Covering argument

II Non-asymptotic random matrix theory (RMT)

We study extreme singular values of random matrix $A \in \mathbb{R}^{m \times n}$.

$$\sigma_1(A) := \|A\| = \max_{x \in S^{n-1}} \|Ax\|_2 \ , \quad \sigma_n(A) = \min_{x \in S^{n-1}} \|Ax\|_2$$

Motivation:

- Condition number: $\frac{\sigma_1(A)}{\sigma_n(A)}$ determines stability in <u>many</u> applications
- Covariance matrix estimation
- techniques generalize. (E.g., what we learn today can be used to prove the restricted isometry property from compressed sensing)

1. Covering arguments

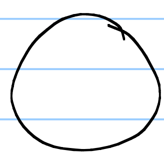
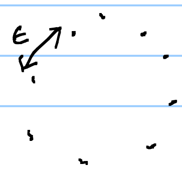
1.1 Simple covering argument

Assume A has iid $N(0,1)$ entries.

$\|A\| = \max_{x \in S^{n-1}} \|Ax\|_2$. i.e., the supremum of a random process indexed by $x \in S^{n-1}$.

Difficulty: $\max_{x \in S^{n-1}}$ is taken over ∞ number of points.

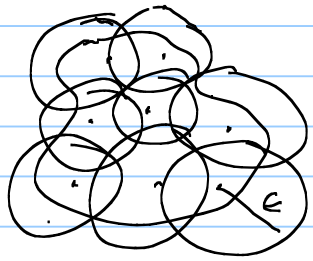
Solution: discretize S^{n-1}

Replace  with ϵ-net 

Def (<u>ϵ-net</u>) An ϵ-net of a subset K of a metric space is a set $\overline{K}$ satisfying $\forall\, x \in K, \exists\, z \in K_\epsilon$ s.t.

$$d(z,k) \leq \epsilon$$

In other words $\overline{K}_\epsilon \supseteq K$



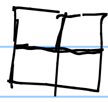
dots: $\overline{K}$
balls: $\overline{K}_\epsilon$

K is covered by $|\overline{K}|$ balls of radius ϵ.

Def (<u>Covering number</u>) The minimal cardinality of an ϵ-net of K, if finite, is denoted $N(K,\epsilon)$ and is called the covering number of K (at scale ϵ).

(Remark: Quantitative version of compactness, dimension, complexity of K.)

Q: $N(B_\infty^n, \frac{1}{2}) = ?$ (Using $\|\cdot\|_\infty$ metric)

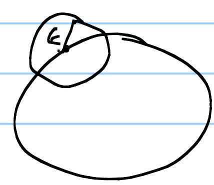
A:  $N(B_\infty^n, \frac{1}{2}) = 2^n$

Our goal: $N(S^{n-1}, \epsilon) \leq ?$ (Using $\|\cdot\|_2$ metric)

Lemma <u>$N(S^{n-1},\epsilon)$</u> S^{n-1} equipped w/ Euclidean metric satisfies

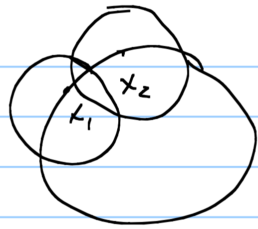
$$N(S^{n-1}, \epsilon) \leq \left(1 + \frac{2}{\epsilon}\right)^n \leq \left(\frac{3}{\epsilon}\right)^n \quad \text{for } 0 < \epsilon \leq 1$$

<u>Proof:</u> Construct ϵ-net:



Pick $x_1 \in S^{n-1}$ and grow by ϵ ball.

Pick $x_2 \in S^{n-1}$ such that $\|x_2 - x_1\|_2 > \epsilon$

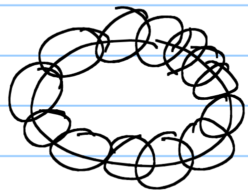


$\vdots$

Pick $x_i \in S^{n-1}$ such that $\|x_i - x_j\|_2 > \epsilon$ for $j = 1, \dots, i-1$

Stop if there are no points left to pick.

(This is known as a maximal packing)



Let $\bar{K} = \{x_1, x_2, \dots x_N\}$ be the set of points picked.

Claim: $\bar{K}$ is an ϵ-net.

Proof: Suppose not. This would violate our stopping criteria.

$|\bar{K}| \le ?$

Volumetric argument:

Observe: 1) $\bar{K}_{\frac{\epsilon}{2}}$ consists of $|\bar{K}|$ disjoint balls of radius $\frac{\epsilon}{2}$.

2) $\bar{K}_{\frac{\epsilon}{2}} \subseteq (1 + \frac{\epsilon}{2}) B_2^n$

$$\Rightarrow |\bar{K}| \operatorname{Vol}\left(\tfrac{\epsilon}{2} B_2^n\right) = \operatorname{Vol}\left(\bar{K}_{\frac{\epsilon}{2}}\right) \le \operatorname{Vol}\left(\left(1 + \tfrac{\epsilon}{2}\right) B_2^n\right)$$

$$\Rightarrow |\bar{K}| \le \frac{\operatorname{Vol}\left(\left(1 + \frac{\epsilon}{2}\right) B_2^n\right)}{\operatorname{Vol}\left(\frac{\epsilon}{2} B_2^n\right)} = \frac{\left(1 + \frac{\epsilon}{2}\right)^n}{\left(\frac{\epsilon}{2}\right)^n} = \left(1 + \frac{2}{\epsilon}\right)^n$$

QED

Prop (Discretization of $\|A\|$) Let $\bar{K}$ be a $\frac{1}{2}$-net of S^{n-1}. Then $\|A\| \le 2 \max_{x \in \bar{K}} \|Ax\|_2$

Proof: For any $x \in S^{n-1}$

$\|Ax\| \le \|Az\|_2 + \|A(z-x)\|_2$ where $z \in \mathcal{K}$, $\|x - z\|_2 \le \frac{1}{2}$

$\le \max_{z \in \mathcal{K}} \|Az\|_2 + \frac{1}{2}\|A\|$

$\Rightarrow \|A\| = \max_{x \in S^{n-1}} \|Ax\| \le \max_{z \in \mathcal{K}} \|Az\|_2 + \frac{1}{2}\|A\|$

$\Rightarrow \frac{1}{2}\|A\| \le \max_{z \in \mathcal{K}} \|Az\|$

QED

$$\boxed{\begin{array}{l}\underline{\text{Thm}}\ \text{Let } A \text{ have iid } N(0,1) \text{ entries.} \\ \quad \text{Then, w/ prob} \ge 1 - \exp(-cn), \\ \quad \|A\| \le C(\sqrt{m} + \sqrt{n})\end{array}}$$

Proof: Let $\mathcal{K}$ be a $\frac{1}{2}$-net for S^{n-1} w/ $|\mathcal{K}| \le 6^n$.
Fix $z \in \mathcal{K}$. $g := Az \sim N(0, I_m)$.

1. Expectation
2. Deviation $\neq$
3. Union bound

1. $\sqrt{\mathbb{E}\|g\|_2^2} = \sqrt{m}$

2. Deviation $\neq$. By Gaussian concentration,

$\|\, \|g\|_2 - \sqrt{m}\, \|_{\psi_2} \le C.. \Rightarrow P(|\|g\| - \sqrt{m}| > t) \le e^{-ct^2}$.

3. Union bound

$$P\left(\max_{z \in \mathcal{K}} |\|Az\|_2 - \sqrt{m}| \ge t\right) \le |\mathcal{K}|\, e^{-ct^2}$$

$$\le 6^n e^{-ct^2}$$

$$= \exp(n \ln(6) - ct^2)$$

$$\le \exp(-cn)$$

by choosing $t^2 \geq \frac{2n}{\ln 6}$

We have shown that w/ prob $\geq 1 - \exp(-cn)$

$$\max_{z \in \mathcal{E}} \|Ax\| \leq \sqrt{m} + C\sqrt{n}$$

Apply the last proposition to complete the proof.

QED

Goal: Bound $\sigma_n(A)$ as well.

- Show that $\left| \|Ax\|_2 - \sqrt{m} \right|$ is small $\forall x \in S^{n-1}$.

Thm Let $A \in \mathbb{R}^{m \times n}$ have iid $N(0,1)$ entries. Then,

w/ prob $\geq$???,

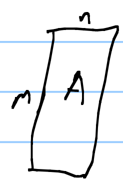
$$\sup_{x \in S^{n-1}} \left| \|Ax\| - \sqrt{m} \right| \leq ???? \quad \forall x \in S^{n-1}$$

$\boxed{\text{HW}}$ Use simple coveringing argument to fill in ?'s. Give tightest result you can (up to abs. const.'s).

$\boxed{\text{HW}}$ Generalize the theorem. What other distributions can A have? Prove more general version of thm using simple covering argument.

Remarks:

- ???? is $< \frac{1}{2}\sqrt{m}$ when $m \geq cn$. Thus, tall skinny random matrices are near isometries



$$\frac{1}{\sqrt{m}}\|Ax\|_2 \approx \|x\|_2$$

- ???? Is not optimal. Simple covering argument is not quite powerful enough.