1 Introduction

Let M_n denote a random $n \times n$ matrix whose entries are iid Rademacher random variables, i.e. every entry is $+1$ or -1 with probability $1/2$, independently over all entries. In this talk, we'll study the *singularity problem* for M_n, which is simply the question of how likely it is for M_n to be singular (non-invertible) for large n. Formally, we define

$$s_n = \Pr(M_n \text{ is singular}),$$

and we wish to understand how s_n behaves as $n \to \infty$.

On the one hand, it stands to reason that $s_n \to 0$ as $n \to \infty$. Indeed, if $\widetilde{M_n}$ is an $n \times n$ matrix drawn from some "usual" random matrix ensemble (e.g. the GUE or GOE ensembles), then $\widetilde{M_n}$ will be invertible a.s. The reason is that the space of singular matrices is a codimension-1 submanifold of $\mathbb{R}^{n^2}$, and so it has Lebesgue measure zero. As the "usual" ensembles have an absolutely continuous distribution, we see that $\widetilde{M_n}$ does not lie on this submanifold a.s. Additionally, we know that many results in random matrix theory exhibit *universality*, meaning that the behavior as $n \to \infty$ stops depending heavily on the precise distribution of the entries. Combining these two, one might expect that $\lim_{n\to\infty} s_n = 0$.

On the other hand, we can easily see that $s_n \neq 0$ for all n. Indeed, there is a 2^{-n} probability that the first two rows of M_n are equal, and if this happens, then M_n is certainly singular, which shows that $s_n \geq 2^{-n}$. More generally, if any two rows of M_n are either equal to one another or negatives of each other, then M_n is singular, and similarly for pairs of columns. This shows that

$$s_n \geq 2 \cdot 2 \cdot \binom{n}{2} \cdot 2^{-n} = (2 - o(1))n^2 2^{-n} = \left(\frac{1}{2} + o(1)\right)^n.$$

Moreover, no one has ever been able to come up with a significantly better lower bound on s_n, suggesting that this may be best possible. In other words, it may be that the "likeliest reason" for why a Rademacher random matrix is singular is that two of its rows or two of its columns are equal or negatives of each other.

Because of this intuition, people have over the years made a number of conjectures about s_n, which I present in below in increasing order of strength.

Conjecture 1.1.

$$s_n \leq \begin{cases} o(1) & \\ e^{-cn} & \text{for some } c > 0 \\ (\frac{1}{2} + o(1))^n & \\ (2 + o(1))n^2 2^{-n} & \end{cases}$$

The first conjecture, that $\lim s_n = 0$, was proven by Komlós in 1967. The first exponential bound, that $s_n < 0.999^n$, was proven by Kahn, Komlós, and Szemerédi in 1995. Their exponential constant was subsequently improved several times by several authors, and finally in 2018 Tikhomirov proved that $s_n \leq (\frac{1}{2} + o(1))^n$, which left only the final, strongest

conjecture open. A proof of this result was announced earlier this year by Irmatov, but my understanding is that it is not yet accepted as correct (and may have serious issues). Just two days ago, Jain, Sah, and Sawhney posted a paper on the arXiv which proves the analogue of this conjecture for a random matrix whose entries come from *any* discrete distribution that is not uniform on its support, thus "almost" proving this conjecture.

In this talk, we'll prove Komlós's theorem, that $s_n = o(1)$. In fact, we'll show a quantitative bound of the form $s_n = O(\log n/\sqrt{n})$, which is roughly the same quantitative result as Komlós's original bound. Komlós's original proof has been simplified many times (including several times by Komlós himself); the argument I'll present today is due to Asaf Ferber, and I learned of it from Matthew Kwan. However, many of the basic ideas go all the way back to Komlós; most notably, Komlós's key insight was that to understand the singularity problem, one needs to understand *anticoncentration*. Essentially every improved upper bound on s_n has involved developing new results on anticoncentration of certain random variables. As such, we'll begin by discussing anticoncentration.

2 Anticoncentration and Erdős–Littlewood–Offord

Before discussing anticoncentration, let's begin with concentration. Generally speaking, a concentration result for some random variable X is a statement of the form

$$\text{There exists an interval } I \text{ of length } \ell \text{ such that } \Pr(X \notin I) \text{ is small.} \tag{$*$}$$

Here, "small" can mean many things, e.g. a small absolute constant, or something decaying to zero as some parameter n tends to ∞, or maybe something decaying exponentially in n. Many concentration results exist, including simple ones like Markov's and Chebyshev's inequalities, ones giving exponential bounds such as Chernoff's, Hoeffding's, and Azuma's inequalities, and more sophisticated ones like Talagrand's inequality. Usually, when applying a concentration inequality, we want the length ℓ to be as small as possible.

Anticoncentration is simply the negation of $(*)$. Namely, an anticoncentration result for X is a statement of the form

$$\text{For any interval } I \text{ of length } \ell, \Pr(X \in I) \text{ is small.}$$

We'll be dealing only with discrete random variables, which means that it will usually be more convenient to think of our anticoncentration results as

$$\text{For any } x \in \mathbb{R}, \Pr(X = x) \text{ is small.}$$

Of course, up to the value of "small", the previous two statements are equivalent for discrete random variables. For example, if X is integer-valued, then we can get from the first statement to the second by taking $\ell = 1$ and using an interval around some integer x. To go from the second statement to the first, we can simply apply the bound on $\Pr(X = x)$ to all integer points $x \in I$, and get the first statement.

The first result in anticoncentration, and the one that we'll need, concerns the Littlewood–Offord problem, which is as follows. Let $\varepsilon_1, \dots, \varepsilon_n$ be iid Rademacher random variables, and let $a_1, \dots, a_n$ be arbitrary non-zero real numbers. Let

$$X = \sum_{i=1}^{n} \varepsilon_i a_i$$

be a random signed sum of the a_i. Littlewood and Offord wanted to find a good antinconcentration result for X, i.e. they wanted to obtain a good upper bound on

$$\sup_{x \in \mathbb{R}} \Pr(X = x).$$

Note that it is important to assume that all the a_i are non-zero, for otherwise we'll have some irrelevant terms in the sum for X, and we might even have that $X = 0$ always. However, other than this constraint, we won't assume anything about the a_i.

Intuitively, it makes sense that the worst case for the Littlewood–Offord problem (i.e. the case when $\sup_x \Pr(X = x)$ is largest) is when all the a_i are equal or close to equal. Indeed, if they have vastly different sizes, then we won't get any collisions among the random signed sums, and X will take on 2^n values, each with probability 2^{-n}. Similarly, since there is a symmetry in the problem, it is natural to guess that the highest point probability will be $\Pr(X = 0)$, or something similar. In case n is even and all the a_i equal 1, then $\Pr(X = 0) = \binom{n}{n/2}/2^n = O(1/\sqrt{n})$. Littlewood and Offord were able to prove a nearly matching bound.

Theorem 2.1 (Littlewood–Offord 1938)**.** *For any non-zero $a_1, \dots, a_n$ and any $x \in \mathbb{R}$,*

$$\Pr\left(\sum_{i=1}^{n} \varepsilon_i a_i = x\right) \leq O\left(\frac{\log n}{\sqrt{n}}\right).$$

Their proof used a quantitative version of the central limit theorem: roughly speaking, for large n, we have that X is close to a Gaussian in some appropriate metric. Since the probability that a Gaussian is equal to any fixed real number is zero, this closeness shows that $\Pr(X = x)$ is small as well. By using an appropriate quantitative CLT, one can obtain the bound above.

However, a few years later, Erdős improved their result, getting an exactly tight bound.

Theorem 2.2 (Erdős 1945)**.** *For any non-zero $a_1, \dots, a_n$ and any $x \in \mathbb{R}$,*

$$\Pr\left(\sum_{i=1}^{n} \varepsilon_i a_i = x\right) \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n} = O\left(\frac{1}{\sqrt{n}}\right)$$

Before we prove this, we'll need a simple but important combinatorial result. We say that a collection $\mathcal{F}$ of subsets of $[n]$ is an *antichain* if $S \not\subseteq T$ for any distinct $S, T \in \mathcal{F}$. In other words, $\mathcal{F}$ consists of incomparable elements in the Boolean lattice $2^{[n]}$.

Theorem 2.3 (Sperner 1928)**.** *If $\mathcal{F} \subset 2^{[n]}$ is an antichain, then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.*

Sperner's theorem is easily seen to be tight, as the collection of all sets of size $\lfloor n/2 \rfloor$ forms an antichain.

Proof (Bollobás 1965). Pick a random permutation π of $[n]$, and consider the probability that some set in $\mathcal{F}$ is an initial segment of π. For a fixed $S \in \mathcal{F}$, the probability that S is an initial segment of π is precisely $1/\binom{n}{|S|}$. Moreover, by the antichain condition, we see that these events are disjoint for distinct $S, T \in \mathcal{F}$, for if S is an initial segment of π, we cannot have T be an initial segment as well. Therefore, the probability that some $S \in \mathcal{F}$ is an initial segment is just $\sum_{S \in \mathcal{F}} \binom{n}{|S|}^{-1}$, and we conclude that

$$1 \geq \sum_{S \in \mathcal{F}} \frac{1}{\binom{n}{|S|}} \geq \sum_{S \in \mathcal{F}} \frac{1}{\binom{n}{\lfloor n/2 \rfloor}} = \frac{|\mathcal{F}|}{\binom{n}{\lfloor n/2 \rfloor}},$$

as claimed. □

Remark. We in fact proved a stronger statement, namely that for any antichain $\mathcal{F}$,

$$\sum_{S \in \mathcal{F}} \frac{1}{\binom{n}{|S|}} \leq 1.$$

This result is known as the LYM inequality, and often gives more refined results than Sperner's theorem itself.

Once we know Sperner's theorem, proving Erdős's bound on the Littlewood–Offord problem is quite straightforward.

Proof of Theorem 2.2. We may assume without loss of generality that all the a_i are positive, since the random variable X is unchanged by replacing a_i with $-a_i$. Fix some $x \in \mathbb{R}$, and let $\mathcal{F}$ consist of all sets $S \subseteq [n]$ such that

$$\sum_{i \in S} a_i - \sum_{j \notin S} a_j = x.$$

Note that $\Pr(X = x) = |\mathcal{F}|/2^n$. We claim that $\mathcal{F}$ is an antichain. Indeed, if $S \subset T$, then

$$\sum_{i \in S} a_i - \sum_{j \notin S} a_j < \sum_{i \in T} a_i - \sum_{j \notin T} a_j,$$

since we assumed that all the a_i were positive. In particular, at most one of the two terms can equal x, and thus at most one of S and T lies in $\mathcal{F}$. But then by Sperner's theorem, we conclude that

$$\Pr\left(\sum_{i=1}^{n} \varepsilon_i a_i = x\right) = \frac{|\mathcal{F}|}{2^n} \leq \frac{\binom{n}{\lfloor n/2 \rfloor}}{2^n}.$$

□

3 Bounding the singularity problem

Before we can prove our upper bound of $s_n = O(\log n/\sqrt{n})$, we will need a simple linear-algebraic lemma.

Lemma 3.1. *Let $V \subseteq \mathbb{R}^n$ be a subspace of dimension d. Then V contains at most 2^d vectors in $\{-1,1\}^n$.*

Proof. Let $v_1, \dots, v_d$ be a basis for V, and let A be the matrix whose columns are $v_1, \dots, v_d$. Since $\operatorname{rank} A = d$, there must exist indices $i_1, \dots, i_d \in [n]$ such that the rows indexed by $i_1, \dots, i_d$ are linearly independent. In particular, every other row is a fixed linear combination of these rows. This implies that for every $u_{i_1}, \dots, u_{i_d} \in \mathbb{R}$, there is a unique $u \in V$ whose $i_1, \dots, i_d$ coordinates are $u_{i_1}, \dots, u_{i_d}$. In particular, there are exactly 2^d vectors in V that have ± 1 in coordinates $i_1, \dots, i_d$, and thus at most 2^d vectors in $V \cap \{-1,1\}^n$. □

Now, let the rows of M_n be $\xi^1, \dots, \xi^n$, where the ξ^i are iid random vectors in $\{-1,1\}^n$. Let $V_i = \operatorname{span}(\xi^1, \dots, \xi^i)$ be the span of the first i rows. If M_n is singular, there must be some linear dependence among its rows, and in particular we must have that $\xi^{i+1} \in V_i$ for some $i \leq n-1$. Therefore, we conclude that

$$s_n \leq \sum_{i=1}^{n-1} \Pr(\xi^{i+1} \in V_i).$$

Additionally, we have that $\dim V_i \leq i$ and that ξ^{i+1} is a vector in $\{-1,1\}^n$. So by Lemma 3.1, we see that $\Pr(\xi^{i+1} \in V_i) \leq 2^{i-n}$. Therefore, for any $1 \leq k \leq n-1$,

$$s_n \leq \sum_{i=1}^{n-k-1} \Pr(\xi^{i+1} \in V_i) + \sum_{i=n-k}^{n-1} \Pr(\xi^{i+1} \in V_i) \leq 2^{-k} + \sum_{i=n-k}^{n-1} \Pr(\xi^{i+1} \in V_i). \tag{1}$$

Thus, we see that the first $n-k-1$ rows contribute very little to the probability of singularity. We will eventually end up taking $k = \log_2 n$, and so the first term above will be of order $1/n$, which is much smaller than the $O(\log n/\sqrt{n})$ bound we are trying to prove. More generally, the moral here is that the vast majority of the rows of M_n won't cause us any problems, but we need to work harder to deal with the last few rows. To do so, we will need one more lemma.

Lemma 3.2 (Ferber)**.** *Let $d \geq \frac{99}{100}n$. With probability at least $1 - 2^{-n/20}$, there is some $w \in V_d^\perp$ with at least $n/5$ non-zero coordinates. In fact, **every** non-zero $w \in \mathbb{Q}^n \cap V_d^\perp$ has at least $n/5$ non-zero coordinates.*

In what follows, we'll only need the "there exists" statement. However, it seems that the easiest way to prove the "there exists" statement is to prove the "for all" statement. While this is somewhat surprising, it's not completely crazy; for instance, in the case $d = n-1$, we will typically have that $\dim V_d^\perp = 1$, meaning that in this case, the "there exists" and the "for all" statements are the same, as there is a unique $w \in \mathbb{Q}^n \cap V_d^\perp$ up to scaling.

Proof. Heuristically, this lemma should be true becuase, for any fixed vector w with few non-zero coordinates, the probability that it lies in $V_d^\perp$ is small. However, as there are infinitely many such possible vectors w, we can't just apply the union bound. Instead, Ferber's clever idea is to reduce this whole problem modulo 3 so that we have a finite set to union-bound over.

Concretely, suppose that there is some $w \in \mathbb{Q}^n \cap V_d^\perp$ with at most $n/5$ non-zero coordinates. By rescaling w by the least common denominator of its entries and reducing every coordinate modulo 3, we end up with a vector $u \in \{0,1,2\}^n$ such that

$$\xi^i \cdot u \equiv 0 \mod 3 \qquad \text{for each } 1 \le i \le d.$$

Note too that u has at most $n/5$ non-zero coordinates, though it may have fewer non-zero coordinates than w, since some of w's coordinates may be divisible by 3. However, u has at least one non-zero coordinate, since otherwise we could have divided every entry of w by 3 before reducing modulo 3.

Next, we claim that for any fixed non-zero $u \in \{0,1,2\}^n$, we have that

$$\Pr(\xi^i \cdot u \equiv 0 \mod 3) \le \frac{1}{2}. \tag{2}$$

Indeed, since u is non-zero, it has at least one non-zero entry, which WLOG is u_1. Suppose that $\xi^i \cdot u \equiv 0 \mod 3$. If we let $\widetilde{\xi^i}$ denote ξ^i but with the first entry negated, then we must have that $\widetilde{\xi^i} \cdot u \not\equiv 0 \mod 3$. Indeed, since $u_1 \neq 0$, negating ξ_1^i adds either 1 or 2 to $\xi^i \cdot u \pmod 3$. Thus, every choice of ξ^i for which $\xi^i \cdot u \equiv 0$ has an equiprobable choice $\widetilde{\xi^i}$ for which $\widetilde{\xi^i} \cdot u \not\equiv 0$, implying (2).

Therefore, we conclude that

$$\Pr(\xi^i \cdot u \equiv 0 \mod 3 \text{ for each } 1 \le i \le d) \le 2^{-d} \le 2^{-.99n}. \tag{3}$$

Recall that this holds for any vector $u \in \{0,1,2\}^n$ with at most $n/5$ non-zero coordinates. The number of choices for such a u is

$$\sum_{m=1}^{n/5} \binom{n}{m} 2^m \le 2^{n/5} \sum_{m=1}^{n/5} \binom{n}{m} \approx 2^{.92n},$$

using the standard approximation for the volume of a Hamming ball in terms of the Shannon entropy of the radius. If you don't know what this is, it should it least be plausible that such a bound holds: the sum of the binomial coefficients up to $n/5$ is the probability that a binomial random variable is at most $2/5$ of its expectation, which should be exponentially small in n. Combining this with (3), we see that

$$\Pr(\exists w \in \mathbb{Q}^n \cap V_d^\perp \text{ with at most } n/5 \text{ non-zero coordinates}) \lesssim 2^{.92n} \cdot 2^{-.99n} \le 2^{-n/20}. \quad \square$$

Now, suppose $i \ge 99n/100$, and fix some $w \in V_i^\perp$ with at least $n/5$ non-zero coordinates, which exists with high probability. Then we find that

$$\Pr(\xi^{i+1} \in V_i) \le \Pr(\xi^{i+1} \cdot w = 0) = \Pr\left(\sum_{j=1}^n \xi_j^{i+1} w_j = 0\right).$$

By assumption, at least $n/5$ of the numbers $w_1, \dots, w_n$ are non-zero, and the coordinates ξ_j^{i+1} are iid Rademacher random variables. Therefore, by Theorem 2.2, we find that

$$\Pr(\xi^{i+1} \in V_i \mid \exists w \in V_i^\perp \text{ with } \geq n/5 \text{ non-zero coordinates}) \leq O\left(\frac{1}{\sqrt{n/5}}\right) = O\left(\frac{1}{\sqrt{n}}\right).$$

We now return to (1), and plug in $k = \log_2 n$. For sufficiently large n, we have that $n - \log_2 n \geq 99n/100$. We then see that

$$\begin{aligned} s_n &\leq \sum_{i=1}^{n-\log_2 n - i} 2^{-i} + \sum_{i=n-\log_2 n}^{n-1} \Pr(\xi^{i+1} \in V_i) \\ &\leq \frac{1}{n} + \left(\log_2 n \cdot 2^{-n/20}\right) + \sum_{i=n-\log_2 n}^{n-1} \Pr(\xi^{i+1} \in V_i \mid \exists \text{ good } w \in V_i^\perp) \\ &\leq \frac{1}{n} + O(2^{-n/30}) + O\left(\frac{\log n}{\sqrt{n}}\right) \\ &= O\left(\frac{\log n}{\sqrt{n}}\right), \end{aligned}$$

as claimed.