High-dimensional Landscapes and Random Matrices

University of Vienna

Peter Mühlbacher

December 12, 2015

Abstract

Contents

Introduction

0.1 Motivation

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

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

where the $J_{i_1,\ldots,i_p}$ are independent centered standard Gaussian random variables and $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N)$ 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 $\mathbb{E}[\mathrm{Crt}_{N,k}(u)]$)**.**

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

and

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

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* $I_1 : (-\infty, E_\infty] \to \mathbb{R}$ *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 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*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\left(\left\{\sum_{i=k}^{\infty} Crt_{N,i}(-E_k - \varepsilon) > 0\right\}\right) < 0, \tag{3}$$

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

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

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. *Their density is given by the Gaussian measure*

$$d\mathbb{P}(H) = Z_N^{-1}\exp\{-\frac{1}{4}trH^2\} \tag{1.2}$$

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

1.3 LDP

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.

define E_GOE, rho and B Borel sets

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*

Kac-Rice formulae, AT07

$$\mathbb{E}[Crt_{N,k}(B)] = 2\sqrt{\frac{2}{p}}(p-1)^{\frac{N}{2}}\mathbb{E}^N_{GOE}\left[e^{-N\frac{p-2}{2p}(\lambda_k^N)^2}\mathbf{1}\left\{\lambda_k^N \in \sqrt{\frac{p}{2(p-1)}}B\right\}\right] \tag{1.3}$$

and

sum up previous eq.

$$\mathbb{E}[Crt_N(B)] = 2N\sqrt{\frac{2}{p}}(p-1)^{\frac{N}{2}}\int_{\sqrt{\frac{p}{2(p-1)}}B} exp\left\{-\frac{N(p-2)x^2}{2p}\right\}\rho_N(x)dx \tag{1.4}$$

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 $\sigma^2N^{-1}(1+\delta_{ij})$ satisfies an LDP with speed N and a good rate function*

define GOE

define good rate function

$$I_k(x;\sigma) = kI_1(x;\sigma) = \begin{cases} k\int_{2\sigma}^x \sigma^{-1}\sqrt{(\frac{z}{2\sigma})^2-1}\,\mathrm{d}z, & \textit{if } x \geq 2\sigma \\ \infty, & \textit{otherwise} \end{cases}. \tag{1.5}$$

Chapter 2

Central Identities

prove the main results presented in the introduction

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 $\mathcal{O}(\sqrt{n})$, so it natural to define the empirical spectral distribution (ESD) as follows:

prove it

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 $\lambda_1(M_n) \leq \cdots \leq \lambda_n(M_n)$ 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.

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 almost surely (and thus in probability) to the Wigner semicircle distribution given by*

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

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 $\mu_n \rightharpoonup \mu$ almost surely, it suffices to show that the respective Stieltjes transforms converge almost surely, pointwise in the upper half plane, i.e. $\mu_n \rightharpoonup \mu \Leftrightarrow \forall z \in \mathbb{C} : Im(z) > 0 s_{\mu_n}(z) \to s_\mu(z)$ almost surely.

3. The Stieltjes transform $s_n := s_{\mu_{M_n/\sqrt{n}}}$ is "stable in n", i.e. $s_n(z) = s_{n-1}(z) + \mathcal{O}(\frac{1}{n})$, where $\mathcal{O}$ can depend on z and even $s_n(z) - \mathbb{E}s_n(z) \to 0$ 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 Khorunzhy (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.*

Proof. ∎

decide whether to take more general approach (T.Tao) or the straightforward

Theorem A.1.3 (McDiarmid's inequality)**.** *Let $X_1, \dots, X_n$ be independent random variables taking values in ranges $R_1, \dots, R_n$ and let $F : R_1 \times \dots \times R_n \to \mathbb{C}$, 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 absolute[1] *constants $c, C > 0$ and $\sigma = \sum_{i=1}^n c_i^2$.*

prove it (Feier)

Lemma A.1.4 (Borel-Cantelli)**.** *Given some random variables $(X_n)_{n=1}^\infty, X$ such that $\sum_n \mathbb{P}(d(X_n, X) \geq \varepsilon) < \infty$ for every $\varepsilon > 0$, then X_n converges almost surely to X (in the topology induced by d).*

prove it

Theorem A.1.5 (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 $F : \mathbb{C}^n \to \mathbb{R}$ 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 $F(X)$.

prove it

Theorem A.1.6 (Wielandt-Hoffmann inequality)**.** *For Hermitian*[2] *A, B, where $\|B\|_F^2 := tr(B^2)^{\frac{1}{2}}$ is the Frobenius norm, we have*

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

prove it

[1] Constants 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.

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

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 $Im \frac{1}{x-z} = \frac{b}{(x-a)^2+b^2} > 0$.*
2. *s_μ is analytic in $\mathbb{C} \setminus supp(\mu) \supset \mathbb{C}_+$.*
3. *We can bound the absolute value as well as the derivatives by $|\frac{d^j}{dz^j} s_\mu(z)| \leq \mathcal{O}(|Im(z)|^{-(j+1)})$ 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 $\frac{1}{x-z} \leq \frac{1}{Im(z)}$ and using Cauchy's integral formula integrating this inequality. □

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)) $Im(s_\mu(.+ib)) \rightharpoonup \pi\mu$ as $b \to 0^+$ in the vague topology or equivalently (by $\overline{s_\mu(z)} = s_\mu(\overline{z})$)*

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

Note that this can also be seen by writing $Im(s_\mu)$ 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 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 $s_{\mu_n}(z) \to s_\mu(z)$ almost surely for every $z \in \mathbb{C}_+$.

Proof.

$$\mathbb{P}(\{\limsup_{n\to\infty} d_v(\mu_n,\mu)=0\})=1$$

$$\mathbb{P}(\{\forall\phi\in C_c(\mathbb{R}):\lim_{n\to\infty}\int_{\mathbb{R}}\phi\,\mathrm{d}\mu_n=\int_{\mathbb{R}}\phi\,\mathrm{d}\mu\})=1$$

$$\forall z\in\mathbb{C}_+:\mathbb{P}(\{\lim_{n\to\infty} s_{\mu_n}(z)=s_\mu(z)\})=1$$

"$\Rightarrow$": If $\mu_n \rightharpoonup \mu$ in the vague topology almost surely against a deterministic limit μ, then $\forall\phi\in C_c(\mathbb{R}):\lim_{n\to\infty}\int_{\mathbb{R}}\phi\,\mathrm{d}\mu_n=\int_{\mathbb{R}}\phi\,\mathrm{d}\mu$ 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 $Im(z)>0$ is bounded and continuous on $\mathbb{R}$ and hence $s_{\mu_n}(z)\to s_\mu(z)$ almost surely.

not sure

"$\Leftarrow$": One can, up to an arbitrary small error $\varepsilon>0$, approximate $\int_{\mathbb{R}}\phi\,\mathrm{d}\mu$ by $\int_{\mathbb{R}}\phi*P_b\,\mathrm{d}\mu=\frac{1}{\pi}\int_{\mathbb{R}}\phi(a)s_\mu(a+ib)\,\mathrm{d}a$ (and analogously for μ_n). Thus we have $\frac{1}{\pi}\int_{\mathbb{R}}\phi(a)(s_\mu(a+ib)-s_{\mu_n}(a+ib))\,\mathrm{d}a$ being equal to the difference (we are interested in) $\int_{\mathbb{R}}\phi\,\mathrm{d}\mu-\int_{\mathbb{R}}\phi\,\mathrm{d}\mu_n$ up to an error ε.

In order not to "lose" the almost sure convergence by integration(a summation over uncountable many summands) we approximate it by a Riemann sum (which is possible since we can choose the support I of the test function ϕ as the boundaries of integration). The error for the middle sum is proportional to $\max_{x\in I}(|f''(x)|)|I|^3n^{-2}$, which (for f being the integrand), can be made arbitrarily small. (To be able to control the $\max_{x\in I}(|f''(x)|)$ one may have to approximate the continuous test functions ϕ by smooth (or at least twice differentiable) ones like in every partial differential course.)

not sure if that's the reason

This discretized sum now goes to zero almost surely. □

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 $s_n(z)=s_{n-1}(z)+\mathcal{O}_z(1/n)$ 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 $\{v \in span\{a_i, \dots, a_n\} : \|v\| = 1\}$, 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

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

□

Remembering the identity $Im(s_{\mu_n(a+ib)} = \pi\mu * P_b(a)$ and that $supp\mu_n$ consists of finitely many points, we have $Im(s_{\mu_n}) = \pi \frac{1}{n} \sum_{\lambda_i} \frac{b}{(\lambda_i - a)^2 + b^2}$ 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.*

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

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

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

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 $s_{n-1}(.)$ by $s_{n-1}(\sqrt{n/(n-1).})$ and hence the statement holds. □

Using McDiarmid's inequality one gets

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

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

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

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

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

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

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

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 $Q := Im \left((\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1})^{-1} \right)$ 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 $Q = Im \int \frac{1}{\mu_M - z} dM(\mu_M)$ 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 $Im(z) > 0$ 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 $R := (\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1})^{-1}$ 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 $\mathcal{O}(1)$. By the strong law of large numbers $\|X\| = O(\sqrt{n})$ almost surely.

in Tao's notes it says with "overwhelming" probability - why does he use Chernoff/Hoeffding instead of the strong law of large numbers?

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

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

for any $\lambda > 0$. On the other hand we have $\sqrt{\langle AX, X \rangle} = O(\|X\|) = O(\sqrt{n})$ almost surely.Hence the median $M\sqrt{\langle AX, X \rangle} = O(\sqrt{n})$ and considering the square $\langle AX, X \rangle$, we conclude that

same as above footnote, since it's using this result

which amounts to multiplying $O(\sqrt{n})$, so that's why it appears on the other side ¡– how to make this formal?

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

with some (possibly different) $c, C > 0$.

We mayreplace the median with the expected value, yielding

why?

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

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 $M = M_+ - M_-$ ($M_+ \geq 0, M_- \geq 0, \|M\|_{OP} = O(1)$) by applying the triangle inequality and ...

how to

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.*

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, $tr(R) = tr\left((M_{n-1}/\sqrt{n} - zI_{n-1})^{-1}\right)$ 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 $n \to \infty$, 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 $\lim_{n\to\infty} \mathbb{E}s_n$ 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 $\lim_{n\to\infty} \mathbb{E}s_n$ 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 $\lim_{a\to\infty} s_\mu(a+ib) = 0$ 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

$$\frac{s(.+ib) - s(.-ib)}{2\pi i} \rightharpoonup \frac{1}{2\pi}\sqrt{4-x^2}_+ \,\mathrm{d}x = \mu_{sc}$$

as $b \to 0^+$, 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.

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.