

High-dimensional Landscapes and Random Matrices

University of Vienna



Peter Mühlbacher

January 14, 2016

Abstract

Contents

0.1	Motivation	2
0.2	Main Results	2
1	Prerequisites	4
1.1	Gaussian Random Fields	4
1.2	GOE	4
1.3	LDP	6
1.4	Morse Theory	7
1.5	Spin Glass Models	7
1.6	Intermediary Results	7
2	Central Identities	9
A	Additional Theorems	10
A.1	Wigner's Semicircle Law	10
A.1.1	Preliminary Reductions	11
A.1.2	Stieltjes Transform	15
A.1.3	Stableness and Concentration of Measure	17
A.1.4	Finding the Semicircle Law	18

Introduction

0.1 Motivation

Given some Hamiltonian $H_{N,p} : S^{N-1} \rightarrow \mathbb{R}$ given by

$$H_{N,p}(\boldsymbol{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{i_1, \dots, i_p=1}^N J_{i_1, \dots, i_p} \sigma_{i_1} \dots \sigma_{i_p},$$

where the J_{i_1, \dots, i_p} are independent centered standard Gaussian random variables and $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_N)$ referred to as states, one is interested in finding the expected number of critical points $Crt_N(u)$ and the restriction to some index k , denoted by $Crt_{N,k}(u)$, in some region in $(-\infty, Nu]$.

0.2 Main Results

Theorem 0.2.1 (Large deviations for $\mathbb{E}[Crt_{N,k}(u)]$).

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[Crt_{N,k}(u)] = \Theta_{k,p}(u) \quad (1)$$

and

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}[Crt_N(u)] = \Theta_p(u), \quad (2)$$

for

$$\Theta_p(u) = \begin{cases} \frac{1}{2} \log(p-1) - \frac{p-2}{4(p-1)} u^2 - I_1(u), & \text{if } u \leq -E_\infty \\ \frac{1}{2} \log(p-1) - \frac{p-2}{4(p-1)} u^2, & \text{if } -E_\infty \leq u \leq 0 \\ \frac{1}{2} \log(p-1), & \text{if } 0 \leq u \end{cases}$$

and

$$\Theta_{k,p}(u) = \begin{cases} \frac{1}{2} \log(p-1) - \frac{p-2}{4(p-1)} u^2 - (k+1)I_1(u), & \text{if } u \leq -E_\infty \\ \frac{1}{2} \log(p-1) - \frac{p-2}{p}, & \text{if } u \geq E_\infty \end{cases},$$

where $E_\infty = E_\infty(p) = 2\sqrt{\frac{p-1}{p}}$ and $I_1 : (-\infty, E_\infty] \rightarrow \mathbb{R}$ is given by

$$I_1(u) = \frac{2}{E_\infty^2} \int_u^{-E_\infty} \sqrt{z^2 - E_\infty^2} dz$$

and is the rate function of the LDP for the smallest eigenvalue of the GOE.

Theorem 0.2.2 (Layered structure). *For all $k \geq 0$ and $\varepsilon > 0$ we have*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left(\left\{ \sum_{i=k}^{\infty} \text{Crt}_{N,i}(-E_k - \varepsilon) > 0 \right\} \right) < 0, \quad (3)$$

where $E_k = E_k(p)$ is chosen (uniquely because of strict monotonicity) such that $\Theta_{k,p}(-E_k) = 0$.

mention
2.12 and
that I
don't
prove it

Chapter 1

Prerequisites

To be able to prove the above results we first need to prove some other theorems which require some

1.1 Gaussian Random Fields

Adler and Taylor (2007) p.121 for

$$\mathbb{E} \left\{ \frac{\partial^k f(s)}{\partial s_{i_1} \dots \partial s_{i_k}} \frac{\partial^k f(t)}{\partial t_{i_1} \dots \partial t_{i_k}} \right\} = \frac{\partial^{2k} C(s, t)}{\partial s_{i_1} t_{i_1} \dots \partial s_{i_k} t_{i_k}} \quad (1.1)$$

1.2 GOE

Definition 1.1. *Symmetric $N \times N$ matrices $H = H_N$ with $\mathbb{E}H_{ij} = 0$ and $\mathbb{E}H_{ij} = 1 + \delta_{ij}$.*

Remark (Density in the space of matrices). *Their density is given by the Gaussian measure*

$$d\mathbb{P}(H) = Z_N^{-1} \exp \left(-N \frac{1}{4} \text{tr} H^2 \right) \quad (1.2)$$

with normalization constant $Z_N = \int d\mathbb{P}(H) \prod_{1 \leq i \leq j \leq N} dH_{ij}$ which is a shorter way of writing $Z_N^{-1} \exp \left\{ -N \frac{1}{4} \left(2 \sum_{i < j}^N H_{ij}^2 + \sum_{i=j}^N H_{ij}^2 \right) \right\}$.

Theorem 1.2.1 (Joint probability density of eigenvalues). *The joint probability density Q_N of the unordered eigenvalues $\{\lambda_i\}_{i=1}^N$ of the GOE is given*

by

$$Q_N(\lambda_1, \dots, \lambda_N) = C_N \prod_{i < j} |\lambda_i - \lambda_j| \exp \left(-N \frac{1}{4} \sum_i \lambda_i^2 \right),$$

where, without loss of generality, we assumed the variances of the entries to be normalized to one and for some $C_N > 0$ which is uniquely determined by normalisation.

Proof. This proof will follow the one given in <http://web.math.princeton.edu/mathlab/projects/ranmatrices/yl/randmtx.PDF>.

Let H be some element of the GOE. Since it is symmetric there is a decomposition in $H = UDU^T$ with U orthogonal and $D = \text{diag}(\lambda_1, \dots, \lambda_N)$. Thus we can write

$$H_{ij} = \sum_k \lambda_i U_{ik} U_{jk} \quad (1.3)$$

and, using the orthogonality of U : $\sum_k U_{ki} U_{kj} = \delta_{ij}$. Using those results one can infer that $\sum_{i,j} H_{ij}^2 = \sum_i \lambda_i^2$.

The key idea is to make use of the change of variables formula to go from $\mathbb{P}(\lambda_1, \dots, \lambda_N, \alpha_1, \dots, \alpha_{N(N-1)/2})$ to $\mathbb{P}(H)$. One should think of the $\{\alpha_i\}_{i=1}^{N(N-1)/2}$ as the parameters that determine the matrix U , which, together with the eigenvalues, uniquely determines H .

To do that we first need some information on the determinant of the Jacobian J of the change of variables: From equation 1.3 we see the linearity of H_{ij} in the eigenvalues λ_k which implies that every entry $\partial H_{ij} / \partial \alpha_k$ is linear in the eigenvalues. Hence, $|J|$ has to be a polynomial of degree $N(N-1)/2$ in each of the eigenvalues. If two eigenvalues coincide U cannot be uniquely determined anymore and thus the inverse of the transformation is not unique, meaning that $|J| = 0$. So the determinant of the Jacobian must vanish for all $\lambda_i = \lambda_j, i \neq j$, which is achieved if it contains a factor $\lambda_i - \lambda_j$. However, there are exactly $N(N-1)/2$ such factors and since that is just the degree of the polynomial it follows that we have completely accounted for J 's dependence on the eigenvalues by writing

$$|J| = \prod_{i < j} (\lambda_i - \lambda_j) h(\alpha_1, \dots, \alpha_{N(N-1)/2}).$$

Now we can write

$$\mathbb{P}(\lambda_1, \dots, \lambda_N, \alpha_1, \dots, \alpha_{N(N-1)/2}) =$$

be careful, Cerny paper uses the convention with N in the exp function!

$$\mathbb{P}(H)|J| = Z_N^{-1} \exp \left(-N \frac{1}{4} \sum_i \lambda_i^2 \right) \left| \prod_{i < j} (\lambda_i - \lambda_j) h(\alpha_1, \dots, \alpha_{N(N-1)/2}) \right|.$$

Integrating out the dependence on the α_k yields the desired result. \square

1.3 LDP

Definition 1.2 (Large deviation principle). *Given some separable completely metrizable topological space¹ X , a sequence of Borel probability measures $\{\mathbb{P}_n\}$ on X is said to satisfy a **large deviation principle** with speed $\{a_n\}$ and rate $I : X \rightarrow [0, \infty]$ if a_n goes to $+\infty$ and I is some lower semi-continuous functional such that for each Borel measurable set $E \subseteq X$ we have*

$$\limsup_n a_n^{-1} \log(\mathbb{P}_n(E)) \leq - \inf_{x \in E} I(x)$$

and

$$\liminf_n a_n^{-1} \log(\mathbb{P}_n(E)) \geq - \inf_{x \in E^\circ} I(x).$$

*The lower semi-continuity implies that the sets $\{x \in X : I(x) \leq c\}$ are closed in X for all $c \geq 0$. If they are also compact for all $c \geq 0$, I is called a **good rate function**.*

We are particularly interested in getting LDPs of the k -th largest eigenvalues of the GOE, but it turns out in order to get those we first need a LDP for the law of the empirical measure $\mu^N = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$.

Theorem 1.3.1 (LDP of Wigner's semicircle law for the GOE). *Let $I(\mu) := \frac{1}{2} \left(\int x^2 d\mu(x) - \Sigma(\mu) - \frac{3}{4} - \frac{1}{2} \log 2 \right)$, for $\Sigma(\mu) := \int \int \log |x - y| d\mu(x) d\mu(y)$. Then:*

1. *I is well defined on the space of probability measures $\mathcal{M}(\mathbb{R})$ on \mathbb{R} endowed with the weak topology and takes values in $[0, \infty]$.*
2. *$I(\mu) = \infty \Leftrightarrow \int x^2 d\mu(x) = \infty$ or $\exists A \subseteq \mathbb{R} : \mu(A) > 0 \wedge \exp\{-\inf_{\nu \in \mathcal{M}(A)} \int \int \log |x - y|^{-1} d\mu(x) d\mu(y)\} = 0$, where the latter condition is usually referred to as the existence of a set of “null logarithmic capacity”.*
3. *I is a good rate function.*
4. *I is a convex function.*

¹Henceforth such a space will be referred to as Polish space.

5. I achieves its unique minimum value at Wigner's semicircle law.

In particular the law of the empirical measure μ^N satisfies a LDP with good rate function I and speed N^2 .

Proof. We will only give a very rough outline of the proof that can be found in Ben Arous and Guionnet (1997). □

should I skip this section?

With this we can state the main result of this section. Note that without loss of generality we set the variances of the entries of the GOE to 1.

Theorem 1.3.2 (LDP for the k -th largest eigenvalue of the GOE). *For each fixed $k \geq 1$, the k -th largest eigenvalue λ_{N-k+1} of the GOE satisfies a LDP with speed N and good rate function*

$$I_k(x) = kI_1(x) = \begin{cases} k \int_2^x \sqrt{\frac{z^2}{4} - 1} dz, & \text{if } x \geq 2 \\ \infty, & \text{else.} \end{cases}$$

Proof. □

prove it

1.4 Morse Theory

1.5 Spin Glass Models

1.6 Intermediary Results

To prove our main results we first need some refinements for the expected values of critical values.

Theorem 1.6.1 (Refinement of $\mathbb{E}[Crt_{N,k}(B)]$). *For all $N, p \geq 2$ and $k \in \{0, \dots, N-1\}$ we have*

$$\mathbb{E}[Crt_{N,k}(B)] = 2\sqrt{\frac{2}{p}}(p-1)^{\frac{N}{2}} \mathbb{E}_{GOE}^N \left[e^{-N\frac{p-2}{2p}(\lambda_k^N)^2} \mathbf{1} \left\{ \lambda_k^N \in \sqrt{\frac{p}{2(p-1)}} B \right\} \right] \quad (1.4)$$

and

$$\mathbb{E}[Crt_N(B)] = 2N\sqrt{\frac{2}{p}}(p-1)^{\frac{N}{2}} \int_{\sqrt{\frac{p}{2(p-1)}} B}^{\infty} \exp \left\{ -\frac{N(p-2)x^2}{2p} \right\} \rho_N(x) dx \quad (1.5)$$

define E_GOE, rho and B Borel sets

Kac-Rice formulae, AT07

sum up previous eq.

Theorem 1.6.2 (LDP for k -th largest eigenvalue). *The k -th largest eigenvalue λ_{N-k+1} of the GOE of dimension N with variance $\sigma^2 N^{-1}(1 + \delta_{ij})$ satisfies an LDP with speed N and a good rate function*

$$I_k(x; \sigma) = kI_1(x; \sigma) = \begin{cases} k \int_{2\sigma}^x \sigma^{-1} \sqrt{\left(\frac{z}{2\sigma}\right)^2 - 1} \, dz, & \text{if } x \geq 2\sigma \\ \infty, & \text{otherwise} \end{cases}. \quad (1.6)$$

define
GOE

define
good rate
function

Chapter 2

Central Identities

prove the
main re-
sults pre-
sented in
the intro-
duction

Appendix A

Additional Theorems

The following section will follow Tao (2010) quite closely while trying to be as self-contained as possible.

A.1 Wigner's Semicircle Law

Definition A.1 (Wigner matrix). *A Wigner matrix M is a complex, Hermitian matrix with independent and identically distributed entries M_{ij} for $i \geq j$ and with mean 0 and variance 1 for $i > j$. The diagonal entries M_{ii} have bounded mean and variance.*

If M_n is an n -dimensional Wigner matrix we know that the operator norm $\|M_n\|_{OP}$ is typically of size $O(\sqrt{n})$, so it is natural to define the empirical spectral distribution (ESD) as follows:

Definition A.2 (ESD).

$$\mu_{\frac{1}{\sqrt{n}}M_n} := \frac{1}{n} \sum_{j=1}^n \delta_{\lambda_j(M_n)/\sqrt{n}},$$

where $\lambda_1(M_n) \leq \dots \leq \lambda_n(M_n)$ are the ordered, real eigenvalues of M_n .

Since we are considering random matrices the ESDs will be random as well and thus it is interesting to ask if there is a measure on the real line μ such that it is the weak limit $\mu_{M_n/\sqrt{n}} \rightharpoonup \mu$ of $\mu_{M_n/\sqrt{n}}$, that is $\int_{\mathbb{R}} \varphi d\mu_{M_n/\sqrt{n}}$ converges in almost surely against $\int_{\mathbb{R}} \varphi d\mu$ for all $\varphi \in C_c(\mathbb{R})$. This can also be derived from the more general definition of convergence in probability or almost surely, but we will not do that here.

Surprisingly such a limit μ exists and is even deterministic.

show the tracy widom law if it is needed and if it doesn't need the semicircle law

Theorem A.1.1 (Wigner’s semicircle law). *Let M_n be the top left $n \times n$ minors of an infinite Wigner matrix, then the ESDs $\mu_{M_n/\sqrt{n}}$ converge in probability¹ to the Wigner semicircle distribution given by*

$$\mu_{sc} := \begin{cases} \frac{1}{2\pi} \sqrt{4 - |x|^2} dx, & \text{if } |x| \leq 2 \\ 0, & \text{else} \end{cases} =: \frac{1}{2\pi} \sqrt{4 - x^2}_+ dx.$$

A rough outline of the proof is given by this list of intermediary results that will be shown:

1. Show that without loss of generality we can set the diagonal elements to zero and bound all other entries by some constant. Additionally some classic concentration of measure results will be shown.
2. To show that $\mu_n \rightarrow \mu$ almost surely, it suffices to show that the respective Stieltjes transforms converge almost surely, pointwise in the upper half plane, i.e. $\mu_n \rightarrow \mu \Leftrightarrow \forall z \in \mathbb{C} : \text{Im}(z) > 0 : s_{\mu_n}(z) \rightarrow s_\mu(z)$ almost surely.
3. The Stieltjes transform $s_n := s_{\mu_{M_n/\sqrt{n}}}$ is “stable in n ”, i.e. $s_n(z) = s_{n-1}(z) + O(\frac{1}{n})$, where O can depend on z and even $s_n(z) - \mathbb{E}s_n(z) \rightarrow 0$ almost surely.
4. Derive the semicircle law by deriving the recursion $\mathbb{E}s_n(z) = -\frac{1}{z + \mathbb{E}s_n(z)} + o(1)$, where, again, $o(1)$ will depend on z and “inverting” the Stieltjes transform.

Remark. *Note that instead of step 4 one could have plugged in the semicircle distribution and simplified the proof by just checking that this is indeed the limit. This is not done here because we want to see how the Stieltjes transform method can be used to derive such a conclusion without knowing about it beforehand.*

Also, there are other proofs (e.g. Boutet de Monvel and Khorunzhiy (2015)) specifically for the GOE/GUE (instead of the more general Wigner matrices) which exploit their symmetries to shorten the proof considerably.

A.1.1 Preliminary Reductions

Lemma A.1.2. *For the matrices M_n as given in A.1.1 it can be assumed without loss of generality that the diagonal entries are zero and the absolute values $|[M_n]_{ij}|$ are bounded by some constant $C > 0$ which does not depend on i, j or n .*

¹Even almost sure convergence can be shown, but we will not do so in this proof.

Proof. For every n define the normalized random variable \bar{X}_n by setting the diagonal elements to zero and bounding every entry by zero², i.e.

$$[\bar{X}_n]_{ij} := \begin{cases} [M_n]_{ij}/\sqrt{n}\mathbf{1}_{|[M_n]_{ij}| \leq C} - \mathbb{E}\left[[M_n]_{ij}\mathbf{1}_{|[M_n]_{ij}| \leq C}\right], & \text{if } i \neq j \\ 0, & \text{if } i = j \end{cases}.$$

Now we want to show that convergence of $\mu_{\bar{X}_n}$ to μ implies convergence of μ_{X_n} to the same limit μ . To do so it suffices to show that $|\int \varphi d\mu_{X_n} - \int \varphi d\mu_{\bar{X}_n}| \rightarrow 0$ for every smooth φ with compact support. In particular every such φ is a Lipschitz function with Lipschitz-constant $\|\varphi'\|_{L^\infty}$. By denoting the eigenvalues of X_n, \bar{X}_n as $\lambda_i, \bar{\lambda}_i$ and invoking the Wielandt-Hoffmann inequality we get the following bound:

$$\begin{aligned} \left| \int \varphi d\mu_{X_n} - \int \varphi d\mu_{\bar{X}_n} \right| &\leq \|\varphi'\|_{L^\infty} \frac{1}{n} \sum_{i=1}^n |\lambda_i - \bar{\lambda}_i| \leq \\ \|\varphi'\|_{L^\infty} \left[\frac{1}{n} \sum_{i=1}^n |\lambda_i - \bar{\lambda}_i|^2 \right]^{1/2} &\leq \|\varphi'\|_{L^\infty} \left[\frac{1}{n} \text{tr}((X_n - \bar{X}_n)^2) \right]^{1/2}. \end{aligned}$$

Now we need to show that for every $\varepsilon, \delta > 0$ we can find a C such that $\mathbb{P}\left(\frac{1}{n} \text{tr}((X_n - \bar{X}_n)^2) > \varepsilon\right) < \delta$ for large enough n .

Using the definitions one sees that

$$\frac{1}{n} \text{tr}((X_n - \bar{X}_n)^2) \leq \frac{1}{n^2} \sum_{i \neq j} \sqrt{n} [X_n - \bar{X}_n]_{ij}^2 + \frac{1}{n} \sum_i [X_n]_{ii}^2,$$

where the second term vanishes almost surely as $n \rightarrow \infty$ according to the strong law of large numbers. So we can use this to bound the probability of the event $\frac{1}{n} \text{tr}((X_n - \bar{X}_n)^2) > \varepsilon$ by

$$\mathbb{P}\left(\frac{1}{n} \text{tr}((X_n - \bar{X}_n)^2)\right) \leq \frac{1}{n^2} \sum_{i \neq j} \mathbb{P}(\sqrt{n} [X_n - \bar{X}_n]_{ij}^2 > \varepsilon).$$

To see that this goes to zero for large C we apply Markov's inequality, yielding

$$\mathbb{P}(\sqrt{n} [X_n - \bar{X}_n]_{ij}^2 > \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}[\sqrt{n} [X_n - \bar{X}_n]_{ij}^2] \leq \frac{\sqrt{n}}{\varepsilon} \left(\mathbb{E}[[X_n]_{ij}^2 \mathbf{1}_{\sqrt{n}|[X_n]_{ij}| > C}] + \mathbb{E}[[X_n]_{ij}^2 \mathbf{1}_{\sqrt{n}|[X_n]_{ij}| \leq C}] \right)$$

²Since the distribution changes we also need to recenter them by subtracting its mean value.

Since the entries $\sqrt{n}[X_n]_{ij} = [M_n]_{ij}$ have finite variance the right hand side will go to zero as $C \rightarrow \infty$ which proves the claim. \square

Theorem A.1.3 (Talagrand's concentration inequality). *Let $K > 0$ and X_1, \dots, X_n be independent complex variables with $|X_i| < K$ for all $1 \leq i \leq n$. Identifying \mathbb{C} with \mathbb{R}^2 , let $F : \mathbb{C}^n \rightarrow \mathbb{R}$ a 1-Lipschitz, convex function. Then for every $\lambda > 0$ one has*

$$\mathbb{P}(|F(X) - MF(X)| \geq \lambda K) \leq C \exp^{-c\lambda^2}$$

and

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq \lambda K) \leq C \exp^{-c\lambda^2}$$

for some absolute constants $c, C > 0$, where $MF(X)$ is the median of $F(X)$.

For the proof refer to <https://terrytao.wordpress.com/2010/01/03/254a-notes-1-concentration-of-measure/#boof0>.

Theorem A.1.4 (Weilandt-Hoffmann inequality). *For Hermitian³ A, B , where $\|B\|_F^2 := \text{tr}(B^2)^{\frac{1}{2}}$ is the Frobenius norm, we have*

$$\sum_{j=1}^n |\lambda_j(A+B) - \lambda_j(A)|^2 \leq \|B\|_F^2.$$

Proof. We fix B to be $B - A$ and show the equivalent statement

$$\sum_{i=1}^n |\lambda_i(A) - \lambda_i(B)|^2 \leq \text{tr}((A-B)^2)$$

and denote by D_M the diagonal matrix $\text{diag}(\lambda_1(M), \dots, \lambda_n(M))$ and U the matrix that diagonalises $B = UD_B U^*$ in the basis of A , such that $\text{tr}(AB) = \text{tr}(D_A U D_B U^*) = \sum_{i,j} \lambda_i(A) \lambda_j(B) |[U]_{ij}|^2$.

Now we have to get some bounds on the “worst case” scenario, that is when $\text{tr}(AB)$ is big. To do so we define $v_{ij} := |[U]_{ij}|^2$ and notice that the above $\text{tr}(AB)$ is linear in v_{ij} . Due to the orthogonality of U one has $\sum_i v_{ij} = \sum_j v_{ij} = 1$ so we can bound $\text{tr}(AB)$ from above by $\sup_{v_{ij} \geq 0, \sum_i v_{ij} = \sum_j v_{ij} = 1} \sum_{i,j} \lambda_i(A) \lambda_j(B) v_{ij}$.

This is an optimization problem of a linear functional over a convex set of doubly stochastic matrices. Hence the maximum is attained at the extreme points which are exactly the permutations. Of all the permutations the identity gives the maximal value which, after writing out the first inequality proves the theorem. \square

³The original Weilandt-Hoffmann inequality holds for normal operators, but we will restrict ourselves to Hermitian ones for simplicity's sake here.

Theorem A.1.5 (Chernoff inequality). *Let X_1, \dots, X_n be independent scalar random variables with $|X_i| \leq K$ almost surely, mean μ_i and variance σ_i^2 . Then for any $\lambda > 0$ there are absolute constants $c, C > 0$ such that*

$$\mathbb{P}(|S_n - \lambda| \geq \lambda\sigma) \leq C \max(e^{-c\lambda^2}, e^{-c\lambda\sigma/K}),$$

where $\lambda := \sum_i \lambda_i$ and $\sigma^2 := \sum_i \sigma_i^2$.

Proof. We begin with some preliminary reductions, namely that it is sufficient to assume X_i to be real valued by taking real and imaginary parts. Furthermore without loss of generality we set $K = 1$ by dividing X_i through K , center X_i (i.e. set μ_i to zero by subtracting its mean), and, by symmetry, only show the upper tail estimate $\mathbb{P}(S_n \geq \lambda\sigma) \leq C \max(e^{-c\lambda^2}, e^{-c\lambda\sigma})$ (with different constants c, C).

To do so we compute the exponential moments $\mathbb{E} \exp(tS_n)$ for some $0 \leq t \leq 1$. By independence of the X_i we have $\mathbb{E} \exp(tS_n) = \prod_{i=1}^n \mathbb{E} \exp(tX_i)$. The $\exp(tX_i)$ can be expanded into a Taylor series $1 + tX_i + O(t^2 X_i^2 \exp(O(t)))$ which, after taking expectation and noting that $|X_i| \leq 1$, yields

$$\mathbb{E} \exp(tX_i) = 1 + O(t^2 \sigma_i^2 \exp(O(t))) = \exp(O(t^2 \sigma_i^2))$$

and hence $\mathbb{E} \exp(tS_n) = \exp(O(t^2 \sigma^2))$. By Markov's inequality we conclude with

$$\mathbb{P}(S_n \geq \lambda\sigma) \leq \exp(O(t^2 \sigma^2) - t\lambda\sigma),$$

which proves the claim after minimising the right hand side in $t \in [0, 1]$. \square

Theorem A.1.6 (McDiarmid's inequality). *Let X_1, \dots, X_n be independent random variables taking values in ranges R_1, \dots, R_n and let $F : R_1 \times \dots \times R_n \rightarrow \mathbb{C}$, such that for every $1 \leq i \leq n$ we have $|F(x_1, \dots, x_i, \dots, x_n) - F(x_1, \dots, x'_i, \dots, x_n)| \leq c_i$. Then for any $\lambda > 0$ one has*

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| > \lambda\sigma) \leq C \exp^{-c\lambda^2},$$

for some absolute⁴ constants $c, C > 0$ and $\sigma = \sum_{i=1}^n c_i^2$.

Proof. Similar as above we may assume that F is real, it suffices to show the upper tail estimate $\mathbb{P}(F(X) - \mathbb{E}F(X) \geq \lambda\sigma^2) \leq C e^{-c\lambda^2}$ and try to bound the exponential moment $\mathbb{E} e^{tF(X)}$.

⁴Constants that maintain the same value wherever they occur. In particular applying McDiarmid's inequality in different settings we do not need to consider C_n, c_n , but can still write C, c .

To get a better idea how the exponential moment behaves we write it in such a way that we can use the fact that F does not fluctuate “too much”, i.e. we consider the conditional expectation $\mathbb{E}(e^{tF(X)}|X_1, \dots, X_{n-1})$ for X_1, \dots, X_{n-1} fixed and write this as

$$\mathbb{E}(e^{tF(X)}|X_1, \dots, X_{n-1}) = \mathbb{E}(e^{tY}|X_1, \dots, X_{n-1})e^{t\mathbb{E}(F(X)|X_1, \dots, X_{n-1})},$$

where $Y := F(X) - \mathbb{E}(F(X)|X_1, \dots, X_{n-1})$.

Now we want to control the first term on the right hand side. Since tY has mean zero and variance $t^2 c_n^2$ we can (just in the proof of Chernoff’s inequality) expand to get a Taylor series and take expectations to get

$$\mathbb{E}(e^{tY}|X_1, \dots, X_{n-1}) \leq e^{O(t^2 c_n^2)}.$$

Integrating out the conditioning (note that X_i are independent) we get the following bound on the exponential moment

$$\mathbb{E}e^{tF(X)} \leq e^{O(t^2 c_n^2)} \mathbb{E}e^{t\mathbb{E}(F(X)|X_1, \dots, X_{n-1})}.$$

Basically we reduced the problem by one dimension and noting that $\mathbb{E}(F(X)|X_1, \dots, X_{n-1})$ is a function $F_{n-1}(X_1, \dots, X_{n-1})$ with the same properties as in the theorem’s assumptions on F the above calculations can be performed iteratively n times to get the upper bound

$$\mathbb{E}e^{tF(X)} \leq \exp\left(\sum_{i=1}^n O(t^2 c_i^2) + t\mathbb{E}F(X)\right).$$

Dividing by $\exp(t\mathbb{E}F(X))$ and applying Markov’s inequality yields

$$\mathbb{P}(F(X) - \mathbb{E}F(X) \geq \lambda\sigma) \leq \exp(O(t^2\sigma^2) - t\lambda\sigma),$$

which, after minimising the right hand side in $0 \leq t \leq 1$, proves the claim. \square

A.1.2 Stieltjes Transform

Definition A.3 (Stieltjes transform). *For a probability measure μ we write s_μ for its Stieltjes transform*

$$\int_{\mathbb{R}} \frac{1}{x - z} d\mu(x).$$

As mentioned above, s_n will be a shorthand for $s_{\mu_{M_n/\sqrt{n}}}$.

Lemma A.1.7 (Properties of the Stieltjes transform). *In the following let μ be some probability measure.*

1. For $z = a + ib$ we have $\text{Im} \frac{1}{x-z} = \frac{b}{(x-a)^2 + b^2} > 0$.
2. s_μ is analytic in $\mathbb{C} \setminus \text{supp}(\mu) \supset \mathbb{C}_+$.
3. We can bound the absolute value as well as the derivatives by $|\frac{d^j}{dz^j} s_\mu(z)| \leq O(|\text{Im}(z)|^{-(j+1)})$ for all $j \in \{0, 1, \dots\}$.

Proof. The first property is trivial, the second one can be seen by integrating s_μ over any contour not containing the support of μ , interchanging the order of integration and noting that integrating $\frac{1}{x-z}$ gives 0 by Cauchy's integral formula ($\frac{1}{x-z}$ being holomorphic outside the support of μ). The Stieltjes transform being holomorphic (and thus analytic) follows by Morera's theorem.

The third property can be obtained by using $\frac{1}{x-z} \leq \frac{1}{\text{Im}(z)}$ and using Cauchy's integral formula integrating this inequality. \square

Corollary A.1.7.1. *From A.1.7.1 it follows that s_μ is a Herglotz function and thus (e.g. Teschl (2009)) $\text{Im}(s_\mu(\cdot + ib)) \rightharpoonup \pi\mu$ as $b \rightarrow 0^+$ in the vague topology or equivalently (by $\overline{s_\mu(z)} = s_\mu(\bar{z})$)*

$$\frac{s_\mu(\cdot + ib) - s_\mu(\cdot - ib)}{2\pi i} \rightharpoonup \mu. \quad (\text{A.1})$$

*Note that this can also be seen by writing $\text{Im}(s_\mu)$ as the convolution $\pi\mu * P_b(a)$ with the Poisson kernels $P_b(x) := \frac{1}{\pi} \frac{b}{x^2 + b^2} = \frac{1}{b} P_1(\frac{x}{b})$ which form a family of approximations to the identity.*

Theorem A.1.8 (Stieltjes continuity theorem). *For μ_n (realisations of) random measures and μ a deterministic measure the following statement holds:*

$\mu_n \rightharpoonup \mu$ almost surely in the vague topology if and only if $s_{\mu_n}(z) \rightarrow s_\mu(z)$ almost surely for every $z \in \mathbb{C}_+$.

Proof. “ \Rightarrow ”: If $\mu_n \rightharpoonup \mu$ in the vague topology almost surely against a deterministic limit μ , then $\forall \phi \in C_c(\mathbb{R}) : \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \phi d\mu_n = \int_{\mathbb{R}} \phi d\mu$ by definition and, by taking the completion, for all bounded, continuous functions vanishing at infinity. The function $x \mapsto \frac{1}{x-z}$ for some $z \in \mathbb{C}$ with $\text{Im}(z) > 0$ is bounded and continuous on \mathbb{R} and hence $s_{\mu_n}(z) \rightarrow s_\mu(z)$ almost surely.

“ \Leftarrow ”: One can, up to an arbitrary small error $\varepsilon > 0$, approximate $\int_{\mathbb{R}} \phi d\mu$ by $\int_{\mathbb{R}} \phi * P_b d\mu = \frac{1}{\pi} \int_{\mathbb{R}} \phi(a) s_{\mu}(a + ib) da$ (and analogously for μ_n). Thus we have $\frac{1}{\pi} \int_{\mathbb{R}} \phi(a) (s_{\mu}(a + ib) - s_{\mu_n}(a + ib)) da$ being equal to the difference (we are interested in) $\int_{\mathbb{R}} \phi d\mu - \int_{\mathbb{R}} \phi d\mu_n$ up to an error ε and by dominated convergence (the Stieltjes transform of a measure is bounded for every $z \in \mathbb{C}_+$ and vanishes outside some compact set) we have convergence in the vague topology. \square

A.1.3 Stableness and Concentration of Measure

In the following we keep using the notation as defined in A.1.1. To show that $s_n(z) = s_{n-1}(z) + O_z(1/n)$ we first need to prove the following theorem:

Theorem A.1.9 (Cauchy’s interlacing theorem). *For any $n \times n$ Hermitian matrix A_n with top left minor A_{n-1} and eigenvalues of descending order ($\lambda_i \geq \lambda_{i+1}$) we have:*

$$\lambda_{i+1}(A_n) \leq \lambda_i(A_{n-1}) \leq \lambda_i(A_n),$$

for all $1 \leq i < n$.

Proof. Using the min-max/max-min theorems ($\lambda_i(A) = \inf_{\dim(V)=n-i+1} \sup_{v \in V, \|v\|=1} \langle Av, v \rangle$ and $\lambda_i(A) = \sup_{\dim(V)=i} \inf_{v \in V, \|v\|=1} \langle Av, v \rangle$ respectively, c.f. Teschl (2009) p.141) and writing S_{n-i+1} for $\{v \in \text{span}\{a_i, \dots, a_n\} : \|v\| = 1\}$, where $A_{n-1}a_j = \lambda_j a_j$ and P an orthogonal projection such that $P^* A_n P = A_{n-1}$ we have

$$\lambda_i(A_{n-1}) = \sup_{v \in S_i, \|v\|=1} v^* A_{n-1} v = \sup_{v \in S_i, \|v\|=1} v^* P^* A_n P v \geq \inf_{\dim(V)=n-i} \sup_{v \in V, \|v\|=1} v^* A_n v = \lambda_{i+1}(A_n),$$

and

$$\lambda_i(A_n) = \inf_{\dim(V)=n-i+1} \sup_{v \in V, \|v\|=1} v^* A_n v \geq \sup_{v \in S_i, \|v\|=1} v^* P^* A_n P v = \sup_{v \in S_i, \|v\|=1} v^* A_{n-1} v = \lambda_i(A_{n-1}).$$

\square

Remembering the identity $\text{Im}(s_{\mu_n(a+ib)}) = \pi \mu * P_b(a)$ and that $\text{supp} \mu_n$ consists of finitely many points, we have $\text{Im}(s_{\mu_n}) = \pi \frac{1}{n} \sum_{\lambda_i} \frac{b}{(\lambda_i - a)^2 + b^2}$ which suggests that it is important to take a closer look at the function $x \mapsto \frac{b}{(x-a)^2 + b^2}$ to compare s_{μ_n} with $s_{\mu_{n-1}}$.

Lemma A.1.10. *For fixed $z \in \mathbb{C}_+$ the Stieltjes transform is “stable” in n , i.e.*

$$s_n(z) = s_{n-1}(z) + O\left(\frac{1}{n}\right)$$

Proof. The idea is to use the Cauchy interlacing law and apply it to the previously mentioned identity by seeing that

$$\sum_{j=1}^{n-1} \frac{b}{\lambda_j(M_{n-1})/\sqrt{n} - a} - \sum_{j=1}^n \frac{b}{\lambda_j(M_n)/\sqrt{n} - a}$$

Up to the dimensional factors these two sums correspond to s_n and s_{n-1} and because of Cauchy's interlacing law this is an alternating sum, giving

$$\sqrt{n(n-1)}s_{n-1}(\sqrt{n/(n-1)}(a+ib)) - ns_n(a+ib) = O\left(\frac{1}{n}\right).$$

Now using the fact that the Stieltjes transform s_n is analytic away from the support of μ_n (A.1.7.2) and using the bound for its derivatives (A.1.7.3) we can approximate $s_{n-1}(\cdot)$ by $s_{n-1}(\sqrt{n/(n-1)}\cdot)$ and hence the statement holds. \square

Using McDiarmid's inequality one gets

$$\mathbb{P}(|s_n(z) - \mathbb{E}s_n(z)| \geq \lambda/\sqrt{n}) \leq C \exp^{-c\lambda^2}, \quad (\text{A.2})$$

for all $\lambda > 0$ and some constants $c, C > 0$.

From the Borel-Cantelli lemma we see that for every z away from the real line $s_n(z) - \mathbb{E}s_n(z)$ converges almost surely to zero since, for some fixed $\varepsilon > 0$, the sum $\sum_n \mathbb{P}(d(s_n - \mathbb{E}s_n, 0) \geq \varepsilon) \leq C \sum_n \exp^{-cn\varepsilon^2} < \infty$ which is obtained by setting $\lambda = \varepsilon\sqrt{n}$.

A.1.4 Finding the Semicircle Law

We start off by using the following identity

$$s_n(z) := \int_{\mathbb{R}} \frac{1}{x-z} d\mu_n(x) = \frac{1}{n} \text{tr} \left(\frac{1}{\sqrt{n}} M_n - zI_n \right)^{-1},$$

which holds for every $z \in \mathbb{C} \setminus \text{supp}(\mu_n)$. Because of the linearity of the trace we also have

$$\mathbb{E}s_n(z) = \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} M_n - zI_n \right)^{-1} \right]_{jj} = \mathbb{E} \left[\left(\frac{1}{\sqrt{n}} M_n - zI_n \right)^{-1} \right]_{nn}, \quad (\text{A.3})$$

where the last equality holds because all of the random variables $[(M_n/\sqrt{n} - zI_n)^{-1}]_{jj}$ have the same distribution.

To calculate one entry of an inverse of a matrix we use Schur's complement, which tells us that (under the assumptions that all the occurring inverse matrices exist)

$$[(M_n/\sqrt{n} - zI_n)^{-1}]_{nn} = - \left(z + \frac{1}{n} X^* \left(\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1} \right)^{-1} X \right)^{-1},$$

where $X \in \mathbb{C}^{n-1}$ is the top right column of M_n with the bottom entry removed and the diagonal elements have been set to zero as justified in A.1.1.

This inverse exists because, for $z \in \mathbb{C}_+$, the imaginary part $Q := \text{Im} \left(\left(\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1} \right)^{-1} \right)$ is positive definite according to the spectral theorem. To see this notice that this holds for arbitrary Hermitian matrices M (instead of $\frac{1}{\sqrt{n}} M_{n-1}$) since their spectrum is on the real line. Thus, by the spectral theorem, we can write $Q = \text{Im} \int \frac{1}{\mu_M - z} dM(\mu_M)$ for some projection valued measure dM and since $x \mapsto \frac{1}{x-z}$ is a Herglotz function its imaginary part will be greater than zero for $z \in \mathbb{C}_+$. As a result the imaginary part of the integrand (which is the imaginary part of the eigenvalues) will be greater than zero. We conclude by noticing that $\text{Im}(z) > 0$ plus something of the form $\langle Qx, x \rangle$ for $Q \geq 0$ will have imaginary part strictly greater than zero and hence the inverse exists.

The next step is to get a better understanding of the resolvent $R := \left(\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1} \right)^{-1}$ and its product $\langle RX, X \rangle$. Clearly R and X are independent, so we may treat R almost like a deterministic matrix. Furthermore, again due to the spectral theorem, $\|R\|_{op}$ is at most $O(1)$. By the strong law of large numbers $\|X\| = O(\sqrt{n})$ almost surely and by Chernoff's inequality this holds with overwhelming probability.

In the following we will show some results for some deterministic matrix A which has roughly the same properties as R (i.e. $A \geq 0$ and $\|A\|_{op} = O(1)$).

Noting that the function $X \mapsto \sqrt{\langle AX, X \rangle}$ is a Lipschitz function with operator norm $O(1)$ and remembering from A.1.1 that we can safely assume the entries to be bounded, we can invoke Talagrand's concentration inequality to get

$$\mathbb{P}(|\sqrt{\langle AX, X \rangle} - \mathbb{M}\sqrt{\langle AX, X \rangle}| \geq \lambda) \leq C \exp^{-c\lambda^2}$$

for any $\lambda > 0$. On the other hand we have $\sqrt{\langle AX, X \rangle} = O(\|X\|) =$

$O(\sqrt{n})$ with overwhelming probability (since the operator norm in the non-deterministic case is only controlled with overwhelming probability). Hence the median $\mathbb{M}\sqrt{\langle AX, X \rangle} = O(\sqrt{n})$ and considering the square $\langle AX, X \rangle$, we conclude that

$$\mathbb{P}(|\langle AX, X \rangle - \mathbb{M}\langle AX, X \rangle| \geq \lambda\sqrt{n}) \leq C \exp^{-c\lambda^2}$$

with some (possibly different) $c, C > 0$, since taking the square of $\sqrt{\langle AX, X \rangle}$ amounts to getting an additional factor $O(\sqrt{n})$ and the median is of the same magnitude as the random variable.

We may replace the median with the expected value, yielding

why?

$$\mathbb{P}(|\langle AX, X \rangle - \mathbb{E}\langle AX, X \rangle| \geq \lambda\sqrt{n}) \leq C \exp^{-c\lambda^2} \quad (\text{A.4})$$

for the the case where A is deterministic and positive definite. One can extend this result to arbitrary matrices of operator norm $O(1)$ by noting that it holds for Hermitian matrices $M = M^* = M_+ - M_-$ ($M_+ \geq 0, M_- \geq 0, \|M\|_{OP} = O(1)$) by applying the triangle inequality and

Remark. By using conditional expectations the above results also hold true for random matrices R with $R \geq 0$ and $\|R\|_{OP} = O(1)$ as long as it is independent of X . The idea is to write all the above statements as $\mathbb{P}(E|\{A = R\}) = \frac{\mathbb{P}(E \cap \{A=R\})}{\mathbb{P}(\{A=R\})} = \mathbb{P}(E)$ for all events E .

polar decomposition doesn't seem to work, for diagonalisable matrices it's clear and the probability of coinciding eigenvalues should go to 0, so maybe that suffices(?)

Now we want to know what $\mathbb{E}\langle RX, X \rangle$ actually is. Because of the linearity of the expectation we write it as $\sum_{i,j=1}^{n-1} \mathbb{E}[\overline{X_i} R_{ij} X_i]$. Since the X_i and R_{ij} are independent we can write that as $\sum_{i,j=1}^{n-1} \mathbb{E}[\overline{X_i} X_i] \mathbb{E}[R_{ij}]$, but as the X_i are iid with mean zero and variance one this double sum simplifies to the expectation of the trace of R

$$\sum_{i=1}^{n-1} \mathbb{E} R_{ii}.$$

Noticing that, up to some “almost correct” normalization factors, $tr(R) = tr((M_{n-1}/\sqrt{n} - zI_{n-1})^{-1})$ is the Stieltjes transform $s_{n-1}(z)$. To be more precise we have

$$tr(R) = n \sqrt{\frac{n}{n-1}} s_{n-1} \left(\sqrt{\frac{n}{n-1}} s_{n-1} z \right),$$

but because of the smoothness of the Stieltjes transform for $z \in \mathbb{C}_+$ these factors do not play a role in the limit $n \rightarrow \infty$, i.e. $tr(R) = n(s_{n-1}(z) + o(1))$.

So using the concentration of measure results for the Stieltjes transform (A.2) and for $\langle AX, X \rangle$ (A.4), remembering that latter also holds for random matrices as long as they are independent (A.1.4), we see that

$$\langle RX, X \rangle = n(s_{n-1}(z) + o(1))$$

with overwhelming probability. Substituting back in Schur's complement (A.3) we get

$$\mathbb{E}s_n(z) = -(z + \mathbb{E}s_n(z))^{-1} + o(1).$$

To say something about the limit we first need to ensure $\lim_{n \rightarrow \infty} \mathbb{E}s_n$ exists. This is indeed the case since $\mathbb{E}s_n$ is locally equicontinuous and locally uniformly bounded away from the real line. Applying the Arzelá-Ascoli theorem we get the existence of a subsequence that converges locally uniformly to a limit s , which is again a Herglotz function. Note that, by the concentration of measure for Stieltjes transforms, there is only one possible limit (so $\lim_{n \rightarrow \infty} \mathbb{E}s_n$ is well defined) and $s_n(z)$ even converges almost surely to $s(z)$. As a further result we get

$$s(z) = (z + s(z))^{-1},$$

where the quadratic formula gives

$$s(z) = -\frac{z \pm \sqrt{z^2 - 4}}{2}$$

From $\lim_{a \rightarrow \infty} s_\mu(a + ib) = 0$ for every Stieltjes transform of a fixed measure μ we see that we need to take $s(z) = \frac{-z + \sqrt{z^2 - 4}}{2}$.

We conclude the proof with the Stieltjes inversion formula, yielding the famous result

$$\frac{s(\cdot + ib) - s(\cdot - ib)}{2\pi i} \rightarrow \frac{1}{2\pi} \sqrt{4 - x^2}_+ dx = \mu_{sc}$$

as $b \rightarrow 0^+$, which can be verified by an application of the Cauchy integral formula.

Bibliography

Adler, R. J. and J. E. Taylor (2007), *Random Fields and Geometry*. Springer-Verlag New York.

Ben Arous, G. and A. Guionnet (1997), “Large deviations for wigner’s law and voiculescu’s non-commutative entropy.” *Probability Theory and Related Fields*, 108, 517–542, URL <http://dx.doi.org/10.1007/s004400050119>.

Boutet de Monvel, A. and A. Khorunzhy (2015), “Some elementary results around the wigner semicircle law.” URL <https://www.physik.uni-bielefeld.de/bibos/old-bibos-site/01-03-035.pdf>.

Tao, T. (2010), “254a, notes 4: The semi-circular law.” URL <http://terrytao.wordpress.com/2010/02/02/254a-notes-4-the-semi-circular-law/>.

Teschl, G. (2009), *Mathematical methods in quantum mechanics*, volume 99 of *Graduate Studies in Mathematics*. American Mathematical Society, URL <http://www.mat.univie.ac.at/~gerald/ftp/book-schroe/schroe.pdf>. With applications to Schrödinger operators.