Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdős–Rényi graphs

Fanny Augeri

October 23, 2018

Abstract

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

In this paper, we will be concerned with the following large deviation question: given a random vector $X$ in $\mathbb{R}^n$ and a measurable 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 [10], 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 [5]), the large deviation of sub-graph counts in Erdős–Rényi graphs - which were until then tackled using the graphon formalism in [11] - and arithmetic progressions. Later, Yan generalized in [36] 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 [18] nonlinear large deviation...
bounds which do not require any second order smoothness on the given function, and with a weaker notion of complexity of the gradient, assessed in terms of Gaussian width of its image. Using these, he partially recovered a known result of Basak and Mukherjee [5] on the mean-field approximation of the partition function of the Ising model, and improved the threshold of sparsity for the upper tail large deviation of triangle counts.

We will improve the error terms in the nonasymptotic bounds of [10], 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 [31] and [7] - 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 satisfy certain assumptions which include star graphs, thus strengthening the previous result of Basak and Mukherjee [5].

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 [19], 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 [13].

1.1 Main results

We begin with introducing some definitions. Let $\mu$ be a probability measure on $\mathbb{R}^n$ whose support is not included in a hyperplane. Let $\Lambda_\mu$ be the logarithmic Laplace transform of $\mu$, 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 $\Lambda_\mu$ is a strictly convex function which is $C^\infty$ on the interior of its domain denoted by $\mathcal{D}_{\Lambda_\mu}$. We define $\Lambda_\mu^*$, the Legendre transform of $\Lambda_\mu$, 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 [33, Section 26], we will say that a convex function $\Lambda : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is essentially smooth if the interior of its domain $\mathcal{D}_\Lambda$ is non-empty, $\Lambda$ is differentiable on $\mathcal{D}_\Lambda^0$, and steep, that is, for any $\lambda_k \in \mathcal{D}_\Lambda^0$, converging to a point on the boundary of $\mathcal{D}_\Lambda$,

$$\lim_{k \to +\infty} ||\nabla \Lambda(\lambda_k)|| = +\infty.$$
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 $\Lambda_\mu$ is \textit{essentially smooth}, then by [33, Theorem 26.5], we know that $\Lambda_\mu^*\ast$ is also an essentially smooth function, the map $\nabla \Lambda_\mu$ is one-to-one onto $D_{\Lambda_\mu}^\circ$, the interior of the domain of $\Lambda_\mu^*$, and the maps $\nabla \Lambda_\mu$ and $\nabla \Lambda_\mu^*$ are inverse maps from one another.

For any $\lambda \in D_{\Lambda_\mu}^\circ$, one can define the measure,

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

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

$$y = \nabla \Lambda_\mu(\lambda) = \int x d\mu_y(x).$$  \hspace{1cm} (2)

Thus, the family of measures $\mu_y$ represents the collection of tilts of $\mu$, indexed by their barycenters.

Our first result is the following non-asymptotic version of Varadhan’s lemma (see [16, Theorem 4.3.1]). This will be instrumental in our treatment of the mean-field behavior of the Ising model, see §1.2.1.

1.1 Proposition. Assume $\mu$ is compactly supported. Denote by $K$ the convex hull of its support, and by $D$ its diameter. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. Let $\delta > 0$ and $D_\delta$ be a $\delta/D$-net of the convex hull of $\nabla f(K)$ for the $\ell^2$-norm. Then,

$$\log \int e^f d\mu \leq \sup \{ f - \Lambda_\mu^* \} + \log |D_\delta| + \delta.$$

Our approximation of the partition function is similar to the one of Chatterjee-Dembo (see [10, Theorem 1.6]), except that we do not have any other error terms but the one coming from the cardinality of the net $D_\delta$. Unlike Eldan’s result (see [18, Corollary 2]), we are considering a stronger notion of complexity - which is of deterministic nature - but in return we can allow general compactly supported measures, which can be indifferently product or non-product measures.

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 a lower bound by just changing the barycenter of the underlying reference measure.

1.2 Lemma. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a non-negative measurable function. Then,

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

where $\mu_y$ is defined in (1).

1.3 Remark. Note the gap between the lower bound of Lemma 1.2 and the main term $\sup \{ f - \Lambda_\mu^* \}$ in the upper bound of Proposition 1.1.

When $\mu$ is the uniform measure on the discrete hypercube, one can consider, given a function $f : \{0, 1\}^n \to \mathbb{R}$, its natural harmonic extension, and the just mentioned gap resolves itself. One can view this fact as at the heart of Eldan’s approach [18]. For further discussion on the general case, we refer the reader to Remark 2.2.
We will now state our nonlinear large deviation upper bound. We recall from [16, Section 1.2] that a sequence of random variables \((Z_n)_{n \in \mathbb{N}}\) taking value in some topological space \(X\) equipped with the Borel \(\sigma\)-field \(\mathcal{B}\), satisfies a large deviations principle (LDP) with speed \(v_n\), and rate function \(I: X \to [0, +\infty]\), if \(I\) is lower semicontinuous, \(v_n\) increases to infinity, and for all \(B \in \mathcal{B}\),

\[
-\inf B^o \leq \liminf_{n \to +\infty} \frac{1}{v_n} \log \mathbb{P} (Z_n \in B) \leq \limsup_{n \to +\infty} \frac{1}{v_n} \log \mathbb{P} (Z_n \in B) \leq -\inf B, \tag{3}
\]

where \(B^o\) denotes the interior of \(B\) and \(\overline{B}\) the closure of \(B\). We recall that \(I\) is lower semicontinuous if its \(t\)-level sets \(\{x \in X : I(x) \leq t\}\) 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 [16, (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 \(\delta > 0\),

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

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

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

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

1.4 Proposition. Let \(X\) be a random vector in \(\mathbb{R}^n\) sampled according to a probability measure \(\mu\) such that \(\Lambda_\mu\) is essentially smooth. Let \(f: \mathbb{R}^n \to \mathbb{R}\) be a measurable function. Define,

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

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

\[
\mathbb{P} (X \notin K) \leq e^{-r}. \tag{4}
\]

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

\[
\forall x, y \in K, \ f(x) - f(y) \leq \sup_{\lambda \in W} \langle \lambda, x - y \rangle,
\]

Denote by \(D\) the diameter of \(K\). Let \(\delta > 0\) and let \(D_\delta\) be a \(\delta/D\)-net of \(W\) for the \(\ell^2\)-norm. Then,

\[
\log \mathbb{P} (f(X) \notin V_\delta (\{I \leq r\})) \leq -r + \log |D_\delta| + \log \left(\frac{\kappa LD}{\delta}\right) + 3,
\]

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

1.5 Remark. The wording of Proposition 1.4 is intended to include non-smooth functions which we will encounter in the applications. It is particularly relevant for functions \(f\) which are linear combinations of non-smooth convex functions, for which we can take \(W\) to be Minkowski sums of sets of subdifferentials of the convex functions involved in the decomposition.
1.6 Remark. The tightness assumption (4) in the Proposition 1.4 is automatically satisfied for product measures. Indeed, we know from [16, Lemma 5.1.14] that if \( \mu \) 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

\[
P(\Lambda_\mu^*(X) > \kappa) \leq e^{-r}.
\]

Turning our attention to the lower bound, one can always get a large deviations lower bound by finding a tilting of the underlying measure \( \mu \) which transforms a large deviation event for \( \mu \) into a typical event for \( \mu_h \). (This idea is prevalent in the large deviations literature, and is also the basis for the proof of Lemma 1.2). The following lower bound, which builds on [17, p. 64] quantifies this idea. First, note that differentiating twice \( \Lambda_\mu \), one obtains the covariance of the measure \( \mu_y \), that is,

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

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

1.7 Lemma. Let \( V \subset \mathbb{R}^n \) be a measurable subset, and \( \lambda \in D^\mu_{\Lambda_\mu} \). Then,

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

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

Proof. Changing the measure \( \mu \) into \( \mu_y \), we have

\[
\mu(V) = \int_V e^{-(\lambda, x - \Lambda_\mu(\lambda))} d\mu_y(x) = e^{-\Lambda_\mu^*(y)} \int_V e^{-(\lambda, x - y)} d\mu_y(x).
\]

By Jensen’s inequality we deduce,

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

Using Cauchy-Schwarz inequality we get,

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

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 1.2 and 1.7) 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 \[5\] [18, section 1.3], [22], [21], 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 \( H_n \) denote the set of Hermitian matrices of size \( n \). For any \( A \in H_n \), we denote by \( \mu_A \) the empirical distribution of its eigenvalues, defined by,

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

where \( \lambda_1, ..., \lambda_n \) are the eigenvalues of \( A \).

**1.8 Proposition.** Let \( A_n \) be a sequence of Hermitian matrices whose empirical distribution of eigenvalues \( \mu_{A_n} \) converges weakly to a Dirac at 0. Assume further that \( \limsup_{n} n^{-1} \text{tr} A_n^2 < +\infty \) and that for any \( i \in \{1, ..., n\} \), \( (A_n)_{i,i} = 0 \). Let \( \mu \) be the uniform probability measure on the spin configurations \( \{-1, 1\}^n \). 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 1.2. Let

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

The following lemma taken from [5, Lemma 3.4, Remark 4.1] computes the complexity of the gradient of \( f \) under the mean-field assumption we made.

**1.9 Lemma** ( [5, Lemma 3.4, Remark 4.1]), Under the assumption of Proposition 1.8, 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 \( \delta > 0 \),

\[
\log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) \leq \sup_{x \in [-1,1]^n} \{ \langle x, A_n x \rangle - \Lambda^*_\mu(x) \} + \log |\mathcal{E}_{\delta\sqrt{n}}| + 4\delta n, \tag{5}
\]

which gives the claim. \( \square \)

**1.10 Remark.** Instead of using Lemma 1.9 to bound the error term in (5), one can also use Sudokov’s minoration (see [30, Theorem 3.18 ]), which gives that there exists a numerical constant \( \kappa > 0 \) such that for any \( \delta > 0 \),

\[
\delta \sqrt{n} \sqrt{\log |\mathcal{E}_{\delta\sqrt{n}}|} \leq \kappa \mathbb{E} \sup_{x \in [-1,1]^n} \langle A_n x, \Gamma \rangle.
\]
where $\Gamma$ is a standard Gaussian vector in $\mathbb{R}^n$. But, using twice Cauchy-Schwarz inequality, one obtains

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

where $|| ||_2$ denotes the Hilbert-Schmidt norm. Thus, for any $\delta > 0$,

$$\log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) \leq \sup_{x \in [-1,1]^n} \{ \langle x, A_n x \rangle - \Lambda^*_\mu(x) \} + \kappa \frac{||A_n||_2^2}{\delta} + 4\delta n.$$

Optimizing on $\delta > 0$, we get

$$\log \int e^{\langle \sigma, A_n \sigma \rangle} d\mu(\sigma) \leq \sup_{x \in [-1,1]^n} \{ \langle x, A_n x \rangle - \Lambda^*_\mu(x) \} + C \left( ||A_n||_2 n \right)^2,$$

for some numerical constant $C > 0$, thus giving another proof of the result of Jain, Koehler and Risteski [21].

### 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 [19], in the context of large deviation of the largest eigenvalue of Wigner matrices.

To state our result we need to introduce some notations. Denote by $\mathcal{H}^{(\beta)}_n$ the set of Hermitian matrices when $\beta = 2$, and symmetric matrices when $\beta = 1$, of size $n$, which we will short-hand 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], [4, Theorem 2.5]), we know that

$$\mu_{X/\sqrt{n}} \xrightarrow{n \to \infty} \mu_{sc},$$

in probability, where $\xrightarrow{\ast}$ denotes the weak convergence, and $\mu_{sc}$ is the semi-circular law defined by

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

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

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

then we have the convergence of the moments of $\mu_{X/\sqrt{n}}$ towards the ones of the semi-circular law, that is for any $d \in \mathbb{N}$,

$$\frac{1}{n} \operatorname{tr} \left( \frac{X}{\sqrt{n}} \right)^d \xrightarrow{n \to +\infty} \begin{cases} C_{d/2} & \text{if } d \text{ is even,} \\ 0 & \text{if } d \text{ is odd,} \end{cases}$$

where $C_{d/2}$ depends on $d$.
in probability, where in the case \( d \) is even, \( C_{d/2} = \frac{1}{d + 1} \binom{d/2}{d/2} \) is the \( \left( \frac{d}{2} \right) \)th Catalan number. The question we investigate is the one of the large deviations of these traces around the moments of the semi-circular 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 Proposition 1.4. To this end, we introduce a class of random Hermitian matrices satisfying a certain convex concentration property.

1.11 Definition. We say that a random Hermitian matrix \( X \) satisfies the convex concentration property with constants \( (\kappa, C) \) for some \( \kappa, C > 0 \), if for any \( f : H_n \to \mathbb{R} \) convex 1-Lipschitz function with respect to the Hilbert-Schmidt norm, \( \mathbb{E} f(X) \) exists and for any \( t > 0 \),

\[
P(\left| f(X) - \mathbb{E} f(X) \right| > t) \leq C \exp\left(-\kappa t^2\right).
\]

1.12 Remark. A random Hermitian matrix whose law satisfies a log-Sobolev inequality has normal concentration by [29, 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 [29, Theorem 5.2]. Due to the tensorization property of the log-Sobolev inequality (see [29, 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 are the one with bounded entries, due either Talagrand’s inequality [29, Corollary 4.10] or due to a result by the transportation method [32, Theorem 8.6] (the latter having the advantage of directly measuring deviations from the mean).

In the following, we denote by \( \Lambda \) the logarithmic Laplace transform of the law of \( X \), that is,

\[
\forall H \in H_n, \quad \Lambda(H) = \log \mathbb{E} e^{\text{tr} X H},
\]

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

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

1.13 Theorem. Let \( d \geq 3 \), and \( l > d \) 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 \( \mathbb{R}^n \),

\[
\lim sup_{n \to +\infty} \frac{1}{n^{l+\frac{1}{2}}} \log P\left( \frac{1}{n} \text{tr}(X/\sqrt{n})^d \in F \right) \leq -\inf_F I_+,
\]

and for any open subset \( O \) of \( \mathbb{R}^n \),

\[
\lim inf_{n \to +\infty} \frac{1}{n^{l+\frac{1}{2}}} \log P\left( \frac{1}{n} \text{tr}(X/\sqrt{n})^d \in O \right) \geq -\inf_O I_-,
\]

where for any \( x \in \mathbb{R} \),

\[
I_+(x) = \sup_{\delta > 0} \liminf_{n \to +\infty} I_{n,\delta}(x),
\]

\[
I_-(x) = \sup_{\delta > 0} \limsup_{n \to +\infty} I_{n,\delta}(x),
\]

8
and for any \( n \in \mathbb{N}, \delta > 0, \)
\[
I_{n,\delta}(x) = \inf \left\{ \frac{\Lambda^* (Y)}{n^{1+\frac{d}{2}}} : \left| \frac{1}{n} \text{tr}(Y/\sqrt{n})^d - \mu_{sc}(x^d) - x \right| < \delta, Y \in \mathcal{H}_n \right\}.
\]

1.14 Remark. Together with the previous Remark 1.12, we see that any Wigner matrix with bounded entries satisfies the assumptions of Theorem 1.13.

We will now specify our result to Wigner matrices with sharp sub-Gaussian tails. Following Guionnet and Husson [19], we say that a real or complex random variable \( \xi \) has sharp sub-Gaussian tails if,
\[
\forall z \in \mathbb{C}, \Lambda_{\xi}(z) = \log \mathbb{E} e^{\Re(z\xi)} \leq \frac{1}{2} \langle z, \Sigma z \rangle,
\]
where \( \Sigma \) is the covariance matrix of \((\Re \xi, \Im \xi)\), and \( \langle.,. \rangle \) is the standard inner product in \( \mathbb{C} \).

1.15 Corollary. Let \( d \geq 3 \). 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 1.13, 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 \( n^{1+\frac{d}{2}} \), and good rate function \( J_d \). If \( d \) is even, \( J_d \) is given by,
\[
\forall x \in \mathbb{R}, J_d(x) = \begin{cases} \frac{\beta}{4} (x - C_{d/2})^\frac{d}{2} & \text{if } x \geq C_{d/2}, \\ +\infty & \text{otherwise,} \end{cases}
\]
where \( C_{d/2} \) denotes the \((\frac{d}{2})^{th}\) Catalan number, and if \( d \) is odd,
\[
\forall x \in \mathbb{R}, J_d(x) = \frac{\beta}{4} |x|^\frac{d}{2},
\]
where \( \beta = 1 \) if \( X \) is real symmetric.

1.16 Remark. 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 [19, 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 \( i \in \{1,\ldots,n\} \).

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 sub-graph counts are well understood now due to the work of Chatterjee and Varadhan [11] on the large deviation 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 [35], [26], [9], [15], that for any $p \geq 1/n$,
\[
\log \mathbb{P}(X_{K_3} \geq u\mathbb{E}X_{K_3}) = -\min(n^2p^2 \log(1/p), n^2p^3),
\]
and then generalized in [14] 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 [24]. In particular, denoting by $\Delta$ is the highest degree in $H$, they proved in [24, Theorems 1.2,1.5], that
\[
-n^2p^2 \log(1/p) \lesssim \log \mathbb{P}(X_H \geq u\mathbb{E}X_H) \lesssim -n^2p^\Delta,
\]
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 [14] by Kahn and DeMarco was recently disproved by Šileikis and Warnke in [34].

Working instead with homomorphism densities, that is, if $H$ has vertex set $V$ and edge set $E$,
\[
t(H, X) = \frac{1}{|V|^2} \sum_{x \in \mathbb{V}} \prod_{(v, w) \in E} X_{x(v)}(x(w)),
\]
Chatterjee and Dembo showed in [10] 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,
\[
-\log \mathbb{P}(t(H, X) \geq u\mathbb{E}t(H, X)) \sim \varphi_{n,H}(u),
\]
with
\[
\varphi_{n,H}(u) = \inf \{ \Lambda^*_n(Y) : t(H, Y) \geq u\mathbb{E}t(H, Y), Y \in \mathcal{H}^0_n \},
\]
where $\mathcal{H}^0_n$ is the set of symmetric matrices of size $n$ with null diagonal coefficients, and $\Lambda^*_n(Y) = \sum_{i<j} I_p(Y_{i,j})$, where
\[
\forall x \in [0,1], I_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p},
\]
and $+\infty$ otherwise. The equivalent (6) was shown to hold in [10, 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 (6) found was $n^{-1/42}$, up to logarithmic factor, and was pushed later on by Eldan in [18] 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 (6) holds as soon as the variational problem (7) gives the right order of the upper tail, that is, with the help of [31], $np \gg \log n$.

Very recently, Cook and Dembo [13] proved that the nonlinear large deviations of $t(H, X)$ holds 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 $d = 3$.

1.17 Proposition. Let $p$ such that $p = o(1)$ and $\log^4 n = o(np^2)$. Denote by $v_n = n^2p^2 \log(1/p)$. For any $t \geq 1$,
\[
\log \mathbb{P}(\text{tr}X^d \geq tn^d p^d) \leq -\varphi_n(t) + o(v_n),
\]
Indeed, we know by the work of Janson, Łuczak and Ruciński [129, 309] observes a change of speed in the large deviation lower tail of cycles counts. Corresponds to pick the top eigenvector. Lengths (with possibly multiple changes of speeds as the sparsity increases). Thus a change of speed happens at

\[ n \]

(\( n \)) 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 (\( n \)))-dimensional sphere, as it corresponds to pick the top eigenvector.

With this knowledge of the optimizers, one can obtain the complementary lower bound, so that together with Proposition 1.17, we get, for any \( t \geq 1 \),

\[
\lim_{n \to +\infty} \frac{1}{n} \log P(\text{tr} X^d \geq tn^2 p^d) = -\Phi(t),
\]

for any \( n^{-1/2} \log n \ll p \ll 1 \). The edge-probability \( n^{-1/2} \) appears to be critical for cycles in many senses. Note that below this threshold, the anti-clique 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 (\( n \))-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 [23], and Janson and Warnke [25], that in the regime where \( p = o(1) \), for any finite graph \( H \), and any \( 0 < u \leq 1 \),

\[
\log P(X_H \leq uEX_H) \asymp -(1 - u)^2 \Phi_H,
\]

where \( \Phi_H = \min \{ EX_J : J \text{ sub-graph of } H \} \). Thus, the speed of the lower tail is driven by the “least expected” sub-graph of \( H \). For example for \( H = K_3 \), this implies that,

\[
\log P(X_H \leq (1 - u)EX_H) \asymp -u^2 \min(n^3 p^3, n^2 p).
\]

Thus a change of speed happens at \( n^{-1/2} \), and similarly for cycles of greater lengths (with possibly multiple changes of speeds as the sparsity increases).

In [13, 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 [27] all the way down to \( p \gg n^{-1/2} \).
2 Nonlinear large deviation upper bounds

We give here a proof of the two Propositions 1.1 and 1.4.

Proof of Proposition 1.1. Let \( x, y \in K \). We have by the mean-value theorem,

\[
f(x) = f(x) - f(y) + \Lambda_\mu^*(y) + f(y) - \Lambda_\mu^*(y) \\
\leq \sup_{t \in K} \langle \nabla f(t), x - y \rangle + \Lambda_\mu^*(y) + \sup \{ f - \Lambda_\mu^* \}.
\]

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,

\[
f(x) \leq h_W(x - y) + \Lambda_\mu^*(y) + \sup \{ f - \Lambda_\mu^* \}.
\]

Optimizing in \( y \), we deduce that

\[
f(x) \leq \inf_{y \in K} \{ h_W(x - y) + \Lambda_\mu^*(y) \} + \sup \{ f - \Lambda_\mu^* \},
\]

Using the minimax theorem (see [12, Theorem 4.36]), we have

\[
\inf_{y \in K} \{ h_W(x - y) + \Lambda_\mu^*(y) \} = \sup_{\lambda \in W} \inf_{y \in K} \{ \langle \lambda, x - y \rangle + \Lambda_\mu^*(y) \}
= \sup_{\lambda \in W} \{ \langle \lambda, x \rangle - \Lambda_\mu(\lambda) \},
\]

where we used the fact that \( \Lambda_\mu^*(y) = +\infty \) if \( y \notin K \) by the Hahn-Banach theorem, and that \( \Lambda_\mu \) is the Legendre transform of \( \Lambda_\mu^* \) by [12, Theorem 4.21]. Therefore,

\[
f(x) \leq \sup_{\lambda \in W} \{ \langle \lambda, x \rangle - \Lambda_\mu(\lambda) \} + \sup \{ f - \Lambda_\mu^* \}.
\]

Thus,

\[
\log \int e^f d\mu \leq \log \int e^{\sup_{\lambda \in W} \{ \langle \lambda, x \rangle - \Lambda_\mu(\lambda) \}} d\mu(x) + \sup \{ f - \Lambda_\mu^* \}. \quad (10)
\]

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 \( \ell^2 \)-norm. Let now \( \delta > 0 \) and let \( \mathcal{D}_\delta \) be a \( \delta/D \)-net of \( W \). 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. \( \square \)

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

\[
\partial_i \Lambda(\lambda) = \int \langle x, e_i \rangle d\mu(x),
\]
with \( y = \nabla \Lambda (\lambda) \) and \( \mu_y \) is defined in (1). Therefore, if \( \text{supp} \mu \subset [0, 1]^n \), \( \partial \Lambda (\lambda) \in [0, 1] \). 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 \( \ell^1 \)-norm. In the last step (10) of the proof of Proposition 1.1, if one takes \( \delta > 0 \) and \( \mathcal{F}_\delta \) to be a \( \delta \)-net for the \( \ell^1 \)-norm, one obtains,
\[
\log \int e^{\sup_{\lambda \in W} \langle \langle \lambda, x \rangle - \Lambda \mu (\lambda) \rangle} \, d\mu (x) \leq \delta + \log |\mathcal{F}_\delta|.
\]
This yields an error term in the approximation of the partition function getting closer to the one recently found in [3, Proposition 5.1] as a consequence of a structural result on probability measures on product spaces.

2.2 Remark. As observed in Remark 1.3, 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 1.2 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 \( \Lambda^* \mu \),
\[
\int f \, d\mu_y \simeq f (y),
\]
where \( \mu_y \) is defined in (1). Indeed, if \( \mu \) is the \( n \) fold product measure \( \nu^n \), with \( \nu \) compactly supported for example, we note that by the mean-value theorem, we have,
\[
\left| \int \left( f \, d\mu_y - f (y) \right) \right| \leq \int \sup_{t \in K} |\langle \nabla f (t), x - y \rangle| \, d\mu_y (x),
\]
where \( K \) denotes the convex hull of the support of \( \mu \). But, as \( \mu_y \) has barycenter \( y \), and is also a product measure with marginals having the same support as the one of \( \nu \), we can deduce by the Majorization Theorem (see [30, Theorem 12.16]),
\[
\left| \int \left( f \, d\mu_y - f (y) \right) \right| \leq c \int \sup_{t \in K} |\langle \nabla f (t), x \rangle| \, d\gamma^n (x),
\]
where \( c \) is a positive constant depending on the diameter of the support of \( \nu \), and \( \gamma^n \) is the standard Gaussian measure on \( \mathbb{R}^n \).

We now turn our attention to the proof of Proposition 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 Proposition 1.1. We will actually refrain ourselves from using Chebytchev’s inequality, and prefer to work directly on the probability of deviations, in contrast with the path followed by Chatterjee and Dembo in [10, section 4, proof of Theorem 1.1].

**Proof of Proposition 1.4.** For simplicity, we assume that \( f \) is continuously differentiable. We also write \( \Lambda, \Lambda^* \) as shorthands for \( \Lambda \mu \) and \( \Lambda^* \mu \) respectively. Let \( r > 0 \) and \( \kappa > 0 \) such that
\[
\mathbb{P}(\Lambda^*(X) > \kappa) \leq e^{-r}.
\]
Let \( K = \{ \Lambda^* \leq \kappa \} \). Define the measure,
\[
\mathbb{P}_K = \mathbb{P}(\cdot \cap \{ X \in K \}).
\]
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,
\[
\mathbb{P}_K \{ f (X) \notin V_\delta (\{ I \leq r \}) \} \leq \mathbb{P}_K \{ \inf_{\Lambda^* (y) \leq r} |f (X) - f (y)| \geq \delta \}.
\]
As $K = \{ \Lambda^* \leq \kappa \}$ is a convex set and $\kappa \geq r$, we get by the mean-value theorem,

$$
\mathbb{P}_K \left( \inf_{\Lambda^{*}(y) \leq r} |f(X) - f(y)| \geq \delta \right) \leq \mathbb{P}_K \left( \inf_{\Lambda^{*}(y) \leq r, \lambda \in W} \sup_{t \in K} |(\nabla f(t), X - y)| \geq \delta \right).
$$

Let $W$ denote the convex hull of $\nabla f(K) \cup (-\nabla f(K))$, so that $W$ is convex and symmetric. We have,

$$
\mathbb{P}_K \left( f(X) \not\in V_{\delta}(\{I \leq r\}) \right) \leq \mathbb{P}_K \left( \inf_{\Lambda^{*}(y) \leq r, \lambda \in W} \sup_{\lambda, X - y \geq \delta} \right).
$$

We claim that we have the inequality,

$$
\mathbb{P}_K \left( \inf_{\Lambda^{*}(y) \leq r} \sup_{\lambda, X - y \geq \delta} \right) \leq \mathbb{P}_K \left( \inf_{z \in \partial W^\circ} \Lambda^*(X - z) \geq r \right), \quad (11)
$$

where $W^\circ$ is the polar set of $W$, that is, denoting by $h_W$ the support function of $W$, 

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

Indeed, arguing by contradiction, if

$$
\inf_{z \in \partial W^\circ} \Lambda^*(X - z) < r,
$$

then by continuity, there is some $z$ such that $h_W(z) < \delta$ and $\Lambda^*(X - z) < r$. Then, setting $y = X - z$, we have for any $\lambda \in W$, 

$$
\langle \lambda, X - y \rangle = \langle \lambda, z \rangle < \delta,
$$

which proves the inequality (11).

Fix $X \in K$. Observe that this implies in particular, taking $z = 0$, that

$$
\inf_{z \in \partial W^\circ} \Lambda^*(X - z) < +\infty.
$$

As $\partial W^\circ$ is a closed set and $\Lambda^*$ has compact level sets, the infimum is achieved at some $z_* \in \partial W^\circ$.

Since $\Lambda^*$ and $h_W$ are both convex functions, we deduce by Kuhn-Tucker Theorem (see [12, 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

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

Evaluating the function on the right-hand side at $z = 0$, we see that the non-triviality condition $(\eta, \theta) \neq (0, 0)$ implies that $\eta = 1$.

We claim now that $z_*$ lies in the interior of $\partial \Lambda^*$. Note first that the barycenter of $\mu$, which we denote by $\overline{\mu}$, is in the interior of $\partial \Lambda^*$. Indeed, by Jensen’s inequality,

$$
\forall \lambda \in \mathbb{R}^n, \Lambda(\lambda) \geq \langle \lambda, \overline{\mu} \rangle.
$$

Thus, $\overline{\mu} \in \partial \Lambda(0)$. Consequently by [33, Theorem 23.5], $0 \in \partial \Lambda^*(\overline{\mu})$. As $\Lambda$ is strictly convex and essentially smooth, $\Lambda^*$ is also strictly convex and essentially smooth by [33, Theorem 26.5]. Therefore $\overline{\mu} \in \partial \Lambda^*$.

Arguing again by contradiction, if $X - z_* \in \partial \Lambda^*$ then, by [33, Theorem 6.1], we can deduce that for any $t \in (0, 1], X - z_* - tu \in \partial \Lambda^*$, where $u = X - z_* - \overline{\mu}$, as we are taking a barycenter between an interior point and a boundary point of a convex set. Let,

$$
\forall z \in \mathbb{R}^n, F(z) = \Lambda^*(X - z) + \theta h_W(z).
$$
By convexity, denoting by \( v_t = X - z^* - tu \), we get
\[
F(z^*) \geq F(z^* + tu) + \langle \nabla \Lambda^*(v_t) - \theta \zeta, tu \rangle,
\]
where \( \zeta \in \partial h_W(z^* + tu) \). As \( \Lambda^* \) is strictly convex and achieves its infimum at \( \overline{m} \), we deduce that \( \langle \nabla \Lambda^*(v_t), u \rangle > 0 \) for any \( t < 1 \).

In the meantime, by Danskin’s formula (see [12, Corollary 10.23]), the subdifferential of the support function \( h_W \), is
\[
\partial h_W(z) = \{ \lambda^* \in W : h_W(z) = \langle \lambda^*, z \rangle \}.
\]
But \( W \) is a bounded set, therefore \( \zeta \) remains bounded as \( t \in (0, 1] \). Since \( \Lambda^* \) is steep, we deduce that for \( t \) close enough to 0,
\[
\langle \nabla \Lambda^*(v_t) - \theta \zeta, u \rangle > 0,
\]
which gives a contradiction.

Therefore, \( X - z^* \) is an interior point of \( D_{\Lambda^*} \). As \( \Lambda^* \) is differentiable on \( D_{\Lambda^*} \) and \( 0 \in \partial F(z^*) \), we have \( 0 \in \{ -\nabla \Lambda^*(X - z^*) \} + \partial h_W(z^*) \) by [12, Theorem 4.10]. Thus, there is a \( \lambda^* \in W \) satisfying
\[
\delta = \langle \lambda^*, z^* \rangle.
\]
Using that \( \nabla \Lambda \) and \( \nabla \Lambda^* \) are inverse maps, and \( 0 \in \delta W^\circ \), we have
\[
\Lambda^*(\nabla \Lambda(\theta \lambda)) = \Lambda^*(X - z^*) \leq \Lambda^*(X) \leq \kappa.
\]
Besides, using the convexity of \( \Lambda^* \), we have,
\[
\Lambda^*(X) - \Lambda^*(X - z^*) \geq \langle \nabla \Lambda^*(X - z^*), z^* \rangle.
\]
Using the fact that \( \langle \lambda^*, z^* \rangle = \delta \), we get
\[
\Lambda^*(X) \geq \theta \delta.
\]
In conclusion, we have \( \theta \leq \kappa/\delta \) and \( \lambda \in W_{\theta} \), where
\[
W_{\theta} = \{ \lambda \in W : \theta \lambda \in D_{\Lambda^*}, \nabla \Lambda(\theta \lambda) \in K \}.
\]
Therefore,
\[
\Lambda^*(X - z^*) = \langle \theta \lambda, X - z^* \rangle - \Lambda(\theta \lambda) \leq \sup_{0 \leq \theta \leq \theta_0} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta \},
\]
where \( \theta_0 = \kappa/\delta \) and where we used the fact that \( \langle \lambda, z^* \rangle = \delta \). Coming back to (11), we obtain
\[
P_K(\{ f(X) \notin V_\delta(\{ I \leq r \}) \}) \leq P_K(\{ \sup_{0 \leq \theta \leq \theta_0} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta \} \geq r \}).
\]
We are now ready to perform a net argument. We start by the net argument on \( \theta \). Let \( \mathcal{E} \) be a \( 1/(LD) \)-net of the interval \( [0, \theta_0] \), where
\[
L = \sup_{x \in K} ||\nabla f(x)||_{L^2} = \sup_{\lambda \in W} ||\lambda||_{L^2},
\]
the last equality being a consequence of the convexity of the norm. One can find a net \( \mathcal{E} \) such that,
\[
|\mathcal{E}| \leq \frac{2\kappa LD}{\delta}.
\]
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 $\theta' \leq \theta$,
\[ G(\theta) - G(\theta') \leq (\theta - \theta')LD. \]  
(14)

Indeed, note first that $(W_\theta)_{\theta \geq 0}$ is non-increasing for the inclusion as, for any $\theta \geq 0$ and $\lambda \in \mathbb{R}^n$ such that $\theta \lambda \in D_\lambda$,
\[ \frac{\partial}{\partial \theta} \Lambda^*(\nabla \Lambda(\theta \lambda)) = \langle \nabla \Lambda^*(\nabla \Lambda(\theta \lambda)), \nabla^2 \Lambda(\theta \lambda) \rangle \lambda \]
\[ = \theta(\lambda, \nabla^2 \Lambda(\theta \lambda) \lambda) \geq 0, \]
where we used the fact that $\nabla \Lambda$ and $\nabla \Lambda^*$ are inverse maps from one another. Let $\theta' \leq \theta$. There is some $\lambda \in W_\theta$ such that,
\[ G(\theta) - G(\theta') \leq (\theta - \theta')(\lambda, X) - \Lambda(\theta \lambda) + \Lambda(\theta' \lambda) - (\theta - \theta')\delta. \]
By convexity,
\[ G(\theta) - G(\theta') \leq (\theta - \theta')(\lambda, X - \nabla \Lambda(\theta' \lambda)). \]

Since $(W_\theta)_{\theta \geq 0}$ is non-increasing and $\lambda \in W_\theta$, we deduce that $\lambda \in W_{\theta'}$, which immediately yields (14). Thus, using a union bound, we get,
\[ \mathbb{P}_K \left( \sup_{\lambda \in W_\theta} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta \} \geq r \right) \]
\[ \leq \sum_{\theta \in \mathcal{E}} \mathbb{P}_K \left( \sup_{\lambda \in W_\theta} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) - \theta \delta \} \geq r - 1 \right). \]

Let us now fix $\theta \in \mathcal{E}$ and perform a net argument on $\lambda$. Let $D_\delta$ be a $\delta/D$-net of $W$ for the $\ell^2$-norm. On the event where $X \in K$, we observe that the concave function $\lambda \mapsto \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda)$ is $\theta D$-Lipschitz w.r.t the $\ell^2$-norm on $W_\theta$, since by definition of $\lambda \in W_\theta$, we have $\nabla \Lambda(\theta \lambda) \in K$. Thus,
\[ \mathbb{P}_K \left( \sup_{\lambda \in W_\theta} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) \} \geq r-1+\theta \delta \right) \leq \mathbb{P}_K \left( \sup_{\lambda \in D_\delta} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) \} \geq r-1 \right). \]

Using finally a union bound we get,
\[ \mathbb{P} \left( \sup_{\lambda \in D_\delta} \{ \langle \theta \lambda, X \rangle - \Lambda(\theta \lambda) \} \geq r-1 \right) \leq |D_\delta| e^{-r+1}, \]
where we used the fact that for any fixed $\xi \in \mathbb{R}^n$, and $t \geq 0$,
\[ \mathbb{P}(\langle \xi, X \rangle - \Lambda(\xi) \geq t) \leq e^{-t}, \]
by Chernoff’s inequality. Coming back to (13),
\[ \mathbb{P}_K \left( f(X) \notin V_\delta(\{ I \leq r \}) \right) \leq |\mathcal{E}||D_\delta| e^{-r+1}. \]
As we assumed that $\mathbb{P}(X \notin K) \leq e^{-r}$, we have
\[ \mathbb{P}(f(X) \notin V_\delta(\{ I \leq r \})) \leq |\mathcal{E}||D_\delta| e^{-r+1} + e^{-r} \leq 2|\mathcal{E}||D_\delta| e^{-r+1}, \]
which gives us the claim.
3 Large deviation of traces of Wigner matrices

In this section, we will give a proof of Theorem 1.13 and Corollary 1.15.

3.1 Large deviation upper bound

We start with the large deviation upper bound of Theorem 1.13. 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 [2, 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 made, we will be able to use efficiently Proposition 1.4.

We denote for any \( k \in \{1, \ldots, n\} \) and \( Y \in \mathcal{H}_n \), by \( \text{tr}_{[k]} Y \) the truncated trace:

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

where \( \lambda_1(Y), \ldots, \lambda_n(Y) \) are the eigenvalues of \( Y \) in non-increasing order. We also define, for \( d \) even,

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

whereas for \( d \) odd,

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

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

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

\[
\lim_{n \to +\infty} \frac{1}{n^{1+\frac{d}{k}}} \log P\left( \left| \frac{1}{n} \text{tr} \left( \frac{X}{\sqrt{n}} \right)^p - f_k(X) - \mu_{sc}(x^p) \right| > t \right) = -\infty.
\]

3.1.1 Concentration inequalities

In order to prove Lemma 3.1, we will develop some deviations 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 1.11. The proof of these inequalities will follow the now classical path lead by the work of Guionnet and Zeitouni in [20].

3.2 Proposition. Let \( X \) be a random Hermitian matrix satisfying the convex concentration property for some constants \((\kappa, C)\). Let \( k \in \{1, \ldots, n\} \) and let \( f: \mathbb{R} \to \mathbb{R} \) be a convex \( 1 \)-Lipschitz function which achieves its maximum on \( \mathbb{R} \). For any \( t > 0 \),

\[
P\left( \left| \text{tr}_{[k]} f(X) - \text{Etr}_{[k]} f(X) \right| > t \right) \leq C \exp \left( -\frac{\kappa t^2}{k} \right),
\]

where \( \text{tr}_{[k]} \) is the truncated trace defined in (15).

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.
3.3 Lemma. Let \( f : \mathbb{R} \to \mathbb{R} \) be a convex function which achieves its infimum on \( \mathbb{R} \). Let \( k \in \{1, ..., n\} \). The function \( T_f \) defined by,

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

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 \( \tilde{f} = f(. + x_0) - f(x_0) \), 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.

3.4 Lemma. Let \( f : \mathbb{R} \to \mathbb{R} \) be a convex function. Assume \( f(0) = 0 \) and \( 0 \in \partial f(0) \). Let \( \zeta : \mathbb{R} \to \mathbb{R} \) be a function such that,

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

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

\[
T_f(X) = \sup_{\text{rank} Y \leq k} \{ \text{tr}(Y) + \text{tr}\zeta(Y)(X - Y) \}.
\]

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 [33, Theorem 23.4] that for every \( x \in \mathbb{R}, \partial f(x) \neq \emptyset \), which justifies the existence of \( \zeta \). Let \( \lambda_1, ..., \lambda_n \) be the eigenvalues of \( X \) such that \( f(\lambda_1) \geq ... \geq f(\lambda_n) \). Let \( u_1, ..., u_n \) 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,

\[
T_f(X) = \text{tr}(Y) + \text{tr}\zeta(Y)(X - Y),
\]

using the orthogonality of the eigenvectors. Therefore only the inequality is left to prove. Let \( Y \in \mathcal{H}_n \) with rank less that \( k \), and denote by \( e_1, ..., e_n \) an orthonormal basis of eigenvectors such that \( e_{k+1}, ..., e_n \) are in the kernel, and by \( \mu_1, ..., \mu_k \) the eigenvalues associated to \( e_1, ..., e_k \). From the convexity of \( f \), we have for any \( j \in \{1, ..., n\}, i \in \{1, ..., k\} \),

\[
f(\lambda_j) \geq f(\mu_i) + \zeta(\mu_i)(\lambda_j - \mu_i).
\]

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

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

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

\[
\text{tr}\zeta(Y)(X - Y) = \sum_{i=1}^n |\langle e_i, \zeta(Y)(X - Y)e_i \rangle|
\]

\[
= \sum_{i=1}^n |\langle \zeta(Y)e_i, (X - Y)e_i \rangle|
\]

\[
= \sum_{i=1}^k |\langle \zeta(Y)e_i, (X - Y)e_i \rangle|,
\]

18
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 (18) now over $i \in \{1, \ldots, k\}$, 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

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

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 (17) holds, and that $T_f$ is a convex function.

Assume further that $f$ is 1-Lipschitz. Observe that 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 (17) that $T_f$ is $\sqrt{k}$-Lipschitz w.r.t the Hilbert-Schmidt norm.

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

**3.5 Proposition.** Let $X$ be a random Hermitian matrix satisfying the convex concentration property with constants $(\kappa, C)$. Denote by $\lambda_1(X)$ its top eigenvalue. For any $M \geq 4\mathbb{E}(\lambda_1(X))_+$, and $1 \leq k \leq n$,

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

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 $1 \leq k \leq n$,

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

**Proof.** Let $f(x) = \frac{1}{M}(x - M/2)_+$ for any $x \in \mathbb{R}$. Since $f(x) \geq 1$ for $x \geq M$, we have, denoting by $\lambda_1(X), \ldots, \lambda_n(X)$ the eigenvalues of $X$ in non-increasing 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
\[
P(N[M, +\infty) \geq k) \leq Ce^{-\frac{\kappa k M^2}{16}},
\] (19)
which gives the first claim.

Since by definition \([k/2] - 1 < k/2\), we have using a union bound,
\[
P(N((-M, M)^c) \geq k) \leq P(N(-\infty, -M] \geq l) + P(N[M, +\infty) \geq l),
\]
where \(l = [k/2]\). For any \(M \geq 4\|X\|\), we obtain by applying inequality (19) to \(X\) and \(-X\) alternatively,
\[
P(N((-M, M)^c) \geq k) \leq 2Ce^{-\frac{\kappa k M^2}{32}},
\]
which gives the second claim.

### 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 \(\frac{n}{\sqrt{d}} = o(k)\) 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
\[
f_{M}(x) = \begin{cases} x^d & \text{for } |x| \leq M, \\ dM^{d-1}(|x| - M) + M^d & \text{for } |x| > M. \end{cases}
\]

Let \(\lambda_1^*, ..., \lambda_n^*\) 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 \(f_{M}\) is \(dM^{d-1}\)-Lipschitz, applying Proposition 3.2, we deduce that for any \(t > 0\),
\[
P\left(\left| \sum_{i=k+1}^{n} f_{M}(\lambda_i^*) - \text{tr} f_{M}(X/\sqrt{n}) - \text{tr}_k f_{M}(X/\sqrt{n}) \right| > tn \right) \leq 2C \exp\left(-\frac{\kappa n^2 t^2}{4d^2 M^2 (d-1)}\right).
\] (20)

On one hand,
\[
\left| \text{tr} \sum_{i=k+1}^{n} \lambda_i^d - \text{tr} f_{M}(X/\sqrt{n}) \right| \leq \text{tr} \sum_{i=k+1}^{n} |\lambda_i^d| 1_{|\lambda_i| \geq M}
\]
\[
\leq \frac{1}{M} \text{tr}|X/\sqrt{n}|^{d+1} = O\left(\frac{n}{M}\right),
\]
where we used the fact that by Wigner’s Theorem, \(\text{tr}|X/\sqrt{n}|^{d+1} = O(n)\).

On the other hand,
\[
\left| \text{tr} \sum_{i=k+1}^{n} \lambda_i^d - \text{tr}(X/\sqrt{n})^d \right| \leq k \text{E}\|X/\sqrt{n}\|^d = O(k),
\]
where we used the fact that by Wigner’s Theorem, \(\text{tr}|X/\sqrt{n}|^{d+1} = O(n)\).
as \( E||X/\sqrt{n}||^d = O(1) \) using the convex concentration property and a net argument as in [1, Exercice 2.1.19]. Therefore, if \( k = o(n) \) and \( M \) goes to infinity with \( n \), then as \( E\frac{1}{n} tr(X/\sqrt{n})^d \) converges to \( \mu_{sc}(x^d) \), we have,

\[
\left| \mu_{sc}(x^d) - E\left( \frac{1}{n} \sum_{i=k+1}^{n} f_M(\lambda_i^*) \right) \right| \xrightarrow{n \to +\infty} 0. \tag{21}
\]

Fix \( t > 0 \). Now let \( M \) go to infinity with \( n \) such that \( n^{1+2/d} = o(n^2/M^2(p-1)) \), that is \( M^2 = o(n^a) \) with \( a = (d-2)/(d(d-1)) \). The above inequality (20) and the convergence (21) give

\[
\lim_{n \to +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}\left( \left| \frac{1}{n} \sum_{i=k+1}^{n} f_M(\lambda_i^*) - \mu_{sc}(x^d) \right| > t \right) = -\infty. \tag{22}
\]

On the other hand, in order that

\[
\lim_{n \to +\infty} \frac{1}{n^{1+\frac{2}{d}}} \log \mathbb{P}\left( N([-R\sqrt{n}, R\sqrt{n}^c]) \geq k \right) = -\infty, \tag{23}
\]

we see from Proposition 3.5, as \( E||X|| = O(\sqrt{n}) \), that it is sufficient to take \( R \) going to \(+\infty\) with \( n \) such that \( n^{2/d} = o(kR^2) \). But we assumed that \( n^{\frac{2}{d}} = o(k) \), 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,

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

With this choice of \( M \), both estimates (22), and (23) with \( R = M \), hold. But then,

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

Therefore,

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

which together with (22) 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 (16), with \( n^{\frac{1}{\frac{d}{d-1}}} = o(k) \), 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 \( \text{tr}'(X^d) \). 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 \( n^{\frac{1}{\frac{d}{d-1}}} = o(k) \) 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 1.13 is a good rate function. It will follow from the next lemma.

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

\[
\bigcup_{n \in \mathbb{N}} F_n(\{c_n \leq \tau\}),
\]

is bounded. Then the functions defined as,

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

and

\[
\forall x \in \mathbb{R}, \ I_+(x) = \sup_{\delta > 0} \liminf_{n \to +\infty} I_{n,\delta}(x),
\]

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 \( x \in \mathbb{R} \),

\[
I^{(n)}(x) \leq \tau \iff \forall \delta > 0, \ \inf_{m \geq n} I_{m,\delta}(x) \leq \tau.
\]

Thus,

\[
\{I^{(n)} \leq \tau\} = \bigcap_{\epsilon > 0} \bigcup_{m \geq n} F_m(\{c_m \leq \tau + \epsilon\}).
\]

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. As \( X \) satisfies the convex concentration property defined in 1.11, its entries are in particular sub-Gaussian. Thus, there exists a constant \( c > 0 \) such that,

\[
\forall Y \in \mathcal{H}_n, \ \Lambda^*(Y) \geq \frac{1}{c^2} \text{tr} Y^2.
\]

But \(|\text{tr} Y|^{d/2} \leq \text{tr}|Y|^{d} \leq (\text{tr} Y^2)^{d/2}\), so that whenever

\[
\Lambda^*(Y) \leq r n^{1+\frac{d}{2}},
\]

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

**Upper bound.** We can now proceed with the proof of the upper bound. By [16, 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 \( (\mu_{\text{sec}}(x^d) + f_k(X))_{n \in \mathbb{N}} \) is exponentially equivalent to \( (\frac{1}{c^2} \text{tr} (X/\sqrt{n})^d)_{n \in \mathbb{N}} \) by Lemma 3.1. We will first make sure that the rate function we are going to obtain by applying Proposition 1.4 is the same as the one we are aiming for, that is:
3.7 Lemma. Assume that $k$ goes to $+\infty$ with $n$. For any $n \in \mathbb{N}$, $\delta > 0$, set

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

and let

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

For any $x \in \mathbb{R}$,

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

where $I_+$ is defined in Theorem 1.13.

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

$$\text{tr} Y^d - \text{tr}_{[k]} Y^d = o(m^{1+\frac{d}{2}}). \quad (25)$$

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

$$\lambda_l \leq l^{-\frac{1}{2}} O(m^{\frac{1}{2} - \frac{d}{2}}).$$

Thus,

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

which gives (25). Let now $\tau > 0$ 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,

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

From the observation (25) above, we deduce that for $n$ large enough,

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

Therefore,

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

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.

We come back to the proof of the upper bound of Theorem 1.13. Let $F$ be a closed subset of $\mathbb{R}$. We want to prove that,

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

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,

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

23
As $J_+$ is a good rate function, we can find a $\delta > 0$ such that

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

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

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

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

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

Again by Lemma 3.6 and arguing as we did before for $I_+$, we deduce that $J^{(n)}$ is a good rate function. Therefore, $\{J^{(n)} \leq r\}$ is a non-increasing sequence of compact subsets, so that for $n$ large enough,

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

Thus,

$$P(\mathbf{1}_{\{J_+ \leq r\}}) \leq P(\mathbf{1}_{\{J \leq r\}}).$$ 

We are now in the position of applying Proposition 1.4. First, we check that the tightness condition (4) is satisfied. By [16, Lemma 5.1.14], we know that for any $i,j$,

$$\mathbb{E}\mathbf{1}_{\Lambda^\ast(X_{i,j})} \leq 4^{(1+\frac{2}{3})},$$

using the fact that $(\mathbb{R}X_{i,j}, 3X_{i,j})$ are independent. Therefore, by Chernoff’s inequality,

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

which proves that (4) is fulfilled.

Let $K = \{\Lambda^\ast \leq 8n^2\}$. We will now bound the increments of $f_k$ on $K$. As noted before, the level sets of $\Lambda^\ast$ are included in Hilbert-Schmidt balls, more precisely,

$$K \subset 4cnB_2,$$

where $c$ is as in (24) and $B_2$ denotes the ball of radius 1 for the Hilbert-Schmidt norm. By Lemma 3.4, we know that if $f : \mathbb{R} \to \mathbb{R}$ is a differentiable convex function such that $f(0) = f'(0) = 0$. Then, the function

$$\forall Y \in \mathcal{H}_n, \ T_f(Y) = \text{tr}[\delta_\Gamma(Y)],$$

is convex and its subdifferential $\partial F(H)$ at some $Y \in \mathcal{H}_n$ satisfies,

$$\{f'(Y) : \text{rank}(Y) \leq k\} \subset \partial T_f(Y).$$

As by definition $f_k$ is a combination of at most two functions $T_f$ associated with the functions $x^d, x^d_\oplus$ and $x^d_\ominus$, we deduce that for any $X, Y \in 4cnB_2$,

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

(29)
where
\[ W = \{ H \in \mathcal{H}_n : \text{rank}(H) \leq 2k, ||H|| \leq c_0 n^{d-1} \}. \tag{30} \]

where \( c_0 \) is some positive constant depending on \( c \).

Note that by (27), the diameter of the level set \( K \) is bounded by \( 8cn \), and by (29), the Lipschitz constant of \( f_k \) on \( K \) is only polynomial in \( n \). Therefore, by Proposition 1.4,
\[
\log \mathbb{P}(f_k(X) \notin V_k([J \leq rn^{1+\frac{d}{2}}])) \leq -rn^{1+\frac{d}{2}} + \log \mathcal{N}(8neV, \delta B_2) + O(\log n), \tag{31}
\]

where \( \mathcal{N}(8neV, \delta B_2) \) denotes the covering number of \( 8neV \) by Hilbert-Schmidt balls of radius \( \delta \). It now remains to compute the covering numbers of \( V \). This is the object of the following lemma.

3.8 Lemma. Let \( k \in \{1, ..., n\} \). Define the set
\[ T_k = \{ Y \in \mathcal{H}_n : \text{rank}(Y) \leq k, ||Y|| \leq 1 \}. \]

For any \( \varepsilon \in (0, 1) \), let \( \mathcal{N}(T_k, \varepsilon B_2) \) be the covering number of \( T_k \) by Hilbert-Schmidt balls of radius \( \varepsilon \). Then,
\[
\log \mathcal{N}(T_k, \varepsilon B_2) \leq 2nk \log \left( \frac{12k}{\varepsilon} \right).
\]

Proof. Let \( Y \in T \) 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 \( \mu_i \) are real numbers, \( v_i \) are unit vectors (not necessarily orthogonal to one another), \( \lambda_1, ..., \lambda_k \) are the possible non-zero eigenvalues of \( Y \) and \( u_1, ..., u_k \) the associated eigenvectors. As the \( v_i \)'s are unit vectors,
\[
||Y - Z||_2 \leq \sum_{i=1}^{k} |\lambda_i - \mu_i| + \sum_{i=1}^{k} |\lambda_i|.||u_i u_i^* - v_i v_i^*||_2.
\]

Since \( ||Y|| \leq 1 \) and \( ||uv^*||_2 = ||u||_2 ||v||_2 \) for any \( u, v \in \mathbb{C}^n \), we get,
\[
||Y - Z|| \leq k \max_{1 \leq i \leq k} |\lambda_i - \mu_i| + 2k \max_{1 \leq i \leq k} ||u_i - v_i||_2. \tag{32}
\]

Let \( \mathcal{E} \) be a \( \varepsilon/2k \)-net of \([-1, 1]\) and \( \mathcal{F} \) a \( \varepsilon/4k \)-net of the unit sphere \( S^{n-1} \). 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 \( \varepsilon \)-net of \( T \) due to the inequality (32). Clearly, we can find a net \( \mathcal{E} \) such that \( |\mathcal{E}| \leq 4k/\varepsilon \). Moreover, by [8, 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 \( \mathcal{D}_\varepsilon \) we get finally,
\[
\log |\mathcal{D}_\varepsilon| \leq 2nk \log \left( \frac{12k}{\varepsilon} \right).
\]

\[ \square \]
Coming back to the inequality (31), we see that we have the inclusion $8cnV \subset n^bT_{2k}$, for some $b > 0$. Thus, by Lemma 3.8,

$$\log N(8cnV, \delta B_2) \leq N(T_{2k}, \delta n^{-b} B_2) = O\left(nk \log \left(\frac{n}{\delta}\right)\right).$$

As we chose $k \log n = o\left(n^{1+\frac{d}{2}}\right)$, we get

$$\limsup_{n \to +\infty} \frac{1}{n^{1+\frac{d}{2}}} \log P\left(\left|\operatorname{tr}(X/\sqrt{n})^d - \operatorname{tr}(X + \varepsilon \Gamma/\sqrt{n})^d\right| > tn\right) = -\infty.$$

As this inequality is true for any $r < \inf F J_+$, we obtain the large deviation upper bound of Theorem 1.13.

### 3.2 Large deviation lower bound

In this section we will use Lemma 1.7 to prove the large deviation lower bound for $(\frac{1}{n} \operatorname{tr}(X/\sqrt{n})^d)_{n \in \mathbb{N}}$. One of the difficulties in the derivation of the lower bound is that the domain of $\Lambda^*$ may be different from $H_n$. To bypass this problem, we will add a small Gaussian noise, as in the proof of Cramer’s theorem (see [16, Proof of Theorem 2.2.30]).

#### 3.2.1 Regularization by Gaussian noise

We recall that $N$ is said to belong 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,$$
\frac{1}{Z_n^{(\beta)}} e^{\frac{-\beta}{4} \operatorname{tr} H^2} d\ell_n^{(\beta)}(H),$$

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

#### 3.9 Lemma. Let $X$ be a Wigner matrix satisfying the convex concentration property 1.11. Let $\Gamma$ be a GOE matrix if $\beta = 1$ or a GUE matrix if $\beta = 2$. For any $t > 0$,

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

Proof. Decomposing into positive and negative parts, we see that it is sufficient to prove that for $\sigma \in \{-1, 1\}$ and any $t > 0$,

$$\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{1+\frac{d}{2}}} \log P\left(\left|\operatorname{tr}(X \sigma/\sqrt{n})^d - \operatorname{tr}(X + \varepsilon \Gamma \sigma/\sqrt{n})^d\right| > tn\right) = -\infty.$$

Fix $\sigma \in \{-1, 1\}$. Applying the convex concentration property to the $d^{th}$ Schatten norm,

$$\|Y\|_d : Y \in H_n \mapsto (\operatorname{tr}|Y|^d)^{1/d},$$

we deduce that $(\frac{1}{n} \operatorname{tr}(X/\sqrt{n})^d)_{n \in \mathbb{N}}$ is exponentially tight at the scale $n^{1+\frac{d}{2}}$. Using further the triangle inequality for the $d^{th}$ Schatten norm, and the fact that $N$ also satisfies the convex concentration property by [29, Theorem 5.2,
4.3], we deduce that for any \( \varepsilon > 0 \), \( \left( \frac{1}{n} \text{tr}(X + \varepsilon \Gamma)^d \right)_{n \in \mathbb{N}} \) is exponentially tight at the scale \( n^{1 + \frac{3}{d}} \) as well. Therefore, it suffices to show that for arbitrary large but fixed \( \tau > 0 \) and any \( t > 0 \),
\[
\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{1 + \frac{3}{d}}} \log \mathbb{P}_A(\left| \mu_{X/\sqrt{n}}(x_d^d) - \mu_{(X + \varepsilon \Gamma)/\sqrt{n}}(x_d^d) \right| > t) = -\infty,
\]
where \( \mathbb{P}_A \) denotes the measure \( \mathbb{P}(\cdot \cap A) \), and \( A \) is the event,
\[
A = \{ \mu_{X/\sqrt{n}}(|x|^d) \leq \tau, \mu_{(X + \varepsilon \Gamma)/\sqrt{n}}(|x|^d) \leq \tau \}.
\]
Using the uniform continuity of \( x \mapsto |x|^{1/d} \) on compact sets, we deduce that it is enough to show for any \( \delta > 0 \),
\[
\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{1 + \frac{3}{d}}} \log \mathbb{P}(\left| (\mu_{X/\sqrt{n}}(x_d^d))^\frac{1}{d} - (\mu_{(X + \varepsilon \Gamma)/\sqrt{n}}(x_d^d))^\frac{1}{d} \right| > \delta) = -\infty.
\]
But, by Minkowski’s inequality,
\[
\left| (\mu_{X/\sqrt{n}}(x_d^d))^\frac{1}{d} - (\mu_{(X + \varepsilon \Gamma)/\sqrt{n}}(x_d^d))^\frac{1}{d} \right| \leq \left( \frac{1}{n} \sum_{i=1}^{n} |(\lambda_i)x - (\mu_i)x|^d \right)^\frac{1}{d},
\]
where \( \lambda_1 \geq \ldots \geq \lambda_n \) and \( \mu_1 \geq \ldots \geq \mu_n \) are the eigenvalues of respectively \( X/\sqrt{n} \) and \( (X + \varepsilon \Gamma)/\sqrt{n} \). But,
\[
\sum_{i=1}^{n} |(\lambda_i) - (\mu_i)|^d \leq \sum_{i=1}^{n} |\lambda_i - \mu_i|^d,
\]
and by Lidskii’s theorem (see [6, Theorem III.4.1]),
\[
\frac{1}{n} \sum_{i=1}^{n} |\lambda_i - \mu_i|^d \leq \mu_1 / \sqrt{n}(|x|^d).
\]
Thus it actually suffices to prove that for any \( \delta > 0 \),
\[
\lim_{\varepsilon \to 0} \limsup_{n \to +\infty} \frac{1}{n^{1 + \frac{3}{d}}} \log \mathbb{P}(\mu_{\varepsilon \Gamma}/\sqrt{n}(|x|^d) > \delta) = -\infty,
\]
But \( \left( \frac{1}{n} \text{tr}(\Gamma/\sqrt{n})^d \right) \) is exponentially tight at the scale \( n^{1 + \frac{3}{d}} \), therefore we get finally the claim. \( \square \)

### 3.2.2 A tilting strategy

By [16, Theorem 4.2.16], it is sufficient to prove the large deviations lower bound for the sequence \( \left( \frac{1}{n} \text{tr}((X + \varepsilon \Gamma)^d)/\sqrt{n} \right)_{n \in \mathbb{N}} \) for any \( \varepsilon > 0 \) small enough. We denote by \( \Lambda^*_\varepsilon \) the Legendre transform of the logarithmic Laplace transform \( \Lambda_\varepsilon \) of \( X + \varepsilon \Gamma \), that is,
\[
\forall H \in \mathcal{H}_n^{(\beta)}, \quad \Lambda_\varepsilon(H) = \log \mathbb{E} e^{\varepsilon \text{tr}H(X + \varepsilon \Gamma)} = \Lambda(H) + \frac{\varepsilon^2 \text{tr}H^2}{\beta}.
\]
Adding this small Gaussian noise yields that the domain of \( \Lambda^*_\varepsilon \) is \( \mathcal{H}_n^{(\beta)} \). We note also for future record that,
\[
\forall Y \in \mathcal{H}_n^{(\beta)}, \quad \Lambda^*_\varepsilon(Y) \leq \Lambda^*(Y). \tag{34}
\]
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 1.13, that we can find a sequence $Y \in \mathcal{H}_n^{(\beta)}$ such that,
\[
\lim_{n \to +\infty} \mu_{sc}(x^d) + \frac{1}{n} \text{tr}(Y/\sqrt{n})^d \xrightarrow{n \to +\infty} x, \quad \lim_{n \to +\infty} \frac{\Lambda^*(Y)}{n^{1+\frac{\varepsilon}{2}}} = I_-(x). \tag{35}
\]
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 $s_n = \mu_{sc}(x^d) + \frac{1}{n} \text{tr}(Y/\sqrt{n})^d$. Let $E$ denote the event,
\[
V = \{A \in \mathcal{H}_n^{(\beta)} : \frac{1}{n} \text{tr}(A/\sqrt{n})^d \in B(s_n, \delta)\}.
\]
Using Lemma 1.7 and (34), we have,
\[
\mathbb{P}(Z \in E) \geq e^{-\Lambda^*(Y)} \mathbb{P}_Y(Z \in E) \exp\left(-\frac{1}{\mathbb{P}_Y(Z \in E)\mathbb{E}(H, \nabla^2 \Lambda_c(H), H)^{1/2}}\right),
\]
where $H = \nabla \Lambda_c^*(Y)$, and under $\mathbb{P}_Y$, $Z$ follows the tilted law,
\[
e^{\text{tr}(HZ) - \Lambda_c(H)} d\mathbb{P}(Z).
\]
To obtain the lower bound, we will prove on one hand,
\[
\langle H, \nabla^2 \Lambda_c(H), H \rangle = O(\Lambda^*(Y)). \tag{36}
\]
and on the other hand, that
\[
\mathbb{P}_H(Z \in V) \longrightarrow 1, \tag{37}
\]
Provided these two claims hold, we get, as $\Lambda^*(Y) = O(n^{1+\frac{\varepsilon}{2}})$,
\[
\lim_{n \to +\infty} \frac{1}{\mathbb{P}_Y(Z \in E)} \log \mathbb{P}\left(\frac{1}{n} \text{tr}(Z/\sqrt{n})^d \in B(x, \delta)\right) \geq -I_-(x),
\]
which gives the lower bound of Theorem 1.13.

Let us now prove (36) and (37). For the first claim, note that for any $A \in \mathcal{H}_n$, $\nabla^2 \Lambda_c(A) \geq \frac{2\varepsilon^2}{\beta}$ for the matrix order. Thus, it is sufficient to prove that,
\[
\text{tr}H^2 = O(\Lambda^*(Y)). \tag{38}
\]
But, denoting by $\Lambda_{c,(i,j)}$ the logarithmic Laplace transform of $X_{i,j} + \varepsilon \Gamma_{i,j}$ for any $i, j \in \{1, ..., n\}$, and integrating the inequalities,
\[
\forall A \in \mathcal{H}_n^{(\beta)}, \forall i < j, \quad \Lambda''_{c,(i,i)}(A_{i,i}) \geq \frac{2\varepsilon^2}{\beta}, \quad \nabla^2 \Lambda_{c,(i,j)}(A_{i,j}) \geq \frac{2\varepsilon^2}{\beta},
\]
(where the last inequality is meant in the matrix order when $\beta = 2$), knowing that $\nabla \Lambda_c(0) = 0$, we obtain
\[
||\nabla \Lambda_c(H)||_2 \geq \varepsilon^2 ||H||_2.
\]
From (24), we deduce that,
\[
\Lambda^*(\nabla \Lambda_c(H)) \geq \frac{1}{\varepsilon^2} \text{tr}(\nabla \Lambda_c(H))^2 \geq \frac{\varepsilon^4}{\varepsilon^2} \text{tr}H^2,
\]
which proves the estimate (38) since by definition $\nabla \Lambda_c(H) = Y$. 

28
For the second claim (37), we observe that under $P_{H}$, $Z$ is a random Hermitian matrix with independent entries up to the symmetry with mean $\nabla \Lambda_{\varepsilon}(H) = Y$. Moreover, if $\sigma_{ij}^{2} = \mathbb{E}_{H}|Z_{i,j} - \mathbb{E}Z_{i,j}|^{2}$ is the variance of the $(i,j)$th coordinate for $i < j$, then

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

$$= |\text{tr}\nabla^{2} \Lambda_{i,j}(0) - \text{tr}\nabla^{2} \Lambda_{i,j}(H_{i,j})|,$$

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 $|1 + \varepsilon^{2} - \sigma_{ij}^{2}| = O(||H_{i,j}||)$, 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

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

(39)

where we further used Cauchy-Schwarz inequality. Therefore, from (38), 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 (35). Thus, we finally obtain,

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

Thus, to show the claim (36), it remains to prove the following lemma, which is where the assumptions we made on the boundedness of the derivatives of $\Lambda$ play further their roles.

3.10 Lemma. Let $d \geq 3$ and $l$ even such that $l > d$. Let $X$ be a random Hermitian matrix such that $(X_{i,j})_{i,j} \leq j$ 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,

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

in probability.

Proof. We know from [2, Lemma 2.1 (9)], that denoting by $\hat{X} = X - \mathbb{E}X$, we have

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

where $|| \cdot ||_{m}$ denotes the $m^{th}$ Schatten norm, which is defined in (33). As $l > d$, we have $||\hat{X}||_{d+1} \leq ||\hat{X}||_{l}$. Using the assumption that $\text{tr}(\mathbb{E}X)^{2} = O(n^{1+\frac{2}{d}})$, we get,

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

29
But as \( l \) is even,
\[
\mathbb{E} \text{tr} \hat{X}^l = O(n^{1+\frac{l}{2}}),
\]
(40)
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
\[
\mathbb{P}(\|\hat{X}\|_\ell \geq (\log n)n^{\frac{1}{2}+\frac{1}{d}}) \longrightarrow 0.
\]
Therefore,
\[
|\text{tr} X^d - \text{tr} \hat{X}^d - \text{tr}(\mathbb{E}X)^d| = O\left((\log n)^d n^{\frac{1}{2}+\frac{1}{d}} n^{\frac{1}{2}+\frac{1}{d}} (d-k)\right),
\]
with probability going to 1. As \( l > d \), the maximum on the right-hand side is achieved at \( k = 1 \), so that
\[
|\text{tr} X^d - \text{tr} \hat{X}^d - \text{tr}(\mathbb{E}X)^d| = o(n^{1+\frac{1}{d}}),
\]
with probability going to 1. It now remains to show that,
\[
\frac{1}{n} \text{tr}(\hat{X} / \sqrt{n})^d \longrightarrow_{n \to +\infty} \mu_{sc}(x^d),
\]
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,
\[
\mu_{\hat{Y} / \sqrt{n}} \overset{\text{w}}{\longrightarrow} \mu_{sc},
\]
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,
\[
\forall \mu, \nu \in \mathcal{P}(\mathbb{R}), \mathcal{W}_2(\mu, \nu) = \inf_{\pi} \int |x - y|^2 d\pi(x, y),
\]
where the infimum runs over all couplings between \( \mu \) and \( \nu \). From the assumption on the variance profile of \( X \), we have by Hoffman-Weilandt inequality (see [6, Theorem VI.4.1]),
\[
\mathbb{E} \mathcal{W}_2(\mu_{\hat{X} / \sqrt{n}}, \mu_{\hat{Y} / \sqrt{n}}) \longrightarrow_{n \to +\infty} 0.
\]
Thus, together with the weak convergence of \( \mu_{\hat{Y} / \sqrt{n}} \) towards the semi-circle law, we deduce that \( \mathcal{W}_2(\mu_{\hat{X} / \sqrt{n}}, \mu_{sc}) \) converges to 0 in probability. But, from (40), we get by Markov’s inequality,
\[
\lim_{\tau \to +\infty} \lim_{n \to +\infty} \mathbb{P}\left(\mu_{\hat{X} / \sqrt{n}}(x^d) \geq \tau\right) \longrightarrow_{n \to +\infty} 0.
\]
Therefore, we can integrate the limit and deduce that
\[
\frac{1}{n} \text{tr}(\hat{X} / \sqrt{n})^d = \mu_{sc}(x^d) + o(1),
\]
in probability.
3.3 Computation of the rate function

In this section we give a proof of Corollary 1.15. 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.

3.11 Lemma. 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,

\[
\forall x \geq 0, \limsup_{n \to +\infty} I_n(x) \leq \frac{\beta}{4} x^{\frac{d}{2}},
\]

where

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

In particular,

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

Proof. Let \( x \geq 0 \). Let \( y = \left(\frac{x}{n^{1+\frac{d}{2}}}\right)^\frac{1}{n} n^{1+\frac{d}{2}}, \) and define

\[
Y = \begin{pmatrix} 0 & y & \ldots & y \\ y & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & y \\ y & \ldots & y & 0 \end{pmatrix} \in \mathcal{H}_n.
\]

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 \( y \sim n^{1-\frac{1}{2}x^{\frac{1}{2}}} \). As \( \Re X_{1,2} \) has variance \( 1/\beta \), we have \( \Lambda^*_{\Re X_{1,2}}(y) = \frac{\beta y^2}{2} + o(y^2) \), thus

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

which ends the proof.

3.12 Remark. 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 (for example). It reads,

\[
\forall x \geq 0, \quad I_-(x + \mu_{sc}(x^d)) \leq \limsup_{n \to +\infty} \frac{\Lambda^\dagger_{1,1}(x^{\frac{d}{2}} n^{1+\frac{d}{2}})}{n^{1+\frac{d}{2}}},
\]

which is nontrivial if \( \Lambda^\dagger_{1,1} \) 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 \( \Lambda_{1,1}(\lambda) \simeq_{+\infty} \frac{\beta}{2} \lambda^2 \), then \( \Lambda^\dagger_{1,1}(x) \simeq_{+\infty} \frac{1}{2\pi} x^2 \), so that,

\[
\forall x \geq 0, \quad I_-(x + \mu_{sc}(x^d)) \leq \frac{1}{2\beta} x^{\frac{d}{2}}.
\]

This bound is somewhat related to the large deviations of the traces of Wigner matrices with entries having super-Gaussian tails (see [2] 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 1.13, this fact immediately implies that a full LDP holds for the traces of this class of Wigner matrices, as announced in Corollary 1.15.

3.13 Proposition. Assume $X$ is a Wigner matrix whose entries have sharp sub-Gaussian tail. Assume further that $\mathbb{E}X^2_{1,1} \leq \frac{2}{\beta}$, 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 1.15, and $I_+, I_-$ are defined in Theorem 1.13.

Proof. First, it is clear that when $d$ is even, $I_+ = I_- = +\infty$ on $(-\infty, \mu_{sc}(x^d))$. As we assumed the entries of $X$ have sharp sub-Gaussian tails, and from our assumptions on the covariance structure, we have for any $Y \in \mathcal{H}_n^{(\beta)}$,

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

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

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

which gives the inequality $I_+ \geq J_d$. As by Lemma 3.11, $I_- \leq J_d$, and $I_+ \leq I_-$, we get the claim. \qed

4 Upper tail of cycle counts in sparse Erdős–Rényi graphs

In this section, we will give a proof of Proposition 1.17. 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

For any $Y \in \mathcal{H}_n$, we denote by $\lambda_1(Y), \ldots, \lambda_n(Y)$ its eigenvalues in non-decreasing order. For any $k \in \{1, \ldots, n\}$, we denote by $\text{tr}[k]Y$ the truncated trace:

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

and we denote by $g_k$ the function,

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

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

32
4.1 Lemma. Denote by \( v_n = n^2 p^2 \log(1/p) \). Assume \( \log(1/p) = o(np^2) \). Let \( k \in \{1, \ldots, n\} \) such that \( \log^4(1/p) = o(k) \), and in the case \( d = 3 \) such that moreover \( k = o((np)^{3/2}) \). For any \( \delta > 0 \),

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

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 \( \beta \)-ensembles (see [2, Proposition 3.5]). In a first step, we will prove that for any \( \delta > 0 \),

\[
\lim_{n \to +\infty} \frac{1}{v_n} \log \mathbb{P}(Z - \mathbb{E}Z > \delta p^d n^d) = -\infty, \tag{41}
\]

where \( Z = \text{tr} X^d - g_k(X) \). We will only consider the case where \( d \) is odd. In order to show (41), we see that it suffices to prove that for \( \sigma \in \{-, +\} \), and for any \( \delta > 0 \),

\[
\lim_{n \to +\infty} \frac{1}{v_n} \log \mathbb{P}(|\sum_{i=k+1}^{n} \lambda_i^d - \mathbb{E} \sum_{i=k+1}^{n} \lambda_i^d| > \delta p^d n^d) = -\infty, \tag{42}
\]

where we use \( \lambda_i \) as a shorthand for \( \lambda_i(X_\sigma) \) for any \( i \in \{1, \ldots, n\} \). Fix \( \sigma \in \{-, +\} \). Let \( M > 0 \), and 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 [32, Theorem 8.6], we know that \( X \) satisfies the convex concentration property as defined in 1.11 with constants \( \kappa = 1/2 \) and \( C = 2 \). Observe that we can write,

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

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

\[
\mathbb{P}
\left(
\left| \sum_{i=k+1}^{n} f_M(\lambda_i) - \mathbb{E} \sum_{i=k+1}^{n} f_M(\lambda_i) \right| > \delta n^d p^d \right) 
\leq 4 \exp\left( -\frac{n^{2d-1} p^{2d} \delta^2}{8d^2 M^{2(d-1)}} \right). \tag{43}
\]

Let \( \alpha_n \) to be a sequence going to \(+\infty\) such that,

\[
\alpha_n^2 = o\left( \frac{\sqrt{np}}{(\log(1/p))^2} \right), \tag{44}
\]

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

\[
M = \sqrt{np} \alpha_n. \tag{45}
\]

Re-writing (43) with this choice of \( M \), we have

\[
\mathbb{P}
\left(
\left| \sum_{i=k+1}^{n} f_M(\lambda_i) - \mathbb{E} \sum_{i=k+1}^{n} f_M(\lambda_i) \right| > \delta n^d p^d \right) 
\leq 4 \exp\left( -\frac{n^{d+1} p^{d+1} \delta^2}{8d^2 \alpha_n^2 (d-1)} \right). \tag{46}
\]

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

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

Proof. We have,

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

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

\[
|\mathbb{E} \sum_{i=k+1}^{n} (\lambda_i^d - f_M(\lambda_i))| \leq n^{d+1} \mathbb{P}(\max(-\lambda_n(X), \lambda_2(X)) \geq M).
\] (47)

Let \( \mathbf{u} \in \mathbb{R}^n \) be the vector \((1, 1, \ldots, 1)\), and and let \( \hat{X} = X - p u u^* \). By Weyl’s inequalities (see [6, Theorem III.2.1]), we have

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

Since \( \log n = o(np) \), we know from [28, 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 \( \alpha \geq c_0 \sqrt{np} \),

\[
\mathbb{P}(||\hat{X}|| \geq c) \leq 2e^{-\alpha c^2}.
\] (48)

As we assumed \( \log n = o(np) \), this bound is exponentially small, which, in view of (47) ends proof of the lemma. \( \square \)

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

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

Therefore, by (46), we have for \( n \) large enough,

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

Since \( d \geq 3 \), we have \( \frac{d^2}{d-2} \leq \frac{5}{2} \). Thus, from the choice of \( \alpha_n \) made in (44),

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

so that

\[
\lim_{n \to +\infty} \frac{1}{v_n} \log \mathbb{P}\left( \left| \sum_{i=k+1}^{n} f_M(\lambda_i) - \mathbb{E} \sum_{i=k+1}^{n} \lambda_i^d \right| > 2\delta n^d p^d \right) = -\infty.
\] (49)

But,

\[
\mathbb{P}\left( \sum_{i=k+1}^{n} \lambda_i^d - f_M(\lambda_i) > \delta p^d n^d \right) \leq \mathbb{P}\left( \sum_{i=k+1}^{n} \lambda_i^d \mathbb{1}_{\lambda_i \geq M} > \delta n^d p^d \right).
\] (50)
Let $R > 0$, and denote by $\mathcal{N}[R, +\infty)$ the number of eigenvalues of $X_\sigma$ in $[R, +\infty)$. By Weyl’s inequalities (see [6, Theorem III.2.1]), we have for any $i \geq 2$,

$$\lambda_i(\tilde{X}) \leq \lambda_i(X) \leq \lambda_{i-1}(\tilde{X}),$$

(51)

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

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

As $E[||\tilde{X}||] = O(\sqrt{mp})$ again by [28, Example 4.10], we deduce by Proposition 3.5 that if $R$ is such that $\sqrt{mp} = o(R)$, then for any $k \geq 2$,

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

(52)

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 $\sqrt{mp} = o(R)$, and we have

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

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

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

(53)

using the fact that $np$ grows polynomially fast to $+\infty$. Thus, in view of (49) and (50), it is sufficient to prove that,

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

(54)

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 $MN \geq R/2$. Therefore,

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

Using a union bound we get,

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

(55)
As $M_n \geq M$ and $\sqrt{np} = o(M)$ by definition of $M$ in (45), we deduce from (52) that for any $m \geq 0$,
\[
P\left(N[M_m, +\infty) > \delta n^d p^d / NM_{m+1}^d\right) \leq 2 \exp\left(-\frac{\delta n^d p^d}{32N R^{d-2}}\right).
\]
But $R = \sqrt{v_n} / \beta_n$, therefore,
\[
P\left(N[M_m, +\infty) > \delta n^d p^d / NM_{m+1}^d\right) \leq 2 \exp(-v_n c_n),
\]
with
\[
c_n = \frac{\delta n^d p^d}{32N \log^2 (1/p)}.
\]
As $d/(d - 2) \leq 3$ since $d \geq 3$, and $N = O(\log \log n)$, we see that with our choice of $\beta_n$ which satisfies $\log^2 (1/p) = o(\beta_n)$, we have in particular
\[
N \frac{\delta n^d p^d}{\log^2 (1/p)} = o(\beta_n).
\]
Thus, $c_n$ goes to $+\infty$. But, from the union bound (55), we have
\[
P\left(\sum_{i=k+1}^{n} \lambda_i^d \mathbb{I}_{M \leq \lambda_i \leq M_N} > \delta n^d p^d\right) \leq N \exp(-v_n c_n).
\]
Therefore,
\[
\lim_{n \to +\infty} \frac{1}{v_n} \log P\left(\sum_{i=k+1}^{n} \lambda_i^d \mathbb{I}_{M \leq \lambda_i \leq M_N} > \delta n^d p^d\right) = -\infty,
\]
which ends the proof of (42), and thus of the claim (41).

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.

4.3 Lemma. Assume $d \geq 4$ and $np^2 \gg 1$, or $d = 3$, $\log n = o(np)$ and $k = o((np)^{3/2})$. Then,
\[
\|E(trX^d - g_k(X))\| = o(n^d p^d).
\]
Proof. We start with the case $d \geq 4$. By Weyl’s inequalities (51), we get
\[
|trX^d - g_k(X)| \leq n||\hat{X}||^d.
\]
But from (48), we know that $E||\hat{X}||^d = O((np)^{d/2})$, thus
\[
|trX^d - g_k(X)| = O(n(np)^{d/2}).
\]
As $d \geq 4$ and $np^2 \to +\infty$, we have that $n^{d-1}p^{d/2}$ 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,
\[
trX^3 - (tr_{[l]}X_+^3 - tr_{[l]}X_-^3) = \sum_{i=l+1}^{\nu} \lambda_i(X)^3,
\]
for some $l, l'$ such that $1 \leq l \leq k$, and $n - l' \leq k$. We have,
\[
|\sum_{i=l+1}^{\nu} \lambda_i(X)^3 - tr(\hat{X})^3| \leq |\sum_{i=l+1}^{\nu} \lambda_i(X)^3 - \sum_{i=l+1}^{\nu} \lambda_i(\hat{X})^3| + 2k||\hat{X}||^3.
\]
By Weyl’s inequalities (51), 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,
\[
0 \leq \sum_{i=l+1}^{l'} \lambda_i(X)^3 - \sum_{i=l+1}^{l'} \lambda_i(\hat{X})^3 \leq 2||\hat{X}||^3.
\]
Thus,
\[
|\mathbb{E} \sum_{i=l+1}^{l'} \lambda_i(X)^3| \leq |\text{Etr}(\hat{X})^3| + 2(k + 1)\mathbb{E}||\hat{X}||^3.
\]
But, as \((\hat{X})_{i,i} = -p\) for any \(i \in \{1, \ldots, n\}\) and the off-diagonal entries of \(\hat{X}\) are centered, we obtain
\[
\text{Etr}(\hat{X})^3 = -np^3 + O(n^2p^2) = o(n^3p^3).
\]
Besides, as \(\log n = o(np)\), we know from (48) that \(\mathbb{E}||\hat{X}||^3 = O((np)^{3/2})\). Thus,
\[
|\mathbb{E} \sum_{i=l+1}^{l'} \lambda_i(X)^3| = O(k(np)^{3/2}) + o(n^3p^3),
\]
which gives the claim. \(\square\)

### 4.2 Proof of Proposition 1.17

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
\[
k \log n = o(np^2 \log(1/p)), \text{ and } \log^4(1/p) = o(k),
\]
(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:

**4.4 Lemma.** 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,
\[
\psi_n(t) \geq \varphi_n \left( t - O(k^{1/2}) \right),
\]
where \(\varphi_n\) is defined in Proposition 1.17.
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 $r = [(t - 1)^{1/d} np]$, that is,

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

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 [31, 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 $i \geq 1$,

$$
\lambda_i(Y) = O\left(\frac{np}{\sqrt{4}}\right).
$$

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.

We can now proceed with the proof of the upper bound of Proposition 1.17. Note that $\Phi$ is lower semi-continuous as one can see from its explicit expression (9). Therefore, by Lemma 4.1 it is sufficient to show that for any $t > 1$,

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

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 $\Lambda_p^*$ is strictly convex, $\psi_n$, defined in Lemma 4.4, is strictly increasing. Therefore we can write,

$$
\{s \geq 1 : \psi_n(s) \leq \psi_n(t - \delta)\} \cap (t - \delta, +\infty) = \emptyset.
$$

Let us extend $\psi_n$ on $(-\infty, 1]$ by setting,

$$
\forall s \leq 1, \quad \psi_n(s) = \inf \{\Lambda_p^*(Y) : g_k(Y) = sn^d p^d, Y \in \mathcal{H}_n^0\}.
$$

Observe that since $\Lambda_p^*$ is convex and $g_k$ is homogeneous, the above equality holds for $s \geq 1$ as well. Therefore,

$$
\{s \in \mathbb{R} : \psi_n(s) \leq \psi_n(t - \delta)\} \cap (t - \delta, +\infty) = \emptyset.
$$

Therefore,

$$
V_\delta\{\{\psi_n \leq \psi_n(t - \delta)\} \cap [t, +\infty) = \emptyset,
$$

38
where we recall that $V_\delta(A)$ denotes the open $\delta$-neighborhood of a subset $A \subset \mathbb{R}$.

Thus,

$$\mathbb{P}(g_k(X) \geq tn^d p^d) \leq \mathbb{P}\left(\frac{1}{n^d p^d} g_k(X) \notin V_\delta(\{\psi_n \leq \psi_n(t-\delta)\})\right).$$

We are now in the position to apply Proposition 1.4. Note that for any $Y \in \mathcal{H}_n([0,1])$, $\Lambda_\ast p(Y) \leq n^2 \log(1/p)$, so that the tightness assumption (4) of Proposition 1.4 is fulfilled for $K = \mathcal{H}_n([0,1])$ and $\kappa = n^2 \log(1/p)$. By Lemma 3.4, we know that for any $f : \mathbb{R} \to \mathbb{R}$ differentiable and convex such that $f(0) = f'(0) = 0$, the function

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

is convex, and its subdifferential at $Y$ satisfies,

$$\{f'(H) : \text{rank}(H) \leq k\} \subset \partial T_f(Y).$$

We deduce that for any $X, Y \in \mathcal{H}_n([0,1])$,

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

where

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

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 Proposition 1.4,

$$\mathbb{P}(g_k(X) \geq tn^d p^d) \leq -\psi_n(t-\delta) + \log \mathcal{N}(W, \delta/nB_2) + O(\log n),$$

where $\mathcal{N}(W, \delta/nB_2)$ is the covering number of $W$ by balls of radius $\delta/n$ for the Hilbert-Schmidt norm. By Lemma 3.8, we have

$$\log \mathcal{N}(W, \delta/nB_2) = O(nk \log n).$$

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

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

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

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

Since this is true for any $\delta > 0$, and $\Phi$ is lower semi-continuous, we get the claim by taking $\delta \to 0$.

**Acknowledgement**

I would like to thank Ofer Zeitouni for his enlightening insight and advice. I am also grateful to Ronen Eldan for some influential discussions and for the remark 1.10, as well as to Anirban Basak for many helpful discussions.
References

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

[2] F. Augeri. On the large deviations of traces of random matrices. *arXiv:1605.03894*, to appear in the *Annales de l’Institut Poincaré*, May 2016.

[3] T. Austin. The structure of low-complexity Gibbs measures on product spaces. *arXiv:1810.07278*.

[4] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[5] A. Basak and S. Mukherjee. Universality of the mean-field for the Potts model. *arXiv:1508.03949*.

[6] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.

[7] B. Bhattacharya, S. Ganguly, E. Lubetzky, and Y. Zhao. Upper tails and independence polynomials in random graphs. *Adv. Math.*, 319:313–347, 2017.

[8] D. Chafaï, O. Guédon, G. Lecué, and A. Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry*, volume 37 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2012.

[9] S. Chatterjee. The missing log in large deviations for triangle counts. *Random Structures Algorithms*, 40(4):437–451, 2012.

[10] S. Chatterjee and A. Dembo. Nonlinear large deviations. *Adv. Math.*, 299:396–450, 2016.

[11] S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős-Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

[12] F. Clarke. *Functional analysis, calculus of variations and optimal control*, volume 264 of *Graduate Texts in Mathematics*. Springer, London, 2013.

[13] N. Cook and A. Dembo. Large deviations of sub-graph counts in sparse Erdős-Rényi. *arXiv:1809.11148*.

[14] B. DeMarco and J. Kahn. Tight upper tail bounds for cliques. *Random Structures Algorithms*, 41(4):469–487, 2012.

[15] B. DeMarco and J. Kahn. Upper tails for triangles. *Random Structures Algorithms*, 40(4):452–459, 2012.

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

[17] R. Dobrushin, P. Groeneboom, and M. Ledoux. *Lectures on probability theory and statistics*, volume 1648 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1996. Lectures from the 24th Saint-Flour Summer School held July 7–23, 1994, Edited by P. Bernard.
[18] R. Eldan. Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. arXiv:1612.04346.

[19] A. Guionnet and J. Husson. Large deviations for the largest eigenvalue of Rademacher matrices. hal-01828877.

[20] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. Electron. Comm. Probab., 5:119–136 (electronic), 2000.

[21] V. Jain, F. Koehler, and A. Risteski. Mean-field approximation, convex hierarchies, and the optimality of correlation rounding: a unified perspective. arXiv:1808.07226.

[22] V. Jain, F. Koehler, and A. Risteski. The Mean-Field Approximation: Information Inequalities, Algorithms, and Complexity. arXiv:1802.06126.

[23] S. Janson, T. Łuczak, and A. Ruciński. An exponential bound for the probability of nonexistence of a specified subgraph in a random graph. Random graphs 87. Wiley, 1990.

[24] S. Janson, K. Oleszkiewicz, and A. Ruciński. Upper tails for subgraph counts in random graphs. Israel J. Math., 142:61–92, 2004.

[25] S. Janson and L. Warnke. The lower tail: Poisson approximation revisited. Random Struct. Alg., 48:219–246, 2016.

[26] J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. Random Structures Algorithms, 24(2):166–174, 2004.

[27] G. Kozma and W. Samotij. private communication.

[28] R. Latała, R. van Handel, and P. Youssef. The dimension-free structure of nonhomogeneous random matrices. arXiv:1711.00807.

[29] M. Ledoux. The concentration of measure phenomenon, volume 89 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2001.

[30] M. Ledoux and M. Talagrand. Probability in Banach spaces, volume 23 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]. Springer-Verlag, Berlin, 1991. Isoperimetry and processes.

[31] E. Lubetzky and Y. Zhao. On the variational problem for upper tails in sparse random graphs. Random Structures Algorithms, 50(3):420–436, 2017.

[32] P. Massart, G. Lugosi, and S. Boucheron. Concentration Inequalities : A Nonasymptotic Theory of Independence. Oxford University Press, 2013.

[33] R. T. Rockafellar. Convex analysis. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997.

[34] M. Šileikis and L. Warnke. A counterexample to the DeMarco-Kahn Upper Tail Conjecture. arXiv:1809.09595.

[35] Van H. Vu. A large deviation result on the number of small subgraphs of a random graph. Combin. Probab. Comput., 10(1):79–94, 2001.

[36] J. Yan. Nonlinear Large Deviations: Beyond the Hypercube. arXiv:1703.08887.