



**HAL**  
open science

# Large deviations of the extreme eigenvalues of random deformations of matrices

Florent Benaych-Georges, Alice Guionnet, Mylène Maïda

► **To cite this version:**

Florent Benaych-Georges, Alice Guionnet, Mylène Maïda. Large deviations of the extreme eigenvalues of random deformations of matrices. 2010. hal-00505502v2

**HAL Id: hal-00505502**

**<https://hal.science/hal-00505502v2>**

Preprint submitted on 31 Aug 2010 (v2), last revised 18 Jun 2011 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# LARGE DEVIATIONS OF THE EXTREME EIGENVALUES OF RANDOM DEFORMATIONS OF MATRICES

F. BENAYCH-GEORGES\*, A. GUIONNET†, M. MAIDA‡.

ABSTRACT. Consider a real diagonal deterministic matrix  $X_n$  of size  $n$  with spectral measure converging to a compactly supported probability measure. We perturb this matrix by adding a random finite rank matrix of a certain form, with delocalized eigenvectors. We show that the joint law of the extreme eigenvalues of the perturbed model satisfies a large deviation principle, in the scale  $n$  with a good rate function given by a variational formula.

We tackle both cases when the extreme eigenvalues of  $X_n$  converge to the edges of the support of the limiting measure and when we allow some eigenvalues of  $X_n$ , that we call outliers, to converge out of the bulk.

We can also generalise our results to the case when  $X_n$  is random, with law proportional to  $e^{-\text{Tr } V(X)} dX$ , for  $V$  growing fast enough at infinity and any perturbation of finite rank with orthonormal eigenvectors.

## CONTENTS

1. Introduction	2
2. Statement of the results	3
3. Scheme of the proofs	8
4. Characterisation of the eigenvalues of $\widetilde{X}_n$ as zeroes of a function $H^n$	9
5. Large deviations for $H^n$ in the case without outliers	11
6. Large deviations for the largest eigenvalues in the case without outliers	19
7. Large deviations for the eigenvalues of Wishart matrices	28
8. Large deviations for $H^n$ in the presence of outliers	30
9. Large deviations principle for the largest eigenvalues in the case with outliers	33
10. Application to $X_n$ random, following some classical matrix distribution	39
11. Appendix	42

---

*Date:* August 31, 2010.

*2000 Mathematics Subject Classification.* 15A52,60F10.

*Key words and phrases.* Random matrices, large deviations.

\* UPMC Univ Paris 6, LPMA, Case courrier 188, 4, Place Jussieu, 75252 Paris Cedex 05, France, florent.benaych@upmc.fr,

† UMPA, ENS Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France, aguionne@umpa.ens-lyon.fr,

‡ Université Paris-Sud, Laboratoire de Mathématiques, Bâtiment 425, Faculté des Sciences, 91405 Orsay Cedex, France, mylene.maida@math.u-psud.fr.

This work was supported by the *Agence Nationale de la Recherche* grant ANR-08-BLAN-0311-03.

## 1. INTRODUCTION

In the last twenty years, many features of the asymptotics of the spectrum of large random matrices have been understood. For a wide variety of classical models of random matrices, it has been shown that the spectral measure converges almost surely (see [32] for Wigner matrices, or [28] for Wishart matrices). The extreme eigenvalues converge for these models to the boundary of the limiting spectral measure (see e.g. [24] or [3]). Fluctuations of the spectral measure and the extreme eigenvalues of these models could also be studied under a fair generality over the entries of the matrices; we refer to [1] and [4] for reviews. Recently, even the fluctuations of the eigenvalues inside the bulk could be studied for rather general entries and were shown to be universal (see e.g. [16] or [30]). Concentration of measure phenomenon and moderate deviations could also be established in [18, 13, 15]. Yet, the understanding of the large deviations of the spectrum of large random matrices is still very scarce and exists only in very specific cases.

Indeed, the spectrum of a matrix is a very complicated function of the entries, so that usual large deviation theorems, based on independence, do not apply easily. Moreover, large deviations rate functions have to depend on the distribution of the entries and only guessing their definition is still a wide open question. In the case of Gaussian Wigner matrices, where the joint law of the eigenvalues is simply given by a Coulomb gas Gibbs measure, things are much easier and a full large deviation principle for the law of the spectral measure of such matrices was proved in [8]. This extends to other ensembles distributed according to similar Gibbs measure, for instance Gaussian Wishart matrices (see [19]). Similar large deviation results hold in discrete situations with a similar Coulomb gas distribution (see [21]). A large deviation principle was also established in [25] for the law of the spectral measure of a random matrix given as the sum of a self-adjoint Gaussian Wigner random matrix and a deterministic self-adjoint matrix (or as a Gaussian Wishart matrix with non trivial covariance matrix). In this case, the proof was using stochastic analysis and Dyson's Brownian motion, as there is no explicit joint law for the eigenvalues, but again relied heavily on the fact that the random matrix has Gaussian entries. The large deviations for the law of the extreme eigenvalues were studied in a slightly more general setting. Again relying on the explicit joint law of the eigenvalues, a large deviation principle was derived in [7] for the same Gaussian type models. The large deviations of extreme eigenvalues of Gaussian Wishart matrices was studied in [31]. In the case where the Wishart matrix is of the form  $XX^*$  with  $X$  a  $n \times r$  rectangular matrix so that the ratio  $r/n$  of its dimensions goes to zero, large deviations of the extreme eigenvalues could be studied under more general assumptions on the entries [22]. Our results allow also to derive a full large deviation for the spectrum of such Wishart matrices when  $r$  is kept fixed while  $n$  goes to infinity (see Section 7).

In this article, we shall be concerned with the effect of finite rank deformations on the deviations of the extreme eigenvalues of random matrices. In fact, such finite rank perturbations do not change the deviations of the spectral measure by Weyl's interlacing property, but it changes the deviations, and even the convergence, of the extreme eigenvalues.

In the case of Gaussian Wishart matrices, the asymptotics of these extreme eigenvalues were established in [6] and a sharp phase transition, known as the BBP transition, was exhibited. According to the strength of the perturbation, the extreme eigenvalues converge to the edge of the bulk or away from the bulk. The fluctuations of these eigenvalues were also shown in [6] to be given either by the Tracy-Widom distribution in the first case, or by the Gaussian distribution in the second case. Universality (and non-universality) of the fluctuations in BBP transition was studied for various models, see e.g. [11, 12, 17, 5]. The goal of this article is to study the large deviations of the extreme eigenvalues of such finite rank perturbation of large random matrices. In [20], a large deviation principle for the largest eigenvalue of matrices of the GOE and GUE deformed by a rank one matrix was obtained by using fine asymptotics of the Itzykson-Zuber-Harich-Chandra (or spherical) integrals. The large deviations of the extreme eigenvalues of a Wigner matrix perturbed by a matrix with finite rank greater than one happened to be much more complicated. One of the goal of this paper is to prove such a large deviation result when the Wigner matrix is Gaussian. In fact, our result will include the more general case where the non-perturbed matrices are taken in some classical matrix ensembles, namely the ones with distribution  $\propto e^{-n \operatorname{tr}(V(X))} dX$ , for which the deviations are well known. We first tackle a closely related question: the large deviation properties of the largest eigenvalues of a deterministic matrix  $X_n$  perturbed by a finite rank random matrix, considering therefore the same kind of models as in [9]. We show that the law of these extreme eigenvalues satisfies a large deviation principle for a fairly general class of random finite rank perturbations. We can then consider random matrices  $X_n$ , independent of the perturbation, by studying the deviations of the perturbed matrix conditionally to the non-perturbed matrices. Even though our rate functions are not very explicit in general, in the simple case where  $X_n = 0$ , we can retrieve more explicit formulae (see Section 7). Our approach is, as in [11, 12, 5], based on the characterisation of the eigenvalues as the zeroes of the characteristic polynomial, which can be seen as a function of the matrix whose entries are the Stieltjes transform of the non-deformed matrix evaluated along the random vectors of the perturbation. We therefore deduce the deviations of the extreme eigenvalues from the deviations of this matrix, which can be obtained by classical large deviations techniques.

Let us now present more precisely the kind of models we will be dealing with.

## 2. STATEMENT OF THE RESULTS

**2.1. The models.** Let  $X_n$  be a real diagonal matrix of size  $n \times n$  with eigenvalues  $\lambda_1^n \geq \lambda_2^n \geq \dots \geq \lambda_n^n$ .

We perturb  $X_n$  by a random matrix whose rank does not depend on  $n$ . More precisely, let  $m, r$  be fixed positive integers and  $\theta_1 \geq \theta_2 \geq \dots \geq \theta_m > 0 > \theta_{m+1} \geq \dots \geq \theta_r$  be fixed, let  $G = (g_1, \dots, g_r)$  be a random vector and  $(G(k) = (g_1(k), \dots, g_r(k)))_{k \geq 1}$  be independent copies of  $G$ . We then define, the  $r$  vectors with dimension  $n$

$$G_1^n := (g_1(1), \dots, g_1(n))^T, \dots, G_r^n := (g_r(1), \dots, g_r(n))^T$$

and study the eigenvalues  $\tilde{\lambda}_1^n \geq \dots \geq \tilde{\lambda}_n^n$  of the deformed matrices

$$\tilde{X}_n = X_n + \frac{1}{n} \sum_{i=1}^r \theta_i G_i^n G_i^{n*}. \quad (1)$$

In the sequel, we will refer to the model (1) as the *i.i.d. case*.

Alternatively, if we assume moreover that the law of  $G$  does not charge any hyperplane, then, for  $n > r$ , the  $r$  vectors  $G_1^n, \dots, G_r^n$  are almost surely linearly independent and we denote by  $(U_i^n)_{1 \leq i \leq r}$  the vectors obtained from  $(G_i^n)_{1 \leq i \leq r}$  by a Gram-Schmidt orthonormalisation procedure with respect to the usual scalar product on  $\mathbb{C}^n$ .

We shall consider the eigenvalues  $\tilde{\lambda}_1^n \geq \dots \geq \tilde{\lambda}_n^n$  of

$$\tilde{X}_n = X_n + \sum_{i=1}^r \theta_i U_i^n U_i^{n*} \quad (2)$$

and refer in the sequel to the model (2) as the *orthonormalized case*.

If  $g_1, \dots, g_r$  are  $r$  independent standard (real or complex) Gaussian variables, it is well known that the law of  $(U_i^n)_{1 \leq i \leq r}$  is the uniform measure on the set of  $r$  orthonormal vectors. The model (2) coincides then with the one introduced in [9].

Our goal will be to examine the large deviations for the  $m$  largest eigenvalues of the deformed matrix  $\tilde{X}_n$ , with  $m$  the number of positive eigenvalues of the random deformation.

**2.2. The assumptions.** Concerning the spectral measure of the full rank deterministic matrix  $X_n$ , we assume the following

**Assumption 2.1.** *The empirical distribution  $\frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i^n}$  of  $X_n$  converges weakly as  $n$  goes to infinity to a compactly supported probability  $\mu$ .*

Concerning the random vector  $G$ , we assume the following:

**Assumption 2.2.**  *$G = (g_1, \dots, g_r)$  is a random vector with entries in  $\mathbb{K} = \mathbb{R}$  or  $\mathbb{C}$  such that there exists  $\alpha > 0$  with  $\mathbb{E}(e^{\alpha \sum_{i=1}^r |g_i|^2}) < \infty$ . If we consider the orthonormalized model, we assume moreover that for any  $\lambda \in \mathbb{K}^r \setminus \{0\}$ ,  $\mathbb{P}(\sum_{i=1}^r \lambda_i g_i = 0) = 0$*

We consider two distinct kind of assumptions on the extreme eigenvalues of  $X_n$ . We will be first interested in the case when these extreme eigenvalues **stick to the bulk** (see Assumption 2.3), and then to the case **with outliers**, when we allow some eigenvalues of  $X_n$  to take their limit outside the support of the limiting measure  $\mu$  (see Assumption 2.5).

**2.3. The results in the case without outliers.** We first consider the case where the extreme eigenvalues of  $X_n$  stick to the bulk.

**Assumption 2.3.** *The largest and smallest eigenvalues of  $X_n$  tend respectively to the upper bound (denoted by  $b$ ) and the lower bound (denoted by  $a$ ) of the support of  $\mu$ .*

Our main theorem is the following (see Theorem 6.1 and Theorem 6.4 for precise statements).

**Theorem 2.4.** *Under Assumptions 2.1, 2.2 and 2.3, the law of the  $m$  largest eigenvalues  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n) \in \mathbb{R}^m$  of  $\tilde{X}_n$  satisfies a large deviation principle (LDP) in the scale  $n$  with a good rate function  $L$ . In other words, for any closed set  $F$  of  $\mathbb{R}^m$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( (\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n) \in F \right) \leq - \inf_F L$$

and for any open set  $O \subset \mathbb{R}^m$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( (\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n) \in O \right) \geq - \inf_O L.$$

Moreover, this rate function achieves its minimum value at a unique  $m$ -tuple  $(\lambda_1^*, \dots, \lambda_m^*)$  towards which  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n)$  converges almost surely.

Theorem 2.4 is true for both the orthonormalized model and the i.i.d. model, but the rate function  $L$  is not the same for both models. As could be expected, the minimum  $(\lambda_1^*, \dots, \lambda_m^*)$  only depends on the  $\theta_i$ 's, on the limiting spectral distribution  $\mu$  of  $X_n$ , and on the covariance matrix of the vector  $G$ .

The rate function  $L$  is not explicit in general. However, in the particular case where  $X_n = 0$  and the vector  $G$  is Gaussian,  $L$  can be computed. It allows us to recover a large deviation principle and to refine a bound of [22] about the eigenvalues of matrices  $W_n = \frac{1}{n} G_n^* \Theta G_n$  for  $G_n$  an  $n \times r$  matrix, with  $r$  fixed and  $n$  growing to infinity (see Section 7).

**2.4. The results in the case with outliers.** We now consider the case where some eigenvalues of  $X_n$  escape from the bulk, so that Assumption 2.3 is not fulfilled. We assume that these eigenvalues, that we call *outliers*, converge:

**Assumption 2.5.** *There exist some non negative integers  $p^+, p^-$  such that for any  $i \leq p^+$ ,  $\lambda_i^n \xrightarrow{n \rightarrow +\infty} \ell_i^+$ , for any  $j \leq p^-$ ,  $\lambda_{n-j+1}^n \xrightarrow{n \rightarrow +\infty} \ell_j^-$ ,  $\lambda_{p^++1}^n \xrightarrow{n \rightarrow +\infty} b$  and  $\lambda_{n-p^-}^n \xrightarrow{n \rightarrow +\infty} a$  with  $-\infty < \ell_1^- \leq \dots \leq \ell_{p^-}^- < a \leq b < \ell_{p^+}^+ \leq \dots \leq \ell_1^+ < \infty$ , where  $a$  and  $b$  denote respectively the lower and upper bounds of the limiting measure  $\mu$ .*

To simplify the notations in the sequel we will use the following conventions :  $\ell_{p^-+1}^- := a$  and  $\ell_{p^++1}^+ := b$ .

In this framework, we will need to make on  $G$  the additional assumption

**Assumption 2.6.**

- The law of the random vector  $G = (g_1, \dots, g_r)$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^r$  or on  $\mathbb{C}^r$ .
- The law of the vector  $\frac{G}{\sqrt{n}}$  satisfies a large deviation principle in the scale  $n$  with a good rate function that we denote by  $I$ .

**Theorem 2.7.** *If Assumptions 2.1, 2.2, 2.5 and 2.6 hold, the law of the  $m + p^+$  largest eigenvalues of  $\tilde{X}_n$  satisfies a large deviation principle with a good rate function  $L^\circ$ .*

Again, Theorem 2.7 is true for both the orthonormalized model and the i.i.d. model, but the rate function is not the same for both models. See Theorem 9.1 for a precise definition of  $L^\circ$ .

Before going any further, let us discuss Assumption 2.6.

- Lemma 2.8.** (1) *If  $G = (g_1, \dots, g_r)$  are i.i.d standard Gaussian variables, Assumption 2.6 holds with  $I(v) = \frac{1}{2}\|v\|_2^2$ .*  
 (2) *If  $G$  is such that for any  $\alpha > 0$ ,  $\mathbb{E}[e^{\alpha \sum_{i=1}^r |g_i|^2}] < \infty$ , then the second point of Assumption 2.6 holds with  $I$  infinite except at 0, where it takes value 0.*

*Proof.* The first result is a direct consequence of Schilder's theorem. For the second, it is enough to notice by Tchebychev's inequality that for all  $L, \delta > 0$ ,

$$\mathbb{P}\left(\max_{1 \leq i \leq r} |g_i|^2 \geq \delta n\right) \leq r e^{-L\delta n} \mathbb{E}(e^{L \sum_{i=1}^r |g_i|^2})$$

so that taking the large  $n$  limit and then  $L$  going to infinity yields

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\max_{1 \leq i \leq r} |g_i/\sqrt{n}|^2 \geq \delta\right) = -\infty$$

thus proving the claim.  $\square$

## 2.5. Large deviations for the largest eigenvalues of perturbed matrix models.

In this section, we apply our results to study the large deviations of the law of the extreme eigenvalues of  $\widetilde{X}_n$ , in the orthonormalized case but where  $X_n$  is now randomly chosen, independently of the  $G_i^n$ 's. We suppose that  $X_n$  is distributed according to the Gibbs measure

$$d\mu_\beta^n(X) = \frac{1}{Z_n} e^{-n \operatorname{tr}(V(X))} dX$$

with  $dX$  the Lebesgue measure on the set of  $n \times n$  Hermitian ( $\beta = 2$  and  $G$  is  $\mathbb{C}^r$ -valued) or symmetric matrices ( $\beta = 1$  and  $G$  is  $\mathbb{R}^r$ -valued). Let us first recall a few facts about this non-perturbed model. It is well known that the law of the eigenvalues of  $X_n$  is given by

$$\mathbb{P}_{V,\beta}^n(d\lambda_1, \dots, d\lambda_n) = \frac{1_{\lambda_1 > \lambda_2 > \dots > \lambda_n}}{Z_{V,\beta}^n} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta e^{-n \sum_{i=1}^n V(\lambda_i)} \prod_{i=1}^n d\lambda_i.$$

A large deviation principle in the scale  $n^2$  for the law of the spectral measure  $n^{-1} \sum_{i=1}^n \delta_{\lambda_i}$  under  $\mathbb{P}_{V,\beta}^n$  was obtained in [8], resulting in particular with the almost sure convergence of the latter to a probability measure  $\mu_V$ . The large deviations for the extreme eigenvalues of  $X_n$  were derived in [7], see also [1, Section 2.6.2]. We give below a slightly more general statement to consider the deviations of the  $p$ th largest eigenvalues (note that the  $p$ th smallest can be considered similarly).

This will be obtained under the following assumption

**Assumption 2.9.** •  *$V$  is continuous and  $\liminf_{|x| \rightarrow \infty} \frac{V(x)}{\beta \log|x|} = +\infty$ .*

• *For all integer numbers  $p$ , the limit*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{Z_{nV/n-p,\beta}^{n-p}}{Z_{V,\beta}^n}$$



exists and is denoted by  $\alpha_{V,\beta}^p$ .

- Under  $\mathbb{P}_{V,\beta}^n$  the largest eigenvalue  $\lambda_1^n$  converges almost surely to the upper boundary  $b_V$  of the support of  $\mu_V$ .

These assumptions hold in a wide generality. It is verified for the law of the GUE ( $\beta = 2$ ,  $V(x) = x^2$ ) and the GOE ( $\beta = 1$ ,  $V(x) = x^2/2$ ) as the second point is verified by Selberg formula whereas the last one is well known (see [1, Section 2.1.6]). It is as well verified for Gaussian Wishart matrices (case  $\beta = 1, 2$  and  $V(x) = \frac{\beta}{2}x^2 - \alpha \log(x)$ ) with the slight extension of taking the integration on the positive real line.

**Theorem 2.10.** *Under Assumption 2.9 the law of the  $p$ th largest eigenvalues ( $\lambda_1^n > \dots > \lambda_p^n$ ) of  $X_n$  satisfies a large deviation principle in the scale  $n$  and with good rate function given by*

$$J^p(x_1, \dots, x_p) = \begin{cases} \sum_{i=1}^p J_V(x_i) + \alpha_{V,\beta}^p, & \text{if } x_1 \geq x_2 \geq \dots \geq x_p, \\ \infty, & \text{otherwise.} \end{cases}$$

with  $J_V(x) = V(x) - \beta \int \log|x-y|d\mu_V(y)$ . Moreover,  $\alpha_{V,\beta}^p = p\alpha_{V,\beta}^1$ .

**Remark 2.11.** *Note that in the case of the GOE and the GUE, see [7],*

$$J_V(x) = \beta \int_2^x \sqrt{(y/2)^2 - 1} dy - \alpha_{V,\beta}^1, \quad \alpha_{V,\beta}^p = p\alpha_{V,\beta}^1.$$

We now go to the perturbed model and consider the orthonormalized case. In this case, due to the rotational invariance of the law of  $X_n$ , one can consider perturbation of a more general form namely

**Assumption 2.12.** *( $U_1^n, \dots, U_r^n$ ) is a family of orthonormal vectors in  $(\mathbb{R}^n)^r$  (resp.  $(\mathbb{C}^n)^r$ ) if  $\beta = 1$  (resp.  $\beta = 2$ ), either deterministic or independent of  $X_n$ .*

Under these assumptions, we know that  $\widetilde{X}_n$  has in law the same eigenvalues as  $D_n + \sum_{i=1}^r \theta_i (O_n U_i^n)(O_n U_i^n)^*$ , with  $D_n$  a real diagonal matrix with eigenvalues of law  $\mathbb{P}_{V,\beta}^n$  and  $O_n$  Haar distributed on the orthogonal (resp. unitary) group of size  $n$  if  $\beta = 1$  (resp.  $\beta = 2$ ). Now, from the well know properties of the Haar measure, if the  $U_i^n$ 's satisfy Assumption 2.12, then the  $O_n U_i^n$ 's are column vectors of a Haar distributed matrix. It means in particular that they can be obtained by the orthonormalisation procedure described in the introduction, with  $G = (g_1, \dots, g_r)$  i.i.d. Gaussian standard variables (which satisfies in particular Assumption 2.6).

With these considerations in mind, we can state the large deviation principle for the extreme eigenvalues of  $\widetilde{X}_n$ . We denote  $b_V$  the rightmost point of the support of  $\mu_V$ .

**Theorem 2.13.** *With  $V$  satisfying Assumption 2.9, we consider the orthonormalized model under Assumption 2.12. Then, for any integer  $k$ , the law of the  $k$ th largest eigenvalues ( $\widetilde{\lambda}_1^n, \dots, \widetilde{\lambda}_k^n$ ) of  $\widetilde{X}_n$  satisfies a large deviation principle in the scale  $n$  and with good rate function given by*

$$\widetilde{J}^k(x_1, \dots, x_k) = \inf_{p \geq 0} \inf_{\ell_1 > \dots > \ell_p > b_V} \{L_{\ell_1, \dots, \ell_p}^0(x_1, \dots, x_k) + J^p(\ell_1, \dots, \ell_p)\},$$

if  $x_1 \geq \dots \geq x_k$ , the function being infinite otherwise.



Here,  $L_{\ell_1, \dots, \ell_p}^0$  is the rate function defined in Theorem 9.1 for the orthonormalized model built on  $G = (g_1, \dots, g_r)$  i.i.d. standard Gaussian variables and  $X_n$  with limiting spectral measure  $\mu_V$  and outliers  $\ell_1, \dots, \ell_p$ .

### 3. SCHEME OF THE PROOFS

The strategy of the proof will be quite similar in both cases (with or without outliers), so, for the sake of simplicity, we will outline it in the present section only in the case without outliers (both orthonormalized and i.i.d. models will be treated simultaneously).

The cornerstone is a nice representation, already crucially used in many papers on finite rank deformations, see e.g. [9, 2], of the eigenvalues  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n)$  as roots of a fixed polynomial of matrices of fixed size, depending only on  $\Theta = \text{diag}(\theta_1, \dots, \theta_r)$  the  $r \times r$  diagonal matrix of the eigenvalues of the perturbation.

Indeed, if  $V$  is the  $n \times r$  matrix equal to  $[U_1^n \cdots U_r^n]$  in the orthonormalized model and to  $[G_1^n \cdots G_r^n]$  in the i.i.d. model, and  $I_n$  the identity in  $n \times n$  matrices, the characteristic polynomial of  $\widetilde{X}_n$  reads

$$\det(zI_n - \widetilde{X}_n) = \det(zI_n - X_n - V\Theta V^*) = \det(zI_n - X_n) \det(I_r - V^*(zI_n - X_n)^{-1}V\Theta) \quad (3)$$

It means that the eigenvalues of  $\widetilde{X}_n$  that are not<sup>1</sup> eigenvalues of  $X_n$  are the zeroes of  $\det(I_r - V^*(zI_n - X_n)^{-1}V\Theta)$ , which is the determinant of a matrix whose size is independent of  $n$ .

Because of the relation between  $V$  and the random vectors  $G_1^n, \dots, G_r^n$ , it is not hard to check that, if we let, for  $z \notin \{\lambda_1^n, \dots, \lambda_n^n\}$ ,  $K^n(z)$  and  $C^n$  be the elements of the set  $\mathbf{H}_r$  of  $r \times r$  Hermitian matrices given, for  $1 \leq i \leq j \leq r$ , by

$$K^n(z)_{ij} = \frac{1}{n} \sum_{k=1}^n \frac{\overline{g_i(k)} g_j(k)}{z - \lambda_k^n} \quad (4)$$

and

$$C^n_{ij} = \frac{1}{n} \sum_{k=1}^n \overline{g_i(k)} g_j(k), \quad (5)$$

we have:

**Proposition 3.1.** *In both i.i.d and orthonormalized situations, there exists a function  $P_{\Theta, r}$  defined on  $\mathbf{H}_r \times \mathbf{H}_r$  which is polynomial in the entries of its arguments and depends only on the matrix  $\Theta$ , such that any  $z \notin \{\lambda_1^n, \dots, \lambda_n^n\}$  is an eigenvalue of  $\widetilde{X}_n$  if and only if*

$$H^n(z) := P_{\Theta, r}(K^n(z), C^n) = 0.$$

Of course, the polynomial  $P_{\Theta, r}$  is different in the i.i.d. and the orthonormalized case. In the i.i.d. model,  $P_{\Theta, r}$  is simpler and does not depend on  $C$ . This proposition characterises the eigenvalues of  $\widetilde{X}_n$  as the zeroes of the random function  $H^n$ , which depends continuously (as a polynomial function) on the random pair  $(K^n(\cdot), C^n)$ . The large deviations of these eigenvalues are therefore inherited from the large deviations of  $(K^n(\cdot), C^n)$ , which we thus

<sup>1</sup> We show in section 11.3 that the spectrum of  $X_n$  and  $\widetilde{X}_n$  is distinct in generic situation.

study in detail before getting into the deviations of the eigenvalues themselves. Because  $K^n(z)$  blows up when  $z$  approaches  $\lambda_1^n$ , which itself converges to  $b$ , we study the large deviations of  $(K^n(z), C^n)$  for  $z$  away from  $b$ . We shall let  $\mathcal{K}$  be a compact interval in  $(b, \infty)$ ,  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r)$  and  $\mathcal{C}(\mathcal{K}, \mathbb{R})$  be the space of continuous functions on  $\mathcal{K}$  taking values respectively in  $\mathbf{H}_r$  and in  $\mathbb{R}$ . We endow the latter set with the uniform topology. We will then prove that (see Theorem 5.1 for a precise statement and a definition of the rate function  $\mathbf{I}$  involved)

**Proposition 3.2.** *The law of  $((K^n(z))_{z \in \mathcal{K}}, C^n)$  on  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  equipped with the uniform topology, satisfies a large deviation principle in the scale  $n$  and with good rate function  $\mathbf{I}$ .*

By the contraction principle, we therefore deduce

**Corollary 3.3.** *The law of  $(H^n(z))_{z \in \mathcal{K}}$  on  $\mathcal{C}(\mathcal{K}, \mathbb{R})$  equipped with the uniform topology, satisfies a large deviation principle in the scale  $n$  and with rate function given, for a continuous function  $f \in \mathcal{C}(\mathcal{K}, \mathbb{R})$ , by*

$$J_{\mathcal{K}}(f) = \inf \{ \mathbf{I}(K(\cdot), C) ; (K(\cdot), C) \in \mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r, P_{\Theta, r}(K(z), C) = f(z) \forall z \in \mathcal{K} \}$$

with  $P_{\Theta, r}$  the polynomial function of Proposition 3.1.

Theorem 2.4 is then a consequence of this corollary with, heuristically,  $L(\alpha)$  the infimum of  $J_{[b, +\infty)}$  on the set of functions which vanish exactly at  $\alpha \in \mathbb{R}^m$ . Of course, the problem is again to deal with deviations to  $b$ . An important technical issue will come from the fact that the set of functions which vanish exactly at  $\alpha$  has an empty interior, which requires extra care for the large deviation lower bound.

The organisation of the paper will follow the scheme we have just described: in the next section, we detail the orthonormalisation procedure and prove Proposition 3.1. Section 5 and Section 6 will then deal more specifically with the case without outliers. In Section 5, we establish the functional large deviation principles for  $(K^n(\cdot), C^n)$  and  $H^n$  and in Section 6 our main results in this case, namely the large deviation principle for the largest eigenvalues of  $\widetilde{X}_n$  and the almost sure convergence to the minimisers of the rate function. We then turn to the case with outliers in Sections 8 and 9. Therein, the proofs will be less detailed, but we will insist on the points that differ from the previous case. The extension to random matrices  $X_n$  given by classical matrix models is presented in Section 10. To make the core of the paper easier to read, we gather some technical results in Section 11.

#### 4. CHARACTERISATION OF THE EIGENVALUES OF $\widetilde{X}_n$ AS ZEROES OF A FUNCTION $H^n$

As will be seen further, the proof of this proposition is straightforward in the i.i.d. case but more involved in the orthonormalized case and we first detail the orthonormalisation procedure.

##### 4.1. The Gram-Schmidt orthonormalisation procedure.

We start this section by detailing the construction of  $(U_i^n)_{1 \leq i \leq r}$  from  $(G_i^n)_{1 \leq i \leq r}$  in the orthonormalized model. The canonical scalar product in  $\mathbb{C}^n$  will be denoted by  $\langle v, w \rangle =$

$v^*w = \sum_{k=1}^n \overline{v_k} w_k$ , and the associated norm by  $\|\cdot\|_2$ . We also recall that  $\mathbf{H}_r$  is the space of  $r \times r$  either symmetric or Hermitian matrices, according to whether  $G$  is a real or complex random vector.

Fix  $1 \leq r \leq n$  and consider a linearly independent family  $G_1, \dots, G_r$  of vectors in  $\mathbb{C}^n$ . Define their Gram matrix (up to a factor  $n$ )

$$C = [C_{ij}]_{i,j=1}^r, \quad \text{with } C_{ij} = \frac{1}{n} \langle G_i, G_j \rangle.$$

We then define

$$q_1 = 1 \text{ and for } i = 2, \dots, r, \quad q_i := \det[C_{kl}]_{k,l=1}^{i-1}. \quad (6)$$

and the lower triangular matrix  $A = [A_{ij}]_{1 \leq j \leq i \leq r}$  as follows : for all  $1 \leq j < i \leq r$ ,

$$A_{ij} = \frac{\det[\gamma_{k,l}^j]_{k,l=1}^{i-1}}{q_i} \text{ with } \gamma_{kl}^j = \begin{cases} C_{kl}, & \text{si } l \neq j \\ -C_{ki}, & \text{si } l = j. \end{cases} \quad (7)$$

Note that by linear independence of the  $G_i$ 's, none of the  $q_i$ 's is zero so that the matrix  $A$  is well defined.

Then the vectors  $W_1, \dots, W_r$  defined, for  $i = 1, \dots, r$ , by

$$W_i = \sum_{l=1}^i A_{il} \frac{G_l}{\sqrt{n}}$$

are orthogonal and the  $U_i$ 's, defined, for  $i = 1, \dots, r$ , by

$$U_i = \frac{W_i}{\|W_i\|_2}$$

are orthonormal. They are said to be the Gram-Schmidt orthonormalized vectors from  $(G_1, \dots, G_r)$ . The following proposition, which can be easily deduced from the definitions we have just introduced, will be useful in the sequel.

**Property 4.1.** *For each  $i_0 = 1, \dots, r$ , there is a real function  $P_{i_0}$ , defined on  $\mathbf{H}_r$ , polynomial in the entries of the matrix, not depending on  $n$  and nor on the  $G_i$ 's, such that*

$$\|q_{i_0} W_{i_0}\|_2^2 = P_{i_0}(C).$$

*Moreover, the polynomial function  $P_{i_0}$  is positive on the set of positive definite matrices.*

The last assertion of the proposition comes from the fact that any positive definite  $r \times r$  Hermitian matrix is the Gram matrix of a linearly independent family of  $r$  vectors of  $\mathbb{C}^r$  (namely the columns of its square root).

Let  $G$  be a random vector satisfying Assumption 2.2 and  $(G^n(k), k \geq 1)$  be i.i.d copies of  $G$ . Let  $G_i^n = (G^n(k)_i)_{1 \leq k \leq n}$  for  $i \in \{1, \dots, r\}$ . One can easily check that if  $n > r$ , these vectors are almost surely linearly independent, so that we can apply Gram-Schmidt orthonormalisation to this family of random vectors. We define the  $r \times r$  matrices  $C^n, A^n$ , the real number  $q_i^n$  and the vectors  $W_1^n, \dots, W_r^n, U_1^n, \dots, U_r^n$  of  $\mathbb{C}^n$  as above. As announced in Section 1, these  $U_i^n$ 's are the Gram-Schmidt orthonormalized of the  $G_i^n$ 's we used to define our model in the introduction.

## 4.2. Characterization of the eigenvalues of $\widetilde{X}_n$ : proof of Proposition 3.1.

As explained in Section 3, a crucial observation (see [9, Proposition 5.1]) is that the eigenvalues of  $\widetilde{X}_n$  can be characterised as the zeroes of a polynomial function of matrices of size  $r \times r$ . This was stated in Proposition 3.1 which we prove below.

*Proof of Proposition 3.1.* We have, from (3),

$$\begin{aligned} \det(zI_n - \widetilde{X}_n) &= \det(zI_n - X_n - V\Theta V^*) \\ &= \det(zI_n - X_n) \det(\Theta) \det(\Theta^{-1} - V^*(zI_n - X_n)^{-1}V) \end{aligned}$$

Hence any  $z \notin \{\lambda_1^n, \dots, \lambda_r^n\}$  is an eigenvalue of  $\widetilde{X}_n$  if and only if

$$D_n(z) := \det(\Theta^{-1} - V^*(zI_n - X_n)^{-1}V) = 0.$$

In the i.i.d. model, since  $V^*(zI_n - X_n)^{-1}V = K^n(z)$ ,

$$H^n(z) := \det(\Theta^{-1} - V^*(zI_n - X_n)^{-1}V)$$

is actually a polynomial, depending on  $\Theta$ , in the entries of  $K^n(z)$ .

Let us now treat the orthonormalized model, where the Gram-Schmidt procedure makes things a bit more involved. If we denote by  $\mathbf{G}$  the  $n \times r$  matrix with column vectors  $(G_i^n)_{1 \leq i \leq r}$ ,  $D$  the  $r \times r$  diagonal matrix given by  $D = \text{diag}(\|W_1^n\|_2, \dots, \|W_r^n\|_2)$  and  $\Sigma = (A^n)^T$ ,  $V$  is equal to  $n^{-1/2}\mathbf{G}\Sigma D^{-1}$ . As  $K^n(z) = \frac{1}{n}\mathbf{G}^*(zI_n - X_n)^{-1}\mathbf{G}$ , we deduce that

$$D_n(z) = \det(\Theta^{-1} - D^{-1}\Sigma^*K_n(z)\Sigma D^{-1}).$$

Now, if we define  $Q = \text{diag}(q_1^n, \dots, q_r^n)$  (recall (6)),  $E = DQ$ ,  $F = \Sigma Q$  and  $H^n(z) := \det(E^*\Theta^{-1}E - F^*K_n(z)F)$  then on one hand, one can check that

$$D_n(z) = (\det E^*E)^{-1}H^n(z),$$

so that any  $z \notin \{\lambda_1^n, \dots, \lambda_r^n\}$  is an eigenvalue of  $\widetilde{X}_n$  if and only if it is a zero of  $H^n$ .

On the other hand,  $H^n(z)$  is obviously a polynomial (depending only on the matrix  $\Theta$ ) of the entries of  $K^n(z)$ ,  $E^*\Theta^{-1}E$  and  $F$ . Furthermore,  $E^*\Theta^{-1}E$  is a diagonal matrix whose  $i$ -th entry is given by  $(E^*\Theta^{-1}E)_i = \theta_i^{-1}\|q_i^n W_i^n\|_2^2 = \theta_i^{-1}P_i(C^n)$  (by Property 4.1) and  $F_{ij} = \det[\gamma_{k,l}^j]_{k,l=1}^{i-1}$  with  $\gamma_{k,l}^j$  defined in (7). This concludes the proof.  $\square$

## 5. LARGE DEVIATIONS FOR $H^n$ IN THE CASE WITHOUT OUTLIERS

We assume throughout this section that Assumptions 2.1, 2.2 and 2.3 hold.

### 5.1. Statement of the result.

In the sequel,  $\mathcal{K}$  will denote any compact interval included in  $(b, \infty)$ , and we denote by  $z^*$  its upper bound. We equip  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  with the uniform topology which is given by the distance  $d$  defined, for  $(K_1, C_1), (K_2, C_2) \in \mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  by

$$d((K_1, C_1), (K_2, C_2)) = \sup_{z \in \mathcal{K}} \|K_1(z) - K_2(z)\|_2 + \|C_1 - C_2\|_2,$$

where  $\|M\|_2 = \sqrt{\text{Tr}(M^2)}$  for all  $M \in \mathbf{H}_r$ .

With  $G = (g_1, \dots, g_r)$  satisfying Assumption 2.2, we define  $Z$  a matrix in  $\mathbf{H}_r$  such that, for  $i \leq j$ ,  $Z_{ij} = \overline{g_i}g_j$  and  $\Lambda$  given, for any  $H \in \mathbf{H}_r$  by

$$\Lambda(H) = \log \mathbb{E} \left( e^{\text{Tr}(HZ)} \right). \quad (8)$$

The goal of this section is to show the following theorem.

**Theorem 5.1.** (1) *The law of  $((K^n(z))_{z \in \mathcal{K}}, C^n)$ , viewed as an element of the space  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  equipped with the uniform topology, satisfies a large deviation principle in the scale  $n$  and with good rate function  $\mathbf{I}$  defined, for  $K \in \mathcal{C}(\mathcal{K}, \mathbf{H}_r)$  and  $C \in \mathbf{H}_r$  by  $\mathbf{I}(K(\cdot), C) = \infty$  if  $z \mapsto K(z)$  is not Lipschitz continuous on  $\mathcal{K}$ . Otherwise,  $K'$  exists Lebesgue-almost everywhere and*

$$\mathbf{I}(K(\cdot), C) = \sup_{P, X, Y} \left\{ \text{Tr} \left( \int K'(z)P(z)dz + K(z^*)X + CY \right) - \Gamma(P, Y, X) \right\}$$

where  $\Gamma(P, Y, X)$  is given by the formula

$$\Gamma(P, Y, X) = \int \Lambda \left( - \int \frac{1}{(z-x)^2} P(z)dz + \frac{1}{z^* - x} X + Y \right) d\mu(x)$$

and the supremum is taken over piece-wise constant functions  $P$  with values in  $\mathbf{H}_r$  and  $X, Y$  in  $\mathbf{H}_r$ .

(2) *The law of  $(H^n(z))_{z \in \mathcal{K}}$  on  $\mathcal{C}(\mathcal{K}, \mathbb{R})$  equipped with the uniform topology, satisfies a large deviation principle in the scale  $n$  and with rate function given, for a continuous function  $f \in \mathcal{C}(\mathcal{K}, \mathbb{R})$ , by*

$$J_{\mathcal{K}}(f) = \inf \{ \mathbf{I}(K(\cdot), C); (K(\cdot), C) \in \mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r, P_{\Theta, r}(K(z), C) = f(z) \forall z \in \mathcal{K} \}$$

with  $P_{\Theta, r}$  the polynomial function of Proposition 3.1.

Since the map  $(K(\cdot), C) \mapsto (P_{\Theta, r}(K(z), C))_{z \in \mathcal{K}}$  from  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  to  $\mathcal{C}(\mathcal{K}, \mathbb{R})$ , both equipped with their uniform topology, is continuous and  $\mathbf{I}$  is a good rate function, the second part of the theorem is a direct consequence of its first part and the contraction principle [14, Theorem 4.2.1].

The remainder of the section will be devoted to the proof of the first part of the theorem and the study of the properties of the rate function  $\mathbf{I}$ , in particular its minimisers.

## 5.2. Proof of Theorem 5.1.

The strategy will be to establish a LDP for finite dimensional marginals of the process  $((K^n(z))_{z \in \mathcal{K}}, C^n)$  based on [27, Theorem 3.2] that we remind in Theorem 11.1 of the Appendix. From that, we will establish a LDP in the topology of point-wise convergence via the Dawson-Gärtner theorem. As  $((K^n(z))_{z \in \mathcal{K}}, C^n)$  will be shown to be exponentially tight for the uniform topology, the LDP will also hold in this latter topology.

**5.2.1. Exponential tightness.** We start with the exponential tightness, stated in the following lemma. As  $\mathcal{K}$  is a compact subset of  $(b, \infty)$  and the largest eigenvalue  $\lambda_1^n$  tends to  $b$ , there exists  $1 > \varepsilon > 0$  (depending only on  $\mathcal{K}$ ) such that for  $n$  large enough, for any  $z \in \mathcal{K}$  and  $1 \leq i \leq n$ ,  $z - \lambda_i^n > \varepsilon$ . We fix hereafter such an  $\varepsilon$ .

For any  $L > 0$ , we set, for any  $p \geq 1$ ,  $\delta(p) = \frac{2\varepsilon}{pL}$  and define

$$\mathcal{K}_L := \bigcap_{p \geq 1} \left\{ \sup_{z \in \mathcal{K}} \|K(z)\|_2 + \|C\|_2 \leq L, \quad \sup_{z, z' \in \mathcal{K}, |z-z'| \leq \delta(p)} \|K(z) - K(z')\|_2 \leq \frac{1}{p} \right\}.$$

We have

**Lemma 5.2.**

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( ((K^n(z))_{z \in \mathcal{K}}, C^n) \in \mathcal{K}_L^c \right) = -\infty.$$

*In particular, the law of  $((K^n(z))_{z \in \mathcal{K}}, C^n)$  is exponentially tight for the uniform topology on  $\mathcal{C}(\mathcal{K}, \mathbb{H}_r) \times \mathbb{H}_r$ .*

*Proof.* We claim that

$$\left\{ \max_{1 \leq i \leq r} C_{ii}^n \leq \frac{\varepsilon L}{2r} \right\} \subset \left\{ ((K^n(z))_{z \in \mathcal{K}}, C^n) \in \mathcal{K}_L \right\}.$$

Indeed, for  $n$  large enough,

$$|K^n(z)_{ij} - K^n(z')_{ij}| \leq \sqrt{C_{ii}^n C_{jj}^n} \frac{|z - z'|}{\varepsilon^2}$$

so that, for  $|z - z'| \leq \delta(p)$ ,

$$\|K^n(z) - K^n(z')\|_2 \leq r \frac{\delta(p)}{\varepsilon^2} \max_{1 \leq i \leq r} C_{ii}^n.$$

Therefore, if  $\max_{1 \leq i \leq r} C_{ii}^n \leq \frac{\varepsilon L}{2r}$ , then  $\|K^n(z) - K^n(z')\|_2 \leq \frac{1}{p}$  whereas

$$\|C^n\|_2 \leq r \max_{1 \leq i \leq r} C_{ii}^n \leq \frac{\varepsilon L}{2} \leq \frac{L}{2}$$

and  $\|K^n(z)\|_2 \leq \frac{1}{\varepsilon} r \max_{1 \leq i \leq r} C_{ii}^n \leq \frac{L}{2}$ .

Now, by Assumption 2.2, let  $\alpha > 0$  be such that  $C := \mathbb{E} \left( e^{\alpha |g_1|^2} \right) < \infty$ .

$$\mathbb{P} \left( \max_{1 \leq i \leq r} C_{ii}^n > \frac{\varepsilon L}{2r} \right) \leq r \mathbb{P} \left( C_{11}^n > \frac{\varepsilon L}{2r} \right) \tag{9}$$

$$\leq r \mathbb{E} \left( e^{\alpha \sum_k |G_1^n(k)|^2} \right) e^{-n\alpha \frac{\varepsilon L}{2r}} \leq r C^n e^{-n\alpha \frac{\varepsilon L}{2r}} \leq e^{-n\alpha \frac{\varepsilon L}{4r}}, \tag{10}$$

where the last inequality holds for  $n$  and  $L$  large enough. This gives

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( ((K^n(z))_{z \in \mathcal{K}}, C^n) \in \mathcal{K}_L^c \right) = -\infty.$$

By the Arzela-Ascoli theorem,  $\mathcal{K}_L$  is a compact subset of  $\mathcal{C}(\mathcal{K}, \mathbb{H}_r) \times \mathbb{H}_r$  for any  $L > 0$ , from which we get immediately the second part of the lemma.  $\square$

5.2.2. *Large deviation principle for finite dimensional marginals.* We now study the finite dimensional marginals of our process. More precisely, we intend to show the following:

**Proposition 5.3.** *Let  $M$  be a positive integer and  $b < z_1 < z_2 < \dots < z_M$ .*

*The law of  $((K^n(z_i))_{1 \leq i \leq M}, C^n)$  viewed as an element of  $\mathbf{H}_r^{M+1}$  satisfies a large deviation principle in the scale  $n$  with good rate function  $I_M^{z_1, \dots, z_M}$  defined, for  $K_1, \dots, K_M, C \in \mathbf{H}_r$  by*

$$I_M^{z_1, \dots, z_M}(K_1, \dots, K_M, C) = \sup_{\Xi_1, \dots, \Xi_M, Y \in \mathbf{H}_r} \left\{ \text{Tr} \left( \sum_{l=1}^M \Xi_l K_l + Y C \right) - \Gamma_M(\Xi_1, \dots, \Xi_M, Y) \right\},$$

with  $\Gamma_M(\Xi_1, \dots, \Xi_M, Y)$  defined by the formula

$$\Gamma_M(\Xi_1, \dots, \Xi_M, Y) = \int \Lambda \left( \sum_{l=1}^M \frac{1}{z_l - x} \Xi_l + Y \right) d\mu(x),$$

with  $\Lambda$  given by (8).

*Proof.* The proof of the proposition is a direct consequence of Theorem 11.1.1. Indeed, let  $Z_1$  be the  $\mathbf{H}_r$ -valued random variable such that for all  $1 \leq i, j \leq r$ ,

$$(Z_1)_{ij} = \overline{g_i(1)} g_j(1)$$

and we define  $f$  the matrix-valued continuous function with values in  $\mathbb{R}^{[(M+1)r] \times r}$  such that, if we denote by  $I_r$  the identity matrix in  $\mathbf{H}_r$ ,

$$f(x) = \begin{pmatrix} \frac{1}{z_1 - x} I_r \\ \vdots \\ \frac{1}{z_M - x} I_r \\ I_r \end{pmatrix}.$$

Now, if  $(Z_k)_{1 \leq k \leq n}$  are iid copies of  $Z_1$  and if we denote by

$$L_n := \frac{1}{n} \sum_{k=1}^n f(\lambda_k^n) \cdot Z_k = \begin{pmatrix} K^n(z_1) \\ \vdots \\ K^n(z_M) \\ C^n \end{pmatrix},$$

from Theorem 11.1.1, we get that  $L_n$  satisfies an LDP in the scale  $n$  with good rate function  $I_M^{z_1, \dots, z_M}$ .  $\square$

5.2.3. *Large deviation principle for the law of  $((K_n(z))_{z \in \mathcal{K}}, C^n)$ .* The next step is to establish a LDP for the law of  $((K^n(z))_{z \in \mathcal{K}}, C^n)$  associated to the topology of point-wise convergence. The following proposition will be a straightforward application of the Dawson-Gärtner theorem on projective limits.

**Proposition 5.4.** *The law of  $((K^n(z))_{z \in \mathcal{K}}, C^n)$  as an element of  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  equipped with the topology of pointwise convergence satisfies a LDP in the scale  $n$  with good rate function  $\mathbf{J}$  defined as follows : for  $K \in \mathcal{C}(\mathcal{K}, \mathbf{H}_r)$  and  $C \in \mathbf{H}_r$ ,*

$$\mathbf{J}(K, C) = \sup_M \sup_{z_1 < \dots < z_M, z_i \in \mathcal{K}} I_M^{z_1, \dots, z_M}(K(z_1), \dots, K(z_M), C).$$



*Proof.* Let  $\mathcal{J}$  be the collection of all finite subsets of  $\mathcal{K}$  ordered by inclusion. For  $j = \{z_1, \dots, z_{|j|}\} \in \mathcal{J}$  and  $f$  a measurable function from  $\mathcal{K}$  to  $\mathbf{H}_r$ ,  $p_j(f) = (f(z_1), \dots, f(z_{|j|})) \in \mathbf{H}_r^{|j|}$ .

We know from Proposition 5.3 that the law of  $(p_j(K^n), C^n)$  satisfies a LDP with good rate function  $I_{|j|}^{z_1, \dots, z_{|j|}}$ . Moreover, one can check that the projective limit of the family  $\mathbf{H}_r^{|j|} \times \mathbf{H}_r$  is  $\mathbf{H}_r^{\mathcal{K}} \times \mathbf{H}_r$  equipped with the topology of pointwise convergence.

Therefore, the Dawson-Gärtner theorem [14, Theorem 4.6.1] gives us Proposition 5.4.  $\square$

The next step is the following : from Proposition 5.4 and Lemma 5.2, and as the topology of uniform convergence is finer than the topology of pointwise convergence, we can apply [14, Corollary 4.2.6] and get that the law of  $((K^n(z))_{z \in \mathcal{K}}, C^n)$  as an element of  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  equipped with the uniform topology satisfies a LDP in the scale  $n$  with good rate function  $\mathbf{J}$ .

### 5.3. Identification and properties of the rate function.

Now, to complete the proof of Theorem 5.1, the last step is to identify the rate function, that is, show that  $\mathbf{I} = \mathbf{J}$ .

We will consider the operator norm, given, for  $H \in \mathbf{H}_r$ , by  $\|H\|_\infty = \sup \langle u, Hu \rangle$ , where the supremum is taken over vectors  $u \in \mathbb{C}^r$  with norm one. We also use the usual order on Hermitian matrices, i.e.  $H_1 \leq H_2$  if and only if  $H_2 - H_1$  is positive semi-definite (respectively  $H_1 < H_2$  if  $H_2 - H_1$  is positive definite).

We first want to check that  $\mathbf{J}(K(\cdot), C) = \infty$  whenever  $K$  is not Lipschitz continuous. We prove the following more complete lemma, which also proves that  $\mathbf{I}$  is a good rate function<sup>2</sup>. Moreover, some of the items will be useful in the sequel. We recall that  $\Lambda$  was defined in (8).

**Lemma 5.5.** (1)  $H \mapsto \Lambda(H)$  is increasing,  $\Lambda(-H) \leq 0$  if  $H \geq 0$  and for all non null positive semi-definite  $H \in \mathbf{H}_r$ ,

$$\lim_{t \rightarrow +\infty} \Lambda(-tH) = -\infty. \quad (11)$$

(2) If we denote by  $(C^*)_{ij} = \mathbb{E}[\bar{g}_i g_j]$ . Then, for any  $H \in \mathbf{H}_r$ ,

$$\Lambda(H) \geq \text{Tr}(HC^*).$$

(3) There exists  $\gamma > 0$  so that

$$B := \sup_{H: \|H\|_\infty \leq \gamma} \Lambda(H) < \infty.$$

(4) If  $\mathbf{I}(K(\cdot), C)$  or  $\mathbf{J}(K(\cdot), C)$  are finite, then  $C > 0$  and  $K(z) > 0$  for any  $z \in \mathcal{K}$ .

(5) For all  $L$ , there exists a finite constant  $M_L$  so that on  $\{\mathbf{I} \leq L\}$  or  $\{\mathbf{J} \leq L\}$ , we have

$$\sup_{z \in \mathcal{K}} \|K(z)\|_\infty \leq M_L, \quad \|C\|_\infty \leq M_L$$

---

<sup>2</sup>Note that this fact could also be deduced from the fact that  $\mathbf{J}$  is a good rate function because of exponential tightness and the identification  $\mathbf{I} = \mathbf{J}$ .

and for all  $z_1, z_2 \in \mathcal{K}$ ,  $\|K(z_2) - K(z_1)\|_\infty \leq M_L |z_1 - z_2|$ .

In particular,  $K'$  exists almost surely and is bounded by  $M_L$ .

- (6) If  $\mathbf{I}(K(\cdot), C)$  or  $\mathbf{J}(K(\cdot), C)$  are finite, then  $K'(z) \leq 0$  for almost any  $z \in \mathcal{K}$  and for any non zero vector  $e$ , there is no non empty interval where the function  $\langle e, K'(\cdot)e \rangle$  vanishes everywhere.
- (7)  $\mathbf{I}$  is a good rate function.

*Proof.*

- (1) The two first assertions of the first point are just based on the fact that almost surely,  $\text{Tr}(HZ) \geq 0$  if  $H \geq 0$ . Let us now prove the last assertion. Consider  $\eta > 0$  and a non vanishing orthogonal projector  $p \in \mathbf{H}_r$  such that  $H \geq \eta p$ . For all  $t > 0$ , we have

$$0 \leq \mathbb{E}[e^{-t \text{Tr}(HZ)}] \leq \mathbb{E}[e^{-t\eta \text{Tr}(pZ)}] = \mathbb{E}[e^{-t\eta \text{Tr}(pGG^*)}] = \mathbb{E}[e^{-t\eta G^* p G}].$$

Since, by dominated convergence,

$$\lim_{t \rightarrow +\infty} \mathbb{E}[e^{-t\eta G^* p G}] = \mathbb{P}\{G^* p G = 0\} = \mathbb{P}\{G \in \ker p\} = 0$$

(where we used Assumption 2.2 in the last equality), we have

$$\lim_{t \rightarrow +\infty} \Lambda(-tH) = \lim_{t \rightarrow +\infty} \log \mathbb{E}[e^{-t \text{Tr}(HZ)}] = -\infty.$$

- (2) The second point follows from Jensen's inequality.
- (3) The third point is due to the fact that  $\text{Tr}(HZ) \leq \|H\|_\infty \sum_{i=1}^r |g_i|^2$  so that by Hölder's inequality,

$$\Lambda(H) \leq \log \mathbb{E}[e^{\|H\|_\infty \sum_{i=1}^r |g_i|^2}] \leq \frac{1}{r} \sum_{i=1}^r \log \mathbb{E}[e^{\|H\|_\infty r |g_i|^2}]$$

which is finite by Assumption 2.2 if  $\|H\|_\infty r \leq \alpha$ .

To prove (4), (5) and (6), we shall consider  $\mathbf{I}$  rather than  $\mathbf{J}$  but proofs are similar for  $\mathbf{J}$  (even slightly simpler).

- (4) Assume  $\mathbf{I}(K(\cdot), C) \leq L < \infty$ . We first prove that  $C$  is positive. We take  $P, X \equiv 0$  to get

$$\sup_{Y \in \mathbf{H}_r} \{\text{Tr}(CY) - \Lambda(Y)\} \leq \mathbf{I}(K, C) \leq L.$$

Suppose now that there exists some vector  $u \in \mathbb{C}^r$  such that  $\langle u, Cu \rangle \leq 0$ , and define, for any  $t > 0$ ,  $Y_t = -t u u^*$ . Then  $\text{Tr}(CY_t) \geq 0$  so that for all  $t > 0$ ,

$$-\Lambda(Y_t) \leq \text{Tr}(CY_t) - \Lambda(Y_t) \leq L,$$

which contradicts Equation (11).

Let  $z_0 \in \mathcal{K}$ . Suppose now that there exists some vector  $u \in \mathbb{C}^r$  such that

$$\langle u, K(z_0)u \rangle \leq 0,$$

and take  $Y = 0$  and  $P(z) = -\mathbb{1}_{z \geq z_0} X$  so that for all  $X \in \mathbf{H}_r$ ,  $K(z^*)X + \int K'(z)P(z)dz = K(z_0)X$ . Now, for  $t > 0$ , choose  $X_t = -t u u^*$ . To exhibit a contradiction, we can then proceed as above, using that, for  $x$  in the support of  $\mu$ ,  $\frac{1}{z_0 - x}$  is bounded below by some positive constant.

- (5) Assume  $\mathbf{I}(K(\cdot), C) \leq L$ . With  $\gamma$  and  $B$  introduced in the third point, we define  $Y = \pm\gamma uu^*$  and take  $P, X \equiv 0$ . We get

$$\gamma|\langle u, Cu \rangle| \leq B + L$$

for all vector  $u$  with norm one, that is  $\|C\|_\infty \leq \gamma^{-1}(L + B)$ . Similar considerations hold for the bound over  $\|K(z)\|_\infty$ .

We next show that  $K$  is Lipschitz. Let  $u, v$  be two vectors of norm 1. For  $z_1 \leq z_2 \in \mathcal{K}$ , if we take  $P_{z_1, z_2}$  equals to  $(z_1 - z_2)^{-1}(uv^* + vu^*)$  on  $[z_1, z_2]$  and vanishing outside this interval, one then checks that there exists a constant  $\kappa$  which does not depend on  $z_1$  and  $z_2$  nor on  $u$  such that for  $\mu$ -almost all  $x$ , the operator norm of the matrix  $\int \frac{P_{z_1, z_2}(z)}{(z-x)^2} dz$  can be bounded by  $\kappa$ . Let  $\eta_0$  be such that  $\eta_0\kappa \leq \gamma$  (with  $\gamma$  satisfying the third point). Then, for any  $z_1 \leq z_2 \in \mathcal{K}$ ,

$$\text{Tr} \left( \int \eta_0 P_{z_1, z_2}(z) K'(z) dz \right) \leq B + L$$

that is

$$(z_2 - z_1)^{-1} 2\Re \langle u, (K(z_2) - K(z_1)) v \rangle \leq \frac{L + B}{\eta_0}$$

which gives the desired control after optimisation over  $u, v$ .

- (6) Let us prove that for any  $z_1, z_2 \in \mathcal{K}$  such that  $z_1 < z_2$ ,  $K(z_2) \leq K(z_1)$  (dividing by  $z_2 - z_1$  will then give the fact that  $K'$  is negative semi-definite where it is defined). So let us fix  $z_1, z_2 \in \mathcal{K}$  such that  $z_1 < z_2$ . Let us fix  $u \in \mathbb{C}^r \setminus \{0\}$ . For all real number  $t \geq 0$ , we have, for  $P_t(z) := t \mathbb{1}_{[z_1, z_2]}(z) uu^*$  and  $X = Y = 0$ ,

$$I(K(\cdot), C) \geq tu^*(K(z_2) - K(z_1))u - \Gamma(P_t, 0, 0).$$

Note that

$$\Gamma(P_t, 0, 0) = \int \Lambda \left( -t \int_{z_1}^{z_2} \frac{dz}{(z-x)^2} u^* u \right) d\mu(x) \leq 0$$

by (1) of this lemma. Thus for all  $t > 0$ ,

$$I(K(\cdot), C) \geq tu^*(K(z_2) - K(z_1))u.$$

It follows that  $u^*(K(z_2) - K(z_1))u$  is non positive by letting  $t$  going to infinity, which completes the proof of this point. Similarly, if  $u$  is a fixed vector and  $\langle u, K'(\cdot)u \rangle$  vanishes on an interval  $[z_1, z_2]$  with  $z_1 < z_2$ ,  $P_t = t \mathbb{1}_{[z_1, z_2]}(z) uu^*$ , and  $X = Y = 0$ ,

$$I(K(\cdot), C) \geq - \int \Lambda \left( -t \int_{z_1}^{z_2} \frac{dz}{(z-x)^2} uu^* \right) d\mu(x)$$

which goes to infinity as  $t$  goes to infinity by the first point.

- (7) Finally, let us show that  $\mathbf{I}$  is a good rate function, namely that  $\{\mathbf{I} \leq L\}$  is compact as a subset of  $\mathcal{C}(\mathcal{K}, \mathbb{H}_r) \times \mathbb{H}_r$  and that  $\mathbf{I}$  is lower semicontinuous. Clearly  $\mathbf{I}$  is non-negative (take  $P, Y, X = 0$ ). Note first that on this set, by the fifth point  $K$  is uniformly Lipschitz with constant  $M_L$  and is also uniformly bounded by the fourth point (recall  $\mathcal{K}$  is compact), hence it belongs to a compact set by the Ascoli-Arzelà theorem. Similarly,  $C$  being uniformly bounded, we conclude that

$\{\mathbf{I} \leq L\}$  is included into a compact set of  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$ . Moreover, for any piece-wise constant function  $P$ ,  $X$  and  $Y$  in  $\mathbf{H}_r$ ,

$$L_{P,X,Y}(K, C) := \text{Tr} \left( \int P(z)K'(z)dz + CY + K(z^*)X \right)$$

is a continuous function of  $K, C$ . Hence,  $\mathbf{I}$  is lower semicontinuous as the supremum of such functions, that is  $\{\mathbf{I} \leq L\}$  is closed. □

We can now identify the rate function itself. Let  $z \mapsto K(z)$  be Lipschitz continuous on  $\mathcal{K}$  taking values in  $\mathbf{H}_r$  and  $C \in \mathbf{H}_r$ .

For any  $M$ , the map  $(X_1, \dots, X_M, Y) \mapsto (-X_1, -(X_1 + X_2), \dots, -(X_1 + \dots + X_M), Y)$  being bijective on  $\mathbf{H}_r^{M+1}$ , one can rewrite, for any  $b < z_1 < \dots < z_M$ ,

$$\begin{aligned} & I_M^{z_1, \dots, z_M}(K(z_1), \dots, K(z_M), C) \\ &= \sup_{\Xi_1, \dots, \Xi_{M-1}, X, Y} \left\{ \text{Tr} \left( \sum_{k=1}^{M-1} (K(z_{k+1}) - K(z_k))\Xi_k + K(z_M)X + CY \right) \right. \\ & \quad \left. - \int \Lambda \left( \sum_{k=1}^{M-1} \left( \frac{1}{z_{k+1} - x} - \frac{1}{z_k - x} \right) \Xi_k + \frac{1}{z_M - x} X + Y \right) d\mu(x) \right\}. \end{aligned} \quad (12)$$

Now, if we define the piece-wise constant function equal to 0 on  $(-\infty, z_1)$ ,  $\Xi_k$  on  $[z_k, z_{k+1})$  for  $1 \leq k \leq M-1$  and  $X$  on  $[z_M, z^*]$ , the functional inside the supremum above is equal to

$$\text{Tr} \left( \int K'(z)P(z)dz + K(z^*)X + CY \right) - \int \Lambda \left( - \int \frac{P(z)}{(z-x)^2} dz + \frac{X}{z^* - x} + Y \right) d\mu(x).$$

Hence the equality  $\mathbf{I} = \mathbf{J}$  is straightforward.

#### 5.4. Study of the minimisers of $\mathbf{I}$ .

We characterise the minima of  $\mathbf{I}$  as follows :

**Lemma 5.6.** *For any compact set  $\mathcal{K}$  of  $(b, \infty)$ , the unique minimizer of  $\mathbf{I}$  on  $\mathcal{C}(\mathcal{K}, \mathbf{H}_r) \times \mathbf{H}_r$  is the pair  $(K^*, C^*)$  given, for  $1 \leq i, j \leq r$ , by*

$$(K^*(z))_{ij} = \int \frac{(C^*)_{ij}}{z - \lambda} d\mu(\lambda), \quad \text{for } z \in \mathcal{K} \quad \text{and} \quad (C^*)_{ij} = \mathbb{E}[\overline{g_i}g_j].$$

*Proof.*  $\mathbf{I}$  vanishes at its minimisers (as a good rate function) and therefore a minimizer  $(K, C)$  satisfies

$$\text{Tr} \left( \int K'(z)P(z)dz + K(z^*)X + CY \right) \leq \Gamma(P, X, Y) \quad (13)$$

for all  $P, X, Y$ .

Now, for any fixed  $(P, X, Y)$ , there exists  $\varepsilon_0 > 0$  such that for any for any  $0 < \varepsilon < \varepsilon_0$ , for

any  $x$  in the support of  $\mu$  we have

$$\varepsilon \left\| - \int \frac{1}{(z-x)^2} P(z) dz + \frac{1}{z^* - x} X + Y \right\|_{\infty} < \alpha,$$

with  $\alpha$  given by Assumption 2.2. Therefore, there exists a constant  $L$  such that for any  $x$  in the support of  $\mu$

$$\left| \mathbb{E} \left( e^{\varepsilon \operatorname{Tr} \left( - \int \frac{1}{(z-x)^2} P(z) dz + \frac{1}{z^* - x} X + Y \right) Z} \right) - \mathbb{E} \left( 1 + \varepsilon \operatorname{Tr} \left( \left( - \int \frac{1}{(z-x)^2} P(z) dz + \frac{1}{z^* - x} X + Y \right) C^* \right) \right) \right| \leq \varepsilon^2 L,$$

so that

$$\Gamma(\varepsilon P, \varepsilon X, \varepsilon Y) = \varepsilon \operatorname{Tr} \left( \int (K^*)'(z) P(z) dz + K^*(z^*) X + C^* Y \right) + O(\varepsilon^2).$$

As a consequence, for any minimizer  $(K, C)$ , we find after replacing  $(P, X, Y)$  by  $\varepsilon(P, X, Y)$ , using (13) and letting  $\varepsilon$  going to zero,

$$\operatorname{Tr} \left( \int K'(z) P(z) dz + K(z^*) X + C Y \right) \leq \operatorname{Tr} \left( \int (K^*)'(z) P(z) dz + K^*(z^*) X + C^* Y \right).$$

Changing  $(P, X, Y)$  in  $-(P, X, Y)$  gives the equality. This implies that

$$C = C^*, \quad K' = (K^*)' \quad \text{a.s. and} \quad K(z^*) = K^*(z^*)$$

and therefore  $(K, C) = (K^*, C^*)$ .  $\square$

## 6. LARGE DEVIATIONS FOR THE LARGEST EIGENVALUES IN THE CASE WITHOUT OUTLIERS

We again assume throughout this section that Assumptions 2.1, 2.2 and 2.3 hold.

### 6.1. Statement of the main result.

For any  $\varepsilon > 0$ , we define the compact set  $\mathcal{K}_\varepsilon := [b + \varepsilon, \varepsilon^{-1}]$ . Let  $s := \operatorname{sign}(\prod_{i=1}^r \theta_i) = (-1)^{r-m}$ .

For  $x \in \mathbb{R}$ , we set  $\mathbb{R}_\downarrow^p(x) = \{(\alpha_1, \dots, \alpha_p) \in \mathbb{R}^p / \alpha_1 \geq \dots \geq \alpha_p \geq x\}$ .

We also denote by  $\omega(g) := \sup_{x \neq y} \frac{|g(x) - g(y)|}{|x - y|} \in [0, \infty]$  the modulus of continuity of a function  $g$ .

For any  $\varepsilon, \gamma > 0$ , and  $\alpha \in \mathbb{R}_\downarrow^p(b + \varepsilon)$ , we put

$$S_{\alpha, \gamma}^\varepsilon := \left\{ f \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R}) : \exists g \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R}) \text{ with } \gamma \leq g \leq \frac{1}{\gamma}, \omega(g) \leq \frac{1}{\gamma} \right. \\ \left. \text{and } f(z) = s.g(z) \prod_{i=1}^p (z - \alpha_i) \right\}$$

Note that in the latter product, the  $\alpha_i$ 's appear with multiplicity. We also denote by

$$C_{k,\gamma}^\varepsilon := \left\{ f \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R}) : \exists p \text{ polynomial of degree } m - k \text{ with } m - k \text{ roots in } \mathcal{K}_\varepsilon \right.$$

and dominant coefficient 1,  $g \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R})$  with  $\gamma \leq g \leq \frac{1}{\gamma}$ ,  $\omega(g) \leq \frac{1}{\gamma}$  and  $f(z) = s.g(z)p(z)$   $\left. \right\}$

and

$$C_\gamma^\varepsilon = \bigcup_{0 \leq k \leq m} C_{k,\gamma}^\varepsilon.$$

We have the following theorem.

**Theorem 6.1.** *Under Assumptions 2.1, 2.2 and 2.3, the law of the  $m$  largest eigenvalues  $\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n$  of  $\tilde{X}_n$  satisfies a large deviation principle in  $\mathbb{R}^m$  in the scale  $n$  and with good rate function  $L$ , defined as follows. For  $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$ , we take  $\alpha_{m+1} = b$  and*

$$L(\alpha) = \begin{cases} \lim_{\varepsilon \downarrow 0} \inf_{\cup_{\gamma > 0} S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon} & \text{if } \alpha \in \mathbb{R}_{\downarrow}^m(b), \alpha_{m-k+1} = b \text{ and } \alpha_{m-k} > b \\ & \text{for a } k \in \{0, \dots, m-1\} \text{ or } \alpha_1 = b, \\ +\infty & \text{otherwise.} \end{cases}$$

**Remark 6.2.** *The function  $L$  is well defined. Indeed, one can easily notice that for all  $\alpha \in \mathbb{R}_{\downarrow}^m(b)$  such that for some  $k \in \{0, \dots, m\}$ ,  $\alpha_{m-k+1} = b$  and  $\alpha_{m-k} > b$ , the map*

$$\varepsilon \longmapsto \inf \{ J_{\mathcal{K}_\varepsilon}(f) ; f \in \cup_{\gamma > 0} S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon \}$$

*is increasing, so that its limits as  $\varepsilon$  decreases to zero exists.*

**Remark 6.3.** *Note that  $J_{\mathcal{K}_\varepsilon}(f)$  is infinite if  $f$  has more than  $r$  zeroes greater than  $b$ . Indeed, by definition, if  $J_{\mathcal{K}_\varepsilon}(f)$  is finite,*

$$f(z) = P_{\Theta, r}(K(z), C) = a \det(A - K(z))$$

*with a non-vanishing constant  $a$  and a self-adjoint matrix  $A$  with eigenvalues  $(\theta_1^{-1}, \dots, \theta_r^{-1})$  and a function  $K$  with values in  $r \times r$  positive self-adjoint matrices so that  $K' \leq 0$  by Lemma 5.5. We may assume without loss of generality that  $f$  vanishes at a point  $x > b$ , since otherwise we are done, so that there exists a non zero  $e \in \mathbb{C}^r$  so that  $K(x)e = Ae$ . There is at most one  $x$  at which  $K(x)e = Ae$ ; otherwise,  $\langle e, K'(\cdot)e \rangle$  would vanish on a non trivial interval which is impossible by Lemma 5.5.6. Moreover, if we let  $P$  be the orthogonal projection onto the orthocomplement of  $e$ ,  $H(z) = \det((1 - P)(A - K(z))(1 - P)) \det(PAP - PK(z)P)$  so that  $H$  vanishes at  $x$  and at the zeroes of  $\det(PAP - PK(z)P)$ . But  $PAP$  and  $PK(z)P$  have the same properties as  $A$  and  $K(z)$  except they have one dimension less. Thus, we can proceed by induction and see that  $f$  can vanish at at most  $r$  points.*

The minimisers are described by the following result.

**Theorem 6.4.** *If we define on  $(b, \infty)$*

$$H(z) = P_{\Theta, r}(K^*(z), C^*)$$

*where  $(K^*, C^*)$  are given in Lemma 5.6 and  $P_{\Theta, r}$  is defined in Proposition 3.1, there exists  $k \in \{0, \dots, m\}$  such that  $H$  has  $m - k$  zeroes  $(\lambda_1^*, \dots, \lambda_{m-k}^*)$  (counted with multiplicity).*

The unique point of  $\mathbb{R}^m$  on which  $L$  vanishes is  $(\lambda_1^*, \dots, \lambda_{m-k}^*, b, \dots, b)$  and consequently  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n)$  converges almost surely to this point as  $n$  grows to infinity.

**Remark 6.5.** In the case when  $(g_1, \dots, g_r)$  are independent centered variables with variance one, one can check that  $C^* = I_r$ ,  $K^*(z) = \int \frac{1}{z-x} d\mu(x) \cdot I_r$  and

$$H(z) = \prod_{i=1}^r \left( \frac{1}{\theta_i} - \int \frac{1}{z-x} d\mu(x) \right)$$

so that we recover [9, Theorem 2.1].

## 6.2. Preliminary remarks and strategy of the proof.

Let us first notice that at most  $m$  eigenvalues of  $\widetilde{X}_n$  can deviate from the bulk since by Weyl's interlacing inequalities

$$\tilde{\lambda}_{m+1}^n \leq \lambda_1^n,$$

which converges to  $b$  as  $n$  goes to infinity.

Secondly, let us state the following lemma.

**Lemma 6.6.** *The law of the sequence  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n)$  of the  $m$  largest eigenvalues of  $\widetilde{X}_n$  is exponentially tight in the scale  $n$ .*

*Proof.* Let us define  $R_n := \widetilde{X}_n - X_n$  and let us denote by  $\|R_n\|_\infty$  the operator norm of the perturbation matrix  $R_n$ . Note that for all  $k$ ,

$$\lambda_k^n - \|R_n\|_\infty \leq \tilde{\lambda}_k^n \leq \lambda_k^n + \|R_n\|_\infty.$$

Since for any fixed  $k$ , the non random sequence  $\lambda_k^n$  tends to  $b$  as  $n$  tends to infinity, it suffices to prove that

$$\limsup_{L \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\|R_n\|_\infty \geq L) = -\infty. \quad (14)$$

For the orthonormalized model, since  $\|R_n\|_\infty = \max\{\theta_1, -\theta_r\}$ , (14) is clear. In the i.i.d. model, we have, for  $\theta := \max_{1 \leq i \leq r} |\theta_i|$ ,

$$\|R_n\|_\infty = \sup_{\|v\|_2=1} |\langle v, R_n v \rangle| \leq \frac{1}{n} \sum_{i=1}^r \theta \|G_i^m\|^2 = \frac{\theta}{n} \sum_{k=1}^n \sum_{i=1}^r |g_i(k)|^2.$$

It implies, by Tchebychev's inequality, that

$$\mathbb{P}(\|R_n\|_\infty \geq L) \leq e^{-\frac{n\alpha L}{\theta}} \mathbb{E} \left[ \exp \left( \alpha \sum_{k=1}^n \sum_{i=1}^r |g_i(k)|^2 \right) \right] = e^{-\frac{n\alpha L}{\theta}} \mathbb{E} \left[ \exp \left( \alpha \sum_{i=1}^r |g_i(1)|^2 \right) \right]^n$$

which allows to conclude by Assumption 2.2.  $\square$

As the law of  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_m^n)$  is exponentially tight, the proof of Theorem 6.1 reduces to establishing a weak LDP. In virtue of [14, Theorem 4.1.11] (see also [1, Corollary D.6]), this weak LDP (and the fact that  $L$  is a rate function) will be a direct consequence of Equation (20) and Lemma 6.8 below. The fact that  $L$  is a good rate function is then implied by exponential tightness [14, Lemma 1.2.18].

Before getting into the proof itself in Section 6.4, we first need to know more about  $H^n$ .



### 6.3. The structure of $H^n$ .

From Proposition 3.1, we know that the  $\tilde{\lambda}_i^n$ 's are essentially the zeroes of  $H^n$ . However,  $H^n$  could a priori have other zeroes than these eigenvalues or take very small values. To control this point, we need to understand better the structure of  $H^n$  and more precisely, we intend to show the following fact

**Lemma 6.7.** *For any  $\varepsilon > 0$  small enough, there exists a positive integer  $n_0(\varepsilon)$ ,  $L(\varepsilon) > 0$  and a deterministic function  $g_n$  such that for any  $z \in \mathcal{K}_\varepsilon$  and  $n \geq n_0(\varepsilon)$ ,*

$$H^n(z) = \begin{cases} s \prod_{i=1}^r \|q_i^n W_i^n\|_2^2 g_n(z) \prod_{i=1}^m (z - \tilde{\lambda}_i^n) & \text{in the orthonormalized model,} \\ s g_n(z) \prod_{i=1}^m (z - \tilde{\lambda}_i^n) & \text{in the i.i.d. model,} \end{cases} \quad (15)$$

$$\text{with } s = (-1)^{r-m}, L(\varepsilon) \leq g_n \leq \frac{1}{L(\varepsilon)} \text{ and } \omega(g_n) \leq \frac{1}{L(\varepsilon)}. \quad (16)$$

In particular, for any  $\varepsilon > 0$ ,

$$\limsup_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( (H^n(z))_{z \in \mathcal{K}_\varepsilon} \in (C_\gamma^\varepsilon)^c \right) = -\infty, \quad (17)$$

and

$$\limsup_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \left( \prod_{i=1}^m (z - \tilde{\lambda}_i^n)^{-1} H^n(z) \right)_{z \in \mathcal{K}_\varepsilon} \in (C_{\gamma,0}^\varepsilon)^c \right) = -\infty. \quad (18)$$

*Proof.* Note that once (16) will be established, (17) and (18) are then a direct consequence of Lemma 11.3. Let us define the random sequence

$$c_n := \begin{cases} s \prod_{i=1}^r \|q_i^n W_i^n\|_2^2 & \text{in the orthonormalized model,} \\ s & \text{in the i.i.d. model.} \end{cases}$$

Going back to the proof of Proposition 3.1, one can easily see that, for any  $z \notin \{\lambda_1^n, \dots, \lambda_n^n\}$ ,

$$H^n(z) = c_n \prod_{i=1}^r \theta_i^{-1} \det(zI_n - X_n)^{-1} \det \left( zI_n - X_n - \sum_{i=1}^r \theta_i U_i^n (U_i^n)^* \right).$$

We can rewrite the above as

$$H^n(z) = c_n g_n(z) \prod_{i=1}^m (z - \tilde{\lambda}_i^n)$$

with

$$g_n(z) := \prod_{i=1}^r |\theta_i|^{-1} \frac{1}{\prod_{i=1}^m (z - \lambda_i^n)} \prod_{i=m+1}^n \left( 1 + \frac{\lambda_i^n - \tilde{\lambda}_i^n}{z - \lambda_i^n} \right)$$

Now, let  $\varepsilon > 0$  be fixed.

As  $\mathcal{K}_\varepsilon$  is compact and the  $\lambda_i^n$  belong to a fixed compact, for  $\varepsilon > 0$  small enough, for any  $i$  and  $z \in \mathcal{K}_\varepsilon$  we have  $z - \lambda_i^n \leq \frac{2}{\varepsilon}$  and  $|\lambda_i^n| \leq \frac{2}{\varepsilon}$  so that

$$0 \leq \sum_{i=m+1}^n (\lambda_{i-m}^n - \lambda_i^n) = \sum_{i=1}^m \lambda_i^n - \sum_{i=n-m}^n \lambda_i^n \leq 2m \frac{2}{\varepsilon}. \quad (19)$$

We choose  $n_0(\varepsilon)$  such that for  $n \geq n_0(\varepsilon)$  and any  $i$  and  $z \in \mathcal{K}_\varepsilon$  we have  $\frac{\varepsilon}{2} \leq z - \lambda_i^n$  so that as  $z - \lambda_i^n \leq \frac{2}{\varepsilon}$ ,

$$0 \leq \frac{\lambda_{i-m}^n - \lambda_i^n}{z - \lambda_i^n} = 1 - \frac{z - \lambda_{i-m}^n}{z - \lambda_i^n} \leq 1 - \frac{\varepsilon^2}{4}.$$

Now, using Weyl's interlacing properties, we have for any  $i \geq m + 1$ ,

$$\tilde{\lambda}_i^n \leq \lambda_{i-m}^n,$$

so that

$$\lambda_i^n - \tilde{\lambda}_i^n \geq -(\lambda_{i-m}^n - \lambda_i^n).$$

For  $0 \leq x \leq 1 - \frac{\varepsilon^2}{4}$ ,  $\log(1 - x) \geq -\frac{4}{\varepsilon^2}x$ , so that we finally get by (19),

$$g_n(z) \geq \prod_{i=1}^r |\theta_i|^{-1} \left(\frac{\varepsilon}{2}\right)^m e^{\frac{4}{\varepsilon^2} \sum_{i=m+1}^n \frac{\lambda_i^n - \tilde{\lambda}_i^n}{z - \lambda_i^n}} \geq \prod_{i=1}^r |\theta_i|^{-1} \left(\frac{\varepsilon}{2}\right)^m e^{-2m \left(\frac{4}{\varepsilon^2}\right)^2}.$$

By very similar arguments (using  $\log(1 + x) \leq x$  for  $x \geq 0$ ), one can also check that for any  $n \geq n_0(\varepsilon)$ ,

$$g_n(z) \leq \prod_{i=1}^r |\theta_i|^{-1} \left(\frac{2}{\varepsilon}\right)^m e^{\frac{4}{\varepsilon^2} 2m}$$

To finish the proof, we have to verify the uniform equicontinuity of  $g_n$  on  $\mathcal{K}_\varepsilon$ . For any  $i$ , we have

$$\left| \frac{1}{z - \lambda_i^n} - \frac{1}{z' - \lambda_i^n} \right| \leq \left(\frac{2}{\varepsilon}\right)^2 |z - z'|$$

Moreover, for any  $f, g$  continuous on  $\mathcal{K}_\varepsilon$ ,  $\omega(fg) \leq \|f\|_\infty \omega(g) + \|g\|_\infty \omega(f)$  so that

$$\omega(g_n) \leq \prod_{i=1}^r |\theta_i|^{-1} \left( m \left(\frac{2}{\varepsilon}\right)^{m+1} e^{\frac{8}{\varepsilon^2} m} + \left(\frac{2}{\varepsilon}\right)^{m+2} e^{\frac{8}{\varepsilon^2} m} \sum_{i=m+1}^n |\lambda_i^n - \tilde{\lambda}_i^n| \right)$$

Now, by Weyl's interlacing inequalities, we have

$$\lambda_{i+r-m}^n \leq \tilde{\lambda}_i^n \leq \lambda_{i-m}^n$$

so that  $|\lambda_i^n - \tilde{\lambda}_i^n| \leq \lambda_{i-m}^n - \lambda_{i+r-m}^n$  and  $\sum_{i=m+1}^n |\lambda_i^n - \tilde{\lambda}_i^n| \leq r \frac{2}{\varepsilon}$ .  $\square$

#### 6.4. Core of the proof.

First, from what we said in the preliminary remarks and the fact that the  $\tilde{\lambda}_i^n$  are ordered in decreasing order, we obviously have that if  $\alpha \notin \mathbb{R}_\downarrow^m(b)$ , one has

$$\begin{aligned} \limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \bigcap_{1 \leq i \leq m} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \right) \\ = \liminf_{\delta \downarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \bigcap_{1 \leq i \leq m} \{|\tilde{\lambda}_i^n - \alpha_i| < \delta\} \right) = -\infty. \end{aligned} \quad (20)$$

The weak LDP will then be a direct consequence of the following lemma, with  $k$  the numbers of eigenvalues going to  $b$ ,

**Lemma 6.8.** *Let  $\alpha \in \mathbb{R}_\downarrow^m$  and  $k$  between 0 and  $m$  such that  $\alpha_{m-k+1} = \dots = \alpha_m = b$  and  $\alpha_{m-k} > b$  if  $k < m$ . We have*

$$\begin{aligned} \lim_{\varepsilon \downarrow 0} \limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon\} \right) \\ = \lim_{\varepsilon \downarrow 0} \liminf_{\delta \downarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon\} \right) = -L(\alpha), \end{aligned}$$

with the obvious convention that  $\bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon\} = \Omega$  if  $k = 0$ .

*Proof.* Let  $\delta$  and  $\varepsilon$  positive small enough so that  $\alpha_{m-k} - \delta \geq b + 2\varepsilon$ . In particular,  $\bigcap_{i=1}^{m-k} [\alpha_i - \delta, \alpha_i + \delta] \subset \mathcal{K}_\varepsilon$ . On the set  $\bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon\}$ , for all  $i \leq m-k$ ,  $\tilde{\lambda}_i^n$  is in  $\mathcal{K}_\varepsilon$ . On the other hand, for  $n$  large enough,  $\{\lambda_1^n, \dots, \lambda_n^n\} \cap \mathcal{K}_\varepsilon = \emptyset$ . Therefore,  $\tilde{\lambda}_i^n \notin \{\lambda_1^n, \dots, \lambda_n^n\}$  for  $i \in \{1, \dots, m\}$  and, by Proposition 3.1, is a zero of  $H^n$ .

Let us next prove the large deviation upper bound and fix  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_{m-k} > b$ . A function  $f \in C_{\gamma, k}^\varepsilon$  which vanishes within a distance  $\delta$  of  $(\alpha_i)_{1 \leq i \leq m-k}$  with  $\delta < \alpha_{m-k} - b$  belongs to the set

$$\begin{aligned} B_{\alpha, \gamma, \delta}^\varepsilon := \left\{ f \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R}) : \exists g \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R}) \text{ with } \frac{1}{\gamma} \leq g \leq \gamma, \omega(g) \leq \frac{1}{\gamma} \right. \\ \left. \text{and } f(z) = s.g(z) \prod_{i=1}^{m-k} (z - \beta_i) \text{ with } \forall i \leq m-k, \beta_i \in [\alpha_i - \delta, \alpha_i + \delta] \right\} \end{aligned} \quad (21)$$

Moreover, writing  $H^n(z) = h^n(z) \prod_{i=1}^m (z - \tilde{\lambda}_i^n)$ , clearly  $H^n$  belongs to  $C_{\gamma, k}^\varepsilon$  as soon as  $h^n \in C_{\gamma', 0}^\varepsilon$  and  $\bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon - \varepsilon'\}$  if  $\gamma' \cdot (\varepsilon')^m > \gamma$  and

$\varepsilon' < \varepsilon$ . As a consequence, we can write

$$\mathbb{P} \left( \bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon - \varepsilon'\} \right) \leq \mathbb{P}(H^n \in B_{\alpha, \gamma, \delta}^\varepsilon) + \mathbb{P}(h^n \in (C_{\gamma', 0}^{\varepsilon'})^c).$$

As a consequence, by [14, Lemma 1.2.15],

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon - \varepsilon'\} \right) \leq \max \left\{ \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(H^n \in B_{\alpha, \gamma, \delta}^\varepsilon); \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(h^n \in (C_{\gamma', 0}^{\varepsilon'})^c) \right\}, \quad (22)$$

Moreover,  $B_{\alpha, \gamma, \delta}^\varepsilon$  is a closed subset of  $\mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R})$ . Indeed, if we take a converging sequence  $f_n(z) = sg_n(z) \prod_{i=1}^{m-k} (z - \beta_i^n)$ , since the  $\beta_i^n, n \geq 0$  belongs to compacts and the  $g_n, n \geq 0$  are tight by Ascoli-Arzelà's theorem, we can always assume up to extraction that  $g_n$  and  $\beta_i^n, 1 \leq i \leq m-k$  converge so that the limit of  $f_n$  belongs to  $B_{\alpha, \gamma, \delta}^\varepsilon$ .

Since  $J_{\mathcal{K}_\varepsilon}$  is a good rate function and  $\bigcap_{\delta > 0} B_{\alpha, \gamma, \delta}^\varepsilon = S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon$ , Theorem 5.1 gives with [14, Lemma 4.1.6] that

$$\limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(H^n \in B_{\alpha, \gamma, \delta}^\varepsilon) \leq - \inf_{S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon}. \quad (23)$$

Taking  $\gamma' = \gamma'_0$  small enough, (18) and (22) give for  $\gamma/(\varepsilon')^m < \gamma'_0$ ,

$$\limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + \varepsilon - \varepsilon'\} \right) \leq - \inf_{S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon} \leq - \inf_{\cup_{\gamma > 0} S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon}.$$

We can finally take  $\gamma = 0$  (nothing depends on it anymore),  $\varepsilon - \varepsilon'$  going to zero, as the left hand side obviously decreases as  $\varepsilon - \varepsilon'$  decreases to 0 and, as we already mentioned it in Remark 6.2, the right hand side increases as  $\varepsilon$  decreases to 0.

We turn to the lower bound, which is a bit more delicate. Let us again consider  $\delta$  and  $\varepsilon$  small enough so that  $\bigcap_{i=1}^{m-k} [\alpha_i - \delta, \alpha_i + \delta] \subset \mathcal{K}_\varepsilon$ . As  $J_{\mathcal{K}_\varepsilon}$  is a good rate function, for all  $\gamma > 0$ , the infimum  $\inf_{S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon}$  is achieved, say at  $f_\gamma^{k, \varepsilon}$ . To complete the proof, we need the following lemma,

**Lemma 6.9.** *Let  $\varepsilon, \gamma$  be fixed and small enough. There exists  $\delta_0$  such that for any  $\delta \leq \delta_0$ , there exists  $\delta'$  such that for any  $n$ ,*

$$\left\{ \tilde{\lambda}_1^n \leq \varepsilon^{-1} \right\} \cap \left\{ \sup_{x \in \mathcal{K}_\varepsilon} |H^n(x) - f_\gamma^{k, \varepsilon}(x)| < \delta' \right\} \subset \bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + 2\varepsilon\}.$$

We postpone the proof of this lemma to the next subsection but it is easy to convince oneself that this lemma holds in the particular case when  $k = 0$  and all the multiplicities

of the  $\alpha_i$ 's are one.

To prove the lower bound in Theorem 5.1, we may assume without loss of generality that

$$J := \lim_{\varepsilon \downarrow 0} \inf_{\cup_{\gamma > 0} S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon} < \infty.$$

Let  $\eta > 0$  be fixed. As

$$\inf_{\cup_{\gamma > 0} S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon} = \inf_{\gamma > 0} \inf_{S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon} = \inf_{\gamma > 0} J_{\mathcal{K}_\varepsilon}(f_\gamma^{k, \varepsilon}),$$

we can choose  $\varepsilon, \gamma$  small enough so that  $J_{\mathcal{K}_\varepsilon}(f_\gamma^{k, \varepsilon}) \leq J + \eta$ . By Lemma 6.6, there exists  $L(\varepsilon)$  going to infinity as  $\varepsilon$  goes to zero so that for  $n$  large enough,

$$\mathbb{P}\left(\tilde{\lambda}_1^n \geq \varepsilon^{-1}\right) \leq e^{-nL(\varepsilon)}.$$

We choose  $\varepsilon$  small enough so that  $L(\varepsilon) > J + 2\eta$ .

The above lemma implies, that for  $\delta \leq \delta_0$ , for  $\delta'$  small enough, for  $n$  large enough,

$$\begin{aligned} & \mathbb{P}\left(\bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + 2\varepsilon\}\right) \\ & \geq \mathbb{P}\left(\sup_{x \in \mathcal{K}_\varepsilon} |H^n(z) - f_\gamma^{k, \varepsilon}(z)| < \delta'\right) - \mathbb{P}\left(\tilde{\lambda}_1^n \geq \varepsilon^{-1}\right) \geq \frac{1}{2}e^{-n(J+2\eta)} \end{aligned}$$

the last inequality following from Theorem 5.1.2. As  $\eta$  can be chosen as small as we want, we have

$$\begin{aligned} \liminf_{\delta \downarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}\left(\bigcap_{1 \leq i \leq m-k} \{|\tilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\tilde{\lambda}_i^n \leq b + 2\varepsilon\}\right) \\ \geq - \inf_{\cup_{\gamma > 0} S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon} J_{\mathcal{K}_\varepsilon}, \end{aligned}$$

and one concludes by the monotonicity of the right hand side with respect to  $\varepsilon$ .  $\square$

**6.5. Proof of Lemma 6.9.** Let  $\varepsilon > 0$  be fixed and  $n_0(\varepsilon)$  as in Lemma 6.7. By definition of  $S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^\varepsilon$ , we know that for any  $z \in \mathcal{K}_\varepsilon$ ,

$$f_\gamma^{k, \varepsilon}(z) = s \cdot g(z)P(z), \quad \text{with } P(z) = \prod_{i=1}^{m-k} (z - \alpha_i),$$

with  $\alpha_i \in \mathcal{K}_\varepsilon$  and  $g \geq \gamma > 0$ .

On the other hand, by Lemma 6.7, for  $n \geq n_0(\varepsilon)$ , there exists a (random) positive number  $c_n$  such that for any  $z \in \mathcal{K}_\varepsilon$ ,

$$H^n(z) = s \cdot c_n g_n(z) P_n(z), \quad \text{with } P_n(z) = \prod_{i=1}^m (z - \tilde{\lambda}_i^n), \quad (24)$$

where  $g_n$  is uniformly bounded above and below. Without loss of generality, one can assume that  $s = 1$  and  $g_n \geq \gamma$  and we choose  $M > 1$  such that  $g_n \leq M$ ,  $g \leq M$  and

$\varepsilon^{-1} - b \leq M$ . We also restrict ourselves to the case where the eigenvalues are smaller than  $\varepsilon^{-1}$  so that if they do not belong to  $\mathcal{K}_\varepsilon$ , they are smaller than  $b + \varepsilon$ .

We also denote by

$$\ell := \min \left( 1, \frac{1}{3} \min_{\alpha_i \neq \alpha_j} |\alpha_i - \alpha_j|, \frac{1}{3} |\alpha_{m-k} - b| \right)$$

and  $\eta := \frac{\varepsilon^{-1} - (b + \varepsilon)}{2(m+1)} \wedge 1$ .

We fix hereafter  $\delta > 0$  such that

$$\delta < \frac{\gamma^2}{2} \left( \frac{\eta \ell}{2} \right)^m \frac{1}{M^{2m+1}}.$$

Obviously, as  $g$  is bounded below by  $\gamma$ , one can choose  $\delta'_1$  such that for any  $\delta' < \delta'_1$ , any  $f$  satisfying that  $\sup_{x \in \mathcal{K}_\varepsilon} |f(x) - f_\gamma^{k,\varepsilon}(x)| < \delta'$  vanishes only on  $\cup_{i=1}^{m-k} [\alpha_i - \delta, \alpha_i + \delta]$ .

Let  $\delta'_2$  be such that

$$0 < \delta'_2 \leq \delta^m \left( \frac{\gamma^2}{2} \frac{(\eta \ell)^m}{M^{m+1}} - \delta(2M)^m \right)$$

Let now  $\delta'$  be strictly smaller than  $\min(\delta'_1, \delta'_2, \frac{\gamma \eta^m}{2})$  and  $H^n$  be a function satisfying (24) and  $\sup_{x \in \mathcal{K}_\varepsilon} |H^n(x) - f_\gamma^{k,\varepsilon}(x)| < \delta'$ .

We first claim that  $c_n$  is bounded below uniformly on this set. Indeed, if we cut  $\mathcal{K}_\varepsilon$  into  $m + 1$  intervals of size  $2\eta$ , at least one of them does not contain any zero of  $P$ . We call  $x_0$  the centre of this interval. Then  $|g(x_0)P(x_0)| \geq \gamma \eta^m$ . Therefore  $|c_n g_n(x_0) P_n(x_0)| \geq \gamma \eta^m - \delta' > \frac{\gamma \eta^m}{2}$  so that  $c_n > \frac{\gamma \eta^m}{2M^{m+1}}$ .

Let  $\alpha_i$  be a zero of  $P$  of multiplicity  $p \geq 1$ . Assume that  $P_n$  has strictly less than  $p$  zeroes on the interval  $[\alpha_i - \delta, \alpha_i + \delta]$ . In this case,

$$|g(\alpha_i + 2\delta)P(\alpha_i + 2\delta)| \leq M(2\delta)^p M^{m-p} \leq \delta^p (2M)^m,$$

and on the other hand

$$|c_n g_n(\alpha_i + 2\delta) P_n(\alpha_i + 2\delta)| \geq \frac{\gamma^2 \eta^m}{2M^{m+1}} \delta^{p-1} \ell^{m-k-p} \geq \frac{\gamma^2 (\eta \ell)^m}{2M^{m+1}} \delta^{p-1},$$

where we used the fact that all the roots of  $P_n$  are in  $\cup_{j=1}^{m-k} [\alpha_j - \delta, \alpha_j + \delta]$  so that the roots that are not  $[\alpha_i - \delta, \alpha_i + \delta]$  are at distance at least  $\ell$  of  $\alpha_i + 2\delta$ . Therefore,

$$\begin{aligned} \sup_{x \in \mathcal{K}_\varepsilon} |H^n(x) - f_\gamma^{k,\varepsilon}(x)| &> \left| |c_n g_n(\alpha_i + 2\delta) P_n(\alpha_i + 2\delta)| - |g(\alpha_i + 2\delta) P(\alpha_i + 2\delta)| \right| \\ &> \delta^{p-1} \left( \frac{\gamma^2 (\eta \ell)^m}{2M^{m+1}} - \delta(2M)^m \right) > \delta', \end{aligned} \quad (25)$$

which is a contradiction. A similar argument holds if we suppose that  $P_n$  has strictly more than  $p$  zeroes on the interval  $[\alpha_i - \delta, \alpha_i + \delta]$ . This finishes the proof since we have proved that inside  $\mathcal{K}_\varepsilon$ ,  $H^n$  vanishes iff  $f_\gamma^{k,\varepsilon}$  vanishes in a very small neighbourhood and with the same multiplicity, and its other zeroes must be smaller than  $b + 2\varepsilon$  on  $\{\tilde{\lambda}_1^n \leq \varepsilon^{-1}\}$ .

### 6.6. Identification of the minimizers.

We prove Theorem 6.4, which is straightforward. Since  $L$  is a good rate function, it vanishes at its minimizers  $(\lambda_1^*, \dots, \lambda_m^*) \in \mathbb{R}_\downarrow^m(b)$ . Putting  $\lambda_0^* = b + 1$ , we know that there exists  $0 \leq k \leq m$  such that  $\lambda_{m-k}^* > b$  and  $\lambda_{m-k+1}^* = b$ . From the definition of  $L$ , for any  $n$  large enough such that  $b + \frac{1}{n} < \lambda_{m-k}^*$ , we can find a function  $f_n$  defined on  $\mathcal{K}_{\frac{1}{n}}$  vanishing at  $(\lambda_1^*, \dots, \lambda_{m-k}^*)$  such that  $J_{\mathcal{K}_{\frac{1}{n}}}(f_n) \leq \frac{1}{n}$ . From the definition of  $J_{\mathcal{K}}$  and the fifth point of Lemma 5.5, all the functions  $f_n$  are in a compact set of  $\mathcal{C}((b, \infty), \mathbb{R})$  so that we can find a function  $f$  vanishing at  $(\lambda_1^*, \dots, \lambda_{m-k}^*)$  so that  $J_{\mathcal{K}_\varepsilon}(f) = 0$  for all  $\varepsilon > 0$ . But the latter also implies that  $f(z) = P_{\Theta, r}(K(z), C)$  with  $(K, C)$  minimising  $\mathbf{I}$ , that is  $(K, C) = (K^*, C^*)$  by Lemma 5.6.

## 7. LARGE DEVIATIONS FOR THE EIGENVALUES OF WISHART MATRICES

In this section, we study the i.i.d. case when  $X_n = 0$ . More precisely, we consider  $G = (g_1, \dots, g_r)$  satisfying Assumption 2.2,  $n \times r$  matrices  $G_n$  whose rows are i.i.d. copies of  $G$ , a diagonal matrix  $\Theta = \text{diag}(\theta_1, \dots, \theta_r)$  and we study the large deviations of Wishart matrices  $W_n = \frac{1}{n} G_n^* \Theta G_n$ . The large deviations for the largest and smallest eigenvalues were already studied in [22] in the case when  $\Theta = (1, \dots, 1)$  and the  $g_i$ 's are i.i.d.

**Proposition 7.1.** *Assume that  $G$  satisfies Assumption 2.2. Let  $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_r)$  be a diagonal matrix with positive entries. Then, the law of the eigenvalues of  $W_n$  satisfies a large deviation principle in the scale  $n$  with rate function which is infinite unless  $\alpha_1 \geq \dots \geq \alpha_r \geq 0$  and in this case given by*

$$L(\alpha_1, \dots, \alpha_r) = \inf \{ J(C) : (\alpha_1, \dots, \alpha_r) \text{ are the eigenvalues of } \Theta^{-\frac{1}{2}} C \Theta^{-\frac{1}{2}} \},$$

with

$$J(C) = \sup_{Y \in \mathbb{H}_r} \{ \text{Tr}(CY) - \log \mathbb{E}[e^{(G, YG)}] \}.$$

The Gaussian case allows an exact computation, given by the following

**Corollary 7.2.** *Assume that  $G = (g_1, \dots, g_r)$  is a Gaussian vector with positive definite covariance matrix  $R$ . Let  $\Theta = \text{diag}(\theta_1, \dots, \theta_r)$  be a diagonal matrix with positive entries. We denote by  $0 < r_1(\Theta) \leq r_2(\Theta) \leq \dots \leq r_r(\Theta)$  the eigenvalues of the matrix  $\Theta^{-1/2} R^{-1} \Theta^{-1/2}$  in increasing order.*

*Then, the law of the eigenvalues of  $W_n$  satisfies a large deviation principle in the scale  $n$  with rate function which is infinite unless  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_r > 0$  and otherwise given by*

$$L(\alpha_1, \dots, \alpha_r) = \frac{1}{2} \sum_{i=1}^r (\alpha_i r_i(\Theta) - 1 - \log(\alpha_i r_i(\Theta))).$$

*In the particular case when the entries are i.i.d. standard normal, the above rate function can be rewritten*

$$L(\alpha_1, \dots, \alpha_r) = \frac{1}{2} \sum_{i=1}^r \left( \frac{\alpha_i}{\theta_i} - 1 - \log \frac{\alpha_i}{\theta_i} \right).$$



Now, by a straightforward use of the contraction principle, we can derive some results about the deviations of the largest eigenvalue. This problem was addressed in particular in [22]. The following corollary holds for the Gaussian case.

**Corollary 7.3.** *Under the assumptions of Corollary 7.2, the law of the largest eigenvalue satisfies a LDP with good rate function*

$$L_{\max}(x) = \begin{cases} \frac{1}{2}(xr_1(\Theta) - 1 - \log(xr_1(\Theta))) & \text{if } x \geq \frac{1}{r_1(\Theta)}, \\ \frac{1}{2} \sum_{i=1}^j (xr_i(\Theta) - 1 - \log(xr_i(\Theta))) & \text{if } \frac{1}{r_{j+1}(\Theta)} < x \leq \frac{1}{r_j(\Theta)}, \end{cases}$$

with the convention that  $r_{r+1}(\Theta) = \infty$ .

In particular, in the i.i.d. standard case when  $\Theta = \text{diag}(1, \dots, 1)$ , we have

$$L_{\max}(x) = \begin{cases} \frac{1}{2}(x - 1) - \frac{1}{2} \log x & \text{if } x \geq 1, \\ \frac{r}{2}(x - 1) - \frac{r}{2} \log x & \text{if } x \in (0, 1), \end{cases}$$

and this allows to retrieve [22, Corollary 2.1] (note that a direct proof based on the formula for the joint law of the eigenvalues is then also available). This is in agreement with the fact that as  $r$  goes to infinity, we expect the deviations below one to be impossible in this scale.

In the general case, we have

**Corollary 7.4.** *Under the assumptions of Proposition 7.1, the law of the largest eigenvalue of  $W_n$  satisfies a large deviation principle in the scale  $n$  with a rate function  $L_{\max}(\alpha)$  which satisfies, for any  $\alpha \in \mathbb{R}$ ,*

$$\begin{aligned} L_{\max}(\alpha) &= \inf\{L(\alpha_1, \dots, \alpha_r) : \max \alpha_i = \alpha\} \\ &\geq \inf_{\|x\|_2=1} \sup_{t \in \mathbb{R}} \{t\alpha - \log E[e^{t\langle G, \Theta^{\frac{1}{2}} Gx \rangle}]\} =: I_{r, \Theta}(\alpha) \end{aligned}$$

From there, one can easily improve the upper bound on the probability of deviations of the largest eigenvalue of [22, Theorem 2.1] :

**Corollary 7.5.** *Assume that  $G$  satisfy Assumption 2.2 and that the  $g_i$ 's are i.i.d. with mean 0 and variance 1. Let  $\Theta = \text{diag}(\theta_1, \theta_2, \dots, \theta_r)$  be a diagonal matrix with positive entries, with  $\theta_1 \geq \theta_2 \geq \dots \geq \theta_r$ . Then we have that, for  $\alpha \geq \theta_1$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\lambda_{\max} \geq \alpha) = I_{r, \Theta}(\alpha).$$

Note that in this case  $I_{r, \Theta}(\alpha) = \inf_{[\alpha, \infty)} L_{\max}$  and in particular  $I_{r, \Theta}$  is not necessarily lower semicontinuous. We refer to [22] for more properties of  $I_{r, \Theta}$ , related results and conjectures.

**Proof of Proposition 7.1.** In the case where  $X_n = 0$ , we can apply Theorem 6.1 with  $P_{\theta, r}(K(z), C) = \det(z - \Theta^{\frac{1}{2}} C \Theta^{\frac{1}{2}})$  and  $\mathbf{I}(C) = J(C)$ . Hence, for  $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_r > 0$ ,  $L(\alpha)$  is the infimum of  $J$  over the nonnegative Hermitian matrices  $C$  such that  $\Theta^{\frac{1}{2}} C \Theta^{\frac{1}{2}}$

has spectrum  $(\alpha_1, \dots, \alpha_r)$ . Note that the case  $\alpha_r = 0$  has infinite entropy by Lemma 5.5.4)  $\square$

**Proof of Corollary 7.2.** In this case,  $\log \mathbb{E}[e^{\langle G, YG \rangle}]$  equals  $\log \det[(R^{-1} - 2Y)^{-\frac{1}{2}} R^{-\frac{1}{2}}]$  if  $R^{-1} - 2Y > 0$ , and is infinite otherwise. A classical saddle point analysis shows that the supremum in  $J$  is taken at

$$C_{ij} = \frac{E[g_i g_j e^{\langle G, YG \rangle}]}{E[e^{\langle G, YG \rangle}]} = ((R^{-1} - 2Y)^{-1})_{ij}$$

which yields

$$J(C) = \frac{1}{2} \operatorname{tr}(CR^{-1} - I) - \frac{1}{2} \log \det(CR^{-1}).$$

We finally take the infimum over  $C$  so that  $\Theta^{\frac{1}{2}} C \Theta^{\frac{1}{2}} = \sum_{i=1}^r \alpha_i e_i e_i^*$  for some orthonormal basis  $(e_i)_{1 \leq i \leq r}$ . This gives

$$\begin{aligned} L(\alpha) &= \inf_{(e_i)_{ONB}} \left\{ \frac{1}{2} \sum_{i=1}^r \alpha_i \langle e_i, \Theta^{-1/2} R^{-1} \Theta^{-1/2} e_i \rangle \right\} - \frac{1}{2} \sum_{i=1}^r \log \alpha_i - \frac{r}{2} - \frac{1}{2} \sum \log r_i(\Theta) \\ &= \frac{1}{2} \sum_{i=1}^r \alpha_i r_i(\Theta) - \frac{1}{2} \sum_{i=1}^r \log(\alpha_i r_i(\Theta)) - \frac{r}{2}. \end{aligned}$$

$\square$

**Proof of Corollary 7.4.** We only need to take  $Y = tvv^*$  if  $C$  has eigenvector  $v$  for its largest eigenvalue to get the bound.  $\square$

**Proof of Corollary 7.5.** The inequality in Corollary 7.4 gives the upper bound and the lower bound is obtained by the same proof as in [22] as for fixed  $x$ ,  $\langle x, W_n x \rangle$  can be also in this case written as a sum of i.i.d. random variables so that Cramer's Theorem apply. By arguments as in [22], one can also check that  $I_{r, \Theta}$  is increasing on  $[\theta_1, \infty)$ , which concludes the proof.  $\square$

## 8. LARGE DEVIATIONS FOR $H^n$ IN THE PRESENCE OF OUTLIERS

Let  $\mathcal{K}^\circ := \bigcup_{i=1}^{p_0} [a_i, b_i]$  a compact subset of  $(b, \infty) \setminus \{\ell_1^+, \dots, \ell_{p^+}^+\}$ . We equip again  $\mathcal{C}(\mathcal{K}^\circ, \mathbf{H}_r) \times \mathbf{H}_r$  with the uniform topology. Hereafter, we denote by  $\ell_i = \ell_i^+$  for  $1 \leq i \leq p^+$  and  $\ell_i = \ell_{p^+ p^- - i + 1}^-$  for  $p^+ + 1 \leq i \leq p^+ + p^-$ .

We recall that  $K^n(z)$  and  $C^n$  were defined in (4) and (5) respectively.

**Theorem 8.1.** *We assume that Assumptions 2.1, 2.2, 2.5 and 2.6 hold.*

- (1) *The law of  $((K^n(z))_{z \in \mathcal{K}^\circ}, C^n)$ , viewed as an element of the space  $\mathcal{C}(\mathcal{K}^\circ, \mathbf{H}_r) \times \mathbf{H}_r$  endowed with the uniform topology, satisfies a large deviation principle in the scale  $n$  with rate function  $\mathbf{I}^\circ$ . For  $K \in \mathcal{C}(\mathcal{K}^\circ, \mathbf{H}_r)$  and  $C \in \mathbf{H}_r$ ,  $\mathbf{I}^\circ(K(\cdot), C)$  is infinite if  $z \rightarrow K(z)$  is not uniformly Lipschitz on  $\mathcal{K}^\circ$ . Otherwise, it is given by*

$$\mathbf{I}^\circ(K(\cdot), C) = \inf \left\{ \Gamma^*(K_0(\cdot), C_0) + \sum_{i=1}^{p^+ + p^-} I^{(Z)}(L_i) \right\},$$

where the infimum is taken over the families  $K_0(\cdot) \in \mathcal{C}(\mathcal{K}^o, \mathbf{H}_r)$ ,  $C_0, L_1, \dots, L_{p^++p^-} \in \mathbf{H}_r$  satisfying the condition

$$K_0(\cdot) + \sum_{i=1}^{p^++p^-} \frac{1}{z - \ell_i} L_i = K(\cdot) \text{ and } C_0 + \sum_{i=1}^{p^++p^-} L_i = C \quad (26)$$

and with

$$\Gamma^*(K(\cdot), C) = \sup_{P, X, Y} \left\{ \text{Tr} \left( \int K'(z) P(z) dz + \sum_{i=1}^{p_0} K(b_i) X_i + CY \right) - \int \Lambda \left( - \int \frac{1}{(z-x)^2} P(z) dz + \sum_{i=1}^{p_0} \frac{1}{b_i - x} X_i + Y \right) d\mu(x) \right\},$$

the supremum being taken over piece-wise constant  $P$  with values in  $\mathbf{H}_r$ ,  $X = (X_1, \dots, X_{p_0}) \in (\mathbf{H}_r)^{p_0}$  and  $Y \in \mathbf{H}_r$ .

- (2) The law of  $(H^n(z))_{z \in \mathcal{K}^o}$  on  $\mathcal{C}(\mathcal{K}^o, \mathbb{R})$  equipped with the uniform topology, satisfies a large deviation principle in the scale  $n$  with rate function given, for a function  $f \in \mathcal{C}(\mathcal{K}^o, \mathbb{R})$ , by

$$J_{\mathcal{K}^o}^o(f) = \inf \{ \mathbf{I}^o(K(\cdot), C); (K(\cdot), C) \in \mathcal{C}(\mathcal{K}^o, \mathbf{H}_r) \times \mathbf{H}_r, P_{\Theta, r}(K(z), C) = f(z) \forall z \in \mathcal{K}^o \}.$$

Note that the function  $\Gamma^*$  is well defined because if  $K$  is uniformly Lipschitz on  $\mathcal{K}^o$ , then so is any  $K_0$  satisfying the compatibility condition (26), so that  $K'_0$  almost surely exists.

Note firstly that under the second assertion of Assumption 2.6, we have the following straightforward application of the contraction principle.

**Lemma 8.2.** *Let  $Z_1$  be the  $\mathbf{H}_r$ -valued random variable such that for  $1 \leq i \leq j \leq r$ ,  $(Z_1)_{ij} = g_i(1)g_j(1)$ . Under Assumption 2.6,  $\frac{Z_1}{n}$  also satisfies a large deviation principle in the scale  $n$  with a good rate function  $I^{(Z)}(M) = \inf \{ I(v) : \bar{v}_i v_j = M_{ij}, 1 \leq i, j \leq r \}$ .*

The proof of Theorem 8.1 follows the same lines as that of Theorem 5.1. The LDP for finite dimensional marginals for our process is described by the second part of Theorem 11.1. It is based on the large deviations for  $K^n$  and  $C^n$  that can be, up to a re-indexation, shown to be exponentially equivalent to

$$K^n(z)_{ij} = \frac{1}{n} \sum_{k=p^++p^-+1}^n \frac{1}{z - \lambda_k^n} \overline{g_i(k)} g_j(k) + \sum_{k=1}^{p^++p^-} \frac{1}{z - \ell_k} \frac{\overline{g_i(k)} g_j(k)}{n}$$

which satisfy a LDP by independence of the  $g_i(k)$ , and large deviations of each parts by Proposition 5.3 and Lemma 8.2. The corresponding rate function will be denoted by  $(I_M^{z_1, \dots, z_M})^o$ . To define this new rate function, we first extend in an obvious way the definition of  $I_M^{z_1, \dots, z_M}$  for  $z_i$ 's in  $\mathcal{K}^o$ . Then one can define, for  $K_1, \dots, K_M, C \in \mathbf{H}_r$ , and  $z_1, \dots, z_M \in \mathcal{K}^o$ ,

$$(I_M^{z_1, \dots, z_M})^o(K_1, \dots, K_M, C) = \inf \left\{ I_M^{z_1, \dots, z_M}(K_{0,1}, \dots, K_{0,M}, C) + \sum_{i=1}^{p^++p^-} I^{(Z)}(L_i) \right\},$$

where the infimum is taken over families

$$(C, K_{0,1}, \dots, K_{0,M}, L_1, \dots, L_{p^++p^-}) \in (\mathbf{H}_r)^{1+M+p^++p^-}$$

under the condition that for all  $1 \leq j \leq M$ ,

$$K_{0,j} + \sum_{i=1}^{p^++p^-} \frac{1}{z_j - \ell_i} L_i = K_j \quad \text{and} \quad C_0 + \sum_{i=1}^{p^++p^-} L_i = C.$$

By Dawson-Gärtner Theorem, we deduce that  $((K^n(z))_{z \in \mathcal{K}^o}, C^n)$  satisfies a LDP for the topology of point-wise convergence with good rate function

$$\mathbf{J}^o(K, C) = \sup_M \sup_{z_1 < \dots < z_M, z_j \in \mathcal{K}^o} (I_M^{z_1, \dots, z_M})^o(K(z_1), \dots, K(z_M), C).$$

Since exponential tightness is clear, this LDP can be reinforced into the uniform topology. We then have to check that  $\mathbf{I}^o = \mathbf{J}^o$ .

From the definition of  $\mathbf{I}^o$ , the first thing to check is that  $\mathbf{J}^o(K(\cdot), C) = \infty$  if  $K$  is not Lipschitz continuous on  $\mathcal{K}^o$ . Assume the latter and choose any  $L > 0$ . We know that there exists  $\eta$  such that for any  $z \in \mathcal{K}^o$  and  $i \leq p^+ + p^-$ ,  $z - \ell_i > \eta$ . As  $I^{(Z)}$  is a good rate function, its level sets are compact and there exists  $R > 0$  such that  $\{I^{(Z)} \leq L\}$  is included in the ball of radius  $\frac{R\eta^2}{(p^++p^-)}$  for the norm  $\|\cdot\|_2$  in  $\mathbf{H}_r$ . Let  $\gamma$  and  $B$  be given by Lemma 5.5.(3). As  $K$  is not Lipschitz continuous, there exists  $z_1 < z_2 \in \mathcal{K}^o$  such that

$$\|K(z_1) - K(z_2)\|_2 \geq \left( \frac{L+B}{\gamma\eta^2} + R \right) |z_1 - z_2|.$$

We will show that  $(I_2^{z_1, z_2})^o(K(z_1), K(z_2), C) \geq L$ . Let  $K_0, L_i$  be such that

$$K_0(z_1) + \sum_{i=1}^{p^++p^-} \frac{1}{z_1 - \ell_i} L_i = K(z_1) \quad \text{and} \quad K_0(z_2) + \sum_{i=1}^{p^++p^-} \frac{1}{z_2 - \ell_i} L_i = K(z_2).$$

Then

$$K_0(z_1) - K_0(z_2) = (z_1 - z_2) \left( \frac{K(z_1) - K(z_2)}{z_1 - z_2} + \sum_{i=1}^{p^++p^-} \frac{1}{(z_1 - \ell_i)(z_2 - \ell_i)} L_i \right)$$

Therefore, there exists  $i_0$  such that  $L_{i_0} > \frac{R\eta^2}{(p^++p^-)}$  or  $\|K_0(z_1) - K_0(z_2)\|_2 \geq \frac{L+B}{\gamma\eta^2} |z_1 - z_2|$ . In the first case,  $I^{(Z)}(L_{i_0}) > L$ . In the second case, we have, from the rewriting (12), that,

$$I_2^{z_1, z_2}(K_0(z_1), K_0(z_2), C) \geq \text{Tr} \left( \pm \gamma \eta^2 \frac{K_0(z_1) - K_0(z_2)}{z_1 - z_2} \right) - \int \Lambda \left( \frac{\pm \gamma \eta^2 I_r}{(z_1 - x)(z_2 - x)} \right) d\mu \geq L,$$

if we choose the right sign.

This gives  $(I_2^{z_1, z_2})^o(K(z_1), K(z_2), C) \geq L$  so that  $\mathbf{J}^o(K(\cdot), C) = \infty$  after optimisation.

We now suppose that  $K$  is Lipschitz continuous on  $\mathcal{K}^o$  and we want to identify the two rate functions. By mimicking<sup>3</sup> the proof at the end of Section 5.3, one can easily show

<sup>3</sup>We just have to be careful in the rewriting to put one border term for each interval involved in  $\mathcal{K}^o$ .

that for  $K$  is Lipschitz continuous on  $\mathcal{K}^o$ ,

$$\sup_M \sup_{z_1, \dots, z_M} I_M^{z_1, \dots, z_M}(K(z_1), \dots, K(z_M)) = \Gamma^*(K, C). \quad (27)$$

Now, in order to achieve this identification, we have to check that we can switch the supremum over  $M$  and the  $z_i$ 's and the infimum over the admissible simultaneous decompositions of  $K$  and  $C$ . It is clear that,

$$\mathbf{J}^o(K, C) \leq \Gamma^*(K_0(\cdot), C_0) + \sum_{i=1}^{p^+ + p^-} I^{(Z)}(L_i)$$

for any admissible choice of  $L_i$ , and therefore  $\mathbf{J}^o \leq \mathbf{I}^o$  after optimisation. We now need the converse inequality. By definition of  $\mathbf{J}^o$ , if it is finite, then for any positive integer  $p$ , there exists  $M(p)$  and  $z_1, \dots, z_{M(p)}$  such that

$$\mathbf{J}^o(K, C) \geq (I_{M(p)}^{z_1, \dots, z_{M(p)}})^o(K(z_1), \dots, K(z_{M(p)}), C) - \frac{1}{p}.$$

Now for each  $z_1, \dots, z_{M(p)}$  we choose an admissible decomposition (according to (26)) of  $K$  so that

$$\mathbf{J}^o(K, C) \geq I_{M(p)}^{z_1, \dots, z_{M(p)}}(K_0^{M(p)}(z_1), \dots, K_0^{M(p)}(z_{M(p)}), C) + \sum_{i=1}^{p^+ + p^-} I^{(Z)}(L_i^{M(p)}) - \frac{1}{p}.$$

Moreover, for each  $M$  and choices of  $z_1 < \dots < z_M$ ,

$$I_M^{z_1, \dots, z_M}(K(z_1), \dots, K(z_M)) = \Gamma^*(K_M^{z_1, \dots, z_M}, C)$$

with  $K_M^{z_1, \dots, z_M}(z) = \sum_{i=1}^M 1_{[z_i, z_{i+1}]} K(z)$ .

By definition, since  $I^{(Z)}$  and  $\Gamma^*$  are good rate functions and as for all  $i$ ,  $I^{(Z)}(L_i^{M(p)})$  and  $\Gamma^*(K_0^{M(p)}(z_1), \dots, K_0^{M(p)}(z_{M(p)}), C)$  are uniformly bounded, it implies that the arguments are tight and we can take a converging subsequence. Let  $K_0$  and  $L_i$  be limits along a subsequence, we get

$$\mathbf{J}^o(K, C) \geq \Gamma^*(K_0(\cdot), C) + \sum_{i=1}^{p^+ + p^-} I^{(Z)}(L_i) - \frac{1}{p}$$

which insures that  $\mathbf{J}^o(K, C) \geq \mathbf{I}^o(K, C)$ . This completes the proof of Theorem 8.1.

## 9. LARGE DEVIATIONS PRINCIPLE FOR THE LARGEST EIGENVALUES IN THE CASE WITH OUTLIERS

We now state the main Theorem of this section, namely an analogue of Theorem 6.1. For any  $\varepsilon, \rho$  small enough, we define the compact sets

$$\mathcal{K}_{\varepsilon, \rho}^o := [b + \varepsilon, \varepsilon^{-1}] \setminus \bigcup_{i=1}^{p^+} (\ell_i^+ - \rho, \ell_i^+ + \rho)$$

and  $\mathcal{K}_\varepsilon^o := [b + \varepsilon, \varepsilon^{-1}]$ . The reason why we distinguish the two parameters  $\varepsilon$  and  $\rho$  will become apparent in the statement and the proof of Lemma 9.7 below. We also define the set  $\{\ell\} := \{\ell_1^+, \dots, \ell_{p^+}^+, b\}$ , and for  $z \notin \{\ell\}$ ,  $R(z) := \prod_{i=1}^{p^+} \frac{1}{z - \ell_i^+}$ . We recall that  $s$  is the sign of the product  $\prod_{i=1}^r \theta_i$ .

For any  $\varepsilon, \rho, \gamma > 0$ , and  $\alpha \in \mathbb{R}_\downarrow^p(b + \varepsilon)$ , we put

$$S_{\alpha, \gamma}^{\varepsilon, \rho, o} := \left\{ f \in \mathcal{C}(\mathcal{K}_{\varepsilon, \rho}^o, \mathbb{R}) : \exists g \in \mathcal{C}(\mathcal{K}_{\varepsilon, \rho}^o, \mathbb{R}) \text{ with } \gamma \leq g \leq \frac{1}{\gamma}, \omega(g) \leq \frac{1}{\gamma} \right. \\ \left. \text{and } f(z) = s.R(z).g(z) \prod_{i=1}^p (z - \alpha_i) \right\}$$

We also denote by

$$C_{k, \gamma}^{\varepsilon, \rho, o} := \left\{ f \in \mathcal{C}(\mathcal{K}_{\varepsilon, \rho}^o, \mathbb{R}) : \exists p \text{ polynomial of degree } m + p^+ - k \text{ with } m + p^+ - k \text{ roots in } \mathcal{K}_\varepsilon^o \right. \\ \left. \text{and dominant coefficient } 1, g \in \mathcal{C}(\mathcal{K}_{\varepsilon, \rho}^o, \mathbb{R}) \text{ with } \gamma \leq g \leq \frac{1}{\gamma}, \omega(g) \leq \frac{1}{\gamma} \right. \\ \left. \text{and } f(z) = s.g(z).R(z).p(z) \right\}$$

and

$$C_\gamma^{\varepsilon, \rho, o} = \bigcup_{0 \leq k \leq m + p^+} C_{k, \gamma}^{\varepsilon, \rho, o}.$$

Then the main statement of this section is the following.

**Theorem 9.1.** *Under Assumptions 2.1, 2.2, 2.5 and 2.6, the law of the  $m + p^+$  largest eigenvalues  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_{m+p^+}^n)$  of  $\tilde{X}_n$  satisfies a large deviation principle in  $\mathbb{R}^{m+p^+}$  with good rate function  $L^o$ . For  $\alpha = (\alpha_1, \dots, \alpha_{m+p^+}) \in \mathbb{R}^{m+p^+}$ , we take  $\alpha_{m+p^++1} = b$  and  $L^o$  is defined as follows :*

$$L^o(\alpha) = \begin{cases} \lim_{\varepsilon \downarrow 0} \lim_{\rho \downarrow 0} \inf_{\cup_{\gamma > 0} S_{(\alpha_1, \dots, \alpha_{m+p^+-k})}^{\varepsilon, \rho, o}, \gamma} J_{\mathcal{K}_{\varepsilon, \rho}^o}^o & \text{if } \alpha \in \mathbb{R}_\downarrow^{m+p^+}(b), \alpha_{m+p^+-k+1} = b, \\ & \alpha_{m+p^+-k} > b \text{ for a } k \in \{0, \dots, m\}, \\ \infty & \text{otherwise.} \end{cases}$$

Even though the rate function  $L^o$  is not very explicit, we show below that it must be infinite if Horn's inequalities are violated.

**Remark 9.2.** *Recall that the eigenvalues  $(\tilde{\lambda}_i^n)_{1 \leq i \leq n}$  of the sum of two Hermitian matrices with eigenvalues  $(\lambda_i^n)_{1 \leq i \leq n}$  and  $\theta := (\theta_1, \dots, \theta_r, 0, \dots, 0)$  satisfy Horn's inequalities and are characterised by the fact that they satisfy such inequalities (see [29] for details). Assume that  $\tilde{\lambda} := (\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m+p^+})$  is at distance of the bulk and of the outliers which is bounded below. We claim that the rate function  $L^o(\tilde{\lambda})$  is infinite if  $(\tilde{\lambda}, \ell, \theta)$  do not satisfy the Horn inequalities. Indeed, if  $L^o(\tilde{\lambda})$  is finite,  $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m+p^+})$  are zeroes of a function  $f$  which can be written*

$$f(z) = P_{\Theta, r}(K(z), C).$$

with  $\mathbf{I}^\circ(K(\cdot), C)$  is finite. It implies that there exists sequences  $\lambda^n \in \mathbb{R}^n, g_j(\cdot) \in \mathbb{C}^n$  so that  $\lambda^n$  satisfies Assumptions 2.1, 2.5 and 2.6 and

$$K^n(z) = \frac{1}{n} \sum_{i=1}^n \frac{\overline{g_i(k)} g_j(k)}{z - \lambda_k^n}, \quad C^n = \frac{1}{n} \sum_{i=1}^n \overline{g_i(k)} g_j(k)$$

converge to  $K(z)$  (uniformly away from the bulk and the outliers) and  $C$  respectively. By definition, there exists a constant  $c$  such that

$$P_{\Theta, r}(K^n(z), C^n) \prod_{i=1}^n (z - \lambda_i^n) = c \det \left( z - \text{diag}(\lambda^n) - \sum_{i=1}^r \theta_i u_i u_i^* \right)$$

with  $u_i = g_i$  in the i.i.d. model and  $u_i$  the Gram-Schmidt orthonormalisation of the vectors  $g_i$  in the orthonormalized model. Hence, the function  $f_n(z) = P_{\Theta, r}(K^n(z), C^n)$  vanishes at the eigenvalues  $(\tilde{\lambda}^n)$  of the sum of the two Hermitian matrices  $\text{diag}(\lambda^n)$  and  $\sum_{i=1}^r \theta_i u_i u_i^*$  (note that we can assume without loss of generality that its zeroes are different from  $\lambda^n$  by Lemma 11.4). Therefore,  $(\lambda^n, \tilde{\lambda}^n, \theta)$  satisfy Horn's inequalities by [29]. Since the  $(\tilde{\lambda}^n)$  are bounded, they are relatively compact and we see that the limit points  $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m+p^+})$  of  $(\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_{m+p^+}^n)$  which stay away from the bulk and the outliers are the zeroes of  $f$ . By passing to the limit in Horn's inequalities, we thus deduce that if the vector  $(\tilde{\lambda}_1, \dots, \tilde{\lambda}_{m+p^+})$  has finite  $L^\circ$ -entropy, and is away from the bulk and the outliers,  $(\tilde{\lambda}, \ell, \theta)$  satisfies Horn's inequalities. It would be interesting to have a direct proof of this fact at the asymptotic level.

### 9.1. Proof of Theorem 9.1.

We now prove Theorem 9.1, following roughly the same lines as for Theorem 6.1.

As in the proof of Theorem 6.1, the crucial point is to use Proposition 3.1. In the sticking case, if  $z \in \mathcal{K}_\varepsilon$ , for  $n$  large enough, the condition that  $z$  should not belong to the set of eigenvalues of  $X_n$  was very easy to check. Here, we need to show the following

**Lemma 9.3.** *Assume that the eigenvalues  $\lambda_1^n, \dots, \lambda_n^n$  of  $X_n$  are pair-wise distinct and that Assumptions 2.1, 2.5 and 2.6 hold, then  $X_n$  and  $\tilde{X}_n$  have no eigenvalue in common.*

The proof of this lemma is postponed to Appendix 11.3. We shall therefore give the proof of the Theorem when the eigenvalues of  $X_n$  are distinct. This is however sufficient to get the large deviation principle without this hypothesis due to the following Lemma.

**Lemma 9.4.** *Let  $X_n$  satisfy Assumptions 2.1 and 2.5 and  $G$  Assumption 2.6. Then, there exists a sequence  $\bar{X}_n$  of matrices satisfying Assumptions 2.1 and 2.5 such that moreover,*

- the eigenvalues of  $\bar{X}_n$  are pair-wise distinct,
- almost surely for all  $i = 1, \dots, m$ , the  $i$ th largest eigenvalue of  $\bar{X}_n$  and the  $i$ th largest eigenvalue of  $\tilde{X}_n$  are within distance  $n^{-1}$ .

Therefore, the random variables  $(\tilde{\lambda}_i^n)_{i \leq m}$  and  $(\tilde{\lambda}_i^n)_{i \leq m}$  are exponentially equivalent and [14, Theorem 4.2.13] asserts that a large deviations principle for the extreme eigenvalues  $(\tilde{\lambda}_i^n)_{i \leq m}$  of  $\tilde{X}_n$  entails the large deviations principle for the law of  $(\tilde{\lambda}_i^n)_{i \leq m}$ .



The proof of Lemma 9.4 follows by taking  $\bar{X}_n$  to be the matrix with the same eigenvectors as  $X_n$  and the same eigenvalues except for those which are stuck together which we separate by an arbitrary small weight  $w_n \leq 1/n$ , much smaller than the minimal distance between two distinct eigenvalues of  $X_n$ , so that the eigenvalues of  $\bar{X}_n$  are distinct and the operator norm of  $X_n - \bar{X}_n$  is bounded above by  $w_n$ . But then also the eigenvalues of  $\widetilde{X}_n$  and  $\bar{X}_n$  differ only at most by  $w_n$  almost surely. It is straightforward to verify Assumptions 2.1 and 2.5 for  $\bar{X}_n$ .

Thus, we shall restrict ourselves in the sequel to the case where  $X_n$  has distinct eigenvalues. By Lemma 9.3, the eigenvalues of  $\widetilde{X}_n$  are then the zeroes of  $H^n$  and we can proceed as in the case without outliers.

We first focus our attention to the function  $H^n$  restricted to  $\mathcal{K}_{\varepsilon, \rho}^o$  and show the counterpart of Lemma 6.7, that is

**Lemma 9.5.** *Let  $\varepsilon, \rho$  be fixed. There exists a positive integer  $n_0(\varepsilon, \rho)$  and  $L(\varepsilon) > 0$  such that for any  $n \geq n_0(\varepsilon, \rho)$ , for any  $z \in \mathcal{K}_{\varepsilon, \rho}^o$ ,*

$$H^n(z) = \begin{cases} s \prod_{i=1}^r \|q_i^n W_i^n\|_2^2 g_n(z) R(z) \prod_{i=1}^m (z - \tilde{\lambda}_i^n) & \text{in the orthonormalized model,} \\ s g_n(z) R(z) \prod_{i=1}^m (z - \tilde{\lambda}_i^n) & \text{in the i.i.d. model,} \end{cases} \quad (28)$$

with  $L(\varepsilon) \leq g_n \leq \frac{1}{L(\varepsilon)}$  and  $\omega(g_n) \leq \frac{1}{L(\varepsilon)}$ .

In particular, for any  $\varepsilon > 0$  and  $\rho > 0$  small enough,

$$\limsup_{\gamma \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \left( \prod_{i=1}^m (z - \tilde{\lambda}_i^n)^{-1} H^n(z) \right)_{z \in \mathcal{K}_{\varepsilon, \rho}^o} \in (C_\gamma^{\varepsilon, \rho, o})^c \right) = -\infty.$$

*Proof.* In this case,

$$g_n(z) := \prod_{i=1}^r |\theta_i|^{-1} \prod_{i=1}^{p^+} \left( 1 + \frac{\lambda_i^n - \ell_i^+}{z - \lambda_i^n} \right) \prod_{i=p^++1}^{m+p^+} \frac{1}{z - \lambda_i^n} \prod_{i=m+p^++1}^n \left( 1 + \frac{\lambda_i^n - \tilde{\lambda}_i^n}{z - \lambda_i^n} \right).$$

The proof is exactly the same as in the sticking case once we have noticed that, from Assumption 2.5, there exists  $n_0(\varepsilon, \rho)$  such that for  $n \geq n_0(\varepsilon, \rho)$ ,  $\prod_{i=1}^{p^+} \left( 1 - \frac{\lambda_i^n - \ell_i^+}{\lambda_i^n - z} \right) \geq \frac{1}{2^{p^+}}$ , so that

$$g_n(z) \geq \prod_{i=1}^r |\theta_i|^{-1} \left( \frac{1}{2} \right)^{p^+} \left( \frac{\varepsilon}{2} \right)^m e^{-2m \left( \frac{4}{\varepsilon^2} \right)^2}.$$

Note that we could similarly show that for  $n \geq n_0(\varepsilon, \rho)$ ,

$$g_n(z) \leq L(\varepsilon) := \left( \frac{3}{2} \right)^{p^+} \left( \frac{2}{\varepsilon} \right)^m e^{\frac{8m}{\varepsilon^2}}. \quad (29)$$

The uniform equicontinuity is also shown very similarly.  $\square$

As in the sticking case, we have the analogue of Lemma 6.8, with  $L^o$  instead of  $L$ . To state more precisely the lemma, we introduce the following notation: we denote by



$G_k(\alpha, \delta, \varepsilon, \rho)$  the set of  $n$ -tuples  $(\tilde{\lambda}_1^n \geq \dots \geq \tilde{\lambda}_n^n)$  such that for all  $i \leq m + p^+ - k$ ,

$$|\tilde{\lambda}_i^n - \alpha_i| \leq \delta \text{ if } \alpha_i \notin \{\ell\} \quad \text{and} \quad |\tilde{\lambda}_i^n - \alpha_i| \leq \rho \text{ if } \alpha_i \in \{\ell\}$$

and for all  $m + p^+ - k + 1 \leq i \leq m + p^+$ ,

$$\tilde{\lambda}_i^n \leq b + \varepsilon.$$

We shall show the following

**Lemma 9.6.** *Let  $\alpha \in \mathbb{R}_{\downarrow}^m$  and  $k$  between 0 and  $m$  such that  $\alpha_{m+p^+-k+1} = \dots = \alpha_{m+p^++1} = b$  and  $\alpha_{m+p^+-k} > b$ . We have*

$$\begin{aligned} & \lim_{\varepsilon \downarrow 0} \lim_{\rho \downarrow 0} \limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( (\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_n^n) \in G_k(\alpha, \delta, \varepsilon, \rho) \right) \\ &= \lim_{\varepsilon \downarrow 0} \lim_{\rho \downarrow 0} \liminf_{\delta \downarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( (\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_n^n) \in G_k(\alpha, \delta, \varepsilon, \rho) \right) = -L^o(\alpha), \end{aligned}$$

with the obvious convention that  $\bigcap_{m+p^+-k+1 \leq i \leq m+p^+} \{\tilde{\lambda}_i^n \leq b + \varepsilon\} = \Omega$  if  $k = 0$ .

The only point that differs substantially is the proof of the analog of Lemma 6.9. We fix  $\varepsilon > 0$  small enough such that  $\alpha_{m-k} > b + \varepsilon$ . Let  $L(\varepsilon)$  be as in (29) (and denote it by  $L$  for simplicity). We assume that  $L > 1$ . We denote by  $\eta := \frac{\varepsilon^{-1} - (b + \varepsilon)}{2(p^+ + m + 1)}$  and

$$t := \min \left( 1, \frac{1}{3} \min_{\alpha_i \neq \alpha_j} |\alpha_i - \alpha_j|, \frac{1}{3} |\alpha_{m-k} - b|, \frac{1}{3} \min_{i,j; \alpha_i \notin \{\ell\}} |\alpha_i - \ell_j| \right).$$

Let  $j$  be such that  $\alpha_i = \ell_j$  and denote by  $q$  the multiplicity of  $\ell_j$  as an outlier. From the definition of  $R$ , one can check that there exists two constants  $c_i$  and  $C_i$  (depending only on  $\varepsilon$ ) such that, for any  $\rho > 0$ ,  $c_i \rho^{-q} \leq |R(\alpha_i + 2\rho)| \leq C_i \rho^{-q}$ . Let  $c := \max_i c_i$  and  $C := \max_i C_i$ .

Let now fix  $\rho$  small enough such that for any  $i$  such that  $\alpha_i \notin \{\ell\}$ ,  $\min_j |\alpha_i - \ell_j| > 3\rho$  and  $\rho < \frac{c\gamma(\eta t)^{m+p^+}}{C \cdot 2^{m+p^++1} L^{2m+1+3p^+}}$

Let  $f_{\gamma}^{\varepsilon, \rho, o}$  be a minimizer of  $J_{\mathcal{K}_{\varepsilon, \rho}^o}$  over  $S_{\alpha, \gamma}^{\varepsilon, \rho, o}$ , then we have

**Lemma 9.7.** *There exists  $\delta_0 > 0$  such that for  $\delta \leq \delta_0$  there exists  $\delta'$  such that for any  $n \geq n_0(\varepsilon, \rho)$ ,*

$$\left\{ \tilde{\lambda}_1^n \leq \varepsilon^{-1} \right\} \cap \left\{ \sup_{z \in \mathcal{K}_{\varepsilon, \rho}^o} |H^n(z) - f_{\gamma}^{\varepsilon, \rho, o}(z)| < \delta' \right\} \subset \left\{ (\tilde{\lambda}_1^n, \dots, \tilde{\lambda}_n^n) \in G_k(\alpha, \delta, \varepsilon, \rho) \right\}$$

The remaining of this section will be devoted to the proof of this lemma and this will complete the study of the case with outliers.

By definition of  $S_{(\alpha_1, \dots, \alpha_{m-k}), \gamma}^{\varepsilon, \rho, o}$ , we know that for any  $z \in \mathcal{K}_{\varepsilon, \rho}^o$ ,

$$f_{\gamma}^{\varepsilon, \rho, o}(z) = s \cdot R(z) \cdot g(z) \cdot P(z) \text{ with } P(z) = \prod_{i=1}^{m-k} (z - \alpha_i),$$

with  $\alpha_i \in \mathcal{K}_\varepsilon^o$  and  $g \geq \gamma > 0$ .

On the other hand, from the study of the structure of  $H^n$ , we know that there exists a (possibly random) positive number  $c_n$  such that for any  $z \in \mathcal{K}_{\varepsilon, \rho}^o$ ,

$$H^n(z) = s \cdot c_n R(z) g_n(z) P_n(z) \text{ with } P_n(z) := \prod_{i=1}^m (z - \tilde{\lambda}_i^n).$$

With no loss of generality, one can assume that  $s = 1$ ,  $g_n \geq \gamma$  and we know that, for  $n$  large enough,  $g_n \leq L(\varepsilon)$ , with  $L(\varepsilon)$  as defined in (29). One can also assume that  $g \leq L(\varepsilon)$  and  $b^* - b \leq L$ .

We fix hereafter  $\delta > 0$  such that

$$\delta < \left( \frac{\gamma^2}{2} \left( \frac{\eta}{2} \right)^{m+p^+} \frac{t^{m+p^+}}{L^{2m+1+4p^+}} \right) \wedge t.$$

As  $g$  is bounded below by  $\gamma$ , one can choose  $\delta'_1$  such that for any  $\delta' < \delta'_1$ , any  $f$  satisfying that  $\sup_{x \in \mathcal{K}_{\varepsilon, \rho}^o} |f(x) - f_\gamma^{\varepsilon, \rho, o}(x)| < \delta'$  vanishes only in

$$\cup_{j, \alpha_j \in \{\ell\}} [\alpha_j - \rho, \alpha_j + \rho] \cup \cup_{j, \alpha_j \notin \{\ell\}} [\alpha_j - \delta, \alpha_j + \delta].$$

Let  $\delta'_2$  be such that

$$0 < \delta'_2 \leq \delta^m \left( \frac{\gamma^2}{2} \frac{(\eta t)^{m+p^+}}{L^{m+1+3p^+}} - \delta \frac{(2L)^{m+p^+}}{t^{p^+}} \right)$$

and

$$0 < \delta'_2 \leq \rho^m \left( c \frac{\gamma^2 (\eta t)^{m+p^+}}{2L^{m+1+2p^+}} - \rho C (2L)^{m+p^+} \right)$$

(it exists thanks to our choices of  $\delta$  and  $\rho$ ).

Let now  $\delta'$  be such that  $\delta' < \min \left( \delta'_1, \delta'_2, \frac{\gamma \eta^m}{2L^{p^+}} \right)$  and  $H^n$  such that

$$\sup_{x \in \mathcal{K}_{\varepsilon, \rho}^o} |H^n(x) - f_\gamma^{\varepsilon, \rho, o}(x)| < \delta'.$$

We first claim that  $c_n$  is bounded below uniformly on this set. Indeed, if we cut  $\mathcal{K}_\varepsilon^o$  into  $p^+ + m + 1$  intervals of size  $2\eta$ , at least one of them do not contain any zero of  $\alpha_i$  nor any  $\ell_i^+$ . We call  $x_0$  the centre of this interval. Then  $|g(x_0)R(x_0)P(x_0)| \geq \gamma \eta^m \frac{1}{L^{p^+}}$ . Therefore  $|c_n g_n(x_0)R(x_0)P_n(x_0)| \geq \gamma \eta^m \frac{1}{L^{p^+}} - \delta' > \frac{\gamma \eta^m}{2L^{p^+}}$  so that  $c_n > \frac{\gamma \eta^{m+p^+}}{2L^{m+1+2p^+}}$ .

Let  $\alpha_i \notin \{\ell\}$  be of multiplicity  $p \geq 1$ . Assume that  $P_n$  has strictly less than  $p$  zeroes on the interval  $[\alpha_i - \delta, \alpha_i + \delta]$ . We recall that  $\min_j |\alpha_i - \ell_j^+| > \rho$ . In this case,

$$|g(\alpha_i + 2\delta)R(\alpha_i + 2\delta)P(\alpha_i + 2\delta)| \leq L(2\delta)^p \frac{(b^* - b)^{m+p^+-p}}{t^{p^+}} \leq \delta^p \frac{(2L)^{m+p^+}}{t^{p^+}},$$

and on the other hand

$$|c_n R(\alpha_i + 2\delta) g_n(\alpha_i + 2\delta) P_n(\alpha_i + 2\delta)| \geq \frac{\gamma^2 \eta^{m+p^+}}{2L^{m+1+2p^+}} \delta^{p-1} \frac{t^{m+p^+-k-p}}{L^{p^+}} \geq \frac{\gamma^2 (\eta t)^{m+p^+}}{2L^{m+1+3p^+}} \delta^{p-1}.$$

where we used the fact that all the roots of  $P_n$  that are not  $[\alpha_i - \delta, \alpha_i + \delta]$  are at distance at least  $t$  of  $\alpha_i + 2\delta$ . Therefore,

$$\begin{aligned} \sup_{x \in \mathcal{K}_{\varepsilon, \rho}^o} |H^n(x) - f_{\gamma}^{\varepsilon, \rho, o}(x)| &\geq \left| |c_n g_n(\alpha_i + 2\delta) P_n(\alpha_i + 2\delta)| - |g(\alpha_i + 2\delta) P(\alpha_i + 2\delta)| \right| \\ &\geq \delta^{p-1} \left( \frac{\gamma^2 (\eta t)^{m+p^+}}{2 L^{m+1+3p^+}} - \delta \frac{(2L)^{m+p^+}}{t^{p^+}} \right) > \delta', \end{aligned} \quad (30)$$

which is a contradiction. Symmetrically, one can do the same kind of reasoning when we suppose that  $P_n$  has strictly more than  $p$  zeroes on the interval  $[\alpha_i - \delta, \alpha_i + \delta]$ .

We now consider the case, specific to the framework with outliers, of an  $\alpha_i \in \{\ell\}$  of multiplicity  $p \geq 1$ . Assume that  $P_n$  has strictly less than  $p$  zeroes on the interval  $[\alpha_i - \rho, \alpha_i + \rho]$ . In this case,  $\alpha_i + 2\rho \in \mathcal{K}_{\varepsilon, \rho}^o$ . We have

$$|g(\alpha_i + 2\rho) R(\alpha_i + 2\rho) P(\alpha_i + 2\rho)| \leq LC \rho^{-q} (2\rho)^p L^{m+p^+-k-p} \leq C(2L)^m \rho^{p-q}.$$

On the other hand,

$$|c_n R(\alpha_i + 2\delta) g_n(\alpha_i + 2\delta) P_n(\alpha_i + 2\delta)| \geq c \frac{\gamma^2 (\eta t)^{m+p^+}}{2L^{m+1+2p^+}} \rho^{p-1-q},$$

so that

$$\sup_{x \in \mathcal{K}_{\varepsilon}} |H^n(x) - f_{\gamma}^{\varepsilon, \rho, o}(x)| \geq \rho^{p-1-s} \left( c \frac{\gamma^2 (\eta t)^{m+p^+}}{2L^{m+1+2p^+}} - \rho C(2L)^{m+p^+} \right) > \delta'.$$

This completes the proof.

## 10. APPLICATION TO $X_n$ RANDOM, FOLLOWING SOME CLASSICAL MATRIX DISTRIBUTION

This section is devoted to the proofs of the results stated in section 2.5.

### 10.1. Proof of Theorem 2.10.

The proof is a slight extension of the proof of [1, Theorem 2.6.6]. We introduce the notations  $\phi(\mu, x) = -V(x) + \beta \int \log |x - y| d\mu(y)$ , and  $\hat{\mu}_n = \frac{1}{n} \sum_{i=p+1}^n \delta_{\lambda_i^n}$ . Then

$$\begin{aligned} \mathbb{P}_{V, \beta}^n(d\lambda_1, \dots, d\lambda_n) &= \\ &= \frac{Z_{nV/(n-p), \beta}^{n-p}}{Z_{V, \beta}^n} e^{n \sum_{i=1}^p \phi(\hat{\mu}_n, \lambda_i) + \beta \sum_{1 \leq i < j \leq p} \log |\lambda_i - \lambda_j|} d\mathbb{P}_{nV/(n-p), \beta}^{n-p}(\lambda_{p+1}, \dots, \lambda_n) d\lambda_1 \cdots d\lambda_p. \end{aligned}$$

By [1, Lemma 2.6.7], the law  $\mathbb{P}_{V, \beta}^n$  is exponentially tight so that it is enough to estimate the probability of a small ball around  $\mathbf{x} = (x_1 \geq x_2 \geq \dots \geq x_p)$ , namely events of the form  $B(x, \delta) := \{\max_{1 \leq i \leq p} |\lambda_i - x_i| \leq \delta, \max |\lambda_i| \leq M\}$ .

We proceed as in [7], using the fact that  $\hat{\mu}_n$  converges to  $\mu_V$  much faster than exponentially under  $\mathbb{P}_{nV/(n-p), \beta}^{n-p}$  (indeed, its LDP is in the scale  $n^2$ ), we can replace  $\phi(\hat{\mu}_n, \lambda_i)$  by  $\phi(\mu_V, x_i)$ , whereas the ratio of partition functions converges by hypothesis.

To be more precise, let us sketch the proof of the upper bound. Note that there exists a constant  $\Phi_M$  such that on  $B(x, \delta)$ ,  $\phi(\mu, x)$  is bounded above by  $\Phi_M$  so that

$$\begin{aligned} \mathbb{P}_{V,\beta}^n(B(x, \delta)) &\leq \frac{Z_{nV/(n-p),\beta}^{n-p}}{Z_{V,\beta}^n} e^{\beta p(p-1)/2 \log(x_1-2)} (e^{np\Phi_M} \mathbb{P}_{nV/(n-p),\beta}^{n-p}(\hat{\mu}^n \in B_\varepsilon(\mu_V)^c) \\ &\quad + (2M)^p e^{n \sum_{i=1}^p \max_{|y-x_i| \leq \delta} \max_{\mu \in B_\varepsilon(\mu_V)} \phi(\mu,y)}) \end{aligned}$$

with  $B_\varepsilon(\mu_V)$  a small ball with radius  $\varepsilon$  around  $\mu_V$  for a distance compatible with the weak topology. As  $n^{-2} \log \mathbb{P}_{nV/(n-p),\beta}^{n-p}(\hat{\mu}^n \in B_\varepsilon(\mu_V)^c)$  is bounded above by a negative real number for all  $\delta > 0$  by the LDP for the law of  $\hat{\mu}^n$ , the first term is negligible as  $n$  goes to infinity. Using the fact that  $(x, \mu) \rightarrow \phi(\mu, x)$  is upper continuous, we obtain the upper bound by letting  $n$  going to infinity, then  $\delta$  decreasing to zero and finally  $\varepsilon$  going to zero. The lower bound is similar to the proof in [1, p. 84]. Indeed, note that we may assume the  $x_p$ 's away from the boundary of the support of  $\mu_V$  since otherwise we may and shall consider only the deviations of the first  $p-1$ th eigenvalues (and proceed by induction till  $p=1$  which is already known). There exists  $x_i^\delta, 1 \leq i \leq p$ , whose small neighbourhood are included in the  $\delta$  neighbourhood of  $x_i, 1 \leq i \leq p$ , and which are distinct, so that for  $\epsilon$  small enough

$$\mathbb{P}_{V,\beta}^n(\max_{1 \leq i \leq p} |\lambda_i - x_i| < \delta) \geq \mathbb{P}_{V,\beta}^n(\max_{1 \leq i \leq p} |\lambda_i - x_i^\delta| < \epsilon)$$

When the  $x_i$ 's are distinct and away from  $b_V$ , their logarithmic interaction is negligible and we refer the reader to [1, p. 84] as the proof is very similar. We get

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_{V,\beta}^n(\max_{1 \leq i \leq p} |\lambda_i - x_i| < \delta) \geq - \sum_{i=1}^p J_V(x_i^\delta) - \alpha_{V,\beta}^p$$

Now,  $J_V$  is continuous away from the support of  $\mu_V$  so that we can conclude by letting  $\delta$  going to zero.  $\square$

## 10.2. Proof of Theorem 2.13.

As explained in Section 2.5, we have to study  $\widetilde{X}_n$ , when  $X_n$  is diagonal with eigenvalues having  $\mathbb{P}_{V,\beta}^n$  as their joint law and the  $U_i$ 's obtained by orthonormalisation procedure from  $G = (g_1, \dots, g_r)$  iid standard Gaussian.

It is clear that since the largest eigenvalues of  $X_n$  are exponentially tight, so are the eigenvalues of  $\widetilde{X}_n$ , and therefore it is enough to prove a weak large deviation principle. We let  $K(L)$  be such that the probability that  $\lambda_1^n$  or  $\tilde{\lambda}_1^n$  is greater than  $K(L)$  is smaller than  $e^{-nL}$ . We first consider the deviations of  $(\tilde{\lambda}_1^n > \dots > \tilde{\lambda}_k^n)$  towards  $x_1 \geq x_2 \geq \dots > x_k > b_V$  with  $b_V$  the rightmost point of the support of  $\mu_V$ . Let  $\varepsilon > 0$  so that  $x_k \geq b_V + \varepsilon$ . The proof will consist in first fixing the possible deviations of the extreme eigenvalues of  $X_n$  (hence providing outliers) and then, being given these outliers, computing the deviations of the eigenvalues of  $\widetilde{X}_n$ . The main point of course is that with exponentially large probability, only a finite number of eigenvalues of  $X_n$  can deviate. For  $B$  an event  $B$  with positive probability, we denote in the sequel by  $\mathbb{P}(A|B)$  the probability of an event  $A$  conditionally to  $B$ .

- We first claim that for any  $(\ell_1, \dots, \ell_p) \in (b_V, +\infty)$  and any  $\eta > 0$ , there exists  $\varepsilon(\eta, \ell), \delta(\eta, \ell) > 0$  so that for  $n$  large enough, for  $\varepsilon < \varepsilon(\eta, \ell), \delta < \delta(\eta, \ell)$ ,

$$\begin{aligned} -L_{\ell_1, \dots, \ell_p}^0(x_1, \dots, x_k) - \eta &\leq \frac{1}{n} \log \mathbb{P} \left( \max_{1 \leq i \leq k} |\tilde{\lambda}_i^n - x_i| \leq \delta \left| \begin{array}{l} \max_{1 \leq i \leq p} |\lambda_i^n - \ell_i| \leq \varepsilon, \\ \lambda_{p+1}^n \leq b_V + \varepsilon \end{array} \right. \right) \quad (31) \\ &\leq -L_{\ell_1, \dots, \ell_p}^0(x_1, \dots, x_k) + \eta \end{aligned}$$

These inequalities were proven for  $\varepsilon$  going to zero a priori with  $n$  in Theorem 9.1. Let  $X_n$  be a matrix such that the event  $\{\max_{1 \leq i \leq p} |\lambda_i^n - \ell_i| \leq \varepsilon\}$  is realised. Let  $X'_n$  be a real diagonal matrix with same eigenvalues of  $X_n$  except its  $k$  largest eigenvalues are equal to the outliers  $(\ell_1, \dots, \ell_p)$ . Then we have  $\|X_n - X'_n\|_\infty \leq \varepsilon$ , so that, with obvious notations,  $\|\tilde{X}_n - \tilde{X}'_n\|_\infty \leq \varepsilon$ , so that the ordered eigenvalues of  $\tilde{X}_n$  and  $\tilde{X}'_n$  differ at most by  $\varepsilon$ . Thus, up to change  $\delta$  into  $\delta \pm \varepsilon$ , Theorem 9.1 gives (31) gives the inequalities for  $X'_n$  for  $\eta/2$ , from which we deduce .

- By Theorem 2.10, there exists  $\varepsilon(\eta, \ell) > 0$  so that for  $n$  large enough, for  $\varepsilon \leq \varepsilon(\eta, \ell)$

$$-J^p(\ell_1, \dots, \ell_p) - \eta \leq \frac{1}{n} \log \mathbb{P} \left( \max_{1 \leq i \leq p} |\lambda_i^n - \ell_i| \leq \varepsilon, \lambda_{p+1}^n \leq b_V + \varepsilon \right) \leq -J^p(\ell_1, \dots, \ell_p) + \eta. \quad (32)$$

For all  $(\ell_1, \dots, \ell_p)$ , we define the set

$$V(\ell_1, \dots, \ell_p) = \{(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n; \max_{1 \leq i \leq p} |\lambda_i - \ell_i| < \varepsilon(\eta, \ell), \lambda_{p+1} < b_V + \varepsilon(\eta, \ell)\}.$$

- We observe, again by Theorem 2.10 that for all  $\kappa > 0$ , the probability that  $\lambda_p^n$  is greater than  $b_V + \delta$  goes to zero as  $e^{-np\varepsilon(\delta)}$  for some  $\varepsilon(\delta) = \inf_{(b_V + \delta, +\infty)} J_V > 0$ .
- To prove the upper bound we can write

$$\mathbb{P} \left( \max_{1 \leq i \leq k} |\tilde{\lambda}_i^n - x_i| \leq \delta \right) \leq \mathbb{P} \left( \max_{1 \leq i \leq k} |\tilde{\lambda}_i^n - x_i| \leq \delta, \lambda_{p+1}^n \leq b_V + \delta \right) + e^{-np\varepsilon(\delta)} \quad (33)$$

We next consider the infinite open covering  $\cup V(\ell_1, \dots, \ell_p)$  of  $[b(V), K(L)]^p$ . We can always extract a finite covering  $\cup_{1 \leq s \leq M(\delta)} V(\ell_1^s, \dots, \ell_p^s)$  as the latter is compact. We then take  $\delta = \min \delta(\eta, \ell^s) > 0$ . Thus, we get by the LDP estimate (32)

$$\begin{aligned} \mathbb{P} \left( \max_{1 \leq i \leq k} |\tilde{\lambda}_i^n - x_i| \leq 2\delta \right) &\leq e^{-nL} + e^{-np\varepsilon(\delta)} + \sum_{s=1}^{M(\delta)} \mathbb{P} \left( \max_{1 \leq i \leq k} |\tilde{\lambda}_i^n - x_i| \leq 2\delta \cap V(\ell_1^s, \dots, \ell_p^s) \right) \\ &\leq e^{-nL} + e^{-np\varepsilon(\delta)} + \sum_{s=1}^{M(\delta)} e^{-nL_{\ell_1^s, \dots, \ell_p^s}^0(x_1, \dots, x_k) - nJ^p(\ell_1^s, \dots, \ell_p^s) + n\eta} \\ &\leq M(\delta) e^{-n \min_{1 \leq s \leq M(\delta)} (L_{\ell_1^s, \dots, \ell_p^s}^0(x_1, \dots, x_k) + J^p(\ell_1^s, \dots, \ell_p^s) - \eta)} + e^{-nL} + e^{-np\varepsilon(\delta)} \\ &\leq 3M(\delta) e^{-n \min\{L, p\varepsilon(\delta), \bar{J}^k(x_1, \dots, x_k)\}} \end{aligned}$$

which gives the announced bound by taking first the limit as  $n$  goes to infinity, then  $\delta$  to zero and  $L$  to infinity.

- The lower bound is easier as we simply write

$$\mathbb{P} \left( \max_{1 \leq i \leq k} |\tilde{\lambda}_i^n - x_i| \leq 2\delta \right) \geq \mathbb{P} \left( \max_{1 \leq i \leq k} |\tilde{\lambda}_i^n - x_i| \leq 2\delta \cap V(\ell_1^s, \dots, \ell_p^s) \right)$$

and use the large deviation theorems.  $\square$

## 11. APPENDIX

### 11.1. An auxiliary large deviations result.

For sake of completeness, we recall the following result, following from [27, Theorem 3.2] (see also [7] and [10]).

**Theorem 11.1.** *Let, for each  $n$ ,  $\lambda_1^n \geq \dots \geq \lambda_n^n$  be a family of deterministic real numbers such that there exists a compactly supported probability measure  $\mu$  on  $\mathbb{R}$  such that*

$$\frac{1}{n} \sum_{i=1}^n \delta_{\lambda_i^n} \xrightarrow[n \rightarrow +\infty]{} \mu \quad \text{weakly,} \quad (34)$$

Let  $(Z_n)_{n \geq 1}$  be a sequence of  $\mathbb{R}^d$ -valued i.i.d. random variables such that  $Z_1$  has some exponential moments. Let  $f : \mathbb{R} \rightarrow \mathbb{R}^{m \times d}$  be a continuous matrix-valued function and consider the sequence of  $\mathbb{R}^m$ -valued random variables

$$L_n := \frac{1}{n} \sum_{i=1}^n f(\lambda_i^n) \cdot Z_i.$$

- (1) *If we assume moreover that*

$$\lambda_1^n \xrightarrow[n \rightarrow +\infty]{} \max(\text{supp}(\mu)), \quad \lambda_n^n \xrightarrow[n \rightarrow +\infty]{} \min(\text{supp}(\mu)). \quad (35)$$

then  $L_n$  satisfies a LDP with good rate function

$$\Gamma^*(y) = \sup_{\lambda \in \mathbb{R}^m} \{ \langle \lambda, y \rangle - \Gamma(\lambda) \},$$

where  $\Gamma(\lambda)$  is defined by the formula

$$\Gamma(\lambda) = \int_{x \in \mathbb{R}} \Lambda \left( \sum_{k=1}^m \lambda_k \cdot k\text{-th row of } f(x) \right) d\mu(x)$$

and  $\Lambda$  is defined by the formula  $\Lambda(\theta) = \log \mathbb{E}(e^{\langle \theta, Z_1 \rangle})$ .

- (2) *If there exists  $p^+, p^- \in \mathbb{N}$  such that for any  $i \leq p^+$ ,  $\lambda_i^n \xrightarrow[n \rightarrow +\infty]{} \ell_i^+$ , for any  $j \leq p^-$ ,*

$$\lambda_{n-j+1}^n \xrightarrow[n \rightarrow +\infty]{} \ell_j^-, \lambda_{p^++1}^n \xrightarrow[n \rightarrow +\infty]{} \max(\mathcal{S}) \text{ and } \lambda_{n-p^-}^n \xrightarrow[n \rightarrow +\infty]{} \min(\mathcal{S}) \text{ with}$$

$$-\infty < \ell_1^- \leq \dots \leq \ell_{p^-}^- < \min(\mathcal{S}) \leq \max(\mathcal{S}) < \ell_{p^+}^+ \leq \dots \leq \ell_1^+ < \infty$$

and if  $\frac{Z_n}{n}$  satisfies the LDP with a good rate function denoted by  $I^{(Z)}$ , then  $L_n$  satisfies a LDP with good rate function

$$I_f(y) = \inf \left\{ \Gamma^*(y_0) + \sum_{i=1}^{p^++p^-} I^{(Z)}(y_i) \middle/ y_0 \in \mathbb{R}^m, y_1, \dots, y_{p^++p^-} \in \mathbb{R}^d, y_0 + \sum_{i=1}^{p^++p^-} f(m_i) \cdot y_i = y \right\},$$

with  $m_i = \ell_i^+$  if  $1 \leq i \leq p^+$  and  $m_i = \ell_i^-$  if  $p^+ + 1 \leq i \leq p^+ + p^-$  and where  $\Gamma^*$  and  $\Lambda$  are defined as above.

**Remark 11.2.** Note that in [27, Theorem 3.2], the assumption that  $\frac{Z_1}{n}$  satisfies the LDP with a good rate function is required in both case but a careful reading of the proof shows that in the case when (35) is fulfilled, this is not a necessary condition.

## 11.2. Proof of a technical lemma.

With the notations of Section 4.1, we have the following result

**Lemma 11.3.** Under Assumption 2.2, for any  $1 \leq i_0 \leq r$ , we have

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\|q_{i_0}^n W_{i_0}^n\|_2^2 \notin [\delta, \frac{1}{\delta}]) = -\infty.$$

*Proof.* To simplify the notations, we shall assume that  $i_0 = r$ .

Recall that the  $G_i^n$ 's were constructed from a family  $(G(k) = (g_1(k), \dots, g_r(k))_{k \geq 1})$  of independent copies of  $G$ , via the formula  $G_i^n := (g_i(1), \dots, g_i(n))^T$ . For  $1 \leq k$ , we consider the random  $r \times r$  Hermitian matrix

$$Z_k = G(k)^* G(k) = [\overline{g_i(k)} g_j(k)]_{1 \leq i, j \leq r} \quad \text{and} \quad L^n = \frac{1}{n} \sum_{k=1}^n Z_k.$$

By Theorem 11.1, we have that the law of  $L^n$  satisfies a LDP with convex good rate function

$$I^{(L)}(y) = \sup_{\lambda \in \mathbb{H}_r} \{\langle \lambda, y \rangle - \Lambda(\lambda)\},$$

where  $\Lambda(\lambda) = \log \mathbb{E}(e^{\langle \lambda, Z_1 \rangle})$  is exactly the function defined in Equation (8).

Note that since for all  $n$ ,  $L^n$  is almost surely a positive semi-definite matrix, by closedness of the set of such matrices, the domain of  $I$  is contained in the set of positive semi-definite matrices.

Let  $P_r$  be the real polynomial function on  $\mathbb{H}_r$  introduced in Proposition 4.1: we have  $\|q_r^n W_r^n\|_2^2 = P_r(L^n)$ . Therefore, if, for any  $\delta > 0$ , we introduce the closed set  $\mathcal{E}_\delta := \{y \in \mathbb{H}_r; P_r(y) \leq \delta\}$ , we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\|q_r^n W_r^n\|_2^2 \leq \delta) \leq - \inf_{y \in \mathcal{E}_\delta} I^{(L)}(y).$$

Let us assume that

$$M := \liminf_{\delta \downarrow 0} \inf_{y \in \mathcal{E}_\delta} I^{(L)}(y) < \infty.$$

Since  $I^{(L)}$  is a good rate function, there exists a compact set  $K$  such that  $\inf_{y \in K^c} I^{(L)}(y) > M$ , so that for all  $\delta > 0$ ,  $\inf_{y \in \mathcal{E}_\delta} I^{(L)}(y) = \inf_{y \in \mathcal{E}_\delta \cap K} I^{(L)}(y)$ . Moreover the infimum on  $\mathcal{E}_\delta$  is reached : let, for all  $n \geq 0$ ,  $y_n$  be an element of  $K$  such that  $I^{(L)}(y_n) = \inf_{y \in \mathcal{E}_{\frac{1}{n}}} I^{(L)}(y)$ . There exists a subsequence  $\varphi(n)$  such that  $y_{\varphi(n)}$  converges, as  $n$  goes to infinity to some  $y_0$ . By continuity of  $P_r$ ,  $P_r(y_0) = \lim_{n \rightarrow \infty} P_r(y_{\varphi(n)}) = 0$ . It follows, by the last part of Proposition 4.1, that  $y_0$  is not positive definite. However, since  $I^{(L)}$  is lower semicontinuous, we

have  $I^{(L)}(y_0) \leq M < \infty$ , which implies that  $y_0$  is a positive semi-definite matrix. Let  $p$  be the orthogonal projection onto  $\ker y_0$ . Note that  $p \neq 0$  and that  $\langle p, y_0 \rangle = \text{Tr}(y_0 p) = 0$ .

$$\begin{aligned} I^{(L)}(y_0) &= \sup_{\lambda \in \mathbb{H}_r} \{ \langle \lambda, y_0 \rangle - \Lambda(\lambda) \} \\ &\geq \sup_{t > 0} \{ \langle -tp, y_0 \rangle - \Lambda(-tp) \} \\ &= \sup_{t > 0} -\Lambda(-tp) \\ &= +\infty \quad \text{by (11),} \end{aligned}$$

which yields a contradiction (as we already proved that  $I^{(L)}(y_0) \leq M$ ).

Similarly, as  $I^{(L)}$  is a good rate function, it has compact level sets and therefore has to be large on the set  $\{y : P_r(y) \geq 1/\delta\}$ . Hence,

$$\limsup_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\|q_r^n W_r^n\|_2^2 \geq \delta^{-1}) = -\infty$$

which completes the proof of the lemma.  $\square$

### 11.3. On the eigenvalues of the deformed matrix.

The goal of this section is to prove Lemma 9.3. In fact, we will prove the slightly more general

**Lemma 11.4.** *Let  $\mathbb{K}$  be either  $\mathbb{R}$  or  $\mathbb{C}$ . Let us fix some positive integers  $n, r$  such that  $n > r$ , a self adjoint  $n \times n$  real matrix  $X$  with eigenvalues  $\lambda_1, \dots, \lambda_n$  and some non null real numbers  $\theta_1, \dots, \theta_r$ . We make the following hypothesis:*

(H)  $\lambda_1, \dots, \lambda_n$  are pair-wise distinct and there are pair-wise distinct indices  $i_1, \dots, i_{r-1} \in \{1, \dots, n\}$  such that  $\{\lambda_{i_1} + \theta_1, \dots, \lambda_{i_{r-1}} + \theta_{r-1}\} \cap \{\lambda_1, \dots, \lambda_n\} = \emptyset$ .

Let us define, for  $g = [g_1, \dots, g_r] \in \mathbb{K}^{n \times r}$ ,

$$\tilde{X}_g := X + \theta_1 u_1 u_1^* + \dots + \theta_r u_r u_r^*,$$

where  $(u_1, \dots, u_r)$  is either the orthonormalized family deduced from the columns of  $g$  by the Gram-Schmidt process or  $\frac{1}{\sqrt{n}}(g_1, \dots, g_r)$ . Then the Lebesgue measure of the set of the  $g$ 's such that  $\tilde{X}_g$  and  $X$  have at least one eigenvalue in common is null.

Now, Lemma 9.3 will be easy to deduce from the above. Indeed, one can check that for  $n$  large enough,  $X_n$  satisfies hypothesis (H). We know that its eigenvalues  $\lambda_1^n, \dots, \lambda_n^n$  are distinct. Moreover, let  $\eta$  be such that  $\eta < \frac{1}{2} \min_{1 \leq i \leq r} |\theta_i|$  and  $\eta < \frac{1}{3} \min_{i \neq j} |\ell_i - \ell_j|$ . From Assumption 2.5, there exists  $n$  large enough so that  $X_n$  has at most  $p^+$  eigenvalues greater than  $b + \eta$ , at most  $p^-$  eigenvalues smaller than  $a - \eta$ , more than  $2r(p^+ + 1)$  eigenvalues in the interval  $(b - \eta, b + \eta)$  and more than  $2r(p^- + 1)$  eigenvalues in  $(a - \eta, a + \eta)$ .

Let us assume that  $\theta_1 > 0$ . Then one can find an eigenvalue  $\lambda_{i_1}$  among the  $p^+ + 1$  greater ones in  $(b - \eta, b + \eta)$  such that  $\lambda_{i_1} + \theta_1$  do not belong to  $\{\lambda_1^n, \dots, \lambda_n^n\}$ . We then forget the  $p^+ + 1$  greater eigenvalues and look at the  $p^+ + 1$  following. Among them, one can find



an eigenvalue  $\lambda_{i_2}$  such that  $\lambda_{i_2} + \theta_2$  do not belong to  $\{\lambda_1^n, \dots, \lambda_n^n\}$ . and so on. For the negative  $\theta_i$ 's, we consider the  $p^- + 1$  smallest eigenvalues in  $(a - \eta, a + \eta)$ .

From Assumption 2.6, the law of  $G$  is absolutely continuous with respect to Lebesgue measure and Lemma 9.3 follows.

We now prove Lemma 11.4.

*Proof.* The idea of the proof is the following. We shall first prove (Step I) that the set of  $g$ 's such that  $\tilde{X}_g$  and  $X$  have at least one eigenvalue in common is, up to a set of null Lebesgue measure, the set of zeroes of a polynomial function. Since it can easily be proved, by induction on the number of variables, that the set of zeroes of any non null polynomial in several real variables has null Lebesgue measure, it will then only remain to prove (Step II) that this function is not identically null to know that the set of such  $g$ 's has vanishing Lebesgue measure.

Let  $\beta$  be either 1 or 2 according to whether  $\mathbb{K}$  is  $\mathbb{R}$  or  $\mathbb{C}$ .

*Step I.* Let us first treat the case where  $(u_1, \dots, u_r) = \frac{1}{\sqrt{n}}(g_1, \dots, g_r)$ . Let us define  $P$  to be the polynomial of  $\beta nr$  real variables which maps  $[g_1, \dots, g_r] \in \mathbb{K}^{n \times r}$  to the resultant of the characteristic polynomials of  $X$  and  $\tilde{X}_g$ . The set of  $g$ 's in  $\mathbb{K}^{n \times r}$  such that  $X$  and  $\tilde{X}_g$  have an eigenvalue in common is exactly the set of  $g$ 's such that  $P(g) = 0$  : Step I is achieved in the case where  $(u_1, \dots, u_r) = \frac{1}{\sqrt{n}}(g_1, \dots, g_r)$ .

Let us now treat the case where  $(u_1, \dots, u_r)$  is the orthonormalized family deduced from the columns of  $g$  by the Gram-Schmidt process. In this case, the resultant of the characteristic polynomials of  $X$  and of  $\tilde{X}_g$  is not anymore a polynomial function of the real coordinates of  $g$ , so we shall use the following trick. It can easily be noticed, through a careful look at the Gram-Schmidt process, that for all  $k \in \{1, \dots, r\}$ , for all  $i, j \in \{1, \dots, n\}$ , there are two polynomial functions of  $D_k, N_{k,i,j}$  of  $\beta nr$  real variables such that the  $i, j$ -th entry of  $u_k u_k^*$  is  $\frac{N_{k,i,j}(g)}{D_k(g)}$  and that  $D_k(g)$  is positive for any  $g \in \mathbb{K}^{n \times r}$  which columns are linearly independent. Let us define the polynomial function of  $\beta nr$  real variables

$$D(g) := \prod_{k=1}^r D_k(g).$$

For any  $g$  such that  $D(g) > 0$  (which is the case for any  $g \in \mathbb{K}^{n \times r}$  which columns are linearly independent),  $X$  and  $\tilde{X}_g$  have no eigenvalue in common if and only if  $D(g)X$  and  $D(g)\tilde{X}_g$  have no eigenvalue in common. Now, the advantage of having replaced  $X$  and  $\tilde{X}_g$  by  $D(g)X$  and  $D(g)\tilde{X}_g$  is that the entries of  $D(g)X$  and  $D(g)\tilde{X}_g$  are polynomial functions of  $g$ . Hence if one defines  $P(g)$  to be the resultant of the characteristic polynomials of  $D(g)X$  and  $D(g)\tilde{X}_g$ ,  $P(g)$  is a polynomial function of the  $\beta nr$  real coordinates of  $g$  and, up to the set (with zero Lebesgue measure) of  $g$ 's in  $\mathbb{K}^{n \times r}$  which columns are linearly independent, the set of  $g$ 's in  $\mathbb{K}^{n \times r}$  such that  $X$  and  $\tilde{X}_g$  have an eigenvalue in common is exactly the set of  $g$ 's such that  $P(g) = 0$  : Step I is achieved in the second case.

*Step II.* Let us now prove that in both cases, the polynomial function  $g \mapsto P(g)$  is not identically null. To treat both cases together, it suffices to prove that there exists  $g = [g_1, \dots, g_r] \in \mathbb{K}^{n \times r}$  with orthonormalized columns such that  $\tilde{X}_g$  and  $X$  have no

eigenvalue in common. One can suppose that  $i_1 = 1, \dots, i_{r-1} = r - 1$ , that  $\lambda_r < \dots < \lambda_n$  and that

$$X = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

We shall choose the  $r - 1$  first columns  $g_1, \dots, g_{r-1}$  of  $g$  to be the  $r - 1$  first elements of the canonical basis and  $g_r$  with null  $r - 1$  first coordinates and unit norm. With such a choice of  $g$ , we have

$$\tilde{X}_g = \begin{bmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_{r-1} & & \\ & & & \lambda_r & \\ & & & & \ddots \\ & & & & & \lambda_n \end{bmatrix} + \begin{bmatrix} \theta_1 & & & & \\ & \ddots & & & \\ & & \theta_{r-1} & & \\ & & & & \\ & & & & \theta_r g_r g_r^* \end{bmatrix}.$$

Let us suppose that  $\theta_r > 0$ . It was shown in [23, Section 3.2] that as  $g_r$  runs through the set of unit norm vectors of  $\mathbb{K}^{n \times 1}$  with null  $r - 1$  first coordinates, the ordered eigenvalues of the  $n - (r - 1) \times n - (r - 1)$  lower right block of  $\tilde{X}_g$  describe the set of families  $\mu_r, \dots, \mu_n$  of real numbers which sum up to  $\lambda_r + \dots + \lambda_n + \theta_r$  and such that

$$\lambda_r \leq \mu_r \leq \lambda_{r+1} \leq \dots \leq \lambda_n \leq \mu_n.$$

One can easily find such a family  $\mu_r, \dots, \mu_n$  such that

$$\{\mu_r, \dots, \mu_n\} \cap \{\lambda_1, \dots, \lambda_n\} = \emptyset,$$

which concludes the proof, by hypothesis (H).  $\square$

## REFERENCES

- [1] G. Anderson, A. Guionnet, O. Zeitouni *An Introduction to Random Matrices*, Cambridge University Press (2009).
- [2] Z. Bai and J-F Yao *Central limit theorems for eigenvalues in a spiked population model*, Ann. Inst. Henri Poincaré Probab. Stat. **44**(2008) 447–474
- [3] Z.D Bai and J. W. Silverstein *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Prob. **26**,(1998) 316–345
- [4] Z.D Bai and J. W. Silverstein *Spectral analysis of large dimensional random matrices*, Springer Series in Statistics, New York (2010)
- [5] Z. D. Bai, J.-F. Yao *Central limit theorems for eigenvalues in a spiked population model*, Ann. I.H.P.-Prob. et Stat., 2008, Vol. 44 No. 3, 447–474.
- [6] J. Baik, G. Ben Arous and S. Péché *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Prob. **33** (2005) 1643-1697.
- [7] G. Ben Arous, A. Dembo, A. and A. Guionnet, *Aging of spherical spin glasses*, Probab. Theory Related Fields, **120**, (2001)1, 1–67.
- [8] G. Ben Arous, A. Guionnet. *Large deviations for Wigner’s law and Voiculescu’s non commutative entropy*, Probability theory and related fields, **108**, 517-542 (1997).
- [9] F. Benaych-Georges, R. N. Rao. *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Preprint.
- [10] B. Bercu and F. Gamboa and A. Rouault, *Large deviations for quadratic forms of stationary Gaussian processes*, Stochastic Process. Appl., **71**(1997)75–90
- [11] M. Capitaine, C. Donati-Martin, D. Féral *The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations*, Ann. Probab. **37**(2009)1–47.

- [12] M. Capitaine, C. Donati-Martin, D. Féral *Central limit theorems for eigenvalues of deformations of Wigner matrices*, Preprint.
- [13] A. Dembo, A. Guionnet and O. Zeitouni, *Moderate Deviations for the Spectral Measure of certain Random Matrices*, Ann. Inst. H. Poincaré Probab. Statist. **39** (2003) 1013.
- [14] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*. Springer.
- [15] H. Döring and P. Eichelsbacher, *Moderate deviations in random graphs and Bernoulli random matrices*, Electron. J. Probab. **14** (2009) 2636.
- [16] L. Erdős, S. Péché, J. Ramírez, B. Schlein, Benjamin and H.T Yau, *Bulk universality for Wigner matrices*, Comm. Pure Appl. Math. **63**(2010) 895–925.
- [17] D. Féral and S. Péché *The largest eigenvalue of rank one deformation of large Wigner matrices*, Comm. Math. Phys. **272** (2007)185–228.
- [18] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Elec. Comm. in Probab. **5** (2000), 119–136.
- [19] F. Hiai and D. Petz *Eigenvalue density of the Wishart matrix and large deviations*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **1** (1998) 633–646
- [20] M. Maïda *Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles*, Elec. J. Probab.
- [21] D. Féral *On large deviations for the spectral measure of discrete Coulomb gas*, Séminaire de probabilités XLI, Lecture Notes in Math. **1934** (2008) 19–49
- [22] A. Fey, R. van der Hofstad, Remco and M. Klok *Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems*, Adv. in Appl. Probab. **40** (2008) 1048–1071
- [23] P. J. Forrester, T. Nagao *Determinantal Correlations for Classical Projection Processes* Arxiv.
- [24] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica **1** (1981),233–241.
- [25] A. Guionnet and O. Zeitouni, *Large deviations asymptotics for spherical integrals*, J. Funct. Anal. **188** (2002)461–515
- [26] R. Horn and C. R. Johnson *Matrix Analysis*, Cambridge University Press (1990)
- [27] M. Maïda, J. Najim, S. Péché *Large deviations for weighted empirical mean with outliers*, Stochastic Process. Appl. **117**1373–1403 (2007)
- [28] L. A. Pastur and V.A Martchenko, *The distribution of eigenvalues in certain sets of random matrices*, Math. USSR-Sbornik **1** (1967) 457–483.
- [29] T. Tao and A. Knutson *Honeycombs and sums of Hermitian matrices*, Notices Amer. Math. Soc. **48** (2001) 175–186
- [30] T. Tao and V. Vu, *Random matrices: Universality of local eigenvalue statistics*, arXiv:0906.0510
- [31] P. Vivo, S. Majumdar and O. Bohigas *Large deviations of the maximum eigenvalue in Wishart random matrices*, J. Phys. A **40** (2007) 4317–4337
- [32] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. Math. **62** (1955), 548–564.