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

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).

© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

La plus petite valeur singulière d’une matrice carrée aléatoire est en $O(n^{-1/2})$. 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).

© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

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:

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

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

$$s_n(A) \sim n^{-1/2} \quad \text{with high probability.}$$  \hspace{1cm} (2)
(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 $s_n(A) = O(n^{-1/2})$, 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 $\delta$ and $B$, and such that

$$
\mathbb{P}(s_n(A) > Kn^{-1/2}) \leq \delta \quad \text{for all } n \geq n_0.
$$

**Remark.** The same result but with the reverse estimate, $\mathbb{P}(s_n(A) < Kn^{-1/2}) \leq \delta$, 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 $\xi$ is called subgaussian if its tail is dominated by that of the standard normal random variable: there exists $B > 0$ such that $\mathbb{P}(|\xi| > t) \leq 2 \exp(-t^2/B^2)$ for all $t > 0$. The minimal $B$ is called the subgaussian moment of $\xi$. The class of subgaussian random variables includes, among others, normal, symmetric $\pm 1$, 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 $K \geq 2$ one has

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

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

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

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 $(X_k^*, X_k) = \delta_{j,k}$ for all $j, k = 1, \ldots, n$. The system is called complete if span$(X_k) = H$. The following notation will be used throughout the paper:

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

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 $\|X_k^*\|_2 = 1/\text{dist}(X_k, H_k)$ for $k = 1, \ldots, n$.

Without loss of generality, we can assume that $n \geq 2$ 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:

\[ \exists x \in \mathbb{R}^n: \| x \|_2 \leq u, \quad \| A^{-1} x \|_2 \geq v n^{1/2} \implies s_n(A) \leq (u/v)n^{-1/2}. \]  

(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

\[ \ker(P_1) = \text{span}(X^*_1). \]  

(6)

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

\[ \mathbb{P}(\| x \|_2 > u) \leq Ce^{-cu^2}, \quad u > 0. \]  

(7)

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

To address the second bound in (5), we write \( A^{-1} x = A^{-1} X_1 - A^{-1} P_1 X_1 = e_1 - A^{-1} P_1 X_1 \). Since \( P_1 X_1 \in H_1 \), the vector \( A^{-1} P_1 X_1 \) is supported in \( \{2, \ldots, n\} \) and hence is orthogonal to \( e_1 \). Therefore

\[ A^{-1} X_1 = \sum_{k=1}^n (A^{-1} P_1 X_1, e_k) e_k + \sum_{k=1}^n (P_1 (A^{-1})^* e_k, X_1) + \sum_{k=1}^n (P_1 X^*_k, X_1). \]

The first term of the last sum is zero since \( P_1 X^*_k = 0 \) by (6). We have proved that

\[ A^{-1} X_1 \geq \sum_{k=2}^n (Y^*_k, X_1)^2, \quad \text{where } Y^*_k := P_1 X^*_k \in H_1, \ k = 2, \ldots, n. \]  

(8)

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

**Proof.** By (8) and (6), \( Y^*_k - X^*_k \in \ker(P_1) = \text{span}(X^*_1) \), 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 \( \text{dim}(H_1) = n - 1 \). \( \square \)

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, \( \| Y^*_k \|_2 = 1/\text{dist}(X_k, H_{1,k}) \). We have therefore proved that

\[ A^{-1} X_1 \geq \sum_{k=2}^n (a_k/b_k)^2, \quad \text{where } a_k = \left| \frac{Y^*_k}{\| Y^*_k \|_2}, X_1 \right|, \ b_k = \text{dist}(X_k, H_{1,k}). \]  

(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

\[ \text{LCD}_{\alpha, \gamma}(a) := \inf \{ \theta > 0: \text{dist}(\theta a, \mathbb{Z}^n) < \min(\gamma \| a \|_2, \alpha) \}. \]

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

\[ \text{LCD}_{\alpha, \gamma}(H) := \inf \{ \text{LCD}_{\alpha, \gamma}(a): a \in H, \| a \|_2 = 1 \}. \]

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

\[ \mathbb{P}\left\{ \text{LCD}_{\alpha, c}(H_{1,2}) \leq e^n \right\} \geq 1 - e^{-cn} \]

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

$$\mathbb{P}(b_2 = \text{dist}(X_2, H_{1,2}) \geq t) \leq Ce^{-ct^2}, \quad t > 0.$$  

Therefore, the event

$$\mathcal{E} := \left\{ \text{LCD}_{a,e}(\perp H_{1,2}) \geq e^{cn}, \ b_2 < t \right\}$$

satisfies

$$\mathbb{P}(\mathcal{E}) \geq 1 - e^{-cn} - Ce^{-ct^2}. \quad (10)$$

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 $Y^*_{a}$ is now fixed. By Lemma 2.1, $Y^*_{a}$ is orthogonal to $(X_j)_{j=3}^n$. Therefore $Y^* := Y^*_{a}/\|Y^*_{a}\|_2 \in (H_{1,2})^\perp$, and because event $\mathcal{E}$ holds, we have

$$\text{LCD}_{a,e}(Y^*) \geq e^{cn}.$$  

Let us write in coordinates $a_2 = |(Y^*, X_1)| = |\sum_{i=1}^n Y^*(i)X_1(i)|$ 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

$$\mathbb{P}(X_1(a_2 \leq \varepsilon) \leq C(\varepsilon + 1/\text{LCD}_{a,e}(Y^*) + e^{-cn}) \leq C(\varepsilon + e^{-cn}). \quad (11)$$

Here the subscript in $\mathbb{P}(X_1)$ means that we take the probability 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 fix all random vectors, i.e. work with $\mathbb{P} = \mathbb{P}_{X_2,\ldots,X_n}$. We have

$$\mathbb{P}(a_2 \leq \varepsilon \text{ or } b_2 \geq t) = \mathbb{E}_{X_2,\ldots,X_n} \mathbb{P}_{X_1}(a_2 \leq \varepsilon \text{ or } b_2 \geq t) \leq \mathbb{E}_{X_2,\ldots,X_n} \mathbb{I}_e \mathbb{P}_{X_1}(a_2 \leq \varepsilon) + \mathbb{P}_{X_2,\ldots,X_n}(\mathcal{E}^c)$$

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

$$\mathbb{P}(a_2 \leq \varepsilon \text{ or } b_2 \geq t) \leq C(\varepsilon + e^{-cn}) + (e^{-cn} + C e^{-ct^2}) = C_1(\varepsilon + e^{-ct^2} + e^{-cn}) := p(\varepsilon, t, n).$$

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

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

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 $Z_k \geq 0, \ k = 1, \ldots, n$, be random variables. Then, for every $\varepsilon > 0$, we have

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

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

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

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

$$\mathbb{P}(s_n(A) \leq (ut/e)n^{-1/2}) \geq \mathbb{P}(\|x\|_2 \leq u, \|A^{-1}x\|_2 \geq (\varepsilon/t)n^{1/2}) \geq 1 - Ce^{-cu^2} - 2p(4\varepsilon, t, n).$$

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.