Non-Asymptotic Theory of Random Matrices
Lecture 7: Largest, smallest singular values of random rectangular matrices

Lecturer: Roman Vershynin Scribe: Blake Hunter

Thursday, January 25, 2007

1 Large Deviation Inequalities

Let A be an $m \times n$ subgaussian matrix with entries that are independent copies of a subgaussian mean zero random variable x with variance $\mathbb{E}[x^2] = 1$.

Problem 1. *Compute $S_1(A)$, $S_n(A)$, the best numbers s.t.*

$$S_n(A)||x||_2 \leq ||Ax||_2 \leq S_1(A)||x||_2 \quad \forall x \in \mathbb{R}^n$$

($S_1(A) = ||A||$, the distortion of distance under A.)

1.) Asymptotic : $n \to \infty$, $\frac{n}{m} \to y > 0$ (General : $\mathbb{E}X^4 < \infty$)

$$S_1(A) \simeq \sqrt{m} + \sqrt{n},$$

$$\text{or} \quad \frac{S_1(A)}{\sqrt{m}} \to 1 + \sqrt{y} \quad a.s.$$

$$S_n(A) \simeq \sqrt{m} - \sqrt{n},$$

$$\text{or} \quad \frac{S_n(A)}{\sqrt{m}} \to 1 - \sqrt{y} \quad a.s.$$

See [1] for the Gaussian case and [3] for the general case. Note that this says if y is small (if A is tall) then A is almost an isometry.

2.) Non-Asymptotic
We proved : $S_1(A) \leq C(\sqrt{m} + \sqrt{n})$ with high probability (w.h.p.). We will improve this to

$$\sqrt{m} - C\sqrt{n} \leq S_1(A) \leq \sqrt{m} + C\sqrt{n} \quad w.h.p.$$

Previous Method:

$$S_1(A) = \max_{x\epsilon S^{n-1}} ||Ax||_2 = \max_{x\epsilon S^{n-1}, y\epsilon S^{n-1}} \langle Ax, y\rangle$$

- Then discretize the spaces (ϵ-nets)
- Prove estimate for fixed x,y in the net
- Union bound

This would not work for lower estimates $\langle Ax, y\rangle = 0$. This will work with $||Ax||_2$. Ax has independent coordinates $\langle \bar{x_k}, x\rangle$, $k = 1, ..., m$ where $\bar{x_k}$ are the rows of A:

$$||Ax||_2^2 = \sum_{k=1}^{m} \langle \bar{x_k}, x\rangle^2.$$

Note that $\langle \bar{x_k}, x\rangle$ are i.i.d. subgaussian r.v.'s and $\langle \bar{x_k}, x\rangle^2$ are not subgaussian. In fact, we actually have

$$X subgaussian \Leftrightarrow X^2 subexponential$$

Definition 2. *X is a subexponential r.v. if*

$$\mathbb{P}(|X| > t) \leq Ce^{-ct} \quad \forall t > 0.$$

Exercise 3. *Deduce a version of Lemma (Tails/Moments/Integer) of lecture 3 for subexponential random variable. In particular, $(\mathbb{E}|X|^P)^{\frac{1}{p}} \leq c_3 p \; \forall p \geq 1$.*

Note that sums of independent subexponential r.v.'s are not necessarily subexponential.

Example 4. *Let $a_k = 0, ..., a_1 = 1$. Then we can calculate the variance of the sum:*

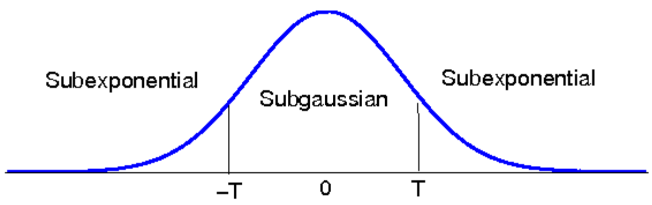
$$\mathbb{E}|\sum_k a_k x_k|^2 = \mathbb{E}(\sum_{k,j} a_k a_j x_k) = \sum_k a_k^2$$
$$= \|a\|^2$$

by the independence and $\mathbb{E}X_k^2 = 1$.

Theorem 5 (Large Deviations [2]). *Let $x_1, ..., x_n$ be i.i.d. subexponential random variables with $\mathbb{E}x_k = 0, a_1, ..., a_n \in \mathbb{R}$. Let $T = \frac{||a||_2}{||a||_\infty}$, (worst case $T = 1$, best case $T < \sqrt{n}$). Then for all $t > 0$,*

$$\mathbb{P}(\sum |a_k x_k| > t||a||_2) \leq \begin{cases} c\mathrm{e}^{-ct^2} & \text{for } t \leq T \\ c\mathrm{e}^{-cTt} & \text{for } t > T \end{cases}$$

This theorem shows we have the following situation:



Proof. Estimate the moment generating function. Without loss of generality, assume $\|a\|_2 = 1$.

Fact 6. *If X is subexponential then $\mathbb{E}\mathrm{e}^{\lambda X} \leq \mathrm{e}^{C_1 \lambda^2} for \lambda \leq C_1$*
(For Gaussian, holds for all λ)

Proof. We use the Taylor Expansion for the exponential: $\mathrm{e}^x = 1 + x + \sum_{p=2}^{\infty} \frac{x^p}{p!}$. We then have

$$\begin{aligned} \mathbb{E}\mathrm{e}^{\lambda X} &= 1 + \mathbb{E}[\lambda X] + \sum_{p=2}^{\infty} \frac{\lambda^2 \mathbb{E}x^p}{p!} \\ &\leq 1 + \sum_{p=2}^{\infty} (c\lambda)^p \leq 1 + c_1\lambda^2 \quad if \quad \lambda \leq c_1 \\ &\leq \mathrm{e}^{c_1\lambda^2} \end{aligned}$$

□

So then

$$p := \mathbb{P}(\sum_k a_k x_k > t) = \mathbb{P}(\mathsf{e}^{\lambda \sum_k a_k x_k} > \mathsf{e}^{\lambda t})$$
$$\leq \mathsf{e}^{-\lambda t} \mathbb{E} \mathsf{e}^{\lambda \sum_k a_k x_k}$$
$$= \mathsf{e}^{-\lambda t} \prod_k \mathbb{E} \mathsf{e}^{\lambda a_k x_k}.$$

If $\lambda \leq \frac{c_1}{\|a\|_\infty}$ then $\lambda a_k < c_1$ and we have

$$= \mathsf{e}^{-\lambda t} \prod_k \mathsf{e}^{c_1 \lambda^2 a_k^2}$$
$$= \mathsf{e}^{-\lambda t} + C_1 \sum_k \lambda^2 a_k^2$$
$$= \mathsf{e}^{-\lambda t} + C_1 \lambda^2$$

Optimize in $\lambda = \frac{t}{2C_1} \Rightarrow p \leq \mathsf{e}\frac{-t^2}{4C_1}$ (subgaussian tail).

1.) **Check:** $\frac{t}{2C_1} \leq \frac{c_1}{\|a\|_\infty} \Leftrightarrow t \leq \frac{2C_1 c_1}{\|a\|_\infty}$
We can choose $2C_1 c_1 = 1$.

Conclusion: if $t \leq \frac{1}{\|a\|_\infty}$ then $p < \mathsf{e}^{-ct^2}$, which proves part 1.

2.) If $t > \frac{1}{\|a\|_\infty}$, then we can choose $\lambda = \frac{c_1}{||a||_\infty}$
Check that $c_1 \lambda^2 \leq \frac{1}{2} \lambda t$.
Hence $p \leq \mathsf{e}^{\frac{-\lambda t}{2}} = \mathsf{e}^{\frac{-t}{2||a||_\infty}}$, a subexponential tail, proving the remainder of the theorem.

□

Now we want to prove with high probability,

$$S_n(A) \leq \|Ax\|_2 \leq S_1(A) \quad \forall x \in S^{n-1},$$

where

$$\|Ax\|_2^2 = \sum_{k=1}^m \langle \bar{x_k}, x \rangle^2.$$

We note that:
1.) $\mathbb{E}||Ax||_2^2 = m \cdot \mathbb{E}\langle \bar{x_k}, x\rangle^2 = m \cdot \|x\|_2^2 = m$
2.) Deviation :

$$\|Ax\|_2^2 - m = \sum_{k=1}^{m}(\langle \bar{x_k}, x\rangle^2 - 1) =: \sum_{k=1}^{m} z_k$$

The $\langle \bar{x_k}, x\rangle$ are subgaussian, by the Lemma, and the $\langle \bar{x_k}, x\rangle^2$ are subexponential.

Corollary 7. *Let $x_1, ..., x_n$ be i.i.d. subexponential mean zero random variables. Then*

$$\mathbb{P}(\frac{1}{\sqrt{n}}|\sum_{k=1}^{n} x_k| > t) \leq \begin{cases} c\mathsf{e}^{-ct^2} & for\ t \leq \sqrt{n} \\ c\mathsf{e}^{-c\sqrt{n}t} & for\ t > \sqrt{n} \end{cases}$$

Apply the corollary for z_k for every $x \in S^{n-1}$

$$\mathbb{P}\left(\frac{1}{\sqrt{m}}\left|\|Ax\|_2^2 - m\right| > t)\right) \leq \mathsf{e}^{-ct^2} \quad if\ t \leq \sqrt{m}$$

Discretize S^{n-1} **:** Let $\mathcal{N} \subseteq S^{n-1}$ be an ϵ-net, $|\mathcal{N}| \leq \left(\frac{3}{\epsilon}\right)^n$. Then

$$\begin{aligned}\mathbb{P}\left(\exists\ x \in \mathcal{N} : \left|\frac{1}{\sqrt{m}}\|Ax\|_2^2 - m\right|\right) &\leq |\mathcal{N}|\mathsf{e}^{-ct^2} \\ &\leq \mathsf{e}^{n\log\frac{3}{\epsilon} - ct^2} \\ &\leq \mathsf{e}^{\frac{-ct^2}{2}} \quad if\ ct^2 > 2n\log\frac{3}{\epsilon}.\end{aligned}$$

Now set

$$t := c\sqrt{n\log\frac{3}{\epsilon}}$$

We proved with probibility that $\forall\ x \in \mathcal{N}$:

$$\left|\frac{\|Ax\|_2^2}{m} - 1\right| < \frac{t}{\sqrt{m}} = c\sqrt{\frac{n}{m}\log\frac{3}{\epsilon}}$$

For the whole S^{n-1} every $z \in S^{n-1}$ can be written as,

$$z = x + u, \quad x \in \mathcal{N}, \quad \|u\|_2 \leq \epsilon.$$

Then

$$||\|Az\|_2 - \|Ax\|_2| \leq \|A(z-x)\|_2 = \|Au\|_2 \leq \|A\| \cdot \|u\|_2 \leq \epsilon\sqrt{m} \cdot 2c$$

since $\|A\| \leq c(\sqrt{m} + \sqrt{n}) \leq 2c\sqrt{m}$ where $m > n$.
So

$$\forall\ z \in S^{n-1},\ \left| \frac{\|Az\|_2^2}{\sqrt{m}} - 1 \right| < c\sqrt{\frac{n}{m} \log \frac{3}{\epsilon}} + 2c\epsilon$$

Finally, we optimize in ϵ, $\epsilon = \sqrt{\frac{n}{m}} < c'\sqrt{\frac{n}{m} \log \frac{3m}{n}}$.

Theorem 8. *For a subgaussian matrix A, $\frac{n}{m} = y$*

$$1 - c\sqrt{y \log \frac{1}{y}} \leq \frac{S_n(A)}{\sqrt{m}} \leq \frac{S_1(A)}{\sqrt{m}} \leq 1 + c\sqrt{y \log \frac{1}{y}}$$

with high probability.

References

[1] J.W. Silverstein. On the weak limit of the largest eigenvalue of a large dimensional sample variance matrix. *Journal of Multivariate Analysis*, 30:1, 1989.

[2] R. Vershynin. On large random almost euclidean bases. *Acta Mathematica Universitatis Comenianae*, 69:137, 2000.

[3] Z.D.Bai ; J.W.Silverstein ; Y.Q.Yin. A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 26:166, 1988.