Non-Asymptotic Theory of Random Matrices
Lecture 18: Strong invertibility of subgaussian matrices and Small ball probability via arithmetic progression

Lecturer: Roman Vershynin Scribe: Yuji Nakatsukasa

Thursday, March 6th, 2007

1 Strong invertibility of subgaussian matrices

In the last lecture, we derived an estimate for the smallest singular value of a subgaussian random matrix;

Theorem 1. *Let A be a $n \times n$ subgaussian matrix. Then, for any $\epsilon > 0$,*

$$\mathbb{P}(s_n(A) < \frac{\varepsilon}{\sqrt{n}}) \leq c\varepsilon + Cn^{\frac{-1}{2}} \tag{1}$$

In particular, this implies $s_n(A) \sim \frac{1}{\sqrt{n}}$ with high probability. However, (1) cannot show $\mathbb{P}(s_n(A) < \frac{\varepsilon}{\sqrt{n}}) \to 0$ as $\epsilon \to 0$ because of the $Cn^{\frac{-1}{2}}$ term. If $s_n(A) \simeq 0$, then the matrix is not invertible. We want to know whether $Cn^{\frac{-1}{2}}$ can be removed or not.

Question. Can the term $Cn^{\frac{-1}{2}}$ in (1) be removed?

- Yes, for Gaussian matrices [3],[9]
- No, for Bernoulli matrices.

It cannot be removed for Bernoulli matrices, since $P_n = \mathbb{P}(s_n(A) = 0) = \mathbb{P}(A$ is singular$) > 0$, because two first row of A are equal with probability $\left(\frac{1}{2}\right)^n$. Therefore, we know

$$P_n \geq \left(\frac{1}{2}\right)^n.$$

Then we want to estimate an upper bound for P_n.

Question. Estimate of an upper bound for P_n?
There is a conjecture for this question by Erdös:
Conjecture. $P_n \leq \left(\frac{1}{2} + o(1)\right)^n$.
It is nontrivial to prove $P_n \to 0$ as $n \to \infty$[6]. This was proved in 1995 in [7]:

Theorem 2.

$$P_n \leq c^n \ for \ some \ constant \ c < 1.$$

So far, the best known bound is by Tao and Vu in [12]:

Theorem 3.

$$P_n \leq \left(\frac{3}{4} + o(1)\right)^n.$$

This bound is much better than $Cn^{\frac{1}{2}}$ seen in (1). Based on these results, Spielman and Teng[10] conjectured that the estimate of $s_n(A)$ can be improved:

Conjecture. For a $n \times n$ Bernoulli random matrix A,

$$P(s_n(A) \leq \frac{\varepsilon}{\sqrt{n}}) \leq \varepsilon + c^n, \quad c < 1.$$

Recently Rudelson and Vershynin[8] proved that this holds for all subgaussian matrices, up to an absolute constant:

Theorem 4 (Strong Invertibility Theorem)**.** *Let A be a $n \times n$ subgaussian matrix. Then,*

$$\mathbb{P}(s_n(A) \leq \frac{\varepsilon}{\sqrt{n}}) \leq C\varepsilon + c^n,$$

where $C > 0$ and $0 < c < 1$.

Letting $\varepsilon = 0$ in this Theorem, we get $\mathbb{P}(A \text{ is nonsingular}) \leq c^n$, which includes the result of [7].
We observe that all these results boil down to small ball probability, which we discuss next.

2 Littlewood-Offord problem

We want to bound from above the small ball probability

$$P_\varepsilon(a) = \sup_{v \in \mathbb{R}} \mathbb{P}(|S - v| \leq \varepsilon),$$

where

$$S = \sum_{k=1}^{n} a_k \xi_k,$$

$\xi_1, \cdots, \xi_n$ are independent identically distributed random varibles, and $(a_1, \cdots, a_n) = a \in \mathbb{R}^n$.

If $P_\epsilon(a)$ is small, that means the random sum S is well spread in $\mathbb{R}$.
For Gaussian ξ_k, we know $P_\varepsilon(a) \sim \varepsilon/||a||_2$.
However, for most other distributions evaluation of $P_\varepsilon(a)$ is hard. For example, for Bernoulli ξ_k, $P_\varepsilon(a)$ depends on a, as follows:

1. $a = (1, 1, 0, 0, \cdots, 0) : P_\varepsilon(a)(= P_0(a)) = \frac{1}{2}$ -this is bad.

2. $a = (1, 1, 1, \cdots, 1) : P_0(a) \sim n^{1/2}$. In fact, a classical result of Littlewood and Offord, strengthened by Erdos[1] proves that if $|a_k| \geq 1 \quad \forall k$, then $P_1(a) \leq\leq n^{-1/2}$. This is sharp for $a_k = 1$.

3. $a = (1, 2, 3, \cdots, n) : P_0(a) \sim n^{3/2}$.

This shows the result in [1] can be further reduced in this case. In [2],[4] it is proved that if $|a_j - a_k| \geq 1$ for $j \neq k$, then the small ball probability can be even smaller:

$$P_1(a) \leq n^{-3/2}.$$

How to further reduce the small ball probability is an open question. Since $P_0(a)$ is big when there are many cancellations in $\sum_{k=1}^{n} a_k \xi_k$, we want to know when this happens. Perhaps this occurs then coefficients a_k are arithmetically comparable. Tao and Vu[11] recently suggested studying the following phenomenon:

If $P_0(a)$ is large, then a has a rich additive structure.

Here, holding a rich additive structure means a enbeds into a short arithmetic progression. Rudelson and Vershynin[8] proved the following:

The coefficients of a are essentially contained in an arithmetic progression of length $\leq \dfrac{1}{P_\varepsilon(a)}$.

Here, "essentially" means most coefficients are near elements of the arithmetic progression.

Example 5.

- $(1, 1, \cdots, 1) \hookrightarrow$ *embeds into arithmetic progression of length 1.*
- $(1, 2, \cdots, n) \hookrightarrow$ *embeds into arithmetic progression of length n.*
- $(1/2, 1/3, 1, \cdots, 1) \hookrightarrow$ *embeds into arithmetic progression of length 6n.*
- $(p_1/q_1, p_2/q_2 \cdots, p_n/q_n) \hookrightarrow$ *embeds into arithmetic progression of length* $LCD(a) \cdot n$.

Here we give the definition of the essential least common denominator of real numbers:

Definition 6 (Essential LCD)**.** *Let $\alpha \in (0,1)$ and $\kappa \geq 0$. The essential least common denominator $D(a) = D_{\alpha,\kappa}(a)$ of a vector $a \in \mathbb{R}^n$ is defined as the infimum of $t > 0$ such that all except κ coordinates of the vector ta are of distance at most α from nonzero integers.*

Theorem 7 (Small Ball Probability[8])**.** *for any random variables $\xi_1, \cdots \xi_n$, Assume that $a = (a_1, a_2, \cdots, a_n)$ satisfies*

$$K_1 \leq |a_k| \leq K_2 \quad \forall k.$$

Then, $\forall \alpha, \kappa, \varepsilon$,

$$P_\varepsilon(a) \leq \frac{1}{\sqrt{\kappa}} \left(\varepsilon + \frac{1}{D_{\alpha,\kappa}(a)} \right) + Ce^{-c\alpha^2\kappa}.$$

Example 8. *Let $\alpha = 0.001, \kappa = 0.001n$.*

1. *$D(a) \leq$ const. $\rightarrow P_0(a) \leq n^{-1/2}$.*
2. *If the values of a are spread uniformly between two variables 1 and 2,*

$$a = (1, 1 + \frac{1}{n}, 1 + \frac{2}{n}, \cdots, 2) \;\rightarrow\; D(a) = n,\; P_\varepsilon(a) \leq n^{-3/2}.$$

3. *If $D(a)$ is larger $\rightarrow$ $P_\varepsilon(a)$ is smaller.*

In order to prove Small Ball Probability, in the next section we introduce Esseen's Lemma.

3 Esseen's Lemma

Esseen's Lemma bounds Small Ball Probability via charateristic functions. The charecteristic function $\phi(t)$ of a random variable X is defined as

$$\phi(t) = \mathbb{E}e^{iXt}.$$

Lemma 9 (Esseen's Lemma[5])**.**

$$\sup_{v \in \mathbb{R}} \mathbb{P}(|x - v| \leq 1) \leq C \int_{-1}^{1} |\phi(t)| dt.$$

Proof. we use Fourier Transform:

$$\hat{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-ixt} dx.$$

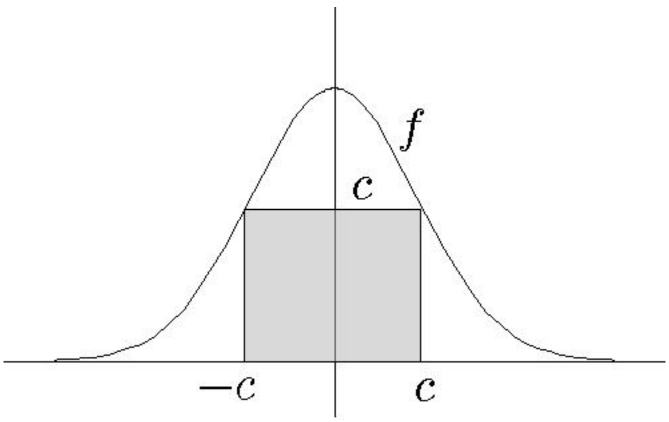
The inverse Fourier Transform is

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \hat{f}(t) e^{ixt} dt.$$

Assume (⋇)

$$f(x) \geq g(x),$$

$$\text{where } g(x) = \left\{ \begin{array}{ll} c, & |x| \leq c \\ c, & |x| > c \end{array} \right\}.$$



Then,

$$\mathbb{E}f(X) \geq \mathbb{E}g(X) = c\mathbb{E}1\{|X| \leq c\} = c\mathbb{P}(|X| \leq c).$$

On the other hand,

$$\begin{aligned} \mathbb{E}f(X) &\sim \mathbb{E} \int_{\mathbb{R}} \hat{f}(t) e^{iXt} dt \\ &= \int_{\mathbb{R}} \hat{f}(t)\phi(t) dt \\ &\lesssim \int_{-1}^{1} |\phi(t)| dt, \end{aligned}$$

where the last inequality holds provided that

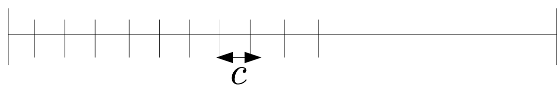
$$supp\hat{f} \subseteq [-1, 1], \ and \ ||\hat{f}||_\infty \leq C. \ (\hat{\divideontimes})$$

Therefore, we have proved :
If $\exists f$ satisfying $(\divideontimes), (\hat{\divideontimes})$, then

$$\mathbb{P}(|X| < c) \lesssim \int_{-1}^{1} |\phi(t)| dt.$$

It is an exersize to prove the existence of a function f satisfying $(\divideontimes), (\hat{\divideontimes})$. In order to complete the proof,

- To prove this for $|X - v|$ instead of $|X|$, we translate f by v, and redo the argument. ($|\hat{f}|$ will not change.)
- To prove for 1 instead of c, divide $[0, 1]$ into $1/c$ intervals of length c, and sum up the Small Ball Probability. ■



In the next lecture, we will apply Esseen's Lemma to prove Small Ball Probability(Theorem (7)).

References

[1] P. Erdös. On a lemma of littlewood and offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[2] P. Erdös and Leo Moser. Elementary problems and solutions: Solutions:. *Amer. Math. Monthly*, 54(4):229–230, 1947.

[3] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[4] A.; Szemeredi E. Erdos, P.; Sarkozi. On divisibility properties of sequences of integers. *Number Theory*, 43:35–49, 1970.

[5] C. G. Esseen. On the concentration function of a sum of independent random variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 9(37):290–308, 1968.

[6] G.Halasz. Estimates for the concentration function of combinatorial number theory and probability. *Period. Math. Hungar.*, 8(3-4):197–211, 1977.

[7] J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random ±1-matrix is singular. *JOT*, 8:223, 1995.

[8] Mark Rudelson and Roman Vershynin. Preprint. 2006.

[9] Daniel A Sankar. A, Spielman and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476, 2006.

[10] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, pages 597–606, Beijing, 2002. Higher Ed. Press.

[11] T. Tao and V. Vu. *Additive Combinatorics*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, NY, 2006.

[12] Terence Tao and Van Vu. On random ±1 matrices: singularity and determinant. *Random Structures Algorithms*, 28(1):1–23, 2006.