

NONLINEAR LARGE DEVIATION BOUNDS WITH APPLICATIONS TO WIGNER MATRICES AND SPARSE ERDŐS–RÉNYI GRAPHS

BY FANNY AUGERI

Faculty of Mathematics and Computer Science, Weizmann Institute of Science, fanny.augeri@weizmann.ac.il

We prove general nonlinear large deviation estimates similar to Chatterjee–Dembo’s original bounds, except that we do not require any second order smoothness. Our approach relies on convex analysis arguments and is valid for a broad class of distributions. Our results are then applied in three different setups. Our first application consists in the mean-field approximation of the partition function of the Ising model under an optimal assumption on the spectra of the adjacency matrices of the sequence of graphs. Next, we apply our general large deviation bound to investigate the large deviation of the traces of powers of Wigner matrices with sub-Gaussian entries and the upper tail of cycles counts in sparse Erdős–Rényi graphs down to the sparsity threshold $n^{-1/2}$.

1. Introduction. We will be concerned with the following large deviation question: given a random vector X in \mathbb{R}^n and a smooth function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, when is *shifting the mean of X* the *optimal* large deviation strategy for the random variable $f(X)$?

In a seminal paper [12], Chatterjee and Dembo provided a sufficient criterion and showed in the case where X is uniformly sampled on the discrete hypercube that, as soon as f has a gradient of *low complexity*, in the sense that it can be encoded by a small number of bits compared to the dimension, then the large deviations of $f(X)$ are only due to changes of the mean of X . Their framework encompasses a large class of large deviation problems as the mean-field approximation of the partition function of the Ising model (see [6]), the large deviation of subgraph counts in Erdős–Rényi graphs—which were until then tackled using the graphon formalism in [13]—and arithmetic progressions. Later, Yan generalized in [37] their nonlinear large deviation estimate to any compactly supported distribution.

As a consequence of powerful structure theorems for probability measures on the discrete hypercube, Eldan obtained in [20] nonlinear large deviation bounds, which do not require any second order smoothness on the given function and where the complexity of the gradient is assessed in terms of the Gaussian mean-width of its image. Using these, he partially recovered a known result of Basak and Mukherjee [6] on the mean-field approximation of the partition function of the Ising model and improved the threshold of sparsity for the large deviation upper tail of triangle counts.

We will improve the error terms in the nonasymptotic bounds of [12] and, in particular, remove the smoothness terms. The motivation behind these improvements is that it allows for obtaining weaker dimension dependence. In particular, this entails that one can consider large deviation speeds much smaller than the dimension. In turn, this allows us to reach the critical sparsity level, as identified in [33] and [8], in the large deviation bounds of cycle counts in sparse Erdős–Rényi graphs.

We will develop nonlinear large deviations estimates for general distributions and propose several applications. As a first example we will apply our results to the mean-field approximation of the partition function of the Ising model on a sequence of graphs, whose spectra

Received October 2018; revised December 2019.

MSC2020 subject classifications. 60F10, 60B20, 05C80, 52A40.

Key words and phrases. Large deviations, convex analysis, mean-field approximation, Wigner matrices, Erdős–Rényi graphs.

satisfy certain assumptions which include star graphs, thus strengthening the previous result of Basak and Mukherjee [6].

Next, we will use our nonlinear large deviation estimate to investigate the large deviation of traces of Wigner matrices with sub-Gaussian entries. Using a truncation argument to reduce the complexity of our function at hand, we show general upper and lower bounds. We will also prove the universality of the rate function for a class of Wigner matrices with *sharp sub-Gaussian tails*. This complements the results of Guionnet and Husson [21], who introduced this class and showed such universality for the large deviation of the largest eigenvalue.

Finally, we estimate the upper tail of the large deviations of cycle counts in sparse Erdős–Rényi graphs, down to the sparsity threshold $n^{-1/2}$, up to logarithmic corrections, improving in the case of triangles the recent result of Cook and Dembo [15].

1.1. *Main results.* We begin with introducing some definitions. Let μ be a probability measure on \mathbb{R}^n whose support is not included in a hyperplane. Let Λ_μ be the *logarithmic Laplace transform* of μ , that is,

$$\forall \lambda \in \mathbb{R}^n, \quad \Lambda_\mu(\lambda) = \log \int e^{\langle \lambda, x \rangle} d\mu(x).$$

It is known that Λ_μ is a strictly convex function which is C^∞ on the interior of its domain denoted by $\mathcal{D}_{\Lambda_\mu}$. We define Λ_μ^* , the *Legendre transform* of Λ_μ , by

$$\forall x \in \mathbb{R}^n, \quad \Lambda_\mu^*(x) = \sup_{\lambda \in \mathbb{R}^n} \{ \langle \lambda, x \rangle - \Lambda_\mu(\lambda) \}.$$

Following [34], Section 26, we say that a convex function $\Lambda : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is *essentially smooth* if the interior of its domain $\text{int}(\mathcal{D}_\Lambda)$ is nonempty, Λ is differentiable on $\text{int}(\mathcal{D}_\Lambda)$ and steep, that is, for any $\lambda_k \in \text{int}(\mathcal{D}_\Lambda)$, converging to a point on the boundary of \mathcal{D}_Λ ,

$$\lim_{k \rightarrow +\infty} \|\nabla \Lambda(\lambda_k)\| = +\infty.$$

The interest of this concept of essential smoothness lies in the fact that the Legendre transformation leaves invariant the class of strictly convex functions on \mathbb{R}^n which are essentially smooth. More precisely, assuming that Λ_μ is *essentially smooth*, then by [34], Theorem 26.5, we know that Λ_μ^* is also an essentially smooth function, the map $\nabla \Lambda_\mu$ is one-to-one onto $\text{int}(\mathcal{D}_{\Lambda_\mu^*})$, the interior of the domain of Λ_μ^* and the maps $\nabla \Lambda_\mu$ and $\nabla \Lambda_\mu^*$ are inverse maps from one another.

For any $\lambda \in \text{int}(\mathcal{D}_{\Lambda_\mu})$, one can define the measure

$$(1) \quad \mu_y = e^{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)} d\mu(x),$$

where $y = \nabla \Lambda_\mu(\lambda)$. Observe that differentiating Λ_μ , one obtains the barycenter of μ_y , that is,

$$(2) \quad y = \nabla \Lambda_\mu(\lambda) = \int x d\mu_y(x).$$

Thus, the family of measures μ_y represents the collection of tilts of μ , indexed by their barycenters.

Our first result is the following nonasymptotic version of Varadhan’s lemma (see [18], Theorem 4.3.1). This will be instrumental in our treatment of the mean-field behavior of the Ising model; see Section 1.2.1.

THEOREM 1.1. *Assume μ is compactly supported. Denote by K the convex hull of its support, and by D its diameter. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuously differentiable function. Let $\delta > 0$ and \mathcal{D}_δ be a δ/D -net of the convex hull of $\nabla f(K)$ for the ℓ^2 -norm. Then,*

$$\log \int e^f d\mu \leq \sup_{y \in \mathbb{R}^n} \{ f(y) - \Lambda_\mu^*(y) \} + \log |\mathcal{D}_\delta| + \delta.$$

Our approximation of the partition function is similar to the one of Chatterjee–Dembo (see [12], Theorem 1.6), except that we do not have any other error terms but the one coming from the cardinality of the net \mathcal{D}_δ . Unlike Eldan’s result (see [20], Corollary 2), which involves the *Gaussian mean width* of the set of gradients, we are assessing the complexity of the gradient in terms of covering numbers. Using Sudakov inequality (see [32], Theorem 3.18), it is possible to compare [20], Corollary 2, with Theorem 1.1. More precisely, we have the following result as a corollary of Theorem 1.1.

COROLLARY 1.2. *Assume μ is compactly supported. Denote by K the convex hull of its support, assume that the diameter of K is \sqrt{n} . Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuously differentiable function and let $V = \nabla f(K)$. There exists a numerical constant $\kappa > 0$ such that*

$$\log \int e^f d\mu \leq \sup_{y \in \mathbb{R}^n} \{f(y) - \Lambda_\mu^*(y)\} + \kappa n^{\frac{1}{3}} g(V)^{\frac{2}{3}},$$

where $g(V)$ is the *Gaussian mean-width* of V , defined as

$$g(V) = \mathbb{E} \sup_{\zeta \in V} \langle \zeta, \Gamma \rangle,$$

with Γ being a standard Gaussian random variable in \mathbb{R}^n .

PROOF. From Theorem 1.1 we have, for any $\delta > 0$,

$$\log \int e^f d\mu \leq \sup_{y \in \mathbb{R}^n} \{f(y) - \Lambda_\mu^*(y)\} + \log N(\text{co}(V), (\delta/\sqrt{n})B_{\ell^2}) + \delta,$$

where $\text{co}(V)$ denotes the convex hull of V and $N(\text{co}(V), (\delta/\sqrt{n})B_{\ell^2})$ the covering number of $\text{co}(V)$ by $(\delta/\sqrt{n})B_{\ell^2}$. By Sudakov inequality (see [32], Theorem 3.18), we know that there exists a numerical constant $c > 0$ such that

$$\log N(\text{co}(V), (\delta/\sqrt{n})B_{\ell^2}) \leq c \frac{ng(\text{co}(V))^2}{\delta^2}.$$

But as the supremum of a convex function on a set is equal to the supremum on its convex hull, we have $g(\text{co}(V)) = g(V)$. Thus,

$$\log \int e^f d\mu \leq \sup_{y \in \mathbb{R}^n} \{f(y) - \Lambda_\mu^*(y)\} + c \frac{ng(V)^2}{\delta^2} + \delta.$$

Optimizing in $\delta > 0$, we obtain that there exists a numerical constant $\kappa > 0$ such that

$$\log \int e^f d\mu \leq \sup_{y \in \mathbb{R}^n} \{f(y) - \Lambda_\mu^*(y)\} + \kappa n^{\frac{1}{3}} g(V)^{\frac{2}{3}}. \quad \square$$

As for the lower bound, it is a classical fact to note that, as a consequence of Jensen’s inequality, one always has the following lower bound by just changing the barycenter of the underlying reference measure.

LEMMA 1.3. *Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a nonnegative measurable function. Then,*

$$\log \int e^f d\mu \geq \sup \left\{ \int f d\mu_y - \Lambda_\mu^*(y), y \in \nabla \Lambda_\mu(\text{int}(\mathcal{D}_{\Lambda_\mu})) \right\},$$

where μ_y is defined in (1).

REMARK 1.1. Note the gap between the lower bound of Lemma 1.3 and the main term $\sup\{f - \Lambda_\mu^*\}$ in the upper bound of Proposition 1.1.

When μ is the uniform measure on the discrete hypercube, one can consider, given a function $f : \{0, 1\}^n \rightarrow \mathbb{R}$, its natural harmonic extension, and the just mentioned gap resolves itself. One can view this fact as at the heart of Eldan’s approach [20]. For further discussion on the general case, we refer the reader to Remark 2.2.

We will now state our nonlinear large deviations upper bound. We recall from [18], Section 1.2, 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} , satisfies a *large deviations principle* (LDP) with *speed* v_n , and *rate function* $I : \mathcal{X} \rightarrow [0, +\infty]$, if I is lower semicontinuous, v_n increases to infinity, and for all $B \in \mathcal{B}$,

$$(3) \quad \begin{aligned} - \inf_{\text{int}(B)} I &\leq \liminf_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(Z_n \in B) \\ &\leq \limsup_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(Z_n \in B) \leq - \inf_{\text{cl}(B)} I, \end{aligned}$$

where $\text{int}(B)$ denotes the interior of B and $\text{cl}(B)$ the closure of B . We recall that I is *lower semicontinuous* if its t -level sets $\{x \in \mathcal{X} : I(x) \leq t\}$ are closed, for any $t \in [0, +\infty)$. Furthermore, if all the level sets are compact, then I is said to be a *good rate function*.

By [18], (1.2.7), for good rate functions I , the large deviations upper bound, that is the right-most inequality in (3), is equivalent to the statement that for any $r > 0$ and $\delta > 0$,

$$\limsup_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(Z_n \notin V_\delta(\{I \leq r\})) \leq -r,$$

where for any subset $A \subset \mathbb{R}$, $V_\delta(A)$ denotes the *open δ -neighborhood of A* , that is,

$$V_\delta(A) = \left\{ x \in \mathbb{R} : \inf_{y \in A} |x - y| < \delta \right\}.$$

With that in mind, the following proposition is our general large deviation upper bound.

THEOREM 1.4. *Let X be a random vector in \mathbb{R}^n with law μ such that Λ_μ is essentially smooth. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a measurable function. Define,*

$$\forall t \in \mathbb{R}, \quad I(t) = \inf\{\Lambda_\mu^*(y) : f(y) = t, y \in \mathbb{R}^n\}.$$

Let $r > 0$. Assume there exists $\kappa \geq r$ such that denoting by $K = \{\Lambda_\mu^ \leq \kappa\}$,*

$$(4) \quad \mathbb{P}(X \notin K) \leq e^{-r}.$$

Let $W \subset \mathbb{R}^n$ be a compact convex set such that,

$$(5) \quad \forall x, y \in K, \quad f(x) - f(y) \leq \sup_{\lambda \in W} \langle \lambda, x - y \rangle.$$

Denote by D the diameter of K . Let $\delta > 0$ and let $\mathcal{D}_{\delta/2}$ be a $\delta/2D$ -net of W for the ℓ^2 -norm. Then,

$$\log \mathbb{P}(f(X) \notin V_\delta(\{I \leq r\})) \leq -r + \log |\mathcal{D}_{\delta/2}| + \log \left(\left(\frac{\kappa L D}{\delta} \right) \vee 1 \right) + 2,$$

where $L = \sup_{\lambda \in V} \|\lambda\|_{\ell^2}$.

REMARK 1.2. The wording of Theorem 1.4 is intended to include nonsmooth functions which we will encounter in the applications. It is particularly relevant for functions f which are linear combinations of nonsmooth convex functions, for which we can take W to be Minkowski sums of sets of subdifferentials of the convex functions involved in the decomposition.

REMARK 1.3. The tightness assumption (4) in the Proposition 1.4 is automatically satisfied for product measures. Indeed, we know from [18], Lemma 5.1.14, that if μ is a probability measure on \mathbb{R} , then for any $\alpha \in (0, 1)$,

$$\mathbb{E}e^{\alpha\Lambda_\mu^*(X_1)} \leq \frac{2}{1-\alpha}.$$

Taking $\alpha = 1/2$, we deduce that for $\kappa = 12(r \wedge n)$, Chernoff’s inequality gives

$$\mathbb{P}(\Lambda_{\mu^n}^*(X) > \kappa) \leq e^{-r}.$$

As a consequence of Theorem 1.4, we obtain a nonlinear large deviations estimate for the upper tail of a function with a low-complexity gradient.

COROLLARY 1.5. *Let X be a random vector in \mathbb{R}^n with law μ such that Λ_μ is essentially smooth. Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a measurable function. Define,*

$$\forall t \in \mathbb{R}, \quad \phi(t) = \inf\{\Lambda_\mu^*(y) : f(y) \geq t, y \in \mathbb{R}^n\}.$$

Let $t, \delta > 0$ and assume that ϕ is increasing on $[t - \delta, +\infty)$. Let $\kappa \geq \phi(t - \delta)$ such that denoting by $K = \{\Lambda_\mu^ \leq \kappa\}$,*

$$(6) \quad \mathbb{P}(X \notin K) \leq e^{-\phi(t-\delta)}.$$

Let $W \subset \mathbb{R}^n$ be a compact convex set such that,

$$\forall x, y \in K, \quad f(x) - f(y) \leq \sup_{\lambda \in W} \langle \lambda, x - y \rangle.$$

Denote by D the diameter of K and let $\mathcal{D}_{\delta/2}$ be a $\delta/2D$ -net of W for the ℓ^2 -norm. Then,

$$\log \mathbb{P}(f(X) \geq t) \leq -\phi(t - \delta) + \log |\mathcal{D}_{\delta/2}| + \log \left(\left(\frac{\kappa LD}{\delta} \right) \vee 1 \right) + 2,$$

where $L = \sup_{\lambda \in V} \|\lambda\|_{\ell^2}$.

PROOF. We claim that our assumption on the strict monotonicity of ϕ yields that

$$(7) \quad \{I \leq \phi(t - \delta)\} \subset (-\infty, t - \delta].$$

Indeed, for any $z \in \mathbb{R}$, if $I(z) \leq \phi(t - \delta)$, then in particular $\phi(z) \leq \phi(t - \delta)$. Since ϕ is increasing on $[t - \delta, +\infty)$, we have $z \leq t - \delta$, which shows (7). Therefore, if $x \in \mathbb{R}^n$ is such that $f(x) \geq t$ then $f(x) \notin V_\delta(\{I \leq \phi(t - \delta)\})$. Thus, we have

$$\mathbb{P}(f(X) \geq t) \leq \mathbb{P}(f(X) \notin V_\delta(\{I \leq \phi(t - \delta)\})),$$

which, by Theorem 1.4, ends the proof. \square

Turning our attention to the lower bound, one can always get a large deviations lower bound by finding a tilting of the underlying measure μ which transforms a large deviations event for μ into a typical event for μ_h . (This idea is prevalent in the large deviations literature, and is also the basis for the proof of Lemma 1.3.) The following lower bound, which builds

on [19], page 64, quantifies this idea. First, note that differentiating twice Λ_μ , one obtains the covariance of the measure μ_y , that is,

$$\forall \eta \in \mathbb{R}^n, \quad \text{Var}_{\mu_y} \langle \eta, X \rangle = \langle \eta, \nabla^2 \Lambda_\mu(\lambda) \cdot \eta \rangle,$$

where $y = \nabla \Lambda_\mu(\lambda)$. With this notation, one obtains the following lemma.

LEMMA 1.6. *Let $V \subset \mathbb{R}^n$ be a measurable subset, and $\lambda \in \text{int}(\mathcal{D}_{\Lambda_\mu})$. Then,*

$$\mu(V) \geq e^{-\Lambda_\mu^*(y)} \mu_y(V) \exp\left(-\frac{1}{\mu_y(V)^{1/2}} \langle \lambda, \nabla^2 \Lambda_\mu(\lambda) \cdot \lambda \rangle^{1/2}\right),$$

where $y = \nabla \Lambda_\mu(\lambda)$, and μ_y is defined in (1).

PROOF. Let $\lambda \in \text{int}(\mathcal{D}_{\Lambda_\mu})$ and $y = \nabla \Lambda_\mu(\lambda)$. Changing the measure μ into μ_y , we have

$$\mu(V) = \int_V e^{-\langle \lambda, x \rangle - \Lambda_\mu(\lambda)} d\mu_y(x) = e^{-\Lambda_\mu^*(y)} \int_V e^{-\langle \lambda, x - y \rangle} d\mu_y(x),$$

where we used the fact that $\Lambda_\mu^*(y) = \langle \lambda, y \rangle - \Lambda_\mu(\lambda)$. By Jensen’s inequality we deduce,

$$\mu(V) \geq e^{-\Lambda_\mu^*(y)} \mu_h(V) \exp\left(-\frac{1}{\mu_y(V)} \int_V \langle \lambda, x - y \rangle d\mu_y(x)\right).$$

Using the Cauchy–Schwarz inequality, we get,

$$\int_V \langle \lambda, x - y \rangle d\mu_y(x) \leq \mu_y(V)^{1/2} \left(\int_V \langle \lambda, x - y \rangle^2 d\mu_y(x) \right)^{1/2}$$

which yields the claim. \square

1.2. *Applications.* In this section we provide several examples in which the gap between the upper bounds (Propositions 1.1 and 1.4) and lower bounds (Lemmas 1.3 and 1.6) is negligible in the large deviation scale.

1.2.1. *Mean-field approximation in the Ising model.* Our first example of application is the accuracy of the mean-field prediction in the Ising model with large degree. This topic is discussed in [6], [20], Section 1.3, [23, 24], and we will improve on some of the results therein.

In particular, we show that the mean-field approximation of the Ising model holds true as soon as the empirical distributions of the eigenvalues of the interaction matrices converge weakly to a Dirac at 0, and the second moments are uniformly bounded.

Let \mathcal{H}_n denote the set of Hermitian matrices of size n . For any $A \in \mathcal{H}_n$, we denote by μ_A the empirical distribution of its eigenvalues, defined by

$$\mu_A = \frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i},$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A .

PROPOSITION 1.7. *Let A_n be a sequence of Hermitian matrices such that, for any $i \in \{1, \dots, n\}$, $(A_n)_{i,i} = 0$. Assume that the empirical distribution of eigenvalues μ_{A_n} converges weakly to a Dirac at 0 and that $\limsup_n n^{-1} \text{tr} A_n^2 < +\infty$. Let μ be the uniform probability measure on $\{-1, 1\}^n$. Then,*

$$\log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) = \sup_{x \in \{-1, 1\}^n} \{ \langle x, A_n x \rangle - \Lambda_\mu^*(x) \} + o(n).$$

PROOF. Note that the lower bound,

$$\log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) \geq \sup_{x \in [-1, 1]^n} \{ \langle x, A_n x \rangle - \Lambda_\mu^*(x) \},$$

follows directly Lemma 1.3. Let

$$\forall x \in \mathbb{R}^n, \quad f(x) = \langle x, A_n x \rangle.$$

The following lemma taken from [6], Lemma 3.4, Remark 4.1, computes the complexity of the gradient of f under the mean-field assumption we made.

LEMMA 1.8 ([6], Lemma 3.4, Remark 4.1). *Under the assumption of Proposition 1.7, for any $\delta > 0$ there exists a $\delta\sqrt{n}$ -net $\mathcal{E}_{\delta\sqrt{n}}$ for the ℓ^2 -norm of the set $A_n([-1, 1]^n)$ such that*

$$\log |\mathcal{E}_{\delta\sqrt{n}}| = o(n).$$

As $\nabla f(x) = 2A_n x$ for any $x \in [-1, 1]^n$, applying Proposition 1.1 and using the above lemma, we deduce that, for any $\delta > 0$,

$$(8) \quad \log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) \leq \sup_{x \in [-1, 1]^n} \{ \langle x, A_n x \rangle - \Lambda_\mu^*(x) \} + \log |\mathcal{E}_{\delta\sqrt{n}}| + 4\delta n$$

which gives the claim. \square

REMARK 1.4. Instead of using Lemma 1.8 to bound the error term in (8), one can use alternatively Corollary 1.2, which gives

$$\log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) \leq \sup_{x \in [-1, 1]^n} \{ \langle x, A_n x \rangle - \Lambda_\mu^*(x) \} + \kappa n^{\frac{1}{3}} \left(\mathbb{E} \sup_{x \in [-1, 1]^n} \langle A_n x, \Gamma \rangle \right)^{\frac{2}{3}},$$

where $\kappa > 0$ and Γ is a standard Gaussian vector in \mathbb{R}^n . But, using twice the Cauchy–Schwarz inequality and the fact that $\mathbb{E} \|A_n \Gamma\|^2 = \|A_n\|_2$, one obtains

$$\mathbb{E} \sup_{x \in [-1, 1]^n} \langle A_n x, \Gamma \rangle \leq \sqrt{n} \|A_n\|_2,$$

where $\|\cdot\|_2$ denotes the Hilbert–Schmidt norm. Thus,

$$\log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) \leq \sup_{x \in [-1, 1]^n} \{ \langle x, A_n x \rangle - \Lambda_\mu^*(x) \} + \kappa (\|A_n\|_2 n)^{\frac{2}{3}},$$

thus giving another proof of the result of Jain, Koehler and Risteski [24].

1.2.2. *Traces of powers of Wigner matrices.* In this section we discuss the large deviations of the normalized traces of powers of Wigner matrices with sub-Gaussian entries. Under technical assumptions we show that we can find large deviation upper and lower bounds, which match in the case where the entries have *sharp sub-Gaussian tails* and have the same covariance structure as the GOE or the GUE. In the latter case the large deviations are universal, that is, the resulting rate function is the same for all such Wigner matrices and coincides with the Gaussian case. This universality phenomenon was first discovered by Guionnet and Husson in [21] in the context of the large deviations of the largest eigenvalue of Wigner matrices.

To state our result, we need to introduce some notation. Denote by $\mathcal{H}_n^{(\beta)}$ the set of Hermitian matrices when $\beta = 2$, and symmetric matrices when $\beta = 1$, of size n , which we will shorthand as \mathcal{H}_n whenever there is no ambiguity. Say that X is a *Wigner matrix* if X is a random Hermitian matrix with independent coefficients (up to the symmetry) such that both

$(X_{i,i})_{1 \leq i \leq n}$ and $(X_{i,j})_{i < j}$ are identically distributed, $\mathbb{E}X = 0$, $(\Re X_{1,2}, \Im X_{1,2})$ are independent and $\mathbb{E}|X_{1,2}|^2 = 1$. By Wigner’s theorem (see [1], Theorem 2.1.1, Exercise 2.1.16, [5], Theorem 2.5), we know that

$$\mu_{X/\sqrt{n}} \underset{n \rightarrow \infty}{\rightsquigarrow} \mu_{\text{sc}}$$

in probability, where \rightsquigarrow denotes the weak convergence and μ_{sc} is the semicircular law defined by

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

If we assume moreover that, for any $d \in \mathbb{N}$,

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

then we have by [1], Lemmas 2.1.6, 2.1.7, the convergence of the moments of $\mu_{X/\sqrt{n}}$ toward the ones of the semicircular law, that is, for any $d \in \mathbb{N}$,

$$\frac{1}{n} \text{tr}(X/\sqrt{n})^d \xrightarrow{n \rightarrow +\infty} \mu_{\text{sc}}(x^d)$$

in probability, where $\mu_{\text{sc}}(x^d) = \int x^d d\mu_{\text{sc}}(x)$. It is known that

$$\mu_{\text{sc}}(x^d) = \begin{cases} C_{d/2} & \text{if } d \text{ is even,} \\ 0 & \text{if } d \text{ is odd,} \end{cases}$$

where in the case d is even, $C_{d/2} = \frac{1}{d/2+1} \binom{d}{d/2}$ is the $(\frac{d}{2})$ th Catalan number. The question we investigate is the one of the large deviations of the these traces around the moments of the semicircular law.

We will actually prove a large deviation result under a more restrictive assumption than just the one of having sub-Gaussian entries, as we will need to use further concentration arguments in addition to the result of Theorem 1.4. To this end, we introduce a class of random Hermitian matrices satisfying a certain *convex concentration property*.

DEFINITION 1.9. We say that a random Hermitian matrix X satisfies the *convex concentration property* if there exists $c > 0$ such that, if for any $f : \mathcal{H}_n \rightarrow \mathbb{R}$ convex 1-Lipschitz function with respect to the Hilbert–Schmidt norm, $\mathbb{E}f(X)$ exists and, for any $t > 0$,

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq c^{-1} \exp(-ct^2).$$

REMARK 1.5. A random Hermitian matrix whose law satisfies a log-Sobolev inequality has normal concentration by [31], Theorem 5.3. In particular, such random matrices satisfy the above convex concentration property. Workable criteria for log-concave measures to satisfy a log-Sobolev inequality include the strict uniform convexity of the potential; see [31], Theorem 5.2. Due to the tensorization property of the log-Sobolev inequality (see [31], Corollary 5.7), any Wigner matrix with entries satisfying a log-Sobolev inequality has the convex concentration property.

Another important family of Wigner matrices satisfying this property is the one with bounded entries, due either to Talagrand’s inequality [31], Corollary 4.10, or due to a result by the transportation method [9], Theorem 8.6 (the latter having the advantage of directly measuring deviations from the mean).

In the following, we denote by Λ the logarithmic Laplace transform of the law of X , that is,

$$\forall H \in \mathcal{H}_n, \quad \Lambda(H) = \log \mathbb{E} e^{\text{tr}(XH)}$$

and by $\Lambda_{i,j}$ the one of the law of each of its entries $X_{i,j}$. Finally, we denote by Λ^* the Legendre transform of Λ .

Using Proposition 1.4, together with a truncation argument which enables us to reduce ourself to a function with a low complexity gradient, we obtain the following large deviation result for traces of Wigner matrices.

THEOREM 1.10. *Let $d \geq 3$, and $\ell > d$ be an even integer. Let X be a Wigner matrix satisfying the convex concentration property such that $\Lambda_{1,1}$ and $\Lambda_{1,2}$ have their derivatives of order 2 to ℓ uniformly bounded. For any closed subset F of \mathbb{R} ,*

$$\limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P} \left(\frac{1}{n} \text{tr}(X/\sqrt{n})^d \in F \right) \leq - \inf_F I_+,$$

and for any open subset O of \mathbb{R} ,

$$\liminf_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P} \left(\frac{1}{n} \text{tr}(X/\sqrt{n})^d \in O \right) \geq - \inf_O I_-,$$

where, for any $x \in \mathbb{R}$,

$$I_+(x) = \sup_{\delta > 0} \liminf_{n \rightarrow +\infty} I_{n,\delta}(x),$$

$$I_-(x) = \sup_{\delta > 0} \limsup_{n \rightarrow +\infty} I_{n,\delta}(x),$$

and, for any $n \in \mathbb{N}$, $\delta > 0$,

$$I_{n,\delta}(x) = \inf \left\{ \frac{\Lambda^*(Y)}{n^{1+\frac{2}{d}}} : \left| \frac{1}{n} \text{tr}(Y/\sqrt{n})^d + \mu_{\text{sc}}(x^d) - x \right| < \delta, Y \in \mathcal{H}_n \right\}.$$

REMARK 1.6. Together with the previous Remark 1.5, we see that any Wigner matrix with bounded entries satisfies the assumptions of Theorem 1.10.

We will now specify our result to Wigner matrices with sharp sub-Gaussian tails. Following Guionnet and Husson [21], we say that a real or complex random variable ξ has *sharp sub-Gaussian tails* if

$$\forall z \in \mathbb{C}, \quad \Lambda_\xi(z) = \log \mathbb{E} e^{\Re(z\xi)} \leq \frac{1}{2} \langle z, \Sigma z \rangle,$$

where Σ is the covariance matrix of $(\Re \xi, \Im \xi)$ and $\langle \cdot, \cdot \rangle$ is the standard inner product in \mathbb{C} .

COROLLARY 1.11. *Let $d \in \mathbb{N}$, $d \geq 3$. Let X be a Wigner matrix such that $\mathbb{E} X_{1,1}^2 \leq 2/\beta$. Assume that X is real symmetric if $\beta = 1$ and, if $\beta = 2$, that $(\Re X_{1,2}, \Im X_{1,2})$ are independent each of variance $1/2$. In addition to the assumptions of Theorem 1.10, assume that the entries of X have sharp sub-Gaussian tails. The sequence $(\frac{1}{n} \text{tr}(X/\sqrt{n})^d)_{N \in \mathbb{N}}$ satisfies a LDP with speed $n^{1+\frac{2}{d}}$ and good rate function J_d . If d is even, J_d is given by*

$$\forall x \in \mathbb{R}, \quad J_d(x) = \begin{cases} \frac{\beta}{4} (x - C_{d/2})^{\frac{2}{d}} & \text{if } x \geq C_{d/2}, \\ +\infty & \text{otherwise,} \end{cases}$$

where $C_{d/2}$ denotes the $(\frac{d}{2})$ th Catalan number and, if d is odd,

$$\forall x \in \mathbb{R}, \quad J_d(x) = \frac{\beta}{4}|x|^{\frac{2}{d}},$$

where $\beta = 1$ if X is real symmetric and $\beta = 2$ if X is complex Hermitian.

REMARK 1.7. Distributions which have sharp sub-Gaussian tails include the Rademacher distribution $\frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$ and uniform probability measures on intervals symmetric around the 0; see [21], Examples 1.2. Therefore, the above LDP holds in particular for Wigner matrices with Rademacher entries or uniformly sampled in $[-\sqrt{3}, \sqrt{3}]$.

1.2.3. *Upper tail of cycle counts in sparse Erdős–Rényi graphs.* Let X be the adjacency matrix of an Erdős–Rényi graph on n vertices with parameter $p_n = p$. More precisely, we assume that X is a symmetric matrix such that $(X_{i,j})_{i < j}$ are independent Bernoulli random variables of parameter p , whereas $X_{i,i} = 0$ for any $i \in \{1, \dots, n\}$.

For a given finite graph H , denote by X_H the number of copies of H in the Erdős–Rényi graph. A general question is to understand the large deviation of X_H of the order of its expectation when n goes to $+\infty$. When p is fixed, the large deviation of these subgraph counts are now well understood, due to the work of Chatterjee and Varadhan [13] on the large deviations of the Erdős–Rényi graph for the cut metric.

When $p \ll 1$, a first question is to estimate the order of the upper tail, that is, for $u \geq 1$,

$$\mathbb{P}(X_H \geq u\mathbb{E}X_H).$$

In the case of triangles $H = K_3$, it was proven in a series of papers [11, 16, 28, 36], that, for any $p \geq 1/n$,

$$\log \mathbb{P}(X_{K_3} \geq u\mathbb{E}X_{K_3}) \asymp -\min(n^2 p^2 \log(1/p), n^3 p^3),$$

and then generalized in [17] for cliques of arbitrary size. For general graphs H the order of the upper tail has been computed up to some $\log(1/p)$ factor, by Janson, Oleszkiewicz and Ruciński in [26]. In particular, denoting by Δ is the highest degree in H , they proved in [26], Theorems 1.2, 1.5, that

$$-n^2 p^\Delta \log(1/p) \lesssim \log \mathbb{P}(X_H \geq u\mathbb{E}X_H) \lesssim -n^2 p^\Delta$$

for $p \geq n^{-1/\Delta}$. Still, the exact order of this tail is not fully understood for small edge probability, as the order conjectured in [17] by Kahn and DeMarco was recently disproved by Šileikis and Warnke in [35].

Working instead with homomorphism densities, that is, if H has vertex set V and edge set E ,

$$t(H, X) = \frac{1}{n^{|V|}} \sum_{\underline{i} \in \{1, \dots, n\}^V} \prod_{(v,w) \in E} X_{\underline{i}(v), \underline{i}(w)},$$

Chatterjee and Dembo showed in [12] that the large deviations of $t(H, X)$ fall into their framework of nonlinear large deviations. The large deviation upper tail is understood by a certain variational problem,

$$(9) \quad -\log \mathbb{P}(t(H, X) \geq u\mathbb{E}t(H, X)) \sim \phi_{n,H}(u),$$

with

$$(10) \quad \phi_{n,H}(u) = \inf\{\Lambda_p^*(Y) : t(H, Y) \geq u\mathbb{E}t(H, Y), Y \in \mathcal{H}_n^0\},$$

where \mathcal{H}_n^0 is the set of symmetric matrices of size n with null diagonal coefficients and $\Lambda_p^*(Y) = \sum_{i < j} I_p(Y_{i,j})$, where

$$\forall x \in [0, 1], \quad I_p(x) = x \log \frac{x}{p} + (1 - x) \log \frac{1 - x}{1 - p},$$

and $+\infty$ otherwise. The equivalent (9) was shown to hold in [12], Theorem 1.2, for $p \geq n^{-\alpha(H)}$, for some explicit constant $\alpha(H) > 0$ depending on H . In the case where $H = K_3$, the threshold of validity of (9) found was $n^{-1/42}$, up to logarithmic factor, and was pushed later on by Eldan in [20] to $n^{-1/18}$. Yet neither of these thresholds are expected to be optimal, and in this case of triangle counts it is conjectured that (9) holds as soon as the variational problem (10) gives the right order of the upper tail, that is, with the help of [33], $np \gg \log n$.

Very recently, Cook and Dembo [15] proved that the nonlinear large deviations of $t(H, X)$ hold for the range $n^{-1/(5\Delta-4)} \ll p \ll 1$ and, more strongly, in the case of d -cycles counts for $n^{-1/2} \log n \ll p \ll 1$ when $d \geq 4$, and $n^{-1/3} \ll p \ll 1$ when $d = 3$.

We will give an alternative proof of their result on cycle counts. In the following proposition we push the estimation of the upper tail for sparsity parameters satisfying $n^{-1/2} \log^4 n \ll p \ll 1$, thus improving on Cook and Dembo’s result in the case $d = 3$.

PROPOSITION 1.12. *Let p such that $p = o(1)$ and $\log^4 n = o(np^2)$. Denote by $v_n = n^2 p^2 \log(1/p)$. For any $t \geq 1$,*

$$\log \mathbb{P}(\text{tr}(X^d) \geq tn^d p^d) \leq -\phi_n(t) + o(v_n),$$

where

$$\phi_n(t) = \inf\{\Lambda_p^*(Y) : \text{tr}(Y^d) \geq t(np)^d, Y \in \mathcal{H}_n^0\}.$$

The proof, as for the traces of Wigner matrices, relies on a truncation argument which enables us to lower the complexity of our function of interest and efficiently apply Proposition 1.4.

In [8], Theorem 1.5, (or [33], Theorem 1.1, for the case $d = 3$), the variational problem was solved asymptotically. They showed that, for $np \gg 1$,

$$(11) \quad \frac{\phi_n(t)}{v_n} \xrightarrow{n \rightarrow +\infty} \Phi(t),$$

where Φ is given by

$$(12) \quad \Phi(t) = \begin{cases} \min\left(\theta_t, \frac{1}{2}(t-1)^{\frac{2}{d}}\right) & \text{if } p \gg n^{-1/2}, \\ \frac{1}{2}(t-1)^{\frac{2}{d}} & \text{if } n^{-1} \ll p \ll n^{-1/2}, \end{cases}$$

with θ_t the unique solution of the equation $P_{C_d}(\theta_t) = t$ where P_{C_d} is the independence polynomial of the d -cycle. The optimizers correspond on one hand to planting a clique of order $np(t-1)^{1/d}$, which gives the $\frac{1}{2}(t-1)^{2/d}$ term, and on the other hand to planting an anticlique of order $np^2\theta_t$ which corresponds to the other θ_t term.

With this knowledge of the optimizers, one can obtain the complementary lower bound so that, together with Proposition 1.12, we get, for any $t \geq 1$,

$$\lim_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(\text{tr}(X^d) \geq tn^d p^d) = -\Phi(t)$$

for any $n^{-1/2} \log^4 n \ll p \ll 1$. The edge probability $n^{-1/2}$ appears to be critical for cycles in different ways. Note that, below this threshold, the anticlique construction is no longer

available. Moreover, for this parameter the order at the exponential scale of the upper tail approaches n . From the point of view of the truncation method we are using, the speed n arises as being impassable (without further dimension reduction): the complexity of the gradient of a truncated trace is at least the one of the $(n - 1)$ -dimensional sphere, as it corresponds to pick the top eigenvector.

Furthermore, the threshold $n^{-1/2}$ is also critical in the sense that one observes a change of speed in the large deviation lower tail of cycles counts. Indeed, we know by the work of Janson, Łuczak and Ruciński [25] and Janson and Warnke [27] that in the regime where $p = o(1)$, for any finite graph H and any $0 < u \leq 1$,

$$\log \mathbb{P}(X_H \leq u \mathbb{E}X_H) \asymp -(1 - u)^2 \Phi_H,$$

where $\Phi_H = \min\{\mathbb{E}X_J : J \text{ subgraph of } H\}$. Thus, the speed of the lower tail is driven by the “least expected” subgraph of H . For example, for $H = K_3$, this implies that

$$\log \mathbb{P}(X_H \leq (1 - u)\mathbb{E}X_H) \asymp -u^2 \min(n^3 p^3, n^2 p).$$

Thus, a change of speed happens at $n^{-1/2}$ and, similarly, for cycles of greater lengths.

In [15], Theorem 1.2, the authors provided sharp lower-tail estimates of homomorphisms densities of d -cycles for sparsity parameters $p \gg n^{-c}$ (up to logarithmic corrections), with $c = \frac{d-2}{2(d-1)}$. We mention also that an entropic perspective on the estimation of the lower tail of triangle counts has been announced by Kozma and Samotij [29] all the way down to $p \gg n^{-1/2}$.

2. Nonlinear large deviation upper bounds. We give here a proof of Theorems 1.1 and 1.4.

PROOF OF PROPOSITION 1.1. Let $x, y \in K$. We have, by the mean-value theorem,

$$\begin{aligned} f(x) &= f(x) - f(y) + \Lambda_\mu^*(y) + f(y) - \Lambda_\mu^*(y) \\ &\leq \sup_{t \in K} \langle \nabla f(t), x - y \rangle + \Lambda_\mu^*(y) + \sup_{z \in \mathbb{R}^n} \{f(z) - \Lambda_\mu^*(z)\}. \end{aligned}$$

Denote by W the convex hull of $\nabla f(K)$ and h_W its support function, that is, $h_W(y) = \sup_{\lambda \in W} \langle \lambda, y \rangle$ for any $y \in \mathbb{R}^n$. With this notation we get

$$f(x) \leq h_W(x - y) + \Lambda_\mu^*(y) + \sup_{z \in \mathbb{R}^n} \{f(z) - \Lambda_\mu^*(z)\}.$$

Optimizing in y , we deduce that

$$f(x) \leq \inf_{y \in K} \{h_W(x - y) + \Lambda_\mu^*(y)\} + \sup_{z \in \mathbb{R}^n} \{f(z) - \Lambda_\mu^*(z)\}.$$

Using the minimax theorem (see [14], Theorem 4.36), we have

$$\begin{aligned} \inf_{y \in K} \{h_W(x - y) + \Lambda_\mu^*(y)\} &= \sup_{\lambda \in W} \inf_{y \in K} \{\langle \lambda, x - y \rangle + \Lambda_\mu^*(y)\} \\ &= \sup_{\lambda \in W} \{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)\}, \end{aligned}$$

where we used the fact that $\Lambda_\mu^*(y) = +\infty$, if $y \notin K$ by the Hahn–Banach theorem, and that Λ_μ is the Legendre transform of Λ_μ^* by [14], Theorem 4.21. Therefore,

$$f(x) \leq \sup_{\lambda \in W} \{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)\} + \sup_{z \in \mathbb{R}^n} \{f(z) - \Lambda_\mu^*(z)\}.$$

Thus,

$$(13) \quad \log \int e^f d\mu \leq \log \int e^{\sup_{\lambda \in W} \{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)\}} d\mu(x) + \sup_{z \in \mathbb{R}^n} \{f(z) - \Lambda_\mu^*(z)\}.$$

Note that $\nabla \Lambda_\mu(\mathbb{R}^n) \subset K$ since any point in $\nabla \Lambda_\mu(\mathbb{R}^n)$ is a barycenter of a probability measure supported on K by (2). Therefore, for any $x \in K$, the function $\lambda \mapsto \langle \lambda, x \rangle - \Lambda_\mu(\lambda)$ is D -Lipschitz with respect to the ℓ^2 -norm. Let now $\delta > 0$, and let \mathcal{D}_δ be a δ/D -net of V . Then,

$$\log \int e^{\sup_{\lambda \in W} \{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)\}} d\mu(x) \leq \delta + \log \int e^{\sup_{\lambda \in \mathcal{D}_\delta} \{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)\}} d\mu(x).$$

Using a union bound, we get

$$\log \int e^{\sup_{\lambda \in \mathcal{D}_\delta} \{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)\}} d\mu(x) \leq \log |\mathcal{D}_\delta|$$

which yields the claim. \square

REMARK 2.1. In the case $\text{supp} \mu \subset [0, 1]^n$, one can get a potentially better bound by using not a net for the ℓ^2 -norm but for the ℓ^1 -norm instead. Indeed, for any $i \in \{1, \dots, n\}$ and $\lambda \in \mathbb{R}^n$,

$$\partial_i \Lambda(\lambda) = \int \langle x, e_i \rangle d\mu_y(x),$$

with $y = \nabla \Lambda(\lambda)$ and μ_y is defined in (1). Therefore, if $\text{supp} \mu \subset [0, 1]^n$, $\partial_i \Lambda(\lambda) \in [0, 1]$. Thus, for fixed $x \in \text{supp} \mu$, the function $\lambda \mapsto \langle \lambda, x \rangle - \Lambda_\mu(\lambda)$ is 1-Lipschitz w.r.t the ℓ^1 -norm. In the last step (13) of the proof of Proposition 1.1, if one takes $\delta > 0$ and \mathcal{F}_δ to be a δ -net for the ℓ^1 -norm, one obtains

$$\log \int e^{\sup_{\lambda \in V} \{\langle \lambda, x \rangle - \Lambda_\mu(\lambda)\}} d\mu(x) \leq \delta + \log |\mathcal{F}_\delta|.$$

This yields an error term in the approximation of the partition function getting closer to the one recently found in [4], Proposition 5.1, as a consequence of a structural result on probability measures on product spaces.

REMARK 2.2. As observed in Remark 1.1, we see that the above proof of Proposition 1.1 yields an approximation of the partition function which seems a bit off from the one Lemma 1.3 proposes.

In general, it is expected that, when f has a gradient of low complexity, this gap becomes negligible, that is, for any y in an appropriate level set of Λ_μ^* ,

$$\int f d\mu_y \simeq f(y),$$

where μ_y is defined in (1). Indeed, if μ is the n -fold product measure ν^n , with ν compactly supported, for example, we note that, by the mean-value theorem, we have

$$\left| \int f d\mu_y - f(y) \right| \leq \int \sup_{t \in K} |\langle \nabla f(t), x - y \rangle| d\mu_y(x),$$

where K denotes the convex hull of the support of μ . But as μ_y has barycenter y and is also a product measure with marginals having the same support as the one of ν , we can deduce, by the majorization theorem (see [32], Theorem 12.16),

$$\left| \int f d\mu_y - f(y) \right| \leq c \int \sup_{t \in K} \langle \nabla f(t), x \rangle d\gamma^n(x),$$

where c is a positive constant depending on the diameter of the support of ν and γ^n is the standard Gaussian measure on \mathbb{R}^n .

We now turn our attention to the proof of Theorem 1.4. As the “rate function” we are aiming for is not a priori convex, we cannot restrict ourselves to estimate the logarithmic Laplace transform of $f(X)$ and use the previous Theorem 1.1. We will actually refrain ourselves from using Chebychev’s inequality and prefer to work directly on the probability of deviations, in contrast with the path followed by Chatterjee and Dembo in [12], Section 4, proof of Theorem 1.1.

PROOF OF THEOREM 1.4. To ease the notation, we write Λ, Λ^* as shorthands for Λ_μ and Λ_μ^* , respectively. The key element of this upper bound is a deterministic result which translates the event where $f(X) \notin V_\delta(\{I \leq r\})$ into a large deviation event for the process

$$((\theta\lambda, X) - \Lambda(\theta\lambda))_{\theta>0, \lambda \in W}.$$

The bound of Theorem 1.4 is then obtained by using a net argument and a union bound to control the deviations of the above process.

Fix $\kappa \geq r > 0$. Denote by $K = \{\Lambda^* \leq \kappa\}$. In a first step we translate in a more geometric language the event where $X \in K$ and $f(X) \notin V_\delta(\{I \leq r\})$ which is the object of the following lemma.

LEMMA 2.1. *Let $x \in K$. If $f(x) \notin V_\delta(\{I \leq r\})$, then*

$$\inf_{z \in \delta W^\circ} \Lambda^*(x - z) \geq r,$$

where W° is the polar set of W , that is, denoting by $h_W(z) = \sup_{\lambda \in W} \langle \lambda, z \rangle$ for $z \in \mathbb{R}^n$ the support function of W ,

$$W^\circ = \{z \in \mathbb{R}^n, h_W(z) \leq 1\}.$$

Observe that, since Theorem 1.4 is only relevant when W is a small set in terms of complexity, its polar δW° can be thought as a large or typical set. The above lemma quantifies the fact that, when $f(x) \notin V_\delta(\{I \leq r\})$, x is far away from δW° with respect to the cost function Λ^* .

PROOF. Let $x \in K$. By definition of I , for any $y \in \mathbb{R}^n$ such that $\Lambda^*(y) \leq r$, we have $I(f(y)) \leq r$. Therefore, if $f(x) \notin V_\delta(\{I \leq r\})$, then

$$\inf_{\Lambda^*(y) \leq r} |f(x) - f(y)| \geq \delta.$$

Since $\kappa \geq r$, we deduce from the assumption (5) that, for any $y \in \mathbb{R}^n$ such that $\Lambda^*(y) \leq r$,

$$|f(x) - f(y)| \leq \sup_{\lambda \in W} \langle \lambda, x - y \rangle.$$

Therefore, we have

$$(14) \quad \inf_{\Lambda^*(y) \leq r} \sup_{\lambda \in W} \langle \lambda, x - y \rangle \geq \delta.$$

Let $z \in \delta W^\circ$, and assume that $\Lambda^*(x - z) < r$. By continuity of h_W and the continuity of Λ^* on its domain, we can assume $h_W(z) < \delta$. Using the above inequality (14) for $y = x - z$, we get that $h_W(x - y) = h_W(z) \geq \delta$ which yields a contradiction. Therefore,

$$\inf_{z \in \delta W^\circ} \Lambda^*(x - z) \geq r. \quad \square$$

As one can observe, Lemma 2.1 involves a convex optimization problem. A specific feature of such an optimization problem is that it admits a dual concave optimization problem which is linked through its multipliers to the initial problem. Using this duality, we obtain the following lemma.

LEMMA 2.2. *Let $x \in K$ and $\theta_0 = \kappa/\delta$. Then,*

$$\inf_{z \in \delta W^\circ} \Lambda^*(x - z) \leq \sup_{\substack{0 \leq \theta \leq \theta_0 \\ \lambda \in W_\theta}} \{ \langle \theta \lambda, x \rangle - \Lambda(\theta \lambda) - \theta \delta \},$$

where, for any $\theta > 0$,

$$(15) \quad W_\theta = \{ \lambda \in W : \theta \lambda \in \text{int}(\mathcal{D}_\Lambda), \nabla \Lambda(\theta \lambda) \in K \}.$$

PROOF. Since $\Lambda^*(x) \leq \kappa$ and $0 \in \delta W^\circ$, we obtain by taking $z = 0$ that

$$(16) \quad \inf_{z \in \delta W^\circ} \Lambda^*(x - z) \leq \kappa < +\infty.$$

As δW° is a closed set and Λ^* has compact level sets by [18], Lemma 2.2.31, the infimum is achieved at some $z_* \in \delta W^\circ$.

Since Λ^* and h_W are both convex functions, we deduce by Kuhn–Tucker Theorem (see [14], Theorem 9.4) that there exists $(\eta, \theta) \neq (0, 0)$ with $\eta \in \{0, 1\}$ and $\theta \geq 0$, such that $\theta(h_W(z_*) - \delta) = 0$ and

$$(17) \quad \eta \Lambda^*(x - z_*) = \inf \{ \eta \Lambda^*(x - z) + \theta(h_W(z) - \delta) : z \in \mathbb{R}^n \}.$$

Evaluating the function on the right-hand side at $z = 0$, we obtain that $\eta \Lambda^*(x - z_*) \leq -\theta \delta$. Therefore, the nontriviality condition $(\eta, \theta) \neq (0, 0)$ implies that $\eta = 1$. By (16) we know that $x - z_* \in \mathcal{D}_\Lambda^*$. We claim that further,

FACT 2.1. *$x - z_*$ lies in the interior of \mathcal{D}_Λ^* .*

The intuition behind this fact is that the gradient of Λ^* blows up around the frontier of its domain whereas the “penalization function” h_W remains Lipschitz, so that it can only be beneficial for the minimum to lie in the interior of \mathcal{D}_Λ^* . We now make this idea rigorous and prove formally Fact 2.1.

PROOF. Note first that the barycenter of μ , which we denote by \bar{m} , is in the interior of \mathcal{D}_Λ^* . Indeed, by Jensen’s inequality

$$(18) \quad \forall \lambda \in \mathbb{R}^n, \quad \Lambda(\lambda) \geq \langle \lambda, \bar{m} \rangle.$$

Therefore, $\Lambda^*(\bar{m}) \leq 0$. Since $\Lambda^* \geq 0$, we deduce that $\Lambda^*(\bar{m}) = 0$ and, in consequence, $0 \in \partial \Lambda^*(\bar{m})$. As Λ is strictly convex and essentially smooth, Λ^* is also strictly convex and essentially smooth by [34], Theorem 26.5. In particular, Λ^* is differentiable on $\text{int}(\mathcal{D}_\Lambda^*)$ and steep, so that any point with a nonempty subdifferential should lie in $\text{int}(\mathcal{D}_\Lambda^*)$. Therefore, $\bar{m} \in \text{int}(\mathcal{D}_\Lambda^*)$.

To prove that $x - z_* \in \text{int}(\mathcal{D}_\Lambda)$, we argue by contradiction. We will show that, if $x - z_*$ lies on the frontier of \mathcal{D}_Λ^* , denoted by $\text{fr}(\mathcal{D}_\Lambda^*)$, then, using the fact that Λ^* is steep, we can find a barycenter of \bar{m} and $x - z_*$ at which the cost function in the optimization problem (17) is smaller than at $x - z_*$, thus violating the optimality of $x - z_*$. Assume now that $x - z_* \in \text{fr}(\mathcal{D}_\Lambda^*)$. Then, by [34], Theorem 6.1, we deduce that for any $t \in (0, 1]$, $x - z_* - tu \in \text{int}(\mathcal{D}_\Lambda^*)$, where $u = x - z_* - \bar{m}$, as we are taking a barycenter between an interior point and a frontier point of a convex set. Define

$$(19) \quad \forall z \in \mathbb{R}^n, \quad F(z) = \Lambda^*(x - z) + \theta h_W(z).$$

For any $t > 0$, there exists $\zeta_t \in W$ such that $h_W(z_*) - h_W(z_* + tu) \geq -t \langle \zeta_t, u \rangle$. Using the convexity of Λ^* , we get

$$F(z_*) \geq F(z_* + tu) + \langle \nabla \Lambda^*(v_t) - \theta \zeta_t, tu \rangle,$$

where $v_t = x - z_* - tu$. Besides, Λ^* is strictly convex and, therefore, achieves its unique minimum at \bar{m} where $\Lambda^*(\bar{m}) = 0$. Thus, for any $t < 1$, $\Lambda^*(v_t) > 0$. Since $v_t = \bar{m} + (1 - t)u$, we deduce by convexity of Λ^* that $\langle \nabla \Lambda^*(v_t), u \rangle > 0$ for any $t < 1$. But W is a bounded set; therefore, ζ_t remains bounded as $t \in (0, 1]$. Since Λ^* is steep, we deduce that, for t close enough to 0,

$$\langle \nabla \Lambda^*(v_t) - \theta \zeta_t, u \rangle > 0$$

which gives the desired contradiction. \square

Since z_* is a global minimum of the function F defined in (19), we have $0 \in \partial F(z_*)$. From Fact 2.1 we know that $x - z_*$ is an interior point of \mathcal{D}_{Λ^*} . As Λ^* is differentiable on $\text{int}(\mathcal{D}_{\Lambda^*})$, we have $0 \in \{-\nabla \Lambda^*(x - z_*)\} + \theta \partial h_W(z_*)$ by [14], Theorem 4.10. By Danskin’s formula (see [14], Theorem 10.22), the subdifferential of the support function h_W is

$$\partial h_W(z) = \{\lambda_* \in W : h_W(z) = \langle \lambda_*, z \rangle\}.$$

Thus, there is a $\lambda \in W$ satisfying $\delta = \langle \lambda, z_* \rangle$ such that

$$(20) \quad \nabla \Lambda^*(x - z_*) = \theta \lambda.$$

Therefore,

$$\Lambda^*(x - z_*) = \langle \theta \lambda, x - z_* \rangle - \Lambda(\theta \lambda) = \langle \theta \lambda, x \rangle - \Lambda(\theta \lambda) - \theta \delta.$$

Note that θ and λ depends on x in an implicit way. However, we will show that θ and λ can be restricted to certain subsets uniformly in x such that $\Lambda^*(x) \leq \kappa$. More precisely, we will prove that $\theta \leq \kappa/\delta$ and $\lambda \in W_\theta$, with W_θ defined in (15), thus ending the proof of Lemma 2.2. We start with the bound on θ . Using the convexity of Λ^* , we have

$$\Lambda^*(x) - \Lambda^*(x - z_*) \geq \langle \nabla \Lambda^*(x - z_*), z_* \rangle.$$

Using that $\langle \lambda, z_* \rangle = \delta$ and that Λ^* is nonnegative, we get $\Lambda^*(x) \geq \theta \delta$. Since $\Lambda^*(x) \leq \kappa$, we obtain that $\theta \leq \kappa/\delta$. To prove that $\lambda \in W_\theta$, we use the fact that $\nabla \Lambda$ and $\nabla \Lambda^*$ are inverse maps and (16) to deduce that

$$\Lambda^*(\nabla \Lambda(\theta \lambda)) = \Lambda^*(x - z_*) \leq \Lambda^*(x) \leq \kappa$$

which finally ends the proof. \square

Combining Lemmas 2.1 and 2.2, we obtain that if $x \in K$ and $f(x) \notin V_\delta(\{I \leq r\})$, then

$$(21) \quad \sup_{\substack{0 \leq \theta \leq \theta_0 \\ \lambda \in W_\theta}} \{\langle \theta \lambda, x \rangle - \Lambda(\theta \lambda) - \theta \delta\} \geq r,$$

where W_θ is defined in (15) for any $\theta > 0$. In order to be allowed to take the probability of this event, we check the measurability of the supremum appearing in the above inequality.

FACT 2.2. *The function $x \in \mathbb{R}^n \mapsto \sup_{\substack{0 \leq \theta \leq \theta_0 \\ \lambda \in W_\theta}} \{\langle \theta \lambda, x \rangle - \Lambda(\theta \lambda) - \theta \delta\}$ is measurable.*

PROOF. Since $(\theta, \lambda) \mapsto \langle \theta \lambda, x \rangle - \Lambda(\theta \lambda) - \theta \delta$ is a continuous function, it suffices to prove that the subset,

$$A := \bigcup_{\theta \leq \theta_0} \{\theta\} \times W_\theta,$$

is separable in the sense that it contains a dense subset. Observe that A is a bounded set since $W_\theta \subset W$ for any $\theta \geq 0$ and W is bounded by assumption. Therefore, its closure, $\text{cl}(A)$, is compact. This entails that A is precompact and, in consequence, that it is separable. \square

Define now the measure

$$\mathbb{P}_K = \mathbb{P}(\cdot \cap \{X \in K\}).$$

From (21) we have the upper bound

$$(22) \quad \mathbb{P}_K(f(X) \notin V_\delta(\{I \leq r\})) \leq \mathbb{P}_K\left(\sup_{\substack{0 \leq \theta \leq \theta_0 \\ \lambda \in W_\theta}} \{\langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta\} \geq r\right).$$

In a last step we control the deviations of the supremum of the process appearing in the right-hand side of (22) using a net argument and performing a union bound.

LEMMA 2.3. *Let D denote the diameter of K . Let $\mathcal{D}_{\delta/2}$ be a $\delta/2D$ -net for the ℓ^2 -norm of W . Then,*

$$\log \mathbb{P}_K\left(\sup_{\substack{0 \leq \theta \leq \theta_0 \\ \lambda \in W_\theta}} \{\langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta\} \geq r\right) \leq -r + \log |\mathcal{D}_{\delta/2}| + \log\left(\left(\frac{\kappa L D}{\delta}\right) \vee 1\right) + 1,$$

where $\theta_0 = \kappa/\delta$, $L = \sup_{\lambda \in W} \|\lambda\|_{\ell^2}$ and for any $\theta \geq 0$,

$$W_\theta = \{\lambda \in W : \theta \lambda \in \text{int}(\mathcal{D}_\Lambda), \nabla \Lambda(\theta \lambda) \in K\}.$$

PROOF. We start by a net argument on θ . For $x \in K$ fixed, define the function

$$G : \theta \in \mathbb{R}_+ \mapsto \sup_{\lambda \in W_\theta} \{\langle \theta \lambda, x \rangle - \Lambda(\theta \lambda) - \theta \delta\}.$$

We claim that, for any $\theta' \leq \theta$,

$$(23) \quad G(\theta') - G(\theta) \leq (\theta - \theta')LD.$$

We observe first the following fact on the monotonicity of the sets W_θ .

FACT 2.3. *The family $(W_\theta)_{\theta \geq 0}$ is nonincreasing for the inclusion.*

PROOF. Let $\theta \geq 0$ and $\lambda \in W_\theta$. Since $K = \{\Lambda^* \leq \kappa\}$, it means that $\theta \lambda \in \text{int}(\mathcal{D}_{\Lambda^*})$ and $\Lambda^*(\nabla \Lambda(\theta \lambda)) \leq \kappa$. We know by (18) that $\bar{m} \in \partial \Lambda(0)$. Since Λ is essentially smooth and Λ has a nonempty subdifferential at 0, the point 0 must lie in the interior of \mathcal{D}_Λ . Since, by [34], Theorem 6.2, $\text{int}(\mathcal{D}_\Lambda)$ is a convex set, we deduce that for any $0 \leq \theta' \leq \theta$, $\theta' \lambda \in \text{int}(\mathcal{D}_\Lambda)$. In addition, for any $\theta' \leq \theta$ we can compute

$$\begin{aligned} \frac{\partial}{\partial \theta'} \Lambda^*(\nabla \Lambda(\theta' \lambda)) &= \langle \nabla \Lambda^*(\nabla \Lambda(\theta' \lambda)), \nabla^2 \Lambda(\theta' \lambda). \lambda \rangle \\ &= \theta' \langle \lambda, \nabla^2 \Lambda(\theta' \lambda). \lambda \rangle \geq 0, \end{aligned}$$

where we used the fact that $\nabla \Lambda$ and $\nabla \Lambda^*$ are inverse maps from one another. Thus, we can conclude that, for any $\theta' \leq \theta$, $\Lambda^*(\theta' \lambda) \leq \Lambda^*(\theta \lambda) \leq \kappa$ and, therefore, $\lambda \in W_{\theta'}$. \square

We now prove (23). Let $\theta' \leq \theta$. By convexity of Λ , we get, for any $\lambda \in W_\theta$,

$$(\langle \theta \lambda, X \rangle - \Lambda(\theta \lambda)) - (\langle \theta' \lambda, X \rangle - \Lambda(\theta' \lambda)) \leq (\theta - \theta') \langle \lambda, \nabla \Lambda(\theta' \lambda) - X \rangle.$$

Since $W_\theta \subset W_{\theta'}$ by Fact 2.3, we have $\nabla \Lambda(\theta'\lambda) \in K$. As we denoted by D the diameter of K , we obtain

$$(\langle \theta\lambda, X \rangle - \Lambda(\theta'\lambda)) - (\langle \theta'\lambda, X \rangle - \Lambda(\theta'\lambda)) \leq (\theta - \theta')LD$$

which holds for any $\theta' \leq \theta$ and $\lambda \in W_\theta \subset W_{\theta'}$. This yields the claim (23).

Let \mathcal{E} be a $1/(LD)$ -net of the interval $[0, \theta_0]$. One can find a net \mathcal{E} such that

$$(24) \quad |\mathcal{E}| \leq (LD\theta_0) \vee 1 = \left(\frac{\kappa LD}{\delta}\right) \vee 1.$$

For any $\theta \in [0, \theta_0]$, one can find $\theta' \in \mathcal{E}$ such that $0 \leq \theta - \theta' \leq 1/LD$. From (23) we deduce that

$$\sup_{\theta \in [0, \theta_0]} G(\theta) \leq \sup_{\theta' \in \mathcal{E}} G(\theta') + 1.$$

Thus, using a union bound, we get

$$(25) \quad \begin{aligned} & \mathbb{P}_K \left(\sup_{\substack{0 \leq \theta \leq \theta_0 \\ \lambda \in W_\theta}} \{ \langle \theta\lambda, X \rangle - \Lambda(\theta\lambda) - \theta\delta \} \geq r \right) \\ & \leq \sum_{\theta \in \mathcal{E}} \mathbb{P}_K \left(\sup_{\lambda \in W_\theta} \{ \langle \theta\lambda, X \rangle - \Lambda(\theta\lambda) - \theta\delta \} \geq r - 1 \right). \end{aligned}$$

Let us now fix $\theta \in \mathcal{E}$, and perform a net argument on $\lambda \in W_\theta$. Fix $x \in K$, and define the function H taking values in $\mathbb{R} \cup \{-\infty\}$,

$$H : \lambda \in \mathbb{R}^n \mapsto \langle \theta\lambda, X \rangle - \Lambda(\theta\lambda).$$

We claim that, for any $\lambda, \lambda' \in W_\theta$,

$$(26) \quad H(\lambda) - H(\lambda') \leq \theta D \|\lambda - \lambda'\|.$$

Indeed, by convexity of Λ we have, for $\lambda, \lambda' \in W_\theta$,

$$H(\lambda) - H(\lambda') \leq \theta \langle \lambda - \lambda', x - \nabla \Lambda(\theta\lambda') \rangle.$$

Since $\lambda' \in W_\theta$, we have $\nabla \Lambda(\theta\lambda') \in K$ which yields (26).

Let \mathcal{F} be a δ/D -net for the ℓ^2 -norm of W_θ such that $\mathcal{F} \subset W_\theta$. Using (26), we obtain that

$$\sup_{\lambda \in W_\theta} H(\lambda) \leq \sup_{\lambda' \in \mathcal{F}} H(\lambda') + \theta\delta.$$

Therefore,

$$\mathbb{P}_K \left(\sup_{\lambda \in W_\theta} \{ \langle \theta\lambda, X \rangle - \Lambda(\theta\lambda) \} \geq r - 1 + \theta\delta \right) \leq \mathbb{P}_K \left(\sup_{\lambda \in \mathcal{F}} \{ \langle \theta\lambda, X \rangle - \Lambda(\theta\lambda) \} \geq r - 1 \right).$$

Performing a union bound, we get

$$\mathbb{P} \left(\sup_{\lambda \in \mathcal{F}} \{ \langle \theta\lambda, X \rangle - \Lambda(\theta\lambda) \} \geq r - 1 \right) \leq |\mathcal{F}| e^{-r+1},$$

where we used the fact that, for any $\xi \in \mathbb{R}^n$ and $t \geq 0$,

$$\mathbb{P}(\langle \xi, X \rangle - \Lambda(\xi) \geq t) \leq e^{-t},$$

by Chernoff's inequality. Therefore,

$$(27) \quad \mathbb{P}_K \left(\sup_{\lambda \in W_\theta} \{ \langle \theta\lambda, X \rangle - \Lambda(\theta\lambda) \} \geq r - 1 + \theta\delta \right) \leq |\mathcal{F}| e^{-r+1}.$$

It remains now to show that we can find a δ/D -net \mathcal{F} of W_θ such that $\mathcal{F} \subset W_\theta$ and relate its cardinal to the one of a net of W . For any $A \subset \mathbb{R}^n$ and $r > 0$, we denote by $N(A, rB_{\ell^2})$ and $\overline{N}(A, rB_{\ell^2})$ the following covering numbers:

$$N(A, rB_{\ell^2}) = \min \left\{ N \in \mathbb{N} : \exists x_1, \dots, x_N \in \mathbb{R}^n, A \subset \bigcup_{i=1}^N B_{\ell^2}(x_i, r) \right\},$$

$$\overline{N}(A, rB_{\ell^2}) = \min \left\{ N \in \mathbb{N} : \exists x_1, \dots, x_N \in A, A \subset \bigcup_{i=1}^N B_{\ell^2}(x_i, r) \right\}.$$

While it is true that the first notion of covering number $N(A, rB_{\ell^2})$ is nondecreasing in A for the inclusion, this fact becomes wrong for $\overline{N}(A, rB_{\ell^2})$. When A is convex, it is known (see [2], Fact 4.1.4) that both covering numbers defined above coincide. Unfortunately, we cannot prove in general that W_θ is convex. But, up to loose a factor 2 in the mesh of the net, we have the following fact.

FACT 2.4. *Let $A \subset B \subset \mathbb{R}^n$ and $r > 0$. Then, $\overline{N}(A, rB_{\ell^2}) \leq N(B, (r/2)B_{\ell^2})$.*

PROOF. For any $A \subset \mathbb{R}^n$ and $r > 0$, let $M(A, rB_{\ell^2})$ be the separation number defined as

$$M(A, rB_{\ell^2}) = \max \{ N \in \mathbb{N} : \exists x_1, \dots, x_N \in A, \forall i \neq j, \|x_i - x_j\| \geq r \}.$$

Let $A \subset B \subset \mathbb{R}^n$. We know by [2], Fact 4.1.11, that $\overline{N}(A, rB_{\ell^2}) \leq M(A, rB_{\ell^2})$. Since $A \subset B$,

$$M(A, rB_{\ell^2}) \leq M(B, rB_{\ell^2}).$$

Using again [2], Fact 4.1.11, we have

$$M(B, rB_{\ell^2}) \leq N(B, (r/2)B_{\ell^2})$$

which ends the proof. \square

Let $\mathcal{D}_{\delta/2}$ be a $\delta/2D$ -net for the ℓ^2 -norm of W . As $W_\theta \subset W$, we deduce from Fact 2.4 that there exists a δ/D -net \mathcal{F} of W_θ for the ℓ^2 -norm such that $\mathcal{F} \subset W_\theta$ and $|\mathcal{F}| \leq |\mathcal{D}_{\delta/2}|$. From the two union bounds (25) and (27), we obtain

$$\log \mathbb{P}_K \left(\sup_{\substack{0 \leq \theta \leq \theta_0 \\ \lambda \in W_\theta}} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta \} \geq r \right) \leq \log |\mathcal{E}| + \log |\mathcal{F}| - r + 1.$$

Since $|\mathcal{E}| \leq (\kappa LD/\delta) \vee 1$ by (24) and $|\mathcal{F}| \leq |\mathcal{D}_{\delta/2}|$, we deduce that

$$\begin{aligned} & \log \mathbb{P}_K \left(\sup_{0 \leq \theta \leq \theta_0, \lambda \in W_\theta} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta \} \geq r \right) \\ & \leq -r + \log |\mathcal{D}_{\delta/2}| + \log \left(\left(\frac{\kappa LD}{\delta} \right) \vee 1 \right) + 1 \end{aligned}$$

which ends the proof. \square

To finalize the proof, we choose $\kappa \geq r$ such that $\mathbb{P}(X \notin K) \leq e^{-r}$. Using (22) and Lemma 2.3, we obtain

$$\begin{aligned} \mathbb{P}(f(X) \notin V_\delta(\{I \leq r\})) & \leq \mathbb{P}_K(f(X) \notin V_\delta(\{I \leq r\})) + e^{-r} \\ & \leq \frac{\kappa LD}{\delta} |\mathcal{D}_{\delta/2}| e^{-r+1} + e^{-r} \\ & \leq \frac{2\kappa LD}{\delta} |\mathcal{D}_{\delta/2}| e^{-r+1}, \end{aligned}$$

which gives us the claim. \square

3. Large deviation of traces of Wigner matrices. In this section we will give a proof of Theorem 1.10 and Corollary 1.11.

3.1. *Large deviation upper bound.* We start with the large deviation upper bound of Theorem 1.10. The strategy for this upper bound will be similar to the one we adopted to investigate the large deviations of the moments of β -ensembles in [3], Section 3. It consists in truncating the spectrum so as to reduce ourselves to study the deviations of a small fraction of the eigenvalues. Once this truncation is made, we will be able to use Theorem 1.4 efficiently.

We introduce some notation we will use throughout this section. Let $Y \in \mathcal{H}_n$. For any $k \in \{1, \dots, n\}$, we denote by $\text{tr}_{[k]}Y$ the truncated trace

$$(28) \quad \text{tr}_{[k]}Y = \sum_{i=1}^k \lambda_i(Y),$$

where $\lambda_1(Y), \dots, \lambda_n(Y)$ are the eigenvalues of Y in nonincreasing order. For any measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$, we define $f(Y)$ by functional calculus as the Hermitan matrix,

$$f(Y) = \sum_{i=1}^n f(\lambda_i(Y))u_i u_i^*,$$

where u_1, \dots, u_n are the eigenvectors of Y associated to $\lambda_1(Y), \dots, \lambda_n(Y)$. We now define, for d even,

$$(29) \quad f_k(Y) = \frac{1}{n} \text{tr}_{[k]}(Y/\sqrt{n})^d,$$

whereas for d odd,

$$f_k(Y) = \frac{1}{n} \text{tr}_{[k]}(Y_+/\sqrt{n})^d - \frac{1}{n} \text{tr}_{[k]}(Y_-/\sqrt{n})^d.$$

The first step toward the proof of the upper bound is the following lemma which will rely on concentration arguments.

LEMMA 3.1. *Assume X is a Wigner matrix satisfying the convex concentration property. Let $k \in \{1, \dots, n\}$ such that $n^{\frac{1}{d-1}} = o(k)$ and $k = o(n)$. For any $t > 0$,*

$$\lim_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P} \left(\left| \frac{1}{n} \text{tr}(X/\sqrt{n})^d - f_k(X) - \mu_{\text{sc}}(x^d) \right| > t \right) = -\infty.$$

3.1.1. *Concentration inequalities.* In order to prove Lemma 3.1, we will develop some deviation inequalities for the number of eigenvalues falling outside an interval and for truncated linear statistics of Hermitian random matrices satisfying the convex concentration property defined in 1.9. The proof of these inequalities will follow the now classical path laid by the work of Guionnet and Zeitouni in [22].

PROPOSITION 3.2. *Let X be a random Hermitian matrix satisfying the convex concentration property for some constant $c > 0$. Let $k \in \{1, \dots, n\}$, and let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a convex 1-Lipschitz function which achieves its infimum on \mathbb{R} . For any $t > 0$,*

$$\mathbb{P}(|\text{tr}_{[k]}f(X) - \mathbb{E}\text{tr}_{[k]}f(X)| > t) \leq c^{-1} \exp\left(-\frac{ct^2}{k}\right),$$

where $\text{tr}_{[k]}$ is the truncated trace defined in (28).

In order to apply our convex concentration property, we need to prove that the truncated linear statistics are convex Lipschitz functions of the entries. This is the object of the following lemma.

LEMMA 3.3. *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function which achieves its infimum on \mathbb{R} . Let $k \in \{1, \dots, n\}$. The function T_f defined by*

$$\forall X \in \mathcal{H}_n, \quad T_f(X) = \text{tr}_{[k]} f(X)$$

is convex. Moreover, if f is 1-Lipschitz, then T_f is \sqrt{k} -Lipschitz with respect to the Hilbert–Schmidt norm.

PROOF. The proof is a variation around the one of Klein’s lemma (see [1], Lemma 4.4.12). Since f achieves its infimum, we know that there exists x_0 such that $0 \in \partial f(x_0)$. Considering $\tilde{f} = f(\cdot + x_0) - f(x_0)$, we may and will assume that $x_0 = 0$ and $f(0) = 0$. We will show the following representation of T_f as supremum of affine functions.

LEMMA 3.4. *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. Assume $f(0) = 0$ and $0 \in \partial f(0)$. Let $\zeta : \mathbb{R} \rightarrow \mathbb{R}$ be a function such that*

$$\forall x \in \mathbb{R}, \quad \zeta(x) \in \partial f(x), \quad \text{and} \quad \zeta(0) = 0.$$

Then, for any $X \in \mathcal{H}_n$,

$$(30) \quad T_f(X) = \sup_{\text{rank} Y \leq k} \{ \text{tr} f(Y) + \text{tr} \zeta(Y)(X - Y) \}.$$

Moreover, if f is 1-Lipschitz, then T_f is \sqrt{k} -Lipschitz with respect to the Hilbert–Schmidt norm.

PROOF. First, we know by [34], Theorem 23.4, that, for every $x \in \mathbb{R}$, $\partial f(x) \neq \emptyset$ which justifies the existence of ζ . Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of X such that $f(\lambda_1) \geq \dots \geq f(\lambda_n)$. Let u_1, \dots, u_n be the associated eigenvectors. Note that, because we assumed $f(0) = 0$, then if we take $Y = \sum_{i=1}^k \lambda_i u_i u_i^*$, we get the equality

$$T_f(X) = \text{tr} f(Y) + \text{tr} \zeta(Y)(X - Y),$$

using the orthogonality of the eigenvectors. Therefore, only the inequality is left to prove. Let $Y \in \mathcal{H}_n$ with rank less than k , and denote by e_1, \dots, e_n an orthonormal basis of eigenvectors such that e_{k+1}, \dots, e_n are in the kernel and by μ_1, \dots, μ_k the eigenvalues associated to e_1, \dots, e_k . From the convexity of f , we have, for any $j \in \{1, \dots, n\}, i \in \{1, \dots, k\}$,

$$f(\lambda_j) \geq f(\mu_i) + \zeta(\mu_i)(\lambda_j - \mu_i).$$

Multiplying the above inequality by $|\langle u_j, e_i \rangle|^2$ and summing over $j \in \{1, \dots, n\}$, we get

$$(31) \quad \sum_{j=1}^n |\langle u_j, e_i \rangle|^2 f(\lambda_j) \geq f(\mu_i) + \sum_{j=1}^n |\langle u_j, e_i \rangle|^2 \zeta(\mu_i)(\lambda_j - \mu_i).$$

Writing the trace of $f'(Y)(X - Y)$ in the basis of the e_i ’s, we have

$$\begin{aligned} \text{tr} \zeta(Y)(X - Y) &= \sum_{i=1}^n \langle e_i, \zeta(Y)(X - Y)e_i \rangle \\ &= \sum_{i=1}^n \langle \zeta(Y)e_i, (X - Y)e_i \rangle \\ &= \sum_{i=1}^k \langle \zeta(Y)e_i, (X - Y)e_i \rangle, \end{aligned}$$

where we used the fact that $\zeta(0) = 0$. Finally, using the spectral decomposition of X , we get

$$\text{tr}\zeta(Y)(X - Y) = \sum_{i=1}^k \sum_{j=1}^n |\langle u_j, e_i \rangle|^2 \zeta(\mu_i)(\lambda_j - \mu_i).$$

Summing (31) now over $i \in \{1, \dots, k\}$, we deduce, writing $\alpha_j = \sum_{i=1}^k |\langle u_j, e_i \rangle|^2$, that

$$\sum_{j=1}^n \alpha_j f(\lambda_j) \geq \text{tr} f(Y) + \text{tr}\zeta(Y)(X - Y).$$

Observe that $\alpha_j \in [0, 1]$ and $\sum_j \alpha_j = k$. The maximum,

$$\max \left\{ \sum_j \alpha_j f(\lambda_j) : \alpha_j \in [0, 1], \sum_j \alpha_j = k \right\},$$

is achieved at the vector $\alpha = \mathbb{1}_J$ where J is the set of indices j corresponding to the k highest values of $f(\lambda_j)$. This shows that the representation (30) holds and that T_f is a convex function.

Assume further that f is 1-Lipschitz. This entails that $|\zeta(x)| \leq 1$ for any $x \in \mathbb{R}$. Consequently, $\|\zeta(Y)\|_2 \leq \sqrt{k}$ for any $Y \in \mathcal{H}_n$ with $\text{rank}(Y) \leq k$. We conclude from the representation (30) that T_f is \sqrt{k} -Lipschitz w.r.t the Hilbert–Schmidt norm. \square

As a consequence of Proposition 3.2, we get the following deviation estimate on the number of eigenvalues present outside the bulk.

PROPOSITION 3.5. *Let X be a random Hermitian matrix satisfying the convex concentration property with constant $c > 0$. Denote by $\lambda_1(X)$ its top eigenvalue and $\|X\|$ its operator norm. For any $M \geq 4\mathbb{E}(\lambda_1(X))_+$ and $1 \leq k \leq n$,*

$$\mathbb{P}(\mathcal{N}[M, +\infty) \geq k) \leq c^{-1} \exp\left(-\frac{cM^2k}{16}\right),$$

where for any $I \subset \mathbb{R}$, $\mathcal{N}(I)$ denotes the number of eigenvalues of X in I . As a consequence, for any $M \geq 4\mathbb{E}\|X\|$ and $1 \leq k \leq n$,

$$\mathbb{P}(\mathcal{N}((-M, M)^c) \geq k) \leq 2c^{-1} \exp\left(-\frac{cM^2k}{32}\right).$$

PROOF. Let $f(x) = \frac{2}{M}(x - M/2)_+$ for any $x \in \mathbb{R}$. Since $f(x) \geq 1$ for $x \geq M$, we have, denoting by $\lambda_1(X), \dots, \lambda_n(X)$, the eigenvalues of X in nonincreasing order,

$$\mathbb{P}(\mathcal{N}[M, +\infty) \geq k) \leq \mathbb{P}\left(\sum_{i=1}^k f(\lambda_i(X)) \geq k\right).$$

But

$$\mathbb{E}\left(\sum_{i=1}^k f(\lambda_i(X))\right) \leq \frac{2}{M}k\mathbb{E}(\lambda_1(X))_+ \leq \frac{k}{2}$$

for $M \geq 4\mathbb{E}(\lambda_1(X))_+$. Thus,

$$\mathbb{P}(\mathcal{N}[M, +\infty) \geq k) \leq \mathbb{P}\left(\sum_{i=1}^k f(\lambda_i(X)) - \mathbb{E}\sum_{i=1}^k f(\lambda_i(X)) \geq k/2\right).$$

Applying Proposition 3.2, we deduce that

$$(32) \quad \mathbb{P}(\mathcal{N}[M, +\infty) \geq k) \leq c^{-1} e^{-\frac{ckM^2}{16}}$$

which gives the first claim.

Since by definition $\lceil k/2 \rceil - 1 < k/2$, we have using a union bound,

$$\mathbb{P}(\mathcal{N}((-M, M)^c) \geq k) \leq \mathbb{P}(\mathcal{N}(-\infty, -M] \geq \ell) + \mathbb{P}(\mathcal{N}[M, +\infty) \geq \ell),$$

where $\ell = \lceil k/2 \rceil$. For any $M \geq 4\mathbb{E}\|X\|$, we obtain by applying inequality (32) to X and $-X$ alternatively,

$$\mathbb{P}(\mathcal{N}((-M, M)^c) \geq k) \leq 2c^{-1} e^{-\frac{ckM^2}{32}}$$

which gives the second claim. \square

3.1.2. An exponential equivalent. In this section, we apply the concentration inequalities we obtained in the previous section to give a proof of the exponential equivalent of Lemma 3.1.

PROOF OF LEMMA 3.1. Let k such that $n^{\frac{1}{d-1}} = o(k)$ and $k = o(n)$. We will prove the claim only in the case where d is even, the case where d is odd being almost the same. Let $M > 0$, and define the truncated power function, by $f_M(x) = x^d$ for $|x| \leq M$, and for $|x| > M$ by,

$$f_M(x) = dM^{d-1}(|x| - M) + M^d.$$

Let $\lambda_1^*, \dots, \lambda_n^*$ be the eigenvalues of X/\sqrt{n} in decreasing absolute values. We can write,

$$\sum_{i=k+1}^n f_M(\lambda_i^*) = \text{tr} f_M(X/\sqrt{n}) - \text{tr}_{[k]} f_M(X/\sqrt{n}).$$

As $f_M(\cdot/\sqrt{n})$ is dM^{d-1}/\sqrt{n} -Lipschitz, applying Proposition 3.2 alternatively to $\text{tr} f_M(X/\sqrt{n})$ and $\text{tr}_{[k]} f_M(X/\sqrt{n})$, and performing a union bound, we deduce that for any $t > 0$,

$$(33) \quad \mathbb{P}\left(\left| \sum_{i=k+1}^n f_M(\lambda_i^*) - \mathbb{E} \sum_{i=k+1}^n f_M(\lambda_i^*) \right| > tn\right) \leq 2c^{-1} \exp\left(-\frac{cn^2t^2}{4d^2M^{2(d-1)}}\right),$$

where we used the fact that $k \leq n$. On one hand,

$$\begin{aligned} \left| \mathbb{E} \sum_{i=k+1}^n \lambda_i^{*d} - \mathbb{E} \sum_{i=k+1}^n f_M(\lambda_i^*) \right| &\leq \mathbb{E} \sum_{i=1}^n |\lambda_i^*|^d \mathbb{1}_{|\lambda_i^*| \geq M} \\ &\leq \frac{1}{M} \mathbb{E} \text{tr} |X/\sqrt{n}|^{d+1}. \end{aligned}$$

Using twice the Cauchy–Schwarz inequality, we get

$$\frac{1}{n} \mathbb{E} \text{tr} |X/\sqrt{n}|^{d+1} \leq \mathbb{E} \left(\frac{1}{n} \text{tr} (X/\sqrt{n})^{2(d+1)} \right)^{\frac{1}{2}} \leq \left(\mathbb{E} \frac{1}{n} \text{tr} (X/\sqrt{n})^{2(d+1)} \right)^{\frac{1}{2}} = O(1),$$

where we used in the last equality [1], Lemma 2.1.6. Therefore,

$$\left| \mathbb{E} \sum_{i=k+1}^n \lambda_i^{*d} - \mathbb{E} \sum_{i=k+1}^n f_M(\lambda_i^*) \right| = O\left(\frac{n}{M}\right).$$

On the other hand,

$$\left| \mathbb{E} \sum_{i=k+1}^n \lambda_i^{*d} - \mathbb{E} \text{tr}(X/\sqrt{n})^d \right| \leq k \mathbb{E} \|X/\sqrt{n}\|^d = O(k),$$

as $\mathbb{E} \|X/\sqrt{n}\|^d = O(1)$ by [1], Exercice 2.1.27. Therefore, if $k = o(n)$ and M goes to infinity with n , then as $\mathbb{E} \frac{1}{n} \text{tr}(X/\sqrt{n})^d$ converges to $\mu_{\text{sc}}(x^d)$, we have

$$(34) \quad \left| \mu_{\text{sc}}(x^d) - \mathbb{E} \left(\frac{1}{n} \sum_{i=k+1}^n f_M(\lambda_i^*) \right) \right| \xrightarrow{n \rightarrow +\infty} 0.$$

Fix $t > 0$. Now, let M go to infinity with n such that $n^{1+2/d} = o(n^2/M^{2(d-1)})$, that is, $M^2 = o(n^a)$ with $a = (d-2)/(d(d-1))$. The above inequality (33) and the convergence (34) give

$$(35) \quad \lim_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P} \left(\left| \frac{1}{n} \sum_{i=k+1}^n f_M(\lambda_i^*) - \mu_{\text{sc}}(x^d) \right| > t \right) = -\infty.$$

But since $\mathbb{E} \|X\| = O(\sqrt{n})$ by [1], Exercice 2.1.29, we deduce from Proposition 3.5 that in order to have that

$$(36) \quad \lim_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}(\mathcal{N}([-R\sqrt{n}, R\sqrt{n}]^c) \geq k) = -\infty,$$

it is sufficient to take R going to $+\infty$ with n such that $n^{2/d} = o(kR^2)$. We assumed that $n^{\frac{1}{d-1}} = o(k)$ which is equivalent to say that

$$n^{\frac{2}{d}} = o(kn^a).$$

Therefore, we can find M going to infinity which satisfies both conditions,

$$M^2 = o(n^a) \quad \text{and} \quad n^{\frac{2}{d}} = o(kM^2).$$

With this choice of M , both estimates (35) and (36) with $R = M$ hold. But then

$$\mathbb{P} \left(\left| \sum_{i=k+1}^n \lambda_i^{*d} - \sum_{i=k+1}^n f_M(\lambda_i^*) \right| > tn \right) \leq \mathbb{P}(\mathcal{N}([-M\sqrt{n}, M\sqrt{n}]^c) \geq k).$$

Therefore,

$$\limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P} \left(\left| \sum_{i=k+1}^n \lambda_i^{*d} - \sum_{i=k+1}^n f_M(\lambda_i^*) \right| > tn \right) = -\infty$$

which together with (35) give the claim. \square

3.1.3. *Proof of the large deviation upper bound.* From Lemma 3.1 we see that it suffices to understand the large deviations of a truncated trace $f_k(X)$ defined in (29), with $n^{\frac{1}{d-1}} = o(k)$ and $k = o(n)$. The point is that this truncation lowers significantly the complexity. Indeed, we will be able to encode by $O(nk \log n)$ bits the “gradient” of the truncated trace $\text{tr}_{[k]}(X^d)$. Thus, Proposition 1.4 will give us a relevant upper bound, with respect to the speed $n^{1+\frac{2}{d}}$, as soon as we can take $k \log n = o(n^{\frac{2}{d}})$. But for $d \geq 3$, we see that

$$\frac{1}{d-1} < \frac{2}{d}.$$

Therefore, we can and will take k which satisfies both conditions $n^{\frac{1}{d-1}} = o(k)$ and $k \log n = o(n^{\frac{2}{d}})$.

Property of the rate function I_+ . We begin with proving that the rate function I_+ defined in Theorem 1.10 is a good rate function. It will follow from the next lemma:

LEMMA 3.6. *Let $c_n : \mathbb{R}^n \rightarrow \mathbb{R}_+$ and $F_n : \mathbb{R}^n \rightarrow \mathbb{R}$ be families of functions indexed by $n \in \mathbb{N}$ such that, for any $r > 0$, the subset*

$$\bigcup_{n \in \mathbb{N}} F_n(\{c_n \leq r\})$$

is bounded. Then, the functions defined as

$$\forall x \in \mathbb{R}, \quad I^{(n)}(x) = \sup_{\delta > 0} \inf_{m \geq n} I_{m,\delta}(x)$$

and

$$\forall x \in \mathbb{R}, \quad I_+(x) = \sup_{\delta > 0} \liminf_{n \rightarrow +\infty} I_{n,\delta}(x)$$

with

$$I_{m,\delta}(x) = \inf\{c_m(h) : |F_m(h) - x| < \delta, h \in \mathbb{R}^m\}$$

are good rate functions.

PROOF. Let $\tau > 0$. We can write, for any $x \in \mathbb{R}$,

$$\begin{aligned} I^{(n)}(x) \leq \tau &\iff \forall \delta > 0, \quad \inf_{m \geq n} I_{m,\delta}(x) \leq \tau \\ &\iff \forall \delta > 0, \forall \varepsilon > 0, \exists m \geq n, \quad I_{m,\delta}(x) < \tau + \varepsilon. \end{aligned}$$

By definition of $I_{m,\delta}$, we get

$$\begin{aligned} I^{(n)}(x) \leq \tau \\ \iff \forall \varepsilon > 0, \forall \delta > 0, \exists m \geq n, \exists h \in \mathbb{R}^m, \quad c_m(h) < \tau + \varepsilon, \quad |F_m(h) - x| < \delta. \end{aligned}$$

Thus,

$$\{I^{(n)} \leq \tau\} = \bigcap_{\varepsilon > 0} \overline{\bigcup_{m \geq n} F_m(\{c_m \leq \tau + \varepsilon\})}.$$

This yields that $I^{(n)}$ is a good rate function. As $I_+ = \sup_{n \in \mathbb{N}} I^{(n)}$, we deduce that I_+ is also a good rate function. \square

Let us now check that the assumptions of Lemma 3.6 are fulfilled in our setting. We first show that X is sub-Gaussian in the sense that there exists $C > 0$ such that

$$(37) \quad \forall Y \in \mathcal{H}_n, \quad \Lambda^*(Y) \geq \frac{1}{2C^2} \text{tr}(Y^2).$$

Applying the convex concentration property defined in 1.9 to the function $f(X) = \text{tr}(XH)$ with $H \in \mathcal{H}_n$ fixed, we obtain that, for any $t > 0$,

$$\mathbb{P}(|\text{tr}(XH)| > t) \leq c^{-1} e^{-\frac{ct^2}{\text{tr}(H^2)}}.$$

We deduce that $\Lambda(H) \leq \frac{C^2}{2} \text{tr}(H^2)$ which gives (37). But using the fact that $x \in \mathbb{R}_+ \mapsto x^{2/d}$ is subadditive, we have for any $Y \in \mathcal{H}_n$, $|\text{tr}(Y^d)| \leq \text{tr}(|Y|^d) \leq (\text{tr}(Y^2))^{d/2}$ so that, whenever

$$\Lambda^*(Y) \leq rn^{1+\frac{2}{d}}$$

for some $r > 0$, then we have $|\frac{1}{n} \text{tr}(Y/\sqrt{n})^d| \leq (2C^2r)^{\frac{d}{2}}$. This shows that we can apply Lemma 3.6 with $c_n(Y) = n^{-(1+\frac{2}{d})} \Lambda^*(Y)$ and $F_n(Y) = \mu_{\text{sc}}(x^d) + \frac{1}{n} \text{tr}(Y/\sqrt{n})^d$ and conclude that I_+ is a good rate function.

Upper bound. We can now proceed with the proof of the upper bound. By [18], Theorem 4.2.13, it is sufficient to prove the large deviations upper bound for the sequence $(f_k(X))_{n \in \mathbb{N}}$, as $(\mu_{sc}(x^d) + f_k(X))_{n \in \mathbb{N}}$ is exponentially equivalent to $(\frac{1}{n} \text{tr}(X/\sqrt{n})^d)_{n \in \mathbb{N}}$ by Lemma 3.1. We will first make sure that the rate function we are going to obtain by applying Theorem 1.4 is the same as the one we are aiming for, that is,

LEMMA 3.7. *Assume that k goes to $+\infty$ with n . For any $n \in \mathbb{N}$, $\delta > 0$, set*

$$\forall x \in \mathbb{R}, \quad J_{n,\delta}(x) = \inf \left\{ \frac{1}{n^{1+\frac{2}{d}}} \Lambda^*(Y) : |f_k(Y) - x| < \delta, Y \in \mathcal{H}_n \right\},$$

and let

$$J_+ = \sup_{\delta > 0} \liminf_{n \rightarrow +\infty} J_{n,\delta},$$

where f_k is defined in (29). For any $x \in \mathbb{R}$,

$$I_+(x + \mu_{sc}(x^d)) = \sup_{\delta > 0} \liminf_{n \rightarrow +\infty} J_{n,\delta}(x),$$

where I_+ is defined as

$$\forall x \in \mathbb{R}, \quad I_+(x) = \sup_{\delta > 0} \liminf_{n \rightarrow +\infty} I_{n,\delta}(x),$$

and for any $n \in \mathbb{N}$, $\delta > 0$,

$$I_{n,\delta}(x) = \inf \left\{ \frac{\Lambda^*(Y)}{n^{1+\frac{2}{d}}} : \left| \frac{1}{n} \text{tr}(Y/\sqrt{n})^d - \mu_{sc}(x^d) - x \right| < \delta, Y \in \mathcal{H}_n \right\}.$$

PROOF. We will prove that $I_+(\cdot + \mu_{sc}(x^d))$ and J_+ have the same level sets. We will first observe that if $Y \in \mathcal{H}_m$ is a nonnegative matrix such that $\text{tr}(Y^2) = O(m^{1+\frac{2}{d}})$, then

$$(38) \quad \text{tr}(Y^d) - \text{tr}_{[k]}(Y^d) = o(m^{1+\frac{d}{2}}).$$

Let $\lambda_1 \geq \dots \geq \lambda_m \geq 0$ be the eigenvalues of Y . From the fact that $\text{tr}(Y^2) = O(m^{1+\frac{2}{d}})$, we deduce that, for any $1 \leq \ell \leq m$,

$$\lambda_\ell \leq \ell^{-\frac{1}{2}} O(m^{\frac{1}{2}+\frac{1}{d}}).$$

Thus,

$$\text{tr}(Y^d) - \text{tr}_{[k]}(Y^d) = \sum_{\ell > k} \lambda_\ell^d = O(m^{1+\frac{d}{2}} k^{1-\frac{d}{2}})$$

which gives (38). Now, let $\tau > 0$ and x such that $J_+(x) \leq \tau$. Let $\delta > 0$. For any $n \in \mathbb{N}$, we have

$$\inf_{m \geq n} J_{m,\delta}(x) \leq \tau.$$

Therefore,

$$\inf_{m \geq n} J_{m,\delta}(x) = \inf_{m \geq n} \inf_{Y \in \mathcal{H}_m} \left\{ \frac{\Lambda^*(Y)}{m^{1+\frac{2}{d}}} : |f_k(Y) - x| < \delta, \Lambda^*(Y) \leq 2\tau m^{1+\frac{2}{d}} \right\}.$$

For any $m \in \mathbb{N}$ and $Y \in \mathcal{H}_n$ such that $\Lambda^*(Y) \leq 2\tau m^{1+\frac{2}{d}}$, we know by (37) that

$$\text{tr}(Y^2) \leq 4C^2 \tau m^{1+\frac{2}{d}}.$$

Applying the observation (38) to Y_+ and Y_- alternatively, we deduce that $\text{tr}(Y^d) - f_k(Y) = o(m^{1+\frac{d}{2}})$. Therefore, for n large enough,

$$\inf_{m \geq n} J_{m,\delta}(x) \geq \inf_{m \geq n} I_{m,2\delta}(x).$$

Therefore,

$$\liminf_{n \rightarrow +\infty} I_{n,2\delta}(x) \leq \tau.$$

As the above inequality is true for any $\delta > 0$, we obtain $I_+(x) \leq \tau$. Inverting the roles of I_+ and J_+ , we get the other inclusion. \square

We come back to the proof of the upper bound of Theorem 1.10. Let F be a closed subset of \mathbb{R} . We want to prove that

$$\limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{d}{2}}} \log \mathbb{P}(f_k(X) \in F) \leq -\inf_F J_+,$$

where J_+ is defined in Lemma 3.7. We can assume without loss of generality that $\inf_F J_+ > 0$. Let $r > 0$ such that $\inf_F J_+ > r$. Put in another way,

$$F \cap \{J_+ \leq r\} = \emptyset.$$

As J_+ is a good rate function, we can find a $\delta > 0$ such that

$$F \cap V_{2\delta}(\{J_+ \leq r\}) = \emptyset.$$

Define for any $n \in \mathbb{N}$,

$$J^{(n)} = \sup_{\delta > 0} \inf_{m \geq n} J_{m,\delta}.$$

Note that, as $J_+ = \sup_{n \in \mathbb{N}} J^{(n)}$,

$$\{J_+ \leq r\} = \bigcap_{n \in \mathbb{N}} \{J^{(n)} \leq r\}.$$

Using (37) and the fact that, for any $Y \in \mathcal{H}_n$,

$$f_k(Y) \leq (\text{tr}(Y^2))^{d/2},$$

we deduce by Lemma 3.6 that $J^{(n)}$ is a good rate function. Therefore, $\{J^{(n)} \leq r\}$ is a nonincreasing sequence of compact subsets so that, for n large enough,

$$\{J^{(n)} \leq r\} \subset V_\delta(\{J_+ \leq r\}).$$

Therefore,

$$F \cap V_\delta(\{J^{(n)} \leq r\}) = \emptyset.$$

But $\{J^{(n)} \leq r\} \supset \{J \leq rn^{1+\frac{2}{d}}\}$, where

$$J(x) = \inf\{\Lambda^*(Y) : f_k(Y) = x, Y \in \mathcal{H}_n\}.$$

Thus,

$$(39) \quad \mathbb{P}(f_k(X) \in F) \leq \mathbb{P}(f_k(X) \notin V_\delta(\{J \leq rn^{1+\frac{2}{d}}\})).$$

We are now in the position of applying Theorem 1.4. First, we check that the tightness condition (4) is satisfied. By [18], Lemma 5.1.14, we know that, for any i, j ,

$$\mathbb{E}e^{\frac{1}{2}\Lambda_{i,j}^*(X_{i,j})} \leq 4^{(1+\mathbb{1}_{i \neq j})},$$

using the fact that $(\Re X_{i,j}, \Im X_{i,j})$ are independent. Therefore, by Chernoff’s inequality

$$\mathbb{P}(\Lambda^*(X) > 8n^2) \leq e^{-4n^2} 4^{n^2} \leq e^{-n^2}$$

which proves that (4) is fulfilled.

Let $K = \{\Lambda^* \leq 8n^2\}$. We will now bound the increments of f_k on K . As noted before, the level sets of Λ^* are included in Hilbert–Schmidt balls, more precisely, we know from (37),

$$(40) \quad K \subset 4CnB_2,$$

where B_2 denotes the ball of radius 1 for the Hilbert–Schmidt norm. By Lemma 3.4 we know that if $f : \mathbb{R} \rightarrow \mathbb{R}$ is a convex differentiable function such that $f(0) = f'(0) = 0$, then the function

$$\forall Y \in \mathcal{H}_n, \quad T_f(Y) = \text{tr}_{[k]} f(Y)$$

admits the variational representation

$$\forall X \in \mathcal{H}_n, \quad T_f(X) = \sup_{Y \in \mathcal{H}_n, \text{rank}(Y) \leq k} \{ \text{tr} f(Y) + \text{tr}(f'(Y)(X - Y)) \}.$$

We deduce that, for any $X, Y \in \mathcal{H}_n$,

$$T_f(X) - T_f(Y) \leq \text{tr}(f'(Z)(X - Y)),$$

where $Z = \sum_{i=1}^k \lambda_i u_i u_i^*$, with $\lambda_1 \geq \dots \geq \lambda_n$ the eigenvalues of X and u_1, \dots, u_n the associated eigenvectors. If we take $f : x \mapsto x_+^d, x \mapsto x_-^d$, or $x \mapsto x^d$ (in the case d is even) and $X \in 4CnB_2$, then we see that $f'(Z)$ is of rank k and $\|f'(Z)\| \leq d\|X\|^{d-1} \leq d(4C)^{d-1}n^{d-1}$. As by definition, f_k is a combination of at most two functions T_f associated with the functions $x \mapsto x_+^d, x \mapsto x_-^d$ or $x \mapsto x^d$, we get that, for any $X, Y \in 4CB_2$,

$$(41) \quad f_k(X) - f_k(Y) \leq \sup_{H \in W} \text{tr} H(X - Y),$$

where

$$(42) \quad W = \{ H \in \mathcal{H}_n : \text{rank}(H) \leq 2k, \|H\| \leq c_0 n^{d-1} \},$$

where c_0 is some positive constant depending on C .

Note that by (40), the diameter of the level set K is bounded by $8Cn$, and, by (41), the Lipschitz constant of f_k on K is only polynomial in n . By Theorem 1.4 we get

$$(43) \quad \log \mathbb{P}(f_k(X) \notin V_\delta(\{J \leq rn^{1+\frac{2}{d}}\})) \leq -rn^{1+\frac{2}{d}} + \log N(8ncW, \delta B_2) + O(\log n),$$

where $N(8ncW, \delta B_2)$ denotes the covering number of $8ncW$ by Hilbert–Schmidt balls of radius δ . It now remains to compute the covering numbers of W . This is the object of the following lemma.

LEMMA 3.8. *Let $k \in \{1, \dots, n\}$. Define the set*

$$\mathcal{M}_k = \{ Y \in \mathcal{H}_n : \text{rank}(Y) \leq k, \|Y\| \leq 1 \}.$$

For any $\varepsilon \in (0, 1)$, let $N(\mathcal{M}_k, \varepsilon B_2)$ be the covering number of \mathcal{M}_k by Hilbert–Schmidt balls of radius ε . Then,

$$\log N(\mathcal{M}_k, \varepsilon B_2) \leq 2nk \log \left(\frac{12k}{\varepsilon} \right).$$

PROOF. Let $Y \in \mathcal{M}_k$ and $Z \in \mathcal{H}_n$. Let us spectrally decompose Y , and write Z as

$$Y = \sum_{i=1}^k \lambda_i u_i u_i^*, \quad Z = \sum_{i=1}^k \mu_i v_i v_i^*,$$

where μ_i are real numbers, v_i are unit vectors (not necessarily orthogonal to one another), $\lambda_1, \dots, \lambda_k$ are the possible nonzero eigenvalues of Y and u_1, \dots, u_k the associated eigenvectors. As the v_i 's are unit vectors,

$$\|Y - Z\|_2 \leq \sum_{i=1}^k |\lambda_i - \mu_i| + \sum_{i=1}^k |\lambda_i| \cdot \|u_i u_i^* - v_i v_i^*\|_2.$$

Since $\|Y\| \leq 1$ and $\|uv^*\|_2 = \|u\|_2 \|v\|_2$ for any $u, v \in \mathbb{C}^n$, we get,

$$(44) \quad \|Y - Z\| \leq k \max_{1 \leq i \leq k} |\lambda_i - \mu_i| + 2k \max_{1 \leq i \leq k} \|u_i - v_i\|_2.$$

Let \mathcal{E} be a $\varepsilon/2k$ -net of $[-1, 1]$ and \mathcal{F} a $\varepsilon/4k$ -net of the unit sphere \mathbb{S}^{n-1} . Then, the set

$$\mathcal{D}_\varepsilon = \left\{ \sum_{i=1}^k \mu_i v_i v_i^* : \mu_i \in \mathcal{E}, u_i \in \mathcal{F} \right\},$$

is an ε -net of \mathcal{M}_k due to the inequality (44). Clearly, we can find a net \mathcal{E} such that $|\mathcal{E}| \leq 2k/\varepsilon$. Moreover, by [10], Lemma 1.4.2, there exists a net \mathcal{F} such that,

$$|\mathcal{F}| \leq \left(\frac{12k}{\varepsilon} \right)^n.$$

From the construction of \mathcal{D}_ε we get, finally,

$$\log |\mathcal{D}_\varepsilon| \leq 2nk \log \left(\frac{12k}{\varepsilon} \right). \quad \square$$

Coming back to the inequality (43), we see that we have the inclusion $8CnW \subset n^b \mathcal{M}_{2k}$, for some $b > 0$. Thus, by Lemma 3.8

$$\log N(8CnW, \delta B_{\ell^2}) \leq N(\mathcal{M}_{2k}, \delta n^{-b} B_{\ell^2}) = O\left(nk \log\left(\frac{n}{\delta}\right)\right).$$

As we chose $k \log n = o(n^{\frac{2}{d}})$, we get

$$\limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}(f_k(X) \notin V_\delta(\{J \leq rn^{1+\frac{2}{d}}\})) \leq -r,$$

which implies, because of (39),

$$\limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}(f_k(X) \in F) \leq -r.$$

As this inequality is true for any $r < \inf_F J_+$, we obtain the large deviation upper bound of Theorem 1.10.

3.2. *Large deviation lower bound.* In this section we will use Lemma 1.6 to prove the large deviation lower bound for $(\frac{1}{n} \text{tr}(X/\sqrt{n})^d)_{n \in \mathbb{N}}$. One of the difficulties in the derivation of the lower bound is that the domain of Λ^* may be different from \mathcal{H}_n . To bypass this problem, we will add a small Gaussian noise, as in the proof of Cramer's theorem (see [18], Proof of Theorem 2.2.30).

3.2.1. *Regularization by Gaussian noise.* Recall that we denote by $\mathcal{H}_n^{(\beta)}$ the set of Hermitian matrices, when $\beta = 2$, and symmetric matrices, when $\beta = 1$, of size n . We say that Γ belongs to the *Gaussian orthogonal ensemble* (GOE), when $\beta = 1$, or to the *Gaussian unitary ensemble* (GUE), when $\beta = 2$, if it distributed according to the law

$$\frac{1}{Z_n^{(\beta)}} e^{-\frac{\beta}{4} \text{tr} H^2} d\ell_n^{(\beta)}(H),$$

where $\ell_n^{(\beta)}$ is the Lebesgue measure on $\mathcal{H}_n^{(\beta)}$.

LEMMA 3.9. *Let X be a Wigner matrix satisfying the convex concentration property 1.9. Let Γ be a GOE matrix if $\beta = 1$ or a GUE matrix if $\beta = 2$. For any $t > 0$,*

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}(|\text{tr}(X/\sqrt{n})^d - \text{tr}((X + \varepsilon\Gamma)/\sqrt{n})^d| > tn) = -\infty.$$

PROOF. For $\sigma \in \{-1, 1\}$ and $Y \in \mathcal{H}_n$, we write $Y_\sigma = Y_+$ if $\sigma = +$ and $Y_\sigma = Y_-$ if $\sigma = -$. With this notation it is sufficient to prove that, for $\sigma \in \{-1, 1\}$ and any $t > 0$,

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}(|\text{tr}(X_\sigma/\sqrt{n})^d - \text{tr}((X + \varepsilon\Gamma)_\sigma/\sqrt{n})^d| > tn) = -\infty.$$

Fix $\sigma \in \{-1, 1\}$. We will first show that for any $\varepsilon \geq 0$, the sequence $(n^{-1} \text{tr}|X + \varepsilon\Gamma|^d)_{n \in \mathbb{N}}$ is exponentially tight at the scale $n^{1+\frac{2}{d}}$. This will allow us to reduce ourselves to prove that $n^{-1/d} \|(X + \varepsilon\Gamma)_\sigma/\sqrt{n}\|_d$ is an exponentially good approximation of $n^{-1/d} \|X_\sigma/\sqrt{n}\|_d$, where $\|\cdot\|_d$ denotes the d th Schatten norm,

$$(45) \quad \|\cdot\|_d : Y \in \mathcal{H}_n \mapsto (\text{tr}|Y|^d)^{1/d}.$$

Since $\|\cdot\|_d$ is a convex 1-Lipschitz function with respect to the Hilbert–Schmidt norm, we obtain by applying the convex concentration property 1.9 that, for any $t > 0$,

$$(46) \quad \mathbb{P}(\|X/\sqrt{n}\|_d - \mathbb{E}\|X/\sqrt{n}\|_d > tn^{1/d}) \leq c^{-1} e^{-ct^2 n^{1+\frac{2}{d}}}.$$

But, using the Cauchy–Schwarz inequality and Jensen’s inequality, we have

$$\mathbb{E}(n^{-1/d} \|X/\sqrt{n}\|_d) \leq \mathbb{E}\left(\frac{1}{n} \text{tr}(X/\sqrt{n})^{2d}\right)^{\frac{1}{2d}} \leq \left(\mathbb{E}\frac{1}{n} \text{tr}(X/\sqrt{n})^{2d}\right)^{\frac{1}{2d}}.$$

By Wigner’s theorem [1], Lemma 2.1.6, we get that $\mathbb{E}\|X/\sqrt{n}\|_d = O(n^{1/d})$. Therefore, we can deduce from (46) that $(\frac{1}{n} \text{tr}|X/\sqrt{n}|^d)_{n \in \mathbb{N}}$ is exponentially tight at the scale $n^{1+\frac{2}{d}}$. Using further the triangle inequality for the d th Schatten norm, and the fact that Γ also satisfies the convex concentration property by [31], Theorem 5.2, 4.3, we deduce that, for any $\varepsilon \geq 0$, $(\frac{1}{n} \text{tr}|(X + \varepsilon\Gamma)/\sqrt{n}|^d)_{n \in \mathbb{N}}$ is exponentially tight at the scale $n^{1+\frac{2}{d}}$. Therefore, it suffices to show that, for arbitrary large but fixed $\tau > 0$ and any $t > 0$,

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}_{A_\tau}(|\mu_{X/\sqrt{n}}(x_\sigma^d) - \mu_{(X+\varepsilon\Gamma)/\sqrt{n}}(x_\sigma^d)| > t) = -\infty,$$

where \mathbb{P}_{A_τ} denotes the measure $\mathbb{P}(\cdot \cap A_\tau)$ and A_τ is the event,

$$A_\tau = \{\mu_{X/\sqrt{n}}(|x|^d) \leq \tau, \mu_{(X+\varepsilon\Gamma)/\sqrt{n}}(|x|^d) \leq \tau\}.$$

Using the uniform continuity of $x \mapsto |x|^{1/d}$ on compact sets, we deduce that it is enough to show, for any $\delta > 0$,

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}(|(\mu_{X/\sqrt{n}}(x_\sigma^d))^{\frac{1}{d}} - (\mu_{(X+\varepsilon\Gamma)/\sqrt{n}}(x_\sigma^d))^{\frac{1}{d}}| > \delta) = -\infty.$$

But, by Minkowski’s inequality,

$$|(\mu_{X/\sqrt{n}}(x^d))^{1/d} - (\mu_{(X+\varepsilon\Gamma)/\sqrt{n}}(x^d))^{1/d}| \leq \left(\frac{1}{n} \sum_{i=1}^n |(\lambda_i)_\sigma - (\mu_i)_\sigma|^d\right)^{1/d},$$

where $\lambda_1 \geq \dots \geq \lambda_n$ and $\mu_1 \geq \dots \geq \mu_n$ are the eigenvalues of, respectively, X/\sqrt{n} and $(X + \varepsilon\Gamma)/\sqrt{n}$. But

$$\sum_{i=1}^n |(\lambda_i)_\sigma - (\mu_i)_\sigma|^d \leq \sum_{i=1}^n |\lambda_i - \mu_i|^d,$$

and by Lidskii’s theorem (see [7], Theorem III.4.1),

$$\frac{1}{n} \sum_{i=1}^n |\lambda_i - \mu_i|^d \leq \mu_{\varepsilon\Gamma/\sqrt{n}}(|x|^d).$$

Thus, it actually suffices to prove that, for any $\delta > 0$,

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}(\mu_{\varepsilon\Gamma/\sqrt{n}}(|x|^d) > \delta) = -\infty.$$

But $(\frac{1}{n} \text{tr}|\Gamma/\sqrt{n}|^d)$ is exponentially tight at the scale $n^{1+\frac{2}{d}}$; therefore, we finally get the claim. \square

3.2.2. *A tilting strategy.* By [18], Theorem 4.2.16, it is sufficient to prove the large deviations lower bound for the sequence $(\frac{1}{n} \text{tr}((X + \varepsilon\Gamma)/\sqrt{n})^d)_{n \in \mathbb{N}}$ for any $\varepsilon > 0$ small enough. We denote by Λ_ε^* the Legendre transform of the logarithmic Laplace transform Λ_ε of $X + \varepsilon\Gamma$, that is,

$$\forall H \in \mathcal{H}_n^{(\beta)}, \quad \Lambda_\varepsilon(H) = \log \mathbb{E} e^{\text{tr}[(X+\varepsilon\Gamma)H]} = \Lambda(H) + \frac{\varepsilon^2 \text{tr}H^2}{\beta}.$$

Adding this small Gaussian noise yields that the domain of Λ_ε^* is $\mathcal{H}_n^{(\beta)}$. We note also for future record that

$$(47) \quad \forall Y \in \mathcal{H}_n^{(\beta)}, \quad \Lambda_\varepsilon^*(Y) \leq \Lambda^*(Y).$$

Let us now proceed with the proof of the large deviation lower bound. Denote by $Z = X + \varepsilon\Gamma$. Let $x \in \mathbb{R}$ such that $I_-(x) < +\infty$. This means, from the definition of I_- in Theorem 1.10, that we can find a sequence $Y \in \mathcal{H}_n^{(\beta)}$ such that

$$(48) \quad \lim_{n \rightarrow +\infty} \mu_{\text{sc}}(x^d) + \frac{1}{n} \text{tr}(Y/\sqrt{n})^d \xrightarrow[n \rightarrow +\infty]{} x, \quad \lim_{n \rightarrow +\infty} \frac{\Lambda^*(Y)}{n^{1+\frac{2}{d}}} = I_-(x).$$

Let $\delta > 0$. For n large enough, we have

$$\mathbb{P}\left(\frac{1}{n} \text{tr}(Z/\sqrt{n})^d \in B(x, 2\delta)\right) \geq \mathbb{P}\left(\frac{1}{n} \text{tr}(Z/\sqrt{n})^d \in B(s_n, \delta)\right),$$

where $s_n = \mu_{\text{sc}}(x^d) + \frac{1}{n} \text{tr}(Y/\sqrt{n})^d$. Let E denote the event

$$E = \left\{ A \in \mathcal{H}_n^{(\beta)} : \frac{1}{n} \text{tr}(A/\sqrt{n})^d \in B(s_n, \delta) \right\}.$$

Using Lemma 1.6 and (47), we have

$$\mathbb{P}(Z \in E) \geq e^{-\Lambda^*(Y)} \mathbb{P}_Y(Z \in E) \exp\left(-\frac{1}{\mathbb{P}_Y(Z \in E)^{1/2}} \langle H, \nabla^2 \Lambda_\varepsilon(H).H \rangle^{1/2}\right),$$

where $H = \nabla \Lambda_\varepsilon^*(Y)$ and, under \mathbb{P}_Y , Z follows the tilted law,

$$e^{\text{tr}(HZ) - \Lambda_\varepsilon(H)} d\mathbb{P}(Z).$$

To obtain the lower bound, we will prove on one hand,

$$(49) \quad \langle H, \nabla^2 \Lambda_\varepsilon(H).H \rangle = O(\Lambda^*(Y))$$

and, on the other hand, that

$$(50) \quad \mathbb{P}_Y(Z \in V) \xrightarrow{n \rightarrow +\infty} 1.$$

Provided these two claims hold, we get, as $\Lambda^*(Y) = O(n^{1+\frac{2}{d}})$,

$$\liminf_{n \rightarrow +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}\left(\frac{1}{n} \text{tr}(Z/\sqrt{n})^d \in B(x, \delta)\right) \geq -I_-(x)$$

which gives the lower bound of Theorem 1.10.

Let us now prove (49) and (50). For the first claim note that, for any $A \in \mathcal{H}_n^{(\beta)}$,

$$(51) \quad \nabla^2 \Lambda_\varepsilon(A) = \nabla^2 \Lambda(A) + \frac{2\varepsilon^2}{\beta} Id$$

and, for any $H \in \mathcal{H}_n^{(\beta)}$,

$$\text{tr}(H \nabla^2 \Lambda(A).H) = \sum_{i=1}^n \frac{\partial^2 \Lambda_{1,1}(A_{i,i})}{\partial A_{1,1}^2} H_{i,i}^2 + \sum_{i \neq j} \langle H_{i,j}, \nabla^2 \Lambda_{1,2}(A_{i,j}).H_{i,j} \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in \mathbb{C}^2 . As we assumed that the second derivatives of $\Lambda_{1,1}$ and $\Lambda_{1,2}$ are bounded, we deduce that there exists $r > 0$ such that, for any $A \in \mathcal{H}_n^{(\beta)}$, $\nabla^2 \Lambda_\varepsilon(A) \leq r$ for the matrix order. Thus, it is sufficient to prove that, for any $Y \in \mathcal{H}_n^{(\beta)}$,

$$(52) \quad \text{tr} H^2 = O(\Lambda^*(Y)),$$

where $H = \nabla \Lambda_\varepsilon^*(Y)$. From (51) we have, in particular, $\nabla^2 \Lambda_\varepsilon(A) \geq \frac{2\varepsilon^2}{\beta}$ for the matrix order. Therefore, denoting by $\Lambda_{\varepsilon,(i,j)}$ the logarithmic Laplace transform of $X_{i,j} + \varepsilon \Gamma_{i,j}$ for any $i, j \in \{1, \dots, n\}$, and integrating the inequalities

$$\forall A \in \mathcal{H}_n^{(\beta)}, \forall i < j, \quad \Lambda''_{\varepsilon,(i,i)}(A_{i,i}) \geq \frac{2\varepsilon^2}{\beta}, \nabla^2 \Lambda_{\varepsilon,(i,j)}(A_{i,j}) \geq \frac{2\varepsilon^2}{\beta}$$

(where the last inequality is meant in the matrix order when $\beta = 2$), knowing that $\nabla \Lambda_\varepsilon(0) = 0$, we obtain

$$\|\nabla \Lambda_\varepsilon(H)\|_2 \geq \varepsilon^2 \|H\|_2.$$

From (37) we deduce that

$$\Lambda^*(\nabla \Lambda_\varepsilon(H)) \geq \frac{1}{2C^2} \text{tr}(\nabla \Lambda_\varepsilon(H))^2 \geq \frac{\varepsilon^4}{2C^2} \text{tr} H^2$$

which proves the estimate (52) since, by definition, $\nabla \Lambda_\varepsilon(H) = Y$.

For the second claim (50), we observe that under \mathbb{P}_Y , Z is a random Hermitian matrix with independent entries up to the symmetry with mean $\nabla \Lambda_\varepsilon(H) = Y$. Moreover, if $\sigma_{i,j}^2 = \mathbb{E}_H |Z_{i,j} - \mathbb{E}Z_{i,j}|^2$ is the variance of the (i, j) th coordinate for $i < j$, then

$$\begin{aligned} |1 + \varepsilon^2 - \sigma_{i,j}^2| &= |\text{tr} \nabla^2 \Lambda_{\varepsilon,(i,j)}(0) - \text{tr} \nabla^2 \Lambda_{\varepsilon,(i,j)}(H_{i,j})| \\ &= |\text{tr} \nabla^2 \Lambda_{i,j}(0) - \text{tr} \nabla^2 \Lambda_{i,j}(H_{i,j})|, \end{aligned}$$

and

$$\sigma_{i,i}^2 = \Lambda''_{i,i}(H_{i,i}) + \frac{2\varepsilon^2}{\beta}.$$

As we assumed the second and third derivatives of $\Lambda_{i,j}$ to be uniformly bounded, we deduce that $|1 + \varepsilon^2 - \sigma_{i,j}^2| = O(|H_{i,j}|)$ and $\sigma_{i,i}^2 = O(1)$. Thus, using the fact $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for $x, y \geq 0$, we get

$$(53) \quad \sum_{i,j} (\sqrt{1 + \varepsilon^2} - \sigma_{i,j})^2 \leq \sum_{i,j} |1 + \varepsilon^2 - \sigma_{i,j}^2| \leq O(n(\|H\|_2 \vee 1)),$$

where we further used the Cauchy–Schwarz inequality. Therefore, from (52) we get

$$\sum_{i,j} (\sqrt{1 + \varepsilon^2} - \sigma_{i,j})^2 = O(n(\Lambda^*(Y)^{1/2} \vee 1)).$$

But $\Lambda^*(Y) = O(n^{1+\frac{2}{d}})$ by (48). Thus, we finally obtain

$$\sum_{i,j} (\sqrt{1 + \varepsilon^2} - \sigma_{i,j})^2 = O(n^{\frac{3}{2}+\frac{1}{d}}) = o(n^2).$$

Thus, to show the claim (50) it remains to prove the following lemma, which is where the assumptions we made on the boundedness of the derivatives of Λ play further their roles.

LEMMA 3.10. *Let $d \geq 3$ and ℓ even such that $\ell > d$. Let X be a random Hermitian matrix such that $(X_{i,j})_{i \leq j}$ are independent. Assume that the ℓ th moments of $X_{i,j}$ are uniformly bounded,*

$$\text{tr}(\mathbb{E}X)^2 = O(n^{1+\frac{2}{d}}) \quad \text{and} \quad \sum_{i,j} (1 - \sigma_{i,j})^2 = o(n^2),$$

where $\sigma_{i,j}^2 = \mathbb{E}|X_{i,j} - \mathbb{E}X_{i,j}|^2$. Then,

$$\left| \frac{1}{n} \text{tr}(X/\sqrt{n})^d - \mu_{\text{sc}}(x^d) - \frac{1}{n} \text{tr}(\mathbb{E}X/\sqrt{n})^d \right| \xrightarrow{n \rightarrow +\infty} 0$$

in probability.

PROOF. Let $\hat{X} = X - \mathbb{E}X$. From [3], Lemma 2.1 (9), we know that expanding $\text{tr}(\hat{X} + \mathbb{E}X)^d$, we obtain by bounding the mixed terms using Hölder’s inequality,

$$|\text{tr}X^d - \text{tr}\hat{X}^d - \text{tr}(\mathbb{E}X)^d| \leq 2^d \max_{1 \leq k \leq d-1} \|\hat{X}\|_{d+1}^k \|\mathbb{E}X\|_2^{d-k},$$

where $\|\cdot\|_m$ denotes the m th Schatten norm which is defined in (45). As $\ell > d$, we have $\|\hat{X}\|_{d+1} \leq \|\hat{X}\|_\ell$. Using the assumption that $\text{tr}(\mathbb{E}X)^2 = O(n^{1+\frac{2}{d}})$, we get

$$|\text{tr}X^d - \text{tr}\hat{X}^d - \text{tr}(\mathbb{E}X)^d| = O\left(\max_{1 \leq k \leq d-1} \|\hat{X}\|_\ell^k n^{(\frac{1}{2}+\frac{1}{d})(d-k)}\right).$$

But, as ℓ is even,

$$(54) \quad \mathbb{E} \text{tr} \hat{X}^\ell = O(n^{1+\frac{\ell}{2}})$$

by expanding the trace and using the fact that the ℓ th moments of the entries X are uniformly bounded. By Markov’s inequality we deduce that

$$\mathbb{P}(\|\hat{X}\|_\ell \geq (\log n)n^{\frac{1}{\ell}+\frac{1}{2}}) \xrightarrow{n \rightarrow +\infty} 0.$$

Therefore,

$$|\operatorname{tr} X^d - \operatorname{tr} \hat{X}^d - \operatorname{tr}(\mathbb{E}X)^d| = O\left((\log n)^d \max_{1 \leq k \leq d-1} n^{(\frac{1}{2} + \frac{1}{\ell})k} n^{(\frac{1}{2} + \frac{1}{d})(d-k)}\right)$$

with probability going to 1. As $\ell > d$, the maximum on the right-hand side is achieved at $k = 1$, so that

$$|\operatorname{tr} X^d - \operatorname{tr} \hat{X}^d - \operatorname{tr}(\mathbb{E}X)^d| = O((\log n)^d n^{1 + \frac{d}{2}} n^{-(\frac{1}{d} - \frac{1}{\ell})}) = o(n^{1 + \frac{d}{2}})$$

with probability going to 1. It now remains to show that

$$\frac{1}{n} \operatorname{tr}(\hat{X}/\sqrt{n})^d \xrightarrow{n \rightarrow +\infty} \mu_{\text{sc}}(x^d)$$

in probability. Let $\Sigma = (\sigma_{i,j})_{i,j}$ and \hat{Y} such that $\hat{X} = \Sigma \circ \hat{Y}$, where \circ denotes the Hadamard product. From Wigner’s theorem we know that, in probability,

$$\mu_{\hat{Y}/\sqrt{n}} \xrightarrow{n \rightarrow +\infty} \mu_{\text{sc}},$$

weakly. Denote by \mathcal{W}_2 the L^2 -Wasserstein distance on the space $\mathcal{P}(\mathbb{R})$ of probability measures on \mathbb{R} , which is defined by

$$\forall \mu, \nu \in \mathcal{P}(\mathbb{R}), \quad \mathcal{W}_2(\mu, \nu) = \inf_{\pi} \int |x - y|^2 d\pi(x, y),$$

where the infimum runs over all couplings between μ and ν . From the assumption on the variance profile of X , we have by Hoffman–Weilandt inequality (see [7], Theorem VI.4.1),

$$\mathbb{E} \mathcal{W}_2(\mu_{\hat{X}/\sqrt{n}}, \mu_{\hat{Y}/\sqrt{n}}) \xrightarrow{n \rightarrow +\infty} 0.$$

Thus, together with the weak convergence of $\mu_{\hat{Y}/\sqrt{n}}$ toward the semicircle law, we deduce that $\mathcal{W}_2(\mu_{\hat{X}/\sqrt{n}}, \mu_{\text{sc}})$ converges to 0 in probability. But from (54), we get, by Markov’s inequality,

$$\lim_{\tau \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \mathbb{P}(\mu_{\hat{X}/\sqrt{n}}(x^\ell) \geq \tau) \xrightarrow{n \rightarrow +\infty} 0.$$

Therefore, we can integrate the limit and deduce that

$$\frac{1}{n} \operatorname{tr}(\hat{X}/\sqrt{n})^d = \mu_{\text{sc}}(x^d) + o(1)$$

in probability. \square

3.3. *Computation of the rate function.* In this section we give a proof of Corollary 1.11. We first prove a general upper bound on the lower bound rate function I_- . This bound corresponds to the strategy of increasing (or decreasing) uniformly the mean of the off-diagonal entries.

LEMMA 3.11. *Assume X is a Wigner matrix such that in the case $\beta = 2$, $(\Re X_{1,2}, \Im X_{1,2})$ are each of variance $1/2$. Then,*

$$\forall x \geq 0, \quad \limsup_{n \rightarrow +\infty} I_n(x) \leq \frac{\beta}{4} x^{\frac{2}{d}},$$

where

$$I_n(x) = \inf \left\{ \frac{\Lambda^*(Y)}{n^{1 + \frac{2}{d}}} : \frac{1}{n} \operatorname{tr}(Y/\sqrt{n})^d = x, Y \in \mathcal{H}_n^{(\beta)} \right\}.$$

In particular,

$$I_-(x) \leq \begin{cases} \frac{\beta}{4}(x - \mu_{\text{sc}}(x^d))^{\frac{2}{d}} & \text{if } d \text{ is even and } x \geq \mu_{\text{sc}}(x^d), \\ \frac{\beta}{4}|x|^{\frac{2}{d}} & \text{if } d \text{ is odd and } x \in \mathbb{R}. \end{cases}$$

PROOF. Let $x \geq 0$. Let $y = (\frac{x}{(n-1)^{\frac{d}{2}+(n-1)}})^{\frac{1}{d}} n^{\frac{1}{2}+\frac{1}{d}}$, and define

$$Y = \begin{pmatrix} 0 & y & \dots & y \\ & \ddots & \ddots & \vdots \\ y & & \ddots & y \\ \vdots & & & \vdots \\ y & \dots & y & 0 \end{pmatrix} \in \mathcal{H}_n.$$

Then, $\text{tr}(Y^d) = xn^{1+\frac{d}{2}}$, and

$$\Lambda^*(Y) = \frac{n(n-1)}{2} \Lambda_{\Re X_{1,2}}^*(y).$$

But $y \sim n^{\frac{1}{d}-\frac{1}{2}} x^{\frac{1}{d}}$. As $\Re X_{1,2}$ has variance $1/\beta$, we claim that the Taylor expansion of Λ^* around 0 gives $\Lambda_{\Re X_{1,2}}^*(y) = \frac{\beta y^2}{2} + o(y^2)$. Indeed, if we let $g = \Lambda_{\Re X_{1,2}}^*$, then one can derive from the identity $(g^*)' \circ g' = Id$ the fact that

$$\forall \lambda \in \mathbb{R}, \quad (g^*)''(g'(\lambda)) = \frac{1}{g''(\lambda)}.$$

Since $\Re X_{1,2}$ is centered and has variance $1/\beta$. We have $g'(0) = 0$ and $g''(0) = 1/\beta$. Thus, $(g^*)''(0) = \beta$ and, therefore, $\Lambda_{\Re X_{1,2}}^*(y) = \frac{\beta y^2}{2} + o(y^2)$ as $y \rightarrow 0$. Thus,

$$\Lambda^*(Y) = \frac{n(n-1)}{2} \left(\frac{\beta}{2} x^{\frac{2}{d}} n^{\frac{2}{d}-1} + o(n^{\frac{2}{d}-1}) \right) = \left(\frac{\beta}{4} x^{\frac{2}{d}} + o(1) \right) n^{1+\frac{d}{2}}$$

which ends the proof. \square

REMARK 3.1. One can produce a second upper bound on I_- which corresponds this time to the strategy of having one very large entry on the diagonal (e.g.). It reads

$$\forall x \geq 0, \quad I_-(x + \mu_{\text{sc}}(x^d)) \leq \limsup_{n \rightarrow +\infty} \frac{\Lambda_{1,1}^*(x^{\frac{1}{d}} n^{\frac{1}{d}+\frac{1}{2}})}{n^{1+\frac{d}{2}}}$$

which is nontrivial if $\Lambda_{1,1}^*$ has a quadratic behavior at infinity. In particular, this bound shows that one can have a rate function very different from the one of the GOE or GUE. Indeed, if $\Lambda_{1,1}(\lambda) \simeq_{+\infty} \frac{B}{2} \lambda^2$, then $\Lambda_{1,1}^*(x) \simeq_{+\infty} \frac{1}{2B} x^2$, so that

$$\forall x \geq 0, \quad I_-(x + \mu_{\text{sc}}(x^d)) \leq \frac{1}{2B} x^{\frac{2}{d}}.$$

This bound is somewhat related to the large deviations of the traces of Wigner matrices with entries having super-Gaussian tails (see [3] for more details), where a heavy-tail phenomenon arises, in the sense that only the large entries of the matrix are controlling the large deviation of the trace.

In the case where the entries of X have sharp sub-Gaussian tails, we show in the following proposition that the upper bound on I_- proved in Lemma 3.11 is also a lower bound on I_+ . By Theorem 1.10 this fact immediately implies that a full LDP holds for the traces of this class of Wigner matrices, as announced in Corollary 1.11.

PROPOSITION 3.12. *Assume X is a Wigner matrix whose entries have sharp sub-Gaussian tail. Assume further that $\mathbb{E}X_{1,1}^2 \leq \beta/2$ and, in the case $\beta = 2$, that $(\Re X_{1,2}, \Im X_{1,2})$ are each of variance $1/2$. Then,*

$$I_+ = I_- = J_d,$$

where J_d is defined in Corollary 1.11 and where, for any $x \in \mathbb{R}$,

$$I_+(x) = \sup_{\delta > 0} \liminf_{n \rightarrow +\infty} I_{n,\delta}(x),$$

$$I_-(x) = \sup_{\delta > 0} \limsup_{n \rightarrow +\infty} I_{n,\delta}(x),$$

and for any $n \in \mathbb{N}$, $\delta > 0$,

$$I_{n,\delta}(x) = \inf \left\{ \frac{\Lambda^*(Y)}{n^{1+\frac{2}{d}}} : \left| \frac{1}{n} \operatorname{tr}(Y/\sqrt{n})^d + \mu_{\text{sc}}(x^d) - x \right| < \delta, Y \in \mathcal{H}_n \right\}.$$

PROOF. Since the entries of X have sharp sub-Gaussian tails, we deduce from our assumptions on the covariance structure that, for any $H \in \mathcal{H}_n^{(\beta)}$, $\Lambda(H) \leq (1/\beta)\operatorname{tr}H^2$. Therefore, we have, for any $Y \in \mathcal{H}_n^{(\beta)}$,

$$\Lambda^*(Y) \geq \frac{\beta}{4} \operatorname{tr}Y^2.$$

But $|\operatorname{tr}(Y^d)| \leq \operatorname{tr}(|Y|^d) \leq (\operatorname{tr}(Y^2))^{d/2}$; therefore,

$$\Lambda^*(Y) \geq \frac{\beta}{4} |\operatorname{tr}(Y^d)|^{\frac{2}{d}}.$$

Thus, for any $x \in \mathbb{R}$,

$$I_+(x) \geq \inf \left\{ \frac{\beta}{4} |y|^{\frac{2}{d}} : y + \mu_{\text{sc}}(x^d) = x, y \in \mathbb{R} \right\}.$$

In the case d is even, we see that the function on the right-hand side is infinite if $x < \mu(x^d)$ and, otherwise, equal to $(\beta/4)(x - \mu_{\text{sc}}(x^d))^{2/d}$. If d is odd, then one obtains $I_+(x) \geq (\beta/4)|x|^{2/d}$. In conclusion, we have the inequality $I_+ \geq J_d$. But, by Lemma 3.11, $I_- \leq J_d$, which completes the proof. \square

4. Upper tail of cycle counts in sparse Erdős–Rényi graphs. In this section we will give a proof of Proposition 1.12. As for the traces of Wigner matrices, the main idea is to reduce the complexity by showing that we can replace the full trace by a truncated version, involving only the eigenvalues at the edges of the spectrum. Then, using standard complexity computations, we apply Proposition 1.4 to obtain the desired upper bound.

4.1. A truncation argument. We recall that, for any $Y \in \mathcal{H}_n$, we denote by $\lambda_1(Y), \dots, \lambda_n(Y)$ its eigenvalues in nondecreasing order, and, for any $k \in \{1, \dots, n\}$, we denote by $\operatorname{tr}_{[k]}Y$ the truncated trace

$$\operatorname{tr}_{[k]}Y = \sum_{i=1}^k \lambda_i(Y).$$

We define g_k the function,

$$\forall Y \in \mathcal{H}_n, \quad g_k(Y) = \begin{cases} \text{tr}_{[k]}(Y^d) & \text{if } d \text{ is even,} \\ \text{tr}_{[k]}(Y_+^d) - \text{tr}_{[k]}(Y_-^d) & \text{if } d \text{ is odd.} \end{cases}$$

Building on the concentration inequalities we proved in Section 3.1.1, we will show the following exponential equivalent.

LEMMA 4.1. *Denote by $v_n = n^2 p^2 \log(1/p)$. Assume $\log(1/p) = o(np^2)$. Let $k \in \{1, \dots, n\}$ such that $\log^4(1/p) = o(k)$, and in the case $d = 3$ such that moreover $k = o((np)^{3/2})$. For any $\delta > 0$,*

$$\lim_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(\text{tr}(X^d) - g_k(X) > \delta p^d n^d) = -\infty.$$

PROOF. The proof will be slightly more involved than its counterpart for Wigner matrices, which we proved in Lemma 3.1. We will actually need here a ‘‘slicing argument’’ as the one used in the proof the large deviations of the moments of β -ensembles (see [3], Proposition 3.5). In a first step we will prove that, for any $\delta > 0$,

$$(55) \quad \lim_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(Z - \mathbb{E}Z > \delta p^d n^d) = -\infty,$$

where $Z = \text{tr}X^d - g_k(X)$. We will only consider the case where d is odd. In order to show (55), we see that it suffices to prove that, for $\sigma \in \{-, +\}$ and for any $\delta > 0$,

$$(56) \quad \lim_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}\left(\left| \sum_{i=k+1}^n \lambda_i^d - \mathbb{E} \sum_{i=k+1}^n \lambda_i^d \right| > \delta p^d n^d\right) = -\infty,$$

where we use λ_i as a shorthand for $\lambda_i(X_\sigma)$ for any $i \in \{1, \dots, n\}$. Fix $\sigma \in \{-, +\}$. Let $M > 0$, define f_M , the truncated power function, by $f_M(x) = x^d$ for $x \in [0, M]$ and, for $x > M$,

$$f_M(x) = dM^{d-1}(x - M) + M^d.$$

Note that f_M is by construction convex and dM^{d-1} -Lipschitz on \mathbb{R}_+ . By [9], Theorem 8.6, we know that X satisfies the convex concentration property, as defined in 1.9 with constants $\kappa = 1/2$ and $C = 2$. Observe that we can write

$$\sum_{i=k+1}^n f_M(\lambda_i) = \text{tr} f_M(X_\sigma) - \text{tr}_{[k]} f_M(X_\sigma).$$

Applying Proposition 3.2 to the function $x \mapsto f_M(x_\sigma)$ which is convex and dM^{d-1} -Lipschitz, we get

$$(57) \quad \mathbb{P}\left(\left| \sum_{i=k+1}^n f_M(\lambda_i) - \mathbb{E} \sum_{i=k+1}^n f_M(\lambda_i) \right| > \delta n^d p^d\right) \leq 4 \exp\left(-\frac{n^{2d-1} p^{2d} \delta^2}{8d^2 M^{2(d-1)}}\right).$$

Let α_n to be a sequence going to $+\infty$ such that

$$(58) \quad \alpha_n^2 = o\left(\frac{\sqrt{np}}{(\log(1/p))^{d-1}}\right),$$

which is possible since we assumed $\log(1/p) = o(np^2)$. We set

$$(59) \quad M = \sqrt{np} \alpha_n.$$

Rewriting (57) with this choice of M , we have

$$(60) \quad \mathbb{P}\left(\left|\sum_{i=k+1}^n f_M(\lambda_i) - \mathbb{E} \sum_{i=k+1}^n f_M(\lambda_i)\right| > \delta n^d p^d\right) \leq 4 \exp\left(-\frac{n^d p^{d+1} \delta^2}{8d^2 \alpha_n^{2(d-1)}}\right).$$

But

LEMMA 4.2. *Assume $\log n = o(np)$. There exists a constant $C_0 > 0$ such that, for any $M \geq C_0 \sqrt{np}$ and $k \geq 1$,*

$$\left|\mathbb{E} \sum_{i=k+1}^n (\lambda_i^d - f_M(\lambda_i))\right| = o(n^d p^d).$$

PROOF. We have

$$\left|\sum_{i=k+1}^n \lambda_i^d - \sum_{i=k+1}^n f_M(\lambda_i)\right| \leq \sum_{i=k+1}^n \mathbb{1}_{\lambda_i \geq M} \lambda_i^d.$$

Using the fact that $\lambda_i \leq n$ since the entries of X are bounded by 1, we deduce that

$$(61) \quad \left|\mathbb{E} \sum_{i=k+1}^n (\lambda_i^d - f_M(\lambda_i))\right| \leq n^{d+1} \mathbb{P}(\max(-\lambda_n(X), \lambda_2(X)) \geq M).$$

Let $u \in \mathbb{R}^n$ be the vector $(1, 1, \dots, 1)$, and let $\hat{X} = X - puu^*$. By Weyl’s inequalities (see [7], Theorem III.2.1), we have

$$\max(-\lambda_n(X), \lambda_2(X)) \leq \|\hat{X}\|.$$

Since $\log n = o(np)$, we know from [30], Example 4.10, that $\mathbb{E}\|\hat{X}\| = O(\sqrt{np})$. As noted before, \hat{X} has the convex concentration property with constants $\kappa = 1/2$ and $C = 2$. Using the fact that the spectral radius of a Hermitian matrix is a convex and 1-Lipschitz function with respect to the Hilbert–Schmidt norm, we deduce that there exist $c_0, \alpha > 0$ such that, for any $c \geq c_0 \sqrt{np}$,

$$(62) \quad \mathbb{P}(\|\hat{X}\| \geq c) \leq 2e^{-\alpha c^2}.$$

As we assumed $\log n = o(np)$, this bound is exponentially small which in view of (61) ends proof of the lemma. \square

By Lemma 4.2 we deduce that, for n large enough, we have

$$\left|\mathbb{E} \sum_{i=k+1}^n (\lambda_i^d - f_M(\lambda_i))\right| \leq \delta n^d p^d.$$

Therefore, by (60) we have, for n large enough,

$$\mathbb{P}\left(\left|\sum_{i=k+1}^n f_M(\lambda_i) - \mathbb{E} \sum_{i=k+1}^n \lambda_i^d\right| > 2\delta n^d p^d\right) \leq 4 \exp\left(-\frac{\delta^2 n^d p^{d+1}}{8d^2 \alpha_n^{2(d-1)}}\right).$$

Since $d \geq 3$, we have $\frac{d-2}{d-1} \leq \frac{1}{2}$. Thus, from the choice of α_n made in (58),

$$\alpha_n^2 = o\left(\frac{n^{\frac{d-2}{d-1}} p}{(\log(1/p))^{\frac{1}{d-1}}}\right),$$

so that

$$(63) \quad \lim_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P} \left(\left| \sum_{i=k+1}^n f_M(\lambda_i) - \mathbb{E} \sum_{i=k+1}^n \lambda_i^d \right| > 2\delta n^d p^d \right) = -\infty.$$

But

$$(64) \quad \mathbb{P} \left(\sum_{i=k+1}^n (\lambda_i^d - f_M(\lambda_i)) > \delta p^d n^d \right) \leq \mathbb{P} \left(\sum_{i=k+1}^n \lambda_i^d \mathbb{1}_{\lambda_i \geq M} > \delta n^d p^d \right).$$

Let $R > 0$, and denote by $\mathcal{N}[R, +\infty)$ the number of eigenvalues of X_σ in $[R, +\infty)$. By Weyl’s inequalities (see [7], Theorem III.2.1), we have, for any $i \geq 2$,

$$(65) \quad \lambda_i(\hat{X}) \leq \lambda_i(X) \leq \lambda_{i-1}(\hat{X}),$$

where $\hat{X} = X - puu^*$ with u being the all-ones vector. Therefore,

$$\mathbb{P}(\mathcal{N}[R, +\infty) \geq k) \leq \mathbb{P}(|\{i : (\lambda_i(\hat{X}))_\sigma \geq R\}| \geq k - 1).$$

As $\mathbb{E}\|\hat{X}\| = O(\sqrt{np})$ again by [30], Example 4.10, we deduce by Proposition 3.5 that if R is such that $\sqrt{np} = o(R)$, then, for any $k \geq 2$,

$$(66) \quad \mathbb{P}(\mathcal{N}[R, +\infty) \geq k) \leq 2 \exp\left(-\frac{R^2 k}{32}\right).$$

Now, take $R = \sqrt{v_n/\beta_n}$, with $\beta_n = o(k)$ and $\beta_n = o(np \log(1/p))$. Our choice of R satisfies $\sqrt{np} = o(R)$, and we have

$$\lim_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(\mathcal{N}[R/2, +\infty) \geq k) = -\infty.$$

Since we assumed that $\log^4(1/p) = o(k)$, we can actually set β_n to satisfy the conditions

$$(67) \quad \log^4(1/p) = o(\beta_n), \quad \beta_n = o(k) \quad \text{and} \quad \beta_n = o(np \log(1/p)),$$

using the fact that np grows polynomially fast to $+\infty$. Thus, in view of (63) and (64) it is sufficient to prove that

$$(68) \quad \limsup_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P} \left(\sum_{i=k+1}^n \lambda_i^d \mathbb{1}_{M \leq \lambda_i \leq R/2} > \delta n^d p^d \right) = -\infty.$$

If $R/2 < M$, then there is nothing left to prove. Assuming that $R/2 \geq M$, we are going to use a “slicing argument” to fill the gap between these two levels. To this end, define an increasing sequence M_m of levels such that $M_0 = M$ and, for any $m \geq 0$,

$$M_{m+1}^d = R^{d-2} M_m^2,$$

that is,

$$\log M_m = \left(\frac{2}{d}\right)^m \log M + \left(1 - \left(\frac{2}{d}\right)^m\right) \log R.$$

Let N be such that $(2/d)^N \leq \log 2/\log R$. As $\log R = O(\log n)$, it is possible to find such N with $N = O(\log \log n)$. For such N , we get $M_N \geq R/2$. Therefore,

$$\mathbb{P} \left(\sum_{i=k+1}^n \lambda_i^d \mathbb{1}_{M \leq \lambda_i \leq R/2} > \delta n^d p^d \right) \leq \mathbb{P} \left(\sum_{i=k+1}^n \lambda_i^d \mathbb{1}_{M \leq \lambda_i \leq M_N} > \delta n^d p^d \right).$$

Using a union bound, we get

$$(69) \quad \mathbb{P}\left(\sum_{i=k+1}^n \lambda_i^d \mathbb{1}_{M \leq \lambda_i \leq M_N} > \delta n^d p^d\right) \leq \mathbb{P}\left(\sum_{m=0}^{N-1} M_{m+1}^d \mathcal{N}[M_m, +\infty) > \delta n^d p^d\right) \\ \leq \sum_{m=0}^{N-1} \mathbb{P}(\mathcal{N}[M_m, +\infty) > \delta n^d p^d / N M_{m+1}^d).$$

As $M_m \geq M$ and $\sqrt{np} = o(M)$ by definition of M in (59), we deduce from (66) that, for any $m \geq 0$,

$$\mathbb{P}(\mathcal{N}[M_m, +\infty) > \delta n^d p^d / N M_{m+1}^d) \leq 2 \exp\left(-\frac{\delta n^d p^d}{32N R^{d-2}}\right).$$

But $R = \sqrt{v_n/\beta_n}$; therefore,

$$\mathbb{P}(\mathcal{N}[M_m, +\infty) > \delta n^d p^d / N M_{m+1}^d) \leq 2 \exp(-v_n c_n),$$

with

$$c_n = \frac{\delta \beta_n^{\frac{d-2}{2}}}{32N \log^{\frac{d}{2}}(1/p)}.$$

As $d/(d-2) \leq 3$ since $d \geq 3$ and $N = O(\log \log n)$, we see that with our choice of β_n , which satisfies $\log^4(1/p) = o(\beta_n)$, we have, in particular,

$$N^{\frac{2}{d-2}} \log^{\frac{d}{d-2}}(1/p) = o(\beta_n).$$

Thus, c_n goes to $+\infty$. But from the union bound (69), we have

$$\mathbb{P}\left(\sum_{i=k+1}^n \lambda_i^d \mathbb{1}_{M \leq \lambda_i \leq M_N} > \delta n^d p^d\right) \leq N \exp(-v_n c_n).$$

Therefore,

$$\lim_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}\left(\sum_{i=k+1}^n \lambda_i^d \mathbb{1}_{M \leq \lambda_i \leq M_N} > \delta n^d p^d\right) = -\infty$$

which ends the proof of (56) and, thus, of the claim (55).

Now, to end the proof of Lemma 4.1, we need the following lemma which estimates the expectation of the truncated trace involving the bulk eigenvalues.

LEMMA 4.3. *Assume $d \geq 4$ and $np^2 \gg 1$, or $d = 3$, $\log n = o(np)$ and $k = o((np)^{3/2})$. Then,*

$$|\mathbb{E}(\text{tr}(X^d) - g_k(X))| = o(n^d p^d).$$

PROOF. We start with the case $d \geq 4$. By Weyl’s inequalities (65), we get

$$|\text{tr}(X^d) - g_k(X)| \leq n \|\hat{X}\|^d.$$

But from (62), we know that $\mathbb{E}\|\hat{X}\|^d = O((np)^{\frac{d}{2}})$, thus,

$$|\text{tr}(X^d) - g_k(X)| = O(n(np)^{\frac{d}{2}}).$$

As $d \geq 4$ and $np^2 \rightarrow +\infty$, we have that $n^{\frac{d}{2}-1} p^{\frac{d}{2}}$ also goes to $+\infty$ which yields the claim.

Assume now that $d = 3$, $\log n = o(np)$ and $k = o((np)^{3/2})$. As $\lambda_1(X) \geq 0$, we have

$$\text{tr}(X^3) - (\text{tr}_{[k]}(X^3_+) - \text{tr}_{[k]}(X^3_-)) = \sum_{i=\ell+1}^{\ell'} \lambda_i(X)^3$$

for some ℓ, ℓ' such that $1 \leq \ell \leq k$ and $n - \ell' \leq k$. We have

$$\left| \sum_{i=\ell+1}^{\ell'} \lambda_i(X)^3 - \text{tr}(\hat{X})^3 \right| \leq \left| \sum_{i=\ell+1}^{\ell'} \lambda_i(X)^3 - \sum_{i=\ell+1}^{\ell'} \lambda_i(\hat{X})^3 \right| + 2k \|\hat{X}\|^3.$$

By Weyl's inequalities (65), we have, for any $i \geq 2$,

$$\lambda_i(\hat{X})^3 \leq \lambda_i(X)^3 \leq \lambda_{i-1}(\hat{X})^3.$$

Therefore,

$$0 \leq \sum_{i=\ell+1}^{\ell'} \lambda_i(X)^3 - \sum_{i=\ell+1}^{\ell'} \lambda_i(\hat{X})^3 \leq 2\|\hat{X}\|^3.$$

Thus,

$$\left| \mathbb{E} \sum_{i=\ell+1}^{\ell'} \lambda_i(X)^3 \right| \leq |\mathbb{E} \text{tr}(\hat{X})^3| + 2(k+1)\mathbb{E}\|\hat{X}\|^3.$$

But, as $(\hat{X})_{i,i} = -p$ for any $i \in \{1, \dots, n\}$ and the off-diagonal entries of \hat{X} are centered, we obtain

$$\mathbb{E} \text{tr}(\hat{X})^3 = -np^3 + O(n^2 p^2) = o(n^3 p^3).$$

Besides, as $\log n = o(np)$, we know from (62) that $\mathbb{E}\|\hat{X}\|^3 = O((np)^{3/2})$. Thus,

$$\left| \mathbb{E} \sum_{i=\ell+1}^{\ell'} \lambda_i(X)^3 \right| = O(k(np)^{3/2}) + o(n^3 p^3)$$

which gives the claim. \square

4.2. *Proof of Proposition 1.12.* From Lemma 4.1 we see that it suffices to consider the upper tail of the truncated trace,

$$g_k(X) = \begin{cases} \text{tr}_{[k]} X^d & \text{if } d \text{ is even,} \\ \text{tr}_{[k]} X^d_+ - \text{tr}_{[k]} X^d_- & \text{if } d \text{ is odd,} \end{cases}$$

for k such that $\log^4(1/p) = o(k)$ and with the additional condition that $k = o((np)^{3/2})$ in the case $d = 3$. As we assume that $\log^4(n) = o(np^2)$, we can find k which satisfies both conditions

$$(70) \quad k \log n = o(np^2 \log(1/p)) \quad \text{and} \quad \log^4(1/p) = o(k)$$

(the additional condition when $d = 3$ being automatically fulfilled).

With this choice of k , we will see that we have reduced the complexity enough to be able to apply Proposition 1.4. Indeed, we will see that we can encode the ‘‘gradient’’ of the truncated trace $\text{tr}_{[k]} X^d$ by $O(nk \log n)$ bits. Thus, the choice of k made above will guarantee us that the main error term in Proposition 1.4 is negligible with respect to the large deviation speed.

First, we will check that the rate function that Proposition 1.4 is giving us is as good as the one we are claiming, that is,

LEMMA 4.4. *Let $t \geq 1$, and define,*

$$\psi_n(t) = \inf\{\Lambda_p^*(Y) : g_k(Y) \geq tn^d p^d, Y \in \mathcal{H}_n^0\}.$$

Then,

$$\psi_n(t) \geq \phi_n(t - O(k^{1-\frac{d}{2}})),$$

where ϕ_n is defined in Proposition 1.12.

PROOF. Note that when d is even, the claim of the above lemma is trivial, so that we will assume for now on that d is odd. Observe also that $\psi_n(t) = O(v_n)$ by taking C to correspond to planting clique of size $r = \lceil (t - 1)^{1/d} np \rceil$, that is,

$$\forall i < j, \quad C_{i,j} = \begin{cases} 1 & \text{if } i \wedge j \leq r, \\ p & \text{if } i \vee j > r. \end{cases}$$

Let then $Y \in \mathcal{H}_n^0$ be such that $\Lambda_p^*(Y) = O(v_n)$ and $g_k(Y) \geq tn^d p^d$. Without loss of generality we can assume that $Y_{i,j} \geq p$ for any $i < j$. By [33], Corollary 3.5, we know that, for any $x \geq 0$,

$$\Lambda_p^*(p + x) \geq x^2(\log(1/p) - o(1)).$$

We deduce, denoting by U the matrix such that $U_{i,j} = 1$ for $i \neq j$ and null diagonal coefficients, that

$$\text{tr}(Y - pU)^2 = O(n^2 p^2).$$

In particular, $\text{tr}Y^2 = O(n^2 p^2)$. We obtain, for any $i \geq 1$,

$$\lambda_i(Y_-) = O\left(\frac{np}{\sqrt{i}}\right).$$

Therefore,

$$\sum_{i=k+1}^n \lambda_i(Y_-)^d = O(n^d p^d k^{1-\frac{d}{2}}).$$

But

$$g_k(Y) \leq \text{tr}(Y^d) + \sum_{i=k+1}^n \lambda_i(Y_-)^d$$

which gives the claim. \square

We can now proceed with the proof of the upper bound of Proposition 1.12. Note that Φ is lower semicontinuous, as one can see from its explicit expression (12). Therefore, by Lemma 4.1 it is sufficient to show that, for any $t > 1$,

$$\limsup_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(g_k(X) \geq tn^d p^d) \leq -\Phi(t),$$

with k such that $\log^4(1/p) = o(k)$ and $k = o(np^2)$.

Fix some $t > 1$ and $\delta > 0$ such that $t - \delta > 1$. Observe that as Λ_p^* is strictly convex, ψ_n , defined in Lemma 4.4, is strictly increasing. Therefore, we can apply Corollary 1.5. Let $\mathcal{H}_n^0([0, 1])$ denote the subset of \mathcal{H}_n^0 which consists of matrix with entries in $[0, 1]$. Observe that, for any $Y \in \mathcal{H}_n^0([0, 1])$,

$$\Lambda_p^*(Y) \leq n^2 \log(1/p),$$

so that the tightness assumption (4) of Proposition 1.4 is fulfilled with $K = \mathcal{H}_n^0([0, 1])$ and $\kappa = n^2 \log(1/p)$. Arguing as in the proof of (41) and using the fact that $\|X\| \leq n$ for any $X \in \mathcal{H}_n^0([0, 1])$, we obtain that, for any $X, Y \in \mathcal{H}_n^0([0, 1])$,

$$g_k(X) - g_k(Y) \leq \sup_{H \in W} \operatorname{tr} H(X - Y),$$

where

$$W = \{H \in \mathcal{H}_n : \operatorname{rank}(H) \leq 2k, \|H\| \leq 2dn^{d-1}\},$$

using the fact that for any $H \in \mathcal{H}_n([0, 1])$, $\|H\| \leq n$. As the Lipschitz constant of g_k on $\mathcal{H}_n([0, 1])$ and the diameter of $\mathcal{H}_n([0, 1])$ are both only polynomial in n , we get, by Corollary 1.5,

$$\mathbb{P}(g_k(X) \geq tn^d p^d) \leq -\psi_n(t - \delta) + \log N(W, (\delta/2n)B_2) + O(\log n),$$

where $N(W, (\delta/2n)B_2)$ is the covering number of W by balls of radius δ/n for the Hilbert–Schmidt norm. By Lemma 3.8 we have

$$\log N(W, (\delta/2n)B_2) = O(nk \log n).$$

As we chose k such that $k \log n = o(np^2 \log(1/p))$, we get

$$\mathbb{P}(g_k(X) \geq tn^d p^d) \leq -\psi_n(t - \delta) + o(v_n).$$

Using Lemma 4.4 and the fact that ϕ_n is increasing, we obtain that, for n large enough, $\psi_n(t - \delta) \geq \phi_n(t - 2\delta)$. Therefore,

$$\limsup_{n \rightarrow +\infty} \frac{1}{v_n} \log \mathbb{P}(g_k(X) \geq tn^d p^d) \leq -\Phi(t - 2\delta).$$

Since this is true for any $\delta > 0$ and Φ is lower semicontinuous, we get the claim by taking $\delta \rightarrow 0$.

Acknowledgements. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 692452).

I would like to thank Ofer Zeitouni for his enlightening insight and advice. I am also grateful to Ronen Eldan for some influential discussions, as well as to Anirban Basak for many helpful discussions. Finally, I thank the anonymous referee for the numerous comments and suggestions.

REFERENCES

- [1] ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics **118**. Cambridge Univ. Press, Cambridge. [MR2760897](#)
- [2] ARTSTEIN-AVIDAN, S., GIANOPOULOS, A. and MILMAN, V. D. (2015). *Asymptotic Geometric Analysis, Part I. Mathematical Surveys and Monographs* **202**. Amer. Math. Soc., Providence, RI. [MR3331351](#) <https://doi.org/10.1090/surv/202>
- [3] AUGERI, F. (2018). On the large deviations of traces of random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** 2239–2285. [MR3865672](#) <https://doi.org/10.1214/17-AIHP870>
- [4] AUSTIN, T. (2019). The structure of low-complexity Gibbs measures on product spaces. *Ann. Probab.* **47** 4002–4023. [MR4038047](#) <https://doi.org/10.1214/19-aop1352>
- [5] BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*, 2nd ed. Springer Series in Statistics. Springer, New York. [MR2567175](#) <https://doi.org/10.1007/978-1-4419-0661-8>
- [6] BASAK, A. and MUKHERJEE, S. (2017). Universality of the mean-field for the Potts model. *Probab. Theory Related Fields* **168** 557–600. [MR3663625](#) <https://doi.org/10.1007/s00440-016-0718-0>

- [7] BHATIA, R. (1997). *Matrix Analysis. Graduate Texts in Mathematics* **169**. Springer, New York. MR1477662 <https://doi.org/10.1007/978-1-4612-0653-8>
- [8] BHATTACHARYA, B. B., GANGULY, S., LUBETZKY, E. and ZHAO, Y. (2017). Upper tails and independence polynomials in random graphs. *Adv. Math.* **319** 313–347. MR3695877 <https://doi.org/10.1016/j.aim.2017.08.003>
- [9] BOUCHERON, S., LUGOSI, G. and MASSART, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univ. Press, Oxford. MR3185193 <https://doi.org/10.1093/acprof:oso/9780199535255.001.0001>
- [10] CHAFAÏ, D., GUÉDON, O., LECUÉ, G. and PAJOR, A. (2012). *Interactions Between Compressed Sensing Random Matrices and High Dimensional Geometry. Panoramas et Synthèses [Panoramas and Synthèses]* **37**. Société Mathématique de France, Paris. MR3113826
- [11] CHATTERJEE, S. (2012). The missing log in large deviations for triangle counts. *Random Structures Algorithms* **40** 437–451. MR2925306 <https://doi.org/10.1002/rsa.20381>
- [12] CHATTERJEE, S. and DEMBO, A. (2016). Nonlinear large deviations. *Adv. Math.* **299** 396–450. MR3519474 <https://doi.org/10.1016/j.aim.2016.05.017>
- [13] CHATTERJEE, S. and VARADHAN, S. R. S. (2011). The large deviation principle for the Erdős–Rényi random graph. *European J. Combin.* **32** 1000–1017. MR2825532 <https://doi.org/10.1016/j.ejc.2011.03.014>
- [14] CLARKE, F. (2013). *Functional Analysis, Calculus of Variations and Optimal Control. Graduate Texts in Mathematics* **264**. Springer, London. MR3026831 <https://doi.org/10.1007/978-1-4471-4820-3>
- [15] COOK, N. and DEMBO, A. Large deviations of sub-graph counts in sparse Erdős–Rényi. Available at [arXiv:1809.11148](https://arxiv.org/abs/1809.11148).
- [16] DEMARCO, B. and KAHN, J. (2012). Upper tails for triangles. *Random Structures Algorithms* **40** 452–459. MR2925307 <https://doi.org/10.1002/rsa.20382>
- [17] DEMARCO, B. and KAHN, J. (2012). Tight upper tail bounds for cliques. *Random Structures Algorithms* **41** 469–487. MR2993131 <https://doi.org/10.1002/rsa.20440>
- [18] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications. Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. MR2571413 <https://doi.org/10.1007/978-3-642-03311-7>
- [19] DOBRUSHIN, R., GROENEBOOM, P. and LEDOUX, M. (1996). *Lectures on Probability Theory and Statistics. Lecture Notes in Math.* **1648**. Springer, Berlin. MR1600892 <https://doi.org/10.1007/BFb0095673>
- [20] ELKAN, R. (2018). Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* **28** 1548–1596. MR3881829 <https://doi.org/10.1007/s00039-018-0461-z>
- [21] GUIONNET, A. and HUSSON, J. Large deviations for the largest eigenvalue of Rademacher matrices. hal-01828877. Available at <https://hal.archives-ouvertes.fr/hal-01828877v2>.
- [22] GUIONNET, A. and ZEITOUNI, O. (2000). Concentration of the spectral measure for large matrices. *Electron. Commun. Probab.* **5** 119–136. MR1781846 <https://doi.org/10.1214/ECP.v5-1026>
- [23] JAIN, V., KOEHLER, F. and RISTESKI, A. The mean-field approximation: Information inequalities, algorithms, and complexity. Available at [arXiv:1802.06126](https://arxiv.org/abs/1802.06126).
- [24] JAIN, V., RISTESKI, A. and KOEHLER, F. (2019). Mean-field approximation, convex hierarchies, and the optimality of correlation rounding: A unified perspective. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* 1226–1236. ACM, New York. MR4003424 <https://doi.org/10.1145/3313276.3316299>
- [25] JANSON, S., ŁUCZAK, T. and RUCIŃSKI, A. (1990). An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. In *Random Graphs '87 (Poznań, 1987)* 73–87. Wiley, Chichester. MR1094125
- [26] JANSON, S., OLESZKIEWICZ, K. and RUCIŃSKI, A. (2004). Upper tails for subgraph counts in random graphs. *Israel J. Math.* **142** 61–92. MR2085711 <https://doi.org/10.1007/BF02771528>
- [27] JANSON, S. and WARNKE, L. (2016). The lower tail: Poisson approximation revisited. *Random Structures Algorithms* **48** 219–246. MR3449596 <https://doi.org/10.1002/rsa.20590>
- [28] KIM, J. H. and VU, V. H. (2004). Divide and conquer martingales and the number of triangles in a random graph. *Random Structures Algorithms* **24** 166–174. MR2035874 <https://doi.org/10.1002/rsa.10113>
- [29] KOZMA, G. and SAMOTIJ, W. Private communication.
- [30] LATAŁA, R., VAN HANDEL, R. and YOUSSEF, P. (2018). The dimension-free structure of nonhomogeneous random matrices. *Invent. Math.* **214** 1031–1080. MR3878726 <https://doi.org/10.1007/s00222-018-0817-x>
- [31] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon. Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. MR1849347

- [32] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* **23**. Springer, Berlin. MR1102015 <https://doi.org/10.1007/978-3-642-20212-4>
- [33] LUBETZKY, E. and ZHAO, Y. (2017). On the variational problem for upper tails in sparse random graphs. *Random Structures Algorithms* **50** 420–436. MR3632418 <https://doi.org/10.1002/rsa.20658>
- [34] ROCKAFELLAR, R. T. (1997). *Convex Analysis. Princeton Landmarks in Mathematics*. Princeton Univ. Press, Princeton, NJ. MR1451876
- [35] ŠILEIKIS, M. and WARNKE, L. (2019). A counterexample to the DeMarco–Kahn upper tail conjecture. *Random Structures Algorithms* **55** 775–794. MR4025388 <https://doi.org/10.1002/rsa.20859>
- [36] VU, V. H. (2001). A large deviation result on the number of small subgraphs of a random graph. *Combin. Probab. Comput.* **10** 79–94. MR1827810 <https://doi.org/10.1017/S0963548399004095>
- [37] YAN, J. Nonlinear large deviations: Beyond the hypercube. Available at [arXiv:1703.08887](https://arxiv.org/abs/1703.08887).