Notes for Random Matrix Theory

Vishwak Srinivasan

Contents

1 Introduction **2**

1.1 Some miscellaneous results 2

2 Random matrices **3**

2.1 Independent entries 3

2.1.1 Auxiliary lemmata 5

2.2 Independent rows 6

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 2.1.3, Step (ii) uses Lemma 2.1.4 and Step (iii) uses Lemma 2.1.5, 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: $\langle Au, v\rangle = \sum_{i=1}^{N}\sum_{j=1}^{n} v_i u_j A_{ij} \sim \mathcal{N}\left(0, \|v\|_2^2\|u\|_2^2\right)$. Similarly, $Y_{u,v} = \langle g, u\rangle + \langle h, v\rangle$ is a Gaussian process because $\langle g, u\rangle \sim \mathcal{N}(0, \|g\|_2^2)$ and $\langle h, v\rangle \sim \mathcal{N}(0, \|h\|_2^2)$.

Gordon's Theorem gives you bounds on the expected singular value. Via a Lipschitz concentration lemma, we can provide high probability estimates for instantiations of such matrices. The Lipschitz concentration lemma is stated below:

Lemma 2.1.1. *X is a standard normal variable. Let f be an L-Lipschitz function i.e., $|f(x) - f(y)| \leq L\|x - y\|_2$ for all $x, y \in \mathbb{R}^n$. Then:*

$$\Pr\left(|f(X) - \mathbb{E}\left[f(X)\right]| > t\right) \leq 2\exp\left(-\frac{t^2}{2L^2}\right)$$

This is useful because $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ are 1-Lipschitz functions of the inputs. The corollary below couples the results from above:

Corollary 2.1.1. *Let A be an $N \times n$ matrix with standard normal entries. Then with probability at least $1 - 2\exp(t^2/2)$, we have that:*

$$\sqrt{N} - \sqrt{n} - t \leq \sigma_{\min}(A) \leq \sqrt{N} - \sqrt{n} + t$$
$$\sqrt{N} + \sqrt{n} + t \leq \sigma_{\max}(A) \leq \sqrt{N} + \sqrt{n} - t$$

We now go back to our older discussion on approximate isometries. A matrix $\bar{A} = \frac{A}{\sqrt{N}}$ is an approximate isometry iff $\bar{A}^T\bar{A}$ is an approximate identity. Why is this true? Recall from earlier:

$$(1-\delta)\|x\|_2^2 \leq x^T\bar{A}^T\bar{A}x \leq (1+\delta)\|x\|_2^2 \Rightarrow |x^T(\bar{A}^T\bar{A} - I)x| \leq \delta\|x\|_2^2$$

where we have used $K = N$. This might seem as a fragile explanation, however, but motivates future discussions about relating $\bar{A}$ and $\bar{A}^T\bar{A}$ in our analyses.

Lemma 2.1.2. *If $B \in \mathbb{R}^{N\times n}$ is a matrix that satisfies:*

$$\|B^TB - I\|_2 \leq \min\{\delta, \delta^2\}$$

Then $(1-\delta) \leq \sigma_{\min}(B) \leq \sigma_{\max}(B) \leq (1+\delta)$

Conversely, if B *satisfies:*

$$(1-\delta) \leq \sigma_{\min}(B) \leq \sigma_{\max}(B) \leq (1+\delta)$$

then $\|B^TB - I\|_2 \leq 3\delta$.

Proof. First note that:

$$\|B^TB - I\|_2 \leq \alpha \Leftrightarrow |\|Bx\|_2^2 - 1| \leq \alpha$$

for any $x \in S^{n-1}$. This can be proven as follows:

$$\|B^TB - I\|_2 \leq \alpha \Leftrightarrow \max_{x\in S^{n-1}} |x^T(B^TB - I)x| \leq \alpha \Leftrightarrow \max_{x\in S^{n-1}} |\|Bx\|_2^2 - 1| \leq \alpha \Leftrightarrow |\|Bx\|_2^2 - 1| \leq \alpha \ \forall x \in S^{n-1}$$

This leads to:

$$1 - \sqrt{\alpha} \leq \sqrt{1-\alpha} \leq \|Bx\|_2 \leq \sqrt{1+\alpha} \leq 1 + \sqrt{\alpha}$$

Now substitute $\alpha = \delta$ and noting that $\delta \leq \sqrt{\delta}$ to complete the proof for the first statement.

For the second statement, we know that $\sigma_{\min}(B) = \sqrt{\lambda_{\min}(B^TB)}$ and $\sigma_{\max}(B) = \sqrt{\lambda_{\max}(B^TB)}$. Therefore, the eigenvalues of B^TB lie in the interval $[(1-\delta)^2, (1+\delta)^2]$. This means that the maximum absolute eigenvalue of $B^TB - I$ must be bounded as $\max\{|(1-\delta)^2 - 1|, |(1+\delta)^2 - 1|\} \leq \max\{2\delta, 3\delta\} = 3\delta$, and this completes the proof. □

2.1.1 Auxiliary lemmata

Lemma 2.1.3. *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:*

$$\mathbb{E}\left[(v^TAu - v'^TAu')^2\right] = \sum_{i=1}^{N}\sum_{j=1}^{n}(v_iu_j - v'_iu'_j)^2$$

Proof.

$$v^TAu - v'^TAu' = \sum_{i=1}^{N}\sum_{j=1}^{n} v_iu_jA_{ij} - v'_iu'_jA_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{n} A_{ij}\Delta_{ij}$$

for $\Delta_{ij} = v_iu_j - v'_iu'_j$.

This implies:

$$\begin{aligned}
(v^TAu - v'^TAu')^2 &= \left(\sum_{i=1}^{N}\sum_{j=1}^{n} A_{ij}\Delta_{ij}\right)^2 \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{k=1}^{N}\sum_{l=1}^{n} A_{ij}\Delta_{ij}A_{kl}\Delta_{kl} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n} A_{ij}^2\Delta_{ij}^2 + \sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{N}\sum_{\substack{l=1\\l\neq j}}^{n} A_{ij}A_{kl}\Delta_{ij}\Delta_{kl} \\
\Rightarrow \mathbb{E}\left[(v^TAu - v'^TAu')^2\right] &= \sum_{i=1}^{N}\sum_{j=1}^{n} \mathbb{E}\left[A_{ij}^2\right]\Delta_{ij}^2 + \sum_{i=1}^{N}\sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{N}\sum_{\substack{l=1\\l\neq j}}^{n} \underbrace{\mathbb{E}\left[A_{ij}A_{kl}\right]}_{=\,0}\Delta_{ij}\Delta_{kl} \\
&= \sum_{i=1}^{N}\sum_{j=1}^{n} \Delta_{ij}^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}(v_iu_j - v'_iu'_j)^2
\end{aligned}$$

$\square$

Lemma 2.1.4. *For any $u, u' \in S^{n-1}$ and $v, v' \in S^{N-1}$, 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. $\square$

Lemma 2.1.5. *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.

$$\begin{aligned}
Y_{u,v} - Y_{u',v'} &= \langle g, u-u'\rangle + \langle h, v-v'\rangle \\
\Rightarrow (Y_{u,v} - Y_{u',v'})^2 &= (\langle g, u-u'\rangle)^2 + (\langle h, v-v'\rangle)^2 + 2\langle g, u-u'\rangle \langle h, v-v'\rangle \\
\Rightarrow \mathbb{E}\left[(Y_{u,v} - Y_{u',v'})^2\right] &= \underbrace{\mathbb{E}\left[(\langle g, u-u'\rangle)^2\right]}_{\mathrm{Var}(\mathcal{N}(0,\|u-u'\|_2^2))} + \underbrace{\mathbb{E}\left[(\langle h, v-v'\rangle)^2\right]}_{\mathrm{Var}(\mathcal{N}(0,\|v-v'\|_2))} + \underbrace{2\mathbb{E}\left[\langle g, u-u'\rangle \langle h, v-v'\rangle\right]}_{=\,0}
\end{aligned}$$

$\square$

2.2 Independent rows

First, we will look at rows that are sub-Gaussian. Recall that a random vector $X \in \mathbb{R}^p$ is sub-Gaussian if for any $v \in S^{p-1}$

$$\Pr(|v^T X| > t) \leq 2\exp\left(-\frac{t^2}{2K^2}\right)$$

Also, a random vector X is said to be *isotropic* if $\Sigma(X) = I$.

The following theorem presents high-probability bounds on the range of singular values, under the aforementioned structure.

Theorem 2.2.1. *Let A be an $N \times n$ matrix, whose rows are independent isotropic sub-Gaussian random variables. For for every $t \geq 0$, we have with probability at least $1 - 2\exp(-ct^2)$ that:*

$$\sqrt{N} - C\sqrt{n} - t \leq \sigma_{\min}(A) \leq \sigma_{\max}(A) \leq \sqrt{N} + C\sqrt{n} + t$$

where $C := C(K)$ where K is the sub-Gaussian norm $K = \max_i \|A_i\|_{\psi_2}$.

Proof. From the approximate isometry property (Lemma 2.1.2), we have for $B = \frac{A}{\sqrt{N}}$:

$$\begin{aligned}
\|B^T B - I\|_2 \leq \max\{\delta, \delta^2\} &\Rightarrow 1 - \delta \leq \sigma_{\min}(B) \leq \sigma_{\max}(B) \leq 1 + \delta \\
&\Rightarrow \sqrt{N} - \sqrt{N}\delta \leq \sigma_{\min}(A) \leq \sigma_{\max}(A) \leq \sqrt{N} + \sqrt{N}\delta
\end{aligned}$$

So, if we able to show that $\|B^T B - I\|_2 \leq \max\{\delta, \delta^2\}$ for $\delta = C\frac{\sqrt{n}}{\sqrt{N}} + \frac{t}{\sqrt{N}}$, the proof is complete.

For this, note that we have to primarily control $\|(B^T B - I)x\|_2$ for $x \in S^{n-1}$. As seen earlier:

$$\max_{x \in S^{n-1}} \|x^T(B^T B - I)x\|_2 \leq \frac{1}{1-2\epsilon}\max_{x \in \mathcal{N}_\epsilon(S^{n-1})}\|x^T(B^T B - I)x\|_2 = \frac{1}{1-2\epsilon}\max_{x \in \mathcal{N}_\epsilon(S^{n-1})}\left|\frac{1}{N}\|Ax\|_2^2 - 1\right|$$

Choose $\epsilon = \frac{1}{4}$, and hence:

$$\max_{x \in S^{n-1}} \|x^T(B^TB - I)x\|_2 \leq 2 \max_{x \in \mathcal{N}_\epsilon(S^{n-1})} \left| \frac{1}{N}\|Ax\|_2^2 - 1 \right|$$

If remains to show that $\max\limits_{x \in \mathcal{N}_\epsilon(S^{n-1})} \left|\frac{1}{N}\|Ax\|_2^2 - 1\right| \leq \frac{\delta}{2}$.

$\|Ax\|_2^2 = \sum\limits_{i=1}^{N} \langle A_i, x\rangle^2 = \sum\limits_{i=1}^{N} Y_i^2$, where $Y_i = \langle A_i, x\rangle$ is a sub-Gaussian as well. The square of a sub-Gaussian random variable is sub-exponential. Additionally, note that Y_is are independent due to the independence of rows of A. Since A_i is isotropic, $\mathbb{E}[Y_i^2] = 1$, and consequently:

$$\mathbb{E}\left[\frac{1}{N}\|Ax\|_2^2\right] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} Y_i^2\right] = 1$$

Therefore, from a sub-exponential concentration lemma:

$$\Pr\left(\left|\frac{1}{N}\|Ax\|_2^2 - 1\right| \geq \epsilon\right) = \Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Z_i^2 - 1)\right| > \epsilon\right) \leq 2\exp\left(-c\min\left(\frac{\epsilon^2}{K_1^2}, \frac{\epsilon}{K_1}\right)N\right)$$

where $K_1 = \max_i \|Z_i^2 - 1\|_{\psi_1}$. We can bound K_1 as:

$$\|Z_i^2 - 1\|_{\psi_1} \overset{(i)}{\leq} \|2Z_i^2\|_{\psi_1} \leq 4\|Z_i\|_{\psi_2}^2 \leq 4K^2$$

where step (i) follows from a centering lemma. □

We also know that:

$$K \geq \|Z_i\|_{\psi_2} \geq \frac{1}{\sqrt{2}}\left(\mathbb{E}[|Z_i|^2]\right)^2 = \frac{1}{\sqrt{2}} \Rightarrow 4K^2 > 1$$

This gives:

$$\Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Z_i^2 - 1)\right| > \epsilon\right) \leq 2\exp\left(-\frac{c}{16K^4}\min\{\epsilon^2, \epsilon\}N\right)$$

If $\max\{\delta, \delta^2\} = \epsilon$, then $\min\{\epsilon, \epsilon^2\} = \delta^2$, this leads to:

$$\Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Z_i^2 - 1)\right| > \epsilon\right) \leq 2\exp\left(-\frac{c}{16K^4}\delta^2 N\right)$$

We also know that $\delta\sqrt{N} = C\sqrt{n} + t \Rightarrow \delta^2 N \geq C^2 n + t^2$.

$$\Rightarrow \Pr\left(\left|\frac{1}{N}\sum_{i=1}^{N}(Z_i^2 - 1)\right| > \epsilon\right) \leq 2\exp\left(-\frac{c'}{K^4}n - \frac{c''}{K^4}t^2\right)$$

Since we only considered a particular $x \in \mathcal{N}_\epsilon(S^{p-1})$, the probability that maximum deviation is greater than ϵ can be given via the union bound:

$$\Pr\left(\max_{x \in \mathcal{N}_\epsilon(S^{p-1})} \left|\frac{1}{N}\|Ax\|_2^2 - 1\right| \geq \epsilon\right) \leq 9^n \cdot 2\exp\left(-\frac{c'}{K^4}n - \frac{c''}{K^4}t^2\right) \leq 2\exp\left(-\frac{c''}{K^4}t^2\right)$$