Non-Asymptotic Theory of Random Matrices
Lecture 17: Invertibility of Subgaussian Matrices; Small Ball Probability via the Central Limit Theorem

Lecturer: Roman Vershynin Scribe: Yuting Yang

Thursday, March 1, 2007

1 Invertibility of Subgaussian Matrices

Let A be an $n \times n$ subgaussian matrix (entries are i.i.d. subgaussian r.v's with variance 1). There are two reasons for the invertibility of A, depending on the nature of the unit vector on which A is acting – either compressible or incompressible. We recall that compressible vectors are those whose distance is at most some constant ρ from the set of (δn)-sparse vectors, and incompressible vectors are those that are not compressible. It is obvious that the unit sphere S^{n-1} is the disjoint union of the compressible vectors (**Comp**) and the incompressible vectors (**Incomp**). We have the following lemmas for A Gaussian.

Lemma 1 (Compressible)**.** $\mathbb{P}(\inf_{x \in Comp} \|Ax\|_2 \leq C\sqrt{n}) \leq \exp(-cn)$.

Lemma 2 (Incompressible)**.** *Let* $X_1, \cdots, X_n$ *be the rows of* A *and* H_n=*span*$(X_1, \cdots, X_{n-1})$. *Then, for* $\epsilon > 0$,

$$\mathbb{P}(\inf_{x \in Incomp} \|Ax\|_2 < \frac{C\epsilon}{\sqrt{n}}) \leq C\ \mathbb{P}(dist(X_n, H_n) < \epsilon).$$

It turns out that both lemmas also hold for subgaussian matrices.
For Gaussian matrices, $\mathbb{P}(dist(X_n, H_n) < \epsilon) \sim \epsilon$, where X_n is a random Gaussian vector and H_n is a hyperplane. Note that X_n and H_n are independent because of the independence of each row. Then, the probability for incompressible vectors is $\leq C\,\epsilon$, and thus

$$\mathbb{P}(s_n(A) < \frac{\epsilon}{\sqrt{n}}) \leq \exp(-cn) + C\,\epsilon.$$

In particular, $s_n(A) \sim 1/\sqrt{n}$ with high probability.
Now, the distance bound has to be proved for subgaussian A. Let X^* be a unit vector orthogonal to H_n (in particular, X^* is orthogonal to $X_1, \cdots, X_{n-1}$).

Such X^* is called a **random normal** vector. We leave the following result as an exercise:

$$dist(X_n, H_n) \geq |\ \langle X^*,\ X_n \rangle\ |. \qquad (Ex)$$

Note that X^* and X_n are independent. We will condition on X^*(i.e. fix X^*).
Let $X^* = (a_1, \cdots, a_n)$ be fixed, $X_n = (\xi_1, \cdots, \xi_n)$ be composed of i.i.d. r.v's. Then,

$$\langle\ X^*, X_n\ \rangle = \sum_{k=1}^{n}\ a_k\, \xi_k$$

is a sum of independent random vectors. Our goal is to find an upper bound for $\mathbb{P}(|\sum_{k=1}^{n}\ a_k\, \xi_k|\ <\ \epsilon)$.

2 Small Ball Probabilities

Consider a sum of independent r.v's $S = \sum_{k=1}^{n}\ a_k\, \xi_k$, where ξ_k are mean zero i.i.d. r.v's, and $a = (a_1, \cdots, a_n)\ \in \mathbb{R}^n$.
Exercise: Consider the special case where $\xi_k = \pm 1$ so that $S = \sum_{k=1}^{n} \pm a_k$. How is S distributed?
• Large Deviation Theory: S is concentrated around its mean (so in this case, around 0).
• Small Ball Probability: This theory gives lower bounds on its mean (anti-concentration). Define the **small ball(ϵ-ball) probability** with respect to a by

$$p_\epsilon(a) := \sup_{v \in \mathbb{R}}\ \mathbb{P}(|S - v| < \epsilon).$$

Then we want to find $p_\epsilon(a) \leq ?$.
For $\xi_k' s$ Gaussian, S is also Gaussian. So $p_\epsilon(a) \leq \sim \epsilon$. However, this estimate fails for $\pm$ 1 sums: take $a = (1, 1, \cdots, 1)$ and $S = \sum_{1}^{n} \pm 1$. Then, $\mathbb{P}(S = 0) \sim \frac{1}{\sqrt{n}}$ because the number of choices for cancelation(half +'s and half -'s) is equal to $\binom{n}{n/2}$. Then,

$$\mathbb{P}(S = 0) = \frac{\binom{n}{n/2}}{2^n}\ \sim\ \frac{1}{\sqrt{n}}.$$

Open Question: Is this the worst case?

We will use the **Central Limit Theorem (CLT)** to approximate the sum S by a Gaussian random variable, for which the small ball probabilities are easy to estimate.

Consider a random sum $S = \sum_{k=1}^{n} \zeta_k$, where ζ_k are centered independent r.v's with finite third moments. Then, the variance of S is $\sigma^2 = \sum_{k=1}^{n} \mathbb{E}|\zeta_k|^2$.

The classic CLT says that if $\sigma = 1$, then $S \sim N(0, 1)$. Here, we will use another version of the CLT.

Let g be standard Gaussian (i.e. $g \sim N(0, 1)$).

Theorem 3 (Central Limit Theorem (Berry-Esseén))**.** *Assume* $\sigma = 1$. *Then* $\forall\, t > 0$*,*

$$|\mathbb{P}(S < t) - \mathbb{P}(g < t)| \le C \sum_{k=1}^{n} \mathbb{E}|\zeta_k|^3.$$

See [2].

We use this for $S = \sum_{k=1}^{n} a_k\, \xi_k$ so that $\zeta_k = a_k \zeta_k$. Then,

$$\sigma^2 = \sum_{k=1}^{n} a_k^2\, \mathbb{E}|\xi_k|^2 = \|a\|_2^2,$$

so that $\sigma = \|a\|_2$.

Also, $\sum_{k=1}^{n} \mathbb{E}|\zeta_k|^3 \lesssim \|a\|_3^3$.

Corollary 4 (CLT)**.** *Assume* $\|a\|_2 = 1$. *Then,*

$$\forall\, t > 0, \qquad |\mathbb{P}(S < t) - \mathbb{P}(g < t)| \le C\|a\|_3^3.$$

Remark: The estimate using the third moment is best when a is well-spread and worst when peaked(in which case $\|a\|_2 = \|a\|_3$).

Exercise: $p_\epsilon(a) \le p_\epsilon(P_\sigma(a))$, where $P_\sigma : \mathbb{R}^n \to \mathbb{R}^\sigma$ is a projection onto coordinates in σ.

Intuitively, CLT is better for spread vectors because $\|a\|_3$ is smaller.

3 Central Limit Theorem for Incompressible Vectors

If a is incompressible, then a has $\sim n$ coordinates $\sim \frac{1}{\sqrt{n}}$. That is, $\exists\ \sigma \subset \{1, \cdots, n\} : |\sigma| \geq cn$ and $\frac{c_1}{\sqrt{n}} \leq |a_k| \leq \frac{c_2}{\sqrt{n}}\ \forall\ k \in \sigma$.
We can restrict a onto σ by the exercise above. Then, if $\|a\|_2 = 1$, we have $\|P_\sigma a\|_2 \sim 1$. (We are restricting incompressible a onto its spread portion.)
By Corollary 4, we have that, in particular,

$$|\mathbb{P}(|S - v| < t) - \mathbb{P}(|g - v| < t)| \leq 2C\|a\|_3^3$$

because the density of g is bounded above by $\frac{1}{\sqrt{2\pi}} \leq 1$ so that $\mathbb{P}(|g - v| < \epsilon) \leq \epsilon$. Then, $\mathbb{P}(|S - v| < \epsilon) \leq \epsilon + 2C\|a\|_3^3$.
Then,

$$p_\epsilon(a) \leq p_\epsilon(P_\sigma a) \leq \epsilon + 2C\|P_\sigma a\|_3^3 \lesssim \epsilon + \frac{1}{\sqrt{n}}$$

since $|a_k| \sim \frac{1}{\sqrt{n}}$.

Corollary 5 (SBP for Incompressible Vectors)**.** *If a is incompressible, then $p_\epsilon(a) \leq C(\epsilon + \frac{1}{\sqrt{n}}$.*

Remark:

(1) This is the best possible result because of the ± 1-sum case.

(2) This fails for sparse vectors: as a counterexample, take $a = (1, 1, 0, 0, \cdots, 0)$. Then, $\mathbb{P}(\sum \pm a_k = 0) = \frac{1}{2}$.

4 Random Normal Vectors are incompressible

We can control random normal X^* via the random matrix $A' := A \backslash$last row $= \text{rows}(X_1, \cdots, X_{n-1})$. Recall that X^* is orthogonal to all $X_1, \cdots, X_{n-1}$. So $X^* \in \ker A'$, and

$$A'\ X^* = 0.$$

Thus, if A' is invertible on some subset S of the unit sphere, then $X^* \notin S$. A', similarly to A, is invertible on the set of compressible vectors. Hence, $X^* \in \mathit{Incomp}$ with high probability.

Corollary 6. *With probability $\geq 1 - \exp(-cn)$, X^* is incompressible.*

5 Distance Bound

$\mathbb{P}(dist(X_n, H_n) < \epsilon) \leq \mathbb{P}(|\langle X^*, X_n \rangle| < \epsilon)$

$= \mathbb{E}_{X_1,\cdots,X_{n-1}} \mathbb{P}_{X_n}(|\langle X^*, X_n \rangle| < \epsilon \;\; and \;\; X^* \in Incomp| \; given \;\; X_1, \cdots, X_{n-1}) + \mathbb{P}(X^* \in Comp).$

By Corollary 6, if $X^* \in Comp$, then $\mathbb{P}(|\langle X^*, X_n \rangle| < \epsilon) \leq C(\epsilon + \frac{1}{\sqrt{n}})$. So,

$$\mathbb{P}(dist(X_n, H_n) < \epsilon) \leq C(\epsilon + \frac{1}{\sqrt{n}} + \exp{(-cn)}.$$

Theorem 7 (Distance)**.** $\mathbb{P}(dist(X_n, H_n) < \epsilon) \leq C(\epsilon + \frac{1}{\sqrt{n}}$.

It follows that

Theorem 8 ($s_n(A)$)**.** *Let A be an $n \times n$ subgaussian random matrix. Then,* $\forall \; \epsilon > 0$,

$$\mathbb{P}(s_n(A) \leq \frac{\epsilon}{\sqrt{n}}) \leq C(\epsilon + \frac{1}{\sqrt{n}}).$$

In particular, $s_n(A) \sim \frac{1}{\sqrt{n}}$ with high probability.

See [1].

References

[1] M.Rudelson; R.Vershynin. Sampling from large matrices: An approach through geometric functional analysis.

[2] Daniel W. Stroock. *Probability theory: an analytic view.* The Press Syndicate of the University of Cambridge, Cambridge ; New York, 1993.