A Sharp Small Deviation Inequality for the Largest Eigenvalue of a Random Matrix

Guillaume Aubrun

Université de Paris 6, Institut de Mathématiques, Équipe d'Analyse Fonctionnelle, Boite 186, 4 place Jussieu, 75005 PARIS

Summary. We prove that the convergence of the largest eigenvalue λ_1 of a $n \times n$ random matrix from the Gaussian Unitary Ensemble to its Tracy-Widom limit holds in a strong sense, specifically with respect to an appropriate Wasserstein-like distance. This unifying approach allows us both to recover the limiting behaviour and to derive the inequality $\mathbb{P}(\lambda_1 \geqslant 2+t) \leqslant C\exp(-cnt^{3/2})$, valid uniformly for all n and t. This inequality is sharp for "small deviations" and complements the usual "large deviation" inequality obtained from the Gaussian concentration principle. Following the approach by Tracy and Widom, the proof analyses several integral operators, which converge in the appropriate sense to an operator whose determinant can be estimated.

Key words: Random matrices, largest eigenvalue, GUE, small deviations.

Introduction

Let $\mathcal{H}_n$ be the set of n-dimensional (complex) Hermitian matrices. The general element of $\mathcal{H}_n$ is denoted $A^{(n)}$, and its entries are denoted $(a_{ij}^{(n)})$.

We exclusively focus on the Gaussian Unitary Ensemble GUE, which can be defined by the data of a probability measure $\mathbb{P}_n$ on $\mathcal{H}_n$ which fulfills the following conditions

1. The n^2 random variables $(a_{ii}^{(n)}), (\Re a_{ij}^{(n)})_{i<j}, (\Im a_{ij}^{(n)})_{i<j}$ are independent,
2. $\forall i, a_{ii}^{(n)}$ follows the Gaussian law $N(0,1/n)$,
3. $\forall i<j, \Re a_{ij}^{(n)}$ and $\Im a_{ij}^{(n)}$ follow the Gaussian law $N(0,1/2n)$.

The measure $\mathbb{P}_n$ is uniquely determined by these three conditions because of the extra symmetry constraint $a_{ij} = \overline{a_{ji}}$; it can also be made explicit. $\mathcal{H}_n$ is a vector space on which the scalar product $\langle u,v\rangle := \mathrm{tr}(uv)$ induces an Euclidean structure, hence a Lebesgue measure. The probability measure $\mathbb{P}_n$ has a density with respect to this Lebesgue measure, which can be shown to equal

$$\mathrm{d}\mathbb{P}_n := \frac{1}{c_n} \exp(-\frac{n}{2} \mathrm{tr} M^2) \mathrm{d}M$$

where c_n is a normalization constant.

The GUE has the wonderful property of invariance under rotation : indeed the measure $\mathbb{P}_n$ is invariant under the conjugation action of the unitary group. This makes calculations easier and is very useful for the study of eigenvalues, which are also invariant under the same action.

From now on, "random matrix" means "element of $(\mathcal{H}_n, \mathbb{P}_n)$" seen as a probability space. Let $(\Omega, \mathbb{P})$ denote the product of all these probability spaces $\prod_{i=n}^{\infty} (\mathcal{H}_n, \mathbb{P}_n)$; an element of Ω is a sequence of random matrices, the nth matrix being of size n. However, for all the questions we shall consider, the relationships between the $\mathcal{H}_n$'s for different n's are immaterial.

For more background, we refer the reader to the monograph [10]. Throughout the argument, $C, c, c', \dots$ will stand for positive universal constants, independent of the dimension and any other parameters that may be involved. The values of these constants may change from occurrence to occurrence.

Let $\lambda_1(A^{(n)}) \geqslant \lambda_2(A^{(n)}) \geqslant \dots \geqslant \lambda_n(A^{(n)})$ be the ordered eigenvalues of a random matrix $A^{(n)}$. The global asymptotic behavior of these eigenvalues is well-known. The most famous result in this topic is the semi-circle law, which can be stated as follows : let $N(A^{(n)})$ be the probability measure on $\mathbb{R}$ derived from the random matrix $A^{(n)}$ in the following way (δ_x denotes the Dirac mass at point x)

$$N(A^{(n)}) := \sum_{k=1}^{n} \delta_{\lambda_k(A^{(n)})}$$

Then, $\mathbb{P}$-almost surely, the sequence of probabilities $(N(A^{(n)})$ converges weakly to a deterministic measure μ_c, with a density with respect to Lebesgue measure given by

$$d\mu_c := \frac{1}{2\pi} 1_{[-2,2]} \sqrt{4 - x^2} dx$$

We are interested here in the asymptotic behavior of the largest eigenvalue $\lambda_1(A^{(n)})$, which is a so-called local problem. Classical results (see e.g. [1], also for precise references to the original articles) claim that

$$\lim_{n \to \infty} \lambda_1(A^{(n)}) = 2 \quad \mathbb{P}\text{-almost surely}$$

The asymptotic behaviour of $\lambda_1(A^{(n)})$ was further clarified by Tracy and Widom, who proved the following result : there exists a continuous decreasing function ψ_{TW} from $\mathbb{R}$ onto $(0, 1)$ such that

$$\lim_{n \to \infty} \mathbb{P}_n(\lambda_1(A^{(n)}) \geqslant 2 + x n^{-2/3}) = \psi_{\text{TW}}(x) \tag{1}$$

This function ψ_{TW} naturally arises as a determinant linked to the so-called "Airy kernel", which will be defined later. The most difficult point of Tracy and Widom's work was to show that this function ψ_{TW} can be written in terms of a Painlevé function (see [17]). From this point one can deduce asymptotic behavior of ψ_{TW} around $+\infty$ and find universal positive constants C, c, C', c' such that for x large enough

$$c' \exp(-C' x^{3/2}) \leqslant \psi_{\mathrm{TW}}(x) \leqslant c \exp(-C x^{3/2}) \tag{2}$$

The remainder of this article is organized as follows : in section 1, we define an appropriate Wasserstein distance and state our main theorem which asserts that Tracy-Widom convergence holds in this strong distance. In section 2, we derive from this theorem the small deviation inequality and compare it with the classical one. Section 3 introduces the needed framework of determinantal kernels, which are classical in this field, and section 4 contains the proof of the main theorem. Finally, section 5 contains an alternative simple deviation of upper bounds (2) for Tracy-Widom distribution.

1 Convergence in Terms of a Wasserstein Distance

We call tail function of a measure μ on $\mathbb{R}$ the function $\psi_\mu : \mathbb{R} \to [0,1]$ defined by $\psi_\mu(x) := \mu((x,+\infty))$. Such a function is decreasing, left-continuous, tends to 1 at $-\infty$ and to 0 at $+\infty$. The tail function just equals 1 minus the cumulative distribution function. The function appearing in the r.h.s. of (1) is the tail function of the Tracy-Widom distribution on $\mathbb{R}$ (we denote this distribution by TW).

We want to prove that the law of the rescaled largest eigenvalue tends to the Tracy-Widom law in a strong sense. As we only focus on the upper tail, we can consider truncated laws, supported on an interval $[a,+\infty)$ for some real a. Let Λ_n^a be the probability measure with tail function defined by

$$\psi_{\Lambda_n^a}(x) = \begin{cases} \mathbb{P}_n(\lambda_1(A^{(n)}) \geqslant 2 + x n^{-2/3}) & \text{if } x \geqslant a \\ 1 & \text{if } x < a \end{cases}$$

Similarly, let TW^a be the truncated Tracy-Widom law defined by

$$\psi_{\mathrm{TW}^a}(x) = \begin{cases} \psi_{\mathrm{TW}}(x) & \text{if } x \geqslant a \\ 1 & \text{if } x < a \end{cases}$$

We are going to show that for any a, Λ_n^a tends to TW^a with respect to the distance defined through a mass transportation problem in its Monge-Kantorovich formulation (see [12]).

A mass transportation problem is the question of optimizing the transshipment from a measure to another with respect to to a given cost. More precisely, let μ and ν be two probability measures on the same space X, and

$c : X \times X \to \mathbb{R}_+$ a symmetric function vanishing on the diagonal ($c(x,y)$ represent the price to pay to transfer a unit of mass from x to y). Ways to carry μ onto ν are represented through probability measures π on the square space $X \times X$ having μ and ν as marginals (this means that for any measurable subset A of X, $\pi(A \times X) = \mu(A)$ and $\pi(X \times A) = \nu(A)$). We denote by $\Pi(\mu,\nu)$ the space of such π.

The Wasserstein distance associated with the problem is the "minimum cost to pay", defined by

$$d(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{X^2} c(x,y) \mathrm{d}\pi(x,y)$$

We are going to consider a very special case of this problem. Let us suppose that $X = \mathbb{R}$ and that the cost c is defined as follows

$$c(x,y) := |\int_x^y w(t) \mathrm{d}t| \tag{3}$$

where w is a positive function.

We can now state the main result of this note

Theorem 1. *Let $w(x) := \exp(\gamma x^{3/2})$ and let d be the Wasserstein distance associated with the cost induced by w via the formula (3). Then, for any fixed $a \in \mathbb{R}$, if $\gamma > 0$ is small enough, Λ_n^a tends to TW^a for the distance d*

$$\lim_{n\to\infty} d(\Lambda_n^a, TW^a) = 0$$

2 The Small Deviation Inequality

The simplest idea to get concentration inequalities for the largest eigenvalue of a GUE random matrix is to use Gaussian concentration ; it is a straightforward consequence of the measure concentration phenomenon in the Gaussian space (see [1]) that

$$\forall t > 0, \forall n, \mathbb{P}_n(\lambda_1(A^{(n)}) \geqslant M_n + t) < \exp(-nt^2/2) \tag{4}$$

where M_n is the median of $\lambda_1(A^{(n)})$ with respect to the probability measure $\mathbb{P}_n$. One has the same upper estimate is the median M_n is replaced by the expected value $\mathbb{E}_n \lambda_1(A^{(n)})$.

The value of M_n can be controlled : for example we have $M_n \leqslant 2 + c/\sqrt{n}$. This will be for example a consequence of our Proposition. Plugging this into the equation (4), we get the following result, where C is a universal constant

$$\forall t > 0, \forall n, \mathbb{P}_n(\lambda_1(A^{(n)}) \geqslant 2 + t) < C \exp(-nt^2/2) \tag{5}$$

We ask the question whether in fact both $\mathbb{E}_n\lambda_1(A^{(n)})$ and M_n are smaller than 2. Note that since the funcion λ_1 is convex, its median with respect to $\mathbb{P}_n$ does not exceed its expected value ([7]). A positive answer to this question would imply that one could choose $C = 1$ in the inequality (5). The answer to the analogous question is known to be positive for the GOE (Gaussian Orthogonal Ensemble), an ensemble of real symmetric matrices defined in a similar way as GUE (see [10] for a precise definition). The argument, due to Gordon, uses a result about comparison of Gaussian processes known as Slepian's lemma (see [1]) and doesn't carry over to the complex setting.

There are similar though not as simple results for $\mathbb{P}_n(\lambda_1(A^{(n)}) \leqslant 2 - t)$, but in this paper we will concentrate on the "upper tail" estimates.

The result of Tracy and Widom (1) shows that the majoration (5) is not optimal for very small values of t. If for example t is equal to $xn^{-2/3}$ for a fixed x, then the right-hand side in concentration inequality (5) tends to 1 when n grows to ∞, whereas the left-hand side tends to $\psi_{\text{TW}}(x)$, which can be very small.

We would like to derive from our Theorem a deviation inequality which would improve the inequality (5) for small values of t. For this purpose, the uniform convergence in (1) (which, by Dini's theorem, follows formally form the pointwise convergence) is not enough. But we will prove in this section that our Theorem implies the following Proposition :

Proposition 1. *There exist positive universal constants C and c such that for every positive t and any integer n*

$$\mathbb{P}_n(\lambda_1(A^{(n)}) \geqslant 2 + t) \leqslant C \exp(-cnt^{3/2}) \tag{6}$$

Of course, by symmetry of the law $\mathbb{P}_n$, similar results are true for the smallest eigenvalue $\lambda_n(A^{(n)})$

$$\mathbb{P}_n(\lambda_n(A^{(n)}) \leqslant -2 - t) \leqslant C \exp(-cnt^{3/2}) \tag{7}$$

Using the fact that for a Hermitian matrix A, the norm equals the maximum absolute value of an eigenvalue, we get a similar estimate for $||A^{(n)}||$

$$\mathbb{P}_n(||A^{(n)}|| \geqslant 2 + t) \leqslant C \exp(-cnt^{3/2}) \tag{8}$$

We need the following lemma to prove the proposition, which will help us to explicitly compute Wasserstein distance

Lemma 1. *Suppose that the measures μ and ν are defined on $\mathbb{R}$, and that the cost c is defined by an integral, as in (3). If μ and ν are regular enough, for example if ψ_μ and ψ_ν are piecewise C^1, then the Wasserstein distance for the cost c equals*

$$d(\mu, \nu) = \int_{-\infty}^{\infty} w(t)|\psi_\mu(t) - \psi_\nu(t)|\mathrm{d}x \tag{9}$$

Proof. In fact, this transportation problem is explicitly solvable. For a one-dimensional problem with a cost satisfying the Monge condition (which is always the case when the cost is defined using an integral as in (3)), the optimal transshipment is achieved through the map T defined as follows (see [12], chapter 3.1)

$$\int_{-\infty}^{x} \mathrm{d}\mu = \int_{-\infty}^{T(x)} \mathrm{d}\nu$$

Thus, we can compute the value of $d(\mu,\nu)$

$$d(\mu,\nu) = \int_0^1 c(\psi_\mu^{-1}(u), \psi_\nu^{-1}(u))\mathrm{d}u$$

Let us consider first the particular case when $\psi_\mu \leqslant \psi_\nu$. This allows us to drop the absolute values in the definition of c (see (3)) and unfold the calculations. Using the appropriate changes of variables, we come to the equality (9).

For general μ and ν, define $\mu \wedge \nu$ and $\mu \vee \nu$ using their tail functions

$$\psi_{\mu\wedge\nu}(x) = \min(\psi_\mu(x), \psi_\nu(x)) \quad \text{and} \quad \psi_{\mu\vee\nu}(x) = \max(\psi_\mu(x), \psi_\nu(x))$$

We easily check that $\psi_{\mu\wedge\nu} \leqslant \psi_{\mu\vee\nu}$, $d(\psi_\mu, \psi_\nu) = d(\psi_{\mu\wedge\nu}, \psi_{\mu\vee\nu})$ and that the value of the r.h.s. of (9) does not change if we replace ψ_μ and ψ_ν by $\psi_{\mu\wedge\nu}$ and $\psi_{\mu\vee\nu}$. This yields the conclusion for the general case. □

Using this lemma, we get from our theorem (with $a = 0$), using the upper bound (2) for ψ_{TW}, the uniform estimate

$$\int_0^\infty w(x)\mathbb{P}_n(\lambda_1(A^{(n)}) \geqslant 2 + xn^{-2/3})\ \mathrm{d}x \leqslant C$$

which implies immediately for $x \geqslant 1$ (keep in mind that ψ_n is decreasing)

$$\psi_n(x) \leqslant C\exp(-\gamma(x-1)^{3/2}) \leqslant C' \exp(-\gamma' x^{3/2}) \tag{10}$$

This is, up to the rescaling $t = xn^{-2/3}$, the content of the proposition.

Now we can also easily show that our theorem implies the Tracy-Widom limit (1) : using the uniform bound (10) and Lebesgue's convergence theorem, we get from the Theorem that ψ_{TW} is the pointwise limit of the ψ_n's on $[a, +\infty)$, and thus on the whole real line if we let a go to $-\infty$.

It should be emphasized that recently (independently from and slightly preceding this work), this small deviations result has been proved by Ledoux in [8] using an argument based on the Harer-Zagier recurrence formula (see [5] for a simple proof of this formula). The same paper by Ledoux contains another proof based on hypercontractivity which gives the result up to a polynomial factor ; this method works also for the Laguerre Unitary Ensemble (see [8] for the definition). However, the existence of a Tracy-Widom limit does not follow from this approach. More generally, many contributions to this and related topics either address the limit behaviour or provide dimension-free bounds, rarely combining the two. Our technique captures both phenomena in a single "stroke".

3 Relation to Determinants

The remainder of this note is devoted to the proof of the main theorem. For simplicity, we will prove only the case $a = 0$, and drop all the superscripts. The proof for a general a requires only routine modifications.

We are first going to express all involved quantities in terms of determinants of certain operators. This is quite classical work due to Gaudin and Mehta (see [10]). Part of the calculations done here are present, at least implicitly, in the paper by Tracy and Widom ([17]).

We need new notation. Let (H_n) be the Hermite polynomials, which are defined by

$$H_n(t) := (-1)^n \exp(t^2)(\frac{d}{dt})^n \exp(-t^2)$$

They are orthogonal for the measure on $\mathbb{R}$ of density $\exp(-x^2)$ with respect to Lebesgue measure. Then we note

$$\phi_n(t) := \frac{1}{\sqrt{d_n}} H_n(t) \exp(-t^2/2) \tag{11}$$

where $d_n := \int_{\mathbb{R}} H_n(x)^2 \mathrm{d}x = 2^n n! \sqrt{\pi}$. The family (ϕ_n) is therefore orthonormal in $L^2(\mathbb{R})$. We introduce

$$k_n(x, y) := \sum_{j=0}^{n-1} \phi_j(x)\phi_j(y)$$

We can associate to k_n an integral operator K_n acting on the Hilbert space $L^2(\mathbb{R})$ in the following way

$$(K_n f)(x) := \int_{\mathbb{R}} k_n(x, y) f(y) \mathrm{d}y \tag{12}$$

This operator K_n is nothing but the orthogonal projection in $L^2(\mathbb{R})$ onto the subspace spanned by $(\phi_j)_{1 \leqslant j \leqslant n}$.

This is a very general setting : if we have a measure space (X, μ) and a "kernel" $k \in L^2(X \times X)$, we can define an operator K on $L^2(X)$ using a formula similar to (12). From now on, all kernels are assumed to belong to

$L^2(X \times X)$ and are denoted by small letters ; associated integral operators are denoted by the corresponding capital letter.

It is straightforward to prove that Hilbert-Schmidt operators on $L^2(X)$ are exactly integral operators with a L^2 kernel. Moreover, the Hilbert-Schmidt norm of the operator and the L^2 norm of the kernel coincide. This fact is proved in [3], which is a good reference for a reader who wants more detail on integral operators. Let us just quote the formula for compositions of operators : if k and l are two kernels on the same space (X, μ), then the operator KL is an integral operator with kernel :

$$(kl)(x,y) = \int_X k(x,z)l(z,y)\mathrm{d}\mu(z) \tag{13}$$

The tail function of $\lambda_1(A^{(n)})$ can now be expressed using the kernel k_n. The key formula is the following (see [10])

$$\forall t \in \mathbb{R}, \mathbb{P}_n(\frac{\sqrt{n}}{\sqrt{2}}\lambda_1(A^{(n)}) \leqslant t) = \det_{[t,\infty)}(\mathrm{Id} - K_n) \tag{14}$$

In the formula (14), the right-hand side must be understood as the determinant of the operator K_n acting on the space $L^2([t,\infty))$ (or equivalently of the operator with kernel equal to is the restriction of k_n to $[t,\infty)^2$). This restricted operator is denoted $K_n^{[t]}$.

It may not be immediately obvious how to define such a determinant, as the operator involved acts on an infinite-dimensional space. However, the operator K_n that we consider here has a finite rank, hence we can define its determinant as if it were acting on a finite-dimensional space.

A problem will arise when we want to consider limits of such operators, which might fail to have a finite rank. Fortunately, a whole theory of determinants (and traces) of integral operators exists (so-called "Fredholm" determinants). In fact, there are several possible ways to extend these concepts to the infinite-dimensional case. We will focus on a more algebraic approach, due to Grothendieck (see [4] or [13] for a complete exposition), which defines determinants of a nuclear (= trace class) perturbation of identity in terms of traces of its exterior powers (here N is a nuclear operator, for which trace is well-defined).

$$\det(\mathrm{Id} + N) := 1 + \sum_{k=1}^{\infty} \mathrm{tr}(\Lambda^k(N))$$

Of course, this definition coincides with the usual one in the finite-dimensional case.

The presence of the factor $\sqrt{n}/\sqrt{2}$ in equation (14) requires an explanation. It arose because there are several normalizations possible. We chose to define the GUE so that the first eigenvalue is about 2, while other authors, as Tracy and Widom in [17], prefer to locate it around $\sqrt{2n}$ (there are still other normalizations but an exhaustive list would be too long). As we kept their notation for the kernels k_n, a scaling factor will appear when we pass from a normalization to the other one.

To get a nontrivial limit, we must replace the t in formula (14) by the following rescaling, as for the Tracy-Widom limit (1)

$$t = \tau_n(x) := \frac{\sqrt{n}}{\sqrt{2}}(2 + \frac{x}{n^{2/3}})$$

Let also $\tilde{k}_n$ be the rescaled kernel

$$\tilde{k}_n(x,y) := \frac{1}{\sqrt{2}n^{1/6}} k_n(\tau_n(x), \tau_n(y))$$

We can see using a change of variable that $\tilde{K}_n^{[x]}$ and $K_n^{[\tau_n(x)]}$ have the same eigenvalues. More precisely, if f is a eigenfunction of $\tilde{K}_n^{[x]}$, then $f \circ \tau_n^{-1}$ is a eigenfunction of $K_n^{[\tau_n(x)]}$, with the same eigenvalue.

Plugging these renormalizations into the formula (14), we obtain

$$\mathbb{P}_n(\lambda_1(A^{(n)}) \leqslant 2 + xn^{-2/3}) = \det_{[x,+\infty)}(\mathrm{Id} - \tilde{K}_n) \tag{15}$$

Using the previous definition for the tail function ψ_n, we can write for a positive s

$$\psi_n(s) = 1 - \det(\mathrm{Id} - \tilde{K}_n^{[s]})$$

The following result was known before Tracy and Widom's work (see for example [2])

$$\lim_{n\to\infty} \tilde{k}_n(x,y) = k(x,y) \tag{16}$$

uniformly on compact subsets in x and y.

Here k is the kernel, often called Airy kernel, defined by

$$k(x,y) := \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y} \tag{17}$$

The kernel k is extended by continuity to the diagonal. The function Ai is called the Airy function. It is very useful in physics and can be defined by several means. One of them is the following integral representation

$$\mathrm{Ai}(z) := \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(\mathrm{i}(zt + t^3/3))\mathrm{d}t$$

It can also be written as a combination of Bessel functions. It satisfies the Airy ODE

$$\frac{\partial^2}{\partial x^2} y(x) = xy(x) \tag{18}$$

The asymptotic behavior of Ai is well-known, for example [16] contains the following formula, valid when x tends to $+\infty$

$$\mathrm{Ai}(x) \sim \frac{1}{2} 3^{-1/4} \sqrt{\pi} x^{-1/4} \exp(-\frac{2}{3^{3/2}} x^{3/2}) \tag{19}$$

The function ψ_{TW} can be defined using this Airy kernel

$$\psi_{\mathrm{TW}}(x) := 1 - \det(\mathrm{Id} - K^{[x]}) \tag{20}$$

In [17] Tracy and Widom found another expression for ψ_{TW}. Let q be the solution of the Painlevé II ODE

$$\frac{\partial^2}{\partial x^2} q(x) = xq(x) + 2q(x)^3$$

which is determined by the asymptotics $q(x) \sim \mathrm{Ai}(x)$ for x close to $+\infty$. Then we have the representation

$$\psi_{\mathrm{TW}}(x) = 1 - \exp(-\int_x^\infty (t-x) q(t)^2 \mathrm{d}t) \tag{21}$$

It is easy to get from (19) and (21) the bounds (2) for the asymptotic behavior of ψ_{TW}. However, as we do not really need all the depth of Tracy and Widom's results and connections to Painlevé functions, we will reprove this fact in a more elementary way at the end of this note.

4 Convergence of the Operators

The convergence in (16) as determined in the existing literature is rather weak ; in particular, it does not imply convergence of the associated integral operators in the Hilbert-Schmidt norm or even in the operator norm on L^2. In particular, we are not *a priori* allowed to exchange limit and determinant in (15) when n tends to infinity.

Our main step will be to show that $\tilde{K}_n$ tends to K with respect to the nuclear (trace class) norm. To that end we need several lemmas.

Lemma 2. *The following equality holds*

$$\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right) k_n(x,y) = -\sqrt{\frac{n}{2}} \left(\phi_n(x)\phi_{n-1}(y) + \phi_{n-1}(x)\phi_n(y)\right)$$

Proof. We start with the Christoffel-Darboux formula (see [16])

$$k_n(x,y) = \sqrt{\frac{n}{2}}\frac{\phi_n(x)\phi_{n-1}(y) - \phi_{n-1}(x)\phi_n(y)}{x-y}$$

Then we apply the operator $\frac{\partial}{\partial x} + \frac{\partial}{\partial y}$ to each term. We use the formula (11) and the following identities (those which are not obvious are shown in [16])

$$\phi_n'(x) = -\frac{\exp(-x^2/2)}{\sqrt{d_n}}(H_n'(x) - xH_n(x))$$

$$H_{n-1}'(x) = 2xH_{n-1}(x) - H_n(x)$$

$$H_n'(x) = 2nH_{n-1}(x)$$

We obtain exactly the expected result. □

Lemma 3. *The following integral representation holds*

$$\tilde{k}_n(x,y) = \frac{n^{1/6}}{2\sqrt{2}}\int_0^\infty \phi_n(\tau_n(x+z))\phi_{n-1}(\tau_n(y+z)) + \phi_{n-1}(\tau_n(x+z))\phi_n(\tau_n(y+z))\mathrm{d}z \tag{22}$$

Proof. If we apply the operator $\frac{\partial}{\partial x} + \frac{\partial}{\partial y}$ to the right-hand side of (22) (there is no trouble with interchanging the operations "$\frac{\partial}{\partial x} + \frac{\partial}{\partial y}$" and "$\int_0^\infty$" since all the functions involved are Schwartz functions), we get after standard calculations

$$\frac{n^{1/6}}{2\sqrt{2}}(\phi_n(\tau_n(x))\phi_{n-1}(\tau_n(y)) + \phi_{n-1}(\tau_n(x))\phi_n(\tau_n(y)))$$

Lemma 2 asserts that we obtain exactly the same expression when we apply the operator $\frac{\partial}{\partial x} + \frac{\partial}{\partial y}$ to the left member of (22). Thus, the two members of the equation are equal modulo a function (say, α) which only depends on $x-y$. But both members tend to zero when x et y tend to infinity in an independent way. Therefore the function α has to vanish identically and the lemma is proved. □

Let us introduce extra notation. The following kernels are defined on $[s,+\infty)^2$, where s is any positive number

$$a_n^{[s]}(x,y) := \frac{n^{1/12}}{2^{1/4}}\phi_n(\tau_n(x+y-s))$$

$$b_n^{[s]}(x,y) := \frac{n^{1/12}}{2^{1/4}} \phi_{n-1}(\tau_n(x+y-s))$$

$$a^{[s]}(x,y) := \mathrm{Ai}(x+y-s)$$

The equality (22) can be translated in terms of operators (this is just a consequence of the formula (13) for the composition of kernels)

$$\tilde{K}_n^{[s]} = \frac{1}{2}(A_n^{[s]} B_n^{[s]} + B_n^{[s]} A_n^{[s]}) \tag{23}$$

A similar equality for the operator K is proved (exactly in the same way) in [17]

$$K^{[s]} = (A^{[s]})^2 \tag{24}$$

We shall subsequently show that (for a fixed s) the operators $A_n^{[s]}$ and $B_n^{[s]}$ tend to $A^{[s]}$ with respect to the Hilbert-Schmidt norm. To that end, we need estimates for ϕ_n contained in two lemmas that follow.

Lemma 4. *The functions ϕ_n, after rescaling, converge to* Ai*, uniformly on compact subsets in y :*

$$\phi_n(\tau_n(y)) 2^{-1/4} n^{1/12} \to \mathrm{Ai}(y) \quad \text{and} \quad \phi_{n-1}(\tau_n(y)) 2^{-1/4} n^{1/12} \to \mathrm{Ai}(y) \tag{25}$$

Proof. This is an immediate consequence of the following asymptotic formulae for Hermite polynomials due to Plancherel and Rotach. They can be found, in a slightly different presentation, in the book by Szegö ([16])

$$\text{If } x = \sqrt{2n+1} + \frac{y}{\sqrt{2} n^{1/6}}, \text{ then } \phi_n(x) = 2^{1/4} n^{-1/12} (\mathrm{Ai}(y) + O(n^{-3/4}))$$

The O holds when n tends to $+\infty$, uniformly in y on compact subsets. □

Lemma 5. *We have a bound for ϕ_n which is uniform in n : there exists a positive constant c such that for any $y > 0$ and any integer n*

$$\begin{cases} n^{1/12} \phi_n(\tau_n(y)) \leqslant C \exp(-c y^{3/2}) \\ n^{1/12} \phi_n(\tau_{n-1}(y)) \leqslant C \exp(-c y^{3/2}) \end{cases} \tag{26}$$

Proof. Let us sketch a proof of the fisrt inequality in (26). We will use the following result, which is an exercise at page 403 in [11]. It is valid for $x \geqslant 1$

$$H_n(\nu x) \leqslant 1.13\sqrt{2\pi}\exp(-\nu^2/4)\nu^{(3\nu^2-1)/6}\exp(\nu^2x^2/2)\left(\frac{\zeta}{x^2-1}\right)^{1/4}\mathrm{Ai}(\nu^{4/3}\zeta)$$

where $\nu := \sqrt{2n+1}$ and

$$\zeta := \left(\frac{3}{4}x\sqrt{x^2-1} - \frac{3}{4}\mathrm{Argch}x\right)^{2/3}$$

Using the definition of ϕ_n given in formula (11) and Stirling's formula to estimate d_n, we obtain

$$n^{1/12}\phi_n(\nu x) \leqslant C\left(\frac{\zeta}{x^2-1}\right)^{1/4}\mathrm{Ai}(\nu^{4/3}\zeta)$$

We deduce from (19) a bound for Ai, and we also use the inequality $\zeta \geqslant c(x-1)$ to get

$$n^{1/12}\phi_n(\sqrt{2n+1}x) \leqslant Cn^{-1/6}\frac{1}{(x-1)^{1/4}}\exp(-c(2n+1)(x-1)^{3/2}) \quad (27)$$

We now return to our notation through the change of variable $\sqrt{2n+1}\,x = \tau_n(y)$. We can estimate x in the following way

$$x \geqslant 1 - \frac{c}{n} + \frac{y}{\sqrt{2}n^{1/6}\sqrt{2n+1}}$$

For y large enough, we even have

$$x \geqslant 1 + c\frac{y}{n^{2/3}} \quad (28)$$

Combining (27) and (28) yields

$$n^{1/12}\phi_n(\tau_n(y)) \leqslant C\frac{1}{y^{1/4}}\exp(-cy^{3/2})$$

The factor $y^{-1/4}$ can be deleted if c is made small enough. This inequality is only true for y large enough, but we keep in mind that convergence in (25) was uniform on compact subsets, so we can extend it to all positive y, and the inequality is proved. The same scheme of demonstration works for the second inequality, with τ_{n-1} instead of τ_n. □

We are now ready to prove our main theorem

Proof. We denote by $||.||_{HS}$ the Hilbert-Schmidt norm and by ν the nuclear norm.

We are first going to estimate the quantity $|\psi_n(s) - \psi_{\mathrm{TW}}(s)| = |\det(\mathrm{Id} - \tilde{K}_n^{[s]}) - \det(\mathrm{Id} - K^{[s]})|$. To reach this goal, we will use the following majoration (see [4]), valid for any two nuclear operators A and B

$$|\det(\mathrm{Id} + A) - \det(\mathrm{Id} + B)| \leqslant \nu(A - B)\mathrm{e}^{1+\nu(A)+\nu(B)} \tag{29}$$

It will be useful to notice that lemma 5 implies in particular the following remark : there is a positive C such that for any $s \geqslant 0$ and any integer n, all the quantities $||A_n^{[s]}||_{HS}$, $||B_n^{[s]}||_{HS}$ and $||A^{[s]}||_{HS}$ are bounded by C (remember that the Hilbert-Schmidt norm is just the L^2-norm of the kernel). Using inequalities (23), (24) and the noncommutative Hölder inequality, we get that $\nu(K^{[s]})$ and $\nu(\tilde{K}_n^{[s]})$ are also bounded by the constant. Hence we can drop the exponential factor in formula (29)

$$|\psi_n(s) - \psi_{\mathrm{TW}}(s)| \leqslant C\nu(\tilde{K}_n^{[s]} - K^{[s]})$$

We need to estimate the quantity $\nu(\tilde{K}_n^{[s]} - K^{[s]})$. The key to do this is to use the equalities (23) et (24) to get

$$\begin{aligned}\tilde{K}_n^{[s]} - K^{[s]} = \frac{1}{4}\Big(&(A_n^{[s]} - A^{[s]})(B_n^{[s]} + A^{[s]}) + (A_n^{[s]} + A^{[s]})(B_n^{[s]} - A^{[s]}) \\ &+ (B_n^{[s]} + A^{[s]})(A_n^{[s]} - A^{[s]}) + (B_n^{[s]} - A^{[s]})(A_n^{[s]} + A^{[s]})\Big)\end{aligned}$$

The non-commutative Hölder inequality yields

$$\begin{aligned}\nu(\tilde{K}_n^{[s]} - K^{[s]}) \leqslant\ & \frac{1}{2}||A_n^{[s]} - A^{[s]}||_{HS}||B_n^{[s]} + A^{[s]}||_{HS} \\ &+ \frac{1}{2}||A_n^{[s]} + A^{[s]}||_{HS}||B_n^{[s]} - A^{[s]}||_{HS}\end{aligned} \tag{30}$$

The factors with a "+" are easy to get rid of : we can use the triangle inequality to write $||A_n^{[s]} + A^{[s]}||_{HS} \leqslant ||A_n^{[s]}||_{HS} + ||A^{[s]}||_{HS}$, which is uniformly bounded according to the remark following formula (29). We obtain

$$\nu(\tilde{K}_n^{[s]} - K^{[s]}) \leqslant C(||A_n^{[s]} - A^{[s]}||_{HS} + ||B_n^{[s]} - A^{[s]}||_{HS})$$

We can now calculate the Wasserstein distance from Λ_n to TW, using the expression given by lemma 1

$$\begin{aligned}d(\Lambda_n, TW) &= \int_0^\infty \exp(\gamma s^{3/2})|\psi_n(s) - \psi_{\mathrm{TW}}(s)|\mathrm{d}s \\ &\leqslant C\int_0^\infty \exp(\gamma s^{3/2})(||A_n^{[s]} - A^{[s]}||_{HS} + ||B_n^{[s]} - A^{[s]}||_{HS})\mathrm{d}s\end{aligned}$$

First deal with the term $||A_n^{[s]} - A^{[s]}||_{HS}$. Using the definition of $A_n^{[s]}$ and $A^{[s]}$ we get

$$\int_0^\infty \exp(\gamma s^{3/2})||A_n^{[s]} - A^{[s]}||_{HS}\mathrm{d}s \tag{31}$$
$$= \sqrt{2}\int_0^\infty \exp(\gamma s^{3/2})\left(\int_0^\infty z\left((\frac{n^{1/12}}{2^{1/4}}\phi_n \circ \tau_n - \mathrm{Ai})(z+s)\right)^2 dz\right)^{1/2} \mathrm{d}s$$

Fix an $\varepsilon > 0$ and use the uniform bound of lemma 5 : we get that for γ small enough, S large enough and any n

$$\sqrt{2}\int_S^\infty \exp(\gamma s^{3/2})\left(\int_0^\infty z\left((\frac{n^{1/12}}{2^{1/4}}\phi_n \circ \tau_n - \mathrm{Ai})(z+s)\right)^2 \mathrm{d}z\right)^{1/2} \mathrm{d}s \leqslant \varepsilon$$

Similarly, for Z large enough, any s smaller than S and any n

$$\left(\int_Z^\infty z\left((\frac{n^{1/12}}{2^{1/4}}\phi_n \circ \tau_n - \mathrm{Ai})(z+s)\right)^2 \mathrm{d}z\right)^{1/2} \leqslant \frac{\varepsilon}{\sqrt{2}S\exp(S^{3/2})}$$

Now we can split the integral in (31) into three terms to get (remember than convergence in lemma 4 was uniform on compact subsets)

$$\int_0^\infty \exp(\gamma s^{3/2})||A_n^{[s]} - A^{[s]}||_{HS}\mathrm{d}s \leqslant 3\varepsilon \qquad \text{for } n \text{ large enough}$$

We can write a similar estimate with B_n instead of A_n. We finally deduce that, for n large enough, $d(\Lambda_n, TW) \leqslant 6C\varepsilon$. Hence Λ_n tends to TW in the Wasserstein sense. This is the announced result. □

5 An Elementary Proof of Asymptotics for ψ_{TW}

To prove our theorem, we needed the upper asymptotics (2) for ψ_{TW}. It is possible to derive them from the representation (21) : keeping in mind that $q \sim \mathrm{Ai}$, we get from (19)

$$\int_x^\infty (t-x)q^2(t)\mathrm{d}t \leqslant C\exp(-cx^{3/2})$$

Hence, (21) yields

$$\psi_{\mathrm{TW}}(x) \leqslant 1 - \exp(-C\exp(-cx^{3/2})) \leqslant C'\exp(-cx^{3/2})$$

However, for sake of completeness, we are going to derive in this section this last result in a more elementary way, i.e. without using the Painlevé representation. To do this, we need some facts about integral operators. Of course, a general integral operator can fail to be nuclear (for example, any Hilbert-Schmidt operator from $L^2(X)$ into itself can be written as an integral operator). Nevertheless, there exist several "nuclearity tests", criteria ensuring that under some conditions, kernels generate nuclear operators ([3],[4]). The main result in this topic is Mercer's theorem, which enables us to expand a continuous self-adjoint kernel (i.e. the associated operator is self-adjoint) as a series of eigenfuctions of the operator. Unfortunately, these results are usually stated when dealing with a compact space of finite measure, and we have to consider half-infinite intervals $[s, +\infty)$. However, the standard proofs work also in this setting with only slight modifications.

A result which fits the present context is the following

Lemma 6. *Let $X = [s, +\infty)$, equipped with the Lebesgue measure, and k be a kernel on $X \times X$ which satisfies the following conditions*

1. *$k \in L^2(X \times X)$*
2. *k is jointly continuous*
3. *K is positive self-adjoint as an operator on $L^2(X)$*
4. *There exists a continous positive function ϱ in $L^2(X)$ such that $|k(x,y)| \leqslant \varrho(x)\varrho(y)$ for every x, y in X.*

Then the operator K is nuclear and the trace formula holds

$$\mathrm{tr}(K) = \int_s^\infty k(x,x)\mathrm{d}x \tag{32}$$

Proof. We are going to derive our result from the classical finite-measure case using a change of density trick. Let μ be the mesaure on X with density ϱ^2 with respect to Lebesgue measure ; we have $\mu(X) < \infty$. If we (isometrically) identify $L^2(X, dx)$ with $L^2(X, \mu)$ sending f to f/ϱ, the integral operator K viewed from $L^2(X, \mu)$ into itself has kernel $k(x,y)/\varrho(x)\varrho(y)$. To get the result we simply apply to the new kernel the following version of Mercer's theorem (it can proved adapting straightforward the classical proof from [14]) : if μ is a finite Borel measure on X and k a continuous bounded positive self-adjoint kernel, then the associated operator K is nuclear and its trace is equal to the integral of the kernel along the diagonal. □

Lemma 7. *The following estimation holds*

$$\exists C, c > 0, \forall s > 0, \quad \psi_{\mathrm{TW}}(s) \leqslant C \exp(-cs^{3/2})$$

Proof. By definition (see [13]), we have

$$\psi_{\mathrm{TW}}(s) = \sum_{k=1}^{\infty} (-1)^{k-1} \mathrm{tr}(\Lambda^k(K^{[s]}))$$

Using the fact that $\mathrm{tr}(\Lambda^k(K^{[s]})) \leqslant \mathrm{tr}(K^{[s]})^k/k!$, we get

$$|\psi_{\mathrm{TW}}(s)| \leqslant \exp(\mathrm{tr}K^{[s]}) - 1$$

Actually, formula (20) shows that ψ_{TW} is positive since we have $0 \leqslant K^{[s]} \leqslant 1$. In the end, the convexity of the exponential function on $[0, \mathrm{tr}K^{[0]}]$ yields for $s \geqslant 0$

$$\psi_{TW}(s) \leqslant C \mathrm{tr} K^{[s]}$$

It is not hard to check that the kernel $k^{[s]}$ satisfies the hypotheses of lemma 6 ; to check condition 4 we can cook up a function ϱ using Ai and its derivative.

Thus we can rewrite the trace of $K^{[s]}$ as an integral

$$\psi_{TW}(s) \leqslant C \int_s^{\infty} \big((\mathrm{Ai}'(x))^2 - x\mathrm{Ai}(x)\big)\, \mathrm{d}x \tag{33}$$

The value of K on the diagonal comes from (17) and the Airy ODE (18).

Using (19), we can write

$$\exists C, c > 0 \quad \forall s \geqslant 0 \quad \mathrm{Ai}(s) \leqslant C \exp(-cs^{3/2}) \tag{34}$$

A similar majoration holds for Ai$'$: we only need to write $\mathrm{Ai}'(s) = \int_s^{\infty} \mathrm{Ai}''(x)\mathrm{d}x$ and to use formulae (18) and (34)

$$\exists C, c > 0 \quad \forall s \geqslant 0 \quad \mathrm{Ai}'(s) \leqslant C \exp(-cs^{3/2}) \tag{35}$$

The conclusion comes when combining formulae (33), (34) and (35). □

Possible Generalizations

Of course, we expect the inequalities analogous to (6) to be true in a much more general setting. Basically, each time a Tracy-Widom-like behavior has been proved or is suspected, we can ask whether such an uniform estimate holds.

The most natural extension would be the setting of general Wigner matrices, for which universality of Tracy-Widom limit has been proved by Soshnikov ([15]). However, the bounds on moments he obtained are not enough to derive the small deviations inequality.

Tracy and Widom proved results similar to (1), involving a different limit law, for the matrix ensembles GOE and GSE (the real orthogonal and the symplectic cases) in [18].

Several authors investigated the behavior of the largest s-number (also called singular value) of a rectangular $m \times n$ matrix with independant entries, when the ratio m/n tends to a limit in $(0, 1)$. The paper [6] contains an result anaologous to (1) for the Gaussian case (the so-called Wishart ensemble). There is strong numerical evidence indicating that a convergence on the scale $n^{-2/3}$ as in Tracy-Widom behavior occurs also universally in this case, for the largest s-number, but also for the smallest one.

Another quantity of interest is the norm of a $n \times m$ random matrix when considered as a operator from ℓ_p^n to ℓ_q^m. Concentration results have been recently obtain in this case by Meckes (cf [9]).

In all these cases, we know concentration inequalities similar to (5), it would be interesting to prove the corresponding small deviation result.

Acknowledgements

I would like to express my sincere thanks and appreciation to my supervisor, Prof. Stanislaw J. Szarek, for our inspiring discussion and for his ongoing support. This paper will be part of my Ph.D. thesis prepared at the Équipe d'Analyse Fonctionnelle of the University Paris VI. I would also like to thank Charles-Antoine Louët for his assiduous proofreading.

References

1. Davidson, K.R. and Szarek, S.J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach Spaces, vol. 1*, pages 317–366.
2. Forrester, P.J. (1993). The spectrum edge of random matrix ensembles *Nucl. Phys.* **B402**, 709–728.
3. Gohberg, I. and Goldberg, S. (1981). *Basic operator theory.* Birkhaüser.
4. Gohberg, I., Goldberg, S. and Krupnik, N. (1998). *Traces and Determinants of Linear Operators.* Birkhaüser.
5. Haagerup, U, and Thorbjørnsen, S (1999). Random matrices with complex Gaussian entries. Preprint available at URL `http://imada.sdu.dk/~haagerup/2000-.html`
6. Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.* **29**, 295–327.
7. Kwapień, S. (1993). A remark on the median and the expecation of convex functions of Gaussian vectors. *Probabilities in Banach spaces 9 (Sandjberg, 1993), Progr. Probab.* **35**, 271–272.
8. Ledoux, M. (2003). A remark on hypercontractivity and tail inequalities for the largest eigenvalues of random matrices. *Séminaire de probabilités XXXVII.* Lecture Notes in Mathematics, Springer.

9. Meckes, M.W. (2003). Concentration of norms and eigenvalues of random matrices. To appear in *J. Funct. Anal.*
10. Mehta, M.L. (1991). *Random Matrices.* Academic Press, Boston, MA, second edition.
11. Olver, F.W.J. (1974). *Asymptotics and Special Functions.* Cambridge University Press.
12. Rachev, S.T. and Rüschendorf, L. (1998). *Mass Transportation Problems.* Vol. 1, Springer-Verlag.
13. Simon, B. (1977). Infinite determinants. *Advances in Math.* **24**, 244–273.
14. Smithies, F. (1959). *Integral equations.* Cambridge University Press.
15. Soshnikov, A. (1999). Universality at the edge of spectrum in Wigner random matrices. *Comm. Math. Phys.* **207**, 697–733.
16. Szegö, G. (1967). *Orthogonal Polynomials.* American Mathematical Society, third edition.
17. Tracy, C.A. and Widom, H. (1994). Level-spacing distributions and the Airy kernel. *Comm. Math. Phys.* **159**, 151–174.
18. Tracy, C.A. and Widom, H. (1996). On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177**, 727–754.