Recall

- Norm Concentration
  - Random matrix $A \in \mathbb{R}^{m \times n}$
    (Independent, mean zero, isotropic, sub-gaussian rows)

    Fix $x \in \mathbb{R}^n$, $\quad \mathbb{E}\,\|Ax\| \approx \sqrt{m}\cdot\|x\|$

    $Ax = \begin{pmatrix} \vdots \\ A_i^T x \\ \vdots \end{pmatrix} \in \mathbb{R}^m$

- Standard gaussian $g \sim N(0, I_n)$

  $\mathbb{E}\,\|g\| \sim \sqrt{n}$

- JL Lemma. Rand Matrix $A \in \mathbb{R}^{m \times n}$. finite set $T \subset \mathbb{R}^n$

$$\mathbb{E}\left[\sup_{z \in T} \left|\|Az\| - \sqrt{m}\|z\|\right|\right] \leq CK^2 \sqrt{\log |T|}$$

Matrix Deviation Inequality

THM: Let A be an $m \times n$ matrix whose rows A_i are ind, isotropic, and sub-gaussian random vectors in $\mathbb{R}^n$. Then, for any subset $T \subset \mathbb{R}^n$, we have

$$\mathbb{E} \sup_{x \in T} \left|\|Ax\| - \sqrt{m}\|x\|\right| \leq CK^2\, \gamma(T).$$

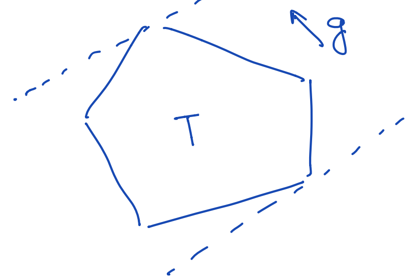
Here, $\gamma(T)$ is the gaussian width of T and $K = \max_i \|A_i\|_{\psi_2}$.

[c.f. Chapter 9, Vershynin's Book
Proof via Talagrand's Comparison Ineq.]

Defn: The gaussian width of a set $T \subset \mathbb{R}^n$ is defined as

$$\gamma(T) := \mathbb{E} \sup_{x \in T} |\langle g, x\rangle|.$$

$g \sim N(0, I_n)$.



Examples

① Unit Sphere S^{n-1}

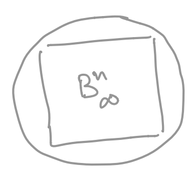
$$\gamma(S^{n-1}) := \mathbb{E} \sup_{x \in S^{n-1}} |\langle g, x\rangle|$$

$$= \mathbb{E} \|g\| = \sqrt{n} \pm C$$

② Cube, $B_\infty^n := [-1, 1]^n$ (unit ball wrt ℓ_∞ norm)

$$\gamma(B_\infty^n) = \mathbb{E} \sup_{x \in B_\infty^n} |\langle g, x\rangle|$$

$$= \mathbb{E} \|g\|_1 = \mathbb{E}\, n. |g_1| = n . \sqrt{\frac{2}{\pi}}$$



③ ℓ_1 ball $B_1^n := \{ x \in \mathbb{R}^n : \|x\|_1 \le 1 \}$

$$\gamma(B_1^n) = \mathbb{E} \sup_{x \in B_1^n} |\langle g, x\rangle|$$

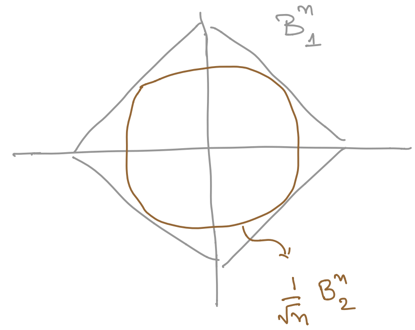
$$= \mathbb{E} \|g\|_\infty = \mathbb{E} \max_i |g_i|$$

$$\approx \sqrt{\log n}$$

④ Finite Point Set $T \subset \mathbb{R}^n$

$$\gamma(T) \leq C \cdot \sqrt{\log |T|} \cdot \mathrm{rad}(T)$$

radius $\left[\mathrm{rad}(T) := \max_{x \in T} \|x\|_2 \right]$



- Upto a log factor the gaussian width of these two sets are the same.
- The ℓ_1 ball, B_1^n, has only $2n$ vertices. The bulk of B_1^n lies within the inscribed ball
- Similar obsv holds even from a volumetric perspective.

Properties of Gaussian Width

1. If T is bounded, then $\gamma(T)$ is finite

2. U be any orthogonal matrix, then
$$\gamma(UT) = \gamma(T)$$

3. Gaussian width is invariant under taking convex hulls
$$\gamma(\mathrm{conv}(T)) = \gamma(T)$$

4. Gaussian width respects Minkowski sum of sets & scaling.
For any $T, S \subset \mathbb{R}^n$ and $a \in \mathbb{R}$ we have
$$\gamma(T+S) = \gamma(T) + \gamma(S)$$
$$\gamma(aT) = |a| \, \gamma(T)$$

5. Gaussian width and radius (diameter)

For any $T \subset \mathbb{R}^n$

$$\frac{1}{\sqrt{2\pi}} \operatorname{rad}(T) \leq \gamma(T) \leq \sqrt{n} \cdot \operatorname{rad}(T)$$

$\left[\operatorname{rad}(T) := \sup_{x \in T} \|x\|_2 \right]$

$$\gamma(T) = \mathbb{E} \sup_{x \in T} |\langle g, x \rangle|$$

$$\leq \sup_{x \in T} \mathbb{E} |\langle g, x \rangle|$$

[Jensen's Ineq]

$$= \sup_{x \in T} \sqrt{\frac{2}{\pi}} \cdot \|x\|$$

$\langle g, x \rangle \sim N(0, \|x\|^2)$

$\mathbb{E} |\langle g, x \rangle| = \sqrt{\frac{2}{\pi}} \cdot \|x\|$

$$= \sqrt{\frac{2}{\pi}} \cdot \operatorname{rad}(T)$$

Upper Bound

$$\gamma(T) = \mathbb{E} \sup_{x \in T} |\langle g, x \rangle|$$

[Cauchy Schwarz]

$$\leq \mathbb{E} \sup_{x \in T} \|g\| \cdot \|x\|$$

$$= \operatorname{rad}(T) \cdot \mathbb{E} \|g\|$$

$$= \operatorname{rad}(T) \sqrt{n}$$

Matrix Deviation Ineq: Tail Bound

THM: Under the conditions of the previous theorem, for any $t \geq 0$, the event

$$\sup_{x \in T} \left| \|Ax\| - \sqrt{m}\, \|x\| \right| \leq CK^2 \left(\gamma(T) + t\, \mathrm{rad}(T) \right)$$

holds with prob. at least $1 - 2\exp(-t^2)$

Applications of Matrix Deviation Inequality

- JL lemma

For finite $T \subset \mathbb{R}^n$

$\gamma(T) = \sqrt{\log |T|}$

- Spectra of Random Matrices.

THM: Let A be a $m \times n$ matrix whose rows A_i are ind, mean zero, isotropic, sub-gaussian rand. vectors in $\mathbb{R}^n$. Then, with high prob., we have

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

Here, $K := \max_i \|A_i\|_{\psi_2}$

$\sigma_1(A)$ = largest singular value of A.

$\|A\| = \sigma_1(A)$ Spectral Norm.

$= \max_{x \in S^{n-1}} \|Ax\|$

Largest Eigenvalue if A is symmetric

Proof: Spectral norm

$\sigma_1(A) = \max_{x \in S^{n-1}} \|Ax\|$

MDI over the sphere.

$$\mathbb{E} \sup_{x \in S^{n-1}} \left| \|Ax\| - \sqrt{m} \cdot \underbrace{1}_{\|x\|} \right| \leq CK^2 \cdot \underbrace{\sqrt{n}}_{\gamma(S^{n-1})}$$

Markov's ineq. ensures that, with const. prob.,

$$\sqrt{m} - CK^2\sqrt{n} \leq \sup_{x \in S^{n-1}} \|Ax\| \leq \sqrt{m} + CK^2\sqrt{n}$$