Journal of Multivariate Analysis 111 (2012) 120–135

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

The singular values and vectors of low rank perturbations of large rectangular random matrices

Florent Benaych-Georges [a,b], Raj Rao Nadakuditi [c,*]

[a] LPMA, UPMC Univ Paris 6, Case courier 188, 4, Place Jussieu, 75252 Paris Cedex 05, France
[b] CMAP, École Polytechnique, route de Saclay, 91128 Palaiseau Cedex, France
[c] Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:
Received 9 March 2011
Available online 3 May 2012

AMS 2000 subject classifications:
15A52
46L54
60F99

Keywords:
Random matrices
Haar measure
Free probability
Phase transition
Random eigenvalues
Random eigenvectors
Random perturbation
Sample covariance matrices

ABSTRACT

In this paper, we consider the singular values and singular vectors of finite, low rank perturbations of large rectangular random matrices. Specifically, we prove almost sure convergence of the extreme singular values and appropriate projections of the corresponding singular vectors of the perturbed matrix.

As in the prequel, where we considered the eigenvalues of Hermitian matrices, the non-random limiting value is shown to depend explicitly on the limiting singular value distribution of the unperturbed matrix via an integral transform that linearizes rectangular additive convolution in free probability theory. The asymptotic position of the extreme singular values of the perturbed matrix differs from that of the original matrix if and only if the singular values of the perturbing matrix are above a certain critical threshold which depends on this same aforementioned integral transform.

We examine the consequence of this singular value phase transition on the associated left and right singular eigenvectors and discuss the fluctuations of the singular values around these non-random limits.

1. Introduction

In many applications, the $n \times m$ signal-plus-noise data or measurement matrix formed by stacking the m samples or measurements of $n \times 1$ observation vectors alongside each other can be modeled as

$$\widetilde{X} = \sum_{i=1}^{r} \sigma_i u_i v_i^* + X, \tag{1}$$

where u_i and v_i are left and right 'signal' column vectors, σ_i are the associated 'signal' values and X is the noise-only matrix of random noises. This model is ubiquitous in signal processing [51,47], statistics [40,2,34] and machine learning [36] and is known under various guises as a signal subspace model [48], a latent variable statistical model [35], or a probabilistic PCA model [50].

Relative to this model, a common application-driven objective is to estimate the signal subspaces $\mathrm{Span}\{u_1, \ldots, u_r\}$ and $\mathrm{Span}\{v_1, \ldots, v_r\}$ that contain signal energy. This is accomplished by computing the *singular value decomposition* (*SVD*,

* Corresponding author.
E-mail addresses: florent.benaych@upmc.fr (F. Benaych-Georges), rajnrao@umich.edu, rajnrao@eecs.umich.edu (R.R. Nadakuditi).
URLs: http://www.cmapx.polytechnique.fr/~benaych/ (F. Benaych-Georges), http://www.eecs.umich.edu/~rajnrao/ (R.R. Nadakuditi).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.04.019

henceforth) of $\widetilde{X}$ and extracting the r largest singular values and the associated singular vectors of $\widetilde{X}$—these are referred to as the r *principal components* [46] and the Eckart–Young–Mirsky theorem states that they provide the best rank-r approximation of the matrix $\widetilde{X}$ for any unitarily invariant norm [24,39]. This theoretical justification along with the fact that these vectors can be efficiently computed using now-standard numerical algorithms for the SVD [28] has led to the ubiquity of the SVD in applications such as array processing [51], genomics [1,52], wireless communications [25], information retrieval [27], to list a few [37,23].

In this paper, motivated by emerging high-dimensional statistical applications [33], we place ourselves in the setting where n and m are large, r is known (or provided by an oracle) and the SVD of $\widetilde{X}$ is used to form estimates of $\{\sigma_i\}$, $\{u_i\}_{i=1}^r$ and $\{v_i\}_{i=1}^r$. We provide a characterization of the relationship between the estimated extreme singular values of $\widetilde{X}$ and the underlying (or latent) 'signal' singular values σ_i (and also the angle between the estimated and true singular vectors).

In the limit of large matrices, the extreme singular values only depend on integral transforms of the distribution of the singular values of the noise-only matrix X in (1) and exhibit a phase transition about a critical value; this critical value depends on integral transforms which arise from rectangular free probability theory [10,11]. The phase transition in the singular value is a new manifestation of the so-called *BBP phase transition*, named after the authors of the seminal paper [5] that first brought into focus this phenomenon for the eigenvalues of a special class of 'spiked' Wishart or sample covariance matrices. In this paper, we also characterize the fluctuations of the singular values about these asymptotic limits. The results obtained are precise in the large matrix limit and, akin to our results in [17], go beyond answers that might be obtained using matrix perturbation theory [49].

Our results are very general in terms of possible distributions for the noise model X, in a sense that which will be made more precise shortly; consequently, our theorems yield as a special case, results found in the literature for the eigenvalues [5,6] and eigenvectors [32,44,42] of $\widetilde{X}\widetilde{X}^*$ in the setting where X in (1) is Gaussian. For the Gaussian setting, we provide a new characterization for the right singular vectors, or equivalently, the eigenvectors of $\widetilde{X}^*\widetilde{X}$.

Such results had already been proved in the particular case where X is a Gaussian matrix, but our approach brings to light a general principle, which can be applied beyond the Gaussian case. Roughly speaking, this principle says that for X a $n \times m$ matrix (with $n, m \gg 1$), if one adds an independent small rank perturbation $\sum_{i=1}^r \sigma_i u_i v_i^*$ to X, then the extreme singular values will move to positions which are approximately the solutions z of the equations

$$\frac{1}{n}\operatorname{Tr}\frac{z}{z^2 I - XX^*} \times \frac{1}{m}\operatorname{Tr}\frac{z}{z^2 I - X^*X} = \frac{1}{\theta_i^2}, \quad (1 \le i \le r).$$

In the case where these equations have no solutions (which means that the θ_i's are below a certain threshold), then the extreme singular values of X will not move significantly. We also provide similar results for the associated left and right singular vectors and give limit theorems for the fluctuations. These expressions provide the basis for the parameter estimation algorithm developed by Hachem et al. in [29].

The papers [17,15] considered the eigenvalues of finite rank perturbations of Hermitian matrices. We employ the strategy developed in these papers for our proofs in this paper. Specifically, we derive master equation representations that implicitly encode the relationship between the singular values and singular vectors of X and $\widetilde{X}$ in terms of the low-rank perturbing matrix. We then employ concentration results to obtain the stated analytical expressions. Of course, because of these similarities in the proofs, we chose to focus, in the present paper, in what differs from [17,15].

At a certain level, our proof also present analogies with the ones of other papers devoted to other occurrences of the BBP phase transition, such as [45,26,20–22,41]. We mention that the approach of the paper [16] could also be used to consider large deviations of the extreme singular values of $\widetilde{X}$.

This paper is organized as follows. We state our main results in Section 2 and provide some examples in Section 3. The proofs are provided in Sections 4–7 with some technical details relegated to the Appendix.

2. Main results

2.1. Definitions and hypotheses

Let X_n be a $n \times m$ real or complex random matrix. Throughout this paper we assume that $n \le m$ so that we may simplify the exposition of the proofs. We may do so without loss of generality because in the setting where $n > m$, the expressions derived will hold for X_n^*. Let the $n \le m$ singular values[1] of X_n be $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$. Let μ_{X_n} be the empirical singular value distribution, *i.e.*, the probability measure defined as

$$\mu_{X_n} = \frac{1}{n}\sum_{i=1}^n \delta_{\sigma_i}.$$

Let m depend on n—we denote this dependence explicitly by m_n which we will sometimes omit for brevity by substituting m for m_n. Assume that as $n \longrightarrow \infty$, $n/m_n \longrightarrow c \in [0, 1]$. In the following, we shall need some of the following hypotheses.

[1] Recall that for $n \le m$, the *singular values* of an $n \times m$ matrix X are the eigenvalues of the $n \times n$ matrix $\sqrt{XX^*}$.

Assumption 2.1. The probability measure μ_{X_n} converges almost surely weakly to a non-random compactly supported probability measure μ_X.

Examples of random matrices satisfying Assumption 2.1 can be found in, for example, [7,18,8,10,3,43]. Note that the question of isolated extreme singular values is not addressed in papers like [7,43]; moreover, in [7,43], the perturbation considered is of unbounded rank.

We now state an assumption about the smallest singular value of X_n. Note that since we assumed that $n \leq m$ and X_n is an $n \times m$ random matrix, whenever X_n has full rank (with high probability), the smallest singular value (which are the square root of the eigenvalues of $X_n X_n^*$—see Footnote 1) will be greater than zero.

Assumption 2.2. Let a be infimum of the support of μ_X. The smallest singular value of X_n converges almost surely to a.

Assumption 2.3. Let b be supremum of the support of μ_X. The largest singular value of X_n converges almost surely to b.

Examples of random matrices satisfying Assumptions 2.2 and 2.3 can be found in e.g. [19,8,3,43].

In this problem, we shall consider the extreme singular values and the associated singular vectors of $\widetilde{X}_n$, which is the random $n \times m$ matrix:

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

where P_n is defined below.

For a given $r \geq 1$, let $\theta_1 \geq \cdots \geq \theta_r > 0$ be deterministic non-zero real numbers, chosen independently of n. For every n, let $G_u^{(n)}$, $G_v^{(n)}$ be two independent matrices with sizes respectively $n \times r$ and $m \times r$, with i.i.d. entries distributed according to a fixed probability measure ν on $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$. We introduce the column vectors $u_1, \ldots, u_r \in \mathbb{K}^{n\times 1}$ and $v_1, \ldots, v_r \in \mathbb{K}^{m\times 1}$ obtained from $G_u^{(n)}$ and $G_v^{(n)}$ by either:

(1) Setting u_i and v_i to equal the i-th column of $\frac{1}{\sqrt{n}} G_u^{(n)}$ and $\frac{1}{\sqrt{m}} G_v^{(n)}$ respectively or,

(2) Setting u_i and v_i to equal to the vectors obtained from a Gram–Schmidt (or QR factorization) of $G_u^{(n)}$ and $G_v^{(n)}$ respectively.

We shall refer to the model (1) as the *i.i.d. model* and to the model (2) as the *orthonormalized model*. With the u_i's and v_i's constructed as above, we define the random perturbing matrix $P_n \in \mathbb{K}^{n\times m}$ as

$$P_n = \sum_{i=1}^{r} \theta_i u_i v_i^*.$$

In the orthonormalized model, the θ_i's are the non zero singular values of P_n and the u_i's and the v_i's are the left and right associated singular vectors.

We make the following hypothesis on the law ν of the entries of $G_u^{(n)}$ and $G_v^{(n)}$ (see [3, Section 2.3.2] for the definition of log-Sobolev inequalities).

Assumption 2.4. The probability measure ν has mean zero, variance one and that satisfies a log-Sobolev inequality.

Remark 2.5. We also note if ν is the standard real or complex Gaussian distribution, then the singular vectors produced using the orthonormalized model will have uniform distribution on the set of r orthogonal random vectors.

Remark 2.6. If X_n is random but has a bi-unitarily invariant distribution and P_n is non-random with rank r, then we are in same setting as the orthonormalized model for the results that follow. More generally, our idea in defining both of our models (the i.i.d. one and the orthonormalized one) was to show that if P_n is chosen independently from X_n in a somehow "isotropic way" (i.e. via a distribution which is not far away from being invariant by the action of the orthogonal group by conjugation), then a BBP phase transition occurs, which is governed by a certain integral transform of the limit empirical singular value distribution of X_n, namely μ_X.

Remark 2.7. We note that there is small albeit non-zero probability that r i.i.d. copies of a random vector are not linearly independent. Consequently, there is a small albeit non-zero probability that the r vectors obtained as in (2) via the Gram–Schmidt orthogonalization may not be well defined. However, in the limit of large matrices, this process produces well-defined vectors with overwhelming probability (indeed, by Proposition A.2, the determinant of the associated $r \times r$ Gram matrix tends to one). This is implicitly assumed in what follows.

Remark 2.8. Our work could easily be adapted to the framework where the distribution of the entries of $G_u^{(n)}$ and the distribution of the entries of $G_u^{(n)}$ are not the same, both satisfying Assumption 2.4.

2.2. Notation

Throughout this paper, for f a function and $d \in \mathbb{R}$, we set

$$f(d^+) := \lim_{z \downarrow d} f(z); \qquad f(d^-) := \lim_{z \uparrow d} f(z),$$

we also let $\xrightarrow{\text{a.s.}}$ denote almost sure convergence. The (ordered) singular values of an $n \times m$ matrix M will be denoted by $\sigma_1(M) \geq \cdots \geq \sigma_n(M)$. Last, for a subspace F of a Euclidean space E and a unit vector $x \in E$, we denote the norm of the orthogonal projection of x onto F by $\langle x, F \rangle$.

2.3. Largest singular values and singular vectors phase transition

In Theorems 2.9–2.11, we suppose Assumptions 2.1, 2.3 and 2.4 to hold.

We define $\overline{\theta}$, the threshold of the phase transition, by the formula

$$\overline{\theta} := (D_{\mu_X}(b^+))^{-1/2},$$

with the convention that $(+\infty)^{-1/2} = 0$, and where D_{μ_X}, the D-transform of μ_X is the function, depending on c, defined by

$$D_{\mu_X}(z) := \left[\int \frac{z}{z^2 - t^2} d\mu_X(t)\right] \times \left[c \int \frac{z}{z^2 - t^2} d\mu_X(t) + \frac{1-c}{z}\right] \quad \text{for } z > b.$$

In the theorems below, $D_{\mu_X}^{-1}(\cdot)$ will denote its functional inverse on $[b, +\infty)$.

Theorem 2.9 (*Largest Singular Value Phase Transition*). *The r largest singular values of the $n \times m$ perturbed matrix $\widetilde{X}_n$ exhibit the following behavior as $n, m_n \to \infty$ and $n/m_n \to c$. We have that for each fixed $1 \leq i \leq r$,*

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

Moreover, for each fixed $i > r$, we have that $\sigma_i(\widetilde{X}_n) \xrightarrow{\text{a.s.}} b$.

Theorem 2.10 (*Norm of Projection of Largest Singular Vectors*). *Consider indices $i_0 \in \{1, \ldots, r\}$ such that $\theta_{i_0} > \overline{\theta}$. For each n, define $\widetilde{\sigma}_{i_0} = \sigma_{i_0}(\widetilde{X}_n)$ and let $\widetilde{u}$ and $\widetilde{v}$ be left and right unit singular vectors of $\widetilde{X}_n$ associated with the singular value $\widetilde{\sigma}_{i_0}$. Then we have, as $n \longrightarrow \infty$,*

(a)

$$|\langle \widetilde{u}, \mathrm{Span}\{u_i \text{s.t.} \theta_i = \theta_{i_0}\}\rangle|^2 \xrightarrow{\text{a.s.}} \frac{-2\varphi_{\mu_X}(\rho)}{\theta_{i_0}^2 D'_{\mu_X}(\rho)}, \tag{2}$$

(b)

$$|\langle \widetilde{v}, \mathrm{Span}\{v_i \text{s.t.} \theta_i = \theta_{i_0}\}\rangle|^2 \xrightarrow{\text{a.s.}} \frac{-2\varphi_{\widetilde{\mu}_X}(\rho)}{\theta_{i_0}^2 D'_{\mu_X}(\rho)}, \tag{3}$$

where $\rho = D_{\mu_X}^{-1}(1/\theta_{i_0}^2)$ is the limit of $\widetilde{\sigma}_{i_0}$ and $\widetilde{\mu}_X = c\mu_X + (1-c)\delta_0$ and for any probability measure μ,

$$\varphi_\mu(z) := \int \frac{z}{z^2 - t^2} d\mu(t). \tag{4}$$

(c) *Furthermore, in the same asymptotic limit, we have*

$$|\langle \widetilde{u}, \mathrm{Span}\{u_i \text{s.t.} \theta_i \neq \theta_{i_0}\}\rangle|^2 \xrightarrow{\text{a.s.}} 0, \quad \text{and} \quad |\langle \widetilde{v}, \mathrm{Span}\{v_i \text{s.t.} \theta_i \neq \theta_{i_0}\}\rangle|^2 \xrightarrow{\text{a.s.}} 0,$$

and

$$\langle \varphi_{\mu_X}(\rho) P_n \widetilde{v} - \widetilde{u}, \mathrm{Span}\{u_i \text{s.t.} \theta_i = \theta_{i_0}\}\rangle \xrightarrow{\text{a.s.}} 0.$$

Theorem 2.11 (*Largest Singular Vector Phase Transition*). *When $r = 1$, let the sole singular value of P_n be denoted by θ. Suppose that*

$$\theta \leq \overline{\theta} \quad \text{and} \quad \varphi'_{\mu_X}(b^+) = -\infty. \tag{5}$$

For each n, let $\widetilde{u}$ and $\widetilde{v}$ denote, respectively, left and right unit singular vectors of $\widetilde{X}_n$ associated with its largest singular value. Then

$$\langle \widetilde{u}, \ker(\theta^2 I_n - P_n P_n^*)\rangle \xrightarrow{\text{a.s.}} 0, \quad \text{and} \quad \langle \widetilde{v}, \ker(\theta^2 I_m - P_n^* P_n)\rangle \xrightarrow{\text{a.s.}} 0,$$

as $n \longrightarrow \infty$.

The following proposition allows us to assert that in many classical matrix models, the threshold $\overline{\theta}$ of the above phase transitions is positive. The proof relies on a straightforward computation which we omit.

Proposition 2.12 (*Edge Density Decay Condition for Phase Transition*)**.** *Assume that the limiting singular distribution μ_X has a density f_{μ_X} with a power decay at b, i.e., that, as $t \to b$ with $t < b$, $f_{\mu_X}(t) \sim M(b-t)^\alpha$ for some exponent $\alpha > -1$ and some constant M. Then*

$$\overline{\theta} = (D_{\mu_X}(b^+))^{-1/2} > 0 \iff \alpha > 0 \quad \text{and} \quad \varphi'_{\mu_X}(b^+) = -\infty \iff \alpha \le 1,$$

so that the phase transitions in Theorems 2.9 and 2.11 manifest for $\alpha = 1/2$.

Remark 2.13 (*Necessity of Singular Value Repulsion for the Singular Vector Phase Transition*)**.** Under additional hypotheses on the manner in which the empirical singular distribution of $X_n \xrightarrow{\text{a.s.}} \mu_X$ as $n \longrightarrow \infty$, Theorem 2.11 can be generalized to any singular value with limit b such that $D'_{\mu_X}(\rho)$ is infinite. The specific hypothesis has to do with requiring the spacings between the singular values of X_n to be more "random matrix like" and exhibit repulsion instead of being "independent sample like" with possible clumping. We plan to develop this line of inquiry in a separate paper.

2.4. Smallest singular values and vectors for square matrices

We now consider the phase transition exhibited by the smallest singular values and vectors. We restrict ourselves to the setting where $\widetilde{X}_n$ is a square matrix; this restriction is necessary because the non-monotonicity of the function D_{μ_X} on $[0, a)$ when $c = \lim n/m < 1$, poses some technical difficulties that do not arise in the square setting. Moreover, in Theorems 2.14–2.16, we assume that Assumptions 2.1, 2.2 and 2.4 hold.

We define $\underline{\theta}$, the threshold of the phase transition, by the formula

$$\underline{\theta} := (\varphi_{\mu_X}(a^-))^{-1},$$

with the convention that $(+\infty)^{-1} = 0$, and where $\varphi_{\mu_X}(z) = \int \frac{z}{z^2-t^2} d\mu(t)$, as in Eq. (4).

In the theorems below, $\varphi_{\mu_X}^{-1}(\cdot)$ will denote the functional inverse of the function $\varphi_{\mu_X}(\cdot)$ on $(0, a)$.

Theorem 2.14 (*Smallest Singular Value Phase Transition for Square Matrices*)**.** *When $a > 0$ and $m = n$, the r smallest singular values of $\widetilde{X}_n$ exhibit the following behavior. We have that for each fixed $1 \le i \le r$,*

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

Moreover, for each fixed $i > r$, we have that $\sigma_{n+1-i}(\widetilde{X}_n) \xrightarrow{a.s.} a$.

Theorem 2.15 (*Norm of Projection of Smallest Singular Vector for Square Matrices*)**.** *Consider indices $i_0 \in \{1, \ldots, r\}$ such that $\theta_{i_0} > \underline{\theta}$. For each n, define $\widetilde{\sigma}_{i_0} = \sigma_{n+1-i_0}(\widetilde{X}_n)$ and let $\widetilde{u}$ and $\widetilde{v}$ be left and right unit singular vectors of $\widetilde{X}_n$ associated with the singular value $\widetilde{\sigma}_{i_0}$. Then we have, as $n \longrightarrow \infty$,*

(a)

$$|\langle \widetilde{u}, \text{Span}\{u_i \text{s.t.} \theta_i = \theta_{i_0}\}\rangle|^2 \xrightarrow{a.s.} \frac{-1}{\varphi'_{\mu_X}(\rho)}, \tag{6}$$

(b)

$$|\langle \widetilde{v}, \text{Span}\{v_i \text{s.t.} \theta_i = \theta_{i_0}\}\rangle|^2 \xrightarrow{a.s.} \frac{-1}{\varphi'_{\mu_X}(\rho)}. \tag{7}$$

(c) *Furthermore, in the same asymptotic limit, we have*

$$|\langle \widetilde{u}, \text{Span}\{u_i \text{s.t.} \theta_i \neq \theta_{i_0}\}\rangle|^2 \xrightarrow{a.s.} 0, \quad \text{and} \quad |\langle \widetilde{v}, \text{Span}\{v_i \text{s.t.} \theta_i \neq \theta_{i_0}\}\rangle|^2 \xrightarrow{a.s.} 0$$

and

$$\langle \varphi_{\mu_X}(\rho) P_n \widetilde{v} - \widetilde{u}, \text{Span}\{u_i \text{s.t.} \theta_i = \theta_{i_0}\}\rangle \xrightarrow{a.s.} 0.$$

Theorem 2.16 (*Smallest Singular Vector Phase Transition*)**.** *When $r = 1$ and $m = n$, let the smallest singular value of $\widetilde{X}_n$ be denoted by $\widetilde{\sigma}_n$ with $\widetilde{u}$ and $\widetilde{v}$ representing the associated left and right unit singular vectors, respectively. Suppose that*

$$a > 0, \qquad \theta \le \underline{\theta} \quad \text{and} \quad \varphi'_{\mu_X}(a^-) = -\infty.$$

Then

$$\langle \widetilde{u}, \ker(\theta^2 I_n - P_n P_n^*)\rangle \xrightarrow{a.s.} 0, \quad \text{and} \quad \langle \widetilde{v}, \ker(\theta^2 I_m - P_n^* P_n)\rangle \xrightarrow{a.s.} 0,$$

as $n \longrightarrow \infty$.

The analogue of Remark 2.13 also applies here.

2.5. The D-transform in free probability theory

The C-*transform with ratio* c of a probability measure μ on $\mathbb{R}_+$, defined as

$$C_\mu(z) = U\left(z(D_\mu^{-1}(z))^2 - 1\right), \tag{8}$$

where the function U, defined as

$$U(z) = \begin{cases} \dfrac{-c-1+\left[(c+1)^2+4cz\right]^{1/2}}{2c} & \text{when } c > 0, \\ z & \text{when } c = 0, \end{cases}$$

is the analogue of the logarithm of the Fourier transform for the *rectangular free convolution with ratio* c (see [12,14] for an introduction to the theory of rectangular free convolution) in the sense described next.

Let A_n and B_n be independent $n \times m$ rectangular random matrices that are invariant, in law, by conjugation by any orthogonal (or unitary) matrix. Suppose that, as $n, m \to \infty$ with $n/m \to c$, the empirical singular value distributions μ_{A_n} and μ_{B_n} of A_n and B_n satisfy $\mu_{A_n} \longrightarrow \mu_A$ and $\mu_{B_n} \longrightarrow \mu_B$. Then by [10], the empirical singular value distribution $\mu_{A_n+B_n}$ of $A_n + B_n$ satisfies $\mu_{A_n+B_n} \longrightarrow \mu_A \boxplus_c \mu_B$, where $\mu_A \boxplus_c \mu_B$ is a probability measure which can be characterized in terms of the C-transform as

$$C_{\mu_A \boxplus \mu_B}(z) = C_{\mu_A}(z) + C_{\mu_B}(z).$$

The coefficients of the series expansion of $U(z)$ are the *rectangular free cumulants with ratio* c of μ (see [11] for an introduction to the rectangular free cumulants). The connection between free rectangular additive convolution and D_μ^{-1} (via the C-transform) and the appearance of D_μ^{-1} in Theorem 2.9 could be of independent interest to free probabilists: the emergence of this transform in the study of isolated singular values completes the picture of [17], where the transforms linearizing additive and multiplicative free convolutions already appeared in similar contexts.

2.6. Fluctuations of the largest singular value

Assume that the empirical singular value distribution of X_n converges to μ_X faster than $1/\sqrt{n}$. More precisely,

Assumption 2.17. We have

$$\frac{n}{m_n} = c + o\left(\frac{1}{\sqrt{n}}\right),$$

$r = 1, \theta := \theta_1 > \overline{\theta}$ and

$$\frac{1}{n}\operatorname{Tr}(\rho^2 I_n - X_n X_n^*)^{-1} = \int \frac{1}{\rho^2 - t^2} d\mu_X(t) + o\left(\frac{1}{\sqrt{n}}\right)$$

for $\rho = D_{\mu_X}^{-1}(1/\theta^2)$ the limit of $\sigma_1(\widetilde{X}_n)$.

We also make the following hypothesis on the law ν. In fact, without the hypothesis on the fourth moment, we would still have a limit theorem on the fluctuations of the largest singular value (this hypothesis is made here to lighten the presentation, since the computations simplify considerably whenever the fourth moment matches that of the Gaussian distribution). The reader who wishes to extend the result in that direction may easily adapt Theorem 3.4 of [15] using the specific arguments developed herein.

Assumption 2.18. If ν is entirely supported by the real line, $\int x^4 d\nu(x) = 3$. If ν is not entirely supported by the real line, the real and imaginary parts of a ν-distributed random variables are independent and identically distributed with $\int |z|^4 d\nu(z) = 2$.

Note that we do not ask ν to be symmetric and make no hypothesis about its third moment. The reason is that the main ingredient of the following theorem is Theorem 6.4 of [15] (or Theorem 7.1 of [9]), where no hypothesis of symmetry or about the third moment is made.

Theorem 2.19. *Suppose Assumptions 2.1, 2.3, 2.4, 2.17 and 2.18 hold. Let $\widetilde{\sigma}_1$ denote the largest singular value of $\widetilde{X}_n$. Then as $n \longrightarrow \infty$,*

$$n^{1/2}\left(\widetilde{\sigma}_1 - \rho\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2),$$

where $\rho = D_{\mu_X}^{-1}(c, 1/\theta^2)$ *and*

$$s^2 = \begin{cases} \dfrac{f^2}{2\beta} & \text{for the i.i.d. model,} \\ \dfrac{f^2-2}{2\beta} & \text{for the orthonormalized model,} \end{cases}$$

with $\beta = 1$ *(or 2) when X is real (or complex) and*

$$f^2 := \frac{\int \frac{d\mu_X(t)}{(\rho^2-t^2)^2}}{\left(\int \frac{d\mu_X(t)}{\rho^2-t^2}\right)^2} + \frac{\int \frac{d\widetilde{\mu}_X(t)}{(\rho^2-t^2)^2}}{\left(\int \frac{d\widetilde{\mu}_X(t)}{\rho^2-t^2}\right)^2} + 2\frac{\int \frac{t^2 d\mu_X(t)}{(\rho^2-t^2)^2}}{\int \frac{\rho d\mu_X(t)}{\rho^2-t^2} \int \frac{\rho d\widetilde{\mu}_X(t)}{\rho^2-t^2}},$$

with $\widetilde{\mu}_X = c\mu_X + (1-c)\delta_0$.

2.7. Fluctuations of the smallest singular value of square matrices

When $m_n = n$ so that $c = 1$, assume that:

Assumption 2.20. For all $n, m_n = n, r = 1, \theta := \theta_1 > \underline{\theta}$ and

$$\frac{1}{n}\operatorname{Tr}(\rho^2 I_n - X_n X_n^*)^{-1} = \int \frac{1}{\rho^2 - t^2} d\mu_X(t) + o\left(\frac{1}{\sqrt{n}}\right)$$

for $\rho := \varphi_{\mu_X}^{-1}(1/\theta)$ the limit of the smallest singular value of $\widetilde{X}_n$.

Theorem 2.21. *Suppose Assumptions 2.1, 2.2, 2.4, 2.18 and 2.20 to hold. Let* $\widetilde{\sigma}_n$ *denote the smallest singular value of* $\widetilde{X}_n$*. Then as* $n \longrightarrow \infty$

$$n^{1/2}\,(\widetilde{\sigma}_n - \rho) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2),$$

where

$$s^2 = \begin{cases} \dfrac{f^2}{2\beta} & \text{for the i.i.d. model} \\ \dfrac{f^2-2}{2\beta} & \text{for the orthonormalized model} \end{cases}$$

with $\beta = 1$ *(or 2) when X is real (or complex) and* $f^2 := 2\theta^2 \int \frac{\kappa^2+t^2}{(\kappa^2-t^2)^2} d\mu_X(t)$.

3. Examples

3.1. Gaussian rectangular random matrices with non-zero mean

Let X_n be an $n \times m$ real (or complex) matrix with independent, zero mean, normally distributed entries with variance $1/m$. It is known [38,8] that, as $n, m \longrightarrow \infty$ with $n/m \to c \in (0, 1]$, the spectral measure of the singular values of X_n converges to the distribution with density

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

where $a = 1 - \sqrt{c}$ and $b = 1 + \sqrt{c}$ are the end points of the support of μ_X. It is known [8] that the extreme eigenvalues converge to the bounds of this support.

Associated with this singular measure, we have, by an application of the result in [13, Section 4.1] and Eq. (8),

$$D_{\mu_X}^{-1}(z) = \sqrt{\frac{(z+1)(cz+1)}{z}},$$

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

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

$$\sigma_i(X_n + P_n) \xrightarrow{\text{a.s.}} \begin{cases} \sqrt{\dfrac{(1+\theta_i^2)(c+\theta_i^2)}{\theta_i^2}} & \text{if } i \leq r \text{ and } \theta_i > c^{1/4} \\ 1+\sqrt{c} & \text{otherwise} \end{cases} \tag{9}$$

as $n \longrightarrow \infty$. As far as the i.i.d. model is concerned, this formula allows us to recover some of the results of [5].

Now, let us turn our attention to the singular vectors. In the setting where $r = 1$, let $P_n = \theta uv^*$. Then, by Theorems 2.10 and 2.11, we have

$$|\langle \widetilde{u}, u\rangle|^2 \xrightarrow{\text{a.s.}} \begin{cases} 1 - \dfrac{c(1+\theta^2)}{\theta^2(\theta^2+c)} & \text{if } \theta \geq c^{1/4}, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

The phase transitions for the eigenvectors of $\widetilde{X}_n^*\widetilde{X}_n$ or for the pairs of singular vectors of $\widetilde{X}_n$ can be similarly computed to yield the expression:

$$|\langle \widetilde{v}, v\rangle|^2 \xrightarrow{\text{a.s.}} \begin{cases} 1 - \dfrac{(c+\theta^2)}{\theta^2(\theta^2+1)} & \text{if } \theta \geq c^{1/4}, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

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 limiting spectral measure

$$\mu_X(x) = \delta_1,$$

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

Associated with this spectral measure, we have (of course, $c = 1$)

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

thus for all $\theta > 0$,

$$D_{\mu_X}^{-1}(1/\theta^2) = \begin{cases} \dfrac{\theta + \sqrt{\theta^2+4}}{2} & \text{if the inverse is computed on } (1, +\infty), \\ \dfrac{-\theta + \sqrt{\theta^2+4}}{2} & \text{if the inverse is computed on } (0, 1). \end{cases}$$

Thus for any $n \times n$, rank r perturbing 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

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

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. Proof of Theorems 2.9 and 2.14

The proofs of both theorems are quite similar. As a consequence, we only prove Theorem 2.9.

The sequence of steps described below yields the desired proof (which is very close to the one of Theorem 2.1 of [17]):

(1) The first, rather trivial, step in the proof of Theorem 2.9 is to use Weyl's interlacing inequalities to prove that any fixed-rank singular value of $\widetilde{X}_n$ which does not tend to a limit $> b$ tends to b.
(2) Then, we utilize Lemma 4.1 below to express the extreme singular values of $\widetilde{X}_n$ as the z's such that a certain random $2r \times 2r$ matrix $M_n(z)$ is singular.
(3) We then exploit convergence properties of certain analytical functions (derived in the Appendix) to prove that almost surely, $M_n(z)$ converges to a certain deterministic matrix $M(z)$, uniformly in z.
(4) We then invoke a continuity lemma (see Lemma A.1 in the Appendix) to claim that almost surely, the z's such that $M_n(z)$ is singular (*i.e.* the extreme singular values of $\widetilde{X}_n$) converge to the z's such that $M(z)$ is singular.

(5) We conclude the proof by noting that, for our setting, the z's such that $M(z)$ is singular are precisely the z's such that for some $i \in \{1, \dots, r\}$, $D_{\mu_X}(z) = \frac{1}{\theta_i^2}$. Part (ii) of Lemma A.1, about the rank of $M_n(z)$, will be useful to assert that when the θ_i's are pairwise distinct, the multiplicities of the isolated singular values are all equal to one.

First, up to a conditioning by the σ-algebra generated by the X_n's, one can suppose them to be deterministic and all the randomness supported by the perturbing matrix P_n.

Second, by Horn and Johnson [31, Theorem 3.1.2], one has, for all $i \geq 1$,

$$\sigma_{i+r}(X_n) \leq \sigma_i(\widetilde{X}_n) \leq \sigma_{i-r}(X_n)$$

with the convention $\sigma_j(X_n) = +\infty$ for $i \leq 0$ and 0 for $i > n$. By the same proof as in [17, Section 6.2.1], it follows that for all $i \geq 1$ fixed,

$$\liminf \sigma_i(\widetilde{X}_n) \geq b \tag{12}$$

and that for all fixed $i > r$,

$$\sigma_i(\widetilde{X}_n) \underset{n\to\infty}{\longrightarrow} b \tag{13}$$

(we insist here on the fact that i has to be fixed, i.e. not to depend on n: of course, for $i = n/2$, (13) is not true anymore in general).

Our approach is based on the following lemma, which reduces the problem to the study of $2r \times 2r$ random matrices. Recall that the constants $r, \theta_1, \dots, \theta_r$, and the random column vectors (which depend on n, even though this dependence does not appear in the notation) $u_1, \dots, v_r, v_1, \dots, v_r$ have been introduced in Section 2.1 and that the perturbing matrix P_n is given by

$$P_n = \sum_{i_1}^{r} \theta_i u_i v_i^*.$$

Recall also that the singular values of X_n are denoted by $\sigma_1 \geq \cdots \geq \sigma_n$. Let us define the matrices

$$\Theta = \operatorname{diag}(\theta_1, \dots, \theta_r) \in \mathbb{R}^{r\times r}, \qquad U_n = \begin{bmatrix} u_1 \cdots u_r \end{bmatrix} \in \mathbb{K}^{n\times r}, \qquad V_m = \begin{bmatrix} v_1 \cdots v_r \end{bmatrix} \in \mathbb{K}^{m\times r}.$$

Lemma 4.1. *The positive singular values of $\widetilde{X}_n$ which are not singular values of X_n are the $z \notin \{\sigma_1, \dots, \sigma_n\}$ such that the $2r \times 2r$ matrix*

$$M_n(z) := \begin{bmatrix} U_n^*(z^2 I_n - X_n X_n^*)^{-1} U_n & U_n^*(z^2 I_n - X_n X_n^*)^{-1} X_n V_m \\ V_m^* X_n^*(z^2 I_n - X_n X_n^*)^{-1} U_n & V_m^*(z^2 I_m - X_n^* X_n)^{-1} V_m \end{bmatrix} - \begin{bmatrix} 0 & \Theta^{-1} \\ \Theta^{-1} & 0 \end{bmatrix}$$

is not invertible.

For the sake of completeness, we provide a proof, even though several related results can be found in the literature (see e.g. [4,16]).

Proof. First, [30, Theorem 7.3.7] states that the non-zero singular values of $\widetilde{X}_n$ are the positive eigenvalues of $\begin{bmatrix} 0 & \widetilde{X}_n \\ \widetilde{X}_n^* & 0 \end{bmatrix}$. Second, for any $z > 0$ which is not a singular value of X_n, by Benaych-Georges et al. [15, Lemma 6.1],

$$\det\left(zI_n - \begin{bmatrix} 0 & \widetilde{X}_n \\ \widetilde{X}_n^* & 0 \end{bmatrix}\right) = \det\left(zI_n - \begin{bmatrix} 0 & X_n \\ X_n^* & 0 \end{bmatrix}\right)^{-1} \times \prod_{i=1}^{r} \theta_i^2 \times \det M_n(z),$$

which allows us to conclude, since by hypothesis, $\det\left(zI_{n+m} - \begin{bmatrix} 0 & X_n \\ X_n^* & 0 \end{bmatrix}\right)^{-1} \neq 0$. $\square$

Note that by Assumption 2.1,

$$\frac{1}{n} \operatorname{Tr} \frac{z}{z^2 I_n - X_n X_n^*} \underset{n\to\infty}{\longrightarrow} \int \frac{z}{z^2 - t^2} d\mu_X(t),$$
$$\frac{1}{m} \operatorname{Tr} \frac{z}{z^2 I_m - X_n^* X_n} \underset{n\to\infty}{\longrightarrow} \int \frac{z}{z^2 - t^2} d\widetilde{\mu}_X(t) \quad (\widetilde{\mu}_X = c\mu_X + (1-c)\delta_0),$$

uniformly on any subset of $\{z \in \mathbb{C} \text{s.t.} \Re(z) > b + \eta\}$, $\eta > 0$. It follows, by a direct application of Ascoli's Theorem and Proposition A.2, that almost surely, we have the following convergence (which is uniform in z)

$$U_n^* \frac{z}{z^2 I_n - X_n X_n^*} U_n \underset{n\to\infty}{\longrightarrow} \left(\int \frac{z}{z^2 - t^2} d\mu_X(t)\right) \cdot I_r,$$
$$V_m^* \frac{z}{z^2 I_m - X_n^* X_n} V_m \underset{n\to\infty}{\longrightarrow} \left(\int \frac{z}{z^2 - t^2} d\widetilde{\mu_X}(t)\right) \cdot I_r.$$

In the same way, almost surely

$$U_n^*(z^2I_n - X_nX_n^*)^{-1}X_nV_m \underset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad V_m^*X_n^*(z^2I_n - X_nX_n^*)^{-1}U_n \underset{n\to\infty}{\longrightarrow} 0.$$

It follows that almost surely,

$$M_n(z) \underset{n\to\infty}{\longrightarrow} M(z) := \begin{bmatrix} \varphi_{\mu_X}(z)I_r & 0 \\ 0 & \varphi_{\widetilde{\mu}_X}(z)I_r \end{bmatrix} - \begin{bmatrix} 0 & \Theta^{-1} \\ \Theta^{-1} & 0 \end{bmatrix}, \tag{14}$$

where φ_{μ_X} and $\varphi_{\widetilde{\mu}_X}$ are the functions defined in the statement of Theorem 2.10.

Now, note that once (12) has been established, our result only concerns the number of singular values of $\widetilde{X}_n$ in $[b+\eta, +\infty)$ (for any $\eta > 0$), hence can be proved via Lemma A.1. Indeed, by Hypothesis 2.3, for n large enough, X_n has no singular value $> b + \eta$, thus numbers $> b + \eta$ cannot be in the same time singular values of X_n and $\widetilde{X}_n$.

In the case where the θ_i's are pairwise distinct, Lemma A.1 allows us to conclude the proof of Theorem 2.9. Indeed, Lemma A.1 says that exactly as much singular values of $\widetilde{X}_n$ as predicted by the theorem have limits $> b$ and that their limits are exactly the ones predicted by the Theorem. The part of the theorem devoted to singular values tending to b can then be deduced from (12) and (13).

In the case where the θ_i's are not pairwise distinct, an approximation approach allows us to conclude (proceed for example as in Section 6.2.3 of [17], using [30, Corollary 7.3.8 (b)] instead of [30, Corollary 6.3.8]).

5. Proof of Theorems 2.10 and 2.15

The proofs of both theorems are quite similar. As a consequence, we only prove Theorem 2.10.

As above, up to a conditioning by the σ-algebra generated by the X_n's, one can suppose them to be deterministic and all the randomness supported by the perturbing matrix P_n.

First, by the Law of Large Numbers, even in the *i.i.d. model*, the u_i's and the v_i's are almost surely asymptotically orthonormalized. More specifically, for all $i \neq j$,

$$\langle u_i, u_j\rangle \underset{n\to\infty}{\longrightarrow} \mathbb{1}_{i=j}$$

(the same being true for the v_i's). As a consequence, it is enough to prove that

(a')

$$\sum_{i \text{ s.t. } \theta_i=\theta_{i_0}} |\langle \widetilde{u}, u_i\rangle|^2 \xrightarrow{\text{a.s.}} \frac{-2\varphi_{\mu_X}(\rho)}{\theta_{i_0}^2 D'_{\mu_X}(\rho)}, \tag{15}$$

(b')

$$\sum_{i \text{ s.t. } \theta_i=\theta_{i_0}} |\langle \widetilde{v}, v_i\rangle|^2 \xrightarrow{\text{a.s.}} \frac{-2\varphi_{\widetilde{\mu}_X}(\rho)}{\theta_{i_0}^2 D'_{\mu_X}(\rho)}, \tag{16}$$

(c')

$$\sum_{i \text{ s.t. } \theta_i\neq\theta_{i_0}} |\langle \widetilde{u}, u_i\rangle|^2 + |\langle \widetilde{v}, v_i\rangle|^2 \xrightarrow{\text{a.s.}} 0, \tag{17}$$

(d')

$$\sum_{i \text{ s.t. } \theta_i=\theta_{i_0}} |\varphi_{\mu_X}(\rho)\theta_{i_0}\langle \widetilde{v}, v_i\rangle - \langle \widetilde{u}, u_i\rangle|^2 \xrightarrow{\text{a.s.}} 0. \tag{18}$$

Again, the proof is based on a lemma which reduces the problem to the study of the kernel of a random $2r \times 2r$ matrix. The matrices Θ, U_n and V_m are the ones introduced before Lemma 4.1.

Lemma 5.1. *Let z be a singular value of $\widetilde{X}_n$ which is not a singular value of X_n and let (u, v) be a corresponding singular pair of unit vectors. Then the column vector*

$$\begin{bmatrix} \Theta V_m{}^* v \\ \Theta U_n{}^* u \end{bmatrix}$$

belongs to the kernel of the $2r \times 2r$ matrix $M_n(z)$ introduced in Lemma 4.1. Moreover, we have

$$v^*P_n^*\frac{z^2}{(z^2I_n - X_nX_n^*)^2}P_nv + u^*P_n\frac{X_n^*X_n}{(z^2I_m - X_n^*X_n)^2}P_n^*u + v^*P_n^*\frac{z}{(z^2I_n - X_nX_n^*)^2}X_nP_n^*u$$
$$+ u^*P_nX_n^*\frac{z}{(z^2I_n - X_nX_n^*)^2}P_nv = 1. \tag{19}$$

Proof. The first part of the lemma is easy to verify with the formula $X_n^* f(X_nX_n^*) = f(X_n^*X_n)X_n^*$ for any function f defined on $[0, +\infty)$. For the second part, use the formulas

$$\widetilde{X}_n\widetilde{X}_n^* u = z^2 u \quad \text{and} \quad X_n^* u = zv - P_n^* u,$$

to establish $u = (z^2 I_n - X_nX_n^*)^{-1}(zP_n v + X_nP_n^* u)$, and then use the fact that $u^*u = 1$. □

Let us consider z_n, $(\widetilde{u}, \widetilde{v})$ as in the statement of Theorem 2.10. Note first that for n large enough, $z_n > \sigma_1(X_n)$, hence Lemma 5.1 can be applied, and the vector

$$\begin{bmatrix} \Theta {V_m}^* \widetilde{v} \\ \Theta {U_n}^* \widetilde{u} \end{bmatrix} = \left[\theta_1\langle v_1, \widetilde{v}\rangle, \dots, \theta_r\langle v_r, \widetilde{v}\rangle, \theta_1\langle u_1, \widetilde{u}\rangle, \dots, \theta_r\langle u_r, \widetilde{u}\rangle\right]^T \tag{20}$$

belongs to $\ker M_n(z_n)$. As explained in the proof of Theorem 2.9, the random matrix-valued function $M_n(\cdot)$ converges almost surely uniformly to the matrix-valued function $M(\cdot)$ introduced in Eq. (14). Hence $M_n(z_n)$ converges almost surely to $M(\rho)$, and it follows that the orthogonal projection on $(\ker M(\rho))^\perp$ of the vector of (20) tends almost surely to zero.

Let us now compute this projection. For x, y column vectors of $\mathbb{K}^r$,

$$\begin{aligned} M(\rho)\begin{bmatrix} x \\ y \end{bmatrix} = 0 &\iff \forall i, \quad y_i = \theta_i\varphi_{\mu_X}(\rho)x_i \quad \text{and} \quad x_i = \theta_i\varphi_{\widetilde{\mu}_X}(\rho)y_i \\ &\iff \forall i, \quad \begin{cases} x_i = y_i = 0 & \text{if } \theta_i^2\varphi_{\mu_X}(\rho)\varphi_{\widetilde{\mu}_X}(\rho) \neq 1, \\ y_i = \theta_i\varphi_{\mu_X}(\rho)x_i & \text{if } \theta_i^2\varphi_{\mu_X}(\rho)\varphi_{\widetilde{\mu}_X}(\rho) = 1. \end{cases} \end{aligned}$$

Note that ρ is precisely defined by the relation $\theta_{i_0}^2\varphi_{\mu_X}(\rho)\varphi_{\widetilde{\mu}_X}(\rho) = 1$. Hence with $\beta := -\theta_{i_0}\varphi_{\widetilde{\mu}_X}(\rho)$, we have

$$\ker M(\rho) = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{K}^{r+r} \text{s.t.} \forall i, x_i = y_i = 0 \text{ if } \theta_i \neq \theta_{i_0} \text{ and } y_i = -\beta x_i \text{ if } \theta_i = \theta_{i_0}\right\},$$

hence

$$(\ker M(\rho))^\perp = \left\{\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{K}^{r+r} \text{s.t.} \forall i, x_i = \beta y_i \text{ if } \theta_i = \theta_{i_0}\right\}$$

and the orthogonal projection of any vector $\begin{bmatrix} x \\ y \end{bmatrix}$ on $(\ker M(\rho))^\perp$ is the vector $\begin{bmatrix} x' \\ y' \end{bmatrix}$ such that for all i,

$$(x_i', y_i') = \begin{cases} (x_i, y_i) & \text{if } \theta_i \neq \theta_{i_0}, \\ \dfrac{\beta x_i + y_i}{\beta^2 + 1}(\beta, 1) & \text{if } \theta_i = \theta_{i_0}. \end{cases}$$

Then, (17) and (18) are direct consequences of the fact that the projection of the vector of (20) on $(\ker M(\rho))^\perp$ tends to zero.

Let us now prove (16). By (19), we have

$$a_n + b_n + c_n + d_n = 1, \tag{21}$$

with

$$a_n = \widetilde{v}^* P_n^* \frac{z_n^2}{(z_n^2 I_n - X_nX_n^*)^2} P_n\widetilde{v} = \sum_{i,j=1}^r \theta_i\theta_j\overline{\langle v_i, \widetilde{v}\rangle}\langle v_j, \widetilde{v}\rangle {u_i}^* \frac{z_n^2}{(z_n^2 I_n - X_nX_n^*)^2} u_j \tag{22}$$

$$b_n = \widetilde{u}^* P_n \frac{X_n^*X_n}{(z_n^2 I_m - X_n^*X_n)^2} P_n^*\widetilde{u} = \sum_{i,j=1}^r \theta_i\theta_j\overline{\langle u_i, \widetilde{u}\rangle}\langle u_j, \widetilde{u}\rangle {v_i}^* \frac{X_n^*X_n}{(z_n^2 I_m - X_n^*X_n)^2} v_j \tag{23}$$

$$c_n = \widetilde{v}^* P_n^* \frac{z_n}{(z_n^2 I_n - X_nX_n^*)^2} X_nP_n^*\widetilde{u} = \sum_{i,j=1}^r \theta_i\theta_j\overline{\langle v_i, \widetilde{v}\rangle}\langle u_j, \widetilde{u}\rangle {u_i}^* \frac{z_n}{(z_n^2 I_n - X_nX_n^*)^2} X_n v_j$$

$$d_n = \widetilde{u}^* P_n X_n^* \frac{z_n}{(z_n^2 I_n - X_nX_n^*)^2} P_n\widetilde{v} = \sum_{i,j=1}^r \theta_i\theta_j\overline{\langle u_i, \widetilde{u}\rangle}\langle v_j, \widetilde{v}\rangle {v_i}^* X_n^* \frac{z_n}{(z_n^2 I_n - X_nX_n^*)^2} u_j.$$

Since the limit of z_n is out of the support of μ_X, one can apply Proposition A.2 to assert that both c_n and d_n have almost sure limit zero and that in the sums (22) and (23), any term such that $i \neq j$ tends almost surely to zero. Moreover, by (17), these sums can also be reduced to the terms with index i such that $\theta_i = \theta_{i_0}$. To sum up, we have

$$a_n = \theta_{i_0}^2 \sum_{i \text{s.t.} \theta_i = \theta_{i_0}} |\langle v_i, \widetilde{v}\rangle|^2 {u_i}^* \frac{z_n^2}{(z_n^2 I_n - X_nX_n^*)^2} u_i + o(1)$$

$$b_n = \theta_{i_0}^2 \sum_{i \text{s.t.} \theta_i = \theta_{i_0}} |\langle u_i, \widetilde{u}\rangle|^2 {v_i}^* \frac{X_n^*X_n}{(z_n^2 I_m - X_n^*X_n)^2} v_i + o(1).$$

Now, note that since z_n tends to ρ,

$$\frac{1}{n}\operatorname{Tr}\frac{z_n^2}{(z_n^2 I_n - X_n X_n^*)^2} \underset{n\to\infty}{\longrightarrow} \int \frac{\rho^2}{(\rho^2-t^2)^2} d\mu_X(t),$$

$$\frac{1}{m_n}\operatorname{Tr}\frac{X_n^* X_n}{(z_n^2 I_m - X_n^* X_n)^2} \underset{n\to\infty}{\longrightarrow} \int \frac{t^2}{(\rho^2-t^2)^2} d\widetilde{\mu}_X(t),$$

hence by Proposition A.2, almost surely,

$$a_n = \theta_{i_0}^2 \int \frac{\rho^2}{(\rho^2-t^2)^2} d\mu_X(t) \sum_{i \text{ s.t. } \theta_i=\theta_{i_0}} |\langle v_i, \widetilde{v}\rangle|^2 + o(1),$$

$$b_n = \theta_{i_0}^2 \int \frac{t^2}{(\rho^2-t^2)^2} d\widetilde{\mu}_X(t) \sum_{i \text{ s.t. } \theta_i=\theta_{i_0}} |\langle u_i, \widetilde{u}\rangle|^2 + o(1).$$

Moreover, by (18), for all i such that $\theta_i = \theta_{i_0}$,

$$|\langle u_i, \widetilde{u}\rangle|^2 = \theta_{i_0}^2 (\varphi_{\mu_{XX}}(\rho))^2 |\langle v_i, \widetilde{v}\rangle|^2 + o(1).$$

It follows that

$$b_n = \theta_{i_0}^4 (\varphi_{\mu_{XX}}(\rho))^2 \int \frac{t^2}{(\rho^2-t^2)^2} d\widetilde{\mu}_X(t) \sum_{i \text{ s.t. } \theta_i=\theta_{i_0}} |\langle v_i, \widetilde{v}\rangle|^2 + o(1).$$

Since $a_n + b_n = 1 + o(1)$, we get

$$\sum_{i \text{ s.t. } \theta_i=\theta_{i_0}} |\langle v_i, \widetilde{v}\rangle|^2 \underset{n\to\infty}{\longrightarrow} \left(\theta_{i_0}^2 \int \frac{\rho^2}{(\rho^2-t^2)^2} d\mu_X(t) + \theta_{i_0}^4 (\varphi_{\mu_X}(\rho))^2 \int \frac{t^2}{(\rho^2-t^2)^2} d\widetilde{\mu}_X(t)\right)^{-1}.$$

The relations

$$\theta_{i_0}^2 \varphi_{\mu_X}(\rho)\varphi_{\widetilde{\mu}_X}(\rho) = 1$$

$$2\int \frac{\rho^2}{(\rho^2-t^2)^2} d\mu_X(t) = \frac{1}{\rho}\varphi_{\mu_X}(\rho) - \varphi'_{\mu_X}(\rho)$$

$$2\int \frac{t^2}{(\rho^2-t^2)^2} d\widetilde{\mu}_X(t) = -\frac{1}{\rho}\varphi_{\widetilde{\mu}_X}(\rho) - \varphi'_{\widetilde{\mu}_X}(\rho)$$

allow to recover the RHS of (16) easily. Via (18), one easily deduces (15).

6. Proof of Theorems 2.11 and 2.16

Again, we shall only prove Theorem 2.11 and suppose the X_n's to be non random.

Let us consider the matrix $M_n(z)$ introduced in Lemma 4.1. Here, $r = 1$, so one easily gets, for each n,

$$\lim_{z\to+\infty} \det M_n(z) = -\theta^{-2}.$$

Moreover, for $b_n := \sigma_1(X_n)$ the largest singular value of X_n, looking carefully at the term in $\frac{1}{z^2-b_n^2}$ in $\det M_n(z)$, it appears that with a probability which tends to one as $n \longrightarrow \infty$, we have

$$\lim_{z\to b_n} \det M_n(z) = +\infty.$$

It follows that with a probability which tends to one as $n \longrightarrow \infty$, the largest singular value $\widetilde{\sigma}_1$ of $\widetilde{X}_n$ is $> b_n$.

Then, one concludes using the second part of Lemma 5.1, as in the proof of Theorem 2.3 of [17].

7. Proof of Theorems 2.19 and 2.21

We shall only prove Theorem 2.19, because Theorem 2.21 can be proved similarly.

We have supposed that $r = 1$. Let us denote $u = u_1$ and $v = v_1$. Then we have

$$P_n = \theta u v^*,$$

with $u \in \mathbb{K}^{n\times 1}$, $v \in \mathbb{K}^{m\times 1}$ random vectors whose entries are ν-distributed independent random variables, renormalized in the *orthonormalized model*, and divided by respectively $\sqrt{n}$ and $\sqrt{m}$ in the *i.i.d. model*. We also have that the matrix $M_n(z)$ defined in Lemma 4.1 is a 2×2 matrix.

Let us fix an arbitrary b^* such that $b < b^* < \rho$. Theorem 2.9 implies that almost surely, for n large enough, $\det[M_n(\cdot)]$ vanishes exactly once in (b^*, ∞). Since moreover, almost surely, for all n,

$$\lim_{z\to+\infty} \det[M_n(z)] = -\frac{1}{\theta^2} < 0,$$

we deduce that almost surely, for n large enough, $\det[M_n(z)] > 0$ for $b^* < z < \widetilde{\sigma}_1$ and $\det[M_n(z)] < 0$ for $\widetilde{\sigma}_1 < z$.

As a consequence, for any real number x, for n large enough,

$$\sqrt{n}(\widetilde{\sigma}_1 - \rho) < x \iff \det M_n\left(\rho + \frac{x}{\sqrt{n}}\right) > 0. \tag{24}$$

Therefore, we have to understand the limit distributions of the entries of $M_n\left(\rho + \frac{x}{\sqrt{n}}\right)$. They are given by the following.

Lemma 7.1. *For any fixed real number x, as $n \longrightarrow \infty$, the distribution of*

$$\Gamma_n := \sqrt{n}\left(M_n\left(\rho + \frac{x}{\sqrt{n}}\right) - \begin{bmatrix} \varphi_{\mu_X}(\rho) & -\theta^{-1} \\ -\theta^{-1} & \varphi_{\widetilde{\mu}_X}(\rho)\end{bmatrix}\right)$$

converges weakly to the one of

$$x\begin{bmatrix} \varphi_{\mu_X}(\rho) & 0 \\ 0 & \varphi_{\widetilde{\mu}_X}(\rho)\end{bmatrix} + \begin{bmatrix} c_1 X & dZ \\ d\overline{Z} & c_2 Y\end{bmatrix},$$

for X, Y, Z (resp. $X, Y, \Re(Z), \Im(Z)$) independent standard real Gaussian variables if $\beta = 1$ (resp. if $\beta = 2$) and for c_1, c_2, d some real constants given by the following formulas:

$$c_1^2 = \begin{cases} \dfrac{2}{\beta}\displaystyle\int \frac{\rho^2}{(\rho^2 - t^2)^2} d\mu_X(t) & \text{in the i.i.d. model,} \\ \dfrac{2}{\beta}\left(\displaystyle\int \frac{\rho^2}{(\rho^2 - t^2)^2} d\mu_X(t) - (\varphi_{\mu_X}(\rho))^2\right) & \text{in the orthonormalized model,}\end{cases} \tag{25}$$

$$c_2^2 = \begin{cases} \dfrac{2}{\beta}\displaystyle\int \frac{\rho^2}{(\rho^2 - t^2)^2} d\widetilde{\mu}_X(t) & \text{in the i.i.d. model,} \\ \dfrac{2}{\beta}\left(\displaystyle\int \frac{\rho^2}{(\rho^2 - t^2)^2} d\widetilde{\mu}_X(t) - (\varphi_{\widetilde{\mu}_X}(\rho))^2\right) & \text{in the orthonormalized model,}\end{cases} \tag{26}$$

$$d^2 = \frac{1}{\beta}\int \frac{t^2}{(\rho^2 - t^2)^2} d\mu_X(t). \tag{27}$$

Proof. Let us define $z_n := \rho + \frac{x}{\sqrt{n}}$. We have

$$\Gamma_n = \sqrt{n}\begin{bmatrix} \dfrac{1}{n}u^*\dfrac{z_n}{z_n^2 I_n - X_n X_n^*}u - \varphi_{\mu_X}(\rho) & \dfrac{1}{\sqrt{nm_n}}u^*(z_n^2 I_n - X_n X_n^*)^{-1}X_n v \\ \dfrac{1}{\sqrt{nm_n}}v^* X_n^*(z_n^2 I_n - X_n X_n^*)^{-1}u & \dfrac{1}{m_n}v^*\dfrac{z_n}{z_n^2 I_m - X_n^* X_n}v - \varphi_{\widetilde{\mu}_X}(\rho)\end{bmatrix}.$$

Let us for example expand the upper left entry of $\Gamma_{n,1,1}$ of Γ_n. We have

$$\begin{aligned}\Gamma_{n,1,1} &= \sqrt{n}\left(\frac{1}{n}u^*\frac{z_n}{z_n^2 I_n - X_n X_n^*}u - \varphi_{\mu_X}(\rho)\right) \\ &= \sqrt{n}\left(\frac{1}{n}u^*\frac{z_n}{z_n^2 I_n - X_n X_n^*}u - \frac{1}{n}\operatorname{Tr}\frac{z_n}{z_n^2 I_n - X_n X_n^*}\right) + \sqrt{n}\left(\frac{1}{n}\operatorname{Tr}\frac{z_n}{z_n^2 I_n - X_n X_n^*} - \varphi_{\mu_X}(z_n)\right) \\ &\quad + \sqrt{n}\left(\varphi_{\mu_X}(z_n) - \varphi_{\mu_X}(\rho)\right).\end{aligned} \tag{28}$$

The third term of the RHS of (28) tends to $x\varphi'_{\mu_X}(\rho)$ as $n \longrightarrow \infty$. By Taylor–Lagrange Formula, there is $\xi_n \in (0, 1)$ such that the second one is equal to

$$\sqrt{n}\left(\frac{1}{n}\operatorname{Tr}\frac{\rho}{\rho^2 I_n - X_n X_n^*} - \varphi_{\mu_X}(\rho)\right) + x\frac{\partial}{\partial z}_{|z=\rho+\xi_n x/\sqrt{n}}\left(\frac{1}{n}\operatorname{Tr}\frac{z}{z^2 I_n - X_n X_n^*} - \varphi_{\mu_X}(z)\right),$$

hence tends to zero, by Assumptions 2.1 and 2.20. To sum up, we have

$$\Gamma_{n,1,1} = \sqrt{n}\left(\frac{1}{n}u^*\frac{z_n}{z_n^2 I_n - X_n X_n^*}u - \frac{1}{n}\operatorname{Tr}\frac{z_n}{z_n^2 I_n - X_n X_n^*}\right) + x\varphi'_{\mu_X}(\rho) + o(1). \tag{29}$$

In the same way, we have

$$\Gamma_{n,2,2} = \sqrt{n}\left(\frac{1}{m_n}v^*\frac{z_n}{z_n^2 I_m - X_n^* X_n}v - \frac{1}{m_n}\operatorname{Tr}\frac{z_n}{z_n^2 I_m - X_n^* X_n}\right) + x\varphi'_{\widetilde{\mu}_X}(\rho) + o(1). \tag{30}$$

Then the "$\kappa_4(\nu) = 0$" case of Theorem 6.4 of [15] allows us to conclude. □

Let us now complete the proof of Theorem 2.19. By the previous lemma, we have

$$\det M_n\left(\rho + \frac{x}{\sqrt{n}}\right) = \det\left(\begin{bmatrix}\varphi_{\mu_X}(\rho) & -\theta^{-1}\\ -\theta^{-1} & \varphi_{\widetilde{\mu}_X}(\rho)\end{bmatrix} + \frac{1}{\sqrt{n}}\begin{bmatrix}x\varphi_{\mu_X}(\rho) + c_1 X_n & dZ_n\\ d\overline{Z_n} & x\varphi_{\widetilde{\mu}_X}(\rho) + c_2 Y_n\end{bmatrix}\right)$$

for some random variables X_n, Y_n, Z_n with converging in distribution to the random variables X, Y, Z of the previous lemma. Using the relation $\varphi_{\mu_X}(\rho)\varphi_{\widetilde{\mu}_X}(\rho) = \theta^{-2}$, we get

$$\det M_n\left(\rho + \frac{x}{\sqrt{n}}\right) = 0 + \frac{1}{\sqrt{n}}\left\{2x\theta^{-2} + \varphi_{\mu_X}(\rho)c_2 Y_n + \varphi_{\widetilde{\mu}_X}(\rho)c_1 X_n + \theta^{-1}d(Z_n + \overline{Z_n})\right\} + O\left(\frac{1}{n}\right).$$

Thus by (24), we have

$$\begin{aligned}\lim_{n\to\infty}\mathbb{P}\{\sqrt{n}(\widetilde{\sigma}_1 - \rho) < x\} &= \lim_{n\to\infty}\mathbb{P}\left\{\det M_n\left(\rho + \frac{x}{\sqrt{n}}\right) > 0\right\}\\ &= \mathbb{P}\left\{-\frac{\theta^2}{2}\left(\varphi_{\mu_X}(\rho)c_2 Y + \varphi_{\widetilde{\mu}_X}(\rho)c_1 X + \theta^{-1}d(Z + \overline{Z})\right) < x\right\}.\end{aligned}$$

It follows that the distribution of $\sqrt{n}(\widetilde{\sigma}_1 - \rho)$ converges weakly to the one of sX, for X a standard Gaussian random variable on $\mathbb{R}$ and

$$s^2 = \frac{\theta^4}{4}\left((\varphi_{\widetilde{\mu}_X}(\rho)c_1)^2 + (\varphi_{\mu_{XX}}(\rho)c_2)^2 + 4(\theta^{-1}d)^2\right).$$

One can easily recover the formula given in Theorem 2.19 for s^2, using the relation $\varphi_{\mu_X}(\rho)\varphi_{\widetilde{\mu}_X}(\rho) = \theta^{-2}$.

Acknowledgments

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 work was supported by an Office of Naval Research Young Investigator Award N00014-11-1-0660. The authors thank the reviewers for their careful reading of the work and for their many excellent suggestions which greatly improved the manuscript.

Appendix

We now state the continuity lemma that we use in the proof of Theorem 2.9. We note that nothing in its hypotheses is random. As hinted earlier, we will invoke it to localize the extreme eigenvalues of $\widetilde{X}_n$.

Lemma A.1. *We suppose the positive real numbers $\theta_1, \ldots, \theta_r$ to be pairwise distinct. Let us fix a real number $0 \le b$ and two analytic functions φ_1, φ_2 defined on $\{z \in \mathbb{C}\text{s.t.}\Re(z) > 0\} \setminus [0, b]$ such that for all $i = 1, 2$,*

(a) $\varphi_i(z) \in \mathbb{R} \iff z \in \mathbb{R}$,
(b) *for all $z > b$, $\varphi_i(z) < 0$,*
(c) $\varphi_i(z) \longrightarrow 0$ *as* $|z| \longrightarrow \infty$.

Let us define the $2r \times 2r$-matrix-valued function

$$M(z) := \begin{bmatrix}0 & \Theta^{-1}\\ \Theta^{-1} & 0\end{bmatrix} - \begin{bmatrix}\varphi_1(z)I_r & 0\\ 0 & \varphi_2(z)I_r\end{bmatrix}$$

and denote by $z_1 > \cdots > z_p$ *the z's in* (b, ∞) *such that* $M(z)$ *is not invertible, where* $p \in \{0, \ldots, r\}$ *is the number of* θ_i*'s such that*

$$\lim_{z \downarrow b} \varphi_1(z)\varphi_2(z) > \frac{1}{\theta_i^2}.$$

Let us also consider a sequence $0 < b_n$ *with limit b and, for each n, a* $2r \times 2r$*-matrix-valued function* $M_n(\cdot)$*, defined on*

$$\{z \in \mathbb{C}\text{s.t.}\Re(z) > 0\} \setminus [0, b_n],$$

which coefficient are analytic functions, such that

(d) *for all* $z \notin \mathbb{R}$, $M_n(z)$ *is invertible,*
(e) *for all* $\eta > 0$, $M_n(\cdot)$ *converges to the function* $M(\cdot)$ *uniformly on* $\{z \in \mathbb{C}\text{s.t.}\Re(z) > b + \eta\}$.

Then

(i) *there exist p real sequences* $z_{n,1} > \cdots > z_{n,p}$ *converging respectively to* $z_1, \ldots, z_p$ *such that for any* $\varepsilon > 0$ *small enough, for n large enough, the z's in* $(b + \varepsilon, \infty)$ *such that* $M_n(z)$ *is not invertible are exactly* $z_{n,1}, \ldots, z_{n,p}$,
(ii) *for n large enough, for each i,* $M_n(z_{n,i})$ *has rank* $2r - 1$.

Proof. To prove this lemma, we use the formula

$$\det \begin{bmatrix} xI_r & \operatorname{diag}(\alpha_1, \ldots, \alpha_r) \\ \operatorname{diag}(\alpha_1, \ldots, \alpha_r) & yI_r \end{bmatrix} = \prod_{i=1}^{r} (xy - \alpha_i^2)$$

in the appropriate place and proceed as the proof of Lemma 6.1 in [17]. □

We also need the following proposition. The u_i's and the v_i's are the random column vectors introduced in Section 2.1.

Proposition A.2. *Let, for each n,* A_n, B_n *be complex* $n \times n$, $n \times m$ *matrices which operator norms, with respect to the canonical Hermitian structure, are bounded independently of n. Then for any* $\eta > 0$*, there exists* $C, \alpha > 0$ *such that for all n, for all* $i, j, k \in \{1, \ldots, r\}$ *such that* $i \neq j$,

$$\mathbb{P}\left\{ \left| \langle u_i, A_n u_i \rangle - \frac{1}{n} \operatorname{Tr}(A_n) \right| > \eta \text{ or } |\langle u_i, A_n u_j \rangle| > \eta \text{ or } |\langle u_i, B_n v_k \rangle| > \eta \right\} \leq Ce^{-n^{\alpha}}.$$

Proof. In the *i.i.d. model*, this result is an obvious consequence of [15, Proposition 6.2]. In the *orthonormalized model*, one also has to use [15, Proposition 6.2], which states that the u_i's (the same holds for the v_i's) are obtained from the $n \times r$ matrix $G_u^{(n)}$ with i.i.d. entries distributed according to ν by the following formula: for all $i = 1, \ldots, r$,

$$u_i = \frac{i\text{th column of } G_u^{(n)} \times (W^{(n)})^T}{\|i\text{th column of } G_u^{(n)} \times (W^{(n)})^T\|_2},$$

where $W^{(n)}$ is a (random) $r \times r$ matrix such that for certain positive constants D, c, κ, for all $\varepsilon > 0$ and all n,

$$\mathbb{P}\left\{ \|W^{(n)} - I_r\| > \varepsilon \text{ or } \max_{1 \leq i \leq r} \left| \frac{1}{\sqrt{n}} \|i\text{th column of } G_u^{(n)} \times (W^{(n)})^T\|_2 - 1 \right| > \varepsilon \right\} \leq D(e^{-cn\varepsilon} + e^{-c\sqrt{n}}). \quad \square$$

References

[1] O. Alter, P.O. Brown, D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling, Proc. Nat. Acad. Sci. USA 97 (18) (2000) 10101.
[2] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, second ed., Wiley, New York, 1984.
[3] G. Anderson, A. Guionnet, O. Zeitouni, An Introduction to Random Matrices, in: Cambridge Studies in Advanced Mathematics, vol. 118, 2009.
[4] P. Arbenz, W. Gander, G.H. Golub, Restricted rank modification of the symmetric eigenvalue problem: theoretical considerations, Linear Algebra Appl. 104 (1988) 75–95.
[5] J. Baik, G. Ben Arous, S. Péché, Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices, Ann. Probab. 33 (5) (2005) 1643–1697.
[6] J. Baik, J.W. Silverstein, Eigenvalues of large sample covariance matrices of spiked population models, J. Multivariate Anal. 97 (6) (2006) 1382–1408.
[7] Z.D. Bai, J.W. Silverstein, On the empirical distriubtion of eigenvalues of a class of large dimensional random matrices, J. Multivariate Anal. (1995) 175–192.
[8] Z.D. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, second ed., Springer, New York, 2009.
[9] Z.D. Bai, J.-F. Yao, Central limit theorems for eigenvalues in a spiked population model, Ann. Inst. Henri Poincaré Probab. Stat. 44 (3) (2008) 447–474.
[10] F. Benaych-Georges, Rectangular random matrices, related convolution, Probab. Theory Related Fields 144 (3–4) (2009) 471–515.
[11] F. Benaych-Georges, Rectangular random matrices, related free entropy and free Fisher's information, J. Operator Theory 23 (2) (2010) 447–465.
[12] F. Benaych-Georges, On a surprising relation between the Marchenko–Pastur law, rectangular and square free convolutions, Ann. Inst. Henri Poincaré Probab. Stat. 46 (3) (2010) 644–652.
[13] F. Benaych-Georges, Infinitely divisible distributions for rectangular free convolution: classification and matricial interpretation, Probab. Theory Related Fields 139 (1–2) (2007) 143–189.

[14] F. Benaych-Georges, Rectangular *R*-transform at the limit of rectangular spherical integrals, J. Theoret. Probab. 24 (4) (2011) 969–987.
[15] F. Benaych-Georges, A. Guionnet, M. Maïda, Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices, Electron. J. Probab. 16 (2011) 1621–1662. Paper no. 60.
[16] F. Benaych-Georges, A. Guionnet, M. Maïda, Large deviations of the extreme eigenvalues of random deformations of matrices, Probab. Theory Related Fields (2010) 1–49.
[17] F. Benaych-Georges, R.R. Nadakuditi, The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices, Adv. Math. 227 (1) (2011) 494–521.
[18] M. Capitaine, M. Casalis, Asymptotic freeness by generalized moments for Gaussian and Wishart matrices, application to beta random matrices, Indiana Univ. Math. J. 53 (2) (2004) 397–431.
[19] M. Capitaine, C. Donati-Martin, Strong asymptotic freeness for Wigner and Wishart matrices, Indiana Univ. Math. J. 56 (2) (2007) 767–803.
[20] M. Capitaine, C. Donati-Martin, D. Féral, The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations, Ann. Probab. 37 (2009) 1–47.
[21] M. Capitaine, C. Donati-Martin, D. Féral, Central limit theorems for eigenvalues of deformations of Wigner matrices, Ann. Inst. Henri Poincaré Probab. Stat. 48 (1) (2012) 107–133.
[22] M. Capitaine, C. Donati-Martin, D. Féral, M. Février, Free convolution with a semi-circular convolution and eigenvalues of spiked deformations of Wigner matrices, Electron. J. Probab. 16 (2011) 1750–1792. Paper no. 64.
[23] P. Drineas, A. Frieze, R. Kannan, S. Vempala, V. Vinay, Clustering large graphs via the singular value decomposition, Mach. Learn. 56 (1) (2004) 9–33.
[24] C. Eckart, G. Young, The approximation of one matrix by another of lower rank, Psychometrika 1 (3) (1936) 211–218.
[25] O. Edfors, M. Sandell, J.J. Van de Beek, S.K. Wilson, P.O. Borjesson, OFDM channel estimation by singular value decomposition, IEEE Trans. Commun. 46 (7) (2002) 931–939.
[26] D. Féral, S. Péché, The largest eigenvalue of rank one deformation of large Wigner matrices, Comm. Math. Phys. 272 (2007) 185–228.
[27] G.W. Furnas, S. Deerwester, S.T. Dumais, T.K. Landauer, R.A. Harshman, L.A. Streeter, K.E. Lochbaum, Information retrieval using a singular value decomposition model of latent semantic structure, in: Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 1988, pp. 465–480.
[28] G. Golub, W. Kahan, Calculating the singular values and pseudo-inverse of a matrix, J. Soc. Ind. Appl. Math.: Ser. B, Numer. Anal. 2 (2) (1965) 205–224.
[29] W. Hachem, P. Loubaton, X. Mestre, J. Najim, P. Vallet, A subspace estimator for fixed rank perturbations of large random matrices, preprint, 2011.
[30] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, 1985.
[31] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, 1991.
[32] D.C. Hoyle, M. Rattray, Statistical mechanics of learning multiple orthogonal signals: asymptotic theory and fluctuation effects, Phys. Rev. E (3) 75 (1) (2007) 016101. 13.
[33] I.M. Johnstone, D.M. Titterington, Statistical challenges of high-dimensional data, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 367 (1906) (2009) 4237.
[34] I. Jolliffe, Principal component analysis, Springer Series in Statistics (2002).
[35] M.I. Jordan, Learning in Graphical Models, Kluwer Academic Publishers, 1998.
[36] R. Kannan, S. Vempala, Spectral Algorithms, Now Publishers Inc., 2009.
[37] V. Klema, A. Laub, The singular value decomposition: its computation and some applications, IEEE Trans. Automat. Control 25 (2) (2002) 164–176.
[38] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (NS) 72 (114) (1967) 507–536.
[39] L. Mirsky, Symmetric gauge functions and unitarily invariant norms, Q. J. Math. 11 (1) (1960) 50–59.
[40] R.J. Muirhead, Aspects of Multivariate Statistical Theory, Wiley, New York, 1982.
[41] R.R. Nadakuditi, J.W. Silverstein, Fundamental limit of sample generalized eigenvalue based detection of signals in noise using relatively few signal-bearing and noise-only samples, IEEE J. Sel. Top. Signal Process. 4 (3) (2010) 468–480.
[42] B. Nadler, Finite sample approximation results for principal component analysis: a matrix perturbation approach, Ann. Statist. 36 (6) (2008) 2791–2817.
[43] G. Pan, Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix, J. Multivariate Anal. 101 (2010) 1330–1338.
[44] D. Paul, Asymptotics of sample eigenstructure for a large dimensional spiked covariance model, Statist. Sinica 17 (4) (2007) 1617–1642.
[45] S. Péché, The largest eigenvalue of small rank perturbations of Hermitian random matrices, Probab. Theory Related Fields 134 (2006) 127–173.
[46] C.R. Rao, The use and interpretation of principal component analysis in applied research, Sankhyā 26 (4) (1964) 329–358.
[47] L.L. Scharf, Statistical Signal Processing: Detection, Estimation, and Time Series Analysis, Addison-Wesley, Reading, Massachusetts, 1991.
[48] R.O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. AP-34 (1986) 276–280.
[49] G.W. Stewart, J.G. Sun, Matrix Perturbation Theory, in: Computer Science and Scientific Computing, Academic Press Inc., Boston, MA, 1990.
[50] M.E. Tipping, C.M. Bishop, Probabilistic principal component analysis, J. R. Stat. Soc. Ser. B Stat. Methodol. 61 (3) (1999) 611–622.
[51] H.L. Van Trees, Detection, Estimation, and Modulation Theory Part IV: Optimum Array Processing, John wiley and Sons, Inc., new York, 2002.
[52] M. Wall, A. Rechtsteiner, L. Rocha, Singular value decomposition and principal component analysis, Pract. Approach to Microarray Data Anal. (2003) 91–109.