
\]

where in the final line we used \( S \) and \( S' \) are smooth on \( [a; b] \) (and therefore uniformly Lipschitz), \( S > 0 \) in a neighborhood of \( [a; b] \), further \( |x - \lambda_i| \leq CN^{\xi-1} \), and \( |J_2| \leq CN \). Moreover, \( g(y) \) is uniformly Lipschitz in \( [a; b] \) with constant \( CN^\xi \), since:

\[
M_N^{(2)} \left( \int_0^1 \frac{S'(t(-y) + y)}{S(t)S(y)} \, dt \right) = (z - y)M_N^{(2)} \left( \int_0^1 \frac{tS''(ut - y) + t(-z + y)}{S(t)S(y)} \, dt \right) \leq (z - y)M_N^{(2)} \leq CN^\xi.
\]

and both terms appearing in \( M_N^{(2)} \) above are of the same form as \( g \) so they are bounded by \( CN^\xi \). Returning to (2.31), we may bound

\[
\left| M_{N}^{(2)} \left( \frac{g(y)}{u - N^\alpha(y - E)} \right) \right| = \left| N \sum_{i \in J_2} \int_{\gamma_i} \frac{g(\lambda_i) - g(x)}{(u - N^\alpha(\lambda_i - E))} \, d\mu_V(x) \right| \leq \int_{[a,b] \cap \{|x - E| \geq \frac{2M}{N} \}} \frac{CN^{2\xi}}{|u - N^\alpha(x - E)|} + \frac{CN^{2\xi+\alpha}}{(u - N^\alpha(x - E))^2} \, dx \leq CN^{2\xi-\alpha} \log N + CN^{2\xi},
\]

uniformly in \( u \in [-M; M] \). Thus (2.31) is bounded by \( CN^{2\xi} \) as \( f \) is bounded.

The remaining term in (2.30) is

\[
N^\alpha \int_{-M}^{M} \frac{f(u)}{\sigma(E + \frac{u}{N^\alpha})} M_N^{(2)} \left( \frac{1}{S(t)S(\lambda_i - E)} \right) M_N^{(2)} \left( \frac{1}{u - N^\alpha(\cdot - E)} \right) \, du. \tag{2.32}
\]
Repeating our argument in the previous paragraph gives:

\[
\left| M_N^{(2)} \left( \frac{1}{S(\cdot)(u - N^\alpha(\cdot - E))} \right) \right| \leq CN^{-\alpha} \log N + CN^\xi, \\
\left| M_N^{(2)} \left( \frac{1}{u - N^\alpha(\cdot - E)} \right) \right| \leq CN^\xi,
\]

where in the first inequality we use $1/S$ is uniformly bounded and uniformly Lipschitz on $[a; b]$. Inserting the bounds into (2.32) gives an upper bound of $CN^{2\xi + \alpha}$, as $f$ is bounded.

Altogether

\[
\left| \Xi^{-1}(f_N)(x) - \Xi^{-1}(f_N)(y) \right| \leq CN^{2\xi + \alpha},
\]

which is $o(1)$ when divided by $N$.

For $(j_1, j_2) = (1, 2)$. By the bounds $|\lambda_j - \gamma_j| \leq CN^{\xi - 1}$, $|\gamma_j - x_j| \leq \frac{C}{N}$ for $x_j \in [\gamma_j; \gamma_{j+1}]$, whenever

\[
N^\alpha |tv(\lambda_{i_1} - x_1) + ut(x_2 - \lambda_{i_2}) + t(x_1 - x_2) + u(\lambda_{i_2} - x_2) + x_2 - E| \geq M + 1, \tag{2.33}
\]

we have

\[
|t(\gamma_{j_1} - \gamma_{j_2}) + (\gamma_{j_2} - E)| + CN^{\xi - 1} \geq M + 1,
\]

by triangle inequality. It follows that for $N$ sufficiently large, uniformly in $x_1 \in [\gamma_{i_1}; \gamma_{i_1+1}]$, $x_2 \in [\gamma_{i_2}; \gamma_{i_2+1}]$, $u, v \in [0; 1]$

\[
\frac{1}{|tv(\lambda_{i_1} - x_1) + ut(x_2 - \lambda_{i_2}) + t(x_1 - x_2) + u(\lambda_{i_2} - x_2) + x_2 - E|} \leq \frac{C}{|t(\gamma_{i_1} - \gamma_{i_2}) + (\gamma_{i_2} - E)|},
\]

where the constant $C$ only depends on $M$. Therefore, whenever (2.33) is satisfied, applying Lemma 2.10 (2.11) yields

\[
\left| \Xi^{-1}(f_N)^{(3)}(tv(\lambda_{i_1} - x_1) + ut(x_2 - \lambda_{i_2}) + t(x_1 - x_2) + u(\lambda_{i_2} - x_2) + x_2) \right| \leq \frac{C}{N^\alpha(t(\gamma_{i_1} - \gamma_{i_2}) + \gamma_{i_2} - E)^4}. \tag{2.34}
\]

Now, let $t \in (0, 1)$ fixed and define the sets

\[
K_t^1 := \left\{ j \in J_2 : t \left( E - \frac{2M}{N^\alpha} - \gamma_j \right) + \gamma_j - E \geq \frac{2M}{N^\alpha} \right\}, \\
K_t^2 := \left\{ j \in J_2 : t \left( E + \frac{2M}{N^\alpha} - \gamma_j \right) + \gamma_j - E \leq -\frac{2M}{N^\alpha} \right\}, \\
K_t := K_t^1 \cup K_t^2.
\]
By construction, if \( i_2 \in K^t_1 \) then

\[
|t(\gamma_{i_1} - \gamma_{i_2}) + (\gamma_{i_2} - E)| \geq \frac{2M}{N^\alpha}
\]

uniformly in \( i_1 \in J_1 \). Thus for such \( i_2 \in K^t_1 \), (2.33) is satisfied for \( N \) sufficiently large (uniformly in \( u, v, x_1 \), and \( x_2 \); also the choice of how large \( N \) must be only depends on \( \xi \) and \( \mu_V \)). The same statement holds for \( K^t_2 \).

We now proceed to bound (2.29) for \( j_1 = 1 \) and \( j_2 = 2 \) by splitting \( J_2 \) into the regions \( K^t_1, K^t_2 \) and \( J_2 \setminus K_t \). We start with \( K^t_1 \) (the argument for \( K^t_2 \) is identical). Our observations from the previous paragraph along with (2.34) gives:

\[
\int_T du dv dt \left| N^2 \sum_{\substack{i_1 \in J_1 \\ i_2 \in K^t_1}}^{\gamma_{i_1}+1} d\mu_V(x_1) \int_{\gamma_{i_1}+1}^{\gamma_{i_2}+1} d\mu_V(x_2) \left\{ (\lambda_{i_1} - x_1)(\lambda_{i_2} - x_2)t(1-t) \right. \\
\quad \times \Xi^{-1}(f_N)^{(3)}(tv(\lambda_{i_1} - x_1) + ut(x_2 - \lambda_{i_2}) + u(\lambda_{i_2} - x_2) + t(x_1 - x_2) + x_2) \left. \right\} \right| \\
\leq \int_0^1 \sum_{\substack{i_1 \in J_1 \\ i_2 \in K^t_1}}^{\gamma_{i_1}+1} \frac{CN^{2\xi-2-\alpha}t(1-t)}{(t(\gamma_{i_1} - \gamma_{i_2}) + (\gamma_{i_2} - E))^4} dt \leq \int_0^1 \sum_{i_2 \in K^t_1}^{\gamma_{i_2}+1} \frac{CN^{2\xi-1-2\alpha}t(1-t)}{((1-t)(\gamma_{i_2} - E) - \frac{2M}{N^\alpha})^4} dt
\]

where in the final line we used \( |J_1| \leq CN^{1-\alpha} \) from Lemma 2.12 (c). Next, note that

\[
\frac{1}{N} \sum_{i_2 \in K^t_1}^{\gamma_{i_2}+1} \frac{1}{(1-t)(\gamma_{i_2} - E) - \frac{2M}{N^\alpha})^4} \leq C \int_{E+\frac{1}{2}(E-a) \wedge (b-E)}^{E+\frac{1}{2}(E-a) \wedge (b-E)} \frac{dx}{(1-t)(x - E) - \frac{2M}{N^\alpha})^4} \\
\quad \leq \frac{CN^{3\alpha}}{1-t},
\]

since, by definition of \( K^t_1 \), \( \gamma_{i_2} \geq E + \frac{2M}{N^\alpha} \left( \frac{1+t}{1-t} \right) \). We conclude,

\[
\int_0^1 \sum_{i_2 \in K^t_1}^{\gamma_{i_2}+1} \frac{CN^{2\xi-1-2\alpha}t(1-t)}{((1-t)(\gamma_{i_2} - E) - \frac{2M}{N^\alpha})^4} dt \leq CN^{2\xi+\alpha}.
\]

We continue with \( J_2 \setminus K_t \). By the same argument as in Lemma 2.12 (c) \( |J_2 \setminus K_t| \leq \frac{CN^{1-\alpha}}{1-t} \) where the constant \( C \) does not depend on \( t \), we use this in addition with Lemma 2.10 (2.11), \( |J_1| \leq CN^{1-\alpha} \), and \( |\lambda_{i_j} - x_j| \leq CN^{\xi-1} \) to obtain the bound

\[
\int_T du dv dt \left| N^2 \sum_{\substack{i_1 \in J_1 \\ i_2 \in J_2 \setminus K_t}}^{\gamma_{i_1}+1} d\mu_V(x_1) \int_{\gamma_{i_1}+1}^{\gamma_{i_2}+1} d\mu_V(x_2) \left\{ (\lambda_{i_1} - x_1)(\lambda_{i_2} - x_2)t(1-t) \right. \\
\quad \times \Xi^{-1}(f_N)^{(3)}(tv(\lambda_{i_1} - x_1) + ut(x_2 - \lambda_{i_2}) + u(\lambda_{i_2} - x_2) + t(x_1 - x_2) + x_2) \left. \right\} \right| \\
\leq C \int_0^1 N^{3\alpha} \log N \times N^{2\xi-2} \times N^{2-2\alpha} t dt \leq CN^{\alpha+2\xi} \log N.
\]
Combining the bounds we have obtained gives
\[
\left| \int \int \frac{- \Xi^{-1} f_N(x) - \Xi^{-1} f_N(y)}{x - y} dM^{(1)}_N(x) dM^{(2)}_N(y) \right| \leq CN^{\alpha + 2\xi} \log N,
\]
which is \(\omega(1)\) when divided by \(N\) for \(\xi\) small enough.

For \(j_1 = 1\) or 2 and \(j_2 = 3\), the proof is similar and we omit the details.

\[\square\]

§2.4 Proof of Theorem 2.5.
We proceed with the proof of Theorem 2.5. As we did in the sketch of the proof, by (2.7) applied to \(h = \Xi^{-1}(f_N)\) yields

\[
F^N_1(\Xi^{-1}(f_N)) = M_N(f_N) + \left(1 - \frac{\beta}{2}\right) L_N((\Xi^{-1}f_N)'),
\]
\[
+ \frac{1}{N} \left[ \beta \int \int \frac{- \Xi^{-1} f_N(x) - \Xi^{-1} f_N(y)}{x - y} dM_N(x) dM_N(y) \right].
\]

Combining Lemma 2.14 (2.22) and (2.24) we can bound the two terms on the right hand side to get

\[
F^N_1(\Xi^{-1}(f_N)) = M_N(f_N) + \omega(1).
\]

We consider an event \(A_1\) of probability higher than \(1 - \epsilon^{-N^c}\) on which

\[
\left| F^N_1(\Xi^{-1}(f_N)) - M_N(f_N) \right| \leq \frac{C}{N^\delta},
\]

for some positive constants \(c, C\) and \(\delta\). Using the first loop equation from Proposition 2.6, and the trivial deterministic bounds

\[
M_N(f_N) = O\left(N\|f\|_\infty\right), \quad F^N_1(\Xi^{-1}(f_N)) = O\left(N(\|f\|_\infty + \|\Xi^{-1}(f_N)\|_{C^1(R)})\right) = O(N^2),
\]

we obtain

\[
0 = E^N_V(F^N_1(\Xi^{-1}(f_N))) = E^N_V(F^N_1(\Xi^{-1}(f_N))1_{A_1}) + E^N_V(F^N_1(\Xi^{-1}(f_N))1_{A_1^c})
\]
\[
= E^N_V(M_N(f_N)1_{A_1}) + o(1) + O(N^2 \mathbb{P}_V(A_1^c))
\]
\[
= E^N_V(M_N(f_N)) + o(1),
\]

and thus

\[
E^N_V(M_N(f_N)) = o(1).
\]

We now show recursively that

\[
F^N_k(\Xi^{-1}(f_N), f_N, \cdots, f_N) = \tilde{M}_N(f_N)^k - (k - 1)\sigma_f^2 \tilde{M}_N(f_N)^{k-2} + \omega(1).
\]
Here, the set on which the bound holds might vary from one \( k \) to another but each bound has probability greater than \( 1 - e^{-N^k} \) for each fixed \( k \).

The bound holds for \( k = 1 \), by (2.35). Now, assume this holds for \( k \geq 1 \). On a set of probability greater than \( 1 - e^{-N^{k+1}} \) we have by the induction hypothesis, Lemma 2.13 (2.18), and Lemma 2.14 (2.23), for some \( \delta > 0 \) and a constant \( C \)

\[
|F_k^N(\Xi^{-1}(f_N), f_N, \ldots, f_N) - \tilde{M}_N(f_N)^k + (k-1)f_2 \tilde{M}_N(f_N)^{k-2}| \leq \frac{C}{N^\delta},
\]

\[
|L_N(\Xi^{-1}(f_N)f_N') + \sigma_I^2| \leq \frac{C}{N^\delta},
\]

\[
|M_N(f_N)| \leq N^{\delta/2k}.
\]

On this set, using the definition of \( F_{k+1}^N \) from Proposition 2.6,

\[
F_{k+1}^N(\Xi^{-1}(f_N), f_N, \ldots, f_N) = \frac{F_k^N(\Xi^{-1}(f_N), f_N, \ldots, f_N)\tilde{M}_N(f_N)^k + \tilde{M}_N(f_N)^{k-1}L_N(\Xi^{-1}(f_N)f_N')}{\tilde{M}_N(f_N)^k} = \tilde{M}_N(f_N)^{k+1} - (k-1)f_2 \tilde{M}_N(f_N)^{k-2} + O\left(\frac{1}{N^\delta}\right)\tilde{M}_N(f_N)
\]

\[
+ \tilde{M}_N(f_N)^{k-1}\left(-\sigma_I^2 + O\left(\frac{1}{N^\delta}\right)\right)
\]

\[
= \tilde{M}_N(f_N)^{k+1} - (k-1)f_2 \tilde{M}_N(f_N)^{k-2} + O\left(\frac{1}{N^\delta}\right)
\]

and this proves the induction. Using the fact that \( F_k \) is bounded polynomially and deterministically as before, we see that for any \( k \geq 1 \)

\[
E_N^V\left(M_N(f_N)^{k+1}\right) = \sigma_I^2 k E_N^V\left(M_N(f_N)^{k-1}\right) + o(1).
\]

Coupled with (2.36), the computation of the moments is then straightforward and we obtain for all \( k \in \mathbb{N} \)

\[
E_N^V\left(M_N(f_N)^{2k}\right) = \sigma_I^{2k} \frac{(2k)!}{2^k k!} + o(1)
\]

\[
E_N^V\left(M_N(f_N)^{2k+1}\right) = o(1)
\]

This concludes the proof of Theorem 2.5.

**Remark 2.15.** The same proof would also show the macroscopic central limit Theorem already shown in [BG13b, Shc13, J] with appropriate decay conditions on \( f \).
PART 3

MARČenko–Pastur Law for Kendall’s Tau

Based on joint work with A. S. Bandeira and P. Rigollet

Acknowledgments: The authors thank Alice Guionnet for helpful comments and discussions. A. S. B. was supported by Grant NSF-DMS-1541100. Part of this work was done while A. S. B. was with the Department of Mathematics at MIT. A. L. was supported by Grant NSF-MS-1307704. P. R was supported by Grants NSF-DMS-1541099, NSF-DMS-1541100 and DARPA-BAA-16-46 and a grant from the MIT NEC Corporation.
§3.1 Introduction

Estimating the association between two random variables $X, Y \in \mathbb{R}$ is a central statistical problem. As such many methods have been proposed, most notably Pearson's correlation coefficient. While this measure of association is well suited to the Gaussian case, it may be inaccurate in other cases. This observation has led statisticians to consider other measures of associations such as Spearman's $\rho$ and Kendall's $\tau$ that can be proved to be more robust to heavy-tailed distributions (see, e.g., [LHY+12]). In a multivariate setting, covariance and correlation matrices are preponderant tools to understand the interaction between variables. They are also used as building blocks for more sophisticated statistical questions such as principal component analysis or graphical models.

The past decade has witnessed an unprecedented and fertile interaction between random matrix theory and high-dimensional statistics (see [PA14] for a recent survey). Indeed, in high-dimensional settings, traditional asymptotics where the sample size tends to infinity fail to capture a delicate interaction between sample size and dimension and random matrix theory has allowed statisticians and practitioners alike to gain valuable insight on a variety of multivariate problems.

The terminology "Wishart matrices" is often, though sometimes abusively, used to refer to $p \times p$ random matrices of the form $X^T X/n$, where $X$ is an $n \times p$ random matrix with independent rows (throughout this paper we restrict our attention to real random matrices). The simplest example arises where $X$ has i.i.d standard Gaussian entries but the main characteristics are shared by a much wider class of random matrices. This universality phenomenon manifests itself in various aspects of the limit distribution, and in particular in the limiting behavior of the empirical spectral distribution of the matrix. Let $W = X^T X/n$ be a $p \times p$ Wishart matrix and denote by $\lambda_1, \ldots, \lambda_p$ its eigenvalues; then the empirical spectral distribution $\hat{\mu}_p$ of $W$ is the distribution on $\mathbb{R}$ defined as the following mixture of Dirac point masses at the $\lambda_j$s:

$$
\hat{\mu}_p = \frac{1}{p} \sum_{k=1}^{p} \delta_{\lambda_k}.
$$

Assuming that the entries of $X$ are independent, centered and of unit variance, it can be shown that $\mu_p$ converges weakly to the Marčenko-Pastur distribution under weak moment conditions (see [EKYY12] for the weakest condition).

While this development alone has led to important statistical advances, it fails to capture more refined notions of correlations, notably more robust ones involving ranks and therefore dependent observations. A first step in this direction was made by [YK86], where the matrix $X$ is assumed to have independent rows with isotropic distribution. More recently, this result was extended in [BZ08, O'R12] and covers for example the case of Spearman’s $\rho$ matrix that is based on ranks, which is also a Wishart matrix of the form $X^T X/n$.

The main contribution of this paper is to derive the limiting distribution of Kendall’s $\tau$ matrix, a cousin of Spearman’s $\rho$ matrix but which is not of the Wishart type but rather a matrix whose entries are $U$-statistics. Kendall’s $\tau$ matrix is a very popular surrogate for correlation matrices but an understanding the fluctuations of its eigenvalues is still missing.
Interestingly, Marčenko-Pastur results have been used as heuristics, without justification, precisely for Kendall’s \(\tau\) in the context of certain financial applications [CCL+15].

As it turns out, the limiting distribution of \(\hat{\mu}_p\) is not exactly Marčenko-Pastur, but rather an affine transformation of it. Our main theorem below gives the precise form of this transformation.

**Theorem 3.1.** Let \(X_1, \ldots, X_n,\) be \(n\) independent random vectors in \(\mathbb{R}^p\) whose components \(X_i(k)\) are independent random variables that have a density with respect to the Lebesgue measure on \(\mathbb{R}\). Then as \(n \to \infty\) and \(\frac{p}{n} \to \gamma > 0\) the empirical spectral distribution of \(\tau\) converges in probability to

\[
\frac{1}{3} Y + \frac{1}{3},
\]

where \(Y\) is distributed according to the standard Marčenko-Pastur law with parameter \(\gamma\) (see Theorem D.1 for the appropriate definition).

**Notation:** For any integer \(k \geq 1\) we write \([k] = \{1, \ldots, k\}\). We denote by \(I_p\) the identity matrix of \(\mathbb{R}^p\). For a vector \(x \in \mathbb{R}^p\), we denote by \(x(j)\) its \(j\)th coordinate. For any \(p \times p\) matrix \(M\), we denote by \(\text{diag} (M)\) the \(p \times p\) diagonal matrix with the same diagonal elements as \(M\) and any real number \(r\), we define \(D_r(M) = M - \text{diag} (M) + rI_p\). In other words, the operator \(D_r\) replaces each diagonal element of a matrix by the value \(r\). We denote sign the sign function with convention that \(\text{sign}(0) = 1\).

We define the Frobenius norm of a \(p \times p\) matrix \(H\) as \(\|H\|_F^2 := \text{Tr} (H^T H)\). Finally, we define \(\text{Unif}([a, b])\) to be the uniform distribution on the interval \([a, b]\).

**§3.2 Kendall’s Tau**

The (univariate) Kendall \(\tau\) statistic [Ess24, Lin25, Lin29, Ken38] is defined as follows. Let \((Y_1, Z_1), \ldots, (Y_n, Z_n)\) be \(n\) independent samples of a pair \((Y, Z) \in \mathbb{R} \times \mathbb{R}\) of real-valued random variables. Then the (empirical) Kendall \(\tau\) between \(Y\) and \(Z\) is defined as

\[
\tau(Y, Z) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign}(Y_i - Y_j) \cdot \text{sign}(Z_i - Z_j).
\]

The statistic \(\tau\) takes values in \([-1, 1]\) and it is not hard to see that it can be expressed as

\[
\tau = \frac{1}{\binom{n}{2}} (\# \{\text{concordant pairs}\} - \# \{\text{discordant pairs}\}),
\]

Where a pair \((i, j)\) is said to be *concordant* if \(Y_i - Y_j\) and \(Z_i - Z_j\) have the same sign and *discordant* otherwise.

It is known that the Kendall \(\tau\) statistic is asymptotically Gaussian (see, e.g., [Ken38]). Specifically, if \(Y\) and \(Z\) are independent, then as \(n \to \infty\),

\[
\sqrt{n} \tau(Y, Z) \Rightarrow \mathcal{N} \left(0, \frac{4}{9}\right).
\]
This property has been central to construct independence tests between two random variables \( X \) and \( Y \) (see, e.g., [KG90]).

Kendall's \( \tau \) statistic can be extended to the multivariate case. Let \( X_1, \ldots, X_n \), be \( n \) independent copies of a random vector \( X \in \mathbb{R}^p \), with independent coordinates \( X(1), \ldots, X(p) \). The (empirical) Kendall \( \tau \) matrix of \( X \) is defined to be the \( p \times p \) matrix whose entries \( \tau_{kl} \) are given by

\[
\tau_{kl} := \tau(X(k), X(l)) = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign}\left((X_i(k) - X_j(k)) \cdot (X_i(l) - X_j(l))\right) 1 \leq k, l \leq p. \tag{3.2}
\]

Note that the \( \tau \) can be written as the sum of \( \binom{n}{2} \) rank-one random matrices:

\[
\tau = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \text{sign}(X_i - X_j) \otimes \text{sign}(X_i - X_j), \tag{3.3}
\]

where the sign function is taken entrywise.

It is easy to see that \( \tau_{ii} = 1 \) for all \( i \). Together with (3.1), it implies that the matrix

\[
\tilde{\tau} = \frac{3}{2} \tau - \frac{1}{2} I_p
\]

is such that \( \mathbb{E}[\sqrt{n} \tilde{\tau}] \to I_p \) and \( \text{Var}[\tilde{\tau}_{ij}] \to 1 \{i \neq j\} \) as \( n \to \infty \). This suggests that if the empirical spectral distribution of \( \tilde{\tau} \) converges to a Marčenko-Pastur distribution, it should be a standard Marčenko-Pastur distribution. This heuristic argument supports the affine transformation arising in Theorem 3.1. However, the matrix \( \tau \) is not Wishart and the Marčenko-Pastur limit distribution does not follow from standard arguments. Nevertheless, Kendall's \( \tau \) is a \( U \)-statistic which are known to satisfy the weakest form of universality, namely a Central Limit Theorem under general conditions [Hoe48, dIPG99]. In this paper, we show that in the case of the Kendall \( \tau \) matrix, this universality phenomenon extends to the empirical spectral distribution.
§3.3 Proof of Theorem 3.1.

For any pair \((i, j)\) such that \(1 \leq i, j \leq n\) and \(i \neq j\), let \(A_{(i,j)}\) be the vector

\[ A_{(i,j)} := \text{sign}(X_i - X_j), \]

and recall from (3.3) that

\[ \tau = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} A_{(i,j)} \otimes A_{(i,j)}. \]

Akin to most asymptotic results on U-statistics, we utilize a variant of Hoeffding’s (a.k.a. Efron-Stein, a.k.a ANOVA) decomposition [Hoe48]:

\[ A_{(i,j)} = \tilde{A}_{(i,j)} + \tilde{A}_{(i,\cdot)} + \tilde{A}_{(\cdot,j)} \tag{3.4} \]

where

\[ \tilde{A}_{(i,\cdot)} := E[A_{(i,j)}|X_i], \quad \tilde{A}_{(\cdot,j)} := E[A_{(i,j)}|X_j] \quad \text{and} \quad \tilde{A}_{(i,j)} := A_{(i,j)} - \tilde{A}_{(i,\cdot)} - \tilde{A}_{(\cdot,j)}. \]

It is easy to check that each of the vectors in the right-hand side of (3.4) are centered and are orthogonal to each other with respect to the inner product \(E[v^T w]\) where \(v, w \in \mathbb{R}^p\). These random vectors can be expressed conveniently thanks to the following Lemma.

**Lemma 3.2.** For \(k \in [p]\), let \(F_k\) denote the cumulative distribution function of \(X(k)\). Fix \(i \in [n]\) and let \(U_i \in \mathbb{R}^p\) is a random vector with kth coordinate given by \(U_i(k) = 2F_k(X_i(k)) - 1 \sim \text{Unif}([-1, 1])\). Then

\[ \tilde{A}_{(i,\cdot)} = -\tilde{A}_{(\cdot,i)} = U_i. \]

**Proof.** For any \(i \in [n]\), observe that since the components of \(X\) have a density, then

\[ P(X_i(k) = X_j(k)|X_i) = 0 \]

so that

\[ E[\text{sign}(X_i(k) - X_j(k))|X_i] = P(X_i(k) > X_j(k)|X_i) - P(X_i(k) < X_j(k)|X_i) = 2F_k(X_i(k)) - 1. \]

The observation that \(\tilde{A}_{(i,\cdot)} = -\tilde{A}_{(\cdot,i)}\) follows from the fact that \(A_{(i,j)} = -A_{(j,i)}\). ■

Using (3.4) we obtain the equality:

\[ A_{(i,j)} \otimes A_{(i,j)} = M^{(1)}_{(i,j)} + M^{(2)}_{(i,j)} + (M^{(2)}_{(i,j)})^T + M^{(3)}_{(i,j)}, \tag{3.5} \]

where

\[ M^{(1)}_{(i,j)} := I_p + D_0[\{\tilde{A}_{(i,\cdot)} + \tilde{A}_{(\cdot,j)}\} \otimes \{\tilde{A}_{(i,\cdot)} + \tilde{A}_{(\cdot,j)}\}], \]

\[ M^{(2)}_{(i,j)} := D_0[\tilde{A}_{(i,j)} \otimes \{\tilde{A}_{(i,\cdot)} + \tilde{A}_{(\cdot,j)}\}], \]

\[ M^{(3)}_{(i,j)} := D_0[\tilde{A}_{(i,j)} \otimes \tilde{A}_{(i,j)}]. \]
By the relation $\tilde{A}_{(i,.)} = -\tilde{A}_{(.,i)}$ from Lemma 3.2 we have

$$
\sum_{1 \leq i < j \leq n} \{\tilde{A}_{(i,.)} + \tilde{A}_{(.,j)}\} \otimes \{\tilde{A}_{(i,.)} + \tilde{A}_{(.,j)}\} =
(n-1) \sum_{i=1}^{n} \tilde{A}_{(i,.)} \otimes \tilde{A}_{(i,.)} - \sum_{(i,j) \in [n]^2; i \neq j} \tilde{A}_{(i,.)} \otimes \tilde{A}_{(j,.)}.
$$

Using Lemma 3.2 yields:

$$
\frac{1}{(n^2)} \sum_{1 \leq i < j \leq n} M_{(i,j)}^{(1)} = I_p + \frac{2}{n} \sum_{i=1}^{n} D_0[U_i \otimes U_i] - \frac{1}{(n^2)} D_0 \left[ \sum_{(i,j) \in [n]^2; i \neq j} U_i \otimes U_j \right]. \quad (3.6)
$$

Next, note that, the coordinates of each $U_i, i = 1, \ldots, n$ are mutually independent so that $E[U_i] = 0$ and

$$
E[U_i \otimes U_i] = E[T^2] I_p = \frac{1}{3} I_p,
$$

where $T \sim Unif([-1, 1])$. Theorem D.1 implies as $n \to \infty$ and $\frac{p}{n} \to \gamma > 0$, the empirical spectral distribution of

$$
\frac{2}{n} \sum_{i=1}^{n} U_i \otimes U_i
$$

converges in probability to $(2/3) Y$, where $Y$ is distributed according to the standard Marčenko-Pastur law with parameter $\gamma$. Moreover,

$$
\frac{1}{p} E \left\| \frac{2}{n} \sum_{i=1}^{n} \operatorname{diag} (U_i \otimes U_i) - \frac{2}{3} I_p \right\|_F^2 = \frac{4}{pn^2} \sum_{k=1}^{p} E \left( \sum_{i=1}^{n} \left\{ U_i(k)^2 - E[U_i(k)^2] \right\} \right)^2 \leq \frac{C}{n} \to 0,
$$

$$
\frac{1}{p} E \left\| \frac{1}{(n^2)} D_0 \left[ \sum_{(i,j) \in [n]^2; i \neq j} U_i \otimes U_j \right] \right\|_F^2 = \frac{1}{p(n^2)^2} \sum_{(k,l) \in [p]^2; k \neq l} E \left\{ \sum_{(i,j) \in [n]^2; i \neq j} U_i(k) U_j(l) \right\}^2 \leq \frac{C_p}{n^2} \to 0,
$$

for some constant $C > 0$ independent of $n$. By Lemma D.2, the normalized Frobenius norm bounds the Lévy distance between spectral measures. An application of Lemma D.2, together with (3.6), triangle inequality and the above bounds yields the following result.

**Proposition 3.3.** As $n \to \infty$ and $\frac{p}{n} \to \gamma > 0$, the empirical spectral distribution $\tilde{\mu}_p$ of

$$
\frac{1}{(n^2)} \sum_{1 \leq i < j \leq n} M_{(i,j)}^{(1)}
$$

is...
converges in probability to the law of \( \frac{2}{3} Y + \frac{1}{3} \), where \( Y \) is distributed according to the standard Marčenko-Pastur law with parameter \( \gamma \).

Let \( \hat{\mu}_p^\tau \) denote the empirical spectral distribution of \( \tau \). Using Lemma D.2 once more, we show that the Lévy distance between \( \hat{\mu}_p^\tau \) and \( \tilde{\mu}_p \) converges to zero. This implies Theorem 3.1 by Proposition 3.3. To that end, observe that by (3.5) and triangle inequality:

\[
\frac{1}{p} \mathbb{E} \left\| \tau - \frac{1}{n^2} \sum 1_{1 \leq i < j \leq n} M_{(i,j)}^{(1)} \right\|_F^2 \leq \frac{2}{p} \mathbb{E} \left\| \frac{1}{n^2} \sum 1_{1 \leq i < j \leq n} M_{(i,j)}^{(2)} \right\|_F^2 + \frac{1}{p} \mathbb{E} \left\| \frac{1}{n^2} \sum 1_{1 \leq i < j \leq n} M_{(i,j)}^{(3)} \right\|_F^2.
\]

To show that (3.8) goes to zero, we notice that the collection of matrices \( M_{(i,j)}^{(3)} \), \( 1 \leq i < j \leq n \) satisfies

\[
\mathbb{E} \text{Tr} \left\{ \left( M_{(i,j)}^{(3)} \right)^T M_{(i',j')}^{(3)} \right\} = \begin{cases} \mathbb{E} \left\| M_{(i,j)}^{(3)} \right\|_F^2 & \text{for } (i,j) = (i',j'), \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}
\]

To see this, expand

\[
\mathbb{E} \text{Tr} \left\{ \left( M_{(i,j)}^{(3)} \right)^T M_{(i',j')}^{(3)} \right\} = \sum_{(k,l) \in [p] \times [p], k \neq l} \mathbb{E} \left[ \left( A_{(i,j)} - U_i + U_j \right) \left( A_{(i',j')} - U_{i'} + U_{j'} \right) \right] \times \mathbb{E} \left[ \left( A_{(i,j)} - U_i + U_j \right) \left( A_{(i',j')} - U_{i'} + U_{j'} \right) \right], \tag{3.10}
\]

and notice that each expectation is zero unless \( (i,j) = (i',j') \) by Tower property and Lemma 3.2. Note that when \( (i,j) = (i',j') \), the expression (3.10) is bounded by \( Cp^2 \) for some \( C > 0 \). The equation (3.9) also holds for the collection of matrices \( \{ M_{(i,j)}^{(2)} \} \) and we also have \( \mathbb{E} \left\| M_{(i,j)}^{(2)} \right\|_F^2 \leq Cp^2 \) by a similar argument. Therefore the right side of (3.8) is bounded by:

\[
\frac{Cp}{n^2} \times \# \{ (i,j,i',j') \in [n]^4 : (i,j) = (i',j') \} \leq \frac{Cp}{n^2},
\]

for some constant \( C > 0 \), which vanishes as \( n \to \infty \). This concludes the proof of Theorem 3.1. \( \blacksquare \)
A. Weak Convergence results

**Theorem A.1** (Martingale Central Limit Theorem) [Bil08, Theorem 35.12]. Let \( M_{k,N}, 1 \leq k \leq N, N \geq 1 \) be a sequence of zero-mean, square-integrable martingales adapted to the filtration \( \mathcal{F}_{k,N} \), and let \( \mathcal{F}_{0,N} \) denote the trivial \( \sigma \)-field. Let \( Y_{k,N} \) denote the martingale difference sequence \( Y_{k,N} = M_{k,N} - M_{k-1,N} \) where \( 1 \leq k \leq N \). Suppose the following conditions hold

\[
\text{for all } \epsilon > 0 \quad \sum_{k=1}^{N} \mathbb{E}[Y_{k,N}^2 \mathbf{1}_{|Y_{k,N}| > \epsilon \mathcal{F}_{k-1,N}}] \to 0 \quad \text{in probability},
\]

\[
\sum_{k=1}^{N} \mathbb{E}[Y_{k,N}^2 | \mathcal{F}_{k-1,N}] \to \sigma^2 \quad \text{in probability}
\]

where \( \sigma^2 \). Then \( M_{N,N} \) converges in distribution to a Gaussian random variable with mean 0 and variance \( \sigma^2 \).

**Theorem A.2** [CK86, Theorem 1; Lemma 1]. Let \( \Phi \) be a space of measurable functions \( \phi(z,w) : H \times C \to C \) such that \( \phi(z,\cdot) \) is continuous for all \( z \in H \). Let \( M \) be a set of measurable functions \( x : H \to C \) such that

\[
\int |\phi(z,x(z))| \, dz < \infty, \quad \phi \in \Phi,
\]

and define the functional

\[
\ell_\phi(\xi_N) = \int_H \phi(z, \xi_N(z)) \, dz.
\]

Suppose \( \{\xi_N : N \in \mathbb{N}_0\} \) is a sequence of stochastic processes \( \xi_N(z) : H \to C \), with paths in \( M \). If the finite dimensional distributions of \( \xi_N \) converge weakly to those of \( \xi_0 \) Lebesgue almost everywhere in \( H \) and for all \( \epsilon > 0 \) and \( \phi \in \Phi \):

\[
\lim_{K \to \infty} \limsup_{N \to \infty} \mathbb{P} \left( \int_H (|\phi(z,\xi_N(z))| - K)^+ \, dz \geq \epsilon \right) = -1,
\]

\[
\inf_{B \subset H, \lambda(B) < \infty} \limsup_{N \to \infty} \mathbb{P} \left( \int_{H-B} |\phi(z,\xi_N(z))| \, dz \geq \epsilon \right) = 0
\]

where \( \lambda \) is the Lebesgue measure on \( H \), then for \( \phi_1, \ldots, \phi_k \in \Phi \), we have that

\[(\ell_{\phi_1}(\xi_N), \ldots, \ell_{\phi_k}(\xi_N)) \Rightarrow (\ell_{\phi_1}(\xi_0), \ldots, \ell_{\phi_k}(\xi_0)).\]
B. Concentration inequalities and bounds on the resolvent

In this section we record some important bounds required in Part 1. We remind the reader the notation used in that Part

\[ \delta_N^{k,n}(z) := h_k^\dagger G_k(z)^n h_k - N^{-1} \text{Tr}(G_k(z_1)^n) \] (B.1)

where \( h_k \) denotes the \( k \)-th column of \( \mathcal{H} \) with \( k \)-th element removed. The matrix \( \mathcal{H}_k \) is defined as the matrix \( \mathcal{H} \) with the \( k \)-th row and column erased and \( G_k(z) := (\mathcal{H}_k - z)^{-1} \) is the resolvent of \( \mathcal{H}_k \).

**Lemma B.1.** Let \( z = E + \tau + \eta \) with fixed \( \eta > 0 \), \( \tau \in \mathbb{R} \) and \( E \in (-2 + \delta, 2 - \delta) \). Fix \( \epsilon' > 0 \). Then for all \( q \geq 1 \), there are positive constants \( c_1, c_2 \) such that

\[ \mathbb{E}|\delta_N^{k,1}(z)|^q \leq c_1 \left( \frac{N\eta}{d_N} \right)^{-\frac{q}{2}} \] (B.2)

\[ \mathbb{E}\left| z + \frac{1}{N} \text{Tr}(G_k(z)) \right|^q \leq c_2 \] (B.3)

\[ \mathbb{E}\left| \frac{1}{z + \frac{1}{N} \text{Tr}(G_k(z))} + s(z) \right|^q \leq c_3 \left( \frac{d_N}{N} \right)^{q-\epsilon'} \] (B.4)

for all \( k, N \).

**Proof.** Inequality (B.2) can be found in Proposition 3.2 of [CMS14] as a consequence of the Hanson-Wright large deviation inequality [RV13]. Such concentration inequalities appear in several other works concerning the local semi-circle law for Wigner matrices, see e.g. [ESY09b, ESY09a, EYY12a, EYY12b]. To prove (B.3), set \( \tilde{b}_k := (z + N^{-1} \text{Tr}(G_k))^{-1} = (z + s(z) + \Lambda_k)^{-1} \) where \( \Lambda_k = N^{-1} \text{Tr}(G_k) - s(z) \). Now we use the elementary identity

\[ \tilde{b}_k := \frac{1}{z + s(z) + \Lambda_k} = \sum_{i=0}^{p} \frac{\Lambda_i}{(z + s(z))^{i+1}} + \frac{\Lambda_i^p}{(z + s(z))^{p}(z + N^{-1} \text{Tr}(G_k))} \] (B.5)

Then it is known that \( |z + s(z)|^{-1} \leq 1 \) uniformly on the upper-half plane \( z \in \mathbb{C}^+ \) (see [BS10a, Eq. 8.1.19]). Furthermore, we have the trivial bound \( |z + N^{-1} \text{Tr}(G_k)|^{-1} \leq (\eta/d_N + N^{-1} \text{Tr}(G_k))^{-1} \leq d_N/\eta \). This gives us

\[ |\tilde{b}_k| \leq \sum_{i=0}^{p} |\Lambda_i|^{i+1} + \frac{d_N}{\eta} |\Lambda_i|^p \]

On the other hand, by (B.2) we know that \( \mathbb{E}|\Lambda_k|^q \leq C'(N\eta/d_N)^{-q} \), so for example

\[ \mathbb{E}|\tilde{b}_k|^q \leq 1 + O((N\eta/d_N)^{-\frac{q}{2}}) + O((N\eta/d_N)^{p-1}) \]

This last error term can be made \( o(1) \) by choosing \( p \) large enough (setting \( d_N = N^\alpha \), one finds the condition \( p + 1 > \alpha/(1 - \alpha) \)). For (B.4), note the identity \( (z + s(z))^{-1} = -s(z) \) and so after subtracting \( -s(z) \) the remaining error terms in (B.5) are smaller than \( O_{L^a}(d_N/\eta)^q \) except the last term. As before this last term is estimated with the trivial bound on \( (z + N^{-1} \text{Tr}(G_k))^{-1} \) and choosing \( p \) large enough. \( \blacksquare \)
Lemma B.2. Let \( z = E + \tau + i\eta \). For all \( q \geq 1 \) with fixed \( \eta > 0 \) and \( \tau \in \mathbb{R} \), there are positive constants \( c, C \) such that

\[
E|d_{N}^{-1} \delta_{N}^{k_2}(z)|^q < C \left( \frac{N}{d_N} \right)^{-q/2} \eta^{-3q/2},
\]

\[
E \left| \frac{\eta}{d_N} \text{Tr}(G_k^2(z)) \right|^q \leq c \max \left( \frac{\eta}{d_N}, \left( \frac{N\eta}{d_N} \right)^{-1} \right)^q,
\]

for all \( k, N \).

Proof. Since \( G_k(z) \) is an analytic function of \( z \) in the upper half plane and \( \frac{d}{dz} G_k(z) = G_k^2(z) \), we write

\[
d_{N}^{-1} \delta_{N}^{k_2} = d_{N}^{-1} \frac{d}{dz} \left( h_k^1 G_k h_k - \frac{1}{N} \text{Tr}(G_k) \right)
\]

\[
= \frac{1}{2\pi i} \oint_{S_z} \frac{1}{(z-w)^2} \left( h_k^1 G_k(w) h_k - \frac{1}{N} \text{Tr}(G_k(w)) \right) dw
\]

where \( S_z \) is a small circle of radius \( \eta/2d_N \) around the point \( z \). Note that \( S_z \) is of distance \( O(\eta/d_N) \) from the real axis and that \( |z - w| = \eta/2d_N \). By Hölder’s inequality and Lemma B.1, we have

\[
E|d_{N}^{-1} \delta_{N}^{k_2}|^q \leq \frac{1}{(2\pi d_{N})^{q}} \oint_{S_z} \ldots \oint_{S_z} E \prod_{i=1}^{q} |\delta_{N}^{k_2}(w_i)| \frac{1}{|z - w_i|^2} |dw_i|
\]

\[
\leq C \left( \frac{N\eta}{d_N} \right)^{-q/2} \frac{1}{(2\pi d_{N})^{q}} \left( \int_{S_z} \frac{1}{|z - w|^2} |dw| \right)^q
\]

\[
= C \left( \frac{N}{d_N} \right)^{-q/2} \eta^{-3q/2}.
\]

As for \( \text{Tr}(G_k^2) \) we similarly have

\[
\frac{\eta}{d_N} \text{Tr}(G_k^2) = \frac{\eta}{d_N} \frac{d}{dz} \text{Tr} G_k = \frac{\eta}{d_N} s'(z) + \frac{\eta}{2\pi id_N} \int_{S_z} \frac{1}{(w-z)^2} (N^{-1} \text{Tr} G_k(w) - m(w)) dw
\]

(B.6)

where \( S_z \) is a small circle of radius \( \eta/2d_N \) with center \( z = z(s) \). Then using the known rigidity estimates we get \( E|N^{-1} \text{Tr}(G_k(w)) - s(z)|^q \leq (N\eta/d_N)^{-q} \) which we use to estimate
Then

\[ E \left| \frac{\eta}{d_N} \text{Tr} \left( G_k^2 \right) \right|^q \]

\[ \leq \sum_{i=0}^{q} \left( \frac{q}{i} \right) \frac{\eta}{d_N} s'(z) \left| \int_{s_z} \frac{1}{(w-z)^2} \left( N^{-1} \text{Tr} G_k(w) - m(w) \right) dw \right|^q, \]

\[ \leq \sum_{i=0}^{q} \left( \frac{q}{i} \right) \frac{\eta}{d_N} s'(z) \left| \left( c_i N \eta \right)^{-i} \right|, \]

\[ \leq c \max \left( \frac{\eta}{d_N}, \left( \frac{N \eta}{d_N} \right)^{-1} \right)^q, \]

as required. \( \blacksquare \)

**Lemma B.3.** Let \( z = E + \frac{\tau+i\eta}{d_N} \) as in the condition of Lemma B.1 and let \( F_{k,N} = \sigma(\{H_{i,j} : 1 \leq i, j \leq N\}) \) where \( H \) is a Wigner Matrix as in Part 1. Define \( E_k[\cdot] := E[\cdot|F_{k,N}] \). We have the following identity:

\[ (E_k - E_{k-1}) \frac{1}{d_N} \left( \frac{1 + H_k^2 G_k(z)^2 h_k}{H_{kk} - z - h_k^2 G_k(z) h_k} \right) = (E_k - E_{k-1}) \frac{1}{d_N} \left( \frac{\partial}{\partial z} \right) \frac{H_{kk} - \delta_N^{k,1}(z)}{z + \frac{1}{N} \text{Tr} (G_k(z))} + \epsilon_{k,N}(z), \]

where

\[ \epsilon_{k,N}(z) := (E_k - E_{k-1}) \frac{1}{d_N} \left( \frac{1}{z + \frac{1}{N} \text{Tr} (G_k(z))} \right)^2 - \frac{1}{d_N} \left( \delta_N^{k,2}(z) (H_{kk} - \delta_N^{k,1}(z)) \right). \]

**Proof.** We omit the explicit \( z \)-dependence on the resolvent and make use of the quantities given in \( \delta_N^{k,n} \) in (B.1) and two further quantities

\[ \tilde{b}_k := \frac{1}{z + \frac{1}{N} \text{Tr} G_k}, \quad G_{kk} = \frac{1}{H_{kk} - z - h_k^2 G_k h_k}. \]

Using the identity \( z + \frac{1}{N} \text{Tr} G_k = (\delta_N^{k,1} + H_{kk} - G_{kk}^{-1}) \) we obtain

\[ \frac{1 + H_k^2 G_k(z)^2 h_k}{H_{kk} - z - h_k^2 G_k h_k} - \frac{1}{z + \frac{1}{N} \text{Tr} G_k} \]

\[ = \frac{(1 + H_k^2 G_k^2 h_k)(H_{kk} - \delta_N^{k,1})}{\tilde{b}_k^2 G_{kk}^{-1}} - \frac{G_{kk}^{-1}(1 + h_k^2 G_k^2 h_k)}{\tilde{b}_k^2 G_{kk}^{-1}} + \frac{G_{kk}^{-1}(1 + N^{-1} \text{Tr} G_k^2)}{G_{kk}^{-1} \tilde{b}_k^2} \]

\[ = G_{kk} \left( 1 + h_k^2 G_k^2 h_k \right) (H_{kk} - \delta_N^{k,1}) \]

\[ = \frac{(1 + H_k^2 G_k^2 h_k)(H_{kk} - \delta_N^{k,1})}{\tilde{b}_k^2 G_{kk}^{-1}} - \delta_N^{k,2} \tilde{b}_k \]

\[ = - \tilde{b}_k^2 (1 + H_k^2 G_k^2 h_k)(H_{kk} - \delta_N^{k,1}) + \tilde{b}_k^2 (1 + H_k^2 G_k^2 h_k)(H_{kk} - \delta_N^{k,1})^2 - \delta_N^{k,2} \tilde{b}_k \]  

\( \text{(B.8)} \)

\[ = - \tilde{b}_k^2 (1 + N^{-1} \text{Tr} G_k^2)(H_{kk} - \delta_N^{k,1}) - \delta_N^{k,2} \tilde{b}_k \]  

\( \text{(B.9)} \)

\[ - \tilde{b}_k^2 \delta_N^{k,2}(H_{kk} - \delta_N^{k,1}) + \tilde{b}_k^2 (1 + H_k^2 G_k^2 h_k)(H_{kk} - \delta_N^{k,1})^2 \]  

\( \text{(B.10)} \)
where to obtain (B.7) we expanded using the simple identity $G_{kk} = -\tilde{b}_k - \tilde{b}_k G_{kk} (\delta_{N}^{k,1} - \mathcal{H}_{kk})$. Then to obtain (B.9) from (B.8) we used that $h_k^1 G_k^2 h_k = N^{-1} \text{Tr} G_k^2 + \delta_{N}^{k,2}$. The two terms in (B.9) combine as an exact derivative

$$\frac{\partial}{\partial z} \tilde{b}_k (\mathcal{H}_{kk} - \delta_{N}^{k,1}) = -\tilde{b}_k^2 (1 + \text{Tr} G_k^2) (\mathcal{H}_{kk} - \delta_{N}^{k,1}) - \delta_{N}^{k,2} \tilde{b}_k$$

while the remaining two terms in (B.10) combine to give the error term $\epsilon_{k,N}(z)$. Finally, we conclude the proof of the identity by applying $(E_k - E_{k-1})$ to (B.7), noting that the second term vanishes.

Proposition B.4. Fix $\tilde{\eta} > 0$. Then under the same conditions as Theorem 1 in Part 1 we have that there exists a positive constants $N_0$, $M_0$, $C$, $c_0$, and $c_1 = c_1(C, c)$ such that

$$E|V_N(z)|^2 = \text{Var} \left\{ \frac{\text{Tr} G(E + z/d_N)}{d_N} \right\} \leq c_1 \eta^{-2},$$

for all $N > N_0$ such that $N \eta/d_N \geq M_0$, $\eta/d_N \leq \tilde{\eta}$, and $|E + t/d_N| \leq 2 + \eta/d_N$.

Proof. By the triangle inequality, we can write

$$E|V_N(z)|^2 = E|(N/d_N) (s_N(E + z/d_N) - E s_N(E + z/d_N))|^2$$

$$\leq 2E|(N/d_N) (s_N(E + z/d_N) - s(E + z/d_N))|^2$$

$$+ 2|(N/d_N) E(s_N(E + z/d_N) - s(E + z/d_N))|^2$$

$$\leq 4E|(N/d_N)(s_N(E + z/d_N) - s(E + z/d_N))|^2$$

where the last line follows from Jensen’s inequality. The latter expectation is by definition the integral

$$\int_0^\infty 2u \mathbf{P}(|s_N(E + z/d_N) - s(E + z/d_N)| > ud_N/N) \, du$$

$$= \eta^{-2} \int_0^\infty 2K \mathbf{P} \left( |s_N(E + z/d_N) - s(E + z/d_N)| > \frac{Kd_N}{\eta N} \right) \, dK$$

It suffices to bound the contribution to the above integral when $K \geq 1$. We apply Theorem *INSERT* from Part 1 with e.g. $q = 3$ to obtain a convergent estimate.

75
C. Function extension

The following Lemma is useful for proving results about the linear spectral statistic of a random matrix via analogous results for the resolvent. It can be found in [AGZ09, Lemma 5.5.5].

**Lemma C.1.** Let \( f : \mathbb{R} \rightarrow \mathbb{R} \) be a compactly supported function with first derivative continuous (denote it by \( f \in C^1_0(\mathbb{R}) \)) and consider an extension \( \Psi_f : \mathbb{R}^2 \rightarrow \mathbb{C} \) that inherits the same regularity and such that \( \Psi_f(x, 0) = f(x) \) for all \( x \) and \( \Im(\Psi_f(x, 0)) = 0 \). Further assume that \( \overline{\partial} \Psi_f(x, y) = O(y) \) as \( y \rightarrow 0 \). Then one has

\[
 f(\lambda) + iH[f] = \frac{1}{\pi} \int_0^\infty \int_{-\infty}^{\infty} \frac{\overline{\partial} \Psi_f(x, y)}{\lambda - x - iy} \, dx \, dy,
\]

where \( \overline{\partial} := \frac{\partial}{\partial x} + i \frac{\partial}{\partial y} \), and \( H : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R}) \) is the Hilbert Transform of \( f \):

\[
 H[f] := \frac{1}{\pi} \mathrm{P.V.} \int_{-\infty}^{\infty} \frac{f(t)}{x - t} \, dt.
\]

**Remark C.2.** Our particular choice of \( \Psi_f \) in this paper will be the following (see [Dav95]):

\[
 \Psi_f(x, y) = (f(x) + i(f(x + y) - f(x)))J(y),
\]

where \( J(y) \) is a smooth function of compact support, equal to 1 in a neighborhood of 0 and equal to 0 if \( y > 1 \).

**Proof.** We make use of the the substitution \( x \rightarrow x + \lambda \) and compute the real part (denoting \( \Psi_f = u + iv \))

\[
 -\frac{1}{\pi} \Re \int_0^\infty \, dy \int_{-\infty}^{\infty} \frac{\overline{\partial} \Psi_f(x + \lambda, y)}{x + iy} =
\]

\[
 -\frac{1}{\pi} \lim_{\epsilon \to 0} \int_{A_{\epsilon, R}} \frac{x}{x^2 + y^2} \left( \frac{\partial u(x, y)}{\partial x} - \frac{\partial v(x, y)}{\partial y} \right) + \frac{y}{x^2 + y^2} \left( \frac{\partial u(x, y)}{\partial y} + \frac{\partial v(x, y)}{\partial x} \right),
\]

where \( A_{\epsilon, R} = \{ \epsilon < \sqrt{x^2 + y^2} < R \} \) and \( R \) is sufficiently large (by the compact support hypothesis). Changing to polar coordinates in the latter integral, a simple computation shows that the integral transforms as

\[
 -\frac{1}{\pi} \lim_{\epsilon \to 0} \int_0^\pi \, d\theta \int_\epsilon^R \, dr \frac{\partial u(r \cos(\theta), r \sin(\theta))}{r} - \frac{\partial}{\partial \theta} \frac{v(r \cos(\theta), r \sin(\theta))}{r} =
\]

\[
 -\frac{1}{\pi} \lim_{\epsilon \to 0} \left( \int_0^\pi \, d\theta u(R \cos(\theta), R \sin(\theta)) - \int_0^\pi \, d\theta u(\epsilon \cos(\theta), \epsilon \sin(\theta)) + \int_{\epsilon |r|<R} \frac{v(r, 0)}{r} \, dr \right)
\]

where in the second line we applied Green's theorem to reduce the double integral to an integral over the boundary of the positively oriented rectangle \( R_{\epsilon, R} \) with vertices \((\epsilon, 0), (R, 0), (R, \pi), (\epsilon, \pi)\). Now choosing \( R \) large enough and using the assumption

\[
 \Im(\Psi_f(r + \lambda, 0)) = v(r, 0) = 0
\]

we see that the above limit is equal to \( u(0, 0) = \Psi_f(\lambda, 0) = f(\lambda) \). The proof for the imaginary part is similar.
D. The Standard Marčenko-Pastur Law.

We include here the definition of the standard Marčenko-Pastur law and a bound on the distance between empirical spectral distributions of two matrices.

**Theorem D.1** (Marčenko-Pastur law) [BS10b, Theorem 3.6]. Let \( X_1, \ldots, X_n \) be independent copies of a random vector \( X \in \mathbb{R}^p \) such that

\[
\mathbb{E}[X] = 0, \quad \mathbb{E}[X \otimes X] = I_p.
\]

Suppose that \( n \to \infty, \frac{p}{n} \to \gamma > 0 \) and define \( a = (1 - \sqrt{\gamma})^2 \), and \( b = (1 + \sqrt{\gamma})^2 \). Then the empirical spectral distribution of the matrix

\[
\frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i,
\]

converges almost surely to the standard Marčenko-Pastur law which has density:

\[
p_\gamma(x) = \begin{cases} 
\frac{1}{2\pi x \gamma} \sqrt{(b - x)(x - a)}, & \text{if } a \leq x \leq b, \\
0, & \text{otherwise},
\end{cases}
\]

and has a point mass \( 1 - \frac{1}{\gamma} \) at the origin if \( \gamma > 1 \).

**Lemma D.2** [BS10b, Corollary A.41]. Let \( A \) and \( B \) be two \( p \times p \) normal matrices, with empirical spectral distributions \( \hat{\mu}^A \) and \( \hat{\mu}^B \). Then

\[
L(\hat{\mu}^A, \hat{\mu}^B)^3 \leq \frac{1}{p} \| A - B \|_F^2,
\]

where \( L(\hat{\mu}^A, \hat{\mu}^B) \) is the Lévy distance between the distribution functions \( \hat{\mu}^A \) and \( \hat{\mu}^B \).
REFERENCES


