THE EIGENVALUES AND EIGENVECTORS OF FINITE, LOW RANK PERTURBATIONS OF LARGE RANDOM MATRICES

FLORENT BENAYCH-GEORGES AND RAJ RAO NADAKUDITI

Abstract. In this paper, we consider the eigenvalues and eigenvectors of finite, low rank perturbations of random matrices. Specifically, we prove almost sure convergence of the extreme eigenvalues and appropriate projections of the corresponding eigenvectors of the perturbed matrix for additive and multiplicative perturbation models.

The limiting non-random value is shown to depend explicitly on the limiting spectral measure and the assumed perturbation model via integral transforms that correspond to very well known objects in free probability theory that linearize non-commutative free additive and multiplicative convolution. Moreover, we uncover a remarkable phase transition phenomenon whereby the large matrix limit of the extreme eigenvalues of the perturbed matrix differs from that of the original matrix if and only if the eigenvalues of the perturbing matrix are above a certain critical threshold. This critical threshold is intimately related to the same aforementioned integral transforms.

We examine the consequence of this eigenvalue phase transition on the associated eigenvectors and generalize our results to examine the singular values and vectors of finite, low rank perturbations of rectangular random matrices. The analysis brings into sharp focus the analogous connection with rectangular free probability. Various extensions of our results are discussed.

1. Introduction

Let X_n be an $n\times n$ matrix with real eigenvalues $\lambda_1(X_n),\ldots,\lambda_n(X_n)$ and P_n be an $n\times n$ matrix with rank $r\leq n$ and real eigenvalues $\theta_1,\ldots,\theta_r$. A fundamental question in matrix analysis is the following [12, 2]:

> How are the eigenvalues and eigenvectors of X_n+P_n related to the eigenvalues and eigenvectors of X_n and P_n?

When X_n and P_n are diagonalized by the same eigenvectors then we have $\lambda_i(X_n+P_n)=\lambda_j(X_n)+\lambda_k(P_n)$ for appropriate choice of indices $i,j,k\in\{1,\ldots,n\}$. In the general setting, however, the answer is much more complicated because the eigenvalues and eigenvectors of their sum depend on the relationship between the eigenspaces of the individual matrices.

Date: October 12, 2009.

2000 *Mathematics Subject Classification.* 15A52, 46L54, 60F99.

Key words and phrases. Random matrices, Haar measure, free probability, phase transition, random eigenvalues, random eigenvectors, random perturbation, sample covariance matrices.

F.B.G's work was partially supported by the *Agence Nationale de la Recherche* grant ANR-08-BLAN-0311-03. R.R.N's research was partially supported by an Office of Naval Research postdoctoral fellowship award and grant N00014-07-1-0269. R.R.N thanks Arthur Baggeroer for his feedback, support and encouragement. We thank Alan Edelman for feedback and encouragement and for facilitating this collaboration by hosting F.B.G's stay at M.I.T. We gratefully acknowledge the Singapore-MIT alliance for funding F.B.G's stay.

In this scenario, one can use Weyl's interlacing inequalities [20] to obtain coarse bounds for the eigenvalues of the sum in terms of the eigenvalues of X_n. When the norm of P_n is small relative to the norm of X_n, tools from perturbation theory (see [20, Chapter 6] or [33]) can be employed to improve the characterization of the bounded set in which the eigenvalues of the sum must lie. Exploiting any special structure in the matrices allows us to refine these bounds [22] but this is pretty much as far as the theory goes. Instead of exact answers we have a system of messy, coupled bounds. The eigenvector story is even more convoluted.

Surprisingly, adding some randomness to the eigenspaces permits further analytical progress. Specifically, if the eigenspaces are assumed to be isotropically random and "in generic position with respect to each other", then analytical elegance returns in the limit of large matrices.

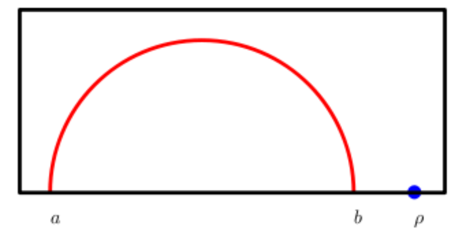
In place of eigenvalue bounds we have simple, exact answers that are to be interpreted probabilistically. The results bring into sharp focus a remarkable phase transition phenomenon of the kind illustrated in Figure 1 for the eigenvalues and eigenvectors of X_n+P_n. In this paper, we also uncover a similar phase transition behavior for the eigenvalues and eigenvectors of $X_n(I+P_n)$ and, in the case where X_n, P_n are rectangular, for the singular values and singular vectors of $X_n + P_n$. A precise statement of the results may be found in Section 2.

Examining the structure of the analytical expression for θ_c and ρ in Figure 1 reveals a common underlying theme in the additive, multiplicative and rectangular cases. The critical values θ_c and ρ in Figure 1 are related to integral transforms of the limiting spectral measure μ_X of X_n. It turns out that these G, T and D integral transforms that emerge in the respective additive, multiplicative and rectangular cases are deeply related to very well known objects in free probability theory [34, 19] that linearize (non-commutative) free additive, multiplicative [34] and rectangular [9, 8] convolutions respectively.

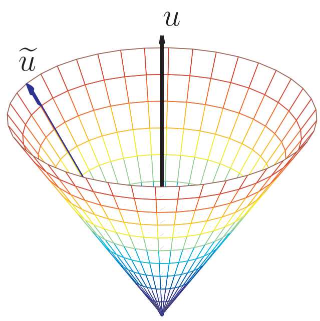
The emergence of these transforms in the context of the study of the extreme/isolated eigenvalue behavior should be of independent interest to free probabilists. This justifies our anointment of the study of finite, low rank perturbations of large random matrices as "spiked" free probability theory. In this framework, regular free probability theory would correspond the study of full rank perturbations of large random matrices.

The development of spiked free probability theory is the main contribution of this paper. In doing so, we dramatically extend the results found in the literature for the eigenvalue phase transition in such finite, low rank perturbation models well beyond the Gaussian [3, 4, 28, 21, 17, 13, 6], Wishart [16, 27, 25] and Jacobi settings [24]. In our situation, the distribution μ_X in Figure 1 can be any probability measure. Consequently, the aforementioned results in the literature can be rederived rather simply using the formulas in Section 2 by substituting μ_X with the semi-circle measure [35] (for Gaussian matrices), the Marčenko-Pastur measure [23] (for Wishart matrices) or the free Jacobi measure (for Jacobi matrices [14]). See Section 3 for some concrete computations.

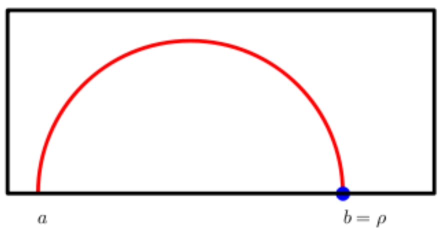
The development of the eigenvector aspect of the story is another important contribution that we would like to highlight. Generally speaking, the eigenvector question has received much less much attention in random matrix theory and in free probability theory despite impressive breakthroughs [11]. A notable exception is the recent body of work on



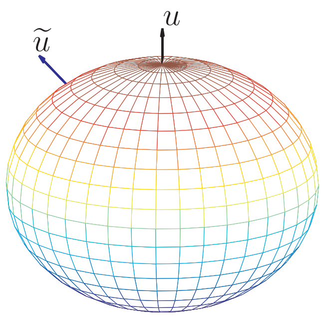
(a) Largest eigenvalue $\rho > b$ in blue when $\theta > \theta_c$



(b) Associated eigenvector when $\theta > \theta_c$



(c) Largest eigenvalue $\rho = b$ in blue when $\theta \leq \theta_c$



(d) Associated eigenvector when $\theta \leq \theta_c$

FIGURE 1. Informally speaking, suppose that the histogram of the eigenvalues of X_n, normalized to have area one, falls on the solid curve μ_X in (a) with largest eigenvalue b. Consider the matrix $P_n := \theta u u^*$ with rank $r = 1$ and largest eigenvalue $\theta (> 0$ say). The vector u is an $n \times 1$ vector chosen uniformly at random from the unit n-sphere. The largest eigenvalue of $X_n + P_n$ will differ from b if and only if θ is greater than some critical value θ_c. In this event, the largest eigenvalue will be concentrated around ρ with high probability. The associated eigenvector $\widetilde{u}$ will, with high probability, lie on a cone around u as in (b). When $\theta \leq \theta_c$, a phase transition occurs so that with high probability the largest eigenvalue of the sum will equal b as in (c) and the corresponding eigenvector will be uniformly distributed on the unit sphere as in (d). For details, see Section 2.

the eigenvectors of spiked Wishart matrices [27, 21, 25] which corresponds to μ_X being the Marčenko-Pastur measure. In this paper, we extend their results for multiplicative models of the kind $X_n(I+P_n)$ to the setting where μ_X is an arbitrary probability measure and obtain new results for the eigenvectors for additive models and the singular vectors for rectangular models.

Our proofs rely heavily on the derivation of master equation representations of the eigenvalues and eigenvectors of the perturbed matrix and the subsequent application of concentration inequalities for random vectors uniformly distributed on high dimensional unit spheres to these implicit master equation representations. Consequently, our technique is simpler, more general and more transparently reveals the source of the phase transition phenomenon than other proofs found in the literature.

The paper is organized as follows. In Section 2, we state the main results and present the G, T and D integral transforms mentioned above, in Section 3, we give examples, and the rest of the paper is devoted to the proofs (sketches of the proofs in Section 4, master equations for the eigenvalues and eigenvectors of perturbed matrices in Section 5, proofs of the main results in Sections 6 to 10 and proofs of technical results needed here in the Appendix).

2. Main results

Let X_n be an $n\times n$ symmetric (or Hermitian) random matrix with eigenvalues $\lambda_1(X_n) \geq \cdots \geq \lambda_n(X_n)$. We denote by μ_{X_n} the empirical distribution on the set of its eigenvalues, *i.e.*, the probability measure defined as

$$\mu_{X_n} = \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j(X_n)}.$$

Assume that the probability measure μ_{X_n} converges almost surely weakly, as n tends to infinity, to a non-random, compactly supported probability measure μ_X. Let the smallest and largest eigenvalues of X_n converge almost surely to a and b which are, respectively, the infimum and supremum of the support of μ_X.

Let us fix a positive integer r and some non-random real numbers $\theta_1 \geq \cdots \geq \theta_r$ not equal to zero. For each n, let P_n be an $n\times n$ symmetric (or Hermitian) random matrix independent of X_n, which non null eigenvalues are $\theta_1,\ldots,\theta_r$. Let $s\in\{0,\ldots,r\}$ where $\theta_1 \geq \cdots \geq \theta_s > 0 > \theta_{s+1} \geq \cdots \geq \theta_r$.

We suppose that either X_n or P_n is invariant, in law, by conjugation by any orthogonal (or unitary) matrix.

For M an Hermitian $n\times n$ matrix, we denoted by $\lambda_1(M) \geq \cdots \geq \lambda_n(M)$ the ordered eigenvalues of M. Let $\xrightarrow{\text{a.s.}}$ denote almost sure convergence. At last, let, for F a subspace of an Euclidian space E and $x\in E$, $\langle x, F\rangle$ denote the norm of the orthogonal projection of x onto F. We are now ready to state our main results.

2.1. Extreme eigenvalues and eigenvectors under additive perturbations.

Let us define

$$\widetilde{X}_n = X_n + P_n.$$

Theorem 2.1 (Eigenvalue phase transition)**.** *The extreme eigenvalues of $\widetilde{X}_n$ exhibit the following behavior as $n\longrightarrow\infty$. We have that for each $i=1,\ldots,s$,*

$$\lambda_i(\widetilde{X}_n) \xrightarrow{a.s.} \begin{cases} G_{\mu_X}^{-1}(1/\theta_i) & \text{if } 1/\theta_i < G_{\mu_X}(b^+), \\ b & \text{otherwise,} \end{cases}$$

while for each $i = s+1, \ldots, r$,

$$\lambda_{n-r+i}(\widetilde{X}_n) \xrightarrow{a.s.} \begin{cases} G_{\mu_X}^{-1}(1/\theta_i) & \text{if } 1/\theta_i > G_{\mu_X}(a^-), \\ a & \text{otherwise.} \end{cases}$$

where

$$G_{\mu_X}(z) = \int \frac{1}{z-t} \mathrm{d}\mu_X(t) \qquad \text{for } z \notin \operatorname{supp} \mu_X,$$

is the Cauchy (or G) transform of μ_X.

Theorem 2.2 (Norm of eigenvector projection)**.** *Consider indices $i_0 \in \{1, \ldots, r\}$ such that $1/\theta_{i_0} \in (G_{\mu_X}(a^-), G_{\mu_X}(b^+))$. For each n, define*

$$z_n := \begin{cases} \lambda_{i_0}(\widetilde{X}_n) & \text{if } \theta_{i_0} > 0, \\ \lambda_{n+r-i_0}(\widetilde{X}_n) & \text{if } \theta_{i_0} < 0, \end{cases}$$

and let x_n be a unit eigenvector of $\widetilde{X}_n$ associated with the eigenvalue z_n. Then we have

(a)

$$\langle x_n, \ker(\theta_{i_0} I_n - P_n)\rangle^2 \xrightarrow{a.s.} \frac{-1}{\theta_{i_0}^2 G'_{\mu_X}(\rho)}$$

where $\rho = G_{\mu_X}^{-1}(1/\theta_{i_0})$ is the limit of z_n;

(b)

$$\langle x_n, \oplus_{j \neq i_0} \ker(\theta_j I_n - P_n)\rangle \xrightarrow{a.s.} 0,$$

as $n \longrightarrow \infty$.

Theorem 2.3 (Eigenvector phase transition)**.** *Suppose here that $r = 1$ and denoted the non-null eigenvalue of P_n by θ. Suppose that*

$$\frac{1}{\theta} \notin (G_{\mu_X}(a^-), G_{\mu_X}(b^+)), \quad \text{and} \quad \begin{cases} G'_{\mu_X}(b^+) = -\infty & \text{if } \theta > 0, \\ G'_{\mu_X}(a^-) = -\infty & \text{if } \theta < 0. \end{cases}$$

For each n, let x_n be a unit eigenvector of $\widetilde{X}_n$ associated with either the largest or smallest eigenvalue depending on whether $\theta > 0$ or $\theta < 0$ respectively. Then we have

$$\langle x_n, \operatorname{ran}(P_n)\rangle \xrightarrow{a.s.} 0$$

as $n \longrightarrow \infty$.

Remark 2.4 (Subtlety regarding the eigenvector phase transition)**.** Let us say a few words about the hypotheses of Theorem 2.3. The strong hypotheses we make here ($r = 1$ and a certain function has infinite integral), compared to the ones of Theorem 2.2, could convey the impression that the phase transition for the localization of the eigenvectors associated with the extreme eigenvalues of $X_n + P_n$ has not been treated in its full generality. Specifically, one might still hope that the statement of Theorem 2.2 remains true for an eigenvalue z_n having limit $\rho \in \{a, b\}$ as long as $-G'_{\mu_X}(\rho)$, which is equal to $\int \frac{\mathrm{d}\mu_X(t)}{(\rho-t)^2}$, is finite and that the statement of Theorem 2.3 might stay true for arbitrary r.

In fact, if stronger hypotheses on the manner in which the spectral measure of X_n tends to μ_X as $n \longrightarrow \infty$ were incorporated, then both of these theorems could be generalized in this direction. Indeed, the key, in the proof of Theorem 2.2, is that

$$\int \frac{\mathrm{d}\mu_{i,j}^{(n)}(t)}{z_n - t} \underset{n\to\infty}{\longrightarrow} \mathbb{1}_{i=j} G_{\mu_X}(\rho), \qquad \int \frac{\mathrm{d}\mu_{i,i}^{(n)}(t)}{(z_n - t)^2} \underset{n\to\infty}{\longrightarrow} \int \frac{\mathrm{d}\mu_X(t)}{(\rho - t)^2} \tag{1}$$

and the key, to generalize the proof of Theorem 2.3 for $r > 1$, would be to have

$$\frac{\int \frac{\mathrm{d}\mu_{i,j}^{(n)}(t)}{(z_n-t)^2}}{\left(\int \frac{\mathrm{d}\mu_{i,i}^{(n)}(t)}{(z_n-t)^2} \int \frac{\mathrm{d}\mu_{j,j}^{(n)}(t)}{(z_n-t)^2}\right)^{\frac{1}{2}}} \underset{n\to\infty}{\longrightarrow} \mathbb{1}_{i=j}. \tag{2}$$

Simulations, however, suggest that stronger hypotheses on the $\lambda_i(X_n)$'s are needed to generalize (1) and (2). That said, it appears that when the spacings between the $\lambda_i(X_n)$'s are more "random matrix like" than "independent sample like", then (1) and (2) seem to hold without imposing any particular requirements on μ_X.

2.2. **Extreme eigenvalues and eigenvectors under multiplicative perturbations.** In this section, the hypotheses are the ones introduced at the beginning of Section 2. We suppose moreover that for all n, X_n is a non-negative definite matrix and that $\mu_X \neq \delta_0$.

Let us define

$$\widetilde{X}_n = X_n(I_n + P_n).$$

Theorem 2.5 (Eigenvalue phase transition)**.** *The extreme eigenvalues of $\widetilde{X}_n$ exhibit the following behavior as $n \longrightarrow \infty$, assuming that $\mu_X \neq \delta_0$. We have that for each $i = 1, \dots, s$,*

$$\lambda_i(\widetilde{X}_n) \xrightarrow{a.s.} \begin{cases} T_{\mu_X}^{-1}(1/\theta_i) & \text{if } 1/\theta_i < T_{\mu_X}(b^+), \\ b & \text{otherwise,} \end{cases}$$

while for each $i = s+1, \dots, r$,

$$\lambda_{n-r+i}(\widetilde{X}_n) \xrightarrow{a.s.} \begin{cases} T_{\mu_X}^{-1}(1/\theta_i) & \text{if } 1/\theta_i > T_{\mu_X}(a^-), \\ a & \text{otherwise,} \end{cases}$$

where

$$T_{\mu_X}(z) = \int \frac{t}{z-t} \mathrm{d}\mu_X(t) \qquad \text{for } z \notin \operatorname{supp} \mu_X,$$

is the T-transform of μ_X (defined in Section 2.4.2).

Theorem 2.6 (Norm of eigenvector projection)**.** *Consider indices $i_0 \in \{1, \dots, r\}$ such that $1/\theta_{i_0} \in (T_{\mu_X}(a^-), T_{\mu_X}(b^+))$. For each n, define*

$$z_n := \begin{cases} \lambda_{i_0}(\widetilde{X}_n) & \text{if } \theta_{i_0} > 0, \\ \lambda_{n+r-i_0}(\widetilde{X}_n) & \text{if } \theta_{i_0} < 0, \end{cases}$$

and let x_n be a unit eigenvector of $\widetilde{X}_n$ associated with the eigenvalue z_n. Then we have *a)*

$$\langle x_n, \ker(\theta_{i_0} I_n - P_n)\rangle^2 \xrightarrow{a.s.} \frac{-1}{\theta_{i_0}^2 \rho T'_{\mu_X}(\rho) + \theta_{i_0}},$$

where $\rho = T_{\mu_X}^{-1}(1/\theta_{i_0})$ *is the limit of* z_n*;*

b)

$$\langle x_n, \oplus_{j \neq i_0} \ker(\theta_j I_n - P_n)\rangle \xrightarrow{a.s.} 0,$$

as $n \longrightarrow \infty$*.*

Theorem 2.7 (Eigenvector phase transition)**.** *Suppose here that* $r = 1$ *and denoted the non-null eigenvalue of* P_n *by* θ*. Suppose that*

$$\frac{1}{\theta} \notin (T_{\mu_X}(a^-), T_{\mu_X}(b^+)), \quad \textit{and} \quad \begin{cases} T'_{\mu_X}(b^+) = -\infty & \textit{if } \theta > 0, \\ T'_{\mu_X}(a^-) = -\infty & \textit{if } \theta < 0. \end{cases}$$

For each n*, let* x_n *be an eigenvector of* $\widetilde{X}_n$ *associated with either the largest or smallest eigenvalue depending on whether* $\theta > 0$ *or* $\theta < 0$*, respectively. Then, by we have*

$$\langle x_n, \operatorname{ran}(P_n)\rangle \xrightarrow{a.s.} 0$$

as $n \longrightarrow \infty$*.*

The analogue of Remark 2.4 also applies here.

Remark 2.8 (Eigenvalues and eigenvectors of a similarity transformation of X)**.** Consider the matrix $S_n = (I_n + P_n)^{1/2} X_n (I_n + P_n)^{1/2}$. The matrix S_n and $\widetilde{X}_n = X_n(I_n + P_n)$ are related by a similarity transformation $S_n = (I + P_n)^{1/2} \widetilde{X}_n (I + P_n)^{-1/2}$ so that they share the same eigenvalues and consequently the same limiting eigenvalue behavior in Theorem 2.5. Additionally, if x_n is a unit norm eigenvector of $\widetilde{X}_n$ then $y_n = (I_n + P_n)^{1/2} x_n / \| (I_n + P_n)^{1/2} x_n \|$ is an eigenvector of S_n. Consequently, we have

$$\langle y_n, \ker(\theta_{i_0} I_n - P_n)\rangle^2 = \frac{(\theta_{i_0} + 1)\langle x_n, \ker(\theta_{i_0} I_n - P_n)\rangle^2}{\theta_{i_0}\langle x_n, \ker(\theta_{i_0} I_n - P_n)\rangle^2 + 1}.$$

It follows that we have the same phase transition and that when $1/\theta_{i_0} \in (T_{\mu_X}(a^-), T_{\mu_X}(b^+))$,

$$\langle y_n, \ker(\theta_{i_0} I_n - P_n)\rangle^2 \xrightarrow{\text{a.s.}} -\frac{\theta_{i_0} + 1}{\theta_{i_0} T'_{\mu_X}(\rho)} \quad \text{and} \quad \langle y_n, \oplus_{j \neq i_0} \ker(\theta_j I_n - P_n)\rangle \xrightarrow{\text{a.s.}} 0,$$

so that we have proved the analogue of Theorems 2.6 and 2.7 for the eigenvectors of S_n.

2.3. **Extreme singular values and singular vectors under additive perturbations.** Let us now treat the problem of the extreme singular values and of the associated singular pairs of vectors for rectangular random matrices. In the particular case where our rectangular matrices appear to be square, our results also allow treat the case of smallest singular values[1].

For this section, we adapt the hypotheses introduced in the beginning of Section 2 to the rectangular setting. For each n, X_n is a real (or complex) $n \times m$ random matrix which

[1]This particularity of the "square case" is due to the fact that for $c = 1$, for any probability measure μ with support contained in $(0, +\infty)$, the function $D_\mu(c, \cdot)$ is positive and increasing between 0 and the minimum of the support of μ (see Section 2.4.3 for more details).

is not anymore symmetric (or Hermitian), with singular values $\sigma_1(X_n) \geq \cdots \geq \sigma_n(X_n)$. The integer $m \geq n$ shall also tend to infinity[2] in such a way that n/m tends to a limit $c \in [0, 1]$.

Assume that the empirical distribution on the set of its singular values, *i.e.*

$$\frac{1}{n}\sum_{j=1}^{n} \delta_{\sigma_j(X_n)}$$

converges almost surely weakly, as n, m tend to infinity, to a non-random, compactly supported probability measure μ_X. Let the smallest and largest singular values of X_n converge almost surely to a and b which are, respectively, the infimum and supremum of the support of μ_X.

For each n, P_n is an $n \times m$ real (or complex) random matrix independent of X_n, which non null singular values are $\theta_1 \geq \cdots \geq \theta_r$ (which are non-random and independent of n, as in the Hermitian context).

We suppose that either X_n or P_n is invariant, in law, by multiplication, on both sides, by any orthogonal (or unitary) matrix.

We define

$$\widetilde{X}_n = X_n + P_n.$$

For X an $n \times m$ matrix, we denoted by $\sigma_1(X) \geq \cdots \geq \sigma_n(X)$ the ordered singular values of X.

2.3.1. *Singular values.*

Theorem 2.9 (Singular value phase transition)**.** *The r largest singular values of the $n \times m$ perturbed matrix $\widetilde{X}_n$ exhibit the following behavior as $n, m \to \infty$ and $n/m \to c$. We have that for each $i = 1, \ldots, r$,*

$$\sigma_i(\widetilde{X}_n) \xrightarrow{a.s.} \begin{cases} D_{\mu_X}^{-1}(c, 1/\theta_i^2) & \text{if } 1/\theta_i^2 < D_{\mu_X}(c, b^+), \\ b & \text{otherwise,} \end{cases}$$

where

$$D_{\mu_X}(c, z) = \left[\int \frac{z}{z^2 - t^2} \mathrm{d}\mu(t)\right] \cdot \left[c \int \frac{z}{z^2 - t^2} \mathrm{d}\mu(t) + \frac{1-c}{z}\right] \qquad \text{for } z > b,$$

is the D-transform of μ_X.

In the special case where $a > 0$ and m is always equal to n, so that $n/m \to 1$, we also have, for each $i \in \{1, \ldots, r\}$,

$$\sigma_{n-1+i}(\widetilde{X}_n) \xrightarrow{a.s.} \begin{cases} D_{\mu_X, [0,a)}^{-1}(1/\theta_i^2) & \text{if } 1/\theta_i^2 < D_{\mu_X}(a^-), \\ a & \text{otherwise.} \end{cases}$$

where $D_{\mu_X,[0,a)}(z)$ is defined as $D_{\mu_X}(1, z)$ above, but with $z \in [0, a)$.

[2]To be as rigorous as possible, one should make m depend on n, thus write m_n. However, in order to lighten the notation, we omit the index n and only write m.

2.3.2. *Brief review on pairs of singular vectors.* Let X be an $n \times m$ matrix with entries in the field $\mathbb{K}$ (which can be either $\mathbb{R}$ or $\mathbb{C}$) with singular values $\sigma_1 \geq \cdots \geq \sigma_n$. For $i = 1, \ldots, n$, a *pair of singular vectors* of X associated with σ_i is a pair (u, v) of non-zero column vectors such that

$$\|u\| = \|v\|, \qquad Xv = \sigma_i u \qquad \text{and} \qquad X^* u = \sigma_i v.$$

These pairs are closely related to the decomposition of

$$X = U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} V^* \qquad (U, V\ \ n \times n,\ m \times n \text{ matrices, } UU^* = I_n = V^* V),$$

since for all $i = 1, \ldots, n$, the i-th columns of U and V form a pair of singular vectors of X associated with σ_i. Note also that for such a pair (u, v),

$$u \in \ker(\sigma_i^2 I_n - XX^*) \quad \text{and} \quad v \in \ker(\sigma_i^2 I_m - X^* X).$$

2.3.3. *Results in the singular vectors.* Let us first consider a singular value θ_{i_0} of P_n which gives rise to a singular value z_n of $X_n + P_n$ with limit $\rho \notin [a, b]$. The following theorem says that a pair of unit-length singular vectors (u_n, v_n) of $X_n + P_n$ associated with z_n somehow "keeps something" from the pairs of singular vectors of P_n associated with θ_{i_0}. Indeed, for all $j = 1, \ldots, r$, the norms of the orthogonal projection of u_n and v_n onto respectively $\ker(\theta_j^2 I_n - P_n P_n^*)$ and $\ker(\theta_j^2 I_m - P_n^* P_n)$ have positive limits for $j = i_0$ and null limits for $j \neq i_0$ and the relation $P_n v_n = \theta_{i_0} u_n$, though false in general, has a "non null true component".

Theorem 2.10 (Norm of projection of singular vectors)**.** *Consider indices $i_0 \in \{1, \ldots, r\}$ such that $1/\theta_{i_0} \in (0, D_{\mu_X}(c, b^+))$. For each n, define $z_n = \sigma_{i_0}(\widetilde{X}_n)$ and let (u_n, v_n) be the corresponding pair of unit singular vectors of $\widetilde{X}_n$. Then we have*

a)

$$\langle u_n, \ker(\theta_{i_0}^2 I_n - P_n P_n^*) \rangle^2 \xrightarrow{a.s.} \frac{-2\varphi_{\mu_X}(\rho)}{\theta_{i_0}^2 \partial_z D_{\mu_X}(c, \rho)}, \tag{3}$$

b)

$$\langle v_n, \ker(\theta_{i_0}^2 I_m - P_n^* P_n) \rangle^2 \xrightarrow{a.s.} \frac{-2\varphi_{\tilde{\mu}_X}(\rho)}{\theta_{i_0}^2 \partial_z D_{\mu_X}(c, \rho)}, \tag{4}$$

as $n \longrightarrow \infty$, where $\rho = D_{\mu_X}^{-1}(c, 1/\theta_{i_0}^2)$ is the limit of z_n, the probability measure $\tilde{\mu}_X = c\mu_X + (1 - c)\delta_0$ and

$$\varphi_\mu(z) = \int \frac{z}{z^2 - t^2} \mathrm{d}\mu(t).$$

Furthermore, in the same asymptotic limit we have

c)

$$\langle u_n, \oplus_{j \neq i_0} \ker(\theta_j^2 I_n - P_n P_n^*) \rangle \xrightarrow{a.s.} 0, \quad \text{and} \quad \langle v_n, \oplus_{j \neq i_0} \ker(\theta_j^2 I_m - P_n^* P_n) \rangle \xrightarrow{a.s.} 0,$$

d)

$$\langle \varphi_{\mu_X}(\rho) P_n(v_n) - u_n, \ker(\theta_{i_0}^2 I_n - P_n P_n^*) \rangle \xrightarrow{a.s.} 0,$$

e) in the case where $a > 0$ and m is always equal to n, a), b), c) and d) are also valid for $z_n = \sigma_{n-1+i_0}(X_n + P_n)$ and $\rho = D^{-1}_{\mu_X,[0,a)}(1/\theta^2_{i_0})$.

Remark 2.11. Note that in the case where $c = 1$ (as in Part e)), the quantities of (3) and (4) are equal and reduce to

$$-\frac{1}{\theta^2_{i_0}\varphi_{\mu_X}(\rho)}.$$

The quantity $\varphi_{\mu_X}(\rho)$, which appears in Part d), also reduces to $1/\theta_{i_0}$.

Let us now consider the pairs of singular vectors corresponding to extreme singular values which are asymptotically in the boundary of the support of μ_X, in the case where $r = 1$. Here, unlike in the previous theorem, these pairs are asymptotically orthogonal to the pairs of singular vectors associated with the non null singular value of P_n.

Theorem 2.12. *Suppose here that $r = 1$ and denote the unique nonzero singular value of P_n by θ. Suppose that*

$$1/\theta^2 \geq D_{\mu_X}(c, b^+) \qquad \text{and} \qquad \varphi'_{\mu_X}(b^+) = -\infty. \tag{5}$$

Let, for each n, (u_n, v_n) be a pair of unit singular vectors of $X_n + P_n$ associated with its larger singular value z_n. Then

$$\langle u_n, \ker(\theta^2 I_n - P_nP_n^*)\rangle \xrightarrow{a.s.} 0, \qquad \text{and} \qquad \langle v_n, \ker(\theta^2 I_m - P_n^*P_n)\rangle \xrightarrow{a.s.} 0.$$

This result stays true for the smallest singular value if m is always equal to n and instead of the hypotheses of (5), one supposes that

$$a > 0, \qquad 1/\theta^2 \geq D_{\mu_X,[0,a)}(a^-) \qquad \text{and} \qquad \varphi'_{\mu_X}(a^-) = -\infty.$$

The analogue of Remark 2.4 also applies here.

2.4. The G, T and D transforms in free probability theory.

2.4.1. *The Cauchy or G-transform and its relation to additive free convolution.* Let μ be a compactly supported law on the real line. Let us define

$$G_\mu(z) = \int \frac{\mathrm{d}\mu(t)}{z - t}$$

for z out of the support of μ. Let $[a, b]$ be the convex hull of the support of μ. Since we have

- $\partial_z G_\mu(z) = -\int \frac{\mathrm{d}\mu(t)}{(z-t)^2}$ out of the support of μ, which implies that G_μ is decreasing on each of the intervals $(-\infty, a)$ and $(b, +\infty)$,
- $G_\mu < 0$ on $(-\infty, a)$ and $G_\mu > 0$ on $(b, +\infty)$,
- $G_\mu(z) \to 0$ as $|z| \to +\infty$,

it follows that $G_\mu(a^-) := \lim_{z\uparrow a} G_\mu(z)$ and $G_\mu(b^+) := \lim_{z\downarrow b} G_\mu(z)$ exist in respectively $[-\infty, 0)$ and $(0, +\infty]$ and G_μ realizes decreasing homeomorphisms from $(-\infty, a)$ onto $(G_\mu(a^-), 0)$ and from $(b, +\infty)$ onto $(0, G_\mu(b^+))$.

In this paper, we shall denote by G_μ^{-1} the inverses of these homeomorphisms, even though G_μ shall sometimes define other homeomorphisms on the holes of the support of μ.

The so-called *R-transform*

$$R_\mu(z) := G_\mu^{-1}(z) - \frac{1}{z}$$

is the analogue of the logarithm of the Fourier transform for the free additive convolution, since for all probability measures μ_1, μ_2, $\mu_1 \boxplus \mu_2$ is characterized by the fact that

$$R_{\mu_1 \boxplus \mu_2}(z) = R_{\mu_1}(z) + R_{\mu_2}(z).$$

The coefficients of the series expansion of $R_\mu(z)$ are the so called *free cumulants* of μ (see [26, 1]).

2.4.2. *The T-transform and its relation to multiplicative free convolution.* Let $\mu \neq \delta_0$ be a law with compact support contained in $[0, +\infty)$. Let us denote by $[a, b]$ the convex hull of the support of μ. Let us define the *T-transform* of μ

$$T_\mu(z) = \int \frac{t}{z-t} \mathrm{d}\mu(t)$$

for z out of the support of μ. Since we have

- $\partial_z T_\mu(z) = -\int \frac{t}{(z-t)^2} \mathrm{d}\mu(t)$ out of the support of μ, which implies that T_μ is decreasing on each of the intervals $(-\infty, a)$ and $(b, +\infty)$,
- $T_\mu < 0$ on $(-\infty, a)$ and $T_\mu > 0$ on $(b, +\infty)$,
- $T_\mu(z) \to 0$ as $|z| \to +\infty$,

it follows that $T_\mu(a^-) := \lim_{z \uparrow a} T_\mu(z)$ and $T_\mu(b^+) := \lim_{z \downarrow b} T_\mu(z)$ exist in respectively $[-\infty, 0)$ and $(0, +\infty]$ and T_μ realizes decreasing homeomorphisms from $(-\infty, a)$ onto $(T_\mu(a^-), 0)$ and from $(b, +\infty)$ onto $(0, T_\mu(b^+))$.

In this paper, we shall denote by T_μ^{-1} the inverses of these homeomorphisms, even though T_μ shall sometimes define other homeomorphisms on the holes of the support of μ.

The so-called *S-transform*

$$S_\mu(z) := \frac{1+z}{z} \cdot \frac{1}{T_\mu^{-1}(z)} \tag{6}$$

is the analogue of the Fourier transform for the free multiplicative convolution, since for all probability measures μ_1, μ_2, $\mu_1 \boxtimes \mu_2$ is characterized by the fact that

$$S_{\mu_1 \boxtimes \mu_2}(z) = S_{\mu_1}(z) S_{\mu_2}(z).$$

2.4.3. *The D-transform and its relation to rectangular additive free convolution.* Let μ be a law with compact support contained in $[0, +\infty)$. Let us denote by $[a, b]$ the convex hull of the support of μ. We define, for any probability measure τ on the real line, $\varphi_\tau(z) = \int \frac{z}{z^2 - t^2} \mathrm{d}\tau(t)$. For $c \in [0, 1]$, the *D-transform with ratio c* of μ is the function

$$D_\mu(c, z) = \varphi_\mu(z) \varphi_{c\mu + (1-c)\delta_0}(z) \qquad (z > b).$$

For any fixed c, we have

- $D_\mu(c, z) > 0$ for all $z \in (b, +\infty)$,

- for all z,

$$-\partial_z D_\mu(c,z) = \varphi_{c\mu+(1-c)\delta_0}(z)\int \frac{z^2+t^2}{(z^2-t^2)^2}\mathrm{d}\mu(t) + \varphi_\mu(z)\int \frac{z^2+t^2}{(z^2-t^2)^2}\mathrm{d}[c\mu+(1-c)\delta_0](t), \quad (7)$$

which implies that $D_\mu(c,\cdot)$ is decreasing on $(b,+\infty)$,
- $D_\mu(c,z) \to 0$ as $|z| \to +\infty$,

it follows that $D_\mu(c,b^+) := \lim_{z\downarrow b} D_\mu(c,z)$ exist in $(0,+\infty]$ and $D_\mu(c,\cdot)$ realizes a decreasing homeomorphism from $(b,+\infty)$ onto $(0, D_\mu(c,b^+))$.

In this paper, we shall denote by $D_\mu(c,\cdot)^{-1}$ the inverses of these homeomorphisms, even though $D_\mu(c,\cdot)$ shall sometimes define other homeomorphisms between other intervals.

In the special setting where $a > 0$, $D_{\mu,[0,a)}(z)$ is defined as $D_\mu(1,z)$ but with $z \in [0,a)$. It defines another increasing homeomorphism from $[0,a)$ onto $[0, D_{\mu,[0,a)}(a^-))$.

The function[3]

$$D_\mu(c,z) = H_\mu(1/z^2) \quad (8)$$

plays a key role in the computation of the rectangular free convolution (see [9, Th. 3.13] or [8, Intro.]). Indeed, the *rectangular R-transform with ratio c* of μ, defined to be

$$C_\mu(z) = U\left(\frac{z}{H_\mu^{-1}(z)} - 1\right),$$

where $U(z) = \frac{-c-1+\left[(c+1)^2+4cz\right]^{1/2}}{2c}$ for $c > 0$ and $U(z) = z$ for $c = 0$, is the analogue of the logarithm of the Fourier transform for the rectangular free convolution with ratio c, since for all probability measures μ_1, μ_2, $\mu_1 \boxplus_c \mu_2$ is characterized by the fact that

$$C_{\mu_1 \boxplus_c \mu_2}(z) = C_{\mu_1}(z) + C_{\mu_2}(z).$$

The coefficients of the series expansion of $C_\mu(z)$ are the so called *rectangular free cumulants with ratio c* of μ (see [7, Eq. (4.1)]).

2.5. **Extensions.**

Remark 2.13 (Phase transition in non-extreme eigenvalues)**.** Theorem 2.1 can easily be adapted to describe the phase transition in the eigenvalues of $X_n + P_n$ which fall in the "holes" of the support of μ_X. Consider $c < d$ such that almost surely, for n large enough, X_n has no eigenvalue in (c,d) (which implies that (c,d) does not intersect the support of μ_X and that G_{μ_X} induces a decreasing homeomorphism $G_{\mu_X,(c,d)}$ from (c,d) onto $(G_{\mu_X}(d^-), G_{\mu_X}(c^+))$). Then almost surely, for n large enough, $X_n + P_n$ has no eigenvalue in (c,d), except if some of the $1/\theta_i$'s are in $(G_{\mu_X}(d^-), G_{\mu_X}(c^+))$, in which case, every such θ_i gives rise to an eigenvalue with almost sure limit $G_{\mu_X,(c,d)}^{-1}(1/\theta_i)$ as n tends to infinity.

Remark 2.14 (Isolated eigenvalues of X_n outside the support of μ_X)**.** When X_n itself has isolated eigenvalues in the sense that the limit a of its smallest eigenvalue and the limit b of its largest eigenvalue are out of the support of μ_X, the phase transition occurs

[3]The functions H_μ and C_μ are often defined only for μ a symmetric measure, but in [10, Sect. 2], the theory is adapted to non-symmetric measures.

at the same values as in Theorem 2.1. Indeed, everything, in the proof, still works, except (22). It follows that for each $i \in \{1, \ldots, s\}$,

$$\lim_{n\to\infty} \lambda_i(X_n + P_n) = G_{\mu_X}^{-1}(1/\theta_i) \qquad \text{if } 1/\theta_i < G_{\mu_X}(b^+),$$
$$\limsup_{n\to\infty} \lambda_i(X_n + P_n) \le b \qquad \text{in the other case,}$$

and that for each $i \in \{s+1, \ldots, r\}$,

$$\lim_{n\to\infty} \lambda_{n-r+i}(X_n + P_n) = G_{\mu_X}^{-1}(1/\theta_i) \qquad \text{if } 1/\theta_i > G_{\mu_X}(a^-),$$
$$\liminf \lambda_{n-r+i}(X_n + P_n) \ge a \qquad \text{in the other case.}$$

Remark 2.15 (Random matrices with Haar-like eigenvectors)**.** Let G be an $n\times m$ Gaussian random matrix with independent real (or complex) entries that are normally distributed with mean 0 and variance 1. The eigenvectors U of the matrix $X = GG^*/m$, will be Haar distributed. Informally speaking, when G is a Gaussian-like matrix in the sense that its entries are i.i.d. with mean zero and variance one, then upon placing adequate restrictions on the higher order moments, we label the eigenvectors of X as being Haar-like. Formally speaking, following the development in [30, 31, 32], when U is Haar-like, then for non-random unit norm vector x_n, the vector $U^* x_n$ will be close to uniformly distributed on the unit real (or complex) sphere. Since our proofs rely heavily on the properties of unit norm vectors uniformly distributed on the n-sphere, they can be easily adapted to the setting where the unit norm vectors are close to uniformly distributed. Hence, we assert without proof, the applicability of our results to the setting where X or P or both have independent Haar-like distributed eigenvectors.

Remark 2.16 (Setting where eigenvalues of P_n are not fixed)**.** Suppose that P_n is a random matrix independent of X_n, with exactly r non-zero eigenvalues values given by $\theta_1^{(n)}, \ldots, \theta_r^{(n)}$. Let $\theta_i^{(n)} \xrightarrow{\text{a.s.}} \theta_i$ as $n \longrightarrow \infty$. Using [20, Cor. 6.3.8] as in Section 6.2, one can easily see that our results will also apply in this case.

The analogues of Remarks 2.13, 2.14, 2.15 and 2.16 for the multiplicative and rectangular settings also hold here.

3. Examples

We now illustrate our results with some concrete computations. The key to applying our results lies in being able to compute the G, T or D transforms of the spectral measure μ_X and their associated functional inverses. In what follows, we focus on settings where the transforms and their inverses can be expressed in closed form. In settings where the transforms are algebraic so that they can be represented as solutions of polynomial equations, the techniques and software developed in [29] can be utilized. In more complicated settings, one will have to resort to numerical techniques.

3.1. **Random Gaussian matrices and the square additive case.** Let X_n be an $n\times n$ symmetric (or Hermitian) matrix with independent, zero mean, normally distributed entries with variance σ^2/n on the diagonal and $\sigma^2/(2n)$ on the off diagonal. It is known that the spectral measure of X_n converges to the famous semi-circle distribution with density

$$\mathrm{d}\mu_X(x) = \frac{\sqrt{4\sigma^2 - x^2}}{2\sigma^2\pi}\mathrm{d}x \qquad \text{for } x \in [-2\sigma, 2\sigma].$$

It is known that the extreme eigenvalues converge to the bounds of this support [1]. Associated with the spectral measure, we have

$$G_{\mu_X}(z) = \frac{z - \operatorname{sgn}(z)\sqrt{z^2 - 4\sigma^2}}{2\sigma^2}, \qquad \text{for } z \in (-\infty, -2\sigma) \cup (2\sigma, +\infty),$$

$G_{\mathrm{sc}_\sigma}(\pm 2\sigma) = \pm\sigma$ and $G_{\mathrm{sc}_\sigma}^{-1}(1/\theta) = \theta + \frac{\sigma^2}{\theta}$.

Thus for P_n with r non-zero eigenvalues $\theta_1 \ge \cdots \ge \theta_s > 0 > \theta_{s+1} \ge \cdots \ge \theta_r$, for any fixed $i \in \{1, \dots, s\}$ (resp. $i \in \{s+1, \dots, r\}$), by Theorem 2.1, we have

$$\lambda_i(X_n + P_n) \text{ (resp. } \lambda_{n+1-i}(X_n + P_n)) \xrightarrow{\text{a.s.}} \begin{cases} \theta_i + \frac{\sigma^2}{\theta_i} & \text{if } |\theta_i| > \sigma \\ 2\sigma(\text{ resp. } -2\sigma) & \text{otherwise,} \end{cases} \tag{9}$$

as $n \longrightarrow \infty$. This result has already been established in [17] for the symmetric case and in [28] for the Hermitian case. Remark 2.15 explains why our results hold for Wigner matrices of the sort considered in [28, 17]. Now, onto the eigenvectors.

In the setting where $r = 1$ and $\theta := \theta_1 > 0$, let x_n be an eigenvector of $X_n + P_n$ associated with its largest eigenvalue. By Theorem 2.2 we have

$$\langle x_n, \ker(\theta I_n - P_n)\rangle^2 \xrightarrow{\text{a.s.}} \begin{cases} 1 - \frac{\sigma^2}{\theta^2} & \text{if } \theta \ge \sigma, \\ 0 & \text{if } \theta < \sigma. \end{cases} \tag{10}$$

Figure 2 illustrates the agreement between theory and experiment and the asymptotic nature of the eigenvector phase transition result.

3.2. **Random sample covariance matrices and the multiplicative case.** Let G_n be an $n \times m$ real (or complex) matrix with independent, zero mean, normally distributed entries with variance 1. Let $X_n = G_n G_n^*/m$. It is known [23, 18] that, as $n, m \longrightarrow \infty$ with $n/m \to c > 0$, the spectral measure of X_n converges to the famous Marčenko-Pastur distribution with density

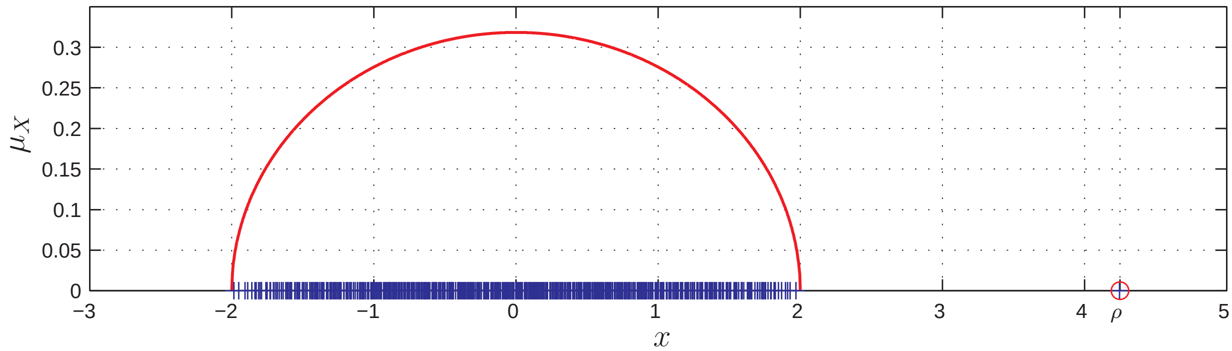
$$\mathrm{d}\mu_X(x) := \frac{1}{2\pi c x}\sqrt{(b-x)(x-a)}\mathbb{1}_{[a,b]}(x)\mathrm{d}x + \max\left(0, 1 - \frac{1}{c}\right)\delta_0$$

where $a = (1 - \sqrt{c})^2$ and $b = (1 + \sqrt{c})^2$ are the end points of the support of μ_X. It is known that the extreme eigenvalues converge to the bounds of this support.

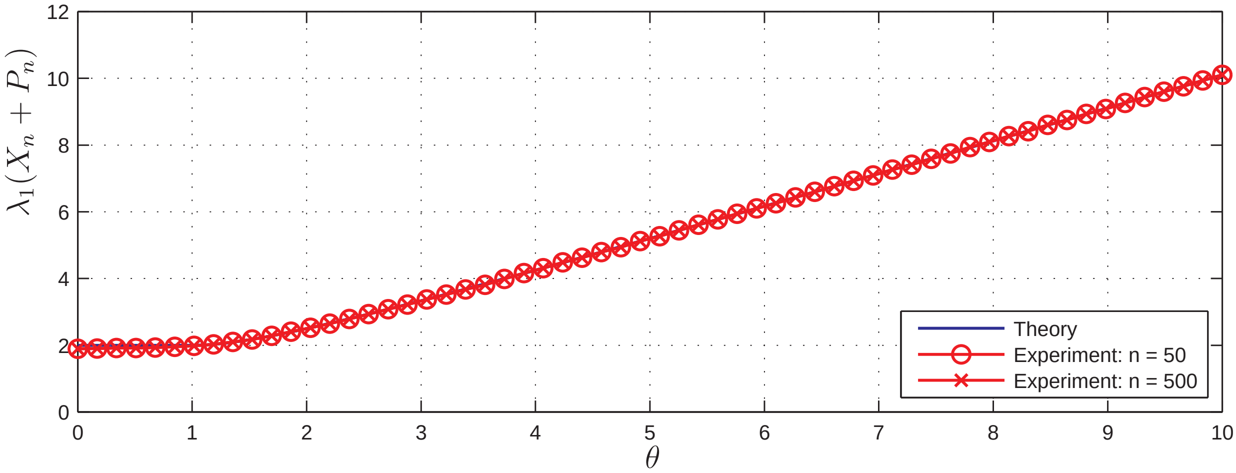
Associated with this spectral measure we have [4]

$$\begin{aligned} T_{\mu_X}^{-1} &= \frac{(z+1)(cz+1)}{z}, \\ T_{\mu_X}(z) &= \frac{z - c - 1 + \varepsilon\sqrt{(z-c-1)^2 - 4c}}{2c} \quad (\varepsilon = 1 \text{ if } z \le a \text{ and } \varepsilon = -1 \text{ if } z \ge b), \\ T_{\mu_X}(b^+) &= 1/\sqrt{c}, \quad T_{\mu_X}(a^-) = -1/\sqrt{c}. \end{aligned}$$

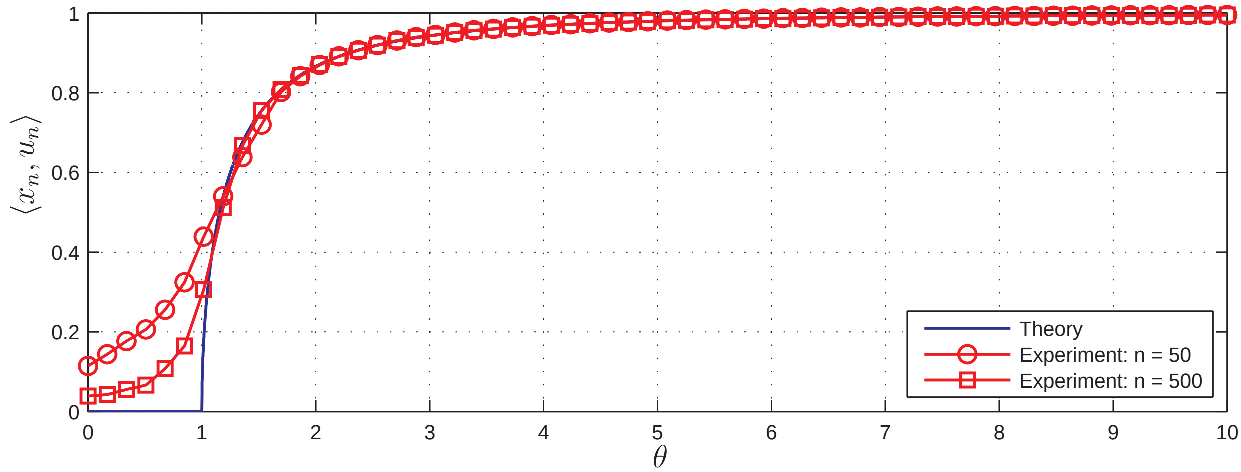
[4]The most easy way to make these computations is to note that $S_{\mu_c}(z) = (1+cz)^{-1}$ (see the proof of Theorem 1 of [8]) and to use (6).



(a) Typical eigenvalues of $X_n + P_n$ with $\theta = 4$, and $n = 500$ when μ_{X_n} converges to the semi-circle distribution.



(b) The empirical averages for $n = 50$ and $n = 500$ versus the theoretical prediction in (9).



(c) The empirical averages for $n = 50$ and $n = 500$ versus the theoretical prediction in (10).

FIGURE 2. The theoretical predictions in Section 3.1 for the eigenvalues and eigenvectors of $X_n + P_n$ are verified using empirical averages of the relevant quantities over 400 Monte-Carlo simulations. Here we have $P_n = \theta u_n u_n^*$ and u_n is a unit norm vector.

When $c > 1$, there is an atom at zero so that the smallest eigenvalue of X equals zero. For simplicity, let us first consider the setting when $c < 1$ so that the extreme eigenvalues of X tend almost surely to a and b. Thus for P_n with r non-zero eigenvalues $\theta_1 \geq \cdots \geq \theta_s > 0 > \theta_{s+1} \geq \cdots \geq \theta_r$, with $l_i := \theta_i - 1$, for any fixed $i \in \{1, \ldots, s\}$ (resp. $i \in \{s+1, \ldots, r\}$) and $c < 1$, by Theorem 2.5, we have

$$\lambda_i(X_n(I_n + P_n)) \text{ (resp. } \lambda_{n+1-i}(X_n(I_n + P_n))) \xrightarrow{\text{a.s.}} \begin{cases} l_i \left(1 + \frac{c}{l_i - 1}\right) & \text{if } |l_i - 1| > \sqrt{c} \\ b \text{ (resp. } a) & \text{otherwise,} \end{cases}$$

as $n \longrightarrow \infty$. Consider the matrix $S_n = (I_n + P_n)^{1/2} X_n (I_n + P_n)^{1/2}$ which can be interpreted as a Wishart distributed sample covariance matrix with "spiked" covariance $I_n + P_n$. By Remark 2.8, the above result applies for the eigenvalues of S_n as well. This result for the largest eigenvalue of spiked sample covariance matrices was established[5] in [3, 27] and for the extreme eigenvalues in [4]. Now, onto the eigenvectors.

In the setting where $r = 1$, let us denote $l_1 = \theta_1 - 1$ by l, and let be x_n a unit norm eigenvector of $X_n(I + P_n)$ associated with its largest (or smallest, according to whether $l > 1$ or $l < 1$) eigenvalue. By Theorem 2.7, we have

$$\langle x_n, \ker(lI_n - \Sigma)\rangle^2 \xrightarrow{\text{a.s.}} \begin{cases} \frac{(l-1)^2 - c}{(l-1)[c(l+1) + l - 1]} & \text{if } |l - 1| \geq \sqrt{c}, \\ 0 & \text{if } |l - 1| < \sqrt{c}. \end{cases} \tag{11}$$

Let y_n be a unit eigenvector of $S_n = (I_n + P_n)^{1/2} X_n (I_n + P_n)^{1/2}$ associated with its largest (or smallest, according to whether $l > 1$ or $l < 1$) eigenvalue. Then, by Theorem 2.7 and Remark 2.8, we have

$$\langle y_n, \ker(lI_n - \Sigma)\rangle^2 \xrightarrow{\text{a.s.}} \begin{cases} \frac{1 - \frac{c}{(l-1)^2}}{1 + \frac{c}{l-1}} & \text{if } |l - 1| \geq \sqrt{c}, \\ 0 & \text{if } |l - 1| < \sqrt{c}. \end{cases} \tag{12}$$

This result has been established in [27]. We generalize it to the case where $\Sigma \leq I_n$.

3.3. **The rectangular case.**

3.3.1. *Gaussian rectangular random matrices with non-zero mean.* Let G_n be an $n \times m$ real (or complex) matrix with independent, zero mean, normally distributed entries with variance $1/m$. It is known [23, 18] that, as $n, m \longrightarrow \infty$ with $n/m \in [0, 1]$, the spectral measure of the singular values of X_n converges to the distribution with density

$$\mathrm{d}\mu_X(x) = \frac{\sqrt{4c - (x^2 - 1 - c)^2}}{\pi c x} \mathbb{1}_{(a,b)}(x) dx,$$

where $a = 1 - \sqrt{c}$ and $b = 1 + \sqrt{c}$ are the end points of the support of μ_X. It is known that the extreme eigenvalues converge to the bounds of this support.

[5]The model considered in [3], which takes into consideration the sample mean, is slightly different from the one of this section, but our results also apply to this model.

Associated with this singular measure, we have, by an application of the result in [7, Sect. 4.1] and the relationship in (8),

$$D_{\mu_X}^{-1}(c,z) = \sqrt{\tfrac{(z+1)(cz+1)}{z}},$$
$$D_{\mu_X}(c,z) = \tfrac{z^2-(c+1)-\sqrt{(z^2-(c+1))^2-4c}}{2c}, \quad D_{\mu_X}(c,b^+) = \tfrac{1}{\sqrt{c}}.$$

Thus for any $n\times m$ deterministic matrix $P_{n,m}$ with r non-zero singular values $\theta_1 \geq \ldots \geq \theta_r$ (r independent of n,m), for any fixed $i\geq 1$, by Theorem 2.9, we have

$$\sigma_i(X_n+P_n) \xrightarrow{\text{a.s.}} \begin{cases} \sqrt{\frac{(1+\theta_i^2)(c+\theta_i^2)}{\theta_i^2}} & \text{if } \theta_i > c^{1/4} \\ 1+\sqrt{c} & \text{otherwise.}\end{cases} \tag{13}$$

as $n\longrightarrow\infty$. This is a new result. Now, onto the singular vectors.

In the setting where $r=1$, let u_n be a unit norm eigenvector of $\tilde{X}_{n,m}\tilde{X}_{n,m}^*$ associated to its largest eigenvalue, then, by Theorems 2.10 and 2.12, we have

$$\langle u_n, \ker(\theta^2 I_n - P_{n,m}P_{n,m}^*)\rangle^2 \xrightarrow{\text{a.s.}} \begin{cases} 1-\frac{c(1+\theta^2)}{\theta^2(\theta^2+c)} & \text{if } \theta\geq c^{1/4}, \\ 0 & \text{otherwise.}\end{cases} \tag{14}$$

The phase transitions for the eigenvectors of $\tilde{X}_{n,m}^*\tilde{X}_{n,m}$ or for the pairs of singular vectors of $\tilde{X}_{n,m}$ can be similarly computed.

3.3.2. *Square Haar unitary matrices.* Let X_n be Haar distributed unitary (or orthogonal) random matrix. All of its singular values are equal to one so that it has the spectral measure

$$\mathrm{d}\mu_X(x) = \delta_1,$$

with $a=b=1$ being the end points of the support of μ_X.

Associated with this spectral measure, we have

$$D_{\mu_X}(1,z) = \frac{z^2}{(z^2-1)^2} \qquad \text{for } z\geq 0,\ z\neq 1,$$

thus for all $\theta>0$,

$$D_{\mu_X}^{-1}(1,1/\theta^2) = \frac{\theta+\sqrt{\theta^2+4}}{2} \quad \text{and} \quad D_{\mu_X,[0,1)}^{-1}(1/\theta^2) = \frac{-\theta+\sqrt{\theta^2+4}}{2}.$$

Thus for any $n\times n$, rank r deterministic matrix P_n with r non-zero singular values $\theta_1\geq\cdots\geq\theta_r$ where neither r, nor the θ_i's depend on n, for any fixed $i=1,\ldots,r$, by Theorem 2.9 we have

$$\sigma_i(X_n+P_n) \xrightarrow{\text{a.s.}} \frac{\theta+\sqrt{\theta^2+4}}{2} \quad \text{and} \quad \sigma_{n+1-i}(X_n+P_n) \xrightarrow{\text{a.s.}} \frac{-\theta+\sqrt{\theta^2+4}}{2}$$

while for any fixed $i\geq r+1$, both $\sigma_i(X_n+P_n)$ and $\sigma_{n+1-i}(X_n+P_n) \xrightarrow{\text{a.s.}} 1$.

4. Sketches of the proofs

Let us explain briefly how Theorems 2.1, 2.2 and 2.3, about extreme eigenvalues and associated eigenvectors of $X+P$, can be proved (the index n in X_n and P_n has been suppressed for brevity). Similar arguments can be used for the multiplicative and rectangular models considered in this paper.

Consider the setting where $r=1$, so that $P=\theta\, uu^*$, with u being a unit norm column vector. Since either X of P is supposed to be invariant, in law, under unitary (or orthogonal) conjugation, one can, without loss of generality, suppose that $X=\operatorname{diag}(\lambda_1,\ldots,\lambda_n)$ and that u is uniformly distributed on the unit n-sphere.

4.1. **Largest eigenvalue phase transition.** For any $z\in\mathbb{C}$ such that z is not an eigenvalue of X, we have

$$z-(X+P)=(z-X)\times(I-(z-X)^{-1}P),$$

so that z is an eigenvalue of $X+P$ if and only if 1 is an eigenvalue of $(z-X)^{-1}P$. But $(z-X)^{-1}P=(z-X)^{-1}\theta\, uu^*$ has rank one, so its only non-null eigenvalue is its trace, which is equal to $\theta G_{\mu_n}(z)$, where μ_n is a "weighted" spectral measure of X, defined by

$$\mu_n=\sum_{k=1}^n|u_k|^2\delta_{\lambda_k}\qquad\text{(the } u_k\text{'s are the coordinates of } u). \tag{15}$$

Thus any z out of the spectrum of X is an eigenvalue of $X+P$ if and only if

$$G_{\mu_n}(z)=\frac{1}{\theta}. \tag{16}$$

Equation (16) describes the relationship between the eigenvalues of $X+P$ and the eigenvalues of X, as illustrated in Figure 3. Here, u is a random vector with uniform distribution on the unit n-sphere. Hence, for large n, we have that $|u_k|^2\approx\frac{1}{n}$ with high probability, so that we have $\mu_n\approx\mu_X$ and consequently $G_{\mu_n}(z)\approx G_{\mu_X}(z)$. Inverting equation (16) after substituting these approximations yields the location of the largest eigenvalue to be $G_{\mu_X}^{-1}(1/\theta)$ as in Theorem 2.1.

The phase transition for the extreme eigenvalues emerges because under our assumption that the limiting probability measure μ_X is compactly supported on $[a,b]$, the Cauchy transform G_{μ_X} is defined *outside* $[a,b]$ and unlike what happens for G_{μ_n}, we do not always have $G_{\mu_X}(b^+)=+\infty$. Consequently, in settings of the sort depicted in Figure 4, when $1/\theta<G_{\mu_X}(b^+)$, we have that $\lambda_1(\widetilde{X})\approx G_{\mu_X}^{-1}(1/\theta)$ as before. However, when $1/\theta\geq G_{\mu_X}(b^+)$ then the phase transition manifests and $\lambda_1(\widetilde{X})\approx\lambda_1(X)=b$.

An extension of these arguments for fixed $r>1$ yields the general result and constitutes the most transparent justification, as sought by the authors in [3], for the emergence of this phase transition phenomenon in such perturbed random matrix models. We rely on concentration inequalities to make the arguments rigorous.

4.2. **Eigenvectors phase transition.** Let x be a unit eigenvector of $X+P$ associated with the eigenvalue z such that $G_{\mu_n}(z)=1/\theta$. From $(X+P)x=zx$, we deduce

$$\begin{aligned}(z-X)x&=\theta uu^*x\\&=(\theta u^*x).u\quad\text{(because } u^*x \text{ is a scalar)},\end{aligned}$$

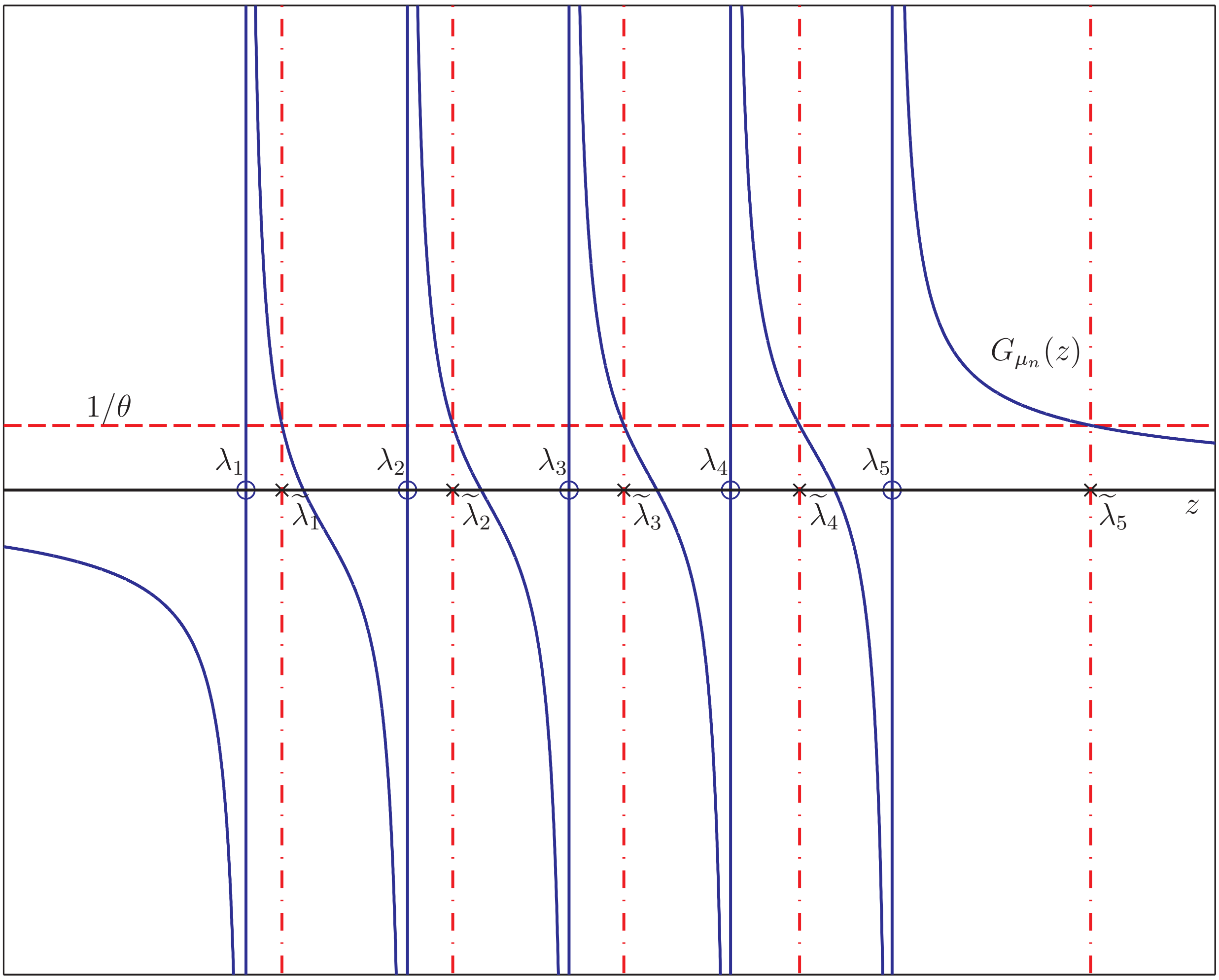


FIGURE 3. Graph illustrating how the relationship between the eigenvalues of $\widetilde{X} = X + \theta uu^*$ and the eigenvalues of X for some θ and unit norm $n \times 1$ vector u is implicitly described by (16). Here, when $n = 5$, $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of X and $\widetilde{\lambda_1}, \ldots, \widetilde{\lambda_n}$ are the eigenvalues of $\widetilde{X}$. The function $G_{\mu_n}(z)$ plotted is the Cauchy or G-transform of the probability measure μ_n given by (15).

implying that x is proportional to $(z - X)^{-1}u$. Hence, since x has norm one and $z - X$ is Hermitian,

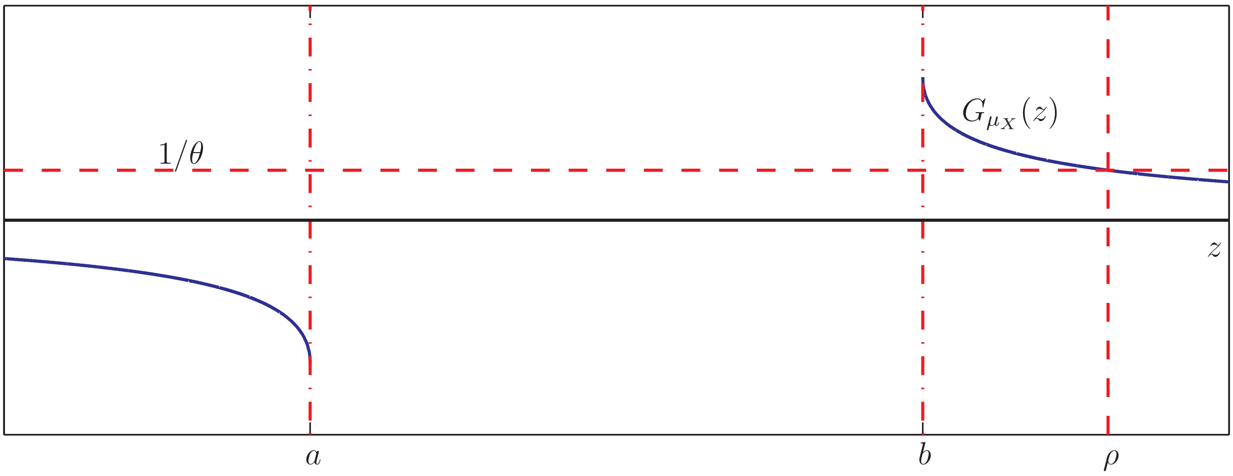
$$x = \frac{(z - X)^{-1}u}{\sqrt{u^*(z - X)^{-2}u}} \tag{17}$$

and

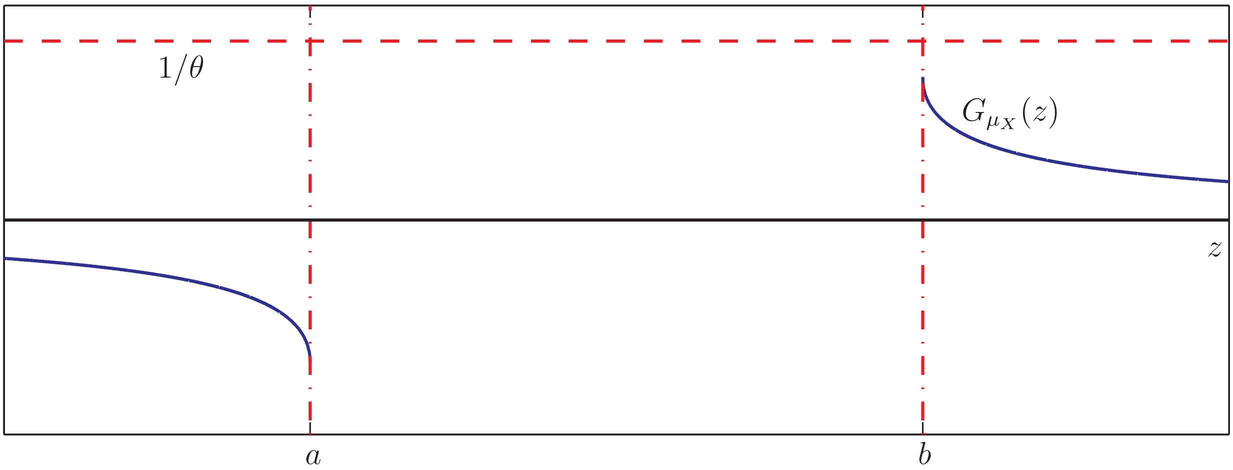
$$\langle x, \ker(\theta I - P)\rangle^2 = |u^* x|^2 = \frac{(u^*(z - X)^{-1}u)^2}{u^*(z - X)^{-2}u} = \frac{G_{\mu_n}(z)^2}{\int \frac{\mathrm{d}\mu_n(t)}{(z-t)^2}} = \frac{1}{\theta^2 \int \frac{\mathrm{d}\mu_n(t)}{(z-t)^2}}. \tag{18}$$

Again, since $\mu_n \approx \mu_X$, if $z \approx \rho > b$, then $\int \frac{\mathrm{d}\mu_n(t)}{(z-t)^2} \approx \int \frac{\mathrm{d}\mu_X(t)}{(\rho-t)^2} < \infty$ and we get

$$\langle x, \ker(\theta I - P)\rangle^2 \xrightarrow{\text{a.s.}} \frac{1}{\theta^2 \int \frac{\mathrm{d}\mu_X(t)}{(\rho-t)^2}} = \frac{-1}{\theta^2 G'_{\mu_X}(\rho)} > 0.$$



(a) When $1/\theta < G_{\mu_X}(b)$, $\lambda_1(X+P) \to \rho = G_{\mu_X}^{-1}(\frac{1}{\theta})$.



(b) When $1/\theta > G_{\mu_X}(b)$, $\lambda_1(X+P) \to b$.

FIGURE 4. Graphical illustration of why a phase transition occurs for the largest eigenvalue of $X+P$. The limiting probability measure μ_X for the eigenvalues of X is supported on $[a,b]$. The Cauchy or G-transform of μ_X is plotted outside the support. Following the argument in Figure 3 and Section 4, the location of the largest eigenvalue is determined by the relationship between $1/\theta$ and $G_{\mu_X}(z)$ as shown. A similar argument can be used to illustrate the phase transition for the smallest eigenvalues and for the multiplicative and rectangular models considered in this paper.

On the other hand, if $z \approx b$ and $G'_{\mu_X}(b^+) = \int \frac{\mathrm{d}\mu_X(t)}{(b-t)^2} = \infty$, then

$$\langle x, \ker(\theta I - P)\rangle^2 \xrightarrow{\text{a.s.}} 0.$$

Again, rigorous arguments rely on concentration inequalities.

5. The exact master equations

In this section, we shall give r-dimensional analogues of the key-formulas (16) and (17) given in the sketches of the proofs.

5.1. Eigenvalues and eigenvectors of $X + P$.

Proposition 5.1. *Let X be an $N \times N$ matrix and $P = V\Theta V^*$, with Θ an $R \times R$ matrix and V an $N \times R$ matrix which columns are orthonormal (i.e. such that $V^*V = I_R$). Let z be a complex number out of the spectrum of X. Let us define the (possibly null) spaces*

$$E_z = \ker[zI_N - (X + P)], \quad F_z = \ker[I_R - V^*(z - X)^{-1}V\Theta].$$

a) Then the mappings

$$\begin{array}{rccl} \varphi_z : & F_z & \to & E_z \\ & v & \mapsto & (z - X)^{-1}V\Theta v \end{array} \qquad \text{and} \qquad \begin{array}{rccl} \psi_z : & E_z & \to & F_z \\ & w & \mapsto & V^*w \end{array}$$

are linear isomorphisms, which are inverses one of each other. Moreover, for all v, w, $\|\varphi_z(v)\| \geq \|v\|$ and $\|\psi_z(w)\| \leq \|w\|$.

b) Let us denote N by n and R by r, lets us suppose that $X = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $\Theta = \operatorname{diag}(\theta_1, \ldots, \theta_r)$ and let us denote the entries of V by $v_{k,l}$ $(1 \leq k \leq n,\ 1 \leq l \leq r)$. Then firstly, any $z \notin \{\lambda_1, \ldots, \lambda_n\}$ is an eigenvalue of $X+P$ if and only if the $r \times r$ matrix

$$I_r - V^*(z - X)^{-1}V\Theta$$

is not invertible and secondly, for all $i, j = 1, \ldots, r$, the (i, j)-th entry of this matrix is

$$\mathbb{1}_{i=j} - \theta_j G_{\mu_{i,j}}(z),$$

where $\mu_{i,j}$ is the complex measure defined by

$$\mu_{i,j} = \sum_{k=1}^{n} \overline{v}_{k,i} v_{k,j} \delta_{\lambda_k}$$

and $G_{\mu_{i,j}}$ is the Cauchy or G-transform of $\mu_{i,j}$.

Proof. Part b) follows directly from a) and from a straightforward computation of the (i, j)-th entry of $V^*(z - X)^{-1}V\Theta$. Part a) is proved, for example, in [2, Th. 2.3] (the assertions about the norms follow from the fact that for all $w \in E_z$, $\|\psi_z(w)\|^2 = |w_1|^2 + \cdots + |w_r|^2$, where $w_1, \ldots, w_n$ are the coordinates of w on an orthonormal basis having the columns of V for first r vectors). $\square$

5.2. Eigenvalues and eigenvectors of $X(I + P)$.

The following proposition can be proved in the same manner as Proposition 5.1.

Proposition 5.2. *Let X be an $n \times n$ matrix and $P = V\Theta V^*$, with Θ an $r \times r$ matrix and V an $n \times r$ matrix which columns are orthogonal (i.e. such that $V^*V = I_r$). Let z be an eigenvalue z of $X(I + P)$ which is not eigenvalue of X. Let us define the (possibly null) spaces*

$$E_z = \ker[zI_n - X(I + P)], \quad F_z = \ker[I_r - V^*(z - X)^{-1}XV\Theta].$$

a) Then the mappings

$$\begin{array}{rcl} \varphi_z : F_z & \to & E_z \\ v & \mapsto & (z-X)^{-1}V\Theta v \end{array} \qquad \text{and} \qquad \begin{array}{rcl} \psi_z : E_z & \to & F_z \\ w & \mapsto & V^* w \end{array}$$

are linear isomorphisms, which are inverses one of each other. Moreover, for all v, w, $\|\varphi_z(v)\| \geq \|v\|$ *and* $\|\psi_z(w)\| \leq \|w\|$.

b) Let us suppose $X = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ *and* $\Theta = \operatorname{diag}(\theta_1, \ldots, \theta_r)$ *and let us denote the entries of* V *by* $v_{k,l}$ $(1 \leq k \leq n,\ 1 \leq l \leq r)$. *Then firstly, any* $z \notin \{\lambda_1, \ldots, \lambda_n\}$ *is an eigenvalue of* $X(I+P)$ *if and only if the* $r \times r$ *matrix*

$$I_r - V^*(z-X)^{-1}XV\Theta$$

is not invertible, and secondly, for all $i, j = 1, \ldots, r$, *the* (i,j)*-th entry of this matrix is*

$$\mathbb{1}_{i=j} - \theta_j T_{\mu_{i,j}}(z),$$

with

$$\mu_{i,j} = \sum_{k=1}^{n} \overline{v}_{k,i} v_{k,j} \delta_{\lambda_k},$$

and $T_{\mu_{i,j}}(z)$ *is the T-transform of* $\mu_{i,j}$.

5.3. **Singular values and singular vectors of $X+P$.** Firstly, let us state the following well known theorem [20, Th. 7.3.7], which relates singular values and singular vectors of non-Hermitian matrices with eigenvalues and eigenvectors of Hermitian matrices (see Section 2.3.2 for the definition of singular vectors).

Theorem 5.3. *For* $\sigma_1, \ldots, \sigma_n$ *nonnegative real numbers, the singular values of* X *are* $\sigma_1, \ldots, \sigma_n$ *if and only if the* $n+m$ *eigenvalues of*

$$\begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix} \tag{19}$$

are $\sigma_1, \ldots, \sigma_n, -\sigma_1, \ldots, -\sigma_n$ *and* $m-n$ *additional zeros. Moreover, for* σ *a nonnegative number, a pair* (u,v) *of unit-length vectors is a pair of singular vectors of* X *associated with the singular value* σ *if and only if* $\begin{bmatrix} u \\ v \end{bmatrix}$ *is an eigenvector of the matrix of* (19) *for the eigenvalue* σ.

The following proposition adapts, via Theorem 5.3, the master equation of Proposition 5.1 to the rectangular setting.

Proposition 5.4. *Let us now denote* $N = n+m$, *with* $1 \leq n \leq m$ *and* $R = 2r$, *let us suppose that there exists* $\sigma = (\sigma_1, \ldots, \sigma_n), \theta = (\theta_1, \ldots, \theta_r)$ *such that for* $D = \operatorname{diag}(\sigma)$ *and* $\Omega = \operatorname{diag}(\theta)$, *we have*

$$X = \begin{bmatrix} 0_{n,n} & D & 0_{n,m-n} \\ D & 0_{n,n} & 0_{n,m-n} \\ 0_{m-n,n} & 0_{m-n,n} & 0_{m-n,m-n} \end{bmatrix}, \Theta = \begin{bmatrix} 0_{r,r} & \Omega \\ \Omega & 0_{r,r} \end{bmatrix}, V = \begin{bmatrix} U & 0_{n,r} \\ 0_{n,r} & K \\ 0_{m-n,r} & L \end{bmatrix}.$$

Then the $2r \times 2r$ matrix $V^*(z-X)^{-1}V\Theta$ can be written with four $r \times r$ blocs

$$\begin{bmatrix} U^* \frac{D}{z^2-D^2} K\Omega & U^* \frac{z}{z^2-D^2} U\Omega \\ \\ K^* \frac{z}{z^2-D^2} K\Omega + \frac{1}{z} L^* L\Omega & K^* \frac{D}{z^2-D^2} U\Omega \end{bmatrix}. \tag{20}$$

(the entries of this matrix can easily be interpreted as in b) of Proposition 5.1, i.e. as integrals of complex measures with support $\{\sigma_1, \ldots, \sigma_n\}$ or $\{\sigma_1, \ldots, \sigma_n, 0\}$.

The proof of the proposition relies on a straightforward computation, inverting $z - X$ via the following elementary lemma, which we shall refer to several times in the paper.

Lemma 5.5. *Let c, d be diagonal $r \times r$ matrices with positive diagonal entries. Then one has*

$$\begin{bmatrix} 0 & c \\ d & 0 \end{bmatrix} = \begin{bmatrix} \sqrt{c} & -\sqrt{c} \\ \sqrt{d} & \sqrt{d} \end{bmatrix} \begin{bmatrix} \sqrt{cd} & 0 \\ 0 & -\sqrt{cd} \end{bmatrix} \begin{bmatrix} \sqrt{c} & -\sqrt{c} \\ \sqrt{d} & \sqrt{d} \end{bmatrix}^{-1} \text{and} \begin{bmatrix} \sqrt{c} & -\sqrt{c} \\ \sqrt{d} & \sqrt{d} \end{bmatrix}^{-1} = \frac{1}{2} \begin{bmatrix} c^{-\frac{1}{2}} & d^{-\frac{1}{2}} \\ -c^{-\frac{1}{2}} & d^{-\frac{1}{2}} \end{bmatrix}.$$

In the particular case where $c = d$, this can be reduced to

$$\begin{bmatrix} 0 & d \\ d & 0 \end{bmatrix} = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} d & 0 \\ 0 & -d \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right), \quad \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right)^2 = I_2. \tag{21}$$

6. Proof of Theorem 2.1 for the extreme eigenvalues of $X_n + P_n$

6.1. **First step: Setting where θ_i's are pairwise distinct.** Note first that by the definition of $\boxplus$ (see also [1, Cor. 5.4.11]), the random probability measure

$$\frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k(X_n+P_n)}$$

converges almost surely to $\mu_X \boxplus \delta_0 = \mu_X$, so for any i fixed independently of n,

$$\liminf_{n\to\infty} \lambda_i(X_n + P_n) \geq b \quad \text{ and } \quad \limsup_{n\to\infty} \lambda_{n-i}(X_n + P_n) \leq a. \tag{22}$$

Let us write, for each n,

$$X_n = U_X^{(n)} \operatorname{diag}(\lambda_1^{(n)}, \ldots, \lambda_n^{(n)}) U_X^{(n)*}, \qquad P_n = U_P^{(n)} \operatorname{diag}(\theta_1, \ldots, \theta_r, 0, \ldots, 0) U_P^{(n)*}.$$

The spectrum of $X_n + P_n$ is the one of

$$\operatorname{diag}(\lambda_1^{(n)}, \ldots, \lambda_n^{(n)}) + \underbrace{U_X^{(n)*} U_P^{(n)}}_{\text{denoted by } U_n} \operatorname{diag}(\theta_1, \ldots, \theta_r, 0, \ldots, 0) U_P^{(n)*} U_X^{(n)},$$

and since either X_n or P_n is invariant in law by conjugation by orthogonal (or unitary) matrices, U_n is Haar-distributed and independent of $\lambda^{(n)} := (\lambda_1^{(n)}, \ldots, \lambda_n^{(n)})$ (see the first paragraph of the proof of [19, Th. 4.3.5] for more details). Let us denote by $[u_{i,j}^{(n)}]_{i,j=1}^n$ the entries of U_n.

Since $\lambda_1(X_n) \xrightarrow{\text{a.s.}} b$ and $\lambda_n(X_n) \xrightarrow{\text{a.s.}} a$, we can focus on the eigenvalues of $X_n + P_n$ which are out of $[\lambda_n(X_n), \lambda_1(X_n)]$. By Proposition 5.1 b), these eigenvalues are the numbers z out of $[\lambda_n(X_n), \lambda_1(X_n)]$ such that the $r \times r$ matrix

$$M(n, z) := I_r - [\theta_j G_{\mu_{i,j}^{(n)}}(z)]_{i,j=1}^r$$

is not invertible, where for all $i, j = 1, \ldots, r$, $\mu_{i,j}^{(n)}$ is the random complex measure defined by

$$\mu_{i,j}^{(n)} = \sum_{k=1}^{n} \overline{u_{k,i}}^{(n)} u_{k,j}^{(n)} \delta_{\lambda_k^{(n)}}.$$

By Proposition 11.3, for all $i \neq j$, the random complex measure $\mu_{i,j}^{(n)}$ converges almost surely weakly to zero and for all i, $\mu_{i,i}^{(n)}$ converges almost surely weakly to μ_X. Thus, by the second statement of Lemma 11.1, for all $\eta > 0$, the matrix-valued function $M(n, \cdot)$ converges to

$$M_{G_{\mu_X}}(\cdot) := \operatorname{diag}(1 - \theta_1 G_{\mu_X}(\cdot), \ldots, 1 - \theta_r G_{\mu_X}(\cdot))$$

uniformly on $\{z \in \mathbb{C}\,;\, \mathrm{d}(z, [a, b]) \geq \eta\}$, almost surely. Now, the conclusion follows exactly from Lemma 11.4, which hypotheses are satisfied (hypotheses a), b), c) follow from the definition of G_{μ_X}, hypothesis d) follows from Proposition 5.1 and from the fact that $X_n + P_n$ is Hermitian and hypothesis e) has been checked above).

6.2. **Second step: Extension to the general case.** We have now to treat the case where the θ_i's are not pairwise distinct. Fix $i_0 \in \{1, \ldots, r\}$ and $\varepsilon > 0$. Assume for example that $i_0 \leq s$ (the other case can be treated in the same way). We denote, for $\theta > 0$,

$$\rho_\theta = \begin{cases} G_{\mu_X}^{-1}(1/\theta) & \text{if } 1/\theta < G_{\mu_X}(b^+), \\ b & \text{in the other case.} \end{cases}$$

Note also that the function G_μ^{-1} is continuous on $(0, G_\mu(b^+))$ and that $\lim_{y \uparrow G_\mu(b^+)} G_\mu^{-1}(y) = b$, hence the function $\theta \longmapsto \rho_\theta$ is continuous on $(0, \infty)$. Hence there is $\eta > 0$ such that for all $\theta > 0$,

$$|\theta - \theta_{i_0}| \leq \eta \Longrightarrow |\rho_\theta - \rho_{\theta_{i_0}}| \leq \varepsilon. \tag{23}$$

Let us consider a family $(\theta'_1 > \cdots > \theta'_r)$, of real numbers such that for all $i = 1, \ldots, r$, $\theta_i \theta'_i > 0$ and

$$\sum_{i=1}^{r} (\theta'_i - \theta_i)^2 \leq \min(\eta^2, \varepsilon^2).$$

It implies that

$$|\rho_{\theta'_{i_0}} - \rho_{\theta_{i_0}}| \leq \varepsilon. \tag{24}$$

Employing the notation in Section 6.1, for each k and each n, we define

$$P'_n = U_P^{(n)} \operatorname{diag}(\theta'_1, \ldots, \theta'_r, 0, \ldots, 0) U_P^{(n)*}.$$

Note that by [20, Cor. 6.3.8], we have, for all n,

$$\sum_{j=1}^{n} (\lambda_j(X_n + P'_n) - \lambda_j(X_n + P_n))^2 \leq \operatorname{Tr}(P'_n - P_n)^2 = \sum_{i=1}^{r} (\theta'_i - \theta_i)^2 \leq \varepsilon^2. \tag{25}$$

Since the theorem can applied to $X_n + P'_n$, it follows that almost surely, for n large enough,

$$|\lambda_{i_0}(X_n + P'_n) - \rho_{\theta'_{i_0}}| \leq \varepsilon. \tag{26}$$

By the triangular inequality, we have

$$|\lambda_{i_0}(X_n + P_n) - \rho_{\theta_{i_0}}| \leq |\lambda_{i_0}(X_n + P_n) - \lambda_{i_0}(X_n + P'_n)| + |\lambda_{i_0}(X_n + P'_n) - \rho_{\theta'_{i_0}}| + |\rho_{\theta'_{i_0}} - \rho_{\theta_{i_0}}|,$$

thus, by (25), (26) and (24), almost surely, for n large enough,

$$|\lambda_{i_0}(X_n + P'_n) - \rho_{\theta_{i_0}}| \leq 3\varepsilon.$$

□

7. Proofs of Theorems 2.2 and 2.3 for the extreme eigenvectors of $X_n + P_n$

We shall use the following elementary lemma several times in the paper.

Lemma 7.1. *Let us fir $r \geq 1$ and let us consider a sequence $M(n)$ of $r \times r$ matrices which converges, as n tends to infinity, to a matrix M. For each n, consider $y_n \in \ker M(n)$ such that $\|y_n\| \leq 1$. Then*

$$\langle y_n, (\ker M)^\perp \rangle \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. Since for all n, $\|y_n\| \leq 1$, one can suppose that y_n converges to a limit y. Then it suffices to prove that $My = 0$, which follows from the fact that $My = \lim M(n)y_n$. □

7.1. Proof of Theorem 2.2.

Let us write, for each n,

$$X_n = U_X^{(n)} \operatorname{diag}(\lambda_1^{(n)}, \ldots, \lambda_n^{(n)}) U_X^{(n)*}, \qquad P_n = U_P^{(n)} \operatorname{diag}(\theta_1, \ldots, \theta_r, 0, \ldots, 0) U_P^{(n)*}.$$

The eigenvectors of $X_n + P_n$ are $U_X^{(n)}$ times the ones of

$$\operatorname{diag}(\lambda_1^{(n)}, \ldots, \lambda_n^{(n)}) + U_X^{(n)*} U_P^{(n)} \operatorname{diag}(\theta_1, \ldots, \theta_r, 0, \ldots, 0) U_P^{(n)*} U_X^{(n)},$$

hence it suffices to prove the result in the case where $X_n = \operatorname{diag}(\lambda_1^{(n)}, \ldots, \lambda_n^{(n)})$ and $P_n = U_n \operatorname{diag}(\theta_1, \ldots, \theta_r, 0, \ldots, 0) U_n^*$, with U_n Haar-distributed. We denote the entries of U_n by $[u_{i,j}^{(n)}]_{i,j=1}^n$, its columns by $C_1^{(n)}, \ldots, C_n^{(n)}$ and the $n \times r$ matrix which columns are respectively $C_1^{(n)}, \ldots, C_r^{(n)}$ by V_n.

Let r_0 be the number of i's such that $\theta_i = \theta_{i_0}$. Up to a reindex of the θ_i's (which is then no longer decreasing, but it is not a problem here), one can suppose that $i_0 = 1$, $\theta_1 = \cdots = \theta_{r_0}$. For each n, $\ker(\theta_1 I_n - P_n)$ is then the linear span of the r_0 first columns of U_n, namely $C_1^{(n)}, \ldots, C_{r_0}^{(n)}$. Since these columns are orthonormal, it suffices to prove that as n tends to infinity,

$$\sum_{j=1}^{r_0} |\langle C_j^{(n)}, x_n \rangle|^2 \quad \text{tends almost surely to} \quad \frac{-1}{\theta_1^2 G'_{\mu_X}(\rho)} = \frac{1}{\theta_1^2 \int \frac{\mathrm{d}\mu_X(t)}{(\rho-t)^2}} \tag{27}$$

and

$$\sum_{j=r_0+1}^{r} |\langle C_j^{(n)}, x_n \rangle|^2 \quad \text{tends almost surely to } 0. \tag{28}$$

Let us introduce, for each n, for all z out of the spectrum of X_n, the $r \times r$ random matrix

$$M(n, z) := I_r - [\theta_j G_{\mu_{i,j}^{(n)}}(z)]_{i,j=1}^r,$$

where, for all $i, j = 1, \ldots, r$, $\mu_{i,j}^{(n)}$ is the random complex measure defined by

$$\mu_{i,j}^{(n)} = \sum_{k=1}^{n} \overline{u_{k,i}}^{(n)} u_{k,j}^{(n)} \delta_{\lambda_k^{(n)}}. \tag{29}$$

As in the proof of Theorem 2.1 (end of the first step), for all $\eta > 0$, the random matrix-valued function $M(n, \cdot)$ converges to

$$M_{G_{\mu_X}}(\cdot) := \operatorname{diag}(1 - \theta_1 G_{\mu_X}(\cdot), \ldots, 1 - \theta_r G_{\mu_X}(\cdot))$$

uniformly on $\{z \in \mathbb{C}\,;\, \mathrm{d}(z, [a, b]) \geq \eta\}$, almost surely. Since z_n tends almost surely to $G_{\mu_X}^{-1}(1/\theta_1)$, which is out of $[a, b]$, it follows that $M(n, z_n)$ tends almost surely to

$$\operatorname{diag}(\underbrace{0, \ldots, 0}_{r_0 \text{ zeros}}, 1 - \frac{\theta_{r_0+1}}{\theta_1}, \ldots, 1 - \frac{\theta_r}{\theta_1})$$

Moreover, by Proposition 5.1 a), for n large enough such that z_n is not an eigenvalue of X_n, $V_n^* x_n$ is a vector of the kernel of the $r \times r$ matrix $M(n, z_n)$ with norm ≤ 1. Since for all n,

$$V_n^* x_n = \begin{bmatrix} \langle C_1^{(n)}, x_n \rangle \\ \vdots \\ \langle C_r^{(n)}, x_n \rangle \end{bmatrix},$$

Lemma 7.1 allows to claim that (28) holds.

Let us now prove (27). By Proposition 5.1 a) again, one has, for all n,

$$\begin{aligned} x_n &= (z_n - X_n)^{-1} V_n \operatorname{diag}(\theta_1, \ldots, \theta_r) V_n^* x_n \\ &= (z_n - X_n)^{-1} \sum_{j=1}^{r} \theta_j \langle C_j^{(n)}, x_n \rangle C_j^{(n)}, \\ &= \underbrace{(z_n - X_n)^{-1} \sum_{j=1}^{r_0} \theta_j \langle C_j^{(n)}, x_n \rangle C_j^{(n)}}_{\text{denoted by } x_n'} + \underbrace{(z_n - X_n)^{-1} \sum_{j=r_0+1}^{r} \theta_j \langle C_j^{(n)}, x_n \rangle C_j^{(n)}}_{\text{denoted by } x_n''}. \end{aligned}$$

Note that z_n tends almost surely to $\rho = G_{\mu_X}^{-1}(1/\theta_1)$, which does not belong to $[a, b]$, and that the upper and lower bounds of the spectrum of X_n tend almost surely respectively to b and a, thus almost surely, the operator norms of $(z_n - X_n)^{-1}$ form a bounded sequence. By (28), it follows that $\|x_n''\|$ tends almost surely to zero. As a consequence, since $\|x_n\| = 1$, $\|x_n'\|$ tends almost surely to one. Since $\theta_1 = \cdots = \theta_{r_0}$ and $z_n - X_n$ is Hermitian,

$$\|x_n'\|^2 = \theta_1^2 \sum_{i,j=1}^{r_0} \overline{\langle C_i^{(n)}, x_n \rangle} \langle C_j^{(n)}, x_n \rangle \underbrace{C_i^{(n)*} (z_n - X_n)^{-2} C_j^{(n)}}_{= \int \frac{1}{(z_n - t)^2} \mathrm{d}\mu_{i,j}^{(n)}(t), \text{ with } \mu_{i,j}^{(n)} \text{ defined by (29)}} \tag{30}$$

By Proposition 11.3, for all $i \neq j$, $\mu_{i,j}^{(n)}$ converges almost surely weakly to zero and for all i, $\mu_{i,i}^{(n)}$ converges almost surely weakly to μ_X. Thus, by the second statement of Lemma

11.1, since z_n converges almost surely to $\rho \notin [a,b]$, for all $i,j = 1,\ldots,r_0$,

$$\int \frac{\mathrm{d}\mu_{i,j}^{(n)}(t)}{(z_n-t)^2} \xrightarrow{\text{a.s.}} \mathbb{1}_{i=j}\int \frac{\mathrm{d}\mu(t)}{(\rho-t)^2}.$$

By (30) and the fact that $\|x_n'\| \xrightarrow{\text{a.s.}} 1$, (27) follows. □

7.2. **Proof of Theorem 2.3.** Let us suppose that $\theta > 0$ (the other case can be treated in the same way). Note first that $G'_{\mu_X}(b^+) = -\infty$ implies that $\int \frac{\mathrm{d}\mu_X(t)}{(b-t)^2} = +\infty$.

As in the proof of Theorem 2.2, it suffices to prove the result in the case where $X_n = \mathrm{diag}(\lambda_1^{(n)},\ldots,\lambda_n^{(n)})$ and $P_n = \theta u^{(n)}u^{(n)*}$, with $u^{(n)}$ a column vector uniformly distributed on the unit real (or complex) n-sphere. We denote by $u_1^{(n)},\ldots,u_n^{(n)}$ the coordinates of the vector $u^{(n)}$ and define, for each n, the random probability measure

$$\mu^{(n)} = \sum_{k=1}^{n} |u_k^{(n)}|^2 \delta_{\lambda_k^{(n)}}.$$

Note that by the $r=1$ case of Proposition 5.1 b), the eigenvalues of $X_n + P_n$ which do not belong to $\{\lambda_1^{(n)},\ldots,\lambda_n^{(n)}\}$ are the solutions of $G_{\mu^{(n)}}(z) = \frac{1}{\theta}$. By the elementary remarks made on the Cauchy transform of a probability measure in Section 2.4.1, since $\lim_{t\downarrow\lambda_1^{(n)}} G_{\mu^{(n)}}(t) = +\infty$, we know that the largest eigenvalue z_n of X_n+P_n is $> \lambda_1^{(n)}$.

Reproducing the arguments leading to (18) in Section (4.2), we get

$$\langle x_n, \ker(\theta I_n - P_n)\rangle^2 = \frac{1}{\theta^2 \int \frac{\mathrm{d}\mu^{(n)}(t)}{(z_n-t)^2}}.$$

Thus it suffices to prove that, as n tends to infinity,

$$\int \frac{\mathrm{d}\mu^{(n)}(t)}{(z_n-t)^2} \quad \text{tends almost surely to} \quad +\infty. \tag{31}$$

Note that by hypothesis, $\frac{1}{n}\sum_{k=1}^{n}\delta_{\lambda_k^{(n)}}$ converges almost surely to μ_X. Since, by Theorem 2.1, $b - z_n$ tends almost surely to zero, it follows that $\frac{1}{n}\sum_{k=1}^{n}\delta_{\lambda_k^{(n)}+b-z_n}$, which is the push-forward of this measure by the map $t \longmapsto t + b - z_n$, converges also almost surely to μ_X (use, for example, the Fourier transform to see it). Hence, by (47),

$$\tilde{\mu}^{(n)} := \sum_{k=1}^{n} |u_k^{(n)}|^2 \delta_{\lambda_k^{(n)}+b-z_n}$$

tends almost surely to μ_X. It implies that for any lower-semicontinuous function $f : \mathbb{R} \to [0,\infty]$,

$$\liminf_{n\to+\infty} \int f(t)\mathrm{d}\tilde{\mu}^{(n)}(t) \geq \int f(t)\mathrm{d}\mu_X(t). \tag{32}$$

Equation (31) follows, by application of (32) for $f(t) = (b-t)^{-2}$. □

8. Proofs of Theorems 2.5–2.7 for the extreme eigenvalues/vectors of $X_n(I+P_n)$

These proofs can easily be obtained adapting to the square multiplicative case the proofs of the square additive case (using the master equation for the identification of the eigenvalues and of the eigenvectors given by Proposition 5.2 instead of the one given by Proposition 5.1).

9. Proof of Theorem 2.9 for the extreme singular values of $X_n + P_n$

9.1. **First step: case where the θ_i's are pairwise distinct.** Let us first suppose the θ_i's to be pairwise distinct.

Note first that by [9, Th. 3.13] (which extends easily to almost sure convergence), the symmetrization of the random probability measure

$$\frac{1}{n}\sum_{k=1}^{n}\delta_{\sigma_k(X_n+P_n)}$$

converges almost surely to $\mu_{X,s}\boxplus_c\delta_0=\mu_{X,s}$, where $\mu_{X,s}$ denotes the symmetrization of μ_X. It follows that for any i fixed independently of n,

$$\liminf_{n\to\infty}\sigma_i(X_n+P_n)\ge b \qquad\text{and}\qquad \limsup_{n\to\infty}\sigma_{n-i}(X_n+P_n)\le a.$$

Let us write, for each n,

$$X_n = U_X^{(n)}\begin{bmatrix}\sigma_1^{(n)} & & & 0 & \cdots & 0\\ & \ddots & & \vdots & & \vdots\\ & & \sigma_n^{(n)} & 0 & \cdots & 0\end{bmatrix}V_X^{(n)*}$$

$$P_n = U_P^{(n)}\begin{bmatrix}\theta_1 & & & & & 0 & \cdots & 0\\ & \ddots & & & & \vdots & & \vdots\\ & & \theta_r & & & 0 & \cdots & 0\\ & & & 0 & & 0 & \cdots & 0\\ & & & & \ddots & \vdots & & \vdots\\ & & & & 0 & 0 & \cdots & 0\end{bmatrix}V_P^{(n)*},$$

with $U_X^{(n)}, U_P^{(n)}$ $n\times n$ orthogonal (or unitary) matrices and $V_X^{(n)}, V_P^{(n)}$ $m\times m$ orthogonal (or unitary) matrices. The singular values of X_n+P_n is the same as the ones of

$$\begin{bmatrix}\operatorname{diag}(\sigma_1^{(n)},\ldots,\sigma_n^{(n)}) & 0_{n,m-n}\end{bmatrix}+U_X^{(n)*}U_P^{(n)}\begin{bmatrix}\operatorname{diag}(\theta_1,\ldots,\theta_r,0,\ldots,0) & 0_{n,m-n}\end{bmatrix}V_P^{(n)*}V_X^{(n)},$$

and since either X_n or P_n are biorthogonally (or biunitary) invariant in law, $U_X^{(n)*}U_P^{(n)}$ and $V_X^{(n)*}V_P^{(n)}$ are Haar-distributed [19, Lem. 4.3.10]. Let us denote by U_n (resp. V_n) the $n\times r$ (resp. $m\times r$) matrix which columns are the first columns of $U_X^{(n)*}U_P^{(n)}$ (resp. of $V_X^{(n)*}V_P^{(n)}$). Note that $\sigma^{(n)}:=(\sigma_1^{(n)},\ldots,\sigma_n^{(n)}), U_n, V_n$ are independent. Thus, from now on, we suppose that for $D_n=\operatorname{diag}(\sigma_1^{(n)},\ldots,\sigma_n^{(n)})$ and $\Omega=\operatorname{diag}(\theta_1,\,\ldots,\theta_r)$, we have

$$X_n=\begin{bmatrix}D_n & 0_{n,m-n}\end{bmatrix},\qquad P_n=U_n\Omega V_n^*,$$

with D_n, U_n, V_n independent and U_n, V_n respectively $n \times r$ and $m \times r$ matrices which columns are the r first columns of uniform orthogonal (or unitary) random matrices.

By Theorem 5.3, the positive singular values of $X_n + P_n$ are the positive eigenvalues of

$$\begin{bmatrix} 0 & X_n + P_n \\ X_n^* + P_n^* & 0 \end{bmatrix}$$

(when $n = m$, one can replace *positive* by *non negative* in this sentence). Moreover,

$$\begin{bmatrix} 0 & X_n + P_n \\ X_n^* + P_n^* & 0 \end{bmatrix} = \begin{bmatrix} 0 & X_n \\ X_n^* & 0 \end{bmatrix} + \begin{bmatrix} U_n & 0 \\ 0 & V_n \end{bmatrix} \begin{bmatrix} 0 & \Omega \\ \Omega^* & 0 \end{bmatrix} \begin{bmatrix} U_n & 0 \\ 0 & V_n \end{bmatrix}^* . \tag{33}$$

Since $\sigma_1(X_n)$ tends to b and $\sigma_n(X_n)$ tends to a almost surely, we can focus on the eigenvalues of the matrix of (33) which are out of $[\sigma_n(X_n), \sigma_1(X_n)]$. By Proposition 5.1 b), first part, and Proposition 5.4, these eigenvalues are the numbers z out of $[\sigma_n(X_n), \sigma_1(X_n)]$ such that the $2r \times 2r$ matrix

$$M(n,z) := I_{2r} - \begin{bmatrix} U_n^* \frac{D_n}{z^2 - D_n^2} K_n \Omega & U_n^* \frac{z}{z^2 - D_n^2} U \Omega \\ K_n^* \frac{z}{z^2 - D_n^2} K_n \Omega + \frac{1}{z} L_n^* L_n \Omega & K_n^* \frac{D_n}{z^2 - D_n^2} U_n \Omega \end{bmatrix} \tag{34}$$

is not invertible, where K_n is the $n \times r$ upper-block of V_n and L_n the $(m-n) \times r$ lower one. As mentioned in Proposition 5.4, the entries of this matrix can be interpreted as integrals of complex measures. Using Proposition 11.3 and the second statement of Lemma 11.1, one easily sees that for all $\eta > 0$, the matrix-valued function $M(n,z)$ converges almost surely uniformly on $\{z \in \mathbb{C}\,;\, \mathrm{d}(z, [a,b] \cup \{0\}) \geq \eta\}$ to

$$M_{\mu_X}(z) := I_{2r} - \begin{bmatrix} 0 & N(z) \\ N_0(z) & 0 \end{bmatrix}, \tag{35}$$

with

$$\begin{aligned} N(z) &= \operatorname{diag}\left(\theta_1 \varphi_{\mu_X}(z), \ldots, \theta_r \varphi_{\mu_X}(z)\right), \\ N_0(z) &= \operatorname{diag}\left(\theta_1 \varphi_{c\mu_X + (1-c)\delta_0}(z), \ldots, \theta_r \varphi_{c\mu_X + (1-c)\delta_0}(z)\right) \end{aligned}$$

(in the case where m is always equal to n, one can replace $[a,b] \cup \{0\}$ above by $[a,b]$).

Now, as in the proof of Theorem 2.1, the conclusion follows exactly from a modified version of Lemma 11.4. Let us explain which version of this lemma we shall need here. To treat the first part of our theorem, firstly replace the interval $[a,b]$ of the lemma by $(-\infty, b]$ and secondly replace the matrix $M_G(z)$ defined in (51) by $M_{\mu_X}(z)$, which, by Lemma 5.5, is actually non invertible if and only if one of the $1/\theta_i^2$'s equals $D_{\mu_X}(c,z)$. To prove the second part of our theorem (which concerns the case where m is always equal to n), apply again a modified version of Lemma 11.4 with the same matrix, where the interval $[a,b]$ is replaced by $(-\infty, 0] \cup [a, +\infty)$ and where the hypothesis c) is replaced by the fact that

$$\lim_{z \downarrow 0} D_{\mu_X}(1,z) = 0.$$

9.2. **Second step: the general case.** The extension to the general case works exactly as in the proof of Theorem 2.1, using [20, Cor. 7.3.8 (b)] instead of [20, Cor. 6.3.8]. $\square$

10. Proofs of Theorems 2.10 and 2.12 for the extreme singular vectors of $X_n + P_n$

10.1. **Proof of Theorem 2.10.** We employ the notation introduced at the beginning of the proof of Theorem 2.9. Since the pairs of singular vectors of $X_n + P_n$ are $U_X^{(n)} \oplus V_X^{(n)*}$ times the ones of

$$D_X^{(n)} + U_X^{(n)*} U_P^{(n)} D_P^{(n)} V_P^{(n)*} U_X^{(n)},$$

from now on, we suppose that for $D_n = \operatorname{diag}(\sigma_1^{(n)}, \ldots, \sigma_n^{(n)})$ and $\Omega = \operatorname{diag}(\theta_1, \ldots, \theta_r)$, we have

$$X_n = \begin{bmatrix} D_n & 0_{n,m-n} \end{bmatrix}, \qquad P_n = U_n \Omega V_n^*,$$

with D_n, U_n, V_n independent and U_n, V_n respectively $n \times r$ and $m \times r$ matrices which columns are the r first columns of uniform orthogonal (or unitary) random matrices.

Let us define, for each n,

$$X_n^h = \begin{bmatrix} 0_{n,n} & D_n & 0_{n,m-n} \\ D_n & 0_{n,n} & 0_{n,m-n} \\ 0_{m-n,n} & 0_{m-n,n} & 0_{m-n,m-n} \end{bmatrix}, \quad \Theta = \begin{bmatrix} 0_{r,r} & \Omega \\ \Omega & 0_{r,r} \end{bmatrix}, \quad W_n = \begin{bmatrix} U_n & 0_{n,r} \\ 0_{n,r} & K_n \\ 0_{m-n,r} & L_n \end{bmatrix},$$

where K_n the $n \times r$ upper-block of V_n and L_n the $(m-n) \times r$ lower one. Then, by Proposition 5.4, we have, for all z out of the spectrum of X_n^h,

$$W_n^*(z - X_n^h)^{-1} W_n \Theta = \begin{bmatrix} U_n^* \frac{D_n}{z^2 - D_n^2} K_n \Omega & U_n^* \frac{z}{z^2 - D_n^2} U_n \Omega \\ K_n^* \frac{z}{z^2 - D_n^2} K_n \Omega + \frac{1}{z} L_n^* L_n \Omega & K_n^* \frac{D_n}{z^2 - D_n^2} U_n \Omega \end{bmatrix}.$$

Let r_0 be the number of i's such that $\theta_i = \theta_{i_0}$. Up to a reindex of the θ_i's (which is no longer decreasing, but it is not a problem here), one can suppose that $i_0 = 1$, $\theta_1 = \cdots = \theta_{r_0}$. Then for each n, $\ker(\theta_1^2 I_n - P_n P_n^*)$ (resp. $\ker(\theta_1^2 I_m - P_n^* P_n)$) is the linear span of the r_0 first columns of U_n (resp. of V_n). We denote the columns of U_n (resp. V_n) by $U_1^{(n)}, \ldots, U_r^{(n)}$ (resp. $V_1^{(n)}, \ldots, V_r^{(n)}$). Since these columns are orthonormal, to prove b), it suffices to prove that as n tends to infinity,

$$\sum_{j=1}^{r_0} |\langle V_j^{(n)}, v_n \rangle|^2 \quad \text{tends almost surely to the RHS of (4).} \tag{36}$$

Moreover, for all column vector x, we have $P_n x = \sum_{j=1}^r \theta_j \langle V_j^{(n)}, x \rangle U_j^{(n)}$, thus to prove d), it suffices to prove that

$$\sum_{j=1}^{r_0} |\theta_1 \varphi_{\mu_X}(\rho) \langle V_j^{(n)}, v_n \rangle - \langle U_j^{(n)}, u_n \rangle|^2 \quad \text{tends almost surely to } 0. \tag{37}$$

Note that a) follows also from (36) and (37). Indeed, (36) and (37) granted, $\sum_{j=1}^{r_0} |\langle U_j^{(n)}, u_n \rangle|^2$ tends to $\theta_1^2 \varphi_{\mu_X}(\rho)^2$ times the RHS of (4), which is equal to the RHS of (3), because $\varphi_\mu(\rho) \varphi_{\tilde{\mu}_X}(\rho) = 1/\theta_{i_0}^2$. To prove d), we have to prove that

$$\sum_{j=r_0+1}^{r} |\langle U_j^{(n)}, u_n \rangle|^2 + \langle V_j^{(n)}, v_n \rangle|^2 \quad \text{tends almost surely to } 0. \tag{38}$$

To prove e), it suffices to remark that the following proof stays valid if $a > 0$, m is always equal to n, $z_n = \sigma_{n+1-i_0}(X_n + P_n)$ and $\rho = D^{-1}_{\mu_X,[0,a)}(1/\theta^2_{i_0})$.

So let us prove (36), (37) and (38).

As in the proof of Theorem 2.9, for all $\eta > 0$, the matrix-valued function

$$M(n, z) := I_{2r} - W_n^*(z - X_n^h)^{-1}W_n\Theta$$

converges almost surely uniformly on $\{z \in \mathbb{C}\,;\, \mathrm{d}(z, [a, b]\cup\{0\}) \geq \eta\}$ to the matrix $M_{\mu_X}(z)$ of (35) (in the case where m is always equal to n, one can replace $[a, b]\cup\{0\}$ above by $[a, b]$). Since z_n tends almost surely to $\rho := D^{-1}_{\mu_X}(c, 1/\theta_1^2)$, which is $> b$ (or in $(0, a) \cup (b, +\infty)$ if $a > 0$ and m is always equal to n), it follows that $M(n, z_n)$ tends almost surely to $M_{\mu_X}(\rho)$.

The space $\ker(M_{\mu_X}(\rho))$ is the set of vectors $(x_1, \ldots, x_r, y_1, \ldots, y_r)$ such that for all $i = 1, \ldots, r$,

$$x_i = \theta_i\varphi_{\mu_X}(\rho)y_i \quad \text{and} \quad y_i = \theta_i\varphi_{\tilde{\mu}_X}(\rho)x_i.$$

Note that $\varphi_{\mu_X}(\rho)\varphi_{\tilde{\mu}_X}(\rho) = D_{\mu_X}(c, \rho) = 1/\theta_1^2$, thus $\ker(M_{\mu_X}(\rho))$ is the set of vectors $(x_1, \ldots, x_r, y_1, \ldots, y_r)$ such that for all $i = 1, \ldots, r_0$, $x_i = \theta_1\varphi_{\mu_X}(\rho)y_i$ and for all $i = r_0 + 1, \ldots, r$, $x_i = y_i = 0$. Hence the orthogonal projection onto $\ker(M_{\mu_X}(\rho))^\perp$ is the map which maps any vector $(x_1, \ldots, x_r, y_1, \ldots, y_r)$ onto

$$(\overline{x}_1, \ldots, \overline{x}_{r_0}, x_{r_0+1}, \ldots, x_r, \overline{y}_1, \ldots, \overline{y}_{r_0}, y_{r_0+1}, \ldots, y_r),$$

where for all i, $(\overline{x}_i, \overline{y}_i) = \frac{x_i-\theta_1\varphi_{\mu_X}(\rho)y_i}{1+\theta_1^2\varphi_{\mu_X}(\rho)^2}(1, -\theta_1\varphi_{\mu_X}(\rho))$.

Let us define $x_n = \begin{bmatrix} u_n \\ v_n \end{bmatrix}$. By the relations between eigenvectors and pairs of singular vectors given in Theorem 5.3 and by Proposition 5.1 a), for n large enough such that z_n is not a singular value of X_n, $W_n^*x_n$ is a vector of $\ker(M(n, z_n))$, with norm ≤ 2. Since for all n, the $2r$ coordinates of $W_n^*x_n$ are respectively

$$\langle U_1^{(n)}, u_n\rangle, \ldots, \langle U_r^{(n)}, u_n\rangle, \langle V_1^{(n)}, v_n\rangle, \ldots, \langle V_r^{(n)}, v_n\rangle,$$

Lemma 7.1 allows to claim that (37) and (38) hold.

Let us now prove (36). Again, by the relations between eigenvectors and pairs of singular vectors given in Theorem 5.3 and by Proposition 5.1 a), one has, for all n,

$$\begin{aligned} x_n &= (z_n - X_n^h)^{-1}W_n\Omega W_n^*x_n \\ &= (z_n - X_n^h)^{-1}\sum_{j=1}^{r}\theta_j\langle V_j^{(n)}, v_n\rangle\tilde{U}_j^{(n)} + \theta_j\langle U_j^{(n)}, u_n\rangle\tilde{V}_j^{(n)}, \end{aligned}$$

where we defined, for $j = 1, \ldots, r$, $\tilde{U}_j^{(n)} := \begin{bmatrix} U_j^{(n)} \\ 0_{m,1} \end{bmatrix}$ and $\tilde{V}_j^{(n)} := \begin{bmatrix} 0_{n,1} \\ V_j^{(n)} \end{bmatrix}$. For each n, we have $x_n = x'_n + x''_n$, with

$$\begin{aligned} x'_n &= \theta_1 (z_n - X_n^h)^{-1} \sum_{j=1}^{r_0} \langle V_j^{(n)}, v_n \rangle (\tilde{U}_j^{(n)} + \theta_1 \varphi_{\mu_X}(\rho) \tilde{V}_j^{(n)}), \\ x''_n &= \theta_1 (z_n - X_n^h)^{-1} \sum_{j=1}^{r_0} (\langle U_j^{(n)}, u_n \rangle - \theta_1 \varphi_{\mu_X}(\rho) \langle V_j^{(n)}, v_n \rangle) \tilde{V}_j^{(n)} \\ &\quad + (z_n - X_n^h)^{-1} \sum_{j=r_0+1}^{r} \theta_j \langle V_j^{(n)}, x_n \rangle \tilde{U}_j^{(n)} + \theta_j \langle U_j^{(n)}, x_n \rangle \tilde{V}_j^{(n)}. \end{aligned}$$

Note that z_n tends almost surely to ρ, which does not belong to $[a, b] \cup \{0\}$, thus almost surely, the operator norms of $(z_n - X_n)^{-1}$ form a bounded sequence. By (37) and (38), it follows that $\|x''_n\|$ tends almost surely to zero. Since u_n has norm one, the norm of the $n \times 1$ upper part of x'_n, that we shall denote by u'_n, tends almost surely to one. By (21),

$$(z_n - X_n^h)^{-1} = \begin{bmatrix} \frac{z_n}{z_n^2 - D_n^2} & \frac{D_n}{z_n^2 - D_n^2} & 0_{n,m-n} \\ \frac{D_n}{z_n^2 - D_n^2} & \frac{z_n}{z_n^2 - D_n^2} & 0_{n,m-n} \\ 0_{m-n,n} & 0_{m-n,n} & z_n^{-1} \end{bmatrix},$$

hence

$$u'_n = \theta_1 \sum_{j=1}^{r_0} \langle V_j^{(n)}, v_n \rangle \left(\frac{z_n}{z_n^2 - D_n^2} U_j^{(n)} + \theta_1 \varphi_{\mu_X}(\rho) \frac{D_n}{z_n^2 - D_n^2} K_j^{(n)} \right),$$

where for all j, $K_j^{(n)}$ is the j-th column of K_n. It follows that

$$\|u'_n\|^2 = \theta_1^2 \sum_{i,j=1}^{r_0} \overline{\langle V_i^{(n)}, v_n \rangle} \langle V_j^{(n)}, v_n \rangle \operatorname{Coef}_n(i, j),$$

where we defined, for all $i, j = 1, \ldots, r_0$,

$$\begin{aligned} \operatorname{Coef}_n(i, j) &:= U_i^{(n)*} \frac{z_n^2}{(z_n^2 - D_n^2)^2} U_j^{(n)} + \theta_1^2 \varphi_{\mu_X}(\rho)^2 K_i^{(n)*} \frac{D_n^2}{(z_n^2 - D_n^2)^2} K_j^{(n)} \\ &\quad + \theta_1 \varphi_{\mu_X}(\rho) U_i^{(n)*} \frac{z_n D_n}{(z_n^2 - D_n^2)^2} K_j^{(n)} + \theta_1 \varphi_{\mu_X}(\rho) K_i^{(n)*} \frac{z_n D_n}{(z_n^2 - D_n^2)^2} U_j^{(n)}, \end{aligned}$$

By Proposition 11.3, for all $i, j = 1, \ldots, r_0$, as n tends to infinity,

$$\operatorname{Coef}_n(i, j) \xrightarrow{\text{a.s.}} \begin{cases} \int \frac{\rho^2}{(\rho^2 - t^2)^2} \mathrm{d}\mu_X(t) + \theta_1^2 \varphi_{\mu_X}(\rho)^2 \int \frac{t^2}{(\rho^2 - t^2)^2} \mathrm{d}\tilde{\mu}_X(t) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Hence $\sum_{j=1}^{r_0} |\langle V_j^{(n)}, v_n \rangle|^2$ tends to

$$\frac{1}{\theta_1^2} \frac{1}{\int_t \frac{\rho^2}{(\rho^2 - t^2)^2} \mathrm{d}\mu_X(t) + \theta_1^2 \varphi_{\mu_X}(\rho)^2 \int \frac{t^2}{(\rho^2 - t^2)^2} \mathrm{d}\tilde{\mu}_X(t)}.$$

The following relations:

$$\begin{aligned}\varphi_{\mu_X}(\rho)\varphi_{\tilde{\mu}_X}(\rho) &= D_{\mu_X}(c,\rho) = 1/\theta_1^2 \\ 2\int \frac{\rho^2}{(\rho^2-t^2)^2}\mathrm{d}\mu_X(t) &= \frac{1}{\rho}\varphi_{\mu_X}(\rho) - \varphi'_{\mu_X}(\rho) \\ 2\int \frac{t^2}{(\rho^2-t^2)^2}\mathrm{d}\tilde{\mu}_X(t) &= -\frac{1}{\rho}\varphi_{\tilde{\mu}_X}(\rho) - \varphi'_{\tilde{\mu}_X}(\rho)\end{aligned}$$

allow to recover the RHS of (3) easily. Thus (36) is proved. □

10.2. **Proof of Theorem 2.12.** As in the proof of Theorem 2.10, the problem can be reduced to the case where

$$X_n = \begin{bmatrix} D_n & 0_{n,m-n}\end{bmatrix}, \qquad P_n = \theta u_n^{(P)} v_n^{(P)*},$$

where $D_n, u_n^{(P)}, v_n^{(P)}$ are independent, $D_n = \operatorname{diag}(\sigma_1^{(n)}, \ldots, \sigma_n^{(n)})$ and $u_n^{(P)}, v_n^{(P)}$ are uniform random vectors of the unit spheres of respectively $\mathbb{R}^n, \mathbb{R}^m$ (or $\mathbb{C}^n, \mathbb{C}^m$). Then, by the relations between eigenvectors and pairs of singular vectors given in Theorem 5.3, by Proposition 5.1 a) and by Proposition 5.4, for n large enough such that z_n is not an singular value of X_n, we have

$$\begin{bmatrix} u_n \\ v_n \end{bmatrix} = \begin{bmatrix} \frac{z_n}{z_n^2-D_n^2} & \frac{D_n}{z_n^2-D_n^2} & 0_{n,m-n} \\ \frac{D_n}{z_n^2-D_n^2} & \frac{z_n}{z_n^2-D_n^2} & 0_{n,m-n} \\ 0_{m-n,n} & 0_{m-n,n} & z_n^{-1} \end{bmatrix} \begin{bmatrix} \theta v_n^{(P)*} v_n u_n^{(P)} \\ \theta u_n^{(P)*} u_n v_n^{(P)} \end{bmatrix}.$$

Let us denote by K_n the column vector of the n first coordinates of $v_n^{(P)}$. We have

$$\begin{aligned} 1 = \|u_n\|^2 &= \theta^2 |v_n^{(P)*} v_n|^2 u_n^{(P)*} \frac{z_n^2}{(z_n^2 - D_n^2)^2} u_n^{(P)} + \theta^2 |u_n^{(P)*} u_n|^2 K_n^* \frac{D_n^2}{(z_n^2 - D_n^2)^2} K_n \\ &\quad + \theta^2 \overline{v_n^{(P)*} v_n} u_n^{(P)*} u_n u_n^{(P)*} \frac{z_n D_n}{(z_n^2 - D_n^2)^2} K_n + \theta^2 \overline{u_n^{(P)*} u_n} v_n^{(P)*} v_n K_n^* \frac{z_n D_n}{(z_n^2 - D_n^2)^2} u_n^{(P)}. \end{aligned}$$

Note that the hypotheses of the theorem imply that $\int \frac{\mathrm{d}\mu_X(t)}{(b^2-t^2)^2} = +\infty$ (and that $\int \frac{\mathrm{d}\mu_X(t)}{(a^2-t^2)^2} = +\infty$ in the last part of the theorem). One concludes as at the end of the proof of Theorem 2.3 that $|u_n^{(P)*} u_n|$ and $|v_n^{(P)*} v_n|^2$ tend almost surely to zero as n tends to infinity, which is the statement of the theorem. □

11. Appendix: technical preliminaries needed for the proofs

11.1. Convergence of weighted spectral measures.

11.1.1. *A few facts about the weak convergence of complex measures.* Recall that a sequence (μ_n) of complex measures on $\mathbb{R}$ is said to *converge weakly* to a complex measure μ on $\mathbb{R}$ if, for any continuous bounded function f on $\mathbb{R}$,

$$\int f(t)\mathrm{d}\mu_n(t) \underset{n\to\infty}{\longrightarrow} \int f(t)\mathrm{d}\mu(t). \tag{39}$$

The following lemma shall be useful. It is well known for probability measures, but since we did not find any reference on its "complex measures version", we give a proof. Recall that a sequence (μ_n) of complex measures on $\mathbb{R}$ is said to be *tight* if

$$\lim_{R\to+\infty} \sup_n |\mu_n|(\{t \in \mathbb{R}\,;\, |t| \geq R\}) = 0.$$

Lemma 11.1. *Let D be a dense subset of the set of continuous functions on $\mathbb{R}$ tending to zero at infinity, endowed with the topology of the uniform convergence. Suppose that the sequence (μ_n) is tight, that $(|\mu_n|(\mathbb{R}))$ is bounded and that (39) holds for any function f in D. Then (μ_n) converges weakly to μ. Moreover, the convergence of (39) is uniform on any set of uniformly bounded and uniformly Lipschitz functions.*

Proof. Firstly, note that using the boundedness of $(|\mu_n|(\mathbb{R}))$, one extends easily (39) to any continuous function tending to zero at infinity.

Let us fix a continuous bounded function f. For any continuous function g tending to zero at infinity, we have

$$\left|\int f\mathrm{d}(\mu-\mu_n)\right| \leq \int |f(1-g)|\mathrm{d}(|\mu|+|\mu_n|) + \left|\int fg\mathrm{d}(\mu-\mu_n)\right|, \tag{40}$$

thus

$$\limsup_{n\to\infty}\left|\int f\mathrm{d}(\mu-\mu_n)\right| \leq \sup_n \int |f(1-g)|\mathrm{d}(|\mu|+|\mu_n|),$$

which, with a good choice of g (relying on the tightness hypothesis), can be made as small as we want. Thus the first statement is proved.

Now, consider a set A of uniformly bounded and uniformly Lipschitz functions and $\varepsilon > 0$. Let $M > 0$ be such that for all $f \in A$, $|f| \leq M$. By tightness, there is $R > 0$ such that for all n, $M(|\mu| + |\mu_n|)(\{t \in \mathbb{R}\,;\, |t| \geq R\}) \leq \varepsilon$. Let g be a continuous compactly supported function such that $0 \leq g \leq 1$ and $g(t) = 1$ for all $t \in [-R, R]$. For all $f \in A$, for all n, $\int |f(1-g)|\mathrm{d}(|\mu|+|\mu_n|) \leq \varepsilon$. Thus by (40), one can suppose that the elements of A have all their supports contained in the same compact set. In this case, the result is an straightforward application of Ascoli's Theorem. □

11.1.2. *Convergence of weighted spectral measures.*

Lemma 11.2. *Let, for each n, $U^{(n)} = (u_1^{(n)}, \ldots, u_n^{(n)})$, $V^{(n)} = (v_1^{(n)}, \ldots, v_n^{(n)})$ be the two first rows of a uniform random orthogonal (resp. unitary) matrix. Let also $x^{(n)} = (x_1^{(n)}, \ldots, x_n^{(n)})$ be a random family of real numbers independent of $(U^{(n)}, V^{(n)})$.*

a) Suppose that for all n, $x^{(n)}$ belongs to the unit euclidian ball of $\mathbb{R}^n$. Then

$$u_1^{(n)}x_1^{(n)} + \cdots + u_n^{(n)}x_n^{(n)} \xrightarrow{a.s.} 0. \tag{41}$$

b) Suppose that for all n, k, $|x_k^{(n)}| \leq 1$ and that $\frac{1}{n}(x_1^{(n)} + \cdots + x_n^{(n)})$ tends almost surely to a deterministic limit l. Then

$$|u_1^{(n)}|^2x_1^{(n)} + \cdots + |u_n^{(n)}|^2x_n^{(n)} \xrightarrow{a.s.} l. \tag{42}$$

c) Suppose that for all n, k, $|x_k^{(n)}| \leq 1$. Then as n tends to infinity,

$$\overline{u}_1^{(n)}{v_1}^{(n)}x_1^{(n)} + \cdots + \overline{u}_n^{(n)}{v_n}^{(n)}x_n^{(n)} \xrightarrow{a.s.} 0. \tag{43}$$

Proof. Firstly, by conditioning, one can suppose the $x^{(n)}$'s to be deterministic.

Let us first prove a) and b). Only $U^{(n)}$ is involved in a) and b), thus since a uniform random vector on the unit sphere of $\mathbb{C}^n$ is a uniform random vector on the unit sphere of $\mathbb{R}^{2n}$, it suffices to treat the real case. Now, let us recall a well-known concentration result [5, Lem. 14.2][6]. Let g be a real-valued 1-Lipschitz function on the unit sphere of $\mathbb{R}^n$. Let m_g be one of its medians: both events $\{g(U^{(n)}) \geq m_g\}$ and $\{g(U^{(n)}) \leq m_g\}$ have probabilities $\geq 1/2$. Then we have, for any $\varepsilon > 0$,

$$\mathbb{P}\{|g(U^{(n)}) - m_g| \geq \varepsilon\} \quad \leq \quad \sqrt{\pi/2}e^{-\varepsilon^2(n-1)/2}.$$

It follows that

$$|\mathbb{E}[g(U^{(n)})] - m_g| \leq \mathbb{E}[|g(U^{(n)}) - m_g|] = \int_{t=0}^{+\infty} \mathbb{P}\{|g(U^{(n)}) - m_g| \geq t\}\mathrm{d}t \leq \frac{\pi}{2\sqrt{n-1}}.$$

Thus, $|g(U^{(n)}) - \mathbb{E}[g(U^{(n)})]| \geq \frac{\varepsilon+\pi}{2\sqrt{n-1}}$ implies that $|g(U^{(n)}) - m_g| \geq \frac{\varepsilon}{2\sqrt{n-1}}$ and the following inequality follows:

$$\mathbb{P}\Big\{|g(U^{(n)}) - \mathbb{E}[g(U^{(n)})]| \geq \frac{\varepsilon+\pi}{2\sqrt{n-1}}\Big\} \leq \sqrt{\pi/2}e^{-\varepsilon^2/8} \tag{44}$$

As a consequence, if, for all n, g_n is a 1-Lipschitz function on the unit sphere of $\mathbb{R}^n$, such that that $\mathbb{E}[g_n(U^{(n)})]$ converges, as n goes to infinity, to a finite limit, then $g_n(U^{(n)})$ converges almost surely to the same limit (indeed, by the Borel Cantelli Lemma, it suffices to prove that for any $\eta > 0$, the series $\sum_n \mathbb{P}\{|g_n(U^{(n)}) - \mathbb{E}[g_n(U^{(n)})]| > \eta\}$ converges, which follows from (44), applied with $\varepsilon = 2\sqrt{n-1}\eta - \pi$). For a), applying this principle for $g_n : u \mapsto \langle x^{(n)}, u\rangle$, which gradient, $x^{(n)}$, actually belongs to the unit euclidian sphere, allows to conclude. For b), use the function $g_n : u \mapsto \frac{1}{2}u^* \operatorname{diag}(x^{(n)})u$, which gradient at any point u of the sphere, is equal to $\operatorname{diag}(x^{(n)})u$, thus actually belongs to the unit euclidian sphere.

To prove c), the strategy is quite different. Let us define the random variable $Z_n = \overline{u}_1^{(n)}{v_1}^{(n)}x_1^{(n)} + \cdots + \overline{u}_n^{(n)}{v_n}^{(n)}x_n^{(n)}$. Since Z_n is centered, by Chebyshev's inequality and Borel Cantelli's Lemma, it suffices to prove that $\mathbb{E}[Z_n^4] = O(n^{-2})$. We have

$$\mathbb{E}[Z_n^4] = \sum_{i,j,k,l=1}^{n} x_i^{(n)}x_j^{(n)}x_k^{(n)}x_l^{(n)}\mathbb{E}[u_i^{(n)}u_j^{(n)}u_k^{(n)}u_l^{(n)}v_i^{(n)}v_j^{(n)}v_k^{(n)}v_l^{(n)}].$$

Note that by definition of the Haar measure, $\mathbb{E}[u_i^{(n)}u_j^{(n)}u_k^{(n)}u_l^{(n)}v_i^{(n)}v_j^{(n)}v_k^{(n)}v_l^{(n)}] = 0$ whenever among i, j, k, l, one is different of all others. It follows that

$$\begin{aligned}\mathbb{E}[Z_n^4] &\leq 3\sum_{i,j} x_i^{(n)^2}x_j^{(n)^2}\mathbb{E}[u_i^{(n)^2}u_j^{(n)^2}v_i^{(n)^2}v_j^{(n)^2}] \\ &\leq 3\sum_{i,j} x_i^{(n)^2}x_j^{(n)^2}\underbrace{(\mathbb{E}[u_i^{(n)^8}]\mathbb{E}[u_j^{(n)^8}]\mathbb{E}[v_i^{(n)^8}]\mathbb{E}[v_j^{(n)^8}])^{\frac{1}{4}}}_{=\mathbb{E}[u_1^{(n)^8}]}.\end{aligned}$$

Since, by [15], $\mathbb{E}[u_1^{(n)^8}] = O(n^{-4})$, the conclusion holds. □

[6]Usually, this result is stated for functions which are 1-Lipschitz with respect to the geodesic distance on the sphere $\mathrm{d}(u, v) = \arccos\langle u, v\rangle$, but using $\langle u, v\rangle = 1 - \|v - u\|^2/2$, it appears that $\|v - u\| \leq \mathrm{d}(u, v)$.

Proposition 11.3. *Let, for each n, $U^{(n)} = (u_1^{(n)}, \ldots, u_n^{(n)})$, $V^{(n)} = (v_1^{(n)}, \ldots, v_n^{(n)})$ be the two first columns of a uniform random orthogonal (resp. unitary) matrix. Consider also $m_n \geq n$ and a uniform random vector $W^{(n)} = (w_1^{(n)}, \ldots, w_{m_n}^{(n)})$ on the unit sphere of $\mathbb{R}^{m_n}$ (resp. $\mathbb{C}^{m_n}$), independent of $U^{(n)}$. Let also $\lambda^{(n)} = (\lambda_1^{(n)}, \ldots, \lambda_n^{(n)})$ be a random family of real numbers independent of $(U^{(n)}, V^{(n)}, W^{(n)})$. We suppose that $n/m_n \underset{n\to\infty}{\longrightarrow} c \in [0,1]$, that*

$$\text{the sequence } (\max_k |\lambda_k^{(n)}|) \text{ is almost surely bounded,} \tag{45}$$

and that there exists a deterministic probability measure μ on $\mathbb{R}$ such that as n tends to infinity,

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k^{(n)}} \text{ converges almost surely weakly to } \mu. \tag{46}$$

Then as n tends to infinity,

$$\mu_{U^{(n)}} := \textstyle\sum_{k=1}^{n} |u_k^{(n)}|^2 \delta_{\lambda_k^{(n)}} \text{ converges almost surely weakly to } \mu, \tag{47}$$

$$\mu_{W^{(n)}} := \textstyle\sum_{k=1}^{n} |w_k^{(n)}|^2 \delta_{\lambda_k^{(n)}} + \sum_{k=n+1}^{m_n} |w_k^{(n)}|^2 \delta_0 \tag{48}$$

$$\text{converges almost surely weakly to } c\mu + (1-c)\delta_0,$$

$$\mu_{U^{(n)},V^{(n)}} := \textstyle\sum_{k=1}^{n} \overline{u}_k^{(n)} v_k^{(n)} \delta_{\lambda_k^{(n)}} \text{ converges almost surely weakly to } 0, \tag{49}$$

$$\mu_{U^{(n)},W^{(n)}} := \textstyle\sum_{k=1}^{n} \overline{u}_k^{(n)} w_k^{(n)} \delta_{\lambda_k^{(n)}} \text{ converges almost surely weakly to } 0. \tag{50}$$

Proof. We shall use the first statement of Lemma 11.1. Note first that by hypothesis (45), almost surely, all these sequences of complex measures are tight. Moreover, we have

$$|\mu_{U^{(n)},V^{(n)}}| = \sum_{k=1}^{n} |u_k^{(n)} v_k^{(n)}| \delta_{\lambda_k^{(n)}},$$

thus, by the Cauchy-Schwartz inequality, we have $|\mu_{U^{(n)},V^{(n)}}|(\mathbb{R}) \leq 1$. In the same way, $|\mu_{U^{(n)},W^{(n)}}|(\mathbb{R}) \leq 1$. Since $\mu_{U^{(n)}}, \mu_{W^{(n)}}$ are probability measures, the same inequality holds obviously for them.

The set of continuous functions on the real line tending to zero at infinity admits a countable dense subset, so it suffices to prove that for a fixed such function f, the convergences of (47), (48), (49) and (50) hold almost surely when applied to f. So let us fix a continuous bounded function f on $\mathbb{R}$. One can suppose that $|f| \leq 1$. The convergences of (47) and (48), applied to f, follow from an application of (42). The convergence of (49), applied to f, follows from (43). At last, the convergence of (50), applied to f, follows from (41). □

11.2. **A technical lemma.** We shall need the following result. Note that nothing, in its hypotheses, is random. We define, for $z \in \mathbb{C}$ and E a closed subset of $\mathbb{R}$, $\mathrm{d}(z, E) = \min_{x\in E} |z - x|$.

Lemma 11.4. *Let us fix a positive integer r, a family $\theta_1, \ldots, \theta_r$ of pairwise distinct nonzero real numbers, two real numbers $a < b$, a function G which is analytic on $\mathbb{C}\backslash[a, b]$ and such that*

a) *G does not take any real value out of $\mathbb{R}\backslash[a,b]$,*
b) *for all $z \in (-\infty, a) \cup (b, +\infty)$, $G(z) \in \mathbb{R}$ and $G'(z) < 0$,*
c) *$G(z)$ tends to zero as $z \in \mathbb{R}$ tends to infinity.*

Let us define, for $z \in \mathbb{C}\backslash[a,b]$, the $r \times r$ matrix

$$M_G(z) = \operatorname{diag}(1 - \theta_1 G(z), \ldots, 1 - \theta_r G(z)), \tag{51}$$

and denote by $z_1 > \cdots > z_p$ the z's such that $M_G(z)$ is not invertible ($p \in \{0, \ldots, r\}$ is the number of θ_i's such that the equation $G(z) = 1/\theta_i$ has a solution).

Let us also consider two sequences a_n, b_n with respective limits a, b and, for each n, a function $M(n, \cdot)$, defined on $\mathbb{C}\backslash[a_n, b_n]$, with values in the set of $r \times r$ complex matrices and which coefficients are analytic functions. We suppose that

d) *for all n, for all $z \in \mathbb{C}\backslash\mathbb{R}$, the matrix $M(n,z)$ is invertible,*
e) *for all $\eta > 0$, $M(n, \cdot)$ converges, as n tends to infinity, to the function $M_G(\cdot)$, uniformly on $\{z \in \mathbb{C}\,;\, \mathrm{d}(z, [a,b]) \geq \eta\}$.*

Then there exists p real sequences $z_{n,1} > \ldots > z_{n,p}$ converging respectively to $z_1, \ldots, z_p$ such that for all $\varepsilon \in (0, \min_i \mathrm{d}(z_i, [a,b]))$, for n large enough, the z's in $\mathbb{R}\backslash[a-\varepsilon, b+\varepsilon]$ such that $M(n,z)$ is not invertible are exactly $z_{n,1}, \ldots, z_{n,p}$. Moreover, for n large enough, for each i, $M(n, z_{n,i})$ has rank $r-1$.

Proof. Note firstly that by c), there exists $R > 0$ such that for $z \in \mathbb{R}$ such that $|z| \geq R$, $|G(z)| \leq \min_i \frac{1}{2|\theta_i|}$. By d) and e), it follows that for n large enough, all the z's such that $M(n,z)$ is not invertible are in $[-R, R]$. To conclude, it suffices to prove that for all $c, d \in \mathbb{R}\backslash([a,b] \cup \{z_1, \ldots, z_p\})$ such that $c < d < a$ or $b < c < d$, we have:

(H) the number $C_{c,d}(n)$ of z's in (c,d) such that $\det M(n,z) = 0$ tends to the cardinality $C_{c,d}$ of the i's in $\{1, \ldots, p\}$ such that $c < z_i < d$.

(The assumption about the ranks following then from the fact that the set of matrices with rank at least $r-1$ is open in the set of $r \times r$ matrices).

To prove (H), by additivity, one can suppose that c and d are close enough to have $C_{c,d} = 0$ or 1. Let us define γ to be the circle with diameter $[c,d]$. By a), $\det M_G(\cdot)$ does not vanish on γ, thus

$$C_{c,d} = \frac{1}{2i\pi} \int_\gamma \frac{\partial_z \det M_G(z)}{\det M_G(z)} \mathrm{d}z = \lim_{n\to\infty} \frac{1}{2i\pi} \int_\gamma \frac{\partial_z \det M(n,z)}{\det M(n,z)} \mathrm{d}z,$$

the last equality following from e). Since $C_{c,d} = 0$ or 1, no ambiguity due to the orders of the zeros has to be taken into account here, and it follows that for n large enough, $C_{c,d}(n) = C_{c,d}$. □

References

[1] G. Anderson, A. Guionnet, and O. Zeitouni. An introduction to random matrices. To appear in Cambridge University Press.
[2] P. Arbenz, W. Gander, and G. H. Golub. Restricted rank modification of the symmetric eigenvalue problem: theoretical considerations. *Linear Algebra Appl.*, 104:75–95, 1988.
[3] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

[4] J. Baik, and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6): 1382–1408, 2006.
[5] A. Barvinok. Measure concentration lecture notes. Available online at `http://www.math.lsa.umich.edu/ barvinok/total710.pdf`
[6] K. E. Bassler, P. J. Forrester, and N. E. Frankel. Eigenvalue separation in some random matrix models. *J. Math. Phys.*, 50(3):033302, 24, 2009.
[7] F. Benaych-Georges Infinitely divisible distributions for rectangular free convolution: classification and matricial interpretation. *Probab. Theory and Related Fields*, 139, 1–2, 2007, 143–189.
[8] F. Benaych-Georges. On a surprising relation between the Marchenko-Pastur law, rectangular and square free convolutions. To appear in *Annales de l'IHP Prob. Stats.* Available online at `http://arxiv.org/abs/0808.3938`, 2009.
[9] F. Benaych-Georges. Rectangular random matrices, related convolution. *Probab. Theory Related Fields*, 144(3-4):471–515, 2009.
[10] F. Benaych-Georges Rectangular R-transform at the limit of rectangular spherical integrals, Available online at `http://arxiv.org/abs/0909.0178`, 2009.
[11] P. Biane. Processes with free increments. *Math. Z.*, 227(1):143–174, 1998.
[12] J. R. Bunch, C. P. Nielsen, and D. C. Sorensen. Rank-one modification of the symmetric eigenproblem. *Numer. Math.*, 31(1):31–48, 1978/79.
[13] M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations. *Ann. Probab.*, 37(1):1–47, 2009.
[14] B. Collins. Product of random projections, Jacobi ensembles and universality problems arising from free probability. *Probab. Theory Related Fields*, 133(3):315–344, 2005.
[15] B. Collins and P. Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Comm. Math. Phys.*, 264(3):773–795, 2006.
[16] N. El Karoui. Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.*, 35(2):663–714, 2007.
[17] D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.*, 272(1):185–228, 2007.
[18] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8, 252–261, 1980.
[19] F. Hiai and D. Petz. *The semicircle law, free random variables and entropy*, volume 77 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2000.
[20] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, Cambridge, 1985.
[21] D. C. Hoyle and M. Rattray. Statistical mechanics of learning multiple orthogonal signals: asymptotic theory and fluctuation effects. *Phys. Rev. E (3)*, 75(1):016101, 13, 2007.
[22] I. C. F. Ipsen and B. Nadler. Refined perturbation bounds for eigenvalues of Hermitian and non-Hermitian matrices. *SIAM J. Matrix Anal. Appl.*, 31(1):40–53, 2009.
[23] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.
[24] R. R. Nadakuditi and J. W. Silverstein. Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples. Available online at `http://arxiv.org/abs/0902.4250v1`, 2009.
[25] B. Nadler. Finite sample approximation results for principal component analysis: a matrix perturbation approach. *Ann. Statist.*, 36(6):2791–2817, 2008.
[26] A. Nica and R. Speicher. *Lectures on the combinatorics of free probability*, volume 335 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2006.
[27] D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica*, 17(4):1617–1642, 2007.
[28] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields*, 134(1):127–173, 2006.
[29] N. R. Rao and A. Edelman. The polynomial method for random matrices. *Found. Comput. Math.*, 8(6), 649–702, 2008.
[30] J. W. Silverstein Some limit theorems on the eigenvectors of large-dimensional sample covariance matrices *J. Multivariate Anal.*, 15(3), 295–324, 1984.

[31] J. W. Silverstein On the eigenvectors of large-dimensional sample covariance matrices *J. Multivariate Anal.*, 30(1), 1–16, 1989.
[32] J. W. Silverstein Weak convergence of random functions defined by the eigenvectors of sample covariance matrices *J. Multivariate Anal.*, 18(3), 1–16, 1990.
[33] G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1990.
[34] D. V. Voiculescu, K. J. Dykema, and A. Nica. *Free random variables*, volume 1 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, 1992. A noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.
[35] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

Florent Benaych-Georges, LPMA, UPMC Univ Paris 6, Case courier 188, 4, Place Jussieu, 75252 Paris Cedex 05, France, and CMAP, École Polytechnique, route de Saclay, 91128 Palaiseau Cedex, France.

E-mail address: florent.benaych@upmc.fr

URL: http://www.cmapx.polytechnique.fr/~benaych/

Raj Rao Nadakuditi, Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, Ann Arbor, MI 48109. USA.

E-mail address: rajnrao@eecs.umich.edu

URL: http://www.eecs.umich.edu/~rajnrao/