MATHEMATICS OF COMPUTATION
Volume 79, Number 272, October 2010, Pages 2333–2352
S 0025-5718(2010)02396-8
Article electronically published on June 4, 2010

SMOOTH ANALYSIS OF THE CONDITION NUMBER AND THE LEAST SINGULAR VALUE

TERENCE TAO AND VAN VU

ABSTRACT. Let x be a complex random variable with mean zero and bounded variance. Let N_n be the random matrix of size n whose entries are iid copies of x and let M be a fixed matrix of the same size. The goal of this paper is to give a general estimate for the condition number and least singular value of the matrix $M + N_n$, generalizing an earlier result of Spielman and Teng for the case when x is gaussian.

Our investigation reveals an interesting fact that the "core" matrix M does play a role on tail bounds for the least singular value of $M+N_n$. This does not occur in Spielman-Teng studies when x is gaussian. Consequently, our general estimate involves the norm $\|M\|$. In the special case when $\|M\|$ is relatively small, this estimate is nearly optimal and extends or refines existing results.

1. INTRODUCTION

Let M be an $n \times n$ matrix and $s_1(M) \geq \cdots \geq s_n(M)$ its singular values. The condition number of A, as defined by numerical analysts, is

$$\kappa(M) := s_1(M)/s_n(M) = \|M\|\|M^{-1}\|.$$

This parameter is of fundamental importance in numerical linear algebra and related areas, such as linear programming. In particular, the value

$$L(M) := \log \kappa(M)$$

measures the (worst case) loss of precision that the equation $Mx = b$ can exhibit [21, 2].

The problem of understanding the typical behavior of $\kappa(M)$ and $L(M)$ when the matrix M is random has a long history. This was first raised by von Neumann and Goldstine in their study of numerical inversion of large matrices [30]. Several years later, the problem was restated in a survey of Smale [21] on the efficiency of algorithm of anaylsis. One of Smale's motivations was to understand the efficiency of the simplex algorithm in linear programming. The problem is also at the core of Demmel's plan on the investigation of the probability that a numerical analysis problem is difficult [8] (see also [18] for a work that inspires this investigation).

To make the problem precise, the most critical issue is to choose a probability distribution for M. A convenient model has random matrices with independent

Received by the editor March 10, 2009.
2010 *Mathematics Subject Classification.* Primary 11B25.
The first author was supported by a grant from the MacArthur Foundation.
The second author was supported by research grants DMS-0901216 and AFOSAR-FA-9550-09-1-0167.

gaussian entries (either real of complex). An essential feature of this model is that here the joint distribution of the eigenvalues can be written precisely:

$$(1) \qquad (\textit{Real Gaussian})\ c_1(n) \prod_{1\le i<j\le n} |\lambda_i - \lambda_j| \exp(-\sum_{i=1}^{n} \lambda_i^2/2).$$

$$(2) \qquad (\textit{Complex Gaussian})\ c_2(n) \prod_{1\le i<j\le n} |\lambda_i - \lambda_j|^2 \exp(-\sum_{i=1}^{n} \lambda_i^2/2).$$

Here $c_1(n), c_2(n)$ are normalization factors whose explicit formulae can be seen in, for example, [16].

Most questions about the spectrum of these random matrices can then be answered by estimating a properly defined integral with respect to these measures. Many advanced techniques have been worked out to serve this purpose (see, for instance, [16]). In particular, the condition number is well understood, thanks to works of Kostlan and Oceanu [12, 21], Edelman [6] and many others (see Section 2).

The gaussian model, however, has serious shortcomings. As pointed out by many researchers (see, for example [3, 23]), the gaussian model does not reflect the arbitrariness of the input. Let us consider, for example, a random matrix with independent real gaussian entries. By sharp concentration results, one can show that the fraction of entries with absolute values at most 1, is, with overwhelming probability, close to the absolute constant $\frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \exp(-t^2/2)dt$. Many classes of matrices that occur in practice simply do not possess this property. This problem persists even when one replaces gaussian by another fixed distribution, such as Bernoulli.

About 10 years ago, Spielman and Teng [23, 24], motivated by Demmel's plan and the problem of understanding the efficiency of the simplex algorithm proposed a new and exciting distribution. Spielman and Teng observed that while the ideal input may be a fixed matrix M, it is likely that the computer will work with a perturbation $M+N$, where N is a random matrix representing random noise. Thus, it raised the issue of studying the distribution of the condition number of $M + N$. This problem is at the heart of the so-called Spielman-Teng *smooth* analysis. (See [23, 24] for a more detailed discussion and [3, 4, 5, 25, 9] for many related works on this topic.) Notice that the special case $M = 0$ corresponds to the setting considered in the previous paragraphs.

The Spielman-Teng model nicely addresses the problem about the arbitrariness of the inputs, as in this model every matrix generates a probability space of its own. In their papers, Spielman and Teng considered mostly gaussian noise (in some cases they also considered other continuous distributions such as uniformity on $[-1, 1]$). However, in the digital world, randomness often *does not* have gaussian nature. To start with, all of the real data are finite. In fact, in many problems (particularly those in integer programming) all entries of the matrix are integers. The random errors made by the digital devices (for example, sometimes a bit gets flipped) are obviously of discrete nature. In other problems, for example, those in engineering, the data may contain measurements where it would be natural to assume gaussian errors. On the other hand, data are usually strongly truncated. For example, if an entry of our matrix represents the mass of an object, then we expect to see a number such as 12.679 (say, tons), rather than 12.6792347043641259. Thus, instead

of the gaussian distribution, we (and/or our computers) often work with a discrete distribution, whose support is relatively small and does not depend on the size of the matrix. (A good example is random a Bernoulli matrix, whose entries take values ± 1 with probability of half.) This leads us to the following problem:

Problem (Smooth analysis of the condition number)**.** Estimate the condition number of a random matrix $M_n := M + N_n$, where M is a fixed matrix of size n, and N_n is a general random matrix.

The goal of this paper is to investigate this question, where, as a generalization of the Spielman-Teng model, we think of N_n as a matrix with independent random entries which (instead as being gaussian) have arbitrary distributions. Our main result will show that with high probability, M_n is well-conditioned. This result could be useful in further studies of smooth analysis in linear programming. The Spielman-Teng smooth analysis of the simplex algorithm [23, 24] was done using gaussian noise. It is a natural and (from the practical point of view) important to repeat this analysis using discrete noise (such as Bernoulli). This problem was posed by Spielman to the authors a few years ago. The paper [23] also contains a specific conjecture on the least singular value of a random Bernoulli matrix.

In connection, we should mention here a recent series of papers by Burgisser, Cucker and Lotz [3, 4, 5], which discussed the smooth analysis of condition number under a somewhat different setting (they considered the notion of *conic* condition number and a different kind of randomness).

Before stating mathematical results, let us describe our notation. We use the usual asymptotic notation $X = O(Y)$ to denote the estimate $|X| \leq CY$ for some constant $C > 0$ (independent of n), $X = \Omega(Y)$ to denote the estimate $X \geq cY$ for some $c > 0$ independent of n, and $X = \Theta(Y)$ to denote the estimates $X = O(Y)$ and $X = \Omega(Y)$ holding simultaneously. In some cases, we write $X \ll Y$ instead of $X = O(Y)$ and $X \gg Y$ instead of $X = \Omega(Y)$. Notation such as $X = O_{x,b}(Y)$ or $X \ll_{a,b} (Y)$ mean that the hidden constant in O or $\ll$ depend on previously defined constants a and b. We use $o(1)$ to denote any quantity that goes to zero as $n \to \infty$. $X = o(Y)$ means that $X/Y = o(1)$.

Recall that

$$\kappa(M) := s_1(M)/s_n(M) = \|M\|\|M^{-1}\|.$$

Since $\|M\|^2 \geq \sum_{ij} |m_{ij}|^2/n$ (where m_{ij} denote the entries of M) it is expected that $\|M\| = n^{\Omega(1)}$. Following the literature, we say that M is well-conditioned (or well-posed) if $\kappa(M) = n^{O(1)}$ or (equivalently) $L(M) = O(\log n)$.

By the triangle inequality, we get

$$\|M\| - \|N_n\| \leq \|M + N_n\| \leq \|M\| + \|N_n\|.$$

Under very general assumptions, the random matrix N_n satisfies $\|N_n\| = n^{O(1)}$ with overwhelming probability (see many estimates in Section 3). Thus, in order to guarantee that $\|M + N_n\|$ is well-conditioned (with high probability), it is natural to assume that

$$\|M\| = n^{O(1)}. \tag{3}$$

This is not only a natural, but fairly safe assumption to make (with respect to the applicability of our studies). Most large matrices in practice satisfy this assumption, as their entries are usually not very large compared to their sizes.

Our main result shows that under this assumption and a very general assumption on the entries of N_n, the matrix $M + N_n$ is well-conditioned, with high probability. This result extends and bridges several existing results in the literature (see the next two sections).

Notice that under assumption (3), if we want to show that $M + N_n$ is typically well-conditioned, it suffices to show that

$$\|(M + N_n)^{-1}\| = s_n(M + N_n)^{-1} = n^{O(1)}$$

with high probability. Thus, we will formulate most results in a form of a tail bound for the least singular value of $M + N_n$. The typical form will be

$$\mathbf{P}(s_n(M + N_n) \leq n^{-B}) \leq n^{-A}$$

where A, B are positive constants and A increases with B. The relation between A and B is of importance and will be discussed in length.

2. Previous results

Let us first discuss the gaussian case. Improving results of Kostlan and Oceanu [21], Edelman [6] computed the limiting distribution of $\sqrt{n}s_n(N_n)$ when N_n is gaussian. His result implies

Theorem 2.1. *There is a constant $C > 0$ such that the following holds. Let x be the real gaussian random variable with mean zero and variance one, let N_n be the random matrix whose entries are iid copies of x. Then for any constant $t > 0$,*

$$\mathbf{P}(s_n(N_n) \leq t) \leq n^{1/2}t.$$

Concerning the more general model $M + N_n$, Sankar, Teng and Spielman [25] proved:

Theorem 2.2. *There is a constant $C > 0$ such that the following holds. Let x be the real gaussian random variable with mean zero and variance one, let N_n be the random matrix whose entries are iid copies of x, and let M be an arbitrary fixed matrix. Let $M_n := M + N_n$. Then for any $t > 0$,*

$$\mathbf{P}(s_n(M_n) \leq t) \leq Cn^{1/2}t.$$

Once we give up the gaussian assumption, the study of the least singular value s_n becomes much harder (in particular, for discrete distributions such as Bernoulli, in which $x = \pm 1$ with equal probability $1/2$). For example, it is already non-trivial to prove that the least singular value of a random Bernoulli matrix is positive with probability $1 - o(1)$. This was first done by Komlós in 1967 [13], but good quantitative lower bounds were not available until recently. In a series of papers, Tao-Vu and Rudelson-Vershynin addressed this question [26, 28, 19, 20] and proved a lower bound of the form $n^{-\Theta(1)}$ for s_n with high probability.

We say that x is *sub-gaussian* if there is a constant $B > 0$ such that

$$\mathbf{P}(|x| \geq t) \leq 2\exp(-t^2/B^2)$$

for all $t > 0$. The smallest B is called the *sub-gaussian moment* of x. The following is a corollary of a more general theorem by Rudelson and Vershynin [20, Theorem 1.2]

Theorem 2.3. *Let x be a sub-gaussian random variable with zero mean, variance one and sub-gaussian moment B, and let A be an arbitrary positive constant. Let N_n be the random matrix whose entries are iid copies of x. Then there is a positive constant C (depending on B) such that for any $t \geq n^{-A}$ we have*

$$\mathbf{P}(s_n(N_n) \leq t) \leq Cn^{1/2}t.$$

We again turn to the general model $M + N_n$. In [28], the present authors proved

Theorem 2.4 ([28, Theorem 2.1])**.** *Let x be a random variable with non-zero variance. Then for any constants $A, C > 0$ there exists a constant $B > 0$ (depending on A, C, x) such that the following holds. Let N_n be the random matrix whose entries are iid copies of x, and let M be any deterministic $n \times n$ matrix with norm $\|M\| \leq n^C$. Then*

$$\mathbf{P}(s_n(M + N_n) \leq n^{-B}) \leq n^{-A}.$$

Notice that this theorem requires very little about the variable x. It does not need to be sub-gaussian nor does it have bounded moments. All we ask is that the variance is bounded from zero, which basically means x is indeed "random". Thus, it guarantees the well-conditionness of $M + N_n$ in a very general setting.

The weakness of this theorem is that the dependence of B on A and C, while explicit, is too generous. The main result of this paper, Theorem 3.2, will improve this dependence significantly and provide a common extension of Theorem 2.4 and Theorem 2.3.

3. Main result

As already pointed out, an important point is the relation between the constants A, B in a bound of the form

$$\mathbf{P}(s_n(M + N_n) \leq n^{-B}) \leq n^{-A}.$$

In Theorem 2.2, we have a simple (and optimal) relation $B = A + 1/2$. It is natural to conjecture that this relation holds for other, non-gaussian, models of random matrices. In fact, this conjecture was our starting point for this study. Quite surprisingly, it turns out not to be the case.

Theorem 3.1. *There are positive constants c_1 and c_2 such that the following holds. Let N_n be the $n \times n$ random Bernoulli matrix with n even. For any $L \geq n$, there is an $n \times n$ deterministic matrix M such that $\|M\| = L$ and*

$$\mathbf{P}(s_n(M + N_n) \leq c_1 \frac{n}{L}) \geq c_2 n^{-1/2}.$$

The assumption n is even is for convenience and can easily be removed by replacing the Bernoulli matrix by a random matrix whose entries take values $0, \pm 1$ with probability $1/3$, say. Notice that if $L = n^D$ for some constant D, then we have the lower bound

$$\mathbf{P}(s_n(M + N_n) \leq c_1 n^{-D+1}) \geq c_2 n^{-1/2},$$

which shows that one cannot expect Theorem 2.2 to hold in general and that the norm of M should play a role in tail bounds of the least singular value.

The main result of this paper is the following.

Theorem 3.2. *Let x be a random variable with mean zero and bounded second moment, and let $\gamma \geq 1/2$, $A \geq 0$ be constants. Then there is a constant c depending on x, γ, A such that the following holds. Let N_n be the random matrix of size n whose entries are iid copies of x, let M be a deterministic matrix satisfying $\|M\| \leq n^\gamma$, and let $M_n := M + N_n$. Then*

$$\mathbf{P}(s_n(M_n) \leq n^{-(2A+1)\gamma}) \leq c\Big(n^{-A+o(1)} + \mathbf{P}(\|N_n\| \geq n^\gamma)\Big).$$

Note that this theorem only assumes bounded second moment on x. The assumption that the entries of N_n are iid is for convenience. A slightly weaker result would hold if one omits this assumption.

Corollary 3.3. *Let x be a random variable with mean zero and bounded second moment, and let $\gamma \geq 1/2$, $A \geq 0$ be constants. Then there is a constant c_2 depending on x, γ, A such that the following holds. Let N_n be the random matrix of size n whose entries are iid copies of x, let M be a deterministic matrix satisfying $\|M\| \leq n^\gamma$, and let $M_n := M + N_n$. Then*

$$\mathbf{P}(\kappa(M_n) \geq 2n^{(2A+2)\gamma}) \leq c\Big(n^{-A+o(1)} + \mathbf{P}(\|N_n\| \geq n^\gamma)\Big).$$

Proof. Since $\kappa(M_n) = s_1(M_n)/s_n(M_n)$, it follows that if $\kappa(M_n) \geq n^{(2A+2)\gamma}$, then at least one of the two events $s_n(M_n) \leq n^{-(2A+1)\gamma}$ and $s_1(M_n) \geq 2n^\gamma$ holds. On the other hand,

$$s_1(M_n) \leq s_1(M) + s_1(N_n) = \|M\| + \|N_n\| \leq n^\gamma + \|N_n\|.$$

The claim follows. □

In the rest of this section, we deduce a few corollaries and connect them with the existing results.

First, consider the special case when x is sub-gaussian. In this case, it is well-known that one can have a strong bound on $\mathbf{P}(\|N_n\| \geq n^\gamma)$ thanks to the following theorem (see [20] for references)

Theorem 3.4. *Let B be a positive constant. There are positive constants C_1, C_2 depending on B such that the following holds. Let x be a sub-gaussian random variable with zero mean, variance one and sub-gaussian moment B and let N_n be the random matrix whose entries are iid copies of x. Then*

$$\mathbf{P}(\|N_n\| \geq C_1 n^{1/2}) \leq \exp(-C_2 n).$$

If one replaces the sub-gaussian condition by the weaker condition that x has fourth moment bounded B, then one has a weaker conclusion that

$$\mathbf{E}(\|N_n\|) \leq C_1 n^{1/2}.$$

From Theorem 3.2 and Theorem 3.4 we see that

Corollary 3.5. *Let A and γ be arbitrary positive constants. Let x be a sub-gaussian random variable with zero mean and variance one and let N_n be the random matrix whose entries are iid copies of x. Let M be a deterministic matrix such that $\|M\| \leq n^\gamma$ and set $M_n = M + N_n$. Then*

$$\mathbf{P}(s_n(M_n) \leq (n^{1/2} + \|M\|)^{-2A-1}) \leq n^{-A+o(1)}. \tag{4}$$

In the case $\|M_n\| = O(n^{1/2})$ (which of course includes the $M_n = 0$ special case), (4) implies

Corollary 3.6. *Let A be an arbitrary positive constant. Let x be a sub-gaussian random variable with zero mean and variance one and let N_n be the random matrix whose entries are iid copies of x. Let M be a deterministic matrix such that $\|M\| = O(n^{1/2})$ and set $M_n = M + N_n$. Then*

$$\mathbf{P}(s_n(M_n) \leq n^{-A-1/2}) \leq n^{-A+o(1)}. \tag{5}$$

Up to a loss of magnitude $n^{o(1)}$, this matches Theorem 2.3, which treated the base case $M = 0$.

If we assume bounded fourth moment instead of sub-gaussian, we can use the second half of Theorem 3.4 to deduce

Corollary 3.7. *Let x be a random variable with zero mean, variance one and bounded fourth moment, and let N_n be the random matrix whose entries are iid copies of x. Let M be a deterministic matrix such that $\|M\| = n^{O(1)}$ and set $M_n = M + N_n$. Then*

$$\mathbf{P}(s_n(M_n) \leq (n^{1/2} + \|M\|)^{-1+o(1)}) = o(1). \tag{6}$$

In the case $\|M\| = O(n^{1/2})$, this implies that almost surely $s_n(M_n) \geq n^{-1/2+o(1)}$. For the special case $M = 0$, this matches (again up to the $o(1)$ term) Theorem [20, Theorem 1.1].

Let us now take a look at the influence of $\|M\|$ on the bound. Obviously, there is a gap between (4) and Theorem 3.1. On the other hand, by setting $A = 1/2$, $L = n^{\gamma}$ and assuming that $\mathbf{P}(\|N_n\| \geq n^{\gamma})$ is negligible (i.e., super-polynomially small in n), we can deduce from Theorem 3.2 that

$$\mathbf{P}(s_n(M_n) \leq c_1 L^{-2}) \leq c_2 n^{-1/2+o(1)}.$$

This, together with Theorem 3.1, suggests that the influence of $\|M\|$ in $s_n(M_n)$ is of polynomial type.

In the next discussion, let us normalize and assume that x has variance one. One can deduce a bound on $\|N_n\|$ from the simple computation

$$\mathbf{E}\|N_n\|^2 \leq \mathbf{E}\ \mathrm{tr} N_n N_n^* = n^2.$$

By Chebyshev's inequality we thus have

$$\mathbf{P}(\|N_n\| \geq n^{1+A/2}) \leq n^{-A}$$

for all $A \geq 0$.

Applying Theorem 3.2 we obtain

Corollary 3.8. *Let x be a random variable with mean zero and variance one and let N_n be the random matrix whose entries are iid copies of x. Then for any constant $A \geq 0$ we get*

$$\mathbf{P}(s_n(N_n) \leq n^{-1-\frac{5}{2}A-A^2}) \leq n^{-A+o(1)}.$$

In particular, $s_n(N_n) \geq n^{-1-o(1)}$ almost surely.

It is clear that one can obtain better bounds for s_n, provided we have better estimates on $\|N_n\|$. The idea of using Chebyshev's inequality is very crude (we just want to give an example) and there are more sophisticated tools. One can, for instance, use higher moments. The expectation of a kth moment can be expressed in a sum of many terms, each corresponding to a certain closed walk of length k on the complete graph of n vertices (see [11, 31]). If the higher moments of N_n (while not bounded) do not increase too fast with n, then the main contribution

in the expectation of the kth moment still come from terms which correspond to walks using each edge of the graph either 0 or 2 times. The expectation of such a term involves only the second moment of the entries in N_n. The reader may want to work this out as an exercise.

One can also use the following nice estimate of Seginer [22]

$$\mathbf{E}\|N_n\| = O(\mathbf{E}\max_{1\le i\le n}\sqrt{\sum_{j=1}^{n} x_{ij}^2} + \mathbf{E}\max_{1\le j\le n}\sqrt{\sum_{i=1}^{n} x_{ij}^2}).$$

The rest of the paper is organized as follows. In the next section, we prove Theorem 3.1. The remaining sections are devoted to the proof of Theorem 3.2. This proof combines several tools that have been developed in recent years. It starts with an ϵ-net argument (in the spirit of those used in [26, 19, 28, 20]. Two important technical ingredients are Theorem 6.8 from [28] and Lemma 9.1 from [20].

4. Theorem 3.1: The influence of M

Let M' be the $n-1\times n$ matrix obtained by concatenating the matrix LI_{n-1} with an all L column, where L is a large number (we will set $L\ge n$). The $n\times n$ matrix M is obtained from M' by adding to it a (first) all zero row; thus,

$$M = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ L & 0 & \dots & 0 & L \\ 0 & L & \dots & 0 & L \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & L & L \end{pmatrix}.$$

It is easy to see that

$$\|M\| = \Theta(L).$$

Now consider $M_n := M + N_n$ where the entries of N_n are iid Bernoulli random variables

$$\mathbf{P}(s_n(M_n) \ll n^{1/4}L^{-1/2}) \gg n^{-1/2}.$$

Let M_n' be the (random) $(n-1)\times n$ matrix formed by the last $n-1$ rows of M_n. Let $v\in\mathbf{R}^n$ be a unit normal vector of the $n-1$ rows of M_n'. By replacing v with $-v$ if necessary we may write v in the form

$$v = \left(\frac{1}{\sqrt{n}} + a_1, \frac{1}{\sqrt{n}} + a_2, \dots, \frac{1}{\sqrt{n}} + a_{n-1}, \frac{-1}{\sqrt{n}} + a_n\right),$$

where $\frac{-1}{\sqrt{n}} + a_n \le 0$.

Let ξ_i be iid Bernoulli random variables. Multiplying v with the first row of M_n', we have

$$\begin{aligned} 0 &= (L+\xi_1)(\frac{1}{\sqrt{n}} + a_1) + (L+\xi_n)(-\frac{1}{\sqrt{n}} + a_n) \\ &= L(a_1+a_n) + \frac{1}{\sqrt{n}}\Big((\xi_1-\xi_n) + \xi_1 a_1 + \xi_n a_n\Big). \end{aligned}$$

Since $|a_i| = O(1)$, it follows that $|a_1+a_n| = O(\frac{1}{L})$. Repeating the argument with all other rows, we conclude that $|a_i+a_n| = O(\frac{1}{L})$ for all $1\le i\le n-1$.

Since v has unit norm, we also have

$$1 = \|v\|^2 = \sum_{i=1}^{n-1} \left(\frac{1}{\sqrt{n}} + a_i\right)^2 + \left(\frac{-1}{\sqrt{n}} + a_n\right)^2,$$

which implies that

$$\frac{2}{\sqrt{n}}(a_1 + \cdots + a_{n-1} - a_n) + \sum_{i=1}^{n} a_i^2 = 0.$$

This, together with the fact that $|a_i + a_n| = O(\frac{1}{L})$ and all $1 \le i \le n-1$, yields

$$na_n^2 - 2na_n(\frac{1}{\sqrt{n}} + \frac{1}{L}) = O(\frac{\sqrt{n}}{L} + \frac{1}{L^2}).$$

Since $-\frac{1}{\sqrt{n}} + a_n \le 0$ and $L \ge n$, it is easy to show from here that $|a_n| = O(\frac{1}{L})$. It follows that $|a_i| = O(\frac{1}{L})$ for all $1 \le i \le n$.

Now consider

$$\|M_n v\| = \left|\sum_{i=1}^{n-1}(\frac{1}{\sqrt{n}} + a_i)\xi_i + (-\frac{1}{\sqrt{n}} + a_n)\xi_n\right|.$$

Since n is even, with probability $\Theta(\frac{1}{\sqrt{n}})$, $\xi_1 + \cdots + \xi_{n-1} - \xi_n = 0$, and in this case

$$\|M_n v\| = \left|\sum_{i=1}^{n} a_i \xi_i\right| = O\left(\frac{n}{L}\right),$$

as desired.

5. Controlled moment

It is convenient to establish some more quantitative control on x. We recall the following notion from [28].

Definition 5.1 (Controlled second moment)**.** Let $\kappa \ge 1$. A complex random variable x is said to have *κ-controlled second moment* if one has the upper bound

$$\mathbf{E}|x|^2 \le \kappa$$

(in particular, $|\mathbf{E}x| \le \kappa^{1/2}$), and the lower bound

$$\mathbf{E}\mathrm{Re}(zx - w)^2 \mathbf{I}(|x| \le \kappa) \ge \frac{1}{\kappa}\mathrm{Re}(z)^2 \tag{7}$$

for all complex numbers z, w.

Example. The Bernoulli random variable ($\mathbf{P}(x = +1) = \mathbf{P}(x = -1) = 1/2$) has 1-controlled second moment. The condition (7) asserts, in particular, that x has variance at least $\frac{1}{\kappa}$, but also asserts that a significant portion of this variance occurs inside the event $|x| \le \kappa$, and also contains some more technical phase information about the covariance matrix of $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$.

The following lemma was established in [28]:

Lemma 5.2 ([28, Lemma 2.4])**.** *Let x be a complex random variable with finite non-zero variance. Then there exists a phase $e^{i\theta}$ and a $\kappa \ge 1$ such that $e^{i\theta}x$ has κ-controlled second moment.*

Since rotation by a phase does not affect the conclusion of Theorem 3.2, we conclude that we can assume, without loss of generality, that x is κ-controlled for some κ. This will allow us to invoke several estimates from [28] (e.g. Lemma 6.2 and Theorem 6.8 below).

Remark 5.3. The estimates we obtain for Theorem 3.2 will depend on κ but will not otherwise depend on the precise distribution of x. It is, in fact, quite likely that the results in this paper can be generalized to random matrices N_n whose entries are independent and are all κ-controlled for a single κ, but do not need to be identical. In order to simplify the exposition, however, we focus on the iid case.

6. Small ball bounds

In this section we give some bounds on the small ball probabilities $\mathbf{P}(|\xi_1 v_1 + \cdots + \xi_n v_n - z| \leq \varepsilon)$ under various assumptions on the random variables ξ_i and the coefficients v_i. As a consequence we shall be able to obtain good bounds on the probability that Av is small, where A is a random matrix and v is a fixed unit vector.

We first recall a standard bound (cf. [28, Lemmas 4.2, 4.3, 5.2]):

Lemma 6.1 (Fourier-analytic bound)**.** *Let $\xi_1, \ldots, \xi_n$ be independent variables. Then we have the bound*

$$\mathbf{P}(|\xi_1 v_1 + \cdots + \xi_n v_n - z| \leq r) \ll r^2 \int_{w \in \mathbf{C}: |w| \leq 1/r} \exp(-\Theta(\sum_{j=1}^n \|w v_j\|_j^2))\, dw$$

for any $r > 0$ and $z \in \mathbf{C}$, and any unit vector $v = (v_1, \ldots, v_n)$, where

$$\|z\|_j := (\mathbf{E}\|\mathrm{Re}(z(\xi_j - \xi_j'))\|_{\mathbf{R}/\mathbf{Z}}^2)^{1/2}, \tag{8}$$

ξ_j' is an independent copy of ξ_j, and $\|x\|_{\mathbf{R}/\mathbf{Z}}$ denotes the distance from x to the nearest integer.

Proof. By the Esséen concentration inequality (see e.g. [29, Lemma 7.17]), we have

$$\mathbf{P}(|\xi_1 v_1 + \cdots + \xi_n v_n - z| \leq r) \ll r^2 \int_{w \in \mathbf{C}: |w| \leq 1/r} |\mathbf{E}(e(\mathrm{Re}(w(\xi_1 v_1 + \cdots + \xi_n v_n))))|\, dw$$

for any $c > 0$, where $e(x) := e^{2\pi i x}$. We can write the right-hand side as

$$r^2 \int_{w \in \mathbf{C}: |w| \leq 1/r} \prod_{j=1}^n f_j(w v_j)^{1/2}\, dw$$

where

$$f_j(z) := |\mathbf{E}(e(\mathrm{Re}(\xi_j z)))|^2 = \mathbf{E}\cos(2\pi \mathrm{Re}(z(\xi_j - \xi_j'))).$$

Using the elementary bound $\cos(2\pi\theta) \leq 1 - \Theta(\|\theta\|_{\mathbf{R}/\mathbf{Z}}^2)$ we conclude that

$$f_j(z) \leq 1 - \Theta(\|z\|_j^2) \leq \exp(-\Theta(\|z\|_j^2))$$

and the claim follows. □

Next, we recall some properties of the norms $\|z\|_j$ in the case when ξ_j is κ-controlled.

Lemma 6.2. *Let $1 \leq j \leq n$, let ξ_j be a random variable, and let $\|\|_j$ be defined by* (8)*.*

(i) *For any $w \in \mathbf{C}$, $0 \leq \|w\|_j \leq 1$ and $\| - w\|_j = \|w\|_j$.*

(ii) *For any* $z, w \in \mathbf{C}$, $\|z+w\|_j \leq \|z\|_j + \|w\|_j$.
(iii) *If* ξ_j *is* κ*-controlled for some fixed* κ*, then for any sufficiently small positive constants* $c_0, c_1 > 0$ *we have* $\|z\|_j \geq c_1 \mathrm{Re}(z)$ *whenever* $|z| \leq c_0$.

Proof. See [28, Lemma 5.3]. □

We now use these bounds to estimate small ball probabilities. We begin with a crude bound.

Corollary 6.3. *Let* $\xi_1, \ldots, \xi_n$ *be independent variables which are* κ*-controlled. Then there exists a constant* $c > 0$ *such that*

$$\mathbf{P}(|\xi_1 v_1 + \cdots + \xi_n v_n - z| \leq c) \leq 1 - c \tag{9}$$

for all $z \in \mathbf{C}$ *and all unit vectors* $(v_1, \ldots, v_n)$.

Proof. Let $c > 0$ be a small number to be chosen later. We divide it into two cases, depending on whether all the v_i are bounded in magnitude by $\sqrt{c}$ or not.

Suppose first that $|v_i| \leq \sqrt{c}$ for all c. Then we apply Lemma 6.1 (with $r := c^{1/4}$) and bound the left-hand side of (9) by

$$\ll c^{1/2} \int_{w \in \mathbf{C}: |w| \leq c^{-1/4}} \exp(-\Theta(\sum_{j=1}^n \|w v_j\|_j^2))\, dw.$$

By Lemma 6.2, if c is sufficiently small, then we have $\|w v_j\|_j \geq c_1 \mathrm{Re}(w v_j)$, for some positive constant c_1. Writing each v_j in polar coordinates as $v_j = r_j e^{2\pi i \theta_j}$, we thus obtain an upper bound of

$$\ll c^{1/2} \int_{w \in \mathbf{C}: |w| \leq c^{-1/4}} \exp(-\Theta(\sum_{j=1}^n r_j^2 \mathrm{Re}(e^{2\pi i \theta_j} w)^2))\, dw.$$

Since $\sum_{j=1}^n r_j^2 = 1$, we can use Hölder's inequality (or Jensen's inequality) and bound this from above by

$$\ll \sup_j c^{1/2} \int_{w \in \mathbf{C}: |w| \leq c^{-1/4}} \exp(-\Theta(\mathrm{Re}(e^{2\pi i \theta_j} w)^2))\, dw$$

which by rotation invariance and scaling is equal to

$$\int_{w \in \mathbf{C}: |w| \leq 1} \exp(-\Theta(c^{-1/4} \mathrm{Re}(w)^2))\, dw.$$

From the monotone convergence theorem (or direct computation) we see that this quantity is less than $1 - c$ if c is chosen sufficiently small. (If necessary, we allow c to depend on the hidden constant in Θ.)

Now suppose instead that $|v_1| > \sqrt{c}$, say. Then by freezing all of the variables $\xi_2, \ldots, \xi_n$, we can bound the left-hand side of (9) by

$$\sup_w \mathbf{P}(|\xi_1 - w| \leq \sqrt{c}).$$

But by the definition of κ-control, one easily sees that this quantity is bounded by $1 - c$ if c is sufficiently small (compared to $1/\kappa$), and the claim follows. □

As a consequence of this bound, we obtain

Theorem 6.4. *Let N_n be an $n \times n$ random matrix whose entries are independent random variables which are all κ-controlled for some constant $\kappa > 0$. Then there are positive constants c, c' such that the following holds. For any unit vector v and any deterministic matrix M,*

$$\mathbf{P}(\|(M + N_n)v\| \leq cn^{1/2}) \leq \exp(-c'n).$$

Proof. Let c be a sufficiently small constant, and let $X_1, \ldots, X_n$ denote the rows of $M + N_n$. If $\|(M + N_n)v\| \leq cn^{1/2}$, then we have $|\langle X_j, v\rangle| \leq c$ for at least $(1 - c)n$ rows. As the events $\mathbf{I}_j := |\langle X_j, v\rangle| \leq c$ are independent, we see from the Chernoff inequality (applied to the sum $\sum_j \mathbf{I}_j$ of indicator variables) that it suffices to show that

$$\mathbf{E}(I_j) = \mathbf{P}(|\langle X_j, v\rangle| \leq c) \leq 1 - 2c,$$

say, for all j. But this follows from Corollary 6.3 (after adjusting c slightly), noting that each X_j is a translate (by a row of M) of a vector whose entries are iid copies of x. □

Now we obtain some statements of inverse Littlewood-Offord type.

Definition 6.5 (Compressible and incompressible vectors)**.** For any $a, b > 0$, let $\mathrm{Comp}(a, b)$ be the set of unit vectors v such that there is a vector v' with at most an non-zero coordinates satisfying $\|v - v'\| \leq b$. We denote by $\mathrm{Incomp}(a, b)$ the set of unit vectors which do not lie in $\mathrm{Comp}(a, b)$.

Definition 6.6 (Rich vectors)**.** For any $\varepsilon, \rho > 0$, let $S_{\varepsilon,\rho}$ be the set of unit vectors v satisfying

$$\sup_{z \in \mathbf{C}} \mathbf{P}(|X \cdot v - z| \leq \varepsilon) \geq \rho,$$

where $X = (x_1, \ldots, x_n)$ is a vector whose coefficients are iid copies of x.

Lemma 6.7 (Very rich vectors are compressible)**.** *For any $\varepsilon, \rho > 0$ we have*

$$S_{\varepsilon,\rho} \subset \mathrm{Comp}\left(O(\frac{1}{n\rho^2}), O(\frac{\varepsilon}{\rho})\right).$$

Proof. We can assume $\rho \gg n^{-1/2}$ since the claim is trivial otherwise. Let $v \in S_{\varepsilon,\rho}$, thus

$$\mathbf{P}(|X \cdot v - z| \leq \varepsilon) \geq \rho$$

for some z. From Lemma 6.1 we conclude that

$$\varepsilon^2 \int_{w \in \mathbf{C}: |w| \leq \varepsilon^{-1}} \exp(-\Theta(\sum_{j=1}^{n} \|wv_j\|_j^2))\, dw \gg \rho. \tag{10}$$

Let $s > 0$ be a small constant (independent of n) to be chosen later, and let A denote the set of indices i for which $|v_i| \geq s\varepsilon$. Then from (10) we have

$$\varepsilon^2 \int_{w \in \mathbf{C}: |w| \leq \varepsilon^{-1}} \exp(-\Theta(\sum_{j \in A} \|wv_j\|_j^2))\, dw \gg \rho.$$

Suppose A is non-empty. Applying Hölder's inequality, we conclude that

$$\varepsilon^2 \int_{w \in \mathbf{C}: |w| \leq \varepsilon^{-1}} \exp(-\Theta(|A|\|wv_j\|_j^2))\, dw \gg \rho$$

for some $j \in A$. By the pigeonhole principle, this implies that

$$|\{w \in \mathbf{C} : |w| \leq \varepsilon^{-1}, |A|\|wv_j\|_j^2 \leq k\}| \gg k^{1/2}\varepsilon^{-2}\rho \tag{11}$$

for some integer $k \geq 1$.

If $|A| \ll k$, then the set in (11) has measure $\Theta(\varepsilon^{-2})$, which forces $|A| \ll \rho^{-2}$. Suppose instead that $k \leq s|A|$ for some small $s' > 0$. Since $|v_j| \geq s\epsilon$, we have $s'/|v_j| \leq s'/s\epsilon$. We will choose s' sufficiently small to make sure that this ratio is smaller than the constant c_0 in Lemma 6.2. By Lemma 6.2, we see that the intersection of the set in (11) with any ball of radius $s'/|v_j|$ has density at most $\sqrt{k/|A|}$, and so by covering arguments we can bound the left-hand side of (11) from above by $\ll k^{1/2}|A|^{-1/2}\varepsilon^{-2}$. Thus we have $|A| \ll \rho^{-2}$ in this case also. Thus we have shown, in fact, that $|A| \ll \rho^{-2}$ in all cases (the case when A is empty being trivial).

Now we consider the contribution of those j outside of A. From (10) and Lemma 6.2 we have

$$\varepsilon^2 \int_{w \in \mathbf{C}: |w| \leq \varepsilon^{-1}} \exp(-\Theta(\sum_{j \notin A} \mathrm{Re}(w v_j)^2))\ dw \gg \rho.$$

Suppose that A is not all of $\{1, \ldots, n\}$. Using polar coordinates $v_j = r_j e^{2\pi i \theta_j}$ as before, we see from Hölder's inequality that

$$\varepsilon^2 \int_{w \in \mathbf{C}: |w| \leq \varepsilon^{-1}} \exp(-\Theta(r^2 \mathrm{Re}(w e^{2\pi i \theta_j})^2))\ dw \gg \rho$$

for some $j \notin A$, where $r^2 := \sum_{j \notin A} r_j^2$. After scaling and rotation invariance, we conclude that

$$\int_{w \in \mathbf{C}: |w| \leq 1} \exp(-\Theta(\frac{r^2}{\varepsilon^2} \mathrm{Re}(w)^2))\ dw \gg \rho.$$

The left-hand side can be computed to be at most $O(\varepsilon/r)$. We conclude that $r \ll \varepsilon/\rho$. If we let v' be the restriction of v to A, we thus have $\|v - v'\| \ll \varepsilon/\rho$, and the claim $v \in \mathrm{Comp}(O(\frac{1}{n\rho^2}), O(\frac{\varepsilon}{\rho}))$ follows. (The case when $A = \{1, \ldots, n\}$ is of course trivial.) □

Roughly speaking, Lemma 6.7 gives a complete characterization of vectors v such that

$$\sup_{z \in \mathbf{C}} \mathbf{P}(|X \cdot v - z| \leq \varepsilon) \geq \rho,$$

where $\rho > Cn^{-1/2}$, for some large constant C. The lemma shows that such a vector v can be approximated by a vector v' with at most $\frac{C'}{\rho^2}$ non-zero coordinates such that $\|v - v'\| \leq \frac{C''\epsilon}{\rho}$, where C', C'' are positive constants.

The dependence of parameters here are sharp, up to constant terms. Indeed, in the Bernoulli case, the vector $v = (1, \ldots, 1, 0, \ldots, 0)$ consisting of k 1's lies in $S_{0,\Theta(1/\sqrt{k})}$ and lies in $\mathrm{Comp}(a, 0)$ precisely when $an \geq k$ (cf. [7]). This shows that the $O(\frac{1}{n\rho^2})$ term on the right-hand side cannot be improved. On the other hand, in the Gaussian case, observe that if $\|v\| \leq b$, then $X \cdot v$ will have magnitude $O(\varepsilon)$ with probability $O(\varepsilon/b)$, which shows that the term $O(\frac{\varepsilon}{\rho})$ cannot be improved.

Lemma 6.7 is only non-trivial in the case $\rho \geq Cn^{-1/2}$, for some large constant C. To handle the case of smaller ρ, we use the following more difficult entropy bound from [28].

Theorem 6.8 (Entropy of rich vectors)**.** *For any ε, ρ, there is a finite set $S'_{\varepsilon,\rho}$ of size at most $n^{-(1/2-o(1))n}\rho^{-n} + \exp(o(n))$ such that for each $v \in S_{\varepsilon,\rho}$, there is $v' \in S'_{\varepsilon,\rho}$ such that $\|v - v'\|_\infty \leq \varepsilon$.*

Proof. See [28, Theorem 3.2]. □

7. Proof of Theorem 3.2: Preliminary reductions

We now begin the proof of Theorem 3.2. Let N_n, M, γ, A be as in that theorem. As remarked in Section 5, we may assume x to be κ-controlled for some κ. We allow all implied constants to depend on κ, γ, A. We may of course assume that n is large compared to these parameters. We may also assume that

$$\mathbf{P}(\|N_n\| \geq n^\gamma) \leq \frac{1}{2} \tag{12}$$

since the claim is trivial otherwise. By decreasing A if necessary, we may further assume that

$$\mathbf{P}(\|N_n\| \geq n^\gamma) \leq n^{-A+o(1)}. \tag{13}$$

It will then suffice to show (assuming (12), (13)) that

$$\mathbf{P}(s_n(M_n) \leq n^{-(2A+1)\gamma}) \ll n^{-A+\alpha+o(1)}$$

for any constant $\alpha > 0$ (with the implied constants now depending on α also), since the claim then follows by sending α to zero very slowly in n.

Fix α, and allow all implied constants to depend on α. By perturbing A and α slightly we may assume that A is not a half-integer; we can also take α to be small depending on A. For example, we can assume that

$$\alpha < \{2A\}/2 \tag{14}$$

where $\{2A\}$ is the fractional part of $2A$.

Using the trivial bound $\|N_n\| \geq \sup_{1\leq i,j\leq n} |x_{ij}|$, we conclude from (12) and (13) that

$$\mathbf{P}(|x_{ij}| \geq n^\gamma \text{ for some } i,j) \leq \min(\frac{1}{2}, n^{-A+o(1)}).$$

Since x_{ij} are iid copies of x, the n^2 events $|x_{ij}| \geq n^\gamma$ are independent with identical probability. It follows that

$$\mathbf{P}(|x| \geq n^\gamma) \leq n^{-A-2+o(1)}. \tag{15}$$

Let F be the event that $s_n(M_n) \leq n^{-(2A+1)\gamma}$, and let G be the event that $\|N_n\| \leq n^\gamma$. In view of (13), it suffices to show that

$$\mathbf{P}(F \wedge G) \leq n^{-A+\alpha+o(1)}.$$

Set

$$b := \beta n^{1/2-\gamma} \tag{16}$$

and

$$a := \frac{\beta}{\log n}, \tag{17}$$

where β is a small positive constant to be chosen later. We then introduce the following events:

- F_{Comp} is the event that $\|M_n v\| \leq n^{-(2A+1)\gamma}$ for some $v \in \mathrm{Comp}(a,b)$.
- F_{Incomp} is the event that $\|M_n v\| \leq n^{-(2A+1)\gamma}$ for some $v \in \mathrm{Incomp}(a,b)$.

Observe that if F holds, then at least one of F_{Comp} and F_{Incomp} holds. Theorem 3.2 then follows immediately from the following two lemmas.

Lemma 7.1 (Compressible vector bound)**.** *If β is sufficiently small, then*

$$\mathbf{P}(F_{\mathrm{Comp}} \wedge G) \leq \exp(-\Omega(n)).$$

Lemma 7.2 (Incompressible vector bound)**.** *We have*

$$\mathbf{P}(F_{\mathrm{Incomp}} \wedge G) \leq n^{-A+o(1)}.$$

In these lemmas we allow the implied constants to depend on β.

The proof of Lemma 7.1 is simple and will be presented in the next section. The proof of Lemma 7.2 is somewhat more involved and occupies the rest of the paper.

8. Treatment of compressible vectors

If $F_{\mathrm{Comp}} \wedge G$ occurs, then by the definition of $\mathrm{Comp}(a,b)$, there are unit vectors v, v' such that $\|M_n v\| \leq n^{-(2A+1)\gamma}$ and v' has support on at most an coordinates and $\|v - v'\| \leq b$.

By the triangle inequality and (16) we have

$$\begin{aligned}\|M_n v'\| &\leq n^{-(2A+1)\gamma} + \|M_n\|\|v - v'\| \\ &\leq n^{-(2A+1)\gamma} + n^{\gamma} b \\ &\leq 2\beta n^{1/2}.\end{aligned}$$

A set $\mathcal{N}$ of unit vectors in $\mathbf{C}^m$ is called a *δ-net* if for any unit vector v, there is a vector w in $\mathcal{N}$ such that $\|v - w\| \leq \delta$. It is well known that for any $0 < \delta < 1$, a δ-net of size $(C\delta^{-1})^m$ exists, for some constant C independent of δ and m.

Using this fact, we conclude that the set of unit vectors with at most an non-zero coordinates admits a b-net $\mathcal{N}$ of size at most

$$|\mathcal{N}| \leq \binom{n}{an} (Cb^{-1})^{an},$$

Thus, if $F_{\mathrm{Comp}} \wedge G$ occurs, then there is a unit vector $v'' \in \mathcal{N}$ such that

$$\|M_n v''\| \leq 2\beta n^{1/2} + \|M_n\| b = 3\beta n^{1/2}.$$

On the other hand, from Theorem 6.4 we see (for $\beta \leq c/3$) that for any fixed v'',

$$\mathbf{P}(\|M_n v''\| \leq 3\beta n^{1/2}) \leq \exp(-c'n),$$

where c and c' are the constants in Theorem 6.4.

By the union bound, we conclude that

$$\mathbf{P}(F_{\mathrm{Comp}} \wedge G) \leq \binom{n}{an} (b^{-1})^{an} \exp(-c'n).$$

But from (16), (17) we see that the right-hand side can be made less than $\exp(-c'n/2)$, given that β is sufficiently small. This concludes the proof of Lemma 7.1.

9. Treatment of incompressible vectors

We now begin the proof of Lemma 7.2. We now fix β and allow all implied constants to depend on β.

Let X_k be the kth row vector of M_n, and let dist_k be the distance from X_k to the subspace spanned by $X_1, \ldots, X_{k-1}, X_{k+1}, \ldots, X_n$. We need the following, which is a slight extension of a lemma from [20].

Lemma 9.1. *For any $\varepsilon > 0$, and any event E, we have*

$$\mathbf{P}(\{\|Mv\| \leq \varepsilon b n^{-1/2} \text{ for some } v \in \mathrm{Incomp}(a,b)\} \wedge E) \leq \frac{1}{an} \sum_{k=1}^{n} \mathbf{P}(\{\mathrm{dist}_k \leq \varepsilon\} \wedge E).$$

Proof. See [20, Lemma 3.5]. The arbitrary event E was not present in that lemma, but one easily verifies that the proof works perfectly well with this event in place. □

Applying this to our current situation with

$$\varepsilon := \frac{1}{\beta} n^{-2A\gamma}, \tag{18}$$

we obtain

$$\mathbf{P}(F_{\mathrm{Incomp}} \wedge G) \ll \frac{\log n}{n} \sum_{k=1}^{n} \mathbf{P}(\{\mathrm{dist}_k \leq \varepsilon\} \wedge G).$$

To prove Lemma 7.2, it therefore suffices (by symmetry) to show that

$$\mathbf{P}(\{\mathrm{dist}_n \leq \varepsilon\} \wedge G) \ll n^{-A+\alpha+o(1)}.$$

Notice that there is a unit vector X_n^* orthogonal to $X_1, \ldots, X_{n-1}$ such that

$$\mathrm{dist}_k = |X_n \cdot X_n^*|. \tag{19}$$

If there are many such X_n^*, choose one arbitrarily. However, note that we can choose X_n^* to depend only on $X_1, \ldots, X_{n-1}$ and thus be independent of X_n.

Let $\rho := n^{-A+\alpha}$. Let X be the random vector of length n whose coordinates are iid copies of x. From Definition 6.6 (and the observation that X_n has the same distribution as X after translating by a deterministic vector (namely the nth row of the deterministic matrix M), we have the conditional probability bound

$$\mathbf{P}(\mathrm{dist}_n \leq \varepsilon | X_n^* \notin S_{\varepsilon,\rho}) \leq \rho = n^{-A+\alpha}.$$

Thus it will suffice to establish the exponential bound

$$\mathbf{P}(\{X_n^* \in S_{\varepsilon,\rho}\} \wedge G) \leq \exp(-\Omega(n)).$$

Let

$$J := \lfloor 2A \rfloor \tag{20}$$

be the integer part of $2A$. Let $\alpha_1 > 0$ be a sufficiently small constant (independent of n and γ, but depending on α, A, J) to be chosen later. Set

$$\varepsilon_j := n^{(\gamma+\alpha_1)j} \varepsilon = \frac{1}{\beta} n^{(\gamma+\alpha_1)j} n^{-2A\gamma} \tag{21}$$

and

$$\rho_j := n^{(1/2-\alpha_1)j} \rho = n^{(1/2-\alpha_1)j} n^{-A+\alpha} \tag{22}$$

for all $0 \leq j \leq J$.

By the union bound, it will suffice to prove the following lemmas.

Lemma 9.2. *If α_1 is sufficiently small, then for any $0 \leq j < J$, we have*

$$\mathbf{P}(\{X_n^* \in S_{\varepsilon_j,\rho_j}\} \wedge \{X_n^* \notin S_{\varepsilon_{j+1},\rho_{j+1}}\} \wedge G) \leq \exp(-\Omega(n)). \tag{23}$$

Lemma 9.3. *If α_1 is sufficiently small, then we have*

$$\mathbf{P}(X_n^* \in S_{\varepsilon_J,\rho_J}) \leq \exp(-\Omega(n)).$$

10. Proof of Lemma 9.2

Fix $0 \leq j < J$. Note that by (14), we have

$$\rho_j \leq n^{(J-1)/2} n^{-A+\alpha} \leq n^{-1/2-\{2A\}/2+\alpha} \leq n^{-1/2}.$$

We can then use Theorem 6.8 to conclude the existence of a set $\mathcal{N}$ of unit vectors such that every vector in S_{ε_j,ρ_j} lies within ε_j in l^∞ norm to a vector in $\mathcal{N}$, and with the cardinality bound

$$|\mathcal{N}| \leq n^{-(1/2-o(1))n} \rho_j^{-n}. \tag{24}$$

Suppose that the event in Lemma 9.2 holds, then we can find $u \in \mathcal{N}$ such that $\|u - X_n^*\|_{l^\infty} \leq \varepsilon_j$, and thus $\|u - X_n^*\| \leq n^{1/2}\varepsilon_j$. On the other hand, since X_n^* is orthogonal to $X_1, \ldots, X_{n-1}$ and $\|M_n\| \ll n^\gamma$, we have

$$\begin{aligned}(\sum_{i=1}^{n-1} |X_i \cdot u|^2)^{1/2} &= (\sum_{i=1}^{n-1} |X_i \cdot (u - X_n^*)|^2)^{1/2} \\ &= \|M(u - X_n^*)\| \\ &\ll n^\gamma n^{1/2} \varepsilon_j \\ &\ll n^{1/2} n^{-\alpha_1} \varepsilon_{j+1}.\end{aligned}$$

On the other hand, from (23) and Definition 6.6 we have

$$\mathbf{P}(|X \cdot X_n^* - z| \leq \varepsilon_{j+1}) \leq \rho_{j+1} \tag{25}$$

for all $z \in \mathbf{C}$, where $X = (x_1, \ldots, x_n)$ consists of iid copies of x.

To conclude the proof, we will need the following lemma.

Lemma 10.1. *If w is any vector with $\|w\|_{l^\infty} \leq 1$, then*

$$\mathbf{P}(|X \cdot w| \geq n^{\gamma+\alpha_1}) \ll n^{-A}.$$

Proof. Write $w = (w_1, \ldots, w_n)$ and $X = (x_1, \ldots, x_n)$. Observe from (13) that with probability $O(n^{-A-1}) = O(n^{-A})$, all the coefficients in X are going to be of magnitude at most n^γ. Thus it suffices to show that

$$\mathbf{P}(|w_1\tilde{x}_1 + \ldots + w_n\tilde{x}_n| \geq n^{\gamma+\alpha_1}) \ll n^{-A}$$

where $\tilde{x}_1, \ldots, \tilde{x}_n$ are iid with law equal to that of x conditioned to the event $|x| \ll n^\gamma$. As x has mean zero and bounded second moment, one verifies from (13) and Cauchy-Schwarz that the mean of the $\tilde{x}_i$ is $O(n^{-(A+2)/2})$. Thus if we let $x_i' := \tilde{x}_i - \mathbf{E}(\tilde{x}_i)$, we see that it suffices to show that

$$\mathbf{P}(|w_1x_1' + \ldots + w_nx_n'| \geq \frac{1}{2}n^{\gamma+\alpha_1}) \ll n^{-A}.$$

We conclude the proof by the moment method, using the estimate

$$\mathbf{E}(|w_1x_1' + \ldots + w_nx_n'|^{2k}) \ll_k n^{2k\gamma}$$

for any integer $k \geq 0$. This is easily verified by a standard computation (using the hypothesis $\gamma \geq 1/2$), since all the x_i' have vanishing first moment, a second moment of $O(1)$, and a jth moment of $O_j(n^{(j-2)\gamma})$ for any $j > 2$. Now take k to be a constant, sufficiently large compared to A/α_1. □

We are now ready to finish the proof of Lemma 9.2. From Lemma 10.1 and the bound $\|u - X_n^*\| \leq \varepsilon_j$ we see that

$$\mathbf{P}(|X \cdot (X_n^* - u)| \geq \varepsilon_{j+1}) \leq n^{-A} \leq \rho_{j+1};$$

combining this with (25) using the triangle inequality, we see that

$$\sup_{z \in \mathbf{C}} \mathbf{P}(|X \cdot u - z| \leq \varepsilon_{j+1}) \ll \rho_{j+1}. \tag{26}$$

We can therefore bound the left-hand side of (23) by

$$\sum_{u \in \mathcal{N}: (26) \text{ holds}} \mathbf{P}\Big((\sum_{i=1}^{n-1} |X_i \cdot u|^2)^{1/2} \ll n^{1/2} n^{-\alpha_1} \varepsilon_{j+1}\Big).$$

Now suppose that $u \in \mathcal{N}$ obeys (26). If we have $\sum_{i=1}^{n-1} |X_i \cdot u|^2)^{1/2} \ll n^{1/2} n^{-\alpha_1} \varepsilon_{j+1}$, then the event $|X_i \cdot u| \leq \varepsilon_{j+1}$ must hold for at least $n - O(n^{1-2\alpha_1})$ values of i. On the other hand, from (26) we see that each of these events $|X_i \cdot u| \leq \varepsilon_{j+1}$ only occurs with probability $O(\rho_{j+1})$. We can thus bound

$$\begin{aligned}\mathbf{P}(\sum_{i=1}^{n-1} |X_i \cdot u|^2)^{1/2} \ll n^{1/2} n^{-\alpha_1} \varepsilon_{j+1}) &\leq \binom{n}{n - O(n^{1-2\alpha_1})} (O(\rho_{j+1}))^{n - O(n^{1-2\alpha_1})} \\ &\ll n^{o(n)} \rho_{j+1}^n.\end{aligned}$$

Applying (24), we can thus bound the left-hand side of (23) by

$$\ll n^{-(1/2 - o(1))n} \rho_j^{-n} \rho_{j+1}^n = n^{-(\alpha_1 - o(1))n}$$

and the claim follows.

11. Proof of Lemma 9.3

Suppose that X_n^* lies in $S_{\varepsilon_J, \rho_J}$. Then by Lemma 6.7, we have

$$X_n^* \subset \mathrm{Comp}(O(\frac{1}{n\rho_J^2}), O(\frac{\varepsilon_J}{\rho_J})).$$

Note from (22) and (20) that

$$\frac{1}{n\rho_J^2} = n^{2A - J - 1 + 2\alpha_1 J - 2\alpha} \leq n^{-\alpha_1}$$

if α_1 is sufficiently small. Thus, by arguing as in Section 8, the set $\mathrm{Comp}(O(\frac{1}{n\rho_J^2}), O(\frac{\varepsilon_J}{\rho_J}))$ has a $O(\frac{\varepsilon_J}{\rho_J})$-net $\mathcal{N}$ in l^2 of cardinality

$$|\mathcal{N}| \ll \binom{n}{\frac{1}{n\rho_J^2}} (O(\frac{\varepsilon_J}{\rho_J}))^{\frac{1}{n\rho_J^2}} = \exp(o(n)).$$

If we let $u \in \mathcal{N}$ be within $O(\frac{\varepsilon_J}{\rho_J})$ of X_n^*, then we have $|X_i \cdot u| \ll \frac{\varepsilon_J}{\rho_J}$ for all $1 \leq i \leq n-1$. Thus we can bound

$$\mathbf{P}(X_n^* \in S_{\varepsilon_J, \rho_J}) \leq \sum_{u \in \mathcal{N}} \mathbf{P}(|X_i \cdot u| \ll \frac{\varepsilon_J}{\rho_J} \text{ for all } 1 \leq i \leq n-1).$$

Now observe from (21), (22), (20) and the hypothesis $\gamma \geq 1/2$ that

$$\frac{\varepsilon_J}{\rho_J} = n^{-\alpha + 2\alpha_1 J} n^{-(2A-J)(\gamma - 1/2)} \leq n^{-\alpha/2},$$

say, if α_1 is sufficiently small. Thus by Corollary 6.3 (or by a minor modification of Theorem 6.4) we see that

$$\mathbf{P}(|X_i \cdot u| \ll \frac{\varepsilon_J}{\rho_J} \text{ for all } 1 \leq i \leq n-1) \ll \exp(-\Omega(n))$$

for each $u \in \mathcal{N}$, and the claim follows.

Acknowledgement

The authors would like to thank the referees for useful comments.

References

[1] Z. Bai and J. Silverstein, Spectral analysis of large dimensional random matrices, Second Edition, Springer, New York, 2010. MR2567175
[2] D. Bau and L. N. Trefethen, Numerical linear algebra, *SIAM* (1997). MR1444820 (98k:65002)
[3] P. T̈urgisser, F. Cucker, M. Lotz, The probability that a slightly perturbed numerical analysis problem is difficult, *Math. Comp.* **77** (2008), 1559-1583. MR2398780 (2009a:65132)
[4] P. Bürgisser, F. Cucker, M. Lotz, General formulas for the smooth analysis of condition numbers, *C. R. Acad. Sci. Paris*, **343** (2006), 145-150. MR2243310 (2007b:65147)
[5] P. Bürgisser, F. Cucker, M. Lotz, Smooth analysis of conic condition numbers, *J. Math. Pure Appl.* **86** (2006), 293-309. MR2257845 (2007h:65040)
[6] A. Edelman, Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** (1988), no. 4, 543–560. MR964668 (89j:15039)
[7] P. Erdös, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51** (1945), 898–902. MR0014608 (7:309j)
[8] J. Demmel, The probability that a numerical analysis problem is difficult, *Math. Comp.* **50** (1988), 449-480. MR929546 (89g:65062)
[9] J. Dunagan, D. A. Spielman and S. H. Teng, Smoothed Analysis of the Renegar's Condition Number for Linear Programming, *preprint*.
[10] G. Golub and C. Van Loan, Matrix computations, Third edition, Johns Hopkins Press, 1996. MR1417720 (97g:65006)
[11] Z. Füredi and J. Komlós, The eigenvalues of random symmetric matrices, *Combinatorica* **1** (1981), no. 3, 233–241. MR637828 (83e:15010)
[12] E. Kostlan, Complexity theory of numerical linear algebra, *J. Comp. Appl. Math.* **22** (1988), no. 2-3, 219–230. MR0956504 (89i:65048)
[13] J. Komlós, On the determinant of $(0, 1)$ matrices, *Studia Sci. Math. Hungar.* **2** (1967), 7-22. MR0221962 (36:5014)
[14] R. Latala, Some estimates of norms of random matrices, *Proc. Amer. Math. Soc.* **133** (2005), 1273–1282. MR2111932 (2005i:15041)
[15] A. 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. MR2146352 (2006g:52009)
[16] M.L. Mehta, Random Matrices and the Statistical Theory of Energy Levels, Academic Press, New York, NY, 1967. MR0220494 (36:3554)
[17] L.A. Pastur, The spectrum of random matrices, *Teoret. Mat. Fiz.* **10** (1973), 102–112. MR0475502 (57:15106)
[18] J. Renegar, On the efficiency of Newton's method in approximating all zeros of a system of complex polynomials, *Math. Oper. Res.* **12** (1987), no. 1, 121–148. MR882846 (88j:65112)
[19] M. Rudelson, Invertibility of random matrices: Norm of the inverse. *Annals of Mathematics* **168** (2008), no. 2, 575–600. MR2434885
[20] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and the condition number of random matrices, *Advances in Mathematics* **218** (2008), no. 2, 600–633. MR2407948
[21] S. Smale, On the efficiency of algorithms of analysis, *Bull. Amer. Math. Soc.* **13** (1985), 87–121. MR799791 (86m:65061)
[22] Y. Seginer, The expected norm of random matrices, *Combin. Probab. Comput.* **9** (2000), no. 2, 149–166. MR1762786 (2001a:62070)

[23] D. A. Spielman and 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. MR1989210 (2004d:90138)
[24] D. A. Spielman and S. H. Teng, Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time, *J. ACM* **51** (2004), no. 3, 385–463. MR2145860 (2006f:90029)
[25] A. Sankar, S. H. Teng, and D. A. Spielman, Smoothed Analysis of the Condition Numbers and Growth Factors of Matrices, *SIAM J. Matrix Anal. Appl.* **28** (2006), no. 2, 446–476. MR2255338 (2008b:65060)
[26] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices, *Annals of Mathematics,* **169** (2009), no. 2, 595–632. MR2480613
[27] T. Tao and V. Vu, The condition number of a randomly perturbed matrix, *STOC'07*, ACM, 248–255, 2007. MR2402448 (2010a:65069)
[28] T. Tao and V. Vu, Random matrices: The circular law, *Communications in Contemporary Mathematics*, **10** (2008), 261-307. MR2409368 (2009d:60091)
[29] T. Tao and V. Vu, Additive Combinatorics, Cambridge Univ. Press, 2006. MR2289012 (2008a:11002)
[30] J. von Neumann and H. Goldstein, Numerical inverting matrices of high order, *Bull. Amer. Math. Soc.* **53** (1947) 1021-1099. MR0024235 (9:471b)
[31] V. Vu, Spectral norm of random matrices, *Combinatorica* **27** (2007), no. 6, 721–736. MR2384414 (2009d:15060)
[32] P. Wigner, On the distribution of the roots of certain symmetric matrices, *Annals of Math.* **67** (1958), 325-327. MR0095527 (20:2029)

Department of Mathematics, UCLA, Los Angeles, California 90095-1555
E-mail address: tao@math.ucla.edu

Department of Mathematics, Rutgers University, Piscataway, New Jersey 08854
E-mail address: vanvu@math.rutgers.edu