A Fourier view on the R-transform and related asymptotics of spherical integrals

A. Guionnet, M. Maïda [1]

Ecole Normale Supérieure de Lyon, Unité de Mathématiques pures et appliquées, UMR 5669, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

Abstract

We estimate the asymptotics of spherical integrals of real symmetric or Hermitian matrices when the rank of one matrix is much smaller than its dimension. We show that it is given in terms of the R-transform of the spectral measure of the full rank matrix and give a new proof of the fact that the R-transform is additive under free convolution. These asymptotics also extend to the case where one matrix has rank one but complex eigenvalue.

Key words: Large deviations, random matrices, non-commutative measure, R-transform.
1991 MSC: 60F10, 15A52, 46L50

1 Introduction

1.1 General framework and statement of the results

In this article, we consider the spherical integrals

$$I_N^{(\beta)}(D_N, E_N) := \int \exp\{N \operatorname{tr}(U D_N U^* E_N)\} dm_N^\beta(U),$$

Email addresses: aguionne@umpa.ens-lyon.fr (A. Guionnet), mmaida@umpa.ens-lyon.fr (M. Maïda).
[1] Present address: Université Paris X-Nanterre, Laboratoire Modal-X, 200 av. de la République, 92001 Nanterre Cedex, France

Preprint submitted to Elsevier *13 February 2012*

where $m_N^{(\beta)}$ denote the Haar measure on the orthogonal group $\mathcal{O}_N$ when $\beta = 1$ and on the unitary group $\mathcal{U}_N$ when $\beta = 2$, and D_N, E_N are $N \times N$ matrices that we can assume diagonal without loss of generality. Such integrals are often called, in the physics literature, Itzykson-Zuber or Harich-Chandra integrals. We do not consider the case $\beta = 4$ mostly to lighten the notations.
The interest for these objects goes back in particular to the work of Harish-Chandra ([13], [14]) who intended to define a notion of Fourier transform on Lie algebras. They have been then extensively studied in the framework of so-called matrix models that are related to the problem of enumerating maps (after [15], it has been developed in physics for example in [26], [18] or [20], in mathematics in [5] or [10]; a very nice introduction to these links is provided in [27]). The asymptotics of the spherical integrals needed to solve matrix models were investigated in [12]. More precisely, when D_N, E_N have N distinct real eigenvalues $(\theta_i(D_N), \lambda_i(E_N))_{1\leqslant i\leqslant N}$ and the spectral measures $\hat{\mu}^N_{D_N} = \frac{1}{N}\sum \delta_{\theta_i(D_N)}$ and $\hat{\mu}^N_{E_N} = \frac{1}{N}\sum \delta_{\lambda_i(E_N)}$ converge respectively to μ_D and μ_E, it is proved in Theorem 1.1 of [12] that

$$\lim_{N\to\infty} \frac{1}{N^2} \log I_N^{(\beta)}(D_N, E_N) = I^{(\beta)}(\mu_D, \mu_E) \tag{1}$$

exists under some technical assumptions and a (complicated) formula for this limit is given.

In this paper, we investigate different asymptotics of the spherical integrals, namely the case where one of the matrix, say D_N, has rank much smaller than N.

Such asymptotics were also already used in physics (see [19], where they consider replicated spin glasses, the number of replica being there the rank of D_N) or stated for instance in [5], section 1, as a formal limit (the spherical integral being seen as a series in θ when $D_N = \mathrm{diag}(\theta, 0, \cdots, 0)$ whose coefficients are converging as N goes to infinity). However, to our knowledge, there is no rigorous derivation of this limit available in the literature. We here study this problem by use of large deviations techniques. The proofs are however rather different from those of [12] ; they rely on large deviations for Gaussian variables and not on their Brownian motion interpretation and stochastic analysis as in [12].

Before stating our results, we now introduce some notations and make a few remarks.
Let $D_N = \mathrm{diag}(\theta, 0, \cdots, 0)$ have rank one so that

$$I_N^{(\beta)}(D_N, E_N) = I_N^{(\beta)}(\theta, E_N) = \int e^{\theta N (U E_N U^*)_{11}} dm_N^{(\beta)}(U). \tag{2}$$

Note that in general, in the case $\beta = 1$, we will omit the superscript (β) in all these notations.

We make the following hypothesis :

Hypothesis 1

(1) $\hat{\mu}^N_{E_N}$ *converges weakly towards a compactly supported measure* μ_E.
(2) $\lambda_{\min}(E_N) := \min_{1\leqslant i\leqslant N}\lambda_i(E_N)$ *and* $\lambda_{\max}(E_N) := \max_{1\leqslant i\leqslant N}\lambda_i(E_N)$ *converge respectively to* $\lambda_{\min}$ *and* $\lambda_{\max}$ *which are finite.*

Note that under Hypothesis 1, the support of μ_E, which we shall denote $\mathrm{supp}(\mu_E)$, is included into $[\lambda_{\min}, \lambda_{\max}]$.

Let us denote, for a probability measure μ_E, its Hilbert transform by H_{μ_E} :

$$\begin{aligned} H_{\mu_E} : I_E := \mathbb{R}\setminus \mathrm{supp}(\mu_E) &\longrightarrow \mathbb{R} \\ z &\longmapsto \int \frac{1}{z-\lambda} d\mu_E(\lambda). \end{aligned} \tag{3}$$

It is easily seen (c.f subsection 1.2 for details) that $H_{\mu_E} : I_E \to H_{\mu_E}(I_E)$ is invertible, with inverse denoted K_{μ_E}. For $z \in H_{\mu_E}(I_E)$, we set $R_{\mu_E}(z) = K_{\mu_E}(z) - z^{-1}$ to be the so-called R-transform of μ_E. In the case of the spectral measure $\hat{\mu}^N_{E_N}$ of E_N, we denote by H_{E_N} its Hilbert transform given by $H_{E_N}(x) = \frac{1}{N}\mathrm{tr}(x-E_N)^{-1} = \frac{1}{N}\sum_{i=1}^N (x-\lambda_i(E_N))^{-1}$.

The central result of this paper can be stated as follows :

Theorem 2 *Let* $\beta = 1$ *or* 2. *If we assume that Hypothesis 1.1 is satified and that there is* $\epsilon > 0$ *such that*

$$\|E_N\|_\infty := \max\{|\lambda_{max}(E_N)|, |\lambda_{min}(E_N)|\} = O\left(N^{\frac{1}{2}-\epsilon}\right), \tag{4}$$

then for θ *small enough so that there exists* $\eta > 0$ *so that*

$$\frac{2\theta}{\beta} \in \bigcup_{N_0\geq 0}\bigcap_{N\geq N_0} H_{E_N}([\lambda_{\min}(E_N)-\eta, \lambda_{\max}(E_N)+\eta]^c), \tag{5}$$

$$I^{(\beta)}_{\mu_E}(\theta) := \lim_{N\to\infty}\frac{1}{N}\log I^{(\beta)}_N(\theta, E_N) = \frac{\beta}{2}\int_0^{\frac{2\theta}{\beta}} R_{\mu_E}(v)dv. \tag{6}$$

Under Hypothesis 1.2, (4) *is obviously satisfied and (5) is equivalent to*

$$\frac{2\theta}{\beta} \in H_{\mu_E}([\lambda_{min}, \lambda_{max}]^c).$$

This result is proved in section 2 and appears in a way as a by-product of Lemma 14. It raises several remarks and generalisations that we shall investigate in this paper.
Note that in Theorems 3, 4 and 5 hereafter we consider the case $\beta = 1$, which requires

simpler notations but every statement could be extended to the case $\beta = 2$. The main difference to extend these theorems to the case $\beta = 2$ is that, following Fact 8, it requires to deal with twice as much Gaussian variables, and hence to consider covariance matrices with twice bigger dimension (the difficulty lying then in showing that these matrices are positive definite).

The first question we can ask is how to precise the convergence (6). Indeed, in the full rank asymptotics, in particular in the framework of [12], the second order term has not yet been rigorously derived. In our case, if d is the Dudley distance between measures (which is compatible with the weak topology) given by

$$d(\mu, \nu) = \sup \left\{ \left| \int f d\mu - \int f d\nu \right| ; |f(x)| \text{ and } \left| \frac{f(x) - f(y)}{x - y} \right| \leqslant 1, \forall x \neq y \right\}, \tag{7}$$

we have

Theorem 3 *Assume Hypothesis 1 and*

$$d(\hat{\mu}^N_{E_N}, \mu_E) = o(\sqrt{N}^{-1}).$$

Let θ be such that $2\theta \in H_{\mu_E}([\lambda_{min}, \lambda_{max}]^c)$.

- *If μ_E is not a Dirac measure at a single point, then, with $v = R_{\mu_E}(2\theta)$,*

$$\lim_{N \to \infty} e^{-N\left(\theta v - \frac{1}{2N} \sum_{i=1}^N \log(1 - 2\theta\lambda_i(E_N) + 2\theta v)\right)} I_N(\theta, E_N) = \frac{2\theta}{\sqrt{Z}},$$

  with $Z := \int \frac{1}{(K_{\mu_E}(2\theta) - \lambda)^2} d\mu_E(\lambda)$.
- *If $\mu_E = \delta_e$ for some $e \in \mathbb{R}$,*

$$\lim_{N \to \infty} e^{-N\theta e} I_N(\theta, E_N) = 1.$$

This theorem gives the second order term for the convergence given in Theorem 2 above. Indeed, with $2\theta \in H_{\mu_E}([\lambda_{min}, \lambda_{max}]^c)$, under Hypothesis 1.2, there exists (c.f. (14) for details) $\eta(\theta) > 0$ so that for N large enough

$$1 - 2\theta\lambda_i(E_N) + 2\theta v > \eta(\theta).$$

Therefore, there exists a finite constant $C(\theta) \leq (\eta(\theta)^{-1} + |\log(\eta(\theta))|)$ such that for N sufficiently large

$$\left| \frac{1}{2N} \sum_{i=1}^N \log(1 - 2\theta\lambda_i(E_N) + 2\theta v) - \frac{1}{2} \int \log(1 - 2\theta\lambda + 2\theta v) d\mu_E(\lambda) \right| \leq C(\theta) d\left(\frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i(E_N)}, \mu_E \right),$$

where d is the Dudley distance.
Moreover, with $v = R_{\mu_E}(2\theta)$, it is easy to see that

$$\theta v - \frac{1}{2}\int \log(1 - 2\theta\lambda + 2\theta v)d\mu_E(\lambda) = \frac{1}{2}\int_0^{2\theta} R_{\mu_E}(u)du,$$

showing how Theorem 3 relates with Theorem 2.

Another remark is that Theorem 2 can be seen as giving an interpretation of the primitive of the R-transform R_{μ_E} as a Laplace transform of $(UE_NU^*)_{11}$ for large N and for compactly supported probability measures μ_E.
A natural question is to wonder whether it can be extended to the case where θ is complex, to get an analogy with the Fourier transform that seems to have originally motivated Harish-Chandra. In the case of the different asymptotics studied in [12], this question is open : in physics, formal analytic extensions of the formula obtained for Hermitian matrices to any matrices are commonly used, but S. Zelditch [25] found that such an extension could be false by exhibiting counter-examples. In the context of the asymptotics we consider here, we shall however see that this extension is valid for $|\theta|$ small enough. Note that, as far as μ_E is compactly supported, R_{μ_E} can be extended analytically at least in a complex neighborhood of the origin (see Proposition 13 for further details).

Theorem 4 *Take $\beta = 1$ and assume that $(E_N)_{N\in\mathbb{N}}$ is a uniformly bounded sequence of matrices satisfying Hypothesis 1.1 where μ_E is not a Dirac mass.*
Assume furthermore that $d(\hat{\mu}^N_{E_N}, \mu_E) = o(\sqrt{N}^{-1})$, where d is the Dudley distance defined by (7).
Then, there exists an $r > 0$ such that, for any $\theta \in \mathbb{C}$, such that $|\theta| \leqslant r$,

$$\lim_{N\to\infty} \frac{1}{N}\log I_N(\theta, E_N) = \theta v(\theta) - \frac{1}{2}\int \log(1 + 2\theta v(\theta) - 2\theta\lambda)d\mu_E(\lambda),$$

where $\log(.)$ *is the main branch of the logarithm in* $\mathbb{C}$ *and* $v(\theta) = R_{\mu_E}(2\theta)$.

Note that the case $\mu_E = \delta_e$ is trivial if we assume additionnally Hypothesis 1.2 with $\lambda_{\min}$ and $\lambda_{\max}$ the edges of the support of μ_E since then $\max_{1\leq i\leq N}|\lambda_i - e|$ goes to zero with N which entails

$$\lim_{N\to\infty} \frac{1}{N}\log I_N(\theta, E_N) = \theta e$$

for all θ in $\mathbb{C}$.

The proof of Theorem 4 will be more involved than the real case treated in sections 2 and 3 and the difficulty lies of course in the fact that the integral is now oscillatory, forcing us to control more precisely the deviations in order to make sure that the term of order one in the large N expansion does not vanish. This is the object of section 4.

Once the view of spherical integrals as Fourier transforms has been justified by the extension to the complex plane, a second natural question is to wonder whether we can use it to see that the R-transform is additive under free convolution. Let us make some reminder about free probability : in this set up, the notion of freeness replaces the standard notion of independence and the R-transform is analogous to the logarithm of the Fourier transform of a measure. Now, it is well known that the log-Laplace (or Fourier) transform is additive under convolution i.e. for any probability measures μ, ν on $\mathbb{R}$ (say compactly supported to simplify), any $\lambda \in \mathbb{R}$, (or $\mathbb{C}$)

$$\log \int e^{\lambda x} d\nu * \mu(x) = \log \int e^{\lambda x} d\mu(x) + \log \int e^{\lambda x} d\nu(x).$$

Moreover, this property, if it holds for λ's in a neighbourhood of the origin, characterizes uniquely the convolution. Similarly, if we denote $\mu \boxplus \nu$ the free convolution of two compactly supported probability measures on $\mathbb{R}$, it is uniquely described by the fact that

$$R_{\mu \boxplus \nu}(\lambda) = R_\mu(\lambda) + R_\nu(\lambda)$$

for sufficiently small λ's. Theorem 2 provides an interpretation of this result. Indeed, Voiculescu [24] proved that if A_N, B_N are two diagonal matrices with spectral measures converging towards μ_A and μ_B respectively, with uniformly bounded spectral radius, then the spectral measure of $A_N + UB_NU^*$ converges, if U follows $m_N^{(2)}$, towards $\mu_A \boxplus \mu_B$. This result extends naturally to the case where U follows $m_N^{(1)}$ (see [6] Theorem 5.2 for instance). Therefore, it is natural to expect the following result :

Theorem 5 *Let $\beta = 1$, $(A_N, B_N)_{N\in\mathbb{N}}$ be a sequence of uniformly bounded real diagonal matrices and V_N following $m_N^{(1)}$.*

(1) Then

$$\lim_{N\to\infty} \left(\frac{1}{N} \log I_N(\theta, A_N + V_N B_N V_N^*) - \int \frac{1}{N} \log I_N(\theta, A_N + V_N B_N V_N^*) dm_N^{(1)}(V_N) \right) = 0 \ \ a.s. \tag{8}$$

(2) If additionnally the spectral measures of A_N and B_N converge respectively to μ_A and μ_B fast enough (i.e. such that $d(\hat{\mu}_{A_N}, \mu_A) + d(\hat{\mu}_{B_N}, \mu_B) = o(\sqrt{N}^{-1})$) and μ_A and μ_B are not Dirac masses at a point, then, for any θ small enough,

$$\lim_{N\to\infty} \frac{1}{N} \log I_N(\theta, A_N + V_N B_N V_N^*) = \lim_{N\to\infty} \frac{1}{N} \log I_N(\theta, A_N) + \lim_{N\to\infty} \frac{1}{N} \log I_N(\theta, B_N) \ \ a.s. \tag{9}$$

Then the additivity of the R-transform (cf. Corollary 22) is a direct consequence of this result together with the continuity of the spherical integrals with respect to the empirical measure of the full rank matrix (which will be shown in Lemma 14).
Note that the case where μ_A or μ_B are Dirac masses is trivial if we assume that the edges of the spectrum of A_N or B_N converge towards this point. The general case could be handled

as well but, since it has no motivation for the R-transform (for which we can always assume that the above condition holds, see Corollary 22), we shall not detail it. Section 6 will be devoted to the proof of this theorem which decomposes mainly in two steps : to get the first point, we establish a result of concentration under $m_N^{(1)}$ that will give us (8); then to prove the second point once we have the first one it is enough to consider the expectation of $\frac{1}{N}\log I_N(\theta, A_N + V_N B_N V_N^*)$ and if one assumes that

$$\lim_{N\to\infty}\frac{1}{N}\int\left(\log\int e^{\theta N(UA_NU^*+UV_NB_NV_N^*U^*)_{11}}dm_N^{(1)}(U)\right)dm_N^{(1)}(V) \\ = \lim_{N\to\infty}\frac{1}{N}\log\int\int e^{\theta N(UA_NU^*+UV_NB_NV_N^*U^*)_{11}}dm_N^{(1)}(U)dm_N^{(1)}(V) \quad (10)$$

the equality (9) follows from the observation that the right hand side equals $N^{-1}\log I_N(\theta, A_N) + N^{-1}\log I_N(\theta, B_N)$.

Note that equation (10) is rather typical to what should be expected for disordered particles systems in the high temperature regime and indeed our proof follows some very smart ideas of Talagrand that he developed in the context of Sherrington-Kirkpatrick model of spin glasses at high temperature (see [22]). This proof is however rather technical because the required control on the L^2 norm of the partition function is based on the study of second order corrections of replicated systems which generalizes Theorem 3.

The next question, that we will actually tackle in section 5, deals with the understanding of the limit (6) for all the values of θ. We find the following result

Theorem 6 *Let $\beta = 1$ or 2. Assume $\hat{\mu}_{E_N}^N$ satisfy Hypothesis 1. If we let $H_{min} := \lim_{z\uparrow\lambda_{min}} H_{\mu_E}(z)$ and $H_{max} := \lim_{z\downarrow\lambda_{max}} H_{\mu_E}(z)$, then*

$$\lim_{N\to\infty}\frac{1}{N}\log I_N^{(\beta)}(\theta, E_N) = I_{\mu_E}^{(\beta)}(\theta) = \theta v(\theta) - \frac{\beta}{2}\int\log\left(1 + \frac{2}{\beta}\theta v(\theta) - \frac{2}{\beta}\theta\lambda\right)d\mu_E(\lambda)$$

with

$$v(\theta) = \begin{cases} R_{\mu_E}\left(\frac{2}{\beta}\theta\right) & \text{if } H_{min} \leqslant \frac{2\theta}{\beta} \leqslant H_{max} \\ \lambda_{max} - \frac{\beta}{2\theta} & \text{if } \frac{2\theta}{\beta} > H_{max} \\ \lambda_{\min} - \frac{\beta}{2\theta}, & \text{if } \frac{2\theta}{\beta} < H_{min}. \end{cases}$$

Note here that the values of $\lambda_{\min}$ and λ_{max} do affect the value of the limit of spherical integrals in the asymptotics we consider here, contrarily to what happens in the full rank asymptotics considered in [12].

As a consequence of Theorem 6, we can see that there are two phase transitions at $H_{max}\beta/2$

and $H_{min}\beta/2$ which are of second order in general (the second derivatives of $I_{\mu_E}(\theta)$ being discontinuous at these points, except when $\lambda_{\max} H'_{\mu_E}(\lambda_{\max}) = 1$ (or similar equation with $\lambda_{\min}$ instead of $\lambda_{\max}$), in which case the transition is of order 3). These transitions can in fact be characterized by the asymptotic behaviour of $(UE_NU^*)_{11}$ under the Gibbs measure

$$d\mu_N^{\beta,\theta}(U) = \frac{1}{I_N^{(\beta)}(\theta, E_N)} e^{N\theta(UE_NU^*)_{11}} dm_N^{(\beta)}(U).$$

For $\theta \in \left[\frac{H_{min}\beta}{2}, \frac{H_{max}\beta}{2}\right]^c$, $(UE_NU^*)_{11}$ saturates and converges $\mu_N^{\beta,\theta}$-almost surely towards $\lambda_{max} - \frac{\beta}{2\theta}$ (resp. $\lambda_{\min} - \frac{\beta}{2\theta}$). Hence, up to a small component of norm of order θ^{-1}, with high probability, the first column vector U_1 of U will align on the eigenvector corresponding to either the smallest or the largest eigenvalue of E_N, whereas for smaller θ's, U_1 will prefer to charge all the eigenspaces of E_N.

Another natural question is to wonder what happens when D_N has not rank one but rank negligible compared to N. It is not very hard to see that in the case where all the eigenvalues of D_N are small enough (namely when they all lie inside $H_{\mu_E}([\lambda_{min}, \lambda_{max}]^c)$), we find that the spherical integral approximately factorizes into a product of integrals of rank one. More precisely,

Theorem 7 *Let $\beta = 1$ or 2. Let $D_N = diag(\theta_1^N, \ldots, \theta_{M(N)}^N, 0, \ldots, 0)$ with $M(N)$ which is $o(N^{\frac{1}{2}-\epsilon})$ for some $\epsilon > 0$. Assume that $\hat{\mu}_{E_N}^N$ fulfills Hypothesis 1.1, that $||E_N||_\infty = o(N^{\frac{1}{2}-\varepsilon})$ for some $\varepsilon > 0$ and that there exists $N_0 \in \mathbb{N}$ and $\eta > 0$ such that, for all $N \geq N_0$ and i from 1 to $M(N)$, $\frac{2\theta_i^N}{\beta} \in H_{E_N}([\lambda_{min}(E_N) - \eta, \lambda_{max}(E_N) + \eta]^c)$.*
Then, if $\frac{1}{M(N)} \sum_{i=1}^{M(N)} \delta_{\theta_i^N}$ converges weakly to μ_D,

$$I_{\mu_E}^{(\beta)}(D) := \lim_{N\to\infty} \frac{1}{NM(N)} \log I_N^{(\beta)}(D_N, E_N)$$

exists and is given by

$$I_{\mu_E}^{(\beta)}(D) = \lim_{N\to\infty} \frac{1}{M(N)} \sum_{i=1}^{M(N)} I_{\mu_E}^{(\beta)}(\theta_i^N) = \int I_{\mu_E}^{(\beta)}(\theta) d\mu_D(\theta). \tag{11}$$

This will be shown at the end of section 2, the proof being very similar to the case of rank one. It relies mainly on Fact 8 hereafter and comes from the fact that in such asymptotics the $M(N)$ first column vectors of an orthogonal or unitary matrix distributed according to the Haar measure behave approximately like independent vectors uniformly distributed on the sphere. This can be compared with the very old result of E. Borel [4] which says that one entry of an orthogonal matrix distributed according to the Haar measure behaves like a Gaussian variable. That kind of considerations finds continuation for example in a recent

work of A. D'Aristotile, P. Diaconis and C. M. Newman [7] where they consider a number of element of the orthogonal group going to infinity not too fast with N. In the same direction, one can also mention the recent work of T. Jiang [16] where he shows that the entries of the first $O(N/\log N)$ columns of an Haar distributed unitary matrix can be simultaneously approximated by independent standard normal variables.
Note that P. Śniady and one of the author could prove by different techniques that the asymptotics we are talking about extend to $M(N) = o(N)$.
Of course we would like to generalize also the full asymptotics we've got in Theorem 6 to the set up of finite rank i.e. in particular consider the case where some (a $o(N)$ number) of the eigenvalues of E_N could converge away from the support. It seems to involve not only the deviations of λ_{max} but those of the first M ones when the rank is M. As it becomes rather complicate and as the proof is already rather involved in rank one, we postpone this issue to further research.

To finish this introduction, we also want to mention that the results we've just presented give (maybe) less obvious relations between the R-transform and Schur functions or vicious walkers.
Indeed, if s_λ denotes the Schur function associated with a Young tableau λ (cf. [21] for more details), then, it can be checked (cf. [11] for instance) that

$$s_\lambda(M) = I_N^{(2)}\left(\log M, \frac{l}{N}\right)\Delta\left(\frac{l}{N}\right)\frac{\Delta(\log M)}{\Delta(M)}$$

with $l_i = \lambda_i + N - i$, $1 \le i \le N$ and $\Delta(M) = \prod_{i<j}(M_i - M_j)$ when $M = \mathrm{diag}(M_1, \cdots, M_N)$. Thus, our results also give the asymptotics of Schur functions when $N^{-1}\delta_{N^{-1}(\lambda_i+N-i)}$ converges towards some compactly supported probability measure μ. For instance, Theorem 2 implies that for θ small enough

$$\lim_{N\to\infty}\frac{1}{N}\log\left(\prod_{i>j}(N^{-1}(\lambda_j - j - \lambda_i + i))^{-1}s_\lambda(e^\theta, 1, \ldots, \ldots, 1)\right) = \int_0^\theta R_\mu(u)du + \log(\theta(e^\theta - 1)^{-1}).$$

Such asymptotics should be more directly related with the combinatorics of the symmetric group and more precisely with non-crossing partitions which play a key role in free convolution.
On the other hand, it is also known that spherical integrals are related with the density kernel of vicious walkers, that is Brownian motions conditionned to avoid each others, either by using the fact that the eigenvalues of the Hermitian Brownian motion are described by such vicious walkers (more commonly named in this context Dyson's Brownian motions) or by applying directly the result of Karlin-McGregor [17]. Hence, the study of the asymptotics of spherical integrals we are considering allows to estimate this density kernel when $N-1$ vicious walkers start at the origin, the last one starting at θ and at time one reach $(x_1, \ldots, x_N)$ whose empirical distribution approximates a given compactly supported probability measure.

Before going into the proofs themselves, we gather here some material and notations that will be useful throughout the paper.

1.2.0.1 Gaussian representation of Haar measure

In the different cases we will develop, the first step will be always the same : we will represent the column vectors of unitary or orthogonal matrices distributed according to Haar measure via Gaussian vectors. To be more precise, we recall the following fact :

Fact 8 *Let $k \leqslant N$ be fixed.*
• *Orthogonal case.*
Let $U = (u_{ij})_{1\leqslant i,j\leqslant N}$ be a random orthogonal matrix distributed according to $m_N^{(1)}$, the Haar measure on $\mathcal{O}_N$. Denote by $(u^{(i)})_{1\leqslant i\leqslant N}$ the column vectors of U.
Let $(g^{(1)}, \ldots, g^{(k)})$ be k independent standard Gaussian vectors in $\mathbb{R}^N$ and let $(\tilde{g}^{(1)}, \ldots, \tilde{g}^{(k)})$ the vectors obtained from $(g^{(1)}, \ldots, g^{(k)})$ by the standard Schmidt orthogonalisation procedure. Then it is well known that

$$(u^{(1)}, \ldots, u^{(k)}) \sim \left(\frac{\tilde{g}^{(1)}}{\|\tilde{g}^{(1)}\|}, \ldots, \frac{\tilde{g}^{(k)}}{\|\tilde{g}^{(k)}\|}\right),$$

where $\|.\|$ denotes the Euclidean norm in $\mathbb{R}^N$ and the equality $\sim$ means that the two $k \times N$-matrices have the same law.
• *Unitary case.*
With the same notations, let U be distributed according to $m_N^{(2)}$, the Haar measure on $\mathcal{U}_N$. Let $(g^{(1),R}, \ldots, g^{(k),R}, g^{(1),I}, \ldots, g^{(k),I})$ be $2k$ independent standard Gaussian vectors in $\mathbb{R}^N$ and let $(\tilde{G}^{(1)}, \ldots, \tilde{G}^{(k)})$ be the k vectors obtained from $(g^{(1),R} + ig^{(1),I}, \ldots, g^{(k),R} + ig^{(k),I})$ by the standard Schmidt orthogonalisation procedure with respect to the usual scalar product in $\mathbb{C}^N$. Then we get that

$$(u^{(1)}, \ldots, u^{(k)}) \sim \left(\frac{\tilde{G}^{(1)}}{\left\|\tilde{G}^{(1)}\right\|}, \ldots, \frac{\tilde{G}^{(k)}}{\left\|\tilde{G}^{(k)}\right\|}\right),$$

where $\|.\|$ denotes the usual norm in $\mathbb{C}^N$.

Note that heuristically, the above representation in terms of Gaussian vectors allows us to understand why the limit in the finite rank case behaves as a sum of functions of each of the eigenvalues of D_N. Indeed, in high dimension, we know that a bunch of k (independent of the dimension) Gaussian vectors are almost orthogonal one from another so that the orthogonalisation procedure let them almost independent.

1.2.0.2 Some properties of the Hilbert and the R-transforms of a compactly supported probability measure on $\mathbb{R}$

Let $\lambda_{\min}(E)$ and $\lambda_{\max}(E)$ be the edges of the support of μ_E. For all $\lambda_{\min} \leqslant \lambda_{\min}(E)$ and $\lambda_{\max} \geqslant \lambda_{\max}(E)$, let us denote by $H_{min} := \lim_{z \uparrow \lambda_{min}} H_{\mu_E}(z)$ and $H_{max} := \lim_{z \downarrow \lambda_{max}} H_{\mu_E}(z)$, where H_{μ_E} was defined in (3).
We sum up the properties of H_{μ_E} that will be useful for us in the following

Property 9 *:*

(1) H_{μ_E} is decreasing and positive on $\{z > \lambda_{max}\}$ and decreasing and negative on $\{z < \lambda_{min}\}$.
(2) Therefore, H_{min} exists in $\mathbb{R}_-^* \cup \{-\infty\}$ and H_{max} exists in $\mathbb{R}_+^* \cup \{+\infty\}$.
(3) H_{μ_E} is bijective from $I = \mathbb{R} \backslash [\lambda_{\min}, \lambda_{\max}]$ onto its image $I' :=]H_{min}, H_{max}[\backslash \{0\}$.
(4) H_{μ_E} is analytic on I and its derivative never cancels on I.

The third point of the property above allows the following

Definition 10 *:*

(1) K_{μ_E} is defined on I' as the functional inverse of H_{μ_E}.
(2) I' does not contain 0 so that, on I', we can define R_{μ_E} given by $R_{\mu_E}(\gamma) = K_{\mu_E}(\gamma) - \frac{1}{\gamma}$ for any $\gamma \in I'$.

We will need to consider the inverse Q_{μ_E} of R_{μ_E}. To define it properly, we have to look more carefully at the properties of R_{μ_E}. We have :

Property 11 *:*

(1) K_{μ_E} and R_{μ_E} are analytic (and in particular continuously differentiable) on I'.
(2) R_{μ_E} is increasing and its derivative never cancels.
(3) $\lim_{\gamma \to 0^-} R_{\mu_E}(\gamma) = \lim_{\gamma \to 0^+} R_{\mu_E}(\gamma) = m := \int \lambda d\mu_E(\lambda)$.
(4) R_{μ_E} is bijective from I' onto its image $I'' := \left]\lambda_{min} - \dfrac{1}{H_{min}}, \lambda_{max} - \dfrac{1}{H_{max}}\right[\backslash \{m\}$ so that we can define its inverse Q_{μ_E} from I'' to I'. Moreover, Q_{μ_E} is differentiable on I''.

The proof of these properties is easy and left to the reader.

The following property deals with the behaviour of these functions on the complex plane. A proof of it can be found for example in [23]. We first extend the definition of the Hilbert transform, that we denote again H_{μ_E} by

$$\begin{aligned} H_{\mu_E}\,:\, \mathbb{C}\setminus \mathrm{supp}(\mu_E) &\longrightarrow \mathbb{C} \\ z &\longmapsto \int \frac{1}{z-\lambda} d\mu_E(\lambda). \end{aligned} \tag{12}$$

Property 12 *:*

(1) There exists a neighbourhood $\mathcal{A}$ of ∞ such that H_{μ_E} is bijective from $\mathcal{A}$ into $H_{\mu_E}(\mathcal{A})$, which is a neighbourhood of 0.
(2) We denote by $K^{(c)}_{\mu_E}$ its functional inverse on $H_{\mu_E}(\mathcal{A})$ and $R^{(c)}_{\mu_E}$ is given by $R^{(c)}_{\mu_E}(\gamma) = K^{(c)}_{\mu_E}(\gamma) - \frac{1}{\gamma}$ for any $\gamma \in H_{\mu_E}(\mathcal{A})$ (that does not contain 0).
(3) $R^{(c)}_{\mu_E}$ is analytic and coincides with R_{μ_E} on $I' \cap H_{\mu_E}(\mathcal{A})$. Therefore, we denote it again R_{μ_E}.

Note that throughout the paper, we will denote $\lambda_i := \lambda_i(E_N)$, $\theta_i := \theta_i(D_N)$ (and even θ will denote $\theta_1(D_N)$ in the case of rank one) and and we recall that H_{E_N} denotes the Hilbert transform of $\hat{\mu}^N_{E_N}$ and is given, for $x \in [\min \lambda_i(E_N), \max \lambda_i(E_N)]^c$, by $H_{E_N}(x) = \dfrac{1}{N}\mathrm{tr}(x - E_N)^{-1}$.

We now state the following property, which will be useful in the proof of Theorem 4 :

Proposition 13 *If $(E_N)_{N\in\mathbb{N}}$ is uniformly bounded and satisfying Hypothesis 1.1, there exists $r > 0$ such that, for any $\theta \in \mathbb{C}$ such that $|\theta| \leqslant r$, there is a solution of*

$$H_{E_N}\left(\frac{1}{2\theta} + v_N(\theta)\right) = 2\theta,$$

such that $v_N(\theta) \xrightarrow[N\to\infty]{} R_{\mu_E}(2\theta)$.

Proof of Proposition 13 : Let $\mathcal{A}_N$ be a neighbourhood of ∞ on which H_{E_N} is invertible ($\mathcal{A}_N$ can be given as $\{z/|z| > R_N\}$, for some R_N). For any $\eta > 0$, we denote by $\mathcal{A}^\eta_N := \{x \in \mathcal{A}_N / d(x, \mathcal{A}^c_N) \geqslant \eta\}$. Let θ be such that there exists $\eta > 0$ such that $2\theta \in \bigcup_{N_0\geqslant 0} \bigcap_{N\geqslant N_0} H_{E_N}(\mathcal{A}^\eta_N)$, we take $v_N(\theta)$ the unique solution in $\mathcal{A}^\eta_N - (2\theta)^{-1}$ of

$$H_{E_N}\left(\frac{1}{2\theta} + v_N(\theta)\right) = 2\theta.$$

Since, for all $\lambda \in \bigcup_{N_0\geqslant 0} \bigcap_{N\geqslant N_0} \mathrm{supp}(\hat{\mu}^N_{E_N})$, the application $z \mapsto (z-\lambda)^{-1}$ is continuous bounded on $\bigcup_{N_0\geqslant 0} \bigcap_{N\geqslant N_0} \mathcal{A}^\eta_N$, under Hypothesis 1.1, $v_N(\theta)$ converges to $R_{\mu_E}(2\theta)$.
Furthermore, the fact that $(E_N)_{N\in\mathbb{N}}$ is uniformly bounded ensures that we can choose the $\mathcal{A}_N$'s such that there exists $r > 0$ such that $\bigcup_{N_0\geqslant 0} \bigcap_{N\geqslant N_0} H_{E_N}(\mathcal{A}^\eta_N) \supset \{\theta/|\theta| \leqslant r\}$. ∎

2 Proof of Theorems 2, 7 and related results

Before going into more details, let us state and prove a lemma which deals with the continuity of I_N and its limit. We state here a trivial continuity in the finite rank matrix but also a weaker continuity result in the spectral measure of the diverging rank matrix, on which the proof of Theorem 2 is based.

Lemma 14 *(1) For any $N \in \mathbb{N}$, any sequence of matrices $(E_N)_{N\in\mathbb{N}}$ with spectral radius $\|E_N\|_\infty$ uniformly bounded by $||E||_\infty$, any Hermitian matrices $(D_N, \tilde{D}_N)_{N\in\mathbb{N}}$,*

$$\left|\frac{1}{N}\log I_N^{(\beta)}(D_N, E_N) - \frac{1}{N}\log I_N^{(\beta)}(\tilde{D}_N, E_N)\right| \leqslant ||E||_\infty tr|D_N - \tilde{D}_N|$$

(2) Let $D_N = diag\ (\theta, 0, \cdots, 0)$. Assume that there is a positive η and a finite integer N_0 such that for $N \geq N_0$, $\frac{2\theta}{\beta} \in H_{E_N}([\lambda_{\min}(E_N) - \eta, \lambda_{\max}(E_N) + \eta]^c)$. We let v_N be the unique solution in $-\beta(2\theta)^{-1} + [\lambda_{\min}(E_N) - \eta\,,\, \lambda_{\max}(E_N) + \eta]^c$ of the equation

$$\frac{\beta}{2\theta} H_{E_N}\left(\frac{\beta}{2\theta} + v_N\right) = 1. \tag{13}$$

Then, $v_N \in [\lambda_{\min}(E_N), \lambda_{\max}(E_N)]$ and for any $\zeta \in (0, \frac{1}{2})$, there exists a finite constant $C(\eta, \zeta)$ depending only on η and ζ such that for all $N \geq N_0$

$$\left|\frac{1}{N}\log I_N^{(\beta)}(\theta, E_N) - \theta v_N + \frac{\beta}{2N}\sum_{i=1}^{N}\log\left(1 + \frac{2\theta}{\beta}v_N - \frac{2\theta}{\beta}\lambda_i\right)\right| \leqslant C(\eta, \zeta) N^{-\frac{1}{2}+\zeta}\|E_N\|_\infty.$$

(3) Let $D_N = diag(\theta, 0, \cdots, 0)$. Let $E_N, \tilde{E}_N$ be two matrices such that

$$d(\hat{\mu}^N_{E_N}, \hat{\mu}^N_{\tilde{E}_N}) \leqslant \delta,$$

where d is the Dudley distance on $\mathcal{P}(\mathbb{R})$ and so that both E_N and $\tilde{E}_N$ satisfy (4). Let $\eta > 0$. Assume that there exists $N_0 < \infty$ so that for $N \geq N_0$, $\frac{2\theta}{\beta} \in H_{E_N}([\lambda_{\min}(E_N) - \eta, \lambda_{\max}(E_N) + \eta]^c) \cap H_{\tilde{E}_N}([\lambda_{\min}(\tilde{E}_N) - \eta, \lambda_{\max}(\tilde{E}_N) + \eta]^c)$. Then, there exists a function $g(\delta, \eta)$ (independent of N) going to zero with δ for any η and such that for all $N \geq N_0$

$$\left|\frac{1}{N}\log I_N^{(\beta)}(D_N, E_N) - \frac{1}{N}\log I_N^{(\beta)}(D_N, \tilde{E}_N)\right| \leqslant g(\delta, \eta)$$

Note that the third point is analogous to the continuity statement obtained in the case where D_N has also rank N in [12], Lemma 5.1. However, let us mention again that there is an important difference here which lies in the fact that the smallest and largest eigenvalues play quite an important role. In fact, it can be seen (see Theorem 6) that if we let one eigenvalue be much larger than the support of the limiting spectral distribution, then the

limit of the spherical integral will change dramatically. However, Lemma 14.3 shows that this limit will not depend on these escaping eigenvalues provided $|\theta|$ is smaller than some critical value $\theta_0(\lambda_{\min}, \lambda_{\max})$ $(= \min(|H_{min}\beta/2|, |H_{max}\beta/2|))$.
Before going into the proof of Lemma 14, let us show that Theorem 2 is a direct consequence of its second point.

Proof of Theorem 2 : Since we assumed that, for N large enough, $2\theta\beta^{-1} \in H_{E_N}([\lambda_{\min}(E_N)-\eta, \lambda_{\max}(E_N)+\eta]^c)$, we can find a v_N satisfying (13). Note that v_N is unique by strict monotonicity of H_{E_N} on $]-\infty\,,\,\lambda_{\min}(E_N)-\eta[$, where it is negative, and on $]\lambda_{\max}(E_N)+\eta\,,\,\infty[$, where it is positive. Therefore,

$$(2\theta)^{-1}+v_N \in [\lambda_{\min}(E_N)-\eta\,,\,\lambda_{\max}(E_N)+\eta]^c$$

ensures that

$$1-\frac{2\theta}{\beta}\lambda_i+\frac{2\theta}{\beta}v_N > \frac{2|\theta|}{\beta}\eta \tag{14}$$

so that, because of the uniform continuity of H_{E_N} on $[\lambda_{\min}(E_N)-\eta, \lambda_{\max}(E_N)+\eta]^c$, as $\hat{\mu}^N_{E_N}$ converges to μ_E, v_N converges to v the solution of $H_{\mu_E}\left(\frac{\beta}{2\theta}+v\right)=\frac{2\theta}{\beta}$ and

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\log\left(1+\frac{2\theta}{\beta}v_N-\frac{2\theta}{\beta}\lambda_i\right)=\int\log\left(1+\frac{2\theta}{\beta}v-\frac{2\theta}{\beta}\lambda\right)d\mu_E(\lambda).$$

Furthermore, the computation of the derivative of $\theta\mapsto\theta v-\frac{\beta}{2}\int\log\left(1+\frac{2\theta}{\beta}v-\frac{2\theta}{\beta}\lambda\right)d\mu_E(\lambda)$, with this particular $v=R_{\mu_E}(2\theta\beta^{-1})$ allows us to get the explicit expression

$$\theta v-\frac{\beta}{2}\int\log\left(1+\frac{2\theta}{\beta}v-\frac{2\theta}{\beta}\lambda\right)d\mu_E(\lambda)=\frac{\beta}{2}\int_0^{\frac{2\theta}{\beta}}R_{\mu_E}(u)du.$$

Therefore, Hypothesis (4) together with Lemma 14.2 finishes the proof of (6).

Now the last point is to check that under Hypothesis 1, the assumption of Lemma 14.2 is equivalent to $2\theta/\beta \in H_{\mu_E}([\lambda_{\min},\lambda_{\max}]^c)$.
Let us first observe that $H_{\mu_E}([\lambda_{\min},\lambda_{\max}]^c)=\bigcup_{\eta>0}H_{\mu_E}([\lambda_{\min}-\eta,\lambda_{\max}+\eta]^c)$ and that, under Hypothesis 1,

$$H_{\mu_E}([\lambda_{\min}-2\eta,\lambda_{\max}+2\eta]^c)\subset\bigcup_{N_0\geq 0}\bigcap_{N\geq N_0}H_{E_N}([\lambda_{\min}(E_N)-\eta,\lambda_{\max}(E_N)+\eta]^c),$$

since, for any $\lambda\in\bigcup_{N_0\geq 0}\bigcap_{N\geq N_0}\mathrm{supp}(\hat{\mu}^N_{E_N})$, the application $z\mapsto(z-\lambda)^{-1}$ is continuous bounded on $[\lambda_{\min}-2\eta,\lambda_{\max}+2\eta]^c$. Therefore, $2\theta/\beta\in H_{\mu_E}([\lambda_{\min},\lambda_{\max}]^c)$ implies the assumption of Lemma 14.2.
Conversely, we get by the same arguments that

$$\bigcup_{N_0\geq 0}\bigcap_{N\geq N_0}H_{E_N}([\lambda_{\min}(E_N)-2\eta,\lambda_{\max}(E_N)+2\eta]^c)\subset H_{\mu_E}([\lambda_{\min}-\eta,\lambda_{\max}+\eta]^c),$$

what completes the proof. ∎

2.1 Proof of Lemma 14

• The first point is trivial since the matrix U is unitary or orthogonal and hence bounded.
• Let us consider the second point. We now stick to the case $\beta = 1$ and will summarize at the end of the proof the changes to perform for the case $\beta = 2$. We can assume that the $\{\lambda_1(E_N), \cdots, \lambda_N(E_N)\}$ is not reduced to a single point $\{e\}$ since otherwise the result is straightforward. We write in short $I_N(\theta, E_N) = I_N^{(1)}(D_N, E_N)$. The ideas of the proof are very close to usual large deviations techniques, and in fact in some sense simpler because strong concentration arguments are available for free (cf. (15)). Following Fact 8, we can write, with $(\lambda_1, \cdots, \lambda_N)$ the eigenvalues of E_N,

$$I_N(\theta, E_N) = \mathbb{E}\left[\exp\left\{N\theta\frac{\sum_{i=1}^N \lambda_i g_i^2}{\sum_{i=1}^N g_i^2}\right\}\right]$$

where the g_i's are i.i.d standard Gaussian variables. Now, writing the Gaussian vector $(g_1, \ldots, g_N)$ in its polar decomposition, we realize of course that the spherical integral does not depend on its radius $r = \|g\|$ which follows the law

$$\rho_N(dr) := Z_N^{-1} r^{N-1} e^{-\frac{1}{2}r^2} dr,$$

with Z_N the appropriate normalizing constant.
The idea of the proof is now that r will of course concentrate around $\sqrt{N}$ so that we are reduced to study the numerator and to make the adequate change of variable so that it concentrates around v_N. For $\kappa < 1/2$, there exists a finite constant $C(\kappa)$ such that

$$\rho_N\left(\left|\frac{r^2}{N} - 1\right| \geqslant N^{-\kappa}\right) \leqslant C(\kappa) e^{-\frac{1}{4}N^{1-2\kappa}}. \tag{15}$$

Such an estimate can be readily obtained by applying standard precise Laplace method to the law $\tilde{\rho}_N$ of $(N-2)^{-1}r^2$ which is given by

$$\tilde{\rho}_N(dx) = \tilde{Z}_N^{-1} 1_{x \geq 0} e^{-\frac{N-2}{2}f(x)} dx$$

with $f(x) = x - \log x$. Indeed, f achieves its minimal value at $x = 1$ so that for any $\epsilon > 0$, there exists $c(\epsilon) > 0$ such that $\tilde{Z}_N \tilde{\rho}_N(|x-1| > \epsilon) \leq e^{-c(\epsilon)N}$. Now, $\sigma_\epsilon = \inf\{f''(x), |x-1| \leq \epsilon\} > 0$ so that Taylor expansion results with

$$\tilde{Z}_N \tilde{\rho}_N(|x-1| \geq N^{-\kappa}) \leq e^{-c(\epsilon)N} + \int_{y \geq N^{-\kappa}} e^{-\frac{N-2}{2}\sigma_\epsilon y^2} dy \leq e^{-\frac{\sigma_\epsilon}{3}N^{1-2\kappa}}$$

where the last inequality holds for N large enough. A lower bound on $\tilde{Z}_N$ is obtained similarly by considering $\tilde{\sigma}_\epsilon = \sup\{f''(x), |x-1| \leq \epsilon\} > 0$ showing that $\tilde{Z}_N \geq \tilde{c}(\epsilon)\sqrt{N}^{-1}$. We conclude by noticing that σ_ϵ goes to one as ϵ goes to zero. Note that such a result can also be seen as a direct consequence of section 3.7 in [8].

From this, if we introduce the event $A_N(\kappa) := \left\{\left|\frac{\|g\|^2}{N} - 1\right| \leqslant N^{-\kappa}\right\}$, it is not hard to see that for any $\kappa < \frac{1}{2}$ and for N large enough (such that $1 - C(\kappa)e^{-\frac{1}{4}N^{1-2\kappa}} > 0$), we have

$$1 \leqslant \frac{I_N(\theta, E_N)}{\mathbb{E}\left[1_{A_N(\kappa)} \exp\left\{N\theta \frac{\sum_{i=1}^N \lambda_i g_i^2}{\sum_{i=1}^N g_i^2}\right\}\right]} \leqslant \delta(\kappa, N)$$

where $\delta(\kappa, N) = \frac{1}{1-C(\kappa)e^{-\frac{1}{4}N^{1-2\kappa}}}$. Therefore,

$$\begin{aligned} I_N(\theta, E_N) &\leqslant \delta(\kappa, N)\mathbb{E}\left[1_{A_N(\kappa)} \exp\left\{N\theta \frac{\sum_{i=1}^N \lambda_i g_i^2}{\sum_{i=1}^N g_i^2}\right\}\right] \\ &\leqslant \delta(\kappa, N) e^{N\theta v + N^{1-\kappa}|\theta|(\|E_N\|_\infty + |v|)} \mathbb{E}\left[1_{A_N(\kappa)} \exp\left\{\theta \sum_{i=1}^N \lambda_i g_i^2 - v\theta \sum_{i=1}^N g_i^2\right\}\right] \end{aligned} \tag{16}$$

for any $v \in \mathbb{R}$. Now,

$$\mathbb{E}\left[1_{A_N(\kappa)} \exp\left\{\theta \sum_{i=1}^N \lambda_i g_i^2 - v\theta \sum_{i=1}^N g_i^2\right\}\right] = \prod_{i=1}^N \left[\sqrt{1 + 2\theta v - 2\theta\lambda_i}\right]^{-1} P_N(A_N(\kappa)) \tag{17}$$

with P_N the probability measure on $\mathbb{R}^N$ given by

$$P_N(dg_1, \ldots, dg_N) = \frac{1}{\sqrt{2\pi}^N} \prod_{i=1}^N \left[\sqrt{1 + 2\theta v - 2\theta\lambda_i}\ e^{-\frac{1}{2}(1+2\theta v - 2\theta\lambda_i)g_i^2} dg_i\right]$$

which is well defined provided we choose v so that

$$1 + 2\theta v - 2\theta\lambda_i > 0 \quad \forall\, i \text{ from 1 to } N. \tag{18}$$

Thus, for any such v's, we get from (16) and (17), that for any $\kappa = \frac{1}{2} - \zeta$ with $\zeta > 0$ and N large enough, since $P_N(A_N(\kappa)) \leq 1$,

$$I_N(\theta, E_N) \leqslant \delta(\kappa, N) \prod_{i=1}^N \left[\sqrt{1 + 2\theta v - 2\theta\lambda_i}\right]^{-1} e^{N\theta v + N^{1-\kappa}|\theta v|} e^{N^{1-\kappa}|\theta|\|E_N\|_\infty}. \tag{19}$$

We similarly obtain the lower bound

$$I_N(\theta, E_N) \geqslant e^{N\theta v - N^{1-\kappa}|\theta|(\|E_N\|_\infty + |v|)} \prod_{i=1}^N \left[\sqrt{1 + 2\theta v - 2\theta\lambda_i}\right]^{-1} P_N(A_N(\kappa))$$

Now, we show that we can choose v wisely so that for $N \geqslant N(\kappa)$,

$$P_N(A_N(\kappa)) = P_N(|N^{-1}||g||^2 - 1| \leqslant N^{-\kappa}) \geqslant \frac{1}{2}. \tag{20}$$

This will finish to prove, with this choice of v, that

$$I_N(\theta, E_N) \geqslant \frac{1}{2} e^{N\theta v - N^{1-\kappa}|\theta|(\|E_N\|_\infty + |v|)} \prod_{i=1}^{N} \left[\sqrt{1 + 2\theta v - 2\theta\lambda_i}\right]^{-1} \tag{21}$$

yielding the desired lower bound.
We know that P_N is a product measure under which

$$\tilde{g}_i = \sqrt{1 + 2\theta v - 2\theta\lambda_i}\; g_i$$

are i.i.d standard Gaussian variables. Let us now choose $v = v_N$ in $-(2\theta)^{-1} + [\lambda_{\min}(E_N) - \eta\,,\, \lambda_{\max}(E_N) + \eta]^c$ satisfying

$$\mathbb{E}_{P_N}\left[\frac{1}{N}\|g\|^2\right] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\tilde{g}_i^2}{1 + 2\theta v_N - 2\theta\lambda_i}\right] = \frac{1}{2\theta} H_{E_N}\left((2\theta)^{-1} + v_N\right) = 1. \tag{22}$$

We recall from (14) that $1 - 2\theta\lambda_i + 2\theta v_N > 2|\theta|\eta > 0$ so that all our computations are validated by this final choice.

With this choice of v_N, we have

$$\mathbb{E}_{P_N}\left[\left(\frac{1}{N}\|g\|^2 - 1\right)^2\right] = \frac{2}{N^2}\sum_{i=1}^{N} \frac{1}{(1 + 2\theta v_N - 2\theta\lambda_i)^2} \leqslant \frac{2}{N\theta^2\eta^2}$$

so that by Chebychev's inequality

$$P_N(|N^{-1}||g||^2 - 1| \geqslant N^{-\kappa}) \leqslant \frac{2}{\eta^2\theta^2} N^{2\kappa - 1},$$

which is smaller than 2^{-1} for sufficiently large N since $2\kappa < 1$, resulting with (20).
Finally, since by definition

$$\frac{1}{N}\sum_{i=1}^{N} \frac{1}{1 - 2\theta\lambda_i + 2\theta v_N} = 1$$

with $(\lambda_i)_{1 \le i \le N}$ which do not all take the same value, there exists i and j so that

$$-2\theta\lambda_i + 2\theta v_N > 0, \quad -2\theta\lambda_j + 2\theta v_N < 0$$

so that $v_N \in [\lambda_{\min}(E_N), \lambda_{\max}(E_N)]$. Thus, (21) together with (19) give the second point of the lemma for $\beta = 1$.

In the case where $\beta = 2$, the g_i^2 have to be replaced everywhere by $g_i^2 + \hat{g}_i^2$ with independent Gaussian variables $(g_i, \hat{g}_i)_{1\leqslant i\leqslant N}$. This time, we can concentrate

$$\frac{1}{N}\|g\|^2 = \frac{1}{N}\sum_{i=1}^N g_i^2 + \frac{1}{N}\sum_{i=1}^N \hat{g}_i^2$$

around 2. Everything then follows by dividing θ by two and noticing that we will get the same Gaussian integrals squared.

• The last point is an easy consequence of the second since, for any $\lambda \in \bigcup_{N_0\geq 0}\bigcap_{N\geq N_0}(\mathrm{supp}(\hat{\mu}^N_{E_N}) \cap \mathrm{supp}(\hat{\mu}^N_{\tilde{E}_N}))$, the application $z \mapsto (z-\lambda)^{-1}$ is continuous bounded (with norm depending on η) on $\bigcup_{N_0\geq 0}\bigcap_{N\geq N_0}[\lambda_{\min}(E_N) - \eta, \lambda_{\max}(E_N) + \eta]^c$.

∎

2.2 Generalisation of the method to the multi-dimensional case

In the sequel, we want to apply the strategy we used above to show Theorem 7, that is to say study the behaviour of the spherical integrals as the rank of D_N remains negligible compared to $\sqrt{N}$. In this case and if all the eigenvalues of D_N are small enough, we show that it behaves like a product, namely that we have the equality (11). To lighten the notations, we let $\theta_i := \theta_i^N$, for all $i \leqslant M(N)$.
We will rely again on Fact 8 and write in the case $\beta = 1$,

$$I_N(D_N, E_N) = \mathbb{E}\left[\exp\left\{N\sum_{m=1}^M \theta_m \frac{\sum_{i=1}^N \lambda_i(\tilde{g}_i^{(m)})^2}{\sum_{i=1}^N (\tilde{g}_i^{(m)})^2}\right\}\right], \tag{23}$$

where the expectation is taken under the standard Gaussian measure and the vectors $(\tilde{g}^{(1)}, \dots, \tilde{g}^{(M)})$ are obtained from the Gaussian vectors $(g^{(1)}, \dots, g^{(M)})$ by a standard Schmidt orthogonalisation procedure.
This means that there exists a lower triangular matrix $A = (A_{ij})_{1\leqslant i,j\leqslant M}$ such that for any integer m between 1 and M,

$$\tilde{g}^{(m)} = g^{(m)} + \sum_{j=1}^{m-1} A_{mj} g^{(j)}$$

and the A_{ij}'s are solutions of the following system : for all p from 1 to $m-1$,

$$\langle g^{(m)}, g^{(p)}\rangle + \sum_{j=1}^{m-1} A_{mj}\langle g^{(j)}, g^{(p)}\rangle = 0, \tag{24}$$

with $\langle .,. \rangle$ the usual scalar product in $\mathbb{R}^N$.

Therefore, if we denote, for i and j between 1 and M, with $i \leqslant j$,

$$X_N^{ij} := \frac{1}{N}\langle g^{(i)}, g^{(j)} \rangle$$

and

$$Y_N^{ij} := \frac{1}{N}\sum_{l=1}^{N} \lambda_l g_l^{(i)} g_l^{(j)},$$

then, for each m from 1 to M, there exists a rational function $F_m : \mathbb{R}^{m(m+1)} \to \mathbb{R}$ such that

$$\frac{\sum_{i=1}^N \lambda_i (\tilde{g}_i^{(m)})^2}{\sum_{i=1}^N (\tilde{g}_i^{(m)})^2} = F_m((X_N^{ij}, Y_N^{ij})_{1\leqslant i\leqslant j\leqslant m}) \tag{25}$$

and a rational function $G_m : \mathbb{R}^{\frac{m(m+1)}{2}} \to \mathbb{R}$ such that

$$\frac{1}{N}\sum_{i=1}^{N} (\tilde{g}_i^{(m)})^2 = G_m((X_N^{ij})_{1\leqslant i\leqslant j\leqslant m}). \tag{26}$$

We now adopt the following system of coordinates in $\mathbb{R}^{MN}$: r_1, $\alpha_1^{(1)}, \ldots, \alpha_{N-1}^{(1)}$ are the polar coordinates of $g^{(1)}$, $r_2 := \|g^{(2)}\|$, β_2 is the angle between $g^{(1)}$ and $g^{(2)}$, $\alpha_1^{(2)}, \ldots, \alpha_{N-2}^{(2)}$ are the angles needed to spot $g^{(2)}$ on the cone of angle β_2 around $g^{(1)}$, then $r_3 := \|g^{(3)}\|$, β_3^i the angle between $g^{(3)}$ and $g^{(i)}$ ($i = 1, 2$) and $\alpha_1^{(3)}, \ldots, \alpha_{N-3}^{(3)}$ the angles needed to spot $g^{(3)}$ on the intersection of the two cones...etc...
Then observe that $F_m((X_N^{ij}, Y_N^{ij})_{1\leqslant i\leqslant j\leqslant m})$ depends only on the α's (because the $\frac{\tilde{g}^{(i)}}{\|\tilde{g}^{(i)}\|}$ do) whereas $G_m((X_N^{ij})_{1\leqslant i\leqslant j\leqslant m})$ depends on the r's and the β's. Therefore, if we consider the event

$$B_N(\kappa) := \left\{\forall i, \quad \left|X_N^{ii} - 1\right| \leqslant N^{-\kappa}, \quad \forall i \neq j \quad \left|X_N^{ij}\right| \leqslant N^{-\kappa}\right\},$$

then, as in the case of rank one, we can write that

$$I_N(D_N, E_N) \leqslant \mathbb{E}\left[1_{B_N(\kappa)} e^{N\theta_m F_m(X_N^{ij}, Y_N^{ij})}\right] + P(B_N(\kappa)^c) I_N(D_N, E_N). \tag{27}$$

Now we claim that, for N large enough, for any $\kappa > 0$, there exists an $\alpha > 0$ such that

$$P(B_N(\kappa)^c) \leqslant C'(\kappa) e^{-\alpha N^{1-2\kappa}}. \tag{28}$$

Indeed, as in (15),

$$\begin{aligned} P(B_N(\kappa)^c) &\leqslant \sum_{i=1}^{M} P\left(\left|X_N^{ii} - 1\right| > N^{-k}\right) + \sum_{i,j=1}^{M} P\left(\left|X_N^{ij}\right| > N^{-k}\right) \\ &\leqslant c_1(\kappa) M e^{-\frac{1}{4}N^{1-2\kappa}} + c_2(\kappa) M^2 e^{-\frac{1}{2}N^{1-2\kappa}}, \end{aligned}$$

what gives immediately (28).
Now, as far as $\kappa < \frac{1}{2}$, (27) together with (28) give

$$1 \leqslant \frac{I_N(D_N, E_N)}{\mathbb{E}\left[1_{B_N(\kappa)} e^{NF_m(X_N^{ij}, Y_N^{ij})}\right]} \leqslant 1 + \epsilon(N, k),$$

with $\epsilon(N, k)$ going to zero.

We now want to expand F_M on $B_N(\kappa)$ as we did in the previous subsection.
As the A_{ij}'s satisfy the linear system (24), we can write the Cramer's formulas corresponding to it and get

$$A_{ij} = \frac{\det(R_N^{kl})_{1\leqslant k,l\leqslant i-1}}{\det(X_N^{kl})_{1\leqslant k,l\leqslant i-1}},$$

where

$$R_N^{kl} = \begin{cases} X_N^{kl}, & \text{if } l \neq j \\ -X_N^{ki} & \text{if } l = j. \end{cases}$$

Now, we look at the denominator and can show that

$$\det(X_N^{kl})_{1\leqslant k,l\leqslant i-1} \geqslant 1 - \sum_{s=1}^{i-1} (MN^{-\kappa})^s \geqslant \frac{1}{2},$$

where the last inequality holds for N large enough as far as $M = o(N^\kappa)$.
We now go to the numerator : expanding over the jth column, we get this time that

$$\det(R_N^{kl})_{1\leqslant k,l\leqslant i-1} \leqslant N^{-\kappa} + (M-1)N^{-2\kappa} \sum_{s=1}^{i-1} (MN^{-\kappa})^s \leqslant cN^{-\kappa},$$

where again the last equality holds as far as $M = o(N^\kappa)$ and c is a fixed constant.
From the two last inequalities, we have that, on $B_N(\kappa)$, $\sup_{i<j} |A_{ij}| \leqslant c' N^{-\kappa}$.
From that we can easily deduce that, for any m less than M, we have

$$\frac{1}{N} \left\|\tilde{g}^{(m)} - g^{(m)}\right\|^2 \leqslant \frac{1}{N} \sum_{i,j=1}^{m-1} |A_{mj} A_{mi}| |\langle g^{(i)}, g^{(j)}\rangle|^2 \leqslant c'' N^{-2\kappa}(M^2 N^{-2\kappa} + M) \leqslant c_3 N^{-\kappa}.$$

From these estimations and (23), for any v_j^N, we get the following upper bound :

$$I_N(D_N, E_N) \leqslant (1 + \epsilon(\kappa, N)) \exp\left\{N \sum_{j=1}^{M} \theta_j v_j^N\right\}$$
$$\mathbb{E}\left[1_{B_N(\kappa)} \prod_{j=1}^{M} \exp\left\{N\theta_j \frac{\frac{1}{N}\sum_{i=1}^N \lambda_i \left(\tilde{g}_i^{(j)}\right)^2 - v_j^N \frac{1}{N}\sum_{i=1}^N \left(\tilde{g}_i^{(j)}\right)^2}{1 + \frac{1}{N}\left(\|\tilde{g}^{(j)}\|^2 - \|g^{(j)}\|^2\right) + \left(\frac{1}{N}\|g^{(j)}\|^2 - 1\right)}\right\}\right]$$

$$\leqslant (1+\epsilon(\kappa,N))\exp\Big\{N\sum_{j=1}^{M}\theta_j v_j^N\Big\}$$
$$\mathbb{E}\left[1_{B_N(\kappa)}\prod_{j=1}^{M}\exp\left\{\left(\theta_j\sum_{i=1}^{N}\lambda_i\left(\tilde{g}_i^{(j)}\right)^2 - v_j^N\theta_j\sum_{i=1}^{N}\left(\tilde{g}_i^{(j)}\right)^2\right)\left[1+c_4N^{-\kappa}\right]\right\}\right]$$

$$\leqslant (1+\epsilon(\kappa,N))\exp\Big\{N\sum_{j=1}^{M}\theta_j v_j^N\Big\}\exp\left\{C\sup|\theta_j|(\|E_N\|_\infty+\sup|v_j^N|)MN^{1-\kappa}\right\}$$
$$\mathbb{E}\left[\prod_{j=1}^{M}\exp\left\{\theta_j\sum_{i=1}^{N}\lambda_i\left(g_i^{(j)}\right)^2 - v_j^N\sum_{i=1}^{N}\left(g_i^{(j)}\right)^2\right\},\right].$$

where C is again a fixed constant.
From the hypotheses of Theorem 7, we know that there exists an N such that $2\theta_j \in H_{E_N}([\lambda_{\min}(E_N)-\eta,\lambda_{\max}(E_N)+\eta]^c)$, from which we can easily deduce that $|2\theta_j|\leqslant \eta^{-1}$. Moreover, as in the proof of Lemma 14.2, $|v_j^N|\leq ||E_N||_\infty$ is uniformly bounded. Therefore, we get
$$\limsup_{N\to\infty}\frac{1}{NM(N)}\log I_N(D_N,E_N)\leqslant \int I_{\mu_E}(\theta)d\mu_D(\theta).$$
We also get a similar lower bound and conclude similarly to the preceding subsection by considering the shifted probability measure $P_N^{\theta_1,\ldots,\theta_M}=\otimes_{j=1}^M P_N^{\theta_j}$ where

$$P_N^{\theta_j}(dg_1,\ldots,dg_N)=\frac{1}{\sqrt{2\pi}^N}\prod_{i=1}^{N}\sqrt{1+2\theta_j v_j^N-2\theta_j\lambda_i}\; e^{-\frac{1}{2}(1+2\theta_j v_j^N-2\theta_j\lambda_i)g_i^2}dg_i.$$

This concludes the proof of Theorem 7. ∎

3 Central limit theorem in the case of rank one

Under the hypotheses of Theorem 2, v_N (defined by (13)) is converging to $v=R_{\mu_E}\left(\frac{2\theta}{\beta}\right)$ and we established that the spherical integral is converging to $\theta v-\frac{\beta}{2}\int\log\left(1+\frac{2\theta}{\beta}v-\frac{2\theta}{\beta}\lambda\right)d\mu_E(\lambda)$. In the case where the fluctuations of the eigenvalues do not interfere, we can get sharper estimates, given, in the case $\beta=1$, by Theorem 3. This section is devoted to its proof, namely the study of the behaviour of $e^{-N\left(\theta R_{\mu_E}(2\theta)-\frac{1}{2N}\sum\log(1+2\theta R_{\mu_E}(2\theta)-2\theta\lambda_i)\right)}I_N(\theta,E_N)$.

Proof of Theorem 3

- We first treat the non degenerate case $\mu_E\neq\delta_e$.

Let us first make an important remark : the hypothesis that $d(\hat{\mu}^N_{E_N}, \mu_E) = o(\sqrt{N}^{-1})$ has the two following consequences :

$$|v - v_N| = o(\sqrt{N}^{-1}) \tag{29}$$

$$\text{and} \quad \lim_{N\to\infty} \sqrt{N}(H_{E_N} - H_{\mu_E})(K_{\mu_E}(2\theta)) = 0. \tag{30}$$

Indeed, since $2\theta \in H_{\mu_E}([\lambda_{min}, \lambda_{max}]^c)$, there is an $\eta > 0$, such that, for N large enough, $2\theta \in H_{E_N}([\lambda_{\min}(E_N) - \eta, \lambda_{\max}(E_N) + \eta]^c)$. Therefore, as for any λ which is in $\mathrm{supp}(\hat{\mu}^N_{E_N})$ for N large enough, $z \mapsto (z-\lambda)^{-1}$ is uniformly bounded and Lipschitz on $\bigcap_{N \geqslant N_0}[\lambda_{\min}(E_N) - \eta, \lambda_{\max}(E_N)+\eta]^c$, we get directly (29), and also (30) as we know that $K_{\mu_E}(2\theta) \in [\lambda_{min}, \lambda_{max}]^c$.

For $v = R_{\mu_E}(2\theta)$, we set

$$\gamma_N = \left(\frac{1}{N} \sum_{i=1}^N g_i^2 - 1 \right) \text{ and } \hat{\gamma}_N = \left(\frac{1}{N} \sum_{i=1}^N \lambda_i g_i^2 - v \right).$$

Let us also define for $\epsilon > 0$

$$I_N^\epsilon(\theta, E_N) := \int_{|\gamma_N| \leqslant \epsilon, |\hat{\gamma}_N| \leqslant \epsilon} \exp\left\{ \theta N \frac{\hat{\gamma}_N + v}{\gamma_N + 1} \right\} \prod_{i=1}^N dP(g_i),$$

with P the standard Gaussian probability measure on $\mathbb{R}$. We claim that, for any $\zeta > 0$, for N large enough,

$$|I_N(\theta, E_N) - I_N^\epsilon(\theta, E_N)| \leqslant e^{-N^{1-\zeta}} I_N(\theta, E_N). \tag{31}$$

Indeed, consider

$$\mu_N^\theta(dg) = \frac{1}{I_N(\theta, E_N)} \exp\left\{ \theta N \frac{\sum_{i=1}^N \lambda_i g_i^2}{\sum_{i=1}^N g_i^2} \right\} \prod_{i=1}^N dP(g_i).$$

(31) is equivalent to

$$\mu_N^\theta\left(|\gamma_N| \geq \epsilon\right) \leq \frac{1}{2} e^{-N^{1-\zeta}} \text{ and } \mu_N^\theta\left(|\hat{\gamma}_N| \geq \epsilon\right) \leq \frac{1}{2} e^{-N^{1-\zeta}} \tag{32}$$

The first inequality is trivial since by (15), for $\kappa < \frac{1}{2}$,

$$\mu_N^\theta\left(|\gamma_N| \geq N^{-\kappa}\right) = \rho_N\left(\left| \frac{r^2}{N} - 1 \right| \geq N^{-\kappa} \right) \leq e^{-\frac{1}{4} N^{1-2\kappa}}.$$

To show the second point, following the proof of Lemma 14, we find a finite constant $C(\kappa)$ so that

$$\mu_N^\theta\left(|\hat{\gamma}_N| \geq \epsilon\right) \leq C(\kappa) e^{C(\kappa) N^{1-\kappa} |\theta| ||E_N||_\infty} P_N\left(|\hat{\gamma}_N| \geq \epsilon\right)$$

where under P_N the g_i are independent centered Gaussian variable with covariance $(1 - 2\theta\lambda_i + 2\theta v_N)^{-1}$. Hence

$$P_N\left(|\hat{\gamma}_N| \geq \epsilon\right) = P^{\otimes N}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_i}{1-2\theta\lambda_i+2\theta v_N}\tilde{g}_i^2 - v\right| \geq \epsilon\right).$$

Let us denote $\tilde{E}_N = \phi_{v_N}(E_N)$ with $\phi_v(x) = x(1-2\theta x+2\theta v)^{-1}$. Then, the spectral measure of $\tilde{E}_N$ converges towards $\mu_{\tilde{E}} := \phi_v \sharp \mu_E$ since v_N converges towards v (see (29)). Moreover $\lambda_{\min}(\tilde{E}_N)$ and $\lambda_{\max}(\tilde{E}_N)$ converge. Hence, we can apply Proposition 18 to obtain a large deviation principle for the law of $\frac{1}{N}\sum_{i=1}^N \lambda_i(\tilde{E}_N)\tilde{g}_i^2$ under $P^{\otimes N}$ with good rate function $L(z) = \inf_u \mathcal{K}(u,z) = \frac{1}{2}h_z(K_{\mu_E}(Q_{\mu_E}(z)))$. Following the proof of Lemma 21, L has a unique minimizer which is

$$z_0 = R_{\mu_{\tilde{E}}}(0) = \int \frac{\lambda}{1-2\theta\lambda+2\theta v}d\mu_E(\lambda) = v.$$

As a consequence, for $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ so that for N large enough

$$P^{\otimes N}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\frac{\lambda_i}{1-2\theta\lambda_i+2\theta v_N}\tilde{g}_i^2 - v\right| > \epsilon\right) \leq e^{-\delta(\epsilon)N}.$$

This completes the proof of (32).

We now deal with $I_N^\epsilon(\theta, E_N)$. We use the expansion $\dfrac{1}{1+\gamma_N} = 1 - \gamma_N + \dfrac{\gamma_N^2}{1+\gamma_N}$ to get that

$$I_N^\epsilon(\theta, E_N) = e^{N\theta v}\int_{|\gamma_N|\leqslant\epsilon, |\hat{\gamma}_N|\leqslant\epsilon} \exp\left\{-\theta N\gamma_N\frac{\hat{\gamma}_N - v\gamma_N}{\gamma_N+1}\right\}\exp\{\theta N(\hat{\gamma}_N - v\gamma_N)\}\prod_{i=1}^N dP(g_i).$$

We note that

$$\exp\{\theta N(\hat{\gamma}_N - v\gamma_N)\}\prod_{i=1}^N dP(g_i) = \prod_{i=1}^N\left[\sqrt{1+2\theta v-2\theta\lambda_i}\right]^{-1}\prod_{i=1}^N dP_i(g_i)$$

with P_i the centered Gaussian probability measure

$$dP_i(x) = \sqrt{(2\pi)^{-1}(1+2\theta v-2\theta\lambda_i)}\ \exp\left\{-\frac{1}{2}(1+2\theta v-2\theta\lambda_i)x^2\right\}dx.$$

We have that

$$1+2\theta v - 2\theta\lambda_i = 2\theta(K_{\mu_E}(2\theta) - \lambda_i) \tag{33}$$

and we know that $K_{\mu_E}(2\theta) \in [\lambda_{\min}, \lambda_{\max}]^c$. Further, arguing as in (14), we find, for any given $\theta > 0$, a constant $\eta_\theta > 0$ such that

$$\inf_{1\leqslant i\leqslant N}(1+2\theta v-2\theta\lambda_i) > \eta_\theta$$

insuring that the P_i are well defined. Therefore,

$$I_N^\epsilon(\theta, E_N) = e^{N\theta v - \frac{N}{2}\int \log\left(2\theta(K_{\mu_E}(2\theta)-\lambda)\right)d\hat{\mu}_{E_N}(\lambda)}$$
$$\int_{|\gamma_N|\leqslant\epsilon, |\hat{\gamma}_N|\leqslant\epsilon} \exp\left\{-\theta N\gamma_N \frac{\hat{\gamma}_N - v\gamma_N}{\gamma_N+1}\right\}\prod_{i=1}^N dP_i(g_i) \quad (34)$$

Now, under $\prod_{i=1}^N dP_i(g_i)$, $\left(\sqrt{N}\gamma_N, \sqrt{N}\hat{\gamma}_N\right)$ converges in law towards a centered two-dimensional Gaussian variables (Γ_1, Γ_2) as soon as their covariances converge. We investigate this convergence.

Hereafter, we shall write $g_i = (1 + 2\theta(v - \lambda_i))^{-\frac{1}{2}}\tilde{g}_i$ with standard independent Gaussian variables $\tilde{g}_i$. Then,

$$\mathbb{E}((\sqrt{N}\gamma_N)^2) = N\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{g}_i^2 - 1}{1 + 2\theta v - 2\theta\lambda_i} + \frac{1}{2\theta}(H_{E_N} - H_{\mu_E})(K_{\mu_E}(2\theta))\right)^2\right]$$

where we used that

$$2\theta = H_{\mu_E}(K_{\mu_E}(2\theta)) = \int \frac{1}{K_{\mu_E}(2\theta) - \lambda} d\mu_E(\lambda)\,, \quad (35)$$

and (33). Equation (30) implies

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{E}((\sqrt{N}\gamma_N)^2) &= \lim_{N\to\infty} N\mathbb{E}\left[\left(\frac{1}{N}\sum_{i=1}^N \frac{\tilde{g}_i^2 - 1}{1 + 2\theta v - 2\theta\lambda_i}\right)^2\right] \\
&= \lim_{N\to\infty} \frac{2}{N}\sum_{i=1}^N \frac{1}{(1 + 2\theta v - 2\theta\lambda_i)^2} \\
&= \frac{1}{2\theta^2}\int \frac{1}{(K_{\mu_E}(2\theta) - \lambda)^2} d\mu_E(\lambda) := \frac{Z}{2\theta^2},
\end{aligned}$$

where the above convergence holds since $K_{\mu_E}(2\theta)$ lies outside $[\lambda_{\min}, \lambda_{\max}]$ and therefore outside the support of μ_E.

Similar computations give that under the same hypotheses,

$$\lim_{N\to\infty} \mathbb{E}((\sqrt{N}\hat{\gamma}_N)^2) = \frac{1}{2\theta^2}\int \frac{\lambda^2}{(K_{\mu_E}(2\theta) - \lambda)^2} d\mu_E(\lambda)$$

and that

$$\lim_{N\to\infty} \mathbb{E}(\sqrt{N}\hat{\gamma}_N \sqrt{N}\gamma_N) = \frac{1}{2\theta^2}\int \frac{\lambda}{(K_{\mu_E}(2\theta) - \lambda)^2} d\mu_E(\lambda)\,.$$

Therefore, provided that the Gaussian integral is well defined, we find that

$$I_N(\theta, E_N) = e^{N\theta v - \frac{N}{2}\int \log\left(2\theta(K_{\mu_E}(2\theta)-\lambda)\right)d\hat{\mu}^N_{E_N}(\lambda)} \int \exp\{-\theta x(y - vx)\} d\Gamma(x,y)(1+o(1)), \quad (36)$$

with Γ a centered Gaussian measure on $\mathbb{R}^2$ with covariance matrix

$$R = \frac{1}{2\theta^2}\begin{bmatrix} \int \frac{1}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) & \int \frac{\lambda}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) \\ \int \frac{\lambda}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) & \int \frac{\lambda^2}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) \end{bmatrix},$$

where we used the notation $K_{\mu_E} := K_{\mu_E}(2\theta)$.

Following [3], we know that there is one step needed to justify this derivation, namely to check that the Gaussian integration in (36) is non-degenerate. If we set $D := 4\theta^4 \det R$, then, using the relation (35), one finds that $D = Z - 4\theta^2$, and that the Gaussian integral in (36) equals

$$\frac{\theta^2}{\pi\sqrt{D}} \int \exp\left(-\frac{1}{2}\sum_{i,j=1}^{2} K_{i,j} x_i x_j\right) dx_1 dx_2\,,$$

where the matrix K equals

$$K = \theta \begin{bmatrix} -2v & 1 \\ 1 & 0 \end{bmatrix} + R^{-1} = \frac{2\theta^2}{D}\begin{bmatrix} \int \frac{\lambda^2}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) - \frac{(K_{\mu_E}-\frac{1}{2\theta})D}{\theta} & -\int \frac{\lambda}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) + \frac{D}{2\theta} \\ -\int \frac{\lambda}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) + \frac{D}{2\theta} & \int \frac{1}{(K_{\mu_E}-\lambda)^2} d\mu_E(\lambda) \end{bmatrix} \quad (37)$$

Our task is to verify that K is positive definite. It is enough to check that $K_{11} > 0$ and $\det K > 0$. Re-expressing K_{11}, one finds that

$$\begin{aligned} K_{11} &= \frac{2\theta^2}{D}\left(1 - 4\theta K_{\mu_E} + K_{\mu_E}^2 Z - \frac{1}{\theta}(Z - 4\theta^2)\left(K_{\mu_E} - \frac{1}{2\theta}\right)\right) \\ &= \frac{2\theta^2}{D}\left(\left(K_{\mu_E} - \frac{1}{2\theta}\right)^2 Z + \frac{Z}{4\theta^2} - 1\right) \end{aligned}$$

But Schwarz's inequality applied to (35) yields that $Z > 4\theta^2$ as soon as μ_E is not degenerate, implying that

$$K_{11} > \left(K_{\mu_E} - \frac{1}{2\theta}\right)^2 Z \geqslant 0\,,$$

as needed. Turning to the evaluation of the determinant, note that

$$\det K = \frac{4\theta^4}{D^2} Z\left(\frac{Z}{4\theta^2} - 1\right) > 0\,,$$

where the last inequality is again due to (35). The proof is complete as our limit is given by $\det(K)^{-1/2}\det(R)^{-1/2}$.

• Let us finally consider the case $\mu_E = \delta_e$. In this case, $H_{\mu_E}(x) = (x-e)^{-1}$ and $K_{\mu_E}(x) = x^{-1} + e$, $v = e$ (note also that Z in Theorem 3.1 is equal to $4\theta^2$). We can follow the previous proof but then

$$\lim_{N\to\infty} \mathbb{E}[(\sqrt{N}(\hat{\gamma}_N - v\gamma_N))^2] = 0.$$

From here, we argue again using [3] that

$$\lim_{N\to\infty} \mathbb{E}[1_{|\gamma_N|\le\epsilon, |\hat{\gamma}_N - v\gamma_N|\le\epsilon}\; e^{-\theta(1+\gamma_N)^{-1}\sqrt{N}\gamma_N\sqrt{N}(\hat{\gamma}_N - v\gamma_N)}] = 1$$

which completes the proof of Theorem 3.

∎

4 Extension of the results to the complex plane

In this section, we would like to extend the results of section 2 to the case where θ is complex, that is to show Theorem 4.

As in the real case, we first would like to write that

$$I_N(\theta, E_N) = \prod_{i=1}^N \sqrt{\zeta_i} \int \exp\left\{\theta N \frac{\sum_{i=1}^N \lambda_i \zeta_i g_i^2}{\sum_{i=1}^N \zeta_i g_i^2} - \frac{1}{2}\sum_{i=1}^N \zeta_i g_i^2\right\} \prod_{i=1}^N dg_i, \tag{38}$$

with $\zeta_i = \dfrac{1}{1 + 2\theta v - 2\theta\lambda_i}$, for v such that $\Re(\zeta_i) > 0, \quad \forall i$ with $1 \le i \leqslant N$.

This is a direct consequence of the following lemma

Lemma 15 *For any function $f : \mathbb{C}^N \longrightarrow \mathbb{C}$ which is invariant by $x \mapsto -x$, analytic outside 0 and bounded on $\{z = x + iy \in \mathbb{C}/|y| < x\}^N$ and for any $(\zeta_1, \dots, \zeta_N)$ such that $\Re(\zeta_i) > 0$ for any i from 1 to N, we have that*

$$\begin{aligned}\mathcal{J}_N &:= \int f(g_1, \dots, g_N) e^{-\frac{1}{2}\sum_{i=1}^N g_i^2} \prod_{i=1}^N dg_i \\ &= \prod_{i=1}^N \sqrt{\zeta_i} \int f(\sqrt{\zeta_1} g_1, \dots, \sqrt{\zeta_N} g_N) e^{-\frac{1}{2}\sum_{i=1}^N \zeta_i g_i^2} \prod_{i=1}^N dg_i,\end{aligned}$$

with $\sqrt{\cdot}$ is the principal branch of the square root in $\mathbb{C}$.

Proof of Lemma 15:

We denote by r_j the modulus of ζ_j and α_j its phase ($\zeta_j = r_j e^{\alpha_j}$).

As f is bounded on $\mathbb{R}^N$, dominated convergence gives that

$$\mathcal{J}_N = \lim_{R\to\infty,\epsilon\to 0} \int_{[-R,R]^N\setminus[-\epsilon,\epsilon]^N} f(g_1,\dots,g_N) e^{-\frac{1}{2}\sum_{i=1}^N g_i^2} \prod_{i=1}^N dg_i.$$

Thanks to invariance of f by $x \mapsto -x$, we also have that

$$\mathcal{J}_N = \lim_{R\to\infty,\epsilon\to 0} 2^N \int_{[\epsilon,R]^N} f(g_1,\dots,g_N) e^{-\frac{1}{2}\sum_{i=1}^N g_i^2} \prod_{i=1}^N dg_i.$$

For each j between 1 and N and $R \in \mathbb{R}^+$, we define the following segments in $\mathbb{C}$:

$$\mathcal{C}^j_{R,\epsilon} := \left\{ r e^{i\frac{\alpha_j}{2}} ; \epsilon \leqslant r \leqslant R \right\},$$

and the following arc of circles

$$\mathcal{D}^j_\epsilon := \left\{ \epsilon e^{i\alpha} ; 0 \leqslant \alpha \leqslant \frac{\alpha_j}{2} \right\} \text{ and } \mathcal{D}^j_R := \left\{ R e^{i\alpha} ; 0 \leqslant \alpha \leqslant \frac{\alpha_j}{2} \right\},$$

so that, for each j, $[\epsilon, R]$ run from ϵ to R followed by $\mathcal{D}^j_R$ run counterclockwise, followed by $\mathcal{C}^j_{R,\epsilon}$ run from $Re^{i\frac{\alpha_j}{2}}$ to $\epsilon e^{i\frac{\alpha_j}{2}}$ followed by $\mathcal{D}^j_\epsilon$ run clockwise form a closed path.
Therefore, if we let

$$\begin{aligned} f_1^{x_2,\dots,x_N} :\ & \mathbb{C} \to \mathbb{C} \\ & x \mapsto f(x, x_2, \dots, x_N), \end{aligned}$$

then for any $(x_2,\dots,x_N) \in \mathbb{C}^{N-1}$, $x \mapsto f_1^{x_2,\dots,x_N}(x) e^{-\frac{1}{2}x^2}$ is analytic inside the contour $[\epsilon, R] \cup \mathcal{D}^j_R \cup \mathcal{C}^j_{R,\epsilon} \cup \mathcal{D}^j_\epsilon$, so that Cauchy's theorem implies

$$\begin{aligned} \int_{[\epsilon,R]} f_1^{g_2,\dots,g_N}(g_1) e^{-\frac{1}{2}g_1^2} dg_1 = & \int_{\mathcal{C}^1_{R,\epsilon}} f_1^{g_2,\dots,g_N}(g_1) e^{-\frac{1}{2}g_1^2} dg_1 \\ & - \int_{\mathcal{D}^1_R} f_1^{g_2,\dots,g_N}(g_1) e^{-\frac{1}{2}g_1^2} dg_1 + \int_{\mathcal{D}^1_\epsilon} f_1^{g_2,\dots,g_N}(g_1) e^{-\frac{1}{2}g_1^2} dg_1. \end{aligned}$$

If we denote by

$$J^1_{N,R} = \int_{[\epsilon,R]^{N-1}} e^{-\frac{1}{2}\sum_{i=2}^N g_i^2} \int_{\mathcal{D}^1_R} f_1^{g_2,\dots,g_N}(g_1) e^{-\frac{1}{2}g_1^2} dg_1 \dots dg_N,$$

we have that

$$\begin{aligned}|J^1_{N,R}| &= \int_{[\epsilon,R]^{N-1}}\int_0^{\frac{\alpha_1}{2}} f(g_1,\dots,g_N)e^{-\frac{1}{2}\sum_{i=2}^N g_i^2} Re^{-\frac{1}{2}R^2\cos(2u_1)}du_1 dg_2\dots dg_N \\ &\leqslant \|f\|_\infty \sqrt{2\pi}^N \frac{\alpha_1}{2} Re^{-\frac{1}{2}R^2\cos(\alpha_1)}.\end{aligned}$$

As $\cos(\alpha_1) > 0$, we have that for any ϵ, $\lim_{R\to\infty} |J^1_{N,R}| = 0$.
In the same way, if we let

$$L^1_{N,\epsilon} = \int_{[\epsilon,R]^{N-1}} e^{-\frac{1}{2}\sum_{i=2}^N g_i^2} \int_{\mathcal{D}^1_\epsilon} f_1^{g_2,\dots,g_N}(g_1)e^{-\frac{1}{2}g_1^2}dg_1\dots dg_N,$$

then we have that

$$|L^1_{N,\epsilon}| \leqslant \|f\|_\infty \sqrt{2\pi}^N \epsilon \frac{\alpha_1}{2},$$

so that $\lim_{\epsilon\to 0}|L^1_{N,\epsilon}| = 0$.
By doing the same computation for each variable, we get that

$$\lim_{R\to\infty,\epsilon\to 0}\int_{[\epsilon,R]^N} f(g_1,\dots,g_N)e^{-\frac{1}{2}\sum_{i=1}^N g_i^2}\prod_{i=1}^N dg_i = \lim_{R\to\infty,\epsilon\to 0}\int_{\prod_{i=1}^N \mathcal{C}^1_{R,\epsilon}} f(g_1,\dots,g_N)e^{-\frac{1}{2}\sum_{i=1}^N g_i^2}\prod_{i=1}^N dg_i.$$

The last step is to make the change of variable in $\mathbb{R}$ which consist in letting $\tilde{g}_j = \sqrt{r_j}g_j$ to get the result announced in the lemma 15 and therefore the formula (38). ∎

We now go back to the **proof of Theorem 4** and proceed as in section 2. We let

$$\gamma_N := \frac{1}{N}\sum_{i=1}^N \zeta_i g_i^2 - 1 \text{ and } \hat{\gamma}_N := \frac{1}{N}\sum_{i=1}^N \lambda_i\zeta_i g_i^2 - v(\theta),$$

with $v(\theta) = R_{\mu_E}(2\theta)$, which, for $|\theta|$ small enough, is well defined and such that $\Re\zeta_i > 0$, by virtue of Property 12 and Proposition 13.

Therefore, we find that

$$I_N(\theta, E_N) = \prod_{i=1}^N \sqrt{\zeta_i}\ e^{N\theta v}\int \exp\left\{N\theta\frac{\gamma_N(v\gamma_N - \hat{\gamma}_N)}{1+\gamma_N}\right\} e^{-\frac{1}{2}\sum_{i=1}^N g_i^2}\prod_{i=1}^N dg_i, \tag{39}$$

which is almost similar to what we got in (34) except that in the complex plane this is not so easy to “localize” the integral around 0 as we did before.
Our goal is now to show that $\lim_{N\to\infty}\int \exp\left\{N\theta\frac{\gamma_N(v(\theta)\gamma_N - \hat{\gamma}_N)}{1+\gamma_N}\right\} e^{-\frac{1}{2}\sum_{i=1}^N g_i^2}\prod_{i=1}^N dg_i$ exists and is not null.

Denote $\gamma_N = u_1^N + iu_2^N - 1$ and $\hat{\gamma}_N = v_1^N + iv_2^N - v(\theta)$, and let

$$X^N := (u_1^N, u_2^N, v_1^N, v_2^N) - X_0 = \Big(\int \zeta_1(\lambda)x^2 d\hat{\mu}^N(x,\lambda), \cdots, \int \zeta_4(\lambda)x^2 d\hat{\mu}^N(x,\lambda)\Big) - X_0$$

with $d\hat{\mu}^N = \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i, g_i}$,

$$\zeta_1(\lambda) = \Re((1+2v(\theta)\theta - 2\theta\lambda)^{-1}),\ \zeta_2(\lambda) = \Im((1+2v(\theta)\theta - 2\theta\lambda)^{-1}),$$

$$\zeta_3(\lambda) = \Re(\lambda(1+2v(\theta)\theta - 2\theta\lambda)^{-1}),\ \zeta_4(\lambda) = \Im(\lambda(1+2v(\theta)\theta - 2\theta\lambda)^{-1})$$

and $X_0 = (1, 0, \Re(v(\theta)), \Im(v(\theta)))$.

Then, we easily see as in [2] (cf Lemma 4.1 therein) that the law of X^N under $\prod_{i=1}^N \sqrt{2\pi}^{-1} e^{-\frac{1}{2}g_i^2} dg_i$ satisfies a large deviation principle on $\mathbb{R}^4$ with rate function

$$\Lambda^*(X) = \sup_{\substack{Y\in\mathbb{R}^4 \\ 1-2\langle\zeta(\lambda),Y\rangle\geq 0 \mu_E \text{ a.s.}}} \left\{\langle Y, X+X_0\rangle + \frac{1}{2}\int \log\left(1-2\langle\zeta(\lambda),Y\rangle\right) d\mu_E(\lambda)\right\},$$

with $\langle\, ,\, \rangle$ the usual scalar product on $\mathbb{R}^4$.
We denote

$$F(X^N) := \theta\frac{\gamma_N(v\gamma_N - \hat{\gamma}_N)}{1+\gamma_N} = F_1(X^N) + iF_2(X^N)$$

with F_1 and F_2 respectively the real and imaginary part of F. With these notations, our problem boils down to show that $\mathbb{E}[e^{NF(X^N)}]$ converges towards a non-zero limit. Following [1], we know that it is enough for us to check that

(1) there is a vector X^* so that $F(X^*) = 0$ and

$$\lim_{M\to\infty}\lim_{N\to\infty}\left(\frac{1}{N}\log\mathbb{E}[e^{NF_1(X^N)}] - \frac{1}{N}\log\mathbb{E}[1_{|X^N - X^*|\leqslant \frac{M}{\sqrt{N}}} e^{NF_1(X^N)}]\right) = 0.$$

To prove this, the main part of the work will be to show that
a) X^* is the unique minimizer of $\Lambda^* - F_1$ (This indeed entails that the expectation can be localized in a small ball around X^*), and then we will check that
b) X^* is a not degenerate minimizer i.e the Hessian of $\Lambda^* - F_1$ is positive definite at X^* (As shown in [3], this will allow us to take this small ball of radius of order $\sqrt{N}^{-1}$).

(2) X^* is also a critical point of F_2. This second point allows to see that there is no fast oscillations which reduces the first order of the integral.

• **Proof of the first point :** To prove a), let us notice that by our choice of $v(\theta)$ (see Proposition 13), Λ^* is minimum at the origin and that the differential of F_1 at the origin is null. Hence, the origin is a critical point of $F_1 - \Lambda^*$ (where this function is null) and we shall now prove that it is the unique one when $|\theta|$ is small enough.
For that, we adopt the strategy used in [2] and consider the joint deviations of the law of $(X^N, \hat{\mu}^N)$. A slight generalization of Lemma 4.1 therein shows that it satisfies a large

deviations principle on $\mathbb{R}^4 \times \mathcal{P}(\mathbb{R})$ with good rate function

$$J(X,\mu) = I(\mu|\mu_E \otimes P) + \tau\left(X + X_0 - \int \zeta(\lambda) x^2 d\mu(\lambda,x)\right),$$

with $I(.|.)$ the usual relative entropy, P a standard Gaussian measure and

$$\tau(X) = \sup_{\alpha \in \mathcal{D}_0} \{\langle \alpha, X\rangle\},$$

where $\mathcal{D}_0 = \{\alpha \in \mathbb{R}^4 : 1 - 2\langle \alpha, \zeta(\lambda)\rangle \geqslant 0 \quad \mu_E \text{ a.s. }\}$. From that and the contraction principle we have that

$$\begin{aligned} I(X) &:= \Lambda^*(X) - F_1(X) \\ &= \inf_{\mu \in \mathcal{P}(\mathbb{R})} \sup_{\alpha : 1 - 2\langle \alpha, \zeta(\lambda)\rangle \geqslant 0 \, \mu_E \text{ a.s.}} \left\{ I(\mu|\mu_E \otimes P) + \langle X + X_0 - \int \zeta(\lambda) x^2 d\mu(\lambda,x), \alpha\rangle - F_1(X)\right\}. \end{aligned} \tag{40}$$

If we set

$$\mu^\alpha(dx, d\lambda) = \frac{1}{Z_\alpha} e^{-\frac{1}{2}(1-2\langle \zeta(\lambda),\alpha\rangle)x^2} dx d\mu_E(\lambda)$$

then

$$I(\mu|\mu^\alpha) = I(\mu|\mu_E \otimes P) - \langle \alpha, \int \zeta(\lambda) x^2 d\mu(\lambda,x)\rangle - \frac{1}{2}\int \log(1 - 2\langle \zeta(\lambda), \alpha\rangle) d\mu_E(\lambda).$$

Thus,

$$I(X) = \inf_{\mu \in \mathcal{P}(\mathbb{R})} \sup_{\alpha} \left\{ I(\mu|\mu^\alpha) + \langle X + X_0, \alpha\rangle + \frac{1}{2}\int \log(1 - 2\langle \zeta(\lambda), \alpha\rangle) d\mu_E(\lambda) - F_1(X)\right\}.$$

Observe that the supremum in $\Lambda^*(X)$ is achieved at some Y^X since $Y \mapsto -\int \log(1 - 2\langle \zeta(\lambda), Y\rangle) d\mu_E(\lambda)$ is lower semicontinuous and $\{Y \in \mathbb{R}^4 : 1 - 2\langle \zeta(\lambda), Y\rangle \geq 0 \quad \mu_E \text{ a.s. }\}$ is compact when μ_E is not a Dirac mass. Indeed, from the definition of $v(\theta)$, we find that $\mu_E(\zeta_i(\lambda) > 0) > 0$ as well as $\mu_E(\zeta_i(\lambda) < 0) > 0$ for $1 \leq i \leq 4$ from which the compactness follows. Moreover Y^X satisfies

$$(X + X_0)_i = \int \frac{\zeta_i(\lambda)}{1 - 2 < \zeta(\lambda), Y^X >} d\mu_E(\lambda), \quad 1 \leq i \leq 4. \tag{41}$$

Consequently,

$$\Lambda^*(X) - F_1(X) = I(X) \geqslant \inf_{\mu \in \mathcal{P}(\mathbb{R})} \left\{ I(\mu|\mu^{Y^X}) + \Lambda^*(X) - F_1(X)\right\}.$$

Since $I(\mu|\mu^{Y^X}) \geqslant 0$, we deduce that the infimum in μ is taken at $\mu = \mu^{Y^X}$. We also check that $\int \zeta(\lambda) x^2 d\mu^{Y^X}(\lambda, x) = X + X_0$ due to (41). Hence, going back to (40), we find that

$I(X) = \mathcal{I}(\mu^{Y^X})$ with

$$\mathcal{I}(\mu) = I(\mu|\mu_E \otimes P) - F_1\left(\int \zeta(\lambda)x^2 d\mu(x,\lambda) - X_0\right).$$

We next show that $\mathcal{I}$ has a unique minimizer for θ small enough, and this minimizer satisfies $\int \zeta(\lambda)x^2 d\mu(x,\lambda) = X_0$. If the infimum is actually reached at a point μ^* such that F_1 is regular enough at the vicinity of $\int \zeta(\lambda)x^2 d\mu^*(x,\lambda) - X_0$ then this saddle point satisfy the equation

$$d\mu(x,\lambda) = \frac{1}{Z_\mu} e^{DF_1(\int \zeta(\lambda)x^2 d\mu(x,\lambda) - X_0)[\zeta(\lambda)x^2] - \frac{1}{2}x^2} dx d\mu_E(\lambda). \tag{42}$$

Before going on the proof, let us justify that it is indeed the case. Note first that as θ goes to zero, $v(\theta)$ goes to $m = \int \lambda d\mu_E(\lambda)$ and $\Re[(1+2\theta v - 2\theta\lambda)^{-1}]$ is bounded below by say 2^{-1}. Consequently, $\Re\gamma_N + 1 \geqslant 2^{-1}\frac{1}{N}\sum_{i=1}^N g_i^2$. The rate function for the deviations of the latest is $x - \log x - 1$ which goes to infinity as x goes to zero as $\log x^{-1}$. Therefore, for θ small enough,

$$\Lambda^*(X) \geqslant \log(2X_1)^{-1}$$

Since $F_1(X)$ is locally bounded , we deduce that the infimum has to be taken on $X_1 \geqslant \epsilon$ for some fixed $\epsilon > 0$. In particular, F_1 is $\mathcal{C}^\infty$ on this set and equation (42) is well defined.

We now want to use this saddlepoint equation to show uniqueness. Suppose that there are two minimizers μ and ν satisfying (42). Then

$$\Delta := \left|\int \zeta(\lambda)x^2 d\mu(x,\lambda) - \int \zeta(\lambda)x^2 d\nu(x,\lambda)\right| \leqslant 4C|\theta| \sup_i \int |\zeta_i(\lambda)|x^2(d\mu(x,\lambda) + d\nu(x,l)),$$

as we have that $y \to DF_1(y)[x]$ is Lipschitz, with Lipschitz norm of order $C|\theta|\|x\|$. We have now to show that for θ small enough, these covariances are uniformly bounded. This can be done using some arguments very similar to the ones we gave above to justify that the critical points are such that $X_1 \geqslant \epsilon$. We let it to the reader. For θ small enough, we obtain a contraction so that $\Delta = 0$, which entails also $\mu = \nu$. It is easy to check that μ such that $\int \zeta(\lambda)x^2 d\mu(x,\lambda) = X_0$ is always a solution to (42), and hence the unique one when θ is small enough. Observe now that by (42), this minimizer is of the form $\mu^* = \mu^{\alpha^*} = \mu^{Y^{X^*}}$, so that $X^* = \int \zeta(\lambda)x^2 d\mu^{\alpha^*}(x,\lambda) - X_0 = 0$ minimizes indeed I and is actually its unique minimizer.

This concludes the proof of point a), which was the hard part of the work.

As we announced at the beginning and following [1], we now have to show b), that is to say to check that this minimizer is non-degenerate. To see that, remark that the second order derivative of F_1 at the origin is simply

$$D^2F_1[0](U,V) = \Re(\theta(U(vU - V))) \leqslant C|\theta|(|U|^2 + |V|^2) = C|\theta|\left(\sum_{i=1}^4 X_i^2\right) \tag{43}$$

with $U = X_1 + iX_2, V = X_3 + iX_4$.
On the other side, observe that, as $d(\hat{\mu}^N_{E_N}, \mu_E) = o(\sqrt{N}^{-1})$, the covariance matrix of $\sqrt{N}(u_1^N, (\Im(\theta))^{-1}u_2^N, v_1^N, (\Im(\theta))^{-1}v_2^N)$ converges as N goes to infinity towards a 4×4 matrix $K(\theta)$ which is positive definite. Now, remark that $v(\theta) = R_{\mu_E}(2\theta)$ implies that $\Re(\theta)(\Im(\theta))^{-1}\Im(v(\theta))$ converges as $|\theta|$ goes to zero, from which we argue that $K(0)$ is positive definite and bounded. By continuity in θ of $K(\theta)$ we deduce that $K(\theta) \leq CI$ for some $C > 0$ and θ small enough. and the limiting covariances $\sqrt{N}(u_1^N, u_2^N, v_1^N, v_2^N)$(which are also given by the second order derivatives of Λ^*) converges towards a matrix $K'(\theta)$ such that

$$D^2\Lambda^*[0](X,X) = \langle X, K'(\theta)^{-1}X\rangle \geq C^{-1}(X_1^2 + X_3^2 + (\Im(\theta))^{-2}X_2^2 + (\Im(\theta))^{-2}X_4^2)$$

and hence, this together with (43) gives that, for $|\theta|$ small enough, $\frac{1}{2}D^2\Lambda^*[0] - D^2F_1[0] \geqslant 0$.

• **Proof of the second point :** To get Theorem 4, the last step is now to establish the second point, namely to check that 0 is also a critical point for F_2, which is straightforward computation since F behaves in the neighborhood of the origin as a sum of monomials of degree 2 in X.

∎

5 Full asymptotics in the real rank one case

The goal of this section is to establish the convergence and to find an explicit expression for $I_{\mu_E}(\theta) := \lim_{N\to\infty} \frac{1}{N} \log I_N(\theta, E_N)$ as far as E_N satisfies Hypothesis 1 but θ do not necessarily satisfy the hypothesis of Theorem 2. This corresponds to show Theorem 6 (in the case $\beta = 1$ to avoid heavy notations).
To simplify a bit further the notations (and without loss of generality as we can see that if we make a shift of all the λ_i's by a constant C, the limit of the integral again exists and is shifted by C), we suppose that $\lambda_{min} = -\lambda_{max}$.

5.1 Large deviations bounds for $\left(\frac{1}{N}\sum_{i=1}^N g_i^2, \frac{1}{N}\sum_{i=1}^N \lambda_i g_i^2\right)$

We denote by $u_N := \frac{1}{N}\sum_{i=1}^N g_i^2$ and $v_N := \frac{1}{N}\sum_{i=1}^N \lambda_i g_i^2$.
We intend to get the following result

Proposition 16 *The joint law $\hat{\mu}^N$ of (u_N, v_N) under the standard N-dimensional Gaussian measure satisfies the following large deviations principle :*

(1) for all closed set F of $\mathbb{R}^2$,

$$\limsup_{N\to\infty} \frac{1}{N} \log \hat{\mu}^N(F) \leqslant - \inf_{(u,v)\in F} k(u,v),$$

(2) for all open set G of $\mathbb{R}^2$,

$$\liminf_{N\to\infty} \frac{1}{N} \log \hat{\mu}^N(G) \geqslant - \inf_{(u,v)\in G} k(u,v),$$

where the rate function k is given by

$$k(u,v) = \sup_{\mathcal{D}} \left[\xi u + \xi' v + \frac{1}{2} \int \log(1 - 2\xi - 2\xi'\lambda) d\mu_E(\lambda) \right], \tag{44}$$

with $\mathcal{D} := \left\{ (\xi, \xi') \in \mathbb{R}^2; \left| \frac{1-2\xi}{2\xi'} \right| \geqslant \lambda_{max} \right\}$.

The **proof of Proposition 16** can be decomposed into two steps. We first consider the log-Laplace transform of the law of $(u_N,\, v_N)$ under the standard Gaussian measure on $\mathbb{R}^N$ namely, for $\Xi = (\xi, \xi') \in \mathbb{R}^2$,

$$\begin{aligned}
\Lambda_N(\Xi) &:= \frac{1}{N} \log \mathbb{E}\left[e^{N\langle \Xi,\, (u_N,\, v_N)\rangle} \right] \\
&= \begin{cases} -\frac{1}{2N} \sum_{i=1}^N \log(1 - 2\xi - 2\xi'\lambda_i), \text{ if for all } i,\ 1 - 2\xi - 2\xi'\lambda_i > 0 \\ +\infty, \text{ otherwise.} \end{cases}
\end{aligned}$$

Therefore, according to the hypotheses of Theorem 6, we have that Λ_N converges to a function Λ given by

$$\Lambda(\Xi) = \begin{cases} -\frac{1}{2} \int \log(1 - 2\xi - 2\xi'\lambda) d\mu_E(\lambda), & \text{if } \Xi \in \mathcal{D}, \\ +\infty, & \text{otherwise.} \end{cases}$$

It is not hard to see that 0 is in the interior of $\mathcal{D}$ and therefore Gärtner-Ellis theorem (see Theorem 2.3.6 in [8]) gives the upper bound 16.1 and the following lower bound : for all open set G of $\mathbb{R}^2$,

$$\liminf_{N\to\infty} \frac{1}{N} \log \hat{\mu}^N(G) \geqslant - \inf_{(u,v)\in G\cap\mathcal{F}} k(u,v),$$

where $\mathcal{F}$ is the set of all exposed points for k.
Now, to conclude the proof of Proposition 16, we have to show the following lemma :

Lemma 17 *Every point* (u,v) *such that* $k(u,v) < \infty$ *is exposed for* k*, with* k *the rate function introduced in equation* (44)*, that is to say, for all* $(u,v) \in \mathbb{R}^2$*, there exists* $(l_1,l_2) \in \mathbb{R}^2$ *such that for all* $(u',v') \neq (u,v)$*,*

$$l_1 u' + l_2 v' - k(u',v') > l_1 u + l_2 v - k(u,v).$$

The proof of this lemma will be the main object of the next section.

5.2 All points are exposed for k : proof of Lemma 17

We will first establish

Proposition 18 *If we denote by* $\mathcal{K}$ *the function defined on* $\mathbb{R}^2$ *by* $\mathcal{K}(u,z) = k(u,zu)$ *, by* $\mathcal{D}' := \{(u,z) \in \mathbb{R}^2 / u \geqslant 0, |z| \leqslant \lambda_{max}\}$ *and if, for* z *such that* $|z| \leqslant \lambda_{max}$ *we define on* I *the function* h_z *given by* $h_z(\kappa) = \int \log\left(\frac{\kappa-\lambda}{\kappa-z}\right) d\mu_E(\lambda)$*, if we denote by* $h^z_{min} = \lim_{\kappa\uparrow -\lambda_{max}} h_z(\kappa)$ *and* $h^z_{max} = \lim_{\kappa\downarrow \lambda_{max}} h_z(\kappa)$*, then* k *can be expressed as follows :*

$$\mathcal{K}(u,z) = \begin{cases} \frac{1}{2}(u - \log u - 1) + \frac{1}{2}\begin{cases} h_z(K_{\mu_E}(Q_{\mu_E}(z))) & \text{if } z \in I'', \\ h^z_{max} & \text{if } z \in I_1, \\ h^z_{min} & \text{if } z \in I_2, \\ 0 & \text{if } z = m, \end{cases} \\ +\infty \qquad\qquad \text{if } (u,z) \neq \mathcal{D}', \end{cases}$$

where $I_1 := \left]\lambda_{max} - \frac{1}{H_{max}}, \lambda_{max}\right[$ *and* $I_2 := \left]-\lambda_{max}, -\lambda_{max} - \frac{1}{H_{min}}\right[$*, and we recall* [2] *that* K_{μ_E} *is the inverse of the Hilbert transform* H_{μ_E} *of* μ_E *whereas* Q_{μ_E} *is the inverse of its R-transform* R_{μ_E} *and* I'' *is the image of* R_{μ_E}*.*

Proof of Proposition 18:

(1) The first step is to show that

$$\mathcal{K}(u,z) = \begin{cases} \frac{1}{2}(u - \log u - 1) + \frac{1}{2}\sup_{|\kappa|>\lambda_{max}} [h_z(\kappa)] \vee 0, & \text{if } (u,z) \in \mathcal{D}', \\ +\infty, & \text{otherwise.} \end{cases} \tag{45}$$

[2] For further details, see Properties 9 and 11 and Definition 10.

We can easily check on equation (44) that $\mathcal{K}(u,z) = +\infty$ outside $\mathcal{D}'$.
Let now (u,z) be in $\mathcal{D}'$.
We consider the case where $\xi' \neq 0$ and denote by $\kappa := \dfrac{1-2\xi}{2|\xi'|}$ and $\sigma := \mathrm{sgn}(\xi')$.
It is not hard to see that

$$\mathcal{D} = \{(\kappa\,,\, |\xi'|\,,\, \sigma);\ \sigma \in \{-1,+1\},\ |\xi'| > 0,\ \kappa > \lambda_{max}\}\,.$$

With these notations,

$$k(u,v) = \sup_{\mathcal{D}} \left[\sigma|\xi'|v + \left(\frac{1}{2} - \kappa|\xi'|\right)u + \frac{1}{2}\log(2|\xi'|) + \frac{1}{2}\int \log(\kappa - \sigma\lambda) d\mu_E(\lambda)\right].$$

We are seeking for $|\xi'|$ that maximises the rate function. We consider the function $g : x \mapsto (\sigma v - \kappa u)x + \frac{1}{2}\log(2x)$ on $\mathbb{R}_+^*$.
As we can easily check, if $u \geqslant 0$ and $\left|\frac{v}{u}\right| \leqslant \lambda_{max}$, on $\mathcal{D}$ we have that $(\sigma v - \kappa u) < 0$ so that g has a unique maximum, reached at $x = \dfrac{1}{2(\kappa u - \sigma v)}$, which is equal to $\dfrac{1}{2}\left(\log(\kappa u - \sigma v) - 1\right)$. Therefore

$$\mathcal{K}(u,z) = \frac{1}{2}(u - \log u - 1) + \frac{1}{2}\sup_{\sigma=\pm 1, \kappa > \lambda_{max}} \left(\int \log\left(\frac{\kappa - \sigma\lambda}{\kappa - \sigma z}\right) d\mu_E(\lambda)\right).$$

Furthermore, for any z in $[-\lambda_{max}, \lambda_{max}]$, we have that

$$\begin{aligned}\sup_{\kappa > \lambda_{max}} \left(\int \log\left(\frac{\kappa + \lambda}{k + z}\right) d\mu_E(\lambda)\right) &= \sup_{-\kappa > \lambda_{max}} \left(\int \log\left(\frac{-\kappa - \lambda}{-\kappa - z}\right) d\mu_E(\lambda)\right) \\ &= \sup_{\kappa < -\lambda_{max}} \left(\int \log\left(\frac{\kappa - \lambda}{k - z}\right) d\mu_E(\lambda)\right),\end{aligned}$$

what concludes the proof of formula (45) in the case where $\xi' \neq 0$.
Otherwise, for $\xi' = 0$,

$$\mathcal{K}(u,z) = \sup_{\xi \leqslant \frac{1}{2}} \left[\xi u + \frac{1}{2}\log(1 - 2\xi)\right] = \frac{1}{2}(u - \log u - 1),$$

so that (45) is proved.

(2) We now turn to the study of the supremum of h_z on $\{\kappa \in \mathbb{R}/|\kappa| > \lambda_{max}\}$.

The first remark is that, for any value of $z \in [-\lambda_{max}, \lambda_{max}]$, the function h_z is analytic on $[-\lambda_{max}, \lambda_{max}]^c$, in particular differentiable and its derivative is given, for any $y \in [-\lambda_{max}, \lambda_{max}]^c$, by

$$h_z'(y) = H_{\mu_E}(y) - \frac{1}{y - z}. \tag{46}$$

We will study the supremum of h_z for the different values of z.
We want to show that if $z \in I''$, the supremum of h_z is reached at $\kappa_0 = K_{\mu_E}(Q_{\mu_E}(z))$.
• The first point is to show that in the case $z \in I''$, there is a unique κ_0 where h'_z cancels. Indeed :

$$h'(\kappa_0) = 0 \Longleftrightarrow H_{\mu_E}(\kappa_0) = \frac{1}{\kappa_0 - z}.$$

This implies that $\frac{1}{\kappa_0 - z} \in I'$ so that, by definition of K_{μ_E},

$$\kappa_0 = K_{\mu_E}\left(\frac{1}{\kappa_0 - z}\right), \tag{47}$$

which gives that $R_{\mu_E}\left(\frac{1}{\kappa_0 - z}\right) = z$.
We know (by point 4. of Property 11) that R_{μ_E} is bijective on I', so that, if $z \in I''$, there exists a unique $\alpha \in I'$ such that $z = R_{\mu_E}(\alpha)$.
But the function $x \mapsto \frac{1}{x - z}$ is also bijective from $\mathbb{R} \setminus \{z\}$ to $\mathbb{R}^*$ so, as I' does not contain 0, there exists a unique $\kappa_0 \neq z$ such that $\alpha = \frac{1}{\kappa_0 - z}$ and by (47), κ_0 is in I and can be expressed as $\kappa_0 = K_{\mu_E}(Q_{\mu_E}(z))$.

• We now want to show that the maximum of h_z is actually reached at κ_0. We claim the following fact :

Lemma 19 *We recall that $h^z_{min} = \lim_{\kappa \uparrow -\lambda_{max}} h_z(\kappa)$ and $h^z_{max} = \lim_{\kappa \downarrow \lambda_{max}} h_z(\kappa)$.
Then h^z has one of the following behaviour on I :*

- *if $\kappa_0 > \lambda_{max}$, h_z is decreasing from 0 to h^z_{min} on $]-\infty, -\lambda_{max}[$, it is increasing from h^z_{max} to $h_z(\kappa_0)$ on $]\lambda_{max}, \kappa_0]$ and then decreasing from $h_z(\kappa_0)$ to 0 on $]\kappa_0, +\infty]$,*
- *if $\kappa_0 < -\lambda_{max}$, h_z is increasing from 0 to $h_z(\kappa_0)$ on $]-\infty, \kappa_0]$ then decreasing from $h_z(\kappa_0)$ to h^z_{min} on $]\kappa_0, -\lambda_{max}[$, it is increasing from h^z_{max} to 0 on $]\lambda_{max}, +\infty[$.*

Proof of lemma 19:
We treat in details the case $\kappa_0 > \lambda_{max}$, the other one being very similar. We recall from Property 11 that I'' is the image of R_{μ_E}.

If $\kappa_0 > \lambda_{max}$, h'_z does not cancel on $]-\infty, -\lambda_{max}[$. We have to determine its sign :

- if $H_{min} = -\infty$, as $z \in I''$, $\lim_{\kappa \to -\lambda_{max}} \frac{1}{\kappa - z} > -\infty$ so that $\lim_{\kappa \to -\lambda_{max}} h'_z(\kappa) = -\infty < 0$,
- similarly if $H_{min} \in \mathbb{R}^-_*$, as $z \in I''$, by property 11,

$$\begin{aligned} z > -\lambda_{max} - \frac{1}{H_{min}} &\Longrightarrow H_{min} - \frac{1}{-\lambda_{max} - z} < 0 \\ &\Longrightarrow \lim_{\kappa \to -\lambda_{max}} h'_z(\kappa) < 0, \end{aligned}$$

so that in both case h'_z is negative on the whole interval $]-\infty, -\lambda_{max}[$. Furthermore, it is not hard to see that $\lim_{\kappa \to -\infty} h_z(\kappa) = 0$ so that h_z is decreasing from 0 to h^z_{min}.

On the other side, we want to find the sign of h'_z on $]\lambda_{max}, +\infty[$ knowing that it cancels at κ_0.
As above, we show that $\lim_{\kappa \to \lambda_{max}} h'_z(\kappa) > 0$ and we deduce from that and the continuity of h'_z, that it is positive till κ_0. Furthermore, h_z is also twice differentiable at κ_0 and

$$\begin{aligned} h''_z(\kappa_0) &= -\int \frac{1}{(\kappa_0 - \lambda)^2} d\mu_E(\lambda) + \Big(\frac{1}{\kappa_0 - z}\Big)^2 \\ &< -\Big(\int \frac{1}{\kappa_0 - \lambda} d\mu_E(\lambda)\Big)^2 + (H_{\mu_E}(\kappa_0))^2 < 0, \end{aligned}$$

where we used Cauchy-Schwarz inequality and the definition of κ_0. Therefore h'_z is negative for $\kappa > \kappa_0$ and the fact that $\lim_{\kappa \to +\infty} h_z(\kappa) = 0$ concludes the proof of Lemma 19. ∎

We now go back to the second step of the **proof of Proposition 18.** From Lemma 19, we can see that in both case, the maximum of h_z is taken at κ_0 and that this maximum is positive, so that

$$\mathcal{K}(u, z) = \frac{1}{2}(u - \log u - 1) + \frac{1}{2} h_z(K_{\mu_E}(Q_{\mu_E}(z))), \text{ if } z \in I''. \tag{48}$$

(3) The next point is look at h_z when z is in $I_1 = \left]\lambda_{max} - \frac{1}{H_{max}}, \lambda_{max}\right[$.
In this case there is no solution to $h'_z(\kappa_0) = 0$ so that the point is to determine the sign of h'_z, on the intervals $]-\infty, -\lambda_{max}[$ and $]\lambda_{max}, +\infty[$.

Note that I_1 is non-empty if and only if $H_{max} < +\infty$. If so, as $z > \lambda_{max} - \dfrac{1}{H_{max}}$, we have that $\lim_{\kappa \to \lambda_{max}} h'_z(\kappa) < 0$ and h_z is decreasing from h^z_{max} to 0 on $]\lambda_{max}, +\infty[$. We also have that h'_z is negative on $]-\infty, -\lambda_{max}[$. Moreover,

$$\begin{aligned} \frac{1}{y - z} &> \frac{1}{y - \lambda_{max} + \frac{1}{H_{max}}} = \frac{1}{y - \lambda_{max} + \frac{1}{\int \frac{1}{\lambda_{max} - \lambda} d\mu_E(\lambda)}} \\ &\geqslant \frac{1}{y - \lambda_{max} + \int (\lambda_{max} - \lambda) d\mu_E(\lambda)} = \frac{1}{\int (y - \lambda) d\mu_E(\lambda)} \geqslant H_{\mu_E}(y), \end{aligned}$$

where we used Jensen's inequality, first with the concavity of $x \mapsto \frac{1}{x}$ on $\mathbb{R}_+^*$ and then with the convexity of $x \mapsto \frac{1}{x}$ on $\mathbb{R}_-^*$.

This implies that h'_z is negative on $]-\infty, -\lambda_{max}[$ so that the maximum of h_z is h^z_{max} and it is again positive, which concludes this second case.

(4) The case $z \in I_2 = \left]-\lambda_{max}, -\lambda_{max} - \frac{1}{H_{min}}\right[$ is very similar to the case of I_1 and we let it to the reader.

(5) The last point to check is when $z = m$.
In this case,

$$h'_z(y) = \int \frac{1}{(y-\lambda)} d\mu_E(\lambda) - \frac{1}{\int (y-\lambda) d\mu_E(\lambda)}.$$

By Jensen's inequality again, this is positive if $y > \lambda_{max}$ and negative for $y < -\lambda_{max}$ so that the supremum of h_m is 0.

This concludes the proof of Proposition 18 and we are now ready to complete the **proof of Lemma 17.**

As $k(u, v) < \infty$ implies that u is positive, we can make the change of variable $z = u/v$ and it is enough to check that $\mathcal{K}$ is strictly convex.

- On $\mathbb{R}_*^+ \times I''$, $\mathcal{K}$ is analytic and as

$$\frac{\partial \mathcal{K}}{\partial z}(u, z) = \frac{1}{2} Q_{\mu_E}(z), \quad \text{so that} \quad \frac{\partial^2 \mathcal{K}}{\partial z^2}(u, z) = \frac{1}{2} \frac{1}{R'_{\mu_E}(Q_{\mu_E}(z))} > 0, \tag{49}$$

  the Hessian (which is diagonal) is positive so that it is strictly convex and therefore each point of this domain is exposed.

- Suppose now that $z \in I_1$ or I_2. We treat in details the case of I_1, that of I_2 being very similar.
  We recall that on $\mathbb{R}_*^+ \times I_1$,

$$\mathcal{K}(u, z) = \frac{1}{2}(u - \log u - 1) - \frac{1}{2} \log(\lambda_{max} - z) + C,$$

  with C a constant (which is the limit of $\int \log(\kappa - \lambda) d\mu_E(\lambda)$ as κ decreases to λ_{max}) so that it is obviously strictly convex.

- The last point is to check that $\lambda_{max} - \frac{1}{H_{max}}$ (and similarly $-\lambda_{max} - \frac{1}{H_{min}}$) are also exposed.
  From (49), we can check that

$$\lim_{z \downarrow \lambda_{max} - \frac{1}{H_{max}}} \frac{\partial \mathcal{K}}{\partial z}(u, z) = \lim_{z \uparrow \lambda_{max} - \frac{1}{H_{max}}} \frac{\partial \mathcal{K}}{\partial z}(u, z) = H_{max}.$$

Therefore, Lemma 17 gives us Proposition 16, that is a full large deviations principle for the joint law of $\left(\frac{1}{N} \sum g_i^2, \frac{1}{N} \sum \lambda_i g_i^2\right)$ and allow us to perform a Laplace method.

Lemma 20 *For any $\theta \in \mathbb{R}$, if $\mathcal{K}$ is the function introduced in Proposition 18, we have*

$$I_{\mu_E}(\theta) = \sup_{u,z}(\theta z - \mathcal{K}(u,z)).$$

Proof of Lemma 20: We have that u_N is almost surely positive and that $\left|\frac{v_N}{u_N}\right|$ is almost surely less than λ_{max} so that we can rewrite

$$I_N(\theta, E_N) = \mathbb{E}\left[e^{N\theta f\left(\frac{v_N}{u_N}\right)}\right],$$

with

$$\begin{aligned} f:\ \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (u,v) &\longmapsto \begin{cases} \frac{v}{u}, & \text{if } u>0 \text{ and } \left|\frac{v}{u}\right| \leqslant \lambda_{max} \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

We first notice that f is lower semi-continuous so that the lower bound in Varadhan's lemma (cf lemma 4.3.4 in [8]) gives that

$$\liminf_{N\to\infty} \frac{1}{N} \log I_N(\theta, E_N) \geqslant \sup\left[\theta\frac{v}{u} - k(u,v)\right]. \tag{50}$$

On the other side, f is bounded above but not upper semi-continuous so to get the limsup, we rewrite, for any $\epsilon > 0$,

$$I_N(\theta, E_N) = \mathbb{E}\left[e^{N\theta f(u_N,v_N)}\mathbf{1}_{u_N>\epsilon}\right] + \mathbb{E}\left[e^{N\theta f(u_N,v_N)}\mathbf{1}_{u_N\leqslant\epsilon}\right],$$

so that

$$\begin{aligned} \limsup_{N\to\infty} \frac{1}{N}\log I_N(\theta, E_N) = \max\Bigg\{ &\limsup_{N\to\infty}\frac{1}{N}\log \mathbb{E}\left[e^{N\theta f(u_N,v_N)}\mathbf{1}_{u_N>\epsilon}\right], \\ &\limsup_{N\to\infty}\frac{1}{N}\log \mathbb{E}\left[e^{N\theta f(u_N,v_N)}\mathbf{1}_{u_N\leqslant\epsilon}\right]\Bigg\}. \end{aligned}$$

But, as f is bounded between $-\lambda_{max}$ and λ_{max},

$$\begin{aligned} \limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{N\theta f(u_N,v_N)}\mathbf{1}_{u_N\leqslant\epsilon}\right] &\leqslant |\theta\lambda_{max}| + \limsup_{N\to\infty}\frac{1}{N}\log P(u_N\leqslant\epsilon) \\ &\leqslant |\theta\lambda_{max}| - \frac{1}{2}\left(1 - \log\frac{1}{\epsilon}\right) \underset{\epsilon\downarrow 0}{\longrightarrow} -\infty \end{aligned}$$

so that, for ϵ small enough,

$$\limsup_{N\to\infty}\frac{1}{N}\log I_N(\theta,E_N)=\limsup_{N\to\infty}\frac{1}{N}\log\mathbb{E}\left[e^{N\theta f(u_N,v_N)}\mathbf{1}_{u_N>\epsilon}\right].$$

Now f is continuous on $\mathcal{D}'\cap\{u>\epsilon\}$ and by Varadhan's lemma again (cf lemma 4.3.5 in [8]), we get

$$\limsup_{N\to\infty}\frac{1}{N}\log I_N(\theta,E_N)\leqslant\sup_{\epsilon>0}\sup_{\mathcal{D}'\cap\{u>\epsilon\}}\left[\theta\frac{v}{u}-k(u,v)\right]\leqslant\sup_{\mathcal{D}'}\left[\theta\frac{v}{u}-k(u,v)\right].$$

This together with (50) gives that

$$I_{\mu_E}(\theta)=\sup\left[\theta\frac{v}{u}-k(u,v)\right]=\sup\left[\theta z-\mathcal{K}(u,z)\right],$$

as was announced in Lemma 20. ∎

5.4 *End of the proof : getting an explicit expression for I_{μ_E}*

We now go back to the **proof of Theorem 6** and, in view of Lemma 20 and Proposition 18, we denote by

$$G(\theta):=\sup_{z\in I''}\left[\theta z-\frac{1}{2}h_z(K_{\mu_E}(Q_{\mu_E}(z)))\right],$$

$$G_1(\theta):=\sup_{z\in I_1}\left[\theta z-\frac{1}{2}h^z_{max}\right],\quad G_2(\theta):=\sup_{z\in I_2}\left[\theta z-\frac{1}{2}h^z_{min}\right].$$

We first remark that $\inf_{u\geqslant 0}(u-\log u-1)=0$.
The most part of the work for this last step will rely on proving

Lemma 21 *With the notations introduced above, we have*[3]

$$G(\theta)=\begin{cases}\frac{1}{2}\int_0^{2\theta}R_{\mu_E}(u)du, & \text{if } 2\theta\in I'\\ \theta\left(-\lambda_{max}-\frac{1}{H_{min}}\right)-\frac{1}{2}\int\log(H_{min}(-\lambda_{max}-\lambda))d\mu_E(\lambda)^\sharp, & \text{if } 2\theta<H_{min}\\ \theta\left(\lambda_{max}-\frac{1}{H_{max}}\right)-\frac{1}{2}\int\log(H_{max}(\lambda_{max}-\lambda))d\mu_E(\lambda)^*, & \text{if } 2\theta>H_{max},\end{cases}$$

[3] $\sharp=-\infty$ if $H_{min}=-\infty$ and otherwise these expressions are well defined in virtue of the fact that $\int_0^1\frac{1}{\lambda}d\mu(\lambda)<+\infty\Rightarrow-\int_0^1\log\lambda d\mu(\lambda)<+\infty$,
$*=-\infty$ if $H_{max}=+\infty$ and otherwise these expressions are well defined for the same reason.

$$G_1(\theta) = \begin{cases} \theta\left(\lambda_{max} - \frac{1}{2\theta}\right) - \frac{1}{2}\int \log(2\theta(\lambda_{max} - \lambda))d\mu_E(\lambda)^*, & \text{if } 2\theta > H_{max} \\ \theta\left(\lambda_{max} - \frac{1}{H_{max}}\right) - \frac{1}{2}\int \log(H_{max}(\lambda_{max} - \lambda))d\mu_E(\lambda)^*, & \text{if } 2\theta < H_{max}, \end{cases}$$

$$G_2(\theta) = \begin{cases} \theta\left(-\lambda_{max} - \frac{1}{2\theta}\right) - \frac{1}{2}\int \log(2\theta(-\lambda_{max} - \lambda))d\mu_E(\lambda)^\sharp, & \text{if } 2\theta < H_{min} \\ \theta\left(-\lambda_{max} - \frac{1}{H_{min}}\right) - \frac{1}{2}\int \log(H_{min}(-\lambda_{max} - \lambda))d\mu_E(\lambda)^\sharp, & \text{if } 2\theta > H_{min}. \end{cases}$$

Proof of Lemma 21 :

- We first study G. This is finding the supremum of $j_\theta(z) := \theta z - \frac{1}{2}h_z(K_{\mu_E}(Q_{\mu_E}(z)))$ on I'. From Definition 10 and Property 11, we have that j_θ is differentiable on I' and an easy computation gives

$$j'_\theta(z) = \frac{1}{2}(2\theta - Q_{\mu_E}(z)).$$

- If $2\theta \in I'$, the unique z_0 such that $j'_\theta(z_0) = 0$ is given by $z_0 = R_{\mu_E}(2\theta)$.
  From point 2 of Property 11, we know that R_{μ_E} is increasing on I' and so is its inverse Q_{μ_E} on I''.
  Therefore j'_θ is decreasing and j_θ is increasing from $\lim_{z\to-\lambda_{max}} j_\theta(z)$ to $j_\theta(z_0)$ and then decreasing so that its maximum is reached at z_0.
  This gives that if $2\theta \in]H_{min}, H_{max}[\backslash\{0\}$,

$$G(\theta) = \frac{1}{2}\left(2\theta R_{\mu_E}(2\theta) - \log(2\theta) - \int \log(K_{\mu_E}(2\theta) - \lambda)d\mu_E(\lambda)\right).$$

  As $|K_{\mu_E}(2\theta)| > \lambda_{max}$, G is analytic on I' and the calculation of its derivative gives $G'(\theta) = R_{\mu_E}(2\theta)$.
  As $\lim_{\theta\to 0} G(\theta) = 0$, we have

$$G(\theta) = \int_0^\theta R_{\mu_E}(2u)du = \frac{1}{2}\int_0^{2\theta} R_{\mu_E}(u)du.$$

- If $H_{min} > -\infty$ and $2\theta < H_{min}$, the equation $j'_\theta(z_0) = 0$ has no solution, we want to determine the sign of j'_θ on I'.
  For all $z \in I''$, $Q_{\mu_E}(z) \in I'$ so that $Q_{\mu_E}(z) > H_{min}$ and therefore $j'_\theta(z) < 0$.
  j_θ is decreasing on I'' and $\lim_{z\uparrow m^-} j_\theta(z) = \lim_{z\downarrow m^+} j_\theta(z)$ so that the supremum is reached at the left boundary $-\lambda_{max} - \dfrac{1}{H_{min}}$ of I'' and is equal to

$$\theta\left(-\lambda_{max} - \frac{1}{H_{min}}\right) - \frac{1}{2}\int \log(H_{min}(-\lambda_{max} - \lambda))d\mu_E(\lambda).$$

- If $H_{max} < +\infty$, a similar treatment in the case $2\theta > H_{max}$ concludes the proof of G.

- We now consider G_1. We recall that, if $H_{max} < +\infty$,

$$G_1(\theta) = \sup_{z \in I_1} \left\{ \theta z - \frac{1}{2} \int \log \left(\frac{\lambda_{max} - \lambda}{\lambda_{max} - z} \right) d\mu_E(\lambda) \right\}.$$

We denote by $l(z) := \theta z - \frac{1}{2} \int \log \left(\frac{\lambda_{max} - \lambda}{\lambda_{max} - z} \right) d\mu_E(\lambda)$ is analytic and derivable and $l'(z) = \theta - \frac{1}{2} \frac{1}{\lambda_{max} - z}$.

- If $2\theta > H_{max}$ then there exists a unique $z_1 = \lambda_{max} - \frac{1}{2\theta}$, such that $l'(z_1) = 0$ and it is easy to see that the maximum of l is reached at z_1 then

$$G_1(\theta) = \theta \left(\lambda_{max} - \frac{1}{2\theta} \right) - \frac{1}{2} \int \log(2\theta(\lambda_{max} - \lambda)) d\mu_E(\lambda).$$

- If $2\theta < H_{max}$, l' is negative on I_1 and the left boundary $\lambda_{max} - \frac{1}{H_{max}}$ on I_1 and

$$G_1(\theta) = \theta \left(\lambda_{max} - \frac{1}{H_{max}} \right) - \frac{1}{2} \int \log(H_{max}(\lambda_{max} - \lambda)) d\mu_E(\lambda).$$

  This concludes the proof of G_1.

- The case G_2 is very similar and this concludes the proof of Lemma 21. ∎

By virtue of Lemmata 20 and 21 and Proposition 18, to finish the proof of Theorem 6, we have now

(1) to compare $G_{|I'}$, $G_{1|I'}$ and $G_{2|I'}$ to get $I_{\mu_E|I'}$.
Since $\lim_{z \uparrow H_{max}} j_\theta(z) = G_1(\theta)$ and $\lim_{z \downarrow H_{min}} j_\theta(z) = G_2(\theta)$ whereas $G(\theta) = \sup_{z \in I'} [j_\theta(z)]$, we get that $I_{\mu_E|I'} = G_{|I'}$.

(2) if $H_{max} < +\infty$, to compare $G_{|\{2\theta > H_{max}\}}$, $G_{1|\{2\theta > H_{max}\}}$ and $G_{2|\{2\theta > H_{min}\}}$ to get $I_{\mu_E||\{2\theta > H_{max}\}}$.
By studying the function $x \mapsto -\frac{\theta}{x} - \frac{1}{2} \log x$, which reaches its maximum at θ, we can easily deduce that $G_{|\{2\theta > H_{max}\}} < G_{1|\{2\theta > H_{max}\}}$.
Moreover $G_{1|\{2\theta > H_{max}\}}$ and $G_{2|\{2\theta > H_{max}\}}$ are the limits of j_θ respectively at $\lambda_{max} - \frac{1}{H_{max}}$ and $-\lambda_{max} - \frac{1}{H_{min}}$ and we know that in the case $2\theta > H_{max}$, j_θ is increasing. This gives $G_{2|\{2\theta > H_{max}\}} < G_{1|\{2\theta > H_{max}\}}$.
In this case we conclude that the maximum is given by $G_{1|\{2\theta > H_{max}\}}$.

(3) Arguing similarly, we can see that in the case where $2\theta < H_{min}$ the maximum is given by $G_{2|\{2\theta < H_{min}\}}$.

To conclude the proof of Theorem 6, we use the continuity of I_{μ_E} with respect to θ given by the first point of Lemma 14 to specify its value at $-\lambda_{max}$, $-\lambda_{max} - \frac{1}{H_{min}}$, $\lambda_{max} - \frac{1}{H_{max}}$ and λ_{max}. ∎

6 Asymptotic independence and free convolution

In this section, we want to prove Theorem 5, that is to say concentration and decorrelation properties for the spherical integrals.

We recall first that as an immediate Corollary of Theorem 5, we get that

Corollary 22 *For θ sufficiently small*

$$R_{\mu_B \boxplus \mu_A}(\theta) = R_{\mu_A}(\theta) + R_{\mu_B}(\theta),$$

where $\boxplus$ denotes the free convolution of measures.

Proof. In fact, being given μ_A, μ_B, we take $\lambda_1(A)$ (resp. $\lambda_1(B)$) to be the lower edge of the support of μ_A (resp. μ_B) and then set for $i \geq 2$

$$\lambda_i(A) = \inf\left\{x \geq \lambda_{i-1}(A) : \mu_A([\lambda_1(A), x]) \geq \frac{i}{N}\right\}, \lambda_i(B) = \inf\left\{x \geq \lambda_{i-1}(A) : \mu_B([\lambda_1(B), x]) \geq \frac{i}{N}\right\}.$$

It is easily seen that with this choice, $A_N = \operatorname{diag}(\lambda_i(A))$ and $B_N = \operatorname{diag}(\lambda_i(B))$ satisfy Hypothesis 1. Since μ_A and μ_B are compactly supported, A_N and B_N have uniformly bounded spectral radius and so does $A_N + UB_NU^*$. Hence, for θ small enough, A_N, B_N and $A_N + UB_NU^*$ satisfy the hypotheses of Theorem 2 (recall that A_N and UB_NU^* are asymptotically free (c.f Theorem 5.2 in [6]) so that $\hat{\mu}^N_{A_N+UB_NU^*}$ converges towards $\mu_B \boxplus \mu_A$). Moreover, we can check that $d(\hat{\mu}^N_{A_N}, \mu_A) \leqslant 2\|A_N\|_\infty N^{-1}$ and similarly for μ_B so that $d(\hat{\mu}^N_{A_N}, \mu_A) + d(\hat{\mu}^N_{B_N}, \mu_B) = o(\sqrt{N}^{-1})$.
Thus, combining Theorem 5.2 and Theorem 2 imply

$$\int_0^{2\theta} R_{\mu_B \boxplus \mu_A}(v)dv = \int_0^{2\theta} R_{\mu_A}(v)dv + \int_0^{2\theta} R_{\mu_B}(v)dv.$$

Differentiating with respect to θ gives Corollary 22.

∎

Since the R-transform is analytic in a neighbourhood of the origin, this entails the famous additivity property of the R-transform. So, Theorem 5 provides a new proof of this property, independent of cumulant techniques.

As announced in the introduction, the first step will be to use a result of concentration for orthogonal matrices.

6.1 *Concentration of measure for orthogonal matrices*

In this section, we prove the first point of Theorem 5 that relies on the following lemma, which is a direct consequence of a theorem due to Gromov and Milman [9]

Lemma 23 *[Gromov-Milman, [9], p. 844 and 846] Let $M_N^{(1)}$ denote the Haar measure on the special orthogonal group $SO(N)$. There exists a positive constant $c > 0$, independent of N, such that for any function $F : SO(N) \to \mathbb{R}$ so that there is a real $||F||_{\mathcal{L}}$ such that, for any $U, U' \in SO(N)$*

$$|F(U) - F(U')| \leqslant ||F||_{\mathcal{L}} \left(\sum_{i,j=1}^{N} |u_{ij} - u'_{ij}|^2 \right)^{\frac{1}{2}},$$

then, for any $\epsilon > 0$,

$$M_N^{(1)} \left(\left| F(U) - \int F(U) dM_N^{(1)}(U) \right| \geq \epsilon \right) \leq e^{-cN||F||_{\mathcal{L}}^{-2} \epsilon^2}.$$

Proof of lemma 23 :
In [9], the authors prove such a lemma using the fact that the Ricci curvature of $SO(N)$ is of order [4] N , and their result holds when F is Lipschitz with respect to the standard bivariant metric which measures the length of the geodesic in $SO(N)$ between two elements $U, U' \in SO(N)$. This distance is of course greater than the length of the geodesic in the whole space of matrices, given by the Euclidean distance, so that Lemma 23 is a direct consequence of [9]. ∎

To get 5.1, we now apply our result with F given by $F(U_N) = \frac{1}{N} \log I_N(\theta, A_N + U_N B U_N^*)$. To get (8), we have to check that this F satisfies the hypotheses of Lemma 23. i.e. that F is Lipschitz.

We have, for any matrices W, $\tilde{W}$ in $M_N := \{W \in \mathcal{M}_N(\mathbb{C}); WW^* \leqslant 1\}$,

[4] In [9], it is reported that the Ricci curvature is given by $N/4$ whereas J.C Sikorav and Y. Ollivier reported to us that it is in fact $(N-2)/2$.

$$\left|\frac{1}{N}\log I_N(\theta, A_N + WB_NW^*) - \frac{1}{N}\log I_N(\theta, A_N + \tilde{W}B_N\tilde{W}^*)\right| \leqslant 2\theta||B||_\infty \sup_{||v||=1} \langle v, |W-\tilde{W}|v\rangle$$
$$\leqslant 2\theta||B||_\infty \left(\sum_{i,j=1}^{N} |w_{ij}-\tilde{w}_{ij}|^2\right)^{\frac{1}{2}}.$$

Moreover, if T is for example the transformation changing the first column vector U_1 of the matrix U into $-U_1$, $O(N) = SO(N) \sqcup T(SO(N))$. Note that

$$F(TU) = \frac{1}{N}\log I_N(\theta, A_N + (TU_N)B_N(TU_N)^*) = \frac{1}{N}\log I_N(\theta, T^*A_NT + U_NB_N(U_N)^*).$$

Now, if we set $E_N = A_N + U_NBU_N^*$ and $E'_N = T^*A_NT + U_NBU_N^*$, we easily see that

$$d(\hat{\mu}^N_{E_N}, \hat{\mu}^N_{E'_N}) \leq \frac{1}{N}\mathrm{tr}|E'_N - E_N| \leq \frac{2||A||_\infty}{N}.$$

Hence, Lemma 14.3 implies that

$$\delta_N = \sup_{U\in SO(N)} |F(U) - F(TU)| \to 0 \text{ as } N \to \infty$$

Since

$$\int_{O(N)} F(U)dm_N^{(1)}(U) = \frac{1}{2}\int_{SO(N)} F(U)dM_N^{(1)}(U) + \frac{1}{2}\int_{SO(N)} F(TU)dM_N^{(1)}(U),$$

we deduce that

$$\left|\int_{O(N)} F(U)dm_N^1(U) - \int_{SO(N)} F(U)dM_N^1(U)\right| \leq \delta_N.$$

Thus, Lemma 23 implies that for $\epsilon > 0$

$$M_N^{(1)}\left(\left|F(U) - \int_{O(N)} F(U)dm_N^{(1)}(U)\right| \geq \epsilon + \delta_N\right) \leq e^{-cN||F||_{\mathcal{L}}^{-2}\epsilon^2} \tag{51}$$

and similarly for $F(TU)$ so that

$$m_N^{(1)}\left(\left|F(U) - \int_{O(N)} F(U)dm_N^{(1)}(U)\right| \geq \epsilon + \delta_N\right) \leq e^{-cN||F||_{\mathcal{L}}^{-2}\epsilon^2},$$

what gives Theorem 5.1. ∎

6.2 Exchanging integration with the logarithm

We are now seeking to establish the second point of Theorem 5. By Jensen's inequality,

$$\mathbb{E}[\log I_N(\theta, A_N + V_N B_N (V_N)^*)] \leqslant \log \mathbb{E}[I_N(\theta, A_N + V_N B_N (V_N)^*)]$$

so that we only need here to prove the converse inequality.

The whole idea to get it is contained in the following

Lemma 24 *For any uniformly bounded sequence of matrices* $(A_N, B_N)_{N\in\mathbb{N}}$ *and* θ *small enough, there exists a finite constant* $C(A,B,\theta)$ *such that for* N *large enough*

$$\frac{\mathbb{E}[I_N(\theta, A_N + V_N B_N (V_N)^*)^2]}{\mathbb{E}[I_N(\theta, A_N + V_N B_N (V_N)^*)]^2} \leqslant C(\theta, A, B)$$

Let us conclude the **proof of Theorem 5.2** before proving this lemma.
Hereafter, $\epsilon > 0$ is fixed. We introduce the event

$$\mathcal{A} = \left\{ I_N(\theta, A_N + V_N B_N (V_N)^*) \geqslant \frac{1}{2}\mathbb{E}[I_N(\theta, A_N + V_N B_N (V_N)^*)] \right\}$$

Following [22], we have, if $I_N := I_N(\theta, A_N + V_N B_N (V_N)^*)$ that

$$\mathbb{E}[I_N] = \mathbb{E}[I_N \mathbf{1}_{\mathcal{A}^c}] + \mathbb{E}[I_N \mathbf{1}_{\mathcal{A}}] \leqslant \frac{1}{2}\mathbb{E}[I_N] + \mathbb{E}[I_N^2]^{\frac{1}{2}} \mathbb{P}(\mathcal{A})^{\frac{1}{2}}$$

so that

$$\frac{1}{4C(A,B,\theta)} \leqslant \mathbb{P}(\mathcal{A}).$$

Furthermore, let

$$t = \frac{1}{N}\log \mathbb{E}\left[\frac{1}{2} I_N(\theta, A_N + V_N B_N (V_N)^*)\right] - \frac{1}{N}\mathbb{E}[\log I_N(\theta, A_N + V_N B_N (V_N)^*)]$$

We can assume that $t \geqslant \delta_N$ (δ_N being given in (51)) since otherwise we are done. We then get by (51) that for any $t \geq \delta_N$ and N large enough,

$$\mathbb{P}(\mathcal{A}) \leqslant \mathbb{P}\Big(\frac{1}{N}\log I_N(\theta, A_N + U B_N U^*) - m_N^{(1)}\Big(\frac{1}{N}\log I_N(\theta, A_N + U B_N U^*)\Big) \geqslant t \Big) \leqslant e^{-cN(t-\delta_N)^2}$$

with $c' = c(2|\theta|||B||_\infty)^{-2}$. As a consequence,

$$\frac{1}{4C(A,B,\theta)} \leqslant e^{-c'N(t-\delta_N)^2}, \quad \text{so that} \quad t \leqslant \delta_N + \Big(\frac{1}{c'N}\log(4C(A,B,\theta))\Big)^{\frac{1}{2}}.$$

Hence, since δ_N goes to zero with N,

$$\lim_{N\to\infty}\left(\frac{1}{N}\log\mathbb{E}\left[\frac{1}{2}I_N(\theta,A_N+V_NB_N(V_N)^*)\right]-\frac{1}{N}\mathbb{E}[\log I_N(\theta,A_N+V_NB_N(V_N)^*)]\right)=0$$

which completes the proof of Theorem 5.2. ∎

We go back to the **proof of Lemma 24.** Observe first that

$$\begin{aligned}
L_N(\theta,A,B) &:= \mathbb{E}[I_N(\theta,A_N+V_NB_N(V_N)^*)^2] \\
&= \int e^{\theta N\left((UAU^*)_{11}+(\tilde{U}A\tilde{U}^*)_{11}+(UV_NB(UV_N)^*)_{11}+(\tilde{U}V_NB(\tilde{U}V_N)^*)_{11}\right)}dm_N^{(1)}(U)dm_N^{(1)}(\tilde{U})dm_N^{(1)}(V_N) \\
&\qquad = \int e^{\theta N\left((UAU^*)_{11}+(\tilde{U}A\tilde{U}^*)_{11}+(VBV^*)_{11}+(\tilde{U}U^*VBV^*U\tilde{U}^*)_{11}\right)}dm_N^{(1)}(V)dm_N^{(1)}(U)dm_N^{(1)}(\tilde{U})
\end{aligned}$$

where we used that $m_N^{(1)}$ is invariant by the action of the orthogonal group. We shall now prove that $L_N(\theta,A,B)$ factorizes. The proof requires sharp estimates of spherical integrals. We already got the kind of estimates we need in section 3. The ideas here will be very similar although the calculations will be more involved.

To rewrite $L_N(\theta,A,B)$ in a more proper way, the key observation is that, if we consider the column vector $W := (V^*U\tilde{U}^*)_1$ then $\langle V_1,W\rangle = \langle U_1,\tilde{U}_1\rangle$ so that we have the decomposition

$$W = \langle U_1,\tilde{U}_1\rangle V_1 + (1-|\langle U_1,\tilde{U}_1\rangle|^2)^{\frac{1}{2}}V_2$$

with (V_1,V_2) orthogonal and distributed uniformly on the sphere.
Therefore,

$$L_N(\theta,A,B) = \mathbb{E}\left[\exp\{N\theta(F_1^N+F_2^N+F_3^N+F_4^N+F_5^N)\}\right]$$

with

$$\begin{aligned}
F_1^N &= \langle U,AU\rangle \\
F_2^N &= \langle \tilde{U},A\tilde{U}\rangle \\
F_3^N &= (1+\langle U,\tilde{U}\rangle^2)\langle V_1,BV_1\rangle \\
F_4^N &= 2(1-|\langle U,\tilde{U}\rangle|^2)^{\frac{1}{2}}\langle U,\tilde{U}\rangle\langle V_1,BV_2\rangle \\
F_5^N &= (1-\langle U,\tilde{U}\rangle^2)\langle V_2,BV_2\rangle
\end{aligned}$$

where U, $\tilde{U}$ are two independent vectors following the uniform law on the sphere of radius $\sqrt{N}$ in $\mathbb{R}^N$ and V_1, V_2 are the two first column vectors of a matrix V following $m_N^{(1)}$, U, $\tilde{U}$ and V being independent.

We now adopt the same strategy as in section 3 to show that the F_i's will become asymptotically independent (or negligible). More precisely, we use again Fact 8 and recall that we can write $U = \frac{g^{(1)}}{\|g^{(1)}\|}$, $\tilde{U} = \frac{g^{(2)}}{\|g^{(2)}\|}$, $V_1 = \frac{g^{(3)}}{\|g^{(3)}\|}$ and $V_2 = \frac{G}{\|G\|}$ with $G = g^{(4)} - \frac{\langle g^{(3)}, g^{(4)} \rangle}{\|g^{(4)}\|^2} g^{(3)}$ where $g^{(1)}$, $g^{(2)}$, $g^{(3)}$ and $g^{(4)}$ are 4 i.i.d standard Gaussian vectors. We now set for $i = 1, 2, 3, 4$, with $\lambda_j^{(i)}$ the eigenvalues of A for $i = 1$ or 2 and of B for $i = 3$ or 4, $v_i = R_{\mu_A}(2\theta)$ for $i = 1$ or 2, $v_i = R_{\mu_B}(2\theta)$ for $i = 3$ or 4,

$$\hat{U}_i^N = \frac{1}{N} \sum_{j=1}^{N} (g_j^{(i)})^2 - 1, \text{ and } \hat{V}_i^N = \frac{1}{N} \sum_{j=1}^{N} \lambda_j^{(i)} (g_j^{(i)})^2 - v_i$$

Moreover, we let for $i = 1$ or 2,

$$\hat{W}_i^N = \frac{1}{N} \sum_{j=1}^{N} \lambda_j^{(i)} g_j^{(2i-1)} g_j^{(2i)} \text{ and } \hat{Z}_i^N = \frac{1}{N} \sum_{j=1}^{N} g_j^{(2i-1)} g_j^{(2i)}.$$

Under the Gaussian measure, all these quantities are going to zero almost surely and we can localize L_N as we made it in section 2, that is to say restrict the integration to the event $A'_N := \left\{ \hat{U}_i^N, \hat{V}_i^N, \hat{W}_i^N, \hat{Z}_i^N \text{ are } o(N^{-\frac{1}{2}+\kappa}) \right\}$, for any $\kappa > 0$. We then express the F_i's as function of these variables and on A'_N we expand them till $o(N^{-1})$. For example, on A'_N,

$$F_1 = \frac{\hat{V}_1^N + v_1}{\hat{U}_1^N + 1} = v_1 + (\hat{V}_1^N - v_1 \hat{U}_1^N) - \hat{U}_1^N (\hat{V}_1^N - v_1 \hat{U}_1^N) + o(N^{-1})$$

and all the calculations go the same way so that we get that the full second order in $\sum_i F_i$ is

$$\Xi^N = -\sum_{i=1}^{4} \hat{U}_i^N (\hat{V}_i^N - v_i \hat{U}_i^N) + 2(\hat{Z}_1^N - \hat{Z}_2^N) \hat{W}_2^N - 2v_2 \hat{Z}_2^N \hat{Z}_1^N + 2v_2 (\hat{Z}_2^N)^2$$

Now, as before, we consider the shifted probability measure P_N (which contains all the first order term above) under which $(\tilde{g}^{(i)})_{i=1,\ldots,4}$ defined by $\tilde{g}_j^{(i)} = \sqrt{1 + 2\theta v_i - 2\theta \lambda_j^{(i)}} g_j^{(i)}$ are i.i.d standard Gaussian vectors.
Under P_N, the $(\hat{U}_i^N, \hat{V}_i^N)_{1 \leqslant i \leqslant 4}$ are still independent with the same law than for the one dimensional case. Moreover, we see that for $i = 1, 2, 3, 4$, $j = 1, 2$,

$$\lim_{N \to \infty} N \mathbb{E}[\hat{U}_i^N \hat{Z}_j^N] = 0, \quad \lim_{N \to \infty} N \mathbb{E}[\hat{U}_i^N \hat{W}_j^N] = 0.$$

Similarly, $(\hat{Z}_i^N, \hat{W}_i^N)_{i=1,2}$ are asymptotically uncorrelated. Moreover, with $\mu_1 = \mu_A$ and $\mu_2 = \mu_B$,

$$\lim_{N\to\infty} N\mathbb{E}[\hat{W}_i^N \hat{Z}_i^N] = \int \frac{x}{(1+2\theta(v_i - x))^2} d\mu_i(x)$$
$$\lim_{N\to\infty} N\mathbb{E}[(\hat{W}_i^N)^2] = \int \frac{x^2}{(1+2\theta(v_i - x))^2} d\mu_i(x)$$
$$\lim_{N\to\infty} N\mathbb{E}[(\hat{Z}_i^N)^2] = \int \frac{1}{(1+2\theta(v_i - x))^2} d\mu_i(x).$$

Thus, with $G_i^N = \theta v_i - \frac{1}{2N}\sum_{j=1}^N \log(1 - 2\theta\lambda_j^{(i)} + 2\theta v_i)$ and if the Gaussian integral is well defined, we have

$$L_N(\theta, A, B) = \frac{e^{2NG_1^N + 2NG_2^N}}{\det(K_A)\det(K_B)} \int \exp\{2\theta(\hat{z}_1 - \hat{z}_2)\hat{w}_2 - 2v_2\theta\hat{z}_2\hat{z}_1 + 2v_2\theta(\hat{z}_2)^2\} \prod_{i=1,2} dP_i(\hat{w}_i, \hat{z}_i)(1 + o(1))$$

with P_i the law of two Gaussian variables with covariance matrix

$$\frac{R_i}{2} = \begin{pmatrix} \int \frac{1}{(1+2\theta(v_i-x))^2} d\mu_i(x) & \int \frac{x}{(1+2\theta(v_i-x))^2} d\mu_i(x) \\ \int \frac{x}{(1+2\theta(v_i-x))^2} d\mu_i(x) & \int \frac{x^2}{(1+2\theta(v_i-x))^2} d\mu_i(x) \end{pmatrix}$$

and K_A and K_B as defined in (37) if we replace μ_E therein respectively by μ_A or μ_B.
We now integrate on the variables $(\hat{z}_2, \hat{w}_2)$ so that the Gaussian computation gives

$$L_N(\theta, A, B) = \frac{e^{2NG_1^N + 2NG_2^N}}{\det(K_A)\det(K_B)^{\frac{3}{2}}} \int \exp\{\theta^2 \langle e, K_B^{-1} e\rangle \hat{z}_1^2\} dP_1(\hat{z}_1, \hat{w}_1)(1 + o(1))$$

with $e = (-v_2, 1)$. To show that the remaining integral is finite it is enough to check that

$$-2\theta^2 \langle e, K_B^{-1} e\rangle + \mathrm{var}\hat{z}_1 \geqslant 0,$$

at least for θ small enough. But we can check that $\theta^2 \langle e, K_B^{-1} e\rangle \approx \theta^2 \sigma_2$, with $\sigma_2 = \int x^2 d\mu_B(x)$ whereas the variance of $\hat{z}_1$ is of order 1.
This finishes to prove that for sufficiently small θ's there exists a finite constant $C(\theta, A, B)$ such that

$$L_N(\theta, A, B) = \frac{e^{2NG_1^N + 2NG_2^N}}{\det(K_A)\det(K_B)} C(\theta, A, B)(1 + o(1))$$

Since on the other hand we have seen in section 3 that

$$I_N(\theta, A) = \frac{e^{NG_1^N}}{\det K_A^{\frac{1}{2}}}(1 + o(1)) \text{ and } I_N(\theta, B) = \frac{e^{NG_2^N}}{\det K_B^{\frac{1}{2}}}(1 + o(1)),$$

we have proved Lemma 24. ∎

Acknowledgments : We are very grateful to O. Zeitouni for helpful discussion at the beginning of this work, which in particular allowed us to obtain the second order correction in the full high temperature region. We would like also to thank P. Śniady for many useful comments during this research. We thank Y. Ollivier for pointing out [9] and showing us how Lemma 23 could be deduced, which simplified a lot the argument. Finally, we are also very grateful to an anonymous referee whose careful reading and useful comments helped us to improve the coherence of the paper. We thank Alexei Onatski for pointing out a typo on the published version of Theorem 3 (value of the first limit).

References

[1] Ben Arous, G. Methods de Laplace et de la phase stationnaire sur l'espace de Wiener. *Stochastics 25*, 3 (1988), 125–153.

[2] Ben Arous, G., Dembo, A., and Guionnet, A. Aging of spherical spin glasses. *Probab. Theory Related Fields 120*, 1 (2001), 1–67.

[3] Bolthausen, E. Laplace approximations for sums of independent random vectors. *Probab. Theory Relat. Fields 72*, 2 (1986), 305–318.

[4] Borel, E. Sur les principes de la théorie cinétique des gaz. *Annales de l'École Normale Supérieure 23* (1906), 9–32.

[5] Collins, B. Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability. *Int. Math. Res. Not.*, 17 (2003), 953–982.

[6] Collins, B., and Sniady, P. Integration with respect to the haar measure on unitary, orthogonal and symplectic group. *arviv http://arxiv.org/abs/math-ph/0402073* (2004).

[7] D'Aristotile, A., Diaconis, P., and Newman, C. M. Brownian motion and the classical groups. In *Probability, statistics and their applications: papers in honor of Rabi Bhattacharya*, vol. 41 of *IMS Lecture Notes Monogr. Ser.* Inst. Math. Statist., Beachwood, OH, 2003, pp. 97–116.

[8] Dembo, A., and Zeitouni, O. *Large deviations techniques and applications*, second ed., vol. 38 of *Applications of Mathematics.* Springer-Verlag, New York, 1998.

[9] Gromov, M., and Milman, V. D. A topological application of the isoperimetric inequality. *Amer. J. Math. 105*, 4 (1983), 843–854.

[10] Guionnet, A. First order asymptotics of matrix integrals ; a rigorous approach towards the understanding of matrix models. *Comm. Math. Phys.* (2003).

[11] Guionnet, A., and Maïda, M. Character expansion method for the first order asymptotics of a matrix integral. Preprint, http://lanl.arXiv.org/abs/math.PR/0401229, 2004.

[12] Guionnet, A., and Zeitouni, O. Large deviations asymptotics for spherical integrals. *J. Funct. Anal. 188*, 2 (2002), 461–515.

[13] Harish-Chandra. Differential operators on a semisimple Lie algebra. *Amer. J. Math. 79* (1957), 87–120.

[14] Harish-Chandra. Fourier transforms on a semisimple Lie algebra. I. *Amer. J. Math. 79* (1957), 193–257.

[15] Itzykson, C., and Zuber, J. B. The planar approximation. II. *J. Math. Phys. 21*, 3 (1980), 411–421.

[16] Jiang, T. How many entries of a typical orthogonal matrix can be approximated by independent normals ? Preprint, 2003.

[17] Karlin, S., and McGregor, J. Coincidence probabilities. *Pacific J. Math. 9* (1959), 1141–1164.

[18] Kazakov, V. Solvable matrix models. In *Random matrix models and their applications*, vol. 40 of *Math. Sci. Res. Inst. Publ.* Cambridge Univ. Press, Cambridge, 2001, pp. 271–283.

[19] Marinari, E., Parisi, G., and Ritort, F. Replica field theory for determinstic models. II. A non-random spin glass with glassy behaviour. *J. Phys. A 27*, 23 (1994), 7647–7668.

[20] Matytsin, A. On the large-N limit of the Itzykson-Zuber integral. *Nuclear Phys. B 411*, 2-3 (1994), 805–820.

[21] Sagan, B. E. *The symmetric group.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1991. Representations, combinatorial algorithms, and symmetric functions.

[22] Talagrand, M. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, 81 (1995), 73–205.

[23] Voiculescu, D. Addition of certain noncommuting random variables. *J. Funct. Anal. 66*, 3 (1986), 323–346.

[24] Voiculescu, D. The analogues of entropy and of Fisher's information measure in free probability theory. I. *Comm. Math. Phys. 155*, 1 (1993), 71–92.

[25] Zelditch, S. Macdonald's identities and the large N limit of YM_2 on the cylinder. *Comm. Math. Phys. 245*, 3 (2004), 611–626.

[26] Zinn-Justin, P., and Zuber, J.-B. On some integrals over the $\mathrm{U}(N)$ unitary group and their large N limit. *J. Phys. A 36*, 12 (2003), 3173–3193. Random matrix theory.

[27] Zvonkin, A. Matrix integrals and map enumeration: an accessible introduction. *Math. Comput. Modelling 26*, 8-10 (1997), 281–304. Combinatorics and physics (Marseilles, 1995).