ANNALES
DE L'INSTITUT
HENRI
POINCARÉ

PROBABILITÉS
ET STATISTIQUES

On the large deviations of traces of random matrices

Fanny Augeri[a]

[a]*Institut de Mathématiques de Toulouse, France. E-mail: faugeri@math.univ-toulouse.fr*

Abstract. We present large deviations principles for the moments of the empirical spectral measure of Wigner matrices and empirical measure of β-ensembles in three cases: the case of β-ensembles associated with a convex potential with polynomial growth, the case of Gaussian Wigner matrices, and the case of Wigner matrices without Gaussian tails, that is Wigner matrices whose entries have tail distributions decreasing as e^{-ct^α}, for some constant $c > 0$ and with $\alpha \in (0, 2)$.

Résumé. Nous proposons des principes de grandes déviations pour les moments de la mesure spectrale empirique de matrices de Wigner et de la mesure empirique de β-ensembles dans trois cas : celui des β-ensembles associés à un potentiel convexe à croissance polynomiale, le cas des matrices de Wigner Gaussiennes, et le cas des matrices de Wigner sans queues Gaussiennes, c'est-à-dire dont les entrées ont une queue de distribution ayant le même comportement que e^{-ct^α}, pour une certaine constante $c > 0$ et $\alpha \in (0, 2)$.

1. Introduction and main results

The study of the traces of random matrices is now a classical tool to understand the behavior of their spectrum. From the original proof of Wigner's theorem by the moments method (see [30]), to the universality results at the edge of Hermitian or covariance random matrices (see for example [15,29]), 'Wigner traces method' has proven extremely effective in the macroscopic, as well as the microscopic study of the spectrum of random matrices.

Starting from Wigner's theorem, which asserts that for a Wigner matrix whose entries are centered and have finite moments, the moments of the empirical spectral measure, or equivalently the normalized traces, converge almost surely to 0, for odd moments, and to the Catalan numbers, for even moments, one can ask about the deviations of these moments around their respective limit value.

The fluctuations of the traces of random matrices have been extensively studied, usually as a first step to get the fluctuations of the linear statistics of the eigenvalues. Originally proven in the context of Wishart matrices in [19], a central limit theorem for the moments of the empirical spectral measure of standard Wigner matrices can be found in [1, Theorem 2.1.31], following Jonsson's strategy of using the moments method and combinatorial techniques. Due to the repulsion of the eigenvalues, one has to multiply by a factor N – instead of $\sqrt{N}$ in the case of independent variables – to see the fluctuations of the centered moments. The development of the combinatorial approach culminated in [27,28], in which the authors show a CLT for the pth moment with p growing with N, $p \ll N^{2/3}$, as well as multivariate version of the CLT for moments, in the case of standard Wigner matrices with symmetric and sub-Gaussian entries.

Regarding the deviations of the moments of the empirical spectral measure, we know from [23, Section 3.1], that the p-Schatten norm of Gaussian Hermitian or symmetric matrices is sub-Gaussian. Still in the Gaussian case, the

estimates of moments of Gaussian chaos of [20] can also provide some concentration inequalities for the moments of the empirical spectral measure. Concentration inequalities for truncated traces of convex perturbation of the GUE multi-matrix model can be found in [18]. More generally, we know by [25], that if the entries of X are bounded or satisfies some logarithmic Sobolev inequalities, then the normalized traces of powers of $X/\sqrt{N}$, say $N^{-1}\operatorname{tr}(X/\sqrt{N})^p$, satisfies a concentration inequality with speed $N^{1+2/p}$. This gives an indication, at least in this case, of the speed of the large deviations of the moments of the empirical spectral measure around the Catalan numbers.

Note that since the map which associates to a probability measure on $\mathbb{R}$, its pth moment is not continuous for the weak topology, one cannot derive, by a contraction principle, large deviations principles for the pth moment of the empirical spectral measure, from the already known large deviations principles for the empirical spectral measure, like in the case of the GUE or GOE due to [4], or in the case of Wigner matrices without Gaussian tails due to [7].

Moderate deviations of certain traces of convex perturbation of the GUE multi-matrix model have been investigated in [14]. In the case where the entries are not centered, some results of large deviations for the moments of the empirical spectral measure are known. In the case of symmetric Bernoulli matrices, we know by [13, Theorem 1.5] that the centered traces satisfy moderate deviations principles with an explicit rate function. A large deviations principle for the traces of Bernoulli matrices is derived in [9, Theorem 4.1], as a consequence of the large deviations principle of Erdös–Renyi graphs with parameter p independent of N, with respect to the cut metric.

1.1. *Main results*

The aim of this paper is to derive large deviations principles for the moments of the empirical (spectral) measure in three cases: the case of β-ensembles for convex potential with polynomial growth in Section 3, the case of Gaussian Wigner matrices in Section 2, and the case of Wigner matrices without Gaussian tails in Section 4.

We recall that a sequence of random variables $(Z_n)_{n\in\mathbb{N}}$ taking value in some topological space $\mathcal{X}$ equipped with the Borel σ-field $\mathcal{B}$, follows a large deviations principle (LDP) with speed $\upsilon:\mathbb{N}\to\mathbb{N}$, and rate function $J:\mathcal{X}\to[0,+\infty]$, if J is lower semicontinuous and υ increases to infinity and for all $B\in\mathcal{B}$,

$$-\inf_{B^\circ} J \leq \liminf_{n\to+\infty}\frac{1}{\upsilon(n)}\log\mathbb{P}(Z_n\in B)$$

$$\leq \limsup_{n\to+\infty}\frac{1}{\upsilon(n)}\log\mathbb{P}(Z_n\in B)$$

$$\leq -\inf_{\overline{B}} J,$$

where B° denotes the interior of B and $\overline{B}$ the closure of B. We recall that J is lower semicontinuous if its t-level sets $\{x\in\mathcal{X}:J(x)\leq t\}$ are closed, for any $t\in[0,+\infty)$. Furthermore, if all the level sets are compact, then we say that J is a good rate function.

We define the β-ensemble associated with the potential V as the following probability measure on $\mathbb{R}^N$,

$$d\mathbb{P}_{V,\beta}^N = \frac{1}{Z_V^N}\prod_{i<j}|\lambda_i-\lambda_j|^\beta e^{-N\sum_{i=1}^N V(\lambda_i)}\prod_{i=1}^N d\lambda_i, \tag{1}$$

where $Z_{V,\beta}^N$ is the partition function, that is,

$$Z_{V,\beta}^N = \int \prod_{i<j}|\lambda_i-\lambda_j|^\beta e^{-N\sum_{i=1}^N V(\lambda_i)}\prod_{i=1}^N d\lambda_i. \tag{2}$$

To make sense of $\mathbb{P}_{V,\beta}^N$, it is usually assumed that V is a continuous function such that there is some $\beta'>1$, $\beta'\geq\beta$, such that

$$\liminf_{|x|\to+\infty}\frac{V(x)}{\beta'\log|x|}>1. \tag{3}$$

It is known (see [1, Theorem 2.6.1] or [4]), that the empirical measure

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

follows, under $\mathbb{P}_{V,\beta}^N$, a LDP with respect to the weak topology, with speed N^2, and good rate function I_β^V. Furthermore, I_β^V achieves its minimum at a unique probability measure σ_β^V, called the equilibrium measure, which is compactly supported (see [1, Lemma 2.6.2]).

In the case of β-ensembles associated with a convex potential with polynomial growth, the following holds.

Theorem 1.1. *Let $\alpha \geq 2$ and $\beta > 0$. Let*

$$\forall y \in \mathbb{R}, \quad V(x) = b|x|^\alpha + w(x), \tag{4}$$

where w is a continuous convex function such that $w(x) = o_{\pm\infty}(|x|^\alpha)$. Let $p \in \mathbb{N}$, $p > \alpha$. For any $\lambda_1, \ldots, \lambda_N \in \mathbb{R}$, we denote by $m_{p,N}$,

$$m_{p,N} = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^p.$$

Under $\mathbb{P}_{V,\beta}^N$, the sequence $(m_{p,N})_{N \geq 1}$ satisfies a large deviations principle with speed $N^{1+\frac{\alpha}{p}}$ and good rate function J_p, where $\mathbb{P}_{V,\beta}^N$ is defined in (1). If p is even,

$$J_p(x) = \begin{cases} b(x - \langle \sigma_\beta^V, x^p \rangle)^{\frac{\alpha}{p}} & \text{if } x \geq \langle \sigma_\beta^V, x^p \rangle, \\ +\infty & \text{otherwise,} \end{cases}$$

where $\langle \sigma_\beta^V, x^p \rangle$ denotes the pth moment of the equilibrium measure of $\mathbb{P}_{V,\beta}^N$, and if p is odd, J_p is defined by,

$$\forall x \in \mathbb{R}, \quad J_p(x) = b\left|x - \langle \sigma_\beta^V, x^p \rangle\right|^{\frac{\alpha}{p}}.$$

Remarks 1.2. (a) If $(X_1, \ldots, X_N)$ are independent and identically distributed with no exponential moments, it is known by the work of Nagaev [26] (see also [16]), that the large deviations of the empirical mean of such a sample, are essentially due to the deviations of order N of *one* variable, and that the speed of deviations depend on the tail distribution of the sample. This can be considered as a instance of some kind of "heavy-tail phenomenon".

Theorem 1.1 deals with the large deviations of the moments in a regime where this "heavy-tail phenomenon" arises. The assumption that the potential V is of polynomial growth is made so as to identify the speed of deviations and the rate function.

Indeed, under our assumptions on the potential V and p, one can see that the variables λ_i^p's do not have any exponential moments under $\mathbb{P}_{V,\beta}^N$. Thus, one can still expect some kind of "heavy-tail phenomenon" to hold despite the strong correlation; so that only a small fraction of the particles λ_i participate to the deviations of the moment $m_{p,N}$. As we will show in Proposition 3.5, the deviations of $m_{p,N}$ are due to the deviations of the $\log N$ largest in absolute value λ_i's, meaning that $m_{p,N}$ is exponentially equivalent to

$$\langle \sigma_\beta^V, x^p \rangle + \frac{1}{N} \sum_{i=1}^{\log N} \lambda_i^{*p},$$

where $\lambda_1^*, \ldots, \lambda_N^*$ is the rearrangement of the λ_i's in decreasing absolute values. To prove this, we will combine a deviation inequality stated in Proposition 3.1, estimating the probability of seeing a given number of particles outside an interval with a concentration inequality of Bobkov and Ledoux [6, Corollary 4.1] (see Proposition 3.7), together

with truncations arguments. This is in view of using this concentration inequality that we made the assumption that V can be written as in (4), with w a convex function.

Analyzing the deviations of this truncated moment where only the $\log N$ largest in absolute value λ_i's remain, one can show (see Proposition 3.9) that the logarithmic interaction is negligible, so that the rate function of $m_{p,N}$ can be realized as the rate function of the LDP of

$$\langle \sigma_\beta^V, x^p \rangle + \frac{1}{N} \sum_{i=1}^{\log N} X_i{}^p,$$

where $(X_i)_{i\geq 1}$ are i.i.d. random variables with law $e^{-NV(x)}\,dx/Z_V$, where we denote by $Z_V = \int e^{-NV(x)}\,dx$.

(b) The fact that the rate function obtained here in the case p is even, is infinite on the set $(-\infty, \langle \sigma_\beta^V, x^p \rangle)$ is not surprising. Indeed, the map which associates to a probability measure μ its pth-moment $\langle \mu, x^p \rangle$, is lower semi-continuous for the weak topology when p is even. Therefore, the deviations of $m_{p,N}$ under its limiting value $\langle \sigma_\beta^V, x^p \rangle$ can be reformulated as deviations of the empirical mean in a closed set which does not contain the equilibrium measure σ_β^V. In particular, these deviations have a infinite cost at the scale $N^{1+\alpha/p}$ (since it is strictly inferior by assumption to N^2). This remark will be relevant for the all subsequent LDPs we will present.

(c) Although it is not in the scope of this article, it should also be possible to derive a large deviations principle for the even moments of the empirical measure, say $m_{2p,N}$ with speed N^2. Firstly, the LDP result of Theorem 1.1 indicates that the rate function at speed N^2 must be 0 on $[\langle \sigma_\beta^V, x^{2p} \rangle, +\infty)$. Secondly, the proof of the large deviations of the empirical measure (see [1, Theorem 2.6.1]) yields the asymptotics of the partition function $Z_{V,\beta}^N$ at the exponential scale N^2,

$$\frac{1}{N^2} \log Z_{V,\beta}^N \xrightarrow[N\to+\infty]{} \inf_\mu \left\{ \int V\,d\mu(x) - \frac{\beta}{2} \int \log|x-y|\,d\mu(x)\,d\mu(y) \right\},$$

where the supremum is over all the probability measures μ on $\mathbb{R}$ satisfying $\int V\,d\mu < +\infty$.

But the logarithmic moment generating function of $m_{2p,N}$ at some tN^2 with $t<0$, can be written as a ratio of partition functions, that is

$$\mathbb{E}_{V,\beta}^N e^{tN^2 m_{p,N}} = \frac{Z_{V-tx^{2p},\beta}^N}{Z_{V,\beta}^N},$$

so that the one can identify the limit of the logarithmic moment generating function at the exponential scale N^2. Gärtner–Ellis theorem (see [11, Theorem 2.3.6]) would thus yield a large deviations principle with speed N^2, after checking the regularity of the limit of the logarithmic moment generating function.

Let us introduce now the model of Wigner matrices. The Wigner matrices and the β-ensembles are linked through the GOE, GUE and GSE, which form β-ensembles for a quadratic potential and $\beta = 1, 2, 4$ respectively. More generally, let $(X_{i,j})_{i<j}$ be independent and identically distributed (i.i.d) complex-valued random variables, and let $(X_{i,i})_{i\geq 1}$ be i.i.d. real-valued random variables. Let $X(N)$ be the $N \times N$ Hermitian matrix with up-diagonal entries $(X_{i,j})_{1\leq i\leq j\leq N}$. We call such a sequence $(X(N))_{N\in\mathbb{N}}$, a Wigner matrix. In the following, we will drop the N and write X instead of $X(N)$.

Assume further that X is centered, for any $k \in \mathbb{N}$,

$$\max\left(\mathbb{E}|X_{1,1}|^k, \mathbb{E}|X_{1,2}|^k\right) < +\infty,$$

and $\mathbb{E}|X_{1,2}|^2 = 1$. Consider now the normalized random matrix $X_N = X/\sqrt{N}$. Let λ_i denote the eigenvalues of X_N, with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. We define L_N the empirical spectral measure of X_N by,

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

Wigner's theorem (see [30], [1, Theorem 2.1.1, Exercice 2.1.16], [3, Theorem 2.5]) states that,

$$L_N \underset{N \to +\infty}{\leadsto} \sigma_{\text{sc}} \quad \text{a.s.,}$$

where σ_{sc} denotes the semicircular law, that is,

$$\sigma_{\text{sc}} = \frac{1}{2\pi} \mathbb{1}_{|x| \leq 2} \sqrt{4 - x^2} \, dx,$$

and for any $p \in \mathbb{N}$, almost surely, it holds

$$\langle L_N, x^p \rangle = \frac{1}{N} \operatorname{tr} X_N^p \underset{N \to +\infty}{\longrightarrow} \begin{cases} C_{p/2} & \text{if } p \text{ is even,} \\ 0 & \text{if } p \text{ is odd,} \end{cases}$$

and in the case p is even, $C_{p/2}$ denotes the $(\frac{p}{2})$th Catalan number, which is also the pth moment of the semicircular law.

In the following we will denote by $\mathcal{M}_N$ the set of matrices of size $N \times N$ with complex coefficients, and the normalized trace $\frac{1}{N} \operatorname{tr}$ by $\operatorname{tr}_N A$. We will also let $\beta \in \{1, 2\}$ and define $\mathcal{H}_N^{(\beta)}$ the set of real symmetric matrices of size N if $\beta = 1$, and complex Hermitian matrices if $\beta = 2$. In the case of Gaussian Wigner matrices, we have the following result.

Theorem 1.3. *Let $p \in \mathbb{N}$, $p \geq 3$. Let X be a centered Wigner matrix with Gaussian entries. We assume that $X_{1,1}$ has positive variance and $X_{1,2}$ is a centered Gaussian variable, possibly complex, such that $\mathbb{E}|X_{1,2}|^2 = 1$. The sequence $(\operatorname{tr}_N X_N^p)_{N \in \mathbb{N}}$, follows a LDP with speed $N^{1 + \frac{2}{p}}$, and good rate function J_p. If p is even, J_p is given by,*

$$\forall x \in \mathbb{R}, \quad J_p(x) = \begin{cases} c(x - C_{p/2})^{\frac{2}{p}} & \text{if } x \geq C_{p/2}, \\ +\infty & \text{if } x < C_{p/2}, \end{cases}$$

where $C_{p/2}$ is the $(p/2)$th Catalan number, whereas if p is odd,

$$J_p(x) = c|x|^{\frac{2}{p}},$$

where

$$c = \inf \left\{ q(H) : \operatorname{tr} H^p = 1, H \in \bigcup_N \mathcal{H}_N^{(\beta)} \right\}, \tag{5}$$

with q the quadratic form such that $Z^{-1} e^{-q} d\ell_N^{(\beta)}$ is the law of X, where $\ell_N^{(\beta)}$ is the Lebesgue measure on $\mathcal{H}_N^{(\beta)}$, and Z the normalizing constant.

Moreover, in the case where the entries of X are real or $(\Re X_{1,2}, \Im X_{1,2})$ are independent with the same variance, we can compute explicitly the constant c appearing in the above theorem by the following lemma.

Lemma 1.4. *Let X be a Gaussian Wigner satisfying the assumptions of Theorem 1.3. If X has real entries or if $(\Re X_{1,2}, \Im X_{1,2})$ are independent with the same variance, then the constant c defined in (5) is,*

$$c = \frac{1}{2} \min\left(\frac{1}{\sigma^2}, \frac{1}{2\beta} \right),$$

where $\sigma^2 = \mathbb{E} X_{1,1}^2$, and $\beta = 1$ if X is a real symmetric matrix and $\beta = 2$ if it is a complex Hermitian matrix.

Remarks 1.5. (a) In this case of Gaussian Wigner matrices, we will revisit the proof of the LDP of Wiener chaoses partially due to Borell [8] and then Ledoux [21]. Indeed, this problem of large deviations of the traces of such matrices can be reformulated as a large deviations' problem of Gaussian chaoses defined on a space of growing dimension. Even though it seems not possible to deduce directly such a LDP from the known LDP of the Wiener chaoses, we will show that the same outline of proof still work in our context, and that in particular, the deviations are also in a certain sense due to translations.

(b) The rate function of Theorem 1.3 actually arises under the following variational form,

$$\forall x \in \mathbb{R}, \quad J_p(x) = \inf\left\{ q(H) : x = \langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}\, H^p, H \in \bigcup_m \mathcal{H}_m^{(\beta)} \right\}, \tag{6}$$

where q is the quadratic form such that $Z^{-1} e^{-q} d\ell_N^{(\beta)}$ is the law of X. One way to interpret this formula, is to say that in order to make a deviation of $\mathrm{tr}_N X_N^p$ around some $x \in \mathbb{R}$, it is sufficient for X to make a translation by some $N^{1/2+1/p} H$ with H such that $x = \langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}\, H^p$, which one pays at the exponential scale $N^{1+2/p}$ by $q(H)$.

We consider now the so-called model of Wigner matrices without Gaussian tails investigated in [7]. We recall in the following definition this model.

Definition 1.6. We say that X is a *Wigner matrix without Gaussian tails*, if X is a Wigner matrix such that there exist $\alpha \in (0, 2)$ and $a, b \in (0, +\infty)$ such that,

$$\begin{aligned}
\lim_{t \to +\infty} -t^{-\alpha} \log \mathbb{P}\big(|X_{1,1}| > t\big) &= b, \\
\lim_{t \to +\infty} -t^{-\alpha} \log \mathbb{P}\big(|X_{1,2}| > t\big) &= a.
\end{aligned} \tag{7}$$

Moreover, we assume that there are two probability measures on $\mathbb{S}^1$, υ_1 and υ_2, and $t_0 > 0$, such that for all $t \geq t_0$ and any measurable subset U of $\mathbb{S}^1$,

$$\begin{aligned}
\mathbb{P}\big(X_{1,1}/|X_{1,1}| \in U, |X_{1,1}| \geq t\big) &= \upsilon_1(U)\mathbb{P}\big(|X_{1,1}| \geq t\big), \\
\mathbb{P}\big(X_{1,2}/|X_{1,2}| \in U, |X_{1,2}| \geq t\big) &= \upsilon_2(U)\mathbb{P}\big(|X_{1,2}| \geq t\big).
\end{aligned}$$

We again denote the normalized matrix $X_N = X/\sqrt{N}$.

With this definition, we can now state the following result.

Theorem 1.7. *Let $p \in \mathbb{N}$, $p \geq 3$. Let X be a Wigner matrix without Gaussian tail. The sequence $(\mathrm{tr}_N X_N^p)_{N \geq 1}$ satisfies a large deviations principle with speed $N^{\alpha(\frac{1}{2}+\frac{1}{p})}$ and good rate function J_p. If p is even, J_p is given by*

$$\forall y \in \mathbb{R}, \quad J_p(x) = \begin{cases} c_p(x - C_{p/2})^{\frac{\alpha}{p}} & \text{if } x \geq C_{p/2}, \\ +\infty & \text{if } x < C_{p/2}, \end{cases}$$

where $C_{p/2}$ denotes the $(\frac{p}{2})$th Catalan number, and if p is odd, the rate function J_p is given by

$$\forall x \in \mathbb{R}, \quad J_p(x) = c_p|x|^{\frac{\alpha}{p}},$$

where c_p is a constant depending on p, α, a and b.

Furthermore, if $\alpha \in (0, 1]$ and p is even, then $c_p = \min(b, 2^{-\alpha/p}a)$.

Remarks 1.8. (a) Note that for $p = 2$, the trace of X^2 is a sum of i.i.d. random variables, so that one can apply Cramer's theorem (see [11, Theorem 2.2.3]) in the case where $|X_{i,j}|^2$ have finite Laplace transform, or Nagaev's results [26].

(b) One can notice that the rate function consists essentially in the cost of deviation of one entry $X_{i,j}$ of order $N^{\frac{1}{2}+\frac{1}{p}}$. Indeed, such a deviation create a deviation of the spectral radius of X_N of order $N^{1/p}$, and therefore a deviation of the normalized trace $\mathrm{tr}_N X_N^p$ of order 1. In a similar fashion as for the LDP of the empirical spectral measure or the largest eigenvalue, the deviations are explained by large deviations of a small proportion of the entries which will here create large deviations of the edges of the spectrum.

(c) As for the Gaussian case, the rate function J_p arises in a variational form, since it is obtained by a contraction principle. One can write

$$\forall s \in \mathbb{R}, \quad J_p(s) = \inf\{I(A) : s = \langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}\, A^p, A \in \mathcal{D}\},$$

where I is defined for any $A \in \bigcup_{n \geq 1} \mathcal{H}_n^{(\beta)}$, by

$$I(A) = b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + a \sum_{i<j} |A_{i,j}|^\alpha,$$

and $\mathcal{D} = \{A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)} : \forall i \leq j, A_{i,j} = 0 \text{ or } A_{i,j}/|A_{i,j}| \in \mathrm{supp}(\nu_{i,j})\}$, with $\nu_{i,j} = \nu_1$ if $i = j$, and $\nu_{i,j} = \nu_2$ if $i < j$, where ν_1 and ν_2 are defined in Definition 1.6.

Thus, the constant c_p is the solution of an optimization problem which is,

$$c_p = \inf\{I(A) : 1 = \mathrm{tr}\, A^p, A \in \mathcal{D}\}.$$

We solve this optimization problem in Section 4.7, in the easiest case when $\alpha \in (0, 1]$ and p is even, and we give a lower and upper bound in the case p is even and $\alpha \in (0, 2)$.

We conclude this presentation of the results by a comment comparing the three theorems. The similarity of the rate functions of these three LDP comes from the fact that the deviations in these three cases are governed by the same "heavy-tail phenomenon". We observe in these three cases that the deviations of the traces are created by large deviations of the edges of the spectrum. For Wigner matrices without Gaussian tails, we are able to show that these deviations are actually explained by a small proportion of the entries. This picture somewhat contrasts with the Gaussian case, where the analysis of the minimizers of (6) (carried out in the proof of Lemma 1.4) reveals that the deviations can actually be due to deviations of *all* the off-diagonal entries, which is why the strategy of proof adopted for the Wigner matrices without Gaussian tails turn out to be ineffective when $\alpha = 2$.

2. The Gaussian case

We study in this section the question of the large deviations of the traces of a Wigner matrix with Gaussian entries. We will use an approach which is greatly inspired from Ledoux's proof of the LDP of Wiener chaoses [12,21], building on the work of Borell [8].

As we will see in the proof, the deviations of the trace are created by translations of X of the form $N^{\frac{1}{2}+\frac{1}{p}}H$, where H is a Hermitian matrix with bounded Hilbert–Schmidt norm. One of the central argument relies on the following lemma.

Lemma 2.1. *Let* $\|\cdot\|_2$ *denote the Hilbert–Schmidt norm on the set* $\mathcal{M}_N$ *of matrices of size* $N \times N$ *with complex coefficients. Let* X *be a Wigner matrix whose entries are centered and have finite* $(p+1)$*th moment. For any* $r > 0$,

$$\sup_{\|H\|_2 \leq r} \left| \mathrm{tr}_N \left(\frac{X}{\sqrt{N}} + N^{1/p} H \right)^p - \langle \sigma_{\mathrm{sc}}, x^p \rangle - \mathrm{tr}\, H^p \right| \underset{N \to +\infty}{\longrightarrow} 0, \tag{8}$$

in probability, where $\langle \sigma_{\mathrm{sc}}, x^p \rangle$ *denotes the* p*th moment of the semicircular law, and* tr_N *the normalized trace* $\frac{1}{N} \mathrm{tr}$ *on* $\mathcal{M}_N$.

Proof. By Wigner's theorem (see [1, Lemmas 2.1.6, 2.1.7]) and Jensen's inequality, we only have to prove that for any $Y, H \in \mathcal{M}_N$,

$$\left| \operatorname{tr}(Y + H)^p - \operatorname{tr} Y^p - \operatorname{tr} H^p \right| \leq 2^p \max_{1 \leq k \leq p-1} \left\{ \left(\operatorname{tr} |Y|^{p+1} \right)^{\frac{k}{p+1}} \left(\operatorname{tr} |H|^2 \right)^{\frac{p-k}{2}} \right\}. \tag{9}$$

Let $Y, H \in \mathcal{M}_N$. Expanding the trace, and using the cyclicity of the trace, it suffices to prove that for any $s \in \{1, \ldots, p\}, n_1, \ldots, n_s \in \mathbb{N}, m_1, \ldots, m_s \in \mathbb{N}$, such that $\sum_{i=1}^s n_i + \sum_{j=1}^s m_j = p$, we have

$$\left| \operatorname{tr}\left(Y^{n_1} H^{m_1} \cdots Y^{n_s} H^{m_s} \right) \right| \leq \left(\operatorname{tr} |Y|^{p+1} \right)^{\frac{k}{p+1}} \left(\operatorname{tr} |H|^2 \right)^{\frac{p-k}{2}},$$

with $k = \sum_{i=1}^s n_i$. Applying Hölder's inequality (see [5, Corollary IV.2.6]) with the exponents $\frac{p+1}{n_1}, \alpha, \frac{p+1}{n_2}, \ldots, \alpha, \frac{p+1}{n_s}$, with α such that

$$\frac{s}{\alpha} = 1 - \sum_{i=1}^s \frac{n_i}{p+1}, \tag{10}$$

we get,

$$\left| \operatorname{tr}\left(Y^{n_1} H^{m_1} \cdots Y^{n_s} H^{m_s} \right) \right| \leq \left(\operatorname{tr} |Y|^{p+1} \right)^{\frac{1}{p+1} \sum_{i=1}^s n_i} \prod_{j=1}^s \left(\operatorname{tr} |H|^{\alpha m_i} \right)^{\frac{1}{\alpha}}.$$

Note that when $s \geq 2$, we have from (10), $\alpha \geq 2$. If $s = 1$ and $m_1 = 1$, then as $p \geq 3$, (10) yields $\alpha m_1 \geq 2$. In any cases, $\alpha m_i \geq 2$ for any $i \in \{1, \ldots, s\}$. Therefore, for all $i \in \{1, \ldots, s\}$,

$$\operatorname{tr} |H|^{\alpha m_i} \leq \left(\operatorname{tr} |H|^2 \right)^{\frac{\alpha m_i}{2}}.$$

Thus,

$$\left| \operatorname{tr}\left(Y^{n_1} H^{m_1} \cdots Y^{n_s} H^{m_s} \right) \right| \leq \left(\operatorname{tr} |Y|^{p+1} \right)^{\frac{1}{p+1} \sum_{i=1}^s n_i} \left(\operatorname{tr} |H|^2 \right)^{\frac{1}{2} \sum_{i=1}^s m_i},$$

which gives the claim. $\qquad \square$

We are now ready to give a proof of Theorem 1.3. As in the proof of the LDP of Wiener chaoses, the upper bound will rely on a reformulation of the deviations of the trace in terms of an enlargement of a properly chosen event, which the Gaussian isoperimetric inequality will allow us to estimate the probability. Similarly as for the Wiener chaoses, the lower bound will make use of the translation formula for the Gaussian measure.

Proof of Theorem 1.3. We closely follow the outline of proof of the large deviations of Weiner chaoses in [12, Section 5, Theorem 5.1]. We will prove that $\operatorname{tr}_N X_N^p$ follows a LDP with speed $N^{1+\frac{2}{p}}$ and rate function

$$\forall s \in \mathbb{R}, \quad J_p(s) = \inf \left\{ q(H) : s = \left\langle \sigma_{\mathrm{sc}}, x^p \right\rangle + \operatorname{tr} H^p, H \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)} \right\},$$

where q is the quadratic form such that $Z^{-1} e^{-q} d\ell_N^{(\beta)}$ is the law of X. By homogeneity of q and of the trace, we see that this definition coincides with the rate function given in Theorem 1.3.

Upper bound. Let A be a closed subset of $\mathbb{R}$. We can assume without loss of generality that $\inf_A J_p > 0$, otherwise there is nothing to prove.

Let $0 < r < \inf_A J_p$. We define for any $N \in \mathbb{N}$,

$$\mathcal{K}_N = \left\{ H \in \mathcal{H}_N^{(\beta)} : q(H) \leq 1 \right\}, \qquad \mathcal{K} = \left\{ H \in \bigcup_{n \geq 1} \mathcal{H}_n^{(\beta)} : q(H) \leq 1 \right\},$$

and

$$\forall H \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}, \quad \varphi(H) = \langle \sigma_{sc}, x^p \rangle + \operatorname{tr} H^p.$$

We claim that,

$$\overline{\varphi(\sqrt{r}\mathcal{K})} \cap A = \varnothing. \tag{11}$$

Indeed, since $\inf_A J_p > r$,

$$\{J_p \le r\} \cap A = \varnothing.$$

By definition of J_p, we have

$$\varphi(\sqrt{r}\mathcal{K}) \subset \{J_p \le r\}.$$

As J_p is lower semi continuous, we deduce

$$\overline{\varphi(\sqrt{r}\mathcal{K})} \subset \{J_p \le r\},$$

which gives the claim.

Because $p \ge 2$, we have for any $H \in \mathcal{H}_N^{(\beta)}$,

$$\left(\operatorname{tr} |H|^p \right)^{2/p} \le \operatorname{tr} H^2 \le \frac{1}{s_1} q(H), \tag{12}$$

where s_1 is the smallest eigenvalue of q. As the covariance matrix of X is diagonal (if $\beta = 1$) or bloc diagonal with 2×2 blocs (if $\beta = 2$), this smallest eigenvalue of q does not depend on N.

We deduce that $\overline{\varphi(\sqrt{r}\mathcal{K})}$ is bounded. Thus it is a compact subset. From (11) we deduce that there is some $\eta > 0$ such that

$$\left(\varphi(\sqrt{r}\mathcal{K}) + B(0, \eta) \right) \cap A = \varnothing.$$

As $\mathcal{K}_N \subset \mathcal{K}$, we have for any $N \in \mathbb{N}$,

$$\left(\varphi(\sqrt{r}\mathcal{K}_N) + B(0, \eta) \right) \cap A = \varnothing.$$

Observe here that η does not depend on N. We deduce that

$$\mathbb{P}\left(\operatorname{tr}_N \left(\frac{X}{\sqrt{N}} \right)^p \in A \right) \le \mathbb{P}\left(\operatorname{tr}_N \left(\frac{X}{\sqrt{N}} \right)^p \notin \varphi(\sqrt{r}\mathcal{K}_N) + B(0, \eta) \right).$$

Let

$$V = \left\{ Y \in \mathcal{H}_N^{(\beta)} : \sup_{H \in \mathcal{K}_N} \left| \operatorname{tr}_N \left(\frac{Y}{\sqrt{N}} + N^{1/p} H \right)^p - \varphi(H) \right| < \eta \right\}.$$

Then,

$$\mathbb{P}\left(\operatorname{tr}_N \left(\frac{X}{\sqrt{N}} \right)^p \notin \varphi(\sqrt{r}\mathcal{K}_N) + B(0, \eta) \right) \le \mathbb{P}\left(X \notin V + \sqrt{r} N^{\frac{1}{2} + \frac{1}{p}} \mathcal{K}_N \right).$$

By Lemma 2.1, we know that for N large enough, $\mathbb{P}(X \in V) \ge 1/2$. The Gaussian isoperimetric inequality (see [12, Theorem 4.3]) yields

$$\mathbb{P}\left(X \notin V + \sqrt{r} N^{\frac{1}{2} + \frac{1}{p}} \mathcal{K}_N \right) \le e^{-rN^{1+2/p}}.$$

Therefore,

$$\mathbb{P}\left(\mathrm{tr}_N\left(\frac{X}{\sqrt{N}}\right)^p \in A\right) \le e^{-rN^{1+2/p}}.$$

Thus,

$$\limsup_{N \to +\infty} \frac{1}{N^{1+2/p}} \log \mathbb{P}\left(\mathrm{tr}_N\left(\frac{X}{\sqrt{N}}\right)^p \in A\right) \le -r.$$

Since the previous inequality is valid for any $0 < r < \inf_A J_p$, this yields the upper bound of the LDP.

Lower bound. Let A be an open subset of $\mathbb{R}$. Let $s \in A$. There is some $\eta > 0$ such that $B(s, \eta) \subset A$. We can assume without loss of generality that $J_p(s) < +\infty$. Define for any $N \in \mathbb{N}$,

$$\forall t \in \mathbb{R}, \quad J_{p,N}(t) = \inf_{H \in \mathcal{H}_N^{(\beta)}} \left\{ q(H) : t = \langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}\, H^p \right\}.$$

Note that

$$J_p = \lim_{N \to +\infty} J_{p,N}. \tag{13}$$

Indeed, if $H \in \mathcal{H}_n^{(\beta)}$, one can construct a sequence $H_N = H \oplus 0_N \in \mathcal{H}_{N+n}^{(\beta)}$, where 0_N denotes the zero matrix of size $N \times N$, which satisfies,

$$q(H_N) = q(H) \quad \text{and} \quad \mathrm{tr}\, H_N^p = \mathrm{tr}\, H^p,$$

so that for any $t \in \mathbb{R}$, $J_{p,N}(t)$ is decreasing in N.

As $J_p(s) < +\infty$, we deduce from (13) that we can find some $r > 0$ such that for N large enough $J_{p,N}(s) \le r^2$. We can then write,

$$J_{p,N}(s) = \inf_{H \in r\mathcal{K}_N} \left\{ q(H) : s = \langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}\, H^p \right\}.$$

Let $H \in r\mathcal{K}_N$ be such that $s = \langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}\, H^p$. Then,

$$\mathbb{P}\left(\mathrm{tr}_N\left(\frac{X}{\sqrt{N}}\right)^p \in A\right) \ge \mathbb{P}\left(\mathrm{tr}_N\left(\frac{X}{\sqrt{N}}\right)^p \in B(s, \eta)\right)$$

$$= \mathbb{P}\left(X \in V + N^{\frac{1}{2}+\frac{1}{p}} H\right),$$

where

$$V = \left\{ Y \in \mathcal{H}_N^{(\beta)} : \left| \mathrm{tr}_N\left(\frac{Y}{\sqrt{N}} + N^{\frac{1}{p}} H\right)^p - \langle \sigma_{\mathrm{sc}}, x^p \rangle - \mathrm{tr}\, H^p \right| < \eta \right\}.$$

But,

$$\mathbb{P}\left(X \in V + N^{\frac{1}{2}+\frac{1}{p}} H\right) = \frac{1}{Z} \int_V e^{-q(Y+N^{\frac{1}{2}+\frac{1}{p}} H)} \, d\ell_N^{(\beta)}(Y),$$

where $d\ell_N^{(\beta)}$ denotes the Lebesgue measure on $\mathcal{H}_N^{(\beta)}$. We re-write this probability as,

$$\mathbb{P}\left(X \in V + N^{\frac{1}{2}+\frac{1}{p}} H\right) = e^{-q(H)N^{1+\frac{2}{p}}} \mathbb{E}\left(\mathbb{1}_{\{X \in V\}} e^{-2N^{\frac{1}{2}+\frac{1}{p}} \Re \psi(H,X)}\right),$$

where ψ is the bilinear (or sesquilinear form if $\beta = 2$) form associated to the quadratic form q. Using Jensen's inequality, we get

$$\mathbb{P}\big(X \in V + N^{\frac{1}{2}+\frac{1}{p}}H\big) \geq e^{-q(H)N^{1+\frac{2}{p}}} \mathbb{P}(X \in V) \exp\left(-2N^{\frac{1}{2}+\frac{1}{p}}\mathbb{E}\left(\Re\psi(H, X)\frac{\mathbb{1}_{\{X \in V\}}}{\mathbb{P}(X \in V)}\right)\right).$$

Using Cauchy–Schwarz inequality yields,

$$\mathbb{E}\left(-\Re\psi(H, X)\frac{\mathbb{1}_{\{X \in V\}}}{\mathbb{P}(X \in V)}\right) \geq -\frac{1}{\mathbb{P}(X \in V)}\big(\mathbb{E}(\Re\psi(X, H))^2\big)^{1/2}.$$

But, as X is a Gaussian matrix with density $Z^{-1}e^{-q}$ with respect to the Lebesgue measure, we have

$$\mathbb{E}\big(\Re\psi(X, H)\big)^2 = \frac{1}{2}q(H).$$

Besides, note that by the second inequality of (12),

$$\mathcal{K}_N \subset \frac{1}{\sqrt{s_1}}B_2,$$

where B_2 denotes the unit ball of $\mathcal{H}_N^{(\beta)}$ for the Hilbert–Schmidt norm. As $H \in r\mathcal{K}_N$, we deduce by Lemma 2.1 that for N large enough, $\mathbb{P}(X \in V) \geq 1/2$. Thus, we have

$$\mathbb{P}\big(X \in V + N^{\frac{1}{2}+\frac{1}{p}}H\big) \geq \frac{1}{2}\exp\big(-q(H)N^{1+\frac{2}{p}} - 2\sqrt{2}q(H)^{1/2}N^{\frac{1}{2}+\frac{1}{p}}\big).$$

Since $H \in r\mathcal{K}_N$, we get

$$\mathbb{P}\big(X \in V + N^{\frac{1}{2}+\frac{1}{p}}H\big) \geq \frac{1}{2}\exp\big(-q(H)N^{1+\frac{2}{p}} - 2\sqrt{2}rN^{\frac{1}{2}+\frac{1}{p}}\big).$$

As the above inequality is true for any $H \in r\mathcal{K}_N$ such that $s = \langle\sigma_{\mathrm{sc}}, x^p\rangle + \mathrm{tr}\,H^p$, we have

$$\mathbb{P}\big(X \in V + N^{\frac{1}{2}+\frac{1}{p}}H\big) \geq \frac{1}{2}\exp\big(-J_{p,N}(s)N^{1+\frac{2}{p}} - 2\sqrt{2}rN^{\frac{1}{2}+\frac{1}{p}}\big).$$

We deduce from (13) that

$$\liminf_{N\to+\infty}\frac{1}{N^{1+2/p}}\log\mathbb{P}\left(\mathrm{tr}_N\left(\frac{X}{\sqrt{N}}\right)^p \in A\right) \geq -J_p(s),$$

which gives the lower bound. $\qquad\square$

We end this section with a proof of Lemma 1.4.

Proof of Lemma 1.4. Let X be a Gaussian Wigner matrix satisfying the assumptions of Theorem 1.3. Assume further that X has real entries or $(\Re X_{1,2}, \Im X_{1,2})$ are independent with variance $1/2$. Then we can write,

$$\forall H \in \mathcal{H}_N^{(\beta)}, \quad q(H) = \frac{1}{2\sigma^2}\sum_i H_{i,i}^2 + \frac{\beta}{2}\sum_{i<j}|H_{i,j}|^2,$$

with $\sigma^2 = \mathbb{E}X_{1,1}^2$. Then, for any $H \in \mathcal{H}_N^{(\beta)}$, we have

$$q(H) \geq \min\left(\frac{1}{2\sigma^2}, \frac{\beta}{4}\right)\mathrm{tr}\,H^2.$$

As $p \geq 2$, we get

$$q(H) \geq \min\left(\frac{1}{2\sigma^2}, \frac{\beta}{4}\right) \left|\operatorname{tr} H^p\right|^{2/p}, \tag{14}$$

which yields,

$$c \geq \min\left(\frac{1}{2\sigma^2}, \frac{\beta}{4}\right),$$

where c is defined in (5). Observe that we always have $c \leq 1/(2\sigma^2)$ by taking a size one matrix. Now define

$$H = \begin{pmatrix} 0 & \lambda & \cdots\cdots & \lambda \\ \lambda & & & \vdots \\ \vdots & & & \lambda \\ \lambda & \cdots\cdots & \lambda & 0 \end{pmatrix} \in \mathcal{H}_N^{(1)},$$

with $\lambda = \left(\frac{1}{(N-1)^p + (N-1)}\right)^{1/p}$. We have $\operatorname{tr} H^p = 1$, and

$$q(H) = \frac{N(N-1)\beta}{4} \left(\frac{1}{(N-1)^p + (N-1)}\right)^{2/p} \xrightarrow[N \to +\infty]{} \frac{\beta}{4}.$$

This yields $c = \min(\frac{1}{2\sigma^2}, \frac{\beta}{4})$, which ends the proof. $\qquad\square$

3. Large deviations of moments of the empirical measure of β-ensembles

We now give a proof of Theorem 1.1. In order to ease the notation, we will write $\mathbb{P}_V^N$ for $\mathbb{P}_{V,\beta}^N$, as well as Z_V^N instead of $Z_{V,\beta}^N$.

3.1. *Deviations inequalities and convergence of the moments*

The first step of the proof of Theorem 1.1 will be to show, under the mild assumption (3), the convergence in expectation, of the moments of the empirical measure towards the moments of the equilibrium measure σ_β^V. To do so, we will need a control on the tail probability of

$$\max_{1 \leq i \leq N} |\lambda_i|,$$

under $\mathbb{P}_V^N$. To this end we prove a more general deviations inequality, which will be crucial later.

Proposition 3.1. *Let $N \in \mathbb{N}$, $N \geq 2$. Under assumption (3), there is a constant $M_0 > 0$, depending only on V and β, such that for any $M \geq M_0$ and $1 \leq k \leq N$,*

$$\mathbb{P}_V^N\left(L_N\left(I_M^c\right) \geq \frac{k}{N}\right) \leq \exp(-CkNV_M),$$

where $I_M = [-M, M]$, C is a positive constant depending on V and β, and where $V_M = \inf_{|\lambda| \geq M} V(\lambda)$.

Remark 3.2. The proof of this proposition looks like the one of the LDP for the largest particle of the β-ensembles (see [1, Theorem 2.6.6]), but note that this inequality is not a consequence of this LDP, as we allow k to take all the range of integers from 1 to n. One can consider this deviation inequality as interpolating deviations inequalities of the largest particle ($k = 1$) and the empirical measure of β-ensembles ($k = n$). Later we will use this inequality with k and M depending on n, this is why we take some care to prove here a *non-asymptotic* deviation inequality.

In order to prove this deviation inequality, we will need a rough control on the ratio of the partition functions Z_V^N and $Z_{\frac{NV}{N-k}}^{N-k}$. This is the object of the following lemma.

Lemma 3.3. *There are some constants c_1, c_2 depending on V and β, such that for any $N \in \mathbb{N}$, and $k \leq N$,*

$$c_1 N k \leq \log \frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} \leq c_2 N k,$$

where Z_V^N, and $Z_{\frac{NV}{N-k}}^{N-k}$ are defined in (2).

Proof. From the invariance under permutation of the coordinates of the measures $\mathbb{P}_V^N$ we have

$$\frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} = \frac{N!}{Z_{\frac{NV}{N-k}}^{N-k}} \int_{|\lambda_1| \geq \cdots \geq |\lambda_N|} e^{-N \sum_{i=1}^N V(\lambda_i)} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^N d\lambda_i.$$

Splitting the λ_i's between the k first largest in absolute value and the rest, and using again the invariance under permutation of the coordinates, we can bring out the measure $\mathbb{P}_{\frac{NV}{N-k}}^{N-k}$, which gives

$$\frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} = \frac{N!}{(N-k)!} \mathbb{E}_{\frac{NV}{N-k}}^{N-k} \left(\int_{|\lambda_1| \geq \cdots \geq |\lambda_k|} e^{-N \sum_{i=1}^k V(\lambda_i)} \prod_{1 \leq i < j \leq k} |\lambda_i - \lambda_j|^\beta \right.$$

$$\left. \times e^{\beta(N-k) \sum_{i=1}^k \langle L_{N-k}, \log |\lambda_i - \cdot| \rangle} \mathbb{1}_{\text{supp}(L_{N-k}) \subset [-\lambda_k, \lambda_k]} \prod_{i=1}^k d\lambda_i \right),$$

where $L_{N-k} = \frac{1}{N-k} \sum_{i=k+1}^N \delta_{\lambda_i}$. We re-write this equality as the following,

$$\frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} = \frac{N!}{(N-k)!} \mathbb{E}_{\frac{NV}{N-k}}^{N-k} \left(\int_{|\lambda_1| \geq \cdots \geq |\lambda_k|} e^{-k^2 \int_{x \neq y} f(x,y) \, dL_k(x) \, dL_k(y)} \right.$$

$$\left. \times e^{-(N-k) \sum_{i=1}^k (V(\lambda_i) - \beta \langle L_{N-k}, \log |\lambda_i - \cdot| \rangle)} \mathbb{1}_{\text{supp}(L_{N-k}) \subset [-\lambda_k, \lambda_k]} \prod_{i=1}^k e^{-V(\lambda_i)} \, d\lambda_i \right),$$

with $L_k = \frac{1}{k} \sum_{i=1}^k \delta_{\lambda_i}$, and $f(x, y) = \frac{1}{2} V(x) + \frac{1}{2} V(y) - \frac{\beta}{2} \log |x - y|$. Note that from the assumption (3) on V, we have

$$c := \inf \{ f(x, y) : x \neq y \} > -\infty,$$

$$c' := \inf \{ V(x) - \beta \log |x - y| : |y| \leq |x| \} > -\infty.$$

Thus,

$$\frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} \leq \binom{N}{k} e^{-k^2 c} e^{-(N-k)kc'} \left(\int e^{-V(x)} \, dx \right)^k.$$

As $\binom{N}{k} \le N^k$, we get

$$\frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} \le e^{c_2 N k},$$

with c_2 some constant depending on V and β.

For the lower bound, we write similarly as for the upper bound,

$$\log \frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} = \log \mathbb{E}_{\frac{NV}{N-k}}^{N-k} \left(\int e^{-(N-1) \sum_{i=1}^k V(\lambda_i)} \prod_{1 \le i < j \le k} |\lambda_i - \lambda_j|^\beta \right.$$

$$\left. \times\, e^{\beta(N-k) \sum_{i=1}^k \langle L_{N-k}, \log|\lambda_i - \cdot| \rangle} \prod_{i=1}^k e^{-V(\lambda_i)} \, d\lambda_i \right).$$

Using twice Jensen's inequality, we get

$$\log \frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} \ge \mathbb{E}_{\frac{NV}{N-k}}^{N-k} \left(\log \int e^{-(N-1) \sum_{i=1}^k V(\lambda_i)} \prod_{1 \le i < j \le k} |\lambda_i - \lambda_j|^\beta \right.$$

$$\left. \times\, e^{\beta(N-k) \sum_{i=1}^k \langle L_{N-k}, \log|\lambda_i - \cdot| \rangle} \prod_{i=1}^k \frac{e^{-V(\lambda_i)}}{\int e^{-V(x)} \, dx} \, d\lambda_i \right) + k \log \left(\int e^{-V(x)} \, dx \right)$$

$$\ge -(N-1)k \left(\int V(\lambda) \frac{e^{-V(\lambda)} \, d\lambda}{\int e^{-V(x)} \, dx} \right) + \frac{k(k-1)\beta}{2} \left(\int \log|\lambda - \mu| \frac{e^{-V(\lambda) - V(\mu)} \, d\lambda \, d\mu}{(\int e^{-V(x)} \, dx)^2} \right)$$

$$+ \beta k(N-k) \mathbb{E}_{\frac{NV}{N-k}}^{N-k} \left(\int \langle L_{N-k}, \log|\lambda - \cdot| \rangle \frac{e^{-V(\lambda)} \, d\lambda}{\int e^{-V(x)} \, dx} \right) + k \log \left(\int e^{-V(x)} \, dx \right).$$

But for any $\mu \in \mathbb{R}$,

$$\int \log|\lambda - \mu| e^{-V(\lambda)} \, d\lambda = \int_0^{+\infty} \log x \left(e^{-V(\mu + x)} + e^{-V(\mu - x)} \right) dx$$

$$\ge \int_0^1 \log x \left(e^{-V(\mu + x)} + e^{-V(\mu - x)} \right) dx.$$

As $\inf V < -\infty$, we have

$$\int \log|\lambda - \mu| e^{-V(\lambda)} \, d\lambda \ge 2 e^{-\inf V} \int_0^1 \log(x) \, dx = -2 e^{-\inf V}.$$

Thus,

$$\mathbb{E}_{\frac{NV}{N-k}}^{N-k} \left(\int \langle L_{N-k}, \log|\lambda - \cdot| \rangle \frac{e^{-V(\lambda)} \, d\lambda}{\int e^{-V(x)} \, dx} \right) \ge -\frac{2 e^{-\inf V}}{\int e^{-V(x)} \, dx}.$$

We can conclude that

$$\log \frac{Z_V^N}{Z_{\frac{NV}{N-k}}^{N-k}} \ge c_1 N k,$$

with c_1 a constant depending on V and β. $\square$

We are now ready to give a proof of Proposition 3.1.

Proof of Proposition 3.1. We can write as in the proof of Lemma 3.3,

$$
\mathbb{P}_V^N\left(L_N\left(I_M^c\right) \geq \frac{k}{N}\right) \leq \frac{N!}{(N-k)!} \frac{Z_{\frac{NV}{N-k}}^{N-k}}{Z_V^N}
$$

$$
\times \mathbb{E}_{\frac{NV}{N-k}}^{N-k}\left(\int_{|\lambda_1| \geq \cdots \geq |\lambda_k| \geq M} e^{-N \sum_{i=1}^k V(\lambda_i)} \prod_{1 \leq i < j \leq k} |\lambda_i - \lambda_j|^\beta\right.
$$

$$
\left.\times e^{\beta(N-k)\sum_{i=1}^k \langle L_{N-k}, \log|\lambda_i - \cdot|\rangle} \mathbb{1}_{\text{supp}(L_{N-k})\subset[-\lambda_k,\lambda_k]} \prod_{i=1}^k d\lambda_i\right),
$$

with $L_{N-k} = \frac{1}{N-k}\sum_{i=k+1}^N \delta_{\lambda_i}$.

As for all $x, y \in \mathbb{R}$, $\log|x-y| \leq \log(1+|x|)+\log(1+|y|)$, and for any $|x| \leq |y|$, $\log|x-y| \leq \log 2 + \log(1+|x|)$, we get

$$
\mathbb{P}_V^N\left(L_N\left(I_M^c\right) \geq \frac{k}{N}\right) \leq \frac{N!}{(N-k)!} \frac{Z_{\frac{NV}{N-k}}^{N-k}}{Z_V^N} e^{k(N-k)\log 2}
$$

$$
\times \mathbb{E}_{\frac{NV}{N-k}}^{N-k}\left(\int_{|\lambda_1| \geq \cdots \geq |\lambda_k| \geq M} e^{-N \sum_{i=1}^k V(\lambda_i)} e^{\beta k \sum_{i=1}^k \log(1+|\lambda_i|)}\right.
$$

$$
\left.\times e^{\beta(N-k)\sum_{i=1}^k \log(1+|\lambda_i|)} \mathbb{1}_{\text{supp}(L_{N-k})\subset[-\lambda_k,\lambda_k]} \prod_{i=1}^k d\lambda_i\right).
$$

From (3), we deduce that there is some $c_0 > 0$, such that for $|y|$ large enough,

$$
V(y) - \beta \log(1+|y|) \geq c_0 V(y).
$$

Thus, for M large enough,

$$
\mathbb{P}_V^N\left(L_N\left(I_M^c\right) \geq \frac{k}{N}\right) \leq \frac{N!}{(N-k)!} \frac{Z_{\frac{NV}{N-k}}^{N-k}}{Z_V^N} \int_{|\lambda_1| \geq \cdots \geq |\lambda_k| \geq M} e^{-c_0 N \sum_{i=1}^k V(\lambda_i)} \prod_{i=1}^k d\lambda_i
$$

$$
= \binom{N}{k} \frac{Z_{\frac{NV}{N-k}}^{N-k}}{Z_V^N}\left(\int_{|\lambda| \geq M} e^{-c_0 N V(\lambda)} d\lambda\right)^k.
$$

But,

$$
\int_{|\lambda| \geq M} e^{-c_0 N V(\lambda)} d\lambda \leq e^{-c_0(N-1)V_M} \int e^{-V(\lambda)} d\lambda \leq c_3 e^{-\frac{c_0}{2} N V_M},
$$

with some constant $c_3 > 0$, and where we used in the last inequality the fact that $N \geq 2$. We deduce from Proposition 3.3 that for M large enough,

$$
\mathbb{P}_V^N\left(L_N\left(I_M^c\right) \geq \frac{k}{N}\right) \leq (c_3 N)^k e^{kNc_2} e^{-\frac{c_0}{2} k N V_M}.
$$

As $\lim_{M\to+\infty} V_M = +\infty$, we can find some constants $M_0 > 0$, and $C > 0$, depending on V and β, such that for any $M > M_0$,

$$\mathbb{P}_V^N \left(L_N(I_M^c) \geq \frac{k}{N} \right) \leq e^{-CkNV_M}.$$
□

As a consequence of the previous Proposition 3.1, we have the convergence of the expectation under $\mathbb{P}_V^N$, of the moments of the empirical measure, as stated in the next corollary.

Corollary 3.4. *Under assumption* (3), *we have for any* $p \in \mathbb{N}$,

$$\mathbb{E}_V^N \langle L_N, x^p \rangle \xrightarrow[N\to+\infty]{} \langle \sigma_\beta^V, x^p \rangle,$$

where $\mathbb{E}_V^N$ *denotes the expectation with respect to* $\mathbb{P}_V^N$.

Proof. Since $(L_N)_{N\geq 1}$ follows a LDP with speed N^2 (see [1, Theorem 2.6.1]), and rate function whose minimum is achieved at σ_β^V, we deduce that $(L_N)_{N\in\mathbb{N}}$ converges weakly in probability to σ_β^V under $\mathbb{P}_V^N$. Thus, it is enough to show that for any $k \in \mathbb{N}$,

$$\sup_{N\geq N_0} \mathbb{E}_V^N \langle L_N, |x|^k \rangle < +\infty,$$

for some $N_0 \geq 1$.

Let $k \in \mathbb{N}$. We have $\langle L_N, |x|^k \rangle \leq \max_{1\leq i\leq N} |\lambda_i|^k$. Besides, we know by Proposition 3.1 that

$$\mathbb{P}_V^N \left(\max_{1\leq i\leq N} |\lambda_i| > M \right) \leq e^{-CNV_M},$$

for any $M > M_0$, where C and M_0 are some positive constants. Thus,

$$\mathbb{E}_V^N \max_{1\leq i\leq N} |\lambda_i|^k \leq M_0^k + \int_{M_0}^{+\infty} kx^{k-1} e^{-CNV_x} \, dx.$$

By assumption we know that for $|x|$ large enough, $V_x \geq \beta' \log |x|$, with $\beta' > 1$, so that for M_0 large enough,

$$\mathbb{E}_V^N \max_{1\leq i\leq N} |\lambda_i|^k \leq M_0^k + \int_{M_0}^{+\infty} kx^{k-1} x^{-C\beta'N} \, dx.$$

We deduce that for $N \geq (k+1)/C\beta'$, and M_0 large enough,

$$\mathbb{E}_V^N \max_{1\leq i\leq N} |\lambda_i|^k \leq M_0^k + \int_{M_0}^{+\infty} kx^{-2} \, dx = M_0^k + \frac{k}{M_0}, \tag{15}$$

which yields the claim.
□

3.2. *An exponential equivalence*

The goal of this section is to prove that the large deviations of $m_{p,N}$ are due to the deviations of the $\log N$ largest in absolute value λ_i's. More precisely, we will prove the following proposition.

Proposition 3.5. *For any* $p \in \mathbb{N}$, $p > \alpha$, *and* $\lambda_1, \ldots, \lambda_N \in \mathbb{R}$, *we denote by* $T_{p,N}$ *the truncated moment*

$$T_{p,N} = \frac{1}{N} \sum_{i=1}^{\log N} \lambda_i^{*p},$$

where $\lambda_1^*, \ldots, \lambda_N^*$ is the rearrangement of the λ_i's by decreasing absolute values. Under the notation and assumption of Theorem 1.1, *we have for any $t > 0$,*

$$\lim_{N \to +\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}_V^N \left(\left| m_{p,N} - \langle \sigma_\beta^V, x^p \rangle - T_{p,N} \right| > t \right) = -\infty.$$

As a consequence of Proposition 3.1 and Corollary 3.4, we have the following result.

Proposition 3.6. *Under assumption* (3), *we have*

$$\frac{1}{N} \mathbb{E}_V^N \left(\sum_{i=\log N+1}^{N} \lambda_i^{*p} \right) \xrightarrow[N \to +\infty]{} \langle \sigma_\beta^V, x^p \rangle.$$

Proof. Due to Corollary 3.4, we only need to prove

$$\frac{1}{N} \sum_{i=1}^{\log N} \mathbb{E}_V^N \lambda_i^{*p} \xrightarrow[N \to +\infty]{} 0.$$

From (15) we have

$$\sup_{N \geq N_0} \mathbb{E}_V^N \left| \lambda_1^* \right|^p < +\infty, \tag{16}$$

with $N_0 \in \mathbb{N}$. Thus for any $N \geq N_0$,

$$\left| \frac{1}{N} \sum_{i=1}^{\log N} \mathbb{E}_V^N \lambda_i^{*p} \right| \leq \frac{\log N}{N} \sup_{N \geq N_0} \mathbb{E}_V^N \left| \lambda_1^* \right|^p \xrightarrow[N \to +\infty]{} 0. \qquad \square$$

Due to the previous proposition, in order to prove Proposition 3.5, it suffices to show that

$$\frac{1}{N} \sum_{i=\log N+1}^{N} \lambda_i^{*p}$$

concentrates at the exponential scale, at a speed higher than $N^{1+\alpha/p}$. To this end, we will use concentration inequalities for measures with α-convex potential from [6]. This is the object of the following proposition.

Proposition 3.7. *Let $\alpha \geq 2$. Let $g : \mathbb{R}^N \to \mathbb{R}$ be a 1-Lipschitz function with respect to $\| \cdot \|_\alpha$. Under the notation and assumption of Theorem* 1.1, *we have for every $t > 0$,*

$$\mathbb{P}_V^N \left(g - \mathbb{E}_V^N g > t \right) \leq \exp\left(-\frac{bNt^\alpha}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}} \right).$$

In particular, if $f : \mathbb{R} \to \mathbb{R}$ is a 1-Lipschitz function, and $l, m \in \{1, \ldots, N\}$, $l \leq m$, then for any $t > 0$,

$$\mathbb{P}_V^N \left(\frac{1}{N} \sum_{i=l}^{m} f(\overline{\lambda}_i) - \frac{1}{N} \mathbb{E}_V^N \sum_{i=l}^{m} f(\overline{\lambda}_i) > t \right) \leq \exp\left(-\frac{bN^2 t^\alpha}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}} \right),$$

where $\overline{\lambda}_1, \ldots, \overline{\lambda}_N$ is the rearrangement of the λ_i's in ascending order.

Proof. Let

$$\forall \lambda \in \mathbb{R}^N, \quad \Phi(\lambda) = N \sum_{i=1}^{N} V(\lambda_i) - \frac{\beta}{2} \sum_{i \neq j} \log |\lambda_i - \lambda_j|.$$

We claim that Φ is α-convex with respect to the norm $\|\cdot\|_\alpha$ on $\mathbb{R}^N$, more precisely we will show that for all $\lambda, \mu \in \mathbb{R}^N$,

$$\Phi(\lambda) + \Phi(\mu) - 2\Phi\left(\frac{\lambda + \mu}{2}\right) \geq \frac{bN}{2^{\alpha-1}} \|\lambda - \mu\|_\alpha^\alpha. \tag{17}$$

Note that for any $k, l \in \{1, \ldots, N\}$,

$$\text{Hess}\left(-\beta \sum_{i \neq j} \log|\lambda_i - \lambda_j|\right)_{k,l} = \begin{cases} -(\lambda_k - \lambda_l)^{-2} & \text{if } k \neq l, \\ \sum_{j \neq k} (\lambda_j - \lambda_k)^{-2} & \text{if } k = l, \end{cases}$$

which defines a non-negative matrix since for any $x \in \mathbb{R}^N$,

$$\sum_{k \neq l} (\lambda_k - \lambda_l)^{-2} x_k^2 - \sum_{k \neq l} (\lambda_k - \lambda_l)^{-2} x_k x_l = \sum_{k < l} (\lambda_k - \lambda_l)^{-2} (x_k - x_l)^2 \geq 0.$$

As by assumption

$$\forall x \in \mathbb{R}, \quad V(x) = b|x|^\alpha + w(x),$$

with w a convex function, we deduce together with the above observation, for any $\lambda, \mu \in \mathbb{R}^N$,

$$\Phi(\lambda) + \Phi(\mu) - 2\Phi\left(\frac{\lambda + \mu}{2}\right) \geq bN\left(\sum_{i=1}^N \lambda_i^\alpha + \sum_{i=1}^N \mu_i^\alpha - 2\sum_{i=1}^N \left(\frac{\lambda_i + \mu_i}{2}\right)^\alpha\right).$$

Since $\alpha \geq 2$, we have for any $x, y \in \mathbb{R}$,

$$\frac{1}{2}x^\alpha + \frac{1}{2}y^\alpha \geq \left(\frac{x+y}{2}\right)^\alpha + \left(\frac{x-y}{2}\right)^\alpha.$$

This yields the desired inequality (17).

We know, by [6, Corollary 4.1], that (17) entails that for any 1-Lipschitz function with respect to $\|\cdot\|_\alpha$, $g : \mathbb{R}^N \to \mathbb{R}$, and every $t > 0$,

$$\mathbb{P}_V^N\left(g - \mathbb{E}_V^N g > t\right) \leq \exp\left(-\frac{bNt^\alpha}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}}\right). \tag{18}$$

Let now $f : \mathbb{R} \to \mathbb{R}$ be a 1-Lipschitz function, and $k, l \in \{1, \ldots, N\}$, $k \leq l$. We set

$$\forall \lambda \in \mathbb{R}^N, \quad g(\lambda) = \frac{1}{N}\sum_{i=l}^m f(\overline{\lambda_i}).$$

For any $\lambda, \mu \in \mathbb{R}^N$, we have by Cauchy–Schwarz inequality

$$g(\lambda) - g(\mu) \leq \frac{1}{N}\sum_{i=l}^m |\overline{\lambda_i} - \overline{\mu_i}| \leq \frac{1}{N^{1/2}}\left(\sum_{i=l}^m |\overline{\lambda_i} - \overline{\mu_i}|^2\right)^{1/2} \leq \frac{1}{N^{1/2}}\left(\sum_{i=1}^N |\lambda_i - \mu_i|^2\right)^{1/2},$$

where we used in the last inequality Hardy–Littlewood–Polyá rearrangement inequality. Thus, by Hölder inequality

$$g(\lambda) - g(\mu) \leq N^{-\frac{1}{\alpha}} \|\lambda - \mu\|_\alpha.$$

This shows that g is $N^{-\frac{1}{\alpha}}$-Lipschitz with respect to the norm $\|\cdot\|_\alpha$. Applying (18) to g gives the second inequality in the statement. $\qquad\square$

In the following proposition, we use the concentration inequalities of Proposition 3.7, together with a truncation procedure and the deviations estimate of Proposition 3.1, to prove that

$$\frac{1}{N} \sum_{i=\log N+1}^{N} \lambda_i^{*p},$$

is exponentially equivalent to its expectation with respect to $\mathbb{P}_V^N$. Combining this with the result of Proposition 3.6,

$$\frac{1}{N} \sum_{i=\log N+1}^{N} \mathbb{E}_V^N \lambda_i^{*p} \xrightarrow[N\to+\infty]{} \langle \sigma_{\mathrm{sc}}, x^p \rangle,$$

we will get Proposition 3.5.

Proposition 3.8. *For any $t > 0$,*

$$\limsup_{N\to+\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}_V^N \left(\left| \sum_{i=\log N+1}^{N} \lambda_i^{*p} - \sum_{i=\log N+1}^{N} \mathbb{E}_V^N \lambda_i^{*p} \right| \geq tN \right) = -\infty,$$

where $\lambda_1^*, \ldots, \lambda_N^*$ denotes the rearrangement of the λ_i's by decreasing absolute values.

Proof. To ease the notation, we set $k = \log N$. The first part of the argument consists in choosing the proper truncation level with respect to our exponential scale $N^{1+\alpha/p}$. For any $M_0 > 0$, we denote by F_{M_0} the function

$$\forall x \in \mathbb{R}, \quad F_{M_0}(x) = \begin{cases} \mathrm{sg}(x)(|x| \wedge M_0)^p & \text{if } p \text{ is odd}, \\ (|x| \wedge M_0)^p & \text{if } p \text{ is even}. \end{cases}$$

Let

$$M_0 = \frac{N^{\frac{1}{\alpha(p-1)}(1-\frac{\alpha}{p})}}{k^{\frac{1}{2\alpha}}}.$$

Note that,

$$\left| \frac{1}{N} \sum_{i=k+1}^{N} \mathbb{E}_V^N \left(\lambda_i^{*p} - F_{M_0}(\lambda_i^*) \right) \right| \leq \frac{1}{N} \sum_{i=k+1}^{N} \mathbb{E}_V^N |\lambda_i^*|^p \mathbb{1}_{|\lambda_i^*| \geq M_0}$$

$$\leq \frac{1}{NM_0} \sum_{i=k+1}^{N} \mathbb{E}_V^N |\lambda_i^*|^{p+1}$$

$$\leq \frac{N-k}{NM_0} \mathbb{E}_V^N |\lambda_1^*|^{p+1} \xrightarrow[N\to+\infty]{} 0,$$

using (16), and the fact that as $p > \alpha$, $M_0 \to +\infty$. Thus, it suffices to prove that for any $t > 0$,

$$\limsup_{N\to+\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}_V^N \left(\left| \sum_{i=k+1}^{N} \lambda_i^{*p} - \sum_{i=k+1}^{N} \mathbb{E}_V^N F_{M_0}(\lambda_i^*) \right| \geq tN \right) = -\infty.$$

Note that,

$$\sum_{i=k+1}^{N} F_{M_0}(\lambda_i^*) = \sum_{j=(k-l)+1}^{N-l} F_{M_0}(\overline{\lambda_i}),$$

where $l = \text{Card}\{i \in \{1, \dots, k\} : \lambda_i^* > 0\}$. Since the function F_{M_0} is pM_0^{p-1}-Lipschitz, we have using a union bound and Proposition 3.7, for any $t > 0$,

$$\mathbb{P}_V^N\left(\left|\sum_{i=k+1}^N F_{M_0}(\lambda_i^*) - \sum_{i=k+1}^N \mathbb{E}_V^N F_{M_0}(\lambda_i^*)\right| > tN\right) \le 2k \exp\left(-\frac{1}{c_\alpha p^\alpha} t^\alpha N^{1+\frac{\alpha}{p}} k^{\frac{p-1}{2}}\right),$$

where c_α is some constant depending on α. We can write,

$$\mathbb{P}_V^N\left(\left|\sum_{i=k+1}^N \lambda_i^{*p} - \sum_{i=k+1}^N \mathbb{E}_V^N F_{M_0}(\lambda_i^*)\right| > Nt\right)$$

$$\le \mathbb{P}_V^N\left(\left|\sum_{i=k+1}^N F_{M_0}(\lambda_i^*) - \sum_{i=k+1}^N \mathbb{E}_V^N F_{M_0}(\lambda_i^*)\right| > tN/2\right) + \mathbb{P}_V^N\left(\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \le |\lambda_i^*|} > tN/2\right).$$

We saw by the concentration inequality above, that the deviations of the truncated moment at the level M_0 around its mean are exponentially negligible at the scale $N^{1+\alpha/p}$. We need now to prove that the contributions in the deviations of the truncated moments of the λ_i's above the level M_0 are also negligible. To do so, we will truncate one more time at a level R, chosen so that the deviation bound of Proposition 3.1 gives the right exponential estimate.

From (4), we have for M large enough,

$$\inf_{|x| \ge M} V(x) \ge \frac{b}{2} M^\alpha.$$

Proposition 3.1 yields that there are some constants $M_0 > 0$, and $C > 0$, depending on V and β, such that for any $M > M_0$, and $k \in \{1, \dots, N\}$,

$$\mathbb{P}_V^N\left(L_N(I_M^c) \ge \frac{k}{N}\right) \le \exp(-CkNM^\alpha). \tag{19}$$

Let $R = e^{-1}\frac{N^{1/p}}{k^{1/2\alpha}}$. We have, with the inequality above, for N large enough,

$$\mathbb{P}_V^N\left(\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \le |\lambda_i^*|} > tN/2\right) \le \mathbb{P}_V^N\left(\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \le |\lambda_i^*| \le R} > tN/2\right) + \mathbb{P}_V^N\left(L_N(I_R^c) \ge \frac{k}{N}\right), \tag{20}$$

where L_N denotes the empirical measure of the λ_i's, and where $I_R = [-R, R]$. From (19), we deduce that,

$$\mathbb{P}_V^N\left(L(I_R^c) \ge \frac{k}{N}\right) \le \exp\left(-Ce^{-\alpha} k^{\frac{1}{2}} N^{1+\frac{\alpha}{p}}\right).$$

We are reduced to show that the event $\{\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \le |\lambda_i^*| \le R} > tN/2\}$ is exponentially negligible at the scale $N^{1+\alpha/p}$. To this end, we will slice up the set $\{\lambda \in \mathbb{R} : M_0 \le |\lambda| \le R\}$ into $\log \log N$ small intervals $\{\lambda \in \mathbb{R} : M_l \le |\lambda| \le M_{l+1}\}$ for which we will use the deviation bound (19). At each step, we choose the largest bound so that the event $\{\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_l \le |\lambda_i^*| \le M_{l+1}} > \frac{tN}{2}\}$ is exponentially negligible by (19). For any $n \ge 1$, we set

$$q_n = \left(1 - \frac{\alpha}{p}\right)\left(\frac{1}{p} + \frac{\alpha}{p^2} + \dots + \frac{\alpha^{n-1}}{p^n} + \frac{\alpha^{n-1}}{p^n(p-1)}\right),$$

and

$$M_n = \frac{N^{q_n}}{k^{1/2\alpha}}.$$

Observe that $q_n \xrightarrow[n \to +\infty]{} \frac{1}{p}$, and

$$\frac{1}{p} - q_n = O\left(\left(\frac{\alpha}{p}\right)^p\right).$$

Let $n = \lfloor c \log \log N \rfloor$ with c such that $q_n \geq \frac{1}{p} - \frac{1}{\log N}$. With this choice, we have

$$M_n \geq R.$$

Thus, slicing up the set $\{\lambda \in \mathbb{R} : M_0 \leq |\lambda| \leq R\}$, we get

$$\mathbb{P}_V^N\left(\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \leq |\lambda_i^*| \leq R} > \frac{tN}{2}\right) \leq \mathbb{P}_V^N\left(\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \leq |\lambda_i^*| \leq M_n} > \frac{tN}{2}\right)$$

$$\leq \mathbb{P}_V^N\left(\sum_{l=0}^{n-1} \sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_l \leq |\lambda_i^*| \leq M_{l+1}} > \frac{tN}{2}\right)$$

$$\leq \mathbb{P}_V^N\left(\sum_{l=0}^{n-1} M_{l+1}^p L_N\left(I_{M_l}^c\right) > \frac{t}{2}\right).$$

Finally, a union bound gives

$$\mathbb{P}_V^N\left(\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \leq |\lambda_i^*| \leq R_N} > \frac{tN}{2}\right) \leq \sum_{l=0}^{n-1} \mathbb{P}_V^N\left(L_N\left(I_{M_l}^c\right) > \frac{t}{2n M_{l+1}^p}\right).$$

Using (19), we get N large enough, and for all $0 \leq l \leq n$,

$$\mathbb{P}_V^N\left(L_N\left(I_{M_l}^c\right) > \frac{t}{2n M_{l+1}^p}\right) \leq \exp\left(-\frac{Ct N^2 M_l^\alpha}{2n M_{l+1}^p}\right) \leq \exp\left(-\frac{Ct N^{2+\alpha q_l - p q_{l+1}} k^{\frac{1}{2}\left(\frac{p}{\alpha}-1\right)}}{2c \log \log N}\right).$$

But

$$\alpha q_l - p q_{l+1} = \left(1 - \frac{\alpha}{p}\right)\left(\frac{\alpha}{p} + \frac{\alpha^2}{p^2} + \cdots + \frac{\alpha^l}{p^l} + \frac{\alpha^l}{p^l(p-1)}\right)$$

$$- \left(1 - \frac{\alpha}{p}\right)\left(1 + \frac{\alpha}{p} + \frac{\alpha^2}{p^2} + \cdots + \frac{\alpha^l}{p^l} + \frac{\alpha^l}{p^l(p-1)}\right)$$

$$= -\left(1 - \frac{\alpha}{p}\right).$$

Therefore,

$$\mathbb{P}_V^N\left(L\left(I_{M_l}^c\right) > \frac{t}{2n M_{l+1}^p}\right) \leq \exp\left(-\frac{Ct N^{1+\frac{\alpha}{p}}(\log N)^\kappa}{2c \log \log N}\right),$$

where $\kappa > 0$ as $p > \alpha$. We can conclude that,

$$\mathbb{P}_V^N\left(\sum_{i=k+1}^N |\lambda_i^*|^p \mathbb{1}_{M_0 \leq |\lambda_i^*| \leq R_N} > tN/2\right) \leq c \log \log N \exp\left(-\frac{Ct N^{1+\frac{\alpha}{p}}(\log N)^\kappa}{2c \log \log N}\right),$$

which ends the proof. $\qquad\square$

3.3. *Large deviations principle for the truncated moments*

Since we know from Proposition 3.5, that $(m_{p,N})_{N \in \mathbb{N}}$ is exponentially equivalent to

$$\left(\langle \sigma_\beta^V, x^p \rangle + T_{p,N}\right)_{N \in \mathbb{N}},$$

we only need to derive a large deviations principle for $(T_{p,N})_{N \in \mathbb{N}}$, in order to get the large deviations principle of $(m_{p,N})_{N \in \mathbb{N}}$ (see [11, Theorem 4.2.13]).

Proposition 3.9. *Under the assumption of Theorem* 1.1 *and the notation of Proposition* 3.5, *the sequence* $(T_{p,N})_{N \in \mathbb{N}}$ *follows a LDP under the law* $\mathbb{P}_V^N$, *with speed* $N^{1 + \frac{\alpha}{p}}$, *and good rate function* I_p. *If* p *is odd,* I_p *is defined by*

$$\forall x \in \mathbb{R}, \quad I_p(x) = b|x|^{\alpha/p},$$

and if p *is even*,

$$\forall x \in \mathbb{R}, \quad I_p(x) = \begin{cases} bx^{\alpha/p} & \text{if } x \geq 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof. Once again, to ease the notation, we set in the following $k = \log N$.

   Exponential tightness. Let

$$\forall \lambda \in \mathbb{R}^N, \quad g(\lambda) = \left(\sum_{i=1}^k |\lambda_i^*|^p \right)^{1/p}.$$

For $\lambda \in \mathbb{R}^N$, we set $l = \text{Card}\{i \in \{1, \ldots, k\} : \lambda_i^* > 0\}$. We can write

$$g(\lambda) = \left(\sum_{i=1}^{k-l} |\overline{\lambda}_i|^p + \sum_{i=N-l+1}^{N} \overline{\lambda}_i^p \right)^{1/p},$$

where $\overline{\lambda}_1, \ldots, \overline{\lambda}_N$ is the rearrangement of the λ_i's in ascending order. When l is fixed, as $p \geq \alpha$, we see that g is 1-Lipschitz with the same argument as in the proof of Proposition 3.7. Using a union bound, we get by Proposition 3.7, for any $t > 0$,

$$\mathbb{P}_V^N\left(g - \mathbb{E}_V^N g > t N^{\frac{1}{p}} \right) \leq k \exp\left(-\frac{bt^\alpha N^{1+\frac{\alpha}{p}}}{2^{\alpha-1}\alpha(\alpha-1)^{\alpha-1}} \right).$$

Besides, by Jensen's inequality

$$\mathbb{E}_V^N\left(\frac{1}{N} \sum_{i=1}^k |\lambda_i^*|^p \right)^{1/p} \leq \left(\mathbb{E}_V^N \frac{1}{N} \sum_{i=1}^k |\lambda_i^*|^p \right)^{1/p} \leq \left(\frac{k}{N} \mathbb{E}_V^N |\lambda_1^*|^p \right)^{1/p}.$$

From (16), we deduce

$$\mathbb{E}_V^N\left(\frac{k}{N} \sum_{i=1}^k |\lambda_i^*|^p \right)^{1/p} \xrightarrow[N \to +\infty]{} 0.$$

From the above concentration inequality, we see that $(T_{p,N})_{N \in \mathbb{N}}$ is exponentially tight.

   Upper bound. Observe that we only have to show that for any $x > 0$,

$$\limsup_{N \to +\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}_V^N(T_{p,N} \geq x) \leq -I_p(x). \tag{21}$$

In the case where p is even, it is clear that (21), is sufficient. In the case p is odd, observe that $\tilde{V}(x) = V(-x)$ satisfies the assumptions of Theorem 1.1. Note also that for any $x > 0$,

$$\mathbb{P}_V^N(T_{p,N} \leq -x) = \mathbb{P}_{\tilde{V}}^N(T_{p,N} \geq x).$$

Therefore, if (21) is proven, and if p odd, then we have also for any $x > 0$,

$$\limsup_{N \to +\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}_V^N(T_{p,N} \leq -x) \leq -I_p(-x).$$

We now prove (21). Since $(\frac{1}{N} \sum_{i=1}^{k} |\lambda_i^*|^p)_{N \in \mathbb{N}}$ is exponentially tight, we only need to show that for any $M > x > 0$, we have

$$\limsup_{N \to +\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}_V^N\left(T_{p,N} \geq x, |\lambda_1^*|^p \leq MN\right) \leq -I_p(x).$$

Let $M > x > 0$. Since the event $\{T_{p,N} \geq x, |\lambda_1^*|^p \leq MN\}$ is invariant under permutation of the λ_i's, we have

$$\mathbb{P}_V^N\left(T_{p,N} \geq x, |\lambda_1^*|^p \leq MN\right) = \frac{N!}{Z_V^N} \int_{\substack{\sum_{i=1}^{k} \lambda_i^p \geq Nx \\ |\lambda_N| \leq \cdots \leq |\lambda_1| \leq (MN)^{1/p}}} e^{-N \sum_{i=1}^{N} V(\lambda_i)} \prod_{i<j} |\lambda_i - \lambda_j|^\beta \prod_{i=1}^{N} d\lambda_i.$$

Bounding the interaction term involving the k largest in absolute value λ_i's, we get

$$\mathbb{P}_V^N\left(T_{p,N} \geq x, |\lambda_1^*|^p \leq MN\right)$$

$$\leq \frac{N!}{(N-k)!} \frac{Z_{NV}^{N-k}}{Z_V^N} \left(2(NM)^{1/p}\right)^{\beta Nk} \int_{\substack{\sum_{i=1}^{k} \lambda_i^p \geq Nx \\ |\lambda_k| \leq \cdots \leq |\lambda_1|}} e^{-N \sum_{i=1}^{k} V(\lambda_i)} \prod_{i=1}^{k} d\lambda_i$$

$$\leq \binom{N}{k} \frac{Z_{NV}^{N-k}}{Z_V^N} \left(2(NM)^{1/p}\right)^{\beta Nk} \int_{\sum_{i=1}^{k} \lambda_i^p \geq Nx} e^{-N \sum_{i=1}^{k} V(\lambda_i)} \prod_{i=1}^{k} d\lambda_i$$

$$= \binom{N}{k} \frac{Z_{NV}^{N-k}}{Z_V^N} \left(2(NM)^{1/p}\right)^{\beta Nk} \left(\int e^{-NV(\lambda)} d\lambda\right)^k \mathbb{P}\left(\frac{1}{N} \sum_{i=1}^{k} X_i^p \geq x\right),$$

where $X_1, \ldots, X_k$ are independent and identically distributed random variables with law $d\mu_V = e^{-NV(x)} \frac{dx}{Z_N}$, where $Z_N = \int e^{-NV(x)} dx$. As

$$\int e^{-NV(x)} dx = e^{O(N)} \quad \text{and} \quad \log \frac{Z_{NV}^{N-k}}{Z_V^N} = O(N \log N),$$

from Lemma 3.3 (recall that $k = \log N$), it only remains to show that

$$\limsup_{N \to +\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}\left(\frac{1}{N} \sum_{i=1}^{\log N} X_i^p \geq x\right) \leq -I_p(x).$$

This is the object of the following lemma.

Lemma 3.10. *Let $(X_j)_{j \geq 1}$ be a sequence of independent and identically distributed random variables with law $d\mu_V = e^{-NV(x)} \frac{dx}{Z_N}$, where $Z_N = \int e^{-NV(x)} dx$, with V as in (4). Let $p \in \mathbb{N}$, $p > \alpha$.*

For any $x > 0$,

$$\limsup_{N \to +\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}\left(\frac{1}{N} \sum_{i=1}^{\log N} X_i^p \geq x\right) \leq -I_p(x),$$

with I_p as in Proposition 3.9.

Proof. Set $k = \log N$. Let $x > 0$. Set $Y_i = N^{1/\alpha} X_i$ for all $i \in \{1, \ldots, \log N\}$. We have

$$\mathbb{P}\left(\frac{1}{N} \sum_{i=1}^{k} X_i^p \geq x\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} Y_i^p \geq x N^{1+\frac{p}{\alpha}}\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} |Y_i|^p \geq x N^{1+\frac{p}{\alpha}}\right).$$

Let $0 < t < 1$. As $\alpha \leq p$, we have $\alpha t / p < 1$. Using the fact that for any $s \in (0, 1)$, $x, y \in \mathbb{R}^+$, $(x + y)^s \leq x^s + y^s$,

$$\mathbb{P}\left(\frac{1}{N} \sum_{i=1}^{k} X_i^p \geq x\right) \leq \mathbb{P}\left(\left(\sum_{i=1}^{k} |Y_i|^p\right)^{\frac{\alpha t}{p}} \geq x^{\frac{\alpha t}{p}} N^{t(1+\frac{\alpha}{p})}\right)$$

$$\leq \mathbb{P}\left(\sum_{i=1}^{k} |Y_i|^{\alpha t} \geq x^{\frac{\alpha t}{p}} N^{t(1+\frac{\alpha}{p})}\right).$$

By Chernoff's inequality we get,

$$\mathbb{P}\left(\frac{1}{N} \sum_{i=1}^{k} X_i^p \geq x\right) \leq e^{-bx^{\frac{\alpha t}{p}} N^{t(1+\frac{\alpha}{p})}} \left(\mathbb{E}\left(e^{b|Y_1|^{\alpha t}}\right)\right)^k. \tag{22}$$

As for any $x \in \mathbb{R}$, $V(x) = b|x|^\alpha + w(x)$,

$$\mathbb{E}\left(e^{b|Y_1|^{\alpha t}}\right) = \frac{1}{Z'_N} \int e^{-b(|x|^\alpha - |x|^{\alpha t}) - Nw(\frac{x}{N^{1/\alpha}})} \, dx,$$

with

$$Z'_N = \int e^{-NV(\frac{x}{N^{1/\alpha}})} \, dx.$$

On one hand,

$$\int e^{-b(|x|^\alpha - |x|^{\alpha t}) - Nw(\frac{x}{N^{1/\alpha}})} \, dx \leq 2e^{N \inf w} \int_0^{+\infty} e^{-b(x^\alpha - x^{\alpha t})} \, dx.$$

Note that as w is convex, $\inf w > -\infty$. On the other hand, $Z'_N = e^{O(N)}$. Therefore,

$$\mathbb{E}\left(e^{b|Y_1|^{\alpha t}}\right) \leq e^{o(\frac{N^{1+\alpha/p}}{k})} \int_0^{+\infty} e^{-b(x^\alpha - x^{\alpha t})} \, dx.$$

As $x \mapsto x^{\alpha-1} - t x^{\alpha t - 1}$ is non-decreasing on $[1, +\infty)$, we have,

$$\int_0^{+\infty} e^{-b(x^\alpha - x^{\alpha t})} \, dx \leq e^b + \frac{1}{\alpha(1-t)} \int_1^{+\infty} \left(\alpha x^{\alpha-1} - \alpha t x^{\alpha t - 1}\right) e^{-b(x^\alpha - x^{\alpha t})} \, dx$$

$$= e^b + \frac{1}{b\alpha(1-t)}.$$

Take $t = t_N = 1 - 1/(\log N)^2$. Then,

$$\mathbb{E}\big(e^{b|Y_1|^\alpha t_N}\big) = e^{o\left(\frac{N^{1+\alpha/p}}{k}\right)}.$$

Together with the bound (22), we get

$$\frac{1}{N^{1+\alpha/p}} \log \mathbb{P}\left(\frac{1}{N} \sum_{i=1}^k X_i^p \geq x\right) \leq -b x^{\frac{\alpha t_N}{p}} N^{-(1-t_N)(1+\frac{\alpha}{p})} + o(1). \tag{23}$$

Taking the limsup as N goes to $+\infty$ we get the claim. $\qquad\square$

Lower bound. Let $x \in \mathbb{R}$. We want to show that

$$\liminf_{\delta \to 0} \liminf_{N \to +\infty} \frac{1}{N^{1+\alpha/p}} \log \mathbb{P}_V^N\big(T_{p,N} \in (x - \delta, x + \delta)\big) \geq -I_p(x). \tag{24}$$

As $T_{p,N}$ converges to 0 in almost surely, it is enough to prove this bound for $x \neq 0$. With the same argument as for the upper bound, it suffices actually prove to the bound above only for $x > 0$.

Let $x > 0$ and $\delta > 0$. We have for N large enough,

$$\mathbb{P}_V^N\big(T_{p,N} \in (x - \delta, x + \delta)\big) \geq \mathbb{P}_{V,\beta}^N\left(\frac{1}{N} \lambda_1^{*p} \in (x - \delta/2, x + \delta/2), \forall i > 1, \big|\lambda_i^*\big| \leq M\right),$$

with $M > 0$. By continuity, there is some $\varepsilon > 0$ such that

$$\mathbb{P}_V^N\big(T_{p,N} \in (x - \delta, x + \delta)\big) \geq \mathbb{P}_V^N\left(\frac{1}{N^{1/p}} \lambda_1^* \in \big(x^{1/p} - \varepsilon, x^{1/p} + \varepsilon\big), \forall i > 1, \big|\lambda_i^*\big| \leq M\right).$$

We have

$$\mathbb{P}_V^N\big(T_{p,N} \in (x - \delta, x + \delta)\big)$$
$$\geq N! \frac{Z_{\frac{NV}{N-1}}^{N-1}}{Z_V^N} \int_{\left|\frac{\lambda_1}{N^{1/p}} - x^{1/p}\right| < \varepsilon} d\lambda_1 e^{-NV(\lambda_1)} \mathbb{E}_{\frac{NV}{N-1}}^{N-1}\big(\mathbb{1}_{L_{N-1} \in \mathcal{E}_M} e^{\beta(N-1)\langle \log(\lambda_1 - \cdot), L_{N-1}\rangle}\big),$$

where $L_{N-1} = \frac{1}{N-1} \sum_{i=2}^N \delta_{\lambda_i}$, and $\mathcal{E}_M = \{\mu \in \mathcal{P}(\mathbb{R}) : \text{supp}(\mu) \subset [-M, M]\}$, with $\mathcal{P}(\mathbb{R})$ the set of probability measures on $\mathbb{R}$. Thus,

$$\mathbb{P}_V^N\big(T_{p,N} \in (x - \delta, x + \delta)\big)$$
$$\geq N! \frac{Z_{\frac{NV}{N-1}}^{N-1}}{Z_V^N} \int_{\left|\frac{\lambda}{N^{1/p}} - x^{\frac{1}{p}}\right| < \varepsilon} e^{-NV(\lambda)} d\lambda e^{\beta(N-1)\log(N^{\frac{1}{p}} x^{\frac{1}{p}} - M - \varepsilon)} \mathbb{P}_{\frac{NV}{N-1}}^{N-1}(L_{N-1} \in \mathcal{E}_M).$$

As $w(y) = o_{\pm\infty}(|y|^\alpha)$, we have

$$\int_{\left|\frac{\lambda}{N^{1/p}} - x^{\frac{1}{p}}\right| < \varepsilon} e^{-NV(\lambda)} d\lambda \geq \int_{\left|\frac{\lambda}{N^{1/p}} - x^{\frac{1}{p}}\right| < \varepsilon} e^{-(b+o(1))N\lambda^\alpha} dy = e^{-(b-o(\varepsilon))N^{1+\frac{\alpha}{p}} x^{\frac{\alpha}{p}}} e^{o(N^{1+\frac{\alpha}{p}})}.$$

Thus,

$$\mathbb{P}_V^N\big(T_{p,N} \in (x - \delta, x + \delta)\big) \geq \frac{Z_{\frac{NV}{N-1}}^{N-1}}{Z_V^N} \mathbb{P}_{\frac{NV}{N-1}}^{N-1}(L_{N-1} \in \mathcal{E}_M) e^{-(b-o(\varepsilon))N^{1+\frac{\alpha}{p}} x^{\frac{\alpha}{p}}} e^{o(N^{1+\frac{\alpha}{p}})}.$$

But from Lemma 3.3 we know that $\log \frac{Z_{NV}^{N-1}}{Z_V^N} = O(N)$. Besides, by Proposition 3.1 (with $k = 1$), we have for M large enough,

$$\mathbb{P}_{\frac{NV}{N-1}}^{N-1}(L_{N-1} \in \mathcal{E}_M) \underset{N \to +\infty}{\longrightarrow} 1.$$

This concludes the proof of the lower bound (24). $\qquad\square$

4. The case of Wigner matrices without Gaussian tails

We will give in this section a proof of Theorem 1.7. The strategy we will follow is in the same spirit as the ones developed in [7,17] and [2] for this model of Wigner matrices without Gaussian tails. We start by a heuristic argument to give a idea of the nature of the deviations of the moments, and of the speed of the deviations.

4.1. *Heuristics*

We show here how one can get the lower bound of the LDP without much effort. A key argument of the proof consists in finding how the trace of a given power behaves under additive perturbations by finite rank matrices. It turns out that if we add to a given Hermitian matrix a low rank Hermitian matrix with not too large operator norm, then the map $A \mapsto \operatorname{tr}_N A^p$ is almost linear. More precisely, we have the following lemma, whose proof is postpone at Section 4.5.3.

Lemma 4.1. *Let $p \geq 2$. Let A and C be two Hermitian matrices of size N. Assume that C is of rank at most r. We have*

$$\left|\operatorname{tr}(A + C)^p - \operatorname{tr} A^p - \operatorname{tr} C^p\right| \leq 2^p r \max_{1 \leq k \leq p-1} \|A\|^k \|C\|^{p-k},$$

where $\|\cdot\|$ denotes the operator norm.

To make the argument clearer, let us assume X has entries $X_{1,1}$ and $X_{1,2}$ distributed according to symmetric exponential laws with parameters b and a respectively. Let $s \neq 0$. Let A be a Hermitian matrix of finite size, say $n > 0$, such that $\operatorname{tr} A^p = s$. With an abuse of notation, we will continue to denote A the Hermitian matrix of size N, $A \oplus 0_{N-n}$, where 0_{N-n} is the null matrix of size $N - n$. Let $X_N^{(n)}$ denote the matrix obtained from X_N by setting all the elements which are not in the principal submatrix of order n to 0, and let $X_N^{(\backslash n)} = X_N - X_N^{(n)}$. We can informally write,

$$\mathbb{P}\big(\operatorname{tr}_N X_N^p \simeq \langle \sigma_{\mathrm{sc}}, x^p \rangle + s\big) \gtrsim \mathbb{P}\big(\operatorname{tr}_N \big(X_N^{(\backslash n)} + N^{1/p} A\big)^p \simeq \langle \sigma_{\mathrm{sc}}, x^p \rangle + s, X_N^{(n)} \simeq N^{1/p} A\big)$$

$$\gtrsim \mathbb{P}\big(\operatorname{tr}_N \big(X_N^{(\backslash n)} + N^{1/p} A\big)^p \simeq \langle \sigma_{\mathrm{sc}}, x^p \rangle + s, \big\|X_N^{(\backslash n)}\big\| \leq c\big)\mathbb{P}\big(X_N^{(n)} \simeq N^{1/p} A\big),$$

with some $c > 2$. As $\|X_N^{(\backslash n)} - X_N\|_2 \to 0$ in probability, and

$$\|X_N\| \underset{N \to +\infty}{\longrightarrow} 2,$$

in probability by [3, Theorem 5.1] (or [1, Theorem 2.1.22, Exercise 2.1.27]), we have

$$\mathbb{P}\big(\big\|X_N^{(\backslash n)}\big\| \leq c\big) \underset{N \to +\infty}{\longrightarrow} 1.$$

By Lemma 4.1, we deduce, as $\operatorname{tr} A^p = s$,

$$\mathbb{P}\big(\operatorname{tr}_N X_N^p \simeq \langle \sigma_{\mathrm{sc}}, x^p \rangle + s\big) \gtrsim \mathbb{P}\big(\operatorname{tr}_N \big(X_N^{(\backslash n)}\big)^p \simeq \langle \sigma_{\mathrm{sc}}, x^p \rangle + s, \big\|X_N^{(\backslash n)}\big\| \leq c\big)\mathbb{P}\big(X_N^{(n)} \simeq N^{1/p} A\big).$$

Since $X_{1,1}$ and $X_{1,2}$ follow the symmetric exponential laws with parameters b and a respectively, we have

$$\mathbb{P}\big(X_N^{(r)} \simeq N^{1/p} A\big) \simeq \exp\big(-N^{\frac{1}{2}+\frac{1}{p}} I(A)\big),$$

where $I(A) = b \sum_i |A_{i,i}| + a \sum_{i<j} |A_{i,j}|$. But $(\mathrm{tr}_N(X_N^{(\backslash r)})^p)_{N\in\mathbb{N}}$ converges to $\langle \sigma_{\mathrm{sc}}, x^p \rangle$ in probability, by Wigner's theorem (see [1, Lemmas 2.1.6, 2.1.7]). Therefore,

$$\mathbb{P}\big(\mathrm{tr}_N X_N^p \simeq \langle \sigma_{\mathrm{sc}}, x^p \rangle + s\big) \gtrsim \exp\big(-N^{\frac{1}{2}+\frac{1}{p}} I(A)\big).$$

As this lower bound is true for any Hermitian matrix A such that $\mathrm{tr}\, A^p = s$, we deduce that the lower bound holds with rate function

$$\forall y \in \mathbb{R}, \quad J_p(y) = \inf\bigg\{ I(A) : \mathrm{tr}\, A^p = \langle \sigma_{\mathrm{sc}}, x^p \rangle + y, A \in \bigcup_{n\in\mathbb{N}} \mathcal{H}_n^{(\beta)} \bigg\},$$

which is precisely the rate function we will give in this case of entries with symmetric exponential laws (see Remark 1.8(c)).

4.2. *Outline of proof*

As suggested by the heuristic argument above, the deviations of $\mathrm{tr}_N X_N^p$ are due to a small proportion of the entries of X_N making deviations of order $N^{1/p}$. Accordingly, we decompose X_N in the following way

$$X_N = A + B^\varepsilon + C^\varepsilon + D^\varepsilon, \tag{25}$$

with

$$A_{i,j} = \frac{X_{i,j}}{\sqrt{N}} \mathbb{1}_{|X_{i,j}|\le(\log N)^d}, \qquad B_{i,j}^\varepsilon = \frac{X_{i,j}}{\sqrt{N}} \mathbb{1}_{(\log N)^d < |X_{i,j}| < \varepsilon N^{\frac{1}{2}+\frac{1}{p}}},$$

$$C_{i,j}^\varepsilon = \frac{X_{i,j}}{\sqrt{N}} \mathbb{1}_{\varepsilon N^{\frac{1}{2}+\frac{1}{p}} \le |X_{i,j}| \le \varepsilon^{-1} N^{\frac{1}{2}+\frac{1}{p}}}, \qquad D_{i,j}^\varepsilon = \frac{X_{i,j}}{\sqrt{N}} \mathbb{1}_{\varepsilon^{-1} N^{\frac{1}{2}+\frac{1}{p}} < |X_{i,j}|},$$

where where d is taken such that $\alpha d > 1$.

In a first phase, we will show that one can neglect in the deviations of $\mathrm{tr}_N X_N^p$ the contributions of the intermediate entries, that is B^ε, and the large entries, that is D^ε, so that $(\mathrm{tr}_N(A + C^\varepsilon)^p)_{N\in\mathbb{N},\varepsilon>0}$ are exponentially good approximations for $(\mathrm{tr}_N X_N^p)_{N\in\mathbb{N}}$.

Then, due to concentration inequalities, we show that the conditional expectation given C^ε, $\mathbb{E}_{C^\varepsilon} \mathrm{tr}_N(H + C^\varepsilon)^p$, where H is a copy of A independent of X, are exponentially good approximations of $(\mathrm{tr}_N X_N^p)_{N\in\mathbb{N}}$. From the choice of the decomposition (25), we deduce that C^ε has only a finite number of non-zero entries at the exponential scale $N^{1+\alpha/p}$. Thus, Lemma 4.1 and Wigner's Theorem allow us to conclude that $(\mathbb{E}_{C^\varepsilon} \mathrm{tr}_N(H + C^\varepsilon)^p)_{N\in\mathbb{N}}$ is exponentially equivalent to $(\langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}_N(C^\varepsilon)^p)_{N\in\mathbb{N}}$. It only remains to show a large deviations principle C^ε, and conclude by contraction principle, with an argument similar as in [2]. The use of the contraction principle is made possible by the fact that C^ε has a finite number of non-zero entries except on an event of exponentially small probability.

4.3. *Concentration inequalities*

In this section, we revisit a concentration inequality from [25] for the trace of powers of sum of a Hermitian matrix with bounded entries with a deterministic Hermitian matrix. This inequality will be crucial to get the exponential tightness and an exponential approximation of $(\mathrm{tr}_N X_N^p)_{N\in\mathbb{N}}$.

Unfortunately, we cannot directly use the concentration inequality of [25, Proposition 4], because of the assumption made on the expectation of the entries. To make the strategy sketched in Section 4.2 work, we need to prove a

concentration inequality for

$$\text{tr}_N \left(\frac{H}{\sqrt{N}} + \frac{C}{\sqrt{N}} \right)^p,$$

where H is a centered matrix with bounded entries, and where C is a deterministic matrix whose entries are of order $N^{1/p+1/2}$. But then,

$$\text{tr}\left(\frac{C}{\sqrt{N}} \right)^{2(p-1)} \leq r^{2(p-1)} N^{\frac{2(p-1)}{p}}, \tag{26}$$

where r is the number of non-zero entries of C, which is a bound too loose to use the concentration inequality of [25, Proposition 4].

However, since we are considering normalized traces, we are looking at deviations of order N of the traces, whereas in [25] the deviations considered were of order 1. Thus, one can expect that there is some room left in the approach of Meckes and Szarek, to get a concentration inequality for $\text{tr}_N(H + C)^p$, with the bound (26).

Proposition 4.2. *Let $p \in \mathbb{N}$, $p \geq 3$. Let H be a centered random Hermitian matrix such that $(H_{i,j})_{i \leq j}$ are independent and bounded by some $\kappa \geq 1$, and let C be a deterministic Hermitian matrix such that $\text{tr}(\frac{C}{\sqrt{N}})^{2(p-1)} \leq m N^{2-\frac{2}{p}}$, where $m \geq 1$. Write $X = H + C$. There are some universal constants $c, c' > 0$, such that for all $t \geq c'(pm^{p-1})^p N^{-\frac{1}{2}(1+\frac{2}{p})}$,*

$$\mathbb{P}\left(\left| \text{tr}_N X^p - \mathbb{E} \text{tr}_N X^p \right| > t N^{p/2} \right) \leq 8 \exp\left(-\frac{N^{1+\frac{2}{p}}}{c\kappa^2} \min\left\{ \left(\frac{t}{p} \right)^{2/p}, \frac{t^2}{p^2 m^{2(d-1)}} \right\} \right).$$

Moreover,

$$\mathbb{P}\left(\left| \text{tr}_N |X|^p - \mathbb{E} \text{tr}_N |X|^p \right| > t N^{p/2} \right) \leq 8 \exp\left(-\frac{N^{1+\frac{2}{p}}}{c\kappa^2} \min\left\{ \left(\frac{t}{p} \right)^{2/p}, \frac{t^2}{p^2 m^{2(d-1)}} \right\} \right).$$

Proof. We follow the same approach as in [25, Proposition 4], with some slight variations at times, but considering deviations of order $N^{1+p/2}$ of the trace of $(H + C)^p$. We will prove only the first inequality, the proof of the second inequality being exactly the same.

Without loss of generality, we can assume $\kappa = 1$. For $\beta \in \{1, 2\}$, we denote by $\mathcal{H}_N^{(\beta)}$ the set of symmetric matrices of size N, when $\beta = 1$, and Hermitian matrices when $\beta = 2$. Note that as H has entries bounded by 1, we know by [22, Corollary 4.10], that for any convex and 1-Lipschitz function $f : \mathcal{H}_N^{(\beta)} \to \mathbb{R}$ with respect to the Hilbert–Schmidt norm, and all $t > 0$,

$$\mathbb{P}\left(\left| f(X) - \mathbb{M} f(X) \right| > t \right) \leq 4 e^{-\frac{t^2}{4}},$$

where $\mathbb{M} f(X)$ denotes the median of $f(X)$. Let $a > 0$. Define

$$K_a = \left\{ Y \in \mathcal{H}_n^{(\beta)} : \|Y\|_{2(p-1)} \leq a \right\},$$

where $\|Y\|_q = (\text{tr} |Y|^q)^{1/q}$ for any matrix Y and $q > 0$. Note that we can write

$$F = F^+ - F^-,$$

with $F^+(Y) = \text{tr} Y_+^p$, and $F^-(Y) = \text{tr} Y_-^p$ for any $Y \in \mathcal{H}_N^{(\beta)}$, where for every $x \in \mathbb{R}$, x_+ and x_- denote the positive and negative parts of x. The functions F^+ and F^- are convex and pa^{p-1}-Lipschitz on K_a. Let F_a^+, F_a^- denote the convex extensions of $F_{|K_a}^+$ and $F_{|K_a}^-$ to $\mathcal{H}_N^{(\beta)}$, which are pa^{p-1}-Lipschitz, as explained in [25, Lemma 5]. Then, for all $t > 0$, we have

$$\mathbb{P}\left(\left| F_a^\sigma(X) - \mathbb{M} F_a^\sigma(X) \right| > t N^{1+p/2} \right) \leq 4 \exp\left(-\frac{t^2 N^{p+2}}{4 p^2 a^{2(d-1)}} \right),$$

with $\sigma \in \{+, -\}$.

Besides $Y \mapsto \|Y\|_{2(d-1)}$ is convex and 1-Lipschitz with respect to the Hilbert–Schmidt norm. From [24, Theorem 8.6], we deduce that for any $t > 0$,

$$\mathbb{P}\big(\|X\|_{2(p-1)} - \mathbb{E}\|X\|_{2(p-1)} > t\big) \leq e^{-\frac{t^2}{32}}.$$

But,

$$\mathbb{E}\|X\|_{2(p-1)} \leq \mathbb{E}\|H\|_{2(p-1)} + \|C\|_{2(p-1)} \leq N^{\frac{1}{2(p-1)}}\mathbb{E}\|H\| + mN^{\frac{1}{2}+\frac{1}{p}},$$

where $\|\cdot\|$ denotes the operator norm, and where we used the fact that $m \geq 1$. But we know, by [1, Exercice 2.1.29, 30], that there is some universal constant $c_1 \geq 1$, such that

$$\mathbb{E}\|H\| \leq c_1\sqrt{N}.$$

Thus, $\mathbb{E}\|X\|_{2(p-1)} \leq 2mc_1 N^{\frac{1}{2}+\frac{1}{p}}$.

Let now $b > 0$, and $a = bN^{\frac{1}{2}+\frac{1}{p}}$. We have, for $b \geq 4mc_1$,

$$\mathbb{P}\big(\|X\|_{2(p-1)} \geq a\big) \leq \mathbb{P}\bigg(\|X\|_{2(p-1)} - \mathbb{E}\|X\|_{2(p-1)} \geq \frac{a}{2}\bigg) \leq \exp\bigg(-\frac{b^2 N^{1+\frac{2}{p}}}{128}\bigg).$$

Besides, with this choice of a, we have for all $t > 0$, and all $\sigma \in \{+, -\}$,

$$\mathbb{P}\bigg(\big|F_a^\sigma(X) - \mathbb{M}F_a^\sigma(X)\big| > \frac{t}{2}N^{1+p/2}\bigg) \leq 4\exp\bigg(-\frac{t^2 N^{1+2/p}}{16p^2 b^{2(p-1)}}\bigg).$$

Thus,

$$\mathbb{P}\bigg(\big|F^\sigma(X) - \mathbb{M}F_a^\sigma(X)\big| > \frac{t}{2}N^{1+p/2}\bigg) \leq \mathbb{P}\bigg(\big|F_a^\sigma(X) - \mathbb{M}F_a^\sigma(X)\big| > \frac{t}{2}N^{1+p/2}\bigg) + \mathbb{P}\big(\|X\|_{2(p-1)} \geq a\big)$$

$$\leq 4\exp\bigg(-\frac{N^{1+2/p}}{128}\min\bigg\{b^2, \frac{t^2}{p^2 b^{2(p-1)}}\bigg\}\bigg).$$

As a consequence, for $b = 4mc_1$, we can find a numerical constant $c_2 \geq 1$, such that for $t = c_2 pN^{-\frac{1}{2}(1+\frac{2}{p})}$, we have

$$\mathbb{P}\big(F^\sigma(X) - \mathbb{M}F_a^\sigma(X) > tN^{1+p/2}\big) < \frac{1}{2}.$$

We deduce that

$$\mathbb{M}F^\sigma(X) \leq \mathbb{M}F_a^\sigma(X) + c_2 pN^{\frac{1}{2}+\frac{p}{2}-\frac{1}{p}}.$$

As F_a^σ is non-decreasing with a, and $F_a^\sigma \leq F^\sigma$ for any $a > 0$, we have for all $b \geq 4mc_1$,

$$\mathbb{M}F^\sigma(X) - c_2 pN^{\frac{1}{2}+\frac{p}{2}-\frac{1}{p}} \leq \mathbb{M}F_a^\sigma(X) \leq \mathbb{M}F^\sigma(X).$$

Thus, for $t \geq 2c_2 pN^{-\frac{1}{2}(1+\frac{2}{p})}$, and any $b \geq 4mc_1$, we deduce that

$$\mathbb{P}\big(\big|F^\sigma(X) - \mathbb{M}F^\sigma(X)\big| > tN^{1+p/2}\big) \leq \mathbb{P}\bigg(\big|F^\sigma(X) - \mathbb{M}F_a^\sigma(X)\big| > \frac{t}{2}N^{1+p/2}\bigg)$$

$$\leq 4\exp\bigg(-\frac{N^{1+2/p}}{128}\min\bigg\{b^2, \frac{t^2}{p^2 b^{2(p-1)}}\bigg\}\bigg).$$

But one can check that,

$$\max_{b \geq 4mc_1} \min\left\{ b^2, \frac{t^2}{p^2 b^{2(p-1)}} \right\} = \min\left\{ \left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{p^2 (mc_1)^{2(p-1)}} \right\}.$$

Optimizing in b in the previous inequality, and setting $c_3 = 128c_1^{2(p-1)}$, we get

$$\mathbb{P}\left(\left|F^\sigma(X) - \mathbb{M}F^\sigma(X)\right| > tN^{1+p/2}\right) \leq 4\exp\left(-\frac{N^{1+\frac{2}{p}}}{c_3} \min\left\{ \left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{p^2 m^{2(p-1)}} \right\}\right).$$

To get the same inequality but with $\mathbb{E}F^\sigma(X)$ instead of $\mathbb{M}F^\sigma(X)$, we integrate by parts the inequality above, and we find that there is some constant $c_4 > 0$, such that

$$\left|\mathbb{E}F^\sigma(X) - \mathbb{M}F^\sigma(X)\right| \leq c_4 m^{p-1} p N^{-\frac{1}{2}(1+\frac{2}{p})}.$$

At the price of taking c_4 larger, we can assume that $c_4 \geq c_2$. Then, for every $t \geq 2c_4 m^{p-1} p N^{-\frac{1}{2}(1+\frac{2}{p})}$,

$$\mathbb{P}\left(\left|F^\sigma(X) - \mathbb{E}F^\sigma(X)\right| > tN^{1+p/2}\right) \leq \mathbb{P}\left(\left|F^\sigma(X) - \mathbb{M}F^\sigma(X)\right| > \frac{t}{2} N^{1+p/2}\right)$$

$$\leq 4\exp\left(-\frac{N^{1+\frac{2}{p}}}{4c_3} \min\left\{ \left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{p^2 m^{2(d-1)}} \right\}\right).$$

As $F = F^+ - F^-$, we have for any $t \geq 2c_4 m^{p-1} p N^{-\frac{1}{2}(1+\frac{2}{p})}$,

$$\mathbb{P}\left(\left|F(X) - \mathbb{E}F(X)\right| > tN^{1+p/2}\right) \leq 8\exp\left(-\frac{N^{1+\frac{2}{p}}}{16c_3} \min\left\{ \left(\frac{t}{p}\right)^{2/p}, \frac{t^2}{m^{2(d-1)}} \right\}\right).$$

Setting $c = 16c_3$, and $c' = 2c_4$, we get the claim. □

4.4. *Exponential tightness*

Throughout the rest of this section, we fix a constant $\gamma > 0$, such that for t large enough,

$$\mathbb{P}\left(|X_{1,1}| > t\right) \vee \mathbb{P}\left(|X_{1,2}| > t\right) \leq e^{-\gamma t^\alpha}. \tag{27}$$

In this section, we will show that the sequence $(\text{tr}_N X_N^p)_{N \in \mathbb{N}}$ is exponentially tight, namely, we have the following proposition.

Proposition 4.3 (Exponential tightness).

$$\lim_{t \to +\infty} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\text{tr}_N |X_N|^p > t\right) = -\infty.$$

Proof of Proposition 4.3. Using the triangular inequality for the p-Schatten norm, we get for any $t > 0$,

$$\mathbb{P}\left(\text{tr}_N |X_N|^p > (4t)^p\right) \leq \mathbb{P}\left(\text{tr}_N |A|^p > t^p\right) + \mathbb{P}\left(\text{tr}_N |B^\varepsilon|^p > t^p\right) + \mathbb{P}\left(\text{tr}_N |C^\varepsilon|^p > t^p\right) + \mathbb{P}\left(\text{tr}_N |D^\varepsilon|^p > t^p\right). \tag{28}$$

This shows that it suffices to estimate at the exponential scale, the probability of each event $\{\text{tr}_N |A|^p > t^p\}$, $\{\text{tr}_N |B^\varepsilon|^p > t^p\}$, $\{\text{tr}_N |C^\varepsilon|^p > t^p\}$, and finally $\{\text{tr}_N |D^\varepsilon|^p > t^p\}$. As a consequence of the concentration inequality of Proposition 4.2, we have the following lemma.

Lemma 4.4.

$$\lim_{t\to+\infty}\limsup_{N\to+\infty}\frac{1}{N^{1+2/p}}\log\mathbb{P}\big(\mathrm{tr}_N\,|A|^p > t\big) = -\infty,$$

where A is as in (25).

Proof. Note that as $p \geq 2$,

$$\mathrm{tr}(\mathbb{E}A)^{2(p-1)} \leq \big(\mathrm{tr}(\mathbb{E}A)^2\big)^{p-1}.$$

Since the entries of X are centered, we get

$$\mathrm{tr}(\mathbb{E}A)^2 = \frac{1}{N}\sum_{1\leq i,j\leq N}\mathbb{E}|X_{i,j}|^2\mathbb{1}_{|X_{i,j}|>(\log N)^d}.$$

Integrating by parts, we have

$$\mathrm{tr}(\mathbb{E}A)^2 = O\big(N^2 e^{-\frac{\gamma}{2}(\log N)^{\alpha d}}\big),$$

where γ is as in (27). As $\alpha d > 1$,

$$\mathrm{tr}(\mathbb{E}A)^{2(p-1)} = o(1). \tag{29}$$

We see that A satisfies the assumptions of Proposition 4.2 with some $m \geq 1$ and $\kappa = (\log N)^d$. We get for any $t > 0$, and N large enough,

$$\mathbb{P}\big(\big|\mathrm{tr}_N\,|A|^p - \mathbb{E}\,\mathrm{tr}_N\,|A|^p\big| > t\big) \leq 8\exp\bigg(-\frac{N^{1+\frac{2}{p}}}{cp^2(\log N)^{2d}}\min\bigg\{t^{2/p}, \frac{t^2}{m^{2(p-1)}}\bigg\}\bigg),$$

which yields, as $\alpha < 2$,

$$\lim_{N\to+\infty}N^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\big(\big|\mathrm{tr}_N\,|A|^p - \mathbb{E}\,\mathrm{tr}_N\,|A|^p\big| > t\big) = -\infty. \tag{30}$$

We know from [1, Theorem 2.1.1, Lemma 2.1.6], that

$$\mathbb{E}\,\mathrm{tr}_N\,|X_N|^p \underset{N\to+\infty}{\longrightarrow} \langle\sigma_{\mathrm{sc}}, |x|^p\rangle. \tag{31}$$

Denoting μ_{X_N} and μ_A the empirical spectral measures of X_N and A respectively, we have using the decreasing coupling and [5, Theorem III 4.4],

$$W_p(\mathbb{E}\mu_{X_N}, \mathbb{E}\mu_A) \leq \big(\mathbb{E}\,\mathrm{tr}_N\,|X_N - A|^p\big)^{1/p}, \tag{32}$$

where W_p denotes the p-Wasserstein distance. As a consequence of the polar decomposition, we can write $|X_N - A|^p = (X_N - A)^p U$, where U is a unitary matrix, so that

$$\mathbb{E}\,\mathrm{tr}\,|X_N - A|^p \leq \frac{1}{N^{p/2}}\sum_{i_1,\ldots,i_{p+1}}\mathbb{E}\prod_{j=1}^{p}|X_{i_j,i_{j+1}}|\mathbb{1}_{|X_{i_j,i_{j+1}}|\leq(\log N)^d}, \tag{33}$$

Hölder's inequality yields,

$$\mathbb{E}\,\mathrm{tr}\,|X_N - A|^p \leq N^{p/2+1}\max\big(\mathbb{E}|X_{1,1}|^p\mathbb{1}_{|X_{1,1}|>(\log N)^p}, \mathbb{E}|X_{1,2}|^p\mathbb{1}_{|X_{1,2}|>(\log N)^p}\big),$$

where we used the fact that the entries of X are centered. Integrating by parts, we get

$$\mathbb{E}\,\mathrm{tr}\,|X_N - A|^p = O\big(N^{p/2+1} e^{-\frac{\gamma}{2}(\log N)^{\alpha d}}\big), \tag{34}$$

where γ is as in (27). As $\alpha d > 1$, we deduce by (32), $W_p(\mathbb{E}\mu_{X_N}, \mathbb{E}\mu_A) = o(1)$, which yields

$$\big|\mathbb{E}\,\mathrm{tr}_N\,|X_N|^p - \mathbb{E}\,\mathrm{tr}_N\,|A|^p\big| = o(1).$$

We can conclude with (30) and (31) that $(\mathrm{tr}_N\,|A|^p)_{N\in\mathbb{N}}$ is exponentially tight. $\qquad\square$

For the second event $\{\mathrm{tr}_N\,|B^\varepsilon|^p > t^p\}$, we have the following lemma.

Lemma 4.5. *For any $\varepsilon > 0$, we have*

$$\lim_{t\to+\infty}\limsup_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\big(\mathrm{tr}_N\big|B^\varepsilon\big|^p > t\big) = -\infty.$$

Proof. Since $p \geq 2$, we have

$$\big(\mathrm{tr}\big|B^\varepsilon\big|^p\big)^{2/p} \leq \mathrm{tr}\big(B^\varepsilon\big)^2.$$

Thus,

$$\mathbb{P}\big(\mathrm{tr}\big|B^\varepsilon\big|^p \geq tN\big) \leq \mathbb{P}\big(\mathrm{tr}\big(B^\varepsilon\big)^2 \geq t^{2/p}N^{2/p}\big).$$

Chernoff's inequality yields for any $\lambda > 0$,

$$\mathbb{P}\bigg(\sum_{1\leq i\leq j\leq N}\big|B^\varepsilon_{i,j}\big|^2 \geq \frac{t^{2/p}}{2}N^{2/p}\bigg) \leq e^{-\frac{\lambda}{2}t^{\frac{2}{p}}N^{\frac{2}{p}+1}}\prod_{1\leq i\leq j\leq N}\mathbb{E}\big(e^{\lambda|X_{i,j}|^2\mathbb{1}_{(\log N)^d < |X_{i,j}| < \varepsilon N^{\frac{1}{2}+\frac{1}{p}}}}\big).$$

Let $1 \leq i \leq j \leq N$. Recall that for μ a probability measure on $\mathbb{R}$ and $g \in C^1$, we have the following integration by parts formula:

$$\int_a^b g(x)\,d\mu(x) = g(a)\mu[a, +\infty) - g(b)\mu(b, +\infty) + \int_a^b g'(x)\mu[x, +\infty)\,dx.$$

Thus, we get for N large enough,

$$\mathbb{E}\big(e^{\lambda|X_{i,j}|^2\mathbb{1}_{(\log N)^d < |X_{i,j}| < \varepsilon N^{\frac{1}{2}+\frac{1}{p}}}}\big) \leq 1 + \int_{(\log N)^d}^{\varepsilon N^{\frac{1}{2}+\frac{1}{p}}} 2\lambda x e^{f(x)}\,dx,$$

with $f(x) = \lambda x^2 - \gamma x^\alpha$, and γ is as in (27). Let

$$\lambda = \frac{\alpha\gamma}{2}\varepsilon^{\alpha-2}N^{-(2-\alpha)(\frac{1}{2}+\frac{1}{p})}.$$

With this choice of λ, one can easily check that f is non-increasing on $[(\log N)^d, \varepsilon N^{\frac{1}{2}+\frac{1}{p}}]$. Thus,

$$\mathbb{E}\big(e^{\lambda|X_{i,j}|^2\mathbb{1}_{(\log N)^d < |X_{i,j}| < \varepsilon N^{\frac{1}{2}+\frac{1}{p}}}}\big) \leq 1 + 2\lambda\varepsilon^2 N^{1+\frac{2}{p}}e^{f((\log N)^d)}$$

$$\leq 1 + \alpha\gamma\varepsilon^\alpha N^{\alpha(\frac{1}{2}+\frac{1}{p})}e^{f((\log N)^d)}.$$

But for N large enough,

$$f\big((\log N)^d\big) = \frac{\alpha\gamma}{2}\varepsilon^{\alpha-2}N^{-(2-\alpha)(\frac{1}{2}+\frac{1}{p})}(\log N)^{2d} - \gamma(\log N)^{\alpha d} \le -\frac{\gamma}{2}(\log N)^{\alpha d}.$$

As $\alpha d > 1$, we get for N large enough,

$$\mathbb{E}\big(e^{\lambda|X_{i,j}|^2\mathbb{1}_{(\log N)^d < |X_{i,j}| < \varepsilon N^{\frac{1}{2}+\frac{1}{p}}}}\big) \le 1 + e^{-\frac{\gamma}{4}(\log N)^{\alpha d}} \le \exp\big(e^{-\frac{\gamma}{4}(\log N)^{\alpha d}}\big).$$

Then,

$$\mathbb{P}\big(\mathrm{tr}\big|B^\varepsilon\big|^p \ge tN\big) \le \exp\bigg(-\frac{\alpha\gamma}{4}\varepsilon^{\alpha-2}N^{\alpha(\frac{1}{2}+\frac{1}{p})}t^{\frac{2}{p}}\bigg)\exp\big(N^2 e^{-\frac{\gamma}{2}(\log N)^{\alpha d}}\big). \tag{35}$$

Since $\alpha d > 1$, we get

$$\lim_{t\to+\infty}\limsup_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\big(\mathrm{tr}\big|B^\varepsilon\big|^p \ge tN\big) = -\infty. \qquad\square$$

We now turn to the event $\{\mathrm{tr}_N |C^\varepsilon|^p > t\}$. As a consequence of Bennett's inequality, we have the following lemma.

Lemma 4.6. *For any $\varepsilon > 0$,*

$$\lim_{t\to+\infty}\lim_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\big(\mathrm{tr}_N\big|C^\varepsilon\big|^p > t\big) = -\infty.$$

To prove this lemma, we will first show that at the exponential scale C^ε has a finite number of non-zero entries.

Proposition 4.7. *For all $\varepsilon > 0$,*

$$\lim_{r\to+\infty}\limsup_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\big(\mathrm{Card}\{(i,j) : C_{i,j}^\varepsilon \ne 0\} \ge r\big) = -\infty,$$

where C^ε is as in (25).

Proof. Let $\varepsilon > 0$. Note that

$$\mathbb{P}\big(\mathrm{Card}\{(i,j) : C_{i,j}^\varepsilon \ne 0\} \ge r\big) \le \mathbb{P}\bigg(\sum_{1\le i\le j\le N}\mathbb{1}_{|X_{i,j}|\ge\varepsilon N^{\frac{1}{2}+\frac{1}{p}}} \ge \frac{r}{2}\bigg).$$

Let $p_{i,j} = \mathbb{P}(|X_{i,j}| \ge \varepsilon N^{\frac{1}{2}+\frac{1}{p}})$, for $i, j \in \{1, 2\}$. From (27), we have

$$p_{1,1} \vee p_{1,2} = o\bigg(\frac{1}{N^2}\bigg).$$

Therefore, it is enough to show that

$$\lim_{r\to+\infty}\limsup_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})}\log\mathbb{P}\bigg(\sum_{1\le i\le j\le N}(\mathbb{1}_{|X_{i,j}|\ge\varepsilon N^{\frac{1}{2}+\frac{1}{p}}} - p_{i,j}) \ge r\bigg) = -\infty.$$

By Bennett's inequality (see [24, Theorem 2.9]) we have,

$$\mathbb{P}\bigg(\sum_{1\le i\le j\le N}(\mathbb{1}_{|X_{i,j}|\ge\varepsilon N^{\frac{1}{2}+\frac{1}{p}}} - p_{i,j}) \ge r\bigg) \le \exp\bigg(-vh\bigg(\frac{r}{v}\bigg)\bigg),$$

with $h(x) = (x+1)\log(x+1) - x$, and $v = \sum_{i \leq j} p_{i,j}$. From (27), we have for N large enough,

$$v \leq N^2 e^{-\gamma \varepsilon^\alpha N^{\alpha(\frac{1}{2}+\frac{1}{p})}}.$$

As $h(x) \underset{+\infty}{\sim} x\log(x)$, we get for N large enough,

$$\mathbb{P}\left(\sum_{1 \leq i,j \leq N} (\mathbb{1}_{|X_{i,j}| \geq \varepsilon N^{\frac{1}{2}+\frac{1}{p}}} - p_{i,j}) \geq r \right) \leq \exp\left(-r\gamma \varepsilon^\alpha N^{\alpha(\frac{1}{2}+\frac{1}{p})}\right) \exp\left(r\log\left(\frac{r}{N^2}\right)\right),$$

which gives the claim. $\qquad\square$

With this result on the number of non-zero entries of C^ε, we can bound the trace of $|C^\varepsilon|^p$ by the polar decomposition, which yields the exponential estimate claimed in Lemma 4.6.

Proof of Lemma 4.6. Using the polar decomposition as in (33), and bounding each coefficient of C^ε by $\varepsilon^{-1} N^{1/p}$, we get,

$$\mathrm{tr}|C^\varepsilon|^p \leq |\mathcal{I}^\varepsilon|^p N \varepsilon^{-p},$$

where $|\mathcal{I}^\varepsilon|$ denotes the number of non-zero entries in C^ε. Due to Lemma 4.7, we get,

$$\lim_{t \to +\infty} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\mathrm{tr}_N |C^\varepsilon|^p > t\right) = -\infty. \qquad\square$$

At last, we prove the following exponential tightness for $\mathrm{tr}_N |D^\varepsilon|^p$.

Lemma 4.8. *It holds*

$$\lim_{\varepsilon \to 0} \limsup_{t \to +\infty} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\mathrm{tr}_N |D^\varepsilon|^p > t\right) = -\infty,$$

with D^ε as in (25).

Proof. A union bound gives for N large enough,

$$\mathbb{P}\left(D^\varepsilon \neq 0\right) \leq N^2 \exp\left(-\gamma \varepsilon^{-\alpha} N^{\alpha(\frac{1}{2}+\frac{1}{p})}\right), \tag{36}$$

with γ as in (27). $\qquad\square$

From (28), Lemmas 4.4, 4.5, and 4.6, we get for any $\varepsilon > 0$,

$$\limsup_{t \to +\infty} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\mathrm{tr}_N |X_N|^p > t\right)$$

$$\leq \limsup_{t \to +\infty} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\mathrm{tr}_N |D^\varepsilon|^p > t\right).$$

Taking the limsup as ε goes to 0, we see that Lemma 4.8 yields the exponential tightness claimed in Proposition 4.3. $\square$

4.5. *Exponential equivalences*

4.5.1. *First step*
We will prove in this section that we can ignore in the deviations of $\mathrm{tr}_N X_N^p$ the contributions of the large entries, that is such that $|X_{i,j}| > \varepsilon^{-1} N^{\frac{1}{2}+\frac{1}{p}}$, and the contributions of the intermediate entries, that is $X_{i,j} \in ((\log N)^d, \varepsilon N^{\frac{1}{2}+\frac{1}{p}})$. More precisely, we will prove the following exponential approximation.

Proposition 4.9. *For any $t > 0$,*

$$\limsup_{\varepsilon \to 0} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\big(\big|\mathrm{tr}_N X_N^p - \mathrm{tr}_N\big(A + C^\varepsilon\big)^p\big| > t\big) = -\infty,$$

with A and C^ε are as in (25). *In other words, $(\mathrm{tr}_N(A + C^\varepsilon)^p)_{N \in \mathbb{N}}$ are exponentially good approximations of $(\mathrm{tr}_N X_N^p)_{N \in \mathbb{N}}$.*

Proof. We will first prove that for any $X, Y \in \mathcal{M}_N$,

$$\big|\mathrm{tr}\, X^p - \mathrm{tr}\, Y^p\big| \le p 2^{p-2}\Big(\big(\mathrm{tr}\, |X|^p\big)^{\frac{p-1}{p}} \big(\mathrm{tr}\, |X - Y|^p\big)^{\frac{1}{p}} + \mathrm{tr}\, |X - Y|^p\Big). \tag{37}$$

Let $X, Y \in \mathcal{M}_N$. By the mean value theorem we have,

$$\big|\mathrm{tr}\, X^p - \mathrm{tr}\, Y^p\big| \le p \sup_{0 \le \theta \le 1} \big|\mathrm{tr}\big(X + \theta(Y - X)\big)^{p-1}(Y - X)\big|.$$

By Hölder's inequality (see [5, Corollary IV.2.6]), we deduce,

$$\big|\mathrm{tr}\, X^p - \mathrm{tr}\, Y^p\big| \le p \sup_{0 \le \theta \le 1} \big(\mathrm{tr}\big|X + \theta(Y - X)\big|^p\big)^{\frac{p-1}{p}} \big(\mathrm{tr}\, |Y - X|^p\big)^{\frac{1}{p}}.$$

Using the triangular inequality for the p-Schatten norm and the convexity of the power function $x \mapsto |x|^{p-1}$, we get

$$\big|\mathrm{tr}\, X^p - \mathrm{tr}\, Y^p\big| \le p 2^{p-2} \sup_{0 \le \theta \le 1} \Big(\big(\mathrm{tr}\, |X|^p\big)^{\frac{p-1}{p}} + \big(\theta\, \mathrm{tr}\, |Y - X|^p\big)^{\frac{p-1}{p}}\Big)\big(\mathrm{tr}\, |Y - X|^p\big)^{\frac{1}{p}},$$

which yields the claim (37).

Let now $\tau > 0$. From (37), we deduce that there exists a non-negative function h depending on τ, satisfying $h(t) \to 0$ as $t \to 0$, such that for any $X, Y \in \mathcal{H}_N^{(\beta)}$, if

$$\mathrm{tr}_N |X|^p \le \tau \quad \text{and} \quad \big|\mathrm{tr}_N X^p - \mathrm{tr}_N Y^p\big| > t,$$

for some $t > 0$, then,

$$\mathrm{tr}_N |X - Y|^p > h(t).$$

But, from Proposition 4.3, we know that $(\mathrm{tr}_N |X_N|^p)_{N \in \mathbb{N}}$ is exponentially tight, therefore, it is enough to show that for any $\tau > 0$,

$$\limsup_{\varepsilon \to 0} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\big(\big|\mathrm{tr}_N X_N^p - \mathrm{tr}_N\big(A + C^\varepsilon\big)^p\big| > t, \mathrm{tr}_N |X_N|^p \le \tau\big) = -\infty.$$

Let $\tau > 0$. With the previous observation, we get for any $t > 0$,

$$\mathbb{P}\big(\big|\mathrm{tr}_N X_N^p - \mathrm{tr}_N\big(A + C^\varepsilon\big)^p\big| > t, \mathrm{tr}_N |X_N|^p \le \tau\big) \le \mathbb{P}\big(\mathrm{tr}_N\big|B^\varepsilon + D^\varepsilon\big|^p > h(t)\big).$$

By the triangular inequality for the p-Schatten norm, we get

$$\mathbb{P}\big(\big|\mathrm{tr}_N X_N^p - \mathrm{tr}_N\big(A + C^\varepsilon\big)^p\big| > t, \mathrm{tr}_N |X_N|^p \le \tau\big) \le \mathbb{P}\bigg(\mathrm{tr}_N\big|B^\varepsilon\big|^p > \frac{h(t)}{2^p}\bigg) + \mathbb{P}\bigg(\mathrm{tr}_N\big|D^\varepsilon\big|^p > \frac{h(t)}{2^p}\bigg). \tag{38}$$

But, on one hand (35) yields

$$\lim_{\varepsilon \to 0} \lim_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\bigg(\mathrm{tr}_N\big|B^\varepsilon\big|^p > \frac{h(t)}{2^p}\bigg) = -\infty,$$

and on the other hand, (36) gives

$$\lim_{\varepsilon \to 0} \lim_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\operatorname{tr}_N |D^\varepsilon|^p > \frac{h(t)}{2^p} \right) = -\infty.$$

This concludes the proof of Proposition 4.9, taking the limsup as N goes to $+\infty$ at the exponential scale, and then the limsup as ε goes to 0 in (38). $\qquad\square$

4.5.2. *Second step*

We show here that in the study of the deviations of $\operatorname{tr}_N(A + C^\varepsilon)^p$, we can replace A by a matrix H independent of X, and that $\operatorname{tr}_N(H + C^\varepsilon)^p$ is exponentially equivalent to its conditional expectation given the σ-algebra $\mathcal{F}$, generated by the $X_{i,j}$ such that $|X_{i,j}| > (\log N)^d$. More precisely, we will prove the following result.

Proposition 4.10. *Let $\mathcal{F}$ be the σ-algebra generated by the variables $X_{i,j}\mathbb{1}_{|X_{i,j}|>(\log N)^d}$. Let H be a random Hermitian matrix independent of X, such that $(H_{i,j})_{i \leq j}$ are independent, and for all $1 \leq i \leq N$, $H_{i,i}$ has the same law as $X_{1,1}/\sqrt{N}$ conditioned on $\{|X_{1,1}| \leq (\log N)^d\}$, and for all $i < j$, $H_{i,j}$ has the same law as $X_{1,2}/\sqrt{N}$ conditioned on $\{|X_{1,2}| \leq (\log N)^d\}$.*

For any $t > 0$,

$$\lim_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\left| \operatorname{tr}_N X_N^p - \mathbb{E}_{\mathcal{F}} \operatorname{tr}_N (H + C^\varepsilon)^p \right| > t \right) = -\infty,$$

where $\mathbb{E}_{\mathcal{F}}$ denotes the conditional expectation given $\mathcal{F}$.

Proof. By Proposition 4.9, we know that $(\operatorname{tr}_N(A + C^\varepsilon)^p)_{N \in \mathbb{N}, \varepsilon > 0}$ are exponentially good approximations of $(\operatorname{tr}_N X_N^p)_{N \in \mathbb{N}}$, therefore it is enough to show that for all $\varepsilon > 0$, and $t > 0$,

$$\lim_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\left| \operatorname{tr}_N (A + C^\varepsilon)^p - \mathbb{E}_{\mathcal{F}} \operatorname{tr}_N (H + C^\varepsilon)^p \right| > t \right) = -\infty.$$

From Proposition 4.6, we see that is actually sufficient to show that for any $r \in \mathbb{N}$,

$$\lim_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\left| \operatorname{tr}_N (A + C^\varepsilon)^p - \mathbb{E}_{\mathcal{F}} \operatorname{tr}_N (H + C^\varepsilon)^p \right| > t, |\mathcal{I}_\varepsilon| \leq r \right) = -\infty,$$

where

$$\mathcal{I}_\varepsilon = \left\{ (i, j) \in \{1, \ldots, N\} \times \{1, \ldots, N\} : C_{i,j}^\varepsilon \neq 0 \right\}.$$

Note that C^ε is $\mathcal{F}$-measurable, and given $\mathcal{F}$, A has independent up-diagonal entries bounded by $(\log N)^d/\sqrt{N}$. Moreover, using the triangle inequality for the $2(p-1)$-Schatten norm, we get

$$\operatorname{tr}(\mathbb{E}A + C^\varepsilon)^{2(p-1)} \leq 2^{2(p-1)} \max\left(\operatorname{tr}(\mathbb{E}A)^{2(p-1)}, \operatorname{tr}(C^\varepsilon)^{2(p-1)} \right).$$

On one hand, we have, expanding the trace and bounding each entry of C^ε by $\varepsilon^{-1} N^{1/p}$,

$$\operatorname{tr}(C^\varepsilon)^{2(p-1)} \leq |\mathcal{I}_\varepsilon|^{2(p-1)} \varepsilon^{-2(p-1)} N^{2-\frac{2}{p}},$$

and on the other hand we have from (29) that $\operatorname{tr}(\mathbb{E}A)^{2(p-1)} = o(1)$. Therefore, we can apply the result of Proposition 4.2 for the trace of $(A + C^\varepsilon)^p$ under the conditional probability given $\mathcal{F}$. As $\alpha < 2$, we get that for any $t > 0$, and $r \in \mathbb{N}$,

$$\lim_{N \to +\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\left(\left| \operatorname{tr}_N (A + C^\varepsilon)^p - \mathbb{E}_{\mathcal{F}} \operatorname{tr}_N (A + C^\varepsilon)^p \right| > t, |\mathcal{I}_\varepsilon| \leq r \right) = -\infty.$$

We will use the same decoupling argument as in [7], to remove the dependency between A and C^ε. Let $I = \{(i, j) : |X_{i,j}| > (\log N)^d\}$. Define A' the $N \times N$ matrix with (i, j)-entry

$$A'_{i,j} = A_{i,j}\mathbb{1}_{(i,j)\notin I} + H_{i,j}\mathbb{1}_{(i,j)\in I}. \tag{39}$$

Note that A' and H are both independent of $\mathcal{F}$ and have the same law. Therefore,

$$\mathbb{E}_\mathcal{F} \operatorname{tr}_N (A' + C^\varepsilon)^p = \mathbb{E}_\mathcal{F} \operatorname{tr}_N (H + C^\varepsilon)^p.$$

Due to the triangular inequality and Lemma 4.6, it only remains to prove that for any $t > 0$, and any $\tau > 0$,

$$\lim_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\big(|\mathbb{E}_\mathcal{F} \operatorname{tr}_N Y^p - \mathbb{E}_\mathcal{F} \operatorname{tr}_N Y'^p| > t, \operatorname{tr}_N |C^\varepsilon|^p \leq \tau\big) = -\infty,$$

where $Y = A + C^\varepsilon$ and $Y' = A' + C^\varepsilon$. Note that using (37) and Hölder's inequality we have,

$$\big|\mathbb{E}_\mathcal{F} \operatorname{tr} Y'^p - \mathbb{E}_\mathcal{F} \operatorname{tr} Y^p\big| \leq p2^{p-2}\big((\mathbb{E}_\mathcal{F} \operatorname{tr}|Y'|^p)^{\frac{p-1}{p}}(\mathbb{E}_\mathcal{F} \operatorname{tr}|A - A'|^p)^{\frac{1}{p}} + \mathbb{E}_\mathcal{F} \operatorname{tr}|A - A'|^p\big).$$

Besides, using again the triangular inequality for the p-Schatten norm, we get

$$\mathbb{E}_\mathcal{F} \operatorname{tr}_N |A' + C^\varepsilon|^p \leq 2^p \max(\mathbb{E} \operatorname{tr}_N |H|^p, \operatorname{tr}_N |C^\varepsilon|^p).$$

With the same argument as in the proof of Lemma 4.4 we have

$$\mathbb{E} \operatorname{tr}_N |H|^p \underset{N\to+\infty}{\longrightarrow} \langle\sigma_{\mathrm{sc}}, |x|^p\rangle.$$

Thus, we see that it is sufficient to show that for any $t > 0$,

$$\lim_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}\big(\mathbb{E}_\mathcal{F} \operatorname{tr}_N |A - A'|^p > t\big) = -\infty.$$

Expanding the trace using the polar decomposition, we get

$$\mathbb{E}_\mathcal{F} \operatorname{tr}_N |A - A'|^p \leq c_0 \frac{|I|^p}{N^{1+p/2}}, \tag{40}$$

where c_0 is constant independent of N such that,

$$\max(\mathbb{E}|\sqrt{N} H_{1,1}|^p, \mathbb{E}|\sqrt{N} H_{1,2}|^p) \leq c_0.$$

Thus, in order to control $\mathbb{E}_\mathcal{F} \operatorname{tr}_N |A - A'|^p$, we need to make sure that I contains no more than $t N^{1+p/2}$ indices, for any $t > 0$, at the exponential scale $N^{\alpha(\frac{1}{2}+\frac{1}{p})}$. By a argument similar as in the proof of Proposition 4.7, we get the following lemma.

Lemma 4.11. *Let $I = \{(i, j) : |X_{i,j}| > (\log N)^d\}$. For $\delta > 0$, we define the event,*

$$F_\delta = \left\{|I| \leq \frac{\delta}{c_0} N^{1+2/p}\right\}.$$

It holds that

$$\lim_{N\to+\infty} N^{-\alpha(\frac{1}{2}+\frac{1}{p})} \log \mathbb{P}(F_\delta^c) = -\infty.$$

Using (40), and Lemma (4.11), we get the claim. □

4.5.3. *Third step*

We showed in Proposition 4.10 that $(\mathbb{E}_{\mathcal{F}} \operatorname{tr}_N (H + C^\varepsilon)^p)_{N \in \mathbb{N}, \varepsilon > 0}$ are exponentially good approximations of $(\operatorname{tr}_N X_N^p)_{N \in \mathbb{N}}$. We will prove now that we can approximate $\mathbb{E}_{\mathcal{F}} \operatorname{tr}_N (H + C^\varepsilon)^p$ at the exponential scale $N^{\alpha(\frac{1}{2} + \frac{1}{p})}$, by $\mathbb{E} \operatorname{tr}_N H^p + \operatorname{tr}_N (C^\varepsilon)^p$, and then by $\langle \sigma_{\mathrm{sc}}, x^p \rangle + \operatorname{tr}_N (C^\varepsilon)^p$. This will give good exponential approximations of $(\operatorname{tr}_N X_N^p)_{N \in \mathbb{N}}$, as stated in the following proposition.

Proposition 4.12. *For any $t > 0$,*

$$\lim_{\varepsilon \to 0} \limsup_{N \to +\infty} N^{-\alpha(\frac{1}{2} + \frac{1}{p})} \log \mathbb{P}\big(\big|\operatorname{tr}_N X_N^p - \langle \sigma_{\mathrm{sc}}, x^p \rangle - \operatorname{tr}_N (C^\varepsilon)^p\big| > t\big) = -\infty,$$

where A and C^ε are as in (25).

In order to prove that $\mathbb{E} \operatorname{tr}_N H^p + \operatorname{tr}_N (C^\varepsilon)^p$ is an exponential equivalent of $\mathbb{E}_{\mathcal{F}} \operatorname{tr}_N (H + C^\varepsilon)^p$, we will need the following deterministic lemma.

Lemma 4.13. *Let $p \geq 2$. Let H and C be two Hermitian matrices of size N. Assume that C is of rank at most r. We have*

$$\big|\operatorname{tr}(H + C)^p - \operatorname{tr} H^p - \operatorname{tr} C^p\big| \leq 2^p r \max_{1 \leq k \leq p-1} \|H\|^k \|C\|^{p-k},$$

where $\|\cdot\|$ denotes the operator norm.

Proof. Expanding the sum we get

$$\operatorname{tr}(H + C)^p = \sum_{k=0}^{p} \sum_{\substack{M^{(i)} \in \{H, C\} \\ |\{i : M^{(i)} = H\}| = k}} \operatorname{tr}\big(M^{(1)} \cdots M^{(p)}\big).$$

Fix $k \in \{1, \dots, p-1\}$, and $M^{(1)}, \dots, M^{(p)}$ matrices such that $M^{(i)} \in \{H, C\}$, and $\operatorname{Card}\{i : M^{(i)} = H\} = k$. Let $(\eta_j)_{1 \leq j \leq N}$ be an orthonormal basis of eigenvectors for C such that $\eta_{r+1}, \dots, \eta_N$ are in the kernel of C. Using the cyclicity of the trace, we can assume $M^{(p)} = C$. Assuming $M^{(p)} = C$, we get

$$\begin{aligned}
\big|\operatorname{tr}\big(M^{(1)} \cdots M^{(p)}\big)\big| &= \left|\sum_{j=1}^{N} \langle M^{(1)} \cdots M^{(p)} \eta_j, \eta_j \rangle\right| \\
&= \left|\sum_{j=1}^{r} \langle M^{(1)} \cdots M^{(p)} \eta_j, \eta_j \rangle\right| \\
&\leq r \|H\|^k \|C\|^{p-k},
\end{aligned}$$

which ends the proof of the claim. $\square$

Proof. Note that the same argument as in the proof of Lemma 4.4 yields

$$\mathbb{E} \operatorname{tr}_N H^p \underset{N \to +\infty}{\longrightarrow} \langle \sigma_{\mathrm{sc}}, x^p \rangle.$$

Therefore, due to Proposition 4.10, we only need to prove that for any $\varepsilon > 0$,

$$\lim_{N \to +\infty} N^{-\alpha(\frac{1}{2} + \frac{1}{p})} \log \mathbb{P}\big(\big|\mathbb{E}_{\mathcal{F}} \operatorname{tr}_N \big(H + C^\varepsilon\big)^p - \mathbb{E} \operatorname{tr}_N H^p - \operatorname{tr}_N \big(C^\varepsilon\big)^p\big| > t\big) = -\infty.$$

Using Lemma 4.1 and the fact that the rank of a matrix is bounded by the number of its non-zero entries, we have

$$\left|\mathbb{E}_{\mathcal{F}}\operatorname{tr}_N\left(H + C^\varepsilon\right)^p - \mathbb{E}\operatorname{tr}_N H^p - \operatorname{tr}_N\left(C^\varepsilon\right)^p\right| \leq \frac{2^p}{N}|\mathcal{I}_\varepsilon| \max_{1\leq k\leq p-1}\left\{\left\|C^\varepsilon\right\|^{p-k}\mathbb{E}\|H\|^k\right\},$$

where $\mathcal{I}_\varepsilon$ denotes the set of indices (i, j) such that $C_{i,j}^\varepsilon \neq 0$. But,

$$\left\|C^\varepsilon\right\| \leq |\mathcal{I}_\varepsilon| \sup_{i,j}|C_{i,j}| \leq |\mathcal{I}_\varepsilon|\varepsilon^{-1} N^{1/p}.$$

Thus,

$$\left|\mathbb{E}_{\mathcal{F}}\operatorname{tr}_N\left(H + C^\varepsilon\right)^p - \mathbb{E}\operatorname{tr}_N H^p - \operatorname{tr}_N\left(C^\varepsilon\right)^p\right| \leq \frac{2^p\varepsilon^{-p+1}}{N^{1/p}}|\mathcal{I}_\varepsilon|^p \max_{1\leq k\leq p-1}\mathbb{E}\|H\|^k.$$

But we know from [1, Theorem 2.1.22, Exercice 2.1.27] that $\|X\|$ converges in all L^p spaces to 2, and we have

$$\mathbb{E}\|X - H\|^p = \mathbb{E}\left\|X - A'\right\|^p \leq \mathbb{E}\operatorname{tr}\left|X - A'\right|^p,$$

where A' is as in (39). With the same argument as in Lemma 4.4, we get

$$\mathbb{E}\operatorname{tr}\left|X - A'\right|^p = o(1).$$

Thus, for any $k \in \{1, \ldots, p\}$, $\mathbb{E}\|H\|^k$ is bounded. We can find a constant $M_p > 0$ such that,

$$\left|\mathbb{E}_{\mathcal{F}}\operatorname{tr}_N\left(A + C^\varepsilon\right)^p - \mathbb{E}\operatorname{tr}_N A^p - \operatorname{tr}_N\left(C^\varepsilon\right)^p\right| \leq M_p|\mathcal{I}_\varepsilon|^p N^{-\frac{1}{p}}.$$

Thus, for any $t > 0$, and $r \in \mathbb{N}$,

$$\lim_{N\to+\infty} N^{-\alpha\left(\frac{1}{2}+\frac{1}{p}\right)} \log\mathbb{P}\left(\left|\mathbb{E}_{\mathcal{F}}\operatorname{tr}_N\left(A + C^\varepsilon\right)^p - \mathbb{E}\operatorname{tr}_N A^p - \operatorname{tr}_N\left(C^\varepsilon\right)^p\right| > t, |\mathcal{I}_\varepsilon| \leq r\right) = -\infty.$$

Invoking Lemma 4.7, we get the claim. □

4.6. A large deviations principle for $\operatorname{tr}_N X_N^p$

We proved in the previous section that $(\langle\sigma_{\mathrm{sc}}, x^p\rangle + \operatorname{tr}_N(C^\varepsilon)^p)_{\varepsilon>0, N\in\mathbb{N}}$ are exponentially good approximations of $(\operatorname{tr}_N X_N^p)_{N\in\mathbb{N}}$ at the exponential scale considered. The aim of this section is to show that we can derive for each $\varepsilon > 0$ a LDP for $(\operatorname{tr}_N(C^\varepsilon)^p)_{N\in\mathbb{N}}$, using the contraction principle, and deduce a LDP for $(\operatorname{tr}_N X_N^p)_{N\in\mathbb{N}}$.

In the view of applying a contraction principle for the sequence $(\operatorname{tr}_N(C^\varepsilon)^p)_{N\in\mathbb{N}}$, we need to find a good space to embed C^ε so that we can define a trace which will be continuous. For every $r \in \mathbb{N}$, we define

$$\mathcal{E}_r = \left\{A \in \bigcup_{n\geq 1}\mathcal{H}_n^{(\beta)} : \operatorname{Card}\{(i, j) : A_{i,j} \neq 0\} \leq r\right\}.$$

For any $n \in \mathbb{N}$, let $\mathcal{S}_n$ be the symmetric group on the set $\{1, \ldots, n\}$. Let $\mathcal{S}$ denote the group $\bigcup_{n\in\mathbb{N}}\mathcal{S}_n$. We denote $\widetilde{\mathcal{E}}_r$ the set of equivalence classes of $\mathcal{E}_r$ under the action of $\mathcal{S}$, which is defined by

$$\forall\sigma \in \mathcal{S}, \forall A \in \mathcal{E}_r, \quad \sigma \cdot A = M_\sigma^{-1} A M_\sigma = (A_{\sigma(i),\sigma(j)})_{i,j},$$

where M_σ denotes the permutation matrix associated with the permutation σ i.e. $M_\sigma = (\delta_{i,\sigma(j)})_{i,j}$.

Let $\mathcal{H}_r^{(\beta)}/\mathcal{S}_r$ be the set of equivalence classes of $\mathcal{H}_r^{(\beta)}$ under the action of the symmetric group $\mathcal{S}_r$. Note that any equivalence class of the action of $\mathcal{S}$ on $\mathcal{E}_r$ has a representative in $\mathcal{H}_r^{(\beta)}$. This defines an injective map from $\widetilde{\mathcal{E}}_r$ into

$\mathcal{H}_r^{(\beta)}/\mathcal{S}_r$. Identifying $\widetilde{\mathcal{E}}_r$ to a subset of $\mathcal{H}_r^{(\beta)}/\mathcal{S}_r$, we equip $\widetilde{\mathcal{E}}_r$ of the quotient topology of $\mathcal{H}_r^{(\beta)}/\mathcal{S}_r$. This topology is metrizable by the distance $\tilde{d}$ given by

$$\forall \tilde{A}, \tilde{B} \in \widetilde{\mathcal{E}}_r, \quad \tilde{d}(\tilde{A}, \tilde{B}) = \min_{\sigma, \sigma' \in \mathcal{S}} \max_{i,j} |B_{\sigma(i),\sigma(j)} - A_{\sigma'(i),\sigma'(j)}|, \tag{41}$$

where A and B are two representatives of $\tilde{A}$ and $\tilde{B}$ respectively.

Since the trace is continuous and invariant by conjugation, we can define the trace on $\mathcal{H}_r^{(\beta)}/\mathcal{S}_r$ and it will be still continuous. Therefore, the trace on $\widetilde{\mathcal{E}}_r$ is continuous for the topology we defined above.

Let $\varepsilon > 0$ and $r > \mathbb{N}$. Let $\mathbb{P}_{N,r}^\varepsilon$ denote the law of $C^\varepsilon/N^{1/p}$ conditioned on the event $\{C^\varepsilon \in \mathcal{E}_r\}$, and $\widetilde{\mathbb{P}}_{N,r}^\varepsilon$ the push-forward of $\mathbb{P}_{N,r}^\varepsilon$ by the projection $\pi : \mathcal{E}_r \to \widetilde{\mathcal{E}}_r$. With these preliminary definitions, we can now state the LDP result for $(\widetilde{\mathbb{P}}_{N,r}^\varepsilon)_{N \in \mathbb{N}}$. The result is almost identical as [2, Proposition 7.1], the only difference being the choice of truncation of the entries. Thus, the rate function is identical, and only the speed is different. We refer the reader to [2] for the proof of the following proposition.

Proposition 4.14. *Let $r \in \mathbb{N}$ and $\varepsilon > 0$. Then $(\widetilde{\mathbb{P}}_{N,r}^\varepsilon)_{N \in \mathbb{N}}$ satisfies a large deviations principle with speed $N^{\alpha(\frac{1}{2}+\frac{1}{p})}$, and good rate function $I_{\varepsilon,r}$ defined for all $\tilde{A} \in \widetilde{\mathcal{E}}_r$ by*

$$I_{\varepsilon,r}(\tilde{A}) = \begin{cases} b \sum_{i \geq 1} |A_{i,i}|^\alpha + \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha & \text{if } A \in \mathcal{D}_{\varepsilon,r}, \\ +\infty & \text{otherwise}, \end{cases} \tag{42}$$

where A is a representative of the equivalence class $\tilde{A}$ and

$$\mathcal{D}_{\varepsilon,r} = \left\{ A \in \mathcal{E}_r : \forall i \leq j, A_{i,j} = 0 \text{ or } \varepsilon \leq |A_{i,j}| \leq \varepsilon^{-1}, \text{ and } A_{i,j}/|A_{i,j}| \in \text{supp}(\nu_{i,j}) \right\},$$

with $\nu_{i,j} = \nu_1$ if $i = j$, and $\nu_{i,j} = \nu_2$ if $i < j$, where ν_1 and ν_2 are defined in Definition 1.6.

We are now ready to use a contraction principle to prove that $(\text{tr}_N(C^\varepsilon)^p)_{N \in \mathbb{N}}$ follows a LDP for any $\varepsilon > 0$. The use of the contraction principle is made possible by the fact that the push-forward of $\widetilde{\mathbb{P}}_{N,r}^\varepsilon$ by the map $A \mapsto \text{tr} A^p$ on $\bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}$, are exponentially good approximations of $(\text{tr}_N(C^\varepsilon)^p)_{N \in \mathbb{N}}$.

Proposition 4.15. *Let $\varepsilon > 0$. The sequence $(\text{tr}_N(C^\varepsilon)^p)_{N \in \mathbb{N}}$ satisfies a large deviations principle of speed $N^{\alpha(\frac{1}{2}+\frac{1}{p})}$, and good rate function J_ε defined for all $x \in \mathbb{R}$ by,*

$$J_\varepsilon(x) = \inf \left\{ I_\varepsilon(A) : x = \text{tr} A^p, A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)} \right\},$$

where for any $A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}$,

$$I_\varepsilon(A) = \begin{cases} b \sum_{i \geq 1} |A_{i,i}|^\alpha + \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha & \text{if } A \in \mathcal{D}_\varepsilon, \\ +\infty & \text{otherwise}, \end{cases} \tag{43}$$

where $\mathcal{D}_\varepsilon = \bigcup_{r \in \mathbb{N}} \mathcal{D}_{\varepsilon,r}$, with $\mathcal{D}_{\varepsilon,r}$ as in Proposition 4.14.

Proof. Let $r \in \mathbb{N}$. We denote by f the function $\tilde{A} \in \widetilde{\mathcal{E}}_r \mapsto \text{tr} A^p$, with A a representative of $\tilde{A}$. As the trace is invariant by conjugation, f is well defined. We define the push-forward of $\tilde{P}_{N,r}^\varepsilon$ by the map f,

$$\nu_{N,r} = \tilde{P}_{N,r}^\varepsilon \circ f^{-1}.$$

Note that $\nu_{N,r}$ is the law of $\text{tr}_N(C^\varepsilon)^p$ conditioned on the event $\{C^\varepsilon \in \mathcal{E}_r\}$. We will show that $(\nu_{N,r})_{N,r \in \mathbb{N}}$ are exponentially good approximations of $(\text{tr}_N(C^\varepsilon)^p)_{N \in \mathbb{N}}$. Let $Y_{N,r}$ be random variable independent of C^ε, and distributed

according to $\nu_{N,r}$. Let

$$Z_{N,r} = \text{tr}_N \left(C^\varepsilon\right)^p \mathbb{1}_{C^\varepsilon \in \mathcal{E}_r} + Y_{N,r} \mathbb{1}_{C^\varepsilon \notin \mathcal{E}_r}.$$

Thus, $Z_{N,r}$ and $Y_{N,r}$ have the same law $\nu_{N,r}$. Furthermore, for any $t > 0$,

$$\mathbb{P}\left(\left|Z_{N,r} - \text{tr}_N \left(C^\varepsilon\right)^p\right| > t\right) \leq \mathbb{P}\left(C^\varepsilon \notin \mathcal{E}_r\right).$$

By Proposition 4.7, we get

$$\lim_{N \to +\infty} N^{-\alpha\left(\frac{1}{2}+\frac{1}{p}\right)} \log \mathbb{P}\left(\left|Z_{N,r} - \text{tr}_N \left(C^\varepsilon\right)^p\right| > t\right) = -\infty,$$

which shows that $(\nu_{N,r})_{N,r \in \mathbb{N}}$ are exponentially good approximations of $(\text{tr}_N(C^\varepsilon)^p)_{N \in \mathbb{N}}$.

For each $r \in \mathbb{N}$, the function f restricted to $\widetilde{\mathcal{E}}_r$ is continuous for the topology we equipped $\widetilde{\mathcal{E}}_r$. Note that as C^ε has entries bounded by $\varepsilon^{-1} N^{1/p}$, $\nu_{N,r}$ is compactly supported uniformly in N. Thus, $(\nu_{N,r})_{N \geq 1}$ is exponentially tight, the contraction principle (see [11, Theorem 4.2.1]) yields that $(\nu_{N,r})_{N \in \mathbb{N}}$ follows a LDP principle with speed $N^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}$ and good rate function $J_{\varepsilon,r}$ given by

$$J_{\varepsilon,r}(x) = \inf\left\{I_{\varepsilon,r}(\tilde{A}) : \tilde{A} \in \widetilde{\mathcal{E}}_r, x = f(\tilde{A})\right\},$$

where $I_{\varepsilon,r}$ is defined in Proposition 4.14. We can re-write this rate function as

$$J_{\varepsilon,r}(x) = \inf\left\{I_\varepsilon(A) : A \in \mathcal{E}_r, x = f(A)\right\},$$

where f denotes as well the function $A \mapsto \text{tr}(A)^p$ on $\bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}$, and where I_ε is defined in (43). By [11, Theorem 4.2.16], we deduce that $(\text{tr}_N(C^\varepsilon)^p)_{N \in \mathbb{N}}$ satisfies a weak LDP with speed $N^{\alpha\left(\frac{1}{2}+\frac{1}{p}\right)}$, and rate function J_ε defined by

$$\forall x \in \mathbb{R}, \quad J_\varepsilon(x) = \sup_{\delta > 0} \liminf_{r \to +\infty} \inf_{|y-x|<\delta} J_{\varepsilon,r}(y).$$

As $J_{\varepsilon,r}$ is non-increasing in r, we have

$$J_\varepsilon(x) = \sup_{\delta > 0} \inf_{r \in \mathbb{N}} \inf_{|y-x|<\delta} J_{\varepsilon,r}(y) = \sup_{\delta > 0} \inf_{|y-x|<\delta} \inf_{r \in \mathbb{N}} J_{\varepsilon,r}(y).$$

Let Φ be the function defined by

$$\forall x \in \mathbb{R}, \quad \Phi(x) = \inf_{r \in \mathbb{N}} J_{\varepsilon,r}(x).$$

Thus,

$$J_\varepsilon(x) = \sup_{\delta > 0} \inf_{|y-x|<\delta} \Phi(y).$$

We see that it suffices to show that Φ is lower semi-continuous to conclude that $J_\varepsilon = \Phi$. We will prove in fact that Φ has compact level sets.

Let $\tau > 0$. Let $x \in \mathbb{R}$, such that $\Phi(x) \leq \tau$. Then

$$\Phi(x) = \left\{I_\varepsilon(A) : x = f(A), I_\varepsilon(A) \leq 2\tau\right\}.$$

But for any $A \in \bigcup_{n \in \mathbb{N}} H_n(\mathbb{C})$ such that $I_\varepsilon(A) < +\infty$, we have

$$\left(b \wedge \frac{a}{2}\right) \varepsilon^\alpha \text{Card}\left\{(i, j) : A_{i,j} \neq 0\right\} \leq I_\varepsilon(A).$$

Thus taking r such that $(b \wedge \frac{a}{2})\varepsilon^{\alpha} \leq \tau$, we get

$$\Phi(x) = \left\{ I_{\varepsilon}(A) : x = f(A), I_{\varepsilon}(A) \leq 2\tau, A \in \mathcal{E}_r \right\}$$
$$= \left\{ I_{\varepsilon,r}(\tilde{A}) : x = f(A), \tilde{A} \in \widetilde{\mathcal{E}}_r \right\}.$$

Since f is continuous on $\widetilde{\mathcal{E}}_r$ and $I_{\varepsilon,r}$ is a good rate function, we have

$$\left\{ x \in \mathbb{R} : \Phi(x) \leq \tau \right\} = \left\{ f(\tilde{A}) : I_{\varepsilon,r}(\tilde{A}) \leq \tau, \tilde{A} \in \widetilde{\mathcal{E}}_r \right\}.$$

As f is continuous on $\mathcal{E}_r$, and $I_{\varepsilon,r}$ is a good rate function, we deduce that the τ-level sets of Φ are compact. Therefore $J_{\varepsilon} = \Phi$. $\qquad\square$

We are now ready to conclude the proof of Theorem 1.7.

Proof of Theorem 1.7 . By Proposition 4.12, $(\langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}_N(C^{\varepsilon})^p)_{N \in \mathbb{N}, \varepsilon > 0}$ are exponentially good approximations of $(\mathrm{tr}_N X_N^p)_{N \in \mathbb{N}}$. We deduce from Proposition 4.15 that for each $\varepsilon > 0$, the sequence $(\langle \sigma_{\mathrm{sc}}, x^p \rangle + \mathrm{tr}_N(C^{\varepsilon})^p)_{N \in \mathbb{N}}$ satisfies a LDP with speed $N^{\alpha(\frac{1}{2}+\frac{1}{p})}$, and with good rate function ψ_{ε} defined by

$$\psi_{\varepsilon}(x) = \begin{cases} J_{\varepsilon}(x - C_{p/2}) & \text{if } p \text{ is even,} \\ J_{\varepsilon}(x) & \text{if } p \text{ is odd,} \end{cases}$$

where J_{ε} is as in Proposition 4.15. Since $(\mathrm{tr}_N X_N^p)_{N \geq 1}$ is exponentially tight by Proposition 4.3, we deduce from [11, Theorem 4.2.16] that $(\mathrm{tr}_N X_N^p)_{N \in \mathbb{N}}$ satisfies a LDP with speed $N^{\alpha(\frac{1}{2}+\frac{1}{p})}$ and rate function J_p defined by

$$\forall x \in \mathbb{R}, \quad J_p(x) = \sup_{\delta > 0} \limsup_{\varepsilon \to 0} \inf_{|y-x| < \delta} \psi_{\varepsilon}(y).$$

Observe that for any $A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}$, $I_{\varepsilon}(A)$ is non-decreasing in ε. Therefore, ψ_{ε} is non-decreasing in ε. Thus,

$$\forall x \in \mathbb{R}, \quad J_p(x) = \sup_{\delta > 0} \inf_{\varepsilon > 0} \inf_{|y-x| < \delta} \psi_{\varepsilon}(y). \tag{44}$$

Let

$$\forall y \in \mathbb{R}, \quad \Phi_p(x) = \begin{cases} \varphi_p(x - C_{p/2}) & \text{if } p \text{ is even,} \\ \varphi_p(x) & \text{if } p \text{ is odd,} \end{cases}$$

with

$$\varphi_p(x) = \inf \left\{ I(A) : x = \mathrm{tr}\, A^p, A \in \mathcal{D} \right\},$$

where I is defined for any $A \in \bigcup_{n \geq 1} \mathcal{H}_n^{(\beta)}$, by

$$I(A) = b \sum_{i=1}^{+\infty} |A_{i,i}|^{\alpha} + a \sum_{i<j} |A_{i,j}|^{\alpha},$$

and $\mathcal{D} = \{A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)} : \forall i \leq j, A_{i,j} = 0 \text{ or } A_{i,j}/|A_{i,j}| \in \mathrm{supp}(\nu_{i,j})\}$. With these notations we have,

$$J_p(x) = \sup_{\delta > 0} \inf_{|x-y| < \delta} \Phi_p(y). \tag{45}$$

As for any $t > 0$, and $A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}$, $I(tA) = t^\alpha I(A)$, and $\mathrm{tr}(tA)^p = t^p \, \mathrm{tr} \, A^p$, we have for p even,

$$\forall y \in \mathbb{R}, \quad \varphi_p(y) = \begin{cases} \varphi_p(1) y^{\alpha/p} & \text{if } y \geq 0, \\ +\infty & \text{otherwise,} \end{cases}$$

and for p odd

$$\forall y \in \mathbb{R}, \quad \varphi_p(y) = \varphi_p(1)|y|^{\alpha/p}.$$

Therefore,

$$\forall y \in \mathbb{R}, \quad \Phi_p(x) = \begin{cases} \varphi_p(1)(x - C_{p/2})^{\alpha/p} & \text{if } p \text{ is even,} \\ +\infty & \text{otherwise,} \end{cases}$$

and if p is odd

$$\forall y \in \mathbb{R}, \quad \Phi_p(x) = \varphi_p(1)|x|^{\alpha/p}.$$

This shows in particular that Φ_p is lower semi-continuous. From (45), we get finally $J_p = \Phi_p$. $\qquad \square$

4.7. *Computation of $J_p(1)$*

We show here that we can compute the constant c_p appearing in Theorem 1.7 when $\alpha \in (0, 1]$ and p is even, and we give a lower bound and upper bound in the case where $\alpha \in (1, 2)$ and p is even.

Theorem 4.16. *With the notations of Theorem 1.7, we have the following*:

(a) *If p is even,*

$$\min\left(b, \frac{a}{2}\right) \leq c_p \leq \min\left(b, 2^{-\alpha/p} a\right).$$

(b) *If $\alpha \in (0, 1]$ and p is even,*

$$c_p = \min\left(b, 2^{-\alpha/p} a\right).$$

Proof. From the proof of Theorem 1.7, we know that

$$c_p = \inf\{I(A) : 1 = \mathrm{tr} \, A^p, A \in \mathcal{D}\}, \tag{46}$$

where I is defined for any $A \in \bigcup_{n \geq 1} \mathcal{H}_n^{(\beta)}$, by

$$I(A) = b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + a \sum_{i<j} |A_{i,j}|^\alpha,$$

and $\mathcal{D} = \{\bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)} : \forall i \leq j, A_{i,j} = 0 \text{ or } A_{i,j}/|A_{i,j}| \in \mathrm{supp}(\nu_{i,j})\}$, with $\nu_{i,j} = \nu_1$ if $i = j$, and $\nu_{i,j} = \nu_2$ if $i < j$, where ν_1 and ν_2 are defined in Definition 1.6.

Note that

$$c_p \leq \min\left(I(s), I\begin{pmatrix} 0 & 2^{-1/p} e^{i\theta} \\ 2^{-1/p} e^{-i\theta} & 0 \end{pmatrix}\right),$$

where $s \in \mathrm{supp}(\nu_1)$, and $\theta \in \mathrm{supp}(\nu_2)$. Thus,

$$c_p \leq \min\left(b, 2^{-\alpha/p} a\right),$$

which proves the upper bound in cases (a) and (b).

On the other hand, we have

$$c_p \geq \inf\left\{ b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha : A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}, 1 = \operatorname{tr} A^p \right\}$$

$$\geq \min\left(b, \frac{a}{2}\right) \inf\left\{ \sum_{i,j} |A_{i,j}|^\alpha : A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)} : \operatorname{tr} A^p = 1 \right\}.$$

Since $\alpha \in (0, 2)$, we know from [31, Theorem 3.32] that for any $A \in \mathcal{H}_n^{(\beta)}$,

$$\sum_{i,j} |A_{i,j}|^\alpha \geq \sum_{i=1}^n |\lambda_i|^\alpha, \tag{47}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. As $\alpha/p \leq 1$, we have

$$\sum_{i=1}^n |\lambda_i|^\alpha \geq \left(\sum_{i=1}^n |\lambda_i|^p \right)^{\alpha/p} = \left(\operatorname{tr} |A|^p \right)^{\alpha/p} \geq \left| \operatorname{tr} A^p \right|^{\alpha/p}.$$

Thus, if $\operatorname{tr} A^p = 1$, we have

$$\sum_{i,j} |A_{i,j}|^\alpha \geq 1.$$

We can deduce that

$$c_p \geq \min\left(b, \frac{a}{2}\right),$$

which proves the lower bound of case (b).

Assume now $\alpha \in (0, 1)$ and p is even. If $A \in \mathcal{H}_n^{(\beta)}$ is such that $\operatorname{tr} A^p = 1$, then

$$\sup_{\operatorname{tr} |B|^q = 1} \operatorname{tr} AB = 1,$$

with $q \geq 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Thus, we can deduce that

$$\forall i \in \{1, \ldots, n\}, \quad |A_{i,i}| \leq 1, \qquad \forall i, j \in \{1, \ldots, n\}, i \neq j, \quad |A_{i,j}| \leq 2^{-1/p}.$$

Then,

$$c_p \geq \inf\left\{ b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha : A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}, 1 = \operatorname{tr} A^p \right\}$$

$$\geq \min\left(b, 2^{-\frac{\alpha}{p}} a\right) \inf\left\{ \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + \frac{1}{2} \sum_{i \neq j} |2^{\frac{1}{p}} A_{i,j}|^\alpha : A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}, 1 = \operatorname{tr} A^p \right\}$$

$$\geq \min\left(b, 2^{-\frac{\alpha}{p}} a\right) \inf\left\{ \sum_{i=1}^{+\infty} |A_{i,i}| + \frac{1}{2} \sum_{i \neq j} |2^{\frac{1}{p}} A_{i,j}| : A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}, 1 = \operatorname{tr} A^p \right\}, \tag{48}$$

where we used in the last inequality the fact the $|A_{i,i}| \leq 1$, and $|A_{i,j}| \leq 2^{-1/p}$ for any $i \neq j$. Thus,

$$c_p \geq \min\big(b, 2^{-\frac{\alpha}{p}} a\big) \inf\left\{ \big(1 - 2^{\frac{1}{p}-1}\big) \sum_{i=1}^{+\infty} |A_{i,i}| + 2^{\frac{1}{p}-1} \sum_{i,j} |A_{i,j}| : A \in \bigcup_{n \in \mathbb{N}} \mathcal{H}_n^{(\beta)}, 1 = \operatorname{tr} A^p \right\}.$$

Using again [31, Theorem 3.36], and the triangular inequality, we get

$$c_p \geq \min\big(b, 2^{-\frac{\alpha}{p}} a\big) \inf_{n \geq 1} \inf\left\{ \big(1 - 2^{\frac{1}{p}-1}\big) \left| \sum_{i=1}^{n} \lambda_i \right| + 2^{\frac{1}{p}-1} \sum_{i=1}^{n} |\lambda_i| : A \in \mathcal{H}_n^{(\beta)}, \sum_{i=1}^{n} \lambda_i^p = 1 \right\}.$$

Let $n \geq 1$. We consider the optimization problem

$$\inf\left\{ \big(1 - 2^{\frac{1}{p}-1}\big) \left| \sum_{i=1}^{n} \lambda_i \right| + 2^{\frac{1}{p}-1} \sum_{i=1}^{n} |\lambda_i| : A \in \mathcal{H}_n^{(\beta)}, \sum_{i=1}^{n} \lambda_i^p = 1 \right\}.$$

Denote for all $\lambda \in \mathbb{R}^n$,

$$\varphi(\lambda) = \big(1 - 2^{\frac{1}{p}-1}\big) \left| \sum_{i=1}^{n} \lambda_i \right| + 2^{\frac{1}{p}-1} \sum_{i=1}^{n} |\lambda_i|.$$

Compactness and continuity arguments show that the infimum is achieved at some $\lambda \in \mathbb{R}^n$. At the price of permuting the coordinates of λ, and taking the opposite of λ, which does not change the value of $\varphi(\lambda)$, we can assume that $\lambda = (\lambda_1, \ldots, \lambda_m, 0, \ldots, 0)$, with $\lambda_1 \neq 0, \ldots, \lambda_m \neq 0$ such that $\sum_{i=1}^{m} \lambda_i \geq 0$. Assume first that $\sum_{i=1}^{m} \lambda_i > 0$. The multipliers rule (see [10, Theorem 9.1]) yields that there is some $\gamma > 0$, such that for any $i \in \{1, \ldots, m\}$,

$$\big(1 - 2^{\frac{1}{p}-1}\big) + 2^{\frac{1}{p}-1} \operatorname{sg}(\lambda_i) = \gamma \lambda_i^{p-1}. \tag{49}$$

Multiplying the above inequality by λ_i, and summing over all $i \in \{1, \ldots, m\}$, we get

$$\gamma = \varphi(\lambda). \tag{50}$$

From (49), we have for all $\in \{1, \ldots, m\}$,

$$\lambda_i = \begin{cases} \gamma^{-\frac{1}{p-1}} & \text{if } \lambda_i > 0, \\ -\gamma^{-\frac{1}{p-1}} (2^{\frac{1}{p}} - 1)^{\frac{1}{p-1}} & \text{if } \lambda_i < 0. \end{cases}$$

Let k denote the number of positive λ_i's, and l the number of negative λ_i's. As $\sum_{i=1}^{m} \lambda_i > 0$, we have $k \geq 1$. Since $\sum_{i=1}^{m} \lambda_i^p = 1$, we have

$$\gamma^{\frac{p}{p-1}} = k + l\big(2^{\frac{1}{p}} - 1\big)^{\frac{p}{p-1}} \geq 1,$$

as $k \geq 1$. Thus, $\varphi(\lambda) \geq 1$.

Assume now that $\sum_{i=1}^{m} \lambda_i = 0$. Then the multipliers rule asserts that there are some $t \in [-1, 1]$ and γ, such that $(t, \gamma) \neq (0, 0)$, and for all $i \in \{1, \ldots, m\}$,

$$\big(1 - 2^{\frac{1}{p}-1}\big)t + 2^{\frac{1}{p}-1} \operatorname{sg}(\lambda_i) = \gamma \lambda_i^{p-1}.$$

At the price of changing λ to $-\lambda$, we can assume $t \geq 0$. As in the previous case, multiplying by λ_i in the above equation and summing over i, yields $\varphi(\lambda) = \gamma$. Note that since $\varphi(1, 0, \ldots, 0) = 1$, we can assume $\gamma \leq 1$. We can write for any $i \in \{1, \ldots, m\}$,

$$\lambda_i = \begin{cases} -\gamma^{-\frac{1}{p-1}} (2^{\frac{1}{p}-1} - (1 - 2^{\frac{1}{p}-1})t)^{\frac{1}{p-1}} & \text{if } \lambda_i < 0, \\ \gamma^{-\frac{1}{p-1}} (2^{\frac{1}{p}-1} + (1 - 2^{\frac{1}{p}-1})t)^{\frac{1}{p-1}} & \text{if } \lambda_i > 0. \end{cases}$$

Let k denotes the number of positive coordinates of λ, and by l the number of negative coordinates. As $\sum_{i=1}^{m} \lambda_i = 0$, we have $k, l \geq 1$, and

$$k\big(2^{\frac{1}{p}-1} + \big(1 - 2^{\frac{1}{p}-1}\big)t\big)^{\frac{1}{p-1}} = l\big(2^{\frac{1}{p}-1} - \big(1 - 2^{\frac{1}{p}-1}\big)t\big)^{\frac{1}{p-1}}.$$

But then,

$$\varphi(\lambda) = 2^{\frac{1}{p}} k \gamma^{-\frac{1}{p-1}} \big(2^{\frac{1}{p}-1} + \big(1 - 2^{\frac{1}{p}-1}\big)t\big)^{\frac{1}{p-1}} \geq 2^{\frac{1}{p}} 2^{-\frac{1}{p}} = 1,$$

as $\gamma \leq 1$. As $\varphi(1, 0, \ldots, 0) = 1$, we can conclude

$$\inf\big\{\varphi(\lambda) : \|\lambda\|_p = 1\big\} = 1.$$

This yields,

$$c_p \geq \min\big(b, 2^{-\frac{\alpha}{p}} a\big),$$

in the case where p is even. $\qquad\square$

Acknowledgements

References

[1] G. W. Anderson, A. Guionnet and O. Zeitouni. *An Introduction to Random Matrices. Cambridge Studies in Advanced Mathematics* **118**. Cambridge University Press, Cambridge, 2010. MR2760897

[2] F. Augeri. Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails. *Electron. J. Probab.* **21** (2016) Paper No. 32, 49. MR3492936

[3] Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, 2nd edition. *Springer Series in Statistics*. Springer, New York, 2010. MR2567175

[4] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields* **108** (1997) 517–542. MR1465640

[5] R. Bhatia. *Matrix Analysis. Graduate Texts in Mathematics* **169**. Springer-Verlag, New York, 1997.

[6] S. G. Bobkov and M. Ledoux. From Brunn–Minkowski to Brascamp–Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.* **10** (5) (2000) 1028–1052. MR1800062

[7] C. Bordenave and P. Caputo. A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.* **42** (6) (2014) 2454–2496.

[8] C. Borell. On polynomial chaos and integrability. *Probab. Math. Statist.* **3** (2) (1984) 191–203.

[9] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős–Rényi random graph. *European J. Combin.* **32** (7) (2011) 1000–1017. MR2825532

[10] F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control. Graduate Texts in Mathematics* **264**. Springer, London, 2013.

[11] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications. Stochastic Modelling and Applied Probability* **38**. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[12] R. Dobrushin, P. Groeneboom and M. Ledoux *Lectures on Probability Theory and Statistics. Lecture Notes in Mathematics* **1648**. Springer-Verlag, Berlin, 1996. Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994. Edited by P. Bernard.

[13] H. Döring and P. Eichelsbacher. Moderate deviations in a random graph and for the spectrum of Bernoulli random matrices. *Electron. J. Probab.* **14** (92) (2009) 2636–2656.

[14] P. Eichelsbacher and J. Sommerauer. Moderate deviations for traces of words in a multi-matrix model. *Electron. Commun. Probab.* **14** (2009) 572–586.

[15] O. N. Feldheim and S. Sodin. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.* **20** (1) (2010) 88–123.

[16] N. Gantert, K. Ramanan and F. Rembart. Large deviations for weighted sums of stretched exponential random variables. *Electron. Commun. Probab.* **19** (41) (2014) 14.

[17] B. Groux. Asymptotic freeness for rectangular random matrices and large deviations for sample covariance matrices with sub-Gaussian tails. Available at arXiv:1505.05733 [math.PR].

[18] A. Guionnet. *Large Random Matrices: Lectures on Macroscopic Asymptotics*. *Lecture Notes in Mathematics* **1957**. Springer-Verlag, Berlin, 2009. Lectures from the 36th Probability Summer School held in Saint-Flour, 2006.

[19] D. Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** (1) (1982) 1–38.

[20] R. Latała. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.* **34** (6) (2006) 2315–2331.

[21] M. Ledoux. A note on large deviations for Wiener chaos. In *Séminaire de Probabilités XXIV, 1988/89* 1–14. *Lecture Notes in Math.* **1426**. Springer, Berlin, 1990.

[22] M. Ledoux. *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. American Mathematical Society, Providence, RI, 2001.

[23] M. Ledoux and M. Talagrand. *Probability in Banach Spaces. Isoperimetry and processes*. *Classics in Mathematics*. Springer-Verlag, Berlin, 2011. Reprint of the 1991 edition.

[24] P. Massart, G. Lugosi and S. Boucheron. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford, 2013. MR3185193

[25] M. W. Meckes and S. J. Szarek. Concentration for noncommutative polynomials in random matrices. *Proc. Amer. Math. Soc.* **140** (5) (2012) 1803–1811.

[26] S. V. Nagaev. Large deviations of sums of independent random variables. *Ann. Probab.* **7** (5) (1979) 745–789. MR0542129

[27] Ya. Sinai and A. Soshnikov. A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices. *Funktsional. Anal. i Prilozhen.* **32** (2) (1998) 56–79, 96.

[28] Ya. Sinai and A. Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Bull. Braz. Math. Soc. (N.S.)* **29** (1) (1998) 1–24.

[29] A. Soshnikov. Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** (3) (1999) 697–733.

[30] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* **67** (1958) 325–327.

[31] X. Zhan. *Matrix Inequalities*. *Lecture Notes in Mathematics* **1790**. Springer-Verlag, Berlin, 2002.