Title

Smoothed analysis of symmetric random matrices with continuous distributions

Journal

Proceedings of the American Mathematical Society, 144(5)

ISSN

0002-9939

Authors

Farrell, Brendan
Vershynin, Roman

Publication Date

2015-10-06

DOI

10.1090/proc/12844

SMOOTHED ANALYSIS OF SYMMETRIC RANDOM MATRICES WITH CONTINUOUS DISTRIBUTIONS

BRENDAN FARRELL AND ROMAN VERSHYNIN

ABSTRACT. We study invertibility of matrices of the form $D+R$ where D is an arbitrary symmetric deterministic matrix, and R is a symmetric random matrix whose independent entries have continuous distributions with bounded densities. We show that $\|(D+R)^{-1}\| = O(n^2)$ with high probability. The bound is completely independent of D. No moment assumptions are placed on R; in particular the entries of R can be arbitrarily heavy-tailed.

1. Introduction

This note concerns the invertibility properties of $n \times n$ random matrices of the type $D+R$, where D is an arbitrary deterministic matrix and R is a random matrix with independent entries. What is the typical value of the spectral norm of the inverse, $\|(D+R)^{-1}\|$?

This question is usually asked in the context of *smoothed analysis* of algorithms [9]. There D is regarded as a given matrix, possibly poorly invertible, and R models random noise. Heuristically, adding noise should improve invertibility properties of D, so the typical value $\|(D+R)^{-1}\|$ should be nicely bounded for any D. Sometimes this is true, but sometimes not quite.

This is indeed the case when R is a real Ginibre matrix, i.e. the entries of R are independent $N(0,1)$ random variables. A result of Sankar, Spielman and Teng [10] states that

$$\mathbb{P}\big\{\|(D+R)^{-1}\| \ge t\sqrt{n}\big\} \le 2.35/t, \quad t>0. \tag{1.1}$$

In particular, $\|(D+R)^{-1}\| = O(\sqrt{n})$ with high probability. Note that this bound is independent of D. It is sharp for $D=0$, since $\|R^{-1}\| \gtrsim \sqrt{n}$ with high probability ([1], see [8]).

For general non-Gaussian matrices R a new phenomenon emerges: *invertibility of $D+R$ can deteriorate as $\|D\| \to \infty$.*

Suppose the entries of R are sub-gaussian[1] i.i.d. random variables with mean zero and variance one. Then a result of Rudelson and Vershynin [6] (as adapted by Pan and Zhou [5]) states that as long as $\|D\| = O(\sqrt{n})$, one has

$$\mathbb{P}\big\{\|(D+R)^{-1}\| \ge t\sqrt{n}\big\} \le C/t + c^n, \quad t>0.$$

Here $C>0$ and $c \in (0,1)$ depend only on a bound on the sub-gaussian moments of the entries of R and on $\|D\|/\sqrt{n}$.

Surprisingly, sensitivity to $\|D\|$ is not an artifact of the proof, but a genuine limitation. Indeed, consider the example where each entry of R equals 1 and -1 with probability 1/4 and

Date: June 2, 2015.

2010 *Mathematics Subject Classification.* 60B20,15B52.

[1]See [14] for an introduction to sub-gaussian distributions. Briefly, a random variable X is sub-gaussian if $p^{-1/2}(\mathbb{E}|X|^p)^{1/p} \le K < \infty$ for all $p \ge 1$; the smallest K can be called the sub-gaussian moment of X.

0 with probability 1/2. Let D be the diagonal matrix with diagonal entries $(0, d, d, \dots, d)$. Then one can show[2] that $\|(D+R)^{-1}\| \gtrsim d/\sqrt{n}$ with probability 1/2. In particular, $\|(D+R)^{-1}\| \gg \sqrt{n}$ as soon as $\|D\| = d \gg n$.

Note however that the typical value of $\|(D+R)^{-1}\|$ remains polynomial in n as long as $\|D\|$ is polynomial in n. This result is due to Tao and Vu [12, 11, 13]; Nguyen [4] proved a similar result for *symmetric* random matrices R.

To summarize, as long as the deterministic part D is not too large, $\|D\| = O(\sqrt{n})$, Sankar-Spielman-Teng's invertibility bound (1.1) remains essentially valid for general random matrices R (with i.i.d. subgaussian entries with zero mean and unit variance). For very large deterministic parts ($\|D\| \gg n$), the bound can fail. It is not clear what happens in the regime $\sqrt{n} \ll \|D\| \lesssim n$.

Taking into account all these results, it would be interesting to describe ensembles of random matrices R for which invertibility properties of $D+R$ are independent of D. In this note we show that if the entries of a symmetric matrix R have *continuous* distributions, then the typical value of $\|(D+R)^{-1}\|$ is polynomially bounded *independently of* D; in particular the bound does not deteriorate as $\|D\| \to \infty$.

Theorem 1.1. *Let A be an $n \times n$ symmetric random matrix in which the entries $\{A_{i,j}\}_{1 \le i \le j \le n}$ are independent and have continuous distributions with densities bounded by K. Then for all $t > 0$,*

$$\mathbb{P}\left\{\|A^{-1}\| \ge n^2 t\right\} \le 8K/t. \tag{1.2}$$

Since we do not assume that the entries have mean zero, this theorem can be applied to matrices of type $A = D + R$, and it yields that $\|(D+R)^{-1}\| = O(n^2)$ with high probability. This bound holds for any deterministic symmetric matrix D, large and small. We conjecture that the bound can be improved to $O(\sqrt{n})$ as in Sankar-Spielman-Teng's result (1.1).

Remark 1.2. We do not place any upper bound assumptions in Theorem 1.1, either on the deterministic part D or the random part R. In particular, the entries of R can be arbitrarily heavy-tailed. The upper bound K on the densities precludes the distributions concentrating near any value, so effectively it is a lower bound on concentration.

Remark 1.3. A result in the same spirit as Theorem 1.1 was proved recently by Rudelson and Vershynin [7] for a different ensemble of random matrices R, namely for *random unitary matrices*. If R is uniformly distributed in $U(n)$ then

$$\mathbb{P}\left\{\|(D+R)^{-1}\| \ge tn^C\right\} \le t^{-c}, \quad t > 0.$$

As in Theorem 1.1, D can be an arbitrary deterministic $n \times n$ matrix; $C, c > 0$ denote absolute constants (independent of D).

Remark 1.4. For the specific class where D is a multiple of identity, sharper results are available than Theorem 1.1. In particular, results by Erdős, Schlein and Yau [2] and Vershynin [15] yield an essentially optimal bound on the resolvent, $\|(D - zI)^{-1}\| = O(\sqrt{n})$. Moreover, the latter estimate does not require that the entries of D have continuous distributions; see [2, 15] for details.

[2] This example is due to M. Rudelson (unpublished); a similar phenomenon was discovered independently by Tao and Vu [13].

Remark 1.5. While Theorem 1.1 is stated for symmetric matrices, it holds as well for Hermitian matrices. The proof for the Hermitian case only requires an easy change to the proof of Lemma 2.1 below.

Remark 1.6. The proof of Theorem 1.1 shows that one can relax the assumption of joint independence of the entries. Is suffices to assume that the individual distribution of each entry A_{ij}, conditioned on all other entries except A_{ji}, has density bounded by K.

In the rest of the paper, we prove Theorem 1.1. The argument is very short and is based on computing the influence of each entry of A on the corresponding entry of A^{-1}.

2. Proof of Theorem 1.1

Recall that the *weak L_p norm* of a random variable X is

$$\|X\|_{p,\infty} := \sup_{t>0} t\,(\mathbb{P}\{|X| > t\})^{1/p}, \quad 0 < p < \infty. \tag{2.1}$$

Lemma 2.1. *Let A be the random matrix defined in Theorem 1.1. Then for all $1 \le i, j \le n$,*

$$\|(A^{-1})_{i,j}\|_{1,\infty} \le 2K.$$

Proof. Let us determine how a single entry of the inverse, say $(A^{-1})_{i,j}$, depends on the corresponding entry of A, i.e. $A_{i,j}$. To this end, let us condition on all entries of A except $A_{i,j}$, thus treating them as constants. We could proceed by the cofactor expansion. But we find it easier to use Jacobi formula, which is valid for an arbitrary square matrix $A = A(t)$ that depends on a parameter t:

$$\frac{d}{dt}|A(t)| = \operatorname{tr}\big[\operatorname{adj}(A(t))\,\frac{dA(t)}{dt}\big].$$

Here and later $|A|$ denotes the determinant and $\operatorname{adj}(A)$ denotes the adjugate matrix of A. Let $A_{(i,j)}$ be the submatrix obtained by removing the i^{th} row and j^{th} column of A, and let $A_{(i,j),(k,l)}$ be the submatrix obtained by removing rows i and k and columns j and l from A.

Consider the off-diagonal case first, where $i \ne j$. The Jacobi formula yields $\frac{d}{dA_{i,j}}|A_{(i,j)}| = (-1)^{i+j}|A_{(i,j),(j,i)}|$, so that

$$|A_{(i,j)}| = (-1)^{i+j}|A_{(i,j),(j,i)}|A_{i,j} + a \tag{2.2}$$

for some constant a (meaning that a does not depend on $A_{i,j}$). Further,

$$\frac{d}{dA_{i,j}}|A| = (-1)^{i+j}(|A_{(i,j)}| + |A_{(j,i)}|) = (-1)^{i+j}2|A_{(i,j)}| = 2|A_{(i,j),(j,i)}|A_{i,j} + (-1)^{i+j}2a.$$

Thus, for some constant b one has

$$|A| = |A_{(i,j),(j,i)}|A_{i,j}^2 + (-1)^{i+j}2aA_{i,j} + b. \tag{2.3}$$

Equations (2.2) and (2.3) and Cramer's rule imply that for all (i,j) there exist constants p, q such that

$$|(A^{-1})_{i,j}| = \left|\frac{|A_{(i,j)}|}{|A|}\right| = \frac{|A_{i,j} + p|}{|(A_{i,j}+p)^2 + q|} = \left|\frac{X}{X^2+q}\right|, \qquad \text{where } X = A_{i,j} + p.$$

First, assume that $q \ge 0$. Then $|(A^{-1})_{i,j}| \le 1/|X|$, and thus we have for all $t > 0$:

$$\mathbb{P}\{|(A^{-1})_{i,j}| > t\} \le \mathbb{P}\{|X| < 1/t\} \le 2K/t. \tag{2.4}$$

Next, assume $0 > q =: -s$; then

$$|(A^{-1})_{i,j}| = \frac{1}{|X - s/X|}.$$

Note that the function $f(x) := x - s/x$ satisfies $f'(x) = 1 + s/x^2 > 1$ for all $x \neq 0$. Thus the set of points $\{x \in \mathbb{R} : |f(x)| < \varepsilon\}$ has diameter at most 2ε for every $\varepsilon > 0$. When $x = X$ is a random variable with density bounded by K, it follows that $\mathbb{P}\{|f(X)| < \varepsilon\} \leq 2K\varepsilon$. Using this for $\varepsilon = 1/t$, we obtain

$$\mathbb{P}\{|(A^{-1})_{i,j}| > t\} \leq \mathbb{P}\{|f(X)| < 1/t\} \leq 2K/t.$$

We have shown that in the off-diagonal case $i \neq j$, the estimate (2.4) always holds.

The diagonal case $i = j$ is similar. The Jacobi formula (or just expanding the determinant along i-th row) shows that $|A| = |A_{(i,i)}|A_{i,i} + c$ for some constant c. Then a similar analysis yields $\mathbb{P}\{|(A^{-1})_{i,j}| > t\} \leq 2K/t$. This completes the proof. $\square$

Proof of Theorem 1.1. Although the weak L_1 norm is not equivalent to a norm, the following inequality holds for any finite sequence of random variables X_i:

$$\Big\| \Big(\sum_i X_i^2 \Big)^{1/2} \Big\|_{1,\infty} \leq 4 \sum_i \|X_i\|_{1,\infty}. \tag{2.5}$$

This inequality is due to Hagelstein (see the proof of Theorem 2 in [3]); it follows by a truncation argument and Chebyshev's inequality. We use (2.5) together with the estimates obtained in Lemma 2.1 to bound the Hilbert-Schmidt norm of A:

$$\big\| \|A^{-1}\|_{\mathrm{HS}} \big\|_{1,\infty} = \Big\| \Big(\sum_{1 \leq i,j \leq n} ((A^{-1})_{i,j})^2 \Big)^{1/2} \Big\|_{1,\infty} \leq 4 \sum_{1 \leq i,j \leq n} \|(A^{-1})_{i,j}\|_{1,\infty} \leq 8Kn^2.$$

The definition of the weak L_1 norm then yields

$$\sup_{t>0} t\mathbb{P}\{\|A^{-1}\|_{\mathrm{HS}} > t\} \leq 8Kn^2.$$

Since $\|A^{-1}\| \leq \|A^{-1}\|_{\mathrm{HS}}$, the proof of Theorem 1.1 is complete. $\square$

Acknowledgments. We thank the referees whose suggestions helped to improve the presentation of this paper.

B. F. was partially supported by Joel A. Tropp under ONR awards N00014-08-1-0883 and N00014-11-1002 and a Sloan Research Fellowship. R. V. was partially supported by NSF grants 1001829, 1265782, and U. S. Air Force Grant FA9550-14-1-0009.

References

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

[2] L. Erdős, B. Schlein, H.-T. Yau,, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. 3 (2010), 436–479.

[3] P. A. Hagelstein, *Weak L^1 norms of random sums*, Proc. Amer. Math. Soc., 133 (2005), 2327–2334.

[4] H. Nguyen, *On the least singular value of random symmetric matrices*, Electron. J. Probab. 17 (2012), 1–19.

[5] G. Pan, W. Zhou, *Circular law, extreme singular values and potential theory*, J. Multivariate Anal. 101 (2010), 645–656.

[6] M. Rudelson, R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600–633.

[7] M. Rudelson, R. Vershynin, *Invertibility of random matrices: unitary and orthogonal perturbations*, J. Amer. Math. Soc. 27 (2014), 293–338.

[8] M. Rudelson, R. Vershynin, *The least singular value of a random square matrix is $O(n^{-1/2})$*, Comptes rendus de l'Académie des sciences - Mathématique 346 (2008), 893–896.

[9] D. Spielman, S.-H. Teng, *Smoothed analysis of algorithms.* Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002), 597–606, Higher Ed. Press, Beijing, 2002.

[10] A. Sankar, D. Spielman, S.-H. Teng, *Smoothed analysis of the condition numbers and growth factors of matrices*, SIAM J. Matrix Anal. Appl. 28 (2006), 446–476.

[11] T. Tao and V. Vu, *The condition number of a randomly perturbed matrix*, STOC07, ACM, 248–255, 2007.

[12] T. Tao and V. Vu, *Random matrices: The circular law*, Commun. Contemp. Math. 10 (2008), 261–307.

[13] T. Tao and V. Vu. *Smooth analysis of the condition number and the least singular value*, Math. Comp. 79 (2010), 2333–2352.

[14] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, Cambridge, 2012.

[15] R. Vershynin, *Invertibility of symmetric random matrices*, Random Structures Algorithms 44 (2014), 135–182.

Computing and Mathematical Sciences, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, U.S.A.

E-mail address: `farrell@cms.caltech.edu`

Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, U.S.A.

E-mail address: `romanv@umich.edu`