Lecture notes on random matrices

Greg W. Anderson       Alice Guionnet
Department of Mathematics   ENS Lyon
University of Minnesota

Ofer Zeitouni
University of Minnesota and Technion

30 November, 2006
©Greg W. Anderson, Alice Guionnet, Ofer Zeitouni
Do not distribute without the written permission of author.
Rather rough - Comments welcomed!
# Contents

1 Random matrices and eigenvalues .......................... 8
   1.1 Eigenvalues ............................................. 8

2 Real and Complex Wigner matrices ....................... 10
   2.1 Real Wigner matrices: traces, moments and combinatorics . 10
      2.1.1 The semicircle distribution, Catalan numbers, and
              Dyck paths ........................................ 11
      2.1.2 Proof #1 of Wigner’s Theorem 2.1.4 ................. 14
      2.1.3 Proof of Lemma 2.1.12 : Words and Graphs .......... 16
      2.1.4 Proof of Lemma 2.1.13 : Sentences and Graphs .... 20
      2.1.5 Some useful approximations .......................... 24
      2.1.6 Maximal eigenvalues and Füredi-Komlós enumeration 26
      2.1.7 Central limit theorems for moments .................. 33
   2.2 Complex Wigner matrices ............................... 39
   2.3 Concentration for functional of random matrices and loga-
        rithmic Sobolev inequalities ............................ 41
      2.3.1 Smoothness properties of linear functions of the em-
              rirical measure .................................... 41
      2.3.2 Concentration inequalities for independent variables
              satisfying logarithmic Sobolev inequalities .......... 42
      2.3.3 Concentration for Wigner-type matrices ............. 45
   2.4 Stieltjes transforms and recursions .................. 46
      2.4.1 Gaussian Wigner matrices ........................... 48
      2.4.2 General Wigner matrices ............................ 50
### 2.5 Joint distribution of eigenvalues in the GOE and the GUE

- **2.5.1** Definition and preliminary discussion of the GOE and the GUE  
- **2.5.2** Proof of the joint distribution of eigenvalues  
- **2.5.3** Selberg’s integral formula and proof of (2.5.4)  
- **2.5.4** Joint distribution of eigenvalues - alternative formulation

### 2.6 Large deviations for random matrices

- **2.6.1** Large deviations for the empirical measure  
- **2.6.2** Large deviations for the top eigenvalue

### 2.7 Bibliographical notes

### 3 Orthogonal polynomials, spacings, and limit distributions for the GUE

#### 3.1 Summary of main results: spacing distributions in the bulk and edge of the spectrum for the GUE

#### 3.2 Hermite polynomials and the GUE

- **3.2.1** The Hermite polynomials and harmonic oscillators  
- **3.2.2** Connection with the GUE

#### 3.3 The semicircle law revisited

- **3.3.1** Calculation of moments of $L_N$  
- **3.3.2** The Harer-Zagier recursion and Ledoux’s argument

#### 3.4 Quick introduction to Fredholm determinants

- **3.4.1** The setting, fundamental estimates, and definition of the Fredholm determinant  
- **3.4.2** Definition of the Fredholm adjugants, Fredholm resolvents, and a fundamental identity

#### 3.5 Gap probabilities at 0 and proof of Theorem 3.1.1

- **3.5.1** The method of Laplace  
- **3.5.2** Evaluation of the scaling limit - proof of Lemma 3.5.2  
- **3.5.3** A complement: determinantal relations

#### 3.6 Analysis of the sine-kernel

- **3.6.1** General differentiation formulae and the Kyoto equations
4 Some generalities

4.1 Joint distribution of eigenvalues in the classical matrix ensembles

4.1.1 Manifolds and the coarea formula

4.1.2 Measures of spheres

4.1.3 Measures of unitary groups

4.1.4 Measures of flag manifolds

4.1.5 The trace-of-powers map

4.1.6 Lie groups and an integration formula of Weyl type

4.1.7 Applications of the Weyl integration formula

4.1.8 The Gaussian ensembles

4.1.9 The Laguerre ensemble

4.1.10 The Jacobi ensemble

4.1.11 Reference notes

4.2 Determinantal processes

4.2.1 Basic definitions

4.2.2 Determinantal projections

4.2.3 Properties of determinantal processes, and the CLT
4.2.4 Translation invariant determinantal processes . . . 177
4.2.5 One dimensional translation invariant determinantal
processes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 182
4.2.6 Convergence issues . . . . . . . . . . . . . . . . . . . . 185
4.2.7 Examples . . . . . . . . . . . . . . . . . . . . . . . . . . 188

4.3 Stochastic analysis for random matrices . . . . . . . . . 193
4.3.1 Dyson’s Brownian motion . . . . . . . . . . . . . . . . 193
4.3.2 Proof #7 of Wigner’s Theorem 2.1.4 . . . . . . . . . 203
4.3.3 Central limit theorem . . . . . . . . . . . . . . . . . . 214

4.4 Concentration inequalities for random matrices . . . . . 220

5 Free probability 221
5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . 221
5.2 Non-commutative laws and noncommutative probability spaces221
5.2.1 Algebraic noncommutative probability spaces and laws221
5.2.2 $C^*$- noncommutative probability spaces and weak topol-
ogy . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 225
5.2.3 $W^*$-noncommutative probability spaces . . . . . . . 243

5.3 Free independence . . . . . . . . . . . . . . . . . . . . . . . 247
5.3.1 Independence and free independence . . . . . . . . . 247
5.3.2 Free independence and combinatorics . . . . . . . . . 250
5.3.3 Consequence of free independence: free convolution . 256
5.3.4 Free central limit theorem . . . . . . . . . . . . . . . . 267
5.3.5 Link with random matrices . . . . . . . . . . . . . . . 268

Appendix 277
A Linear algebra preliminaries . . . . . . . . . . . . . . . . . 277
A.1 Identities and bounds . . . . . . . . . . . . . . . . . . . 277
A.2 Perturbations for normal and Hermitian matrices . . . . 278
A.3 Brief review of resultants and discriminants . . . . . . . 279

B Topological Preliminaries . . . . . . . . . . . . . . . . . . . 280
B.1 Generalities . . . . . . . . . . . . . . . . . . . . . . . . . . . 280
B.2 Topological Vector Spaces and Weak Topologies . . . . 283
B.3 Banach and Polish Spaces . . . . . . . . . . . . . . . . . . 284
C Probability measures on Polish spaces . . . . . . . . . . . . 286
C.1 Generalities . . . . . . . . . . . . . . . . . . . . . . 286
C.2 Weak Topology . . . . . . . . . . . . . . . . . . . . . 287
D Basic notions of large deviations . . . . . . . . . . . . . . 289
E The skew field \( H \) of quaternions, and matrix terminology in \( F \) 292
E.1 Matrix terminology in \( F \) . . . . . . . . . . . . . . . . 293
E.2 The \( \bullet \)-construction . . . . . . . . . . . . . . . . 295
E.3 Matrix factorization theorems . . . . . . . . . . . . . . 296
E.4 The spectral theorem and key corollaries . . . . . . . . . 296
E.5 Proof of the spectral theorem . . . . . . . . . . . . . . 298
E.6 Some specialized results on projectors . . . . . . . . . . 299
F Manifolds embedded in Euclidean space . . . . . . . . . . . 301
G Appendix on Operators algebras . . . . . . . . . . . . . . . 305
G.1 Basic definitions . . . . . . . . . . . . . . . . . . . . . 305
G.2 Spectral properties . . . . . . . . . . . . . . . . . . . . 307
G.3 States and positivity . . . . . . . . . . . . . . . . . . . . 309
G.4 von Neumann algebras . . . . . . . . . . . . . . . . . . 309
G.5 Non-commutative functional calculus . . . . . . . . . . . 310
G.6 Riesz representation Theorem . . . . . . . . . . . . . . 311
H Stochastic calculus notions . . . . . . . . . . . . . . . . . . 311

References 315
Notations - This will move

\[ M_1(S) \quad \text{space of probability measures on a topological space} \ S \]
\[ C(S), C_b(S) \quad \text{Space of continuous (bounded) functions on} \ S \]
\[ \langle \mu, f \rangle = \int f(x) \mu(dx), \text{for} \ f \ \text{measurable}, \mu \in M_1(S). \]
\[ \epsilon(\sigma) \quad \text{the signature of a permutation} \ \sigma \]
\[ \det(M) = \sum \epsilon(\sigma) \prod M_{i,\sigma(i)} \text{where the sum runs over all permutation of} \ \{1, \cdots, n\}. \]

Unless stated otherwise, for \( S \) a Polish space, \( M_1(S) \) is given the topology of weak convergence, that makes it into a Polish space.

When we write \( a(s) \sim b(s), \) we assert that there exists \( c(s) \) defined for \( s \gg 0 \) such that \( \lim_{s \to \infty} c(s) = 1 \) and \( c(s)a(s) = b(s) \) for \( s \gg 0. \) We use the notation \( a_n \sim b_n \) for sequences in the analogous sense.
Chapter 1

Random matrices and eigenvalues

Throughout this book, we let \( \mathcal{H}_N^{(1)} \) be the space of (real) symmetric \( N \times N \) matrices, and let \( \mathcal{H}_N^{(2)} \) be the space of (complex) Hermitian \( N \times N \) matrices. One can always consider \( \mathcal{H}_N^{(\beta)} \), \( \beta = 1, 2 \), as a submanifold of an appropriate Euclidean space, and equip it with the induced topology.

1.1 Eigenvalues

Our interest will be in the study of eigenvalues of matrices in \( \mathcal{H}_N^{(\beta)} \). For \( H \in \mathcal{H}_N^{(\beta)} \) let \( \lambda_1(H) \leq \cdots \leq \lambda_N(H) \) be the eigenvalues of \( H \). The following lemma assures that the eigenvalues are continuous functions in \( H \), and hence can be treated as random variables as soon as a probability measure is put on \( \mathcal{H}_N^{(\beta)} \). For a strengthening of this result, see Appendix A.2.

**Lemma 1.1.1** For \( i = 1, \ldots, N \), the eigenvalue \( \lambda_i(H) \) is a continuous (and a fortiori measurable) function of \( H \).

**Proof:** For \( H \in \mathcal{H}_N^{(\beta)} \) put \( \| H \| := \sqrt{\text{Tr}H^2} \), thus defining a Euclidean metric on \( \mathcal{H}_N^{(\beta)} \). Note that \( \| H \| \geq \max(\lambda_N(H), -\lambda_1(H)) \). For any positive integer \( k \) we have

\[
\lambda_N(H) + \| H \| \leq \sqrt[k]{\text{Tr}(H + \| H \| I)^k} \leq \sqrt[k]{N(\lambda_N(H) + \| H \|)},
\]
hence
\[
\left| \sqrt[k]{\text{Tr}(H + \|H\|)^k} - \|H\| \lambda_N(H) \right| \leq 2(\sqrt[N]{N} - 1)\|H\|, \\
\]
and hence
\[
\lambda_N(H) = \lim_{k \to \infty} \sqrt[k]{\text{Tr}(H + \|H\|I)^k} - \|H\|
\]
uniformly on compact subsets of $\mathcal{H}_N^{(B)}$, which proves continuity of $\lambda_N(\cdot)$. Continuity of $\lambda_{i+1}(\cdot), \ldots, \lambda_N(\cdot)$ granted, by a similar analysis we have
\[
\lambda_i(H) = \lim_{k \to \infty} \left( \text{Tr}(H + \|H\|I)^k - \sum_{j=i+1}^{N} (\lambda_j(H) + \|H\|)^k \right)^{1/k} - \|H\|
\]
uniformly on compact subsets of $\mathcal{H}_N^{(B)}$, which proves continuity of $\lambda_i(\cdot)$. $\Box$

Add here that in fact $C^\infty$ in (open) region where eigenvalues are distinct.
Chapter 2

Real and Complex Wigner matrices

2.1 Real Wigner matrices: traces, moments and combinatorics

We introduce in this section a basic model of random matrices. Nowhere do we attempt to provide the weakest assumption of sharpest results available. We point out in the bibliographical notes (Section 2.7) some places where the interested reader can find finer results.

Start with two independent families of i.i.d., zero mean, real valued random variables \( \{Z_{i,j}\}_{1 \leq i < j} \) and \( \{Y_i\}_{1 \leq i} \), such that \( EZ_{i,2}^2 = 1 \) and, for all integer \( k \geq 1 \),

\[
r_k := \max (E|Z_{1,2}|^k, E|Y_1|^k) < \infty .
\]

We call such a matrix a \emph{Wigner matrix}, and if the random variables \( Z_{i,j} \) and \( Y_i \) are Gaussian, we use the term \emph{Gaussian Wigner matrix}. The case of Gaussian Wigner matrices in which \( EY_1^2 = 2 \) is of particular importance, and for reasons that will become clearer in Chapter 3, such matrices are referred to as GOE (Gaussian Orthogonal Ensemble) matrices.

Let \( \lambda_1^N \) denote the (real) eigenvalues of \( X_N \), with \( \lambda_1^N \leq \lambda_2^N \leq \ldots \leq \lambda_N^N \), and define the \emph{empirical distribution} of the eigenvalues as the probability
measure on $\mathbb{R}$ defined by

$$L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i^N}.$$  

Define the standard semicircle distribution as the probability distribution $\sigma(x)dx$ on $\mathbb{R}$ with density

$$\sigma(x) = \frac{1}{2\pi} \sqrt{4 - x^2} 1_{|x|\leq 2}. \quad (2.1.3)$$

The following theorem, contained in [Wig55], can be considered the starting point of Random Matrix Theory (RMT).

**Theorem 2.1.4 (Wigner)** For a Wigner matrix, the empirical measure $L_N$ converges weakly, in probability, to the standard semicircle distribution.

The statement in Theorem 2.1.4 is that for any $f \in C_b(\mathbb{R})$, and any $\varepsilon > 0$,

$$\lim_{N \to \infty} P(|\langle L_N, f \rangle - \langle \sigma, f \rangle| > \varepsilon) = 0.$$  

**Remark 2.1.5** The assumption (2.1.1) that $r_k < \infty$ for all $k$ is not really needed, see Theorem 2.1.46 in Section 2.1.5.

We will see many proofs of Wigner’s Theorem 2.1.4. In this section, we give a direct combinatorics based proof, mimicking the original argument of Wigner. Before doing so, however, we need to discuss some properties of the semicircle distribution.

### 2.1.1 The semicircle distribution, Catalan numbers, and Dyck paths

Define the moments $m_k := \langle \sigma, x^k \rangle$. By symmetry, $m_{2k+1} = 0$. Recall the Catalan numbers\footnote{There is a slight ambiguity in the literature concerning the numbering of Catalan numbers. Thus, [Aig79, Pg 85] denotes by $c_k$ what we denote by $C_{k-1}$. Our notations follow [Sta97].}

$$C_k = \frac{\binom{2k}{k}}{k + 1}.$$
A calculus exercise in integration by parts shows that \( m_{2k} = C_k \); indeed,

\[
m_{2k} = \int_{-2}^{2} x^{2k} \sigma(x) \, dx = \frac{2 \cdot 2^{2k}}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2k}(\theta) \cos^{2}(\theta) \, d\theta
\]

\[
= \frac{2 \cdot 2^{2k}}{\pi} \int_{-\pi/2}^{\pi/2} \sin^{2k}(\theta) \, d\theta = (2k + 1)m_{2k}.
\]

Hence,

\[
(2k + 2)m_{2k} = 2 \cdot 2^{2k} \int_{-\pi/2}^{\pi/2} \sin^{2k}(\theta) \, d\theta = (2k + 1)m_{2k} - (2k + 1)m_{2k}.
\]

from which, together with \( m_0 = 1 \), one concludes that

\[
m_{2k} = \frac{4(2k - 1)}{(2k + 2)} m_{2k-2}, \quad (2.1.6)
\]

leading to the claimed conclusion that \( m_{2k} = C_k \).

The Catalan numbers possess many combinatorial interpretations. To introduce a first one, say that an integer valued sequence \( \{S_n\}_{0 \leq n \leq \ell} \) is a Bernoulli walk of length \( \ell \) if \( S_0 = 0 \) and \( \lvert S_{t+1} - S_t \rvert = 1 \) for \( t \leq \ell - 1 \). Of particular relevance here is the fact that \( C_k \) counts the number of Dyck Paths of length \( 2k \), that is the number of nonnegative Bernoulli walks of length \( 2k \) that terminate at \( 0 \). Indeed, let \( \beta_k \) denote the number of such paths. A classical exercise in combinatorics is the

**Lemma 2.1.7** \( \beta_k = C_k < 4^k \). Further, the generating function \( \hat{\beta}(z) := 1 + \sum_{k=1}^{\infty} z^k \beta_k \) satisfies, for \( |z| < 1/4 \),

\[
\hat{\beta}(z) = \frac{1 - \sqrt{1 - 4z}}{2z}. \quad (2.1.8)
\]

**Proof of Lemma 2.1.7** Let \( B_k \) denote the number of Bernoulli walks \( \{S_n\} \) of length \( 2k \) that satisfy \( S_{2k} = 0 \), and let \( \bar{B}_k \) denote the number of Bernoulli walks \( \{S_n\} \) of length \( 2k \) that satisfy \( S_{2k} = 0 \) and \( S_t < 0 \) for some \( t < 2k \). Then, \( \beta_k = B_k - \bar{B}_k \). By reflection at the first hitting of \( -1 \), one sees that \( \bar{B}_k \) equals the number of Bernoulli walks \( \{S_n\} \) of length \( 2k \) that satisfy \( S_{2k} = -2 \). \(^{3}\) Hence,

\[
\beta_k = B_k - \bar{B}_k = \binom{2k}{k} - \binom{2k}{k-1} = C_k.
\]

\(^2\)There does not seem to be a clear convention as to whether such paths should be called Dyck path of length \( 2k \) or of length \( k \). Our choice is consistent with our notion of length of Bernoulli walks.

\(^3\)This is an instance of the application of reflection principle. See [Fel57, Ch. III.2]
Turning to the evaluation of \( \hat{\beta}(z) \), considering the first return time to 0 of the Bernoulli walk \( \{S_n\} \) gives the relation

\[
\beta_k = \sum_{j=1}^{k} \beta_{k-j} \beta_{j-1}, \quad k \geq 1
\]  \hspace{1cm} (2.1.9)

with the convention that \( \beta_0 = 1 \). Because the number of Bernoulli walks of length \( 2k \) is bounded by \( 4^k \), one has that \( \beta_k \leq 4^k \), and hence the function \( \hat{\beta}(z) \) is well defined and analytic for \( |z| < 1/4 \). But, substituting (2.1.9),

\[
\hat{\beta}(z) - 1 = \sum_{k=1}^{\infty} z^k \sum_{j=1}^{k} \beta_{k-j} \beta_{j-1} = z \sum_{k=0}^{\infty} z^k \sum_{j=0}^{k} \beta_{k-j} \beta_{j}
\]

while

\[
\hat{\beta}(z)^2 = \sum_{k,k' = 0}^{\infty} z^{k+k'} \beta_k \beta_{k'} = \sum_{q=0}^{\infty} \sum_{\ell=0}^{q} z^q \beta_{q-\ell} \beta_{\ell}.
\]

Combining the last two equations, one sees that

\[
\hat{\beta}(z) = 1 + z \hat{\beta}(z)^2,
\]

from which (2.1.8) follows (using that \( \hat{\beta}(0) = 1 \) to choose the correct branch of the square-root). \( \square \)

We note in passing that expanding (2.1.8) in power series in \( z \) in a neighborhood of zero, one gets (for \( |z| < 1/4 \))

\[
\hat{\beta}(z) = \frac{2 \sum_{k=1}^{\infty} \frac{z^k (2k-2)!}{2k (k-1)!}}{2z} = \sum_{k=0}^{\infty} \frac{(2k)!}{k!(k+1)!} z^k = \sum_{k=0}^{\infty} z^k C_k,
\]

which provides an alternative proof to the fact that that \( \beta_k = C_k \).

Another useful interpretation of the Catalan numbers is that \( C_k \) counts the number of rooted planar trees with \( k \) edges, with ordering given at each vertex (a rooted planar tree with ordering at vertices is a planar graph with no cycles, with one distinguished vertex, and with a choice of ordering at each vertex; the ordering defines a way to “explore” the tree, starting at the root). It is not hard to check that the Dyck paths of length \( 2k \) are in bijection with such rooted planar trees. The interested reader is referred to the proof of Lemma 2.1.12 in Section 2.1.3 for a formal construction of this bijection.

We note in closing that a third interpretation of the Catalan numbers, particularly useful in the context of Chapter 5, is that they enumerate the number of non-crossing partitions of the ordered set \( \mathcal{K}_k := \{1, 2, \ldots, k\} \).
Definition 2.1.10 A partition of the set $\mathcal{K}_k := \{1, 2, \ldots, k\}$ is called crossing if there exists a quadruple $(a, b, c, d)$ with $1 \leq a < b < c < d \leq k$ such that $a, c$ belong to the same part while $b, d$ belong to a different same part. A partition with no crossings is a non-crossing partition.

The collection of non-crossing partitions form a lattice with respect to inclusion. A look at Figure 2.1.1 should explain the terminology “non-crossing”: one puts the points $1, \ldots, k$ on the circle, and connects by an (internal) path each point with the next (in cyclic order) member of its part. Then, the partition is non-crossing if this can be achieved without arcs crossing each other.

![Figure 2.1.1: Non-crossing (left, (1, 4), (2, 3), (5, 6)) and crossing (right, (1, 5), (2, 3), (4, 6)) partitions of the set $\mathcal{K}_6$.](image)

It is not hard to check that $C_k$ is indeed the number $\gamma_k$ of non-crossing partitions of $\mathcal{K}_k$. To see that, let $\pi$ be a noncrossing partition of $\mathcal{K}_k$ and let $j$ denote the smallest element connected to 1 (with $j = 1$ if the part containing 1 is the set $\{1\}$). Then, by the definition, the partition $\pi$ induces non-crossing partitions on the sets $\{1, \ldots, j\}$ and $\{j, \ldots, k\}$. Therefore, $\gamma_k = \sum_{j=1}^{k} \gamma_{k-j} \gamma_{j-1}$. With $\gamma_1 = 1$, and comparing with (2.1.9), one sees that $\beta_k = \gamma_k$.

Exercise 2.1.11 Prove directly that for $z \in \mathbb{C}$ so that $z \not\in [-2, 2]$,

$$G(z) = \int \frac{1}{1-z\lambda} d\sigma(\lambda) = \frac{1 - \sqrt{1-4z}}{2z}$$

Hint: Use the residue theorem.

2.1.2 Proof #1 of Wigner’s Theorem 2.1.4

Define the probability distribution $L_N = E L_N$ by the relation $\langle L_N, f \rangle = E \langle L_N, f \rangle$ for all $f \in C_b$, and set $m_k^N := \langle L_N, x^k \rangle$. Theorem 2.1.4 follows
from the following two lemmas.

**Lemma 2.1.12** For any \( k \in \mathbb{N} \),
\[
\lim_{N \to \infty} m_k^N = m_k.
\]

**Lemma 2.1.13** For any \( k \in \mathbb{N} \) and \( \epsilon > 0 \),
\[
\lim_{N \to \infty} P \left( \left| \langle L_N, x^k \rangle - \langle L_N, x^k \rangle \right| > \epsilon \right) = 0.
\]

Indeed, assume that Lemmas 2.1.12 and 2.1.13 have been proved. To conclude the proof of Theorem 2.1.4, one needs to check that for any bounded continuous function \( f \),
\[
\lim_{N \to \infty} \langle L_N, f \rangle = \langle \sigma, f \rangle,
\]
in probability. (2.1.14)

Toward this end, note first that an application of the Chebycheff inequality yields
\[
P \left( \langle L_N, x^k 1_{|x| > B} \rangle > \epsilon \right) \leq \frac{1}{\epsilon} E \langle L_N, x^k 1_{|x| > B} \rangle \leq \frac{\langle L_N, x^k \rangle}{\epsilon B^k}.
\]
Hence, by Lemma 2.1.12,
\[
\limsup_{N \to \infty} P \left( \langle L_N, x^k 1_{|x| > B} \rangle > \epsilon \right) \leq \frac{\langle \sigma, x^k \rangle}{\epsilon B^k} \leq \frac{4^k}{\epsilon B^k},
\]
where we used that \( C_k \leq 4^k \). Thus, with \( B = 5 \), it follows, noting that the left hand side above is increasing in \( k \),
\[
\limsup_{N \to \infty} P \left( \langle L_N, x^k 1_{|x| > B} \rangle > \epsilon \right) = 0.
\] (2.1.15)

In particular, when proving (2.1.14), we may and will assume that \( f \) is supported on the interval \([-5, 5]\).

Fix next such an \( f \) and \( \delta > 0 \). One can then find a polynomial \( Q_\delta(x) = \sum_{i=0}^L c_i x^i \) such that
\[
\sup_{x:|x| \leq B} |Q_\delta(x) - f(x)| \leq \frac{\delta}{8}.
\]
Then,
\[
P \left( \left| \langle L_N, f \rangle - \langle \sigma, f \rangle \right| > \delta \right) \leq P \left( \left| \langle L_N, Q_\delta \rangle - \langle L_N, Q_\delta \rangle \right| > \frac{\delta}{8} \right) + P \left( \left| \langle L_N, Q_\delta \rangle - \langle \sigma, Q_\delta \rangle \right| > \frac{\delta}{8} \right) + P \left( \left| \langle L_N, Q_\delta 1_{|x| > B} \rangle - \langle \sigma, Q_\delta \rangle \right| > \frac{\delta}{8} \right)
\]
\[=: P_1 + P_2 + P_3.\]
By an application of Lemma 2.1.13, $P_1 \to_{N \to \infty} 0$. Lemma 2.1.12 implies that $P_2 \to_{N \to \infty} 0$, while (2.1.15) implies that $P_3 \to_{N \to \infty} 0$. This completes the proof of Theorem 2.1.4. □

2.1.3 Proof of Lemma 2.1.12 : Words and Graphs

The starting point toward the proof of Lemma 2.1.12 is the following identity:

$$\langle \bar{L}_N, x^k \rangle = \frac{1}{N} ET_{X_N^k}$$

$$= \frac{1}{N} \sum_{i_1, \ldots, i_k = 1}^{N} EX_N(i_1, i_2)X_N(i_2, i_3) \cdots X_N(i_{k-1}, i_k)X_N(i_k, i_1)$$

$$=: \frac{1}{N} \sum_{i_1, \ldots, i_k = 1}^{N} ET_{i_1}^{N} =: \frac{1}{N} \sum_{i_1, \ldots, i_k = 1}^{N} \bar{T}_{i_1}^{N}, \quad (2.1.16)$$

where we use the notation $i = (i_1, \ldots, i_k)$.

The proof of Lemma 2.1.12 now proceeds by considering what terms contribute to (2.1.16). Let us provide first an informal sketch that explains the emergence of the Catalan numbers, followed by a formal proof. For the purpose of this sketch, assume that $Y_1$ possess the same law as $Z_1, Z_2$, and that the law of $Z_{1,2}$ is symmetric, so that all odd moments vanish (and in particular, $\langle \bar{L}_N, x^k \rangle = 0$ for $k$ odd).

A first step in the sketch (that is fully justified in the actual proof below) is to check that the only terms in (2.1.16) that survive the passage to the limit involve only second moments of $Z_{i,j}$, because there are an order of $N^{k/2+1}$ of non-zero terms but only order of $N^{k/2}$ at most terms that involve moments higher than 4. One then sees that

$$\langle \bar{L}_N, x^{2k} \rangle = (1 + O(N^{-1})) \frac{1}{N} \sum_{(i_p, i_{p+1}) = (i_j, i_{j+1}) \text{ or } (i_{j+1}, i_j)} \bar{T}_{i_1, \ldots, i_{2k}}^{N} \cdot (2.1.17)$$

Considering the index $j$ such that either $(i_j, i_{j+1}) = (i_2, i_1)$ or $(i_j, i_{j+1}) = (i_1, i_2)$, one obtains

$$\langle \bar{L}_N, x^{2k} \rangle = (1 + O(N^{-1})) \frac{1}{N} \sum_{j=2}^{2k} \sum_{i_1, i_2 = 1}^{N} \sum_{i_3, \ldots, i_{2k} \setminus \{i_1, i_{j+1}\} = 1}^{N}$$

$$\left(EX_N(i_2, i_3) \cdots X_N(i_{j-1}, i_2)X_N(i_1, i_{j+2}) \cdots X_N(i_{2k}, i_1)$$

$$+ EX_N(i_2, i_3) \cdots X_N(i_{j-1}, i_1)X_N(i_2, i_{j+2}) \cdots X_N(i_{2k}, i_1) \right).$$
Hence, if we took for granted that $E[(L_N - \bar{L}_N, x^k)]^2 = O(N^{-2})$ and hence
\[ E[(L_N, x^j)(L_N, x^{2k-j-2})] = (\bar{L}_N, x^j)(\bar{L}_N, x^{2k-j-2})(1 + O(N^{-1})) , \]
we obtain
\[ (\bar{L}_N, x^{2k}) = (1 + O(N^{-1})) \sum_{j=0}^{2(k-1)} \left( \frac{1}{N^2} E[\text{Tr}(X_N^j)] E[\text{Tr}(X_N^{2k-j-2})] \right) \]
\[ = (1 + O(N^{-1})) \sum_{j=0}^{2k-2} \frac{1}{N^2} E[\text{Tr}(X_N^{j})] E[\text{Tr}(X_N^{2k-j-2})] \]
\[ = (1 + O(N^{-1})) \sum_{j=0}^{k-1} \frac{1}{N^2} E[\text{Tr}(X_N^{j})] E[\text{Tr}(X_N^{2(k-j-1)})] \quad (2.1.19) \]
where we used that by induction $E[\frac{1}{N} \text{Tr}(X_N^{2k-2})]$ is uniformly bounded and also the fact that odd moments vanish. Further,
\[ (\bar{L}_N, x^2) = \frac{1}{N} \sum_{i,j=1}^{N} E X_N(i,j)^2 \to_{N \to \infty} 1 = C_1 . \quad (2.1.20) \]

Thus, we conclude from (2.1.19) by induction that $(\bar{L}_N, x^{2k})$ converges to a limit $a_k$ with $a_0 = a_1 = 1$, and further the family $\{a_k\}$ satisfies the recursions $a_k = \sum_{j=1}^{k} a_{k-j} a_{j-1}$. Comparing with (2.1.9), one deduces that $a_k = C_k$, as claimed.

We turn next to the actual proof. To handle the summation in expressions like (2.1.16), it is convenient to introduce some combinatorial machinery that will serve us also in the sequel. We thus first digress and discuss the combinatorics intervening in the evaluation of the sum in (2.1.16). This is then followed by the actual proof of Lemma 2.1.12.

For the purpose of this section, the reader may think of $\mathcal{S}$ as a subset of the integers in the following:

**Definition 2.1.21 ($\mathcal{S}$-Words)** Given a set $\mathcal{S}$, an $\mathcal{S}$-letter $s$ is simply an element of $\mathcal{S}$. An $\mathcal{S}$-word $w$ is a finite sequence of letters $s_1, \ldots, s_n$, at least one letter long. An $\mathcal{S}$-word $w$ is closed if its first and last letters are the same. Two $\mathcal{S}$-words $w_1, w_2$ are called equivalent, denoted $w_1 \sim w_2$, if there is a bijection on $\mathcal{S}$ that maps one into the other.

When $\mathcal{S} = \{1, \ldots, N\}$ for some finite $N$, we use the term $N$-word. Otherwise, if the set $\mathcal{S}$ is clear from the context, we refer to a $\mathcal{S}$-word simply as a word.
For any $S$-word $w = (s_1, \ldots, s_k)$, we use $\ell(w) = k$ to denote the length of $w$, and define the weight $\text{wt}(w)$ as the number of distinct elements of the set $\{s_1, \ldots, s_k\}$, and the support of $w$, denoted $\text{supp} w$, as the set of letters appearing in $w$. To any word $w$ we may associate an undirected graph, with $\text{wt}(w)$ vertices and $k$ edges, as follows.

**Definition 2.1.22 (Graph associated to an $S$-word)** Given a word $w = (s_1, \ldots, s_k)$, we set $G_w = (V_w, E_w)$ to be the graph with set of vertices $V_w = \text{supp} w$ and (undirected) edges $E_w = \{e \in E_w : e = \{u, u\}, u \in V_w\}$ and the set of connecting edges as $E_c^w = E_w \setminus E_s^w$.

The word $w$ defines a path in the connected graph $G_w$, which starts and terminates at the same vertex if the word is closed. For $e \in E_w$, we use $N_{w}^e$ to denote the number of times this path traverses the edge $e$ (in any direction). We note that equivalent words generate the same graphs (up to graph isomorphism) $G_w$ and the same passage counts $N_{w}^e$.

Coming back to the evaluation of $\bar{T}_N^N$, see (2.1.16), note that any $k$-tuple of integers $i$ defines a closed word $w_i = (i_1, i_2, \ldots, i_k, i_1)$ of length $k + 1$. We write $\text{wt}_i = \text{wt}(w_i)$, which is nothing but the number of distinct integers in $i$. Then,

$$\bar{T}_N^N = \frac{1}{N^{k/2}} \prod_{e \in E_{w_1}} E(Z_{1,2}^N) \prod_{e \in E_{w_1}} E(Y_{1}^{N_{w}^e}) \cdot (2.1.23)$$

In particular, $\bar{T}_N^N = 0$ unless $N_{w_1}^e \geq 2$ for all $e \in E_{w_1}$, which implies that $\text{wt}_i \leq k/2 + 1$. Also, (2.1.23) shows that if $w_i \sim w_{i'}$ then $\bar{T}_N^N = \bar{T}_N^{N'}$. Further, if $N \geq t$ then there are exactly

$$C_{N,t} := N(N-1)(N-2)\cdots(N-t+1)$$

$N$-words that are equivalent to a given $N$-word of weight $t$. We set, with $N > t$,

$$W_{k,t}$$

denotes the equivalent classes of closed $N$-words $w$ of length $k + 1$ and weight $t$ with $N_{w}^e \geq 2$ for each $e \in E_w$ (2.1.24)

(noting that $|W_{k,t}|$ does not depend on $N!$), one deduces from (2.1.16) and (2.1.23) that

$$\langle L_N, x^k \rangle = \sum_{t=1}^{[k/2]+1} \frac{C_{N,t}}{N^{k/2+1}} \sum_{w \in W_{k,t}} \prod_{e \in E_w} E(Z_{1,2}^N) \prod_{e \in E_w} E(Y_{1}^{N_{w}^e}), \quad (2.1.25)$$

where the sum is over a set of representatives, belonging to $W_{k,t}$, of equivalent classes of words.
Note that the cardinality of $W_{k,t}$ is bounded by the number of closed $S$-words of length $k + 1$ when the cardinality of $S$ is $t \leq k$, that is $|W_{k,t}| \leq t^k \leq k^k$. Thus, (2.1.25) and the finiteness of $r_k$, see (2.1.1), imply that
\[
\lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle = 0, \text{ if } k \text{ is odd},
\]
while, for $k$ even,
\[
\lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle = \sum_{w \in W_{k,k/2+1}} \prod_{e \in E_w} E(Z_{\bar{w}^w_{i+1}}) \prod_{e \in E_w} E(Y_{\bar{w}^w_i}) \tag{2.1.26}
\]
We have now motivated the following definition. Note that for the purpose of this section, the case $k = 0$ in definition 2.1.27 is not really needed. It is introduced in this way here in anticipation of the analysis in Section 2.1.6.

**Definition 2.1.27** A closed word $w$ of length $k + 1 \geq 1$ is called a Wigner word if either $k = 0$ or $k$ is even and $w \in W_{k,k/2+1}$.

We next note that if $w \in W_{k,k/2+1}$ then $G_w$ is a tree: indeed, $G_w$ is a connected graph with $|V_w| = k/2 + 1$, hence $|E_w| \geq k/2$, while the condition $N^w_e \geq 2$ for each $e \in E_w$ implies that $|E_w| \leq k/2$. Thus, $|E_w| = |V_w| - 1$, implying that $G_w$ is a tree, that is a connected graph with no loops. Further, the above implies that $E^w_e$ is empty for $w \in W_{k,k/2+1}$, and thus,
\[
\lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle = |W_{k,k/2+1}|. \tag{2.1.28}
\]

We may now complete the:

**Proof of Lemma 2.1.12** Let $k$ be even. Each element $w \in W_{k,k/2+1}$ determines a path $v_1, v_2, \ldots, v_k, v_{k+1} = v_1$ of length $k + 1$ on the tree $G_w$. We refer to this path as the exploration process associated with $w$. Let $d(v, v')$ denotes the distance between vertices $v, v'$ on the tree $G_w$, i.e. the shortest path on the tree beginning at $v$ and terminating at $v'$. Setting $x_i = d(v_{i+1}, v_1)$, one sees that each word $w \in W_{k,k/2+1}$ defines a Dyck path $x_1, x_2, \ldots, x_k = 0$ of length $k$. Conversely, given a Dyck path $x_1, \ldots, x_k$, one may construct a word $w \in W_{k,k/2+1}$ by recursively constructing an increasing (with respect to inclusion) sequence of trees $T_i, i = 1, 2, \ldots, k+1$, and an exploration process $v_i, i = 1, \ldots, k+1$, on $T_{k+1}$, as follows. The tree $T_1$ consists of a single vertex $v$ labelled 1, and we set $v_1 = v$. Given a tree $T_i$ with vertices labelled $1, 2, \ldots, j_i$, and a vertex $v_i \in T_i$, if $x_{i+1} = x_i + 1$ then $T_{i+1}$ consists of the tree $T_i$ augmented with a new vertex $v$ labelled $j_i + 1$, and we set $v_{i+1} = v$. If $x_{i+1} = x_i - 1$, then $T_{i+1} = T_i$ but $v_{i+1}$ is the ancestor of $v_i$ in $T_i$. Taking $w = (s_1, s_2, \ldots, s_{k+1})$ to be the sequence of labels attached to the sequence $\{v_i\}$, one obtains an element of $W_{k,k/2+1}$. 
By construction, the Dyck path corresponding to \((s_1, \ldots, s_{k+1})\) coincides with \((x_1, \ldots, x_k)\). This establishes a bijection between Dyck paths of length \(k\) and \(W_{k,k/2+1}\). Lemma 2.1.7 then establishes that

\[ |W_{k,k/2+1}| = C_{k/2}. \]  (2.1.29)

This completes the proof of Lemma 2.1.12.

\[ \square \]

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.1.2.png}
\caption{Coding of the word \(w = (1, 2, 3, 2, 4, 2, 5, 2, 1)\) into a tree and a Dyck path of length 8. Note that \(\ell(w) = 9\) and \(\text{wt}(w) = 5\).}
\end{figure}

### 2.1.4 Proof of Lemma 2.1.13: Sentences and Graphs

By Chebycheff’s inequality, it is enough to prove that

\[ \lim_{N \to \infty} |E \left( (L_N, x^k)^2 \right) - (\bar{L}_N, x^k)^2| = 0. \]

Proceeding as in (2.1.16), one has

\[ E((L_N, x^k)^2) - (\bar{L}_N, x^k)^2 = \frac{1}{N^2} \sum_{i_1, \ldots, i_k = 1}^{N} \bar{T}^N_{i_1 \ldots i_k}, \]  (2.1.30)

where

\[ \bar{T}^N_{i_1 \ldots i_k} = [ET^N_{i_1 \ldots i_k} - ET^N_{i_1 \ldots i_k}] \]  (2.1.31)
The role of words in the proof of Lemma 2.1.12 is now played by pairs of words, which is a particular case of a sentence.

Definition 2.1.32 (S-Sentences) Given a set \( S \), an \( S \)-sentence \( a \) is a finite sequence of \( S \)-words \( w_1, \ldots, w_n \), at least one word long. Two \( S \)-sentences \( a_1, a_2 \) are called equivalent, denoted \( a_1 \sim a_2 \), if there is a bijection on \( S \) that maps one into the other.

As with words, for a sentence \( a = (w_1, w_2, \ldots, w_n) \), we define the support as \( \text{supp}(a) = \bigcup_{i=1}^n \text{supp}(w_i) \), and its weight \( \text{wt}(a) \) as the cardinality of \( \text{supp}(a) \).

Definition 2.1.33 (Graph associated to an \( S \)-sentence) Given a sentence \( a = (w_1, \ldots, w_k) \), with \( w_i = (s_{i1}, s_{i2}, \ldots, s_{i\ell(w_i)}) \), we set \( G_a = (V_a, E_a) \) to be the graph with set of vertices \( V_a = \text{supp}(a) \) and (undirected) edges

\[
E_a = \{ \{ s_{ij}, s_{ij+1} \}, j = 1, \ldots, \ell(w_i) - 1, i = 1, \ldots, k \}.
\]

We define the set of self edges as \( E_s^a = \{ e \in E_a : e = \{u, u\}, u \in V_a \} \) and the set of connecting edges as \( E_c^a = E_a \setminus E_s^a \).

In words, the graph associated with a sentence \( a = (w_1, \ldots, w_k) \) is obtained by piecing together the graphs of the individual words \( w_i \) (and in general, it differs from the graph associated with the word obtained by concatenating the words \( w_i \)!). Unlike the graph of a word, the graph associated with a sentence may be disconnected. Note that the sentence \( a \) defines \( k \) paths in the graph \( G_a \). For \( e \in E_a \), we use \( N_e^a \) to denote the number of times the union of these paths traverses the edge \( e \) (in any direction). We note that equivalent sentences generate the same graphs \( G_a \) and the same passage counts \( N_e^a \).

Coming back to the evaluation of \( \bar{T}_{i^N} \), see (2.1.30), recall the closed words \( w_i, w_i' \) of length \( k + 1 \), and define the two-words sentence \( a_{i^N} = (w_i, w_i') \). Then,

\[
\bar{T}_{i^N} = \frac{1}{N^k} \left[ \prod_{e \in E_{a_{i^N}}} E(Z_{1,2}^{N_e^a}) \prod_{e \in E_{a_{i^N}}} E(Y_1^{N_e^a}) - \prod_{e \in E_{a_i}} E(Z_{1,2}^{N_e}) \prod_{e \in E_{a_i}} E(Y_1^{N_e}) \prod_{e \in E_{a_i'}} E(Z_{1,2}^{N_e'}) \prod_{e \in E_{a_i'}} E(Y_1^{N_e'}) \right].
\]

(2.1.34)

In particular, \( \bar{T}_{i^N} = 0 \) unless \( N_e^a \geq 2 \) for all \( e \in E_{a_{i^N}} \). Also, \( \bar{T}_{i^N} = 0 \) unless \( E_{w_i} \cap E_{w_i'} \neq \emptyset \). Further, (2.1.34) shows that if \( a_{i^N} \sim a_{j^N} \) then...
\( T_1^N = T_1^{N'} \). Finally, if \( N \geq t \) then there are exactly \( C_{N,t} \) \( N \)-sentences that are equivalent to a given \( N \)-sentence of weight \( t \). We set, with \( N > t \),

\[ W_{k,t}^{(2)} \] denotes the equivalent classes of sentences \( a \) of weight \( t \) consisting of two closed words \((w_1, w_2)\), each of length \( k + 1 \), with \( N_a^c \geq 2 \) for each \( c \in E_a \), and \( E_{w_1} \cap E_{w_2} \neq \emptyset \).

One deduces from (2.1.30) and (2.1.34) that

\[ E((L_N, x^k)^2) - \langle L_N, x^k \rangle^2 \]

\[ = \sum_{t=1}^{2k} \frac{C_{N,t}}{N^{k+2}} \sum_{a=(w_1, w_2) \in W_{k,t}^{(2)}} \left[ \prod_{e \in E_a} E(Z_{1,2}^{N_a}) \prod_{e \in E_a} E(Y_1^{N_a}) \right. \]

\[ - \prod_{a \in E_{u_1}} E(Z_{1,2}^{N_{u_1}}) \prod_{a \in E_{u_2}} E(Y_1^{N_{u_2}}) \prod_{a \in E_{u_2}} E(Z_{1,2}^{N_{u_2}}) \prod_{a \in E_{u_2}} E(Y_1^{N_{u_2}}) \]

where the sum is over a set of representatives, belonging to \( W_{k,t}^{(2)} \) of equivalent classes of sentences. We have completed the preliminaries to:

**Proof of Lemma 2.1.13** In view of (2.1.36), since \( W_{k,t}^{(2)} \) does not depend on \( N \) for \( N \geq t \), it suffices to check that \( W_{k,t}^{(2)} \) is empty for \( t \geq k + 2 \). Since we need it later, we prove a slightly stronger claim, namely that \( W_{k,t}^{(2)} \) is empty for \( t \geq k + 1 \).

Toward this end, note that if \( a \in W_{k,t}^{(2)} \) then \( G_a \) is a connected graph, with \( t \) vertices and at most \( k \) edges (since \( N_a^c \geq 2 \) for \( c \in E_a \)), which is impossible when \( t > k + 1 \). Considering the case \( t = k + 1 \), it follows that \( G_a \) is a tree, and each edge must be visited by the paths generated by \( a \) exactly twice. Because the path generated by \( w_1 \) in the tree \( G_a \) starts and end at the same vertex, it must visit each edge an even number of times. Thus, the set of edges visited by \( w_1 \) is disjoint from the set of edges visited by \( w_2 \), contradicting the definition of \( W_{k,t}^{(2)} \).

\[ \square \]

**Remark 2.1.37** Note that in the course of the proof of Lemma 2.1.13, we actually showed that for \( N > 2k \),

\[ E((L_N, x^k)^2) - \langle L_N, x^k \rangle^2 \]

\[ = \sum_{t=1}^{2k} \frac{C_{N,t}}{N^{k+2}} \sum_{a=(w_1, w_2) \in W_{k,t}^{(2)}} \left[ \prod_{e \in E_a} E(Z_{1,2}^{N_a}) \prod_{e \in E_a} E(Y_1^{N_a}) \right. \]

\[ - \prod_{a \in E_{u_1}} E(Z_{1,2}^{N_{u_1}}) \prod_{a \in E_{u_2}} E(Y_1^{N_{u_2}}) \prod_{a \in E_{u_2}} E(Z_{1,2}^{N_{u_2}}) \prod_{a \in E_{u_2}} E(Y_1^{N_{u_2}}) \].
Exercise 2.1.39 Check that Theorem 2.1.4 still holds for symmetric random matrices \( X_N \) if the zero mean independent random variables \( \{ X_N(i,j) \}_{1 \leq i \leq j \leq N} \) are not assumed identically distributed, but rather one assumes that for all \( \varepsilon > 0 \),
\[
\lim_{N \to \infty} \frac{\# \{(i,j) : |1 - NEX_N(i,j)|^2 < \varepsilon \}}{N^2} = 1,
\]
and for all \( k \geq 1 \), there exists a finite \( r_k \) independent of \( N \) such that
\[
\sup_{1 \leq i \leq j \leq N} E \left| N X_N(i,j) \right|^k \leq r_k.
\]

Exercise 2.1.40 Check that the conclusion of Theorem 2.1.4 remains true when convergence in probability is replaced by almost sure convergence. Hint: Using Chebycheff’s inequality and the Borel-Cantelli lemma, it is enough to verify that for all integer \( k \), there exists a constant \( C = C(k) \) such that
\[
|E \left( (L_N, x^k)^2 \right) - (L_N, x^k)^2| \leq \frac{C}{N^2}.
\]

Exercise 2.1.41 We develop in this exercise the limit theory for Wishart matrices. Let \( M = M(N) \) be a sequence of integers such that \( \lim_{N \to \infty} M(N)/N = \alpha \in [1, \infty) \). Consider an \( N \times N \) matrix \( Y_N \) with i.i.d. entries of mean zero and variance \( 1/N \), and such that \( E \left( N^{k/2} |Y_N(1,1)|^k \right) \leq r_k < \infty \). Define the \( N \times N \) Wishart matrix as \( W_N = Y_N Y_N^* \), and let \( L_N \) denote the empirical measure of the eigenvalues of \( W_N \). Set \( L_N = E L_N \).
(i) Write \( N \langle L_N, x^k \rangle \) as
\[
\sum_{i_1, \ldots, i_k, j_1, \ldots, j_k} EY_{i_1,j_1}Y_{i_2,j_2}Y_{i_3,j_3} \cdots Y_{i_k,j_k}Y_{i_1,j_k}
\]
and show that the only contributions to the sum that survive the passage to the limit are those in which each term appears exactly twice.
(ii) Code the contributions as Dyck paths, where the even heights correspond to \( i \) indices and the odd heights correspond to \( j \) indices. Let \( \ell = \ell(i,j) \) denote the number of times the excursion makes a descent from an odd height to an even height (this is the number of distinct \( j \) indices in the tuple!), and show that the combinatorial weight of such a path is asymptotic to \( N^{k+1} \alpha^\ell \).
(iii) Let \( \ell \) denote the number of times the excursion makes a descent from an even height to an odd height, and set
\[
\beta_k = \sum_{\text{Dyck paths of length } 2k} \alpha^\ell, \quad \gamma_k = \sum_{\text{Dyck paths of length } 2k} \alpha^\ell.
\]
(The $\beta_k$ are the $k$-th moments of any weak limit of $\bar{L}_N$.) Prove that

$$\beta_k = \alpha \sum_{j=1}^{k} \gamma_{k-j} \beta_{j-1}, \gamma_k = \sum_{j=1}^{k} \beta_{k-j} \gamma_{j-1}, k \geq 1.$$  

(iv) Setting $\hat{\beta}_\alpha(z) = \sum_{k=0}^{\infty} z^k \beta_k$, prove that $\hat{\beta}_\alpha(z) = 1 + z \hat{\beta}_\alpha(z)^2 + (\alpha - 1)z \hat{\beta}_\alpha(z)$, and thus the limit $F_\alpha$ of $\bar{L}_N$ possesses the Stieltjes transform $-z^{-1} \hat{\beta}_\alpha(1/z)$, where

$$\hat{\beta}_\alpha(z) = \frac{1 - (\alpha - 1)z - \sqrt{1 - 4z \left[ \frac{\alpha+1}{2} - \frac{(\alpha - 1)^2}{4} \right]}}{2z}.$$  

(v) Conclude that $F_\alpha$ possesses a density $f_\alpha$ supported on $[b_-, b_+]$, with $b_- = (1 - \sqrt{\alpha})^2 / \alpha$, $b_+ = (1 + \sqrt{\alpha})^2 / \alpha$, satisfying

$$f_\alpha(x) = \frac{\alpha \sqrt{(x - b_-)(b_+ - x)}}{2\pi x}, \quad x \in [b_-, b_+]. \quad (2.1.42)$$  

(vi) Prove the analog of Lemma 2.1.13 for Wishart matrices, and deduce that $L_N \to F_\alpha$ weakly, in probability.

(vii) Note that $F_1$ is the image of the semi-circle distribution under the transformation $x \mapsto x^2$.

### 2.1.5 Some useful approximations

This section is devoted to the following simple observation, that often allows one to considerably simplify arguments concerning the convergence of empirical measures.

**Lemma 2.1.43** Let $A, B$ be $N \times N$ symmetric matrices, with eigenvalues $\lambda_1^A \leq \lambda_2^A \leq \ldots \leq \lambda_N^A$ and $\lambda_1^B \leq \lambda_2^B \leq \ldots \leq \lambda_N^B$. Then,

$$\sum_{i=1}^{N} |\lambda_i^A - \lambda_i^B|^2 \leq \text{Tr}(A - B)^2.$$  

**Proof:** Note that $\text{Tr}A^2 = \sum_i (\lambda_i^A)^2$ and $\text{Tr}B^2 = \sum_i (\lambda_i^B)^2$. Let $U$ denote the matrix diagonalizing $B$ written in the basis determined by $A$, and let $D_A, D_B$ denote the diagonal matrices with diagonal elements $\lambda_i^A, \lambda_i^B$ respectively. Then,

$$\text{Tr}AB = \text{Tr}D_AUD_BU^T = \sum_{i,j} \lambda_i^A \lambda_j^B u_{ij}^2.$$
The last sum is linear in the coefficients $v_{ij} = u_{ij}^2$, and the orthogonality of $U$ implies that $\sum_j v_{ij} = 1, \sum_i v_{ij} = 1$. Thus,

$$\text{Tr}AB \leq \sup_{v_{ij} \geq 0, \sum_j v_{ij} = 1, \sum_i v_{ij} = 1} \sum_{i,j} \lambda_i^A \lambda_j^B v_{ij}.$$  \hfill (2.1.44)

But this is a maximization of a linear functional over the set of doubly stochastic matrices, and the maximum is obtained at the extreme points, which are well known to correspond to permutations\(^4\). The maximum among permutations is then easily checked to be $\sum_{i,j} \lambda_i^A \lambda_j^B$. Collecting these facts together implies Lemma 2.1.43. Alternatively\(^5\), one sees directly that a maximizing $V = \{v_{ij}\}$ in (2.1.44), is the identity matrix. Indeed, assume w.l.o.g. that $v_{11} < 1$. We then construct a matrix $\bar{V} = \{\bar{v}_{ij}\}$ with $\bar{v}_{11} = 1$ and $\bar{v}_{ii} = v_{ii}$ for $i > 1$ such that $\bar{V}$ is also a maximizing matrix. Indeed, because $v_{11} < 1$, there exist a $j$ and a $k$ with $v_{1j} > 0$ and $v_{kj} > 0$. Set $v = \min(v_{1j}, v_{kj}) > 0$ and define $\bar{v}_{11} = v_{11} + v$, $\bar{v}_{kj} = v_{kj} + v$ and $\bar{v}_{1j} = v_{1j} - v$, $\bar{v}_{k1} = v_{k1} - v$, and $\bar{v}_{ab} = v_{ab}$ for all other pairs $ab$. Then,

$$\sum_{i,j} \lambda_i^A \lambda_j^B (\bar{v}_{ij} - v_{ij}) = v(\lambda_i^A \lambda_j^B + \lambda_k^A \lambda_j^B - \lambda_k^A \lambda_i^B - \lambda_i^A \lambda_k^B)$$
$$= v(\lambda_i^A - \lambda_k^A)(\lambda_j^B - \lambda_j^B) \geq 0.$$

Thus, $\bar{V} = \{\bar{v}_{ij}\}$ satisfies the constraints, is also a maximum, and the number of zero elements in the first row and column of $\bar{V}$ is larger by 1 at least from the corresponding one for $V$. If $\bar{v}_{11} = 1$, the claims follows, while if $\bar{v}_{11} < 1$, one repeats this (at most $2N - 2$ times) to conclude. Proceeding in this manner with all diagonal elements of $V$, one sees that indeed the maximum of the right hand side of (2.1.44) is $\sum_i \lambda_i^A \lambda_i^B$, as claimed. \( \Box \)

**Remark 2.1.45** One notes that the statement and proof of Lemma 2.1.43 carry over to the case where $A$ and $B$ are both Hermitian matrices.

Lemma 2.1.43 allows one to perform all sorts of truncations when proving convergence of empirical measures. For example, let us prove the following variant of Wigner’s Theorem 2.1.4.

**Theorem 2.1.46** Assume $X_N$ is as in (2.1.2), except that instead of (2.1.1), only $r_2 < \infty$ is assumed. Then, the conclusion of Theorem 2.1.4 still holds.

\(^4\)This theorem is usually attributed to G. Birkhoff. See [Chv83] for a proof and a historical discussion which attributes this result to D. Konig.

\(^5\)Thanks to Hongjie Dong for pointing this out.
Proof: Fix a constant $C$ and consider the matrix $\hat{X}_N$ whose elements satisfy, for $i \leq j$ and $i = 1, \ldots, N$,

$$\hat{X}_N(i, j) = X_N(i, j)1_{\sqrt{N}|X_N(i, j)| \leq C} - E(X_N(i, j))1_{\sqrt{N}|X_N(i, j)| \leq C}.$$

Then, with $\hat{\lambda}_N^N$ denoting the eigenvalues of $\hat{X}_N$, ordered, it follows from Lemma 2.1.43 that

$$\frac{1}{N} \sum_{i=1}^{N} |\lambda_i^N - \hat{\lambda}_i^N|^2 \leq \frac{1}{N} \text{Tr}(X_N - \hat{X}_N)^2.$$

But,

$$W_N := \frac{1}{N} \text{Tr}(X_N - \hat{X}_N)^2 \leq \frac{1}{N^2} \sum_{i,j} \left[ \sqrt{N}X_N(i, j)1_{\sqrt{N}|X_N(i, j)| \geq C} - E(\sqrt{N}X_N(i, j))1_{\sqrt{N}|X_N(i, j)| \geq C} \right]^2.$$

Since $r_2 < \infty$, and the involved random variables are identical in law to either $Z_{1,2}$ or $Y_1$, it follows that $E[(\sqrt{N}X_N(i, j))^21_{\sqrt{N}|X_N(i, j)| \geq C}]$ converges to 0 uniformly in $N, i, j$, when $C$ converges to infinity. Hence, one may chose for each $\varepsilon$ a $C$ large such that $P(|W_N| > \varepsilon) < \varepsilon$. Further, let

$$\text{Lip}(\mathbb{R}) = \{ f \in C_b(\mathbb{R}) : \sup_x |f(x)| \leq 1, \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|} \leq 1 \}.$$

Then, on the event $\{|W_N| < \varepsilon\}$, it holds that for $f \in \text{Lip}(\mathbb{R})$,

$$|\langle L_N, f \rangle - \langle \hat{L}_N, f \rangle| \leq \frac{1}{N} \sum_i |\lambda_i^N - \hat{\lambda}_i^N| \leq \sqrt{\varepsilon},$$

where $\hat{L}_N$ denotes the empirical measure of the eigenvalues of $\hat{X}_N$, and Jensen’s inequality was used in the second inequality. This, together with the weak convergence in probability of $L_N$ toward the semicircle law assured by Theorem 2.1.4, and the fact that weak convergence is equivalent to convergence with respect to the Lipschitz bounded metric, see Theorem C.8, complete the proof of Theorem 2.1.46.

2.1.6 Maximal eigenvalues and Füredi-Komlós enumeration

Wigner’s theorem asserts the weak convergence of the empirical measure of eigenvalues to the compactly supported semi-circle law. One immediately
is led to suspect that the maximal eigenvalue of $X_N$ should converge to the value $2$, the largest element of the support of the semi-circle distribution. This fact, however, does not follow from Wigner’s theorem. However, the combinatorial techniques we already saw allow one to prove the following, where we use the notation introduced in (2.1.1) and (2.1.2).

**Theorem 2.1.47 (Maximal eigenvalue)** Consider a Wigner matrix $X_N$ satisfying $r_k \leq k^Ck$ for some constant $C$ and all integer $k$. Then, $\lambda_N^N$ converges to $2$ in probability.

**Remark:** The assumption of Theorem 2.1.47 holds if the random variables $|Z_{1,2}|$ and $|Y_1|$ possess a finite exponential moment.

**Proof of Theorem 2.1.47** Fix $\delta > 0$ and let $g : \mathbb{R} \to \mathbb{R}_+$ be a continuous function supported on $[2 - \delta, 2]$, with $\langle g, \sigma \rangle = 1$. Then, applying Wigner’s theorem 2.1.4,

$$P(\lambda_N^N < 2 - \delta) \leq P(|\langle L_N, g \rangle - \langle \sigma, g \rangle| > \frac{1}{2}) \to_{N \to \infty} 0.$$  

(2.1.48)

We thus need to provide a complimentary estimate on the probability that $\lambda_N^N$ is large. We do that by estimating $\langle \bar{L}_N^N, x^{2k} \rangle$ for $N$-dependent $k$, using the bounds on $r_k$ provided in the assumptions. The key step is contained in the following combinatorial lemma, that gives information on the sets $W_{k,t}$, see (2.1.24).

**Lemma 2.1.49** For all integers $k > 2t - 2$ and $N > t$ one has the estimate

$$|W_{k,t}| \leq 2^k k^{3(k-2t+2)}.$$  

(2.1.50)

The proof of Lemma 2.1.49 is deferred to the end of this section.

Equipped with Lemma 2.1.49, we have for $2k < N$, using (2.1.25),

$$\langle L_N^N, x^{2k} \rangle \leq \sum_{t=1}^{k+1} N^{t-(k+1)} |W_{2k,t}| \sup_{w \in W_{2k,t}} \prod_{e \in E_w^c} E(Z_{1,2}^N) \prod_{e \in E_w} E(Y_1^N)$$  

$$\leq 4^k \sum_{t=1}^{k+1} \left(\frac{2k}{N}\right)^{k+1-t} \sup_{w \in W_{2k,t}} \prod_{e \in E_w^c} E(Z_{1,2}^N) \prod_{e \in E_w} E(Y_1^N).$$

To evaluate the last expectation, fix $w \in W_{2k,t}$, and let $l$ denote the number of edges in $E_w^c$ with $N_e^w = 2$. Hölder’s inequality then gives

$$\prod_{e \in E_w^c} E(Z_{1,2}^N) \prod_{e \in E_w} E(Y_1^N) \leq r_{2k-2t},$$
with the convention that $r_0 = 1$. Since $G_v$ is connected, $|E_v| \geq |V_v| - 1 = t - 1$. On the other hand, by noting that $N_v^e \geq 3$ for $|E_v^e| - l$ edges, one has $2k \geq 3(|E_v^e| - l) + 2l + 2|E_v^e|$. Hence, $2k - 2l \leq 6(k + 1 - t)$. Since $r_{2q}$ is a non-decreasing function of $q$ bounded below by 1, we get, substituting back in (2.1.51), that for some constant $c_1 = c_1(C)$ and all $k < N$,

$$\langle \bar{L}_N, x^{2k} \rangle \leq 4^k \sum_{t=1}^{k+1} \left[ \frac{(2k)^6}{N} \right] r_0^{k+1-t} \leq 4^k \sum_{t=1}^{k+1} \left[ \frac{(2k)^6(6(k + 1 - t))}{N} \right]^{k+1-t} \leq 4^k \sum_{i=0}^{k} \left[ \frac{k^{c_1}}{N} \right]^i.$$  \hspace{1cm} (2.1.52)

Choose next a sequence $k(N) \to N \to \infty$ such that $k(N)^{c_1}/N \to \infty$ but $k(N)/\log N \to \infty$. Then, for any $\delta > 0$, and all $N$ large,

$$P(\lambda_N^N > (2 + \delta)) \leq P(N\langle \bar{L}_N, x^{2k(N)} \rangle > (2 + \delta)^{2k(N)}) \leq N\langle \bar{L}_N, x^{2k(N)} \rangle \leq \frac{2N4^{k(N)}}{(2 + \delta)^{2k(N)}} \to N \to \infty 0,$$

completing the proof of Theorem 2.1.47. \hfill $\Box$

**Proof of Lemma 2.1.49** The idea of the proof it to keep track of the number of possibilities to prevent words in $W_{k,t}$ from having weight $\lfloor k/2 \rfloor + 1$. Toward this end, let $w \in W_{k,t}$ be given. A parsing of the word $w$ is a sentence $a_w = (w_1, \ldots, w_n)$ such that the word obtained by concatenating the words $w_i$ is $w$. One can imagine creating a parsing of $w$ by introducing commas between parts of $w$.

We say that a parsing $a = a_w$ of $w$ is an FK parsing, and call the sentence $a$ an FK sentence, if the graph associated with $a$ is a tree, if $N^a_e \leq 2$ for all $e \in E_a$, and if for any $i = 1, \ldots, n - 1$, the first letter of $w_{i+1}$ belongs to $\cup_{j=1}^{i} \text{supp} w_j$. If the one word sentence $a = w$ is an FK parsing, we say that $w$ is an FK word. Note that the constituent words in an FK parsing are FK words.

As will become clear next, the graph of an FK word consists of trees whose edges have been visited twice by $w$, glued together by edges that have been visited only once. Recalling that a Wigner word is either a one letter word or a closed word of odd length and maximal weight, this leads to the following lemma.
Lemma 2.1.53 Each FK word can be written in a unique way as a concatenation of pairwise disjoint Wigner words. Further, there are at most $2^{n-1}$ equivalence classes of FK words of length $n$.

Proof of Lemma 2.1.53 Let $w = (s_1, \ldots, s_n)$ be an FK word of length $n$. By definition, $G_w$ is a tree. Let $\{s_{i_j}, s_{i_j+1}\}_{j=1}^r$ denote those edges of $G_w$ visited only once by the walk induced by $w$. Defining $i_0 = 1$, one sees that the words $\bar{w}_j = (s_1, \ldots, s_{i_j-1})$, $j \geq 1$, are closed, disjoint, and visit each edge in the tree $G_{\bar{w}_j}$ exactly twice. In particular, with $l_j := i_j - i_{j-1} - 1$, it holds that $l_j$ is even (possibly, $l_j = 0$ if $\bar{w}_j$ is a one letter word), and further if $l_j > 0$ then $\bar{w}_j \in W_{l_j, l_j/2+1}$. This decomposition being unique, one concludes that for any $z$, with $N_n$ denote the number of equivalence classes of FK words of length $n$, and with $|W_{0,1}| := 1$, |

\[ \sum_{n=1}^{\infty} N_n z^n = \sum_{r=1}^{\infty} \sum_{(t_j)_{j=1}^r} \prod_{j=1}^{r} z^{l_j+1} |W_{l_j, l_j/2+1}| \]

in the sense of formal power series. By the proof of Lemma 2.1.12, $|W_{2l, l+1}| = C_l = \beta_l$. Hence, by Lemma 2.1.7, for $|z| < 1/4$,

\[ z + \sum_{l=1}^{\infty} z^{2l+1} |W_{2l, l+1}| = z \hat{\beta}(z^2) = \frac{1 - \sqrt{1 - 4z^2}}{2z} \]

Substituting in (2.1.54), one sees that (again, in the sense of power series)

\[ \sum_{n=1}^{\infty} N_n z^n = \frac{z \hat{\beta}(z^2)}{1 - z \hat{\beta}(z^2)} = \frac{1 - \sqrt{1 - 4z^2}}{2z - 1 + \sqrt{1 - 4z^2}} = \frac{1}{2} + \frac{z + 1}{\sqrt{1 - 4z^2}} \]

Using that

\[ \sqrt{\frac{1}{1-t}} = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left( \begin{array}{c} 2k \\ k \end{array} \right) \]

one concludes that

\[ \sum_{n=1}^{\infty} N_n z^n = \frac{z}{2} (1 + 2z) \sum_{n=1}^{\infty} z^{2n} \left( \begin{array}{c} 2n \\ n \end{array} \right) , \]

from which Lemma 2.1.53 follows. \(\Box\)

Our interest in FK parsings is the following FK parsing $w'$ of a word $w = (s_1, \ldots, s_n)$. Declare an edge $e$ of $G_w$ to be new (relative to $w$) if for
some index $1 \leq i < n$ we have $e = \{s_i, s_{i+1}\}$ and $s_{i+1} \not\in \{s_1, \ldots, s_i\}$. If the edge $e$ is not new, then it is *old*. Define $w'$ to be the sentence obtained by breaking (that is, “inserting a comma”) at all visits to old edges of $G_w$ and at third and subsequent visits to new edges of $G_w$.

![Figure 2.1.3](image)

Figure 2.1.3: Two inequivalent FK sentences $[x_1, x_2]$ corresponding to (solid line) $b = 141252363$ and (dashed line) $c = 1712$ (in left) $\sim 3732$ (in right).

Since a word $w$ can be recovered from its FK parsing by omitting the extra commas, and since the number of equivalent classes of FK words is estimated by Lemma 2.1.53, one could hope to complete the proof of Lemma 2.1.49 by controlling the number of possible parsed FK sequences. A key step toward this end is the following lemma, that clarifies the picture of an FK sentence as FK words glued together on edges visited once by each word. Recall that any FK word $w$ can be written in a unique way as a concatenation of disjoint Wigner words $w_i$, $i = 1, \ldots, r$. With $s_i$ denoting the first (and last) letter of $w_i$, define the *skeleton* of $w$ as the word $s_1, \ldots, s_r$. Finally, for a sentence $a$ with graph $G_a$, let $G^1_a = (V^1_a, E^1_a)$ be the graph with vertex set $V_a = V^1_a$ and edge set $E^1_a = \{e \in E_a : N^1_e = 1\}$. Clearly, when $a$ is an FK sentence, $G^1_a$ is always a forest, that is a disjoint union of trees.

**Lemma 2.1.55** Suppose $b$ is an FK sentence with $n - 1$ words and $c$ is an FK word with skeleton $s_1, \ldots, s_r$ with $s_1 \in \text{supp}(b)$. Let $\ell$ be the largest index with $s_\ell \in \text{supp} b$, and set $d = s_1, \ldots, s_\ell$. Then $a = (b, c)$ is an FK sentence only if $\sup b \cap \text{supp} c = \text{supp} d$ and $d$ is a geodesic in $G^1_b$.

(A *geodesic* connecting $x, y \in G^1_b$ is a path of minimal length starting at $x$ and terminating at $y$).

---

\[^6\text{In fact, if and only if. The sufficiency, which we do not need, is straight forward to check.}\]
Before providing the proof of Lemma 2.1.55, we explain how it leads to the
Completion of proof of Lemma 2.1.49 Let $\Gamma(t, \ell, m)$ denote the set of equivalence classes of FK sentences $a = (w_1, \ldots, w_m)$ consisting of $m$ words, with total length $\sum_{i=1}^{m} \ell(w_i) = \ell$ and $\text{wt}(a) = t$. An immediate corrolary of Lemma 2.1.55 is that
\[
|\Gamma(t, \ell, m)| \leq 2^\ell \cdot m^{2(m-1)} \left( \frac{\ell - 1}{m - 1} \right).
\] (2.1.56)
Indeed, there are $c_{\ell, m} := \left( \frac{\ell - 1}{m - 1} \right)$ $m$-tuples of integers summing to $\ell$, and thus at most $2^\ell \cdot m^{2(m-1)}$ equivalence classes of $m$ FK words with sum of lengths equal to $\ell$. Lemma 2.1.55 then shows that there are at most $\ell^2(m-1)$ ways to “glue these words into an FK sentence”, whence (2.1.56) follows.

For any FK sentence $a$ consisting of $m$ words with total length $\ell$, we have that
\[
m = |E_a^1| - 2\text{wt}(a) + 2 + \ell.
\] (2.1.57)
Indeed, the word obtained by concatenating the words of $a$ generates a list of (not necessarily distinct) $\ell - 1$ unordered pairs of adjoining letters, out of which $m - 1$ correspond to commas in the FK sentence $a$ and $2|E_a^1| - |E_a^1|$ correspond to edges of $G_a$. Using that $|E_a| = |V_a| - 1$, (2.1.57) follows. But then, any word $w \in \mathcal{W}_{k,t}$ can be parsed into an FK sentence $w'$ consisting of $m$ words. Note that if an edge $e$ is retained in $G_w'$, then no comma is inserted at $e$ at the first and second passage on $e$ (but is introduced if there are further passages on $e$). Therefore, $E_{w'}^1 = \emptyset$. By (2.1.57), this implies that for such words, $m - 1 = k + 2 - 2t$. Equation (2.1.56) then allows one to conclude the proof of Lemma 2.1.49.

**Proof of Lemma 2.1.55** Assume $a$ is an FK sentence. Then, $G_a$ is a tree, and since the Wigner words composing $c$ are disjoint, $d$ is the unique geodesic in $G_c \subset G_a$ connecting $s_1$ to $s_t$. Hence, it is also the unique geodesic in $G_b \subset G_a$ connecting $s_1$ to $s_t$. But $d$ visits only edges of $G_b$ that have been visited exactly once by the constituent words of $b$, for otherwise $(b, c)$ would not be an FK sentence (that is, a comma would need to be inserted to split $c$). Thus, $E_d \subset E_b^1$. Since $c$ is an FK word, $E_c^1 = E_c = E_{(s_1, \ldots, s_t)}$. Since $a$ is an FK sentence, $E_b \cap E_c = E_b^1 \cap E_c^1$. Thus, $E_b \cap E_c = E_d$. But, recall that $G_a$, $G_b$, $G_c$, $G_d$ are trees, and hence
\[
|V_a| = 1 + |E_a| = 1 + |E_b| + |E_c| - |E_b \cap E_c| = 1 + |E_b| + |E_c| - |E_d|
\]
\[
= 1 + |E_b| + 1 + |E_c| - 1 - |E_d| = |V_b| + |V_c| - |V_d|.
\]
Since $|V_b| + |V_c| - |V_b \cap V_c| = |V_a|$, it follows that $|V_d| = |V_b \cap V_c|$. Since $V_d \subset V_b \cap V_c$, one concludes that $V_d = V_b \cap V_c$, as claimed. \qed
Remark 2.1.58 The result described in Theorem 2.1.47 is not optimal, in the sense that even with uniform bounds on the (rescaled) entries, i.e. $r_k$ uniformly bounded, the estimate one gets on the displacement of the maximal eigenvalue to the right of 2 is $O(n^{-1/6} \log n)$, whereas the true displacement is known to be of order $n^{-1/6}$ ([TW96], [Sos99]). For more on that in the context of complex Gaussian Wigner matrices, see Theorem 3.1.10.

Exercise 2.1.59 Prove that the conclusion of Theorem 2.1.47 holds with convergence in probability replaced by either almost sure convergence or $L^p$ convergence.

Exercise 2.1.60 Prove that the statement of Theorem 2.1.47 can be strengthened to yield that for some constant $\delta = \delta(C) > 0$, $N^{-1/6}(\lambda_N - 2)$ converges to 0, almost surely.

Exercise 2.1.61 Assume that for some constants $\lambda > 0$, $C$, the independent (but not necessarily identically distributed) entries $\{X_N(i,j)\}_{1 \leq i \leq j \leq N}$ of the symmetric matrices $X_N$ satisfy

$$\sup_{i,j,N} E(e^{\lambda \sqrt{N} |X_N(i,j)|}) \leq C.$$ 

Prove that there exists a constant $c_1 = c_1(C)$ such that $\limsup_{N \to \infty} \lambda_N \leq c_1$, almost surely, and $\limsup_{N \to \infty} E\lambda_N \leq c_1$.

Exercise 2.1.62 We develop in this exercise an alternative proof, that avoids moment computations, to the conclusion of Exercise 2.1.61, under the stronger assumption that for some $\lambda > 0$,

$$\sup_{i,j,N} E(e^{M \sqrt{N} |X_N(i,j)|}) \leq C.$$ 

a) Prove (using Chebycheff’s inequality and the assumption) that there exists a constant $c_0$ independent of $N$ such that for any fixed $z \in \mathbb{R}^N$, and all $C$ large enough,

$$P(\|z^T X_N z\|_2 > C) \leq e^{-c_0 C^2 N}.$$ 

b) Let $N_\delta = \{z_i\}_{i=1}^{N_\delta}$ be a minimal deterministic net in the unit ball of $\mathbb{R}^N$, that is $\|z_i\|_2 = 1$, $\sup \|z - z_i\|_2 \leq \delta$, and $N_\delta$ is the minimal integer with the property that such a net can be found. Check that

$$(1 - \delta^2) \sup_{\|z\|_2 = 1} z^T X_N z \leq \sup_{\|z\|_2 = 1} z_i^T X_N z_i + 2 \sup_{z \in N_\delta} \sup_{\|z - z_i\|_2 \leq \delta} z^T X_N z_i.$$ 

c) Combine steps a) and b) and the estimate $N_\delta \leq c_\delta^2$, valid for some $c_\delta > 0$, to conclude that there exists a constant $c_2$ independent of $N$ such that for all $C$ large enough, independently of $N$,

$$P(\lambda_N^2 > C) = P(\sup_{\|z\|_2 = 1} z^T X_N z > C) \leq e^{-c_2 C^2 N}.$$
2.1.7 Central limit theorems for moments

Our goal here is to derive a simple version of a central limit theorem (CLT) for linear statistics of the eigenvalues of Wigner matrices. With $X_N$ a Wigner matrix and $L_N$ the associated empirical measure of its eigenvalues, set $Z_{N,k} := N[(L_N, x^k) - \langle L_N, x^k \rangle]$. Let

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

denote the Gaussian distribution. We set $\sigma_k^2$ as in (2.1.75) below, and prove the following.

**Theorem 2.1.65** The law of the sequence of random variables $Z_{N,k}/\sigma_k$ converges weakly to the standard Gaussian distribution. More precisely,

$$\lim_{N \to \infty} P \left( \frac{Z_{N,k}}{\sigma_k} \leq x \right) = \Phi(x). \quad (2.1.66)$$

**Proof of Theorem 2.1.65** Most of the proof consists of a variance computation. The reader interested only in a proof of convergence to a Gaussian distribution (without worrying about the actual variance) can skip to the text following equation (2.1.75).

Recall the notation $W^{(2)}_{k,k}$ introduced in the course of proving Lemma 2.1.13. Using (2.1.38), we have

$$\lim_{N \to \infty} E(Z_{N,k}^2) \quad (2.1.67)$$

$$= \lim_{N \to \infty} N^2 \left[ E((L_N, x^k)^2) - \langle L_N, x^k \rangle^2 \right]$$

$$= \sum_{a = (w_1, w_2) \in W^{(2)}_{k,k}} \left[ \prod_{e \in E^+_a} E(Z_{1/2}^{N_{e^+}^a}) \prod_{e \in E^-_a} E(Y_{1/2}^{N_{e^+}^a}) \right. \prod_{e \in E^+_a} E(Z_{1/2}^{N_{e^+}^a}) \prod_{e \in E^-_a} E(Y_{1/2}^{N_{e^+}^a}) \bigg].$$

We note next that if $a = (w_1, w_2) \in W^{(2)}_{k,k}$ then $G_a$ is connected and possesses $k$ vertices and at most $k$ edges, each visited at least twice by the paths generated by $a$. Hence, with $k$ vertices, $G_a$ possesses either $k - 1$ or $k$ edges. Let $W^{(2)}_{k,k,+}$ denote the subset of $W^{(2)}_{k,k}$ such that $|E_a| = k$ (that is, $G_a$ is unicyclic, i.e. “possesses one edge too much to be a tree”) and let $W^{(2)}_{k,k,-}$ denote the subset of $W^{(2)}_{k,k}$ such that $|E_a| = k - 1$.

Suppose first $a \in W^{(2)}_{k,k,-}$. Then, $G_a$ is a tree, $E_a^s = 0$, and necessarily $G_{w_1}$ is a subtree of $G_a$. This implies that $k$ is even and that $|E_{w_1}| \leq k/2$. 


In this case, for $E_{w_1} \cap E_{w_2} \neq \emptyset$ one must have $|E_{w_i}| = k/2$, which implies that all edges of $G_{w_i}$ are visited twice by the walk generated by $w_i$, and exactly one edge is visited twice by both $w_1$ and $w_2$. In particular, $w_i$ are both closed Wigner words of length $k + 1$. The emerging picture is of two trees with $k/2$ edges each “glued together” at one edge. Since there are $C_{k/2}$ ways to choose each of the trees, $k/2$ ways of choosing (in each tree) the edge to be glued together, and $2$ possible orientations for the glueing, we deduce that

$$|W^{(2)}_{k,k,-}| = 2 \left(\frac{k}{2}\right)^2 C_{k/2}^2.$$  \hspace{1cm} (2.1.68)

Further, for each $a \in W^{(2)}_{k,k,-}$,

$$\left[ \prod_{e \in E_{a_1}} E(Z_{1,2}^{N_a}) \prod_{e \in E_{a_1}} E(Y_{1}^{N_a}) \right] - \prod_{e \in E_{a_1}} E(Z_{1,2}^{N_a}) \prod_{e \in E_{a_1}} E(Y_{1}^{N_a}) \prod_{e \in E_{a_2}} E(Z_{1,2}^{N_a}) \prod_{e \in E_{a_2}} E(Y_{1}^{N_a}) \right] = E(Z_{1,2})[E(Z_{1,2})]^{k-2} - |E(Z_{1,2})|^k = E(Z_{1,2}) - 1.$$ \hspace{1cm} (2.1.69)

We next turn to consider $W^{(2)}_{k,k,+}$. In order to do so, we need to understand the structure of unicyclic graphs.

**Definition 2.1.70** A graph $G = (V, E)$ is called bracelet if there exists an enumeration $\alpha_1, \alpha_2, \ldots, \alpha_r$ of $V$ such that

$$E = \begin{cases} 
\{\{\alpha_1, \alpha_1\}\} & \text{if } r = 1, \\
\{\{\alpha_1, \alpha_2\}\} & \text{if } r = 2, \\
\{\{\alpha_1, \alpha_2\}, \{\alpha_2, \alpha_3\}, \{\alpha_3, \alpha_4\}\} & \text{if } r = 3, \\
\{\alpha_1, \alpha_2\}, \{\alpha_2, \alpha_3\}, \{\alpha_3, \alpha_4\}, \{\alpha_4, \alpha_1\}\} & \text{if } r = 4, 
\end{cases}$$

and so on. We call $r$ the circuit length of the bracelet $G$.

We need the following elementary lemma, allowing one to decompose a unicyclic graph as a bracelet and its associated pendant trees. Recall that a graph $G = (V, E)$ is unicyclic if it is connected and $|E| = |V|$.

**Lemma 2.1.71** Let $G = (V, E)$ be a unicyclic graph. Let $Z$ be the subgraph of $G$ consisting of all $e \in E$ such that $G \setminus e$ is connected, along with all attached vertices. Let $r$ be the number of edges of $Z$. Let $F$ be the graph obtained from $G$ by deleting all edges of $Z$. Then, $Z$ is a bracelet of circuit length $r$, $F$ is a forest with exactly $r$ connected components, and $Z$ meets each connected component of $F$ in exactly one vertex. Further, $r = 1$ if $E^* \neq \emptyset$ while $r \geq 3$ otherwise.
We call $Z$ the *bracelet* of $G$. We call $r$ the *circuit length* of $G$, and each of the components of $F$ we call a *pendant tree*.

![Figure 2.1.4: The bracelet 1234 of circuit length 4, and the pendant trees, associated with the unicyclic graph corresponding to \[1256572341, 2383412\]](image)

Coming back to $a \in W_{k,k,+}^{(2)}$, let $Z_a$ be the associated bracelet (with circuit length $r = 1$ or $r \geq 3$). Note that for any $e \in E_a$ one has $N_e^a = 2$. We claim next that $e \in Z_a$ if and only if $N_e^{w_1} = N_e^{w_2} = 1$: indeed, if $e \in Z_a$ then $(V_a, E_a \setminus e)$ is a tree. If one of the paths determined by $w_1$ and $w_2$ fail to visit $e$ then all edges visited by this path determine a walk on a tree and therefore the path visits each edge exactly twice. This then implies that the set of edges visited by the walks are disjoint, a contradiction. On the other hand, if $e = (x,y)$ and $N_e^{w_i} = 1$ then all vertices in $V_{w_i}$ are connected to $x$ and to $y$ by a path using only edges from $E_{w_i} \setminus e$. Hence, $(V_a, E_a \setminus e)$ is connected, and thus $e \in Z_a$.

Thus, any $a = (w_1, w_2) \in W_{k,k,+}^{(2)}$ with bracelet length $r$ can be constructed from the following data: the pendant trees $\{T_j\}_{j=1}^r$ (possibly empty) associated to each word $w_i$ and each vertex $j$ of the bracelet $Z_a$, the starting point for each word $w_i$ on the graph consisting of the bracelet $Z_a$ and trees $\{T_j\}$, and whether $Z_a$ is traversed by the words $w_i$ in the same or in opposing directions (in case $r \geq 3$). In view of the above, counting the number of ways to attach trees to a bracelet of length $r$, and then the distinct number of non-equivalent ways to choose starting points for the paths on the resulting graph, there are exactly

$$
\frac{2^{1+r}k^2}{r} \left( \sum_{\kappa_i \geq k} \prod_{i=1}^r C_{k_i} \right)^2
$$

(2.1.72)
elements of $W^{(2)}_{k,k,+}$ with bracelet of length $r$. Further, for $a \in W^{(2)}_{k,k,+}$ we have

$$
\left[ \prod_{e \in E_a} E(Z_{1,2}^{N_a}) \prod_{e \in E_{a_1}} E(Y_{1,2}^{N_{a_1}}) \right] - \prod_{e \in E_{a_2}} E(Z_{1,2}^{N_{a_2}}) \prod_{e \in E_{a_3}} E(Y_{1,2}^{N_{a_3}}) \prod_{e \in E_{a_4}} E(Y_{1,2}^{N_{a_4}})
$$

$$
= \left\{ \begin{array}{ll}
(E(Z_{1,2}^2))^k - 0 & \text{if } r \geq 3 \\
(E(Z_{1,2}^2))^{k-1} EY_1^2 - 0 & \text{if } r = 1 \\
1 & \text{if } r \geq 3 \\
EY_1^2 & \text{if } r = 1
\end{array} \right.
$$

Combining (2.1.67), (2.1.68), (2.1.69), (2.1.72) and (2.1.73), and setting $C_x = 0$ if $x$ is not an integer, one obtains, with

$$
\sigma_k^2 = k^2 C_2 \frac{1}{2} EY_1^2 + \frac{k^2}{2} C_2 \left[ EZ_{1,2}^1 - 1 \right] + \sum_{r=3}^{\infty} \frac{2k^2}{r} \left( \sum_{k_i \geq 0} \prod_{i=1}^{r} C_{k_i} \right)^2,
$$

that

$$
\sigma_k^2 = \lim_{N \to \infty} EZ_{N,k}^2.
$$

(2.1.74)

(2.1.75)

The rest of the proof consists in verifying that, for $j \geq 3$,

$$
\lim_{N \to \infty} E \left( \frac{Z_{N,k}}{\sigma_k} \right)^j = \left\{ \begin{array}{ll}
0, & j \text{ is odd}, \\
(j-1)!!, & j \text{ is even}.
\end{array} \right.
$$

(2.1.76)

where $(j-1)!! = (j-1)(j-3) \cdots 1$. Indeed, this completes the proof of the theorem since the right hand side of (2.1.76) coincides with the moments of the Gaussian distribution $\Phi$, and the latter moments determine the Gaussian distribution by an application of Carleman’s theorem (see, e.g., [Dur96]), since $\sum_{n=1}^{\infty} [(2j-1)!!(1/2j)] = \infty$.

To see (2.1.76), recall, for a multi-index $i = (i_1, \ldots, i_k)$, the terms $T_{i,N}$ of (2.1.23), and the associated closed word $w_i$. Then, as in (2.1.30), one has

$$
E(Z_{N,k}^j) = \sum_{i_1^n, \ldots, i_k^n = 1}^{N} T_{i_1^n, \ldots, i_k^n} N \prod_{n=1}^{j} (T_{i_1^n}^N - ET_{i_1^n}^N)
$$

(2.1.77)

where

$$
T_{i_1^n, \ldots, i_k^n} = E \left[ \prod_{i_1^n}^{j} (T_{i_1^n}^N - ET_{i_1^n}^N) \right].
$$

(2.1.78)
Note that $\bar{T}_{N,i_1,i_2,...,i_j} = 0$ if the graph generated by any word $w_n := w_t$ does not have an edge in common with any graph generated by the other words $w_{n'}, n' \neq n$. Motivated by that and our variance computation, we set, for $N > t$,

$W_{k,t}^{(j)}$ denotes the equivalent classes of sentences $a$ of weight $t$ consisting of $j$ closed words $(w_1, w_2, ..., w_j)$, each of length $k + 1$, with $N^a \geq 2$ for each $e \in E_a$, and such that for each $n$ there is an $n' = n''(n) \neq n$ such that $E_{w_n} \cap E_{w_{n'}} \neq \emptyset$.

(2.1.79)

As in (2.1.36), one obtains

\[
E(Z_{j,N,k}) = \sum_{t=1}^{jk} C_{N,t} \sum_{a=(w_1,w_2,...,w_j) \in W_{k,t}^{(j)}} \bar{T}_{w_1,w_2,...,w_j} := \sum_{t=1}^{jk} \frac{C_{N,t}}{N^{jk/2}} \sum_{a \in W_{k,t}^{(j)}} \bar{T}_a.
\]

(2.1.80)

The next lemma, whose proof is deferred to the end of the section, is concerned with a study of $W_{k,t}^{(j)}$.

**Lemma 2.1.81** Let $c$ denote the number of connected components of $G_a$ for $a \in \cup_t W_{k,t}^{(j)}$. Then, $c \leq \lfloor j/2 \rfloor$ and $\text{wt}(a) \leq c - j + \lfloor (k + 1)j/2 \rfloor$.

In particular, Lemma 2.1.81 and (2.1.80) imply that

\[
\lim_{N \to \infty} E(Z_{j,N,k}) = \begin{cases} 0 & \text{if } j \text{ is odd} \\ \sum_{a \in W_{k,\lfloor j/2 \rfloor}^{(j)}} \bar{T}_a & \text{if } j \text{ is even} \end{cases}.
\]

(2.1.82)

By Lemma 2.1.81, if $a \in W_{k,\lfloor j/2 \rfloor}^{(j)}$ for $j$ even then $G_a$ possesses exactly $j/2$ connected components. This is possible only if there exists a permutation $\pi : \{1, ..., j\} \mapsto \{1, ..., j\}$, all of whose cycles having length 2 (that is, a matching), such that the connected components of $G_a$ are the graphs $\{G_{(w_i, w_{\pi(i)})}\}$. Letting $\Sigma^m_j$ denote the collection of all possible matchings, one thus obtains that for $j$ even,

\[
\sum_{a \in W_{k,\lfloor j/2 \rfloor}^{(j)}} \bar{T}_a = \prod_{\pi \in \Sigma^m_j} \sum_{i=1}^{j/2} \bar{T}_{w_i, w_{\pi(i)}} = \sum_{\pi \in \Sigma^m_j} \sigma^{j-1}_k \sigma^{j}_k = \sigma^{j}_k (j-1)!!,
\]

(2.1.83)

which, together with (2.1.82), completes the proof of Theorem 2.1.65. □

**Proof of Lemma 2.1.81** That $c \leq \lfloor j/2 \rfloor$ is immediate from the fact that the subgraph corresponding to any word in $a$ must have at least one edge
in common with at least one subgraph corresponding to another word in $a$. 

Next, put

$$a = [\alpha_n]_{n=1}^k; \quad I = \bigcup_{i=1}^j \{i\} \times \{1, \ldots, k\}, \quad A = [\{\alpha_i,\alpha_i+1\}]_{(i,n) \in I}.$$ 

We visualize $A$ as a left-justified table of $j$ rows. Let $G' = (V', E')$ be any spanning forest in $G_a$, with $c$ connected components. Since every connected component of $G'$ is a tree, we have

$$\text{wt}(a) = c + |E'|.$$  \hspace{1cm} (2.1.84)

Now let $X = \{X_{i,n}\}_{(i,n) \in I}$ be a table of the same "shape" as $A$, but with all entries equal either to 0 or 1. We call $X$ an edge-bounding table under the following conditions:

- For all $(i,n) \in I$, if $X_{i,n} = 1$, then $A_{i,n} \in E'$.
- For each $e \in E'$ there exist distinct $(i_1,n_1), (i_2,n_2) \in I$ such that $X_{i_1,n_1} = X_{i_2,n_2} = 1$ and $A_{i_1,n_1} = A_{i_2,n_2} = e$.
- For each $e \in E'$ and index $i \in \{1, \ldots, j\}$, if $e$ appears in the $i^{th}$ row of $A$ then there exists $(i,n) \in I$ such that $A_{i,n} = e$ and $X_{i,n} = 1$.

For any edge-bounding table $X$ the corresponding quantity $\frac{1}{2} \sum_{(i,n) \in I} X_{i,n}$ bounds $|E'|$. At least one edge-bounding table exists, namely the table with a 1 in position $(i,n)$ for each $(i,n) \in I$ such that $A_{i,n} \in E'$ and 0's elsewhere. Now let $X$ be an edge-bounding table such that for some index $i_0$ all the entries of $X$ in the $i_0^{th}$ row are equal to 1. Then the closed word $w_{i_0}$ is a walk in $G'$, and hence every entry in the $i_0^{th}$ row of $A$ appears there an even number of times and a fortiori at least twice. Now choose $(i_0, n_0) \in I$ such that $A_{i_0,n_0} \in E'$ appears in more than one row of $A$. Let $Y$ be the table obtained by replacing the entry 1 of $X$ in position $(i_0, n_0)$ by the entry 0. Then $Y$ is again an edge-bounding table. Proceeding in this way we can find an edge-bounding table with 0 appearing at least once in every row, and hence we have $|E'| \leq \lfloor \frac{|I|}{2} \rfloor$. Together with (2.1.84) and the definition of $I$, this completes the proof. \hspace{1cm} $\square$

Exercise 2.1.85 (from [AZ05]) Prove that the random vector $\{Z_{N,i}\}_{i=1}^k$ satisfies a multidimensional CLT (as $N \to \infty$).

Remark: see Exercise 2.3.16 for an extension of this result.
2.2 Complex Wigner matrices

In this section we describe the (minor) modifications needed when one considers the law of analogue of Wigner’s theorem for Hermitian matrices. Compared with (2.1.2), we will have complex valued random variables $Z_{i,j}$.

That is, start with two independent families of i.i.d. random variables \( \{Z_{i,j}\}_{1 \leq i < j} \) (complex valued) and \( \{Y_i\}_{1 \leq i} \) (real valued), zero mean, such that \( EZ_{1,2}^2 = 0, E|Z_{1,2}|^2 = 1 \) and, for all integer \( k \geq 1 \),

\[
    r_k := \max \left( E|Z_{1,2}|^k, E|Y_1|^k \right) < \infty. \tag{2.2.1}
\]

Consider the (Hermitian) \( N \times N \) matrix \( X_N \) with entries

\[
    X_N^*(j,i) = X_N(i,j) = \begin{cases} 
        Z_{i,j}/\sqrt{N}, & \text{if } i < j, \\
        Y_i/\sqrt{N}, & \text{if } i = j.
    \end{cases} \tag{2.2.2}
\]

We call such a matrix a Hermitian Wigner matrix, and if the random variables \( Z_{i,j} \) and \( Y_i \) are Gaussian, we use the term Gaussian Hermitian Wigner matrix. The case of Gaussian Hermitian Wigner matrices in which \( EY_i^2 = 1 \) is of particular importance, and for reasons that will become clearer in Chapter ??, such matrices are referred to as GUE (Gaussian Unitary Ensemble) matrices.

Let \( \lambda_i^N \) denote the (real) eigenvalues of \( X_N \), with \( \lambda_1^N \leq \lambda_2^N \leq \ldots \leq \lambda_N^N \), and define the empirical distribution of the eigenvalues as the probability measure on \( \mathbb{R} \) defined by

\[
    L_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i^N}.
\]

The following is the analogue of Theorem 2.1.4.

**Theorem 2.2.3 (Wigner)** For a Hermitian Wigner matrix, the empirical measure \( L_N \) converges weakly, in probability, to the standard semicircle distribution.

As in Section 2.1.2, the proof of Theorem 2.2.3 is a direct consequence of the following two lemmas.

**Lemma 2.2.4** For any \( k \in \mathbb{N} \),

\[
    \lim_{N \to \infty} m_k^N = m_k.
\]

**Lemma 2.2.5** For any \( k \in \mathbb{N} \) and \( \varepsilon > 0 \),

\[
    \lim_{N \to \infty} P \left( \left| \langle L_N, x^k \rangle - \langle \bar{L}_N, x^k \rangle \right| > \varepsilon \right) = 0.
\]
Proof of Lemma 2.2.4 We recall the machinery introduced in Section 2.1.3. Thus, an $N$-word $w = (s_1, \ldots, s_k)$ defines a graph $G_w = (V_w, E_w)$ and a path on the graph. For our purpose, it is convenient to keep track of the direction in which edges are traversed by the path. Thus, given an edge $e = \{s, s'\}$, with $s < s'$, we define $N_e^{w,+}$ as the number of times the edge is traversed from $s$ to $s'$, and we set $N_e^{w,-} = N_e^{w,+}$ as the number of times it is traversed in the reverse direction.

Recalling the equality (2.1.16), we now have instead of (2.1.23) the equation

$$\bar{T}^{N_1} = \frac{1}{N_{k/2}} \prod_{e \in E_{w_1}} E(Z_{1,2}^{N_e^{w,+}} (Z_{1,2}^*)^{N_e^{w,-}}) \prod_{e \in E_{w_1}} E(Y_{1,i}^{N_e^w}).$$

(2.2.6)

In particular, $\bar{T}^{N_1} = 0$ unless $N_e^w \geq 2$ for all $e \in E_1$, and, since $EZ_{1,2}^2 = 0$, if $N_e^w = 2$ then $N_e^{w,+} = 1$.

A slight complication occurs since the function $g_w(N_e^{w,+}, N_e^{w,-}) := E(Z_{1,2}^{N_e^{w,+}} (Z_{1,2}^*)^{N_e^{w,-}})$ is not constant over equivalent classes of words (since changing the letters determining $w$ may switch the role of $N_e^{w,+}$ and $N_e^{w,-}$ in the above expression). Note however that for any $w \in \mathcal{W}_{k,t}$, one has

$$|g_w(N_e^{w,+}, N_e^{w,-})| \leq E(|Z_{1,2}|^{N_e^w}).$$

On the other hand, any $w \in \mathcal{W}_{k,k/2+1}$ satisfies that $G_w$ is a tree, with each edge visited exactly nstarts and ends at the same vertex, one has $N_e^{w,+} = N_e^{w,-} = 1$ for each $e \in E_w$. Thus, repeating the argument in Section 2.1.3, the finiteness of $r_k$ implies that

$$\lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle = 0, \text{ if } k \text{ is odd},$$

while, for $k$ even,

$$\lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle = |\mathcal{W}_{k,k/2+1}| g_w(1,1).$$

(2.2.7)

Since $g_w(1,1) = 1$, the proof is completed by applying (2.1.29). \hfill \Box

Proof of Lemma 2.2.5 The proof is a rerun of the proof of Lemma 2.1.13, using the functions $g_w(N_e^{w,+}, N_e^{w,-})$, defined in the course of proving Lemma 2.2.4. The proof boils down to showing that $\mathcal{W}_{k,k+2}^{(2)}$ is empty, a fact that was established in the course of proving Lemma 2.1.13. \hfill \Box
2.3 Concentration for functional of random matrices and logarithmic Sobolev inequalities

In this short section we digress slightly and prove that certain functionals of random matrices have the concentration property, namely with high probability these functionals are close to their mean value. A more complete treatment of concentration inequalities and their application to random matrices is postponed to Section 4.4.

2.3.1 Smoothness properties of linear functions of the empirical measure

Let us recall that if $X$ is a symmetric (Hermitian) matrix and $f$ is a bounded measurable function, $f(X)$ is defined as the matrix with the same eigenvectors than $X$ but with eigenvalues which are the image by $f$ of those of $X$; namely, if $e$ is an eigenvector of $X$ with eigenvalue $\lambda$, $Xe = \lambda e$, $f(X)e := f(\lambda)e$. In terms of the spectral decomposition $X = UDU^*$ with $U$ orthogonal (unitary) and $D$ diagonal real, one has $f(X) = Uf(D)U^*$ with $f(D)_{ii} = f(D_{ii})$. For $M \in \mathbb{N}$, we denote by $\langle \cdot, \cdot \rangle$ the Euclidean scalar product on $\mathbb{R}^M$ (or $\mathbb{C}^M$), $\langle x, y \rangle = \sum_{i=1}^M x_i y_i$ ($\langle x, y \rangle = \sum_{i=1}^M x_i y_i^*$), and by $\| \cdot \|_2$ the associated norm $\|x\|_2^2 = \langle x, x \rangle$.

General functions of independent random variables need not, in general, satisfy a concentration property. Things are different when the functions involved satisfy certain regularity conditions. It is thus reassuring to see that linear functionals of the empirical measure, viewed as functions of the matrix entries, do possess some regularity properties.

Throughout this section, we denote the Lipschitz constant of a function $G : \mathbb{R}^M \to \mathbb{R}$ by

$$|G|_L := \sup_{x \neq y \in \mathbb{R}^M} \frac{|G(x) - G(y)|}{\|x - y\|_2},$$

and call $G$ a Lipschitz function if $|G|_L < \infty$. The following lemma is an immediate application of Lemma 2.1.43. In its statement, we identify $\mathbb{C}$ with $\mathbb{R}^2$.

**Lemma 2.3.1** Let $g : \mathbb{R}^N \to \mathbb{R}$ be Lipschitz with Lipschitz constant $|g|_L$. Then, with $X$ denoting the Hermitian matrix with entries $X(i, j)$, the map $\{X(i, j)\}_{1 \leq i \leq j \leq N} \mapsto g(\lambda_1(X), \ldots, \lambda_N(X))$ is a Lipschitz function on $\mathbb{R}^{N^2}$ with Lipschitz constant bounded by $\sqrt{2} |g|_L$. In particular, if $f$ is a Lipschitz
function on $\mathbb{R}$, \( \{X(i,j)\}_{1 \leq i \leq j \leq N} \mapsto \text{Tr}(f(X)) \) is a Lipschitz function on $\mathbb{R}^{N(N+1)}$ with Lipschitz constant bounded by $\sqrt{2N} |f|_{C}$.

### 2.3.2 Concentration inequalities for independent variables satisfying logarithmic Sobolev inequalities

We derive in this section concentration inequalities based on the logarithmic Sobolev inequality.

To begin with, recall that a probability measure $P$ on $\mathbb{R}$ is said to satisfy the logarithmic Sobolev inequality (LSI) with constant $c$ if, for any differentiable function $f$, 

$$\int f^2 \log \left( \frac{f^2}{\int f^2 dP} \right) dP \leq 2c \int |f'|^2 dP.$$ 

It is not hard to check, by induction, that if $P_i$ satisfy the (LSI) with constant $c$ and if $P_M = \otimes_{i=1}^{M} P_i$ denotes the product measure on $\mathbb{R}^M$, then $P_M$ satisfies the (LSI) with constant $c$ in the sense that for every differentiable function $F$ on $\mathbb{R}^M$, with $\nabla F$ denoting the gradient of $F$, 

$$\int F^2 \log \left( \frac{F^2}{\int F^2 dP_M} \right) dP_M \leq 2c \int ||\nabla F||^2 dP_M. \tag{2.3.2}$$

This fact and a general study of the LSI may be found in [GZ03] or [Led01].

We note that if the law of a random variable $X$ satisfies the LSI with constant $c$, then for any fixed $\alpha \neq 0$, the law of $\alpha X$ satisfies the LSI with constant $\alpha^2 c$.

Before discussing consequences of the logarithmic Sobolev inequality, we quote from [BL00] a general sufficient condition for it to hold.

**Lemma 2.3.3** Let $V : \mathbb{R}^M \to \mathbb{R} \cup \infty$ satisfy that for some constant $C$, $V(x) - ||x||_2^2/2C$ is convex. Then, the probability measure $\nu(dx) = Z^{-1} e^{-V(x)} dx$ where $Z = \int e^{-V(x)} dx$, satisfies the logarithmic Sobolev inequality with constant $C$. In particular, the standard Gaussian law on $\mathbb{R}^M$ satisfies the logarithmic Sobolev inequality.

We note in passing that on $\mathbb{R}$, a criterion for a measure to satisfy the logarithmic Sobolev inequality was developed by Bobkov and Götze [BG99]. In particular, any probability measure on $\mathbb{R}$ possessing a bounded above and below density with respect to the measures $\nu(dx) = Z^{-1} e^{-|x|^\alpha} dx$ for $\alpha \geq 2$, where $Z = \int e^{-|x|^\alpha} dx$, satisfies the LSI, see [Led01], [GZ03, Property 4.6].
The interest in the logarithmic Sobolev inequality, in the context of concentration inequalities, lies in the following argument, that among other things, shows that LSI implies sub-Gaussian tails.

**Lemma 2.3.4 (Herbst)** Assume that \( P \) satisfies the LSI on \( \mathbb{R}^M \) with constant \( c \). Let \( G \) be a Lipschitz function on \( \mathbb{R}^M \), with Lipschitz constant \( |G|_L \).

Then for all \( \lambda \in \mathbb{R} \),

\[
E_P[e^{\lambda(G-E_P(G))}] \leq e^{c\lambda^2|G|^2_L/2}, \tag{2.3.5}
\]

and so for all \( \delta > 0 \)

\[
P(|G - E_P(G)| \geq \delta) \leq 2e^{-\delta^2/2|G|^2_L}. \tag{2.3.6}
\]

Note that part of the statement in Lemma 2.3.4 is that \( E_PG \) is finite.

**Proof of Lemma 2.3.4** Note first that (2.3.6) follows from (2.3.5). Indeed, by Chebycheff’s inequality, for any \( \lambda > 0 \),

\[
P(|G - E_P(G)| > \delta) \leq e^{-\lambda \delta} E_P[e^{\lambda(G-E_P(G))}] \\
\leq e^{-\lambda \delta} (E_P[e^{\lambda(G-E_P(G))}] + E_P[e^{-\lambda(G-E_P(G))}]) \\
\leq 2e^{-\lambda \delta} e^{|G|^2_L \lambda^2/2}
\]

Optimizing with respect to \( \lambda \) (by taking \( \lambda = \delta/|G|^2_L \)) yields the bound (2.3.6).

Turning to the proof of (2.3.5), let us first assume that \( G \) is a bounded differentiable function such that

\[
|||\nabla G||^2||_\infty := \sup_{x \in \mathbb{R}^M} \sum_{i=1}^M (\partial_{x_i} G(x))^2 < \infty.
\]

Define

\[
X_\lambda = \log E_P e^{2\lambda(G-E_P(G))}.
\]

Then, taking \( F = e^{\lambda(G-E_P(G))} \) in (2.3.2), some algebra reveals that for \( \lambda > 0 \),

\[
\frac{d}{d\lambda} \left( \frac{X_\lambda}{\lambda} \right) \leq 2c|||\nabla G||^2||_\infty.
\]

Now, because \( G - E_P(G) \) is centered,

\[
\lim_{\lambda \to 0^+} \frac{X_\lambda}{\lambda} = 0
\]
and hence integrating with respect to $\lambda$ yields

$$X_\lambda \leq 2c\||\nabla G||^2_\infty\lambda^2,$$

first for $\lambda \geq 0$ and then for any $\lambda \in \mathbb{R}$ by considering the function $-G$ instead of $G$. This completes the proof of (2.3.5) in case $G$ is bounded and differentiable.

Let us now assume only that $G$ is Lipschitz with $|G|_\mathcal{L} < \infty$. For $\epsilon > 0$, define $G_\epsilon = G \wedge (-1/\epsilon) \vee (1/\epsilon)$, and note that $|G_\epsilon|_\mathcal{L} \leq |G|_\mathcal{L} < \infty$. Consider the regularization $G_\epsilon(x) = p_\epsilon * G_\epsilon(x) = \int G_\epsilon(y)p_\epsilon(x-y)dy$ with the Gaussian density $p_\epsilon(x) = e^{-|x|^2/2\epsilon}dx/\sqrt{(2\pi\epsilon)^M}$ such that $p_\epsilon(x)dx$ converges weakly towards the atomic measure $\delta_0$ as $\epsilon$ converges to 0. Since for any $x \in \mathbb{R}^M$,

$$|G_\epsilon(x) - G_\epsilon(x)| \leq |G|_\mathcal{L} \int ||y||_2 p_\epsilon(y)dy = M|G|_\mathcal{L} \sqrt{\epsilon},$$

$G_\epsilon$ converges pointwise towards $G$. Moreover, $G_\epsilon$ is Lipschitz, with Lipschitz constant bounded by $|G|_\mathcal{L}$ independently of $\epsilon$. $G_\epsilon$ is also continuously differentiable and

$$\|||\nabla G_\epsilon||^2_\infty = \sup_{x \in \mathbb{R}^M} \sup_{u \in \mathbb{R}^M} \{2\langle \nabla G_\epsilon(x), u \rangle - ||u||^2_2\}$$

$$\leq \sup_{u,x \in \mathbb{R}^M} \sup_{\delta > 0} \{2\delta^{-1}(G_\epsilon(x + \delta u) - G_\epsilon(x)) - ||u||^2_2\}$$

$$\leq \sup_{u \in \mathbb{R}^M} \{2|G|_\mathcal{L}||u||_2 - ||u||^2_2\} = |G|^2_\mathcal{L}. \quad (2.3.7)$$

Thus, we can apply the previous result to find that for any $\epsilon > 0$ and all $\lambda \in \mathbb{R}$

$$E_P[e^{\lambda G_\epsilon}] \leq e^{\lambda E_P G_\epsilon} e^{\lambda^2|G|^2_\mathcal{L}/2}. \quad (2.3.8)$$

Therefore, by Fatou’s lemma,

$$E_P[e^{\lambda G}] \leq e^{\liminf_{\epsilon \to 0} \lambda E_P G_\epsilon} e^{\lambda^2|G|^2_\mathcal{L}/2}. \quad (2.3.9)$$

We next show that $\lim_{\epsilon \to 0} E_P G_\epsilon = E_P G$, which, in conjunction with (2.3.7), will conclude the proof. Indeed, (2.3.8) implies that

$$P(|G_\epsilon - E_P G_\epsilon| > \delta) \leq 2e^{-\delta^2/2|G|^2_\mathcal{L}}. \quad (2.3.10)$$

Consequently,

$$E[(G_\epsilon - E_P G_\epsilon)^2] = 2 \int_0^\infty xP(|G_\epsilon - E_P G_\epsilon| > x)dx \leq 4 \int_0^\infty xe^{-\frac{x^2}{2|G|^2_\mathcal{L}}}dx = 4c|G|^2_\mathcal{L}. \quad (2.3.11)$$
so that the sequence \((G_\epsilon - EP G_\epsilon)_\epsilon \geq 0\) is uniformly integrable. Now, \(G_\epsilon\) converges pointwise towards \(G\) and therefore there exists a constant \(K\), independent of \(\epsilon\), such that for \(\epsilon < \epsilon_0\), \(P(|G_\epsilon| \leq K) \geq \frac{3}{4}\). On the other hand, (2.3.10) implies that \(P(|G_\epsilon - EP G_\epsilon| \leq r) \geq \frac{3}{4}\) for some \(r\) independent of \(\epsilon\). Thus,

\[\{ |G_\epsilon - EP G_\epsilon| \leq r \} \cap \{ |G_\epsilon| \leq K \} \subset \{ |EP G_\epsilon| \leq K + r \}\]

is not empty, providing a uniform bound on \((EP G_\epsilon)_\epsilon < \epsilon_0\). We thus deduce from (2.3.11) that \(\sup_{\epsilon < \epsilon_0} EP G_\epsilon^2\) is finite, and hence \((G_\epsilon)_\epsilon < \epsilon_0\) is uniformly integrable. In particular,

\[
\lim_{\epsilon \to 0} EP G_\epsilon = EP G < \infty
\]

which finishes the proof. \(\Box\)

2.3.3 Concentration for Wigner-type matrices

The concentration results, an immediate corollary of Lemma 2.3.1 and (2.3.6), are obtained for symmetric matrices with independent entries on or above the diagonal. Thus, we consider in this section (symmetric) matrices \(X_N\) with independent (and not necessarily identically distributed) entries \(\{X_N(i,j)\}_{1 \leq i \leq j \leq N}\).

**Theorem 2.3.12** Suppose that the laws of the independent entries \(\{X_N(i,j)\}_{1 \leq i \leq j \leq N}\) all satisfy the (LSI) with constant \(c/N\). Then, for any Lipschitz function \(f\) on \(\mathbb{R}\), for any \(\delta > 0\),

\[
P\left( |\text{Tr}(f(X_N)) - E[\text{Tr}(f(X_N))]| \geq \delta N \right) \leq 2e^{-\frac{\delta^2}{4N^2}}. \tag{2.3.13}
\]

Further, for any \(k \in \{1, \ldots, N\}\),

\[
P\left( |f(\lambda_k(X_N)) - Ef(\lambda_k(X_N))| \geq \delta \right) \leq 2e^{-\frac{\delta^2}{4N^2}}. \tag{2.3.14}
\]

We note that under the assumptions of Theorem 2.3.12, \(E\lambda_1(X_N)\) is uniformly bounded, see Exercise 2.1.61 or Exercise 2.1.62.

**Proof of Theorem 2.3.12** To see (2.3.13), take \(G(X_N(i,j), 1 \leq i \leq j \leq N) = \text{Tr}(f(X_N))\). By Lemma 2.3.1, we see that if \(f\) is Lipschitz, \(G\) is also Lipschitz with constant bounded by \(\sqrt{2N||f||_L}\) and hence Lemma 2.3.4 with \(M = N(N + 1)/2\) allows one to conclude. To see (2.3.14), apply the same argument to the function \(G(X_N(i,j), 1 \leq i \leq j \leq N) = f(\lambda_k(X_N))\). \(\Box\)
Remark 2.3.15 The assumption of Theorem 2.3.12 are satisfied for Gaussian matrices with independent on or above the diagonal entries whose variance is bounded by $c/N$. In particular, the assumptions hold for Gaussian Wigner matrices. We emphasize that Theorem 2.3.12 applies also when the variance of $X_N(i,j)$ depends on $i,j$, e.g. when $X_N(i,j) = a_N(i,j)Y_N(i,j)$ with $Y_N(i,j)$ i.i.d with law $P$ satisfying the log-Sobolev inequality and $a(i,j)$ uniformly bounded (since if $P$ satisfies the log-Sobolev inequality with constant $c$, the law of $ax$ under $P$ satisfies it also with a constant bounded by $a^2c$).

Exercise 2.3.16 [From [AZ05]] Using Exercise 2.1.85, prove that if $X_N$ is a Gaussian Wigner matrix and $f : \mathbb{R} \to \mathbb{R}$ is a $C^1_b$ function, then $N[\langle f, L_N \rangle - \langle f, \bar{L}_N \rangle]$ satisfies a CLT.

2.4 Stieltjes transforms and recursions

We begin by recalling some classical results concerning the Stieltjes transform of a probability measure.

Definition 2.4.1 Let $\mu$ be a probability measure on the real line. The Stieltjes transform of $\mu$ is the function

$$S_\mu(z) := \int_\mathbb{R} \frac{\mu(dx)}{x-z}, \ z \in \mathbb{C} \setminus \mathbb{R}. $$

Note that for $z \in \mathbb{C} \setminus \mathbb{R}$, both the real and imaginary parts of $1/(x-z)$ are continuous bounded functions of $x \in \mathbb{R}$. This crucial observation is used repeatedly in what follows.

Remark 2.4.2 It is interesting to note that $\hat{\beta}(z)$, see (2.1.8), is closely related to the Stieltjes transform of the semi-circle distribution $\sigma$: that is, note that for $|z| < 1/4$,

$$\hat{\beta}(z) = \sum_{k=0}^\infty z^k \int x^{2k} \sigma(x) dx = \int \left( \sum_{k=0}^\infty (zx^2)^k \right) \sigma(x) dx$$

$$= \int \frac{1}{1-zx^2} \sigma(x) dx$$

$$= \int \frac{1}{1-\sqrt{z}x} \sigma(x) dx = \frac{-1}{\sqrt{z}} S_\sigma(1/\sqrt{z}).$$

where the third equality uses that the support of $\sigma$ is the interval $[-2,2]$, and the fourth uses the symmetry of $\sigma$. 

Stieltjes transforms can be inverted. In particular, one has

**Theorem 2.4.3** For any open interval \( I \) with endpoints not on an atom of \( \mu \),

\[
\mu(I) = \lim_{\epsilon \to 0} \frac{1}{\pi} \int_I \frac{S_\mu(\lambda + i\epsilon) - S_\mu(\lambda - i\epsilon)}{2i} d\lambda
\]

\[
= \lim_{\epsilon \to 0} \frac{1}{\pi} \int_I \Im S_\mu(\lambda + i\epsilon) d\lambda. \tag{2.4.4}
\]

**Proof:** Let \( X \) be distributed according to \( \mu \), and denote by \( C_\epsilon \) a Cauchy variable of parameter \( \epsilon \), i.e. a random variable of density

\[
Ca_\epsilon(dx) = \frac{\epsilon dx}{\pi(x^2 + \epsilon^2)}, \tag{2.4.5}
\]

independent of \( X \). Then, \( \Im S_\mu(\lambda + i\epsilon)/\pi \) is nothing but the density (with respect to Lebesgue measure) of the law of \( X + C_\epsilon \) evaluated at \( \lambda \in \mathbb{R} \). The convergence in (2.4.4) is then just a rewriting of the weak convergence of the law of \( X + C_\epsilon \) to that of \( X \), as \( \epsilon \to 0 \). \( \square \)

Theorem 2.4.3 allows for the reconstruction of a measure from its Stieltjes transform. Further, one has the following

**Theorem 2.4.6** Let \( \mu_n \) denote a sequence of probability measures.

a) If \( \mu_n \) converges weakly to a probability measure \( \mu \) then \( S_{\mu_n}(z) \) converges to \( S_{\mu}(z) \) for each \( z \in \mathbb{C} \setminus \mathbb{R} \).

b) If \( S_{\mu_n}(z) \) converges for each \( z \in \mathbb{C} \setminus \mathbb{R} \) to a limit \( S(z) \), then \( S(z) \) is the Stieltjes transform of a sub-probability measure \( \mu \), and \( \mu_n \) converges vaguely to \( \mu \).

**Proof:** Part a) is a restatement of the notion of weak convergence. To see part b), let \( \{X_n\}_{n \geq 1} \) denote a sequence of random variables, each distributed according to \( \mu_n \), and denote by \( C_\epsilon \) a Cauchy random variable of parameter \( \epsilon \) and law \( Ca_\epsilon \) as in (2.4.5), independent of \( \{X_n\}_{n \geq 1} \). Then, \( \Im S_{\mu_n}(\lambda + i\epsilon)/\pi \) is nothing but the density \( p_\epsilon^\mu(\lambda) \) (with respect to Lebesgue measure) of the law of \( X_n + C_\epsilon \) evaluated at \( \lambda \in \mathbb{R} \), and the convergence of the Stieltjes transforms thus implies the pointwise convergence of the density \( p_\epsilon^\mu(\cdot) \) to a (bounded) function \( p_\infty^\mu(\cdot) \). With the sequence \( p_\epsilon^\mu \) being uniformly bounded, this and Fatou's theorem imply the convergence of the distribution function \( F_n^\epsilon \) of \( X_n + C_\epsilon \) to an increasing, continuous function \( F_\infty^\epsilon \), with

\[
\lim_{y \to -\infty} F_n^\epsilon(y) \geq 0, \quad \lim_{y \to \infty} F_n^\epsilon(y) \leq 1. \tag{2.4.7}
\]

Since \( C_\epsilon \to \epsilon \to 0 \) in probability, one concludes that \( F_n \) converges pointwise to a limit \( F_\infty \) at all continuity points of the increasing function \( F_\infty \), that still
satisfies the relations (2.4.7). This completes the proof of vague convergence of $\mu_n$ to a sub-probability measure $\mu$. To see that $S(z)$ is the Stieltjes transform of $\mu$, note that the function $x \mapsto 1/(z - x)$ is, for each $z \in \mathbb{C} \setminus \mathbb{R}$, bounded and continuous, and decays to 0 as $x \to \infty$. \hfill \Box

For a matrix $X$, define $S_X(z) := (X - zI)^{-1}$. Taking $A = X$ in the Matrix inversion lemma A.2, one gets

$$S_X(z) = z^{-1}(XS_X(z) - I), \quad z \in \mathbb{C} \setminus \mathbb{R}. \quad (2.4.8)$$

Note that with $L_N$ denoting the empirical measure of the eigenvalues of $X_N$,

$$S_{L_N}(z) = \frac{1}{N} \text{Tr}S_N(z), \quad S_{\mathbb{E}N}(z) = \frac{1}{N} E\text{Tr}S_N(z).$$

### 2.4.1 Gaussian Wigner matrices

We consider in this section the case of $X_N$ being a Wigner Gaussian matrix, providing a

**Proof #2 of Theorem 2.1.4:** ($X_N$ a Gaussian Wigner matrix).

Recall first the following identity, that characterizes the Gaussian distribution, and that is proved by integration by parts:

**Lemma 2.4.9** If $\zeta$ is a zero mean Gaussian random variable, then for $f$ differentiable, with polynomial growth,

$$E(\zeta f(\zeta)) = E(f'(\zeta))E(\zeta^2).$$

Define next the matrix $\Delta^{i,k}_N$ as the symmetric $N \times N$ matrix satisfying

$$\Delta^{i,k}_N(j, l) = \begin{cases} 1, & (i, k) = (j, l) \text{ or } (i, k) = (l, j), \\ 0, & \text{otherwise}. \end{cases}$$

Then, with $X$ an $N \times N$ symmetric matrix,

$$\frac{\partial}{\partial X(i, k)} S_X(z) = -S_X(z) \Delta^{i,k}_N S_X(z). \quad (2.4.10)$$
Using now (2.4.8) in the first equality and Lemma 2.4.9 and (2.4.10) (conditioning on all entries of $X_N$ but one) in the second, one concludes that

$$
\frac{1}{N} \text{Tr}S_{X_N}(z) = -\frac{1}{z} + \frac{1}{z} E(\text{Tr}S_{X_N}S_{X_N}(z))
$$

(2.4.11)

$$
= -\frac{1}{z} - \frac{1}{zN^2} E \left( \sum_{i,k} [S_{X_N}(z)(i,i)S_{X_N}(z)(k,k) + S_{X_N}(z)(i,k)^2] \right)
\quad - \frac{3}{zN^2} \sum_i \left( (EX_N(i,i)^2 - 1) ES_{X_N}(z)(i,i)^2 \right)

= -\frac{1}{z} - \frac{1}{z} E[(L_N, (x-z)^{-1})^2] - \frac{1}{zN}(\hat{L}_N, (x-z)^{-2})
\quad - \frac{3}{zN^2} \sum_i \left( (EX_N(i,i)^2 - 1) ES_{X_N}(z)(i,i)^2 \right).
$$

Since $(x-z)^{-1}$ is a Lipschitz function for any fixed $z \in \mathbb{C} \setminus \mathbb{R}$, it follows from Theorem 2.3.12 and Remark 2.3.15 that

$$
|E[(L_N, (x-z)^{-1})^2] - (\hat{L}_N, (x-z)^{-1})^2| \to_{N \to \infty} 0.
$$

This, and the boundedness of $1/(z-x)^2$ for a fixed $z$ as above, imply the existence of a sequence $\epsilon_N(z) \to_{N \to \infty} 0$ such that, letting $\bar{S}_{N}(z) := N^{-1}\text{Tr}S_{X_N}(z)$, one has

$$
\bar{S}_{N}(z) = -\frac{1}{z} - \frac{1}{z} \bar{S}_{N}(z)^2 + \epsilon_N(z).
$$

Thus any limit point $s(z)$ of $\bar{S}_{N}(z)$ satisfies

$$
\bar{s}(z)(z + s(z)) + 1 = 0.
$$

(2.4.12)

Further, let $\mathbb{C}_+ = \{ z \in \mathbb{C} : \text{Im}z > 0 \}$. Then, for $z \in \mathbb{C}_+$, by its definition, $s(z)$ must have a non-negative imaginary part, while for $z \in \mathbb{C} \setminus (\mathbb{R} \cup \mathbb{C}_+)$, $s(z)$ must have a non-positive imaginary part. Hence, for all $z \in \mathbb{C}$, with the choice of the branch of the square-root dictated by the last remark,

$$
s(z) = -\frac{1}{2} \left[ z - \sqrt{z^2 - 4} \right].
$$

(2.4.13)

Applying again Theorem 2.3.12 and Remark 2.3.15, it follows that $S_{L_N}(z)$ converges in probability to $s(z)$, solution of (2.4.13), for all $z \in \mathbb{C} \setminus \mathbb{R}$. Comparing with (2.1.8) and using Remark 2.4.2, one deduces that $s(z)$ is the Stieltjes transform of the semi-circle law $\sigma$, since $s(z)$ coincides with the latter for $|z| < 1/4$ and hence for all $z \in \mathbb{C} \setminus \mathbb{R}$ by analyticity. $\square$
2.4.2 General Wigner matrices

We consider in this section the case of $X_N$ being a Wigner matrix. We give now a:

**Proof #3 of Theorem 2.1.4**: ($X_N$ a Wigner matrix).

We begin again by a general fact valid for arbitrary symmetric matrices.

**Lemma 2.4.14** Let $W$ be an $N \times N$ symmetric matrix, and let $w_i$ denote the $i$-th column of $W$ with the entry $W(i,i)$ removed (i.e., $w_i$ is an $N-1$-dimensional vector). Let $W(i)$ denote the $(N-1) \times (N-1)$ dimensional matrix obtained by erasing the $i$-th column and row from $W$. Then, for every $z \in \mathbb{C} \setminus \mathbb{R}$,

$$W - zI)^{-1}(i,i) = \frac{1}{W(i,i) - z - w_i^T(W(i) - zI_{N-1})^{-1}w_i}.$$  \hspace{1cm} (2.4.15)

**Proof of Lemma 2.4.14**: Note first that from Cramer’s rule,

$$(W - zI)^{-1}(i,i) = \frac{\det(W(i) - zI_{N-1})}{\det(W - zI)}.$$ \hspace{1cm} (2.4.16)

Write next

$$W - zI = \begin{pmatrix} W^{(N)} - zI_{N-1} & w_N \\ w_N^T & W(N,N) - z \end{pmatrix},$$

and use the matrix identity (A.1) with $A = W^{(N)} - zI_{N-1}$, $B = w_N$, $C = w_N^T$ and $D = W(N,N) - z$ to conclude that

$$\det(W - zI) = \det(W^{(N)} - zI_{N-1}) \det \left[ W(N,N) - z - w_N^T(W^{(N)} - zI_{N-1})^{-1}w_N \right].$$

The last formula holds in the same manner with $W(i)$, $w_i$ and $W(i,i)$ replacing $W^{(N)}$, $w_N$ and $W(N,N)$ respectively. Substituting in (2.4.16) completes the proof of Lemma 2.4.14. \hfill \Box

We are now ready to return to the proof of Theorem 2.1.4. Note first that by the truncation argument used in the proof of Theorem 2.1.46, we may assume that $X_N(i,i) = 0$ for all $i$ and that for some constant $C$ independent of $N$, it holds that $|\sqrt{N}X_N(i,j)| \leq C$ for all $i,j$. Define $\bar{\alpha}_k(i) = X_N(i,k)$, i.e. $\bar{\alpha}_k$ is the $k$-th column of the matrix $X_N$. Let $\alpha_k$ denote the $N-1$ dimensional vector obtained from $\bar{\alpha}_k$ by erasing the entry $\alpha_k(k) = 0$. Denote by $X_N^{(k)}$ the $(N-1) \times (N-1)$ dimensional matrix
consisting of the matrix $X_N$ with the $k$-th row and column removed. By Lemma 2.4.14, one gets that

$$\frac{1}{N} \text{Tr} S_{X_N}(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{-z - \alpha_i^T (X_N^{(i)} - z I_{N-1})^{-1} \alpha_i}$$

$$= \frac{1}{z + N^{-1} \text{Tr} S_{X_N}(z)} - \delta_N(z) ,$$

(2.4.17)

where

$$\delta_N(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{\epsilon_{i,N}}{-z - N^{-1} \text{Tr} S_{X_N}(z) + \epsilon_{i,N}}(-z - N^{-1} \text{Tr} S_{X_N}(z)) ,$$

(2.4.18)

and

$$\epsilon_{i,N} = N^{-1} \text{Tr} S_{X_N}(z) - \alpha_i^T (X_N^{(i)} - z I_{N-1})^{-1} \alpha_i .$$

(2.4.19)

Our next goal is to prove the convergence in probability of $\delta_N(z)$ to zero for each fixed $z \in \mathbb{C} \setminus \mathbb{R}$ with $|Im z| = \delta_0 > 0$. Toward this end, note that the denominator of each term in the right hand side of (2.4.18) has modulus at least $\delta_0^2$, since $|Im z| = \delta_0$ and all eigenvalues of $X_N$ and $X_N^{(i)}$ are real. Thus, it suffices to prove the convergence of $\sup_{i \leq N} |\epsilon_{i,N}|$ to zero in probability. Toward this end, let $\bar{X}_N^{(i)}$ denote the matrix $X_N$ with the $i$-th column and row set to zero. Then, the eigenvalues of $\bar{X}_N^{(i)}$ and $X_N^{(i)}$ coincide except that $\bar{X}_N^{(i)}$ has one more zero eigenvalue. Hence,

$$\frac{1}{N} |\text{Tr} S_{\bar{X}_N^{(i)}}(z) - \text{Tr} S_{X_N^{(i)}}(z)| \leq \frac{1}{\delta_0 N} ,$$

whereas, with the eigenvalues of $\lambda_N^{(i)}$ denoted $\lambda_1^{(i)} \leq \lambda_2^{(i)} \leq \ldots \leq \lambda_N^{(i)}$, and those of $X_N$ denoted $\lambda_1^{N} \leq \lambda_2^{N} \leq \ldots \leq \lambda_N^{N}$, one has

$$\frac{1}{N} |\text{Tr} S_{\bar{X}_N^{(i)}}(x) - \text{Tr} S_{X_N}(x)| \leq \frac{1}{\delta_0^2 N} \sum_{k=1}^{N} |\lambda_k^{(i)} - \lambda_k^{N}|$$

$$\leq \frac{1}{\delta_0^2} \left( \frac{1}{N} \sum_{k=1}^{N} |\lambda_k^{(i)} - \lambda_k^{N}|^2 \right)^{1/2}$$

$$\leq \frac{1}{\delta_0^2} \left( \frac{2}{N} \sum_{k=1}^{N} X_N(i,k)^2 \right)^{1/2} ,$$

where Lemma 2.1.43 was used in the last inequality. Since $|\sqrt{N}X_N(i,j)| \leq C$, we get that $\sup_i |S_{\bar{X}_N^{(i)}}(z) - S_{X_N}(z)|$ converges to zero (deterministically). Combining the above it follows that to prove the convergence of
\[ \sup_{i \leq N} |\varepsilon_{i,N}| \] to zero in probability, it is enough to prove the convergence to 0 in probability of \( \sup_{i \leq N} |\bar{\varepsilon}_{i,N}| \), where

\[ \bar{\varepsilon}_{i,N} = \frac{1}{N} \text{Tr} B_N^{(i)}(z) - \alpha_i^T B_N^{(i)}(z) \alpha_i \]

(2.4.20)

\[ = \frac{1}{N} \sum_{k=1}^{N} \left( \left( \sqrt{N} \alpha_i(k) \right)^2 - 1 \right) B_N^{(i)}(z)(k,k) + \sum_{k,k'=1, k \neq k'}^{N} \alpha_i(k) \alpha_i(k') B_N^{(i)}(z)(k,k') \]

=: \bar{\varepsilon}_{i,N}(1) + \bar{\varepsilon}_{i,N}(2),

where \( B_N^{(i)}(z) = (X_N^{(i)} - zI_{N-1})^{-1} \). Noting that \( \alpha_i \) is independent of \( B_N^{(i)}(z) \), and possesses zero mean independent entries of variance \( 1/N \), one observe by conditioning on the sigma-field \( \mathcal{F}_{i,N} \) generated by \( X_N^{(i)} \) that \( E \bar{\varepsilon}_{i,N} = 0 \).

Further, since

\[ N^{-1} \text{Tr} \left( B_N^{(i)}(z)^2 \right) \leq \frac{1}{\delta_0}, \]

and the random variables \( |\sqrt{N} \alpha_i(k)| \) are uniformly bounded, it follows that

\[ E|\bar{\varepsilon}_{i,N}(1)|^4 \leq \frac{c_1}{N^2}; \]

for some constant \( c_1 \) that depends only on \( \delta_0 \) and \( C \). Similarly, one checks that

\[ E|\bar{\varepsilon}_{i,N}(2)|^4 \leq \frac{c_2}{N^2}; \]

for some constant \( c_2 \) depending only on \( C, \delta_0 \). One concludes then, by Chebycheff’s inequality, the claimed convergence of \( \sup_{i \leq N} |\varepsilon_{i,N}(z)| \) to 0 in probability.

The rest of the argument is similar to what has already been done in Section 2.4.1, and is omitted. \( \square \)

**Remark 2.4.21** We note that reconstruction and continuity results that are stronger than those contained in Theorems 2.4.3 and 2.4.6 are available. An accessible introduction to these and their use in RMT can be found in [Bai99]. For example, in Theorem 2.4.3, if \( \mu \) possesses a Hölder continuous density \( m \) then for \( \lambda \in \mathbb{R} \),

\[ S_\mu(\lambda + i0) := \lim_{\epsilon \downarrow 0} S_\mu(\lambda + \epsilon) = i\pi m(\lambda) + \text{P.V.} \int_{\mathbb{R}} \frac{\mu(dx)}{x - \lambda} \]

(2.4.22)

exists, where the notation P.V. stands for “principal value”. Also, in the context of Theorem 2.4.6, if the \( \mu \) and \( \nu \) are probability measures supported on \([-B, B] \), \( a, \gamma \) are constants satisfying

\[ \gamma := \frac{1}{\pi} \int_{|u| \leq a} \frac{1}{u^2 + 1} \, du > \frac{1}{2}; \]


and $A$ is a constant satisfying
\[ \kappa := \frac{4B}{\pi (A - B)(2\gamma - 1)} \in (0, 1), \]
then for any $\nu > 0$,
\[
\pi (1 - \kappa)(2\gamma - 1) \sup_{|x| \leq B} |\mu([-B, x]) - \nu([-B, x])| \leq \frac{1}{\nu} \sup_{x} \int_{|y| \leq 2\nu} |\mu([-B, x+y]) - \mu([-B, x])|dy.
\]

In the context of random matrices, equation (2.4.23) is useful in obtaining rate of convergence in the convergence of $L_N$ to its limit, but we will not discuss here at all this issue.

2.5 Joint distribution of eigenvalues in the GOE and the GUE

We are going to calculate the joint distribution of eigenvalues of a random symmetric or Hermitian matrix under a special type of probability law which displays a high degree of symmetry but still makes on-or-above-diagonal entries independent so that the theory of Wigner matrices applies.

2.5.1 Definition and preliminary discussion of the GOE and the GUE

Let $\{\xi_{i,j}, \eta_{i,j}\}_{i,j=1}^{\infty}$ be an i.i.d. family of real mean 0 variance 1 Gaussian random variables. We define
\[
P_{2}^{(1)}, P_{3}^{(1)}, \ldots
\]
to be the laws of the random matrices
\[
\begin{bmatrix}
\sqrt{2}\xi_{1,1} & \xi_{1,2} \\
\xi_{1,2} & \sqrt{2}\xi_{2,2}
\end{bmatrix} \in \mathcal{H}_{2}^{(1)},
\begin{bmatrix}
\sqrt{2}\xi_{1,1} & \xi_{1,2} & \xi_{1,3} \\
\xi_{1,2} & \sqrt{2}\xi_{2,2} & \xi_{2,3} \\
\xi_{1,3} & \xi_{2,3} & \sqrt{2}\xi_{3,3}
\end{bmatrix} \in \mathcal{H}_{3}^{(1)}, \ldots,
\]
respectively. We define
\[
P_{2}^{(2)}, P_{3}^{(2)}, \ldots
\]
to be the laws of the random matrices

\[
\begin{pmatrix}
\xi_{1,1} & \frac{\xi_{1,2} + i\eta_{1,2}}{\sqrt{2}} \\
\frac{\xi_{1,2} - i\eta_{1,2}}{\sqrt{2}} & \xi_{2,2}
\end{pmatrix} \in \mathcal{H}_2^{(2)}, \quad \begin{pmatrix}
\xi_{11} & \frac{\xi_{12} + i\eta_{12}}{\sqrt{2}} & \frac{\xi_{13} + i\eta_{13}}{\sqrt{2}} \\
\frac{\xi_{12} - i\eta_{12}}{\sqrt{2}} & \xi_{22} & \frac{\xi_{23} + i\eta_{23}}{\sqrt{2}} \\
\frac{\xi_{13} - i\eta_{13}}{\sqrt{2}} & \frac{\xi_{23} - i\eta_{23}}{\sqrt{2}} & \xi_{33}
\end{pmatrix} \in \mathcal{H}_3^{(2)}, \ldots,
\]

respectively. A random matrix \( X \in \mathcal{H}_N^{(\beta)} \) with law \( P_N^{(\beta)} \) is said to belong to the Gaussian orthogonal ensemble (GOE) or the Gaussian unitary ensemble (GUE) according as \( \beta = 1 \) or \( \beta = 2 \), respectively. The theory of Wigner matrices developed in previous sections of this book applies here. In particular, for fixed \( \beta \), given for each \( N \) a random matrix \( X(N) \in \mathcal{H}_N^{(\beta)} \) with law \( P_N^{(\beta)} \), the empirical distribution of the eigenvalues of \( X_N := X(N)/\sqrt{N} \) tends to the semicircle law of mean 0 and variance 1.

So what’s special about the law \( P_N^{(\beta)} \) within the class of laws of Wigner matrices? The law \( P_N^{(\beta)} \) is highly symmetrical. To explain the symmetry, as well as to explain the presence of the terms “orthogonal” and “unitary” in our terminology, let us calculate the density of \( P_N^{(\beta)} \) with respect to Lebesgue measure \( \ell_N^{(\beta)} \) on \( \mathcal{H}_N^{(\beta)} \). To fix \( \ell_N^{(\beta)} \) unambiguously (rather than just up to a positive constant factor) we use the following procedure. In the case \( \beta = 1 \), consider the one-to-one onto mapping \( \mathcal{H}_N^{(1)} \to \mathbb{R}^{N(N+1)/2} \) defined by taking on-or-above-diagonal entries as coordinates, and normalize \( \ell_N^{(1)} \) by requiring it to push forward to Lebesgue measure on \( \mathbb{R}^{N(N+1)/2} \). Similarly, in the case \( \beta = 2 \), consider the one-to-one onto mapping \( \mathcal{H}_N^{(2)} \to \mathbb{R}^N \times \mathbb{C}^{N(N-1)/2} = \mathbb{R}^{N^2} \) defined by taking on-or-above-diagonal entries as coordinates, and normalize \( \ell_N^{(2)} \) by requiring it to push forward to Lebesgue measure on \( \mathbb{R}^{N^2} \). Let \( H_{i,j} \) denote the entry of \( H \in \mathcal{H}_N^{(\beta)} \) in row \( i \) and column \( j \). Note that

\[
\text{Tr} H^2 = \text{Tr} HH^* = \sum_{i=1}^N H_{i,i}^2 + 2 \sum_{1 \leq i < j \leq N} |H_{i,j}|^2.
\]

It is a straightforward matter now to verify that

\[
\frac{dP_N^{(\beta)}}{d\ell_N^{(\beta)}}(H) = \begin{cases}
2^{-N/2}(2\pi)^{-N(N+1)/4} \exp(-\text{Tr} H^2/4) & \text{if } \beta = 1, \\
2^{-N/2} \pi^{-N^2/2} \exp(-\text{Tr} H^2/2) & \text{if } \beta = 2.
\end{cases}
\]

The latter formula clarifies the symmetry of \( P_N^{(\beta)} \). The main thing to notice is that the density at \( H \) depends only on the eigenvalues of \( H \). It follows that if \( X \) is a random element of \( \mathcal{H}_N^{(1)} \) with law \( P_N^{(1)} \), then for any \( N \) by
$N$ orthogonal matrix $U$, again $UXU^*$ has law $P_N^{(1)}$; and similarly, if $X$ is a random element of $\mathcal{H}_N^{(2)}$ with law $P_N^{(2)}$, then for any $N$ by $N$ unitary matrix $U$, again $UXU^*$ has law $P_N^{(2)}$. As we already observed (see Lemma 1.1.1), for random $X \in \mathcal{H}_N^{(1)}$ it makes sense to talk about the joint distribution of the eigenvalues $\lambda_1(X) \leq \cdots \leq \lambda_N(X)$. The main result in this section is the following.

**Theorem 2.5.2 (Joint distribution of eigenvalues: GOE and GUE)**

Let $X \in \mathcal{H}_N^{(1)}$ be random with law $P_N^{(1)}$, $\beta = 1, 2$. The joint distribution of the eigenvalues $\lambda_1(X) \leq \cdots \leq \lambda_N(X)$, has density with respect to Lebesgue measure which equals

$$N! C_N^{(1)} \prod_{1 \leq i < j \leq N} |x_i - x_j|^\beta \prod_{i=1}^N e^{-\beta x_i^2/4},$$

where

$$N! C_N^{(1)} = \frac{\Gamma(\beta/2)}{\Gamma(N/2) \Gamma(\beta/2)} \prod_{j=1}^N \Gamma(\beta/2) / \Gamma(j\beta/2).$$

Here, for any positive real $s$,

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx$$

is Euler’s Gamma function.

**Remark 2.5.6** We refer to the probability measure $P_N^{(1)}$ on $\mathbb{R}^N$ with density with respect to the Lebesgue measure $\text{Leb}_N$ on $\mathbb{R}^N$

$$\frac{dP_N^{(1)}}{d\text{Leb}_N} = \prod_{1 \leq i < j \leq N} |x_i - x_j|^\beta \prod_{i=1}^N e^{-\beta x_i^2/4},$$

where $C_N^{(1)}$ is given in (2.5.7), as the law of the unordered eigenvalues of the GOE/GUE.

### 2.5.2 Proof of the joint distribution of eigenvalues

We present in this section a proof of Theorem 2.5.2 that has the advantage of being direct, elementary, and not requiring much in terms of computations. On the other hand, this proof is not enough to provide one with
the evaluation of the normalization constant $\bar{C}_N^\beta$ in (2.5.4). The evaluation of the latter is postponed to subsection 2.5.3, where the Selberg integral formula is derived. Another approach to evaluating the normalization constants, in the case of the GUE, is provided in Section 3.2.2.

The idea behind the proof of 2.5.2 is as follows. Since $X \in H_N^\beta$, there exist a decomposition $X = UDU^*$, with where the eigenvalues matrix $D \in D_N$ denotes diagonal matrices with real entries, and with the eigenvectors matrix $U \in U_N^\beta$, where $U_N^\beta$ denotes the collection of orthogonal matrices (when $\beta = 1$) or unitary matrices (when $\beta = 2$). Suppose this map was a bijection (which it is not, at least at the matrices $X$ that do not possess all distinct eigenvalues) and that one could parametrize $U_N^\beta$ using $\beta N(N-1)/2$ parameters in a smooth way (which one cannot). An easy computation shows that the Jacobian of the transformation would then be a polynomial in the eigenvalues with coefficients that are functions of the parametrization of $U_N^\beta$, of degree $\beta N(N-1)/2$. Since the bijection must break down when $D_{ii} = D_{jj}$ for some $i \neq j$, the Jacobian must vanish on that set; Symmetry and degree considerations then show that the Jacobian must be proportional to the factor $\prod_{1 \leq i < j \leq n} |x_i - x_j|^\beta$. Integrating over the parametrization of $U_N^\beta$ then yields (2.5.3).

Unfortunately, in order to make the above construction work, we need to throw away subsets of $H_N^\beta$, that fortunately turn out to have zero Lebesgue measure. Toward this end, we say that $U \in U_N^\beta$ is normalized if every diagonal entry of $U$ is strictly positive real. We say that $U \in U_N^\beta$ is of general type if it is normalized and every entry of $U$ is nonzero. The collection of general type matrices is denoted $U_N^{\beta,g}$. We also say that $D \in D_N$ is distinct if its entries are all distinct, and denote by $D_N^d$ the collection of distinct matrices, and by $D_N^{d_0}$ the subset of matrices with decreasing entries, that is $D_N^{d_0} = \{ D \in D_N : D_{i,i} > D_{i+1,i+1} \}$.

Let $H_N^{\beta,dg}$ denote the subset of $H_N^{\beta}$ consisting of those matrices that possess a decomposition $X = UDU^*$ where $D \in D_N^d$ and $U \in U_N^{\beta,g}$. The first step is contained in the following lemma.

**Lemma 2.5.8** $H_N^{\beta} \backslash H_N^{\beta,dg}$ has null Lebesgue measure. Further, the map $(D_N^d, U_N^{\beta,g}) \rightarrow H_N^{\beta,dg}$ given by $(D, U) \mapsto UDU^*$ is a one-to-one, while $(D_N^{d_0}, U_N^{\beta,g}) \rightarrow H_N^{\beta,dg}$ given by the same map is $N!$-to-one.

**Proof of Lemma 2.5.8** In order to prove the first part of the lemma, we note that for any non-vanishing polynomial function $p$ of the entries of $X$, the set $\{ X : p(X) = 0 \}$ is closed and has zero Lebesgue measure (this fact can be checked by applying Fubini’s theorem). So it is enough to exhibit a
non-vanishing polynomial \( p \) with \( p(X) = 0 \) if \( X \in \mathcal{H}_N^{(\beta)} \setminus \mathcal{H}_N^{(\beta),dg} \).

Given any \( n \times n \) matrix \( H \), for \( i, j = 1, \ldots, n \) let \( H^{(i,j)} \) be the \( (n-1) \times (n-1) \) matrix gotten by striking the \( i^{th} \) row and \( j^{th} \) column of \( H \), and write \( H^{(k)} \) for \( H^{(k,k)} \). We begin by proving that if \( X = UDU^* \) with \( D \in \mathcal{D}_N^{(\beta)} \) and \( X \) and \( X^{(k)} \) do not have eigenvalues in common for any \( k = 1, 2, \ldots, N \), then all entries of \( U \) are non-zero. Indeed, let \( \lambda \) be an eigenvalue of \( X \), set \( A = X - \lambda I \), and define \( A^{adj} \) as the \( N \times N \) matrix with \( A^{adj}_{i,j} = (-1)^{i+j} \det(A^{(i,j)}) \). Using the identity \( A^2 - \lambda I \), one concludes that \( A^{adj} = 0 \). Since the eigenvalues of \( X \) are assumed distinct, the null space of \( A \) has dimension 1, and hence all columns of \( A^{adj} \) are scalar multiple of some vector \( v_\lambda \), which is then an eigenvector of \( X \) corresponding to the eigenvalue \( \lambda \). Since \( A^{adj}_{i,i} = \det(X^{(i)} - \lambda I) \neq 0 \) by assumption, it follows that all entries of \( v_\lambda \) are non-zero. But each column of \( U \) is a non-zero scalar multiple of some \( v_\lambda \), leading to the conclusion that all entries of \( U \) do not vanish.

We recall, see Appendix A.3, that the resultant of the characteristic polynomials of \( X \) and \( X^{(k)} \), which can be written as a polynomial in the entries of \( X \) and \( X^{(k)} \), and hence as a polynomial \( P_1 \) in the entries of \( X \), vanishes if and only if \( X \) and \( X^{(k)} \) have a common eigenvalue. Further, the discriminant of \( X \), which is a polynomial \( P_2 \) in the entries of \( X \), vanishes if not all eigenvalues of \( X \) are distinct. Taking \( p(X) = P_1(X)P_2(X) \), one obtains a nonzero polynomial \( p \) with \( p(X) = 0 \) if \( X \in \mathcal{H}_N^{(\beta)} \setminus \mathcal{H}_N^{(\beta),dg} \). This completes the proof of the first part of the lemma.

The second part of the lemma is immediate since the eigenspace corresponding to each eigenvalue is of dimension 1, the eigenvectors are fixed by the normalization condition, and the multiplicity arises from the possible permutations of the order of the eigenvalues. □

Next, we say that \( U \in \mathcal{U}_N^{(\beta),sg} \) is of strongly general type if all minors of \( U \) have non-vanishing determinant. Let \( \mathcal{U}_N^{(\beta),sg} \) denote the collection of strongly general type matrices. The interest in such matrices is that they possess a particularly nice parametrization.

**Lemma 2.5.9** The maps \( T : \mathcal{U}_N^{(\beta),sg} \to \mathbb{R}^{\beta N(N-1)/2} \) defined by

\[
T(U) = \left( \frac{U_{1,2}}{U_{1,1}}, \ldots, \frac{U_{1,N}}{U_{1,1}}, \frac{U_{2,3}}{U_{2,2}}, \ldots, \frac{U_{2,N}}{U_{2,2}}, \ldots, \frac{U_{N-1,N}}{U_{N-1,N-1}} \right)
\]

(where \( \mathbb{C} \) is identified with \( \mathbb{R}^2 \) in case \( \beta = 2 \)) is one-to-one with smooth inverse. Further, the set \( T\left((\mathcal{U}_N^{(\beta),sg})^c\right) \) is closed and has zero Lebesgue measure.

**Proof of Lemma 2.5.9** We begin with the first part. The proof is by an
inductive construction. Clearly, \( U_{1,1}^{-2} = 1 + \sum_{j=2}^{N} |U_{1,j}|^2 / |U_{1,1}|^2 \). So suppose that \( U_{i,j} \) are given for \( 1 \leq i \leq i_0 \) and \( 1 \leq j \leq N \). Let \( v_i = (U_{i,1}, \ldots, U_{i,i_0}) \), \( i = 1, \ldots, i_0 \). One then solves the equation

\[
\begin{pmatrix}
v_1 \\
v_2 \\
\vdots \\
v_{i_0}
\end{pmatrix}
Z = -
\begin{pmatrix}
U_{1,i_0+1} + \sum_{i=i_0+2}^{N} U_{1,i} U_{i_0+1,i} \\
U_{2,i_0+1} + \sum_{i=i_0+2}^{N} U_{2,i} U_{i_0+1,i} \\
\vdots \\
U_{i_0,i_0+1} + \sum_{i=i_0+2}^{N} U_{i_0,i} U_{i_0+1,i}
\end{pmatrix}.
\]

The strong general type condition on \( U \) ensures that the vector \( Z \) is uniquely determined by this equation, and one then sets \( U_{i_0+1,j} = Z_j / \sqrt{\sum_{k=1}^{N} |Z_k|^2} \) for \( 1 \leq j \leq i_0 \) and \( U_{i_0+1,i_0+1} = 1 - \sqrt{\sum_{k=1}^{i_0} |U_{i_0+1,k}|^2} \) (all entries \( U_{i_0+1,j} \) with \( j > i_0 + 1 \) are then determined by \( T(U) \)). This completes the proof of the first part. To see the second part, note first that \( T(U_N^{(3),sg}) = T(Z_N^{(3)}) \) where \( Z_N \) is the space of matrices whose columns are orthogonal, whose diagonal entries all equal to 1, and all of whose minors have non-vanishing determinants. Applying the previous constructions, one immediately obtains a polynomial type condition for a point in \( \mathbb{R}^{3N(N-1)/2} \) to not belong to the set \( T(Z_N^{(3)}) \).

Let \( \mathcal{H}_N^{(3),sg} \) denote the subset of \( \mathcal{H}_N^{(3),dg} \) consisting of those matrices \( X \) that can be written as \( X = UDU^* \) with \( D \in \mathcal{D}_N^d \) and \( U \in \mathcal{U}_N^{(3),sg} \).

**Lemma 2.5.10** The Lebesgue measure of \( \mathcal{H}_N^{(3),sg} \setminus \mathcal{H}_N^{(3),sg} \) is zero.

**Proof of Lemma 2.5.10:** We identify a subset of \( \mathcal{H}_N^{(3),sg} \) which we will prove to be of full Lebesgue measure. We say that a matrix \( D \in \mathcal{D}_N^d \) is *strongly distinct* if for any integer \( r = 1, 2, \ldots, N - 1 \) and subsets \( I, J \) of \( \{1, 2, \ldots, N\} \),

\[
I = \{i_1 < \cdots < i_r\}, \quad J = \{j_1 < \cdots < j_r\}
\]

with \( I \neq J \), it holds that \( \prod_{i \in I} D_{i,i} \neq \prod_{i \in J} D_{i,i} \). We consider the subset \( \mathcal{H}_N^{(3),sdg} \) of \( \mathcal{H}_N^{(3),sg} \) consisting of those matrices \( X = UDV^* \) with \( D \) strongly distinct and \( U \in \mathcal{U}_N^{(3),sdg} \).

Given an integer \( r \) and subsets \( I, J \) as above, put

\[
(\bigwedge_r X)_{IJ} := \prod_{\mu, \nu=1}^{r} X_{i_\mu, j_\nu},
\]

thus defining a square matrix \( \bigwedge_r X \) with rows and columns indexed by \( r \)-element subsets of \( \{1, \ldots, n\} \). If we replace each entry of \( X \) by its complex
conjugate, we replace each entry of $\Lambda^r X$ by its complex conjugate. If we replace $X$ by its transpose, we replace $\Lambda^r X$ by its transpose. Given another $N$ by $N$ matrix $Y$ with complex entries, by the Cauchy-Binet Theorem A.3 we have $\Lambda^r(XY) = (\Lambda^r X)(\Lambda^r Y)$. Thus, if $U \in \mathcal{U}_N^{(\beta)}$ then $\Lambda^r U \in \mathcal{U}_N^{(\beta)}$ where $c_r^\beta = N!/(N - r)!r!$. We thus obtain that if $X = UD_U^*$ then $\Lambda^r X$ can be decomposed as $\Lambda^r X = (\Lambda^r U)(\Lambda^r D)(\Lambda^r U^*)$. In particular, if $D$ is not strongly distinct then for some $r$, $\Lambda^r X$ does not possess all eigenvalues distinct. Similarly, if $D$ is strongly distinct but $U \notin \mathcal{U}_N^{(\beta),sg}$, then some entry of $\Lambda^r U$ vanishes. Repeating the argument presented in the proof of the first part of Lemma 2.5.8, we conclude that the Lebesgue measure of $\mathcal{H}_N^{(\beta)} \setminus \mathcal{H}_N^{(\beta),sg}$ vanishes. This completes the proof of the lemma. \qed

We are now ready to provide the

**Proof of (2.5.3):** Recall the map $\hat{T}$ introduced in Lemma 2.5.9, and define, the map $\hat{T} : T(\mathcal{U}_N^{(\beta),sg}) \times \mathbb{R}^N \to \mathcal{H}_N^{(\beta)}$ by setting, for $\lambda \in \mathbb{R}^N$ and $z \in T(\mathcal{U}_N^{(\beta),sg})$, $D \in D_N$ with $D_{r,i} = \lambda_i$ and $\hat{T}(\lambda, z) = T^{-1}(z)DT^{-1}(z)^*$. By Lemma 2.5.9, $\hat{T}$ is smooth, whereas by Lemma 2.5.8, it is $N!$-to-1 on a set of full Lebesgue measure and is locally one-to-one on a set of full Lebesgue measure. Letting $J\hat{T}$ denote the Jacobian of $\hat{T}$, we note that $J\hat{T}(\lambda, z)$ is a homogeneous polynomial in $\lambda$ of degree $\beta N(N - 1)/2$ with coefficients that are functions of $z$ (since derivatives of $\hat{T}(\lambda, z)$ with respect to the variables $z$ do not depend on $\lambda$, while derivatives with respect to the $z$ variables are linear in $\lambda$). Since $\hat{T}$ fails to be locally one-to-one only when $\lambda_i = \lambda_j$ for some $i \neq j$, it follows by the implicit function theorem that $J\hat{T}$ vanishes exactly at such points. Hence, $\Delta(\lambda) := \prod_{i < j}(\lambda_i - \lambda_j)$ is a factor of $J\hat{T}$, and by symmetry under permutations, it follows that $J\hat{T}(\lambda, z)$ must be proportional to an integer power of $\Delta(\lambda)$, which can only vanish when $\Delta(\lambda)$ vanishes. Since $\Delta(\lambda)$ is a polynomial of degree $N(N - 1)/2$, it follows that $J\hat{T}(\lambda, z) = g(z)\Delta(\lambda)\beta$ for some (continuous, hence measurable) function $g$. By Lemma 2.5.10, we conclude that for any function $f$ that depends only on the eigenvalues of $X$, it holds that

$$N! \int f(H)dP_N^{(\beta)} = \int g(z)dz \int f(\lambda)|\Delta(\lambda)|^\beta \prod_{i=1}^N e^{-\beta \lambda_i^2/4}d\lambda_i.$$ 

Up to the normalization constant $(\int g(z)dz)/N!$, this is precisely (2.5.3).

### 2.5.3 Selberg’s integral formula and proof of (2.5.4)

We begin by stating Selberg’s integral formula. We then describe in Corollary 2.5.13 a couple of limiting cases of Selberg’s formula. The evaluation of the normalizing constant in (2.5.4) is immediate from Corollary 2.5.13.
Theorem 2.5.11 (Selberg’s integral formula) For all positive numbers \(a, b\) and \(c\) we have
\[
\frac{1}{n!} \int_0^1 \cdots \int_0^1 \Delta(x)^{2c} \prod_{i=1}^n x_i^{a_i-1}(1-x_i)^{b-1} dx_i = \prod_{j=0}^{n-1} \frac{\Gamma(a+jc)\Gamma(b+jc)\Gamma((j+1)c)}{\Gamma(a+b+(n+j-1)c)\Gamma(c)}. \tag{2.5.12}
\]

Corollary 2.5.13 For all positive numbers \(a\) and \(c\) we have
\[
\frac{1}{n!} \int_0^\infty \cdots \int_0^\infty \Delta(x)^{2c} \prod_{i=1}^n x_i^{a_i-1} e^{-x_i} dx_i = \prod_{j=0}^{n-1} \frac{\Gamma(a+jc)\Gamma((j+1)c)}{\Gamma(c)}, \tag{2.5.14}
\]
and
\[
\frac{1}{n!} \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty \Delta(x)^{2c} \prod_{i=1}^n e^{-x_i^2/2} dx_i = (2\pi)^{n/2} \prod_{j=0}^{n-1} \frac{\Gamma((j+1)c)}{\Gamma(c)}. \tag{2.5.15}
\]

Remark 2.5.16 The identities in Theorem 2.5.11 and Corollary 2.5.13 hold under rather less stringent conditions on the parameters \(a, b\) and \(c\). For example, one can allow \(a, b,\) and \(c\) to be complex with positive real parts. We refer to the bibliographical notes for references. We note also that only (2.5.15) is directly relevant to the study of the normalization constants for the GOE and GUE. The usefulness of the other more complicated formulas will become apparent in Section ??.

We will prove Theorem 2.5.11 following Anderson’s method [And91], after first explaining how to deduce Corollary 2.5.13 from (2.5.12) by means of the Stirling approximation, which we recall is the statement
\[
\Gamma(s) = \sqrt{\frac{2\pi}{s}} \left(\frac{s}{e}\right)^s (1 + o_{s\to+\infty}(1)), \tag{2.5.17}
\]
where \(s\) tends to \(+\infty\) along the positive real axis. (The proof of (2.5.17) is an application of Laplace’s method, see Exercise 3.5.15.)

Proof of Corollary 2.5.13 We denote the left side of (2.5.12) by \(S_n(a, b, c)\). Consider first the integral
\[
I_s = \frac{1}{n!} \int_0^s \cdots \int_0^s \Delta(x)^{2c} \prod_{i=1}^n x_i^{a_i-1}(1-x_i/s)^s dx_i
\]
where $s$ is a large positive number. By monotone convergence, the left side of (2.5.14) equals $\lim_{s \to \infty} I_s$. By rescaling the variables of integration, we find that
\[ I_s = s^{n(a+(n-1)c)} S_n(a, s+1, c). \]
From (2.5.17) we deduce the formula
\[ \frac{\Gamma(s+1+A)}{\Gamma(s+1+B)} = s^{A-B} (1 + o_{s \to +\infty}(1)), \quad (2.5.18) \]
in which $A$ and $B$ are any real constants. Finally, (2.5.12) taken for granted, we can evaluate $\lim_{s \to \infty} I_s$ with the help of (2.5.18), thus verifying (2.5.14).

Turning to the proof of (2.5.15), consider the integral
\[ J_s = \frac{1}{n!} \int_{-\sqrt{2s}}^{\sqrt{2s}} \cdots \int_{-\sqrt{2s}}^{\sqrt{2s}} \Delta(x)^{2c} \left( 1 - \frac{x^2}{2s} \right)^s \prod_{i=1}^n dx_i \]
where $s$ is a large positive number. By monotone convergence the left side of (2.5.15) equals $\lim_{s \to \infty} J_s$. By shifting and rescaling the variables of integration, we find that
\[ J_s = \frac{2^{3n(n-1)/2+3n/2+2ns} s^{n(n-1)c/2+n/2}}{\pi^{n/2}} S_n(s+1, s+1, c). \]
From (2.5.17) we deduce the formula
\[ \frac{\Gamma(2s+2+A)}{\Gamma(s+1+B)} = 2^{A+3/2+2s} e^{A-2B+1/2} \sqrt{2\pi} \frac{1}{\Gamma(A+1)} (1 + o_{s \to +\infty}(1)) \quad (2.5.19) \]
where $A$ and $B$ are any real constants. Finally, (2.5.12) taken for granted, we can evaluate $\lim_{s \to \infty} J_s$ with the help of (2.5.19), thus verifying (2.5.15).

Before providing the proof of Theorem 2.5.11, we note the following identity involving the beta integral in the left side:
\[ \int_{\{x \in \mathbb{R}^n : \min_{i=1}^n x_i > 0, \sum_{i=1}^n x_i = 1\}} \left( 1 - \sum_{i=1}^n x_i \right)^{s_i-1} \prod_{i=1}^n x_i \, dx_i = \frac{\Gamma(s_1) \cdots \Gamma(s_{n+1})}{\Gamma(s_1 + \cdots + s_{n+1})} \quad (2.5.20) \]
The identity (2.5.20) is proved by substituting $u_1 = tx_1, \ldots, u_n = tx_n, u_{n+1} = t(1-x_1-\cdots-x_n)$ in the integral
\[ \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{n+1} u_i^{s_i-1} e^{-u_i} du_i. \]
and applying Fubini’s theorem both before and after the substitution.

**Proof of Theorem 2.5.11** We aim now to rewrite the left side of (2.5.12) in an intuitive way, see Lemma 2.5.23 below. Toward this end, we introduce some notation.

Let $\mathcal{D}_n$ be the space consisting of monic polynomials $P(t)$ of degree $n$ in a variable $t$ with real coefficients such that $P(t)$ has $n$ distinct real roots. More generally, given an open interval $I \subset \mathbb{R}$, let $\mathcal{D}_n I \subset \mathcal{D}_n$ be the subspace consisting of polynomials with $n$ distinct roots in $I$. Given $x \in \mathbb{R}^n$, let $P_x(t) = t^n + \sum_{i=1}^{n} (-1)^i x_{n-i} t^{n-i}$. For any open interval $I \subset \mathbb{R}$, the set $\{x \in \mathbb{R}^n \mid P_x \in \mathcal{D}_n I\}$ is open, since the perturbation of a degree $n$ polynomial by the addition of a degree $n-1$ polynomial with small real coefficients does not destroy the property of having $n$ distinct real roots, nor does it move the roots very much. By definition a set $A \subset \mathcal{D}_n$ is measurable if and only if $\{x \in \mathbb{R}^n \mid P_x \in A\}$ is Lebesgue measurable. Let $\ell_n$ be the measure on $\mathcal{D}_n$ obtained by pushing Lebesgue measure on the open set $\{x \in \mathbb{R}^n \mid P_x \in \mathcal{D}_n\}$ forward to $\mathcal{D}_n$ via $x \mapsto P_x$ (that is, under $\ell_n$, monic polynomials of degree $n$ have coefficients that are jointly Lebesgue distributed). Given $P \in \mathcal{D}_n$, we define $\sigma_k(P) \in \mathbb{R}$ for $k = 0, \ldots, n$ by the rule $P(t) = \sum_{k=0}^{n} (-1)^k \sigma_k(P) t^{n-k}$. Equivalently, if $\alpha_1 < \cdots < \alpha_n$ are the roots of $P \in \mathcal{D}_n$, we have $\sigma_0(P) = 1$ and

$$\sigma_k(P) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} \alpha_{i_1} \cdots \alpha_{i_k}$$

for $k = 1, \ldots, n$. The map $(P \mapsto (\sigma_1(P), \ldots, \sigma_n(P))) : \mathcal{D}_n \to \mathbb{R}^n$ inverts the map $(x \mapsto P_x) : \{x \in \mathbb{R}^n \mid P_x \in \mathcal{D}_n\} \to \mathcal{D}_n$. Let $\tilde{\mathcal{D}}_n \subset \mathbb{R}^n$ be the open set consisting of $n$-tuples $(x_1, \ldots, x_n)$ such that $x_1 < \cdots < x_n$. Finally, for $P \in \mathcal{D}_n$ with roots $\alpha = (\alpha_1 < \cdots < \alpha_n)$, we set $D(P) = \prod_{i \neq j} (\alpha_i - \alpha_j) = \Delta(\alpha)^2$.

**Lemma 2.5.21** For $k, \ell = 1, \ldots, n$ and $x = (x_1, \ldots, x_n) \in \tilde{\mathcal{D}}_n$ put

$$\tau_k = \tau_k(x_1, \ldots, x_n) = \sum_{1 \leq i_1 < \cdots < i_k \leq n} x_{i_1} \cdots x_{i_k}, \quad \tau_{k, \ell} = \frac{\partial \tau_k}{\partial x_\ell}.$$

Then,

$$\left| \det_{k, \ell=1}^{n} \tau_{k, \ell} \right| = \prod_{1 \leq i < j \leq n} \left| x_i - x_j \right| = |\Delta(x)|. \quad (2.5.22)$$

**Proof:** We have

$$\tau_{k, \ell} = \sigma_{k-1} \left( \prod_{\ell \in \{1, \ldots, n\} \setminus \{\ell\}} (t - x_i) \right),$$

for
whence follows the identity
\[
\sum_{m=1}^{n} (-1)^{m-1} x_k^{n-m} \tau_{m,t} = \delta_{k,t} \prod_{i \in \{1, \ldots, n\} \setminus \{t\}} (x_t - x_i).
\]

This last is equivalent to a matrix identity $AB = C$ where $\det A$ up to a sign equals the Vandermonde determinant $\det_{i,j=1}^{n} x_j^{n-1}$, $\det B$ is the determinant we want to calculate, and $\det C$ up to a sign equals $(\det A)^2$. Formula (2.5.22) follows. \( \square \)

(See Exercise 2.5.32 for an alternative proof of Lemma 2.5.21.)

We can now state and prove the representation formula alluded to above.

**Lemma 2.5.23** The left side of (2.5.12) equals
\[
\int_{D_n(0,1)} |P(0)|^{a-1} |P(1)|^{b-1} D(P)^{c-1/2} d\ell_n(P). \tag{2.5.24}
\]

**Proof:** We prove a slightly more general statement: for any nonnegative $\ell_n$-measurable function $f$ on $D_n$, we have
\[
\int_{D_n} f d\ell_n = \int_{D_n} f(\prod_{i=1}^{n} (t - x_i)) \Delta(x) dx_1 \cdots dx_n, \tag{2.5.25}
\]
from which (2.5.24) follows by taking $f(P) = |P(0)|^{a-1} |P(1)|^{b-1} D(P)^{c-1/2}$. To see (2.5.25), put $g(x) = f(P_x)$ for $x \in \mathbb{R}^n$ such that $P_x \in D_n$. We have
\[
\int_{\{x \in \mathbb{R}^n | P_x \in D_n\}} g(x_1, \ldots, x_n) dx_1 \cdots dx_n
= \int_{D_n} g(\tau_1, \ldots, \tau_n) \left| \det_{k,\ell=1}^{n} \tau_{k,\ell} \right| dx_1 \cdots dx_n \tag{2.5.26}
\]
by the usual formula for changing variables in a multivariable integral. The left sides of (2.5.25) and (2.5.26) are equal by definition; the right sides are equal by (2.5.22). \( \square \)

We next transform some naturally occurring integrals on $D_n$ to beta integrals, see Lemma 2.5.30 below. This involves some additional notation. Let $\mathcal{E}_n \subset D_n \times D_{n+1}$ be the subset consisting of pairs $(P, Q)$ such that the roots $\alpha_1 < \cdots < \alpha_n$ of $P$ and the roots $\beta_1 < \cdots < \beta_{n+1}$ of $Q$ are interlaced, i.e., $\alpha_i \in (\beta_i, \beta_{i+1})$ for $i = 1, \ldots, n$. More generally, given an interval $I \subset \mathbb{R}$, let $\mathcal{E}_n I = \mathcal{E}_n \cap (D_n I \times D_{n+1} I)$. 


Lemma 2.5.27 Fix \( Q \in \mathcal{D}_{n+1} \) with roots \( \beta_1 < \cdots < \beta_{n+1} \). Fix real numbers \( \gamma_1, \ldots, \gamma_{n+1} \) and let \( P(t) \) be the unique polynomial in \( t \) of degree \( \leq n \) with real coefficients such that the partial fraction expansion
\[
P(t) = \sum_{i=1}^{n+1} \frac{\gamma_i}{t - \beta_i}
\]
holds. Then the following statements are equivalent:

(I) \((P, Q) \in \mathcal{E}_n\).

(II) \(\min_{i=1}^{n+1} \gamma_i > 0 \) and \(\sum_{i=1}^{n+1} \gamma_i = 1\).

Proof: (I\(\Rightarrow\)II) The numbers \( P(\beta_i) \) do not vanish and their signs alternate. Similarly, the numbers \( Q'(\beta_i) \) do not vanish and their signs alternate. By L’Hôpital’s Rule, we have \( \gamma_i = P(\beta_i)/Q'(\beta_i) \) for \( i = 1, \ldots, n + 1 \). Thus all the quantities \( \gamma_i \) are nonzero and have the same sign. The quantity \( P(t)/Q'(t) \) depends continuously on \( t \) in the interval \( [\beta_{n+1}, \infty) \), does not vanish in that interval, and tends to \( 1/(n+1) \) as \( t \to +\infty \). Thus \( \gamma_{n+1} \) is positive. Since the signs of \( P(\beta_i) \) alternate, and so do the signs of \( Q'(\beta_i) \), it follows that \( \gamma_i = P(\beta_i)/Q'(\beta_i) > 0 \) for all \( i \). Because \( P(t) \) is monic, the numbers \( \gamma_i \) sum to 1. Thus condition (II) holds.

(II\(\Rightarrow\)I) Because the signs of the numbers \( Q'(\beta_i) \) alternate, we have sufficient information to force \( P(t) \) to change sign \( n + 1 \) times, and thus to have \( n \) distinct real roots interlaced with the roots of \( Q(t) \). And because the numbers \( \gamma_i \) sum to 1, the polynomial \( P(t) \) must be monic in \( t \). Thus condition (I) holds. \( \square \)

Lemma 2.5.28 Fix \( Q \in \mathcal{D}_{n+1} \) with roots \( \beta_1 < \cdots < \beta_{n+1} \). Then we have
\[
\ell_n\{P \in \mathcal{D}_n \mid (P, Q) \in \mathcal{E}_n\} = \frac{1}{n!} \prod_{j=1}^{n+1} |Q'(\beta_j)|^{1/2} = \frac{D(Q)^{1/2}}{n!}. \tag{2.5,29}
\]

Proof: Consider the set
\[
A = \{x \in \mathbb{R}^n \mid (P_x, Q) \in \mathcal{E}_n\}.
\]
By definition the left side of (2.5.29) equals the Lebesgue measure of \( A \). Consider the polynomials \( Q_j(t) = Q(t)/(t - \beta_j) \) for \( j = 1, \ldots, n + 1 \). By Lemma 2.5.27, for all \( x \in \mathbb{R}^n \), we have \( x \in A \) if and only if \( P_j(t) = \sum_{i=1}^{n+1} \gamma_i Q_i(t) \) for some real numbers \( \gamma_i \) such that \( \min \gamma_i > 0 \) and \( \sum \gamma_i = 1 \), or equivalently, \( A \) is the interior of the convex hull of the points
\[
(\tau_{2,j}(\beta_1, \ldots, \beta_{n+1}), \ldots, \tau_{n+1,j}(\beta_1, \ldots, \beta_{n+1})) \in \mathbb{R}^n \quad \text{for} \quad j = 1, \ldots, n + 1,
\]
where the $\tau$'s are defined as in Lemma 2.5.21 (but with $n$ replaced by $n + 1$). Noting that $\tau_{1,\ell} \equiv 1$ for $\ell = 1, \ldots, n + 1$, the Lebesgue measure of $A$ equals the absolute value of $\frac{1}{n!} \det_{k,\ell=1}^{n+1} \tau_{k,\ell}(\beta_1, \ldots, \beta_{n+1})$ by the determinantal formula for computing the volume of a simplex in $\mathbb{R}^n$. Finally, we get the claimed result by (2.5.22).

□

Lemma 2.5.30 Fix $Q \in D_{n+1}$ with roots $\beta_1 < \cdots < \beta_{n+1}$. Fix positive numbers $s_1, \ldots, s_{n+1}$. Then we have

$$\int \{|P(\beta)|^{s_i-1}d\ell_n(P)|_{P,Q} \in E_n\} = \frac{\prod_{i=1}^{n+1} |Q'(\beta)|^{s_i-1/2}\Gamma(s_i)}{\Gamma(\sum_{i=1}^{n+1} s_i)}.$$  (2.5.31)

Proof: Because $Q(x)/(t-x)Q'(x)|_{x=\beta} = 1$, it follows from Lemma 2.5.27 that the integral

$$\int \{|P(\beta)|^{s_i-1}d\ell_n(P)|_{P,Q} \in E_n\}$$

is just a linear change of coordinates away from the beta integral (2.5.20). Thus (2.5.31) holds after multiplying the expression on the right by a positive number $C$ independent of $s_1, \ldots, s_{n+1}$. Finally, by evaluating the left side of (2.5.31) for $s_1 = \cdots = s_{n+1} = 1$ by means of Lemma 2.5.28 (and recalling the $\Gamma(n+1) = n!$) we find that $C = 1$. □

We may now complete the proof of Theorem 2.5.11. Recall that the integral on the left side of (2.5.12), denoted as above by $S_n(a,b,c)$, can be represented as the integral (2.5.24). Consider the double integral

$$K_n(a,b,c) = \int_{E_n(0,1)} |Q(0)|^{a-1}|Q(1)|^{b-1}|R(P,Q)|^{c-1}d\ell_n(P)d\ell_{n+1}(Q),$$

where $R(P,Q)$ denotes the resultant of $P$ and $Q$, see Appendix A.3. We will apply Fubini’s theorem in both possible ways. On the one hand, we have

$$K_n(a,b,c) = \int_{D_{n+1}(0,1)} |Q(0)|^{a-1}|Q(1)|^{b-1}$$

$$\times \left(\int_{\{P \in D_n(0,1)\}|P,Q} \in E_n\} |R(P,Q)|^{c-1}d\ell_n(P)\right) d\ell_{n+1}(Q)$$

$$= S_{n+1}(a,b,c) \frac{\Gamma(c^{n+1})}{\Gamma((n+1)c)}$$
via Lemma 2.5.30. On the other hand, writing $\tilde{P} = t(t - 1)P$, we have

$$K_n(a, b, c) = \int_{D_n(0, 1)} \left( \int_{\{Q \in D_{n+1}((Q, \tilde{P}) \in E_{n+2}\}} |Q(0)|^{a-1}|Q(1)|^{b-1}|R(P, Q)|^{c-1} \, d\nu_{n+1}(Q) \right) \, d\nu_n(P)$$

$$= \int_{D_n(0, 1)} |\tilde{P}'(0)|^{a-1/2}|\tilde{P}'(1)|^{b-1/2}|R(P, \tilde{P}')|^{c-1/2} \, d\nu_{n}(P) \frac{\Gamma(a)\Gamma(b)\Gamma(c)^n}{\Gamma(a + b + nc)}$$

$$= S_n(a + c, b + c, c) \frac{\Gamma(a)\Gamma(b)\Gamma(c)^n}{\Gamma(a + b + nc)}$$

by another application of Lemma 2.5.30. This proves (2.5.12) by induction on $n$; the induction base $n = 1$ is an instance of (2.5.20).

Exercise 2.5.32 Provide an alternative proof of Lemma 2.5.21 by noting that the determinant in the left side of (2.5.22) is a polynomial of degree $n(n - 1)/2$ that vanishes whenever $x_i = x_j$ for some $i \neq j$, and thus, must equal a constant multiple of $\Delta(x)$.

2.5.4 Joint distribution of eigenvalues - alternative formulation

It is sometimes useful to represent the formulae for the joint distribution of eigenvalues as integration formulae for functions that depend only on the eigenvalues. We develop this correspondence now.

Let $f : \mathcal{H}^N \rightarrow [0, \infty]$ be a Borel function such that $f(H)$ depends only on the sequence of eigenvalues $\lambda_1(H) \leq \cdots \leq \lambda_N(H)$. In this situation, for short, we say that $f(H)$ depends only on the eigenvalues of $H$. (note that the definition implies that $f$ is a symmetric function of the eigenvalues of $H$). Let $X \in \mathcal{H}^{(2)}$ be random with law $P_N^{(2)}$. Theorem 2.5.2 granted, we have

$$Ef(X) = \frac{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_N) \prod_{1 \leq i < j \leq N} |x_i - x_j| \beta \prod_{i=1}^{N} e^{-\beta x_i^2/4} \, dx_i}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{1 \leq i < j \leq N} |x_i - x_j| \beta \prod_{i=1}^{N} e^{-\beta x_i^2/4} \, dx_i}$$

(2.5.33)

where $f(x_1, \ldots, x_N)$ denotes the value of $f$ at the diagonal matrix with diagonal entries $x_1, \ldots, x_N$. Conversely, formula (2.5.33) granted, we immediately verify that (2.5.3) is proportional to the joint density of the eigenvalues $\lambda_1(X), \ldots, \lambda_N(X)$ by taking $f(H) = 1_{\{\lambda_1(H), \ldots, \lambda_N(H)\} \in A}$ where $A \subset \mathbb{R}^N$ is any Borel set. In turn, to prove (2.5.33), it suffices to prove the
general integration formula

\[
\int f(H) \ell_N^{(\beta)}(dH) = C_N^{(\beta)} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_N) \prod_{1 \leq i < j \leq N} |x_i - x_j|^\beta \prod_{j=1}^{N} dx_i
\]

where

\[
C_N^{(\beta)} = \begin{cases} 
\frac{1}{N!} \prod_{k=1}^{N} \frac{\Gamma(1/2)^k}{\Gamma(k/2)} & \text{if } \beta = 1, \\
\frac{1}{N!} \prod_{k=1}^{N} \frac{\pi^{k-1}}{(k-1)!} & \text{if } \beta = 2,
\end{cases}
\]

and as in (2.5.33), the integrand \( f(H) \) is nonnegative, Borel measurable, and depends only on the eigenvalues of \( H \). Moreover, integration formula (2.5.34) granted, it follows by taking \( f(H) = \exp(-a \text{Tr}(H^2)/2) \) with \( a > 0 \) and using Gaussian integration that

\[
\frac{1}{N!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{1 \leq i < j \leq N} |x_i - x_j|^\beta \prod_{i=1}^{N} e^{-ax_i^2/2} dx_i
\]

\[
= (2\pi)^{N/2} a^{-\beta N(N-1)/4-N/2} \prod_{j=1}^{N} \frac{\Gamma(j\beta/2)}{\Gamma(j/2)} =: \frac{1}{N!C_N^{(\beta)}}.
\]

Thus, Theorem 2.5.2 is equivalent to integration formula (2.5.34).

**Exercise 2.5.36** Guide the students through the analysis of the Gaussian symplectic ensemble (GSE).

### 2.6 Large deviations for random matrices

In this section, we consider \( N \) random variables \((\lambda_1, \ldots, \lambda_N)\) with law

\[
P_{V,\beta}^N(d\lambda_1, \cdots, d\lambda_N) = (Z_{V,\beta}^N)^{-1} |\Delta(\lambda)|^\beta e^{-N \sum_{i=1}^{N} V(\lambda_i)} \prod_{i=1}^{N} d\lambda_i,
\]

for a continuous function \( V : \mathbb{R} \to \mathbb{R} \) such that

\[
\liminf_{|x| \to \infty} \frac{V(x)}{\beta \log |x|} > 1
\]
and a positive real number $\beta$. Here, $\Delta(\lambda) = \prod_{1 \leq i < j \leq N} (\lambda_i - \lambda_j)$ and

$$Z_{V,\beta}^N = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} |\Delta(\lambda)|^\beta e^{-N \sum_{i=1}^N V(\lambda_i)} \prod_{i=1}^N d\lambda_i. \quad (2.6.3)$$

When $V(x) = \beta x^2 / 4$, and $\beta = 1, 2$, we saw in Section 2.5 that $P_{\beta x^2/4, \beta}$ is the law of the eigenvalues of a $N \times N$ GOE matrix when $\beta = 1$, and of a GUE matrix when $\beta = 2$. It also follows from the general results in Section 4.1 that the case $\beta = 4$ corresponds to another matrix ensemble, namely the GSE. In view of these and application to certain problems in physics, we consider in this section the slightly more general model. We emphasize however that the distribution (2.6.1) precludes us from considering random matrices with independent non Gaussian entries.

We have proved earlier in this chapter (for the GOE, see Section 2.1, and for the GUE, see Section 2.2) that the spectral measure $L_N = N^{-1} \sum_{i=1}^N \delta_{\lambda_i}$ converges in probability (and almost surely, under appropriate moment assumptions), and we studied its fluctuations around its mean. We have also considered the convergence of the top eigenvalue $\lambda_N^N$. Such results did not depend much on the Gaussian nature of the entries.

We address here a different type of question. Namely, we study the probability that $L_N$, or $\lambda_N^N$, take a very unlikely value. This was already considered in our discussion of concentration inequalities, c.f. Section 2.3, where the emphasis was put on obtaining upper bounds on the probability of deviation. In contrast, the purpose of the analysis here is to exhibit a precise estimate on these probabilities, or at least on their logarithmic asymptotics. The appropriate tool for handling such questions is large deviations theory, and we bring in Appendix D a concise introduction to that theory and related definitions, together with related references.

### 2.6.1 Large deviations for the empirical measure

Endow $M_1(\mathbb{R})$ with the usual weak topology, compatible with the Lipschitz bounded metric, see (C.9). Our goal is to estimate the probability $P_{V,\beta}(L_N \in A)$, for measurable sets $A \subset M_1(\mathbb{R})$. Of particular interest is the case where $A$ does not contain the limiting distribution of $L_N$.

Define the non-commutative entropy $\Sigma : M_1(\mathbb{R}) \to [-\infty, \infty]$, as

$$\Sigma(\mu) = \begin{cases} \int \int \log |x - y| d\mu(x) d\mu(y) & \text{if } \int \log |x| d\mu(x) < \infty, \\ -\infty & \text{else}. \end{cases} \quad (2.6.4)$$
Set next
\[
I_{\beta}^V(\mu) = \begin{cases} 
\int V(x)d\mu(x) - \frac{\beta}{2}\Sigma(\mu) - c_{\beta}^V & \text{if } \int V(x)d\mu(x) < \infty \\
\infty & \text{else},
\end{cases}
\]  
with \(c_{\beta}^V = \inf_{\nu \in M_1(\mathbb{R})} \{ \int V(x)d\nu(x) - \frac{\beta}{2}\Sigma(\nu) \} \). \hfill (2.6.5)

**Theorem 2.6.6** Let \( L_N = N^{-1} \sum_{i=1}^{N} \delta_{\lambda_iN} \) where the random variables \( \{\lambda_iN\}_{i=1}^{N} \) are distributed according to the law \( P_{V,\beta}^N \) with potential \( V \) satisfying (2.6.2), see (2.6.1). Then, the family of random measures \( L_N \) satisfies, in \( M_1(\mathbb{R}) \) equipped with the weak topology, a large deviation principle in the scale \( N^2 \), with good rate function \( I_{\beta}^V \). That is, \( I_{\beta}^V : M_1(\mathbb{R}) \to [0, \infty] \) possesses compact level sets \( \{ \nu : I_{\beta}^V(\nu) \leq M \} \) for all \( M \in \mathbb{R}_+ \), and

for any open set \( O \subset M_1(\mathbb{R}) \),
\[
\liminf_{N \to \infty} \frac{1}{N^2} \log P_{V,\beta}^N (L_N \in O) \geq - \inf_O I_{\beta}^V,
\]  
and

for any closed set \( F \subset M_1(\mathbb{R}) \),
\[
\limsup_{N \to \infty} \frac{1}{N^2} \log P_{V,\beta}^N (L_N \in F) \leq - \inf_F I_{\beta}^V.
\]  

The proof of Theorem 2.6.6 relies on the properties of the function \( I_{\beta}^V \) collected in Lemma 2.6.9 below. Define the logarithmic capacity of a measurable set \( A \subset \mathbb{R} \) as
\[
\gamma(A) := \exp \left\{ - \inf_{\nu \in M_1(A)} \int \int \log \frac{1}{|x-y|}d\nu(x)d\nu(y) \right\}.
\]

**Lemma 2.6.9**

a. \( I_{\beta}^V \) is well defined on \( M_1(\mathbb{R}) \) and takes its values in \([0, +\infty] \).

b. \( I_{\beta}^V(\mu) \) is infinite as soon as \( \mu \) satisfies one of the following conditions
   b.1 \( \int V(x)d\mu(x) = +\infty \).
   b.2 There exists a set \( A \subset \mathbb{R} \) of positive \( \mu \) mass but null logarithmic capacity, i.e. a set \( A \) such that \( \mu(A) > 0 \) but \( \gamma(A) = 0 \).

c. \( I_{\beta}^V \) is a good rate function.

d. \( I_{\beta}^V \) is a strictly convex function on \( M_1(\mathbb{R}) \).
e. $I^V_\beta$ achieves its minimum value at a unique probability measure $\sigma^V_\beta$ on $\mathbb{R}$ characterized by

$$V(x) - \beta \int \log |y - x| \, d\sigma^V_\beta(y) = \inf_{\nu \in M_1(\mathbb{R})} \left( \int V \, d\nu - \beta \Sigma(\nu) \right), \sigma^V_\beta \text{ a.s.,}$$

(2.6.10)

and, for all $x$ except possibly on a set with null logarithmic capacity,

$$V(x) - \beta \int \log |y - x| \, d\sigma^V_\beta(y) \geq \inf_{\nu \in M_1(\mathbb{R})} \left( \int V \, d\nu - \beta \Sigma(\nu) \right),$$

(2.6.11)

with the inequality strict outside the support of $\sigma^V_\beta$.

As an immediate corollary of Theorem 2.6.6 and of part e. of Lemma 2.6.9 we have the following.

**Corollary 2.6.12 (Proof #5? of Wigner's theorem)** Under $P^N_{V,\beta}$, $L_N$ converges almost surely towards $\sigma^V_\beta$.

**Proof of Lemma 2.6.9** If $I^V_\beta(\mu) < \infty$, since $V$ is bounded below by assumption (2.6.2), $\Sigma(\mu) > -\infty$ and therefore also $\int V \, d\mu < \infty$. This proves that $I^V_\beta(\mu)$ is well defined (and by definition non negative), yielding point a.

Set

$$f(x, y) = \frac{1}{2} V(x) + \frac{1}{2} V(y) - \beta \log |x - y|.$$  

(2.6.13)

Note that $f(x, y)$ goes to $+\infty$ when $x, y$ do by (2.6.2) since $\log |x - y| \leq \log(|x| + 1) + \log(|y| + 1)$ implies

$$f(x, y) \geq \frac{1}{2} (V(x) - \beta \log(|x| + 1)) + \frac{1}{2} (V(y) - \beta \log(|y| + 1)).$$

(2.6.14)

Further, $f(x, y)$ goes to $+\infty$ when $x, y$ approach the diagonal $\{x = y\}$. Therefore, for all $L > 0$, there exist constants $K(L)$ (going to infinity with $L$) such that, with $B_L := \{(x, y) : |x - y| < L^{-1}\} \cup \{(x, y) : |x| > L\} \cup \{(x, y) : |y| > L\},$

$$B_L \subset \{(x, y) : f(x, y) \geq K(L)\}.$$  

(2.6.15)

Since $f$ is continuous on the compact set $B_L^c$, we conclude that $f$ is bounded below on $\mathbb{R}^2$, and denote by $b_f > -\infty$ a lower bound. Therefore, since for
any measurable subset $A$ of $\mathbb{R}$,

$$I^Y_\beta(\mu) = \int \int (f(x, y) - b_f) d\mu(x) d\mu(y) + b_f - c^Y_\beta$$

$$\geq \int_A \int_A (f(x, y) - b_f) d\mu(x) d\mu(y) + b_f - c^Y_\beta$$

$$\geq \frac{\beta}{2} \int_A \int_A \log |x - y|^{-1} d\mu(x) d\mu(y) + \inf_{x \in \mathbb{R}} V(x) \mu(A)^2 - |b_f| - c^Y_\beta$$

$$\geq -\frac{\beta}{2} \mu(A)^2 \log(\gamma(A)) - |b_f| - c^Y_\beta + \inf_{x \in \mathbb{R}} V(x) \mu(A)^2$$

one concludes that if $I^Y_\beta(\mu) < \infty$, and $A$ is a measurable set with $\mu(A) > 0$, then $\gamma(A) > 0$. This completes the proof of point b.

We now show that $I^Y_\beta$ is a good rate function, and first that its level sets $\{I^Y_\beta \leq M\}$ are closed, that is that $I^Y_\beta$ is lower semi-continuous. Indeed, by the monotone convergence theorem,

$$I^V_\beta(\mu) = \int \int f(x, y) d\mu(x) d\mu(y) - c^Y_\beta$$

$$= \sup_{M \geq 0} \int \int (f(x, y) \wedge M) d\mu(x) d\mu(y) - c^Y_\beta$$

But $f^M = f \wedge M$ is bounded continuous and so for $M < \infty$,

$$I^{V,M}_\beta(\mu) = \int \int (f(x, y) \wedge M) d\mu(x) d\mu(y)$$

is bounded continuous on $M_1(\mathbb{R})$. As a supremum of the continuous functions $I^{V,M}_\beta$, $I^V_\beta$ is lower semi-continuous. Hence, by Theorem C.10, to prove that $\{I^Y_\beta \leq L\}$ is compact, it is enough to show that $\{I^Y_\beta \leq L\}$ is included in a compact subset of $M_1(\mathbb{R})$ of the form

$$K_\epsilon = \cap_{B \in \mathbb{N}} \{\mu \in M_1(\mathbb{R}) : \mu([-B, B]) \leq \epsilon(B)\}$$

with a sequence $\epsilon(B)$ going to zero as $B$ goes to infinity.

Arguing as in (2.6.15), there exist constants $K'(L)$ going to infinity as $L$ goes to infinity, such that

$$\{(x, y) : |x| > L, |y| > L\} \subset \{(x, y) : f(x, y) \geq K'(L)\}. \quad (2.6.16)$$
Hence, for any $L > 0$ large,
\[
\mu (|x| > L)^2 = \mu \otimes \mu (|x| > L, |y| > L) \leq \mu \otimes \mu (f(x, y) \geq K'(L)) \leq \frac{1}{K'(L) - b_f} \int \int (f(x, y) - b_f)d\mu(x)d\mu(y) = \frac{1}{K'(L) - b_f} (I_\beta^V (\mu) + c_V^\beta - b_f)
\]

Hence, with $\epsilon(B) = \sqrt{(M + c_V^\beta - b_f)}/\sqrt{(K'(B) - b_f)} \wedge 1$ going to zero when $B$ goes to infinity, one has that $\{I_\beta^V \leq M\} \subset K_\epsilon$. This completes the proof of point c.

Since $I_\beta^V$ is a good rate function, it achieves its minimal value. Let $\sigma_\beta^V$ be a minimizer. Then, for any signed measure $\nu(dx) = \phi(x)\sigma_\beta^V(dx) + \psi(x)dx$ with two bounded measurable compactly supported functions $(\phi, \psi)$ such that $\psi \geq 0$ and $\nu(\mathbb{R}) = 0$, for $\epsilon > 0$ small enough, $\sigma_\beta^V + \epsilon \nu$ is a probability measure so that
\[
I_\beta^V (\sigma_\beta^V + \epsilon \nu) \geq I_\beta^V (\sigma_\beta^V) \quad (2.6.17)
\]
which implies
\[
\int \left( V(x) - \beta \int \log |x - y|d\sigma_\beta^V(y) \right) d\nu(x) \geq 0.
\]

Taking $\psi = 0$, we deduce by symmetry that there is a constant $C_\beta^V$ such that
\[
V(x) - \beta \int \log |x - y|d\sigma_\beta^V(y) = C_\beta^V \quad \text{a.s.}, \quad (2.6.18)
\]
which implies that $\sigma_\beta^V$ is compactly supported (as $V(x) - \beta \int \log |x - y|d\sigma_\beta^V(y)$ goes to infinity when $x$ does). Taking $\phi(x) = -\int \psi(y)dy$, we then find that
\[
V(x) - \beta \int \log |x - y|d\sigma_\beta^V(y) \geq C_\beta^V \quad (2.6.19)
\]
Lebesgue almost surely, and then everywhere outside of the support of $\sigma_\beta^V$ by continuity. By (2.6.18) and (2.6.19) we deduce that
\[
C_\beta^V = \inf_{\nu \in M_1(\mathbb{R})} \left\{ \int (V(x) - \beta \int \log |x - y|d\sigma_\beta^V(y))d\nu(x) \right\}.
\]

This completes the proof of (2.6.10) and (2.6.11), except that the strict inequality in (2.6.11) will follow from the uniqueness of $\sigma_\beta^V$, by noting that
due to such uniqueness, the inequality (2.6.17) becomes a strict inequality as soon as \( \bar{\nu} \) is non-trivial.

The claimed uniqueness of \( \sigma_V^\beta \), and hence the completion of the proof of part e., will then follow from the strict convexity claim (point d. of the lemma), which we turn to next. Note first that we can rewrite \( I^V_\beta \) as

\[
I^V_\beta(\mu) = -\frac{\beta}{2}\Sigma(\mu - \sigma^V_\beta) + \int \left( V - \beta \int \log |x - y|d\sigma^V_\beta(y) - C^V_\beta \right) d\mu(x).
\]

The fact that \( I^V_\beta \) is strictly convex will follow as soon as we show that \( \Sigma \) is strictly concave. Toward this end, note the formula

\[
\log |x - y| = \int_0^\infty \frac{1}{2t} \left( \exp\left\{ -\frac{1}{2t} \right\} - \exp\left\{ -\frac{|x - y|^2}{2t} \right\} \right) dt,
\]

which follows from the equality

\[
1 = \frac{1}{2} \int_0^\infty e^{-u^2/2} du
\]

by the change of variables \( u \mapsto |x - y|^2/t \) and integration. Now, (2.6.20) implies that for any \( \mu \in M_1(\mathbb{R}) \),

\[
\Sigma(\mu - \sigma^V_\beta) = -\int_0^\infty \frac{1}{2t} \left( \int \exp\left\{ -\frac{|x - y|^2}{2t} \right\} d(\mu - \sigma^V_\beta)(x)d(\mu - \sigma^V_\beta)(y) \right) dt.
\]

Indeed, one may apply Fubini’s theorem when \( \mu, \sigma^V_\beta \) are supported in \([-\frac{1}{2}, \frac{1}{2}] \) since then \( \mu \otimes \sigma^V_\beta \left( \exp\left\{ -\frac{1}{2t} \right\} - \exp\left\{ -\frac{|x - y|^2}{2t} \right\} \right) \leq 0 \). One then deduces (2.6.21) for any compactly supported probability measure \( \mu \) by scaling and finally for all probability measures by approximations. The fact that for all \( t \geq 0 \),

\[
\int \int \exp\left\{ -\frac{|x - y|^2}{2t} \right\} d(\mu - \sigma^V_\beta)(x)d(\mu - \sigma^V_\beta)(y)
\]

\[
= \sqrt{\frac{1}{2\pi}} \int_{-\infty}^{+\infty} \left| \int \exp\{i\lambda x\}d(\mu - \sigma^V_\beta)(x) \right|^2 \exp\left\{ -\frac{t\lambda^2}{2} \right\} d\lambda
\]

therefore entails that \( \Sigma \) is concave since \( \mu \mapsto \int \exp\{i\lambda x\}d(\mu - \sigma^V_\beta)(x) \) is convex for all \( \lambda \in \mathbb{R} \). Strict convexity comes from the fact by the Cauchy-Schwarz inequality, \( \Sigma(\alpha \mu + (1 - \alpha)\nu) = \alpha \Sigma(\mu) + (1 - \alpha)\Sigma(\nu) \) if and only if \( \Sigma(\nu - \mu) = 0 \) which implies that all the Fourier transforms of \( \nu - \mu \) are null, and hence \( \mu = \nu \). This completes the proof of part d and hence of the lemma. \( \square \)
Proof of Theorem 2.6.6: To begin, let us remark that with \( f \) as in (2.6.13),

\[
P_{V,\beta}^N(d\lambda_1, \cdots, d\lambda_N) = (Z_{N}^{\beta,V})^{-1} e^{-N^2 \int_{x \neq y} f(x,y)dL_N(x)dL_N(y)} \prod_{i=1}^{N} e^{-V(\lambda_i)}d\lambda_i.
\]

Hence, if \( \mu \rightarrow \int_{x \neq y} f(x,y)d\mu(x)d\mu(y) \) was a bounded continuous function, the proof would follow from a standard application of Varadhan’s lemma, Theorem D.11. The main point will therefore be to overcome the singularity of this function, with the most delicate part being overcoming the singularity of the logarithm.

Following Appendix D (see Corollary D.9 and Definition D.6), a full large deviation principle can be proved by proving that exponential tightness holds, as well as estimating the probability of small balls. We follow these steps below.

- **Exponential tightness** Observe that by Jensen’s inequality,

\[
\log Z_{N}^{\beta,V} \geq N \log \int e^{-V(x)}dx
\]

\[
- N^2 \int \left( \int_{x \neq y} f(x,y)dL_N(x)dL_N(y) \right) \prod_{i=1}^{N} \frac{e^{-V(\lambda_i)}d\lambda_i}{\int e^{-V(x)}dx} \geq -CN^2
\]

with some finite constant \( C \). Moreover, by (2.6.14) and (2.6.2), there exist constants \( a > 0 \) and \( c > -\infty \) so that

\[
f(x,y) \geq a|V(x)| + a|V(y)| + c
\]

from which one concludes that for all \( M \geq 0 \),

\[
P_{V,\beta}^{\mathbb{M}} \left( \int |V(x)|dL_N \geq M \right) \leq e^{-2aN^2M+(C-c)N^2} \left( \int e^{-V(x)}dx \right)^N.
\]

Since \( V \) goes to infinity at infinity, \( K_M = \{ \mu \in \mathbb{M} \mathbb{R}^2 : \int |V|d\mu \leq M \} \) is a compact set for all \( M < \infty \), so that we have proved that the law of \( L_N \) under \( P_{V,\beta}^N \) is exponentially tight.

- **Large deviations upper bound** Recall that \( d \) denotes the Lipschitz bounded metric, see (C.9). We prove here that for any \( \mu \in \mathbb{M} \mathbb{R}^2 \), if we set \( \bar{P}_{V,\beta}^{N} = Z_{N}^{\beta,V} P_{V,\beta}^{N} \)

\[
\lim_{\epsilon \to 0} \limsup_{N \to \infty} \frac{1}{N^2} \log \bar{P}_{V,\beta}^{N} (d(L_N, \mu) \leq \epsilon) \leq - \int f(x,y)d\mu(x)d\mu(y). \tag{2.6.22}
\]
For any \( M \geq 0 \), set \( f_M(x, y) = f(x, y) \wedge M \). Then, the following bound holds

\[
\bar{P}_{V,\beta}^N (d(L_N, \mu) \leq \epsilon) \leq e^{-N^2 \int f_M(x,y) dL_N(x) dL_N(y)} \prod_{i=1}^N e^{-V(\lambda_i)} d\lambda_i.
\]

Since under the product Lebesgue measure, the \( \lambda_i \)'s are almost surely distinct, it holds that \( L_N \otimes L_N(x = y) = N^{-1} \), \( \bar{P}_{V,\beta}^N \) almost surely. Thus, we deduce that

\[
\int f_M(x, y) dL_N(x) dL_N(y) = \int_{x \neq y} f_M(x, y) dL_N(x) dL_N(y) + M N^{-1},
\]

and so

\[
\bar{P}_{V,\beta}^N (d(L_N, \mu) \leq \epsilon) \leq e^{MN} \int_{d(L_N, \mu) \leq \epsilon} e^{-N^2 \int f_M(x,y) dL_N(x) dL_N(y)} \prod_{i=1}^N e^{-V(\lambda_i)} d\lambda_i.
\]

Since \( f_M \) is bounded and continuous, \( I^{V, M}_\beta : \nu \mapsto \int f_M(x, y) d\nu(x) d\nu(y) \) is a continuous functional, and therefore we deduce that

\[
\lim_{\epsilon \to 0} \limsup_{N \to \infty} \frac{1}{N^2} \log \bar{P}_{V,\beta}^N (d(L_N, \mu) \leq \epsilon) \leq - I^{V, M}_\beta(\mu).
\]

We finally let \( M \) go to infinity and conclude by the monotone convergence theorem. Note that the same argument shows that

\[
\lim_{\epsilon \to 0} \limsup_{N \to \infty} \frac{1}{N^2} \log Z_N^{\beta, V} \leq - \inf_{\mu \in M_1(\mathbb{R})} \int f(x, y) d\mu(x) d\mu(y). \tag{2.6.23}
\]

\textbf{• Large deviations lower bound.} We prove here that for any \( \mu \in M_1(\mathbb{R}) \)

\[
\lim_{\epsilon \to 0} \inf_{N \to \infty} \frac{1}{N^2} \log \bar{P}_{V,\beta}^N (d(L_N, \mu) \leq \epsilon) \geq - \int f(x, y) d\mu(x) d\mu(y). \tag{2.6.24}
\]

Note that we can assume without loss of generality that \( I^V_{\beta}(\mu) < \infty \), since otherwise the bound in trivial, and so in particular, we may and will assume that \( \mu \) has no atoms. We can also assume that \( \mu \) is compactly supported since if we consider \( \mu_M = \mu([-M, M])^{-1} 1_{|x| \leq M} d\mu(x) \), clearly \( \mu_M \) converges towards \( \mu \) and by the monotone convergence theorem, one checks that, since \( f \) is bounded below,

\[
\lim_{M \to \infty} \int f(x, y) d\mu_M(x) d\mu_M(y) = \int f(x, y) d\mu(x) d\mu(y)
\]
which insures that it is enough to prove the lower bound for \((\mu, M, I) \in \mathbb{R}, I^1_N (\mu) < \infty\), and so for compactly supported probability measures with finite entropy.

The idea is to localize the eigenvalues \((\lambda_i)_{1 \leq i \leq N}\) in small sets and to take advantage of the fast speed \(N^2\) of the large deviations to neglect the small volume of these sets. To do so, we first remark that for any \(\nu \in M_1(\mathbb{R})\) with no atoms if we set

\[
x_i^{1,N} = \inf \left\{ x \mid \nu([-\infty, x]) \geq \frac{1}{N+1} \right\}
x_{i+1,N}^{1} = \inf \left\{ x \geq x_{i,N} \mid \nu([x_{i,N}, x]) \geq \frac{1}{N+1} \right\}
\]

for any real number \(\eta\), there exists an integer number \(N(\eta)\) such that, for any \(N\) larger than \(N(\eta)\),

\[
d\left( \nu, \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{i,N}} \right) < \eta.
\]

In particular, for \(N \geq N(\frac{\delta}{2})\),

\[
\left\{ (\lambda_i)_{1 \leq i \leq N} \mid |\lambda_i - x_{i,N}^{1,N}| < \frac{\delta}{2} \forall i \in [1, N] \right\} \subset \{ (\lambda_i)_{1 \leq i \leq N} \mid d(L_N, \nu) < \delta \}
\]

so that we have the lower bound

\[
P^N_{V,\beta} (d(L_N, \mu) \leq \epsilon) \geq \int_{\bigcap_{i \leq j} |\lambda_i - x_{i,N}^{1,N}| < \frac{\delta}{2}} e^{-N^2 \int_{\mathbb{R}^2} f(x,y) dL_N(x) dL_N(y)} \prod_{i=1}^{N} e^{-V(\lambda_i)} d\lambda_i
\]

\[
= \int_{\bigcap_{i \leq j} \{|\lambda_i| < \frac{\delta}{2}\}} \prod_{i \leq j} |x_i^{1,N} - x_j^{1,N} + \lambda_i - \lambda_j|^\beta e^{-N \sum_{i=1}^{N} V(\lambda_i)} \prod_{i=1}^{N} d\lambda_i
\]

\[
\geq \left( \prod_{i \leq j} |x_i^{1,N} - x_j^{1,N}|^\beta \prod_{i} |x_i^{1,N} - x_{i+1,N}|^\beta e^{-N \sum_{i=1}^{N} V(\lambda_i)} \right)^{\frac{\beta}{2}}
\]

\[
\times \left( \int_{\bigcap_{\lambda_1 < \lambda_i < \lambda_{i+1}}} \prod_{i} |\lambda_i - \lambda_{i+1}|^\beta e^{-N \sum_{i=1}^{N} [V(x_{i,N}^{1,N} + \lambda_i) - V(x_{i,N}^{1,N})]} \prod_{i=1}^{N} d\lambda_i \right)
\]

\[
=: P_{N,1} \times P_{N,2}
\]

(2.6.25)
where we used that \( |x_i^{i,N} - x_j^{i,N} + \lambda_i - \lambda_j| \geq |x_i^{i,N} - x_j^{i,N}| \vee |\lambda_i - \lambda_j| \) when \( \lambda_i \geq \lambda_j \) and \( x_i^{i,N} \geq x_j^{i,N} \). To estimate \( P_{N,2} \), note that since we assumed that \( \mu \) is compactly supported, the \((x_i^{i,N}, 1 \leq i \leq N)_{N \in \mathbb{N}}\) are uniformly bounded and so by continuity of \( V \)
\[
\lim_{N \to \infty} \sup_{N \in \mathbb{N}} \sup_{1 \leq i \leq N} |V(x_i^{i,N} + x) - V(x_i^{i,N})| = 0.
\]
Moreover, writing \( u_1 = \lambda_1, u_{i+1} = \lambda_{i+1} - \lambda_i, \)
\[
\int_{|\lambda_i| < \frac{\delta}{\lambda_j}} \prod_{i=1}^N |\lambda_i - \lambda_{i+1}| \prod_{i=1}^N d\lambda_i \geq \int_{0 < u_i < \frac{\delta}{\lambda_j}} \prod_{i=1}^N \frac{u_i^2}{\lambda_j} \prod_{i=1}^N du_i
\]
\[
\geq \left( \frac{\delta}{(\beta + 2)N} \right)^{N(\frac{\beta}{2} + 1)}.
\]
Therefore,
\[
\lim_{\delta \to 0} \lim_{N \to \infty} \frac{1}{N^2} \log P_{N,2} \geq 0. \quad (2.6.26)
\]
To handle the term \( P_{N,1} \), the uniform boundedness of the \( x_i^{i,N} \)'s and the convergence of their empirical measure towards \( \mu \) imply that
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N V(x_i^{i,N}) = \int V(x) d\mu(x). \quad (2.6.27)
\]
Finally since \( x \to \log(x) \) increases on \( \mathbb{R}^+ \), we notice that
\[
\int_{x_i^{i,N} \leq x < y \leq x^{i,N}} \log(y - x) d\mu(x) d\mu(y)
\]
\[
\leq \sum_{1 \leq i \leq j \leq N-1} \int_{x_i^{i+1,N} < x^{i,N}} \log(x^{i+1,N} - x^{i,N}) d\mu(x) d\mu(y)
\]
\[
= \frac{1}{(N+1)^2} \sum_{i<j} \log|x_i^{j,N} - x_j^{j+1,N}| + \frac{1}{2(N+1)^2} \sum_{i=1}^{N-1} \log|x_i^{i+1,N} - x_i^{i,N}|.
\]
Since \( \log |x - y| \) is bounded when \( x, y \) are in the support of the compactly supported measure \( \mu \), the monotone convergence theorem implies that the left side in the last display converges towards \( \int \int \log |x - y| d\mu(x) d\mu(y) \). Thus, with (2.6.27), we have proved
\[
\liminf_{N \to \infty} \frac{1}{N^2} \log P_{N,1} \geq \int_{x < y} \log(y - x) d\mu(x) d\mu(y) - \int V(x) d\mu(x)
\]
which concludes, with (2.6.25) and (2.6.26), the proof of (2.6.24).

- **Conclusion of the proof** By (2.6.24), for all $\mu \in M_1(\mathbb{R})$,
\[
\liminf_{N \to \infty} \frac{1}{N^2} \log Z_{\beta,V}^N \geq \lim_{\epsilon \to 0} \liminf_{N \to \infty} \frac{1}{N^2} \log P_{V,\beta}^N (d(L_N, \mu) \leq \epsilon)
\geq -\int f(x,y) d\mu(x) d\mu(y)
\]
and so optimizing with respect to $\mu \in M_1(\mathbb{R})$ and with (2.6.23),
\[
\lim_{N \to \infty} \frac{1}{N^2} \log Z_{\beta,V}^N = -\inf_{\mu \in M_1(\mathbb{R})} \left\{ \int f(x,y) d\mu(x) d\mu(y) \right\} = -c_{V,\beta}^V.
\]
Thus, (2.6.24) and (2.6.22) imply the weak large deviation principle, i.e. that for all $\mu \in M_1(\mathbb{R})$,
\[
\lim_{\epsilon \to 0} \liminf_{N \to \infty} \frac{1}{N^2} \log P_{V,\beta}^N (d(L_N, \mu) \leq \epsilon)
= \lim_{\epsilon \to 0} \limsup_{N \to \infty} \frac{1}{N^2} \log P_{V,\beta}^N (d(L_N, \mu) \leq \epsilon)
= -I_{V,\beta}(\mu).
\]
This, together with the exponential tightness property proved above, completes the proof of Theorem 2.6.6. $\Box$

### 2.6.2 Large deviations for the top eigenvalue

We consider next the large deviations for the maximum $\lambda_N^* = \max_{i=1}^N \lambda_i$, of random variables that possess the joint law (2.6.1). These will be obtained under the following assumption.

**Assumption 2.6.28** The normalization constants $Z_{V,\beta}^N$ satisfy
\[
\lim_{N \to \infty} \frac{1}{N} \log \frac{Z_{N\beta/V(N-1),\beta}^{N-1}}{Z_{V,\beta}^N} = \alpha_{V,\beta} \quad (2.6.29)
\]
and
\[
\limsup_{N \to \infty} \frac{1}{N} \log \frac{Z_{V,\beta}^{N-1}}{Z_{V,\beta}^N} < \infty. \quad (2.6.30)
\]
It is immediate from (2.5.15) that if $V(x) = \beta x^2/4$ then Assumption 2.6.28 holds, with $\alpha_{V,\beta} = -\beta/4$. Other situations where Assumption 2.6.28 is satisfied are explored in Section ??.

Assumption 2.6.28 is crucial in deriving the following LDP.
Theorem 2.6.31 Let \((\lambda_1^N, \ldots, \lambda_N^N)\) be distributed according to the joint law \(P_{V,\beta}^N\) of (2.6.1), with continuous potential \(V\) that satisfies (2.6.2) and Assumption 2.6.28. Let \(\sigma^N_Y\) be the maximizing measure of Lemma 2.6.9, and set \(\bar{x} = \max\{x : x \in \text{supp}\sigma^N_Y\}\). Then, \(\lambda_N^N = \max_{i=1}^N \lambda_i^N\) satisfies the LDP in \(\mathbb{R}\) with speed \(N\) and good rate function

\[
J_Y^V(x) = \begin{cases} 
\beta \int \log |x - y| \sigma^N_Y (dy) - V(x) - \alpha_{V,\beta}, & x \geq \bar{x} \\
\infty, & \text{else}
\end{cases}
\]

Proof of Theorem 2.6.31 Since \(J_Y^V(x)\) is continuous on \((\bar{x}, \infty)\) and \(J_Y^V(x) \to x \to \infty \infty\), it is a good rate function. Therefore, the stated LDP follows as soon as we show that for any \(x < \bar{x}\),

\[
\limsup_{N \to \infty} \frac{1}{N} \log P_{V,\beta}^N(\lambda_N^N \leq x) = -\infty, \tag{2.6.32}
\]

for any \(x > \bar{x}\),

\[
\limsup_{N \to \infty} \frac{1}{N} \log P_{V,\beta}^N(\lambda_N^N \geq x) \leq -J_Y^V(x), \tag{2.6.33}
\]

and

\[
\liminf_{\delta \to 0} \limsup_{N \to \infty} \frac{1}{N} \log P_{V,\beta}^N(\lambda_N^N \in (x - \delta, x + \delta)) \geq -J_Y^V(x). \tag{2.6.34}
\]

The limit (2.6.32) follows immediately from the LDP (at speed \(N^2\)) for the empirical measure, Theorem 2.6.6: indeed, the event \(\lambda_N^N \leq x\) implies that \(L_N((x, \bar{x}]) = 0\). Hence, one can find a bounded continuous function \(f\) with support in \((x, \bar{x}]\), independent of \(N\), such that \(\langle L_N, f \rangle = 0\) but \(\langle \sigma^N_Y, f \rangle > 0\). Theorem 2.6.6 implies that this event has probability that decays exponentially (at speed \(N^2\)), whence (2.6.32) follows.

The following lemma, whose proof is deferred, will allow for a proper truncation of the top eigenvalue. (The reader interested only in the GOE or GUE setups can note that Lemma 2.6.35 is then a consequence of Exercise 2.1.62.)

Lemma 2.6.35 Under the assumptions of Theorem 2.6.31, we have

\[
\lim_{M \to \infty} \limsup_{N \to \infty} \frac{1}{N} \log P_{V,\beta}^N(\lambda_N^N > M) = -\infty. \tag{2.6.36}
\]

Equipped with Lemma 2.6.35, we may complete the proof. We begin with the upper bound (2.6.33). Note that for any \(M > x\),

\[
P_{V,\beta}^N(\lambda_N^N \geq x) \leq P_{V,\beta}^N(\lambda_N^N > M) + P_{V,\beta}^N(\lambda_N^N \in [x, M]). \tag{2.6.37}
\]
By choosing $M$ large enough, the first term in the right side of (2.6.37) can be made smaller than $e^{-N\cdot J^\beta_N(x)}$, for all $N$ large. In the sequel, we fix an $M$ such that the above is satisfied and also

$$\left[ \beta \int \log |x-y| \sigma^\beta_N( dy) - V(x) \right] > \sup_{z \in [M, \infty)} \left[ \beta \int \log |z-y| \sigma^\beta_N( dy) - V(z) \right].$$

(2.6.38)

Set, for $z \in [-M,M]$ and $\mu$ supported on $[-M,M]$,

$$\Phi(z, \mu) = \beta \int \log |z-y| \mu( dy) - V(z) \leq \beta \log(2M) + V_- =: \Phi_M,$$

where $V_- = -\inf_{x \in \mathbb{R}} V(x) < \infty$. Setting $B(\delta)$ as the ball of radius $\delta$ around $\sigma^\beta_N$, $B_M(\delta)$ as those probability measures in $B(\delta)$ with support in $[-M,M]$, and writing

$$\zeta_N = \frac{Z_{N\cdot V/(N-1), \beta}^{N-1}}{Z_{V, \beta}^N}, \quad I_M = [-M, M]^{N-1},$$

we get

$$P_{V, \beta}(\lambda^*_N \in [x, M]) \leq N\zeta_N \int_x^M d\lambda_N \int_{I_M} e^{(N-1)\Phi(\lambda_N, L_{N-1})} P_{N\cdot V/(N-1), \beta}^{N-1}(d\lambda_1, \ldots, d\lambda_{N-1})$$

$$\leq N\zeta_N \int_x^M e^{(N-1)\sup_{\mu \in B_M(\delta)} \Phi(z, \mu)} dz + Me^{(N-1)\Phi_M} P_{N\cdot V/(N-1), \beta}^{N-1}(L_{N-1} \notin B(\delta)).$$

(2.6.39)

(The choice of metric in the definition of $B(\delta)$ plays no role in our argument, as long as it is compatible with weak convergence.) Noting that the perturbation involving the multiplication of $V$ by $N/(N-1)$ introduces only an exponential in $N$ factor, we get from the LDP for the empirical measure, Theorem 2.6.6, that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P_{N\cdot V/(N-1), \beta}^{N-1}(L_{N-1} \notin B(\delta)) < 0,$$

and hence, for any fixed $\delta > 0$,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P_{N\cdot V/(N-1), \beta}^{N-1}(L_{N-1} \notin B(\delta)) = -\infty.$$
We conclude from (2.6.39) and (2.6.40) that
\[ \limsup_{N \to \infty} \frac{1}{N} P_{V,\beta}^N(\lambda_N \in [x, M]) \]
\[ \leq \limsup_{N \to \infty} \frac{1}{N} \log \zeta_N + \lim_{\delta \to 0} \sup_{z \in [x, M], \mu \in B_{M}(\delta)} \Phi(z, \mu) \]
\[ = \alpha_{V,\beta} + \lim_{\delta \to 0} \sup_{z \in [x, M], \mu \in B_{M}(\delta)} \Phi(z, \mu). \]

Since \( \Phi(z, \mu) = \inf_{\eta > 0} [\log(|z-y|)\eta(dy) - V(z) \right) \), it holds that \((z, \mu) \mapsto \Phi(z, \mu)\) is upper semi-continuous on \([-M, M] \times M \times [-M, M]\). Therefore, using (2.6.38) in the last equality,
\[ \lim_{\delta \to 0} \sup_{z \in [x, M], \mu \in B_{M}(\delta)} \Phi(z, \mu) = \sup_{z \in [x, M]} \Phi(z, \sigma_{\beta}^{V}) = \Phi(z, \sigma_{\beta}^{V}). \]

Combining the last equality with (2.6.41) and (2.6.37), we obtain (2.6.33).

We finally prove the lower bound (2.6.34). Let \( 2 \delta < x - x^* \) and fix \( r \in (x^*, x - 2\delta) \). Then, with \( I_r = (-r, r)^{N-1} \),
\[ P_{V,\beta}^N(\lambda_N \in (x - \delta, x + \delta)) \]
\[ \geq P_{V,\beta}^N(\lambda_N \in (x - \delta, x + \delta), \max_{i=1}^{N-1} |\lambda_i| < r) \]
\[ = \zeta_N \int_{x-\delta}^{x+\delta} d\lambda_N \int_{I_r} e^{(N-1)\Phi(\lambda_N, L_{N-1}^{\lambda_N})} P_{N/(N-1),\beta}^{N-1}(d\lambda_1, \ldots, d\lambda_{N-1}) \]
\[ \geq 2\delta \zeta_N \exp \left( (N-1) \inf_{z \in (x-\delta, x+\delta), \mu \in B_\delta} \Phi(z, \mu) \right) \]
\[ P_{N/(N-1),\beta}^{N-1}(L_{N-1} \in B_\delta). \]

Recall from the upper bound (2.6.33), applied both to the function \( V(x) \) and \( V(-x) \), and the strict inequality in (2.6.11) of Lemma 2.6.9, that
\[ \limsup_{N \to \infty} P_{N/(N-1),\beta}^{N-1}(\max_{i=1}^{N-1} |\lambda_i| \geq r) = 0. \]

Combined with (2.6.40), we get by substituting in (2.6.42) that
\[ \lim_{\delta \to 0} \inf_{N \to \infty} \frac{1}{N} \log P_{V,\beta}^N(\lambda_N \in (x - \delta, x + \delta)) \geq \alpha_{V,\beta} + \lim_{\delta \to 0} \inf_{z \in (x-\delta, x+\delta), \mu \in B_\delta} \Phi(z, \mu) \]
\[ = \alpha_{V,\beta} + \Phi(x, \sigma_{\beta}^{V}), \]
where in the last step we used the continuity of \((z, \mu) \mapsto \Phi(z, \mu)\) on \([x - \delta, x + \delta] \times M \times [-r, r]\). The bound (2.6.34) follows. \( \square \)
Proof of Lemma 2.6.35 For $|x| > M$, $M$ large and $\lambda_i \in \mathbb{R}$, for some constants $a_\beta, b_\beta$,
\[ |x - \lambda_i|^{\beta} e^{-V(\lambda_i)} \leq a_\beta (|x|^{\beta} + |\lambda_i|^{\beta}) e^{-V(\lambda_i)} \leq b_\beta |x|^{\beta} e^{V(x)} \,.
\]
Therefore,
\[
P_{V,\beta}^N (\lambda_N^* > M) \\
\leq \frac{N Z_{V,\beta}^{N-1}}{Z_N^{V,\beta}} \int_{\zeta}^{\infty} e^{-N V(\lambda_N)} d\lambda_N \int_{\mathbb{R}}^{N-1} \prod_{i=1}^{N-1} (|x - \lambda_i|^{\beta} e^{-V(\lambda_i)}) dP_{V,\beta}^{N-1} \\
\leq \frac{N b_\beta^{N-1} e^{-N V(M)/2} Z_{V,\beta}^{N-1}}{Z_N^{V,\beta}} \int_{\zeta}^{\infty} e^{-V(\lambda_N)} d\lambda_N
\]
implying, together with (2.6.30), that
\[
\lim_{M \to \infty} \limsup_{N \to \infty} \frac{1}{N} \log P_{V,\beta}^N (\lambda_N^* > M) = -\infty \,.
\]

2.7 Bibliographical notes

Wigner’s theorem was presented in [Wig55], and proved there using the method of moments developed in Section 2.1. Since, this result was extended in many directions. In particular, under appropriate moment conditions, a.s. result hold true, see [Arn67] for an early result in that direction. Relaxation of moment conditions, requiring only the existence of third moments of the variables, is described by Bai and co-workers, using a mixture of combinatorial, probabilistic, and complex-analytic techniques. For a review, we refer to [Bai99]. It is important to note that one cannot hope to forgo the assumption of finiteness of second moments, see [?]. It is worthwhile to point out that the study of Wigner matrices is closely related to the study of Wishart matrices, discussed in Section ???. We refer to the bibliographical notes there for a discussion of this link.

The study of the distribution of the maximal eigenvalue of Wigner matrices by combinatorial techniques was initiated by [Juh81], and extended by [FK81] (whose treatment we essentially follow). See also [Gem80] for the analogous results for Wishart matrices. The method was recently amazingly extended in the papers [SS98a], [SS98b], [Sos99], allowing one to derive much finer behavior on the law of the largest eigenvalue(s), see the discussion in Section 3.7. Some extensions of the Füredi-Komlós and Sinai-Soshnikov techniques can be found in [Kho01].
The study of central limit theorems for traces of powers of random matrices can be traced back to [Jon82], in the context of Wishart matrices. Our presentation follows to a large extent his method, which allows one to derive a CLT for polynomial functions. A by-product of [SS98a] is a CLT for $\text{Tr} f(X_N)$ for analytic $f$, under a symmetry assumption on the moments. The paper [AZ05] greatly generalizes these results, allowing for differentiable functions $f$ and for non-constant variance of the independent entries. For functions of the form $f(x) = \sum a_i/(z_i - x)$ where $z_i \in \mathbb{C} \setminus \mathbb{R}$, and matrices of Wigner type, CLT statements can be found in [KKP96], with somewhat sketchy proofs. A complete treatment for $f$ analytic in a domain including the support of the limit of the empirical distribution of eigenvalues is given in [BY03] for matrices of Wigner type, and in [BS04] for matrices of Wishart type under a certain restriction on fourth moments. Much more is known for restricted classes of matrices: [Joh98] uses an approach based on the explicit joint density of the eigenvalues, see Section ??, and characterizes completely those functions $f$ for which a CLT holds. For Gaussian matrices, an approach based on the stochastic calculus introduced in Section ?? can be found in [CD01] and [Gui02]. Recent extensions and reinterpretation of this work, using the notion of second order freeness, see Section ??, can be found in [MS].

The study of spectra of random matrices via the Stieltjes transform (resolvent functions) was pioneered by Pastur co-workers, and greatly extended by Bai and co-workers. See [PM67] for an early reference. Our derivation is based Our development is based on [KKP96], [Bai99], and [SB95].

We presented in Section 2.3 a very brief introduction to concentration inequalities. This topic is picked up again in Section 4.4, to which we refer the reader for a complete introduction to different concentration inequalities and their application in RMT, and for bibliographical notes. Good references for the logarithmic Sobolev inequalities used in Section 2.3 are [Led01] and [ABC00]. Our treatment is based on [Led01] and [GZ00]. Lemma 2.3.3 is taken from [BL00, Proposition 3.1].

The basic results on joint distribution of eigenvalues in the GOE and GUE presented in Section 2.5 as well as an extensive list of integral formulas similar to (2.5.4) are given in [Meh91], [?]. We took however a quite different approach to all these topics based on the elementary proof of the Selberg integral formula [Sel44], see [?], given in [And91]. The proof of [And91] is based on a similar proof [And90] of some trigonometric sum identities. The proof of [And91] is similar also in spirit to the proofs in [Gus90] of much more elaborate identities. We note in passing that the Gamma function possesses a much richer theory than we needed (or exposed) here, in particular when viewed as a function over the right half of
the complex plane, see [?, ?, ?].

Theorem 2.6.6 is stated in [BG97, Theorem 5.2] under the additional assumption that $V$ does not grow faster than exponentially and proved there in details when $V(x) = x^2$. In [HP00b], the same result is obtained when the topology over $M_1(\mathbb{R})$ is taken to be the weak topology with respect to polynomial test functions instead of bounded continuous functions. Large deviations for the spectral measure of random matrices with complex eigenvalues were considered in [BZ98] (where non self-adjoint matrices with independent Gaussian entries were studied) and in [HP00a] (where Haar unitary distributed matrices are considered). This strategy can also be used when one is interested in discretized version of the law $P^{N,V}$ as they appear in the context of Young diagrams, see [GM05]. The LDP for the maximal eigenvalue described in Theorem 2.6.31 is based on [?].

We note that many of the results described in this chapter (except for Sections 2.3 and 2.6) can also be found in the book [Gir90], a translation of a 1975 Russian edition, albeit with somewhat sketchy and incomplete proofs.
Chapter 3

Orthogonal polynomials, spacings, and limit distributions for the GUE

In this chapter, we present the analysis of asymptotics for the joint eigenvalues distribution for the GUE. As it turns out, the analysis takes a particularly simple form for this ensemble, because the process of eigenvalues is a determinantal process. We postpone to Section 4.2 a discussion of general determinantal processes, opting to present here all computations “with bare hands”. In keeping with our goal of making this chapter accessible with minimal background, we postpone generalizations to other ensembles, and refinements, to Chapter 4.

3.1 Summary of main results: spacing distributions in the bulk and edge of the spectrum for the GUE

Here is a summary of the main results in this chapter. Note that the $N$ eigenvalues of the GUE are spread out on an interval of width roughly equal to $4\sqrt{N}$, and hence the spacing between adjacent eigenvalues is expected to be of order $1/\sqrt{N}$. Using the determinantal structure of the eigenvalues $\{\lambda_1^N, \ldots, \lambda_N^N\}$ of the GUE, developed in Sections 3.2-3.4, we prove the following.
Theorem 3.1.1 For any compact set \( A \subset \mathbb{R} \),
\[
\lim_{N \to \infty} P[\sqrt{N} \lambda_1^N, \ldots, \sqrt{N} \lambda_N^N \notin A] = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_A \cdots \int_A \det_{i,j=1}^k 1 + \frac{\sin(x_i-x_j)}{x_i-x_j} \prod_{j=1}^k dx_j.
\]

As a consequence of Theorem 3.1.1, we will show that the theory of integrable system applies and yields the following fundamental result concerning the behavior of spacings between eigenvalues in the bulk.

Theorem 3.1.3 One has
\[
\lim_{N \to \infty} P[\sqrt{N} \lambda_1^N, \ldots, \sqrt{N} \lambda_N^N \notin (-t/2, t/2)] = 1 - F(t),
\]
with
\[
F(x) = 1 - \exp \left( \int_0^x \frac{\sigma(t)}{t} dt \right) \quad \text{for } x \geq 0.
\]
with \( \sigma \) solution of
\[
(t\sigma'')^2 + 4(t\sigma' - \sigma)(t\sigma' - \sigma + (\sigma')^2) = 0.
\]
so that
\[
\sigma = -\frac{t}{\pi} - \frac{t^2}{\pi^2} - \frac{t^3}{\pi^3} + O(t^4) \quad \text{as } t \downarrow 0. \quad (3.1.4)
\]
The differential equation satisfied by \( \sigma \) is the Jimbo-Miwa-Okamoto form of Painlevé V. Note that Theorem 3.1.3 implies that \( F(t) \to t \to 0 0 \). Additional analysis (see Remark 3.6.33 in subsection 3.6.3) yields that also \( F(t) \to t \to \infty 1 \), showing that \( F \) is the distribution function of a probability distribution on \( \mathbb{R} \).

We then turn our attention to the edge of the spectrum.

Definition 3.1.5 The Airy function is defined by the formula
\[
\text{Ai}(x) = \frac{1}{2\pi i} \int_C e^{\zeta^3/3 - x\zeta} d\zeta \quad (3.1.6)
\]
where \( C \) is the contour in the \( \zeta \)-plane consisting of the ray joining \( e^{-\pi i/3}\infty \) to the origin plus the ray joining the origin to \( e^{\pi i/3}\infty \).
The Airy kernel is defined by
\[
A(x,y) := \frac{\text{Ai}(x) \text{Ai}'(y) - \text{Ai}'(x) \text{Ai}(y)}{x-y}.
\]
Various properties of the Airy function and kernel are summarized in subsection 3.7.1. The fundamental result concerning the eigenvalues of the GUE at the edge of the spectrum, due to Tracy and Widom, is the following.

**Theorem 3.1.7** For all \(-\infty < t \leq t' \leq \infty\),

\[
\lim_{N \to \infty} P \left[ \frac{N^{2/3}}{\sqrt{N}} \left( \frac{\lambda_i^N}{\sqrt{N}} - 2 \right) \notin [t, t'] \right] = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_t^{t'} \cdots \int_t^{t'} \det_{i,j=1}^k A(x_i, x_j) \prod_{j=1}^k dx_j
\]

with \(A\) the Airy kernel. In particular,

\[
\lim_{N \to \infty} P \left[ \frac{N^{2/3}}{\sqrt{N}} \left( \frac{\lambda_i^N}{\sqrt{N}} - 2 \right) \leq t \right] = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_t^{\infty} \cdots \int_t^{\infty} \det_{i,j=1}^k A(x_i, x_j) \prod_{j=1}^k dx_j =: F_2(t),
\]

The function \(F_2(\cdot)\) is the Tracy-Widom distribution. Note that the statement of Theorem 3.1.7 does not ensure that \(F_2(-\infty) = 0\), that is, it only implies the vague convergence, not the weak convergence, of the random variables \(\lambda_i^N / \sqrt{N} - 2\). The latter convergence, as well as a representation of the Tracy-Widom distribution \(F_2\), are contained in the following.

**Theorem 3.1.10** The function \(F_2(\cdot)\) is a distribution function that admits the representation

\[
F_2(t) = \exp \left( - \int_t^{\infty} (x-t)q(x)^2 dx \right),
\]

where \(q\) satisfies

\[
q'' = tq + 2q^3.
\]

Equation (3.1.12) is the Painlevé II equation.

### 3.2 Hermite polynomials and the GUE

In this section we review Hermite polynomials and harmonic oscillator wave-functions, and then derive a Fredholm determinant representation for certain probabilities connected with the GUE. Throughout, we use the notation \(\langle f, g \rangle_g = \int_R f(x)g(x)e^{-x^2/2}dx\).
3.2.1 The Hermite polynomials and harmonic oscillators

Definition 3.2.1 The \( n \)th Hermite polynomial \( H_n(x) \) is defined as

\[
H_n(x) := (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}.
\]

Lemma 3.2.2 The sequence of polynomials \( \{H_n(x)\}_{n=0}^{\infty} \) has the following properties.

1. \( H_0(x) = 1 \), \( H_1(x) = x \) and \( H_{n+1}(x) = xH_n(x) - H_n'(x) \).
2. \( H_n(x) \) is a polynomial of degree \( n \) with leading term \( x^n \).
3. \( H_n(x) \) is even or odd according as \( n \) is even or odd.
4. \( \langle x, H_n^2 \rangle_G = 0 \).
5. \( \langle H_k, H_k \rangle_G = \sqrt{2\pi} k! \delta_{kk} \).
6. \( \langle f, H_n \rangle_G = 0 \) for all polynomials \( f(x) \) of degree \( < n \).
7. \( xH_n(x) = H_{n+1}(x) + nH_n(x) \) for \( n \geq 1 \).
8. \( H_n'(x) = nH_{n-1}(x) \).
9. \( H_n''(x) - xH_n'(x) + nH_n(x) = 0 \).
10. For \( x \neq y \),

\[
\sum_{k=0}^{n-1} \frac{H_k(x)H_k(y)}{k!} = \frac{(H_n(x)H_{n-1}(y) - H_n(x)H_{n-1}(y))}{(n-1)!(x-y)}.
\]

Property 2 shows that \( \{H_n(x)\}_{n=0}^{\infty} \) is a basis of polynomial functions, whereas property 5 implies that it is an orthogonal basis for the scalar product \( \langle f, g \rangle_G \) defined on \( L^2(e^{-x^2/2}dx) \).

Remark 3.2.3 Properties 7 and 10 are the three-term recurrence and the Christoffel-Darboux identity satisfied by the Hermite polynomials, respectively. One has identities of this type in any system of polynomials orthogonal with respect to a given weight on the real line, see [?] for further details.
**Proof of Lemma 3.2.2:** Properties 1, 2 and 3 are clear. To prove property 5, use integration by parts to get that

\[
\int H_k(x)H_l(x)e^{-x^2/2}dx = (-1)^l \int H_k(x) \frac{d^l}{dx^l}(e^{-x^2/2})dx
\]

vanishes if \( l > k \) (since the degree of \( H_k \) is strictly less than \( l \)), and is equal to \( k! \) if \( k = l \), by property 2. Then, we deduce property 4 since by property 3, \( H_k^2 \) is an even function and so is the function \( e^{-x^2/2} \). Properties 2 and 5 suffice to prove property 6. To prove property 7, we proceed by induction on \( n \). By properties 2 and 5 we have, for \( n \geq 1 \),

\[
xH_n(x) = \sum_{k=0}^{n+1} \frac{\langle xH_n, H_k \rangle_G}{\langle H_k, H_k \rangle_G} H_k(x).
\]

By property 6 the \( k^{th} \) term on the right vanishes unless \( |k - n| \leq 1 \), by property 4 the \( n^{th} \) term vanishes, and by property 2 the \( (n+1)^{st} \) term equals 1. To get the \( (n-1)^{st} \) term we observe that

\[
\frac{\langle xH_n, H_{n-1} \rangle_G}{\langle H_{n-1}, H_{n-1} \rangle_G} = \frac{\langle xH_n, H_{n-1} \rangle_G}{\langle H_n, H_n \rangle_G} \cdot \frac{\langle H_n, H_n \rangle_G}{\langle H_{n-1}, H_{n-1} \rangle_G} = 1 \cdot n = n
\]

by induction on \( n \) and property 5. Thus property 7 is proved. Property 8 is a direct consequence of properties 1 and 7, and property 9 is obtained by differentiating the last identity in property 1 and using property 8. To prove property 10, call the left side of the claimed identity \( F(x,y) \) and the right side \( G(x,y) \). It is not hard to show that the integral

\[
\int \int e^{-x^2/2-y^2/2}H_k(x)H_l(y)F(x,y)(x-y)dxdy
\]

equals the analogous integral with \( G(x,y) \) replacing \( F(x,y) \); we leave the proof to the reader. Equality of these integrals granted, property 10 follows since \( \{H_k\}_{k \geq 0} \) being a basis of the set of polynomials, it implies almost sure equality and hence equality by continuity of \( F,G \). Thus the claimed properties of Hermite polynomials are proved. \( \Box \)

For the calculations to follow a “normalized” version of the Hermite polynomials is going to be more convenient.

**Definition 3.2.4** The \( n^{th} \) normalized harmonic oscillator wave-function is the function

\[
\psi_n(x) = \frac{e^{-x^2/4}H_n(x)}{\sqrt{\sqrt{2} \pi n!}},
\]
Basic properties of the harmonic oscillators are contained in the following easy corollary of Lemma 3.2.2.

Lemma 3.2.5
1. \( \int \psi_k(x)\psi_l(x)dx = \delta_{kl} \).
2. \( x\psi_n(x) = \sqrt{n+1}\psi_{n+1}(x) + \sqrt{n}\psi_{n-1}(x) \).
3. \( \sum_{k=0}^{n-1} \psi_k(x)\psi_k(y) = \sqrt{n}(\psi_n(x)\psi_{n-1}(y) - \psi_{n-1}(x)\psi_n(y))/(x - y) \).
4. \( \psi'_n(x) = -\frac{\sqrt{n}}{2}\psi_n(x) + \sqrt{n}\psi_{n-1}(x) \).
5. \( \psi''_n(x) + (n + \frac{1}{2} - \frac{x^2}{4})\psi_n(x) = 0 \).

We remark that the last relation above is the one-dimensional Schrödinger equation for the eigenstates of the one-dimensional quantum-mechanical harmonic oscillator. This explains the terminology.

3.2.2 Connection with the GUE

We now show that the joint distribution of the eigenvalues following the GUE has a nice determinantal form expressed in terms of the harmonic oscillators \( (\psi_k(x))_{k \geq 0} \), see Lemma 3.2.6 below. We use then this formula in order to deduce an Fredholm determinant expression for the probability that no eigenvalues belong to a given interval, see Lemma 3.2.15.

Throughout this section, we shall consider the eigenvalues of GUE matrices with complex Gaussian entries of variance identity as in Theorem 2.5.2, and normalize later the eigenvalues to study convergence issues. We shall be interested in symmetric statistics of the eigenvalues. For \( p \leq N \), recalling the joint distributions \( \mathcal{P}_N^{(2)} \) of the unordered eigenvalues of the GUE, c.f. Remark 2.5.6, we call its marginal \( \mathcal{P}_{p,N} \) on \( p \) coordinates the distribution of \( p \) unordered eigenvalues of the GUE. More explicitly, \( \mathcal{P}_{p,N}^{(2)} \) is the probability measure on \( \mathbb{R}^p \) so that for any \( f \in C_0(\mathbb{R}^p) \),

\[
\int f(\theta_1, \ldots, \theta_p) d\mathcal{P}_{p,N}^{(2)}(\theta_1, \ldots, \theta_p) = \int f(\theta_1, \ldots, \theta_p) d\mathcal{P}_N^{(2)}(\theta_1, \ldots, \theta_N)
\]

(recall that \( \mathcal{P}_{p,N}^{(2)} \) is the law of the unordered eigenvalues.) Clearly, one also has

\[
\int f(\theta_1, \ldots, \theta_p) d\mathcal{P}_{p,N}^{(2)}(\theta_1, \ldots, \theta_p) = \frac{(N - p)!}{N!} \sum_{\sigma \in S_{p,N}} \int f(\theta_{\sigma(1)}, \ldots, \theta_{\sigma(p)}) d\mathcal{P}_N^{(2)}(\theta_1, \ldots, \theta_N),
\]
where $S_{p,N}$ is the set of injective maps from $\{1, \cdots, p\}$ into $\{1, \cdots, N\}$. Note that we automatically have $P^{(2)}_{N,N} = P^{(2)}_N$.

**Lemma 3.2.6** For any $p \leq N$, the law $P^{(2)}_{p,N}$ is absolutely continuous with respect to Lebesgue measure, with density
\[
\rho_{p,N}^{(2)}(\theta_1, \cdots, \theta_p) = \frac{(N - p)!}{p!} \det_{k,l=1}^p K^{(N)}(\theta_k, \theta_l),
\]
where
\[
K^{(N)}(x, y) = \sum_{k=0}^{N-1} \psi_k(x)\psi_k(y) \tag{3.2.7}
\]

**Proof:** Theorem 2.5.2 shows that $\rho_{p,N}^{(2)}$ exists and equals, if $x_i = \theta_i$ for $i \leq p$ and $\zeta_i$ for $i > p$, to
\[
\rho_{p,N}^{(2)}(\theta_1, \cdots, \theta_p) = C_{N,p} \int \prod_{1 \leq i < j \leq N} |x_i - x_j|^2 \prod_{i=1}^p e^{-x_i^2/2} \prod_{i=p+1}^N d\zeta_i
\]
for some constant $C_{N,p}$. The fundamental remark is that this density depends on the Vandermonde determinant
\[
\prod_{1 \leq i < j \leq N} (x_j - x_i) = \prod_{i,j=1}^N x_i^{j-1} = \prod_{i,j=1}^N \delta_{j-1}(x_i)
\]
where we used in the last equality property 2 of Lemma 3.2.2.

We begin by considering $p = N$, writing $\rho_{N,N}$ for $\rho_{p,N}^{(2)}$. Then,
\[
\rho_{N}(\theta_1, \cdots, \theta_N) = C_{N,N} \left( \prod_{i,j=1}^N \delta_{j-1}(\theta_i) \right)^2 \prod_{i=1}^N e^{-\theta_i^2/2} \tag{3.2.8}
\]
\[
= \tilde{C}_{N,N} \left( \prod_{i,j=1}^N \psi_{j-1}(\theta_i) \right)^2 \prod_{i=1}^N K^{(N)}(\theta_i, \theta_j),
\]
where in the last line we used that $\det(AB) = \det(A) \det(B)$, that is the Cauchy-Binet Theorem A.3 with $m = n = k = r = N$, $A = B^* = (\psi_{j-1}(\theta_i))_{i,j=1}^N$. Here, $\tilde{C}_{N,N} = \prod_{k=0}^{N-1} (\sqrt{2\pi k!}) C_{N,N}$.

Of course, from (2.5.4) we know that $C_{N,N} = \hat{C}_N^{(2)}$. We provide now yet another direct evaluation of the normalization constant, following [Meh91]. We introduce a trick that will be very often applied in the sequel.
Lemma 3.2.9 For any square-integrable functions $f_1, \ldots, f_n$ and $g_1, \ldots, g_n$ on the real line, we have

\[
\frac{1}{n!} \int \cdots \int \frac{\det}{\det_i, j=1} \left( \sum_{k=1}^n f_k(x_i)g_k(x_j) \right) \prod_{i=1}^n dx_i = \frac{1}{n!} \int \cdots \int \frac{\det}{\det_i, j=1} f_i(x_j) \cdot \det_{i, j=1} g_i(x_j) \prod_{i=1}^n dx_i = \frac{1}{n!} \int f_i(x)g_i(x)dx.
\] (3.2.10)

Proof: Using the identity $\det(AB) = \det(A) \det(B)$ applied to $A = \{f_k(x_i)\}$ and $B = \{g_k(x_j)\}$, we get

\[
\int \cdots \int \frac{\det}{\det_i, j=1} \left( \sum_{k=1}^n f_k(x_i)g_k(x_j) \right) \prod_{i=1}^n dx_i = \int \cdots \int \frac{\det}{\det_i, j=1} f_i(x_j) \cdot \det_{i, j=1} g_i(x_j) \prod_{i=1}^n dx_i.
\]

Therefore, expanding the determinants we have

\[
\frac{1}{n!} \int \cdots \int \frac{\det}{\det_i, j=1} \left( \sum_{k=1}^n f_k(x_i)g_k(x_j) \right) \prod_{i=1}^n dx_i = \int \cdots \int \frac{\det}{\det_i, j=1} f_i(x_j) \prod_{j=1}^n g_j(x_j) \prod_{i=1}^n dx_i = \int \cdots \int \frac{\det}{\det_i, j=1} f_i(x_j) \prod_{i=1}^n g_i(x_i) dx_i = \frac{\det}{\det_i, j=1} f_i(x)g_j(x)dx.
\] (3.2.11)

\[\square\]

Substituting $f_i = g_i = \psi_{i-1}$ and $n = N$ in Lemma 3.2.9 we deduce that

\[
\int \frac{\det}{\det_{i, j=1}} K^{(N)}(\theta_i, \theta_j) \prod_{i=1}^N d\theta_i = N!
\] (3.2.12)

which completes the evaluation of $C_{N,N}$ and the proof of Lemma 3.2.6 for $p = N$.

For $p < N$, using (3.2.8) we find, with $x_i = \theta_i$ if $i \leq p$ and $\zeta_i$ otherwise,
that for some constant \( \tilde{C}_{N,p} \),

\[
p^{(2)}_{p,N}(\theta_1, \ldots, \theta_p) = \tilde{C}_{N,p} \int \left( \det_{i,j=1}^{N} \psi_{j-1}(x_i) \right)^2 \prod_{i=p+1}^{N} d\zeta_i
\]

\[
= \tilde{C}_{N,p} \sum_{\sigma,\tau \in S_N} \varepsilon(\sigma)\varepsilon(\tau) \int \prod_{j=1}^{N} \psi_{\tau(j)-1}(x_j) \prod_{i=p+1}^{N} d\zeta_i
\]

\[
= \tilde{C}_{N,p} \sum_{1 \leq \nu_1 < \ldots < \nu_p \leq N} \sum_{\sigma,\tau \in S(p,\nu)} \varepsilon(\sigma)\varepsilon(\tau) \prod_{i=1}^{p} \psi_{\tau(i)-1}(\theta_i) \prod_{i=p+1}^{N} d\zeta_i
\]

\[
= \tilde{C}_{N,p} \sum_{1 \leq \nu_1 < \ldots < \nu_p \leq N} \left( \prod_{i,j=1}^{p} \psi_{\nu_j-1}(\theta_i) \right)^2,
\]

(3.2.13)

where in the third line \( S(p,\nu) \) denotes those bijections \( \tau, \sigma \) of \( \{1, \ldots, p\} \) into \( \{\nu_1, \ldots, \nu_p\} \). This equality is due to the fact that the orthogonality of the \( \psi_j \) implies that the contribution comes only from permutations of \( S_N \) so that \( \tau(i) = \sigma(i) \) for \( i > p \) and we put \( \{\nu_1, \ldots, \nu_p\} = \{\tau(1), \ldots, \tau(p)\} = \{\sigma(1), \ldots, \sigma(p)\} \). Using the Cauchy-Binet Theorem A.3 with \( A = B^* \) and \( A_{i,j} = \psi_{\nu_j-1}(\theta_i) \), we get that

\[
\tilde{C}_{N,p} = \left( \frac{N-p}{N} \right)!.
\]

To compute \( \tilde{C}_{N,p} \), note that by integrating both sides of (3.2.13), we obtain

\[
1 = \tilde{C}_{N,p} \sum_{1 \leq \nu_1 < \ldots < \nu_p \leq N} \int \left( \prod_{i,j=1}^{p} \psi_{\nu_j-1}(\theta_i) \right)^2 d\theta_1 \cdots d\theta_p,
\]

(3.2.14)

whereas Lemma 3.2.9 implies that for all \( \{\nu_1, \ldots, \nu_p\} \),

\[
\int \left( \prod_{i,j=1}^{p} \psi_{\nu_j-1}(\theta_i) \right)^2 d\theta_1 \cdots d\theta_p = p!.
\]

Thus, since there are \( (N)!/(N-p)!p! \) terms in the sum at the right hand of (3.2.14), we conclude that \( \tilde{C}_{N,p} = (N-p)!/N! \). \( \Box \)

Now we arrive at the main point, on which the study of the local properties of the GUE will be based.

**Lemma 3.2.15** For any measurable subset \( A \) of \( \mathbb{R} \),

\[
P^{(2)}_N(\cap_{i=1}^{N} \{ \lambda_i \in A \}) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K^{(N)}(x_i, x_j) \prod_{i=1}^{k} dx_i.
\]

(3.2.16)
The last expression appearing in (3.2.16) is a Fredholm determinant.

**Proof.** By using in the first equality Lemmas 3.2.6 and 3.2.9, and using the orthogonality relations (point 1. of Lemma 3.2.5) in the second equality, we have

\[
P[\lambda_i \in A, i = 1, \ldots, N] = \frac{N}{N-1} \det_{i,j=0} A \int_A \psi_i(x)\psi_j(x)dx = \frac{N-1}{N-1} \left( \delta_{ij} - \int_{A^c} \psi_i(x)\psi_j(x)dx \right)
\]

\[
= 1 + \sum_{k=1}^N (-1)^k \sum_{0 \leq \nu_1 < \cdots < \nu_k \leq N-1} \det_{i,j=0} \left( \int_{A^c} \psi_\nu_i(x)\psi_\nu_j(x)dx \right),
\]

Therefore,

\[
P[\lambda_i \in A, i = 1, \ldots, N] = 1 + \sum_{k=1}^N \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \sum_{0 \leq \nu_1 < \cdots < \nu_k \leq N-1} \left( \det_{i,j=1} \psi_\nu_i(x_j) \right)^2 \prod_{i=1}^k dx_i
\]

\[
= 1 + \sum_{k=1}^N \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1} K^{(N)}(x_i, x_j) \prod_{i=1}^k dx_i
\]

\[
= 1 + \sum_{k=1}^\infty \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1} K^{(N)}(x_i, x_j) \prod_{i=1}^k dx_1,
\]

where the first equality uses (3.2.10) with \( g_i(x) = f_i(x) = \psi_\nu_i(x)1_{A^c}(x) \), the second uses the Cauchy-Binet Theorem A.3, and the last step is trivial since the determinant \( \det_{i,j=1} K^{(N)}(x_i, x_j) \) has to vanish identically for \( k > N \) because the rank of \( \{K^{(N)}(x_i, x_j)\}_{i,j=1}^k \) is at most \( N \).

\[\square\]

**3.3 The semicircle law revisited**

Let \( X_N \in \mathcal{H}_N^{(2)} \) be a random hermitian matrix from the GUE with eigenvalues \( \lambda_1^N \leq \cdots \leq \lambda_N^N \), and let the empirical distribution of the rescaled eigenvalues \( \lambda_i^N / \sqrt{N}, \ldots, \lambda_N^N / \sqrt{N} \) be denoted by

\[
L_N = (\delta_{\lambda_1^N / \sqrt{N}} + \cdots + \delta_{\lambda_N^N / \sqrt{N}}) / N.
\]

Let \( L_N \) thus corresponds to the eigenvalues of a Gaussian Wigner matrix.

We are going to make the average empirical distribution \( L_N \) explicit in terms of Hermite polynomials, calculate the moments of \( L_N \) explicitly, check that the moments of \( L_N \) converge to those of the semicircle law, and thus...
provide an alternative proof of Lemma 2.1.13. We also derive a recursion for the moments of \( \bar{L}_N \) due to Harer and Zagier by a method described by Haagerup and Thorbjørnsen, estimate by means of an observation of Ledoux the order of fluctuation of the renormalized maximum eigenvalue \( \lambda_N^2 / \sqrt{N} \) above the spectrum edge, an observation that will be useful in Section 3.7.

### 3.3.1 Calculation of moments of \( \bar{L}_N \)

By Lemma 3.2.6,

\[
\langle \bar{L}_N, \phi \rangle = \frac{1}{N} \int_{-\infty}^{\infty} \phi \left( \frac{x}{\sqrt{N}} \right) K^{(N)}(x, x) \, dx = \int_{-\infty}^{\infty} \phi(x) \frac{K^{(N)}(\sqrt{N}x, \sqrt{N}x)}{\sqrt{N}} \, dx.
\]

This last identity is enough to show that \( \bar{L}_N \) is absolutely continuous with respect to Lebesgue measure, with density

\[
K^{(n)}(x, x)/\sqrt{n} = \frac{\psi_n(x)\psi_{n-1}(y) - \psi_{n-1}(x)\psi_n(y)}{x - y},
\]

hence by L’Hôpital’s rule

\[
K^{(n)}(x, x)/\sqrt{n} = \psi''_n(x)\psi_{n-1}(x) - \psi'_{n-1}(x)\psi_n(x).
\]

Therefore,

\[
\frac{d}{dx} K^{(n)}(x, x)/\sqrt{n} = \psi''_n(x)\psi_{n-1}(x) - \psi'_{n-1}(x)\psi_n(x) = -\psi_n(x)\psi_{n-1}(x).
\]

By (3.3.2) the function \( K^{(N)}(\sqrt{N}x, \sqrt{N}x)/\sqrt{N} \) is the Radon-Nikodym derivative of \( \bar{L}_N \) with respect to Lebesgue measure and hence we have the following representation of the moment-generating function of \( \bar{L}_N \):

\[
\langle \bar{L}_N, e^{sx} \rangle = \frac{1}{N} \int_{-\infty}^{\infty} e^{sx/\sqrt{N}} K^{(N)}(x, x) \, dx.
\]

Integrating by parts once and then applying (3.3.3), we find that

\[
\langle \bar{L}_N, e^{sx} \rangle = \frac{1}{8} \int_{-\infty}^{\infty} e^{sx/\sqrt{N}} \psi_N(x)\psi_{N-1}(x) \, dx.
\]

Thus the calculation of the moment generating function of \( \bar{L}_N \) boils down to the problem of evaluating the integral on the right.
By Taylor’s theorem it follows from Lemma 3.2.2, part 8, that for any \( n \),
\[
\mathcal{H}_n(x + t) = \sum_{k=0}^{n} \binom{n}{k} \mathcal{H}_{n-k}(x) t^k = \sum_{k=0}^{n} \binom{n}{k} \mathcal{H}_k(x) t^{n-k}.
\]

Let \( S^n_t := \int_{-\infty}^{\infty} e^{tx} \psi_n(x) \psi_{n-1}(x) dx \). By the preceding identity and orthogonality we have
\[
S^n_t = \frac{\sqrt{n}}{n! \sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{H}_n(x) \mathcal{H}_{n-1}(x) e^{-x^2/2 + tx} dx = \frac{\sqrt{n} e^{t^2/2}}{n! \sqrt{2\pi}} \int_{-\infty}^{\infty} \mathcal{H}_n(x + t) \mathcal{H}_{n-1}(x + t) e^{-x^2/2} dx
\]
\[
= e^{t^2/2} \sqrt{\pi} \sum_{k=0}^{n-1} \frac{k!}{n!} \binom{n-1}{k} \binom{n-1}{n-1-k} t^{2n-1-2k}
\]
\[
= e^{t^2/2} \sqrt{\pi} \sum_{k=0}^{n-1} \frac{(n-1-k)!}{n!} \binom{n-1}{n-1-k} \binom{n-1}{k+1} t^{2k+1}
\]
\[
= e^{t^2/2} \sqrt{\pi} \sum_{k=0}^{n-1} \binom{n-1}{k+1} \frac{(n-1-k)!}{n!} t^{2k+1}.
\]

From the last calculation combined with (3.3.5) and after a further bit of rearrangement we obtain the following result:
\[
\langle \mathcal{L}_N, e^{s} \rangle = e^{s^2/(2N)} \sum_{k=0}^{N-1} \frac{1}{k+1} \binom{2k}{k} \frac{(N-1) \cdots (N-k)}{N^k} \frac{s^{2k}}{(2k)!}
\]
(3.3.6)

We can now present another:
**Proof of Lemma 2.1.13 - Gaussian Wigner matrices.** We have written the moment generating function in the form (3.3.6), making it obvious that as \( N \to \infty \) the moments of \( \mathcal{L}_N \) tend to the moments of the semicircle distribution. \( \square \)

### 3.3.2 The Harer-Zagier recursion and Ledoux’s argument

Our goal in this section is to provide the proof of the following lemma.

**Lemma 3.3.7 (Ledoux’s bound)** There exist positive constants \( c' \) and \( C' \) such that
\[
P \left( \frac{\lambda_N^N}{2\sqrt{N}} \geq e^{-N^{2/3} \epsilon} \right) \leq C' e^{-c' \epsilon}
\]
(3.3.8)

for all \( N \geq 1 \) and \( \epsilon > 0 \).
Roughly speaking the last inequality says that fluctuations of $\frac{\lambda N}{2N^2} - 1$ above 0 are of order of magnitude $N^{-2/3}$. This is an a-priori indication that the random variables $N^{2/3}\left(\frac{\lambda N}{2N^2} - 1\right)$ converge in distribution, as stated in Theorems 3.1.7 and 3.1.10. In fact, (3.3.8) is going to play a role in the proof of Theorem 3.1.7, see subsection 3.7.2.

The proof of Lemma 3.3.7 is based on a recursion satisfied by the moments of $\bar{L}_N$. We thus first introduce this recursion in Lemma 3.3.9 below, prove it, and then show how to deduce from it Lemma 3.3.7.

Write

$$\langle \bar{L}_N, e^s \rangle = \sum_{k=0}^{\infty} \frac{b_k^{(N)}}{k+1} \binom{2k}{k} \frac{s^{2k}}{(2k)!}.$$  

**Lemma 3.3.9 (Harer-Zagier recursions)** For any integer numbers $k$ and $N$,

$$b_{k+1}^{(N)} = b_k^{(N)} + \frac{k(k+1)}{4N^2} b_{k-1}^{(N)}$$

(3.3.10)

where if $k = 0$ we ignore the last term.

**Proof of Lemma 3.3.9**

Define the function

$$F_n(t) = F\left(\frac{1-n}{2}, t\right) := \sum_{k=1}^{\infty} \frac{(-1)^k}{(k+1)!} \binom{n-1}{k} t^k,$$

(3.3.11)

and note that

$$\left(t\frac{d^2}{dt^2} + (2-t)\frac{d}{dt} + (n-1)\right) F_n(t) = 0.$$  

(3.3.12)

By rearranging (3.3.6) it follows that

$$\langle \bar{L}_N, e^s \rangle = \Phi_N \left( -\frac{s^2}{N} \right)$$  

(3.3.13)

where

$$\Phi_n(t) = e^{-t/2} F_n(t).$$

From (3.3.12) it follows that

$$\left(4t\frac{d^2}{dt^2} + 8\frac{d}{dt} + 4n - t\right) \Phi_n(t) = 0.$$  

(3.3.14)
Write next $\Phi_n(t) = \sum_{k=0}^{\infty} a_k^{(n)} t^k$. By (3.3.14) we have

$$0 = 4(k+2)(k+1)a_{k+1}^{(n)} + 4na_k^{(n)} - a_k^{(n)}$$

where if $k = 0$ we ignore the last term. Clearly we have, taking $n = N$,

$$\frac{(-1)^k a_k^{(N)}(2k)!}{N^k} = \frac{b_k^{(N)}}{k+1} \binom{2k}{k} = \langle L_N, x^{2k} \rangle.$$

The lemma follows. \qed

**Proof of Lemma 3.3.7** From (3.3.10) and the definitions we obtain the inequalities

$$0 \leq b_k^{(N)} \leq b_{k+1}^{(N)} \leq \left( 1 + \frac{k(k+1)}{4N^2} \right) b_k^{(N)}$$

for $N \geq 1$ and $k \geq 0$. By Stirling’s formula we have

$$\sup_{k=0}^{\infty} \frac{k^{3/2}}{2^{2k}(k+1)} \binom{2k}{k} < \infty.$$

It follows that there exist positive constants $c$ and $C$ such that

$$P \left( \frac{\lambda_N}{2\sqrt{N}} \geq \epsilon t \right) \leq E \left( \frac{\lambda_N}{2\sqrt{N}c} \right)^{2k} \leq e^{-2\epsilon t N b_k^{(N)}} \frac{2k}{k} \leq CNt^{-3/2}e^{-2\epsilon t + ct^3/N^2}$$

for all $N \geq 1$, $k \geq 0$ and real numbers $\epsilon, t > 0$ such that $k = \lfloor t \rfloor$. Taking $t = N^{2/3}$ and substituting $N^{-2/3} \epsilon$ for $\epsilon$ yields the lemma. \qed

**Exercise 3.3.16** Prove that in the setup of this section, for every integer $k$ it holds that

$$\lim_{N \to \infty} E\langle L_N, x^k \rangle^2 = \lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle^2.$$ 

Using that the moments of $L_N$ converge to the moments of the semicircle distribution, complete yet another proof of Wigner’s Theorem 2.1.4 in the Gaussian complex case.

**Hint:** Deduce from (3.3.2) that

$$\langle \bar{L}_N, x^k \rangle = \frac{1}{N^{k/2+1}} \int x^k K^{(N)}(x, x)dx.$$
Also, rewrite $E(L_N, x^k)^2$ as

$$\frac{1}{N^{2+k}N!} \int \cdots \int \left( \sum_{i=1}^{N} x_i^k \right)^2 \det_{i,j=1}^{N} K^{(N)}(x_i, x_j) \prod_{j=1}^{N} dx_j$$

$$= \frac{(N-1)!}{N^{k+1}N!} \int x^{2k} K^{(N)}(x, x) dx$$

$$+ \frac{(N-1)(N-2)!}{N^{k+1}N!} \int x^k y^k K^{(N)} \left( \begin{array}{cc} x & y \\ x & y \end{array} \right) dxdy$$

$$= \frac{1}{N^{k+2}} \int x^{2k} K^{(N)}(x, y)^2 dxdy + \frac{1}{N^{k+2}} \left( \int x^k K^{(N)}(x, x) dx \right)^2$$

$$= \langle \bar{L}_N, x^k \rangle^2 + I^{(N)}$$

where $I^{(N)}$ is equal to

$$\frac{1}{N^{k+3/2}} \int \int \frac{x^{2k} - x^k y^k}{x-y} (\psi_N(x)\psi_{N-1}(y) - \psi_{N-1}(x)\psi_N(y)) K(x, y) dxdy.$$
(i) Verify the following generalization of (3.3.12):
\[
\frac{d}{dt} \left( t \frac{d}{dt} + b_1 - 1 \right) \cdots \left( t \frac{d}{dt} + b_q - 1 \right) F \left( \begin{array}{c} a_1 \cdots a_p \\ b_1 \cdots b_q \end{array} \middle| t \right) \\
= \left( t \frac{d}{dt} + a_1 \right) \cdots \left( t \frac{d}{dt} + a_p \right) F \left( \begin{array}{c} a_1 \cdots a_p \\ b_1 \cdots b_q \end{array} \middle| t \right).
\]

(ii) (Proposed by D. Stanton) Check that \( F_n(t) \) in (3.3.11) is a Laguerre polynomial.

3.4 Quick introduction to Fredholm determinants

In this section we review key definitions and facts concerning Fredholm determinants. We make no attempt to achieve great generality. In particular we do not touch here on any functional analytic aspects of the theory of Fredholm determinants. The reader interested only in the proof of Theorem 3.1.1 may skip subsection 3.4.2 in a first reading.

3.4.1 The setting, fundamental estimates, and definition of the Fredholm determinant

We fix a continuous nonnegative function \( A \) on the real line such that
\[
\int_{A > 0} A(x)^{-2} dx < \infty.
\]
We fix a bounded complex-valued Lebesgue-measurable function \( \lambda' \) on the real line such that \( \lambda' \neq 0 \subset \{ A > 0 \} \). For all Lebesgue-integrable functions \( f \) on the real line we write
\[
\int f(x) d\lambda(x) = \int_{A > 0} f(x)\lambda'(x) dx.
\]

**Definition 3.4.1** A kernel is a continuous complex-valued function \( K(x,y) \) defined on the set
\[
\{ A > 0 \} \times \{ A > 0 \} \subset \mathbb{R}^2
\]
such that
\[
\| K \|_A := \sup_{(x,y) \in \{ A > 0 \} \times \{ A > 0 \}} A(x)A(y) |K(x,y)| < \infty.
\]
The trace of a kernel \( K(x,y) \) is
\[
\text{Tr}(K) = \int K(x,x) d\lambda(x). \tag{3.4.2}
\]
Given two kernels \( K(x, y) \) and \( L(x, y) \), define their composition as
\[
(K \ast L)(x, y) = \int K(x, z)L(z, y)d\lambda(z). \tag{3.4.3}
\]
Note that the trace in (3.4.2) is well defined because the integrand is dominated by \( \|K\|_A/\lambda^2(x) \). Similarly, since the integrand in (3.4.3) is dominated by \( \|K\|_A\|L\|_A^{-1}\lambda^{-2}A^{-1}A^{-1}\), it follows that \( K \ast L \) is well-defined on \( \{A > 0\}^2 \) and satisfies the conditions for being a kernel. Note that by Fubini’s theorem, for any three kernels \( K, L \) and \( M \), we have
\[
\text{Tr}(K \ast L) = \text{Tr}(L \ast K) \quad \text{and} \quad (K \ast L) \ast M = K \ast (L \ast M).
\]
We turn next to the evaluation of some basic estimates.

**Lemma 3.4.4** Fix \( n > 0 \).

1. For any kernels \( K_1(x, y), \ldots, K_n(x, y) \) we have a bound
\[
\left| \det_{i,j=1}^n K_i(x_i, y_j) \right| \leq \frac{n^{n/2} \prod_{i=1}^n \|K_i\|_A \prod_{i=1}^n A(x_i) \cdot \prod_{j=1}^n A(y_j)}{\prod_{i=1}^n \|K_i\|_A^{-1} A(x_i)} \cdot \det[v_1 \ldots v_N] \tag{3.4.5}
\]
2. For any two kernels \( F(x, y) \) and \( G(x, y) \) we have
\[
\left| \det_{i,j=1}^n F(x_i, y_j) - \det_{i,j=1}^n G(x_i, y_j) \right| \leq \frac{n^{1+n/2} \|F - G\|_A \max(\|F\|_A, \|G\|_A)^{n-1} \prod_{i=1}^n A(x_i) \cdot \prod_{j=1}^n A(y_j)}{\prod_{i=1}^n A(x_i) \cdot \prod_{j=1}^n A(y_j)} \tag{3.4.6}
\]
The factor \( n^{n/2} \) in (3.4.5) comes from Hadamard’s inequality (Theorem A.5). In view of Stirling’s formula
\[
\lim_{n \to \infty} \sqrt{2\pi n} n^{n+1/2} e^{-n}/n! = 1, \tag{3.4.7}
\]
it is clear that the Hadamard bound is much better than the bound \( n! \) we would get just by counting terms. The Hadamard bound is ultimately what gives the “edge” to the Fredholm theory.

**Proof:** To prove (3.4.5), we take \( v_i(j) = K_i(x_i, y_j)A(x_i)A(y_j)\|K_i\|_A^{-1}, 1 \leq i, j \leq N, \) in Hadamard’s inequality (Theorem A.5) to get
\[
\left| \det_{i,j=1}^n K_i(x_i, y_j) \right| = \prod_{i=1}^n \frac{\|K_i\|_A}{A(x_i)A(y_i)} \|\det[v_1 \ldots v_N]\|
\leq \prod_{i=1}^n \frac{\|K_i\|_A}{A(x_i)A(y_i)} \prod_{i=1}^n \sqrt{\|v_i\|v_i} \leq \prod_{i=1}^n \frac{\|K_i\|_A}{A(x_i)A(y_i)} n^{n/2},
\]
where we ultimately used that $|v_i(j)| \leq 1$ for all $i, j \in \{1, \ldots, n\}$. In particular, taking $K_i = K$, we have

$$\left| \prod_{i,j=1}^n K(x_i, y_j) \right| \leq \frac{n^{n/2} \|K\|_A^n}{\prod_{i=1}^n A(x_i) \cdot \prod_{j=1}^n A(y_j)} \quad (3.4.8)$$

To prove (3.4.6), define

$$H^{(k)}_i(x, y) = \begin{cases} G(x, y) & \text{if } i < k, \\ F(x, y) - G(x, y) & \text{if } i = k, \\ F(x, y) & \text{if } i > k, \end{cases}$$

noting that

$$\prod_{i,j=1}^n F(x_i, y_j) - \prod_{i,j=1}^n G(x_i, y_j) = \sum_{k=1}^n \prod_{i,j=1}^n H^{(k)}_i(x_i, y_j).$$

Now apply estimate (3.4.5) to the right side of the equation above, completing the proof of (3.4.6).

We are now finally ready to define the Fredholm determinant associated to a kernel $K(x, y)$. Set

$$K \begin{pmatrix} x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \end{pmatrix} = \prod_{i,j=1}^n K(x_i, y_j). \quad (3.4.9)$$

For $n > 0$, put

$$\Delta_n = \Delta_n(K) = \int \cdots \int K \begin{pmatrix} \xi_1 & \cdots & \xi_n \\ \xi_1 & \cdots & \xi_n \end{pmatrix} d\lambda(\xi_1) \cdots d\lambda(\xi_n),$$

setting $\Delta_0 = \Delta_0(K) = 1$. We have

$$\int \cdots \int K \begin{pmatrix} \xi_1 & \cdots & \xi_n \\ \xi_1 & \cdots & \xi_n \end{pmatrix} d\lambda(\xi_1) \cdots d\lambda(\xi_n) \leq C^n \|K\|_A^n n^{n/2}, \quad (3.4.10)$$

where

$$C = \text{(essential supremum of } |\lambda'| \text{)} \cdot \int_{A > 0} A^{-2}(x) dx.$$

So, $\Delta_n$ is well-defined.

**Definition 3.4.11** The Fredholm determinant associated with the kernel $K$ is defined as

$$\Delta = \Delta(K) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \Delta_n.$$
In view of Stirling’s formula (3.4.7) and estimate (3.4.10), the series in Definition 3.4.11 converges absolutely (and quite rapidly), and so $\Delta$ is well-defined.

**Remark 3.4.12** Here’s some motivation for calling $\Delta$ a determinant. Let $f_1(x), \ldots, f_N(x), g_1(x), \ldots, g_N(x)$ be continuous functions on the real line. Put

$$K(x, y) = \sum_{i=1}^{N} f_i(x) g_i(y).$$

Assume further that

$$\max_i \sup_x f_i(x) A(x) < \infty, \quad \max_j \sup_y g_j(y) A(y) < \infty.$$

Then $K(x, y)$ is a kernel and so fits into the theory developed thus far. We claim that

$$\Delta(K) = \det_{i,j=1}^{N} \left( \delta_{ij} - \int f_i(x) g_j(x) d\lambda(x) \right) \tag{3.4.13}$$

which can be proved as in Lemma 3.2.15.

The determinants $\Delta$ inherit good continuity properties with respect to the $\| \cdot \|_A$ norm.

**Lemma 3.4.14** For any two kernels $K(x, y)$ and $L(x, y)$ we have

$$|\Delta(K) - \Delta(L)| \leq \left( \sum_{n=1}^{\infty} n^{1+n/2} C^n \frac{\max(\|K\|_A, \|L\|_A)^{n-1}}{n!} \right) \cdot \|K - L\|_A, \tag{3.4.15}$$

with a finite constant $C$ which depends only on $A$ and $\lambda$.

**Proof:** By (3.4.6), we get the estimate

$$|\Delta_n(K) - \Delta_n(L)| \leq n^{1+n/2} C^n \max(\|K\|_A, \|L\|_A)^{n-1}$$

where $C$ is as in (3.4.10). From here, (3.4.15) is derived by summation. □

In particular, with $K$ held fixed, and with $L$ varying in such a way that $\|K - L\|_A \to 0$, it follows that $\Delta(L) \to \Delta(K)$. This is the only thing we shall need to obtain the convergence in law of the spacing distribution of the eigenvalues of the GUE, Theorem 3.1.1. On the other hand, the next subsections will be useful in the proof of Theorem 3.1.3.
3.4.2 Definition of the Fredholm adjugants, Fredholm resolvents, and a fundamental identity

Throughout, we fix a kernel \( K(x, y) \) and put \( \Delta = \Delta(K) \). All the constructions under this heading depend on \( K(x, y) \) but we suppress reference to this dependence in the notation in order to control clutter. Put, for \( p \geq 0 \) integer,

\[
H_n \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right) = \int \cdots \int K \left( \begin{array}{cccc} x_1 & \cdots & x_p & \xi_1 & \cdots & \xi_n \\ y_1 & \cdots & y_p & \xi_1 & \cdots & \xi_n \end{array} \right) d\lambda(\xi_1) \cdots d\lambda(\xi_n)
\]

for \( n > 0 \) and

\[
H_0 \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right) = K \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right).
\]

We then have

\[
\left| H_n \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right) \right| \leq \frac{\|K\|^{n+p} C^n (u+p)^{(n+p)/2}}{\prod_{i=1}^p A(x_i) \cdot \prod_{j=1}^p A(y_j)}.
\]

(3.4.17)

where \( C \) is as in (3.4.10). Continuity of the function \( H_n(\cdot) \) on \( \{ A > 0 \}^{2p} \) can be verified by a straightforward application of dominated convergence.

**Definition 3.4.18** The higher Fredholm adjugants of the kernel \( K(x, y) \) are the functions

\[
H \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} H_n \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right).
\]

(3.4.19)

If \( \Delta(K) \neq 0 \) we define the higher resolvents of the kernel \( K(x, y) \) as the functions

\[
R \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right) = H \left( \begin{array}{cccc} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{array} \right) / \Delta.
\]

(3.4.20)

Finally, the Fredholm adjugant \( H(x, y) \) and the Fredholm resolvent \( R(x, y) \) of \( K(x, y) \) are defined as

\[
H(x, y) = H \left( \begin{array}{cccc} x \\ y \end{array} \right), \quad R(x, y) = R \left( \begin{array}{cccc} x \\ y \end{array} \right).
\]
By (3.4.17), the series in (3.4.19) converges absolutely and uniformly on compact subsets of \( \{ A > 0 \}^{2p} \). Therefore \( H(\cdot) \) is well-defined and continuous on \( \{ A > 0 \}^{2p} \). The main fact to bear in mind as we proceed is that

\[
\sup A(x_1) \cdots A(x_p)A(y_1) \cdots A(y_p) \left| F \left( \begin{array}{c} x_1 \cdots x_p \\ y_1 \cdots y_p \end{array} \right) \right| < \infty \tag{3.4.21}
\]

for \( F = K, H, R \). These bounds are sufficient to guarantee the absolute convergence of all the integrals we will encounter in the remainder of Section 3.4. Also it bears emphasizing that the two-variable functions \( H(x, y) \) (resp., \( R(x, y) \) if defined) are kernels.

The following lemma, whose proof is postponed to the end of this section, shows that all the higher order Fredholm resolvents \( R \) possess a determinantal representation.

**Lemma 3.4.22** Let \( K \) be a kernel with \( \Delta(K) \neq 0 \) so that all the higher resolvents \( R(\cdot) \) associated to \( K(x, y) \) are defined. Then,

\[
R \left( \begin{array}{c} x_1 \cdots x_p \\ y_1 \cdots y_p \end{array} \right) = \det_{i,j=1}^p R(x_i, y_j). \tag{3.4.23}
\]

We next prove a fundamental identity relating the Fredholm adjugant and determinant associated with a kernel \( K \).

**Lemma 3.4.24 (The fundamental identity)** Let \( H(x, y) \) the Fredholm adjugant of the kernel \( K(x, y) \). Then,

\[
\int K(x, z)H(z, y)d\lambda(z) = H(x, y) - \Delta \cdot K(x, y) = \int H(x, z)K(z, y)d\lambda(z), \tag{3.4.25}
\]

and hence (equivalently)

\[
K \ast H = H - \Delta \cdot K = H \ast K. \tag{3.4.26}
\]

**Remark 3.4.27** Before proving the fundamental identity (3.4.26), we make some amplifying remarks. If \( \Delta \neq 0 \) and hence the resolvent \( R(x, y) = H(x, y)/\Delta \) of \( K(x, y) \) is defined, then the fundamental identity takes the form

\[
\int K(x, z)R(z, y)d\lambda(z) = R(x, y) - K(x, y) = \int R(x, z)K(z, y)d\lambda(z)
\]
and hence (equivalently)

\[ K \ast R = R - K = R \ast K. \]

It is helpful if not perfectly rigorous to rewrite the last formula as the operator identity

\[ 1 + R = (1 - K)^{-1}. \]

Rigor is lacking here because we have not taken the trouble to associate linear operators to our kernels, nor are we going to. Lack of rigor notwithstanding, the last formula makes it clear that \( R(x, y) \) deserves to be called the resolvent of \( K(x, y) \). Moreover, this formula is useful for discovering composition identities which one can then verify directly and rigorously.

**Proof of Lemma 3.4.24:** Here are two reductions to the proof of the fundamental identity. Firstly, it is enough just to prove the first of the equalities claimed in (3.4.25) because the second is proved similarly. Secondly, proceeding term by term, it is enough to prove that

\[
\frac{(-1)^{n-1}}{(n-1)!} \int K(x, z)H_{n-1}(z, y)d\lambda(z) = \frac{(-1)^n}{n!} (H_n(x, y) - \Delta_n \cdot K(x, y))
\]

or, equivalently,

\[
H_n(x, y) = \Delta_n \cdot K(x, y) - n \int K(x, z)H_{n-1}(z, y)d\lambda(z) \tag{3.4.28}
\]

for \( n > 0 \), where \( H_n(x, y) \) is associated to \( K(x, y) \) by the procedure of Section 3.4.2 and \( \Delta_n = \Delta_n(K) \).

Now we can quickly give the proof of the fundamental identity (3.4.26). Expanding by minors of the first row, we find that

\[
K \left( \begin{array}{cccc}
  x & \xi_1 & \cdots & \xi_n \\
  y & \xi_1 & \cdots & \xi_n \\
\end{array} \right)
\]

\[
= K(x, y)K \left( \begin{array}{cccc}
  \xi_1 & \cdots & \xi_n \\
  \xi_1 & \cdots & \xi_n \\
\end{array} \right)
\]

\[
+ \sum_{j=1}^{n} (-1)^j K(x, \xi_j)K \left( \begin{array}{cccc}
  \xi_1 & \cdots & \xi_{j-1} & \xi_j & \xi_{j+1} & \cdots & \xi_n \\
  y & \xi_1 & \cdots & \xi_{j-1} & \xi_{j+1} & \cdots & \xi_n \\
\end{array} \right)
\]

\[
= K(x, y)K \left( \begin{array}{cccc}
  \xi_1 & \cdots & \xi_n \\
  \xi_1 & \cdots & \xi_n \\
\end{array} \right)
\]

\[
- \sum_{j=1}^{n} K(x, \xi_j)K \left( \begin{array}{cccc}
  \xi_j & \xi_1 & \cdots & \xi_{j-1} & \xi_{j+1} & \cdots & \xi_n \\
  y & \xi_1 & \cdots & \xi_{j-1} & \xi_{j+1} & \cdots & \xi_n \\
\end{array} \right).
\]
Integrating out the variables $\xi_1, \ldots, \xi_n$ in evident fashion, we obtain (3.4.28). Thus the fundamental identity is proved.

We extract two further benefits from the proof of the fundamental identity.

**Corollary 3.4.29** (i) For all $n \geq 0$,

$$
\frac{(-1)^n}{n!} H_n(x, y) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \Delta_k \cdot (K \star \cdots \star K)(x, y) .
$$

(ii) Recalling the abbreviated notation $\Delta(K)$ and $\Delta_n(K)$, it holds that

$$
\frac{(-1)^n}{n!} \Delta_{n+1}(K) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \Delta_k(K) \text{Tr}(K \star \cdots \star K) .
$$

In particular, the sequence of numbers

$$
\text{Tr}(K), \; \text{Tr}(K \star K), \; \text{Tr}(K \star K \star K), \; \ldots
$$

uniquely determines the Fredholm determinant $\Delta(K)$.

**Proof:** Part (i) follows from (3.4.28) by employing an induction on $n$. We leave the details to the reader. Part (ii) follows by putting $x = \xi$ and $y = \xi$ in (3.4.30), and integrating out the variable $\xi$.

We can now complete the

**Proof of Lemma 3.4.22:** Let $H_n(\cdot)$ be the many-variable function associated to $K(x, y)$ by the procedure of Section 3.4.2. We proceed by induction on $p$. The case $p = 1$ is trivial, so we assume for the rest of the proof that $p > 1$. Expanding by minors of the first row we have

$$
K \left( \begin{array}{cccc}
  x_1 & \ldots & x_p & \xi_1 & \ldots & \xi_n \\
  y_1 & \ldots & y_p & \xi_1 & \ldots & \xi_n \\
\end{array} \right)
$$

$$
= \sum_{i=1}^{p} (-1)^{i+1} K(x_1, y_i)K \left( \begin{array}{cccc}
  x_2 & \ldots & \ldots & \ldots & x_p & \xi_1 & \ldots & \xi_n \\
  y_1 & \ldots & y_{i-1} & y_{i+1} & \ldots & y_p & \xi_1 & \ldots & \xi_n \\
\end{array} \right)
$$

$$
- \sum_{j=1}^{n} K(x_1, \xi_j)K \left( \begin{array}{cccc}
  \xi_j & x_2 & \ldots & x_p & \xi_1 & \ldots & \xi_{j-1} & \xi_{j+1} & \ldots & \xi_n \\
  y_1 & y_2 & \ldots & y_p & \xi_1 & \ldots & \xi_{j-1} & \xi_{j+1} & \ldots & \xi_n \\
\end{array} \right).
$$
Integrating out the variables $\xi_1, \ldots, \xi_n$ we get the relation

$$H_n \left( \begin{array}{c} x_1 \ldots x_p \\ y_1 \ldots y_p \end{array} \right)$$

$$= \sum_{i=1}^{p} (-1)^{i+1} K(x_1, y_i) H_n \left( \begin{array}{c} x_2 \ldots \ldots \ldots x_p \\ y_1 \ldots y_{i-1} y_{i+1} \ldots y_p \end{array} \right)$$

$$- n \int K(x_1, z) H_{n-1} \left( \begin{array}{c} z \ldots x_p \\ y_1 \ldots y_{i-1} z \ldots y_p \end{array} \right) d\lambda(z).$$

Now multiply both sides of the last identity above by $\frac{(-1)^n}{n!}$ and sum on $n$ from 0 to $\infty$, thus obtaining the relation

$$R \left( \begin{array}{c} x_1 \ldots x_p \\ y_1 \ldots y_p \end{array} \right)$$

$$= \sum_{i=1}^{p} (-1)^{i+1} K(x_1, y_i) R \left( \begin{array}{c} x_2 \ldots \ldots \ldots x_p \\ y_1 \ldots y_{i-1} y_{i+1} \ldots y_p \end{array} \right)$$

$$+ \int K(x_1, z) R \left( \begin{array}{c} z \ldots x_p \\ y_1 \ldots y_{i-1} z \ldots y_p \end{array} \right) d\lambda(z).$$

Next, replace $x_1$ in the identity above by another variable, say $t$, then multiply both sides of the relation above by $R(x_1, t)$, integrate the variable $t$ out, and apply the fundamental identity on the right side. We get

$$\int R(x_1, t) R \left( \begin{array}{c} t \ldots x_p \\ y_1 \ldots y_p \end{array} \right) d\lambda(t)$$

$$= \sum_{i=1}^{p} (-1)^{i+1} (R - K)(x_1, y_i) \cdot R \left( \begin{array}{c} x_2 \ldots \ldots \ldots x_p \\ y_1 \ldots y_{i-1} y_{i+1} \ldots y_p \end{array} \right)$$

$$+ \int (R - K)(x_1, z) \cdot R \left( \begin{array}{c} z \ldots x_p \\ y_1 \ldots y_p \end{array} \right) d\lambda(z).$$

Finally, we replace the dummy variable $t$ on the left side of the last identity by $z$, then we add the last two identities and make the evident cancellations, thus obtaining the relation

$$R \left( \begin{array}{c} x_1 \ldots x_p \\ y_1 \ldots y_p \end{array} \right)$$

$$= \sum_{i=1}^{p} (-1)^{i+1} R(x_1, y_i) R \left( \begin{array}{c} x_2 \ldots \ldots \ldots x_p \\ y_1 \ldots y_{i-1} y_{i+1} \ldots y_p \end{array} \right).$$
By induction on $p$ the sum coincides with the expansion of the determinant

$$R \begin{pmatrix} x_1 & \cdots & x_p \\ y_1 & \cdots & y_p \end{pmatrix}$$

by minors of its first row. The claim (3.4.23) is proved. \qed

3.5 Gap probabilities at 0 and proof of Theorem 3.1.1.

In the remainder of this chapter, we let $X_N \in \mathcal{H}_N^{(2)}$ be a random hermitian matrix from the GUE with eigenvalues $\lambda_N^1 \leq \cdots \leq \lambda_N^N$. We initiate in this section the study of the spacings between eigenvalues of $X_N$. We focus on those eigenvalues that lie near 0, and seek, for a fixed $t > 0$, to evaluate the limit

$$\lim_{N \to \infty} P[\sqrt{N}\lambda_N^1, \ldots, \sqrt{N}\lambda_N^N \notin (-t/2, t/2)],$$

see the statement of Theorem 3.1.1. We note that a-priori, because of Theorems 2.1.4 and 2.1.47, the limit in (3.5.1) has some chance of being non-degenerate because the $N$ random variables $\sqrt{N}\lambda_N^1, \ldots, \sqrt{N}\lambda_N^N$ are spread out over an interval very nearly of length $4N$. As we will show in Section 4.2, the computation of the limit in (3.5.1) allows one to evaluate other limits, such as the limit of the empirical measure of the spacings in the bulk of the spectrum.

As in (3.2.7), set

$$K^{(n)}(x, y) = \sum_{k=0}^{n-1} \psi_k(x)\psi_k(y) = \sqrt{n} \frac{\psi_n(x)\psi_n(y) - \psi_{n-1}(x)\psi_n(y)}{x - y}$$

where the $\psi_k(x)$ are the normalized harmonic oscillator wave-functions introduced in Definition 3.2.4. Set

$$S^{(n)}(x, y) = \frac{1}{\sqrt{n}} K^{(n)} \left( \frac{x}{\sqrt{n}}, \frac{y}{\sqrt{n}} \right).$$

A crucial step in the proof of Theorem 3.1.1 is the following lemma, whose proof, which takes most of the analysis in this section, is deferred.

**Lemma 3.5.2** With the above notation, it holds that

$$\lim_{n \to \infty} S^{(n)}(x, y) = \frac{1}{\pi} \frac{\sin(x - y)}{x - y}$$

holds uniformly on each bounded subset of the $(x, y)$-plane.
Proof of Theorem 3.1.1 Recall that by Lemma 3.2.15,
\[
P[\sqrt{n}\lambda_1^{(n)}, \ldots, \sqrt{n}\lambda_n^{(n)} \notin A] = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{\sqrt{n}-1}^A \cdots \int_{\sqrt{n}-1}^A \det_{i,j=1}^{k} K^{(n)}(x_i, x_j) \prod_{j=1}^{k} dx_j.
\]
Lemma 3.5.2 together with Lemma 3.4.14 complete the proof of the theorem. 

The proof of Lemma 3.5.2 takes up the rest of this section. We begin by bringing, in subsection 3.5.1, a quick introduction to Laplace’s method for the evaluation of asymptotics of integrals, which will be useful for other asymptotic computations, as well. We then apply it in subsection 3.5.2 to conclude the proof.

Remark 3.5.4 We remark that one is naturally tempted to guess that the random variable \( W_N = \text{"width of the largest open interval symmetric about the origin containing none of the eigenvalues } \sqrt{n}\lambda_1^{N}, \ldots, \sqrt{n}\lambda_n^{N} \text{"} \) should possess a limit in distribution. Note however that we do not a priori have tightness for that random variable. But, as we show in the next section, we do have tightness (see (3.6.34) below) a posteriori. In particular, in the next section we prove Theorem 3.1.3, which provides an explicit expression for the limit distribution of \( W_N \).

3.5.1 The method of Laplace

Laplace’s method deals with the asymptotic (as \( s \to \infty \)) evaluation of integrals of the form
\[
\int f(x)^s g(x) dx.
\]
We will be concerned with the situation in which the function \( f \) possesses a global maximum at some point \( a \), and behaves quadratically in a neighborhood of that maximum. More precisely, let \( f : \mathbb{R} \to \mathbb{R}_+ \) be given, and for some constant \( a \) and positive constants \( s_0, K, L, M \), let \( G(a, s_0, f(\cdot), K, L, M) \) be the class of measurable functions \( g : \mathbb{R} \to \mathbb{R} \) satisfying the following conditions:

1. \( |g(a)| \leq K \).
2. \( \sup_{0 < |x-a| \leq \tau_0} \left| \frac{g(x)-g(a)}{x-a} \right| \leq L \).
3. \( \int f(x)^s |g(x)| dx \leq M. \)

We then have the following.

**Theorem 3.5.5 (Laplace)** Let \( f : \mathbb{R} \mapsto \mathbb{R}_+ \) such that, for some \( a \in \mathbb{R} \) and some positive constants \( \varepsilon_0, c \), the following hold.

1. \( f(x) \leq f(x') \) if either \( a - \varepsilon_0 \leq x \leq x' \leq a \) or \( a \leq x' \leq x \leq a + \varepsilon_0 \).
2. For all \( \varepsilon < \varepsilon_0 \), \( \sup_{|x-a|>\varepsilon} f(x) \leq f(a) - c\varepsilon^2 \).
3. \( f(x) \) has two continuous derivatives in the interval \( (a-2\varepsilon_0, a+2\varepsilon_0) \).
4. \( f(a) > 0 \) and \( f''(a) < 0 \).

Then, for any function \( g \in G(a, \varepsilon_0, s_0, f(\cdot), K, L, M) \), we have

\[
\lim_{s \to \infty} \sqrt{s} f(a)^{-s} \int f(x)^s g(x) dx = \sqrt{-\frac{2\pi f(a)}{f''(a)}} g(a),
\]

(3.5.6)

and moreover, for fixed \( f, a, \varepsilon_0, s_0, K, L, M \), the convergence is uniform over the class \( G(a, \varepsilon_0, s_0, f(\cdot), K, L, M) \).

The intuition here is that as \( s \) tends to infinity the function \( (f(x)/f(a))^s \) near \( x = a \) peaks more and more sharply and looks at the microscopic level more and more like a bell-curve, whereas \( f(x)^s \) elsewhere becomes negligible. Formula (3.5.6) is arguably the simplest nontrivial application of Laplace’s method. Later we are going to encounter more sophisticated applications.

**Proof of Theorem 3.5.5:** Let \( \varepsilon(s) \) be a positive function defined for \( s \geq s_0 \) such that \( \varepsilon(s) \to_{s \to \infty} 0 \) and \( \varepsilon(s)^2 / \log s \to_{s \to \infty} \infty \), while \( \varepsilon = \sup_{s \geq s_0} \varepsilon(s) \).

For example we could take \( \varepsilon(s) = c \cdot (s_0/s)^{1/4} \). For \( s \geq s_0 \) write

\[
\int f(x)^s g(x) dx = g(a) I_1 + I_2 + I_3
\]

where

\[
I_1 = \int_{|x-a| \leq \varepsilon(s)} f(x)^s dx,
I_2 = \int_{|x-a| \leq \varepsilon(s)} f(x)^s (g(x) - g(a)) dx,
I_3 = \int_{|x-a| > \varepsilon(s)} f(x)^s g(x) dx.
\]

For \( |t| < 2\varepsilon \) put

\[
h(t) = \int_0^1 (1 - r)(\log f)''(a + rt) dr
\]
thus defining a continuous function of $t$ such that $h(0) = f''(a)/2f(a)$ and which by Taylor’s theorem satisfies

$$f(x) = f(a) \exp(h(x - a)(x - a)^2).$$

for $|x - a| < 2\epsilon$. We then have

$$I_1 = \frac{f(a)}{\sqrt{s}} \int_{|t| \leq \epsilon(s) \sqrt{s}} \exp \left( h \left( \frac{t}{\sqrt{s}} \right) t^2 \right) dt,$$

and hence

$$\lim_{s \to \infty} \sqrt{s}f(a)^{-s}I_1 = \int_{-\infty}^{\infty} \exp \left( h(0) t^2 \right) dt = \sqrt{\frac{2\pi f(a)}{f''(a)}}.$$

We have $|I_2| \leq L\epsilon(s)I_1$ and hence

$$\lim_{s \to \infty} \sqrt{s}f(a)^{-s}I_2 = 0.$$

We have, as soon as $\epsilon(s) < \epsilon_0$,

$$|I_3| \leq M \sup_{x:|x-a|>\epsilon(s)} |f(x)|^{s-s_0} \leq Mf(a)^{s-s_0} \left( 1 - \frac{\epsilon\epsilon(s)^2}{f(a)} \right)^{s-s_0},$$

and hence

$$\lim_{s \to \infty} \sqrt{s}f(a)^{-s}I_3 = 0.$$

This is enough to prove that the limit formula (3.5.6) holds and enough also to prove the uniformity of convergence over all functions $g(x)$ of the class $G$.

\[\square\]

### 3.5.2 Evaluation of the scaling limit - proof of Lemma 3.5.2

Recall that

$$S^{(n)}(x, y) = \frac{\psi_n(\sqrt{n}s)\psi_{n-1}(\sqrt{n}s) - \psi_{n-1}(\sqrt{n}s)}{x-y} \psi_n(\sqrt{n}s).$$

Noting that for any differentiable functions $f, g$ on $\mathbb{R}$,

$$\frac{f(x)g(y) - f(y)g(x)}{x-y} = \left( \frac{f(x) - f(y)}{x-y} \right) g(y) + f(y) \left( \frac{g(y) - g(x)}{x-y} \right)$$

(3.5.7)

$$= g(y) \int_0^1 f'(tx + (1-t)y) dt - f(y) \int_0^1 g'(tx + (1-t)y) dt.$$
we deduce

\begin{align*}
S^{(n)}(x, y) &= \psi_{n-1}(\frac{y}{\sqrt{n}}) \int_0^1 \psi'_n(t \frac{x}{\sqrt{n}} + (1-t) \frac{y}{\sqrt{n}}) dt \\
&\quad - \psi_n(\frac{y}{\sqrt{n}}) \int_0^1 \psi'_{n-1}(t \frac{x}{\sqrt{n}} + (1-t) \frac{y}{\sqrt{n}}) dt \\
&= \psi_{n-1}(\frac{y}{\sqrt{n}}) \int_0^1 (\sqrt{n} \psi_{n-1}(z) - \frac{z}{2} \psi_n(z))_{z=t\frac{x}{\sqrt{n}}+(1-t)\frac{y}{\sqrt{n}}} dt \\
&\quad - \psi_n(\frac{y}{\sqrt{n}}) \int_0^1 (\sqrt{n} - 1 \psi_{n-2}(z) - \frac{z}{2} \psi_{n-1}(z))_{z=t\frac{x}{\sqrt{n}}+(1-t)\frac{y}{\sqrt{n}}} dt \\
&= \psi_{n-1}(\frac{y}{\sqrt{n}}) \int_0^1 (\sqrt{n} \psi_{n-1}(z) - \frac{z}{2} \psi_n(z))_{z=t\frac{x}{\sqrt{n}}+(1-t)\frac{y}{\sqrt{n}}} dt \\
&\quad - \psi_n(\frac{y}{\sqrt{n}}) \int_0^1 (\sqrt{n} - 1 \psi_{n-2}(z) - \frac{z}{2} \psi_{n-1}(z))_{z=t\frac{x}{\sqrt{n}}+(1-t)\frac{y}{\sqrt{n}}} dt \\
&= \psi_{n-1}(\frac{y}{\sqrt{n}}) \int_0^1 (\sqrt{n} \psi_{n-1}(z) - \frac{z}{2} \psi_n(z))_{z=t\frac{x}{\sqrt{n}}+(1-t)\frac{y}{\sqrt{n}}} dt \\
&\quad - \psi_n(\frac{y}{\sqrt{n}}) \int_0^1 (\sqrt{n} - 1 \psi_{n-2}(z) - \frac{z}{2} \psi_{n-1}(z))_{z=t\frac{x}{\sqrt{n}}+(1-t)\frac{y}{\sqrt{n}}} dt
\end{align*}

(3.5.8)

where we used in the last line part 4 of Lemma 3.2.5. Let

\[ \Psi_\nu(t) = n^{\frac{1}{2}} \psi_{\nu}\left(\frac{t}{\sqrt{n}}\right) \]

with \( \nu \) a quantity whose difference from \( n \) is fixed. We shall prove below that uniformly for \( t \) in a fixed bounded interval,

\[ \lim_{n \to \infty} |\Psi_\nu(t) - \frac{1}{\sqrt{\pi}} \cos\left(t - \frac{\pi \nu}{2}\right)| = 0. \]

(3.5.9)

Plugging this result into (3.5.8) and elementary trigonometric formulas show that

\begin{align*}
S^{(n)}(x, y) &\sim \frac{1}{\pi} \int_0^1 \cos\left(tx + (1-t)y - \frac{\pi(n-1)}{2}\right) dt \\
&\quad - \cos(y - \frac{\pi n}{2}) \int_0^1 \cos\left(tx + (1-t)y - \frac{\pi(n-1)}{2}\right) dt \\
&\sim \frac{1}{\pi} \frac{\sin(x-y)}{x-y}
\end{align*}

which completes the proof of the Lemma.

It thus only remains to prove (3.5.9), with the asserted uniformity. Recall the Fourier transform identity

\[ e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \int e^{-\xi^2/2 - i\xi x} d\xi. \]

Differentiating under the integral, we find that

\[ \delta_n(x)e^{-x^2/2} = (-1)^n \frac{d^n}{dx^n} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \int (i\xi)^n e^{-\xi^2/2 - i\xi x} d\xi, \]
or equivalently
\[
\psi_\nu(x) = \frac{i\nu e^{x^2/4}}{(2\pi)^{3/4}\sqrt{\nu!}} \int \xi^\nu e^{-\xi^2/2-i\xi x} d\xi.
\] (3.5.10)

We use the letter \(\nu\) here instead of \(n\) to help avoid confusion at the next step. As a consequence, setting \(C_{\nu,n} = i\nu \sqrt{n/(2\pi)}\), we have
\[
\Psi_\nu(t) = \frac{i\nu e^{t^2/(4n)}n^{1/4}}{(2\pi)^{3/4}\sqrt{n!}} \int \xi^\nu e^{-\xi^2/2-it\xi} d\xi.
\]

Let \(n \to \infty\) in such a way that \(i^n\) is fixed. Note that \(f(\xi) = \xi e^{-\xi^2/2}\) achieves its maximal value at \(\xi = 1\) and

\[
f(1) = e^{-1/2}, \quad f'(1) = 0, \quad f''(1) = -2e^{-1/2}.
\]

Hence, we can apply Laplace’s method to find that
\[
\Psi_\nu(t) \sim \frac{(2\pi)^{1/4}C_{\nu,n}n^{1/4+n/2}}{\sqrt{n!}} \int (\xi e^{-\xi^2/2})^n e^{-it\xi} \xi^{\nu-n} d\xi
\]
\[
\sim C_{\nu,n}e^{n/2} \int |\xi^{-\xi^2/2}|^n (\text{sign} \xi)^\nu e^{-it\xi} |\xi^{\nu-n}| d\xi
\]
\[
\to -n \to \infty \frac{i\nu}{2\sqrt{\pi}} (e^{-it} + (-1)^\nu e^{it}) = \frac{1}{\sqrt{\pi}} \cos \left(t - \frac{\pi\nu}{2}\right)
\]

by using Stirling’s formula (3.4.7) in the second line and (3.5.6) in the third. (The latter is actually applied twice, once to each of the two critical points.) Moreover the convergence here is uniform for \(t\) in a fixed bounded interval, as follows from the uniformity asserted for convergence in limit formula (3.5.6).

\[\square\]

3.5.3 A complement: determinantal relations

Let integers \(\ell_1, \ldots, \ell_p \geq 0\) and bounded disjoint Borel sets \(A_1, \ldots, A_p\) be given. Put
\[
P_N(\ell_1, \ldots, \ell_p; A_1, \ldots, A_p)
= P \left[ \ell_i = \# \left\{ \sqrt{\lambda_1^N}, \ldots, \sqrt{\lambda_n^N} \right\} \cap A_i \right] \text{ for } i = 1, \ldots, p.
\]

We have:
Lemma 3.5.11 Let \( s_1, \ldots, s_p \) be independent complex variables and let

\[
\varphi = (1 - s_1)1_{A_1} + \cdots + (1 - s_p)1_{A_p}.
\]

Then, the limit

\[
P(\ell_1, \ldots, \ell_p; A_1, \ldots, A_p) = \lim_{N \to \infty} P_N(\ell_1, \ldots, \ell_p; A_1, \ldots, A_p). \tag{3.5.12}
\]

exists and satisfies

\[
\sum_{\ell_1=0}^{\infty} \cdots \sum_{\ell_p=0}^{\infty} P(\ell_1, \ldots, \ell_p; A_1, \ldots, A_p) s_1^{\ell_1} \cdots s_p^{\ell_p} = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int \cdots \int \det_{i,j=1}^{k} \frac{1}{\pi} \sin(x_i - x_j) \prod_{j=1}^{k} \varphi(x_j) dx_j. \tag{3.5.13}
\]

That is, the generating function in the left side of (3.5.12) can be represented in terms of a Fredholm determinant. We note that this is no coincidence and holds in far greater generality, see Section 4.2.

Proof: The proof is a slight modification of the method presented in subsection 3.5.2. Note that the right side of (3.5.13) defines by the fundamental estimate (3.4.10) an entire function of the complex variables \( s_1, \ldots, s_p \), whereas the left side defines a function analytic in a domain containing the product of \( p \) copies of the unit disc centered at the origin. Clearly we have

\[
E \prod_{i=1}^{N} \left( 1 - \varphi \left( \sqrt{N} \chi_i^N \right) \right) = \sum_{\ell_1, \ldots, \ell_p \geq 0} P_N(\ell_1, \ldots, \ell_p; A_1, \ldots, A_p) s_1^{\ell_1} \cdots s_p^{\ell_p}. \tag{3.5.14}
\]

The function of \( s_1, \ldots, s_p \) on the right is simply a polynomial, whereas the expectation on the left can be represented as a Fredholm determinant. From this, the lemma follows after representing the probability \( P_N(\ell_1, \ldots, \ell_p; A_1, \ldots, A_p) \) as a \( p \)-dimensional Cauchy integral. □

Exercise 3.5.15 Use Laplace's method (Theorem 3.5.5) with \( a = 1 \) to prove (2.5.17): as \( s \to \infty \) along the positive real axis,

\[
\Gamma(s) = \int_{x=0}^{\infty} x^s e^{-x} \frac{dx}{x} = s^{s} \int_{x=0}^{\infty} (xe^{-x})^s \frac{dx}{x} \sim \sqrt{2\pi} s^{-1/2} e^{-s}.
\]

This recovers in particular Stirling’s formula (3.4.7).
3.6 Analysis of the sine-kernel

We more or less follow [TW93] in our exposition of the fundamental results of [JMMS80] concerning the sine-kernel and Painlevé V. But we alter the operator-theoretic viewpoint of [TW93] to a “matrix algebra” viewpoint consistent with that taken in our general discussion in Section 3.4 of Fredholm determinants.

3.6.1 General differentiation formulae and the Kyoto equations

Recalling the setting of our general discussion of Fredholm determinants in Section 3.4, we fix a bounded open interval \((a, b) \subset \mathbb{R}\) and we choose the function \(A\) so that \((a, b) = \{ A > 0 \}\). We fix real numbers \(a < t_1 < \cdots < t_n < b\) in the interval \((a, b)\) and complex numbers \(s_1, \ldots, s_{n-1}, s_0 = 0 = s_n\), and we set

\[\lambda' = s_11_{(t_1, t_2)} + \cdots + s_{n-1}1_{(t_{n-1}, t_n)}\).

One may think of \(\lambda'\) as the density (with respect to Lebesgue’s measure) of a positive measure \(\lambda\) on \(\mathbb{R}\). We then have, for \(f \in L^1((a, b)]\),

\[\langle f, \lambda' \rangle = \int f(x)d\lambda(x) = \sum_{i=1}^{n-1} s_i \int_{t_i}^{t_{i+1}} f(x)dx\.

for all integrable functions \(f(x)\) on \((a, b)\). Motivated by Theorem 3.1.1, we fix the function

\[S(x, y) = \frac{\sin(x - y)}{x - y}\]

on \((a, b)^2\) as our kernel. As usual \(\Delta = \Delta(S)\) denotes the Fredholm determinant associated to \(S\) and the signed measure \(\lambda\). We assume that \(\Delta \neq 0\) so that the Fredholm resolvent \(R(x, y)\) is also defined.

We consider the parameters \(t_1, \ldots, t_n\) variable, whereas we consider the kernel \(S(x, y)\) and the parameters \(s_1, \ldots, s_{n-1}\) to be fixed. Our aim is to understand the variation of \(\Delta\) and related quantities with \(t_1, \ldots, t_n\). Toward this end, set

\[Q(x) = \sin x + \int R(x, y) \sin y \, d\lambda(y), \quad P(x) = \cos x + \int R(x, y) \cos y \, d\lambda(y)\).

(3.6.1)
The main result of this section, of which Theorem 3.1.3 is an easy corollary, is the following system of differential equations, first discovered by the authors of [JMMS80], which we refer to as the Kyoto equations in their honor.

**Theorem 3.6.2 (Kyoto equations)** With the above notation, put, for \(i,j = 1,\ldots,n\),

\[
p_i = P(t_i), \quad q_i = Q(t_i), \quad R_{ij} = R(t_i, t_j).
\]

Then, for \(i,j = 1,\ldots,n\) with \(i \neq j\), we have the following equations:

\[
R_{ij} = \frac{q_j p_i - q_i p_j}{(t_i - t_j)} \quad \partial q_j / \partial t_i = -(s_i - s_{i-1}) \cdot R_{ji} q_i \quad \partial p_j / \partial t_i = -(s_i - s_{i-1}) \cdot R_{ji} p_i
\]

\[
\partial q_i / \partial t_i = +p_i + \sum_{k \neq i} (s_k - s_{k-1}) \cdot R_{ik} q_k \tag{3.6.3}
\]

\[
\partial p_i / \partial t_i = -q_i + \sum_{k \neq i} (s_k - s_{k-1}) \cdot R_{ik} p_k
\]

\[
R_{ii} = p_i \partial q_i / \partial t_i - q_i \partial p_i / \partial t_i
\]

\[
(\partial / \partial t_i) \log \Delta = (s_i - s_{i-1}) \cdot R_{ii}
\]

The proof of the Kyoto equations is completed in subsection 3.6.2. In the rest of this subsection, we derive a fundamental differentiation formula, see (3.6.7), and derive several relations concerning the functions \(P, Q\) introduced in (3.6.1), and the resolvent \(R\).

Recall from (3.4.16) that

\[
H_\ell \left( \begin{array}{c} x_1 \\ y_1 \\ \vdots \\ x_k \\ y_k \end{array} \right) = \sum_{i_1=1}^{n-1} \cdots \sum_{i_\ell=1}^{n-1} s_{i_1} \cdots s_{i_\ell} \\
\int_{t_1}^{t_{i_1+1}} \cdots \int_{t_\ell}^{t_{i_\ell+1}} S \left( \begin{array}{c} x_1 \\ y_1 \\ \vdots \\ x_p \\ y_p \\ \xi_1 \\ \cdots \xi_\ell \end{array} \right) d\xi_1 \cdots d\xi_\ell
\]
Therefore, by the fundamental theorem of calculus,

\[
\frac{\partial}{\partial t_i} H_t \left( \begin{array}{c} x_1 \\ y_1 \\ \vdots \\ x_k \\ y_k \end{array} \right) = \sum_{j=1}^{\ell} \sum_{t_{i_1}=1}^{t_i} \cdots \sum_{t_{i_{j-1}}=1}^{t_{i_{j-1}}+1} \sum_{t_{i_{j+1}}=1}^{t_{i_{j+1}}-1} \sum_{t_{i_{\ell}}=1}^{t_{i_{\ell}}-1} \sum_{s_{i_{j}} \neq t_{i_{j}}} s_{i_{j}} \left( s_{i_{j}} - s_{i_{j}-1} \right) \\
\cdot \int_{t_{i_{j-1}}}^{t_{i_{j}}} \cdots \int_{t_{i_{j+1}}}^{t_{i_{j+1}}+1} \int_{t_{i_{\ell}}}^{t_{i_{\ell}}+1} S \left( \begin{array}{c} x_1 \\ y_1 \\ \vdots \\ x_k \\ y_k \\ \xi_1 \\ \xi_1 - 1 \\ t_i \\ \xi_{i+1} \\ \xi_{i+1} - 1 \\ \xi_{i} \end{array} \right) \prod_{j=1}^{s} d\xi_j \\
= \ell (s_i - s_{i-1}) H_{t-1} \left( \begin{array}{c} x_1 \\ y_1 \\ \vdots \\ x_k \\ y_k \\ t_i \end{array} \right) \\
= \ell (s_i - s_{i-1}) \frac{\partial}{\partial t_i} \log \Delta = \ell (s_i - s_{i-1}) \frac{\partial}{\partial t_i} \log \Delta.
\]  

(3.6.4)

Multiplying by \((-1)^{\ell}\) and summing, using the estimate (3.4.10) and dominated convergence, we find that

\[
\frac{\partial}{\partial t_i} H \left( \begin{array}{c} x_1 \\ y_1 \\ \vdots \\ x_k \\ y_k \\ t_i \end{array} \right) = (s_i - s_{i-1}) H \left( \begin{array}{c} x_1 \\ y_1 \\ \vdots \\ x_k \\ y_k \\ t_i \end{array} \right). 
\]  

(3.6.5)

(3.6.5)

From (3.6.5) in the case \(k = 0\) we get immediately

\[
\frac{\partial}{\partial t_i} \log \Delta = (s_i - s_{i-1}) R(t_i, t_i).
\]  

(3.6.6)

From (3.6.5) in the case \(k = 1\) we get immediately

\[
\frac{\partial}{\partial t_i} R(x, y) = (s_i - s_{i-1}) R \left( \begin{array}{c} x \\ y \\ t_i \end{array} \right) - R(x, y) \frac{\partial}{\partial t_i} \log \Delta
\]

and if we then also apply (3.4.23) and (3.6.6), we get

\[
\frac{\partial}{\partial t_i} R(x, y) = (s_{i-1} - s_i) R(x, t_i) R(t_i, y).
\]  

(3.6.7)

(3.6.7)

It is for the sake of easily laying hands on the last differentiation formula that we took the trouble to prove the general determinantal formula (3.4.23).

The next lemma will play an important role in the proof of Theorem 3.6.2.

**Lemma 3.6.8** The functions \(P, Q, R\) satisfy the following relations:

\[
R(x, y) = \frac{Q(x)P(y) - Q(y)P(x)}{x - y} = R(y, x),
\]  

(3.6.9)
\begin{equation}
R(x, x) = Q'(x)P(x) - Q(x)P'(x), \quad (3.6.10)
\end{equation}

\begin{equation}
\frac{\partial}{\partial t_i}Q(x) = (s_{i-1} - s_i)R(x, t_i)Q(t_i), \quad (3.6.11)
\end{equation}

and similarly

\begin{equation}
\frac{\partial}{\partial t_i}P(x) = (s_{i-1} - s_i)R(x, t_i)P(t_i). \quad (3.6.12)
\end{equation}

**Proof:** Recall the fundamental identity (3.4.26):

\[
\int R(x, z)S(z, y)\,d\lambda(z) = R(x, y) - S(x, y) = \int S(x, z)R(z, y)\,d\lambda(z),
\]
which for convenience we write in the abbreviated form

\[R \ast S = R - S = S \ast R.
\]

To abbreviate notation further, put

\[\tilde{R}(x, y) = (x - y)R(x, y), \quad \tilde{S}(x, y) = (x - y)S(x, y).
\]

From the fundamental identity we deduce that

\[\tilde{R} \ast \tilde{S} + R \ast \tilde{S} = \tilde{R} - \tilde{S}.
\]

Applying the operation \((\cdot) \ast R\) on both sides, we get

\[\tilde{R} \ast (R - S) + R \ast \tilde{S} \ast R = \tilde{R} \ast R - \tilde{S} \ast R.
\]

Adding the last two relations and making the obvious cancellations and rearrangements, we get

\[\tilde{R} = (1 + R) \ast \tilde{S} \ast (1 + R).
\]

Written out “in longhand” this yields (3.6.9). Using the trigonometric identity

\[\sin(x - y) = \sin x \cos y - \sin y \cos x
\]

as well as the symmetry

\[S(x, y) = S(y, x), \quad R(x, y) = R(y, x),
\]

equation (3.6.10) follows. Finally, by (3.6.7) and the definitions we obtain

\[
\frac{\partial}{\partial t_i}Q(x) = (s_{i-1} - s_i)R(x, t_i) \left( \sin t_i + \int R(t_i, y) \sin y d\lambda(y) \right) = (s_{i-1} - s_i)R(x, t_i)Q(t_i)
\]

yielding (3.6.11). Equation (3.6.12) is obtained similarly. \(\square\)
**Remark 3.6.13** Disregarding issues of rigor, view our kernels as operators, writing multiplication instead of the \( \ast \) operation. Then we have

\[
(1 - S)^{-1} = 1 + R, \quad \tilde{S} = [M, S], \quad \tilde{R} = [M, R]
\]

where \( M \) is the operation of multiplication by \( x \). Note also that under our special assumptions

\[
\tilde{S}(x, y) = \sin x \cos y - \sin y \cos x,
\]

and hence the operator \( \tilde{S} \) is of rank 2. But we have

\[
\tilde{R} = [M, R] = [M, (1 - S)^{-1}] = -(1 - S)^{-1}[M, 1 - S](1 - S)^{-1} = (1 + R)\tilde{S}(1 + R),
\]

and hence \( \tilde{R} \) is also of rank 2. This is some sort of heuristic explanation for (3.6.9).

**Exercise 3.6.14** An alternative to the elementary calculus used in deriving (3.6.4) and (3.6.5), which is useful in obtaining higher order derivatives of the determinants, resolvents, and adjugants, is sketched in this exercise.

(i) Let \( D \) be a domain (connected open subset) in \( \mathbb{C}^n \). With \( X \) a measure space, let \( f(x, \zeta) \) be a measurable function on \( X \times D \), depending analytically on \( \zeta \) for each fixed \( x \) and satisfying the condition

\[
\sup_{\zeta \in K} \int |f(x, \zeta)|d\mu(x) < \infty
\]

for all compact subsets \( K \subset D \). Prove that the function

\[
F(\zeta) = \int f(x, \zeta)d\mu(x)
\]

is analytic in \( D \) and that for each index \( i = 1, \ldots, n \) and all compact \( K \subset D \),

\[
\sup_{\zeta \in K} \int \left| \frac{\partial}{\partial \zeta_i} f(x, \zeta) \right| d\mu(x) < \infty.
\]

Further, applying the Cauchy theorem to turn the derivative into an integral, and then Fubini’s theorem, prove the identity of functions analytic in \( D \):

\[
\frac{\partial}{\partial \zeta_i} F(\zeta) = \int \left( \frac{\partial}{\partial \zeta_i} f(x, \zeta) \right) d\mu(x).
\]

(ii) Using that the kernel \( S \) is an entire function, extend the definitions of \( H_\ell \), \( H \) and \( \Delta \) in the setup of this section to analytic functions in the parameters \( t_1, \ldots, t_n, s_1, \ldots, s_{n-1} \).

(iii) View the signed measure \( \lambda \) as defining a family of distributions \( \lambda' \) (in
the sense of Schwartz) on the interval $(a, b)$ depending on the parameters $t_1, \ldots, t_n$, by the formula

$$\langle \varphi, \lambda' \rangle = \sum_{i=1}^{n-1} s_i \int_{t_{i-1}}^{t_i} \varphi(x) dx,$$

valid for any smooth function $\varphi(x)$ on $(a, b)$. Show that $\partial \lambda'/\partial t_i$ is a distribution satisfying

$$\frac{\partial}{\partial t_i} \lambda' = (s_{i-1} - s_i) \delta_{t_i}$$

(3.6.15)

for $i = 1, \ldots, n$, and that the distributional derivative $\lambda''$ of $\lambda'$ satisfies

$$\lambda'' = \sum_{i=1}^{n} (s_i - s_{i-1}) \delta_{t_i} = -\sum_{i=1}^{n} \frac{\partial \lambda'}{\partial t_i}.$$  

(3.6.16)

(iv) Use (3.6.15) to justify (3.6.4) and step (i) to justify (3.6.5).

3.6.2 Derivation of the Kyoto equations: proof of Theorem 3.6.2

To proceed farther we need means for differentiating $Q(x)$ and $P(x)$ both with respect to $x$ and with respect to the parameters $t_1, \ldots, t_n$. To this end we introduce the further abbreviated notation

$$S'(x, y) = \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) S(x, y) = 0, \quad R'(x, y) = \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) R(x, y)$$

and

$$(F \ast' G)(x, y) = \int F(x, z)G(z, y) d\lambda'(z) = \sum_{i=1}^{n} (s_i - s_{i-1}) F(x, t_i) G(t_i, y),$$

which can be taken as the definition of $\lambda'$. (This notation is consistent with the distributional language introduced in Exercise 3.6.14.) Below we persist for a while in writing $S'$ instead of just automatically putting $S' = 0$ everywhere in order to keep the structure of the calculations clear. From the fundamental identity

$$R \ast S = R - S = S \ast R$$

we deduce

$$R' \ast S + R \ast' S + R \ast S' = R' - S'.$$
Applying the operation $\star R$ on both sides of the last equation we find that

$$R' \star (R - S) + R' \star (R - S) + R \star S' \star S = R' \star R - S' \star R.$$  

Adding the last two equations and then making the obvious cancellations (including now the cancellation $S' = 0$) we find that

$$R' = R \star S' \star R.$$  

Written out “in longhand” the last equation says that

$$(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}) R(x, y) = \sum_{i=1}^{n} (s_i - s_{i-1}) R(x, t_i) R(t_i, y). \quad (3.6.17)$$

Now we can differentiate $Q(x)$ and $P(x)$. We have

$$Q'(x) = (\sin x)' + \int \frac{\partial}{\partial x} R(x, y) \sin y \, d\lambda(y) dy$$

$$= (\sin x)' - \int \frac{\partial}{\partial y} R(x, y) \sin y \, d\lambda(y)$$

$$+ \int \left( \int R(x, t) R(t, y) d\lambda'(t) \right) \sin y \, d\lambda(y)$$

$$= (\sin x)' + \int R(x, y) (\sin y)' \, d\lambda(y) + \int R(x, y) \sin y \, d\lambda'(y)$$

$$+ \int \left( \int R(x, t) R(t, y) \lambda'(t) dt \right) \sin y \, d\lambda(y)$$

$$= (\sin x)' + \int R(x, y) (\sin y)' \, d\lambda(y)$$

$$+ \int R(x, t) \left( \sin t + \int R(t, y) \sin y \, d\lambda(y) \right) \, d\lambda'(t)$$

$$= P(x) + \sum_{k=1}^{n} (s_k - s_{k-1}) R(x, t_k) Q(t_k) \quad (3.6.18)$$

and similarly

$$P'(x) = -Q(x) + \sum_{k=1}^{n} (s_k - s_{k-1}) R(x, t_k) P(t_k). \quad (3.6.19)$$

Observing now that

$$\frac{\partial}{\partial t_i} Q(t_i) = Q'(t_i), \quad \frac{\partial}{\partial t_i} P(t_i) = P'(t_i)$$

$$\frac{\partial}{\partial t_i} Q(x) \bigg|_{x=t_i}, \quad \frac{\partial}{\partial t_i} P(x) \bigg|_{x=t_i}$$
and adding (3.6.18) and (3.6.11) we have
\[ \frac{\partial}{\partial t_i} Q(t_i) = P(t_i) + \sum_{k \in \{1, \ldots, n\} \setminus \{i\}} (s_k - s_{k-1}) R(t_i, t_k) Q(t_k). \] (3.6.20)

Similarly by adding (3.6.19) and (3.6.12) we have
\[ \frac{\partial}{\partial t_i} P(t_i) = -Q(t_i) + \sum_{k \in \{1, \ldots, n\} \setminus \{i\}} (s_k - s_{k-1}) R(t_i, t_k) P(t_k). \] (3.6.21)

It follows also via (3.6.10) and (3.6.11) that
\[ R(t_i, t_i) = P(t_i) \frac{\partial}{\partial t_i} Q(t_i) - Q(t_i) \frac{\partial}{\partial t_i} P(t_i) \] (3.6.22)
(note that the terms involving \( \frac{\partial Q(x)}{\partial t_i} \big|_{x=t_i} \) cancel out to yield the above equality). Unraveling the definitions, this completes the proof of (3.6.3) and hence of Theorem 3.6.2.

\[ \square \]

**3.6.3 Reduction of the Kyoto equations to Painlevé V**

In what follows, we complete the proof of Theorem 3.1.3. We take in Theorem 3.6.2 the values \( n = 2, s_1 = s \). Our goal is to figure out the ordinary differential equation we get by reducing still farther to the case \( t_1 = -t/2 \) and \( t_2 = t/2 \). That is, we prove the following.

**Lemma 3.6.23** Set \( \Delta = \Delta(S) \) with \( \lambda' = s \) and \( \sigma = \sigma(t) = t \frac{d}{dt} \log \Delta \). Then,
\[ (t\sigma'')^2 + 4(t\sigma' - \sigma)(t\sigma' - \sigma + (\sigma')^2) = 0, \] (3.6.24)
and, for each fixed \( s \),
\[ \Delta = 1 - st + O(t^4), \quad \sigma = -st - s^2t^2 - s^3t^3 + O(t^4) \text{ as } t \downarrow 0. \] (3.6.25)

**Proof:** We have from Theorem 3.6.2 specialized to \( n = 2 \) that
\[
\begin{align*}
R_{21} &= (q_2 p_1 - q_1 p_2)/(t_2 - t_1) = R_{12} \\
\partial q_1/\partial t_2 &= +s R_{12} q_2 \\
\partial q_2/\partial t_1 &= -s R_{21} q_1 \\
\partial p_1/\partial t_2 &= +s R_{12} p_2 \\
\partial p_2/\partial t_1 &= -s R_{21} p_1 \\
\partial q_1/\partial t_1 &= +p_1 - s R_{12} q_2 \\
\partial q_2/\partial t_2 &= +p_2 + s R_{21} q_1 \\
\partial p_1/\partial t_1 &= -q_1 - s R_{12} p_2 \\
\partial p_2/\partial t_2 &= -q_2 + s R_{21} p_1 \\
(\partial/\partial t_1) \log \Delta &= +s(p_2^2 + q_2^2 + s(t_1 - t_2)R_{22}^2) \\
(\partial/\partial t_2) \log \Delta &= -s(p_2^2 + q_2^2 + s(t_1 - t_2)R_{21}^2)
\end{align*}
\] (3.6.26)
For convenience we note the following consequences of the preceding equations:

\[
\frac{1}{2} \left( \frac{\partial}{\partial t_2} - \frac{\partial}{\partial t_1} \right) \log \Delta = -\frac{1}{2} s(p_1^2 + q_1^2 + p_2^2 + q_2^2) + s^2(t_2 - t_1) R_{12}^2 \\
\frac{1}{2} \left( \frac{\partial q_1}{\partial t_2} - \frac{\partial q_1}{\partial t_1} \right) = -p_1/2 + sR_{12}q_2, \\
\frac{1}{2} \left( \frac{\partial p_1}{\partial t_2} - \frac{\partial p_1}{\partial t_1} \right) = +q_1/2 + sR_{12}p_2. 
\]

(3.6.27)

We now analyze symmetry. Temporarily we write

\[
\Delta(t_1, t_2), \ p_1(t_1, t_2), \ q_1(t_1, t_2), \ p_2(t_1, t_2), \ q_2(t_1, t_2), \ R_{12}(t_1, t_2)
\]

in order to emphasize the roles of the parameters \( t_1 \) and \( t_2 \). To begin with, since

\[
S(x + c, y + c) = S(x, y)
\]

for any constant \( c \) we have

\[
\Delta(t_1, t_2) = \Delta(t_1 + c, t_2 + c) = \Delta(-t_2, -t_1). 
\]

(3.6.28)

Further, we have

\[
p_1(t_1, t_2) = \cos t_1 + \frac{1}{\Delta(t_1, t_2)} \sum_{n=0}^{\infty} \frac{(-1)^n s^{n+1}}{n!} \\
\int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} S \left( \begin{array}{c} t_1 \\ x_1 \\
\vdots \\
x_n \\ y \\
\end{array} \right) \cos y dx_1 \cdots dx_n dy \\
= \cos(-t_1) + \frac{1}{\Delta(-t_2, -t_1)} \sum_{n=0}^{\infty} \frac{(-1)^n s^{n+1}}{n!} \\
\int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} S \left( \begin{array}{c} -t_1 \\ -x_1 \\
\vdots \\
-x_n \\ -y \\
\end{array} \right) \cos y dx_1 \cdots dx_n dy \\
= \cos(-t_1) + \frac{1}{\Delta(-t_2, -t_1)} \sum_{n=0}^{\infty} \frac{(-1)^n s^{n+1}}{n!} \\
\int_{-t_2}^{-t_1} \cdots \int_{-t_2}^{-t_1} S \left( \begin{array}{c} -t_1 \\ y \\
\vdots \\
x_n \\
\end{array} \right) \cos y dx_1 \cdots dx_n dy \\
= p_2(-t_2, -t_1). 
\]

(3.6.29)

Similarly we have

\[
q_1(t_1, t_2) = -q_2(-t_2, -t_1). 
\]

(3.6.30)
Now we are ready to reduce to the one-dimensional situation. We specialize as follows. Put
\[ p = p(t) = p_1(-t/2, t/2) = p_2(-t/2, t/2), \]
\[ q = q(t) = q_1(-t/2, t/2) = -q_2(-t/2, t/2), \]
\[ R = R(t) = R_{12}(-t/2, t/2) \quad (3.6.31) \]
\[ \sigma = \sigma(t) = t \frac{d}{dt} \log \Delta(-t/2, t/2). \]

Then via (3.6.26) and (3.6.27) and the preceding analysis of symmetry we have
\[ R = -2pq/t \]
\[ \sigma = -st(p^2 + q^2) + 4s^2q^2p^2, \]
\[ q' = -p/2 + 2spq^2/t, \]
\[ p' = +q/2 - 2sp^2q/t, \]
\[ \sigma' = -s(p^2 + q^2), \]
\[ t\sigma'' = 4s^2(p^3q - q^3p), \]
\[ (3.6.32) \]

at which point, after some computation that eliminates \( p \) and \( q \), we discover that
\[ 4t(\sigma')^3 + 4t^2(\sigma')^2 - 4\sigma(\sigma')^2 + 4\sigma^2 + (t\sigma'')^2 - 8t\sigma\sigma' = 0 \]
or equivalently, we get (3.6.24), which is the Jimbo-Miwa-Okamoto \( \sigma \)-form of Painlevé V. Note that the differential equation is independent of \( s \).

Turning to the proof of (3.6.25), we explicitly compute a few terms of the expansion of
\[ \Delta = 1 + \sum_{k=1}^{\infty} \frac{(-s)^k}{k!} \int_{-t/2}^{t/2} \cdots \int_{-t/2}^{t/2} \frac{k}{\sin(x_i - x_j)} \prod_{j=1}^{k} dx_j \]
in powers of \( t \). Indeed,
\[ \int_{-t/2}^{t/2} dx = t, \quad \int_{-t/2}^{t/2} \cdots \int_{-t/2}^{t/2} \frac{k}{\sin(x_i - x_j)} \prod_{j=1}^{k} dx_j = O(t^4) \text{ for } k \geq 2 \]
and hence (3.6.25) follows. More Taylor coefficients of \( \Delta \) and \( \sigma \) can be computed with the help of a computer algebra system and (as protection against sign errors and so forth) it can be confirmed by brute force that \( \sigma \) does indeed satisfy (3.6.24).

**Proof of Theorem 3.1.3** We use Lemma 3.6.23. Take \( s = 1/\pi \) and set
\[ F(t) = 1 - \Delta = 1 - \exp \left( \int_0^t \frac{\sigma(u)}{u} du \right) \text{ for } t \geq 0. \]
Then by (3.1.2) we have

\[ 1 - F(t) = \lim_{N \to \infty} P[\sqrt{N}\lambda_1^N, \ldots, \sqrt{N}\lambda_N^N \notin (-t/2, t/2)], \]

completing the proof of the theorem.

Remark 3.6.33 We emphasize that we have not yet proved that the function \( F(\cdot) \) in Theorem 3.1.3 is a distribution function, that is, we have not shown tightness for the sequence of gaps around 0. From the expansion at 0 of \( \sigma(t) \), see (3.1.4), it follows immediately that \( \lim_{t \to 0} F(t) = 0 \). To show that \( F(t) \to 1 \) as \( t \to \infty \) requires more work. One approach, that uses careful and non-trivial analysis of the resolvent equation, see see [Wid94] for the first rigorous proof, shows that in fact

\[ \sigma(t) \sim -t^2/4 \text{ as } t \to +\infty, \]

implying that \( \lim_{t \to \infty} F(t) = 1 \). An easier approach, which does not however yield such precise information, proceeds from the CLT for determinantal processes developed in Section 4.2: indeed, it is straightforward to verify, see Exercise 4.2.55, that for the determinantal process determined by the sign kernel, the expected number of points in an interval of length \( L \) around 0 increases linearly in \( L \), while the variance increases only logarithmically in \( N \). This is enough to show that with \( A = [-t/2, t/2] \), the right side of (3.1.2) decreases to 0 as \( t \to \infty \), which implies that \( \lim_{t \to \infty} F(t) = 1 \). In particular, it follows that the random variable giving the width of the largest open interval centered at the origin in which no eigenvalue of \( \sqrt{N}X \) appears is weakly convergent as \( N \to \infty \) to a random variable with distribution \( F \).

3.7 Edge-scaling: Proof of Theorem 3.1.7

Our goal in this section is to study the spacing of eigenvalues at the edge of the spectrum. The main result is the proof of Theorem 3.1.7, which is completed in subsection 3.7.2, after we first derive in subsection 3.7.1 some properties of the Airy kernel.

3.7.1 A preliminary view of the Airy function

We first note that by parameterizing the contour \( C \) appearing in Definition 3.1.5 of the Airy function in evident fashion, we obtain the formula

\[ \text{Ai}(x) = \frac{1}{2\pi i} \int_0^\infty \exp \left( -\frac{t^3}{3} \right) \left( \exp \left( -xt - \frac{\pi i}{3} \right) - \exp \left( -xt + \frac{\pi i}{3} \right) \right) dt. \]
The latter makes it clear that $\text{Ai}(x)$ is an entire analytic function of $x$. By differentiating under the integral and then integrating by parts, it follows that $\text{Ai}(x)$ satisfies the Airy equation:

$$\frac{d^2 y}{dx^2} - xy = 0. \quad (3.7.2)$$

We remark that by replacing the contour $C$ in (3.1.6) by another, say the contour consisting of the ray joining $-\infty$ to the origin plus the ray joining the origin to $e^{\pi i/3}\infty$, one can obtain a second linearly independent solution to the Airy equation (3.7.2). The function $\text{Ai}(x)$ is distinguished in the two-dimensional space of solutions of the Airy equation up to a constant factor by its rapid decay along the positive real axis and picked out uniquely by the normalization

$$\text{Ai}(0) = \frac{i}{3^{1/3} \Gamma(1/3)} = \frac{1}{3^{1/3} \sqrt[3]{\pi}} = 0.3550\ldots \quad (3.7.3)$$

It is also worth remarking that $\text{Ai}(x) > 0$ and $\text{Ai}'(x) < 0$ for $x > 0$. The function $\text{Ai}(x)$ comes up often in applied mathematics and for example is implemented in the MAPLE software package under the name “AiryAi”. The following lemma, whose proof is deferred to subsection 3.7.3, summarizes basic properties of the Airy kernel needed in the proof of Theorem 3.1.7.

**Lemma 3.7.4**

$$\sup_{x,y \in \mathbb{R}} e^{x+y} |A(x, y)| < \infty. \quad (3.7.5)$$

### 3.7.2 Vague convergence of the rescaled largest eigenvalue: proof of Theorem 3.1.7

Again we let $X_N \in \mathcal{H}^{(2)}_N$ be a random hermitian matrix from the GUE with eigenvalues $\lambda^N_1 \leq \cdots \leq \lambda^N_N$. We now present the:

**Proof of Theorem 3.1.7** As before put

$$K^{(n)}(x, y) = \sqrt{n} \frac{\psi_n(x) \psi_{n-1}(y) - \psi_{n-1}(x) \psi_n(y)}{x - y}$$

where the $\psi_n(x)$ is the normalized harmonic oscillator wave-function. Define

$$A^{(n)}(x, y) = \frac{1}{n^{1/6}} K^{(n)} \left( 2\sqrt{n} + \frac{x}{n^{1/6}}, 2\sqrt{n} + \frac{y}{n^{1/6}} \right) \quad (3.7.6)$$

In view of the basic estimate (3.4.10) in the theory of Fredholm determinants and the crude bound (3.7.5) for the Airy kernel we can by dominated
convergence integrate to the limit on the right side of (3.1.8). By the bound (3.3.8) of Ledoux type, if the limit
\[
\lim_{t' \to +\infty} \lim_{N \to \infty} P \left[ N^{2/3} \left( \frac{\lambda N}{\sqrt{N}} - 2 \right) \not\in (t, t') \text{ for } i = 1, \ldots, N \right]
\] (3.7.7)
exists then the limit (3.1.9) also exists and both limits are equal. Therefore we can take the limit as \( t' \to \infty \) on the left side of (3.1.8) inside the limit as \( n \to \infty \) in order to conclude (3.1.9). We thus concentrate in the sequel on proving (3.1.8) for \( t' < \infty \).

We begin by extending by analyticity the definition of \( K^{(n)} \) and \( A^{(n)} \) to the complex plane \( \mathbb{C} \). Our goal will be to prove the convergence of \( A^{(n)} \) to \( A \) on compact sets of \( \mathbb{C} \), which will imply also the convergence of derivatives. Recall that by part 4 of Lemma 3.2.5,
\[
K^{(n)}(x, y) = \frac{\psi_n(x)\psi'_n(y) - \psi_n(y)\psi'_n(x)}{x - y} + \frac{1}{2} \psi_n(x)\psi_n(y)
\]
so that if we set
\[
\Psi_n(x) := n^{1/12} \psi_n(2\sqrt{n} + \frac{x}{n^{1/6}})
\]
then
\[
A^{(n)}(x, y) = \frac{\Psi_n(x)\Psi'_n(y) - \Psi_n(y)\Psi'_n(x)}{x - y} + \frac{1}{2n^{1/3}} \Psi_n(x)\Psi_n(y).
\]
The following lemma plays the role of Lemma 3.5.2 in the study of the spacing in the bulk. Its proof takes up most of the rest of this section and is deferred.

**Lemma 3.7.8** \( \Psi_n \) is analytic in an open neighborhood of the real line. Further, fix a number \( C > 1 \). Then,
\[
\lim_{n \to \infty} \sup_{u \in \mathbb{C}: |u| < C} |\Psi_n(u) - A_i(u)| = 0.
\] (3.7.9)

The analyticity statement in Lemma 3.7.8 entails the uniform convergence of \( \Psi_n \) to \( A_i \). Together with Lemma 3.4.14, this completes the proof of the theorem. \( \square \)

**Remark 3.7.10** A analysis similar to, but more elaborate than, the proof of Theorem 3.1.7 shows that
\[
\lim_{N \to \infty} P \left[ N^{2/3} \left( \frac{\lambda N}{\sqrt{N}} - 2 \right) \leq t \right]
\]
exists for each positive integer $\ell$ and real number $t$. In other words, the suitably rescaled $\ell$th largest eigenvalue converges vaguely and in fact weakly by elaboration of the analysis given in the next section.

### 3.7.3 Steepest descent: proof of Lemma 3.7.8 and properties of the Airy function

In this subsection, we use the steepest descent method to prove both Lemma 3.7.8 as well as the well known properties of the Airy function that are summarized in the following.

**Lemma 3.7.11** The Airy function $\text{Ai}(x)$ and its derivative $\text{Ai}'(x)$ possess the following asymptotics.

$$\text{Ai}(x) = \frac{e^{-\frac{2}{3}x^{3/2}}}{2\sqrt{\pi}x^{1/4}} \left(1 + O\left(\frac{1}{x^{1/2}}\right)\right) \text{ as } x \uparrow \infty, \quad (3.7.12)$$

$$\text{Ai}'(x) = -\frac{e^{-\frac{2}{3}x^{3/2}}}{2\sqrt{\pi}x^{1/4}} \left(1 + O\left(\frac{1}{x^{1/2}}\right)\right) \text{ as } x \uparrow \infty, \quad (3.7.13)$$

$$\text{Ai}(x) = \frac{\sin\left(\frac{2}{3}|x|^{3/2} + \frac{\pi}{4}\right)}{\sqrt{\pi}|x|^{1/4}} + O\left(\frac{1}{|x|^{7/4}}\right) \text{ as } x \downarrow -\infty. \quad (3.7.14)$$

$$\text{Ai}'(x) = -\frac{\cos\left(\frac{2}{3}|x|^{3/2} + \frac{\pi}{4}\right)|x|^{1/4}}{\sqrt{\pi}} + O\left(\frac{1}{|x|^{5/4}}\right) \text{ as } x \downarrow -\infty. \quad (3.7.15)$$

The steepest descent method is a general, more elaborate version of the method of Laplace discussed in subsection 3.5.1, which is inadequate when oscillatory integrands are involved. Indeed, consider the evaluation of integrals of the form

$$\int f(x)^s g(x)dx,$$

see (3.5.6), in the situation where $f$ and $g$ are analytic functions and the integral is a contour integral. The oscillatory nature of $f$ prevents the use of Laplace’s method. Instead, the oscillatory integrand is tamed by modifying the contour of integration in such a way that $f$ can be written along the contour as $e^f$ with $f$ real, and the oscillations of $g$ at a neighborhood of the critical points of $f$ are slow. In practice, one needs to consider slightly more general versions of this example, in which $g$ itself may depend (weakly) on $s$. Instead of developing a general theory, we provide two examples of computations involving steepest descent arguments, in the proofs of Lemmas 3.7.8 and 3.7.11.
Proof of Lemma 3.7.8 Throughout, we let

\[ x = 2n^{1/2} + \frac{u}{n^{1/6}} = 2n^{1/2} \left(1 + \frac{u}{2n^{2/3}}\right), \quad \Psi_n(u) = n^{1/12}\psi_n(x). \]

We assume throughout the proof that \( n \) is large enough so that \( |u| < C < n^{2/3} \).

Let \( \zeta \) be a complex variable. By reinterpreting formula (3.5.10) above as a contour integral we get the formula

\[ \psi_n(x) = \frac{e^{x^2/4}}{i(2\pi)^{3/4}\sqrt{n!}} \int_L \zeta^n e^{\zeta^2/2 - \zeta \zeta \frac{x}{2}} d\zeta \tag{3.7.16} \]

where \( L \) is the imaginary axis oriented so that height above the real axis is increasing. The main effort in the proof is to modify the contour integral in the formula above in such a way that the leading asymptotic order of all terms in the integrand match, and then keep track of the behavior of the integrand near its critical point. To carry out this program, note that by Cauchy's theorem, we may take \( L \) in (3.7.16) to be any straight line in the complex plane with slope of absolute value greater than 1 oriented so that height above the real axis is increasing (the condition on the slope is to ensure that no contribution appears from the contour near \( \infty \)). Since \( \Re(x) > 0 \) under our assumptions concerning \( u \) and \( n \), we may take \( L \) in (3.7.16) to be the perpendicular bisector of the line segment joining \( x \) to the origin, that is, replace \( \zeta \) by \( (x/2)(1 + \zeta) \), to obtain

\[ \psi_n(x) = \frac{e^{-x^2/8}(x/2)^{n+1}}{i(2\pi)^{3/4}\sqrt{n!}} \int_{-i\infty}^{i\infty} (1 + \zeta)^n e^{(x/2)^2(\zeta^2/2 - \zeta)} d\zeta. \tag{3.7.17} \]

Let \( \log \zeta \) be the principal branch of the logarithm, i.e., the branch real on the interval \((0, \infty)\) and analytic in the complement of the interval \((-\infty, 0]\), and set

\[ F(\zeta) = \log(1 + \zeta) + \zeta^2/2 - \zeta. \tag{3.7.18} \]

Note that the leading term in the integrand in (3.7.17) has the form \( e^{nF(\zeta)} \), where \( \Re(F) \) has a maximum along the contour of integration at \( \zeta = 0 \), and a Taylor expansion starting with \( \zeta^3/3 \) in a neighborhood of that point (this explains the particular scaling we took for \( u \)). Put

\[ \omega = \left(\frac{x}{2}\right)^{2/3}, \quad u' = \omega^2 - n/\omega, \]

where to define fractional powers of complex numbers such as that figuring in the definition of \( \omega \) we follow the rule that \( \zeta^a = \exp(a \log \zeta) \) whenever \( \zeta \) is in the domain of our chosen branch of the logarithm. We remark that
as $n \to \infty$ we have $u' \to u$ and $\omega \sim n^{1/3}$, uniformly for $|u| < C$. Now rearrange (3.7.17) to the form
\[
\Psi_n(u) = \frac{(2\pi)^{1/4}n^{1/12}(x/2)^{n+1/3}e^{-x^2/8}}{\sqrt{n!}} I_n(u)
\]
where
\[
I_n(u) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \omega e^{\omega^2 F(\zeta) - u' \omega \log(1+\zeta)} d\zeta.
\]
To prove (3.7.9) it is enough to prove that
\[
\lim_{n \to \infty} \sup_{|u| < C} |I_n(u) - \text{Ai}(u)| = 0
\]
because we have
\[
\log \frac{n^{1/12}(x/2)^{n+1/3}e^{-x^2/8}}{e^{-n^{2/3}n^{1/2+1/4}}} = \left( n + \frac{1}{3} \right) \log \left( 1 + \frac{u}{2n^{2/3}} \right) = \frac{n^{1/3}u}{2} - \frac{u^2}{8n^{1/3}}
\]
and hence
\[
\lim_{n \to \infty} \sup_{|u| < C} \left| \frac{(2\pi)^{1/4}n^{1/12}(x/2)^{n+1/3}e^{-x^2/8}}{\sqrt{n!}} - 1 \right| = 0
\]
by Stirling’s formula (3.4.7) and some calculus.

To prove (3.7.21), we proceed by a saddle point analysis near the critical point $\zeta = 0$ of $\Re(F(\zeta))$. The goal is to replace complex integration with real integration. This is achieved by making a change of contour of integration so that $F$ is real along that contour. Ideally, we seek a contour so that the maximum of $F$ is achieved at a unique point along the contour. We proceed to find such a contour now, noting that since the maximum of $\Re(f)(\zeta)$ along the imaginary axis is 0 and is achieved at $\zeta = 0$, we may seek contours that pass through 0 and such that $F$ is strictly negative at all other points of the contour.

Turning to the actual construction, consider the wedge-shaped closed set
\[ S = \{ r e^{i\theta} | r \in [0, \infty), \theta \in [\pi/3, \pi/2] \} \]
in the complex plane with “corner” at the origin. For each $\rho > 0$ let $S_\rho$ be the intersection of $S$ with the closed disk of radius $\rho$ centered at the origin and let $\partial S_\rho$ be the boundary of $S_\rho$. For each $t > 0$ and all sufficiently large $\rho$ the curve $F(\partial S_\rho)$ winds exactly once about the point $-t$ and hence by the argument principle there exists a unique solution $\gamma(t) \in S$ of the equation $F(\zeta) = -t$ (see Figure 3.7.1). Clearly $\gamma(0) = 0$ is the unique solution of the equation $F(\zeta) = 0$ in $S$. We have the following.
Lemma 3.7.22 The function $\gamma : [0, \infty) \to S$ has the following properties.

(i) $\lim_{t \to \infty} |\gamma(t)| = \infty$.

(ii) $\gamma(t)$ is continuous for $t \geq 0$ and real analytic for $t > 0$.

(iii) $\gamma(t) = O(t^{1/2})$ as $t \uparrow \infty$.

Proof: (i) follows by noting that $F$ restricted to $S$ is proper, that is for any sequence $z_n \in S$ with $|z_n| \to \infty$ as $n \to \infty$, it holds that $|F(z_n)| \to \infty$.

The real analyticity claim in (ii) follows from the implicit function theorem.

(iii) follows from a direct computation, and together with $\gamma(0) = 0$ implies the continuity claim in (ii).

From Lemma 3.7.22 we obtain the formula

$$I_n(u) = \frac{1}{2\pi i} \int_0^\infty \omega e^{-\omega^3 t} \left( (1 + \gamma(t))^{-\omega u} \gamma'(t) - (1 + \bar{\gamma}(t))^{-\omega u} \bar{\gamma}'(t) \right) dt.$$  

by deforming the contour $-i\infty \to i\infty$ in (3.7.20) to $\gamma - \bar{\gamma}$. After replacing $t$ by $t^3/3n$ in the integral above we obtain the formula

$$I_n(u) = \frac{1}{2\pi i} \int_0^\infty (A_n(t,u) - B_n(t,u))dt$$  

(3.7.23)
where

\[ A_n(t,u) = \omega \exp \left( -\frac{\omega^3 t^3}{3n} \right) \left( 1 + \gamma \left( \frac{t^3}{3n} \right) \right) \gamma' \left( \frac{t^3}{3n} \right) \frac{t^2}{n}, \]

\[ B_n(t,u) = \omega \exp \left( -\frac{\omega^3 t^3}{3n} \right) \left( 1 + \bar{\gamma} \left( \frac{t^3}{3n} \right) \right) \gamma' \left( \frac{t^3}{3n} \right) \frac{t^2}{n}. \]

Put

\[ A(t,u) = \exp \left( -\frac{t^3}{3} - e^{\pi i/3} tu + \pi i/3 \right), \]
\[ B(t,u) = \exp \left( -\frac{t^3}{3} - e^{-\pi i/3} tu - \pi i/3 \right). \]

According to (3.7.1) we have

\[ A_i(u) = \frac{1}{2\pi i} \int_0^\infty (A(t,u) - B(t,u))dt. \tag{3.7.24} \]

A calculus exercise reveals that for any positive constant \( c \),

\[ \lim_{n \to \infty} \sup_{0 \leq t \leq t_0} \sup_{|u| < c} \left| \frac{A_n(t,u)}{A(t,u)} - 1 \right| = 0 \tag{3.7.25} \]

for each \( t_0 \geq 0 \) and clearly the analogous limit formula linking \( B_n(t,u) \) to \( B(t,u) \) holds also. There exist positive constants \( c_1 \) and \( c_2 \) such that

\[ |\log(1 + \gamma(t))| \leq c_1 t^{1/3}, \quad |\gamma'(t)| \leq c_2 \max(t^{-2/3}, t^{-1/2}) \]

for all \( t > 0 \). There exists a positive constant \( n_0 \) such that

\[ \Re(\omega^3) \geq n/2, \quad |\omega| \leq 2n^{1/3}, \quad |u'| < 2c \]

for all \( n \geq n_0 \) and \( |u| < c \). Also there exists a positive constant \( c_3 \) such that

\[ e^{c_3 t^{1/3}} \geq t^{1/6} \]

for \( t \geq 1 \). Consequently there exist positive constants \( c_4 \) and \( c_5 \) such that

\[ |\omega e^{\omega^3 (1 + \gamma(t))^{-\omega'} \gamma'(t)}| \leq c_4 n^{1/3} e^{-ntl/2 + c_5 n^{1/3} t^{-2/3}} \]

and hence

\[ |A_n(t,u)| \leq c_4 e^{-t^{1/6} + c_5 t} \tag{3.7.26} \]

for all \( n \geq n_0, t > 0 \) and \( |u| < c \). Clearly we have the same majorization for \( |B_n(t,u)| \). Integral formula (3.7.24), uniformity of convergence (3.7.25)
and majorization (3.7.26) together are enough to finish the proof of limit formula (3.7.21) and hence of limit formula (3.7.9).

Our second application of the steepest descent method is in the following.

**Proof of Lemma 3.7.11** We explain in some detail how to prove (3.7.12) and then we briefly indicate the modifications of the method needed to prove the remaining three formulas.

Recall (3.1.6) in Definition 3.1.5, and by replacing \( \zeta \) by \( x^{1/2} \zeta \) there, deduce the integral representation

\[
\text{Ai}(x) = \frac{x^{1/2}}{2\pi i} \int_C e^{x^{3/2}H(\zeta)} d\zeta \quad \text{for } x > 0
\]  

(3.7.27)

where

\[ H(\zeta) = \zeta^{3/3} - \zeta. \]

Consider the image

\[ H(C) = \{-t^{3/3} - e^{\pi i/3}t | t \geq 0\} \cup \{-t^{3/3} - e^{-\pi i/3}t | t \geq 0\} \]

of the contour \( C \) under the map \( H \). The contour \( H(C) \) is supported in the left half-plane, passes through the origin, and is oriented so that the branch below (resp., above) the real axis is traversed left to right (resp., right to left). As in the proof of Lemma 3.7.8, we want to modify the contour \( C \) to another contour, say \( C' \), so that \( \Im(H(C')) \) is constant, and the deformed contour \( C' \) “snags” a critical point \( \zeta_0 \) of \( H \), so that the image contour \( H(C') \) satisfies \( \sup \Re(H(C')) \leq \Re(H(\zeta_0)) \). Now we have \( H'(\zeta) = \zeta^2 - 1 \), hence \( \zeta = \pm 1 \) are the critical points of \( H \) and \( H(\pm 1) = \mp 2/3 \) are the corresponding critical values. The best contour \( C' \) for our purposes “snags” the critical point \( \zeta = 1 \) and has image \( H(C') \) running on the real axis from \(-\infty \) to \(-2/3 \) and back.

Again, as in the proof of Lemma 3.7.8, we define the contour \( C' \) implicitly. Consider the closed set

\[ S' = \{1 + re^{i\theta} | r \geq 0, \ \theta \in [\pi/3, \pi/2]\}. \]

For each \( \rho > 0 \) let \( S'_\rho \) be piece-of-pie-shaped closed set obtained by intersecting \( S' \) with the closed disk of radius \( \rho \) about 1 and let \( \partial S'_\rho \) be the boundary of \( S'_\rho \). For all \( t > 0 \) and for all sufficiently large \( \rho \), the image curve \( H(\partial S'_\rho) \) winds exactly once around the point \(-2/3 - t \), see Figure 3.7.2. Consequently, by the argument principle, the equation \( H(\zeta) = -2/3 - t \) has for each \( t > 0 \) a unique solution \( \gamma(t) \in S' \). Also it is clear that \( \gamma(0) = 1 \) is the only solution of the equation \( H(\zeta) = -2/3 \) in \( S' \). We then have the following analog of Lemma 3.7.22, whose proof is similar and therefore omitted.
Lemma 3.7.28 The function \( \gamma : [0, \infty) \to S' \) has the following properties.

(i) \( \lim_{t \to \infty} |\gamma(t)| = \infty \).
(ii) \( \gamma(t) \) is continuous for \( t \geq 0 \) and real analytic for \( t > 0 \).
(iii) \[
\begin{align*}
  \gamma'(t) &= \frac{1}{1 - \gamma(t)^2} \quad \text{for } t > 0 \\
  \gamma(t) &= e^{\pi^2/3} t^{1/3} + O(t^{-2/3}) \quad \text{as } t \uparrow \infty \\
  \gamma'(t) &= e^{\pi^2/3} t^{-2/3} + O(t^{-5/3}) \quad \text{as } t \uparrow \infty \\
  \gamma(t) &= 1 + it^{1/2} + O(t^{3/2}) \quad \text{as } t \downarrow 0 \\
  \gamma'(t) &= it^{-1/2} + O(t^{1/2}) \quad \text{as } t \downarrow 0
\end{align*}
\]

Let \( C' \) be the contour consisting of \( \gamma \) with positive orientation plus \( \bar{\gamma} \) with the opposite orientation. If we replace \( C \) by \( C' \) we leave the integral (3.7.27) unchanged. It follows that

\[
Ai(x) = \frac{e^{-2x^{3/2}/3} x^{1/2}}{2\pi i} \int_0^\infty e^{-x^{3/2} t} (\gamma'(t) - \bar{\gamma}'(t)) dt \quad \text{for } x > 0, \quad (3.7.29)
\]

whence (3.7.12) by an application of the Laplace transform asymptotics, Lemma D.12. Differentiation of (3.7.29) on both sides with respect to \( x \) leads in very similar fashion to (3.7.13). We remark that by developing \( \gamma'(t) \)
in a series of half-integral powers of $t$ for small $t$ one can refine (3.7.12) and (3.7.13) to asymptotic expansions.

The starting point for the proof of (3.7.14) and (3.7.15) is the integral formula

$$\text{Ai}(x) = \frac{|x|^{1/2}}{2\pi i} \int_C e^{\frac{|x|^{3/2} G(\zeta)}{2\pi i}} d\zeta \quad \text{for } x < 0 \tag{3.7.30}$$

where

$$G(\zeta) = \zeta^3/3 + \zeta.$$

To find the best contour $C'$ with which to replace $C$ consider the wedge-shaped closed sets

$$S_1 = \{ i + re^{i\theta} | t \in [0, \infty), \theta \in [\pi/4, \pi/3] \},$$

$$S_2 = \{ i + re^{i\theta} | t \in [0, \infty), \theta \in [\pi, \pi + \pi/4] \}$$

in the complex plane. One then have the following lemma, whose proof is left to Exercise 3.7.33.

**Lemma 3.7.31**

(i) the equation $G(\zeta) = 2t/3 - t$ has for each $t \geq 0$ a unique solution $\gamma_j(t) \in S_j$ for $j = 1, 2$.

(ii) the functions $\gamma_j : [0, \infty) \to S_j$ are well-behaved and in particular have easily worked out asymptotic behavior as $t \downarrow 0$ and $t \uparrow \infty$. work this out and write here the results! ??

(iii) under replacement of the contour $C$ by the contour $C' = \gamma_1 - \gamma_2 + \gamma_2 - \gamma_1$ the integral (3.7.30) does not change.

The proof of (3.7.14) and (3.7.15) follows from Lemma 3.7.31 by the same recipe that led to the proof of (3.7.12).

**Remark 3.7.32** In contrast to the proof of (3.7.12) and (3.7.13) which involves only one critical point of $F(\zeta)$, the proof of (3.7.14) and (3.7.15) involves both critical points of $G(\zeta)$. In more general applications of steepest descent it is possible for several critical points to come into play; also one can have critical points of higher multiplicity.

**Proof of Lemma 3.7.4** By (3.7.12), (3.7.13), (3.7.14), (3.7.15), and the Airy differential equation there exists a positive constant $C$ such that

$$\max(|\text{Ai}(x)|, |\text{Ai}'(x)|, |\text{Ai}''(x)|) \leq Ce^{-x}$$

for all real $x$ and hence

$$|x - y| \geq 1 \Rightarrow |A(x, y)| \leq 2Ce^{-x-y}.$$
for all real \(x\) and \(y\). But by the variant (3.5.7) of Taylor’s theorem noted above we also have
\[
|x - y| < 1 \Rightarrow |A(x, y)| \leq 2C^2e^2e^{-x-y}.
\]
Thus the lemma is proved. \(\square\)

**Exercise 3.7.33** Complete the proof of Lemma 3.7.31.

### 3.8 Analysis of the Tracy-Widom distribution and proof of Theorem 3.1.10

We will study the Fredholm determinant
\[
\Delta = \Delta(t) := 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_t^\infty \int_t^\infty A\left( \begin{array}{c} x_1 \ldots x_k \\ x_1 \ldots x_k \end{array} \right) \prod_{j=1}^{k} dx_j
\]
where \(A(x, y)\) is the Airy kernel and as before we write
\[
A\left( \begin{array}{c} x_1 \ldots x_k \\ y_1 \ldots y_k \end{array} \right) = \det_{i,j=1}^{k} A(x_i, y_j).
\]

We are going to explain why \(\Delta(t)\) is a distribution function, which, together with Theorem 3.1.7, will complete our proof of weak convergence of \(n^{2/3}\left(\frac{\lambda(n)}{\sqrt{n}} - 2\right)\). Further, we are going to link \(\Delta(t)\) to the Painlevé II differential equation.

We begin by putting the study of the Tracy-Widom distribution \(\Delta(t)\) into a framework compatible with the general theory of Fredholm determinants developed in Section 3.4. Let \(\lambda\) denote the unique measure on the real line such that
\[
\int f(x) d\lambda(x) = \int_t^\infty f(x) dx
\]
for all integrable \(f\) on the line (although \(\lambda\) depends on \(t\), we suppress this dependence from the notation). We have then
\[
\Delta = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int \cdots \int A\left( \begin{array}{c} x_1 \ldots x_k \\ x_1 \ldots x_k \end{array} \right) \prod_{j=1}^{k} d\lambda(x_j).
\]

Put
\[
H(x, y) = A(x, y) + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int \cdots \int A\left( \begin{array}{c} x \ldots x_k \\ y \ldots x_k \end{array} \right) \prod_{j=1}^{k} d\lambda(x_j).
\]
In view of the basic estimate (3.4.10) and the crude bound (3.7.5) for the Airy kernel, we must have $\Delta \to 1$ as $t \uparrow \infty$. Similarly we have

$$\sup_{t \geq t_0} \sup_{x,y \in \mathbb{R}} e^{x+y}|H(x,y)| < \infty$$

for each real $t_0$ and

$$\lim_{t \uparrow \infty} \sup_{x,y \in \mathbb{R}} e^{x+y}|H(x,y) - A(x,y)| = 0.$$

Note that because $\Delta$ can be extended to a not-identically-vanishing entire analytic function of $t$, it follows that $\Delta$ vanishes only for isolated real values of $t$. Put

$$R(x,y) = H(x,y)/\Delta,$$

provided of course that $\Delta \neq 0$; a similar proviso applies to each of the following definitions since each involves $R(x,y)$. Put

$$Q(x) = Ai(x) + \int R(x,y) Ai(y)d\lambda(y),$$

$$P(x) = Ai'(x) + \int R(x,y) Ai'(y)d\lambda(y),$$

$$q = Q(t), \quad p = P(t),$$

$$u = \int Q(x) Ai(x)d\lambda(x), \quad v = \int Q(x) Ai'(x)d\lambda(x) = \int P(x) Ai(x)d\lambda(x),$$

last equality by symmetry $R(x,y) = R(y,x)$. Convergence of all these integrals is easy to check. Note that each of the quantities $q$, $p$, $u$ and $v$ tends to 0 as $t \uparrow \infty$. More precisely, $q$ is asymptotic to $Ai(t)$ as $t \uparrow \infty$.

### 3.8.1 The first standard moves of the game

We follow the trail blazed in the discussion of the sine-kernel in Section 3.6. The first few steps we can get through quickly by analogy. We have

$$\frac{\partial}{\partial t} \chi = -\delta_t, \quad \chi' = \delta_t, \quad (3.8.1)$$

$$\frac{\partial}{\partial t} \log \Delta = R(t,t), \quad (3.8.2)$$

$$\frac{\partial}{\partial t} R(x,y) = -R(x,t)R(t,y). \quad (3.8.3)$$
As before we have a relation
\[ R(x, y) = \frac{Q(x)P(y) - Q(y)P(x)}{x - y} = R(y, x) \]  
(3.8.4)
and hence by L'Hôpital's Rule we have
\[ R(x, x) = Q'(x)P(x) - Q(x)P'(x). \]  
(3.8.5)

We have differentiation formulas
\[ \frac{\partial}{\partial t} Q(x) = -R(x,t)Q(t) = -Q(t)R(t,x), \]  
(3.8.6)
\[ \frac{\partial}{\partial t} P(x) = -R(x,t)P(t) = -P(t)R(t,x). \]  
(3.8.7)

Here the Airy function and its derivative are playing the roles previously played by sine and cosine, but otherwise to this point our calculation is running just as before. Actually the calculation to this point is simpler since we are focusing on a single interval of integration rather than on several.

### 3.8.2 The wrinkle in the carpet

As before we introduce the abbreviated notation
\[ A'(x, y) = \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) A(x, y), \quad R'(x, y) = \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) R(x, y), \]

\[ (F \star G)(x, y) = \int F(x, z)G(z, y)d\lambda'(z) = F(x, t)G(t, y). \]

Here’s the wrinkle in the carpet that changes the game in a critical way: \( A' \) does not vanish identically. Instead we have
\[ A'(x, y) = -Ai(x)Ai(y), \]  
(3.8.8)
which is an immediate consequence of the Airy differential equation \( y'' - xy = 0 \). Calculating as before but this time not putting \( A' \) to zero we find that
\[ R' = R \star R + A' + A' \star R + R \star A' \star R. \]

Written out “in longhand” the last equation says that
\[ \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right) R(x, y) = R(x, t)R(t, y) - Q(x)Q(y). \]  
(3.8.9)
The wrinkle “propagates” to produce the extra term on the right. We now have

\[ Q'(x) = \text{Ai}(x) + \int \left( \frac{\partial}{\partial x} R(x, y) \right) \text{Ai}(y) d\lambda(y) \]

\[ = \text{Ai}(x) - \int \left( \frac{\partial}{\partial y} R(x, y) \right) \text{Ai}(y) d\lambda(y) + R(x, t) \int R(t, y) \text{Ai}(y) d\lambda(y) - Q(x) u \]

\[ = \text{Ai}(x) + \int R(x, y) \text{Ai}'(y) d\lambda(y) + R(x, t) \int R(t, y) \text{Ai}(y) d\lambda(y) - Q(x) u \]

\[ = P(x) + R(x, t)Q(t) - Q(x) u. \]  \hspace{1cm} (3.8.10)

Similarly,

\[ P'(x) = x \text{Ai}(x) + \int \left( \frac{\partial}{\partial x} R(x, y) \right) \text{Ai}'(y) d\lambda(y) \]

\[ = x \text{Ai}(x) - \int \left( \frac{\partial}{\partial y} R(x, y) \right) \text{Ai}'(y) d\lambda(y) + R(x, t) \int R(t, y) \text{Ai}'(y) d\lambda(y) - Q(x) v \]

\[ = x \text{Ai}(x) + \int R(x, y) \text{Ai}(y) d\lambda(y) + \int R(x, y) \text{Ai}'(y) d\lambda(y) + R(x, t) \int R(t, y) \text{Ai}'(y) d\lambda(y) - Q(x) v \]

\[ = xQ(x) + \int (P(x)Q(y) - Q(x) P(y)) \text{Ai}(y) d\lambda(y) + R(x, t)P(t) - Q(x) v \]

\[ = xQ(x) + R(x, t)P(t) + P(x)u - 2Q(x)v. \]  \hspace{1cm} (3.8.11)

This is more or less in analogy with the sine-kernel case. But the wrinkle continues to propagate, producing the extra terms involving the quantities \( u \) and \( v \).
3.8.3 Linkage to Painlevé II

The derivatives of the quantities $p$, $q$, $u$ and $v$ with respect to $t$ we denote simply by a prime. We calculate these derivatives as follows. Observe that

$$q' = \frac{\partial}{\partial t} Q(x) \bigg|_{x=t} + Q'(t), \quad p' = \frac{\partial}{\partial t} P(x) \bigg|_{x=t} + P'(t).$$

By adding (3.8.6) to (3.8.10) and (3.8.7) to (3.8.11) we have

$$q' = p - qu, \quad p' = tq + pu - 2qv.$$  \hspace{1cm} (3.8.12)

It follows also via (3.8.5) that

$$\frac{\partial}{\partial t} \log \Delta(t) = R(t, t) = q'p - pq = p^2 - tq^2 - 2pqu + 2q^2v.$$  \hspace{1cm} (3.8.13)

We have

$$u' = \int \left( \frac{\partial}{\partial x} Q(x) \right) \text{Ai}(x) d\lambda(x) + \int Q(x) \text{Ai}(x) d \left( \frac{\partial \lambda}{\partial t} \right)(x) = -Q(t) \int R(t, x) \text{Ai}(x) d\lambda(x) - Q(t) \text{Ai}(t) = -q^2,$$

$$v' = \int \left( \frac{\partial}{\partial x} Q(x) \right) \text{Ai}'(x) d\lambda(x) + \int Q(x) \text{Ai}'(x) d \left( \frac{\partial \lambda}{\partial t} \right)(x) = -Q(t) \int R(t, x) \text{Ai}'(x) d\lambda(x) - Q(t) \text{Ai}'(t) = -pq.$$

We have a first integral

$$u^2 - 2v = q^2;$$

at least it is clear that the $t$-derivative here vanishes, but then then constant of integration has to be 0 because all the functions here tend to 0 as $t \uparrow \infty$. Finally

$$q'' = (p - qu)' = p' - qu' = tq + pu - 2qv - (p - qu)u - q(-q^2)$$

$$= tq + pu - 2qv - pu + qu^2 + q^3 = tq + 2q^3,$$  \hspace{1cm} (3.8.14)

which is Painlevé II.

It remains to prove that $F_2$ is a distribution function. By adding equations (3.8.9) and (3.8.3) we get

$$\left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial t} \right) R(x, y) = -Q(x)Q(y).$$  \hspace{1cm} (3.8.15)

By evaluating both sides at $x = t = y$ and also using (3.8.2) we get

$$\frac{\partial^2}{\partial t^2} \log \Delta = -q^2.$$  \hspace{1cm} (3.8.16)
Let us now write \( q(t) \) and \( \Delta(t) \) to emphasize the \( t \)-dependence. In view of the rapid decay of \( \Delta(t) - 1 \), \( (\log \Delta(t))^\prime \) and \( q(t) \) as \( t \uparrow \infty \) we must have

\[
\Delta = \exp \left( - \int_t^\infty (x - t)q(x)^2 \, dx \right),
\]

whence (finally!) the conclusion that \( F_2(t) = \Delta(t) \) is a distribution function. Together with (3.8.14) and Theorem 3.1.7, this completes the proof of Theorem 3.1.10.

\[\square\]

### 3.9 Bibliographical notes

a lot to do here; what follows is just a random collection moved from the main body of the text.

The book [Wil78] contains an excellent short introduction to orthogonal polynomials as presented in Section 3.2.

Our treatment of Fredholm determinants in Section 3.4 is for the most part adapted from [Tri85]. The latter gives an excellent short introduction to Fredholm determinants and integral equations from the classical viewpoint.

Section 3.3.1 follows the very nicely written paper [HT98]. Formula (3.3.13) is due to Haagerup and Thorbjørnsen, as is the observation that differential equation (3.3.14) implies a recursion for the moments of \( L_N \) discovered by [HZ86] in the course of the latter’s investigation of the moduli space of curves.

The beautiful set of nonlinear partial differential equations (3.6.3), contained in Theorem 3.6.2, is one of the great discoveries reported in [JMMS80]. We’ll call these the Kyoto equations. To derive the equations we followed the simplified approach of [TW93]. The differential equations have a Hamiltonian structure discussed briefly in [TW93]. The same system of partial differential equations is discussed in [Mos80] in a wider geometrical context. See also [HTW93].

Limit formula (3.7.9) appears in the literature as [Sze75, Eq. 8.22.14, p. 201] but is stated there without much in the way of proof. The important paper [PR29] comes very close to supplying a proof of (3.7.9) but strict speaking misses the mark since it is devoted to the asymptotic behavior of the Hermite polynomials \( \delta_n(x) \) for real positive \( x \) only. But still, the ideas of the latter paper are very robust, and it is on these ideas that we base the relatively short self-contained proof of (3.7.9) presented in Section 3.7.3. [I still must have a look at the Bateman Manuscript Project; perhaps the