Lecture 15

Last time

- Covering numbers via volume
- Detour: error correcting codes

Today

- Two-sided bounds on the singular values of sub-Gaussian matrices

Bounds on the singular values of sub-Gaussian matrices

We started our investigation by thinking that if

$$(\times) \qquad \max_{v \in \mathbb{S}^m} \|Av\|_2 \gtrsim \max_{v \in \mathcal{N}} \|Av\|_2$$

then we can apply union bound to get high probability bounds $\|A\|_{op}$. Today we will execute this plan. Our first result formalizes $(\times)$.

Lemma: Let A be a $n \times m$ matrix and $\varepsilon \in [0,1)$. Then, for any ε-net $\mathcal{N}$ of the sphere $\mathbb{S}^{m-1}$, we have

$$\sup_{x \in \mathcal{N}} \|Ax\|_2 \leq \|A\|_{op} \leq \frac{1}{1-\varepsilon} \sup_{x \in \mathcal{N}} \|Ax\|_2 .$$

Proof: The lower bound is trivial since

$\mathcal{N} \subseteq \mathbb{S}^{m-1}$! To prove the upper bound, take $v \in \mathbb{S}^{m-1}$ s.t. $\|A\|_{op} = \|Ax\|_{op}$. Since $\mathcal{N}$ is an ε-net, there exists $y \in \mathcal{N}$ s.t. $\|y - x\| \leq \varepsilon$. Then

$$\|Ax\| \leq \|Ax - Ay\| + \|Ay\|$$
$$\leq \varepsilon \|A\|_{op} + \|Ay\|.$$

Thus, rearranging

$$\|A_{op}\| \leq \frac{1}{1-\varepsilon} \|Ay\| \leq \frac{1}{1-\varepsilon} \max_{z \in \mathcal{N}} \|Az\|_2. \qquad \square$$

We shall use a slightly different version of this lemma.

Lemma (Max): Let $A \in S^n$ be a sym. matrix and $\varepsilon \in [0,1)$. Then, for any ε-net $\mathcal{N}$ of the sphere $\mathbb{S}^{n-1}$ ~~would~~ have

$$\sup_{x \in \mathcal{N}} \langle x, Ax \rangle \leq \|A\|_{op} \leq \frac{1}{1-2\varepsilon} \sup_{x \in} \langle x, Ax \rangle. \qquad \dashv$$

Try to modify our previous proof to derive this result.

We say that a random vector z in $\mathbb{R}^n$ is **sub-Gaussian** if

$$\|z\|_{\psi_2} := \sup_{v \in \mathbb{S}^{n-1}} \|\langle z, v \rangle\|_{\psi_2} < \infty.$$

We say that z is isotropic if

$$\mathbb{E} z z^T = I.$$

Lemma: Show that $z \sim N(0, I)$ is sub-Gaussian and isotropic. ⊣

We are ready for the main result of today.

Theorem: Let A be an $n \times m$ (with $m \leq n$) random matrix with independent, mean zero, sub-Gaussian, isotropic rows A_i. Then, for any $t \geq 0$ we have

$$\sqrt{n} - CK^2(\sqrt{m} + t) \leq \sigma_m(A) \leq \sigma_1(A) \leq \sqrt{n} + CK^2(\sqrt{m} + t)$$

with probability at least $1 - 2\exp(-t^2)$, where $K = \max_i \|A_i\|$. ⊣

Proof: We shall use the following Lemma to rewrite the conclusion.

Lemma (Iso): For $A \in \mathbb{R}^{n \times m}$ with $m \leq n$ and a scalar $\varepsilon > 0$, the following are equivalent:

(a) $\|A^T A - I\|_{op} \leq \varepsilon$.

(b) $(1 - \varepsilon)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \varepsilon)\|x\|_2^2 \quad \forall x.$

(c) $1-\varepsilon \le \sigma_m(A)^2 \le \sigma_1(A)^2 \le 1+\varepsilon$.

Proof of Lemma (Iso): (a) $\Leftrightarrow$ (b) WLOG assume $\|x\|_2 = 1$ in (b). Then

$$\|A^TA - I\|_{op} = \max_{x\in S^{m-1}} |x^T(A^TA - I)x|$$
$$= \max_{x\in S^{m-1}} \left| \|Ax\|_2^2 - 1 \right|$$

Thus, this being bounded by ε is equivalent to (b).

(b) $\Leftrightarrow$ (c) This follows from the variational characterization

$$\sigma_1(A) = \max_{x\in S^{m-1}} \|Ax\| \quad \text{and} \quad \sigma_m(A) = \min_{x\in S^{m-1}} \|Ax\|.$$

$\square$

Recall from Lecture 8 that

$$|t^2 - 1| \le \delta \vee \delta^2 \Rightarrow |t-1| \le \delta \quad \forall\, t, \delta \ge 0.$$

Thus, to prove the result it suffices to show that (check!)

(iii) $\left\| \frac{1}{m} A^TA - I \right\|_{op} \le K^2(\delta \vee \delta^2)$ with $\delta = C\left(\sqrt{\frac{m}{n}} + \frac{t}{\sqrt{n}}\right)$.

The proof follows in three steps.

Step 1: Approximation. Using Corollary 0 from Lecture 14 allows us to construct an $\frac{1}{4}$-net $\mathcal{N}$ of S^{m-1} with cardinality $|\mathcal{N}| \le 9^m$. Then, by Lemma (Max) we have

$$\left\| \frac{1}{n} A^T A - I \right\|_{op} \le 2 \max_{x \in \mathcal{N}} \left| \left\langle \left(\frac{1}{n} A A^T - I \right) x, x \right\rangle \right|$$

$$= 2 \max_{x \in \mathcal{N}} \left| \frac{1}{n} \|Ax\|_2^2 - 1 \right|$$

So to prove (ii) it suffices to show

$$\max_{x \in \mathcal{N}} \left| \frac{1}{n} \|Ax\|_2^2 - 1 \right| \le \frac{\varepsilon}{2} \quad \text{with } \varepsilon = K^2 (\delta \vee \delta^2).$$

Step 2: Concentration. Fix $x \in \mathcal{N}$ and express $\|Ax\|_2^2$ as

$$\|Ax\|_2^2 = \sum_{i=1}^{n} \underbrace{\langle A_i, X \rangle^2}_{X_i}.$$

By our assumptions, X_i are independent sub-Gaussian random variables with $\mathbb{E} X_i^2 = 1$ and and $\|X_i\|_{\psi_2} \le K$ (Check!). This makes $X_i^2 - 1$ independent, mean zero subexponential random variables with $\|X_i^2 - 1\|_{\psi_1} \leq CK^2$.

By Berenstein's inequality we get

$$\mathbb{P}\left(\left|\frac{1}{n}\|Ax\|_2^2 - 1\right| \geq \frac{\varepsilon}{2}\right)$$

$$= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^2 - 1\right| \geq \frac{\varepsilon}{2}\right)$$

$$\leq 2\exp\left(-c_1\left(\frac{\varepsilon^2}{K^4} \wedge \frac{\varepsilon}{K^2}\right)n\right)$$

$$= 2\exp\left(-c_1 \delta^2 n\right) \quad \leftarrow \text{Since } \frac{\varepsilon}{K^2} = \delta \vee \delta^2$$

$$\leq 2\exp\left(-c_1 C^2 (n + t^2)\right). \quad \leftarrow (a+b)^2 \geq a^2 + b^2 \;\; \forall\, a, b \geq 0.$$

Step 3: Union bound. Now we can bootstrap the result for fixed x to cover all x. Recall that $|\mathcal{N}| \leq 9^n$, then

$$\mathbb{P}\left(\max_{x \in \mathcal{N}} \left|\frac{1}{n}\|Ax\|_2^2 - 1\right| \geq \frac{\varepsilon}{2}\right)$$

$$\leq 9^n \cdot 2\exp\left(-c_1 C^2 (n + t^2)\right)$$

$$= 2\exp\left(n \log 9 - c_1 C^2 (n + t^2)\right)$$

$$\leq 2\exp\left(n(\log 9 - c_1 C^2) - t^2\right) \quad \leftarrow C \geq 1.$$

$$\leq 2\exp(-t^2) \quad \leftarrow C^2 \geq \frac{\log 9}{c_1}.$$

Since have the freedom to pick c, we ensure

$$c^2 \geq 1 \vee \frac{\log 9}{c_1},$$

which concludes the argument. $\square$

Remark: 1) Note that we never computed expectations of norms for this proof. Nonetheless, we can use this high probability bounds to derive

$$\mathbb{E}\,\|A\|_{op} \leq CK(\sqrt{n}+\sqrt{m}) \quad \text{and}$$

$$\mathbb{E}\,\left\|\frac{1}{n}A^T A - I\right\| \leq CK^2\left(\sqrt{\frac{n}{m}}+\frac{n}{m}\right).$$

2) The bound on the operator norm is optimal. See Exercise 4.42 of Vershynin's.

3) Something remarkable about this result is that we only need independence among rows (the entries per row might depend on each other).