SINGULAR VALUES OF RANDOM MATRICES

DJALIL CHAFAÏ

Abstract. These notes are an expanded version of short courses given at the occasion of a school held in Université Paris-Est Marne-la-Vallée, 16–20 November 2009, by Djalil Chafaï, Olivier Guédon, Guillaume Lecué, Alain Pajor, and Shahar Mendelson. The central motivation is compressed sensing, involving interactions between empirical processes, high dimensional geometry, and random matrices.

Contents

The extremal singular values of a matrix are very natural geometrical quantities concentrating an essential information on the invertibility and stability of the matrix. This chapter aims to provide an accessible introduction to the notion of singular values of matrices and their behavior when the entries are random, including quite recent striking results from random matrix theory and high dimensional geometric analysis.

For every square matrix $A \in \mathcal{M}_{n,n}(\mathbb{C})$, we denote by $\lambda_1(A), \ldots, \lambda_n(A)$ the eigenvalues of A which are the roots in $\mathbb{C}$ of the characteristic polynomial $\det(A - ZI) \in \mathbb{C}[Z]$. We label the eigenvalues of A so that $|\lambda_1(A)| \geqslant \cdots \geqslant |\lambda_n(A)|$. In all this chapter, K stands for $\mathbb{R}$ or $\mathbb{C}$, and we say that $U \in \mathcal{M}_{n,n}(K)$ is K-unitary when $UU^* = I$.

1. Singular values of deterministic matrices

This section gathers a selection of classical results from linear algebra. We begin with the Singular Value Decomposition (SVD), a fundamental tool in matrix

Date: 16–20 November 2009.

analysis, which expresses a diagonalization up to unitary transformations of the space.

Theorem 1.1 (Singular Value Decomposition)**.** *For every $A \in \mathcal{M}_{m,n}(K)$, there exists a couple of K–unitary matrices U $(m \times m)$ and V $(n \times n)$ and a sequence of real numbers $s_1 \geqslant \cdots \geqslant s_{m\wedge n} \geqslant 0$ such that*

$$U^* A V = \mathrm{diag}(s_1, \ldots, s_{m\wedge n}) \in \mathcal{M}_{m,n}(K).$$

This sequence of real numbers does not depend on the particular choice of U, V.

Proof. Let $v \in K^n$ be such that $|Av|_2 = \max_{|x|_2=1} |Ax|_2 = |A|_{2\to 2} = s$. If $|v|_2 = 0$ then $A = 0$ and the desired result is trivial. If $s > 0$ then let us define $u = Av/s$. One can find a K-unitary matrix U of size $m \times m$ with first column equal to u, and a K-unitary matrix V of size $n \times n$ with first column equal to v. It follows that

$$U^* A V = \begin{pmatrix} s & w^* \\ 0 & B \end{pmatrix} = A_1$$

for some $w \in \mathcal{M}_{n-1,1}(K)$ and $B \in \mathcal{M}_{m-1,n-1}(K)$. If t is the first row of A_1 then $|A_1 t^*|_2^2 \geqslant (s^2 + |w|_2^2)^2$ and therefore $|A_1|_{2\to 2}^2 \geqslant s^2 + |w|_2^2 \geqslant |A|_{2\to 2}^2$. On the other hand, since A and A_1 are unitary equivalent, we have $|A_1|_{2\to 2} = |A|_{2\to 2}$. Therefore $w = 0$, and the desired decomposition follows by a simple induction. □

The numbers $s_k(A) := s_k$ for $k \in \{1, \ldots, m \wedge n\}$ are called the *singular values* of A. The columns of U and V are the eigenvectors of AA^* $(m \times m)$ and A^*A $(n \times n)$. These two positive semidefinite Hermitian matrices share the same sequence of eigenvalues, up to the multiplicity of the eigenvalue 0, and for every $k \in \{1, \ldots, m \wedge n\}$,

$$s_k(A) = \lambda_k(\sqrt{AA^*}) = \sqrt{\lambda_k(AA^*)} = \sqrt{\lambda_k(A^*A)} = \lambda_k(\sqrt{A^*A}) = s_k(A^*).$$

Actually, if one sees the diagonal matrix $D := \mathrm{diag}(s_1(A)^2, \ldots, s_{m\wedge n}(A)^2)$ as an element of $\mathcal{M}_{m,m}(K)$ or $\mathcal{M}_{n,n}(K)$ by appending as much zeros as needed, we have

$$U^* A A^* U = D \quad \text{and} \quad V^* A^* A V = D.$$

When A is normal (i.e. $AA^* = A^*A$) then $m = n$ and $s_k(A) = |\lambda_k(A)|$ for every $k \in \{1, \ldots, n\}$. For any $A \in \mathcal{M}_{m,n}(K)$, the eigenvalues of the $(m + n) \times (m + n)$ Hermitian matrix

$$H = \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix} \tag{1}$$

are given by

$$+s_1(A), -s_1(A), \ldots, +s_{m\wedge n}(A), -s_{m\wedge n}(A), 0, \ldots, 0$$

where the notation $0, \ldots, 0$ stands for a sequence of 0's of length

$$m + n - 2(m \wedge n) = (m \vee n) - (m \wedge n).$$

One may deduce the singular values of A from the eigenvalues of H. Note that when $m = n$ and $A_{i,j} \in \{0, 1\}$ for all i, j, then A is the adjacency matrix of an oriented graph, and H is the adjacency matrix of a compagnon nonoriented bipartite graph.

For any $A \in \mathcal{M}_{m,n}(K)$, the matrices A, $\bar{A}$, $A^\top$, A^*, WA, AW' share the same sequences of singular values, for any K–unitary matrices W, W'. If $u_1 \perp \cdots \perp u_m \in K^m$ and $v_1 \perp \cdots \perp v_n \in K^n$ are the columns of U, V then for every $k \in \{1, \ldots, m \wedge n\}$,

$$A v_k = s_k(A) u_k \quad \text{and} \quad A^* u_k = s_k(A) v_k \tag{2}$$

while $Av_k = 0$ and $A^*u_k = 0$ for $k > m \wedge n$. The SVD gives an intuitive geometrical interpretation of A and A^* as a dual correspondence/dilation between two orthonormal bases known as the left and right eigenvectors of A and A^*. Additionally, A has exactly $r = \mathrm{rank}(A)$ nonzero singular values $s_1(A), \dots, s_r(A)$ and

$$A = \sum_{k=1}^{r} s_k(A) u_k v_k^* \quad \text{and} \quad \begin{cases} \mathrm{kernel}(A) & = \mathrm{span}\{v_{r+1}, \dots, v_n\}, \\ \mathrm{range}(A) & = \mathrm{span}\{u_1, \dots, u_r\}. \end{cases}$$

We have also $s_k(A) = |Av_k|_2 = |A^*u_k|_2$ for every $k \in \{1, \dots, m \wedge n\}$. It is well known that the eigenvalues of a Hermitian matrix can be expressed in terms of the entries of the matrix via minimax variational formulas. The following theorem is the counterpart for the singular values, and can be deduced from its Hermitian cousin.

Theorem 1.2 (Courant–Fischer variational formulas for singular values)**.** *For every $A \in \mathcal{M}_{m,n}(K)$ and every $k \in \{1, \dots, m \wedge n\}$,*

$$s_k(A) = \max_{V \in \mathcal{V}_k} \min_{\substack{x \in V \\ |x|_2 = 1}} |Ax|_2 = \min_{V \in \mathcal{V}_{n-k+1}} \max_{\substack{x \in V \\ |x|_2 = 1}} |Ax|_2$$

where $\mathcal{V}_k$ is the set of subspaces of K^n of dimension k. In particular, we have

$$s_1(A) = \max_{\substack{x \in K^n \\ |x|_2 = 1}} |Ax|_2 \quad \textit{and} \quad s_{m \wedge n}(A) = \max_{V \in \mathcal{V}_{n, m \wedge n}} \min_{\substack{x \in V \\ |x|_2 = 1}} |Ax|_2.$$

We have also the following alternative formulas, for every $k \in \{1, \dots, m \wedge n\}$,

$$s_k(A) = \max_{\substack{V \in \mathcal{V}_k \\ W \in \mathcal{V}_k}} \min_{\substack{(x,y) \in V \times W \\ |x|_2 = |y|_2 = 1}} \langle Ax, y \rangle.$$

As an exercise, one can check that if $A \in \mathcal{M}_{m,n}(\mathbb{R})$ then the variational formulas for $K = \mathbb{C}$, if one sees A as an element of $\mathcal{M}_{m,n}(\mathbb{C})$, coincide actually with the formulas for $K = \mathbb{R}$. Geometrically, the matrix A maps the Euclidean unit ball to an ellipsoid, and the singular values of A are exactly the half lengths of the $m \wedge n$ largest principal axes of this ellipsoid, see figure 1. The remaining axes have a zero length. In particular, for $A \in \mathcal{M}_{n,n}(K)$, the variational formulas for the extremal singular values $s_1(A)$ and $s_n(A)$ correspond to the half length of the longest and shortest axes.

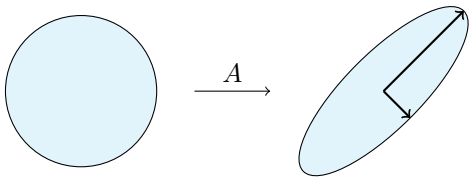

FIGURE 1. Largest and smallest singular values of $A \in \mathcal{M}_{2,2}(\mathbb{R})$.

From the Courant–Fischer variational formulas, the largest singular value is the operator norm of A for the Euclidean norm $|\cdot|_2$, namely

$$s_1(A) = |A|_{2 \to 2}.$$

The map $A \mapsto s_1(A)$ is Lipschitz and convex. In the same spirit, if U, V are the couple of K–unitary matrices from an SVD of A, then for any $k \in \{1, \dots, \mathrm{rank}(A)\}$,

$$s_k(A) = \min_{\substack{B \in \mathcal{M}_{m,n}(K) \\ \mathrm{rank}(B) = k-1}} |A - B|_{2 \to 2} = |A - A_k|_{2 \to 2} \quad \text{where} \quad A_k = \sum_{i=1}^{k-1} s_i(A) u_i v_i^*$$

with u_i, v_i as in (2). Let $A \in \mathcal{M}_{n,n}(K)$ be a square matrix. If A is invertible then the singular values of A^{-1} are the inverses of the singular values of A, in other words

$$\forall k \in \{1, \ldots, n\}, \quad s_k(A^{-1}) = s_{n-k+1}(A)^{-1}.$$

Moreover, a square matrix $A \in \mathcal{M}_{n,n}(K)$ is invertible iff $s_n(A) > 0$, and in this case

$$s_n(A) = s_1(A^{-1})^{-1} = \left|A^{-1}\right|_{2\to 2}^{-1}.$$

Contrary to the map $A \mapsto s_1(A)$, the map $A \mapsto s_n(A)$ is Lipschitz but is not convex. Regarding the Lipschitz nature of the singular values, the Courant–Fischer variational formulas provide the following more general result, which has a Hermitian couterpart.

Theorem 1.3 (Weyl additive perturbations)**.** *If $A, B \in \mathcal{M}_{m,n}(K)$ then for every $i, j \in \{1, \ldots, m \wedge n\}$ with $i + j \leqslant 1 + (m \wedge n)$,*

$$s_{i+j-1}(A) \leqslant s_i(B) + s_j(A - B).$$

In particular, the singular values are uniformly Lipschitz functions since

$$\max_{1 \leqslant k \leqslant m \wedge n} |s_k(A) - s_k(B)| \leqslant |A - B|_{2\to 2}.$$

From the Courant–Fischer variational formulas we obtain also the following result.

Theorem 1.4 (Cauchy interlacing by rows deletion)**.** *Let $A \in \mathcal{M}_{m,n}(K)$ and $k \in \{1, 2, \ldots\}$ with $1 \leqslant k \leqslant m \leqslant n$ and let $B \in \mathcal{M}_{m-k,n}(K)$ be a matrix obtained from A by deleting k rows. Then for every $i \in \{1, \ldots, m - k\}$,*

$$s_i(A) \geqslant s_i(B) \geqslant s_{i+k}(A).$$

In particular we have $[s_{m-k}(B), s_1(B)] \subset [s_m(A), s_1(A)]$. Row deletions produce a sort of compression of the singular values interval. Another way to express this phenomenon consists in saying that if we add a row to B then the largest singular value increases while the smallest singular value is diminished. From this point of view, the worst case corresponds to square matrices. Closely related, the following result on finite rank additive perturbations can be proved by using interlacing inequalities for the eigenvalues of Hermitian matrices and their principal submatrices.

Theorem 1.5 (Interlacing for finite rank additive perturbations [Tho76])**.** *For any $A, B \in \mathcal{M}_{n,n}(K)$ with $\operatorname{rank}(A - B) \leqslant k$, we have, for any $i \in \{1, \ldots, n\}$,*

$$s_{i-k}(A) \geqslant s_i(B) \geqslant s_{i+k}(A)$$

where $s_r = +\infty$ if $r \leqslant 0$ and $s_r = 0$ if $r \geqslant n+1$. Conversely, any sequences of non negative real numbers which satisfy to these interlacing inequalities are the singular values of matrices A and B with $\operatorname{rank}(A - B) \leqslant k$.

In particular, we have $[s_{n-k}(B)\,,\, s_{k+1}(B)] \subset [s_n(A)\,,\, s_1(A)]$. It is worthwhile to observe that the interlacing inequalities of theorem 1.5 give neither an upper bound for the largest singular values $s_1(B), \ldots, s_k(B)$ nor a lower bound for the smallest singular values $s_{n-k+1}(B), \ldots, s_n(B)$, even when $k = 1$.

Remark 1.6 (Hilbert-Schmidt norm)**.** *For every $A \in \mathcal{M}_{m,n}(K)$ we set*

$$\|A\|_{\mathrm{HS}}{}^2 := \operatorname{Tr}(AA^*) = \sum_{i,j=1}^{n} |A_{i,j}|^2 = s_1(A)^2 + \cdots + s_{m\wedge n}(A)^2.$$

This defines the so called Hilbert–Schmidt or Frobenius norm $\|\cdot\|_{\mathrm{HS}}$. We have always

$$|A|_{2\to 2} \leqslant \|A\|_{\mathrm{HS}} \leqslant \sqrt{\operatorname{rank}(A)}\, |A|_{2\to 2}$$

where equalities are achieved when $\mathrm{rank}(A) = 1$ *and* $A = I \in \mathcal{M}_{m,n}(K)$ *respectively. The advantage of* $\|\cdot\|_{\mathrm{HS}}$ *over* $|\cdot|_{2\to 2}$ *lies in its convenient expression in terms of the matrix entries. Actually, the Frobenius norm is Hilbertian for the Hermitian product*

$$\langle A, B\rangle = \mathrm{Tr}(AB^*).$$

Let us mention a result on the Frobenius Lipschitz norm of the singular values, due to Wielandt and Hoffman [HW53]*, which says that if* $A, B \in \mathcal{M}_{m,n}(K)$ *then*

$$\sum_{k=1}^{m\wedge n} (s_k(A) - s_k(B))^2 \leqslant \|A - B\|_{\mathrm{HS}}{}^2.$$

We end up by a result related to the Frobenius norm, due to Eckart and Young [EY39]*. We have seen that a matrix* $A \in \mathcal{M}_{m,n}(K)$ *has exactly* $r = \mathrm{rank}(A)$ *non zero singular values. More generally, if* $k \in \{0, 1, \ldots, r\}$ *and if* $A_k \in \mathcal{M}_{m,n}(K)$ *is obtained from the SVD of* A *by forcing* $s_i = 0$ *for all* $i > k$ *then*

$$\min_{\substack{B\in\mathcal{M}_{m,n}(K)\\ \mathrm{rank}(B)=k}} \|A - B\|_{\mathrm{HS}}{}^2 = \|A - A_k\|_{\mathrm{HS}}{}^2 = s_{k+1}(A)^2 + \cdots + s_r(A)^2.$$

Remark 1.7 (Norms and unitary invariance)**.** *For every* $k \in \{1, \ldots, m \wedge n\}$ *and any real number* $p \geqslant 1$*, the map* $A \in \mathcal{M}_{m,n}(K) \mapsto (s_1(A)^p + \cdots + s_k(A)^p)^{1/p}$ *is a unitary invariant norm on* $\mathcal{M}_{m,n}(K)$*. We recover the operator norm* $|A|_{2\to 2}$ *for* $k = 1$ *and the Frobenius norm* $\|A\|_{\mathrm{HS}}$ *for* $(k, p) = (m \wedge n, 2)$*. The special case* $(k, p) = (m \wedge n, 1)$ *is known as the Ky Fan norm of order* k*, while the special case* $k = m\wedge n$ *is known as the Schatten* p*-norm. For more material, see* [Bha97, Zha02].

1.1. **Condition number.** The condition number of $A \in \mathcal{M}_{n,n}(K)$ is given by

$$\kappa(A) = |A|_{2\to 2}\big|A^{-1}\big|_{2\to 2} = \frac{s_1(A)}{s_n(A)}.$$

The condition number quantifies the numerical sensitivity of linear systems involving A. For instance, if $x \in K^n$ is the solution of the linear equation $Ax = b$ then $x = A^{-1}b$. If b is known up to precision $\delta \in K^n$ then x is known up to precision $A^{-1}\delta$. Therefore, the ratio of relative errors for the determination of x is given by

$$R(b, \delta) = \frac{\big|A^{-1}\delta\big|_2/\big|A^{-1}b\big|_2}{|\delta|_2/|b|_2} = \frac{\big|A^{-1}\delta\big|_2}{|\delta|_2}\frac{|b|_2}{|A^{-1}b|_2}.$$

Consequently, we obtain

$$\max_{b\neq 0,\delta\neq 0} R(b, \delta) = \big|A^{-1}\big|_{2\to 2}|A|_{2\to 2} = \kappa(A).$$

Geometrically, $\kappa(A)$ measures the "spherical defect" of the ellipsoid in figure (1).

1.2. **Basic relationships between eigenvalues and singular values.** We know that if $A \in \mathcal{M}_{n,n}(K)$ is normal (i.e. $AA^* = A^*A$) then $s_k(A) = |\lambda_k(A)|$ for every $k \in \{1, \ldots, n\}$. Beyond normal matrices, for every $A \in \mathcal{M}_{n,n}(K)$ with rows $R_1, \ldots, R_n$, we have, by viewing $|\det(A)|$ as the volume of a hyperparallelogram,

$$|\det(A)| = \prod_{k=1}^{n} |\lambda_k(A)| = \prod_{k=1}^{n} s_k(A) = \prod_{k=1}^{n} \mathrm{dist}(R_k, \mathrm{span}\{R_1, \ldots, R_{k-1}\}) \qquad (3)$$

The following result, due to Weyl, is less global and more subtle.

Theorem 1.8 (Weyl inequalities [Wey49])**.** *If* $A \in \mathcal{M}_{n,n}(K)$*, then*

$$\forall k \in \{1, \ldots, n\}, \quad \prod_{i=1}^{k} |\lambda_i(A)| \leqslant \prod_{i=1}^{k} s_i(A) \quad \textit{and} \quad \prod_{i=k}^{n} s_i(A) \leqslant \prod_{i=k}^{n} |\lambda_i(A)| \qquad (4)$$

Moreover, for every increasing function φ from $(0,\infty)$ to $(0,\infty)$ such that $t \mapsto \varphi(e^t)$ is convex on $(0,\infty)$ and $\varphi(0) := \lim_{t\to 0^+} \varphi(t) = 0$, we have

$$\forall k \in \{1,\dots,n\}, \quad \sum_{i=1}^{k} \varphi(|\lambda_i(A)|^2) \leqslant \sum_{i=1}^{k} \varphi(s_i(A)^2). \tag{5}$$

Observe that from (5) with $\varphi(t) = t$ for every $t > 0$ and $k = n$, we obtain

$$\sum_{k=1}^{n} |\lambda_k(A)|^2 \leqslant \sum_{k=1}^{n} s_k(A)^2 = \mathrm{Tr}(AA^*) = \sum_{i,j=1}^{n} |A_{i,j}|^2 = \mathrm{Tr}(AA^*) = \|A\|_{\mathrm{HS}}{}^2. \tag{6}$$

The following result, due to Horn, constitutes a converse to Weyl inequalities 3. It explains why so many generic relationships between eigenvalues and singular values are consequences of (3), for instance via majorization inequalities and techniques.

Theorem 1.9 (Sherman inverse problem [Hor54])**.** *Let $(\lambda, s) \in \mathbb{C}^n \times \mathbb{R}^n$ be such that $|\lambda_1| \geqslant \cdots \geqslant |\lambda_n|$ and $s_1 \geqslant \cdots \geqslant s_n \geqslant 0$. If these numbers satisfy additionally to all the Weyl relationships (3) then there exists $A \in \mathcal{M}_{n,n}(\mathbb{C})$ such that $\lambda_i(A) = \lambda_i$ and $s_i(A) = s_i$ for every $i \in \{1,\dots,n\}$.*

From (3) we get $s_n(A) \leqslant |\lambda_n(A)| \leqslant |\lambda_1(A)| \leqslant s_1(A)$ for any $A \in \mathcal{M}_{n,n}(K)$. In particular, we have the following comparison between the spectral radius and the operator norm:

$$\rho(A) = |\lambda_1(A)| \leqslant s_1(A) = |A|_{2\to 2}.$$

In this spirit, the following result, due to Gelfand, allows to estimate the spectral radius $\rho(A)$ with the singular values of the powers of A.

Theorem 1.10 (Gelfand spectral radius formula [Gel41])**.** *Let $|\cdot|$ be a submultiplicative matrix norm on $\mathcal{M}_{n,n}(K)$ such as the operator norm $|\cdot|_{2\to 2}$ or the Frobenius norm $\|\cdot\|_{\mathrm{HS}}$. Then for every matrix $A \in \mathcal{M}_{n,n}(K)$ we have*

$$\rho(A) := |\lambda_1(A)| = \lim_{k\to\infty} \sqrt[k]{|A^k|}.$$

The eigenvalues of non normal matrices are far more sensitive to perturbations than the singular values, and this is captured by the notion of pseudo spectrum, which bridges eigenvalues and singular values, see for instance the book [TE05].

1.3. **Relation with rows distances.** The following couple of lemmas relate the singular values of matrices to distances between rows (or columns). For square random matrices, they provide a convenient control on the operator norm and Frobenius norm of the inverse respectively. The first lemma can be found in the work of Rudelson and Vershynin while the second appears in the work of Tao and Vu.

Lemma 1.11 (Rudelson-Vershynin [RV09])**.** *If $A \in \mathcal{M}_{m,n}(K)$ has rows $R_1,\dots,R_m$, then, denoting $R_{-i} = \mathrm{span}\{R_j : j \neq i\}$, we have*

$$m^{-1/2} \min_{1\leqslant i\leqslant m} \mathrm{dist}_2(R_i, R_{-i}) \leqslant s_{m\wedge n}(A) \leqslant \min_{1\leqslant i\leqslant m} \mathrm{dist}_2(R_i, R_{-i}).$$

Proof. Since A and $A^\top$ have same singular values, we can prove the statement for the columns of A. For every vector $x \in K^n$ and every $i \in \{1,\dots,n\}$, the triangle inequality and the identity $Ax = x_1C_1 + \cdots + x_nC_n$ give

$$|Ax|_2 \geqslant \mathrm{dist}_2(Ax, C_{-i}) = \min_{y\in C_{-i}} |Ax - y|_2 = \min_{y\in C_{-i}} |x_iC_i - y|_2 = |x_i|\mathrm{dist}_2(C_i, C_{-i}).$$

If $|x|_2 = 1$ then necessarily $|x_i| \geqslant n^{-1/2}$ for some $i \in \{1,\dots,n\}$, and therefore

$$s_{m\wedge n}(A) = \min_{|x|_2=1} |Ax|_2 \geqslant n^{-1/2} \min_{1\leqslant i\leqslant n} \mathrm{dist}_2(C_i, C_{-i}).$$

Conversely, for any $i \in \{1,\dots,n\}$, there exists a vector $y \in K^n$ with $y_i = 1$ such that

$$\mathrm{dist}_2(C_i, C_{-i}) = |y_1 C_1 + \cdots + y_n C_n|_2 = |Ay|_2 \geqslant |y|_2 \min_{|x|_2=1} |Ax|_2 \geqslant s_{m\wedge n}(A)$$

where we used the fact that $|y|_2^2 = |y_1|^2 + \cdots + |y_n|^2 \geqslant |y_i|^2 = 1$. □

Lemma 1.12 (Tao-Vu [TV10]). *Let $1 \leqslant m \leqslant n$ and $A \in \mathcal{M}_{m,n}(K)$ with rows $R_1,\dots,R_m$. If $\mathrm{rank}(A) = m$ then, denoting $R_{-i} = \mathrm{span}\{R_j : j \neq i\}$, we have*

$$\sum_{i=1}^{m} s_i^{-2}(A) = \sum_{i=1}^{m} \mathrm{dist}_2(R_i, R_{-i})^{-2}.$$

Proof. The orthogonal projection of R_i on R_{-i} is $B^*(BB^*)^{-1}BR_i^*$ where B is the $(m-1)\times n$ matrix obtained from A by removing the row R_i. In particular, we have

$$|R_i|_2^2 - \mathrm{dist}_2(R_i, R_{-i})^2 = \left|B^*(BB^*)^{-1}BR_i^*\right|_2^2 = (BR_i^*)^*(BB^*)^{-1}BR_i^*$$

by the Pythagoras theorem. On the other hand, the Schur bloc inversion formula states that if M is an $m\times m$ matrix then for every partition $\{1,\dots,m\} = I \cup I^c$,

$$(M^{-1})_{I,I} = (M_{I,I} - M_{I,I^c}(M_{I^c,I^c})^{-1}M_{I^c,I})^{-1}.$$

Now we take $M = AA^*$ and $I = \{i\}$, and we note that $(AA^*)_{i,j} = R_i \cdot R_j$, which gives

$$((AA^*)^{-1})_{i,i} = (R_i \cdot R_i - (BR_i^*)^*(BB^*)^{-1}BR_i^*)^{-1} = \mathrm{dist}_2(R_i, R_{-i})^{-2}.$$

The desired formula follows by taking the sum over $i \in \{1,\dots,m\}$. □

1.4. **Algorithm for the computation of the SVD.** To compute the SVD of $A \in \mathcal{M}_{m,n}(K)$ one can diagonalize both AA^* and A^*A or diagonalize the matrix H defined in (1). Unfortunately, this approach can lead to a loss of information numerically. In practice, and up to machine precision, the SVD is better computed with a two step algorithm such as (the real world algorithm is a bit more involved):

(1) Unitary bidiagonalization. Compute a couple of K–unitary matrices W, W' such that $B = WAW'$ is bidiagonal. Both W, W' are product of Householder reflections, see [GVL96]. One can also use Gram–Schmidt orthonormalization of the rows. It is worthwhile to mention that a very similar method allows also the tridiagonalization of Hermitian matrices (in this case we have $W = W'$).
(2) Iterative algorithm for bidiagonal matrices. Compute the SVD of B up to machine precision with a variant of the QR algorithm due to Golub and Kahan. Note that the standard QR iterative algorithm allows the iterative numerical computation of the eigenvalues of arbitrary square matrices.

The `svd` command of Matlab, GNU Octave, GNU R, and Scilab allows the numerical computation of the SVD. At the time of writing, the GNU Octave and GNU R version is based on LAPACK. The GNU Scientific Library (GSL) offers an algorithm based on Jacobi orthogonalization. There exists many other algorithms/variants for the numerical computation of the SVD, see [GVL96, sections 5.4.5 and 8.6].

```
octave> A = rand(5,3) % Generate a random 5x3 matrix
A =  0.3479368   0.7948432   0.0011214
     0.4912752   0.6836159   0.8509682
     0.0315889   0.9831456   0.3328946
     0.3665785   0.9985220   0.6228932
     0.2481886   0.5890069   0.2542045
```

```
octave> [U,D,V] = svd(A) % Compute SVD up to machine prec.
U =  -0.351343   -0.557528    0.667944   -0.165446   -0.303643
     -0.509631    0.708933    0.144001   -0.438839   -0.156123
     -0.448938   -0.414874   -0.704250   -0.353062   -0.075585
     -0.563423    0.084485   -0.082364    0.808135   -0.124705
     -0.312799   -0.085546    0.174251   -0.048095    0.928527
D = Diagonal Matrix
      2.18534         0         0
            0   0.61541         0
            0         0   0.31967
            0         0         0
            0         0         0
V =  -0.30703    0.24525    0.91956
     -0.83093   -0.54016   -0.13337
     -0.46400    0.80503   -0.36963
octave> norm(U*D*V'-A,"fro") % Quality check (Frobenius)
ans =  6.0189e-16
octave> norm(U*U'-eye(5,5),"fro") % Quality check (Frobenius)
ans =  8.2460e-16
octave> norm(V*V'-eye(3,3),"fro") % Quality check (Frobenius)
ans =  3.3309e-16
```

1.5. **Some concrete applications of the SVD.** The SVD is typically used for dimension reduction and for regularization. For instance, the SVD allows to construct the so called Moore–Penrose pseudoinverse [Moo20, Pen56] of a matrix by replacing the non null singular values by their inverse while leaving in place the null singular values. Generalized inverses of integral operators were introduced earlier by Fredholm in [Fre03]. Such generalized inverse of matrices provide for instance least squares solutions to degenerate systems of linear equations. A diagonal shift in the SVD is used in the so called Tikhonov regularization [Tik43, Tar05] or ridge regression for solving over determined systems of linear equations. The SVD is at the heart of the so called principal component analysis (PCA) technique in applied statistics for multivariate data analysis, see for instance the book [Jol02]. The partial least squares (PLS) regression technique is also connected to PCA/SVD. In the last decade, the PCA was used together with the so called kernel methods in learning theory. Certain generalizations of the SVD are used for the regularization of ill posed inverse problems such as X ray tomography, emission tomography, inverse diffraction and inverse source problems, and the linearized inverse scattering problem, see for instance the book [BB98]. The application of the SVD to compressed sensing is under development and some few devoted books will appear in the near future.

2. Singular values of Gaussian random matrices

In the sequel, the standard Gaussian on K is $\mathcal{N}(0,1)$ if $K=\mathbb{R}$ and $\mathcal{N}(0,\frac{1}{2}I_2)$ if $K=\mathbb{C}\equiv\mathbb{R}^2$. If Z is a standard Gaussian random variable on K then

$$\mathrm{Var}(Z):=\mathbb{E}(|Z-\mathbb{E}Z|^2)=\mathbb{E}(|Z|^2)=1.$$

2.1. **Matrix model.** Let $(G_{i,j})_{i,j\geqslant 1}$ be an infinite matrix of i.i.d. standard Gaussian random variables on K. For every $m,n\in\{1,2,\ldots\}$, the random $m\times n$ matrix

$$G:=(G_{i,j})_{1\leqslant i\leqslant m,\,1\leqslant j\leqslant n}$$

has Lebesgue density, in $\mathcal{M}_{m,n}(K) \equiv K^{nm}$, proportional to

$$G \mapsto \exp\left(-\frac{\beta}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}|G_{i,j}|^2\right) = \exp\left(-\frac{\beta}{2}\mathrm{Tr}(GG^*)\right) = \exp\left(-\frac{\beta}{2}\|G\|_{\mathrm{HS}}{}^2\right)$$

where

$$\beta := \begin{cases} 1 & \text{if } K = \mathbb{R}, \\ 2 & \text{if } K = \mathbb{C}. \end{cases}$$

The law of G is K–unitary invariant since $UGV \overset{d}{=} G$ for every deterministic K–unitary matrices U $(m \times m)$ and V $(n \times n)$. For $K = \mathbb{C}$ we have

$$G = \frac{1}{\sqrt{2}}\,(G_1 + \sqrt{-1}\,G_2)$$

where G_1 and G_2 are i.i.d. copies of the case $K = \mathbb{R}$. The law of G is also known as the K Ginibre ensemble, see [Gin65, Meh04]. The symplectic case where K is the quaternions ($\beta = 4$) is not considered in these notes. The columns $C_1, \ldots, C_n$ of the random matrix G are i.i.d. standard Gaussian random column vectors of K^m with i.i.d. standard Gaussian coordinates. Their *empirical covariance matrix* is

$$\frac{1}{n}\sum_{k=1}^{n} C_k C_k^* = \frac{1}{n}\,GG^*.$$

The strong law of large numbers gives $\lim_{n\to\infty} n^{-1}GG^* = I_m$ a.s. We are interested in the sequel in asymptotics when both n and m tend to infinity. The random matrix GG^* is $m \times m$ Hermitian positive semidefinite. If $m > n$ then the random matrix GG^* is singular with probability one, as a linear combination of $n < m$ rank one $m \times m$ matrices $C_1C_1^*, \ldots, C_nC_n^*$. If $m \leqslant n$ then the random matrix GG^* is invertible with probability one (comes from the diffuse nature of Gaussian measures), and

$$\forall k \in \{1, \ldots, m\}, \quad s_k(G)^2 = \lambda_k(GG^*) = n\lambda_k\left(\frac{1}{n}\,GG^*\right).$$

2.2. **Unitary bidiagonalization.** Let us consider the K–unitary bidiagonalization of section 1.4, for the Gaussian matrix G. Assume for convenience that $m \leqslant n$. One can find random K–unitary matrices W $(m \times m)$ and W' $(n \times n)$ such that $B := WGW'$ is bidiagonal with

$$B = \frac{1}{\sqrt{\beta}}\begin{pmatrix} S_n & 0 & 0 & 0 & & & \cdots & & & 0 \\ T_{m-1} & S_{n-1} & 0 & 0 & & & \cdots & & & 0 \\ 0 & T_{m-2} & S_{n-2} & 0 & & & \cdots & & & 0 \\ 0 & 0 & \vdots & \vdots & & & \cdots & & & 0 \\ \vdots & \vdots & & & & & \cdots & & & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & T_1 & S_{n-(m-1)} & 0 \cdots & 0 \end{pmatrix}. \tag{7}$$

Following Silverstein [Sil85] the random variables $S_n, \ldots, S_{n-(m-1)}, T_{m-1}, \ldots, T_1$ are independent with laws given by $S_k^2 \sim \chi^2(\beta k)$ for every $k \in \{n-(m-1), \ldots, n\}$ and $T_k^2 \sim \chi^2(\beta k)$ for every $k \in \{1, \ldots, m-1\}$. The random matrices B and G share the same sequence of singular values. Such an explicit bidiagonalization has an amazing consequence for the simulation of the singular values of G. It allows to reduce the dimension from nm to $2(n \wedge m) - 1$.

2.3. **Densities.** The random Hermitian positive semidefinite $m \times m$ matrix GG^* can be seen as a random vector of $\mathbb{R}^m \times K^{(m^2-m)/2}$. If $m \leqslant n$, the law of GG^* is a Wishart distribution with Lebesgue density in $\mathbb{R}^m \times K^{(m^2-m)/2}$ proportional to

$$W \mapsto \det(W^{\beta(n-m+1)/2-1}) \exp\left(-\frac{\beta}{2}\mathrm{Tr}(W)\right) \tag{8}$$

on the cone of Hermitian positive semidefinite matrices. This Wishart law is also known as the β–Laguerre ensemble or Laguerre Orthogonal Ensemble (LOE) for $\beta = 1$ and Laguerre Unitary Ensemble (LUE) for $\beta = 2$. The correlation between the entries is captured by the determinental term, which surprisingly vanishes when $n = m + 2\beta^{-1} - 1$. In the SVD of G, one can take U, V distributed according to the normalized Haar measure on the K–unitary group, and independent of the singular values. As a consequence, the same holds true for the K–unitary diagonalization of the positive semidefinite Hermitian matrix GG^*. When $m \leqslant n$, this diagonalization, seen as a change of variable, followed by the partial integration over the K–unitary group of (8) with respect to the eigenvectors, gives the expression of the density of $\lambda_1(GG^*), \ldots, \lambda_m(GG^*)$, which turns out to be proportional to

$$\lambda \mapsto \exp\left(-\frac{\beta}{2}\sum_{i=1}^m \lambda_i\right) \prod_{i=1}^m \lambda_i^{\beta(n-m+1)/2-1} \prod_{1 \leqslant i < j \leqslant m} |\lambda_i - \lambda_j|^\beta \tag{9}$$

on $\{\lambda \in [0,\infty)^m : \lambda_1 \geqslant \cdots \geqslant \lambda_n\}$. The normalizing constant is a Selberg integral, and can be explicitly computed [Meh04]. The (repulsive) correlation is captured by the Vandermonde determinant, which comes from the Jacobian of the change of variable (unitary diagonalization). If $m = n = 1$ then (8,9) are identical (χ^2 law). The formulas (8,9) were considered by e.g. Wishart [Wis28] and James [Jam60]. For a modern presentation, see e.g. Edelman and Rao [ER05] or Haagerup and Thorbjørnsen [HT03].

2.4. **Orthogonal polynomials.** Set $K = \mathbb{C}$. If $m \leqslant n$ then the density (9) of the eigenvalues of the $m \times m$ random matrix GG^* turns out to be proportional to

$$\lambda \mapsto \det\left[(S(\lambda_i, \lambda_j))_{1 \leqslant i,j \leqslant m}\right] \quad \text{with} \quad S(x,y) := \sqrt{g(x)g(y)} \sum_{k=0}^{m-1} P_k(x)P_k(y) \tag{10}$$

where $(P_k)_{k \geqslant 0}$ are the Laguerre orthonormal polynomials [Sze75] relative to the Gamma law on $[0,\infty)$ with density g proportional to $x \mapsto x^{n-m}\exp(-x)$. Both g and $(P_k)_{k \geqslant 0}$ depend on m, n. These determinental/polynomial formulas appear in various works, see e.g. Deift [Dei99], Forrester [For10], and Mehta [Meh04], Haagerup and Thorbjørnsen [HT03] and Ledoux [Led04]. When $m \leqslant n$, and at the formal level, it follows from this determinental/polynomial expression of the density that for any Borel symmetric function $F : [0,\infty)^m \to \mathbb{R}$, the expectation

$$\mathbb{E}[F(\lambda_1(GG^*), \ldots, \lambda_m(GG^*))]$$

can be expressed in terms of the determinant (10). The behavior of such averaged symmetric functions when n and m tend to infinity is related to the asymptotics of the Laguerre polynomials $(P_k)_{k \geqslant 0}$. Useful symmetric functions of the eigenvalues include

(1) $F(\lambda_1, \ldots, \lambda_m) = f(\lambda_1) + \cdots + f(\lambda_m)$ for some fixed $f : [0,\infty) \to \mathbb{R}$;
(2) $F(\lambda_1, \ldots, \lambda_m) = \min(\lambda_1, \ldots, \lambda_m)$;
(3) $F(\lambda_1, \ldots, \lambda_m) = \max(\lambda_1, \ldots, \lambda_m)$.

The real case $K = \mathbb{R}$ is similar but trickier due to $\beta = 1$ in (9), see [For10].

2.5. **Behavior of the singular values.** We begin our tour of horizon with the behavior of the counting probability measure of the eigenvalues of $n^{-1}GG^*$. It is customary in random matrix theory to speak about the "bulk behavior", in contrast with the "edge behavior" which concerns the extremal eigenvalues. When $m \leqslant n$, this corresponds to the counting probability measure of the squared singular values of $n^{-1/2}G$. The first version of the following theorem is due to Marchenko and Pastur [MP67].

Theorem 2.1 (Bulk behavior)**.** *If $m = m_n \to \infty$ with $\lim_{n\to\infty} m_n/n = y \in (0, \infty)$ then a.s. the spectral counting probability measure*

$$\mu_{n^{-1}GG^*} := \frac{1}{m} \sum_{k=1}^{m} \delta_{\lambda_k(n^{-1}GG^*)}$$

converges narrowly to the Marchenko–Pastur law

$$\mathcal{L}_{\mathrm{MP}} = \left(1 - \frac{1}{y}\right)^+ \delta_0 + \frac{1}{2\pi y} \frac{\sqrt{(b-x)(x-a)}}{x} \mathbb{1}_{[a,b]}(x)\, dx$$

where

$$a = (1 - \sqrt{y})^2 \quad \textit{and} \quad b = (1 + \sqrt{y})^2.$$

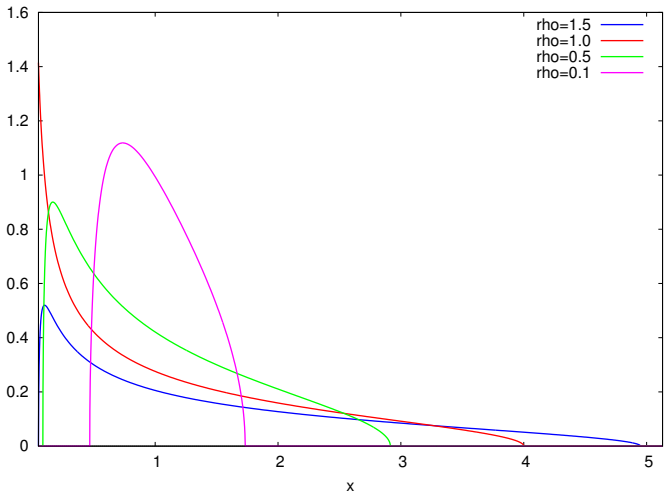

FIGURE 2. Density of the limiting law of the empirical singular values distribution $\frac{1}{m}\sum_{k=1}^{m} \delta_{s_k(n^{-1/2}G)}$ when $m = m_n$ with $\lim_{n\to\infty} m_n/n = y$, for different values of y (theorem 2.1). This is nothing else but the density of the absolutely continuous part of the image law of $\mathcal{L}_{\mathrm{MP}}$ by the map $x \mapsto \sqrt{x}$. The case $y = 1$ corresponds to the so called quartercircular law. This graphics was obtained by using the wxMaxima software.

Idea of the proof. When $K = \mathbb{C}$, the result can be obtained by using the determinental/polynomial approach. Namely, following Haagerup and Thorbjørnsen [HT03] or Ledoux [Led04], for every Borel function $f : [0, \infty) \to \mathbb{R}$, we have,

$$\int f(x)\, d\mu_{n^{-1}GG^*}(x) = \left(1 - \frac{n}{m}\right)^+ f(0) + \int_0^\infty f(x) \frac{1}{m} S(x, x)\, dx$$

where S is as in (10). Note that the left hand side is a symmetric function of the eigenvalues. The Dirac mass at point 0 in the first term of the right hand

side above comes from the fact that if $m > n$ then $m - n$ eigenvalues of GG^* are necessarily zero (the remaining eigenvalues are the square of the singular values of G). The convergence to $\mathcal{L}_{\mathrm{MP}}$ is a consequence of the behavior of $m^{-1}S(x,x)$ related to classical equilibrium measures of orthogonal polynomials, see [Led04, pages 191–192]. In this approach, $\mathcal{L}_{\mathrm{MP}}$ is recovered as a mixture of uniform and arcsine laws.

Another approach is the so called trace/moments method, based on the identity

$$\int x^r \, d\mu_{n^{-1}GG^*}(x) = \frac{1}{nm^r}\mathrm{Tr}((GG^*)^r)$$

valid for every $r \in \{0,1,2,\ldots\}$. The expansion of the right hand side in terms of the entries of G allows to show that the moments of $\mu_{n^{-1}GG^*}$ converge to the moments of $\mathcal{L}_{\mathrm{MP}}$. The Gaussian nature of the entries allows to use the Wick formula in order to simplify the computations. There is also an approach based on the Cauchy–Stieltjes transform, or equivalently the trace–resolvent, see [HP00] or [Bai99]. This gives a recursive equation obtained by bloc matrix inversion, which leads to a fixed point problem. Here again, the Gaussian integration by parts may help. The trace/moment and the Cauchy–Stieltjes trace/resolvent methods are "universal" in the sense that they are still available when G is replaced by a random matrix with non Gaussian i.i.d. entries. The determinental/polynomial approach is rigid, and relies on the determinental nature of the law of G, which comes from the unitary invariance of G. It remains available beyond the Gaussian case, provided that G has a unitary invariant density proportional to $G \mapsto \exp(-\mathrm{Tr}(V(GG^*)))$ for a potential $V : \mathbb{R} \to \mathbb{R}$.

There is finally a more original approach based on large deviations via a Varadhan like lemma, which exploits the explicit expression of the law of the eigenvalues. We recover $\mathcal{L}_{\mathrm{MP}}$ as a minimum of the logarithmic energy with Laguerre external field. □

The limiting distribution is a mixture of a Dirac mass at zero (when $y > 1$) with an absolutely continuous compactly supported distribution known as the Marchenko–Pastur law. The presence of this Dirac mass is due to the fact that if $y > 1$ then a.s. the random matrix $n^{-1}GG^*$ is not of full rank for large enough n. The a.s. weak convergence in theorem 2.1 says that for any $x \in \mathbb{R}$, $x \neq 0$ if $y > 1$, denoting $I = (-\infty, x]$,

$$\frac{\big|\big\{k \in \{1,\ldots,m\} \text{ such that } \lambda_k(n^{-1}GG^*) \in I\big\}\big|}{m} \xrightarrow[n\to\infty]{\text{a.s.}} \mathcal{L}_{\mathrm{MP}}(I).$$

This convergence implies immediately the following corollary.

Corollary 2.2 (Edge behavior implied by bulk behavior)**.** *If $m = m_n \to \infty$ with $\lim_{n\to\infty} m_n/n = y \in (0,\infty)$ then a.s.*

$$\liminf_{n\to\infty} \lambda_1(n^{-1}GG^*) \geqslant (1+\sqrt{y})^2.$$

Moreover, if $y \leqslant 1$ then a.s.

$$\limsup_{n\to\infty} \lambda_{m_n}(n^{-1}GG^*) \leqslant (1-\sqrt{y})^2.$$

In particular, if $m_n = n$ then $y = 1$ and a.s.

$$\frac{1}{\sqrt{n}}\, s_n(G) = \sqrt{\lambda_n(n^{-1}GG^*)} \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

It is then natural to ask about the convergence of the extremal eigenvalues of $n^{-1}GG^*$ to the edge of the limiting support. In a sense, the left edge a is "soft" if $y < 1$ and "hard" if $y = 1$. The term "soft" means that the fluctuation may hold in both sides while "hard" means that the fluctuation is confined in a single side. The

right edge b is "soft" regardless of y. We will see that the nature of the fluctuations of the extremal singular values depends on the hard/soft nature of the edge.

Theorem 2.3 (Convergence of smallest singular value)**.** *If $m = m_n \to \infty$ with $m_n \leqslant n$ and $\lim_{n\to\infty} m_n/n = y \in (0, 1]$ then*

$$\left(\frac{1}{\sqrt{n}}\, s_{m_n}(G)\right)^2 = \lambda_{m_n}(n^{-1}GG^*) \xrightarrow[n\to\infty]{a.s.} (1-\sqrt{y})^2.$$

Idea of the proof. Corollary 2.2 reduces immediately the problem to show that a.s.

$$\liminf_{n\to\infty} \lambda_m(n^{-1}GG^*) \geqslant (1-\sqrt{y})^2.$$

Following Silverstein [Sil85], we have $\lambda_m(GG^*) = \lambda_m(BB^*)$ where B is as in (7). Observe that BB^* is tridiagonal. One can then control $\lambda_m(BB^*)$ by using the law of B and the Geršgorin disks theorem which states that if $A \in \mathcal{M}_{n,n}(K)$ then

$$\{\lambda_1(A), \ldots, \lambda_n(A)\} \subset \bigcup_{i=1}^{n} \{z \in \mathbb{C}; |z - A_{i,i}| \leqslant r_i\} \quad \text{where} \quad r_i := \sum_{j\neq i} |A_{i,j}|.$$

When $K = \mathbb{C}$, an alternative approach is based on the determinental/polynomial formula for the density of the eigenvalues of GG^*, and can be found in [Led04]. □

Theorem 2.4 (Fluctuation of smallest singular value for hard edge)**.** *Assume that $m = n$.*

- *If $K = \mathbb{C}$ then for every $n \in \{1, 2, \ldots\}$, the random variable $(\sqrt{n}\, s_n(G))^2$ follows an exponential law of unit mean with Lebesgue density $x \mapsto \exp(-x)$. In other words, for every $n \in \{1, 2, \ldots\}$ and any real number $t \geqslant 0$,*

$$\mathbb{P}\big(\sqrt{n}\, s_n(G) \geqslant t\big) = \exp\big(-t^2\big).$$

- *If $K = \mathbb{R}$ then the random variable $(\sqrt{n}\, s_n(G))^2$ converges in distribution as $n \to \infty$ to the law with Lebesgue density density*

$$x \mapsto \frac{1+\sqrt{x}}{2\sqrt{x}} \exp\left(-\frac{1}{2}x - \sqrt{x}\right).$$

  In other words, for every real number $t \geqslant 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(\sqrt{n}\, s_n(G) \geqslant t\big) = \exp\left(-\frac{1}{2}t^2 - t\right).$$

Idea of the proof. When $K = \mathbb{C}$, it suffices to integrate (9) over all but the smallest eigenvalue. This gives that the random variable $n\lambda_n(GG^*) = ns_n(G)^2$ follows an exponential law with unit mean. This is immediate when $n = 1$ from (9). When $K = \mathbb{R}$, one can proceed as for the complex case, but with this time $\beta = 1$. This makes the computations non explicit for a fixed n due to the factors $\lambda_i^{-1/2}$ which were not present for $K = \mathbb{C}$. However, following Edelman [Ede88], for every n,

$$\lambda_n(GG^*) = s_n(G)^2 \quad \text{has density proportional to} \quad x \mapsto \frac{1}{\sqrt{x}} U_n\Big(\frac{x}{2}\Big) \exp\left(-\frac{1}{2}nx\right)$$

where U_n is the Tricomi function, unique solution of the Kummer differential equation

$$2xU_n''(x) - (1+2x)U_n'(x) - (n-1)U_n(x) = 0$$

with boundary conditions $2U_n(0)\Gamma(1+n/2) = \sqrt{\pi}$ and $U_n(\infty) = 0$. The Tricomi function admits an integral representation, and is also known as the Gordon function or the confluent hypergeometric function of the second kind, see [AS64, Chapter 13.6]. The behavior of the Tricomi function gives the limiting law of $\sqrt{n}\, s_n(G)$. □

Theorem 2.5 (Convergence of largest singular value)**.** *If $m = m_n \to \infty$ with $\lim_{n\to\infty} m_n/n = y \in (0,\infty)$ then*

$$\left(\frac{1}{\sqrt{n}}\, s_1(G)\right)^2 = \frac{1}{n}\lambda_1(GG^*) \xrightarrow[n\to\infty]{a.s.} (1+\sqrt{y})^2.$$

Idea of the proof. Corollary 2.2 reduces the problem to show that

$$\limsup_{n\to\infty} \lambda_1(n^{-1}GG^*) \leqslant (1+\sqrt{y})^2.$$

This was proved in turn by Geman [Gem80], following an idea of Grenander. The method, which can be seen as an instance of the so called power method, consists in the control of the expected operator norm of a power of $n^{-1}GG^*$ with the expected Frobenius norm, and then in the usage of expansions in terms of the matrix entries via the trace formula for the Frobenius norm. This method does not rely on explicit Gaussian computations. When $K = \mathbb{C}$, one can use the determinental/polynomial formula for the density of the eigenvalues of GG^* as in the work of Ledoux [Led04]. □

Gaussian exponential bounds for the tail of the singular values of G are also available, and can be found for instance in the work of Szarek [Sza91], Davidson and Szarek [DS01], Haagerup and Thorbjørnsen [HT03], and Ledoux [Led04].

The fluctuation of the smallest singular value in the hard edge case given by theorem 2.4 can be also expressed in terms of a Bessel kernel, see for instance the work of Forrester [For10]. Let us consider now the fluctuation of the largest singular value around its limit. The famous Tracy–Widom laws TW_1 and TW_2 are known to describe the fluctuation of the largest eigenvalue in the ensembles of square Hermitian Gaussian random matrices (GOE for $K = \mathbb{R}$ and GUE for $K = \mathbb{C}$), see [TW02]. One can ask if these laws still describe the fluctuations of the largest singular values of the Gaussian matrix G. By definition, TW_1 and TW_2 are the probability distributions on $\mathbb{R}$ with cumulative distribution functions F_1 and F_2 given for every $s \in \mathbb{R}$ by

$$F_2(s) = \exp\left(-\int_s^\infty (x-s)q(x)^2\,dx\right) \quad \text{and} \quad F_1(s)^2 = F_2(s)\exp\left(-\int_s^\infty q(x)\,dx\right)$$

where q is the solution of the Painlevé II differential equation

$$q''(x) = xq(x) + 2q^3(x)$$

with boundary condition $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$, where Ai is the Airy function

$$\mathrm{Ai}(x) := \frac{1}{\pi}\int_0^\infty \cos\left(\frac{1}{3}t^3 + xt\right) dt.$$

The Airy function Ai is also uniquely defined by the properties

$$\mathrm{Ai}''(x) = x\mathrm{Ai}(x) \quad \text{and} \quad \mathrm{Ai}(x) \sim_{x\to\infty} \frac{1}{2\sqrt{\pi}x^{1/4}}\,\exp\left(-\frac{2}{3}\,x^{3/2}\right).$$

The function F_2 can be expressed as a Fredholm determinant: $F_2(s) = \det(I - A_s)$ where A_s is the Airy operator on square integrable functions on (s,∞), with kernel

$$A_s(x,y) := \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}.$$

See for instance [Dei99, Dei07] and [For10] for more information, and [ER05] and [Joh01] for the numerical evaluation of F_1 and F_2.

Theorem 2.6 (Fluctuation of largest singular value)**.** *If $m = m_n \to \infty$ with $m_n \leqslant n$ and $\lim_{n\to\infty} m_n/n = y \in (0,1]$ then, by denoting*

$$\mu_{\beta,n} := \Big(\sqrt{n+\beta-2} + \sqrt{m}\Big)^2 \quad \text{and} \quad \sigma_{\beta,n} := \sqrt{\mu_{\beta,n}}\Big(\frac{1}{\sqrt{n+\beta-2}} + \frac{1}{\sqrt{m}}\Big)^{1/3},$$

the random variable

$$\frac{s_1(G)^2 - \mu_{\beta,n}}{\sigma_{\beta,n}}$$

converges narrowly as $n \to \infty$ to the Tracy–Widom law TW_β. Moreover, if $y > 1$ then the result remains true up to the swap of the roles of m and n in the formulas (recall that G and G^* have same singular values).

For $K = \mathbb{C}$ and $m = n$, we have $\beta = 2$ and $\mu_{2,n} = 4n$ while $\sigma_{2,n} = (16n)^{1/3}$. The Tracy–Widom fluctuation based on the Airy kernel describes also the fluctuation of the smallest singular value in the soft edge regime ($y < 1$), see for instance the book by Forrester [For10] and the approach of Ledoux [Led04].

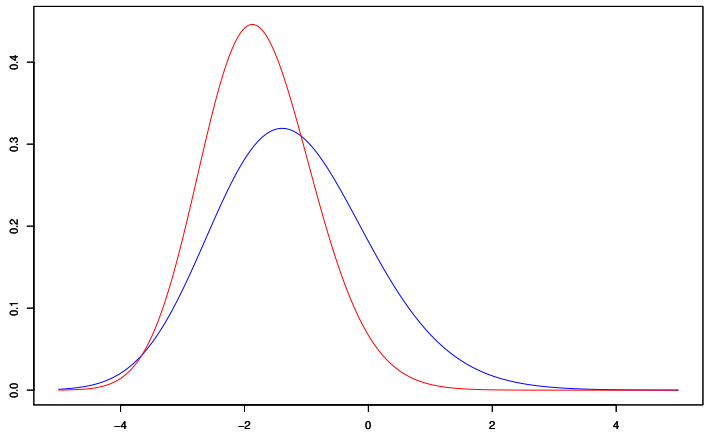

FIGURE 3. Density of TW_β for $\beta = 1$ (blue) and $\beta = 2$ (red), obtained by using the GNU-R package RMTstat.

Idea of the proof. The proofs of Johnstone [Joh01] and Johansson [Joh00] are based on the determinental/polynomial approach. Let us give the first steps when $K = \mathbb{C}$. If S is as in (10), then for every Borel function $f : [0,\infty) \to \mathbb{R}$,

$$\mathbb{E}\left[\prod_{k=1}^{m} (1 + f(\lambda_k(GG^*)))\right] = c_{n,m}\, \det(I + Sf)$$

where $c_{n,m}$ is a normalizing constant. Here one must see S as an integral operator. For the particular choice $f = -\mathbb{1}_{[t,\infty)}$ for some fixed $t \geqslant 0$, this gives

$$\mathbb{P}\Big(\max_{1\leqslant k\leqslant m} \lambda_k(GG^*) \geqslant t\Big) = c_{n,m}\, \det(I - S\mathbb{1}_{[t,\infty)}).$$

Now the Tracy and Widom [TW02] heuristics says that the determinant in the right hand side satisfies to a differential equation, which is Painlevé II as $n \to \infty$. See also the work of Borodin and Forrester [BF03, For10], and the work of Ledoux [Led04] inspired from the work of Haagerup and Thorbjørnsen [HT03].

An alternative approach, based on the bidiagonalization trick (7), was provided by Ramírez, Rider, and Virág [RRV09]. This can be viewed as the β–Laguerre (LUE and LOE) analogue of the work of Edelman and Sutton [ES05] for β–Hermite ensembles (GUE and GOE). In particular, it provides the convergence of the rescaled extremal singular values to a Schrödinger operator. □

The largest eigenvalue of such matrices can be seen as the maximum of a random vector with correlated coordinates (Vandermonde repulsion). Here the asymptotic fluctuation is not captured by classical extreme values theory for i.i.d. samples (Gnedenko–Fréchet–Fisher–Tippett–Gumbel theorem, see [Res08]). The laws TW_1 and TW_2 are unimodal, asymmetric, with exponentially light tails. For instance, TW_1 has a left tail $\approx \exp\left(-\frac{1}{24}|x|^3\right)$ and a right tail $\approx \exp\left(-\frac{2}{3}|x|^{3/2}\right)$, see [Joh01].

The study of the extremal singular values $s_1(G), s_n(G)$ and the condition number $\kappa(G) = s_1(G)/s_n(G)$ of the random Gaussian matrix G was motivated at the origin by the behavior of numerical algorithms with random inputs. This goes back at least to von Neumann and his collaborators [vN63, vNG47], Smale [Sma85], Demmel [Dem88], and Kostlan [Kos88]. Note that if $m = n$ then

$$\kappa(G) = \frac{s_1(G)}{s_n(G)} = \sqrt{\frac{\lambda_1(GG^*)}{\lambda_n(GG^*)}} = \sqrt{\kappa(GG^*)} = \sqrt{n}\,\kappa\left(\frac{1}{n}\,GG^*\right).$$

An elementary result on $\kappa(G)$ is captured by the following corollary. For sharp estimates on the tails of $\kappa(G)$, see for instance the work of Edelman and Sutton [ES05], Szarek [Sza91], Azaïs and Wschebor [AW05], and also Chen and Dongarra [CD05]. These sharp bounds involve the control of the joint law of the extremal singular values. This joint law can be expressed with zonal polynomials and hypergeometric functions [Mui82, RVA05]. This expression is difficult to exploit. The approach of Azaïs and Wschebor [AW05] is based on Rice formulas for Gaussian processes extrema, see [AW09]. For the case $\beta \notin \{1, 2\}$, see for instance [DK08] and references therein.

Corollary 2.7 (Condition number, $m = n$)**.** *If $m = n$ then $n^{-1}\,\kappa(G)$ converges in distribution as $n \to \infty$ to a law with Lebesgue density*

$$x \mapsto \begin{cases} \dfrac{2(x+1)}{x^3}\,\exp\left(-\dfrac{1}{2x^2} - \dfrac{1}{x}\right) & \text{if } K = \mathbb{R}, \\ \\ \dfrac{4}{x^3}\,\exp\left(-\dfrac{1}{x^2}\right) & \text{if } K = \mathbb{C}. \end{cases}$$

Proof. Theorem 2.5 gives that a.s. $s_1(G) = (2 + o(1))\,\sqrt{n}$ as $n \to \infty$. We conclude by using the Slutsky lemma and the limiting law of $\sqrt{n}\,s_n(G)$ (theorem 2.4). □

3. Universality of the Gaussian case

Gaussian random matrices with i.i.d. entries such as G have the advantage to allow explicit computations. But one can ask if such Gaussian matrices are good enough for the modelling of random inputs of algorithms. For instance, the support of such random matrices is essentially concentrated on a centered Frobenius ball, which can be seen as a drawback. More generally, let us consider a sample of n i.i.d. random column vectors of K^m. One can ask about the behavior of the eigenvalues of their empirical covariance matrix in the situation where the common law of the vectors...

- is not centered
- is not a tensor product
- has only few finite moments (heavy tails)

- does not have a density (for instance Bernoulli or Rademacher entries).

It is rather difficult to give a comprehensive account on the available literature in few pages. Regarding independent column vectors, a whole line of research is based on tools and concepts from high dimensional geometric analysis, such as the work of Mendelson and Pajor [MP06] and the work of Adamczak, Guédon, Litvak, Pajor, and Tomczak-Jaegermann [AGL+08]. In the sequel, we restrict our attention on some few results regarding the singular values of random matrices with i.i.d. entries.

Many results for random matrices with i.i.d. Gaussian entries remain valid for non Gaussian entries when the moments match the Gaussian moments up to some order. This is referred as "universality". Let $(X_{i,j})_{i,j\geqslant 1}$ be an infinite matrix with i.i.d. entries in K. We consider in the sequel the $m \times n$ random matrix

$$X := (X_{i,j})_{1\leqslant i\leqslant m,\ 1\leqslant j\leqslant n}.$$

When $X_{1,1}$ is a standard Gaussian random variable then $X \stackrel{d}{=} G$ where G is the Gaussian random matrix of the preceding section. Note that if the law of $X_{1,1}$ has atoms, then XX^* is singular with positive probability, even if $m \leqslant n$. Moreover, if $X_{1,1}$ is not standard Gaussian, the law of X is no longer K–unitary invariant, and the law of the eigenvalues of XX^* is not explicit in general. One of the first universal version of theorem 2.1 is due to Wachter [Wac78]. See also the review article of Bai [Bai99]. For the version given below, see the book of Bai and Silverstein [BS10], and the article by Bai and Yin [BY93] on the behavior at the edge.

Theorem 3.1 (Universality for bulk and edges convergence)**.** *If $X_{1,1}$ has mean $\mathbb{E}[X_{1,1}] \in K$ and variance $\mathbb{E}[|X_{1,1} - \mathbb{E}[X_{1,1}]|^2] = 1$, and if $m = m_n \to \infty$ with $m_n/n \to y \in (0,\infty)$, then the conclusion of theorem 2.1 remain valid if we replace G by X. Moreover, if $\mathbb{E}[X_{1,1}] = 0$ and $\mathbb{E}[|X_{1,1}|^4] < \infty$ then the conclusion of theorem 2.3 (when $y \in (0,1)$) and theorem 2.5 remain valid if we replace G by X. If however $\mathbb{E}[|X_{1,1}|^4] = \infty$ or $\mathbb{E}[X_{1,1}] \neq 0$ then a.s.*

$$\limsup_{n\to\infty} \lambda_1(n^{-1}XX^*) = \infty.$$

The bulk behavior is not sensitive to the mean $\mathbb{E}[X]$, and this can be understood from the decomposition $X = X - \mathbb{E}[X] + \mathbb{E}[X]$ where $\mathbb{E}[X] = \mathbb{E}[X_{1,1}](1 \otimes 1)$ has rank at most 1, by using the Thompson theorem 1.5. Regarding empirical covariance matrices, many other situations are considered in the literature, for instance in the work of Bai and Silverstein [BS98], Dozier and Silverstein [DS07a, DS07b], Hachem, Loubaton, and Najim [HLN06], with various concrete motivations ranging from asymptotic statistics to information theory and signal processing.

The universality of the fluctuation of the smallest and largest eigenvalues of empirical covariances matrices was studied for instance by Soshnikov [Sos02], Baik, Ben Arous, and Péché [BBAP05], Ben Arous and Péché [BAP05], El Karoui [EK07], Féral and Péché [FP09], Péché [Péc09], Tao and Vu [TV09b], and by Feldheim and Sodin [FS10]. The following theorem on the soft edge case, due to Feldheim and Sodin [FS10], includes partly the results of Soshnikov [Sos02] and Péché [Péc09].

Theorem 3.2 (Universality for soft edges)**.** *If the law of $X_{1,1}$ is symmetric about 0, with sub-Gaussian tails, and first two moments identical to the ones of $G_{1,1}$, and if $m = m_n \to \infty$ with $m_n \leqslant n$ and $\lim_{n\to\infty} m_n/n = y \in (0,\infty)$, then $s_1(X)^2$ has the Tracy–Widom rates and fluctuations of $s_1(G)^2$ as in the Gaussian theorem 2.6. If moreover $y < 1$ then the same holds true for the smallest singular value $s_{m_n}(X)^2$.*

The following theorem concerns the universality of the fluctuations of the smallest singular value in the hard edge regime, recently obtained by Tao and Vu [TV09b].

Theorem 3.3 (Universality for hard edge)**.** *If $m = n$ and if $X_{1,1}$ has first two moments identical to the ones of $G_{1,1}$, and if $\mathbb{E}[|X_{1,1}|^{10000}] < \infty$, then the random variable $(\sqrt{n}\, s_n(X))^2$ converges as $n \to \infty$ to the limiting law of the Gaussian case which appears in theorem 2.4.*

Actually, a stronger version of theorems 3.2 and 3.3 is available, expressing the fact that for every fixed k, the top and bottom k singular values are identical in law asymptotically to the corresponding quantities for the Gaussian model.

Following [TV09b], theorem 3.3 implies in particular that if $X_{1,1}$ follows the symmetric Rademacher law $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ then, for $m = n$, for all $t > 0$,

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leqslant t) = \int_0^{t^2} \frac{1+\sqrt{x}}{2\sqrt{x}} e^{-\frac{1}{2}x-\sqrt{x}}\, dx + o(1) = 1 - e^{-\frac{1}{2}t^2-t} + o(1).$$

In [TV09b], the $o(1)$ error term is shown to be of the form $O(n^{-c})$ uniformly over t. This is close to the statement of a conjecture by Spielman and Teng on the invertibility of random sign matrices stating the existence of a constant $c \in (0,1)$ such that

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leqslant t) \leqslant t + c^n \tag{11}$$

for every $t \geqslant 0$. The c^n is due to the fact that X has a positive probability of being singular (e.g. equality of two rows). In 2008, Spielman and Teng were awarded the Gödel Prize for their work on smoothed analysis of algorithms [ST03, ST02]. Actually, it has been conjectured years ago that

$$\mathbb{P}(s_n(X) = 0) = \left(\frac{1}{2} + o(1)\right)^n.$$

This intuition comes from the probability of equality of two rows, which implies that $\mathbb{P}(s_n(X) = 0) \geqslant (1/2)^n$. Many authors contributed to this difficult non-linear discrete problem, such as Komlós [Kom67], Kahn, Komlós, and Szemerédi [KKS95], Rudelson [Rud08], Bruneau and Germinet [BG09], Tao and Vu [TV06, TV07, TV09a], and Bourgain, Vu, and Wood [BVW10] who proved that

$$\mathbb{P}(s_n(X) = 0) \leqslant \left(\frac{1}{\sqrt{2}} + o(1)\right)^n \quad \text{for large enough } n.$$

Back to theorem 3.1, the bulk behavior when $X_{1,1}$ has an infinite variance was recently investigated by Belinschi, Dembo, and Guionnet [BDG09], using (1). They considered heavy tailed laws similar to α–stable laws ($0 < \alpha \leqslant 2$), with polynomial tails. For simulations, if U and ε are independent random variables with U uniform on $[0,1]$ and ε Rademacher $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, then the random variable $T = \varepsilon(U^{-1/\alpha} - 1)$ has a symmetric bounded density and $\mathbb{P}(|T| > t) = (1+t)^{-\alpha}$ for any $t \geqslant 0$. In this situation, the normalization n^{-1} in $n^{-1}XX^*$ must be replaced by $n^{-2/\alpha}$. The limiting spectral distribution is no longer a Marchenko–Pastur distribution, and has heavy tails. In the case where $X_{1,1}$ is Cauchy distributed, it is known that the largest eigenvalues are distributed according to a Poisson statistics, see the work of Soshnikov and Fyodorov [SF05] and the review article by Soshnikov [Sos06]. On can ask about the invertibility of such random matrices with heavy tailed i.i.d. entries. The following lemma gives a rather crude lower bound on the smallest singular value of random matrices with i.i.d. entries with bounded density. However, it shows that the invertibility of these random matrices can be controlled without moments assumptions. Arbitrary heavy tails are therefore allowed, but Dirac masses are not allowed.

Lemma 3.4 (Polynomial lower bound on s_n for bounded densities)**.** *Assume that $X_{1,1}$ is absolutely continuous with bounded density f. If $m = n$ then there exists*

an absolute constant $c > 0$ such that for every $n \in \{1, 2, \ldots\}$ and $u \geqslant 0$,

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leqslant u) \leqslant c n^{\frac{3}{2}} |f|_\infty u^\beta.$$

From the first Borel–Cantelli lemma, it follows that there exists $b > 0$ such that a.s. for large enough n, we have $s_n(X) > n^{-b}$.

Proof. Let $R_1, \ldots, R_n$ be the rows of X. From lemma 1.11 we have

$$\min_{1 \leqslant i \leqslant n} \mathrm{dist}_2(R_i, R_{-i}) \leqslant \sqrt{n}\, s_n(X).$$

Consequently, by the union bound and the exchangeability, for any $u \geqslant 0$,

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leqslant u) \leqslant n \mathbb{P}(\mathrm{dist}_2(R_1, R_{-1}) \leqslant u).$$

Let Y be a unit normal vector to R_{-1}. Such a vector is not unique, but we just pick one which is measurable with respect to $R_2, \ldots, R_n$. This defines a random variable on the unit sphere $S_2(K^n) = \{x \in K^n : |x|_2 = 1\}$, independent of R_1. By the Cauchy–Schwarz inequality, we have $|R_1 \cdot Y| \leqslant |p(R_1)|_2 |Y|_2 = \mathrm{dist}_2(R_1, R_{-1})$ where $p(\cdot)$ is the orthogonal projection on the orthogonal space of R_{-1}. Actually, since the law of $X_{1,1}$ is diffuse, the matrix X is a.s. invertible, the subspace R_{-1} is a hyperplane, and $|R_1 \cdot Y| = \mathrm{dist}_2(R_1, R_{-1})$, but this is useless in the sequel. Let ν be the distribution of Y on $S_2(K^n)$. Since Y and R_1 are independent, for any $u \geqslant 0$,

$$\mathbb{P}(\mathrm{dist}_2(R_1, R_{-1}) \leqslant u) \leqslant \mathbb{P}(|R_1 \cdot Y| \leqslant u) = \int_{S_2(K^n)} \mathbb{P}(|R_1 \cdot y| \leqslant u)\, d\nu(y).$$

Consider some $y \in S_2(K^n)$. Since $|y|_2 = 1$, there exists some $i \in \{1, \ldots, n\}$ such that $|y_i| > 0$ and $|y_i|^{-1} \leqslant \sqrt{n}$. The random variable $X_{i,i}\, \overline{y_i}$ is absolutely continuous with density $|y_i|^{-1} f(\overline{y_i}^{-1} \cdot)$. Now, the random variable $R_1 \cdot y$ is a sum of independent random variables $X_{1,1}\, \overline{y_1}, \ldots, X_{1,n}\, \overline{y_n}$, and one of them is absolutely continuous with a density bounded above by $\sqrt{n}\, |f|_\infty$. Consequently, by a basic property of convolutions of probability distributions, the random variable $R_1 \cdot y$ is itself absolutely continuous with a density φ bounded above by $\sqrt{n}\, |f|_\infty$. Therefore, we have,

$$\mathbb{P}(|R_1 \cdot y| \leqslant u) = \int_{\{z \in K ; |z| \leqslant u\}} \varphi(s)\, ds \leqslant \begin{cases} 2u\sqrt{n}\, |f|_\infty & \text{if } K = \mathbb{R}, \\ \pi u^2 \sqrt{n}\, |f|_\infty & \text{if } K = \mathbb{C}. \end{cases}$$

□

The proof of lemma 3.4 above is quite instructive. Let us focus on the control of $\mathbb{P}(|R_1 \cdot y| \leqslant u)$ when u is small. In the case where $\mathcal{L}$ is Gaussian, the rotational invariance of the distribution of R_1 implies that the quantity $\mathbb{P}(|R_1 \cdot y| \leqslant u)$ does not depend on y and is of order u (take for y an element of the canonical basis and use the fact that $\mathcal{L}$ has a bounded density). However, when $\mathcal{L}$ is not Gaussian, the quantity $\mathbb{P}(|R_1 \cdot y| \leqslant u)$ depends heavily on $\mathcal{L}$. Recall that the simple lemma above does not allow atoms in $\mathcal{L}$. In particular, it does not cover the case where $\mathcal{L}$ is Rademacher $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$. Such discrete matrices have a positive probability of being singular. In this discrete case, the quantity $\mathbb{P}(|R_1 \cdot y| \leqslant u)$ depends heavily on the arithmetic and sparsity structure of the coordinates of y. For instance, if $y = n^{-1/2}(e_1 + \cdots + e_n)$, then by the central limit theorem, the quantity $\mathbb{P}(|R_1 \cdot y| \leqslant u)$ is of order u as $n \to \infty$, whereas if $y = 2^{-1/2}(e_1 + e_2)$ then we get a completely different behavior:

$$\mathbb{P}(|R_1 \cdot y| \leqslant u) \geqslant \mathbb{P}(X_{1,1} = 0) = 1/2.$$

Also, one should restart from $s_n(X) = \min_{|x|_2 = 1} |Xx|_2$ and partition the unit sphere into "compressible" and "incompressible" vectors. This leads to the use of ε–nets

techniques and to the consideration of Littlewood–Offord type problems for the control of small balls probabilities. In this direction, an important step was first made by Rudelson [Rud06]. Later, Rudelson and Vershynin [RV08b, RV08a] have shown that if $X_{1,1}$ has zero mean, unit variance, and finite fourth moment, then for any fixed $t > 0$ (recall that $m = n$),

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leqslant t) \leqslant f(t) + o(1) \quad \text{and} \quad \mathbb{P}(\sqrt{n}\, s_n(X) \geqslant t) \leqslant g(t) + o(1)$$

where f, g do not depend on n and $f(t), g(t) \to 0$ as $t \to 0$, and where $o(1)$ is relative to $n \to \infty$. Moreover, if the entries are additionally sub-Gaussian then there exist constants $C > 0$, $c \in (0, 1)$ depending only on the moments such that for any $t \geqslant 0$,

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leqslant t) \leqslant C\, t + c^n. \tag{12}$$

Since the Rademacher law is sub-Gaussian, the remarkable bound (12) proves, up to the multiplicative constant C, the conjecture of Spielman and Teng (11). The proof of Rudelson and Vershynin has many ingredients, including an upper bound on the right tail of the largest singular value and a lower bound on the smallest singular value of rectangular matrices obtained in [LPRTJ05] (see also the more recent work [RV09]). Regarding moments, Tao and Vu have shown [TV08, TV09a] that under the sole assumption that $X_{1,1}$ has non zero finite variance then for any constants $a, c > 0$ there exists $b > 1/2$ depending on a, c and the law of $X_{1,1}$ such that for any $n \times n$ deterministic matrix Y with $|Y|_{2\to 2} \leqslant n^c$,

$$\mathbb{P}(\sqrt{n}\, s_n(X + Y) \leqslant n^{-b}) \leqslant n^{-a}. \tag{13}$$

Actually, one can find in [TV09c] many bounds of this flavor. For instance, under the sole assumptions that $X_{1,1}$ has zero mean and unit variance, for any fixed $a > 0$,

$$\mathbb{P}(\sqrt{n}\, s_n(X) \leqslant n^{-\frac{1}{2}-\frac{5}{2}a-a^2}) \leqslant n^{-a+o(1)}. \tag{14}$$

The bounds (13–14) are less precise than the bound (12) but do not rely on moments assumptions beyond the finite variance. With lemma 3.4 in mind, one can ask if the finite moment assumption in (13) can be weakened in order to allow for instance heavy tailed non centered discrete laws such as the Zipf type law

$$\zeta(s)^{-1} \sum_{n=0}^{\infty} n^{-s} \delta_n$$

where $s > 0$, and where ζ is the Riemann zeta function.

4. Comments

The singular values of deterministic matrices are studied in many books such as [HJ90, HJ94], [Bha97], and [Zha02]. For the algorithmic aspects, we recommend [GVL96] and [CG05]. The singular values of random matrices are studied in the books [Meh04], [Dei99], [For10], [AGZ09].

References

[AGL+08] Radosław Adamczak, Olivier Guédon, Alexander Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann, *Smallest singular value of random matrices with independent columns*, C. R. Math. Acad. Sci. Paris **346** (2008), no. 15-16, 853–856. MR MR2441920 (2009h:15009)

[AGZ09] Greg Anderson, Alice Guionnet, and Ofer Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press series, 2009.

[AS64] Milton Abramowitz and Irene A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, National Bureau of Standards Applied Mathematics Series, vol. 55, For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964. MR MR0167642 (29 #4914)

[AW05] Jean-Marc Azaïs and Mario Wschebor, *Upper and lower bounds for the tails of the distribution of the condition number of a Gaussian matrix*, SIAM J. Matrix Anal. Appl. **26** (2004/05), no. 2, 426–440 (electronic). MR MR2124157 (2005k:15016)

[AW09] ———, *Level sets and extrema of random processes and fields*, John Wiley & Sons Inc., Hoboken, NJ, 2009. MR MR2478201

[Bai99] Z. D. Bai, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica **9** (1999), no. 3, 611–677, With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author. MR MR1711663 (2000e:60044)

[BAP05] G. Ben Arous and S. Péché, *Universality of local eigenvalue statistics for some sample covariance matrices*, Comm. Pure Appl. Math. **58** (2005), no. 10, 1316–1357. MR MR2162782 (2006h:62072)

[BB98] Mario Bertero and Patrizia Boccacci, *Introduction to inverse problems in imaging*, Institute of Physics Publishing, Bristol, 1998. MR MR1640759 (99k:78027)

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. **33** (2005), no. 5, 1643–1697. MR MR2165575 (2006g:15046)

[BDG09] Serban Belinschi, Amir Dembo, and Alice Guionnet, *Spectral measure of heavy tailed band and covariance random matrices*, Comm. Math. Phys. **289** (2009), no. 3, 1023–1055. MR MR2511659

[BF03] Alexei Borodin and Peter J. Forrester, *Increasing subsequences and the hard-to-soft edge transition in matrix ensembles*, J. Phys. A **36** (2003), no. 12, 2963–2981, Random matrix theory. MR MR1986402 (2004m:82060)

[BG09] Laurent Bruneau and François Germinet, *On the singularity of random matrices with independent entries*, Proc. Amer. Math. Soc. **137** (2009), no. 3, 787–792. MR MR2457415

[Bha97] Rajendra Bhatia, *Matrix analysis*, Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997. MR MR1477662 (98i:15003)

[BS98] Z. D. Bai and Jack W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Ann. Probab. **26** (1998), no. 1, 316–345. MR MR1617051 (99b:60041)

[BS10] Zhidong Bai and Jack W. Silverstein, *Spectral analysis of large dimensional random matrices*, second ed., Springer Series in Statistics, Springer, New York, 2010. MR MR2567175

[BVW10] Jean Bourgain, Van H. Vu, and Philip Matchett Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. **258** (2010), no. 2, 559–603. MR MR2557947

[BY93] Z. D. Bai and Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Ann. Probab. **21** (1993), no. 3, 1275–1294. MR MR1235416 (94j:60060)

[CD05] Zizhong Chen and Jack J. Dongarra, *Condition numbers of Gaussian random matrices*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 3, 603–620 (electronic). MR MR2208323 (2006k:15057)

[CG05] Moody T. Chu and Gene H. Golub, *Inverse eigenvalue problems: theory, algorithms, and applications*, Numerical Mathematics and Scientific Computation, Oxford University Press, New York, 2005. MR MR2263317 (2007i:65001)

[Dei99] P. A. Deift, *Orthogonal polynomials and random matrices: a Riemann-Hilbert approach*, Courant Lecture Notes in Mathematics, vol. 3, New York University Courant Institute of Mathematical Sciences, New York, 1999. MR MR1677884 (2000g:47048)

[Dei07] Percy A. Deift, *Universality for mathematical and physical systems*, International Congress of Mathematicians. Vol. I, Eur. Math. Soc., Zürich, 2007, pp. 125–152. MR MR2334189 (2008g:60024)

[Dem88] James W. Demmel, *The probability that a numerical analysis problem is difficult*, Math. Comp. **50** (1988), no. 182, 449–480. MR MR929546 (89g:65062)

[DK08] Ioana Dumitriu and Plamen Koev, *Distributions of the extreme eigenvalues of beta-Jacobi random matrices*, SIAM J. Matrix Anal. Appl. **30** (2008), no. 1, 1–6. MR MR2399565 (2009a:65089)

[DS01] Kenneth R. Davidson and Stanislaw J. Szarek, *Local operator theory, random matrices and Banach spaces*, Handbook of the geometry of Banach spaces, Vol. I, North-Holland, Amsterdam, 2001, pp. 317–366. MR MR1863696 (2004f:47002a)

[DS07a] R. Brent Dozier and Jack W. Silverstein, *Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices*, J. Multivariate Anal. **98** (2007), no. 6, 1099–1122. MR MR2326242 (2008h:62056)

[DS07b] ———, *On the empirical distribution of eigenvalues of large dimensional information-plus-noise-type matrices*, J. Multivariate Anal. **98** (2007), no. 4, 678–694. MR MR2322123 (2009d:60078)

[Ede88] Alan Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. **9** (1988), no. 4, 543–560. MR MR964668 (89j:15039)

[EK07] Noureddine El Karoui, *Tracy-Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices*, Ann. Probab. **35** (2007), no. 2, 663–714. MR MR2308592 (2007m:60057)

[ER05] Alan Edelman and N. Raj Rao, *Random matrix theory*, Acta Numer. **14** (2005), 233–297. MR MR2168344 (2008f:15081)

[ES05] Alan Edelman and Brian D. Sutton, *Tails of condition number distributions*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 2, 547–560 (electronic). MR MR2179688 (2006j:15018)

[EY39] Carl Eckart and Gale Young, *A principal axis transformation for non-hermitian matrices*, Bull. Amer. Math. Soc. **45** (1939), no. 2, 118–121. MR MR1563923

[For10] Peter Forrester, *Log-Gases and Random Matrices*, Princeton University Press, 2010.

[FP09] Delphine Féral and Sandrine Péché, *The largest eigenvalues of sample covariance matrices for a spiked population: diagonal case*, J. Math. Phys. **50** (2009), no. 7, 073302, 33. MR MR2548630

[Fre03] Ivar Fredholm, *Sur une classe d'équations fonctionnelles*, Acta Math. **27** (1903), no. 1, 365–390. MR MR1554993

[FS10] Ohad N. Feldheim and Sasha Sodin, *A universality result for the smallest eigenvalues of certain sample covariance matrices*, Geom. Funct. Anal. **20** (2010), no. 1, 88–123.

[Gel41] I. Gelfand, *Normierte Ringe*, Rec. Math. [Mat. Sbornik] N. S. **9 (51)** (1941), 3–24. MR MR0004726 (3,51f)

[Gem80] Stuart Geman, *A limit theorem for the norm of random matrices*, Ann. Probab. **8** (1980), no. 2, 252–261. MR MR566592 (81m:60046)

[Gin65] Jean Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, J. Mathematical Phys. **6** (1965), 440–449. MR MR0173726 (30 #3936)

[GVL96] Gene H. Golub and Charles F. Van Loan, *Matrix computations*, third ed., Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, 1996. MR MR1417720 (97g:65006)

[HJ90] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990, Corrected reprint of the 1985 original. MR MR1084815 (91i:15001)

[HJ94] ———, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994, Corrected reprint of the 1991 original. MR MR1288752 (95c:15001)

[HLN06] W. Hachem, P. Loubaton, and J. Najim, *The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile*, Ann. Inst. H. Poincaré Probab. Statist. **42** (2006), no. 6, 649–670. MR MR2269232 (2009b:60010)

[Hor54] Alfred Horn, *On the eigenvalues of a matrix with prescribed singular values*, Proc. Amer. Math. Soc. **5** (1954), 4–7. MR MR0061573 (15,847d)

[HP00] Fumio Hiai and Dénes Petz, *The semicircle law, free random variables and entropy*, Mathematical Surveys and Monographs, vol. 77, American Mathematical Society, Providence, RI, 2000. MR MR1746976 (2001j:46099)

[HT03] Uffe Haagerup and Steen Thorbjørnsen, *Random matrices with complex Gaussian entries*, Expo. Math. **21** (2003), no. 4, 293–337. MR MR2022002 (2005b:46142)

[HW53] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J. **20** (1953), 37–39. MR MR0052379 (14,611b)

[Jam60] Alan T. James, *The distribution of the latent roots of the covariance matrix*, Ann. Math. Statist. **31** (1960), 151–158. MR MR0126901 (23 #A4195)

[Joh00] Kurt Johansson, *Shape fluctuations and random matrices*, Comm. Math. Phys. **209** (2000), no. 2, 437–476. MR MR1737991 (2001h:60177)

[Joh01] Iain M. Johnstone, *On the distribution of the largest eigenvalue in principal components analysis*, Ann. Statist. **29** (2001), no. 2, 295–327. MR MR1863961 (2002i:62115)

[Jol02] I. T. Jolliffe, *Principal component analysis*, second ed., Springer Series in Statistics, Springer-Verlag, New York, 2002. MR MR2036084 (2004k:62010)

[KKS95] Jeff Kahn, János Komlós, and Endre Szemerédi, *On the probability that a random* ± 1*-matrix is singular*, J. Amer. Math. Soc. **8** (1995), no. 1, 223–240. MR MR1260107 (95c:15047)

[Kom67] J. Komlós, *On the determinant of* (0, 1) *matrices*, Studia Sci. Math. Hungar **2** (1967), 7–21. MR MR0221962 (36 #5014)

[Kos88] Eric Kostlan, *Complexity theory of numerical linear algebra*, J. Comput. Appl. Math. **22** (1988), no. 2-3, 219–230, Special issue on emerging paradigms in applied mathematical modelling. MR MR956504 (89i:65048)

[Led04] M. Ledoux, *Differential operators and spectral distributions of invariant ensembles from the classical orthogonal polynomials. The continuous case*, Electron. J. Probab. **9** (2004), no. 7, 177–208 (electronic). MR MR2041832 (2004m:60070)

[LPRTJ05] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), no. 2, 491–523. MR MR2146352 (2006g:52009)

[Meh04] Madan Lal Mehta, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004. MR MR2129906 (2006b:82001)

[Moo20] E. H. Moore, *On the reciprocal of the general algebraic matrix*, Bulletin of the American Mathematical Society **26** (1920), 394–395.

[MP67] V. A. Marchenko and L. A. Pastur, *Distribution of eigenvalues in certain sets of random matrices*, Mat. Sb. (N.S.) **72 (114)** (1967), 507–536. MR MR0208649 (34 #8458)

[MP06] Shahar Mendelson and Alain Pajor, *On singular values of matrices with independent rows*, Bernoulli **12** (2006), no. 5, 761–773. MR MR2265341 (2008a:60160)

[Mui82] Robb J. Muirhead, *Aspects of multivariate statistical theory*, John Wiley & Sons Inc., New York, 1982, Wiley Series in Probability and Mathematical Statistics. MR MR652932 (84c:62073)

[Péc09] Sandrine Péché, *Universality results for the largest eigenvalues of some sample covariance matrix ensembles*, Probab. Theory Related Fields **143** (2009), no. 3-4, 481–516. MR MR2475670 (2009m:60013)

[Pen56] R. Penrose, *On best approximation solutions of linear matrix equations*, Proc. Cambridge Philos. Soc. **52** (1956), 17–19. MR MR0074092 (17,536d)

[Res08] Sidney I. Resnick, *Extreme values, regular variation and point processes*, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2008, Reprint of the 1987 original. MR MR2364939 (2008h:60002)

[RRV09] José Ramírez, Brian Rider, and Báling Virág, *Beta ensembles, stochastic airy spectrum, and a diffusion*, preprint `arXiv:math/0607331v4 [math.PR]`, 2009.

[Rud06] M. Rudelson, *Lower estimates for the singular values of random matrices*, C. R. Math. Acad. Sci. Paris **342** (2006), no. 4, 247–252. MR MR2196007 (2007h:15029)

[Rud08] ———, *Invertibility of random matrices: norm of the inverse*, Ann. of Math. (2) **168** (2008), no. 2, 575–600. MR MR2434885

[RV08a] M. Rudelson and R. Vershynin, *The least singular value of a random square matrix is $O(n^{-1/2})$*, C. R. Math. Acad. Sci. Paris **346** (2008), no. 15-16, 893–896. MR MR2441928 (2009i:60104)

[RV08b] ———, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633. MR MR2407948

[RV09] Mark Rudelson and Roman Vershynin, *Smallest singular value of a random rectangular matrix*, Comm. Pure Appl. Math. **62** (2009), no. 12, 1707–1739. MR MR2569075

[RVA05] T. Ratnarajah, R. Vaillancourt, and M. Alvo, *Eigenvalues and condition numbers of complex random matrices*, SIAM J. Matrix Anal. Appl. **26** (2004/05), no. 2, 441–456 (electronic). MR MR2124158 (2005k:62142)

[SF05] Alexander Soshnikov and Yan V. Fyodorov, *On the largest singular values of random matrices with independent Cauchy entries*, J. Math. Phys. **46** (2005), no. 3, 033302, 15. MR MR2125574 (2005k:82034)

[Sil85] Jack W. Silverstein, *The smallest eigenvalue of a large-dimensional Wishart matrix*, Ann. Probab. **13** (1985), no. 4, 1364–1368. MR MR806232 (87b:60050)

[Sma85] Steve Smale, *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc. (N.S.) **13** (1985), no. 2, 87–121. MR MR799791 (86m:65061)

[Sos02] Alexander Soshnikov, *A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices*, J. Statist. Phys. **108** (2002), no. 5-6, 1033–1056, Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays. MR MR1933444 (2003h:62108)

[Sos06] ———, *Poisson statistics for the largest eigenvalues in random matrix ensembles*, Mathematical physics of quantum mechanics, Lecture Notes in Phys., vol. 690, Springer, Berlin, 2006, pp. 351–364. MR MR2234922 (2007f:82036)

[ST02] Daniel A. Spielman and Shang-Hua Teng, *Smoothed analysis of algorithms*, Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002) (Beijing), Higher Ed. Press, 2002, pp. 597–606. MR MR1989210 (2004d:90138)

[ST03] ———, *Smoothed analysis: motivation and discrete models*, Algorithms and data structures, Lecture Notes in Comput. Sci., vol. 2748, Springer, Berlin, 2003, pp. 256–270. MR MR2078601 (2005c:68304)

[Sza91] Stanisław J. Szarek, *Condition numbers of random matrices*, J. Complexity **7** (1991), no. 2, 131–149. MR MR1108773 (92i:65086)

[Sze75] Gábor Szegő, *Orthogonal polynomials*, fourth ed., American Mathematical Society, Providence, R.I., 1975, American Mathematical Society, Colloquium Publications, Vol. XXIII. MR MR0372517 (51 #8724)

[Tar05] Albert Tarantola, *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005. MR MR2130010 (2007b:62011)

[TE05] Lloyd N. Trefethen and Mark Embree, *Spectra and pseudospectra*, Princeton University Press, Princeton, NJ, 2005, The behavior of nonnormal matrices and operators. MR MR2155029 (2006d:15001)

[Tho76] R. C. Thompson, *The behavior of eigenvalues and singular values under perturbations of restricted rank*, Linear Algebra and Appl. **13** (1976), no. 1/2, 69–78, Collection of articles dedicated to Olga Taussky Todd.

[Tik43] A. N. Tikhonov, *On the stability of inverse problems*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **39** (1943), 176–179. MR MR0009685 (5,184e)

[TV06] T. Tao and V. Vu, *On random ± 1 matrices: singularity and determinant*, Random Structures Algorithms **28** (2006), no. 1, 1–23. MR MR2187480 (2006g:15048)

[TV07] ———, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628 (electronic). MR MR2291914 (2008h:60027)

[TV08] ———, *Random matrices: the circular law*, Commun. Contemp. Math. **10** (2008), no. 2, 261–307. MR MR2409368 (2009d:60091)

[TV09a] ———, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) **169** (2009), no. 2, 595–632. MR MR2480613

[TV09b] ———, *Random matrices: The distribution of the smallest singular values*, preprint `arXiv:0903.0614 [math.PR]`, 2009.

[TV09c] ———, *Smooth analysis of the condition number and the least singular value*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, Lecture Notes in Computer Science, vol. 5687, Springer, 2009.

[TV10] ———, *Random matrices: Universality of ESDs and the circular law*, Annals of Probability **38** (2010), no. 5, 2023–2065.

[TW02] Craig A. Tracy and Harold Widom, *Distribution functions for largest eigenvalues and their applications*, Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002) (Beijing), Higher Ed. Press, 2002, pp. 587–596. MR MR1989209 (2004f:82034)

[vN63] John von Neumann, *Collected works. Vol. V: Design of computers, theory of automata and numerical analysis*, General editor: A. H. Taub. A Pergamon Press Book, The Macmillan Co., New York, 1963. MR MR0157875 (28 #1104)

[vNG47] John von Neumann and H. H. Goldstine, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc. **53** (1947), 1021–1099. MR MR0024235 (9,471b)

[Wac78] Kenneth W. Wachter, *The strong limits of random matrix spectra for sample matrices of independent elements*, Ann. Probability **6** (1978), no. 1, 1–18. MR MR0467894 (57 #7744)

[Wey49] Hermann Weyl, *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proc. Nat. Acad. Sci. U. S. A. **35** (1949), 408–411. MR MR0030693 (11,37d)

[Wis28] J. Wishart, *The generalized product moment distribution in samples from a normal multivariate population*, Biometrika **20** (1928), no. A, 32–52.

[Zha02] Xingzhi Zhan, *Matrix inequalities*, Lecture Notes in Mathematics, vol. 1790, Springer-Verlag, Berlin, 2002. MR MR1927396 (2003h:15030)

Laboratoire d'Analyse et de Mathématiques Appliquées, UMR CNRS 8050, 5 boulevard Descartes, Champs-sur-Marne, F-77454 Marne-la-Vallée, Cedex 2, France

E-mail address: `djalil.chafai(at)univ-mlv.fr`

URL: `http://perso-math.univ-mlv.fr/users/chafai.djalil/`