Notes for Random Matrix Theory

Vishwak Srinivasan

Contents

1 Introduction

A useful fact is the following: *tall matrices are approximate isometries.* Let's parse this statement.

- Tall matrices are those matrices $A \in \mathbb{R}^{N\times n}$ where $N \gg n$.
- Approximate isometries: consider the vector space $\mathbb{R}^n$ and $\mathbb{R}^N$. A tall matrix transforms a vector $x \in \mathbb{R}^n$ to $Ax \in \mathbb{R}^N$. Mathematically:

  $$(1-\delta)K\|x\|_2 \leq \|Ax\|_2 \leq (1+\delta)K\|x\|_2$$

  where K is a normalization factor and $\delta \ll 1$. This looks like a version of Johnson-Lindenstrauss.

Now divide by $\|x\|_2$ to get:

$$(1-\delta)K \leq \frac{\|Ax\|_2}{\|x\|_2} \leq (1+\delta)K \Rightarrow (1-\delta)K \leq \sigma_{\min}(A) \leq \sigma_{\max}(A) \leq (1+\delta)K$$

and this tells us that the range of singular values is small. Furthermore, the condition number be bounded as: $\kappa(A) \leq \frac{1+\delta}{1-\delta} \approx 1$ for the specified bound on δ. Therefore, tall matrices are well conditioned *always.*

1.1 Some miscellaneous results

ϵ-nets are a neat way of computing quantities that can be expressed over balls. An ϵ-net of a set S is a set of points $\mathcal{N}_\epsilon(S)$ that approximates a point in the original set. That is to say, for every $x \in S$, there exists $y \in \mathcal{N}_\epsilon(S)$ such that $\|y - x\| \leq \epsilon$. As you can see, it is define w.r.t a norm. The cardinality of the $\mathcal{N}_\epsilon(S)$ is called the ϵ-covering number of S (w.r.t. a norm).

Now, let's look at an application of nets in computing the ℓ_2-norm of a vector.

Lemma 1.1.1. *Let $x \in \mathbb{R}^p$. Then:*

$$\|x\|_2 \leq \frac{1}{1-\epsilon} \sup_{v\in\mathcal{N}_\epsilon(S^{p-1})} |v^T x|$$

Proof. Let v be the unit vector that results in $|v^T x| = \|x\|_2$ (this is just unit vector in the direction of x). Now, let $w \in \mathcal{N}_\epsilon(S^{p-1})$ be the closest element to v. By Cauchy-Schwarz:

$$|(v-w)^T x| \leq \|v-w\|_2\|x\|_2 \leq \epsilon\|x\|_2$$

Hence, by triangle inequality:

$$|w^T x| \geq |v^T x| - |(w-v)^T x| \geq \|x\|_2 - \epsilon\|x\|_2 \Rightarrow \|x\|_2 \leq \frac{1}{1-\epsilon}|w^T x| \leq \frac{1}{1-\epsilon}\sup_{w\in\mathcal{N}_\epsilon(S^{p-1})} |w^T x|$$

□

Remark. An important takeaway is that an optimization problem over an uncountable set, has now be reduced to a countable set, at the cost of some sub-optimality. The ϵ-covering number of S^{p-1} is $\left(1+\frac{2}{\epsilon}\right)^p$.

Back to random matrix theory, let's use an idea from the above lemma to compute the spectral norm of a matrix A.

Lemma 1.1.2. *Let $A \in \mathbb{R}^{N\times n}$. Then:*

$$\|A\|_2 \leq \frac{1}{1-\epsilon} \sup_{v\in\mathcal{N}_\epsilon(S^{n-1})} \|Av\|_2$$

Proof. The proof is the same from earlier. Let v be unit vector leading to $\|Av\|_2 = \|A\|_2$. Choose $w \in \mathcal{N}_\epsilon(S^{n-1})$ closest to v. Using the fact that $\|Ax\|_2 \le \|A\|_2\|x\|_2$:

$$\|A(w-v)\| \le \|A\|_2\|w-v\|_2 \le \|A\|_2\epsilon$$

By triangle inequality:

$$\|Aw\|_2 \ge \|Av\|_2 - \|A(w-v)\|_2 \ge \|A\|_2 - \epsilon\|A\|_2 \Rightarrow \|Aw\|_2 \le \frac{1}{1-\epsilon}\|A\|_2 \le \frac{1}{1-\epsilon}\sup_{w\in\mathcal{N}_\epsilon(S^{n-1})}\|Aw\|_2$$

□

2 Random matrices

2.1 Independent entries

We will consider the following setup: entries of the matrices being analyzed are independent and zero mean.

Let's look at a asymptotic result, famously known as the Bai-Yin's law:

Theorem 2.1.1 (Bai-Yin's law)**.** *Let A be an $N \times n$ matrix whose entries are independent copies of a random variable with zero mean, unit variance and finite 4^{th} moment. Consider the scenario where $N, n \to \infty$ while $\frac{n}{N} = c \in [0, 1]$. Then:*

$$\sigma_{\min}(A) = \sqrt{N} - \sqrt{n} + o(\sqrt{n}) \qquad \sigma_{\max}(A) = \sqrt{N} + \sqrt{n} + o(\sqrt{n})$$

almost surely.

To revisit notation, $f(n) \in o(\sqrt{n})$ is such that $\lim\limits_{n\to\infty} \frac{f(n)}{\sqrt{n}} = 0$. A non-asymptotic version of the Bai-Yin's law is as follows:

Theorem 2.1.2 (Gordon's Theorem)**.** *Let A be an $N \times n$ matrix whose entries are independent standard normal variables. Then:*

$$\sqrt{N} - \sqrt{n} \le \mathbb{E}[\sigma_{\min}(A)] \le \mathbb{E}[\sigma_{\max}(A)] \le \sqrt{N} + \sqrt{n}$$

Proof. For our proof, we will use the following theorem due to Sudakov and Fernique.

Theorem 2.1.3 (Sudakov-Fernique Theorem)**.** *Let $(X_t)_{t\in T}$ and $(Y_t)_{t\in T}$ be two Gaussian processes satisfying:*

$$\mathbb{E}\left[(X_s - X_t)^2\right] \le \mathbb{E}\left[(Y_s - Y_t)^2\right]$$

for all $s, t \in T$ and T being an abstract set. Then:

$$\mathbb{E}\left[\sup_{t\in T} X_t\right] \le \mathbb{E}\left[\sup_{t\in T} Y_t\right]$$

First note that:

$$\sigma_{\max}(A) = \sup_{u\in S^{n-1}} \sup_{v\in S^{N-1}} \langle Au, v\rangle$$

Therefore, for a Gaussian process defined as $X_{u,v} = \langle Au, v\rangle$, we have that:

$$\mathbb{E}\left[\sup_{t\in T} X_{u,v}\right] = \mathbb{E}[\sigma_{\max}(A)]$$

for $T = S^{N-1} \times S^{n-1}$. To complement this, we define another Gaussian process $Y_{u,v} = \langle g, u \rangle + \langle h, v \rangle$, where g and h are Gaussian random vectors. For this Gaussian process, we have:

$$\mathbb{E}\left[\sup_{t\in T} Y_{u,v}\right] = \mathbb{E}\left[\|g\|_2 + \|h\|_2\right] = \sqrt{N} + \sqrt{n}$$

It remains to check if $\mathbb{E}\left[(X_{u,v} - X_{u',v'})^2\right] \leq \mathbb{E}\left[(Y_{u,v} - Y_{u',v'})^2\right]$.

$$\mathbb{E}\left[(X_{u,v} - X_{u',v'})^2\right] \overset{(i)}{=} \sum_{i=1}^{N}\sum_{j=1}^{n}(v_i u_j - v'_i u'_j)^2 \overset{(ii)}{\leq} \|u-u'\|_2^2 + \|v-v'\|_2^2 \overset{(iii)}{=} \mathbb{E}\left[(Y_{u,v} - Y_{u',v'})^2\right]$$

Step (i) uses Lemma 2.1.1, Step (ii) uses Lemma 2.1.2 and Step (iii) uses Lemma 2.1.3, and therefore the upper bound is complete.

For $\sigma_{\min}(A)$, the respective program is:

$$\sigma_{\min}(A) = \inf_{u\in S^{n-1}} \sup_{v \in S^{N-1}} \langle Au, v \rangle$$

There exists a result that generalizes the Sudakov-Fernique theorem to min-max objectives, and consequently results in the lower bound. □

Remark. $X_{u,v}$ a Gaussian process because: $\langle Au, v \rangle = \sum_{i=1}^{N}\sum_{j=1}^{n} v_i u_j A_{ij} \sim \mathcal{N}\left(0, \|v\|_2^2\|u\|_2^2\right)$.

2.1.1 Auxiliary lemmata

Lemma 2.1.1. *Let A be a matrix whose entries are independent standard normal random variables. Then for any $u, u' \in \mathbb{R}^n$ and $v, v' \in \mathbb{R}^N$, we have:*

$$\mathbb{E}\left[(v^T A u - v'^T A u')^2\right] = \sum_{i=1}^{N}\sum_{j=1}^{n}(v_i u_j - v'_i u'_j)^2$$

Proof. □

Lemma 2.1.2. *For any $u, u' \in \mathbb{R}^n$ and $v, v' \in \mathbb{R}^N$, we have:*

$$\sum_{i=1}^{N}\sum_{j=1}^{n}(v_i u_j - v'_i u'_j)^2 \leq \|u-u'\|_2^2 + \|v-v'\|_2^2$$

Proof. □

Lemma 2.1.3. *Let g, h be random normal vectors with zero mean. Define $Y_{u,v} = \langle g, u \rangle + \langle h, v \rangle$. Then for any $u, u' \in \mathbb{R}^n$ and $v, v' \in \mathbb{R}^N$, we have:*

$$\mathbb{E}\left[(Y_{u,v} - Y_{u',v'})^2\right] = \|u-u'\|_2^2 + \|v-v'\|_2^2$$

Proof. □