A large deviation principle for Wigner matrices

Tomasz Tkocz*

Abstract

In this note we present a large deviation principle for spectral measures of Wigner's random matrices, which is a result due to G. Ben Arous and A. Guionnet.

The note is an assessed essay for the course *Large deviations and statistical mechanics* given by S. Adams at the University of Warwick, Term 1, 2012/2013.

1 Introduction

Random matrices proved their usefulness in physics and beyond. For instance, in nuclear physics a quantum system, which in the simplest case consists of one heavy atom, is described by a Hamiltonian $\hat{H}$ which is a Hermitian operator acting on a Hilbert space. The eigenvalues are possible energy levels of the nucleus, as it is asserted by the Schrödinger equation. Since $\hat{H}$ acts on an infinite dimensional space, to make the model more tractable, it is assumed that $\hat{H}$ is a finite but large Hermitian matrix. The brilliant idea goes back to E. Wigner who proposed to take for $\hat{H}$ a Gaussian random matrix for in high dimensions such randomness should reveal the properties of generic Hamiltonians which are complicated. This paradigm is now the crux of the theory, and turns out to be very effective (see, e.g. [M]).

Let $\{X_{kl}, Y_{kl}\}_{k \le l \le N}$ be a family of i.i.d. real mean 0 variance 1 Gaussian random variables. An $N \times N$ Hermitian matrix $H_N = [H_{kl}]_{k,l \le N}$, where

$$H_{ij} = \begin{cases} X_{kk}, & \text{if } k = l, \\ (X_{kl} + iY_{kl})/\sqrt{2}, & \text{if } k < l \end{cases}$$

*PhD student under the supervision of Prof. K. Ball and Prof. N. O'Connell, Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK, `t.tkocz@warwick.ac.uk`

is called a *GUE (Gaussian Unitary Ensemble) matrix*, and it is Wigner's model of Hamiltonians of heavy nuclei. The rescaled matrix $\frac{1}{\sqrt{N}}H_N$ is sometimes referred as to a *Gaussian Hermitian Wigner matrix*. Let us denote its eigenvalues, which are real, by $\lambda_1^N, \ldots, \lambda_N^N$, and introduce their empirical measure $L_N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i^N}$. The celebrated Wigner's theorem (see, e.g. [AGZ, Theorem 2.2.1]), which holds in much more general settings as well, states that L_N converges weakly, in probability, to the semicircle law σ,

$$\mathrm{d}\sigma(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{\{|x|\leq 2\}}\mathrm{d}x. \tag{1}$$

In this note we would like to study fluctuations of L_N around σ in terms of large deviations, i.e. what is the probability, on the logarithmic scale, that L_N takes *extreme* values. The relevant result was obtained by G. Ben Arous and A. Guionnet [AG], and it is nicely put forward in [AGZ, Section 2.6.1]. We shall follow the latter. A model which is discussed there is slightly more general than just the model of GUE matrices. For our purpose though, we shall present the proof in the GUE case, and we hope it still suffices to show the main ideas behind large deviations for spectral measures of random matrices. Another result in the spirit of large deviations has been recently obtained in [ChV], where both the different scaling ($1/n$ instead of $1/\sqrt{n}$) and the different ensembles of random matrices are investigated.

In the rest of this section we recall necessary facts on GUE matrices and large deviations, and we set up the notation. In the next sections we state the main result and provide its proof. We finish the note with indicating how one can recover the aforementioned Wigner's theorem.

It is known that the law of eigenvalues $\lambda_1, \ldots, \lambda_N$ of an $N \times N$ GUE matrix rescaled by $1/\sqrt{N}$ is given by

$$\mathbb{P}((\lambda_1, \ldots, \lambda_N) \in A) = \int_A \frac{1}{Z_N}|\Delta(\lambda)|^2 e^{-N\sum_{i=1}^N \lambda_i^2/2}\mathrm{d}\lambda, \tag{2}$$

where $\Delta(x) = \Pi_{1\leq i<j\leq N}(x_i - x_j)$ is the Vandermonde determinant, and Z_N is the normalization constant, computable e.g. thanks to the Selberg integrals

$$Z_N = \left(\frac{2\pi}{N}\right)^{N/2}\prod_{j=1}^N j!. \tag{3}$$

The empirical distribution of the eigenvalues $L_N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ can be seen as a random variable taking values in the space $M_1(\mathbb{R})$ of Borel probability measures on $\mathbb{R}$. We endow this space with the usual weak topology which is compatible with the metric

$$d(\mu,\nu) = \sup \left| \int_{\mathbb{R}} f \mathrm{d}\nu - \int_{\mathbb{R}} f \mathrm{d}\mu \right|,$$

where the supremum is subject to all 1-Lipschitz functions $f\colon \mathbb{R} \longrightarrow \mathbb{R}$ bounded by 1.

Let us now collect some facts on large deviations theory. We refer for instance to [DZ] as a proper exposition of the theory. Given a sequence of random variables $(X_N)_{N\geq 1}$ taking values in some Polish space V, we say that it satisfies a *large deviation principle (LDP)* with *speed* a_N, going to infinity with N, and *rate function* I if

$$I\colon V \longrightarrow [0,\infty] \text{ is lower semicontinuous,} \tag{L}$$

$$\varliminf_{N\to\infty} \frac{1}{a_N} \ln \mathbb{P}(X_N \in G) \geq -\inf_G I, \quad \text{for any open set } G \subset V, \tag{D}$$

$$\varlimsup_{N\to\infty} \frac{1}{a_N} \ln \mathbb{P}(X_N \in F) \leq -\inf_F I, \quad \text{for any closed set } F \subset V. \tag{P}$$

Rate function I is called *good* if its level sets $\{\nu;\ I(\nu) \leq t\}$ are compact. It is not inconceivable that to establish LDP it suffices to estimate the probabilities of small balls as long as we know that the random variables X_N posses some regularity. We say that the sequence $X_1, X_2, \ldots$ is *exponentially tight* if for any $E > 0$ there exists a compact set $K_E \subset V$ such that

$$\varlimsup_{N\to\infty} \frac{1}{a_N} \ln \mathbb{P}(X_N \notin K_E) < -E. \tag{T}$$

The usefulness of this notion is revealed in the following

Theorem 1. *Let $(X_N)_{N\geq 1}$ be a sequence of random variables taking values in some Polish space V. Suppose that it is exponentially tight. If there exists a lower semicontinuous function $I\colon V \longrightarrow [0,\infty]$ such that for all $x \in V$ the following estimates of small ball probabilities hold*

$$\varlimsup_{\epsilon\to 0} \varlimsup_{N\to\infty} \frac{1}{a_N} \ln \mathbb{P}(X_N \in B(x,\epsilon)) \leq -I(x), \tag{Upp}$$

$$\varliminf_{\epsilon\to 0} \varliminf_{N\to\infty} \frac{1}{a_N} \ln \mathbb{P}(X_N \in B(x,\epsilon)) \geq -I(x), \tag{Low}$$

then $(X_N)_{N\geq 1}$ satisfies LDP with rate function I which is good.

Therefore, a usual strategy to prove a LDP is to guess a rate function, first establish the so-called *weak LDP*, i.e. verify lower and upper bounds (Low), (Upp), and at the end check the exponential tightness.

2 Main result

Let us define the function $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R} \cup \{\infty\}$,

$$f(x, y) = \frac{x^2 + y^2}{4} - \ln|x - y|. \tag{4}$$

It is not hard to see that f is bounded below. We set

$$c = \inf_{\mu \in M_1(\mathbb{R})} \int_{\mathbb{R}^2} f(x, y) \mathrm{d}\mu(x) \mathrm{d}\mu(y). \tag{5}$$

We also define $I\colon M_1(\mathbb{R}) \longrightarrow [0, \infty]$

$$I(\mu) = \int_{\mathbb{R}^2} f(x, y) \mathrm{d}\mu(x) \mathrm{d}\mu(y) - c. \tag{6}$$

Observe that $I(\mu) = \int_{\mathbb{R}} \frac{x^2}{2} \mathrm{d}\mu(x) - \Sigma(\mu) - c$, where

$$\Sigma(\mu) = \int_{\mathbb{R}^2} \ln|x - y| \mathrm{d}\mu(x) \mathrm{d}\mu(y) \tag{7}$$

is *Voiculescu's noncommutative entropy* of μ.

The following technical lemma asserts that I is a perfect candidate for a rate function

Lemma 1. *(i) I is well defined.*

(ii) I is lower semicontinuous and good.

(iii) I is a strictly convex function on $M_1(\mathbb{R})$.

(iv) I achieves its minimum value at a unique probability measure on $\mathbb{R}$ which is the Wigner semicircle law σ, (1).

Now we are ready to state the main result

Theorem 2. *Let L_N be a spectral measure of an $N \times N$ GUE matrix rescaled by the factor $1/\sqrt{N}$, $N = 1, 2, \ldots$. Then $(L_N)_{N \geq 1}$ viewed as a sequence of random variables taking values in $M_1(\mathbb{R})$ endowed with the weak topology satisfies LDP with speed N^2 and rate function I defined by* (6).

3 Proofs

We skip the proof of Lemma 1. Though it involves quite cute calculations, it is long. The interested reader may want to consult [AGZ, Lemma 2.6.2] for parts (i) - (iii). We comment on (iv) in section 4.

The proof of Theorem 2 will proceed via the strategy described at the very end of Section 1. In the following subsections we carry out the main steps: bounds (Low) and (Upp), and the exponential tightness of $(L_N)_{N\geq 1}$.

3.1 Upper bound (Upp)

First let us notice that by the definition of L_N,

$$\frac{N-1}{2}\sum_{i=1}^{N}\frac{\lambda_i^2}{2} - \ln\prod_{1\leq i<j\leq N}|\lambda_i-\lambda_j|^2 = \sum_{i\neq j}\frac{\lambda_i^2+\lambda_j^2}{4} - \ln\prod_{i\neq j}|\lambda_i-\lambda_j|$$
$$= \sum_{i\neq j} f(\lambda_i,\lambda_j) = N^2\int_{x\neq y} f(x,y)\mathrm{d}L_N(x)\mathrm{d}L_N(y).$$

As a consequence, we can rewrite the density (2) of the random vector λ,

$$\mathbb{P}(\mathrm{d}\lambda) = \frac{1}{Z_N}e^{-N^2\int_{x\neq y} f(x,y)\mathrm{d}L_N(x)\mathrm{d}L_N(y)}\prod_{i=1}^{N}e^{-\lambda_i^2/2}\mathrm{d}\lambda. \tag{8}$$

Fix $\mu\in M_1(\mathbb{R})$ and $\epsilon>0$. Our goal is to estimate $\mathbb{P}(d(L_N,\mu)\leq\epsilon)$. To deal with the singularities of $\ln|x-y|$ we truncate $f_M = f\wedge M$, $M\geq 0$. It is convenient to introduce and work with the nonnormalized measure $\bar{\mathbb{P}}(\cdot) = Z_N\mathbb{P}(\cdot)$. Since $f_M\leq f$, we have

$$\bar{\mathbb{P}}(d(L_N,\mu)\leq\epsilon) \leq \int_{d(L_N,\mu)\leq\epsilon} e^{-N^2\int_{x\neq y} f_M(x,y)\mathrm{d}L_N(x)\mathrm{d}L_N(y)}\prod_{i=1}^{N}e^{-\lambda_i^2/2}\mathrm{d}\lambda.$$

To lighten the notation we denote any product measure $\nu\otimes\nu$ by ν^2. Note that $L_N^2(x=y) = 1/N$, $\mathbb{P}$ almost surely as under the Lebesgue measure λ_i's are almost surely distinct. So,

$$\int f_M\mathrm{d}L_N^2 = \int_{x\neq y} f_M\mathrm{d}L_N + M/N,$$

hence,

$$\bar{\mathbb{P}}(d(L_N,\mu)\leq\epsilon) \leq e^{MN}\int_{d(L_N,\mu)\leq\epsilon} e^{-N^2\int f_M\mathrm{d}L_N^2}\prod e^{-\lambda_i^2/2}\mathrm{d}\lambda$$
$$\leq e^{MN}e^{-N^2\inf_{d(\nu,\mu)\leq\epsilon}\int f_M\mathrm{d}\nu^2}\int\prod e^{-\lambda_i^2/2}\mathrm{d}\lambda.$$

Taking the logarithm we obtain

$$\overline{\lim_{\epsilon\to 0}}\,\overline{\lim_{N\to\infty}}\,\frac{1}{N^2}\bar{\mathbb{P}}(d(L_N,\mu)\le\epsilon)\le -\underline{\lim}_{\epsilon\to 0}\inf_{d(\nu,\mu)\le\epsilon}\int f_M\mathrm{d}\nu^2 = -\int f_M\mathrm{d}\mu^2,$$

where the last equality holds because f_M is continuous and bounded, and therefore $\nu\mapsto\int f_M\mathrm{d}\nu^2$ is continuous with respect to the weak topology. Applying the Lebesgue monotone convergence theorem ($f_M\nearrow f$, and f, f_M are bounded below!) we get $\int f_M\mathrm{d}\mu^2\nearrow\int f\mathrm{d}\mu^2$.

Note that formally, $Z_N=\bar{\mathbb{P}}(d(L_N,\mu)\le\epsilon=\infty)$, thus taking above $\epsilon=\infty$ instead of $\overline{\lim}_{\epsilon\to 0}$ we find that

$$\overline{\lim}(1/N^2)\ln Z_N\le -\inf_{\mu\in M_1(\mathbb{R})}\int f_M\mathrm{d}\mu^2.$$

For a fixed $\delta>0$, for each M we can find a measure $\mu_{M,\delta}$ such that

$$-\inf_{\mu\in M_1(\mathbb{R})}\int f_M\mathrm{d}\mu^2<\delta-\int f_M\mathrm{d}\mu^2_{M,\delta}.$$

As a consequence, $\int f_M\mathrm{d}\mu^2_{M,\delta}\le\delta+\inf_{\mu\in M_1(\mathbb{R})}\int f\mathrm{d}\mu^2=\text{const}<\infty$. Using this it can be shown (exercise!) that the sequence $(\mu_{M,\delta})_{M\ge 1}$ is tight, so by Prokhorov's theorem we can assume without loss of generality that $\mu_{M,\delta}\longrightarrow\mu_\delta$ weakly. Then the monotonicity $f_M\le f_{M+1}$ yields $\int f_M\mathrm{d}\mu^2_{M,\delta}\ge\int f_{M_0}\mathrm{d}\mu^2_{M,\delta}\longrightarrow\int f_{M_0}\mathrm{d}\mu^2_\delta\longrightarrow\int f\mathrm{d}\mu^2_\delta\ge\inf_{\mu\in M_1(\mathbb{R})}\int f\mathrm{d}\mu^2$. Since δ is arbitrary, we obtain

$$\overline{\lim}(1/N^2)\ln Z_N\le -\inf_{\mu\in M_1(\mathbb{R})}\int f\mathrm{d}\mu^2.$$

Summarizing, we have shown that

$$\overline{\lim_{\epsilon\to 0}}\,\overline{\lim_{N\to\infty}}\,\frac{1}{N^2}\bar{\mathbb{P}}(d(L_N,\mu)\le\epsilon)\le -\int f\mathrm{d}\mu^2, \tag{9}$$

$$\overline{\lim_{N\to\infty}}\,\frac{1}{N^2}\ln Z_N\le -c. \tag{10}$$

We will conclude desired bound (Upp) for $\mathbb{P}$ when we establish the analogous estimates from below for Z_N in the next subsection.

3.2 Lower bound (Low)

We prove that for all $\mu\in M_1(\mathbb{R})$

$$\underline{\lim}_{\epsilon\to 0}\,\underline{\lim}_{N\to\infty}\,\frac{1}{N^2}\ln\bar{\mathbb{P}}(d(L_N,\mu)\le\epsilon)\ge -\int f\mathrm{d}\mu^2. \tag{11}$$

Incidentally, since $Z_N \geq \bar{\mathbb{P}}(d(L_N, \mu) \leq \epsilon)$ this immediately implies that

$$\varliminf_{N\to\infty} \frac{1}{N^2} \ln Z_N \geq -c. \tag{12}$$

Fix $\mu \in M_1(\mathbb{R})$ and $\epsilon > 0$. Without loss of generality we assume that $\int f \mathrm{d}\mu^2 < \infty$. Obviously it implies that μ has no atoms. Moreover, since

$$f(x, y) \geq (x^2 + y^2)/8 - 4, \tag{13}$$

which follows by $\ln|x-y| \leq \ln(|x|+1)+\ln(|y|+1) \leq |x|+|y|$,the assumption of a nice integrability $\int f \mathrm{d}\mu^2 < \infty$ also implies that $\int x^2 \mathrm{d}\mu(x) < \infty$.

Now we approximate μ with a discrete measure. Given N let us define the sequence $(x_{i,N})_{i\leq N}$

$$\begin{aligned} x_{1,N} &= \inf\left\{x;\ \mu(-\infty, x] \geq 1/(N+1)\right\}, \\ x_{i+1,N} &= \inf\left\{x \geq x_{i,N};\ \mu(x_{i,N}, x] \geq 1/(N+1)\right\}, \quad i \leq N-1, \end{aligned}$$

i.e. $\{(i/(N+1), x_{i,N}), i \leq N\}$ is a discrete approximation of the *inverse* of the distribution function of μ. Since μ has no atoms, eventually

$$d\left(\mu, \frac{1}{N}\sum_{i=1}^{N} \delta_{x_{i,N}}\right) < \epsilon/2.$$

Thus,

$$A = \{\lambda;\ |\lambda_i - x_{i,N}| < \epsilon/2, i \leq N\} \subset \{\lambda;\ d(L_N, \mu) \leq \epsilon\},$$

which intuitively means that if the atoms of measure L_N are close to the atoms of the approximation of μ, then μ itself is close to L_N. Therefore,

$$\bar{\mathbb{P}}\left(d(L_N, \mu) \leq \epsilon\right) \geq \int_A \prod_{i<j} |\lambda_i - \lambda_j|^2 e^{-N\sum \lambda_i^2/2} \mathrm{d}\lambda.$$

Shifting the variables $\lambda_i \mapsto \lambda_i + x_{i,N}$ we get

$$\bar{\mathbb{P}}\left(d(L_N, \mu) \leq \epsilon\right) \geq \int_{\bigcap_i\{|\lambda_i|<\epsilon/2\}} \prod_{i<j} |x_{i,N} - x_{j,N} + \lambda_i - \lambda_j|^2 e^{-N\sum (x_{i,N}+\lambda_i)^2/2} \mathrm{d}\lambda.$$

Note that $(x_{i,N})$ is increasing. On the set $B = \{\lambda_1 < \ldots < \lambda_N\}$ we thus have $|x_{i,N} - x_{j,N} + \lambda_i - \lambda_j| \geq |x_{i,N} - x_{j,N}| \vee |\lambda_i - \lambda_j|$ for $i < j$, so splitting the product $\prod_{1\leq i<j\leq N} = \prod_{i\leq N-1, j=i+1} \times \prod_{2\leq i+1<j\leq N}$ we obtain on B

$$\prod_{i<j} |x_{i,N} - x_{j,N} + \lambda_i - \lambda_j|^2 \geq \prod_{i\leq N-1} |x_{i,N} - x_{i+1,N}| \cdot |\lambda_i - \lambda_{i+1}| \times \prod_{i+1<j} |x_{i,N} - x_{j,N}|^2.$$

As a result,

$$\bar{\mathbb{P}}\left(d(L_N,\mu)\le\epsilon\right)\ge\left(\prod_{i+1<j}|x_{i,N}-x_{j,N}|^2\prod_{i\le N-1}|x_{i,N}-x_{i+1,N}|e^{-N\sum x_{i,N}^2/2}\right)\times$$
$$\left(\int_{B\cap\bigcap_i\{|\lambda_i|<\epsilon/2\}}\prod_{i\le N-1}|\lambda_i-\lambda_{i+1}|e^{-N\sum((x_{i,N}+\lambda_i)^2-x_{i,N}^2)/2}\mathrm{d}\lambda\right)=Q_N\times R_N$$

Let us deal with the second term R_N. Clearly, $N\sum|(x_{i,N}+\lambda_i)^2-x_{i,N}^2|/2\le N(\epsilon/2)\sum|x_{i,N}|+N^2\epsilon^2/8$ when $|\lambda_i|<\epsilon/2$. Moreover, thanks to $\int|x|\mathrm{d}\mu\le\sqrt{\int|x|^2\mathrm{d}\mu}<\infty$, it is not hard to see that by the construction of the sequence $(x_{i,N})$ we can write $\frac{1}{N+1}\sum|x_{i,N}|\le\int|x|\mathrm{d}\mu+o(1)$. Thus

$$\underline{\lim}_{N\to\infty}\frac{1}{N^2}\ln R_N\ge-\frac{\epsilon^2}{8}-\frac{\epsilon}{2}\int|x|\mathrm{d}\mu$$
$$+\underline{\lim}_{N\to\infty}\frac{1}{N^2}\ln\int_{B\cap\bigcap_i\{|\lambda_i|<\epsilon/2\}}\prod_{i\le N-1}|\lambda_i-\lambda_{i+1}|\mathrm{d}\lambda.$$

The last integral against $\mathrm{d}\lambda$ can be simply estimated. Introducing $u_i=\lambda_{i+1}-\lambda_i$ and noticing that $B\cap\bigcap_i\{|\lambda_i|<\epsilon/2\}\supset\bigcap_i\{0<u_i<\epsilon/(2N)\}=C$ we find

$$\int_{B\cap\bigcap_i\{|\lambda_i|<\epsilon/2\}}\prod_{i\le N-1}|\lambda_i-\lambda_{i+1}|\mathrm{d}\lambda\ge\int_C\prod_{i\le N-1}u_i\mathrm{d}u=\left(\frac{\epsilon^2}{4N^2}\right)^{N-1}\frac{\epsilon}{2N}.$$

This yields

$$\underline{\lim}_{\epsilon\to0}\,\underline{\lim}_{N\to\infty}\frac{1}{N^2}\ln R_N\ge0.$$

Now we handle the first term Q_N,

$$\frac{1}{N^2}\ln Q_N=\frac{2}{N^2}\sum_{i<j\le N-1}\ln|x_{i,N}-x_{j+1,N}|+\frac{1}{N^2}\sum_{i\le N-1}\ln|x_{i,N}-x_{i+1,N}|$$
$$-\frac{1}{N}\sum_{i\le N}\frac{x_{i,N}^2}{2}.$$

Again, the construction of the approximating sequence $(x_{i,N})$ and the nice integrability of μ assure us that $\frac{1}{N+1}\sum x_{i,N}^2/2\le\int(x^2/2)\mathrm{d}\mu+o(1)$. In fact,

$\int |x|^2 \mathrm{d}\mu(x) < \infty$ also implies that $\Sigma(\mu) < \infty$ (recall (7) for the definition!) as $\ln|x-y| \le \ln(|x|+1) + \ln(|y|+1) \le |x|+|y|$. Observe that

$$\begin{aligned}
&\frac{1}{(N+1)^2} \sum_{i<j\le N-1} \ln|x_{i,N} - x_{j+1,N}| + \frac{1}{2(N+1)^2} \sum_{i\le N-1} \ln|x_{i,N} - x_{i+1,N}| \\
&= \sum_{1\le i\le j\le N-1} \ln(x_{j+1,N} - x_{i,N}) \int_{\substack{x\in[x_{i,N},x_{i+1,N}] \\ y\in[x_{j,N},x_{j+1,N}]}} \mathbf{1}_{\{x<y\}} \mathrm{d}\mu(x)\mathrm{d}\mu(y) \\
&\ge \sum_{1\le i\le j\le N-1} \int_{\substack{x\in[x_{i,N},x_{i+1,N}] \\ y\in[x_{j,N},x_{j+1,N}]}} \mathbf{1}_{\{x<y\}} \ln(y-x) \mathrm{d}\mu(x)\mathrm{d}\mu(y) \\
&= \int_{x_{1,N}\le x<y\le x_{N,N}} \ln(y-x)\mathrm{d}\mu(x)\mathrm{d}(y).
\end{aligned}$$

By the Lebesgue monotone convergence theorem, the right hand side tends to $\Sigma(\mu)/2$, hence taking $\underline{\lim}$ we get

$$\underline{\lim}_{N\to\infty} \frac{1}{N^2} \ln Q_N \ge \Sigma(\mu) - \int \frac{x^2}{2} \mathrm{d}\mu(x) = -\int f \mathrm{d}\mu^2.$$

This finishes the proof of (9).

3.3 Conclusion of the proof of the upper and lower bounds

Recall that $\bar{\mathbb{P}}(\cdot) = Z_N \mathbb{P}(\cdot)$. Combining (10) and (12) yields

$$\lim_{N\to\infty} (1/N^2) \ln Z_N = -c.$$

This along with (9) easily imply (Upp), and similarly, (11) implies (Low).

3.4 Exponential tightness (T)

It is a nice exercise to prove that

$$\frac{1}{N^2} \ln Z_N \xrightarrow[N\to\infty]{} -1, \tag{14}$$

knowing (3) (e.g., one may find the Stolz-Cesàro theorem useful). Hence, $Z_N \ge e^{-2N^2}$ eventually.

Note that trivially,

$$2\int x^2 \mathrm{d}L_N = \int (x^2+y^2)\mathrm{d}L_N^2 \le \int_{x\ne y} (x^2+y^2)\mathrm{d}L_N^2 + \frac{1}{N}\int 2x^2 \mathrm{d}L_N.$$

Thus for $N \geq 2$, $\int x^2 \mathrm{d}L_N \leq \int_{x \neq y} (x^2+y^2) \mathrm{d}L_N^2$. Now fix $t > 0$. With the aid of (13), $x^2 + y^2 \leq 8(f(x,y)+4)$, so

$$\mathbb{P}\left(\int x^2 \mathrm{d}L_N > t\right) \leq \mathbb{P}\left(\int_{x \neq y} f(x,y) \mathrm{d}L_N^2 > t/8 - 4\right).$$

Using nice formula (8) for the density of λ we get

$$\mathbb{P}\left(\int x^2 \mathrm{d}L_N > t\right) \leq e^{-N^2(t/8-4)} e^{2N^2} (\sqrt{2\pi})^N.$$

We would like to show (T). It suffices to take $K_E = \left\{\mu;\ \int x^2 \mathrm{d}\mu \leq t(E)\right\}$ for $t(E)$ large enough. (K_E is a closed set as it is the intersection of closed sets $\left\{\mu;\ \int (x^2 \wedge n) \mathrm{d}\mu \leq t(E)\right\}$, $n \geq 1$; moreover if $\mu_m \in K_E$, then it is not hard to see that the sequence $(\mu_m)_{m \geq 1}$ is tight, so by Prokhorov's theorem we get compactness.)

4 Wigner's theorem

Suppose we know that the semicircle law σ is the unique minimum of I. Then for a fixed $\epsilon > 0$ applying (P) for the set $F = \{d(\mu,\sigma) \geq \epsilon\}$ (σ is compactly supported, thus F is closed) we immediately get that $\mathbb{P}(d(L_N,\sigma) \geq \epsilon) \leq e^{-\delta N^2}$, where $\delta = \delta(\epsilon) = \inf_{d(\mu,\sigma)\geq\epsilon} I(\mu)$ is a positive constant. Therefore L_N weakly converges to σ, in probability (with rate e^{-N^2}).

This *short* argument justifying Wigner's theorem hinges on (iv) of Lemma 1. Let us briefly sketch the idea of the proof of the latter. Knowing that there exists the unique minimum $\tilde{\sigma}$ of I, which is guaranteed by strict convexity, it is rather straightforward to give a characterization of $\tilde{\sigma}$. This is a compactly supported measure such that

$$\int \ln|x-y| \mathrm{d}\tilde{\sigma}(y) \leq \frac{x^2}{2} - 1,$$

with the equality iff $x \in \operatorname{supp} \tilde{\sigma}$ (see [AGZ, Lemma 2.6.2 (e)] for the proof). Thus, in order to establish that σ is the unique minimum, it is enough to verify that σ satisfies this inequality. To achieve this, it seems that some cumbersome calculations cannot be omitted; the interested reader is referred to [AG, Lemma 2.7].

References

[AGZ] G. W. Anderson, A. Guionnet and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, 118, Cambridge Univ. Press, Cambridge, 2010.

[AG] G. Ben Arous and A. Guionnet, Large deviations for Wigner's law and Voiculescu's non-commutative entropy, *Probab. Theory Related Fields* **108** (1997), no. 4, 517–542.

[ChV] S. Chatterjee and S. R. S. Varadhan, Large deviations for random matrices, *Commun. Stoch. Anal.* **6** (2012), no. 1, 1–13.

[DZ] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, corrected reprint of the second (1998) edition, Stochastic Modelling and Applied Probability, 38, Springer, Berlin, 2010.

[M] M. L. Mehta, *Random matrices*, third edition, Pure and Applied Mathematics (Amsterdam), 142, Elsevier/Academic Press, Amsterdam, 2004.