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

$$\mathbb{P}\{|| (D + R)^{-1} || \geq t\sqrt{n}\} \leq 2.35/t, \quad t > 0.$$ (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\footnote{See [14] for an introduction to sub-gaussian distributions. Briefly, a random variable $X$ is sub-gaussian if $p^{-1/2}(E|X|^p)^1/p \leq K < \infty$ for all $p \geq 1$; the smallest $K$ can be called the sub-gaussian moment of $X$.} 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

$$\mathbb{P}\{|| (D + R)^{-1} || \geq t\sqrt{n}\} \leq C/t + c^n, \quad t > 0.$$ (1.2)

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...
0 with probability $1/2$. Let $D$ be the diagonal matrix with diagonal entries $(0, d, d, \ldots, d)$. Then one can show\footnote{This example is due to M. Rudelson (unpublished); a similar phenomenon was discovered independently by Tao and Vu \cite{13}.} 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 \cite{12, 11, 13}; Nguyen \cite{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 \cite{11} 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 $\{ A_{i,j} \}_{1 \leq i \leq j \leq n}$ are independent and have continuous distributions with densities bounded by $K$. Then for all $t > 0$,

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

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 \cite{11}.

**Remark 1.2.** We do not place any upper bound assumptions in Theorem \cite{11} 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 \cite{11} was proved recently by Rudelson and Vershynin \cite{7} for a different ensemble of random matrices $R$, namely for random unitary matrices. If $R$ is uniformly distributed in $U(n)$ then

$$
P \left\{ \| (D + R)^{-1} \| \geq tnC \right\} \leq t^{-c}, \quad t > 0.
$$

As in Theorem \cite{11} $D$ 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 \cite{11}. In particular, results by Erdős, Schlein and Yau \cite{2} and Vershynin \cite{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 \cite{2, 15} for details.
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. It 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

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

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

$$
\|(A^{-1})_{ij}\|_{1,\infty} \leq 2K. \quad (2.2)
$$

Proof. Let us determine how a single entry of the inverse, say $(A^{-1})_{ij}$, depends on the corresponding entry of $A$, i.e. $A_{ij}$. To this end, let us condition on all entries of $A$ except $A_{ij}$, 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 $t$:

$$
\frac{d}{dt}(A(t)) = \text{tr}\left[\text{adj}(A(t)) \frac{dA(t)}{dt}\right].
$$

Here and later $|A|$ denotes the determinant and $\text{adj}(A)$ 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 $i \neq j$. The Jacobi formula yields

$$
\frac{d}{dA_{ij}}|A_{(i,j)}| = (-1)^{i+j}|A_{(i,j),(j,i)}|A_{ij} + a \quad (2.2)
$$

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

$$
\frac{d}{dA_{ij}}|A| = (-1)^{i+j}(|A_{ij}| + |A_{(j,i)}|) = (-1)^{i+j}2|A_{(i,j)}| = 2|A_{(i,j),(j,i)}|A_{ij} + (-1)^{i+j}2a. \quad (2.3)
$$

Thus, for some constant $b$ one has

$$
|A| = |A_{(i,j),(j,i)}|A_{ij}^2 + (-1)^{i+j}2aA_{ij} + b. \quad (2.4)
$$

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

$$
|(A^{-1})_{ij}| = \frac{|A_{(ij)}|}{|A|} = \frac{|A_{ij} + p|}{|A_{ij} + p|^2 + q} = \frac{|X|}{X^2 + q}, \quad \text{where } X = A_{ij} + p. \quad (2.5)
$$

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

$$
\mathbb{P}\{|(A^{-1})_{ij}| > t\} \leq \mathbb{P}\{|X| < 1/t\} \leq 2K/t. \quad (2.6)
$$


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 $2\varepsilon$ 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 $X_i$:

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

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 $A$:

$$\|A^{-1}\|_{\text{HS}} \leq \left\| \left( \sum_{1 \leq i,j \leq n} ((A^{-1})_{i,j})^2 \right)^{1/2} \right\|_{1,\infty} \leq 4 \sum_{1 \leq i,j \leq n} \|(A^{-1})_{i,j}\|_{1,\infty} \leq 8Kn^2.$$  

The definition of the weak $L_1$ norm then yields

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

Since $\|A^{-1}\| \leq \|A^{-1}\|_{\text{HS}}$, 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**


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