Probab. Theory Relat. Fields (2012) 154:703–751
DOI 10.1007/s00440-011-0382-3

Large deviations of the extreme eigenvalues of random deformations of matrices

F. Benaych-Georges · A. Guionnet · M. Maida

Received: 26 October 2010 / Revised: 17 June 2011 / Published online: 6 July 2011

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, 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 $e^{-n \operatorname{Tr} V(X)}\,\mathrm{d}X$, for V growing fast enough at infinity and any perturbation of finite rank.

Keywords Random matrices · Large deviations

Mathematics Subject Classification (2000) 15A52 · 60F10

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

F. Benaych-Georges (✉)
UPMC Université Paris 6, LPMA, Case courier 188, 4 Place Jussieu, 75252 Paris Cedex 05, France
e-mail: florent.benaych@upmc.fr

A. Guionnet
UMPA, ENS Lyon, 46 allée d'Italie, 69364 Lyon Cedex 07, France
e-mail: aguionne@umpa.ens-lyon.fr

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

Contents

1 Introduction

In the last 20 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 (the canonical examples hereafter will be Wigner matrices [36], or Wishart matrices [31]), it has been shown that the spectral measure converges almost surely. The extreme eigenvalues converge for most of these models to the boundaries of the limiting spectral measure (see e.g. [4] or [25]). 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 [2,32], or [1,5] 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. [18] or [34]). Concentration of measure phenomenon and moderate deviations could also be established in [15,17,20].

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, mainly based on independence, do not apply. Moreover, large deviations rate functions have to depend on the distribution of the entries and only guessing their definition is still a widely 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 (LDP) for the law of the spectral measure of such matrices was proved in [9]. This extends to other ensembles distributed according to similar Gibbs measure, for instance Gaussian Wishart matrices [21]. Similar large deviation results hold in discrete situations with a Coulomb gas distribution [22]. A LDP was also established in [26] 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 uses stochastic analysis and Dyson's Brownian motion, as there is no explicit joint law for the eigenvalues, but again relies 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 LDP was derived in [8] for the same Gaussian type models.

The large deviations of extreme eigenvalues of Gaussian Wishart matrices were studied in [35]. 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 bounds for the extreme eigenvalues could be derived under more general assumptions on the entries in [23]. Our approaches allow also to obtain a full large deviation for the spectrum of such Wishart matrices when r is kept fixed while n goes to infinity (see Sect. 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, using Weyl's interlacing property, it is easy to check that such finite rank perturbations do not change the deviations of the spectral measure. But it strongly affects the behavior of a few extreme eigenvalues, not only at the level of deviations but also as far as convergence and fluctuations are concerned. In the case of Gaussian Wishart matrices, the asymptotics of these extreme eigenvalues were established in [7]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 [7] 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. [6,13,14,19].

The goal of this article is to study the large deviations of the extreme eigenvalues of such finite rank perturbations of large random matrices. In [28], a LDP 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 outcomes 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 $\propto e^{-n\,\mathrm{tr}(V(X))}\,\mathrm{d}X$, for which the deviations are well known (see Theorem 2.10). 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. We show that the law of these extreme eigenvalues satisfies a LDP 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 Sect. 7). In fact, even in this simple case of sample covariance matrices with non-Gaussian entries, our large deviation result seems to be new and improves on [23].

Our approach is based, as in [6,13,14], on the characterization of the eigenvalues via the determinant of a matrix with fixed size: it is an $r \times r$ matrix whose entries are the Stieltjes transforms of the non-deformed matrix evaluated along the random vectors of the perturbation. We obtain a LDP for the law of this characteristic polynomial (seen as a continuous function outside of the spectrum of the deterministic matrix) by classical large deviation techniques. Even though the application which associate to a function its zeroes is not continuous for the weak topology, we deduce from the latter a LDP for the law of the zeroes of this characteristic polynomial, that is the extreme eigenvalues of the deformed matrix model.

2 Statement of the results

2.1 The models

Let X_n be a real diagonal matrix of size $n \times n$ with eigenvalues $\lambda_1^n \geq \lambda_2^n \geq \cdots \geq \lambda_n^n$.

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 $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_m > 0 > \theta_{m+1} \geq \cdots \geq \theta_r$ be fixed, let $G = (g_1, \ldots, g_r)$ be a random vector and $(G(k) = (g_1(k), \ldots, g_r(k)))_{k \geq 1}$ be independent copies of G. We then define the r vectors with dimension n

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

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

$$\widetilde{X_n} = X_n + \frac{1}{n} \sum_{i=1}^{r} \theta_i G_i^n G_i^{n*}. \tag{1}$$

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

Alternatively, if we assume moreover that the law of G does not charge any hyperplane, then, for $n > r$, the r vectors $G_1^n, \ldots, G_r^n$ 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 then consider the eigenvalues $\widetilde{\lambda}_1^n \geq \cdots \geq \widetilde{\lambda}_n^n$ of

$$\widetilde{X_n} = X_n + \sum_{i=1}^{r} \theta_i U_i^n U_i^{n*} \tag{2}$$

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

If $g_1, \ldots, g_r$ 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 [11].

Our goal will be to examine the large deviations for the m largest eigenvalues of the deformed matrix $\widetilde{X_n}$, 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 μ.

Concerning the random vector G, we make the following assumption. It allows to claim that with probability one, the column vectors $G_1^n, \ldots, G_r^n$ are linearly independent and is technically needed in the proof of Lemma 11.1. It is also the reason why we say that the column vectors $G_1^n, \ldots, G_r^n$ or $U_1^n, \ldots, U_r^n$ are *delocalized* with respect to the eigenvectors of X_n. Indeed, the eigenvectors of X_n are the vectors of the canonical basis, whereas we know that with probability one, none of the entries of the G_i^n's (or of the U_i^n's) is zero. The i.i.d. feature of the $G(k)$'s allows even to assert that all entries of each G_i^n's (or of the U_i^n's) have the same distribution.

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

The law of G could also depend on n provided it satisfies the above hypothesis uniformly on n and converges in law as n goes to infinity.

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 μ (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 μ.

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

Theorem 2.4 *Under Assumptions 2.1, 2.2 and 2.3, the law of the m largest eigenvalues $(\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_m^n) \in \mathbb{R}^m$ of $\widetilde{X_n}$ satisfies a LDP in the scale n with a good rate function L. In other words, for any $K \in \mathbb{R}^+$, $\{L \le K\}$ is a compact subset of $\mathbb{R}^m$, for any closed set F of $\mathbb{R}^m$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}((\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_m^n) \in F) \le -\inf_F L$$

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

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}((\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_m^n) \in O) \geq -\inf_O L.$$

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

Theorem 2.4 is true for both the i.i.d. perturbation model and the orthonormalized perturbation model, but the exact expression of the rate function L is not the same for both models. As could be expected, the minimum $(\lambda_1^*, \ldots, \lambda_m^*)$ only depends on the θ_i's, on the limiting spectral distribution μ of X_n, and on the covariance matrix of the vector G, this latter dependence coming from the fact that the rate function involves a Laplace transform of the law of G and its behavior near the extremum will generically be governed by the second derivatives, that is the covariance.

The rate function L is not explicit in general. However, in the particular case where $X_n = 0$, L can be evaluated. It amounts to consider the large deviations of 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. L is very explicit when G is Gaussian but even when the entries are not Gaussian, we can recover a LDP and refine a bound of [23] about the deviations of the largest eigenvalue (see Sect. 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 $i \leq p^+$, $\lambda_i^n \underset{n\to+\infty}{\longrightarrow} \ell_i^+$, for any $j \leq p^-$, $\lambda_{n-j+1}^n \underset{n\to+\infty}{\longrightarrow} \ell_j^-$, $\lambda_{p^++1}^n \underset{n\to+\infty}{\longrightarrow} b$ and $\lambda_{n-p^-}^n \underset{n\to+\infty}{\longrightarrow} a$ with $-\infty < \ell_1^- \leq \cdots \leq \ell_{p^-}^- < a \leq b < \ell_{p^+}^+ \leq \cdots \leq \ell_1^+ < \infty$, where a and b denote respectively the lower and upper bounds of the support of the limiting measure μ.

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 following assumption.

Assumption 2.6 The law of the vector $\frac{G}{\sqrt{n}}$ satisfies a LDP 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 $\widetilde{X_n}$ satisfies a LDP with a good rate function L^o.*

Again, Theorem 2.7 is true for both i.i.d. perturbation model and orthonormalized perturbation model, but the rate function is not the same for both models. A precise definition of L^o will be given in Theorem 9.1.

Before going any further, let us discuss Assumption 2.6. On one side, let us give some natural examples for which the assumption is fulfilled.

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 Assumption 2.6 holds with I infinite except at 0, where it takes value 0.*

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

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

so that taking the large n limit and then L going to infinity yields for any $\delta > 0$

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

thus proving the claim. □

On the other side, we want to emphasize that in the case with outliers, the individual LDP stated in Assumption 2.6 will be crucial. To understand more deeply this phenomenon, we refer the interested reader to some couterexamples when this assumption is not fulfilled that are studied in [30, Sect. 2.3] and a related discussion in the introduction of [29].

2.5 Large deviations for the largest eigenvalues of perturbed matrix models

We apply hereafter the results above to study the large deviations of the law of the extreme eigenvalues of perturbations of randomly chosen matrices X_n distributed according to the Gibbs measure

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

with $d^\beta X$ the Lebesgue measure on the set of $n \times n$ Hermitian matrices if $\beta = 2$ (corresponding to G $\mathbb{C}^r$-valued) or $n \times n$ symmetric matrices if $\beta = 1$ (corresponding to G $\mathbb{R}^r$-valued).

Let us first recall a few facts about the non-perturbed model. It is well known that if X_n is distributed according to μ_β^n, the law of the eigenvalues of X_n is given by

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

We will make on the potential V the following assumptions:

Assumption 2.9 (i) V is continuous with values in $\mathbb{R} \cup \{+\infty\}$ and

$$\liminf_{|x|\to\infty} \frac{V(x)}{\beta \log|x|} > 1.$$

(ii) For all integer numbers p, the limit

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{Z^{n-p}_{nV/n-p,\beta}}{Z^n_{V,\beta}}$$

exists and is denoted by $\alpha^p_{V,\beta}$.

(iii) Under $\mathbb{P}^n_{V,\beta}$, the largest eigenvalue λ^n_1 converges almost surely to the upper boundary b_V of the support of μ_V.

Under part i) of the assumption, one can get a LDP in the scale n^2 for the law of the spectral measure $n^{-1}\sum_{i=1}^n \delta_{\lambda_i}$ under $\mathbb{P}^n_{V,\beta}$ (see [9]), resulting in particular with the almost sure convergence of the spectral measure to a probability measure μ^β_V. If we add part ii) and iii), one can derive the large deviations for the extreme eigenvalues of X_n (see [8], and also [1, Sect. 2.6.2][1]). We give below a slightly more general statement to consider the deviations of the pth largest eigenvalues (note that the pth smallest can be considered similarly).

One can notice that these assumptions hold in a wide generality. In particular, they are satisfied for the law of the GUE ($\beta = 2$, $V(x) = x^2$) and the GOE ($\beta = 1$, $V(x) = x^2/2$) as part ii) is verified by Selberg formula whereas part iii) is well known (see [1, Sect. 2.1.6]). For the case of Gaussian Wishart matrices, we know (see e.g. [1, p. 190]) that the joint law of the eigenvalues can be written as $\mathbb{P}^n_{V_{p,n},\beta}$ with $V_{p,n}(x) = \frac{\beta}{4}x - (\beta[1-\frac{p}{n}+\frac{1}{n}]-\frac{1}{n})\log x$ on $(0,\infty)$. If the ratio $\frac{p}{n}$ converges to α, one can easily show that the law of the largest eigenvalues are exponentially equivalent under $\mathbb{P}^n_{V_{p,n},\beta}$ and under $\mathbb{P}^n_{V,\beta}$, with $V(x) = \frac{\beta}{4}x - \beta(1-\alpha)\log x$ on $(0,\infty)$, for which the assumptions are satisfied.

Theorem 2.10 *Under Assumption 2.9, the law of the p largest eigenvalues ($\lambda^n_1 > \cdots > \lambda^n_p$) of X_n satisfies a LDP in the scale n and with good rate function given by*

$$J^p(x_1,\ldots,x_p) = \begin{cases} \sum_{i=1}^p J_V(x_i) + p\alpha^1_{V,\beta}, & \text{if } x_1 \ge x_2 \ge \cdots \ge x_p, \\ \infty, & \text{otherwise,} \end{cases}$$

with $J_V(x) = V(x) - \beta\int \log|x-y|\, d\mu_V(y)$.

[1] Note that in the published version of [1], part iii) was not mentioned but it appears in the errata sheet available online: http://www.wisdom.weizmann.ac.il/~zeitouni/cormat.pdf.

Remark 2.11 Note that in the case of the GOE and the GUE (see [8]),

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

Let us now go to the perturbed model. An important remark is that, due to the rotational invariance of the law of X_n, one can in fact consider very general orthonormal perturbations. We make the following

Assumption 2.12 $(U_1^n, \ldots, U_r^n)$ 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.

Indeed, under these assumptions, $\widetilde{X_n}$ 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 $\mathbb{P}^n_{V,\beta}$-distributed eigenvalues and O_n Haar distributed on the orthogonal (resp. unitary) group of size n if $\beta = 1$ (resp. $\beta = 2$), independent of $\{D_n\} \cup \{U_1^n, \ldots, U_r^n\}$. 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. In particular they can be obtained by the orthonormalization procedure described in the introduction, with $G = (g_1, \ldots, g_r)$ a vector whose components are i.i.d. Gaussian standard variables (which satisfies in particular Assumption 2.6).

With these considerations in mind, we can state the LDP for the extreme eigenvalues of $\widetilde{X_n}$. We recall that b_V is the rightmost point of the support of μ_V.

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

$$\tilde{J}^k(x_1, \ldots, x_k) = \inf_{p \geq 0} \inf_{\ell_1 \geq \cdots \geq \ell_p > b_V} \{L^0_{\ell_1,\ldots,\ell_p}(x_1, \ldots, x_k) + J^p(\ell_1, \ldots, \ell_p)\},$$

if $x_1 \geq \cdots \geq x_k$, the function being infinite otherwise. Here, $L^0_{\ell_1,\ldots,\ell_p}$ is the rate function defined in Theorem 9.1 for the orthonormalized perturbation model built on $G = (g_1, \ldots, g_r)$ i.i.d. standard Gaussian variables and X_n with limiting spectral measure μ_V and outliers $\ell_1, \ldots, \ell_p$.

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 the i.i.d. perturbation model and the orthonormalized perturbation model will be treated simultaneously).

The cornerstone is a nice representation, already crucially used in many papers on finite rank deformations (see e.g. [3,11]), of the eigenvalues $(\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_m^n)$ as zeroes of a fixed deterministic polynomial in the entries of matrices of size r depending only on the resolvent of X_n and the random vectors $(G_i^n)_{1\le i\le r}$.

Indeed, if V is the $n\times r$ matrix with column vectors $[U_1^n\cdots U_r^n]$ in the orthonormalized perturbation model and$[G_1^n\cdots G_r^n]$ in the i.i.d. perturbation model, Θ the matrix $\mathrm{diag}(\theta_1,\dots,\theta_r)$ and I_n the identity in $n\times n$ matrices, the characteristic polynomial of $\widetilde{X_n}$ reads

$$\begin{aligned}\det(zI_n-\widetilde{X_n}) &= \det(zI_n-X_n-V\Theta V^*)\\ &= \det(zI_n-X_n)\det(I_r-V^*(zI_n-X_n)^{-1}V\Theta)\end{aligned}\tag{3}$$

It means that the eigenvalues of $\widetilde{X_n}$ that are not[2] 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 H_r of $r\times r$ Hermitian matrices given, for $1\le i\le j\le r$, by

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

and

$$C_{ij}^n=\frac{1}{n}\sum_{k=1}^n\overline{g_i(k)}g_j(k),\tag{5}$$

we have (see Sect. 4 for details):

Proposition 3.1 *In both i.i.d and orthonormalized perturbation models, there exists a function $P_{\Theta,r}$ defined on $\mathsf{H}_r\times\mathsf{H}_r$ which is polynomial in the entries of its arguments and depends only on the matrix Θ, 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. perturbation model and the orthonormalized perturbation model. In the i.i.d. perturbation model, $P_{\Theta,r}$ is simpler and does not depend on C. This proposition characterizes 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 study in detail before getting into the deviations of the eigenvalues themselves. Because $K^n(z)$ blows up when z approaches λ_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,∞), $\mathcal{C}(\mathcal{K},\mathsf{H}_r)$ and $\mathcal{C}(\mathcal{K},\mathbb{R})$ be the space of continuous functions on $\mathcal{K}$ taking

[2] We show in Sect. 11.2 that the spectra of X_n and $\widetilde{X_n}$ are disjoint in generic situation.

values respectively in 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* $\mathcal{C}(\mathcal{K}, \mathsf{H}_r) \times \mathsf{H}_r$ *equipped with the uniform topology, satisfies a LDP 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 LDP in the scale n and with rate function given, for a continuous function* $f \in \mathcal{C}(\mathcal{K}, \mathbb{R})$*, by*

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

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 $\alpha \in \mathbb{R}^m$. An important technical issue will come from the fact that the set of functions which vanish exactly at α 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 orthonormalization procedure and prove Proposition 3.1. Sections 5 and 6 will then deal more specifically with the case without outliers. In Sect. 5, we establish the functional LDPs for $(K^n(\cdot), C^n)$ and H^n, whereas Sect. 6 is devoted to the proof of our main results in this case, namely the LDP for the largest eigenvalues of $\widetilde{X}_n$ and the almost sure convergence to the minimisers of the rate function. In Sect. 7, we will see that the rate function can be studied further in the special case when $X_n = 0$. We then turn to the case with outliers in Sects. 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 Sect. 10. To make the core of the paper easier to read, we gather some technical results in Sect. 11.

4 Characterisation of the eigenvalues of $\tilde{X}_n$ as zeroes of a function H^n

The goal of this section is to prove Proposition 3.1. As will be seen further, the proof of this proposition is straightforward in the i.i.d. perturbation model but more involved in the orthonormalized perturbation model and we first detail the orthonormalization procedure.

4.1 The Gram–Schmidt orthonormalisation procedure

We start 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 perturbation 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 H_r is the space of $r \times r$ either symmetric or Hermitian matrices, according to whether G is a real ($\mathbb{K} = \mathbb{R}$) or complex ($\mathbb{K} = \mathbb{C}$) random vector.

Fix $1 \le r \le n$ and consider a linearly independent family $G_1, \ldots, G_r$ of vectors in $\mathbb{K}^n$. 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

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

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

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

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 $W_1, \ldots, W_r$ defined, for $i = 1, \ldots, r$, by

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

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

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

are orthonormal. They are said to be the Gram–Schmidt orthonormalized vectors from $(G_1, \ldots, G_r)$. 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 $i_0 = 1, \ldots, r$, there is a real function P_{i_0}, defined on 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 $\mathbb{K}^r$ (namely the columns of its square root).

Let now G be a random vector satisfying Assumption 2.2 and $(G(k), k \geq 1)$ be i.i.d. copies of G. Let $G_i^n = (G(k)_i)_{1\leq k\leq n}$ 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 $\mathbb{K}^n$ as above. As announced in Sect. 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 $\tilde{X}_n$: proof of Proposition 3.1

As explained in Sect. 3, a crucial observation (see [11, Proposition 5.1]) is that the eigenvalues of $\widetilde{X_n}$ can be characterized 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 first recall (3), that is for $z \notin \{\lambda_1^n, \dots, \lambda_n^n\}$,

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

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

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

We denote by $\mathbf{G}$ the $n \times r$ matrix with column vectors $(G_i^n)_{1\leq i\leq r}$, so that $K^n(z) = \frac{1}{n}\mathbf{G}^*(zI_n - X_n)^{-1}\mathbf{G}$.

In the i.i.d. perturbation model, as $V = \mathbf{G}$, Proposition 3.1 follows immediately with

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

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

In the orthonormalized perturbation model, the Gram–Schmidt procedure makes things a bit more involved.

If we denote by D the $r\times r$ diagonal matrix given by $D = \text{diag}(\|W_1^n\|_2, \dots, \|W_r^n\|_2)$ and $\Sigma = (A^n)^T$, then V is equal to $n^{-1/2}\mathbf{G}\Sigma D^{-1}$ and 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 $z \notin \{\lambda_1^n, \dots, \lambda_n^n\}$ 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

Θ) of the entries of $K^n(z)$, $E^*\Theta^{-1}E$ and F. Furthermore, $E^*\Theta^{-1}E$ is a diagonal matrix whose ith 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. □

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, ∞), and we denote by z^* its upper bound. We equip $\mathcal{C}(\mathcal{K}, \mathsf{H}_r) \times \mathsf{H}_r$ with the uniform topology which is given by the distance d defined, for $(K_1, C_1), (K_2, C_2) \in \mathcal{C}(\mathcal{K}, \mathsf{H}_r) \times \mathsf{H}_r$ 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 $\|M\|_2 = \sqrt{\mathrm{Tr}(M^2)}$ for all $M \in \mathsf{H}_r$.

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

$$\Lambda(H) = \log \mathbb{E}(e^{\mathrm{Tr}(HZ)}). \tag{8}$$

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 $\mathcal{C}(\mathcal{K}, \mathsf{H}_r) \times \mathsf{H}_r$ equipped with the uniform topology, satisfies a LDP in the scale n and with good rate function $\mathbf{I}$ which is infinite if K is not Lipschitz continuous and otherwise defined, for $K \in \mathcal{C}(\mathcal{K}, \mathsf{H}_r)$ and $C \in \mathsf{H}_r$, by*

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

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

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

and the supremum is taken over piecewise constant functions P with values in H_r and X, Y in 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 LDP in the scale n and with rate function given, for a continuous function*

$f \in \mathcal{C}(\mathcal{K}, \mathbb{R})$, by

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

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

Since the map $(K(\cdot), C) \longmapsto (P_{\Theta,r}(K(z), C))_{z\in\mathcal{K}}$ from $\mathcal{C}(\mathcal{K}, \mathsf{H}_r)\times\mathsf{H}_r$ 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 [16, Theorem 4.2.1].

The reminder 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 [30, Theorem 2.2] (see also [8,12]). From that, we will establish a LDP in the topology of pointwise 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, ∞) and the largest eigenvalue λ_1^n tends to b, there exists $0 < \varepsilon < 1$ (depending only on $\mathcal{K}$) such that for n large enough, for any $z \in \mathcal{K}$ and $1 \le i \le n$, $z - \lambda_i^n > \varepsilon$. We fix hereafter such an ε.

For any $L > 0$, we define

$$\mathcal{C}_{\mathcal{K},L} := \left\{ (K, C) \in \mathcal{C}(\mathcal{K}, \mathsf{H}_r) \times \mathsf{H}_r\,;\ \sup_{z\in\mathcal{K}} \|K(z)\|_2 + \|C\|_2 \le L,\, K \text{ is } \frac{L}{2\varepsilon}\text{-Lipschitz} \right\}.$$

We have

Lemma 5.2

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(((K^n(z))_{z\in\mathcal{K}}, C^n) \in \mathcal{C}^c_{\mathcal{K},L}) = -\infty.$$

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

Proof We claim that

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

Indeed, for n large enough,

$$|K^n(z)_{ij} - K^n(z')_{ij}| \le \sqrt{C^n_{ii}C^n_{jj}}\frac{|z-z'|}{\varepsilon^2},$$

whereas since $|C^n_{ij}|^2 \le C^n_{ii}C^n_{jj}$, $\|C^n\|_2 \le r\max_{1\le i\le r} C^n_{ii}$ and $\|K^n(z)\|_2 \le \frac{1}{\varepsilon}r \max_{1\le i\le r} C^n_{ii}$.

Now, by Assumption 2.2, let $\alpha > 0$ be such that $C := \mathbb{E}(e^{\alpha\sum_{i=1}^r |g_i|^2}) < \infty$.

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

$$\le r\mathbb{E}\left(e^{\alpha\sum_k |G^n_1(k)|^2}\right)e^{-n\alpha\frac{\varepsilon L}{2r}} \le rC^n e^{-n\alpha\frac{\varepsilon L}{2r}} \le e^{-n\alpha\frac{\varepsilon L}{4r}}, \tag{10}$$

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

$$\limsup_{L\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(((K^n(z))_{z\in\mathcal{K}}, C^n) \in \mathcal{C}^c_{\mathcal{K},L}) = -\infty.$$

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

5.2.2 LDP 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 $b < z_1 < z_2 < \cdots < z_M$. The law of $((K^n(z_i))_{1\le i\le M}, C^n)$ viewed as an element of H_r^{M+1} satisfies a LDP in the scale n with good rate function $I_M^{z_1,\ldots,z_M}$ defined, for $K_1,\ldots,K_M, C \in \mathsf{H}_r$ by*

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

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

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

Λ being given by (8).

Proof The proof of the proposition is a direct consequence of Theorem 2.2 of [30]. Indeed, let Z_1 be the H_r-valued random variable such that for all $1 \le i, j \le r$,

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

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

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

A slight problem is that $\frac{1}{n}\sum_{i=1}^{n}\delta_{\lambda_i^n}$ do not fulfill Assumption A.1 in [30] in the sense that this assumption requires that for all i, λ_i^n belongs to the support of the limiting measure μ. Nevertheless, it is easy to construct (as was done in the proof of Theorem 3.2 in [29]) a sequence $\bar{\lambda}_i^n$ such that $\frac{1}{n}\sum_{i=1}^{n}\delta_{\bar{\lambda}_i^n}$ fulfills Assumption A.1 in [30] and $\bar{L}_n := \frac{1}{n}\sum_{k=1}^{n} f(\bar{\lambda}_k^n)$ is exponentially equivalent to L_n. Then from Theorem 2.2 of [30], we get that L_n satisfies an LDP in the scale n with good rate function $I_M^{z_1,\ldots,z_M}$. □

5.2.3 LDP 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 with the topology of pointwise 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* $\mathcal{C}(\mathcal{K},\mathsf{H}_r)\times\mathsf{H}_r$ *equipped with the topology of pointwise convergence satisfies a LDP in the scale* n *with good rate function* $\mathbf{J}$ *defined as follows: for* $K\in\mathcal{C}(\mathcal{K},\mathsf{H}_r)$ *and* $C\in\mathsf{H}_r$,

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

Moreover $\mathbf{J}$ *equals the rate function* $\mathbf{I}$ *given in Theorem 5.1(1).*

Proof Let $\mathcal{J}$ be the collection of all finite subsets of $\mathcal{K}$ ordered by inclusion. For $j=\{z_1,\ldots,z_{|j|}\}\in\mathcal{J}$ and f a measurable function from $\mathcal{K}$ to H_r, $p_j(f)=(f(z_1),\ldots,f(z_{|j|}))\in\mathsf{H}_r^{|j|}$. We know from Proposition 5.3 that the law of $(p_j(K^n), C^n)$ satisfies a LDP with good rate function $I_{|j|}^{z_1,\ldots,z_{|j|}}$. Moreover, one can check that the projective limit of the family $\mathsf{H}_r^{|j|}\times\mathsf{H}_r$ is $\mathsf{H}_r^{\mathcal{K}}\times\mathsf{H}_r$ equipped with the topology of pointwise convergence. Therefore, the Dawson–Gärtner theorem [16, Theorem 4.6.1] proves the LDP with rate function $\mathbf{J}$. The identification of $\mathbf{J}$ as $\mathbf{I}$ is straightforward as by a simple change of variables, $\mathbf{J}$ is the supremum of

$$J(\Xi, M, z) := \mathrm{Tr}\left(\sum_{l=1}^{M-1} \Xi(z_l)(K(z_{l+1}) - K(z_l)) + K(z_M)\Xi(z_M) + CY\right)$$
$$- \int \Lambda\left(\sum_{l=1}^{M-1} \Xi(z_l)\left(\frac{1}{z_{l+1}-x} - \frac{1}{z_l - x}\right) + \frac{\Xi(z_M)}{z_M - x} + Y\right) d\mu(x)$$

over the choices of Ξ, M, z. We may assume without loss of generality that $z_M = z^*$. Putting $P(z) = \sum_{l=1}^{M-1} \Xi(z_l)\mathbb{1}_{[z_l, z_{l+1}]}$ and $X = \Xi(z_M)$, we identify $\mathbf{J}$ and $\mathbf{I}$. Thus the proof of the proposition is complete. □

To complete the proof of Theorem 5.1(1), we now need to show that the LDP is also true for the uniform topology. 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 [16, Corollary 4.2.6] and get that the law of $((K^n(z))_{z\in\mathcal{K}}, C^n)$ as an element of $\mathcal{C}(\mathcal{K}, \mathsf{H}_r) \times \mathsf{H}_r$ equipped with the uniform topology satisfies a LDP in the scale n with good rate function $\mathbf{J}$.

5.3 Properties of the rate function

To finish the proof of Theorem 5.1(1), the last thing to check is that $\mathbf{I}(K(\cdot), C)$ is infinite whenever K is not Lipschitz continuous. This is the object of this section [see Lemma 5.5(6)], together with providing further information on the functions (K, C) with finite $\mathbf{I}$ that will be useful in the sequel.

We will consider the operator norm, given, for $H \in \mathsf{H}_r$, 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 recall that Λ was defined in (8).

Lemma 5.5 1. *$H \mapsto \Lambda(H)$ is increasing, $\Lambda(-H) \leq 0$ if $H \geq 0$.*
2. *If we denote by $(C^*)_{ij} = \mathbb{E}[\overline{g_i} g_j]$. Then, for any $H \in \mathsf{H}_r$,*

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

If we assume moreover that G satisfies the first part of Assumption 2.2 (existence of some exponential moments), we have the following properties.
3. *There exists $\gamma > 0$ so that*

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

4. *If* $\mathbf{I}(K(\cdot), C)$ *is finite,* $C \geq 0$ *and* $K(z) \geq 0$*, for any* $z \in \mathcal{K}$*. Moreover, for all* L*, there exists a finite constant* M_L *so that on* $\{\mathbf{I} \leq L\}$*, we have*

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

5. *If* $\mathbf{I}(K(\cdot), C)$ *or* $\mathbf{J}(K(\cdot), C)$ *are finite, then* $z \to K(z)$ *is non-increasing.*
6. *For all* L*, there exists a finite constant* M_L *so that on* $\{\mathbf{I} \leq L\}$*, we have*

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

In particular, K' *exists almost surely and is bounded by* M_L*. If we assume now that* G *satisfies both parts of Assumption* 2.2 (*the law of* G *does not put mass on hyperplanes*)*, we then have the following additional properties.*

7. *For all non-null positive semi-definite* $H \in \mathsf{H}_r$*,*

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

8. *If* $\mathbf{I}(K(\cdot), C)$ *is finite, then* $C > 0$ *and* $K(z) > 0$ *for any* $z \in \mathcal{K}$*. Moreover, for almost any* $z \in \mathcal{K}$ *and for any non-zero vector* e*, there is no interval with non-empty interior on which the function* $\langle e, K'(.)e\rangle$ *vanishes everywhere.*

Proof 1. The first point is just based on the fact that almost surely, $\mathrm{Tr}(HZ) \geq 0$ if $H \geq 0$.

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

4. To prove the fourth point let $(C, K) \in \{\mathbf{I} \leq L\}$. We first show that $C \geq 0$. We take $P, X \equiv 0$ to get

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

Suppose now that there exists some vector $u \in \mathbb{C}^r$ such that $\langle u, Cu\rangle = \alpha < 0$, and define, for any $t > 0$, $Y_t = -t\, uu^*$. Then $\Lambda(Y_t) \leq 0$ by the first point and $\mathrm{Tr}(CY_t) = -\alpha t$ so that for all $t > 0$,

$$-\alpha t \leq \mathrm{Tr}(CY_t) - \Lambda(Y_t) \leq L.$$

Letting t going to infinity gives a contradiction. The same proof holds for $K(z)$ by taking $P(z) = -\mathbb{1}_{z \geq z_0} X$ and $X = -t\, uu^*$ if $\langle u, K(z_0)u\rangle = \alpha < 0$. We finally bound K and C. With γ and B introduced in the third point, we define $Y = \pm\gamma uu^*$ and take $P, X \equiv 0$. We get

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

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

5. We next prove that $z \to K(z)$ is non-increasing when the entropy is finite. Let us prove that for any $z_1, z_2 \in \mathcal{K}$ such that $z_1 < z_2$, $K(z_2) \le K(z_1)$ (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 $u \in \mathbb{C}^r \backslash \{0\}$. For all real number $t \ge 0$, we have, for $P_t(z) := t\,\mathbb{1}_{[z_1,z_2]}(z)uu^*$ and $X = Y = 0$,

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

Note that

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

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

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

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.

6. Take $P = -(z_2 - z_1)^{-1}1_{[z_1,z_2]}uu^*$, $Y = -uu^* \max_{x\in\mathrm{supp}(\mu)} \int (z-x)^{-2}(z_2 - z_1)^{-1}1_{[z_1,z_2]}(z)\,dz$ and $X = 0$ to get, since then $\tilde{\Gamma}(P, Y, X) \le 0$ by the first point,

$$\langle u, -\mathrm{Tr}((K(z_2) - K(z_1))(z_2 - z_1)^{-1})u\rangle \le L + r\varepsilon^{-2}\|C\|_\infty$$

where we used that Y is bounded by ε^{-2}. This provides the expected bound by the fourth point.

7. Consider $\eta > 0$ and a non-vanishing orthogonal projector $p \in \mathsf{H}_r$ such that $H \ge \eta p$. For all $t > 0$, we have

$$0 \le \mathbb{E}[e^{-t\,\mathrm{Tr}(HZ)}] \le \mathbb{E}[e^{-t\eta\,\mathrm{Tr}(pZ)}] = \mathbb{E}[e^{-t\eta\,\mathrm{Tr}(pGG^*)}] = \mathbb{E}[e^{-t\eta G^*pG}].$$

Since, by dominated convergence,

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

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

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

8. We already proved that K is non-increasing and almost surely differentiable, so that $K'\le 0$ almost surely. Moreover, if u is a fixed vector and $\langle u, K'(\cdot)u\rangle$ vanishes on an interval $[z_1,z_2]$ with $z_1<z_2$, taking $P_t=t\,\mathbb{1}_{[z_1,z_2]}(z)uu^*$, and $X=Y=0$, yields

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

which goes to infinity as t goes to infinity by the previous consideration. Thus, this is not possible. As we have already seen that $K(a)\ge 0$ for all $a\in\mathcal{K}$, we see that $K(a')>0$ for $a'<a$ unless there exists e so that $\langle e,(a-a')^{-1}(K(a)-K(a'))e\rangle$ vanishes, which is impossible by the above.

5.4 Study of the minimisers of I

We characterise the minima of **I** as follows:

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

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

Proof **I** vanishes at its minimisers (as a good rate function) and therefore a minimizer (K,C) satisfies for all P,X,Y,

$$\mathrm{Tr}\left(\int K'(z)P(z)\,\mathrm{d}z+K(z^*)X+CY\right)\le\Gamma(P,X,Y). \tag{12}$$

Now, for any fixed (P,X,Y), there exists $\varepsilon_0>0$ such that for any $0<\varepsilon<\varepsilon_0$, for any x in the support of μ we have

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

with α given by Assumption 2.2. Therefore, there exists a constant L such that for any x in the support of μ

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

so that

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

As a consequence, for any minimizer (K, C), we find after replacing (P, X, Y) by $\varepsilon(P, X, Y)$, using (12) and letting ε going to zero, that

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

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 $(K, C)=(K^*, C^*)$. □

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 $\mathbb{R}_{\downarrow}^p(x)=\{(\alpha_1, \ldots, \alpha_p) \in \mathbb{R}^p / \alpha_1 \geq \cdots \geq \alpha_p \geq x\}$.

We also denote by $\omega(g):=\sup_{x \neq y} \frac{|g(x)-g(y)|}{|x-y|} \in[0, \infty]$ the Lipschitz constant of a function g. For any $\varepsilon, \gamma>0$, and $\alpha \in \mathbb{R}_{\downarrow}^p(b+\varepsilon)$, we put

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

Note that in the latter product, the α_i's appear with multiplicity. $S^{\varepsilon}_{\emptyset,\gamma}$ will denote the set of functions as above but with no zeroes on $\mathcal{K}_\varepsilon$. We have the following theorem.

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

$$L(\alpha) := \begin{cases} \lim_{\varepsilon\downarrow 0} \inf_{\cup_{\gamma>0} S^{\varepsilon}_{(\alpha_1,\ldots,\alpha_{m-k}),\gamma}} J_{\mathcal{K}_\varepsilon} & \text{if } \alpha \in \mathbb{R}^m_{\downarrow}(b), \alpha_{m-k+1} = b \text{ and } \alpha_{m-k} > b \\ & \quad \text{for some } k \in \{0, \ldots, m-1\}, \\ \lim_{\varepsilon\downarrow 0} \inf_{\cup_{\gamma>0} S^{\varepsilon}_{\emptyset,\gamma}} J_{\mathcal{K}_\varepsilon} & \text{if } \alpha_1 = \alpha_2 = \cdots \alpha_m = b \\ +\infty & \text{otherwise.} \end{cases}$$

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

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

is increasing, so that its limits as ε 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,

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

with a non-vanishing constant c and a self-adjoint matrix A with eigenvalues $(\theta_1^{-1}, \ldots, \theta_r^{-1})$ and a function K with values in the set of $r \times r$ positive self-adjoint matrices so that $K' \le 0$ 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 (7) of Lemma 5.5. Moreover, if we let P be the orthogonal projection onto the orthocomplement of e, the function $H(z) = \det((1-P)(A-K(z))(1-P)) \det(PAP - PK(z)P)$ 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, ∞)

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

where (K^*, C^*) *are given in Lemma 5.6 and* $P_{\Theta,r}$ *is defined in Proposition 3.1, there exists* $k \in \{0, \ldots, m\}$ *such that H has* $m - k$ *zeroes* $(\lambda_1^*, \ldots, \lambda_{m-k}^*)$ *(counted with multiplicity). The unique point of* $\mathbb{R}^m$ *on which L vanishes is* $(\lambda_1^*, \ldots, \lambda_{m-k}^*, b, \ldots, b)$ *and consequently* $(\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_m^n)$ *converges almost surely to this point as n grows to infinity.*

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

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

so that we recover [10, Theorem 1.3] or [11, 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 (see e.g. [27, Sect. 4.3])

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

which converges to b as n goes to infinity.

Secondly, let us state the following lemma.

Lemma 6.6 *The law of the sequence* $(\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_m^n)$ *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 denote by $\|R_n\|_\infty$ the operator norm of the perturbation matrix R_n. Note that for all k,

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

Since for any fixed k, the non-random sequence λ_k^n converges to b as n tends to infinity, it suffices to prove that

$$\limsup_{L\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\|R_n\|_\infty \geq L) = -\infty. \tag{13}$$

For the orthonormalized perturbation model, since $\|R_n\|_\infty = \max\{\theta_1, -\theta_r\}$, (13) is clear. In the i.i.d. perturbation model, we have, for $\theta := \max_{1\leq i\leq r} |\theta_i|$,

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

It implies, by Tchebychev's inequality, that

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

which allows to conclude by Assumption 2.2. □

As the law of $(\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_m^n)$ is exponentially tight, the proof of Theorem 6.1 reduces to establishing a weak LDP. In virtue of [16, 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 Eq. (18) and Lemma 6.9 below. The fact that L is a good rate function is then implied by exponential tightness [16, Lemma 1.2.18].

6.3 The structure of H^n

From Proposition 3.1, we know that the $\widetilde{\lambda}_i^n$'s are essentially the zeroes of H^n. However, H^n could a priori have other zeroes than these eigenvalues or take arbitrary small values. To control this point, we need to understand better the structure of H^n. Let

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

and

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

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 sequence of random functions (g_n) such that for any $z \in \mathcal{K}_\varepsilon$ and $n \ge n_0(\varepsilon)$,*

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

with

$$s = (-1)^{r-m}, \quad L(\varepsilon) \le g_n \le \frac{1}{L(\varepsilon)} \quad \textit{and} \quad \omega(g_n) \le \frac{1}{L(\varepsilon)}. \tag{14}$$

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

$$\limsup_{\gamma \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\left(\prod_{i=1}^m (z - \widetilde{\lambda}_i^n)^{-1} H^n(z) \right)_{z \in \mathcal{K}_\varepsilon} \in (C_{0,\gamma}^\varepsilon)^c \right) = -\infty. \tag{15}$$

and

$$\limsup_{\gamma \downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(H^n \in (C_\gamma^\varepsilon)^c) = -\infty. \tag{16}$$

Proof Let us define the random sequence

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

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

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

Now, for $\varepsilon > 0$ fixed, we shall bound g_n and its Lipschitz constant on $\mathcal{K}_\varepsilon$.

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

$$0 \le \sum_{i=m+1}^n (\lambda_{i-m}^n - \lambda_i^n) = \sum_{i=1}^m \lambda_i^n - \sum_{i=n-m}^n \lambda_i^n \le 2m \frac{2}{\varepsilon}. \tag{17}$$

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

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

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

$$\widetilde{\lambda}_i^n \le \lambda_{i-m}^n, \text{ so that } \lambda_i^n - \widetilde{\lambda}_i^n \ge -(\lambda_{i-m}^n - \lambda_i^n).$$

For $0 \le x \le 1 - \frac{\varepsilon^2}{4}$, $\log(1-x) \ge -\frac{4}{\varepsilon^2}x$, so that we finally get by (17),

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

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

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

The proof of the uniform equicontinuity of g_n on $\mathcal{K}_\varepsilon$ is left to the reader as the arguments are very similar since $z \to (z - \lambda_i^n)^{-1}$ is uniformly continuous on $\mathcal{K}_\varepsilon$ for n large enough.

To prove (16) and (15), it is therefore sufficient to prove that, with probability greater than $1 - e^{-cn}$ for some $c > 0$, we have that c_n and c_n^{-1} are bounded which is a direct consequence of Lemma 11.1, and that, $\tilde{\lambda}_n^1 \le \varepsilon^{-1}$ for small ε, which is proved in Lemma 6.6. □

The main application of the previous Lemma will be the following continuity properties of the zeroes of functions in C_γ^ε.

Lemma 6.8 *Let $\varepsilon > 0$ be fixed, $\gamma > 0$ small enough, and $k \in \mathbb{N}$ be fixed. Let $\alpha_1^0 \ge \cdots \ge \alpha_k^0 \in \mathcal{K}_\varepsilon$ and $f_0(z) = h_0(z)\prod_{i=1}^{k}(z - \alpha_i^0) \in C_{k,\gamma}^\varepsilon$, be given. Then, for all $\delta > 0$ there exists $\delta' > 0$ so that*

$$\left\{ f \in C_\gamma^\varepsilon : \sup_{x \in \mathcal{K}_\varepsilon} |f(x) - f_0(x)| < \delta' \right\}$$
$$\subset \left\{ z \mapsto h(z)\prod_{i=1}^{m}(z - \alpha_i) : h \in C_{0,\gamma}^\varepsilon, \max_{1 \le i \le k} |\alpha_i - \alpha_i^0| \le \delta, \max_{i > k} \alpha_i \le b + 2\varepsilon \right\}$$

Proof This amounts to show that if $f^n \in C_\gamma^\varepsilon$ is a sequence converging (for the uniform topology on $\mathcal{K}_\varepsilon$) to $f \in C_{k,\gamma}^\varepsilon$, $m - k$ zeroes of the functions f_n will be below $b + 2\varepsilon$ and the others will converge to the zeroes of f. Indeed, if we take a sequence $f^n \in C_\gamma^\varepsilon$, we can always denote it $f^n(z) = h^n(z)\prod_{i=1}^{m}(z - \alpha_i^n)$ (with possibly some $\alpha_i^n \in (b + \varepsilon/4, b + 3\varepsilon/4)$ if $f^n \in C_{k,\gamma}^\varepsilon$ with $k < m$), as this amounts at the worst to change h and take γ smaller. Then, the crucial point is that h^n is tight by Arzela–Ascoli theorem so that we can consider a converging subsequence. As the α_i^n belong to $[b, 1/\varepsilon]$, we can also consider converging subsequences. Thus, f^n converges along subsequences to a function $\tilde{f}$ with $\tilde{f}(z) = h(z)\prod_{i=1}^{m}(z - \alpha_i)$ on $\mathcal{K}_\varepsilon$ with $\alpha_i \in [b, 1/\varepsilon]$. But then we must have $f = \tilde{f}$ which allows in particular to identify k limit points with the zeroes of f, the others being below $b + 2\varepsilon$.

6.4 Core of the proof

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

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\bigcap_{1\le i\le m} \{|\widetilde{\lambda}_i^n - \alpha_i| \le \delta\} \right) = \liminf_{\delta\downarrow 0} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\bigcap_{1\le i\le m} \{|\widetilde{\lambda}_i^n - \alpha_i| < \delta\} \right) = -\infty. \tag{18}$$

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

Lemma 6.9 *Let $\alpha \in \mathbb{R}_\downarrow^m$ and k between 0 and m such that $\alpha_{m-k+1} = \ldots = \alpha_m = b$ and $\alpha_{m-k} > b$ if $k < m$. We have*

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

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

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

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

$$B_{\alpha,\gamma,\delta}^\varepsilon := \Bigg\{ f \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R}) : \exists g \in \mathcal{C}(\mathcal{K}_\varepsilon, \mathbb{R}) \text{ with } \frac{1}{\gamma} \le g \le \gamma, \omega(g) \le \frac{1}{\gamma}$$
$$\text{and } f(z) = s.g(z) \prod_{i=1}^{m-k} (z - \beta_i) \text{ with } \forall i \le m-k, \beta_i \in [\alpha_i - \delta, \alpha_i + \delta] \Bigg\} \tag{19}$$

Moreover, writing $H^n(z) = h^n(z) \prod_{i=1}^m (z - \tilde{\lambda}_i^n)$ by Lemma 6.7, clearly H^n belongs to $B_{\alpha,\gamma,\delta}^\varepsilon$ as soon as for some $\varepsilon' < \varepsilon$ and $\gamma' \cdot (\varepsilon')^m > \gamma$, $h^n \in C_{0,\gamma'}^{\varepsilon'}$ and $\bigcap_{1\le i\le m-k} \{|\widetilde{\lambda}_i^n -$

$\alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\widetilde{\lambda}_i^n \leq b + \varepsilon - \varepsilon'\}$ holds. As a consequence, we can write

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

Then, by [16, Lemma 1.2.15],

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\bigcap_{1 \leq i \leq m-k} \{|\widetilde{\lambda}_i^n - \alpha_i| \leq \delta\} \bigcap_{m-k+1 \leq i \leq m} \{\widetilde{\lambda}_i^n \leq b + \varepsilon - \varepsilon'\} \right) \\ \leq \max\left\{ \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(H^n \in B^{\varepsilon}_{\alpha,\gamma,\delta}); \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(h^n \in (C^{\varepsilon'}_{0,\gamma'})^c) \right\}, \tag{20}$$

Moreover, $B^{\varepsilon}_{\alpha,\gamma,\delta}$ 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–Arzela'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 $B^{\varepsilon}_{\alpha,\gamma,\delta}$.

Since $J_{\mathcal{K}_\varepsilon}$ is a good rate function, $(B^{\varepsilon}_{\alpha,\gamma,\delta})_{\delta>0}$ is a nested family and $\cap_{\delta>0} B^{\varepsilon}_{\alpha,\gamma,\delta} = S^{\varepsilon}_{(\alpha_1,\ldots,\alpha_{m-k}),\gamma}$, Theorem 5.1 gives with [16, Lemma 4.1.6] that

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(H^n \in B^{\varepsilon}_{\alpha,\gamma,\delta}) \leq - \inf_{S^{\varepsilon}_{(\alpha_1,\ldots,\alpha_{m-k}),\gamma}} J_{\mathcal{K}_\varepsilon}. \tag{21}$$

Taking $\gamma' = \gamma'_0$ small enough, (15) and (20) give for $\gamma/(\varepsilon')^m < \gamma'_0$,

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

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 ε decreases to 0.

We turn to the lower bound, which is a bit more delicate. Let us again consider δ and ε small enough so that $\cap_{i=1}^{m-k} [\alpha_i - \delta, \alpha_i + \delta] \subset \mathcal{K}_\varepsilon$. As $J_{\mathcal{K}_\varepsilon}$ is a good rate function and $S^{\varepsilon}_{\alpha,\gamma}$ is closed, for all $\gamma > 0$, the infimum $\inf_{S^{\varepsilon}_{(\alpha_1,\ldots,\alpha_{m-k}),\gamma}} J_{\mathcal{K}_\varepsilon}$ is achieved, say at $f^{k,\varepsilon}_\gamma$. To complete the proof, we need the following lemma, based on the structure of H_n and whose proof is a direct application of Lemma 6.8.

Lemma 6.10 *Let ε, γ be fixed and small enough. There exists δ_0 such that for any $\delta \le \delta_0$, there exists $\delta' > 0$ such that for any n,*

$$\left\{H^n \in C^\varepsilon_\gamma\right\} \cap \left\{\sup_{x \in \mathcal{K}_\varepsilon} |H^n(x) - f^{k,\varepsilon}_\gamma(x)| < \delta'\right\} \subset \bigcap_{1 \le i \le m-k} \{|\widetilde{\lambda}^n_i - \alpha_i| \le \delta\} \bigcap_{m-k+1 \le i \le m} \{\widetilde{\lambda}^n_i \le b + 2\varepsilon\}.$$

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

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

Let $\eta > 0$ be fixed. As

$$\inf_{\cup_{\gamma>0} S^\varepsilon_{(\alpha_1,\ldots,\alpha_{m-k}),\gamma}} J_{\mathcal{K}_\varepsilon} = \inf_{\gamma>0} \inf_{S^\varepsilon_{(\alpha_1,\ldots,\alpha_{m-k}),\gamma}} J_{\mathcal{K}_\varepsilon} = \inf_{\gamma>0} J_{\mathcal{K}_\varepsilon}(f^{k,\varepsilon}_\gamma),$$

we can choose ε, γ small enough so that $J_{\mathcal{K}_\varepsilon}(f^{k,\varepsilon}_\gamma) \le J + \eta$. By (16), there exists $L(\gamma, \varepsilon)$ going to infinity as γ, ε go to zero so that for n large enough,

$$\mathbb{P}(H^n \in (C^\varepsilon_\gamma)^c) \le e^{-nL(\varepsilon,\gamma)}.$$

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

Lemma 6.10 implies, that for $\delta \le \delta_0$, for δ' small enough, $\eta > 0$, for n large enough,

$$\mathbb{P}\left(\bigcap_{1 \le i \le m-k} \{|\widetilde{\lambda}^n_i - \alpha_i| \le \delta\} \bigcap_{m-k+1 \le i \le m} \{\widetilde{\lambda}^n_i \le b + 2\varepsilon\}\right) \ge \mathbb{P}(\sup_{z \in \mathcal{K}_\varepsilon} |H^n(z) - f^{k,\varepsilon}_\gamma(z)| < \delta') - \mathbb{P}(H^n \in (C^\varepsilon_\gamma)^c) \ge \frac{1}{2} e^{-n(J+2\eta)}$$

the last inequality following from Theorem 5.1.(2). As η can be chosen as small as we want, we conclude by taking first n going to infinity, and then $\delta, \varepsilon, \eta$ to zero.

6.5 Identification of the minimizers

We prove Theorem 6.4, which is straightforward. Since L is a good rate function, it vanishes at its minimizers $(\lambda^*_1, \ldots, \lambda^*_m) \in \mathbb{R}^m_\downarrow(b)$. Putting $\lambda^*_0 = b + 1$, we know that there exists $0 \le k \le m$ 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 $(\lambda^*_1, \ldots, \lambda^*_{m-k})$ such that $J_{\mathcal{K}_{\frac{1}{n}}}(f_n) \le \frac{1}{n}$. From the definition of

$J_{\mathcal{K}}$ and the fourth and sixth 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 $(\lambda_1^*,\ldots,\lambda_{m-k}^*)$ 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. perturbation model when $X_n=0$. More precisely, we consider $G=(g_1,\ldots,g_r)$ satisfying Assumption 2.2, $n\times r$ matrices G_n whose rows are i.i.d. copies of G, a diagonal matrix $\Theta=\mathrm{diag}(\theta_1,\ldots,\theta_r)$ and we study the large deviations of Wishart matrices $W_n=\frac{1}{n}G_n\Theta G_n^*$. This matrix has zero as an eigenvalue with multiplicity at least $n-r$ and we refer in the whole section to the r eigenvalues of W_n that can be non-zero as "the eigenvalues of W_n". The large deviations for the largest and smallest such eigenvalues were already studied in [23] in the case when $\Theta=(1,\ldots,1)$ and the g_i's are i.i.d.

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

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

with

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

Note that the previous proposition could also have been deduced directly from Cramér's theorem and the contraction principle.

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

Corollary 7.2 *Assume that $G=(g_1,\ldots,g_r)$ is a Gaussian vector with positive definite covariance matrix R. Let $\Theta=\mathrm{diag}(\theta_1,\ldots,\theta_r)$ be a diagonal matrix with positive entries. We denote by $0<r_1(\Theta)\le r_2(\Theta)\le\cdots\le r_r(\Theta)$ 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 LDP in the scale n with rate function which is infinite unless $\alpha_1\ge\alpha_2\ge\cdots\ge\alpha_r>0$ and otherwise given by*

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

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

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

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 [23]. 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 $\Theta=\operatorname{diag}(1,\ldots,1)$, 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 [23, 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 LDP in the scale n with a rate function $L_{max}(\alpha)$ which satisfies, for any $\alpha\in\mathbb{R}$,*

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

From there, one can easily improve the upper bound on the probability of deviations of the largest eigenvalue of [23, Theorem 2.1]:

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

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

Note that when $\alpha\geq\theta_1$, $I_{r,\Theta}(\alpha)=\inf_{[\alpha,\infty)}L_{max}$ and in particular $I_{r,\Theta}$ is not necessarily lower semicontinuous. We refer to [23] 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 $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_r > 0$, $L(\alpha)$ is the infimum of J over the non-negative Hermitian matrices C such that $\Theta^{\frac{1}{2}} C \Theta^{\frac{1}{2}}$ has spectrum $(\alpha_1, \ldots, \alpha_r)$. □

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

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

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 (ONB) $(e_i)_{1\leq i\leq r}$. This gives

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

□

Proof of Corollary 7.4 We only need to take, in the definition of $J(C)$, $Y = tvv^*$ if C has eigenvector v for its largest eigenvalue to get a lower bound on $J(C)$, and thus on L. □

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 [23], that is by noticing that

$$\mathbb{P}(\lambda_{\max} \geq \alpha) = \mathbb{P}\left(\sup_{\|x\|_2=1} \langle x, W_n x\rangle \geq \alpha\right) \geq \sup_{\|x\|_2=1} \mathbb{P}(\langle x, W_n x\rangle \geq \alpha)$$

and that for fixed x, $\langle x, W_n x\rangle = n^{-1}\sum_{j=1}^n (\langle x, \Theta^{\frac{1}{2}} G_n^j\rangle)^2$ is a sum of i.i.d. random variables so that Cramer's theorem apply. By arguments as in [23], one can also check that $I_{r,\Theta}$ is increasing on $[\theta_1, \infty)$, which concludes the proof. □

8 Large deviations for H^n in the presence of outliers

We now go to the proof of the LDP in the presence of outliers, that will be stated in details in Theorem 9.1. The proof follows the same lines as in the case without outliers and starts therefore with the study of the deviations of H^n.

Let $\mathcal{K}^o := \bigcup_{i=1}^{p_0}[a_i, b_i]$ a compact subset of $(b,\infty)\setminus\{\ell_1^+,\ldots,\ell_{p^+}^+\}$. We equip again $\mathcal{C}(\mathcal{K}^o,\mathsf{H}_r)\times\mathsf{H}_r$ with the uniform topology. Hereafter, we denote by $\ell_i=\ell_i^+$ for $1\le i\le p^+$ and $\ell_i=\ell^-_{p^++p^--i+1}$ for $p^++1\le i\le p^++p^-$.
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 $\big((K^n(z))_{z\in\mathcal{K}^o}, C^n\big)$, viewed as an element of the space $\mathcal{C}(\mathcal{K}^o,\mathsf{H}_r)\times\mathsf{H}_r$ endowed with the uniform topology, satisfies a LDP in the scale n with rate function $\mathbf{I}^o$. For $K\in\mathcal{C}(\mathcal{K}^o,\mathsf{H}_r)$ and $C\in\mathsf{H}_r$, $\mathbf{I}^o(K(\cdot),C)$ is infinite if $z\to K(z)$ is not uniformly Lipschitz on $\mathcal{K}^o$. Otherwise, it is given by*

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

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

$$K_0(\cdot)+\sum_{i=1}^{p^++p^-}\frac{1}{\cdot-\ell_i}L_i=K(\cdot)\quad\text{and}\quad C_0+\sum_{i=1}^{p^++p^-}L_i=C \tag{22}$$

and with

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

the supremum being taken over piecewise constant P with values in H_r, $X=(X_1,\ldots,X_{p_0})\in(\mathsf{H}_r)^{p_0}$ and $Y\in\mathsf{H}_r$.
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 LDP in the scale n with rate function given, for a function $f\in\mathcal{C}(\mathcal{K}^o,\mathbb{R})$, by*

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

Note that the function Γ^* is well defined because if K is uniformly Lipschitz on $\mathcal{K}^o$, then so is any K_0 satisfying the compatibility condition (22), so that K_0' almost surely exists.

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* H_r*-valued random variable such that for* $1 \leq i \leq j \leq r$, $(Z_1)_{ij} = \overline{g_i(1)}g_j(1)$. *Under Assumption* 2.6, $\frac{Z_1}{n}$ *also satisfies a LDP in the scale* n *with a good rate function* $I^{(Z)}(M) = \inf\{I(v) : \overline{v_i}v_j = M_{ij}, 1 \leq i, j \leq r\}$.

The proof of Theorem 8.1 follows the same lines as that of Theorem 5.1, except that the LDP for finite dimensional marginals for our process is described by Theorem 3.2 of [29] instead of Theorem 2.2 of [30]. 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 $(I_M^{z_1,\ldots,z_M})^o$. To define this new rate function, we first extend in an obvious way the definition of $I_M^{z_1,\ldots,z_M}$ for z_i's in $\mathcal{K}^o$. Then one can define, for $K_1, \ldots, K_M, C \in \mathsf{H}_r$, and $z_1, \ldots, z_M \in \mathcal{K}^o$,

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

where the infimum is taken over families

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

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 pointwise convergence with good rate function

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

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 on the event $\{\mathbf{J}^o(K(\cdot), C) < \infty\}$, K is Lipschitz continuous on $\mathcal{K}^o$. The proof is similar to that of Lemma 5.5 as, once the L_i are given, K is Lipschitz on $\mathcal{K}^o$ as soon as K_0 is.

We now suppose that K is Lipschitz continuous on $\mathcal{K}^o$ and we want to identify the two rate functions. By mimicking[3] the proof at the end of Sect. 5.2, one can easily show that for K is Lipschitz continuous on $\mathcal{K}^o$,

$$\sup_M \sup_{z_1,\ldots,z_M} I_M^{z_1,\ldots,z_M}(K(z_1),\ldots,K(z_M)) = \Gamma^*(K, C). \tag{23}$$

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 $z_1, \ldots, z_{M(p)}$ such that

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

Now for each $z_1, \ldots, z_{M(p)}$ we choose an admissible decomposition [according to (22)] of K so that

$$\mathbf{J}^o(K, C) \geq I_{M(p)}^{z_1,\ldots,z_{M(p)}} \left(K_0^{M(p)}(z_1), \ldots, K_0^{M(p)}(z_{M(p)}), C\right) + \sum_{i=1}^{p^+ + p^-} I^{(Z)}(L_i^{M(p)}) - \frac{2}{p}.$$

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

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

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

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

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

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 LDP 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 ε, ρ small enough, we define the compact sets

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

and $\mathcal{K}^o_\varepsilon := [b + \varepsilon, \varepsilon^{-1}]$. 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 $\alpha \in \mathbb{R}^p_\downarrow(b + \varepsilon)$, we put

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

We also denote by

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

and

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

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 $(\widetilde{\lambda}^n_1, \dots, \widetilde{\lambda}^n_{m+p^+})$ of $\widetilde{X}_n$ satisfies a LDP in $\mathbb{R}^{m+p^+}$ with good rate* *function L^o. For $\alpha = (\alpha_1, \ldots, \alpha_{m+p^+}) \in \mathbb{R}^{m+p^+}$, we take $\alpha_{m+p^++1} = b$ and L^o is defined as follows*:

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

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 $(\widetilde{\lambda}^n_i)_{1 \le i \le n}$ of the sum of two Hermitian matrices with eigenvalues $(\lambda^n_i)_{1 \le i \le n}$ and $\theta := (\theta_1, \ldots, \theta_r, 0, \ldots, 0)$ satisfy Horn's inequalities and are characterised by the fact that they satisfy such inequalities (see [33] for details). Assume that $\widetilde{\lambda} := (\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{m+p^+})$ is at distance of the bulk and of the outliers which is bounded below. We claim that the rate function $L^o(\widetilde{\lambda})$ is infinite if $(\widetilde{\lambda}, \ell, \theta)$ do not satisfy the Horn inequalities. Indeed, if $L^o(\widetilde{\lambda})$ is finite, $(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{m+p^+})$ are zeroes of a function f which can be written

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

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

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

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

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

with $u_i = g_i$ in the i.i.d. perturbation model and u_i the Gram–Schmidt orthonormalization of the vectors g_i in the orthonormalized perturbation model. Hence, the function $f_n(z) = P_{\Theta,r}(K^n(z), C^n)$ vanishes at the eigenvalues $(\widetilde{\lambda}^n)$ of the sum of the two Hermitian matrices $\operatorname{diag}(\lambda^n)$ 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 λ^n by Lemma 11.2). Therefore, $(\lambda^n, \widetilde{\lambda}^n, \theta)$ satisfy Horn's inequalities by [33]. Since the $(\widetilde{\lambda}^n)$ are bounded, they are relatively compact and we see that the limit points $(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{m+p^+})$ of $(\widetilde{\lambda}^n_1, \ldots, \widetilde{\lambda}^n_{m+p^+})$ 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 $(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_{m+p^+})$ has finite L^o-entropy, and is away from the bulk and the outliers, $(\widetilde{\lambda}, \ell, \theta)$ satisfies Horn's inequalities. It would be interesting to have a direct proof of this fact.

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 make sure that the eigenvalues are not exactly equal to the outliers to use our strategy. We show the following

Lemma 9.3 *Assume that the eigenvalues $\lambda_1^n, \ldots, \lambda_n^n$ of X_n are pairwise distinct and that Assumptions 2.1 and 2.5 hold, then X_n and $\widetilde{X}_n$ have no eigenvalue in common for almost all G.*

The proof of this lemma is postponed to Appendix 11.2. We shall therefore give the proof of the Theorem when the eigenvalues of X_n are distinct. This is however sufficient to get the LDP without this hypothesis due to the following Lemma.

Lemma 9.4 *Let X_n satisfy Assumptions 2.1 and 2.5. Then, there exists a sequence $\bar{X}_n$ of matrices with pairwise distinct eigenvalues satisfying Assumptions 2.1 and 2.5 such that, if we define $\widetilde{\bar{X}}_n$ be the perturbation of $\bar{X}_n$ by the i.i.d. or the orthonormalized vectors constructed on the law $\mu_n = \mu * \gamma_n$ of $G + \varepsilon(n)A$ with A r independent standard normal variables and $\varepsilon(n)$ going to zero with n fast enough, then, with $(\widetilde{\bar{\lambda}}_i^n)_{i\le m}$ the extreme eigenvalues of $\widetilde{\bar{X}}_n$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(\max_{1\le i\le m} |\widetilde{\lambda}_i^n - \widetilde{\bar{\lambda}}_i^n| \ge \frac{1}{n}\right) = -\infty.$$

Proof We take $\bar{X}_n$ to be the matrix with the same eigenvectors as X_n and the same eigenvalues except for those which are sticked together which we separate by an arbitrary small weight $w_n \le 1/n$, 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. It is straightforward to verify Assumptions 2.1 and 2.5 for $\bar{X}_n$. Now, if we add the same perturbation to X_n and $\bar{X}_n$, respectively, their eigenvalues will differ at most by w_n almost surely. Then adding a Gaussian vector of variance $\varepsilon(n)^2$ to G will not change the eigenvalues by more than $\sqrt{\varepsilon(n)}$ with probability greater than $1 - e^{-\varepsilon(n)^{-1}n}$ as the empirical covariance matrix of this additional term is bounded by $C\sqrt{\varepsilon(n)}$ with such a probability. We conclude by choosing $\varepsilon(n)$ such that $\sqrt{\varepsilon(n)} < 1/n$. □

Lemma 9.4 means in particular the random variables $(\widetilde{\bar{\lambda}}_i^n)_{i\le m}$ and $(\widetilde{\lambda}_i^n)_{i\le m}$ are exponentially equivalent and [16, Theorem 4.2.13] asserts that a LDP for the extreme eigenvalues $(\widetilde{\bar{\lambda}}_i^n)_{i\le m}$ of $\widetilde{\bar{X}}_n$ entails the LDP for the law of $(\widetilde{\lambda}_i^n)_{i\le m}$ with the same rate function. Therefore, the proof of Theorem 2.5 can be done for the eigenvalues of $\widetilde{\bar{X}}_n$, the main advantage being that, from Lemma 9.3 above, we get that $\bar{X}_n$ and $\widetilde{\bar{X}}_n$ have almost surely no eigenvalue in common and we can proceed as in the case without outliers.

From now on, we assume that X_n satisfies Assumptions 2.1 and 2.5 and has pairwise distinct eigenvalues and that G satisfies Assumptions 2.2 and 2.6 and that its law is absolutely continuous with respect to Lebesgue measure.

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

Lemma 9.5 *Let ε, ρ 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 $z \in \mathcal{K}^o_{\varepsilon,\rho}$,*

$$H^n(z) = \begin{cases} s \prod_{i=1}^r \|q_i^n W_i^n\|_2^2 \, g_n(z) R(z) \prod_{i=1}^m (z - \widetilde{\lambda}_i^n) & \text{in the orth. perturb. model,} \\ s \; g_n(z) R(z) \prod_{i=1}^m (z - \widetilde{\lambda}_i^n) & \text{in the i.i.d. perturb. model,} \end{cases} \tag{24}$$

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,

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

Proof In this case,

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

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 $n \geq n_0(\varepsilon, \rho)$, $\prod_{i=1}^{p^+}(1 - \frac{\lambda_i^n - \ell_i^+}{\lambda_i^n - z}) \geq \frac{1}{2^{p^+}}$, 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 $n \geq n_0(\varepsilon, \rho)$,

$$g_n(z) \leq L(\varepsilon) := \left(\frac{3}{2}\right)^{p^+} \left(\frac{2}{\varepsilon}\right)^m e^{\frac{8m}{\varepsilon^2}}. \tag{25}$$

The uniform equicontinuity is also shown very similarly. □

As in the sticking case, we have the analogue of Lemma 6.9, 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 $(\widetilde{\lambda}_1^n \geq \cdots \geq \widetilde{\lambda}_n^n)$ such that for all $i \leq m + p^+ - k$,

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

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

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

Because of Lemma 9.5, H_n belong to the set of functions $f(z) = h(z)\prod_{i=1}^m (z - \alpha_i)R(z)$ with a bounded positive constant h on $\mathcal{K}^o_{\varepsilon,\rho}$ with values in $[\gamma, \gamma^{-1}]$ with overwhelming probability. But on this set also the zeroes α_i are continuous function of the functions f and therefore we can proceed exactly as in the case without outliers.

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

$$\lim_{\varepsilon\downarrow 0}\lim_{\rho\downarrow 0}\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left((\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_n^n) \in G_k(\alpha, \delta, \varepsilon, \rho)\right)$$
$$= \lim_{\varepsilon\downarrow 0}\lim_{\rho\downarrow 0}\liminf_{\delta\downarrow 0}\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}((\widetilde{\lambda}_1^n, \ldots, \widetilde{\lambda}_n^n) \in G_k(\alpha, \delta, \varepsilon, \rho)) = -L^o(\alpha),$$

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

The proof is similar to the case without outliers.

10 Application to X_n random, following some classical matrix distribution

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

10.1 Proof of Theorem 2.10

Theorem 2.10 is a slight extension of [1, Theorem 2.6.6] and the proof will therefore follow the same lines. We introduce the notations $\phi(\mu, x) = -V(x) + \beta\int\log|x - y|\,d\mu(y)$ (for x greater or equal the right edge of the support of μ) and $\hat{\mu}_n = \frac{1}{n-p}\sum_{i=p+1}^n \delta_{\lambda_i^n}$. Then

$$\mathbb{P}^n_{V,\beta}(\mathrm{d}\lambda_1, \ldots, \mathrm{d}\lambda_n) = \frac{Z^{n-p}_{nV/(n-p),\beta}}{Z^n_{V,\beta}} e^{n\sum_{i=1}^p \phi(\hat{\mu}^n,\lambda_i)+\beta\sum_{1\leq i<j\leq p}\log|\lambda_i-\lambda_j|}$$
$$\times\, \mathrm{d}\mathbb{P}^{n-p}_{nV/(n-p),\beta}(\lambda_{p+1}, \ldots, \lambda_n)\,\mathrm{d}\lambda_1\cdots\mathrm{d}\lambda_p.$$

By [1, Lemma 2.6.7], if parts $i)$ and $ii)$ of Assumption 2.9 hold, the law $\mathbb{P}^n_{V,\beta}$ is exponentially tight so that it is enough to estimate the probability of a small ball around $\mathbf{x} = (x_1 \geq x_2 \geq \cdots \geq x_p)$ (with $x_p \geq b_V$), namely events of the form $B(\mathbf{x}, \delta) := \{\max_{1\leq i\leq p} |\lambda_i - x_i| \leq \delta, \max |\lambda_i| \leq M\}$.

As in [8], a crucial observation is the fact that $\hat{\mu}_n$ converges to μ_V much faster than exponentially under $\mathbb{P}^{n-p}_{nV/(n-p),\beta}$ (its LDP is indeed in the scale n^2). We can therefore 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 first sketch the proof of the upper bound. Note that there exists a constant Φ_M such that on $B(\mathbf{x}, \delta)$, ϕ is bounded above by Φ_M so that

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

with $B_\varepsilon(\mu_V)$ a small ball with radius ε around μ_V for a distance compatible with the weak topology. As $n^{-2} \log \mathbb{P}^{n-p}_{nV/(n-p),\beta}(\hat{\mu}^n \in B_\varepsilon(\mu_V)^c)$ 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 $(\mu, x) \to \phi(\mu, x)$ is upper continuous, we obtain the upper bound by first letting n go to infinity, then letting δ decrease to zero and finally letting ε go to zero. Notice again that in the proof of this upper bound, we use part $i)$ of Assumption 2.9 to get the LDP for $\hat{\mu}_n$ and $ii)$ to control the ratio of the partition functions.

The lower bound is similar to the proof in [1, p. 84], which corresponds to $p = 1$. We proceed by induction on p and we can therefore assume that p is the smallest integer such that $x_p > b_V$. There exists x_i^δ, $1 \leq i \leq p$, whose small neighbourhood are included in the δ neighbourhood of x_i, $1 \leq i \leq p$, and which are distinct, so that for ε small enough

$$\mathbb{P}^n_{V,\beta}\left(\max_{1\leq i\leq p} |\lambda_i - x_i| < \delta\right) \geq \mathbb{P}^n_{V,\beta}\left(\max_{1\leq i\leq p} |\lambda_i - x_i^\delta| < \varepsilon, \lambda_i < x_p - \delta - \varepsilon, \forall i > p\right) \\ \geq \frac{Z^{n-p}_{nV/(n-p),\beta}}{Z^n_{V,\beta}} \exp\left((n-p) \inf_{\substack{|y_i - x_i^\delta| < \varepsilon \\ \mu\in B_{[-M,x_p-\delta-\varepsilon]}(\mu_V,\varepsilon)}} \phi(y_i, \mu)\right) \\ \times \mathbb{P}^{n-p}_{nV/(n-p)}(\hat{\mu}^n \in B_{[-M,x_p-\delta-\varepsilon]}(\mu_V, \varepsilon)),$$

with $B_{[-M,x_p-\delta-\varepsilon]}(\mu_V, \varepsilon))$ the set of probability measures in $B_\varepsilon(\mu_V)$ with support in $[-M, x_p - \delta - \varepsilon]$. When the x_i's are distinct and away from b_V, their logarithmic interaction is negligible; moreover, part $iii)$ of Assumption 2.9 allows to claim that the last term in the lower bound above converges to one. We therefore get

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n_{V,\beta}\left(\max_{1\leq i\leq p} |\lambda_i - x_i| < \delta\right) \geq -\sum_{i=1}^p J_V(x_i^\delta) - \alpha^p_{V,\beta}$$

Now, J_V is continuous away from the support of μ_V so that we can conclude by letting δ going to zero.

Then to get the correct expression of the rate function, we just have to check that $\alpha^p_{V,\beta} = p\alpha^1_{V,\beta}$, which is easy and left to the reader. □

10.2 Proof of Theorem 2.13

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

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.

- More precisely, we observe that, by Theorem 2.10, for all $p \in \mathbb{N}^*$, the probability that λ^n_p is greater than $b_V + \delta$ is less than $e^{-np\varepsilon(\delta)}$ for $\varepsilon(\delta) = \inf_{(b_V+\delta,+\infty)} J_V$. The only point to check is that $\inf_{(b_V+\delta,+\infty)} J_V > 0$, which is a consequence of part $iii)$ of Assumption 2.9.
- The deviations of the eigenvalues of X_n are controlled by Theorem 2.10 : there exists $\varepsilon(\eta, \ell) > 0$ so that for n large enough, for $\varepsilon \le \varepsilon(\eta, \ell)$

$$-J^p(\ell_1, \dots, \ell_p) - \eta \le \frac{1}{n} \log \mathbb{P}\left(\max_{1\le i\le p} |\lambda^n_i - \ell_i| \le \varepsilon, \lambda^n_{p+1} \le b_V + \varepsilon \right) \le -J^p(\ell_1, \dots, \ell_p) + \eta. \quad (26)$$

For all $(\ell_1, \dots, \ell_p)$ and $\eta > 0$, we define the set

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

- Now, knowing the deviations of the eigenvalues of X_n, one can treat them as outliers and deal with the eigenvalues of the perturbed model. We have that, for any $(\ell_1, \dots, \ell_p) \in (b_V, +\infty)^p$ and any $\eta > 0$, there exists $\varepsilon(\eta, \ell), \delta(\eta, \ell) > 0$ so that for n large enough, for $\varepsilon < \varepsilon(\eta, \ell)$, $\delta < \delta(\eta, \ell)$,

$$\begin{aligned} &-L^0_{\ell_1,\dots,\ell_p}(x_1, \dots, x_k) - \eta \\ &\quad\le \frac{1}{n} \log \mathbb{P}\left(\max_{1\le i\le k} |\widetilde{\lambda}^n_i - x_i| \le \delta \,\middle|\, \begin{matrix} \max_{1\le i\le p} |\lambda^n_i - \ell_i| \le \varepsilon, \\ \lambda^n_{p+1} \le b_V + \varepsilon \end{matrix} \right) \\ &\quad\le -L^0_{\ell_1,\dots,\ell_p}(x_1, \dots, x_k) + \eta \end{aligned} \quad (27)$$

These inequalities are a consequence of Theorem 9.1. Indeed, let X_n be a matrix such that the event $\{\max_{1\le i\le p}|\lambda_i^n-\ell_i|\le\epsilon\}$ holds. Let X_n' be a real diagonal matrix with same eigenvalues of X_n except its k largest eigenvalues are equal to the outliers $(\ell_1,\ldots,\ell_p)$. Then we have $\|X_n-X_n'\|_\infty\le\epsilon$, so that, with obvious notations, $\|\widetilde{X_n}-\widetilde{X_n}'\|_\infty\le\epsilon$, so that the ordered eigenvalues of $\widetilde{X_n}$ and $\widetilde{X_n}'$ differ at most by ϵ. Thus, up to change δ into $\delta\pm\varepsilon$, Theorem 9.1 gives (27).

- We have now all the ingredients to prove the LDP. 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 LDP. We let $K(L)$ be such that the probability that λ_1^n or $\widetilde{\lambda}_1^n$ is greater than $K(L)$ is smaller than e^{-nL}.
- To prove the upper bound we can write, for any $p\ge k$, any $\eta>0$, $\delta>0$,

$$\mathbb{P}\left(\max_{1\le i\le k}|\widetilde{\lambda}_i^n-x_i|\le\delta\right)\le\mathbb{P}\left(\max_{1\le i\le k}|\widetilde{\lambda}_i^n-x_i|\le\delta,\ \lambda_{p+1}^n\le b_V+\delta\right)+e^{-np\varepsilon(\delta)} \tag{28}$$

We fix $\eta>0$. As $[b_V,K(L)]^p$ is compact, from its infinite open covering $\cup V_\eta(\ell_1,\ldots,\ell_p)$, one can always extract a finite covering $\cup_{1\le s\le M(\eta)} V_\eta(\ell_1^s,\ldots,\ell_p^s)$. We then take $\delta=\min\delta(\eta,\ell^s)>0$. Thus, we get by the LDP estimate (26)

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

which gives the announced bound by taking first the limit as n goes to infinity, then L, p to infinity and finally δ and η to zero.

- The lower bound is easier as we simply write

$$\mathbb{P}\left(\max_{1\le i\le k}|\widetilde{\lambda}_i^n-x_i|\le 2\delta\right)\ge\mathbb{P}\left(\max_{1\le i\le k}|\widetilde{\lambda}_i^n-x_i|\le 2\delta\cap V(\ell_1^s,\ldots,\ell_p^s)\right)$$

and use the large deviation theorems. □

11 Appendix

11.1 Proof of a technical lemma

With the notations of Sect. 4.1, we have the following result

Lemma 11.1 *Under Assumption 2.2, for any $1 \leq i_0 \leq r$,we have*

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

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 Cramér's Theorem [16], we have that the law of L^n satisfies a LDP with convex good rate function

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

where $\Lambda(\lambda) = \log \mathbb{E}(e^{\langle \lambda, Z_1 \rangle})$ is exactly the function defined in Eq. (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 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 $\mathcal{E}_\delta := \{y \in \mathsf{H}_r\,;\ P_r(y) \leq \delta\}$, we have

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

Let us assume that

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

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, $P_r(y_0) = \lim_{n \to \infty} P_r(y_{\varphi(n)}) = 0$. 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 $\langle p, y_0 \rangle = \mathrm{Tr}(y_0 p) = 0$.

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

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,

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

which completes the proof of the lemma.

11.2 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 11.2 *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 $\lambda_1, \ldots, \lambda_n$ and some non-null real numbers $\theta_1, \ldots, \theta_r$. We make the following hypothesis: (H) $\lambda_1, \ldots, \lambda_n$ are pairwise distinct and there are pairwise distinct indices $i_1, \ldots, i_{r-1} \in \{1, \ldots, n\}$ such that $\{\lambda_{i_1} + \theta_1, \ldots, \lambda_{i_{r-1}} + \theta_{r-1}\} \cap \{\lambda_1, \ldots, \lambda_n\} = \emptyset$.*

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

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

where $(u_1, \ldots, u_r)$ is either the orthonormalized family deduced from the columns of g by the Gram–Schmidt process or $\frac{1}{\sqrt{n}}(g_1, \ldots, g_r)$. Then the Lebesgue measure of the set of the g's such that $\widetilde{X}_g$ 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 $\lambda_1^n, \ldots, \lambda_n^n$ are distinct. Moreover, let η 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 λ_{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 ones. Among them, one can find an eigenvalue λ_{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 θ_i's, we consider the $p^- + 1$ smallest eigenvalues in $(a - \eta, a + \eta)$. We now prove Lemma 11.2.

Proof The idea of the proof is the following. We shall first prove (in Step I) that the set of g's such that $\widetilde{X}_g$ 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, proving (in Step II) that this function is not identically null will then imply that the set of such g's has vanishing Lebesgue measure.

Let β 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 βnr 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 $\widetilde{X}_g$. The set of g's in $\mathbb{K}^{n\times r}$ such that X and $\widetilde{X}_g$ 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 $\widetilde{X}_g$ 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 βnr real variables such that the i, jth 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 βnr 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 $\widetilde{X}_g$ have no eigenvalue in common if and only if $D(g)X$ and $D(g)\widetilde{X}_g$ have no eigenvalue in common. Now, the advantage of having replaced X and $\widetilde{X}_g$ by $D(g)X$ and $D(g)\widetilde{X}_g$ is that the entries of $D(g)X$ and $D(g)\widetilde{X}_g$ are polynomial functions of g. Hence if one defines $P(g)$ to be the resultant of the characteristic polynomials of $D(g)X$ and $D(g)\widetilde{X}_g$, $P(g)$ is a polynomial function of the βnr 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 $\widetilde{X}_g$ 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 $g \longmapsto P(g)$ 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 $\widetilde{X}_g$ and X have no eigenvalue in common. One can suppose that $i_1 = 1, \dots, i_{r-1} = r - 1$, that

$\lambda_r < \cdots < \lambda_n$ and that

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

We shall choose the $r-1$ first columns $g_1, \ldots, g_{r-1}$ 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

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

Let us suppose that $\theta_r > 0$. It was shown in [24, Sect. 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 $\widetilde{X}_g$ describe the set of families $\mu_r, \ldots, \mu_n$ of real numbers which sum up to $\lambda_r + \cdots + \lambda_n + \theta_r$ and such that

$$\lambda_r \le \mu_r \le \lambda_{r+1} \le \cdots \le \lambda_n \le \mu_n.$$

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

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

which concludes the proof, by hypothesis (H). □

References

1. Anderson, G., Guionnet, A., Zeitouni, O.: An introduction to random matrices. In: Cambridge Studies in Advanced Mathematics, vol. 118. Cambridge University Press, Cambridge (2010)
2. Anderson, G., Zeitouni, O.: A CLT for a band matrix model. Probab. Theory Relat. Fields **134**, 283–338 (2006)
3. Bai, Z., Yao, J.-F.: Central limit theorems for eigenvalues in a spiked population model. Ann. Inst. Henri Poincaré Probab. Stat. **44**, 447–474 (2008)
4. Bai, Z.D., Silverstein, J.W.: No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. Ann. Prob. **26**, 316–345 (1998)
5. Bai, Z.D., Silverstein, J.W.: Spectral analysis of large dimensional random matrices. In: Springer Series in Statistics. New York (2010)
6. Bai, Z.D., Yao, J.-F.: Central limit theorems for eigenvalues in a spiked population model. Ann. I.H.P.-Prob. et Stat. **44**, 447–474 (2008)
7. Baik, J., Ben Arous, G., Péché, S.: Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. Ann. Prob. **33**, 1643–1697 (2005)

8. Ben Arous, G., Dembo, A., Guionnet, A.A.: Aging of spherical spin glasses. Probab. Theory Relat. Fields **120**, 1–67 (2001)
9. Ben Arous, G., Guionnet, A.: Large deviations for Wigner's law and Voiculescu 's non commutative entropy. Probab. Theory Relat. Fields **108**, 517–542 (1997)
10. Benaych-Georges, F., Guionnet, A., Maïda, M.: Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. Electron. J. Probab. (2011, in press)
11. Benaych-Georges, F., Rao, R.N.: The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Adv. Math. **227**(1), 494–521 (2011)
12. Bercu, B., Gamboa, F., Rouault, A.: Large deviations for quadratic forms of stationary Gaussian processes. Stoch. Process Appl. **71**, 75–90 (1997)
13. Capitaine, M., Donati-Martin, C., Féral, D.: The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. Ann. Probab. **37**, 1–47 (2009)
14. Capitaine, M., Donati-Martin, C., Féral, D.: Central limit theorems for eigenvalues of deformations of Wigner matrices. Ann. Inst. H. Poincaré Probab. Stat. (2011, in press)
15. Dembo, A., Guionnet, A., Zeitouni, O.: Moderate deviations for the spectral measure of certain random matrices. Ann. Inst. H. Poincaré Probab. Stat. **39**, 1013–1042 (2003)
16. Dembo, A., Zeitouni, O.: Large deviation techniques and applications. In: Applications of Mathematics, vol. 38. Springer, New York (1998)
17. Döring, H., Eichelsbacher, P.: Moderate deviations in random graphs and Bernoulli random matrices. Electron. J. Probab. **14**, 2636–2656 (2009)
18. Erdős, L., Péché, S., Ramírez, J., Schlein, B., Yau, H.T.: Bulk universality for Wigner matrices. Commun. Pure Appl. Math. **63**, 895–925 (2010)
19. Féral, D., Péché, S.: The largest eigenvalue of rank one deformation of large Wigner matrices. Commun. Math. Phys. **272**, 185–228 (2007)
20. Guionnet, A., Zeitouni, O.: Concentration of the spectral measure for large matrices. Electron. Commun. Probab. **5**, 119–136 (2000)
21. Hiai, F., Petz, D.: Eigenvalue density of the Wishart matrix and large deviations. Infin. Dimens. Anal. Quantum Probab. Relat. Top. **1**, 633–646 (1998)
22. Féral, D.: On large deviations for the spectral measure of discrete Coulomb gas. In: Séminaire de probabilités XLI. Lecture Notes in Math. vol. 1934, pp. 19–49 (2008)
23. Fey, A., van der Hofstad, R., Klok, M.: Large deviations for eigenvalues of sample covariance matrices with applications to mobile communication systems. Adv. Appl. Probab. **40**, 1048–1071 (2008)
24. Forrester, P.J., Nagao, T.: Determinantal Correlations for Classical Projection Processes. arxiv
25. Füredi, Z., Komlós, J.: The eigenvalues of random symmetric matrices. Combinatorica **1**, 233–241 (1981)
26. Guionnet, A., Zeitouni, O.: Large deviations asymptotics for spherical integrals. J. Funct. Anal. **188**, 461–515 (2002)
27. Horn, R., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (1990)
28. Maïda, M.: Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. Electron. J. Probab. **12**, 1131–1150 (2007)
29. Maïda, M., Najim, J., Péché, S.: Large deviations for weighted empirical mean with outliers. Stoch. Process. Appl. **117**, 1373–1403 (2007)
30. Najim, J.: A Cramér type theorem for weighted random variables. Electron. J. Probab. **7**, 1–32 (2002)
31. Pastur, L.A., Martchenko, V.A.: The distribution of eigenvalues in certain sets of random matrices. Math. USSR Sbornik **1**, 457–483 (1967)
32. Soshnikov, A.: Universality at the edge of the spectrum in Wigner random matrices. Commun. Math. Phys. **207**, 697–733 (1999)
33. Tao, T., Knutson, A.: Honeycombs and sums of Hermitian matrices. Not. Am. Math. Soc. **48**, 175–186 (2001)
34. Tao, T., Vu, V.: Random matrices: Universality of local eigenvalue statistics. Acta Math (in press)
35. Vivo, P., Majumdar, S., Bohigas, O.: Large deviations of the maximum eigenvalue in Wishart random matrices. J. Phys. A **40**, 4317–4337 (2007)
36. Wigner, E.P.: Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math. **62**, 548–564 (1955)