

## Large deviations for the largest eigenvalue of matrices with variance profiles\*

Jonathan Husson<sup>†</sup>

### Abstract

In this article we consider Wigner matrices  $(X_N)_{N \in \mathbb{N}}$  with variance profiles which are of the form  $X_N(i, j) = \sigma(i/N, j/N)a_{i,j}/\sqrt{N}$  where  $\sigma$  is a symmetric real positive function of  $[0, 1]^2$ , either continuous or piecewise constant and where the  $a_{i,j}$  are independent, centered of variance one above the diagonal. We prove a large deviation principle for the largest eigenvalue of those matrices under the condition that they have sharp sub-Gaussian tails and under some additional assumptions on  $\sigma$ . These sub-Gaussian bounds are verified for example for Gaussian variables, Rademacher variables or uniform variables on  $[-\sqrt{3}, \sqrt{3}]$ . This result is new even for Gaussian entries.

**Keywords:** random matrices; large deviations; largest eigenvalue.

**MSC2020 subject classifications:** 60B20; 60F10.

Submitted to EJP on September 4, 2020, final version accepted on April 29, 2022.

Supersedes arXiv:2002.01010v3.

## 1 Introduction

One of the key results in random matrix theory is Wigner's theorem: it establishes the convergence of the empirical measure of the eigenvalues of Wigner matrices towards the semi-circular measure [37]. These Wigner matrices are a model of real or complex self-adjoint random matrices with independent centered subdiagonal entries of variance  $1/N$  and independent centered diagonal entries of variance  $O(1/N)$ . Later Füredi and Komlós proved that the largest eigenvalue of such matrices converges almost surely toward 2 [25] under an assumption of boundedness on the moments of the entries. This moment hypothesis was then relaxed to an hypothesis of boundedness for the fourth moment by Vu in [36] which was later proved to be necessary by Lee and Yin in [31]. Similar results also exist for Wishart matrices (that is matrices of the form  $\frac{1}{M}X^*X$  where  $X$  is a  $M \times N$  random matrix with i.i.d. centered entries of variance 1) and for matrices with variance profiles (that is self-adjoint random matrices whose diagonal and

---

\*This work was partially supported by ERC Project LDRAM: ERC-2019-ADG Project 884584.

<sup>†</sup>University of Michigan E-mail: [jhusson@umich.edu](mailto:jhusson@umich.edu)

subdiagonal entries are independent centered but whose variance of the entries may not be constant up to a factor  $1/N$ ). In that case the limit of the empirical measure depends on the profile [26].

Once one knows the limits of the empirical measure and the largest eigenvalue, one can wonder how the probability that they are away from these limits behaves. These questions are of great importance for instance in mobile communication systems [18, 24] and in the study of the energy landscape of disordered systems [10, 32]. In the case of matrices from the Gaussian Orthogonal Ensemble or the Gaussian Unitary Ensemble, thanks to the the orthogonal or unitary invariance of the distributions, the joint law of the eigenvalues is explicitly known (see for example [33]) and the spectrum behaves like a so-called  $\beta$ -ensemble. By Laplace's principle, once one takes care of the singularities, those formulas lead to large deviation principles both for the empirical measure [16] and the largest eigenvalue [15].

In the case of general distributions, since eigenvalues are complicated functions of the entries, large deviations remain mysterious. Concentration of measure results were obtained in compactly supported and log-Sobolev settings by Guionnet and Zeitouni [29]. Several recent breakthroughs proved large deviation principles for matrices with entries with distributions whose tails are heavier than Gaussian both for the empirical measure and the largest eigenvalue respectively by Bordenave and Caputo and by Augeri [11, 19]. Those results rely on the fact that the large deviation behaviour comes from a small number of large entries. These ideas are further generalized to the questions of subgraphs counts and eigenvalues of random graphs in [12, 21, 17]. In the case of sub-Gaussian entries, a large deviation principle for the largest eigenvalue of matrices with Rademacher-distributed entries was proved by Guionnet and the author in [27] using the asymptotics of Itzykson-Zuber integrals computed by Guionnet and Maïda in [28]. Indeed, one obtains the large deviations by tilting the measure by spherical integrals. Under this tilted law, the matrix is roughly distributed as a sum  $W_N + R$  where  $W_N$  is a Wigner matrix and  $R$  a deterministic matrix of rank one. Then, the largest eigenvalue of such a deformed model is well known and follows the phenomenon of BBP transition (coined after Bai, Ben Arous and Péché who observed it in the case of deformations of sample covariance matrices [14]). Notably the rate function for the large deviation principle of the largest eigenvalue of a matrix with Rademacher entries is the same as for the GOE. The crucial hypothesis verified by the Rademacher law that assure that the upper and lower large deviations bounds both coincides with those of the Gaussian case is the so-called sharp sub-Gaussianity. This property of the Rademacher law is expressed in terms of its Laplace transform:

$$\forall t \in \mathbb{R}, \int \exp(tx) d\mu(x) \leq \frac{\exp(\mu(x^2)t^2)}{2}$$

For distributions that are sub-Gaussian but not sharply so, large deviation lower and upper bounds were also proved by Augeri, Guionnet and the author in [13] for large values and values near the bulk of the limit measure. In this case though our rate function near infinity can be strictly smaller to the rate function for the GOE.

Wigner's original approach to determine the limit of the empirical measure was to estimate the trace of moments of Wigner matrices but a more modern approach is to estimate the resolvent  $(z - W_N)^{-1}$  using the Schur complement formula. One then finds that the Stieltjes transform  $m$  of the limit measure must be a solution of the so called Dyson equation:

$$\frac{1}{m(z)} = z - m(z), \forall z \in \mathbb{C} \setminus \mathbb{R}$$

with the convention that the Stieltjes transform of  $\mu$  is  $z \mapsto \mu((z - x)^{-1})$ . Furthermore, if

for  $z \in \mathbb{H}^+$  (where  $\mathbb{H}^+$  is  $\{z \in \mathbb{C} : \Im(z) > 0\}$ ) we set the condition  $\Im m(z) < 0$ , then the only solution to this equation for  $z \in \mathbb{H}^+$  is  $m(z) = (z - \sqrt{z^2 - 4})/2$  which is the Stieltjes transform of the semicircular measure  $\sqrt{(4 - x^2)_+} dx/2\pi$ . In the case of matrices with variance profiles, it can be computed again by the Schur complement formula applied on the resolvent  $G(z) = (z - W_N)^{-1}$ , which shows that up to an error term, its diagonal terms satisfies the following equation which admits only one solution of negative imaginary part:

$$\frac{1}{G_{i,i}(z)} = z - \sum_j s_{i,j} G_{j,j}(z), \forall z \in \mathbb{H}^+$$

where  $s_{i,j} = N\mathbb{E}[|X_N(i,j)|^2]$ .

Then, using that the Stieltjes transform of the empirical measure is  $N^{-1} \sum_i G_{i,i}(z)$  one can find the limit measure. This equation has been used to study those matrices for instance by Girko in [26] by Khorunzy and Pastur in [30], Anderson and Zeitouni in [8] and Schlyakhtenko in [34]. It was extensively studied in itself by Alt, Erdős, Ajanki, Kreuger and Schröder in a series of articles where it is used to prove local laws and universality of the local eigenvalue statistics both on the bulk, the cusp and the edge of the spectrum [4, 22, 20, 1, 3, 2, 5]. One may want to look at [23] for a more thorough review on the subject.

In this article, we will use the techniques developed in [27] and apply them to random matrices with variance profiles to prove a large deviation principle for the largest eigenvalue. We will place ourselves in the same context of entries with sharp sub-Gaussian law. Such a result is new, even for matrices with Gaussian entries and once again our rate function will not depend on the laws of the entries. We will consider a symmetric (or Hermitian) matrix model  $(X_N)_{N \in \mathbb{N}}$  with independent sub-diagonal entries with a variance profile  $\Sigma_N(i,j) = N^{1/2} \sqrt{\mathbb{E}[|X_N(i,j)|^2]}$ . We will consider a piecewise constant case where  $\Sigma_N$  is equal to some  $\sigma_{k,l}$  on squares of the form  $I_k^{(N)} \times I_l^{(N)}$  for  $k, l = 1, \dots, n$  where  $(I_k^{(N)})_{1 \leq k \leq n}$  is a collection of disjoint intervals covering  $\llbracket 1, N \rrbracket$  and such that  $I_k^{(N)}/N$  converges to some non-trivial interval  $I_k$  of  $[0, 1]$ . In this case we will define  $\sigma$  to be the piecewise constant function equal to  $\sigma_{k,l}$  on  $I_k \times I_l$ . We will also consider the case of a variance profile which converges toward a continuous function  $\sigma$  in the sense that  $\lim_N \sup_{i,j} |\Sigma_N(i,j) - \sigma(i/N, j/N)| = 0$ . In both cases the empirical measure converges to a measure  $\mu_\sigma$  characterized by the fact that its Stieltjes transform  $m$  is equal to  $m(z) = \int_0^1 \mathfrak{m}(x, z) dx$  where  $\mathfrak{m}$  is the only solution of the equation (see [26]):

$$\frac{1}{\mathfrak{m}(x, z)} = z - \int \sigma^2(y, x) \mathfrak{m}(y, z) dy, \forall z \in \mathbb{H}^+$$

with the condition that  $\Im \mathfrak{m}(x, z) < 0$  for for all  $x \in [0, 1], z \in \mathbb{H}^+$ . We will find then that large deviations of the largest eigenvalue occur when we tilt our measure so that  $X_N$  has roughly the same law than  $\tilde{X}_N + R$  where  $\tilde{X}_N$  is a random matrix with the same variance profile as  $X_N$  and where  $R$  is deterministic and of finite rank. Since  $\tilde{X}_N$  will not be a Wigner matrix, finding the correct tilt will be more involved than in the classical case and will require some additional hypothesis on the variance profiles in order for the tilt to yield the desired lower bound.

First, in section 2 we will introduce the rate function and the assumption on the variance profile we will need in order for our large deviation lower bound to coincide with our upper bound. In sections 3 to 5 we will treat the case of matrices with piecewise constant variance profile which bears the most similarities with the models treated in [27]. In these sections we will insist on the differences with [27] while redirecting the reader to it for the parts of the proofs that stay the same. We will first prove a large

deviations upper bound using an annealed spherical integral in section 4. We will then tilt our initial measure to prove the lower bound in section 5. There we will use the assumption made in section 2 to prove we can find a good tilt. In section 6 we will approximate the case of a continuous variance profile using piecewise constant ones. We will have to prove the convergence of the rate functions of the approximations. Since the approximations will only satisfy our lower bound up to an error term, we will also prove that this error can ultimately be neglected. In section 7 we will illustrate the cases where our result applies in the simple context of a piecewise constant variance profile with four blocks. In the same section we will illustrate the limits of our approach and the necessity to make some assumptions concerning the variance profiles, with an example of a matrix whose variance profile does not satisfy our assumptions and such that the rate function for the large deviations of the largest eigenvalue does not match our rate function. Finally, in section 8 we will discuss the explicit value of the rate function and in particular we will present a condition that when verified assures us that the rate function does depend on the variance profile only through the limit measure of the matrix model.

**Acknowledgement**

The author would like to thank Alice Guionnet for her help proofreading this article and particularly the introduction, Ion Nechita for bringing to his attention the example in Remark 2.6 and the referees for their careful reading and suggestions. This work was partially supported by the ERC Advanced Grant LDRaM (ID:884584).

**1.1 Variance profiles**

In the rest of the article, a real  $x$  is said to be non-negative if  $x \geq 0$ ,  $\mathbb{R}^+$  is the set  $\{x \in \mathbb{R} : x \geq 0\}$  and  $\mathbb{R}^{+,*} = \mathbb{R}^+ \setminus \{0\}$ . Our random matrix model will be of the form  $W_N \odot \Sigma_N$ , where  $W_N$  is either a real or a complex Wigner matrix,  $\Sigma_N$  is a real symmetric matrix and  $\odot$  is the entrywise product.  $\mathcal{P}(A)$ , where  $A$  is a measurable space, will denote the set of probability measures on  $A$ . We denote for  $n \in \mathbb{N}$  and a set  $A$ ,  $\mathcal{S}_n(A)$  the set of symmetric matrices with entries in  $A$ . First of all, we describe the matrices  $\Sigma_N$  we will be using. These matrices will converge as piecewise constant functions of  $[0, 1]^2$  to some function  $\sigma$  on  $[0, 1]^2$  called the variance profile. We will consider here two cases: the case where  $\sigma$  is piecewise constant and the case where it is continuous.

**Piecewise constant variance profile:** We consider a variance profile piecewise constant on rectangular blocks. Let  $n \in \mathbb{N}^*$ ,  $\Sigma = (\sigma_{i,j})_{i,j \in \llbracket 1, n \rrbracket}$  a real symmetric  $n \times n$  matrix with non-negative coefficients and  $\vec{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$  such that for every  $i$ ,  $\alpha_i > 0$  and  $\alpha_1 + \dots + \alpha_n = 1$ . In this context we will consider  $\Sigma_N$  defined by block by:

$$\Sigma_N(i, j) = \sigma_{k,l} \text{ if } i \in I_k^{(N)} \text{ and } j \in I_l^{(N)}$$

where for all  $N \in \mathbb{N}$ ,  $\{I_1^{(N)}, \dots, I_n^{(N)}\}$  is a partition of  $\llbracket 1, N \rrbracket$  such that for all  $j \in \llbracket 1, n \rrbracket$ ,  $I_j^{(N)} = \llbracket a_j^N + 1, a_{j+1}^N \rrbracket$  where for every  $N$ , the  $a_j^N$  are such that:

$$0 = a_1^N \leq a_2^N \leq \dots \leq a_{n+1}^N = N$$

and such that for  $j \in \llbracket 1, n \rrbracket$ :

$$\lim_N \frac{|I_j^{(N)}|}{N} = \alpha_j$$

We then define  $(\gamma_i)_{1 \leq i \leq n}$  and  $(I_i)_{1 \leq i \leq n}$  by:

$$\gamma_0 = 0 \text{ and } \forall j \neq 0, \gamma_j := \sum_{i=1}^j \alpha_i \text{ and } I_i = [\gamma_{i-1}, \gamma_i[.$$

We shall also denote  $\sigma : [0, 1]^2 \rightarrow \mathbb{R}^+$  the piecewise constant function defined by

$$\sigma(x, y) = \sigma_{k,l} \text{ if } (x, y) \in I_k \times I_l.$$

This setting will be referred as the case of a piecewise constant variance profile associated to the parameters  $\Sigma$  and  $\vec{\alpha}$ .

**Continuous variance profile:** In this case, we will consider a real non-negative symmetric continuous function  $\sigma : [0, 1]^2 \rightarrow \mathbb{R}^+$  and for every  $N$ , we will consider a symmetric matrix with non negative entries  $\Sigma_N$  such that the sequence  $\Sigma_N$  satisfies:

$$\lim_{N \rightarrow \infty} \sup_{1 \leq i, j \leq N} \left| \Sigma_N(i, j) - \sigma\left(\frac{i}{N}, \frac{j}{N}\right) \right| = 0.$$

In both cases, we will call  $\sigma$  the variance profile of the matrix model.

### 1.2 The generalized Wigner matrix model

For the Wigner matrix  $W_N$ , we will consider two cases, a real symmetric one when  $\beta = 1$  and a complex Hermitian one when  $\beta = 2$ . For every  $N$ , we will consider a family of independent random variables  $(a_{i,j}^{(\beta)})_{1 \leq i \leq j \leq N}$  such that  $a_{i,j}^{(\beta)}$  has the distribution  $\mu_{i,j}^N$ . For  $\beta = 1$ ,  $a_{i,j}^{(\beta)}$  is a real random variable for all  $i, j$  and for  $\beta = 2$ ,  $a_{i,j}^{(\beta)}$  will be a complex random variable for  $i \neq j$  and a real random variable for  $i = j$ . These will be the unrenormalized entries of  $W_N$ . We will assume that all the  $\mu_{i,j}^N$  are centered:

$$\int x d\mu_{i,j}^N(x) = 0, \forall 1 \leq i \leq j \leq N.$$

For  $\beta = 1$  we assume that off-diagonal entries have variance 1 and diagonal entries have variance 2:

$$\mu_{i,j}^N(x^2) = \int_{\mathbb{R}} x^2 d\mu_{i,j}^N(x) = 1, \forall 1 \leq i < j \leq N, \quad \mu_{i,i}^N(x^2) = 2, \quad \forall 1 \leq i \leq N.$$

For  $\beta = 2$ , if we denote  $x$  the function  $z \mapsto \Re z$  and  $y$  the function  $z \mapsto \Im z$ , we assume the following conditions on the variances of the entries:

$$\mu_{i,j}^N(x^2) = \mu_{i,j}^N(y^2) = \frac{1}{2}, \mu_{i,j}^N(xy) = 0, \forall 1 \leq 1 < j \leq N, \quad \mu_{i,i}^N(x^2) = 1, \forall 1 \leq i \leq N.$$

If  $\mu$  is a probability measure on some  $\mathbb{R}^d$  with a covariance matrix  $C$ , we say that  $\mu$  has a *sharp sub-Gaussian Laplace transform* if

$$\forall u \in \mathbb{R}^d, T_\mu(u) := \int_{\mathbb{R}^d} \exp(\langle u, x \rangle) d\mu(x) \leq \exp \frac{\langle u, Cu \rangle}{2}.$$

For  $\mu$  a measure on  $\mathbb{C}$ , we can identify  $\mathbb{R}^2$  and  $\mathbb{C}$  and then the Laplace transform  $T_\mu$  can be expressed for  $z \in \mathbb{C}$  as

$$T_\mu(z) = \int_{\mathbb{C}} \exp(\Re \bar{z}x) d\mu(x).$$

We will need to make the following assumption on the  $\mu_{i,j}^N$ :

**Assumption 1.1.** For every  $N \in \mathbb{N}$  and  $1 \leq i \leq j \leq N$ , the distribution  $\mu_{i,j}^N$  has sub-Gaussian Laplace transforms.

In particular, we can notice that in the complex case, it implies that for  $i \neq j, \forall z \in \mathbb{C}$ :

$$T_{\mu_{i,j}^N}(z) \leq \exp\left(\frac{|z|^2}{4}\right).$$

Examples of distributions that satisfy this sharp sub-Gaussian bound in  $\mathbb{R}$  are the (centered) Gaussian laws the Rademacher laws  $\frac{1}{2}(\delta_{-1} + \delta_1)$  and the uniform law on a centered interval. On  $\mathbb{C}$ , if  $X$  is a random variable such that  $\Re(X)$  and  $\Im(X)$  are independent and have sharp sub-Gaussian Laplace transform, then  $X$  has a sharp sub-Gaussian Laplace transform.

**Remark 1.2.** From the sharp sub-Gaussian bound, we have the following bound on the moments of  $\mu_{i,j}^N$  if Assumption 1.1 is verified for  $\beta = 1$  and  $X$  is a random variable of distribution  $\mu_{i,j}^N$ :

$$\mathbb{E}[X^{2k}] \leq (2k)!(T_{\mu_{i,j}^N}(-1) + T_{\mu_{i,j}^N}(1))/2 \leq (2k)!e^{\mu_{i,j}^N(x^2)/2}$$

and

$$\mathbb{E}[|X|^{2k+1}] \leq \mathbb{E}[X^{2k+2}]^{2k+1/2k+2} \leq ((2k+2)!e^{\mu_{i,j}^N(x^2)/2})^{2k+1/2k+2}.$$

We have a bound of the form:

$$\mathbb{E}[|X|^k] \leq Ck!$$

for some universal constant  $C$ . From this bound, we have that for every  $\delta > 0$ , there exists  $\epsilon > 0$  that does not depend on the laws  $\mu_{i,j}^N$  such that for  $|t| \leq \epsilon$ .

$$T_{\mu_{i,j}^N}(t) \geq \exp\left\{\frac{(1-\delta)t^2\mu_{i,j}^N(x^2)}{2}\right\}.$$

We have also that the  $T_{\mu_{i,j}^N}$  are uniformly  $C^3$  in a neighbourhood of the origin: for  $\epsilon > 0$  small enough  $\sup_{|t| \leq \epsilon} \sup_{i,j,N} |\partial_t^3 \ln T_{\mu_{i,j}^N}(t)|$  is finite. In the complex case, with the same method we have a similar result, that is that for every  $\delta > 0$ , there is  $\epsilon > 0$  such that for every  $z \in \mathbb{C}$  such that  $|z| \leq \epsilon$ :

$$T_{\mu_{i,j}^N}(z) \geq \exp\left\{\frac{(1-\delta)|z|^2}{4}\right\}.$$

for  $i \neq j$ .

For both those cases, we will need to use concentration inequalities to ensure that at the exponential scale we consider, the empirical measure of our matrices can be approximated by their typical value. To this we will need this classical assumption.

**Assumption 1.3.** *There exists a compact set  $K$  such that the support of all  $\mu_{i,j}^N$  is included in  $K$  for all  $i, j \in \llbracket 1, N \rrbracket$  and all integer number  $N$ , or all  $\mu_{i,j}^N$  satisfy a log-Sobolev inequality with the same constant  $c$  independent of  $N$ . In the complex case, we will suppose also that for all  $(i, j)$ , if  $Y$  is a random variable of law  $\mu_{i,j}$ , there is a complex  $a \neq 0$  such that  $\Re(aY)$  and  $\Im(aY)$  are independent.*

Now for  $\beta = 1$  or  $2$  and  $N \in \mathbb{N}$ , given the family  $(a_{i,j}^{(\beta)})$ , we define the following Wigner matrices:

$$W_N^{(\beta)}(i, j) = \begin{cases} \frac{a_{i,j}^{(\beta)}}{\sqrt{N}} & \text{when } i \leq j, \\ \frac{a_{j,i}^{(\beta)}}{\sqrt{N}} & \text{when } i > j. \end{cases}$$

From these definitions we define  $X_N^{(\beta)}$  a real (if  $\beta = 1$ ) or complex (if  $\beta = 2$ ) matrix with variance profile  $\Sigma_N$  as:

$$X_N^{(\beta)} := W_N^{(\beta)} \odot \Sigma_N$$

where for two matrices  $A = (a_{i,j})_{i,j \in \llbracket 1, n \rrbracket}$ ,  $B = (b_{i,j})_{i,j \in \llbracket 1, n \rrbracket}$ ,  $A \odot B$  is the matrix  $(a_{i,j} b_{i,j})_{i,j \in \llbracket 1, n \rrbracket}$ .

If  $A$  is a self-adjoint  $N \times N$  matrix, we denote  $\lambda_{\min}(A) = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N = \lambda_{\max}(A)$  its eigenvalues and  $\mu_A$  its empirical measure:

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

For  $X_N^{(\beta)}$ , we will abbreviate  $\mu_{X_N^{(\beta)}}$  in  $\hat{\mu}_N$  throughout the article.

### 1.3 Statement of the results

First of all with this matrix model, we will state with the following theorem the existence of a limit in probability of the empirical measure  $\hat{\mu}_N$ . This limit, which depends only on the limit  $\sigma$  of the variance profile is described in more detail in the Appendix A where this theorem is proved:

**Theorem 1.4.** *Both in the piecewise constant and in the continuous case, the empirical measure  $\hat{\mu}_N$  converges weakly in probability toward a compactly supported measure  $\mu_\sigma$  which only depends on  $\sigma$ .*

This theorem is in fact an almost direct consequence from [26, Theorem 1.1]. It can also be obtained in the piecewise case and in the continuous case with the additional assumption that  $\sigma$  is 1/2-Hölder by applying Lemma 9.2 from [6] which itself uses stability results for the Dyson equation. Since here the Dyson equation of our setting is simpler, we present a more elementary proof of this result using rougher stability results in Appendix A. We denote by  $r_\sigma$  the rightmost point of the support of  $\mu_\sigma$ . First of all, we have the following result for the convergence of the largest eigenvalue of  $X_N^{(\beta)}$ .

**Theorem 1.5.** *Suppose that Assumption 1.1 holds. Both in the piecewise constant case and the continuous case, we have that  $\lambda_{\max}(X_N^{(\beta)})$  converges almost surely toward  $r_\sigma$ .*

This theorem is a generalization of the result of convergence of the largest eigenvalue toward 2 in the Wigner case which was proved by Füredi and Komlós [25] for distributions with moments such that  $\mathbb{E}[|a_{i,j}^{(\beta)}|^k] \leq k^{Ck}$  for some  $C > 0$  and then by Vu for distributions with finite fourth moment [36]. For this result, we need only to have a bound of the form  $\mathbb{E}[|a_{i,j}^{(\beta)}|^k] \leq r_k$  for some sequence  $(r_k)_{k \in \mathbb{N}}$  (this hypothesis is automatically verified with our sharp sub-Gaussian bound). There are numerous similar results of convergence for the largest eigenvalue in the literature for models similar to this one, unfortunately, to the author knowledge none seem to quite correspond to the level of generality we are going for in this paper (the most similar to our model would be Theorem 2.7 from [7] but here we would like to allow for rectangular blocks). Therefore we will be using here a stronger kind of results, which are the local law results from [2] (corollary 2.10) in the case of a positive piecewise variance profile. The non-negative case as well as the continuous case will be proven by approximation, the only technicality is to prove that when we approximate a variance profile  $\sigma$  by a sequence of variance profiles  $(\sigma_n)$ , the rightmost point of the support of  $\mu_{\sigma_n}$  converges toward the rightmost point of the support of  $\mu_\sigma$  (see Lemma 6.4). Although using the local law may seem excessive for the purpose of proving the convergence of the largest eigenvalue, its anisotropic version will end up being used in the large deviation lower bound in section 5.

For the following theorem, which states a large deviation principle for  $\lambda_{\max}(X_N^{(\beta)})$ , we will need Assumptions 2.4 and 2.8 respectively for the case of a continuous variance

profile and for the case of a piecewise constant variance profile. These assumptions are more thoroughly discussed in section 2. Assumption 2.8 states that the following optimization problem for  $\psi \in \mathcal{P}([0, 1])$ :

$$\sup_{\psi \in \mathcal{P}([0,1])} \left\{ \frac{\theta^2}{\beta} \int_{[0,1]^2} \sigma^2(x, y) d\psi(x) d\psi(y) - \frac{\beta}{2} D(\text{Leb} || \psi) \right\}$$

has a determination of its maximum argument that is continuous in  $\theta$ .

Similarly, Assumption 2.4 states that the following optimization problem for  $\psi \in (\mathbb{R}^+)^n$  such that  $\sum \psi_i = 1$ :

$$\sup_{\psi \in (\mathbb{R}^+)^n, \sum \psi_i = 1} \left\{ \frac{\theta^2}{\beta} \sum_{i,j=1}^n \sigma_{i,j}^2 \psi_i \psi_j + \frac{\beta}{2} \sum_{i=1}^n \alpha_i (\log \psi_i - \log \alpha_i) \right\}$$

has a determination of its maximum argument that is continuous in  $\theta$ . Both assumptions are necessary to obtain the large deviation lower bound.

**Theorem 1.6.** *Suppose Assumptions 1.1, 1.3 hold. Furthermore suppose that Assumption 2.4 holds in the piecewise constant case or that Assumption 2.8 holds in the continuous case. Then, the law of the largest eigenvalue  $\lambda_{\max}(X_N^{(\beta)})$  of  $X_N^{(\beta)}$  satisfies a large deviation principle with speed  $N$  and good rate function  $I^{(\beta)}$  which is infinite on  $(-\infty, r_\sigma)$ .*

*In other words, for any closed subset  $F$  of  $\mathbb{R}$ ,*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left( \lambda_{\max}(X_N^{(\beta)}) \in F \right) \leq - \inf_F I^{(\beta)},$$

*whereas for any open subset  $O$  of  $\mathbb{R}$ ,*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left( \lambda_{\max}(X_N^{(\beta)}) \in O \right) \geq - \inf_O I^{(\beta)}.$$

*The same result holds for the opposite of the smallest eigenvalue  $-\lambda_{\min}(X_N^{(1)})$ . Furthermore  $I^{(2)} = 2I^{(1)}$*

The rate functions  $I^{(\beta)}$  are defined in section 2. Examples of variance profiles that satisfy our Assumptions 2.4 and 2.8 are also given in section 2.

## 2 The rate function

We will now define the rate function  $I^{(\beta)}$  in Theorem 1.6. This is in fact done the same way as in [27] with the supremum  $\sup_{\theta \geq 0} (J(\mu_\sigma, \theta, x) - F(\theta))$ .

In this formula,  $J(\mu_\sigma, \theta, x)$  is the limit of  $N^{-1} \log \mathbb{E}[\exp(N\theta \langle e, A_N e \rangle)]$  where  $e$  is a unitary vector taken uniformly on the sphere and  $A_N$  is a sequence of matrices such that the empirical measures converge weakly to  $\mu_\sigma$  and such that the sequence of the largest eigenvalues of  $A_N$  converges to  $x$ .  $F(\theta)$  is the limit of  $N^{-1} \log \mathbb{E}[\exp(N\theta \langle e, X_N e \rangle)]$  where the expectation is taken both in  $X_N$  and  $e$ . We will first describe the quantity  $F(\theta)$ .

### 2.1 The asymptotics of the annealed spherical integral

For  $\sigma : [0, 1]^2 \rightarrow \mathbb{R}^+$  a bounded measurable function and  $\psi$  a probability measure on  $[0, 1]$ , let us denote:

$$P(\sigma, \psi) := \int_0^1 \int_0^1 \sigma^2(x, y) d\psi(x) d\psi(y)$$

and for  $\theta > 0$ :



$$\Psi(\theta, \sigma, \psi) := \frac{\theta^2}{\beta} P(\sigma, \psi) - \frac{\beta}{2} D(\text{Leb}||\psi)$$

where  $D(\cdot||\cdot)$  is the Kullback-Leibler divergence, that is for  $\lambda, \mu \in \mathcal{P}([0, 1])$ :

$$D(\lambda||\mu) = \begin{cases} \int_0^1 \log\left(\frac{d\lambda}{d\mu}(x)\right) d\lambda(x) & \text{if } \lambda \text{ is absolutely continuous with respect to } \mu \\ +\infty & \text{if this is not the case} \end{cases}$$

and  $\text{Leb}$  is the Lebesgue measure on  $[0, 1]$ .

We consider here the following optimization problem with parameter  $\theta > 0$  on the set  $\mathcal{P}([0, 1])$ :

$$F(\sigma, \theta) := \sup_{\mu \in \mathcal{P}([0,1])} \left\{ \frac{\theta^2}{\beta} P(\sigma, \mu) - \frac{\beta}{2} D(\text{Leb}||\mu) \right\}. \tag{2.1}$$

First, let us study this problem with the following lemma:

**Lemma 2.1.** *If  $\sigma$  is bounded and continuous, the supremum is achieved in (2.1). Furthermore, in both the continuous and the piecewise cases, the function  $F$  is continuous in  $\theta$ .*

*Proof.* Let us take  $\mu_n$  a sequence of measures such that  $\frac{\theta^2}{\beta} P(\sigma, \mu_n) - \frac{\beta}{2} D(\text{Leb}||\mu_n)$  converges toward  $F(\sigma, \theta)$ . By compactness of  $\mathcal{P}([0, 1])$  for the weak topology we can assume that this sequence converges weakly to some  $\mu$ . Since we assume  $\sigma$  continuous,  $P(\sigma, \cdot)$  is continuous for the weak topology and so,  $\lim_n P(\sigma, \mu_n) = P(\sigma, \mu)$ . Furthermore, since  $(\lambda, \mu) \mapsto D(\lambda||\mu)$  is lower semi-continuous, we have  $\liminf_n D(\text{Leb}||\mu_n) \geq D(\text{Leb}||\mu)$  so that

$$\frac{\theta^2}{\beta} P(\sigma, \mu) - \frac{\beta}{2} D(\text{Leb}||\mu) \geq \limsup_n \left\{ \frac{\theta^2}{\beta} P(\sigma, \mu_n) - \frac{\beta}{2} D(\text{Leb}||\mu_n) \right\} = F(\sigma, \theta).$$

Furthermore, we have for every  $\mu \in \mathcal{P}([0, 1])$ ,  $|\Psi(\theta, \sigma, \mu) - \Psi(\theta', \sigma, \mu)| \leq \|\sigma^2\|_\infty |\theta^2 - \theta'^2|/\beta$  and so  $|F(\sigma, \theta) - F(\sigma, \theta')| \leq \|\sigma^2\|_\infty |\theta^2 - \theta'^2|/\beta$ .  $\square$

In section 4 we will prove that the following limit:

$$\lim_{N \rightarrow \infty} N^{-1} \log \mathbb{E}_{e, X_N^{(\beta)}} [\exp(N\theta \langle e, X_N^{(\beta)} e \rangle)] = F(\sigma, \theta)$$

holds in the piecewise constant case.

In the piecewise constant case, that is when  $\sigma$  is defined with a matrix  $(\sigma_{i,j})_{1 \leq i,j \leq n}$  and parameters  $\vec{\alpha}$ , the optimization problem that defines  $F$  is a simpler one. Indeed, if we denote for  $\vec{\psi} = (\psi_1, \dots, \psi_n) \in \mathbb{R}$ :

$$\vec{P}(\sigma, \psi_1, \dots, \psi_n) = \sum_{i,j=1}^n \sigma_{i,j}^2 \psi_i \psi_j$$

and

$$\vec{\Psi}(\theta, \sigma, \vec{\psi}) := \frac{\theta^2}{\beta} \vec{P}(\sigma, \psi_1, \dots, \psi_n) + \frac{\beta}{2} \left( \sum_{i=1}^n \alpha_i \log \psi_i - \sum_{i=1}^n \alpha_i \log \alpha_i \right).$$

We have easily, replacing  $\mu$  by  $\sum_{i=1}^n \alpha_i^{-1} \mu(I_i) \text{Leb}_{I_i}$  that

$$F(\sigma, \theta) = \max_{\psi_i \geq 0, \sum_1^n \psi_i = 1} \vec{\Psi}(\theta, \sigma, \vec{\psi}) \tag{2.2}$$

where  $\text{Leb}_{I_i}$  is the Lebesgue measure restricted to the interval  $I_i$ .

**2.2 Definition of the rate functions**

Now, in order to introduce our rate functions we need first to introduce the function  $J$ . This function is linked to the asymptotics of the following spherical integrals:

$$I_N(X, \theta) = \mathbb{E}_e[e^{\theta N \langle e, Xe \rangle}]$$

where the expectation holds over  $e$  which follows the uniform measure on the sphere  $\mathbb{S}^{\beta N - 1}$  of radius one (taken in  $\mathbb{R}^N$  when  $\beta = 1$  and  $\mathbb{C}^N$  when  $\beta = 2$ ). Denoting  $J_N$  the following quantity:

$$J_N(X, \theta) = \frac{1}{N} \log I_N(X, \theta)$$

the following theorem was proved in [28]:

**Theorem 2.2.** [28, Theorem 6]

If  $(E_N)_{N \in \mathbb{N}}$  is a sequence of  $N \times N$  real symmetric matrices when  $\beta = 1$  and complex Hermitian matrices when  $\beta = 2$  such that:

- The sequence of the empirical measures  $\mu_{E_N}$  of  $E_N$  weakly converges to a compactly supported measure  $\mu$ ,
- There are two reals  $\lambda_{\min}(E), \lambda_{\max}(E)$  such that  $\lim_{N \rightarrow \infty} \lambda_{\min}(E_N) = \lambda_{\min}(E)$  and  $\lim_{N \rightarrow \infty} \lambda_{\max}(E_N) = \lambda_{\max}(E)$ ,

and  $\theta \geq 0$ , then:

$$\lim_{N \rightarrow \infty} J_N(E_N, \theta) = J(\mu, \theta, \lambda_{\max}(E))$$

The limit  $J$  is defined as follows. For a compactly supported probability measure we define its Stieltjes transform  $G_\mu$  by

$$G_\mu(z) := \int_{\mathbb{R}} \frac{1}{z - t} d\mu(t)$$

We assume hereafter that  $\mu$  is supported on a compact  $[a, b]$ . Then  $G_\mu$  is a bijection from  $\mathbb{R} \setminus [a, b]$  to  $]G_\mu(a), G_\mu(b)[ \setminus \{0\}$  where  $G_\mu(a), G_\mu(b)$  are taken as the limits of  $G_\mu(t)$  when  $t \rightarrow a^-$  and  $t \rightarrow b^+$ . We denote by  $K_\mu$  its inverse and let  $R_\mu(z) := K_\mu(z) - 1/z$  be its  $R$ -transform as defined by Voiculescu in [35] (both defined on  $]G_\mu(a), G_\mu(b)[$  and  $G_\mu(a)$  and/or  $G_\mu(b)$  if they are finite). Let us denote by  $r(\mu)$  the right edge of the support of  $\mu$ .  $J$  is defined for any  $\theta \geq 0$ , and  $\lambda \geq r(\mu)$  by,

$$J(\mu, \theta, \lambda) := \theta v(\theta, \mu, \lambda) - \frac{\beta}{2} \int \log \left( 1 + \frac{2}{\beta} \theta v(\theta, \mu, \lambda) - \frac{2}{\beta} \theta y \right) d\mu(y),$$

with

$$v(\theta, \mu, \lambda) := \begin{cases} R_\mu(\frac{2}{\beta} \theta), & \text{if } 0 \leq \frac{2\theta}{\beta} \leq G_\mu(\lambda), \\ \lambda - \frac{\beta}{2\theta}, & \text{if } \frac{2\theta}{\beta} > G_\mu(\lambda), \end{cases}$$

In both the piecewise constant and the continuous case, we introduce our rate function  $I^{(\beta)}$  as

$$I^{(\beta)}(\sigma, x) = -\infty \text{ for } x \in ]-\infty, r_\sigma[$$

and

$$I^{(\beta)}(\sigma, x) = \max_{\theta \geq 0} (J(\mu_\sigma, \theta, x) - F(\sigma, \theta))$$

where  $\mu_\sigma$  is the limit measure of  $X_N^{(\beta)}$ , our Wigner matrix whose variance profile converges toward  $\sigma$ .

**Lemma 2.3.** For  $\beta = 1, 2$ ,  $I^{(\beta)}(\sigma, \cdot)$  is a good rate function. Furthermore  $I^{(2)}(\sigma, \cdot) = 2I^{(1)}(\sigma, \cdot)$

*Proof.* As a supremum of continuous functions,  $I^{(\beta)}(\sigma, \cdot)$  is lower semi-continuous. We want to prove that the level sets of  $I^{(\beta)}(\sigma, \cdot)$  that is the  $\{x \in \mathbb{R} : I^{(\beta)}(\sigma, x) \leq M\}$  are compact. It is sufficient to show that  $\lim_{x \rightarrow +\infty} I^{(\beta)}(\sigma, x) = +\infty$ . For any fixed  $\theta > 0$ , we have  $\lim_{x \rightarrow \infty} J(\mu_\sigma, \theta, x) = +\infty$ . And so since we have  $I^{(\beta)}(\sigma, x) \geq J(\mu_\sigma, \theta, x) - F(\sigma, \theta)$ ,  $I^{(\beta)}$  is a good rate function. With the change of variables  $\theta' = \theta/2$  in the case  $\beta = 2$ , we have that  $I^{(2)}(\sigma, \cdot) = 2I^{(1)}(\sigma, \cdot)$ .  $\square$

### 2.3 Assumptions on the variance profile $\sigma$

In order to prove the large deviation lower bound in the piecewise constant case, we will need the following assumption on  $\sigma$ :

**Assumption 2.4.** There exists some continuous  $\theta \mapsto (\psi_i^\theta)_{i \in \llbracket 1, n \rrbracket}$  with values in  $(\mathbb{R}^+)^n \cap \{\psi : \psi_1 + \dots + \psi_n = 1\}$  such that  $\psi^\theta$  is a maximal argument of the equation (2.2), that is:

$$\frac{\theta^2}{\beta} \bar{P}(\sigma, \psi_1^\theta, \dots, \psi_n^\theta) + \frac{\beta}{2} \left( \sum_{i=1}^n \alpha_i \log \psi_i^\theta - \sum_{i=1}^n \alpha_i \log \alpha_i \right) = F(\sigma, \theta).$$

As a more practical example, the following assumption implies Assumption 2.4:

**Assumption 2.5.** The function  $\psi \mapsto \langle \psi, \sigma^2 \psi \rangle$  is concave on the set of  $\psi \in \mathbb{R}^n$  such that  $\sum_{i=1}^n \psi_i = 1$ . Equivalently, for all  $\psi \in \mathbb{R}^n$  such that  $\sum_{i=1}^n \psi_i = 0$ ,  $\langle \psi, \sigma^2 \psi \rangle \leq 0$  (where  $\sigma^2$  is the matrix  $(\sigma_{i,j}^2)_{1 \leq i, j \leq n}$ ).

**Remark 2.6.** Examples of variance profiles that satisfies this assumption are the variance profiles associated to some parameters  $(\alpha_1, \dots, \alpha_n) \in (\mathbb{R}^{+,*})^n$  and  $\sigma_{i,j} = \mathbb{1}_{i \neq j}$  (where  $\mathbb{1}_{i \neq j}$  is the indicator function equal to 1 if  $i \neq j$  and 0 if  $i = j$ ). In the case  $n = 2$  this a linearization of a Wishart matrix as in [27].

**Lemma 2.7.** Assumption 2.5 implies Assumption 2.4.

*Proof.* The function  $\vec{\psi} \mapsto \frac{\theta^2}{\beta} \bar{P}(\sigma, \vec{\psi}) + \frac{\beta}{2} \sum_{i=1}^n \alpha_i \log \psi_i$  is strictly concave and since it tends to  $-\infty$  on the boundary of the domain, it admits a unique maximal argument  $\psi^\theta$  which is also the unique solution to the following critical point equation:

$$f(\psi) = \frac{2\theta^2}{\beta} \left( \sum_{j=1}^n \sigma_{i,j}^2 \psi_j \right)_{i=1, \dots, n} + \frac{\beta}{2} \left( \frac{\alpha_i}{\psi_i} \right)_{i=1, \dots, n} \in \text{Vect}(1, \dots, 1)$$

where  $\text{Vect}(1, \dots, 1)$  is the subspace of  $\mathbb{R}^n$  spanned by the vector whose coordinates are all 1. We now want to apply the implicit function theorem to prove that  $\theta \mapsto \psi^\theta$  is analytic. First of all, the equation above can be rewritten  $\Pi f(\psi) = 0$  where  $\Pi$  is the orthogonal projection on  $\text{Vect}(1, \dots, 1)^\perp$ . We have that for every  $u \in \mathbb{R}^n$ :

$$\forall i = 1, \dots, n, (df_\psi(u))_i = \frac{2\theta^2}{\beta} (Su)_i - \frac{\beta u_i}{2\psi_i^2}.$$

where we denote  $S = (\sigma_{i,j}^2)_{1 \leq i, j \leq n}$ .

It suffices to show that  $d(\Pi f)_\psi(u) = \Pi df_\psi(u) \neq 0$  for  $u \in \text{Vect}(1, \dots, 1)^\perp$ , that is  $df_\psi(u) \notin \text{Vect}(1, \dots, 1)$ . For such a  $u$ , we have

$$\langle u, df_\psi(u) \rangle = \frac{2\theta^2}{\beta} \langle u, Su \rangle - \frac{\beta}{2} \sum_{j=1}^n \frac{u_j^2}{\psi_j^2}$$

Since  $u \in \text{Vect}(1, \dots, 1)^\perp$  we have by Assumption 2.5  $\langle u, Su \rangle \leq 0$  and therefore  $\langle u, df_\psi(u) \rangle < 0$ . So  $df_\psi(u) \notin \text{Vect}(1, \dots, 1)$  and we can apply the implicit function theorem.  $\square$

Examples of variance profiles that satisfies Assumption 2.4 but not Assumption 2.5 are provided in section 7. In the same section, we will also show that without any assumptions on  $\sigma$ , the method used in this article may fail as we can have a large deviation principle but with a rate function different from  $I$ .

In the continuous case, we will need the following assumption:

**Assumption 2.8.** *There exists some continuous  $\theta \mapsto \psi^\theta$  (for the weak topology) from  $\mathbb{R}^+$  to  $\mathcal{P}([0, 1])$  such that  $\psi^\theta$  is a maximal argument of (2.1) that is:*

$$F(\sigma, \theta) = \Psi(\theta, \sigma, \psi^\theta)$$

As for the piecewise constant case, the following assumption implies 2.8

**Assumption 2.9.** *The function  $P(\sigma, \cdot)$  is concave on the set  $\mathcal{P}([0, 1])$  of probability measures on  $[0, 1]$ .*

**Lemma 2.10.** *Assumption 2.9 implies 2.8.*

**Remark 2.11.** A family of  $\sigma$  satisfying Assumption 2.9 is given by  $\sigma^2(x, y) = |f(x) - f(y)| + C$  where  $f$  is an increasing continuous function and  $C \in \mathbb{R}^+$ . Indeed, if  $f$  is an increasing and continuous function on  $[0, 1]$ , there is a positive measure  $\nu$  on  $[0, 1]$  such that  $f(x) - f(0) = \int_0^x d\nu(t)$  and we have  $\sigma^2(x, y) = C + |\int_x^y d\nu(t)| = \int_0^1 \tau_t(x, y) d\nu(t) + C$  where  $\tau_t(x, y) = \mathbb{1}_{x \leq t < y} + \mathbb{1}_{y \leq t < x}$  and so

$$P(\sigma, \psi) = \int_0^1 P(\tau_t, \psi) d\nu(t) + C.$$

Since  $P(\tau_t, \psi) = 2\psi([0, t])(1 - \psi([0, t]))$ ,  $P(\tau_t, \cdot)$  is concave and so is  $P(\sigma, \cdot)$ .

### 3 Scheme of the proof

The proof of Theorem 1.6 will follow a path similar to [27] for the piecewise constant case and then for  $\sigma$  continuous, we will approximate it by a sequence of piecewise constant profiles. In the piecewise constant case, we will insist on the differences with [27] and novelties brought by the introduction of a variance profile and we will refer the reader to the relevant parts of [27] for further details on the proofs that stay similar. First of all, we will prove that the sequence of distributions of the largest eigenvalue of  $X_N^{(\beta)}$  is exponentially tight.

#### 3.1 Exponential tightness

We will prove the following lemma of exponential tightness:

**Lemma 3.1.** *For  $\beta = 1, 2$ , assume that the distribution of the entries  $a_{i,j}^{(\beta)}$  satisfy Assumption 1.1. Then:*

$$\lim_{K \rightarrow +\infty} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) > K] = -\infty.$$

*Similar results hold for  $\lambda_{\min}(X_N^{(\beta)})$ .*

We will in fact prove a stronger and slightly more quantitative result that will also be useful when we will approximate continuous variance profiles using piecewise constant ones (we recall that  $\odot$  is the entrywise product of matrices):

**Lemma 3.2.** *Let  $\beta = 1, 2$  and let us assume that the distribution of the entries  $a_{i,j}^{(\beta)}$  satisfy Assumption 1.1. Let  $\mathcal{A}_N$  be the following subset of symmetric matrices:*

$$\mathcal{A}_N := \{A \in \mathcal{S}_N(\mathbb{R}^+) | \forall i, j \in \llbracket 1, N \rrbracket, A(i, j) \leq 1\}.$$

For every  $M > 0$  there exists  $B > 0$  such that:

$$\limsup_N \frac{1}{N} \sup_{A \in \mathcal{A}_N} \log \mathbb{P}[\|A \odot W_N^{(\beta)}\| \geq B] \leq -M.$$

*Proof.* We will use a standard net argument that we recall here for the sake of completeness. Let us denote:

$$Y_N^{(\beta)} := A \odot W_N^{(\beta)}.$$

Where  $A \in \mathcal{A}_N$ . If  $R_N$  is a  $1/2$ -net of  $\mathcal{S}^{\beta N-1}$  for the classical Euclidian norm, using a classical argument (see the proof of Lemma 1.8 from [27]), we have:

$$\mathbb{P}[\|Y_N^{(\beta)}\| \geq 4K] \leq 9^{\beta N} \sup_{u,v \in R_N} \mathbb{P}[\langle Y_N^{(\beta)} u, v \rangle \geq K]. \tag{3.1}$$

We next bound the probability of deviations of  $\langle Y_N^{(\beta)} v, u \rangle$  by using Tchebychev's inequality. For  $\theta \geq 0$  we indeed have

$$\begin{aligned} \mathbb{P}[\langle Y_N^{(\beta)} u, v \rangle \geq K] &\leq \exp\{-NK\} \mathbb{E}[\exp\{N \langle Y_N^{(\beta)} u, v \rangle\}] \\ &\leq \exp\{-NK\} \mathbb{E}\left[\exp\left\{\sqrt{N} \left(2 \sum_{i < j} \Re(A(i, j) a_{i,j}^{(\beta)} u_i \bar{v}_j) \right. \right. \right. \\ &\quad \left. \left. + \sum_i A(i, i) a_{i,i} u_i v_i\right)\right\}] \\ &\leq \exp\{-NK\} \exp\left(\frac{N}{\beta} \left(2 \sum_{i < j} |u_i|^2 |v_j|^2 + \sum_i |u_i|^2 |v_i|^2\right)\right) \end{aligned} \tag{3.2}$$

$$\leq \exp\left(N \left(\frac{1}{\beta} - K\right)\right) \tag{3.3}$$

where we used that the entries have a sharp sub-Gaussian Laplace transform and that  $|A(i, i)| \leq 1$ . This complete the proof of the Lemma with (3.1).  $\square$

With this result, we conclude that the sequence of the distributions of the largest eigenvalue of  $X_N^{(\beta)}$  in Lemma 3.1 is indeed exponentially tight. Therefore it is enough to prove a weak large deviation principle. In the following we summarize the assumptions on the distribution of the entries as follows:

**Assumption 3.3.** *Either the  $\mu_{i,j}^N$  are uniformly compactly supported in the sense that there exists a compact set  $K$  such that the support of all  $\mu_{i,j}^N$  is included in  $K$ , or the  $\mu_{i,j}^N$  satisfy a uniform log-Sobolev inequality in the sense that there exists a constant  $c$  independent of  $N$  such that for all smooth function  $f$ :*

$$\int f^2 \log \frac{f^2}{\mu_{i,j}^N(f^2)} d\mu_{i,j}^N \leq c \mu_{i,j}^N(\|\nabla f\|_2^2).$$

Additionally, the  $\mu_{i,j}^N$  satisfy Assumption 1.1.

### 3.2 Large deviation upper and lower bounds

To use the result of Lemma A.9 in appendix A which states convergence of the largest eigenvalue toward the edge of the support, as well as the isotropic local laws we will need the following positivity assumption (which is mainly technical and will be relaxed later by approximation):

**Assumption 3.4.** *In the piecewise constant case,  $\forall i, j \in \llbracket 1, n \rrbracket, \sigma_{i,j} > 0$ .*

We shall first prove that we have a weak large deviation upper bound similar to theorem 1.9 in [27]:

**Theorem 3.5.** *Assume that we have a piecewise constant variance profile  $\sigma$  and that Assumption 3.3 holds. Let  $\beta = 1, 2$ . Then, for any real number  $x$ ,*

$$\lim_{\delta \rightarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} \left( \left| \lambda_{\max}(X_N^{(\beta)}) - x \right| \leq \delta \right) \leq -I^{(\beta)}(\sigma, x).$$

The lower bound will however be slightly different since we need to take into account the error term  $E$ .

**Theorem 3.6.** *Assume that we are in the case of a piecewise constant variance profile  $\sigma$  and that assumptions 3.3 and 3.4 hold. Let  $E : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  be a non-negative function. Suppose that there exist continuous functions  $\theta \mapsto (\psi_i^{E,\theta})_{i \in \llbracket 1, n \rrbracket}$  such that:*

$$\vec{\Psi}(\theta, \sigma, \psi^{E,\theta}) \geq F(\sigma, \theta) - E(\theta)$$

then, if we let  $\tilde{I}(\sigma, x) := \sup_{\theta \geq 0} [J(\mu_\sigma, \theta, x) - F(\sigma, \theta) + E(\theta)]$ , we have for every  $x \geq r_\sigma$  and any  $\delta > 0$ :

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} [|\lambda_{\max}(X_N^{(\beta)}) - x| \leq \delta] \geq -\tilde{I}(\sigma, x).$$

Then, we will show that when Assumption 2.4 is verified, we can take  $E = 0$  and the main theorem follows. However, when we deal with the continuous case, we will approximate  $\sigma$  by piecewise constant functions  $\sigma^n$ . But for  $\sigma^n$  Assumption 2.4 will be verified only up to an error term  $E^n$  that can be neglected when  $n$  tends to infinity.

Proving that Lemma 3.5 is verified for  $x < r_\sigma$  is done as in [27, Corollary 1.12] using the following Lemma and saying that a deviation of  $\lambda_{\max}(X_N^{(\beta)})$  below  $r_\sigma$  imply a deviation of  $\hat{\mu}_N$  (which cannot occur with probability larger than the exponential scale we are interested in).

**Lemma 3.7.** *Assume that the  $\mu_{i,j}^N$  are uniformly compactly supported or satisfy a uniform log-Sobolev inequality. Then, for  $\beta = 1, 2$ , there exists some sequence  $\xi(N)$  converging to 0 such that*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P} (d(\hat{\mu}_N, \mu_\sigma) > \xi(N)) = -\infty.$$

with  $d$  the Dudley distance

The sequence  $\xi(N)$  in this lemma depends on the the quantities  $|I_i^{(N)}|/N$  and on how fast they converge to  $\alpha_i$ . The proof of this lemma is in Appendix A. The asymptotics of

$$J_N(X, \theta) = \frac{1}{N} \log I_N(X, \theta)$$

are given by Theorem 2.2. We will also need the following lemma, which is a result of continuity for the  $J_N$  and where we denote by  $\|A\|$  the operator norm of the matrix  $A$  given by  $\|A\| = \sup_{\|u\|_2=1} \|Au\|_2$  where  $\|u\|_2 = \sqrt{\sum |u_i|^2}$ .

**Lemma 3.8.** Given  $\mu \in \mathcal{P}([0, 1])$  a compactly supported probability measure,  $\xi : \mathbb{N} \rightarrow \mathbb{R}^+$  an arbitrary sequence tending to 0, for every  $\theta > 0$ , every  $M > 0$ , every  $\rho \geq r(\mu)$  there exists a function  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  going to 0 at 0 such that for any  $\delta > 0$ , if we denote by  $\mathcal{B}_N$  the set of real symmetric or complex Hermitian matrices  $B_N$  such that  $d(\mu_{B_N}, \mu) < \xi(N)$ ,  $|\lambda_{\max}(B_N) - \rho| < \delta$ , and  $\sup_N \|B_N\| \leq M$ , for  $N$  large enough, we have:

$$\limsup_{N \rightarrow \infty} \sup_{B_N \in \mathcal{B}_N} |J_N(B_N, \theta) - J(\mu, \theta, \rho)| \leq g(\delta).$$

*Proof.* Given that  $x \mapsto J(\mu, \theta, x)$  is continuous on  $[r(\mu), +\infty[$  (for fixed  $\theta$  and  $\mu$ ), we can choose  $g(\delta) > 0$  such that for  $|\rho - \rho'| \leq \delta$ ,  $|J(\mu, \theta, \rho) - J(\mu, \theta, \rho')| \leq g(\delta)/2$  and  $\lim_{\delta \rightarrow 0} g(\delta) = 0$ . Then, if we assume by the absurd that the lemma is false, there exists a sequence of matrices  $(A_N)_{N \in \mathbb{N}}$  such that  $\|A_N\| \leq M$ ,  $d(\mu_{A_N}, \mu_\sigma) \leq \xi(N)$ ,  $|\lambda_{\max}(A_N) - \rho| \leq \delta$  and  $|J_N(A_N, \theta) - J(\mu, \theta, \rho)| > g(\delta)$ . But then by compactness, we can assume that up to extraction  $\lim_{N \rightarrow \infty} \lambda_{\max}(A_N) = \rho' \in [\rho - \delta, \rho + \delta]$ ,  $\lim_{N \rightarrow \infty} \lambda_{\min}(A_N) = \rho'' \geq -M$  and then applying Theorem 2.2, we have that  $\lim_{N \rightarrow \infty} J_N(A_N, \theta) = J(\mu_\sigma, \theta, \rho')$ , implying  $|J(\mu, \theta, \rho) - J(\mu, \theta, \rho')| \geq g(\delta)$  which is absurd.  $\square$

Using Lemma 3.1 and Lemma 3.7, and defining

$$\mathcal{A}_{x,\delta}^M = \{X : |\lambda_{\max}(X) - x| < \delta\} \cap \{X : d(\hat{\mu}_N, \mu_\sigma) < \xi(N)\} \cap \{X : \|X\| \leq M\},$$

we have that for any  $L > 0$ , for  $M$  large enough and for  $N$  large enough

$$\mathbb{P} \left[ |\lambda_{\max}(X_N^{(\beta)}) - x| < \delta \right] = \mathbb{P} \left[ X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M \right] + O(e^{-NL}).$$

Therefore it is enough to study the probability of deviations on the set where  $J_N$  is continuous. The last item we need to this end is the asymptotics of the annealed version of the spherical integral defined by

$$F_N(\theta, \beta) = \frac{1}{N} \log \mathbb{E}_{X_N^{(\beta)}} \mathbb{E}_e [\exp(N\theta \langle e, X_N^{(\beta)} e \rangle)]$$

where  $e$  is a unit vector independent of  $X_N^{(\beta)}$  and where we take both the expectation  $\mathbb{E}_e$  over  $e$  and the expectation  $\mathbb{E}_{X_N^{(\beta)}}$  over  $X_N^{(\beta)}$ .

**Theorem 3.9.** Suppose Assumption 3.3 holds and that  $\sigma$  is a piecewise constant variance profile. For  $\beta = 1, 2$  and  $\theta \geq 0$ ,

$$\lim_{N \rightarrow \infty} F_N(\theta, \beta) = F(\sigma, \theta)$$

where we recall that  $F(\sigma, \theta)$  is defined in equation (2.1).

This counterpart to [27, Theorem 1.17] will be proven in section 4. We are now in position to get an upper bound for  $\mathbb{P} \left[ X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M \right]$ . In fact, using the result of Lemma 3.8, for any  $\theta \geq 0$ ,

$$\begin{aligned} \mathbb{P} \left[ X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M \right] &= \mathbb{E} \left[ \frac{I_N(X_N^{(\beta)}, \theta)}{I_N(X_N^{(\beta)}, \theta)} \mathbb{1}_{X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M} \right] \\ &\leq \mathbb{E} [I_N(X_N^{(\beta)}, \theta)] \exp \left\{ -N \inf_{X \in \mathcal{A}_{x,\delta}^M} J_N(X, \theta) \right\} \\ &\leq \exp \{ N(F(\sigma, \theta) - J(\mu_\sigma, \theta, x) + g(\delta) + o(1)) \} \end{aligned} \tag{3.4}$$

(where  $o(1)$  is some quantity converging to 0 as  $N \rightarrow +\infty$ ). Taking the log, dividing by  $N$  and optimizing in  $\theta \geq 0$  then gives Lemma 3.5.

To prove the complementary lower bound, we shall prove the following limit:

**Lemma 3.10.** For  $\beta = 1, 2$ , with the assumptions and notations of Theorem 3.6, for any  $x > r_\sigma$ , there exists  $\theta = \theta_x \geq 0$  such that for any  $\delta > 0$  and  $N$  large enough,

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \frac{\mathbb{E}[\mathbb{1}_{X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)}, \theta)]}{\mathbb{E}[I_N(X_N^{(\beta)}, \theta)]} \geq -E(\theta_x).$$

This lemma is proved by showing in section 5 that the matrix whose law has been tilted by the spherical integral is approximately a finite rank perturbation of a matrix with the same variance profile, from which we can use the techniques developed to study the famous BBP transition [14]. The conclusion follows since then

$$\begin{aligned} \mathbb{P}\left[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M\right] &\geq \frac{\mathbb{E}[\mathbb{1}_{X_N^\delta \in \mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)}, \theta_x)]}{\mathbb{E}[I_N(X_N^{(\beta)}, \theta_x)]} \mathbb{E}[I_N(X_N^{(\beta)}, \theta_x)] \exp\{-N \sup_{X \in \mathcal{A}_{x,\delta}^M} J_N(X, \theta_x)\} \\ &\geq \exp\{N(g(\delta) + F(\theta_x, \beta) - E(\theta_x) - J(\mu_\sigma, \theta_x, x) + o(1))\} \\ &\geq \exp\{-N(\tilde{I}_\beta(x) + o(1))\} \end{aligned}$$

where we used Theorem 3.9 and Lemma 3.10. The Theorem 1.6 follows in the case of piecewise constant variance profile satisfying Assumption 3.4 by noticing that if Assumption 2.4 is verified then we can choose  $E = 0$ . We will relax the Assumption 3.4 by approximation in the same time we will treat the continuous case.

#### 4 Proof for the asymptotics of the annealed integral in Theorem 3.9

In this section we determine that the limit of  $F_N(\theta, \beta)$  is  $F(\sigma, \theta)$  as in Theorem 3.9. In fact, we prove the following refinement of this theorem which shows that with our assumption of sharp sub-Gaussian tails, the vectors  $e$  that make the dominant contributions are delocalized.

**Proposition 4.1.** Suppose Assumption 1.1 holds. For  $j \in \llbracket 1, n \rrbracket$ , let  $\psi_j := \sum_{i \in I_j^{(N)}} |e_i|^2$  and  $V_N^\epsilon = \{e \in \mathbb{S}^{\beta N - 1} : \forall j \in \llbracket 1, n \rrbracket, \forall i \in I_j^{(N)}, |e_i| \leq \sqrt{\psi_j} N^{-1/4 - \epsilon}\}$ . Then, for  $\epsilon \in (0, \frac{1}{4})$ ,

$$\begin{aligned} F(\sigma, \theta) &= \lim_{N \rightarrow +\infty} F_N(\theta, \beta) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_e[\mathbb{1}_{e \in V_N^\epsilon} \mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta \langle e, X_N^{(\beta)} e \rangle)]] \end{aligned}$$

*Proof.* There again, the proof is very similar to the proof of Theorem 1.17 in [27]. By denoting  $L_\mu = \log T_\mu$ , we have with  $e \in \mathbb{S}^{\beta N - 1}$  fixed and by expanding the scalar product  $\langle e, X_N^{(\beta)} e \rangle$ :

$$\begin{aligned} \mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta \langle e, X_N^{(\beta)} e \rangle)] &= \exp\left\{\sum_{i < j} L_{\mu_{i,j}^N} (2\Sigma_N(i, j)\theta \bar{e}_i e_j \sqrt{N})\right. \\ &\quad \left. + \sum_i L_{\mu_{i,i}^N} (\Sigma_N(i, i)\theta |e_i|^2 \sqrt{N})\right\} \\ &\leq \exp\left\{\frac{2N\theta^2}{\beta} \sum_{i < j} \Sigma_N(i, j)^2 |e_i|^2 |e_j|^2 + \frac{N\theta^2}{\beta} \sum_i \Sigma_N(i, i)^2 |e_i|^4\right\} \end{aligned}$$

where we used the independence of the  $(a_{i,j}^{(\beta)})_{i \leq j}$  and their sub-Gaussian character. Let us recall  $\psi_j = \sum_{i \in I_j^{(N)}} |e_i|^2$  and

$$\vec{P}(\sigma, \vec{\psi}) = \sum_{i,j=1}^n \sigma_{i,j}^2 \psi_i \psi_j.$$



We deduce:

$$\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^{(\beta)}e \rangle)] \leq \exp\left(N\frac{\theta^2}{\beta}\vec{P}(\sigma, \psi_1, \dots, \psi_n)\right).$$

But since  $e$  is taken uniformly on the sphere, the vector  $\psi = (\psi_1, \dots, \psi_n)$  follows a Dirichlet law of parameters  $\left(\frac{\beta\alpha_1 N}{2}, \dots, \frac{\beta\alpha_n N}{2}\right) + o(N)$ . We have the following large deviation principle for this Dirichlet law:

**Lemma 4.2.** *Let  $n \in \mathbb{N}^*$ , and  $(\alpha_N)_{N \geq 0} = (\alpha_{1,N}, \dots, \alpha_{n,N})_{N \geq 0} \in ((\mathbb{R}^{+,*})^n)^{\mathbb{N}}$  be a sequence of vectors such that  $\lim_{N \rightarrow \infty} \frac{\alpha_N}{N} = (\alpha_1, \dots, \alpha_n)$  and  $\alpha_i > 0$  for all  $i \in \llbracket 1, n \rrbracket$ . The sequence of Dirichlet laws  $Dir_N = Dir(\alpha_{1,N}, \dots, \alpha_{n,N})$  satisfies a large deviation principle with good rate function  $I(x_1, \dots, x_n) = \sum_{i=1}^n \alpha_i(\log x_i - \log \alpha_i)$ .*

*Proof.* We denote  $f_N$  and  $f$  the functions defined on  $D = \{x \in (\mathbb{R}^{+,*})^n : x_1 + \dots + x_n = 1\}$  by

$$f_N(x) = \sum_i \frac{\alpha_{i,N} - 1}{N} \log x_i$$

$$f(x) = \sum_i \alpha_i \log x_i.$$

For  $x \in D$ , let's denote  $\tilde{x} = (x_1, \dots, x_{n-1})$  and  $\tilde{D}$  the image of  $D$  by this application (so that  $\tilde{D} = \{x \in (\mathbb{R}^{+,*})^{n-1} : x_1 + \dots + x_{n-1} \leq 1\}$ ). We have  $Dir_N(d\tilde{x}) = Z_N^{-1} \exp(Nf_N(x_1, \dots, x_{n-1}, (1 - x_1 - \dots - x_{n-1})))d\tilde{x}$  where

$$Z_N = \int_{\tilde{D}} \exp(Nf_N(x_1, \dots, x_{n-1}, (1 - x_1 - \dots - x_{n-1})))dx_1 \dots dx_{n-1}.$$

We have that on every compact of  $\tilde{D}$ ,  $f_N(\tilde{x}, 1 - \sum_{i=1}^{n-1} x_i)$  converges uniformly toward  $f(\tilde{x}, 1 - \sum_{i=1}^{n-1} x_i)$  (which is continuous) and furthermore, for every  $M > 0$  there is a compact  $K$  of  $\tilde{D}$  such that for  $N$  large enough  $f_N(\tilde{x}, 1 - \sum_{i=1}^{n-1} x_i) \leq -M$  for  $x \notin K$ . With both those remarks we deduce via a classical Laplace method that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N = \max_{x \in D} f(x) = \sum_{i=1}^n \alpha_i \log \alpha_i.$$

Using again classical Laplace methods and the fact that  $x \mapsto \tilde{x}$  is a homeomorphism between  $D$  and  $\tilde{D}$ , we have that the uniform convergence of  $f_N$  and the continuity of the limit  $f$  gives a weak large deviation principle with rate function  $f(x) - \sum_{i=1}^n \alpha_i \log \alpha_i$  and the bound outside compacts gives the exponential tightness. The large deviation principle is proved.  $\square$

Using Lemma 4.2 and Varadhan's lemma, we have since  $\vec{P}$  is continuous that:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_e \left[ \exp\left(N\frac{\theta^2}{\beta}\vec{P}(\sigma, \psi_1, \dots, \psi_n)\right) \right] =$$

$$\sup_{\psi_1, \dots, \psi_n \in [0,1], \psi_1 + \dots + \psi_n = 1} \left\{ \frac{2\theta^2}{\beta}\vec{P}(\sigma, \psi_1, \dots, \psi_n) - \sum_{i=1}^n \frac{\beta\alpha_i}{2} \log(\psi_i) - \sum_{i=1}^n \frac{\beta\alpha_i}{2} \log(\alpha_i) \right\} = F(\sigma, \theta)$$

so that we have proved the following upper bound:

$$\limsup_{N \rightarrow \infty} F_N(\theta, \beta) \leq F(\sigma, \theta). \tag{4.1}$$

For the lower bound, we then again follow [27] and we use that if  $e \in V_N^\epsilon$  then:

$$\lim_{N \rightarrow +\infty} \sup_{i,j \in [1,N]} \sup_{e \in V_N^\epsilon} |2\sqrt{N}\Sigma_N(i,j)\theta e_i e_j| = 0.$$

We can then use the Taylor expansions of  $L_{\mu_{i,j}}$  near 0 to prove that for any  $\delta > 0$ :

$$\mathbb{E}_e[\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle e, X_N^\beta e \rangle)]] \geq \mathbb{E}_e[\mathbb{1}_{e \in V_N^\epsilon} e^{N\frac{\theta^2}{\beta}\vec{P}(\sigma, \psi_1, \dots, \psi_n)(1-\delta)}]. \tag{4.2}$$

We shall then use the following lemma:

**Lemma 4.3.** *For any  $\epsilon \in (0, 1/4)$  we have*

$$\lim_{N \rightarrow \infty} \mathbb{P}_e[e \in V_N^\epsilon] = 1$$

and that the event  $\{e \in V_N^\epsilon\}$  is independent of the vector  $(\psi_1, \dots, \psi_n)$ . As a consequence, we deduce from (4.2) that for any  $\delta > 0$  and  $N$  large enough

$$\liminf_{N \rightarrow \infty} F_N(\theta, \beta) \geq \liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_e \left[ \exp \left( (1 - \delta)N \frac{\theta^2}{\beta} \vec{P}(\sigma, \psi_1, \dots, \psi_n) \right) \right].$$

and then we let  $\delta$  tends to 0. In order to see that the event  $\{e \in V_N^\epsilon\}$  is independent of  $\psi$  and to prove Lemma 4.3, we say that if we denote  $e^{(j)} = (e_i)_{i \in I_j^{(N)}}$ ,  $f^{(j)} := e^{(j)} / \|e^{(j)}\|$  is a uniform unit vector on the sphere of dimension  $\beta|I_j^{(N)}| - 1$ . Furthermore all these  $f^{(j)}$  together with the random vector  $(\psi_1, \dots, \psi_n)$  form a family of independent variables. Indeed, if we construct  $e$  as a renormalized standard Gaussian variable, that is  $e = g / \|g\|$  then one can see that  $f^{(j)} = g^{(j)} / \|g^{(j)}\|$  and  $\psi_j = \|g^{(j)}\|^2 / \|g\|^2$  where  $g^{(j)}$  is defined from  $g$  the same way as  $e^{(j)}$  is defined from  $e$ . The independence of the  $f^{(j)}$  and  $\psi$  then comes from a classical change of variables. We notice then that in term of the  $f^{(j)}$ , we have:

$$\{e \in V_N^\epsilon\} = \bigcap_{j=1}^n \{\forall i \in I_j^{(N)}, |f_i^{(j)}| \leq N^{-1/4-\epsilon}\}.$$

Then using the independence of the  $f^{(j)}$ :

$$\mathbb{P}_e[e \in V_N^\epsilon] = \prod_{j=1}^n \mathbb{P}[\forall i \in I_j^{(N)}, |f_i^{(j)}| \leq N^{-1/4-\epsilon}].$$

The result follows since each of these terms converges to 1. □

## 5 Large deviation lower bound

We will now prove Theorem 3.6. For a vector  $e$  of the sphere  $\mathbb{S}^{\beta N - 1}$  and  $X$  a random symmetric or Hermitian matrix, we denote by  $\mathbb{P}_N^{(e,\theta)}$  the tilted probability measure defined by:

$$d\mathbb{P}_N^{(e,\theta)}(X) = \frac{\exp(N\theta\langle X e, e \rangle)}{\mathbb{E}_{X_N^{(\beta)}}[\exp(N\theta\langle X_N^{(\beta)} e, e \rangle)]} d\mathbb{P}_N(X)$$

where  $\mathbb{P}_N$  is the law of  $X_N^{(\beta)}$ . Let us show that we only need to prove the following lemma:

**Lemma 5.1.** *Let us assume that the hypotheses of Theorem 3.6 hold. Let  $(\delta_N)_{N \in \mathbb{N}}$  be some arbitrary sequence of positive real numbers converging to 0,  $W_N$  the subset of the sphere  $\mathbb{S}^{\beta N - 1}$  defined by:*

$$W_N := \{e \in \mathbb{S}^{\beta N - 1} : \forall i, \left| \|e^{(i)}\|^2 - \psi_i^{E, \theta} \right| \leq \delta_N \}$$

where  $E$  and  $\psi^{E, \theta}$  are as in the hypotheses of Theorem 3.6 and  $e^{(i)} = (e_j)_{j \in I_i^{(N)}}$  is the  $i$ -th block of entries of  $e$ . For any  $x \geq r_\sigma$ , there is  $\theta_x$  such that:

$$\lim_{N \rightarrow \infty} \inf_{e \in V_N^\epsilon \cap W_N} \mathbb{P}^{(e, \theta_x)} [X_N^{(\beta)} \in \mathcal{A}_{x, \delta}^M] = 1$$

where we recall that

$$\mathcal{A}_{x, \delta}^M = \{X : |\lambda_{\max}(X) - x| < \delta\} \cap \{X : d(\hat{\mu}_N, \mu_\sigma) < \xi(N)\} \cap \{X : \|X\| \leq M\}.$$

*Proof that Lemma 5.1 implies Theorem 3.6.* In the rest of the proof, we will abbreviate  $\mathbb{E}_{X_N^{(\beta)}}$  in  $\mathbb{E}_X$ . We only need to prove that if there exists  $E, \psi_i^{E, \theta}$  that satisfy the hypotheses of Theorem 3.6, for every  $x \geq r_\sigma$ , there exists  $\theta_x \geq 0$  such that:

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \frac{\mathbb{E}_X [\mathbb{1}_{X_N^{(\beta)} \in \mathcal{A}_{x, \delta}^M} I_N(X_N^{(\beta)}, \theta_x)]}{\mathbb{E}_X [I_N(X_N^{(\beta)}, \theta_x)]} \geq -E(\theta_x).$$

We have

$$\mathbb{E}_X [\mathbb{1}_{X_N^{(\beta)} \in \mathcal{A}_{x, \delta}^M} I_N(X_N^{(\beta)}, \theta_x)] = \mathbb{E}_e [\mathbb{P}_N^{(e, \theta_x)} [X_N^{(\beta)} \in \mathcal{A}_{x, \delta}^M] \mathbb{E}_X [\exp(N\theta_x \langle X_N^{(\beta)} e, e \rangle)]].$$

For  $\delta > 0$ , let:

$$W_N^\delta := \{e \in \mathbb{S}^{\beta N - 1} : \forall i, \left| \|e^{(i)}\|^2 - \psi_i^{\theta_x, E} \right| \leq \delta\}.$$

We have, using Lemma 4.2 that:

$$\lim_{\delta \rightarrow 0} \liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}[e \in W_N^\delta] = \frac{2}{\beta} \sum_{i=1}^n \alpha_i (\log \psi_i^{\theta_x, E} - \log \alpha_i).$$

Let  $(\delta_N)_{N \in \mathbb{N}}$  be a sequence converging to 0 such that:

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{P}[e \in W_N^{\delta_N}] \geq \frac{2}{\beta} \sum_{i=1}^n \alpha_i (\log \psi_i^{\theta_x, E} - \log \alpha_i)$$

and let:

$$W_N := \{e \in \mathbb{S}^{\beta N - 1} : \forall i, \left| \|e^{(i)}\|^2 - \psi_i^{\theta_x, E} \right| \leq \delta_N\}.$$

We have then, using the Taylor expansions of the  $L_{\mu_{i,j}^{N}}$  and that  $\delta_N \rightarrow 0$  as in equation (4.2), the following limit uniformly in  $e \in W_N \cap V_N^\epsilon$ :

$$\lim_{N \rightarrow \infty} \sup_{e \in W_N \cap V_N^\epsilon} \left| \frac{1}{N} \log \mathbb{E}_X [\exp(N\theta_x \langle X_N^{(\beta)} e, e \rangle)] - \theta_x^2 \bar{P}(\sigma, \psi^{E, \theta_x}) \right| = 0.$$

Then we have:

$$\begin{aligned} \mathbb{E}_e[\mathbb{P}_N^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] \mathbb{E}_X[\exp(N\theta_x \langle X_N^{(\beta)}, e \rangle)]] &\geq \mathbb{E}_e[\mathbb{1}_{e \in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] \\ &\quad \mathbb{E}_X[\exp(N\theta_x \langle X_N^{(\beta)}, e \rangle)]] \\ &\geq \mathbb{E}_e[\mathbb{1}_{e \in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] \\ &\quad e^{-N\theta_x^2 \bar{P}(\sigma, \psi^{E,\theta_x}) + o(N)}] \end{aligned}$$

so we have that:

$$\begin{aligned} \frac{\mathbb{E}_X[\mathbb{1}_{X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M} I_N(X_N^{(\beta)}, \theta_x)]}{\mathbb{E}_X[I_N(X_N^{(\beta)}, \theta_x)]} &\geq \mathbb{E}_e[\mathbb{1}_{e \in V_N^\epsilon \cap W_N} \mathbb{P}^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] e^{-N(\theta_x^2 \bar{P}(\sigma, \psi^{E,\theta_x}) - F(\theta_x) + o(1))}] \\ &\geq \mathbb{P}_e[e \in V_N^\epsilon \cap W_N] \inf_{f \in V_N^\epsilon \cap W_N} \mathbb{P}^{(f,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] \\ &\quad e^{-N(\theta_x^2 \bar{P}(\sigma, \psi^{E,\theta_x}) - F(\theta_x) + o(1))} \\ &\geq \mathbb{P}_e[e \in V_N^\epsilon] \mathbb{P}_e[e \in W_N] \inf_{f \in V_N^\epsilon \cap W_N} \mathbb{P}^{(f,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] \\ &\quad e^{-N(\theta_x^2 \bar{P}(\sigma, \psi^{E,\theta_x}) - F(\theta_x) + o(1))} \\ &\geq e^{-N(E(\theta_x) + o(1))} \end{aligned}$$

where we used that  $\{e \in V_N^\epsilon\}$  and  $\{e \in W_N\}$  are independent (since  $\{e \in V_N^\epsilon\}$  only depends on the  $f^{(j)}$  and  $W_N$  only depends on the  $\psi_i$ ) and that  $\frac{1}{N} \log \inf_{e \in V_N^\epsilon \cap W_N} \mathbb{P}^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M]$  converges to 0. So we have our lower bound.  $\square$

And so it remains to prove the Lemma 5.1. More precisely, we will show that for  $\epsilon \in (\frac{1}{8}, \frac{1}{4})$ , for any  $x > r_\sigma$  (the rightmost point of the support of  $\mu_\sigma$ ) and  $\delta > 0$  we can find  $\theta_x \geq 0$  so that for  $M$  large enough,

$$\lim_{N \rightarrow \infty} \inf_{e \in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}[X_N^{(\beta)} \in \mathcal{A}_{x,\delta}^M] = 1. \tag{5.1}$$

To prove (5.1), we have to show that uniformly on  $e \in V_N^\epsilon \cap W_N$  we still have that  $\lim_{N \rightarrow \infty} \mathbb{P}_N^{(e,\theta)}[d(\hat{\mu}_N, \mu_\sigma) \geq \xi(N)] = 0$  and that for  $K$  large enough  $\lim_{N \rightarrow \infty} \mathbb{P}_N^{(e,\theta)}[\|X_N^{(\beta)}\| \geq K] = 0$ . This is done as in Lemma 5.1 in [27]. Hence, the main point of the proof is to show that:

**Lemma 5.2.** *Pick  $\epsilon \in ]\frac{1}{8}, \frac{1}{4}[$ . For any  $x > r_\sigma$ , there exists  $\theta_x$  such that for every  $\eta > 0$ ,*

$$\lim_{N \rightarrow \infty} \sup_{e \in V_N^\epsilon \cap W_N} \mathbb{P}_N^{(e,\theta_x)}[|\lambda_{\max} - x| \geq \eta] = 0.$$

*Proof.* For  $e \in V_N^\epsilon$  fixed, let  $X^{(e),N}$  be a matrix with law  $\mathbb{P}_N^{(e,\theta)}$ . We will prove that  $X^{(e),N}$  can be written as an additive perturbation of a random self-adjoint matrix  $\tilde{X}^{(e),N}$  with independent sub-diagonal entries with the same variance profile as  $X_N^{(\beta)}$ :

$$X^{(e),N} = \tilde{X}^{(e),N} + ESE^* + \Delta^{(e),N} + Y^{(e),N}.$$

Here  $\Delta^{(e),N} = \mathbb{E}[X^{(e),N}] - ESE^*$  and  $Y^{(e),N} = (X^{(e),N} - \mathbb{E}[X^{(e),N}]) \odot (I_N - V^{(e),N})$  where for all  $1 \leq i, j \leq N$ :

$$V_{i,j}^{(e),N} = \frac{(1 + \mathbb{1}_{i=j, \beta=1}) \Sigma_N(i, j)}{\sqrt{N \text{Var}(X_{i,j}^{(e),N})}}.$$

The deterministic matrix  $S$  is defined by  $S := (\sigma_{i,j}^2)_{1 \leq i,j \leq n}$  and  $E$  is the  $N \times n$  matrix defined by:

$$E = \begin{pmatrix} 0 & \dots & 0 \\ e^{(1)} & \vdots & \vdots \\ 0 & 0 & \dots & 0 \\ \vdots & e^{(2)} & \dots & \vdots \\ 0 & & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \\ \vdots & \vdots & & e^{(n)} \\ 0 & 0 & \dots & \end{pmatrix}$$

where we recall the  $e^{(i)}$  are the vectors defined by  $e^{(i)} = (e_j)_{j \in I_j^{(N)}}$ . In particular, one can notice that the entries of  $ESE^*$  are given by  $(ESE^*)_{i,j} = \Sigma_N^2(i,j)e_i\bar{e}_j$ . Furthermore the terms  $\Delta^{(e),N}$  and  $Y^{(e),N}$  are negligible in operator norm for large  $N$  in the sense that:

- There is a constant  $C$  such that  $\|\Delta^{(e),N}\| \leq CN^{-1/2-4\epsilon}$ .
- For every  $\delta > 0$ :

$$\lim_{N \rightarrow +\infty} \sup_{e \in V_N^\epsilon} \mathbb{P}_N^{(e),N} [\|Y^{(e),N}\| > \delta] = 0.$$

Those estimates revolve around the Taylor expansion of the  $L_{\mu_{i,j}}$  and the uniform bound on their derivatives near 0 given by Remark 1.2. Here we will only expose the computation justifying that the entries of  $\Delta^{(e),N}$  and  $Y^{(e),N}$  tend to 0. For how to refine this estimates and obtain that  $\Delta^{(e),N}$  and  $Y^{(e),N}$  are negligible in operator norm, we refer the reader to the subsection 5.1 of [27].

We can express the density of  $\mathbb{P}_N^{(e,\theta)}$  as the following product:

$$\frac{d\mathbb{P}_N^{(e,\theta)}}{d\mathbb{P}_N}(X) = \prod_{i \leq j} \exp(2^{1_{i \neq j}} \theta \Sigma_N(i,j) \sqrt{N} \Re(e_i \bar{e}_j a_{i,j}^{(\beta)}) - L_{\mu_{i,j}}(2^{1_{i \neq j}} \theta \Sigma_N(i,j) \sqrt{N} e_i \bar{e}_j))$$

where the  $a_{i,j}^{(\beta)}$  are defined as in the introduction. Since the  $a_{i,j}^{(\beta)}$  independent (for  $i \leq j$ ), the entries  $X_{i,j}^{(e),N}$  remain so and their mean is given by the derivative of  $L_{\mu_{i,j}}$ :

$$\begin{aligned} (\mathbb{E}[X^{(e),N}])_{i,j} &= \frac{\Sigma_N(i,j) L'_{\mu_{i,j}}(2\sqrt{N}\Sigma_N(i,j)\theta e_i \bar{e}_j)}{\sqrt{N}} \\ &= \frac{2\theta}{\beta} \Sigma_N^2(i,j) e_i \bar{e}_j + \delta(2\Sigma_N(i,j)\sqrt{N}\theta e_i \bar{e}_j) \frac{\Sigma_N^3(i,j) N \theta^2 |e_i|^2 |e_j|^2}{\sqrt{N}} \end{aligned}$$

if  $i \neq j$ , and if  $i = j$

$$\begin{aligned} (\mathbb{E}[X^{(e),N}])_{i,i} &= \frac{\Sigma_N(i,i) L'_{\mu_{i,i}}(\sqrt{N}\Sigma_N(i,i)\theta |e_i|^2)}{\sqrt{N}} \\ &= \frac{2\theta}{\beta} \Sigma_N^2(i,i) e_i \bar{e}_i + \delta(\sqrt{N}\Sigma_N(i,i)\theta |e_i|^2) \frac{\Sigma_N^3(i,i) N \theta^2 |e_i|^4}{\sqrt{N}} \end{aligned}$$

where we used that by centering and variance one,  $L'_{\mu_{i,j}^N}(0) = 0$ ,  $HessL_{\mu_{i,j}^N}(0) = \frac{1}{\beta}Id$  for all  $i \neq j, N$ ,  $L''_{\mu_{i,i}^N}(0) = \frac{2}{\beta}$  for all  $i, N$ , and where

$$|\delta(t)| \leq 4 \sup_{|u| < |t|} \max_{i,j,N} \{ |L_{\mu_{i,j}^N}^{(3)}(u)| \}.$$

In the complex case, the notation  $|L^{(3)}|$  just means

$$|L^{(3)}(u)| = \max_{|z|=1} \left| \frac{d^3}{dt^3} L(u + tz)_{t=0} \right|.$$

Hence, we have

$$\Delta_{i,j}^{(e),N} = \Sigma_N^3(i, j) \delta(2\sqrt{N}\Sigma_N(i, j)\theta e_i \bar{e}_j) \sqrt{N}\theta^2 |e_i|^2 |e_j|^2, 1 \leq i, j \leq N.$$

Furthermore, when we identify  $\mathbb{C}$  to  $\mathbb{R}^2$  when  $X_{i,j}^{(e),N}$  is a complex variable the covariance matrix of  $X_{i,j}^{(e),N}$  is given by the Hessian of  $L_{\mu_{i,j}}$  so that the variance of  $X_{i,j}^{(e),N}$  is given by the Laplacian of  $L_{\mu_{i,j}}$  (i.e.  $\partial_z \partial_{\bar{z}} L_{\mu_{i,j}}$ ):

$$\begin{aligned} Var(X_{i,j}^{(e),N}) &= \frac{\Sigma_N(i, j)^2}{N} \partial_z \partial_{\bar{z}} L_{\mu_{i,j}^N}(\theta \Sigma_N(i, j) e_i \bar{e}_j \sqrt{N}) \\ &= \left(\frac{2}{\beta}\right)^{\mathbf{1}_{i=j}} \frac{\Sigma_N(i, j)^2}{N} \left(1 + \sqrt{N}\Sigma_N(i, j) |e_i e_j| \delta(\theta \sqrt{N}\Sigma_N(i, j) |e_i e_j|)\right) \end{aligned}$$

And so:

$$(I_N - V^{(e),N})_{i,j} = 1 - \frac{1}{\sqrt{1 + \sqrt{N}\Sigma_N(i, j) |e_i e_j| \delta(\theta \sqrt{N}\Sigma_N(i, j) |e_i e_j|)}}$$

So that  $\lim_{N \rightarrow \infty} \max_{i,j} (I_N - V^{(e),N})_{i,j} = 0$ .

And so, to conclude we need only to identify the limit of  $\lambda_{\max}(\tilde{X}^{(e),N} + \frac{2\theta}{\beta}ESE^*)$ . The eigenvalues of  $\tilde{X}^{(e),N} + \frac{2\theta}{\beta}ESE^*$  satisfy the following equation in  $z$

$$0 = \det(z - \tilde{X}^{(e),N} - \frac{2\theta}{\beta}ESE^*) = \det(z - \tilde{X}^{(e),N}) \det(1 - \frac{2\theta}{\beta}(z - \tilde{X}^{(e),N})^{-1}ESE^*)$$

and therefore  $z$  is an eigenvalue away from the spectrum of  $\tilde{X}^{(e),N}$  if and only if

$$\det(1 - \frac{2\theta}{\beta}(z - \tilde{X}^{(e),N})^{-1}ESE^*) = 0.$$

Recall that if  $\mathbb{K}$  is a field and  $A, B$  are two matrices respectively in  $\mathcal{M}_{n,p}(\mathbb{K})$  and  $\mathcal{M}_{p,n}(\mathbb{K})$  then we have  $\det(I_n + AB) = \det(I_p + BA)$ . Using this, we have that the preceding equality is equivalent to

$$\det(I_n - \theta' E^*(z - \tilde{X}^{(e),N})^{-1}ES) = 0$$

where  $\theta' = 2\theta/\beta$ .

**Lemma 5.3.** For  $i, j \in \llbracket 1, n \rrbracket$ ,  $\eta > 0$ ,  $a > r_\sigma$ , we have:

$$\sup_{e \in V_\sigma^e \cap W_N} \mathbb{P} \left[ \sup_{z \geq a} \left| (E^*(z - \tilde{X}^{(e),N})^{-1}ES)_{i,j} - \sigma_{i,j}^2 \psi_j^\theta m_j(z) \right| \geq \eta \right] \rightarrow 0.$$

Where  $m$  is the solution of the canonical equation  $K_\sigma$  and  $m_i$  the value taken by  $m$  on the interval  $I_i$  (see Appendix A for the definition of  $K_\sigma$  and  $m$ ).

*Proof.* Let  $G_N(z) := (z - \tilde{X}^{(e),N})^{-1}$ ,  $M_N(z) = \text{diag}(m_{1,N}(z), \dots, m_{N,N}(z))$ , where  $m_x^N := \sum_{i=1}^N \mathbb{1}_{Nx \in ]i-1, i]} m_{i,N}$  and  $m^N$  is the solution of  $K_{\sigma_N}$  (the canonical equation defined in Appendix A with  $\sigma_N$  being the approximation of  $\sigma$  defined in Theorem A.4). If we denote  $\tilde{e}^k = (\mathbb{1}_{j \in I_k^{(N)}} e_j)_{j=1, \dots, N}$  the vector  $e$  where we replaced all entries by 0 except for the  $k$ -th block.

$$(E^*(z - \tilde{X}^{(e),N})^{-1}ES)_{i,j} = \sum_{k=1}^n (\tilde{e}^i)^* G_N(z) \sigma_{k,j}^2 \tilde{e}^k$$

So since  $\lim_{N \rightarrow \infty} \sup_{e \in V_N^\epsilon \cap W_N} \left| \|\tilde{e}^k\|^2 - \psi_k^\theta \right| = 0$  we only need to prove that for  $k, l \in \llbracket 1, n \rrbracket$ :

$$\lim_{N \rightarrow \infty} \sup_{e \in \mathbb{S}^{\beta N-1}} \sup_{z \geq a} \mathbb{P} [ |(\tilde{e}^k)^* G_N(z) \tilde{e}^l - \delta_{k,l} m_k(z) \psi_k^\theta| \geq \eta ] = 0.$$

To that end, we want to apply the anisotropic local law from [2] but in order to do so, we need to check its assumptions. (A) is verified since the variance profiles are uniformly bounded. (B) is verified with the Assumption 3.4. (D) is verified with the sub-Gaussian bound. To verify (C), we apply [4, Theorem 6.1]. Thanks to [2, Theorem 1.13], if we fix some  $\gamma > 0$ ,  $D > 0$ ,  $\epsilon > 0$ , for  $N$  large enough:

$$\sup_{e, f \in \mathbb{S}^{\beta N-1}} \sup_{z \in \mathbb{C}, \Im z \geq N^{\gamma-1}} \mathbb{P} [ |e^* G_N(z) f - e^* M_N(z) f| \geq N^{-1/10} ] \leq N^{-D}.$$

Furthermore following Theorem A.9, we have that for  $a' \in ]r_\sigma, a[$ ,  $D > 0$ ,  $N$  large enough

$$\mathbb{P} [ \lambda_{\max}(\tilde{X}^{(e),N}) \geq a' ] \leq N^{-D}.$$

Let  $e, f \in \mathbb{S}^{\beta N-1}$  and  $h : z \mapsto e^* G_N(z) f$  and  $k : z \mapsto e^* M_N(z) f$ . On the event  $\{ \lambda_{\max}(\tilde{X}^{(e),N}) < a' \}$ , we have that  $|h(z)|, |k(z)| \leq \frac{1}{(\Re z - a')}$  and  $|h'(z)|, |k'(z)| \leq \frac{1}{(\Re z - a')^2}$  for  $\{z : \Re z > a'\}$  and therefore, for  $\gamma < 1/10$ , we can in fact assume that our bound holds for any  $z$  such that  $\Re(z) > a$  and in particular for  $z$  real (up to some multiplicative constant  $C$  before the  $N^{-1/10}$ ). Let

$$A_N := \{ a + k/N : k \in [0, N^2] \}.$$

By union bound, we have that for  $N$  large enough:

$$\mathbb{P} [ \sup_{z \in A_N} |e^* G_N(z) f - e^* M_N(z) f| \geq CN^{-1/10} ] \leq N^{-D+2}.$$

Combining this again with the bounds of the derivative of  $h$  and  $k$  and the bound in modulus that is derived from the bound on  $\lambda_{\max}(\tilde{X}^{(e),N})$ , we get for some  $C' > 0$ :

$$\mathbb{P} [ \sup_{z > a} |e^* G_N(z) f - e^* M_N(z) f| \geq C' N^{-1/10} ] \leq N^{-D+2}$$

for  $N$  large enough. Furthermore, this bound is uniform in  $e$  and  $f$ . We then use Theorem A.7 and the bounds on the derivatives of  $M_N(z)$  the same way to conclude that for any  $\eta > 0$ , for  $N$  large enough and  $e \in V_N^\epsilon \cap W_N$  we have that (for  $N$  large enough):

$$\mathbb{P} [ \sup_{z > a} |(\tilde{e}^k)^* (G_N(z)) \tilde{e}^l - \delta_{k,l} \psi_k^\theta m_k(z)| \geq \eta ] \leq N^{-D+2}$$

where  $m$  is the solution of  $K_\sigma$  and  $m_i$  is the value taken by  $m$  on the interval  $I_i$ . And so we have:

$$\mathbb{P} [ \sup_{z > a} |(E^* G_N(z) SE)_{i,j} - m_j(z) \sigma_{i,j}^2 \psi_j^\theta| > \eta ] \leq N^{-D+2} \text{ for every } i, j \in \llbracket 1, n \rrbracket. \quad \square$$

Let's denote  $D(\theta, z)$  the diagonal  $n \times n$  matrix  $\text{diag}(m_1(z)\psi_1^\theta, \dots, m_n(z)\psi_n^\theta)$ , we have that the above limit can be rewritten  $SD(\theta, z)$ . From the preceding lemma we have that for  $\eta > 0$  uniformly in  $e \in \mathbb{S}^{\beta N-1}$  that

$$\mathbb{P}[\sup_{z>a} |\det(I_n - \theta' E^*(z - \tilde{X}^{(e),N})^{-1} ES) - \det(I_n - \theta' SD(\theta, z))| \geq \eta] \leq N^{-D}$$

for  $N$  large enough.

So since  $\lim_{z \rightarrow \infty} \det(I_n - \theta' SD(\theta, z)) = 1$ , all that remains is to solve the determinantal equation:

$$\det(I_n - \theta' SD(\theta, z)) = 0$$

and the largest solution  $z > r_\sigma$ , if it exists, will be the the limit of  $\lambda_{\max}$ . We can rewrite this equation:

$$\det(I_n - \theta' \sqrt{D(\theta, z)} S \sqrt{D(\theta, z)}) = 0. \tag{5.2}$$

Let  $\rho(\theta, z)$  be the largest eigenvalue of  $\sqrt{D(\theta, z)} S \sqrt{D(\theta, z)}$ . Then, the largest  $z$  solution of equation (5.2) is the unique solution of:

$$\theta' \rho(\theta, z) = 1 \tag{5.3}$$

one  $]r_\sigma, +\infty[$ . Indeed, with  $\theta$  fixed, if  $\theta' \rho(\theta, z) = 1$  then  $z$  is a solution of (5.2). Since the  $z \mapsto m_i(z)$  are strictly decreasing, so is  $\rho(\theta, \cdot)$ . So for  $z' > z$ ,  $\theta' \rho(\theta, z') < 1$  and so  $z'$  cannot be solution of (5.2) for the same  $\theta$ . Similarly, if  $z$  is a solution of (5.2) then  $\theta' \rho(\theta, z) \geq 1$ . If  $\theta' \rho(\theta, z) > 1$  then since  $z \mapsto \theta' \rho(\theta, z)$  is continuous and decreasing toward 0, there exists  $z' > z$  such that  $\theta' \rho(\theta, z') = 1$  and  $z'$  is therefore a solution of (5.2) strictly larger than  $z$ .

Therefore, it suffices to prove that for any  $x > r_\sigma$  there is at least one  $\theta_x$  such that

$$\theta'_x \rho(\theta_x, x) = 1$$

Here, the Assumption 2.4 is crucial. Indeed, we need this assumption to suppose that the function  $\theta \mapsto D(\theta, z)$  is continuous. This continuity implies the continuity of  $\theta \mapsto \rho(\theta, z)$ . For  $\theta = 0$  the lefthand side is 0 and for  $\theta \rightarrow \infty$ , since  $\max_i \psi_i^\theta \geq n^{-1}$  we have that

$$\rho(\theta, z) \geq \max_{i,j} (\sqrt{D(\theta, z)} S \sqrt{D(\theta, z)})_{i,j} \geq n^{-1} (\min_i m_i(z)) (\min_{i,j} \sigma_{i,j}^2)$$

Therefore since  $M := n^{-1} (\min_i m_i(x)) (\min_{i,j} \sigma_{i,j}^2)$  is such that  $\rho(\theta, x) \geq M$ , we have  $\theta' \rho(\theta, x) \xrightarrow{\theta \rightarrow \infty} +\infty$ . By continuity, there is at least one  $\theta_x$  such that  $\theta'_x \rho(\theta_x, x) = 1$  and so Theorem 3.6 is proved.  $\square$

## 6 Approximation of continuous and non-negative variance profiles

We now choose  $\sigma : [0, 1]^2 \mapsto \mathbb{R}^+$  continuous and symmetric and consider the random matrix model  $X_N^{(\beta)} := \Sigma_N \odot W_N^{(\beta)}$  where

$$\lim_{N \rightarrow \infty} \left| \Sigma_N(i, j) - \sigma\left(\frac{i}{N}, \frac{j}{N}\right) \right| = 0$$

In order to prove a large deviation principle for  $X_N$ , we will approximate the variance profile by a piecewise constant  $\sigma^n$ . Namely, for  $n \in \mathbb{N}$  we let  $\sigma^n$  be the following  $n \times n$  matrix:

$$\sigma_{i,j}^n = n^2 \int_{\frac{i-1}{n}}^{\frac{i}{n}} \int_{\frac{j-1}{n}}^{\frac{j}{n}} \sigma(x, y) dx dy + \frac{1}{n+1}.$$



The term  $1/(n + 1)$  is here to ensure that the approximating variance profile is positive so that Assumption 3.4 is satisfied. Let's denote  $X_N^{(\beta),n} = \Sigma_N^{(n)} \odot W_N^{(\beta)}$  the random matrix constructed with the same family of random variables  $a_{i,j}^{(\beta)}$  but with the piecewise constant variance profile associated with the matrix  $\sigma^n$  and the vector of parameters  $(\frac{1}{n}, \dots, \frac{1}{n})$ . Let  $F^n = F(\sigma^n, \cdot)$ ,  $\mu^n := \mu_{\sigma^n}$ . We will also denote  $F = F(\sigma, \cdot)$  and  $I = I^{(\beta)}(\sigma, \cdot)$ . Even if we suppose that Assumption 2.8 holds in the case of the continuous variance profile  $\sigma$ , we don't necessarily have Assumption 2.4 for the variance profiles  $\sigma^n$  and so we don't necessarily have a sharp lower bound. To this end we will need to introduce an error term  $E^n$  that will be negligible as  $n$  tends to  $\infty$ .

In the first subsection, we will prove that there exist for every  $n$  a function from  $\mathbb{R}^+$  to itself  $E^n$  and a function  $\theta \mapsto \psi^{\theta, E^n}$  from  $\mathbb{R}^+$  to  $\{x \in (\mathbb{R}^+)^n : x_1 + \dots + x_n = 1\}$  such that:

$$\vec{\Psi}(\sigma^n, \theta, \psi^{\theta, E^n}) = F^n(\theta) - E^n(\theta)$$

and such that  $\lim_{n \rightarrow \infty} \sup_{\theta > 0} E^n / \theta^2 = 0$ . In the second subsection, we will prove that the upper and lower large deviation bounds we get for  $X_N^{(\beta),n}$  from Theorems 3.5 and 3.6 (which will be denoted respectively  $I^{(n)}$  and  $\tilde{I}^{(n)}$ ) both converges toward the rate function defined in section 2.

### 6.1 Existence of an error negligible toward infinity

**Lemma 6.1.** *Recalling the definition for  $F$ :*

$$F(\theta) = F(\sigma, \theta) = \sup_{\mu \in \mathcal{P}([0,1])} \left\{ \frac{\theta^2}{\beta} P(\sigma, \mu) - \frac{\beta}{2} D(\text{Leb} || \mu) \right\}.$$

and recalling that  $F^n$  is defined the same way by replacing  $\sigma$  by  $\sigma^n$ , we have the following limit.

$$\lim_{n \rightarrow \infty} \sup_{\theta > 0} \frac{|F^n(\theta) - F(\theta)|}{\theta^2} = 0$$

*Proof.*

$$\begin{aligned} & \left| \int_0^1 \int_0^1 (\sigma^n)^2(x, y) d\mu(x) d\mu(y) - \int_0^1 \int_0^1 \sigma^2(x, y) d\mu(x) d\mu(y) \right| \\ & \leq \int_0^1 \int_0^1 |(\sigma^n)^2(x, y) - \sigma^2(x, y)| d\mu(x) d\mu(y). \end{aligned}$$

Since  $\lim_{n \rightarrow \infty} \sup_{x,y} |(\sigma^n)^2(x, y) - \sigma^2(x, y)| = 0$ , we have  $\lim_{n \rightarrow \infty} \sup_{\psi} |\vec{P}(\sigma, \psi) - \vec{P}(\sigma^n, \psi)| = 0$ . The result follows easily by the definitions of  $F^n$  and  $F$ .  $\square$

**Lemma 6.2.** *Let us recall the definition of  $\vec{\Psi}$ :*

$$\vec{\Psi}(\theta, \sigma, \vec{\psi}) := \frac{\theta^2}{\beta} \vec{P}(\sigma, \psi_1, \dots, \psi_n) + \frac{\beta}{2} \left( \sum_{i=1}^n \alpha_i \log \psi_i - \sum_{i=1}^n \alpha_i \log \alpha_i \right)$$

*If the Assumption 2.8 is true, then for every  $\epsilon > 0$ , there is a sequence of functions  $E^n$  and continuous  $\theta \mapsto (\psi_i^{\theta, E^n})_{i \in [1,n]}$  such that:*

$$\vec{\Psi}(\sigma^n, \theta, \psi^{\theta, E^n}) = F^n(\theta) - E^n(\theta)$$

and there is a  $n_0$  such that for  $n \geq n_0$ :

$$\forall \theta > 0, E^n(\theta) \leq \epsilon \theta^2.$$

*Proof.* Since assumption 2.8 is verified, there is some measure valued continuous function  $\theta \mapsto \psi^\theta$  such that  $F(\theta) = \theta^2 P(\sigma, \psi^\theta) / \beta - \beta D(\text{Leb} || \psi^\theta) / 2$ . Let  $\psi^{\theta, \epsilon} := K_\star(\psi^\theta \star \tau_\epsilon)$  where  $\star$  is the convolution,  $K_\star$  is the push-forward by the application  $K$ ,  $\tau_\epsilon$  the probability measure whose density is a triangular function of support  $[-\epsilon, \epsilon]$  and  $K$  the function defined by  $K(x) = x$  if  $x \in [0, 1]$ ,  $K(x) = -x$  if  $x \in [-1, 0]$  and  $K(x) = 2 - x$  if  $x \in [1, 2]$ . The function  $K$  is needed here in order for the support of  $\psi^{\theta, \epsilon}$  to still be  $[0, 1]$  since the convolution with  $\tau_\epsilon$  enlarges the support to  $[-\epsilon, 1 + \epsilon]$ . Let us now denote

$$\psi_i^{\theta, \epsilon, n} := \psi^{\theta, \epsilon} \left( \left[ \frac{i-1}{n}, \frac{i}{n} \right] \right)$$

We have that for  $i = 1, \dots, n$ :

$$\psi_i^{\theta, \epsilon, n} := \int_{\mathbb{R}} (\mathbb{1}_{[(i-1)/n, i/n]} + \mathbb{1}_{[-i/n, (1-i)/n]} + \mathbb{1}_{[2-i/n, 2-(1-i)/n]})(x) d(\psi^\theta \star \tau_\epsilon)(x).$$

So for  $i = 1, \dots, n$ :

$$\psi_i^{\theta, \epsilon, n} := \int_{\mathbb{R}} (\mathbb{1}_{[(i-1)/n, i/n]} + \mathbb{1}_{[-i/n, (1-i)/n]} + \mathbb{1}_{[2-i/n, 2-(1-i)/n]}) \star \tau_\epsilon(x) d\psi^\theta(x)$$

Since  $x \mapsto (\mathbb{1}_{[(i-1)/n, i/n]} + \mathbb{1}_{[-i/n, (1-i)/n]} + \mathbb{1}_{[2-i/n, 2-(1-i)/n]}) \star \tau_\epsilon(x)$  is continuous and  $\theta \mapsto \psi^\theta$  is continuous for the weak topology then  $\theta \mapsto \psi_i^{\theta, \epsilon, n}$  is continuous for  $i = 1, \dots, n$ .

Let us prove the following lemma:

**Lemma 6.3.** *For every  $\eta > 0$ , there is  $\epsilon > 0, n_0 > 0$  such that for every  $\theta > 0, n \geq n_0$*

$$\left| \sum_{i,j=1}^n (\sigma_{i,j}^n)^2 \psi_i^{\theta, \epsilon, n} \psi_j^{\theta, \epsilon, n} - \int_0^1 \int_0^1 \sigma^2(x, y) d\psi^\theta(x) d\psi^\theta(y) \right| \leq \eta$$

and,

$$\frac{1}{n} \sum_i^n \left( \log \psi_i^{\theta, \epsilon, n} - \log n \right) \geq -D(\text{Leb} || \psi^\theta).$$

*Proof of the lemma.* Let  $\eta > 0$  and let us find  $\epsilon > 0$  such that:

$$\left| \int_0^1 \int_0^1 \sigma^2(x, y) d\psi^{\theta, \epsilon}(x) d\psi^{\theta, \epsilon}(y) - \int_0^1 \int_0^1 \sigma^2(x, y) d\psi^\theta(x) d\psi^\theta(y) \right| \leq \eta.$$

Let us take  $X, Y, U_\epsilon, V_\epsilon$  independent random variables of law respectively,  $\psi^\theta, \psi^\theta$  and  $\tau_\epsilon, \tau_\epsilon$ . Then we have

$$\begin{aligned} \left| \int_0^1 \int_0^1 \sigma^2(x, y) d\psi^{\theta, \epsilon}(x) d\psi^{\theta, \epsilon}(y) - \int_0^1 \int_0^1 \sigma^2(x, y) d\psi^\theta(x) d\psi^\theta(y) \right| = \\ \left| \mathbb{E}[\sigma^2(K(X + U_\epsilon), K(Y + V_\epsilon)) - \sigma^2(X, Y)] \right|. \end{aligned}$$

Using the uniform continuity of  $\sigma^2$ , and that  $|K(X + U_\epsilon) - X|, |K(Y + V_\epsilon) - Y| \leq \epsilon$  almost surely, we have that there exists an  $\epsilon > 0$  such that the difference is smaller than  $\eta$ . This bound does not depend on  $\theta$ .

Now, let us find  $n_0$  such that for  $n \geq n_0$ ,

$$\left| \sum_{i,j=1}^n (\sigma_{i,j}^n)^2 \psi_i^{\theta, \epsilon, n} \psi_j^{\theta, \epsilon, n} - \int_0^1 \int_0^1 \sigma^2(x, y) d\psi^{\theta, \epsilon}(x) d\psi^{\theta, \epsilon}(y) \right| \leq \eta.$$

We have

$$\begin{aligned} & \left| \sum_{i,j=1}^n (\sigma_{i,j}^n)^2 \psi_i^{\theta,\epsilon,n} \psi_j^{\theta,\epsilon,n} - \int_0^1 \int_0^1 \sigma^2(x,y) d\psi^{\theta,\epsilon}(x) d\psi^{\theta,\epsilon}(y) \right| \\ & \leq \int_0^1 \int_0^1 |(\sigma^n(x,y))^2 - \sigma^2(x,y)| d\psi^{\theta,\epsilon}(x) d\psi^{\theta,\epsilon}(y) \end{aligned}$$

where we recall that  $(x,y) \mapsto \sigma^n(x,y)$  is the discretized version of  $\sigma$ . There again, using the uniform continuity of  $\sigma$ , we have for every  $\epsilon > 0$  the existence of  $n_0$  such that for  $n \geq n_0$ , for all  $x, y \in [0, 1]$ ,  $|(\sigma^n(x,y))^2 - \sigma^2(x,y)| \leq \eta$ . Combining these two inequalities we get the first point. Then let us show that:

$$-D(\text{Leb}||\psi^{\theta,\epsilon}) \geq -D(\text{Leb}||\psi^\theta).$$

Let  $f_\epsilon(x) = \max\{0, \epsilon^{-1} - \epsilon^{-2}|x|\}$  and

$$g_\epsilon(x,y) = f_\epsilon(x-y) + f_\epsilon(y+x) + f_\epsilon(2-x+y).$$

We have that:

$$\frac{d\psi^{\theta,\epsilon}}{dx}(x) = \int_{[0,1]} g_\epsilon(x,y) d\psi^\theta(y).$$

Let us notice that  $\int_{[0,1]} g_\epsilon(x,y) dy = \int_{[0,1]} g_\epsilon(y,x) dy = 1$ . We have

$$\begin{aligned} -D(\text{Leb}||\psi^{\theta,\epsilon}) &= \int_0^1 \log\left(\frac{d\psi^{\theta,\epsilon}}{dx}\right) dx \\ &= \int_0^1 \log\left(\int_0^1 g_\epsilon(x,y) d\psi^\theta(y)\right) dx \\ &\geq \int_0^1 \log\left(\int_0^1 g_\epsilon(x,y) \frac{d\psi^\theta(y)}{dx} dy\right) dx \\ &\geq \int_0^1 \int_0^1 g_\epsilon(x,y) \log\left(\frac{d\psi^\theta(y)}{dx}\right) dy dx \\ &\geq -D(\text{Leb}||\psi^\theta) \end{aligned}$$

where we used the concavity of  $\log$ . Finally, using again the concavity, we have for every  $i \in [1, n]$

$$\begin{aligned} n \int_{(i-1)/n}^{i/n} \log\left(\frac{d\psi^{\theta,\epsilon}(x)}{dx}\right) dx &\leq \log\left(n \int_{(i-1)/n}^{i/n} \frac{d\psi^{\theta,\epsilon}(x)}{dx}\right) \\ &\leq \log\left(n\psi_i^{\epsilon,\theta}\right). \end{aligned}$$

Summing over  $i$  gives us the result. □

Therefore, using this lemma for  $\epsilon > 0$ , there is  $\epsilon' > 0$  such that

$$\inf_{\theta \geq 0} \frac{\vec{\Psi}(\theta, \sigma^n, \psi^{\theta,\epsilon',n}) - F(\theta)}{\theta^2} > -\epsilon$$

for  $n$  large enough and so:

$$\inf_{\theta \geq 0} \frac{\vec{\Psi}(\theta, \sigma^n, \psi^{\theta, \epsilon', n}) - F^n(\theta)}{\theta^2} > -\epsilon.$$

For  $n$  large enough. Therefore, taking

$$E^n(\theta) := F^n(\theta) - \vec{\Psi}(\theta, \sigma^n, \psi^{\theta, \epsilon', n})$$

our result is proven. □

### 6.2 Convergence of large deviation bounds toward the rate function

We can now introduce  $I_\beta^n$  and  $\tilde{I}_\beta^n$  defined on  $[r_\sigma, +\infty[$  the rate functions for the upper and lower bound of the piecewise constant approximations

$$I_\beta^n(x) := \sup_{\theta \geq 0} [J(\mu_{\sigma^n}, \theta, x) - F^n(\theta)]$$

$$\tilde{I}_\beta^n(x) := \sup_{\theta \geq 0} [J(\mu_{\sigma^n}, \theta, x) - F^n(\theta) + E^n(\theta)].$$

To prove that those two functions converge toward  $I(x) = I^{(\beta)}(\sigma, \cdot)$  we will need the following result:

**Lemma 6.4.** *We recall that by definition  $\mu^n = \mu_{\sigma^n}$  and  $\mu = \mu_\sigma$ . Then*

$$\lim_{n \rightarrow \infty} \mu^n = \mu$$

and if we denote  $r^{(n)}$  the upper bound of the support of  $\mu^n$  and  $l^{(n)}$  its lower bound, we have:

$$\lim_{n \rightarrow \infty} r^{(n)} = r_\sigma$$

$$\lim_{n \rightarrow \infty} l^{(n)} = l_\sigma.$$

*Proof.* The first point is a consequence of Theorem A.5. Let  $\Delta_N^n := X_N^{(\beta)} - X_N^{(\beta),n}$ . We have  $\Delta_N^n := (\Sigma_N - \Sigma_N^{(n)}) \odot W_N^{(\beta)}$ . Using Lemma 3.2 and the fact that

$$\limsup_N \max_{i,j} |(\Sigma_N - \Sigma_N^{(n)})_{i,j}| \xrightarrow{n \rightarrow \infty} 0$$

we have that for every  $\epsilon > 0$  there is  $n_0$  such that for any  $n \geq n_0$ :

$$\mathbb{P}[\|\Delta_N^n\| > \epsilon] \xrightarrow{N \rightarrow \infty} 0$$

In particular if we denote  $\lambda_{1,N} < \dots < \lambda_{N,N}$  the eigenvalues of  $X_N^{(\beta)}$  and  $\lambda_{1,N}^{(n)} < \dots < \lambda_{N,N}^{(n)}$  these of  $X_N^{(\beta),n}$ , on the event  $\{\|\Delta_N^n\| \leq \epsilon\}$  we have  $\max_{i=1}^N |\lambda_{i,N} - \lambda_{i,N}^{(n)}| \leq \epsilon$ . And so, on this event, we have that for any  $a \in \mathbb{R}$ :

$$\#\{i \in \llbracket 1, N \rrbracket : \lambda_{i,N}^{(n)} < a - \epsilon\} \leq \#\{i \in \llbracket 1, N \rrbracket : \lambda_{i,N} < a\} \leq \#\{i \in \llbracket 1, N \rrbracket : \lambda_{i,N} < a + \epsilon\}$$

If we denote for  $t \in \mathbb{R}$ ,  $F(t) := \mu(\cdot - \infty, t)$  and  $F^{(n)}(t) := \mu^n(\cdot - \infty, t)$ , using the convergence in probability of the eigenvalue distribution, this implies that for every  $t \in \mathbb{R}$ :

$$F^{(n)}(t - 2\epsilon) \leq F(t) \leq F^{(n)}(t + 2\epsilon).$$

This then easily implies that  $|r^{(n)} - r_\sigma| \leq 2\epsilon$ . □

This result enables us to finally prove the complete version of Theorem 1.5. Indeed using the Theorem 3.7 we have that for every  $\epsilon > 0$ ,  $\mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) \leq r_\sigma - \epsilon] = o(e^{-N})$ . It suffices to show that for all  $\epsilon > 0$ ,  $\mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) \geq r_\sigma + \epsilon] = O(N^{-2})$ . In both the continuous and the piecewise constant case that does not satisfy Assumption 3.4, we can approximate  $\sigma$  by  $\sigma^n$  positive. And so the results of Lemma 6.4 hold, that is for  $n$  large enough, we have  $r^{(n)} \leq r_\sigma + \epsilon/2$ . For  $n$  large enough, we have  $\mathbb{P}[|\Delta_N^n| \geq \epsilon/2] = o(\exp(-N))$ . So we have:

$$\begin{aligned} \mathbb{P}[\lambda_{\max}(X_N^{(\beta)}) \geq r_\sigma + \epsilon] &\leq \mathbb{P}[|\Delta_N^n| \geq \epsilon/2] + \mathbb{P}[\lambda_{\max}(X_N^{(n)}) \geq r_\sigma + \epsilon/2] \\ &= O(N^{-2}) \end{aligned}$$

where we used Theorem A.9 for  $X_N^n$ . And so Theorem 1.5 is proved.

**Lemma 6.5.** • For every  $x > r_\sigma$ , the function  $\theta \mapsto J(\mu^n, \theta, x)$  converges uniformly on all compact sets of  $\mathbb{R}^+$  towards  $\theta \mapsto J(\mu, \theta, x)$ .

- For every  $x > r_\sigma$

$$I_\beta^n(x) \rightarrow I(x)$$

$$\tilde{I}_\beta^n(x) \rightarrow I(x).$$

where we recall that  $I$  is equal to  $I^{(\beta)}(\sigma, \cdot)$ .

*Proof.* For the first point of the lemma, let's first prove that for every  $x \geq r_\sigma$ ,  $\theta \mapsto J(\mu^n, \theta, x)$  converges uniformly on every compact towards the function  $\theta \mapsto J(\mu, \theta, x)$ . Let  $l < r$  be two reals. For  $\mu$  a probability measure on  $\mathbb{R}$  whose support is a subset of  $]l, r[$ , let  $Q_\mu$  be the function defined on  $\mathcal{D}_{r,\epsilon} = \{(\theta, u) \in \mathbb{R}^+ \times ]r, +\infty[ : \frac{2\theta}{\beta}(r-u) \leq 1 - \epsilon\}$  by

$$Q_\mu(\theta, u) = \int_{y \in \mathbb{R}} \log \left( 1 + \frac{2\theta}{\beta}(u-y) \right) d\mu(y)$$

$Q_\mu$  is continuous in  $(\theta, u)$  and for  $K \subset \mathcal{D}_{r,\epsilon}$  a compact we have that the function  $\mu \rightarrow Q_{\mu|_K}$  mapping  $\mu$  to the restriction of  $Q_\mu$  on  $K$  is continuous in  $\mu$  for the weak topology and  $\mu$  such that their support is a subset of  $]l, r[$  when the arrival space is the set of functions on  $K$  endowed with the uniform norm (this is a consequence of Ascoli's theorem). Let  $x > r_\sigma$  and  $r, l$  such that  $l < l_\sigma < r_\sigma < r < x$ . For  $n$  large enough the support of  $\mu^n$  is in  $]l, r[$ . We have that the sequence of functions  $\theta \mapsto v(\theta, \mu^n, x)$  converges to  $\theta \mapsto v(\theta, \mu, x)$ . Indeed if  $\frac{2\theta}{\beta} > G_\mu(x)$ , then since  $\lim_{n \rightarrow \infty} G_{\mu^n}(x) = G_\mu(x)$ ,  $\frac{2\theta}{\beta} > G_{\mu^n}(x)$  for  $n$  large enough and the result is immediate.

If  $\frac{2\theta}{\beta} < G_\mu(x)$  then  $\frac{2\theta}{\beta} < G_{\mu^n}(x)$  for  $n$  large enough.  $G_{\mu^n}$  converge towards  $G_\mu$  on  $]r, +\infty[$ , for  $\epsilon > 0$ ,  $K_{\mu^n}$  is defined on  $]0, G_\mu(r) - \epsilon[$  for  $n$  large enough and  $K_{\mu^n}$  converges toward  $K_\mu$  and therefore  $R_{\mu^n}(\frac{2\theta}{\beta})$  converges towards  $R_\mu(\frac{2\theta}{\beta})$ . For  $\theta = G_\mu(x)$  we use that  $v(\theta, \mu, x) = R_{\mu^n}(\frac{2\theta}{\beta})$  if  $\frac{2\theta}{\beta} \leq G_{\mu^n}(x)$  and  $v(\theta, \mu, x) = (x - \frac{\beta}{2\theta})$  if  $\frac{2\theta}{\beta} \geq G_{\mu^n}(x)$  and that the limits in both cases are  $v(\theta, \lambda, x)$ .

Then we have that for  $2\theta/\beta \leq G_\mu(x)$

$$\begin{aligned} \frac{2\theta}{\beta}(r - v(\theta, \mu, x)) &= \frac{2\theta}{\beta}(r - R_\mu(2\theta/\beta)) \\ &= \frac{2\theta}{\beta}(r - K_\mu(\beta/2\theta) + \beta/2\theta) \\ &\leq 1 - \frac{2\theta}{\beta}(r - K_\mu(\beta/2\theta)). \end{aligned}$$

Writing  $\frac{2\theta}{\beta} = G_\mu(y)$  with  $y > x$  we have

$$\frac{2\theta}{\beta}(r - v(\theta, \mu, x)) \leq 1 - G_\mu(y)(y - r) \leq 1 - G_\mu(x)(x - r)$$

where we used that  $y \mapsto G_\mu(y)(y - r)$  is increasing.

For  $2\theta/\beta \geq G_\mu(x)$

$$\begin{aligned} \frac{2\theta}{\beta}(r - v(\theta, \mu, x)) &= \frac{2\theta}{\beta}(r - x) + 1 \\ &\leq 1 - G_\mu(x)(x - r). \end{aligned}$$

Taking  $\epsilon > 0$  such that  $\epsilon < G_\mu(x)(x - r)$  and using the continuity in  $\mu$  of  $\mu \mapsto G_\mu(x)(x - r)$ , we have for every compact  $K' \subset \mathbb{R}^+$  and  $n$  large enough:

$$\sup_{\theta \in K'} \frac{2\theta}{\beta}(r - v(\theta, \mu^n, x)) \leq 1 - \epsilon.$$

Therefore, using the convergence of  $v(\theta, \mu^n, x)$  and the uniform convergence of  $Q_{\mu^n}$  on the compacts of  $\mathcal{D}_{r,\epsilon}$ , since:

$$J(\mu^n, \theta, x) = \theta v(\mu^n, \theta, x) - \frac{\beta}{2} Q_{\mu^n}(\theta, v(\mu^n, \theta, x))$$

we have that  $J(\mu^n, \theta, x)$  converges towards  $J(\mu_\sigma, \theta, x)$ . Furthermore, since  $\theta \mapsto J(\mu_\sigma, \theta, x)$  are continuous increasing functions, by Dini's theorem the convergence is uniform on all compact.

We now prove the convergence of  $I^n$  towards  $I$ . Let us prove that there is  $A > 0$  and  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$  and  $\theta > 0$

$$F^n(\theta) - E^n(\theta) \geq A\theta^2.$$

We have

$$\begin{aligned} F^n(\theta) &\geq \frac{\theta^2}{\beta} P(\sigma^n, Leb) \\ &\geq \frac{\theta^2}{\beta} \int_0^1 \int_0^1 \sigma^2(x, y) dx dy \end{aligned}$$

and  $-E^n(\theta) \geq \epsilon\theta^2$  for  $n$  large enough. Choosing  $\epsilon < \int_0^1 \int_0^1 \sigma^2(x, y) dx dy$  we have our result.

Then given that  $J(\mu^n, \theta, x) \leq \theta \max(r_n, x)$  we have that for any  $r > r_\sigma, x > r_\sigma, \theta > 0$  for  $n$  large enough:

$$J(\mu^n, \theta, x) - F^n(\theta) + E^n(\theta) \leq r\theta x - A\theta^2$$

Since  $\lim_{\theta \rightarrow \infty} r\theta x - A\theta^2 = -\infty$ . and that  $\theta \mapsto J(\mu^n, \theta, x) - F^n(\theta) + E^n(\theta)$  converges toward  $\theta \mapsto J(\mu_\sigma, \theta, x) - F(\theta)$  on every compact of  $\mathbb{R}^+$ , we deduce that for every  $x > r_\sigma$ :

$$\lim_{n \rightarrow \infty} \tilde{I}_\beta^n(x) = I(x)$$

and in the same way with  $E^n = 0$ :

$$\lim_{n \rightarrow \infty} I_\beta^n(x) = I(x). \quad \square$$

### 6.3 Conclusion

We will now prove that the difference between  $X_N^{(\beta),n}$  and  $X_N^{(\beta)}$  is negligible at the exponential scale.

**Lemma 6.6.** *For every  $\epsilon > 0$  and every  $A > 0$ , there exists some  $n_0 \in \mathbb{N}$  such that for  $n \geq n_0$ :*

$$\limsup_N \frac{1}{N} \log \mathbb{P}[\|X_N^{(\beta),n} - X_N^{(\beta)}\| \geq \epsilon] \leq -A$$

*Proof.* We can write that

$$X_N^{(\beta),n} - X_N^{(\beta)} = \Delta_N^n \odot W_N^{(\beta)}$$

where

$$\Delta_N^n = \Sigma_N^{(n)} - \Sigma_N$$

Let

$$M_n := \sup_{i,j} |(\Delta_N^n)_{i,j}|$$

We have that:

$$\lim_{n \rightarrow \infty} M_n = 0$$

Following Lemma 3.2, we can write that for every  $n \in \mathbb{N}$ ,  $A > 0$  there is  $B > 0$  such that

$$\limsup_N \frac{1}{N} \log \mathbb{P}[(M_n)^{-1} \|X_N^{(\beta),n} - X_N^{(\beta)}\| > B] \leq -A.$$

For  $n_0 \in \mathbb{N}$  such that  $M_n B \leq \epsilon$  for all  $n \geq n_0$ , our upper bound is verified.  $\square$

Therefore, since both  $I_\beta^{(n)}(x)$  and  $\tilde{I}_\beta^{(n)}(x)$  converge toward  $I_\beta(x)$ , we have a weak large deviation principle with rate function  $I_\beta$ . Furthermore since we also have exponential tightness, we have that Theorem 1.6 holds.

It only remains to relax the positivity assumption 3.4 for the piecewise constant case. Let  $\sigma$  be a piecewise variance profile. We can approximate  $\sigma$  by  $\sigma^n := \sqrt{\sigma^2 + \frac{1}{n+1}}$ . We notice then that with this choice of  $\sigma^n$ :

$$\vec{\Psi}(\theta, \sigma^n, \vec{\psi}) = \frac{\theta^2}{(n+1)\beta} + \vec{\Psi}(\theta, \sigma, \vec{\psi})$$

so that if 2.4 is verified for  $\sigma$ , it is verified for  $\sigma^n$ . And so, as we have just done for the continuous case, we can prove the same way that the rate functions  $I(\sigma^n, \cdot)$  converges to  $I(\sigma, \cdot)$  and that the large deviation principle holds with  $I(\sigma, \cdot)$ .

## 7 The case of matrices with $2 \times 2$ block variance profiles

In this section, we will discuss the case of piecewise constant variance profiles with 4 blocks (which are not necessarily of equal sizes) and determine what are the cases where the Assumption 2.4 holds. In particular, we will provide examples where the maximum argument of Assumption 2.4 can be taken continuous without the need for the concavity assumption.

Let's take a piecewise constant variance profile defined by  $\vec{\alpha} = (\alpha, 1 - \alpha)$  and  $\sigma_{1,1} = a, \sigma_{2,2} = b, \sigma_{1,2} = \sigma_{2,1} = c$ . In order to apply Theorem 1.6 we need to study the maximum argument for  $\theta$  fixed of:

$$\psi(x, \theta) = \vec{\Psi}(\sigma, \theta, (x, (1-x))) = \frac{\theta^2}{\beta} [a^2 x^2 + b^2 (1-x)^2 + 2c^2 x(1-x)] + \frac{\beta}{2} [\alpha \log x + (1-\alpha) \log(1-x)].$$

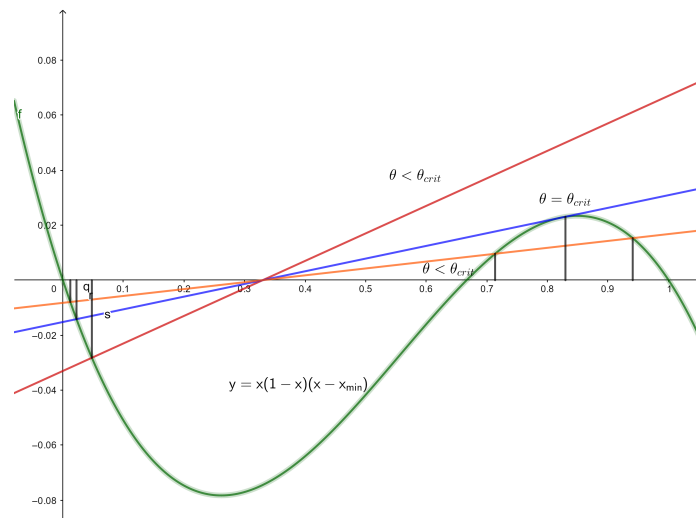


Figure 1: Case  $x_{\min} \geq 1/2$

Since we can change  $\alpha$  in  $1 - \alpha$  by switching  $a$  and  $b$ , we can suppose without loss of generality that  $\alpha \leq 1/2$ .

We have

$$\partial_x \psi(x, \theta) := \frac{2\theta^2}{\beta} [a^2 x - b^2(1-x) + c^2(1-2x)] + \frac{\beta}{2} \left( \frac{\alpha}{x} - \frac{1-\alpha}{1-x} \right).$$

Let  $\Phi(x, \theta) := x(1-x)\partial_x \psi(x, \theta)$ , so that  $\Phi(\cdot, \theta)$  vanishes at the critical points of  $\psi(\cdot, \theta)$ . We have that:

$$\Phi(x, \theta) = (a^2 + b^2 - 2c^2) \frac{2\theta^2}{\beta} x(x - x_{\min})(1-x) + \frac{\beta}{2} (\alpha - x).$$

where

$$x_{\min} := \frac{c^2 - b^2}{2c^2 - a^2 - b^2}$$

$$x_\theta = \operatorname{argmax}_{x \in [0,1]} \psi(x, \theta).$$

### 7.1 Case with $(a^2 + b^2 - 2c^2) \leq 0$

In the case  $(a^2 + b^2 - 2c^2) \leq 0$  we have the  $\psi(\cdot, \theta)$  is strictly concave and therefore  $\theta \mapsto x_\theta$  is analytical and assumption 2.4 is satisfied and the large deviation principle applies.

From now on, we assume  $(a^2 + b^2 - 2c^2) > 0$ .

### 7.2 Case $x_{\min} \geq 1/2$

We look for the zeros of  $\Phi(\cdot, \theta)$  in  $[0, 1]$ . To this end, we look for the intersection points of the curve of equation  $y = x(1-x)(x - x_{\min})$  and the line  $y = A_\theta(x - \alpha)$  where  $A_\theta = \frac{\beta^2}{4\theta^2(a^2 + b^2 - 2c^2)}$ .

We notice that there is a critical value  $\theta_{crit}$  such that for  $\theta \leq \theta_{crit}$ , there is only one critical point  $x_1^\theta$  which is on  $]0, 1/2[$ . For  $\theta > \theta_{crit}$  we have the apparition of two other



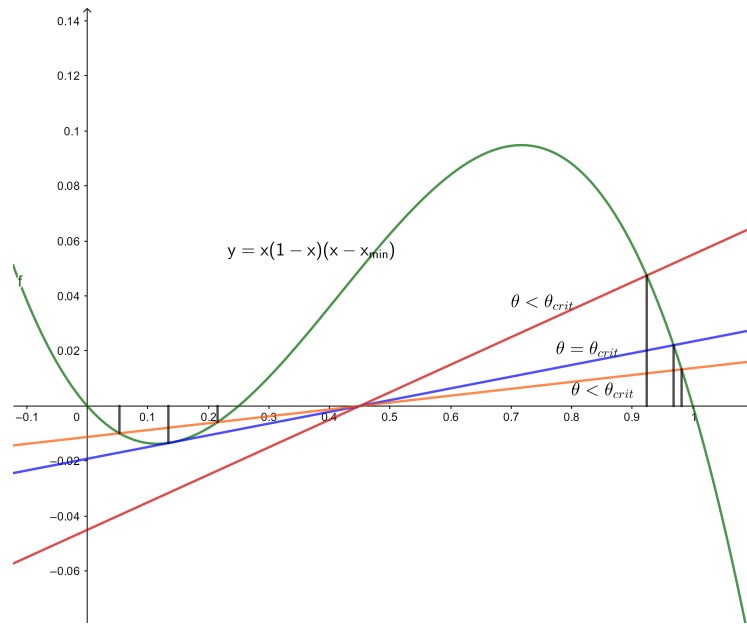


Figure 2: Case  $x_{\min} \leq \alpha$

critical points  $x_2^\theta$  and  $x_3^\theta$  that are such that  $1/2 < x_2^\theta < x_3^\theta$  with  $\psi(x_2^\theta, \theta)$  being a local minimum and  $\psi(x_3^\theta, \theta)$  a local maximum. For  $x \in ]0, 1[$ , we have:

$$\psi(x, \theta) - \psi(1 - x, \theta) = \frac{\beta}{2}(1 - 2\alpha)(\log(1 - x) - \log x) + \frac{\theta^2}{\beta}(a^2 + b^2 - 2c^2)[(x - x_{\min})^2 - (1 - x - x_{\min})^2].$$

For  $x < 1/2$ , we have  $\psi(x, \theta) > \psi(1 - x, \theta)$  and so the absolute maximum of  $\psi(\cdot, \theta)$  is attained at  $x = x_1^\theta$ . Furthermore, we notice the line  $y = A_\theta(x - \alpha)$  is never tangent to the graph  $y = x(x - x_{\min})(1 - x)$  in the point of first coordinate  $x_1^\theta$ , we have  $\partial_x \Phi(x_1^\theta, \theta) \neq 0$ . Now using the implicit function theorem, we have  $\theta \mapsto x_1^\theta$  is analytical and so Assumption 2.4 is verified.

### 7.3 Case $x_{\min} \leq \alpha$

There is again a critical value  $\theta_{crit}$  such that for  $\theta \leq \theta_{crit}$ , there is only one critical point  $x_1^\theta$  which is on  $] \alpha, 1[$  and for  $\theta > \theta_{crit}$  we have the apparition of two other critical points  $x_2^\theta$  and  $x_3^\theta$  such that  $x_3^\theta < x_2^\theta \leq \alpha$ . We have furthermore:

$$\begin{aligned} \psi(x, \theta) - \psi(2\alpha - x, \theta) &= \frac{\beta}{2}[\alpha \log x - (1 - \alpha) \log(1 - 2\alpha + x)] + \frac{\beta}{2}[(1 - \alpha) \log(1 - x) - \alpha \log(2\alpha - x)] \\ &\quad + \frac{\theta^2}{2}(a^2 + b^2 - 2c^2)[(x - x_{\min})^2 - (2\alpha - x - x_{\min})^2]. \end{aligned}$$

For  $x < \alpha$ ,  $\psi(x, \theta) < \psi(2\alpha - x, \theta)$  and there the absolute maximum of  $\psi(\cdot, \theta)$  is attained on  $] \alpha, 1[$ , so for  $x_1^\theta$ . Since  $\theta \mapsto x_1^\theta$  is analytical, Assumption 2.4 is verified.

**7.4 Case  $\alpha = 1/2$  and  $x_{\min} = 1/2$**

Then  $x \mapsto \psi(x, \theta)$  is symmetrical in  $1/2$ . Looking at the zeros of  $\Phi(\cdot, \theta)$  we have that if we set  $\theta_{crit} := \beta\sqrt{2/(a^2 + b^2 - 2c^2)}$  for  $\theta \leq \theta_{crit}$  there is only one zero  $x = 1/2$  and for  $\theta \geq \theta_{crit}$  there is three zeros in  $x = 1/2$  and  $x = \frac{1}{2} \pm \delta(\theta)$  where  $\delta(\theta) = \frac{\beta}{2\theta} \sqrt{\frac{2}{(a^2 + b^2 - 2c^2)}}$ . Furthermore, for  $\theta \leq \theta_{crit}$ ,  $\psi(\cdot, \theta)$  has its maximum in  $x = 1/2$  and for  $\theta \geq \theta_{crit}$ ,  $\psi(\cdot, \theta)$  has its maximum at both points  $x = 1/2 \pm \delta(\theta)$  where  $\delta(\theta) = \frac{1}{2} \sqrt{1 - \frac{2\beta^2}{\theta^2(a^2 + b^2 - 2c^2)}}$ . Therefore the function  $\theta \mapsto 1/2$  for  $\theta \leq \theta_{crit}$  and  $\theta \mapsto 1/2 + \delta(\theta)$  for  $\theta \geq \theta_{crit}$  is a continuous determination of the maximal argument of  $\psi(\cdot, \theta)$  and so Assumption 2.4 is verified and the large deviation principle holds. This gives an example where the maximum argument in Assumption 2.4 is neither unique nor  $C^1$  but where we can still derive a large deviation principle.

**7.5 Case  $x_{\min} \in ]\alpha, 1/2[$  and pathological cases**

The graph we obtain is similar to the graph of the first case. In this case, we also have a  $\theta_{crit}$  such that for  $\theta \leq \theta_{crit}$ , there is only one critical point  $x_1^\theta$  which is in  $]0, \alpha[$  and then the apparition of two other critical points  $x_2^\theta$  and  $x_3^\theta$  that are such that  $1/2 < x_2^\theta < x_3^\theta$ ,  $\psi(x_2^\theta, \theta)$  being a local minimum and  $\psi(x_3^\theta, \theta)$  a local maximum. But in this case for high values of  $\theta$ , we have that the maximum is attained near 1 and so for these high values  $x_3^\theta$  is the maximum argument. We have a discontinuity in the maximum argument and so Assumption 2.4 is not verified.

Let us now show that if  $x_{\min} \in ]\alpha, 1/2[$  and  $c = 0$ , the largest eigenvalue satisfies a large deviation principle but with a rate function  $J$  different from  $I$ .

Our matrix  $X_N^{(\beta)}$  then looks like:

$$\begin{pmatrix} \sqrt{a\alpha}T_{\alpha N}^{(\beta)} & 0 \\ 0 & \sqrt{b(1-\alpha)}U_{(1-\alpha)N}^{(\beta)} \end{pmatrix}$$

where  $(T_N^{(\beta)})_N$  and  $(U_N^{(\beta)})_N$  are independent Wigner matrices. We have:

$$\lambda_{\max}(X_N^{(\beta)}) = \max\{\sqrt{a\alpha}\lambda_{\max}(T_{\alpha N}^{(\beta)}), \sqrt{b(1-\alpha)}\lambda_{\max}(U_{(1-\alpha)N}^{(\beta)})\}.$$

But both these quantities satisfy large deviation principles, more precisely, if  $I_\beta$  is the rate function for the large deviation principle of [27] for a Wigner matrix,  $\lambda_{\max}(\sqrt{a\alpha}T_{\alpha N}^{(\beta)})$  follows a large deviation principle with rate function  $\alpha I_\beta(\frac{x}{\sqrt{a\alpha}})$  and  $\lambda_{\max}(\sqrt{b(1-\alpha)}U_{(1-\alpha)N}^{(\beta)})$  follows a large deviation principle with rate function  $(1-\alpha)I_\beta(\frac{x}{\sqrt{b(1-\alpha)}})$ . Now  $\lambda_{\max}(X_N^{(\beta)})$  follows a large deviation principle with rate function  $J^\beta$  which is:

$$J^\beta(x) := \min\{\alpha I_\beta(\frac{x}{\sqrt{a\alpha}}), (1-\alpha)I_\beta(\frac{x}{\sqrt{b(1-\alpha)}})\}$$

In particular, if we choose  $\alpha, a, b$  such that  $a\alpha > b(1-\alpha)$  and  $b > a$ , we notice that  $J^\beta(x) = \alpha I_\beta(\frac{x}{\sqrt{a\alpha}})$  for  $x$  near  $a\alpha$  and  $J^\beta(x) = (1-\alpha)I_\beta(\frac{x}{\sqrt{b(1-\alpha)}})$  for large  $x$ . In this case one can notice that  $J^\beta$  is not a convex function and therefore cannot be obtained as  $\sup_\theta \{J(x, \mu_\sigma, \theta) - F(\theta)\}$  since it is a sup of convex functions. We have  $J \neq I$ . For  $c > 0$  but small enough we can also conclude that the large deviation principle still does not hold. Indeed, if we denote  $I_c$  the rate function we expect using the formula of section 2. Since  $I_0$  still provides a large deviation upper bound, we have  $J \geq I_0$  and so let  $x_0 \in \mathbb{R}^+$  such that  $J(x_0) \geq I_0(x_0) + \eta$  for some  $\eta > 0$  ( $x_0$  does exist since  $J \neq I_0$ ). Using the same

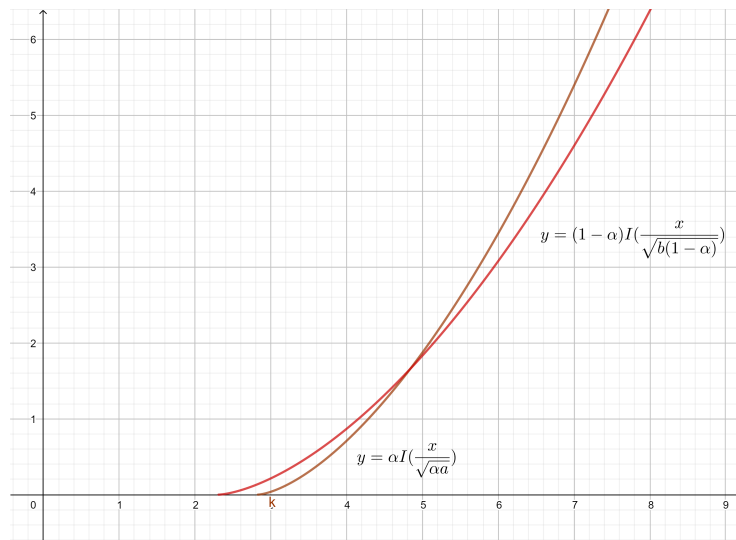


Figure 3: Graph of the rate function  $\alpha I(x/\sqrt{a\alpha})$  and  $(1-\alpha)I(x/\sqrt{b(1-\alpha)})$  with  $\alpha = 2/3$ ,  $a = 3$ ,  $b = 4$

approximation arguments as in section 6, we have that there exists  $\epsilon > 0$  such that for  $c < \epsilon$ ,  $I_c(x_0) < I_0(x_0) + \eta/3$  and:

$$\lim_{\delta \rightarrow 0} \limsup_N -\frac{1}{N} \log \mathbb{P}[\lambda_{\max}(X_N) \in [x_0 - \delta, x_0 + \delta]] \geq J(x_0) - \eta/3 \geq I_0(x_0) + 2\eta/3 \geq I_c(x_0) + \eta/3$$

Since  $I_c$  is continuous in  $x_0$ , we have that there cannot be a large deviation principle with the rate function  $I_c$ .

### 8 Looking for an expression of the rate function

In this section we will present a method to explicitly compute the rate function  $I$  in the piecewise constant case under some hypothesis on the behavior of  $F$ . First, let us describe  $F$  in a neighbourhood of  $\theta = 0$ .

**Theorem 8.1.** *Let  $\sigma$  be a continuous or piecewise constant variance profile, there is  $\theta_0 > 0$  such that for  $\theta \leq \theta_0$ :*

$$F(\sigma, \theta) = \frac{\beta}{2} \int_0^{\frac{2}{\beta}\theta} R(w)dw$$

Where  $R$  is the  $R$ -transform of the measure  $\mu_\sigma$ .

*Proof.* This result was proved in the case of a linearisation of a Wishart matrix in section 4.2 of [27]. For the sake of completeness, we will reproduce here this proof. For the lower bound, we have for  $M > r_\sigma$  and  $2\theta/\beta \leq G(M)$  (where  $G$  is the Stieltjes transform of the measure  $\mu_\sigma$ ):

$$F(\sigma, \theta) \geq \liminf_{N \rightarrow \infty} \frac{1}{N} \mathbb{E}_{X_N} [\mathbb{1}_{\lambda_{\max}(X_N) < M} I_N(X_N, \theta)] \geq \frac{\beta}{2} \int_0^{\frac{2}{\beta}\theta} R(w)dw$$

This is due to the fact that for  $2\theta/\beta \leq G(x)$ ,  $J(\theta, x, \mu) = \frac{\beta}{2} \int_0^{\frac{2\theta}{\beta}} R(\theta)d\theta$ . For the upper bound, we write:

$$\begin{aligned}
 \mathbb{E}_{X_N}[I_N(X_N, \theta)] &\leq \sum_{n=1}^{+\infty} \mathbb{E}_{X_N}[\mathbb{1}_{(n-1)M \leq \lambda_{\max}(X_N) < nM} I_N(X_N, \theta)] \\
 &\leq \mathbb{E}_{X_N}[\mathbb{1}_{\lambda_{\max}(X_N) < M} I_N(X_N, \theta)] + \sum_{n=2}^{+\infty} \mathbb{E}_{X_N}[\mathbb{1}_{(n-1)M \leq \lambda_{\max}(X_N) < nM} I_N(X_N, \theta)] \\
 &\leq \mathbb{E}_{X_N}[\mathbb{1}_{\lambda_{\max}(X_N) < M} I_N(X_N, \theta)] + \sum_{n=2}^{+\infty} \exp(-K(n-1)M + nNM\theta) \\
 &\leq \mathbb{E}_{X_N}[\mathbb{1}_{\lambda_{\max}(X_N) < M} I_N(X_N, \theta)] + e^{N\theta M} \frac{\exp(M(\theta - K)N)}{1 - \exp((M\theta - K)N)}
 \end{aligned}$$

Where we used that for  $N$  large enough, we have for every  $N$ ,  $\mathbb{P}[\lambda_{\max}(X_N) \geq M] \leq \exp(-KM)$  for some  $K > 0$  and that for  $\lambda_{\max}(X_N) \leq M$ ,  $I_N(X_N, \theta) \leq e^{M\theta}$ . Now, by choosing  $\theta$  small enough such that  $(2\theta - K) < 0$ , we have the upper bound.  $\square$

The main results of this section is the following:

**Theorem 8.2.** *If the function  $\theta \mapsto F(\sigma, \theta)$  is analytic, then the  $R$  transform of  $\mu_\sigma$  has an analytic extension on  $\mathbb{R}^+$  and then the rate function  $I(\sigma, \cdot)$  only depends on  $\mu_\sigma$ .*

*Proof.* Since  $F(\sigma, \cdot)$  is analytic and so is  $R$  and since we have  $F'(\sigma, \theta) = R(\frac{2\theta}{\beta})$  for small  $\theta$ ,  $F$  depends only on  $R$  that is on  $\mu_\sigma$  and  $F'(\frac{\beta x}{2})$  extends  $R$  on  $\mathbb{R}^+$ . Then, looking at the expression of  $I(\sigma, \cdot)$ , it only depends on  $\mu_\sigma$ .  $\square$

**Remark 8.3.** Without any condition on the variance profile  $\sigma$ , we do not have that  $I(\sigma, \cdot)$  depends only on  $\mu_\sigma$ . For instance if we take  $X_N$  and  $X'_N$  independent matrices both with the same variance profile  $\sigma$ ,  $\alpha, \beta > 0$  such that  $\alpha > \beta$  and  $\alpha + \beta = 1$ , then the following matrix has a variance profile:

$$Y_N = \begin{pmatrix} X_{\alpha N} & 0 \\ 0 & X'_{\beta N} \end{pmatrix}.$$

And then  $\lambda_{\max}(Y_N) = \max(\lambda_{\max}(X_{\alpha N}), \lambda_{\max}(X'_{\beta N}))$ . We have that  $\lambda_{\max}(Y_N)$  satisfy a large deviation principle with rate function  $\beta I(\sigma, \cdot)$  whereas this matrix has for limit measure  $\mu_\sigma$  whatever the choice of  $\beta$ .

In the case of a piecewise constant variance profile, the same concavity hypothesis as before implies the analyticity of the function  $F(\sigma, \cdot)$  (this is due to the fact that with the implicit function theorem, the maximum argument is indeed analytic in  $\theta$ ).

**Proposition 8.4.** *If the Assumption 2.5 holds in the case of a piecewise constant case then the function  $\theta \mapsto F(\sigma, \theta)$  is analytic.*

We will now shortly discuss how we can obtain an explicit expression of the rate function in the context of a piecewise constant variance profile which satisfies the hypothesis of Theorem 8.2. For this, we will need the following proposition:

**Proposition 8.5.** *If the hypothesis of Theorem 8.2 are verified and if  $\theta \mapsto R(\theta) + \frac{1}{\theta}$  is strictly increasing on  $[G(r_\sigma), +\infty[$ , then we have:*

$$I(\sigma, x) = \frac{\beta}{2} \int_{r_\sigma}^x (\bar{G}(u) - G(u)) du.$$

where we analytically extended  $R$  on  $\mathbb{R}^+$ , where  $G(r_\sigma) = \lim_{x \rightarrow r_\sigma^+} G(x)$  and where  $\bar{G}$  is the inverse function of  $\theta \mapsto R(\theta) + \frac{1}{\theta}$  on  $[r_\sigma, +\infty[$ .

*Proof.* for  $x \geq r_\sigma$ , we have that:

$$J(\mu_\sigma, \theta, x) - F(\sigma, \theta) = \begin{cases} 0 & \text{for } 2\theta/\beta \leq G(x) \\ \theta x - \frac{\beta}{2} - \frac{\beta}{2} \left( \log \left( \frac{2\theta}{\beta} \right) \right) - \frac{\beta}{2} \int \log(x - y) d\mu_\sigma(y) - \frac{\beta}{2} \int_0^{\frac{2\theta}{\beta}} R(w) dw & \text{for } 2\theta/\beta \geq G(x). \end{cases}$$

Differentiating in  $\theta$ , we have:

$$\frac{\partial}{\partial \theta} \left( J(\mu_\sigma, \theta, x) - F(\sigma, \theta) \right) = \begin{cases} 0 & \text{for } 2\theta/\beta \leq G(x) \\ x - \frac{\beta}{2\theta} - R\left(\frac{2\theta}{\beta}\right) & \text{for } 2\theta/\beta \geq G(x) \end{cases} .$$

And so, the maximum is realized for  $\theta_x > \beta G(x)/2$  such that  $x = \frac{\beta}{2\theta_x} + R\left(\frac{2\theta_x}{\beta}\right)$ . By hypothesis, this is equivalent to  $\theta_x = \frac{\beta}{2} \bar{G}(x)$ . And so we have for  $x > r_\sigma$

$$\frac{\partial}{\partial x} I(\sigma, x) = \theta_x - \frac{\beta}{2} G(x)$$

We deduce the result by integrating. □

**Remark 8.6.** In practice, in the case of a piecewise constant variance profile the equations of section A give that  $G(z)$  is an algebraic function, that is a root of some polynomial  $P(\cdot, z)$ . So we have, for  $\theta \leq G(r_\sigma)$ ,  $P(\theta, R(\theta) + \theta^{-1}) = 0$ . Using the analytical extension of  $R$  on  $\mathbb{R}^+$ , this stays true for any  $\theta > 0$  and therefore  $P(\bar{G}(z), z) = 0$ . So  $\bar{G}$  naturally presents itself as a conjugate root of  $G(z)$ . For example, in the Wigner case we have  $G(z) = \frac{z - \sqrt{z^2 - 4}}{2}$  and  $\bar{G}(z) = \frac{z + \sqrt{z^2 - 4}}{2}$ , and we end up with  $I(x) = \frac{\beta}{2} \int_2^x \sqrt{u^2 - 4} du$ . In the case of the linearisation of of Wishart matrix (see section 4.2 of [27]), we have:

$$G(x) = \frac{2\alpha}{1 + \alpha} \frac{x^2 + 1 - \alpha - \sqrt{(x^2 - 1 - \alpha)^2 - 4\alpha}}{2x} + \frac{1 - \alpha}{1 + \alpha} \frac{1}{x}$$

and

$$\bar{G}(x) = \frac{2\alpha}{1 + \alpha} \frac{x^2 + 1 - \alpha + \sqrt{(x^2 - 1 - \alpha)^2 - 4\alpha}}{2x} + \frac{1 - \alpha}{1 + \alpha} \frac{1}{x}$$

and so we have  $I(x) = \frac{\beta\alpha}{1 + \alpha} \int_{r_\sigma}^x \frac{\sqrt{(u^2 - 1 - \alpha)^2 - 4\alpha}}{2u} du$ .

## A Appendix: The limit of the empirical measure

In this section, we describe the limit of the empirical measure  $\mu_\sigma$  of the matrices  $X_N$ . We will also discuss the stability of this measure in function of  $\sigma$ . Under assumptions of positivity for the variance profile, we will prove that the largest eigenvalue converges toward the rightmost point of the support of  $\mu_\sigma$ . We denote  $\mathbb{H}^+$  the complex upper half-plane  $\{z \in \mathbb{C} : \Im z > 0\}$  and  $\mathbb{H}^-$  the complex lower half-plane  $\{z \in \mathbb{C} : \Im z < 0\}$ . To describe the limit of the empirical measure we need the following definition for the so-called canonical equation (also called quadratic vector equation). The following definition takes into account both the piecewise constant and the continuous case:

**Definition A.1.** Let  $\sigma : [0, 1]^2 \rightarrow \mathbb{R}^+$  be a bounded symmetric measurable function. We call canonical equation  $K_\sigma$  the following equation where  $m$  is a function from  $\mathbb{H}^+$  into  $\mathcal{H}$ , where  $\mathcal{H}$  is the set of measurable and bounded functions  $m$  from  $[0, 1]$  to  $\mathbb{H}^-$ ,

$$\frac{1}{m(z)} = z - Sm(z) \tag{K_\sigma}$$

Here  $S$  is the following kernel operator for  $f \in \mathcal{H}$ :

$$Sf(x) := \int_0^1 \sigma^2(x, y) f(y) dy$$

If  $w$  is a function from  $[0, 1]$  to  $\mathbb{C}$ , we denote  $\|w\| = \sup_{x \in [0, 1]} |w_x|$ . If  $S$  is an operator on the space of functions from  $[0, 1]$  to  $\mathbb{C}$ , we denote  $\|S\|$  operator norm corresponding to the previous norm. If  $m$  is a function from  $\mathbb{H}^+$  to  $\mathcal{H}$ , we denote  $m_x$  the function  $z \mapsto m(z)(x)$ . We then have the following result concerning the solution of the equation:

**Theorem A.2.** *The equation  $K_\sigma$  has a unique solution  $m$  which is analytic in  $z$ . Moreover for every  $x \in [0, 1]$ , the function  $m_x = m(\cdot)(x)$  is the Stieltjes transform of a probability measure  $p_x$  on  $\mathbb{R}$ .*

This theorem is in fact a direct application of Theorem 2.1 from [4] which states that the equation  $K_\sigma$  always has a solution in a more general context where we replace  $[0, 1]$  by a probability space  $\mathfrak{X}$  and  $S$  is a symmetric, positivity preserving operator on the space of uniformly bounded complex functions on  $\mathfrak{X}$ .

**Remark A.3.** If  $\sigma$  is a piecewise constant variance profile with parameters  $\alpha_1, \dots, \alpha_n$  and  $(\sigma_{i,j})_{1 \leq i, j \leq n}$ , then the solution of  $(K_\sigma)$  is piecewise constant on the intervals  $I_j$ . This can be viewed directly from  $K_\sigma$  by noticing that  $Sf$  is always piecewise constant on the intervals  $I_j$ .

We will denote  $\mu_\sigma := \int_0^1 p_x dx$  where  $p_x$  is given by the preceding theorem. And so we have that the Stieltjes transform of  $\mu_\sigma$  is  $\int_0^1 m_x dx$ . This measure  $\mu_\sigma$  will be the limit of the empirical measures  $\hat{\mu}_N$  of the matrices  $X_N^{(\beta)}$ . To prove this, we will use the following result which is a reformulation of Girko's result [26, Theorem 1.1].

**Theorem A.4.** *Let us denote  $\sigma_N : [0, 1]^2 \rightarrow \mathbb{R}^+$  the function  $(x, y) \mapsto \Sigma_N([Nx], [Ny])$ . When  $N$  tends to infinity, for almost every  $x$  we have that with probability 1:*

$$\lim_{N \rightarrow \infty} \left| \hat{\mu}_N([-\infty, x]) - \int_{-\infty}^x d\mu_{\sigma_N}(x) \right| = 0$$

*Proof.* First, since for all  $N$ ,  $\Sigma_N$  is bounded and since the coefficients of  $X_N^{(\beta)}$  are centered, hypothesis (1.1) and (1.2) are verified. Then, since we have a sharp sub-Gaussian bound on the entries of  $X_N^{(\beta)}$ , we can easily verify the Lindeberg's condition (1.3). Furthermore, since  $\sigma_N$  is piecewise constant, the solution  $m$  of  $K_{\sigma_N}$  is piecewise constant on the intervals  $[i/N, (i + 1)/N]$ . Making the change of variables  $c_i(z) = m_{(2i-1)/2}(z)$  we have that the equation  $K_{\sigma_N}$  is equivalent, up to the sign since our Stieltjes transform convention is different, to the matricial equation given in [26, Theorem 1.1].  $\square$

And so we are left with determining the convergence of the measure  $\mu_{\sigma_N}$  when  $N$  tends to  $+\infty$ . To that end, we will need the following rough stability results.

**Theorem A.5.** *Let  $\sigma : [0, 1]^2 \rightarrow \mathbb{R}^+$  be a bounded measurable function. For every open neighbourhood  $\mathcal{V}$  of  $\mu_\sigma$  in  $\mathcal{P}([0, 1])$  for the vague topology, there is  $\eta > 0$  such that for every  $\tilde{\sigma}$  bounded measurable function such that  $\sup_{x, y \in [0, 1]^2} |\sigma^2(x, y) - \tilde{\sigma}^2(x, y)| \leq \eta$ ,  $\mu_{\tilde{\sigma}} \in \mathcal{V}$ .*

*Proof.* This proof is inspired from the proof of [1, Proposition 2.1]. We let  $\tilde{S}$  be the kernel operator with kernel  $\tilde{\sigma}^2$ . Let  $\mathbb{H}_\eta^+ = \{z \in \mathbb{C} : \Im z \geq \eta, |z| \leq \eta^{-1}\}$ ,  $\mathbb{H}_\eta^- = \{z \in \mathbb{C} : \Im z \leq$

$-\eta, |z| \leq \eta^{-1}$  and  $D$  the function defined on  $(\mathbb{H}^-)^2$  by

$$D(\zeta, \omega) = \frac{|\zeta - \omega|^2}{\Im \zeta \Im \omega}$$

then  $d := \operatorname{arcosh}(1 + D)$  is the hyperbolic distance on  $\mathbb{H}^-$ . For  $u$  a function from  $\mathbb{H}_\eta^+$  to  $\mathcal{H}$  we let  $\Phi(u)$  and  $\tilde{\Phi}(u)$  be functions from  $\mathbb{H}_\eta^+$  to  $\mathcal{H}$  defined as follows for  $z \in \mathbb{H}_\eta^+$ :

$$\begin{aligned} \Phi(u)(z) &:= \frac{1}{z - Su(z)} \\ \tilde{\Phi}(u)(z) &:= \frac{1}{z - \tilde{S}u(z)}. \end{aligned}$$

If  $\mathcal{B}_\eta := \{u : \mathbb{H}_\eta^+ \rightarrow \mathcal{H} : \sup_{x \in [0,1], z \in \mathbb{H}_\eta^+} \Im u_x(z) \leq -\frac{\eta^3}{(2 + \min\{\|S\|, \|\tilde{S}\|\})^2}, \sup_{x \in [0,1], z \in \mathbb{H}_\eta^+} |u_x(z)| \leq \eta^{-1}\}$ , then following the proof of [1, Proposition 2.1], if  $u$  is such that for all  $z \in \mathbb{H}_\eta^+$ , and all  $x \in [0, 1]$ ,  $u_x(z) \in \mathbb{H}_\eta^-$ , then  $\Phi(u) \in \mathcal{B}_\eta$ . Then, if  $\eta < 2 + \min\{\|S\|, \|\tilde{S}\|\}$ ,  $\frac{\eta^3}{(2 + \min\{\|S\|, \|\tilde{S}\|\})^2} \leq \eta$ ,  $\Phi$  maps  $\mathcal{B}_\eta$  onto itself and so on for  $\tilde{\Phi}$ .

For  $x \in [0, 1]$ ,  $z \in \mathbb{H}_\eta^+$  and  $\delta \geq \sup_{x,y} |\sigma^2(x, y) - \tilde{\sigma}^2(x, y)|$

$$\begin{aligned} D(\Phi(u)_x(z), \tilde{\Phi}(u)_x(z)) &= D((Su)_x(z) - z, (\tilde{S}u)_x(z) - z) \\ &\leq \frac{|(S - \tilde{S})u_x(z)|^2}{\Im z^2} \\ &\leq \frac{\delta^2 \sup_{x \in [0,1]} |u_x(z)|^2}{\eta^2} \\ &\leq \frac{\delta^2}{\eta^4}. \end{aligned}$$

Let  $m$  be the solution of  $K_\sigma$ , that is the fixed point of  $\Phi$ . For every  $n \in \mathbb{N}$  let  $v^{(n)} = \tilde{\Phi}^{(n)}(m)$ . We have for  $z \in \mathbb{H}_\eta^+$ :

$$\sup_{x \in [0,1]} D(m_x(z), v_x^{(1)}(z)) \leq \frac{\delta^2}{\eta^4}$$

and following again [1],

$$\begin{aligned} \sup_{x \in [0,1]} D(v_x^{(n+1)}(z), v_x^{(n)}(z)) &= \sup_{x \in [0,1]} D(\tilde{\Phi}(v^{(n)})_x(z), v_x^{(n)}(z)) \\ &\leq \left(1 + \frac{\eta^2}{\|\tilde{S}\|}\right)^{-2} \sup_{x \in [0,1]} D(v_x^{(n)}(z), v_x^{(n-1)}(z)) \end{aligned}$$

and so we have:

$$\sup_{x \in [0,1]} D(v_x^{(n+1)}(z), v_x^{(n)}(z)) \leq \frac{\delta^2}{\eta^4} \left(1 + \frac{\eta^2}{\|\tilde{S}\|}\right)^{-2n}.$$

And so for  $\delta$  small enough,  $(v^{(n)})_{n \in \mathbb{N}}$  is a Cauchy sequence for the distance  $d_{\mathbb{H}_\eta^+}(u, v) = \sup_{x \in [0,1], z \in \mathbb{H}_\eta^+} \operatorname{arcosh}(1 + D(u_x(z), v_x(z)))$  which is converging toward  $\tilde{m}$  the fixed point of  $\tilde{\Phi}$  and

$$d_{\mathbb{H}_\eta^+}(m, \tilde{m}) \leq \sum_{n=0}^{+\infty} \operatorname{arcosh} \left(1 + \frac{\delta^2}{\eta^4} \left(1 + \frac{\eta^2}{\|\tilde{S}\|}\right)^{-2n}\right).$$

Therefore for every  $\epsilon > 0$  and  $\eta > 0$ , there is  $\delta > 0$  small enough such that  $\sup_{x,y} |\sigma^2(x,y) - \bar{\sigma}^2(x,y)| \leq \delta$  implies  $d_{\mathbb{H}_\eta^+}(m, \tilde{m}) \leq \epsilon$ . Since a base of neighbourhood of  $\mu_\sigma$  for the vague topology is given by:

$$\mathcal{V}_\eta := \{\lambda \in \mathcal{P}(\mathbb{R}) : \forall z \in \mathbb{H}_\eta^+, |G_{\mu_\sigma}(z) - G_\lambda(z)| \leq \eta\}$$

and because the vague topology and the weak topology are equal on the set  $\mathcal{P}(\mathbb{R})$  of probability measure on  $\mathbb{R}$ , we have our result since  $G_{\mu_\sigma} = \int_0^1 m_x dx$  and  $G_{\mu_{\bar{\sigma}}} = \int_0^1 \tilde{m}_x dx$   $\square$

**Corollary A.6.** *In the case of a continuous variance profile  $\sigma$ ,  $\hat{\mu}_N$  converges in probability towards  $\mu_\sigma$ .*

*Proof.* This is a consequence from Theorem A.4 and Theorem A.5 by noticing that  $\lim_{N \rightarrow \infty} \sup_{x,y \in [0,1]} |(\sigma_N)^2(x,y) - \sigma^2(x,y)| = 0$ .  $\square$

We will also need a similar result for the piecewise constant case.

**Theorem A.7.** *Let  $s = (s_{i,j})_{i,j=1}^n \in \mathcal{S}_n(\mathbb{R}^+)$  and  $\vec{\alpha}, \vec{\beta} \in (\mathbb{R}^{+,*})^n$  be two vectors of positive coordinates summing to one and let  $\gamma_i = \sum_{j=0}^i \alpha_j$  and  $\tilde{\gamma}_i = \sum_{j=0}^i \beta_j$ . Let  $\sigma$  and  $\bar{\sigma}$  and the two piecewise constant variance profiles associated respectively with the couples  $(s, \vec{\alpha})$  and  $(s, \vec{\beta})$  and  $v$  and  $\tilde{v}$  the solutions respectively of  $K_\sigma$  and  $K_{\bar{\sigma}}$ . For  $i \in \llbracket 1, n \rrbracket$  let also  $m_i$  and  $\tilde{m}_i$  be the holomorphic functions given by  $v_x = \sum_{i=1}^n \mathbb{1}_{\gamma_{i-1} \leq x < \gamma_i} m_i$  and  $\tilde{v}_x = \sum_{i=1}^n \mathbb{1}_{\tilde{\gamma}_{i-1} \leq x < \tilde{\gamma}_i} \tilde{m}_i$ . Then for every  $\eta > 0$  there is  $\epsilon > 0$  such that if  $\sup_i |\alpha_i - \beta_i| \leq \epsilon$ , then for all  $z \in \mathbb{H}_\eta^+$ , we have  $\sup_i |m_i(z) - \tilde{m}_i(z)| \leq \eta$ .*

Note that if we impose  $s_{i,j} > 0$  for all  $i, j$ , this result is a particular case of Proposition 10.1 [6]. However, since we would like for  $s_{i,j}$  to be potentially 0, we present the following proof which while not at the same level of generality and quantitative bounds will be sufficient enough.

*Proof.* We use the same notations as in the previous proof. Since  $v$  is the solution of  $K_\sigma$ , the  $m_i$  satisfy the following system:

$$m_i = \frac{1}{z - \sum_{j=1}^n \alpha_j s_{i,j}^2 m_j},$$

For the  $\tilde{m}_i$ , we have:

$$\tilde{m}_i = \frac{1}{z - \sum_{j=1}^n \beta_j s_{i,j}^2 \tilde{m}_j}.$$

For  $u$  a function from  $\mathbb{H}_\eta^+$  to  $(\mathbb{H}^-)^n$ , we let  $\Phi(u)$  and  $\tilde{\Phi}(u)$  be two functions from  $\mathbb{H}_\eta^+$  to  $(\mathbb{H}^-)^n$  defined for  $z \in \mathbb{H}_\eta^+$  and  $i = 1, \dots, n$  by:

$$(\Phi(u)(z))_i = \frac{1}{z - (Su)_i(z)}$$

and

$$(\tilde{\Phi}(u)(z))_i = \frac{1}{z - (\tilde{S}u)_i(z)}$$

where  $S$  and  $\tilde{S}$  are the linear applications defined by

$$\forall i = 1, \dots, n, (Su)_i = \sum_{j=1}^n \alpha_j s_{i,j}^2 u_j \text{ and } (\tilde{S}u)_i = \sum_{j=1}^n \beta_j s_{i,j}^2 u_j$$



## Large deviations for matrices with variance profiles

As in the previous proof, if  $\mathcal{B}_\eta := \{u : \mathbb{H}_\eta^+ \rightarrow (\mathbb{H}^-)^n : \sup_{z \in \mathbb{H}_\eta^+} \Im u(z) \leq \frac{-\eta^3}{(2 + \min\{\|S\|, \|\tilde{S}\|\})^2}, \sup_{z \in \mathbb{H}_\eta^+} \|u(z)\| \leq \eta^{-1}\}$ ,  $\Phi$  and  $\tilde{\Phi}$  maps  $\mathcal{B}_\eta$  onto itself for  $\eta$  small enough. For  $u \in \mathcal{B}_\eta$ , we have as before if  $\delta \geq (\sup_{i,j} s_{i,j}^2)(\sup_k |\alpha_k - \beta_k|)$ , for all  $i$ :

$$\begin{aligned} D(\Phi(u)_i(z), \tilde{\Phi}(u)_i(z)) &\leq \frac{|(S - \tilde{S})(u)_i(z)|}{\Im z^2} \\ &\leq \frac{\delta^2}{\eta^4} \end{aligned}$$

Then, using the same reasoning as in the previous case, we have that for every  $\eta > 0$ , there is  $\delta' > 0$  such that if  $\sup_i |\alpha_i - \beta_i| \leq \delta'$  then  $\sup_{z \in \mathbb{H}_\eta^+} \sup_i |m_i(z) - \tilde{m}_i(z)| \leq \eta$ .  $\square$

**Remark A.8.** A more elementary proof of the convergence of the measure can also be obtained since we have bounds on the moments of the entries in our case via a classical moments methods. Let us consider for any  $k \in \mathbb{N}$ , we consider  $W_k$  the set of words  $w = (w_1, \dots, w_{2k+1})$  on  $\llbracket 1, k+1 \rrbracket$  such that  $\{w_1, \dots, w_{2k+1}\} = \llbracket 1, k+1 \rrbracket$ ,  $w_1 = w_{2k+1}$  and such that for any  $i \in \llbracket 1, k+1 \rrbracket$ , there is exactly one  $j \in \llbracket 1, k+1 \rrbracket \setminus \{i\}$  such that  $\{w_i, w_{i+1}\} = \{w_j, w_{j+1}\}$ . For such words  $w$ , we define  $E_w := \{\{w_i, w_{i+1}\} : i \in \llbracket 1, 2k+1 \rrbracket\}$ . On this set, we can define a relation of equivalence  $\sim$  by letting  $w \sim w'$  if there is a permutation  $f$  of  $\llbracket 1, k+1 \rrbracket$  such that  $f(w_i) = w'_i$  for every  $i$ . We can then define  $\mathcal{W}_k$  an arbitrary set of representative of  $W_k$  for the equivalence relation  $\sim$ . Then using classical arguments for the computation of moments of  $\hat{\mu}_N$ , one can see, using that the  $k$ -th moment of the entries of  $X_N^{(\beta)}$  is bounded uniformly in  $N$ , that we have that for  $k \geq 0$ :

$$\mathbb{E}[\hat{\mu}_N(x^k)] = \frac{1}{N} \mathbb{E}[\text{Tr}(X_N^{(\beta)})^k] = c_k^N + O(N^{-1/2})$$

and

$$\text{Var}(\hat{\mu}_N(x^k)) = O(N^{-1/2})$$

where  $c_k^N = 0$  for  $k$  odd and

$$c_{2k}^N = N^{-k-1} \sum_{w \in \mathcal{W}_k} \sum_{\substack{i_1, \dots, i_{k+1}=1 \\ \text{pairwise distinct}}}^N \prod_{\{k,l\} \in E_w} \Sigma_N^2(i_k, i_l)$$

We redirect the reader to [9, Section 2] for instance to get an overview of such methods in the case of classical Wigner matrices. One can then find that in the piecewise constant case:

$$\lim_{N \rightarrow \infty} c_{2k}^N = \sum_{w \in \mathcal{W}_k} \sum_{i_1, \dots, i_{k+1}=1}^n \prod_{j=1}^{k+1} \alpha_{i_j} \prod_{\{k,l\} \in E_w} \sigma_{i_k, i_l}^2$$

and in the continuous case:

$$\lim_{N \rightarrow \infty} c_{2k}^N = \sum_{w \in \mathcal{W}_k} \int_{[0,1]^{k+1}} \prod_{\{k,l\} \in E_w} \sigma^2(x_k, x_{k+1}) dx_1 \dots dx_l.$$

So if we denote  $c_k = \lim_{N \rightarrow \infty} c_k^N$ , we have that in probability  $\hat{\mu}_N$  converges toward a measure  $\mu_\sigma$  whose moments are the  $c_k$ .

In order to apply the full results of [2], we will need the positivity assumption 3.4 for the piecewise constant variance profile. We then have the following convergence result:

**Theorem A.9.** *If we are in the piecewise constant case with Assumption 3.4 satisfied, if we let  $l_N$  and  $r_N$  be respectively the left and right edge of the support of  $\hat{\mu}_N$ , that is respectively the smallest eigenvalue and largest eigenvalue of  $X_N^{(\beta)}$  and  $l_\sigma$  and  $r_\sigma$  the left and right edges of the support of  $\mu_\sigma$ , we have for every  $\delta > 0$ ,  $D > 0$ ,*

$$\mathbb{P}[r_N \geq r_\sigma + \delta \text{ or } l_N \leq l_\sigma - \delta] \leq N^{-D}$$

for  $N$  large enough.

*Proof.* This is in fact an application of corollary 1.10 from [2] which states that the extreme eigenvalues cannot leave the neighborhood of the support of  $\mu_{\sigma_N}$  where  $\sigma_N$  is the same as in Theorem A.4. We need only to check the hypothesis (A) to (D). Up to multiplication by a scalar, our matrix model satisfies the boundedness condition (A) and the Assumption 3.4 gives us the positivity hypothesis (B). The sharp-sub Gaussian hypothesis gives (D). For the boundedness condition on the Stieltjes transform (C) we can use Theorem 6.1 from [4]. Our kernel operator  $S$  satisfy assumptions A1,A2 and B1. Particularly we can use remark 6.2 and 6.3 to bound  $m$  respectively away and near 0. Then, we need only to prove that  $r_{\sigma_N}$  and  $l_{\sigma_N}$  converges toward  $r_\sigma$  and  $l_\sigma$ . This can be done for instance by looking at the expression on the moments of  $\mu_{\sigma_N}$  and  $\mu_\sigma$  given in Remark A.8.  $\sigma_N$  and  $\sigma$  are both piecewise constant functions with  $\sigma$  being associated with the parameters  $(\sigma_{i,j})_{i,j \in [1,n]}$  and  $(\alpha_i)_{i \in [1,n]}$  and  $\sigma_N$  being associated with the parameters  $(\sigma_{i,j})_{i,j \in [1,n]}$  and  $(\alpha_i^N)_{i \in [1,n]}$  where the  $\alpha_i^N$  are defined by

$$\alpha_i^N := \frac{|I_i^{(N)}|}{N}$$

where we remind that the  $I_i^{(N)}$  are defined in subsection 1.1. Since  $\lim_{N \rightarrow \infty} \alpha_i^N = \alpha_i$  for any  $\epsilon > 0$  we have that for  $N$  large enough:

$$\forall i \in [1, n], (1 - \epsilon)\alpha_i^N \leq \alpha_i \leq (1 + \epsilon)\alpha_i^N$$

Then using the formula of Remark A.8 which gives the moments of  $\mu_{\sigma_N}$  and  $\mu_\sigma$  in terms of the  $\sigma_{i,j}$  and the  $\alpha_i$  and  $\alpha_i^N$  we see that for every  $\epsilon > 0$ , for  $N$  large enough, we have that for every  $k \in \mathbb{N}$ ,  $(1 - \epsilon)^{k+1} \mu_{\sigma_N}(x^{2k}) \leq \mu_\sigma(x^{2k}) \leq (1 + \epsilon)^{k+1} \mu_{\sigma_N}(x^{2k})$ . We conclude using that since the  $\mu_\sigma$  and  $\mu_{\sigma_N}$  are symmetric,  $r_{\sigma_N} = -l_{\sigma_N} = \lim_{k \rightarrow \infty} \mu_{\sigma_N}(x^{2k})^{1/2k}$  and  $r_\sigma = -l_\sigma = \lim_{k \rightarrow \infty} \mu_\sigma(x^{2k})^{1/2k}$ .  $\square$

## B Appendix: Proof of Lemma 3.7

This section is devoted to the proof of Lemma 3.7. For this, we will use a concentration results respectively from [29] and Theorem A.4

**Theorem B.1.** *By [29, (Corollary 1.4 a)] (for the compact case) and [29, Corollary 1.4 b)] (for the logarithmic Sobolev case), we have for  $\beta = 1, 2$ , and for  $N$  large enough*

$$\limsup_{N \rightarrow \infty} \frac{1}{N^{7/6}} \log \mathbb{P}[d(\hat{\mu}_N, \mathbb{E}[\hat{\mu}_N]) > N^{-1/6}] < 0$$

where  $d$  is the Wasserstein distance distance defined for two measure  $\mu, \nu \in \mathcal{P}(\mathbb{R})$  by

$$d(\mu, \nu) = \sup_{f \in \mathcal{F}_{Lip}} \left| \int f d\mu - \int f d\nu \right|$$

with  $\mathcal{F}_{Lip}$  being the set of bounded Lipschitz real function  $f$  of  $\mathbb{R}$ , such that  $\|f\|_\infty + \|f\|_{Lip} \leq 1$ , where

$$\|f\|_{Lip} = \sup_{\substack{x, y \in \mathbb{R} \\ x \neq y}} \frac{|f(x) - f(y)|}{|x - y|}$$

Therefore we only need to show that

$$\lim_{N \rightarrow \infty} d(\mathbb{E}[\hat{\mu}_N], \mu_\sigma) = 0$$

Using Theorem A.7, we have that  $\lim_{N \rightarrow +\infty} d(\mu_{\sigma_N}, \mu_\sigma) = 0$  and therefore, using Theorem A.4 we have in probability  $\lim_{N \rightarrow +\infty} d(\hat{\mu}_N, \mu_\sigma) = 0$ . From this, we can deduce that  $\lim_{N \rightarrow \infty} d(\mathbb{E}[\hat{\mu}_N], \mu_\sigma) = 0$ . Indeed, if  $f$  is a continuous function bounded in absolute norm by 1, we have in probability that  $\lim_{N \rightarrow \infty} \hat{\mu}_N(f) = \mu_\sigma(f)$  and so since  $|\hat{\mu}_N(f)| \leq 1$ , we have  $\lim_{N \rightarrow \infty} \mathbb{E}[\hat{\mu}_N](f) = \mu_\sigma(f)$ .

## References

- [1] Oskari H. Ajanki, Torben Krüger, and László Erdős. Singularities of solutions to quadratic vector equations on the complex upper half-plane. *Communications on Pure and Applied Mathematics*, 70(9):1672–1705, 2017. MR3684307
- [2] Oskari H. Ajanki, László Erdős, and Torben Krüger. Universality for general Wigner-type matrices. *Probability Theory and Related Fields*, 169(3):667–727, 2017. MR3719056
- [3] Oskari H. Ajanki, László Erdős, and Torben Krüger. Stability of the matrix Dyson equation and random matrices with correlations. *Probability Theory and Related Fields*, 173(1-2):293–373, 2018. MR3916109
- [4] Oskari H. Ajanki and Torben Krüger. Quadratic Vector Equations on Complex Upper Half-plane. *Memoirs of the American Mathematical Society. American Mathematical Society*, 261(1261), 2019. MR4031100
- [5] Johannes Alt, László Erdős, Torben Krüger, and Dominik Schröder. Correlated Random Matrices: Band Rigidity and Edge Universality. *Ann. Probab.*, 48(2):963–1001, 2020. MR4089499
- [6] Johannes Alt, Laszlo Erdos, and Torben Krüger. The Dyson equation with linear self-energy: spectral bands, edges and cusps. *Documenta Mathematica*, 25:1421–1539, 2018. MR4164728
- [7] Johannes Alt, László Erdős, Torben Krüger, and Yuriy Nemish. Location of the spectrum of Kronecker random matrices. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 55(2):661–696, 2019. MR3949949
- [8] Greg Anderson and Ofer Zeitouni. A CLT for a band matrix model. *Probability Theory and Related Fields*, 134, Feb 2005. MR2222385
- [9] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. MR2760897
- [10] Gerard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica. The landscape of the spiked tensor model. *Communications on Pure and Applied Mathematics*, 72(11):2282–2330, 2019. MR4011861
- [11] Fanny Augeri. Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails. *Electron. J. Probab.*, 21:32–49, 2016. MR3492936
- [12] Fanny Augeri. Nonlinear large deviation bounds with applications to traces of Wigner matrices and cycles counts in Erdos-Renyi graphs. *Ann. Probab.*, 48(5):404–2448, 2020. MR4152647
- [13] Fanny Augeri, Alice Guionnet, and Jonathan Husson. Large deviations for the largest eigenvalue of sub-Gaussian matrices. *Comm. Math. Phys.*, 383(2):997–1050, 2021. MR4239836
- [14] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005. MR2165575
- [15] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1):1–67, 2001. MR1856194
- [16] Gérard Ben Arous and Alice Guionnet. Large deviations for Wigner’s law and Voiculescu’s non-commutative entropy. *Probability Theory and Related Fields*, 108(4):517–542, 1997. MR1465640

- [17] Bhaswar B Bhattacharya and Shirshendu Ganguly. Upper tails for edge eigenvalues of random graphs. *SIAM Journal on Discrete Mathematics*, 34(2):1069–1083, 2020. MR4083586
- [18] Pascal Bianchi, Merouane Debbah, Mylène Maïda, and Jamal Najim. Performance of statistical tests for single-source detection using random matrix theory. *IEEE Transactions on Information theory*, 57(4):2400–2419, 2011. MR2809098
- [19] Charles Bordenave and Pietro Caputo. A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.*, 42(6):2454–2496, 2014. MR3265172
- [20] Giorgio Cipolloni, László Erdős, Torben Krüger, and Dominik Schröder. Cusp universality for random matrices, ‘II: The real symmetric case. *Pure and Applied Analysis*, 1(4):615–707, 2019. MR4026551
- [21] Nicholas Cook and Amir Dembo. Large deviations of subgraph counts for sparse Erdős–Rényi graphs. *Advances in Mathematics*, 373:107289, 2020. MR4130460
- [22] László Erdős, Torben Krüger, and Dominik Schröder. Cusp universality for random matrices I: Local law and the complex Hermitian case. *Communications in Mathematical Physics*, 378(2):1203–1278, 2020. MR4134946
- [23] László Erdős. The matrix Dyson equation and its applications for random matrices. *arXiv preprint arXiv:1903.10060*, 2019. MR3971154
- [24] Anne Fey, Remco van der Hofstad, and Marten J. Klok. Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Advances in Applied Probability*, 40(4):1048–1071, Dec 2008. MR2488532
- [25] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981. MR637828
- [26] Vyacheslav L. Girko. *Canonical Equation K1*, pages 1–24. Springer Netherlands, Dordrecht, 2001. MR1887675
- [27] Alice Guionnet and Jonathan Husson. Large deviations for the largest eigenvalue of Rademacher matrices. *Annals of Probability*, 48(3):1436–1465, 2020. MR4112720
- [28] Alice Guionnet and Mylène Maïda. A Fourier view on the R-transform and related asymptotics of spherical integrals. *Journal of Functional Analysis*, 222(2):435–490, 2005. MR2132396
- [29] Alice Guionnet and Ofer Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136, 2000. MR1781846
- [30] Alexei M Khorunzhy and Leonid A Pastur. On the eigenvalue distribution of the deformed wigner ensemble of random matrices. *Advances in Soviet Mathematics*, 19:97–127, 1994. MR1298444
- [31] Ji Oon Lee and Jun Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Math. J.*, 163(1):117–173, 2014. MR3161313
- [32] Antoine Maillard, Gérard Ben Arous, and Giulio Biroli. Landscape complexity for the empirical risk of generalized linear models. In *Mathematical and Scientific Machine Learning*, pages 287–327. PMLR, 2020.
- [33] Madan Lal Mehta. *Random matrices*, volume 142. Elsevier, 2004. MR2129906
- [34] Dimitri Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *International Mathematics Research Notices*, 1996(20):1013–1025, Jan 1996. MR1422374
- [35] Dan Voiculescu. The analogues of entropy and of fisher’s information measure in free probability theory, V. Noncommutative Hilbert transforms. *Inventiones mathematicae*, 132:189–227, 1994. MR1618636
- [36] Van H Vu. Spectral norm of random matrices. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 423–430, 2005. MR2181644
- [37] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals Math.*, 67:325–327, 1958. MR95527