Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 346 (2008) 893–896
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Probability Theory

The least singular value of a random square matrix is $O(n^{-1/2})$

Mark Rudelson [a,1], Roman Vershynin [b,2]

[a] Department of Mathematics, University of Missouri, Columbia, MO 65211, USA
[b] Department of Mathematics, University of California, Davis, CA 95616, USA

Received 28 May 2008; accepted 10 July 2008

Presented by Gilles Pisier

Abstract

Let A be a matrix whose entries are real i.i.d. centered random variables with unit variance and suitable moment assumptions. Then the smallest singular value $s_n(A)$ is of order $n^{-1/2}$ with high probability. The lower estimate of this type was proved recently by the authors; in this Note we establish the matching upper estimate. ***To cite this article: M. Rudelson, R. Vershynin, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Résumé

La plus petite valeur singulière d'une matrice carrée aléatoire est en $\mathbf{O(n^{-1/2})}$ Soit A une matrice dont les entrées sont des variables aléatoires centrées réelles i.i.d. de variance 1 vérifiant une hypothèse adéquate de moment. Alors la plus petite valeur singulière $s_n(A)$ est de l'ordre de $n^{-1/2}$ avec grande probabilité. La minoration de $s_n(A)$ a été récemment obtenue par les auteurs ; dans cette Note, nous prouvons la majoration. ***Pour citer cet article : M. Rudelson, R. Vershynin, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

1. Introduction

Let A be an $n \times n$ matrix whose entries are real i.i.d. centered random variables with suitable moment assumptions. Random matrix theory studies the distribution of the *singular values* $s_k(A)$, which are the eigenvalues of $|A| = \sqrt{A^*A}$ arranged in the non-increasing order. In this paper we study the magnitude of the smallest singular value $s_n(A)$, which can also be viewed as the reciprocal of the spectral norm:

$$s_n(A) = \inf_{x:\ \|x\|_2=1} \|Ax\|_2 = 1/\|A^{-1}\|. \tag{1}$$

Motivated by numerical inversion of large matrices, von Neumann and his associates speculated that

$$s_n(A) \sim n^{-1/2} \quad \text{with high probability.} \tag{2}$$

E-mail addresses: rudelson@math.missouri.edu (M. Rudelson), vershynin@math.ucdavis.edu (R. Vershynin).

[1] M.R. was supported by NSF DMS grant 0652684.
[2] R.V. was supported by the Alfred P. Sloan Foundation and by NSF DMS grant 0652617.

1631-073X/$ – see front matter © 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crma.2008.07.009

(See [4, pp. 14, 477, 555].) A more precise form of this estimate was conjectured by Smale and proved by Edelman [1] for Gaussian matrices A. For general matrices, conjecture (2) had remained open until we proved in [2] the lower bound $s_n(A) = \Omega(n^{-1/2})$. In the present paper, we shall prove the corresponding upper bound $s_n(A) = \mathrm{O}(n^{-1/2})$, thereby completing the proof of (2).

Theorem 1.1 *(Fourth moment). Let A be an $n \times n$ matrix whose entries are i.i.d. centered random variables with unit variance and fourth moment bounded by B. Then, for every $\delta > 0$ there exist $K > 0$ and n_0 which depend (polynomially) only on δ and B, and such that*

$$\mathbb{P}\big(s_n(A) > Kn^{-1/2}\big) \leqslant \delta \quad \textit{for all } n \geqslant n_0.$$

Remark. The same result but with the reverse estimate, $\mathbb{P}(s_n(A) < Kn^{-1/2}) \leqslant \delta$, was proved in [2]. Together, these two estimates amount to (2).

Under more restrictive (but still quite general) moment assumptions, Theorem 1.1 takes the following sharper form. Recall that a random variable ξ is called *subgaussian* if its tail is dominated by that of the standard normal random variable: there exists $B > 0$ such that $\mathbb{P}(|\xi| > t) \leqslant 2\exp(-t^2/B^2)$ for all $t > 0$. The minimal B is called the *subgaussian moment* of ξ. The class of subgaussian random variables includes, among others, normal, symmetric ± 1, and in general all bounded random variables.

Theorem 1.2 *(Subgaussian). Let A be an $n \times n$ matrix whose entries are i.i.d. centered random variables with unit variance and subgaussian moment bounded by B. Then for every $K \geqslant 2$ one has*

$$\mathbb{P}\big(s_n(A) > Kn^{-1/2}\big) \leqslant (C/K)\log K + c^n, \tag{3}$$

where $C > 0$ and $c \in (0,1)$ depend (polynomially) only on B.

Remark. A reverse result was proved in [2]: for every $\varepsilon \geqslant 0$, one has $\mathbb{P}(s_n(A) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon + c^n$.

Our argument is an application of the small ball probability bounds and the structure theory developed in [2] and [3]. We shall give a complete proof of Theorem 1.2 only; we leave to the interested reader to modify the argument as in [2] to obtain Theorem 1.1.

2. Proof of Theorem 1.2

By $(e_k)_{k=1}^n$ we denote the canonical basis of the Euclidean space $\mathbb{R}^n$ equipped with the canonical inner product $\langle \cdot, \cdot \rangle$ and Euclidean norm $\| \cdot \|_2$. By $C, C_1, c, c_1, \ldots$ we shall denote positive constants that may possibly depend only on the subgaussian moment B.

Consider vectors $(X_k)_{k=1}^n$ and $(X_k^*)_{k=1}^n$ an n-dimensional Hilbert space H. Recall that the system $(X_k, X_k^*)_{k=1}^n$ is called a *biorthogonal system* in H if $\langle X_j^*, X_k \rangle = \delta_{j,k}$ for all $j, k = 1, \ldots, n$. The system is called *complete* if $\mathrm{span}(X_k) = H$. The following notation will be used throughout the paper:

$$H_k := \mathrm{span}(X_i)_{i \neq k}, \quad H_{j,k} := \mathrm{span}(X_i)_{i \notin \{j,k\}}, \quad j, k = 1, \ldots, n. \tag{4}$$

The next proposition summarizes some elementary and known properties of biorthogonal systems:

Proposition 2.1 *(Biorthogonal systems).* 1. *Let A be an $n \times n$ invertible matrix with columns $X_k = Ae_k$, $k = 1, \ldots, n$. Define $X_k^* = (A^{-1})^* e_k$. Then $(X_k, X_k^*)_{k=1}^n$ is a complete biorthogonal system in $\mathbb{R}^n$.*

2. *Let $(X_k)_{k=1}^n$ be a linearly independent system in an n-dimensional Hilbert space H. Then there exist unique vectors $(X_k^*)_{k=1}^n$ such that $(X_k, X_k^*)_{k=1}^n$ is a biorthogonal system in H. This system is complete.*

3. *Let $(X_k, X_k^*)_{k=1}^n$ be a complete biorthogonal system in a Hilbert space H. Then $\|X_k^*\|_2 = 1/\mathrm{dist}(X_k, H_k)$ for $k = 1, \ldots, n$.*

Without loss of generality, we can assume that $n \geqslant 2$ and that A is a.s. invertible (by adding independent normal random variables with small variance to all entries of A).

Let $u, v > 0$. By (1), the following implication holds:

$$\exists x \in \mathbb{R}^n\colon\ \|x\|_2 \leqslant u,\ \left\|A^{-1}x\right\|_2 \geqslant vn^{1/2} \quad \text{implies} \quad s_n(A) \leqslant (u/v)n^{-1/2}. \tag{5}$$

We will now describe how to find such x. Consider the columns $X_k = Ae_k$ of A and the subspaces H_k, $H_{j,k}$ defined in (4). Let P_1 denote the orthogonal projection in $\mathbb{R}^n$ onto H_1. We define the vector

$$x := X_1 - P_1 X_1.$$

Define $X_k^* = (A^{-1})^* e_k$. By Proposition 2.1 $(X_k, X_k^*)_{k=1}^n$ is a complete biorthogonal system in $\mathbb{R}^n$, so

$$\ker(P_1) = \operatorname{span}\bigl(X_1^*\bigr). \tag{6}$$

Clearly, $\|x\|_2 = \operatorname{dist}(X_1, H_1)$. Conditioning on H_1 and using a standard concentration bound, we obtain

$$\mathbb{P}\bigl(\|x\|_2 > u\bigr) \leqslant C\mathrm{e}^{-cu^2}, \quad u > 0. \tag{7}$$

This settles the first bound in (5) with high probability.

To address the second bound in (5), we write $A^{-1}x = A^{-1}X_1 - A^{-1}P_1X_1 = e_1 - A^{-1}P_1X_1$. Since $P_1X_1 \in H_1$, the vector $A^{-1}P_1X_1$ is supported in $\{2, \ldots, n\}$ and hence is orthogonal to e_1. Therefore

$$\left\|A^{-1}x\right\|_2^2 > \left\|A^{-1}P_1X_1\right\|_2^2 = \sum_{k=1}^n \bigl\langle A^{-1}P_1X_1, e_k\bigr\rangle^2 = \sum_{k=1}^n \bigl\langle P_1\bigl(A^{-1}\bigr)^* e_k, X_1\bigr\rangle^2 = \sum_{k=1}^n \bigl\langle P_1X_k^*, X_1\bigr\rangle^2.$$

The first term of the last sum is zero since $P_1X_1^* = 0$ by (6). We have proved that

$$\left\|A^{-1}x\right\|_2^2 \geqslant \sum_{k=2}^n \bigl\langle Y_k^*, X_1\bigr\rangle^2, \quad \text{where } Y_k^* := P_1X_k^* \in H_1,\ k = 2, \ldots, n. \tag{8}$$

Lemma 2.1. $(Y_k^*, X_k)_{k=2}^n$ *is a complete biorthogonal system in* H_1.

Proof. By (8) and (6), $Y_k^* - X_k^* \in \ker(P_1) = \operatorname{span}(X_1^*)$, so $Y_k^* = X_k^* - \lambda_k X_1^*$ for some $\lambda_k \in \mathbb{R}$ and all $k = 2, \ldots, n$. By the orthogonality of X_1^* to all of X_k, $k = 2, \ldots, n$, we have $\langle Y_j^*, X_k\rangle = \langle X_j^*, X_k\rangle = \delta_{j,k}$ for all $j, k = 2, \ldots, n$. The biorthogonality is proved. The completeness follows since $\dim(H_1) = n - 1$. □

In view of the uniqueness in Part 2 of Proposition 2.1, Lemma 2.1 has the following crucial consequence:

Corollary 2.2. *The system of vectors* $(Y_k^*)_{k=2}^n$ *is uniquely determined by the system* $(X_k)_{k=2}^n$. *In particular, the system* $(Y_k^*)_{k=2}^n$ *and the vector* X_1 *are statistically independent.*

By Part 3 of Proposition 2.1, $\|Y_k^*\|_2 = 1/\operatorname{dist}(X_k, H_{1,k})$. We have therefore proved that

$$\left\|A^{-1}x\right\|_2^2 \geqslant \sum_{k=2}^n (a_k/b_k)^2, \quad \text{where } a_k = \left|\left\langle \frac{Y_k^*}{\|Y_k^*\|_2}, X_1\right\rangle\right|, \quad b_k = \operatorname{dist}(X_k, H_{1,k}). \tag{9}$$

We will now need to bound a_k above and b_k below. Without loss of generality, we will do this for $k = 2$.

We are going to use a result of [3] that states that random subspaces have no additive structure. The amount of structure is formalized by the concept of the least common denominator. Given parameters $\alpha > 0$ and $\gamma \in (0, 1)$, the *least common denominator* of a vector $a \in \mathbb{R}^n$ is defined as

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf\bigl\{\theta > 0\colon\ \operatorname{dist}\bigl(\theta a, \mathbb{Z}^N\bigr) < \min\bigl(\gamma\|\theta a\|_2, \alpha\bigr)\bigr\}.$$

The least common denominator of a subspace H in $\mathbb{R}^n$ is then defined as

$$\mathrm{LCD}_{\alpha,\gamma}(H) = \inf\bigl\{\mathrm{LCD}_{\alpha,\gamma}(a)\colon\ a \in H,\ \|a\|_2 = 1\bigr\}.$$

Since $H_{1,2}$ is the span of $n - 2$ random vectors with i.i.d. coordinates, Theorem 4.3 of [3] yields that

$$\mathbb{P}\bigl\{\mathrm{LCD}_{\alpha,c}\bigl((H_{1,2})^\perp\bigr) \geqslant \mathrm{e}^{cn}\bigr\} \geqslant 1 - \mathrm{e}^{-cn}$$

where $\alpha = c\sqrt{n}$, and $c > 0$ is some constant that may only depend on the subgaussian moment B.

On the other hand, note that the random vector X_2 is statistically independent of the subspace $H_{1,2}$. So, conditioning on $H_{1,2}$ and using the standard concentration inequality, we obtain

$$\mathbb{P}\big(b_2 = \operatorname{dist}(X_2, H_{1,2}) \geqslant t\big) \leqslant C\mathrm{e}^{-ct^2}, \quad t > 0.$$

Therefore, the event

$$\mathcal{E} := \big\{\mathrm{LCD}_{\alpha,c}\big((H_{1,2})^{\perp}\big) \geqslant \mathrm{e}^{cn},\ b_2 < t\big\} \quad \text{satisfies} \quad \mathbb{P}(\mathcal{E}) \geqslant 1 - \mathrm{e}^{-cn} - C\mathrm{e}^{-ct^2}. \tag{10}$$

Note that the event $\mathcal{E}$ depends only on $(X_j)_{j=2}^n$. So let us fix a realization of $(X_j)_{j=2}^n$ for which $\mathcal{E}$ holds. By Corollary 2.2, the vector Y_2^* is now fixed. By Lemma 2.1, Y_2^* is orthogonal to $(X_j)_{j=3}^n$. Therefore $Y^* := Y_2^*/\|Y_2^*\|_2 \in (H_{1,2})^{\perp}$, and because event $\mathcal{E}$ holds, we have

$$\mathrm{LCD}_{\alpha,c}\big(Y^*\big) \geqslant \mathrm{e}^{cn}.$$

Let us write in coordinates $a_2 = |\langle Y^*, X_1\rangle| = |\sum_{i=1}^n Y^*(i)X_1(i)|$ and recall that $Y^*(i)$ are fixed coefficients with $\sum_{i=1}^n Y^*(i)^2 = 1$, and $X_1(i)$ are i.i.d. random variables. We can now apply Small Ball Probability Theorem 3.3 of [3] (in dimension $m = 1$) for this random sum. It yields

$$\mathbb{P}_{X_1}(a_2 \leqslant \varepsilon) \leqslant C\big(\varepsilon + 1/\mathrm{LCD}_{\alpha,c}\big(Y^*\big) + \mathrm{e}^{-c_1 n}\big) \leqslant C\big(\varepsilon + \mathrm{e}^{-c_2 n}\big). \tag{11}$$

Here the subscript in $\mathbb{P}_{X_1}$ means that we the probability is with respect to the random variable X_1 while the other random variables $(X_j)_{j=2}^n$ are fixed; we will use similar notations later.

Now we unfix all random vectors, i.e. work with $\mathbb{P} = \mathbb{P}_{X_1,\dots,X_n}$. We have

$$\mathbb{P}(a_2 \leqslant \varepsilon \text{ or } b_2 \geqslant t) = \mathbb{E}_{X_2,\dots,X_n}\mathbb{P}_{X_1}(a_2 \leqslant \varepsilon \text{ or } b_2 \geqslant t) \leqslant \mathbb{E}_{X_2,\dots,X_n}\mathbf{1}_{\mathcal{E}}\mathbb{P}_{X_1}(a_2 \leqslant \varepsilon) + \mathbb{P}_{X_2,\dots,X_n}\big(\mathcal{E}^c\big)$$

because $b_2 < t$ on $\mathcal{E}$. By (11) and (10), we continue as

$$\mathbb{P}(a_2 \leqslant \varepsilon \text{ or } b_2 \geqslant t) \leqslant C\big(\varepsilon + \mathrm{e}^{-c_2 n}\big) + \big(\mathrm{e}^{-cn} + C\mathrm{e}^{-ct^2}\big) = C_1\big(\varepsilon + \mathrm{e}^{-c_3 t^2} + \mathrm{e}^{-cn}\big) := p(\varepsilon, t, n).$$

Repeating the above argument for any $k \in \{2, \dots, n\}$ instead of $k = 2$, we conclude that

$$\mathbb{P}(a_k/b_k \leqslant \varepsilon/t) \leqslant p(\varepsilon, t, n) \quad \text{for } \varepsilon > 0,\ t > 0,\ k = 2, \dots, n. \tag{12}$$

From this we can easily deduce the lower bound on the sum of $(a_k/b_k)^2$, which we need for (9). This can be done using the following elementary observation proved by applying Markov's inequality twice.

Proposition 2.2. *Let $Z_k \geqslant 0$, $k = 1, \dots, n$, be random variables. Then, for every $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^n Z_k \leqslant \varepsilon\right) \leqslant \frac{2}{n}\sum_{k=1}^n \mathbb{P}(Z_k \leqslant 2\varepsilon).$$

We use Proposition 2.2 for $Z_k = (a_k/b_k)^2$, along with the bounds (12). In view of (9), we obtain

$$\mathbb{P}\big(\big\|A^{-1}x\big\|_2 \leqslant (\varepsilon/t)n^{1/2}\big) \leqslant 2p(4\varepsilon, t, n). \tag{13}$$

Estimates (7) and (13) settle the desired bounds in (5), and therefore we conclude that

$$\mathbb{P}\big(s_n(A) \leqslant (ut/\varepsilon)n^{-1/2}\big) \geqslant \mathbb{P}\big(\|x\|_2 \leqslant u,\ \big\|A^{-1}x\big\|_2 \geqslant (\varepsilon/t)n^{1/2}\big) \geqslant 1 - C\mathrm{e}^{-cu^2} - 2p(4\varepsilon, t, n).$$

This estimate is valid for all $\varepsilon, u, t > 0$. Choosing $\varepsilon = 1/K$, $u = t = \sqrt{\log K}$, the proof of Theorem 1.2 is complete.

References

[1] A. Edelman, Eigenvalues and condition numbers of random matrices, SIAM J. Matrix Anal. Appl. 9 (1988) 543–560.

[2] M. Rudelson, R. Vershynin, The Littlewood–Offord problem and invertibility of random matrices, Adv. Math. 218 (2008) 600–633.

[3] M. Rudelson, R. Vershynin, The smallest singular value of a random rectangular matrix, submitted for publication.

[4] J. von Neumann, in: A.H. Taub (Ed.), Collected Works, vol. V: Design of Computers, Theory of Automata and Numerical Analysis, A Pergamon Press Book The Macmillan Co., New York, 1963.