Chapter 6

Gaussian Processes

- Concentration of Lipschitz functions of the uniform spherical random vector
- Basics of Gaussian processes.
- Slepian's Inequality.
- Sudakov-Fernique Inequality
- Sudakov's Minoration Inequality
- Gaussian Width and Random Projections
- Chaining and Dudley's Integral
- Improved Uniform Laws
- Generic Chaining and Talagrand's comparison inequality
- Chevet Inequality

In this chapter, we explore a number of constructions built from gaussians. We begin with the following

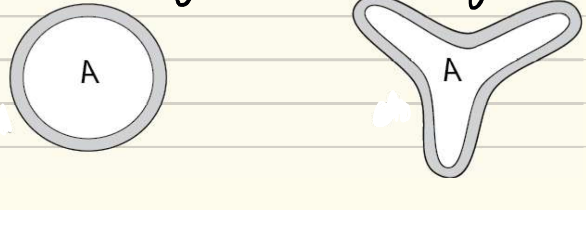
Thm ⊛: Let $f: \sqrt{n}\, S^{n-1} \to \mathbb{R}$ be Lipschitz with constant L_f and let $x \sim \mathrm{Unif}(\sqrt{n}\, S^{n-1})$. Then

$$\mathbb{P}\left[\,|f(x) - \mathbb{E}f(x)| \geq t\,\right] \leq 2\exp\left(-\frac{ct^2}{L_f^2}\right)$$

The proof will be based on the isoparametric inequality

Thm: Let $\varepsilon > 0$. Then among all sets $A \in S^{n-1}$ with prescribed area $\mu^{n-1}(A)$, the spherical caps minimize the area of the blow-up

$$A_\varepsilon = \{x \in S^{n-1} : \exists y \in A \text{ s.t. } \|x-y\|_2 \leq \varepsilon\}$$



Recall that $\sqrt{n}\,S^{n-1}$ is a natural scaling of the sphere because

$\underline{X} \sim \mathrm{Unif}(\sqrt{n}\,S^{n-1})$ is isotropic

and C-subGaussian for a constant C.

[Vershynin Thm 3.4.6: $\underline{X} \sim \sqrt{n}\,\frac{g}{\|g\|_2}$ where $g \sim N(0, I)$ and $\|g\|_2 \sim \sqrt{n}$ w.h.p.]

<u>Lemma</u>: Fix $A \subset \sqrt{n}\,S^{n-1}$ with $\mu^{n-1}(A) \geq \frac{1}{2}$ where μ^{n-1} denotes the normalized area on $\sqrt{n}\,S^{n-1}$. Then

$$\mu^{n-1}(A_t) \geq 1 - 2e^{-ct^2} \qquad \forall t \geq 0.$$

pf: Define

$$H := \{x \in \sqrt{n}\,S^{n-1} : x_1 \leq 0\}$$

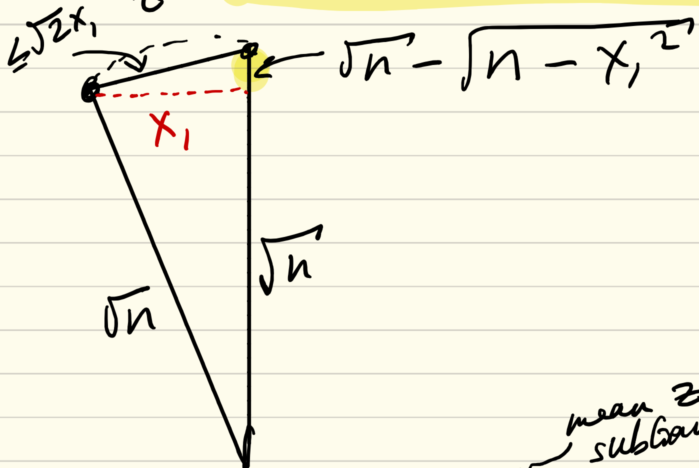
By assumption $\mu^{n-1}(A) \geq \frac{1}{2} = \mu^{n-1}(H)$

$\Rightarrow \mu^{n-1}(A_t) \geq \mu^{n-1}(H_t) \leftarrow$ let's compute

$$\mu^{n-1}(H_t) = \mathbb{P}[X \in H_t]$$

Notice

$$H_t \supset \left\{ x \in \sqrt{n}\, S^{n-1} : x_1 \le \frac{t}{\sqrt{2}} \right\}$$



mean zero subGaussian

$$\Rightarrow \mu^{n-1}(H_t) \ge \mathbb{P}\left[X_1 \le \frac{t}{\sqrt{2}}\right] \ge 1 - 2\exp(-ct^2)$$ $\boxtimes$

pf of Thm ⊛:

Let M denote a median of $f(X)$, namely

$$\mathbb{P}[f(X) \le M] \ge \frac{1}{2} \text{ and } \mathbb{P}[f(X) \ge M] \le \frac{1}{2}$$

Define

$$A := \{x \in \sqrt{n}\, S^{n-1} : f(x) \leq M\}$$

Since $\mathbb{P}[X \in A] \geq \frac{1}{2}$, we just showed

$$\mathbb{P}[X \in A_t] \geq 1 - 2\exp(-ct^2)$$

Notice

$$\mathbb{P}[X \in A_t] \leq \mathbb{P}[f(X) \leq M + L_f t]$$

since for $\forall x \in A_t \; \exists y \in A$ s.t.

$$f(x) \leq f(y) + L_f \|x-y\| \leq M + L_f t$$

So $\mathbb{P}[f(X) \leq M + L_f t] \leq 1 - 2\exp(-ct^2)$

Replace f by $-f$ proves the lower deviation inequality. Finally, the centering inequality (Vershynin 2.6.8)

$$\Rightarrow \|f(X) - \mathbb{E} f(X)\|_{\psi_2} = \|(f(X) - M) - \mathbb{E}(f(X) - M)\|_{\psi_2}$$

$$\leq C \|f(X) - M\|_{\psi_2} \leq C L_f$$

$\square$

Cor: If $f: S^{n-1} \to \mathbb{R}$ is Lipschitz and $X \sim \mathrm{Unif}(S^{n-1})$, then

$$\mathbb{P}[|f(X) - \mathbb{E} f(X)| \geq t] \leq 2\exp\left(-\frac{cnt^2}{L_f^2}\right)$$

This result extends to many other settings

Thm: Let $\varepsilon > 0$. Then among all sets $A \in \mathbb{R}^n$ with prescribed Gaussian measure $\gamma_n(A)$, the half-spaces minimize the Gaussian measure $\gamma_n(A_\varepsilon)$

Thm: (HW) Consider a $X \sim N(0, I)$ and a Lipschitz $f: \mathbb{R}^n \to \mathbb{R}$. Then

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

See Vershynin 5.2 for other examples of probability spaces on which functions concentrate.

Typical results require $\mathbb{X}$ to be very symmetric. One important exception is the following.

<u>Thm</u> (Talagrand)

Consider a random $\mathbb{X} = (X_1, \dots, X_n) \in \mathbb{R}^n$ whose coordinates are independent and satisfy $|\mathbb{X}_i| \leq 1$ almost surely. Then for any <u>convex Lipschitz</u> $f : [0,1]^n \to \mathbb{R}$ it holds

$$\mathbb{P}\left[\left| f(\mathbb{X}) - \mathbb{E} f(\mathbb{X}) \right| \geq t \right] \leq 2 \exp\left(- \frac{c n t^2}{L_f^2} \right)$$

[See Boucheron, Lugosi, Massart for a proof]

Defn: A random process is a collection of random variables $\{X_t : t \in T\}$ indexed by some set $t \in T$.

Let $\{X_t\}_{t\in T}$ be a random process. Assume for simplicity $\mathbb{E}X_t = 0 \ \forall_{t \in T}$. Define the covariance function

$$\Sigma(t,s) := \operatorname{cov}(X_t, X_s) = \mathbb{E}X_t X_s$$

and the increments

$$d(t,s) := \sqrt{\mathbb{E}(X_t - X_s)^2}$$

Rem: $d(\cdot,\cdot)$ is a pseudometric on T.

Rem: $d^2(t,s) = \Sigma(t,t) + \Sigma(s,s) - 2\Sigma(t,s)$

<u>Defn</u>: $\{X_t\}_{t\in T}$ is called a <u>Gaussian process</u> if for any finite $T_0 \subset T$, the random vector $(X_t)_{t\in T_0}$ has a Gaussian distribution.

Main example is the <u>canonical Gaussian Process indexed by $T \subseteq \mathbb{R}^n$</u>:

$$X_t = \langle g, t\rangle \quad \forall t \in T$$

where $g \sim N(0, I)$.

Notice

$$d(t,s) = \sqrt{\mathbb{E}\langle g, t-s\rangle^2} = \sqrt{\mathbb{E}\,\mathrm{tr}\left((t-s)(t-s)^T g g^T\right)} = \|t-s\|_2$$

Our goal is to estimate

$$\mathbb{E} \sup_{t\in T} X_t$$

or more formally

$$\sup_{T_0 \subset T \text{ finite}} \mathbb{E} \sup_{t\in T_0} X_t.$$

<u>Thm</u> (Slepian's inequality)

Let $\{X_t\}_{t\in T}$, $\{Y_t\}_{t\in T}$ be mean-zero Gaussian processes. Assume $\forall t,s\in T$:

$$\mathbb{E}X_t^2 = \mathbb{E}Y_t^2 \quad \text{and} \quad \mathbb{E}(X_t - X_s)^2 \leq \mathbb{E}(Y_t - Y_s)^2$$

Then

$$\mathbb{P}\left[\sup_{t\in T} X_t \geq \delta\right] \leq \mathbb{P}\left[\sup_{t\in T} Y_t \geq \delta\right] \ \forall \delta$$

Therefore

$$\mathbb{E}\sup_{t\in T} X_t \leq \mathbb{E}\sup_{t\in T} Y_t.$$

<u>Proof Strategy:</u>

Assume T is finite and set $X = (X_t)$, $Y = (Y_t)$,

$$Z(u) = \sqrt{u}\,X + \sqrt{1-u}\,Y \quad \text{for } u\in[0,1].$$

We will show that

$u \mapsto \mathbb{P}[\max_i Z_i(u) \geq \delta]$ is nonincreasing

Lemma (HW)
Let $X \sim N(0, \Sigma)$. Then for any differentiable $f: \mathbb{R}^n \to \mathbb{R}$, it holds:

$$\mathbb{E}\, X f(X) = \Sigma \cdot \mathbb{E}\, \nabla f(X)$$

or equivalently

$$\mathbb{E}\, X_i f(X) = \sum_{j=1}^{n} \Sigma_{ij}\, \mathbb{E} \frac{\partial f}{\partial X_j}(X), \quad \forall i = 1, \dots, n$$

Lemma: (Interpolation) Consider independent $X \sim N(0, \Sigma^X)$, $Y \sim N(0, \Sigma^Y)$, and set

$$Z(u) = \sqrt{u}\, X + \sqrt{1-u}\, Y \quad \text{for } u \in [0,1].$$

Then for any twice-differentiable $f: \mathbb{R}^n \to \mathbb{R}$ it holds:

$$\frac{d}{du} \mathbb{E} f(Z(u)) = \frac{1}{2} \left\langle \Sigma^X - \Sigma^Y, \mathbb{E}\, \nabla^2 f(Z(u)) \right\rangle \quad \forall u \in (0,1)$$

pf: Chain rule:

$$\frac{d}{du} \mathbb{E} f(Z(u)) = \mathbb{E} \left\langle \frac{d}{du} Z(u), \nabla f(Z(u)) \right\rangle$$

$$= \frac{1}{2\sqrt{u}} \underbrace{\mathbb{E} \langle X, \nabla f(Z(u)) \rangle}_{=(1)} - \frac{1}{2\sqrt{1-u}} \underbrace{\mathbb{E} \langle Y, \nabla f(Z(u)) \rangle}_{=(2)}$$

$$① = \sum_{i=1}^{n} \mathbb{E} X_i \frac{\partial f}{\partial x_i}(\sqrt{u}\, X + \sqrt{1-u}\, Y)$$

$$= \mathbb{E}_Y \sum_{i=1}^{n} \mathbb{E}_X\, X_i \frac{\partial f}{\partial x_i}(\sqrt{u}\, X + \sqrt{1-u}\, Y)$$

$$= \mathbb{E}_Y \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \Sigma_{ij}^{X}\, \mathbb{E}_X \sqrt{u} \cdot \frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u)) \right)$$

$$= \sqrt{u}\; \mathbb{E} \langle \Sigma^X, \nabla^2 f(Z(u)) \rangle$$

Similar computation gives

$$② = \sqrt{1-u}\; \mathbb{E} \langle \Sigma^Y, \nabla^2 f(Z(u)) \rangle \quad \square$$

<u>Lemma</u>: Let $X \sim N(0, \Sigma^X)$, $Y \sim N(0, \Sigma^Y)$ be independent. Assume

$$\mathbb{E} X_i^2 = \mathbb{E} Y_i^2, \quad \mathbb{E}(X_i - X_j)^2 \le \mathbb{E}(Y_i - Y_j)^2 \quad \forall i,j$$

Consider $f: \mathbb{R}^n \to \mathbb{R}$ with $[\nabla^2 f]_{ij} \ge 0 \;\; \forall i \ne j$

Then
$$\mathbb{E} f(X) \ge \mathbb{E} f(Y)$$

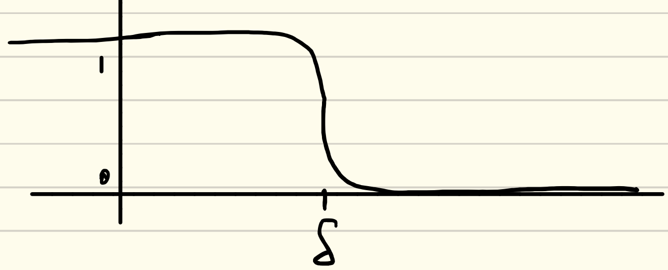
pf: Notice
$$\Sigma_{ii}^X = \Sigma_{ii}^Y \quad \text{and} \quad \Sigma_{ij}^X \geq \Sigma_{ij}^Y$$
We can assume X and Y are independent. Apply the previous lemma. $\boxtimes$

Finally, we are ready to prove Slepian's inequality.

<u>Thm</u> (Slepian's inequality)

Let $\{X_t\}_{t\in T}$, $\{Y_t\}_{t\in T}$ be mean-zero Gaussian processes. Assume $\forall t,s \in T$:
$$\mathbb{E} X_t^2 = \mathbb{E} Y_t^2 \quad \text{and} \quad \mathbb{E}(X_t - X_s)^2 \leq \mathbb{E}(Y_t - Y_s)^2$$
Then
$$\mathbb{P}\left[\sup_{t\in T} X_t \geq \delta\right] \leq \mathbb{P}\left[\sup_{t\in T} Y_t \geq \delta\right] \quad \forall \delta$$
Therefore $\mathbb{E} \sup_{t\in T} X_t \leq \mathbb{E} \sup_{t\in T} Y_t$.

pf: Let $h : \mathbb{R} \to [0,1]$ be the function



We can ensure h is a C^2-smooth approximation to $\mathbb{1}_{(-\infty, \delta)}$.

Define $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x) = h(x_1) \cdots h(x_n)$

[Note f approximates $\mathbb{1}_{\{\max_i x_i \le \delta\}}$]

Note
$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \underbrace{h'(x_i)}_{\le 0} \cdot \underbrace{h'(x_j)}_{\le 0} \underbrace{\prod_{k \notin \{i,j\}} h(x_k)}_{\ge 0} \ge 0$$

So
$$\mathbb{E} f(X) \ge \mathbb{E} f(Y)$$

Pass to the limit [e.g. dominated convergence]

$\boxtimes$

Thm (Sudakov-Fernique)

Let $\{X_t\}_{t\in T}$, $\{Y_t\}_{t\in T}$ be two mean-zero Gaussian processes. Assume

$$\mathbb{E}(X_t - X_s)^2 \le \mathbb{E}(Y_t - Y_s)^2 \quad \forall_{s,t\in T}.$$

Then

$$\mathbb{E}\sup_{t\in T} X_t \le \mathbb{E}\sup_{t\in T} Y_t$$

pf: As before assume $X \sim N(0,\Sigma^X)$
$Y \sim N(0,\Sigma^Y)$

Fix $\beta > 0$ and define

$$f_\beta(x) = \frac{1}{\beta}\log\sum_{i=1}^n e^{\beta x_i}$$

You can check

$$\max_{i=1,\dots,n} x_i \le f_\beta(x) \le \max_{i=1,\dots,n} x_i + \frac{\log n}{\beta}$$

Apply Gaussian interpolation to deduce

$$\frac{d}{du}\mathbb{E} f_\beta(Z(u)) \le 0 \quad (HW)$$

$\square$

Cor: (Gaussian Contraction Inequality)
Fix a set $T \subset \mathbb{R}^n$ and let $g_1, \ldots, g_n \overset{iid}{\sim} N(0,1)$
Let $\phi_i : \mathbb{R} \to \mathbb{R}$ be 1-Lipschitz.
Then

$$\mathbb{E} \sup_{t \in T} \sum_{i=1}^{n} g_i \phi_i(t_i) \leq \mathbb{E} \sup_{t \in T} \sum_{i=1}^{n} g_i t_i$$

pf: Apply Sudakov-Fernique. $\blacksquare$

We will now use Sudakov-Fernique to prove a useful lower bound on

$$\sup_{t \in T} X_t$$

Recall: $d(t,s) = \sqrt{\mathbb{E}(X_t - X_s)^2}$
defines a pseudo-metric on T.

Thm: (Sudakov's minoration inequality)
Let $\{X_t\}_{t\in T}$ be a mean-zero Gaussian process. Then

$$\mathbb{E}\sup_{t\in T} X_t \geq C\varepsilon\sqrt{\log_2 N(T,d,\varepsilon)} \quad \forall_{\varepsilon>0},$$

where $N(T,d,\varepsilon)$ denotes the covering number of T is $d(\cdot,\cdot)$ metric.

pf: Suppose $N := N(T,d,\varepsilon)$ is finite. You'll consider the infinite case for HW. Let M be a maximal ε-packing. Then we know $|M| \geq N$. It suffices to show

$$\mathbb{E}\sup_{t\in M} X_t \geq C\varepsilon\sqrt{\log N}$$

Define $\{Y_t\}_{t\in M}$ by $Y_t = \frac{\varepsilon}{\sqrt{2}} g_t$ where $g_t \sim N(0,1)$ independent.

Fix $t, s \in M$. Then

$$\mathbb{E}(X_t - X_s)^2 = d(t,s)^2 \geq \varepsilon^2$$

while

$$\mathbb{E}(Y_t - Y_s)^2 = \frac{\varepsilon^2}{2}\mathbb{E}(g_t - g_s)^2 = \varepsilon^2$$

Therefore Sudakov-Fernique

$$\Rightarrow \mathbb{E}\sup_{t \in M} X_t \geq \mathbb{E}\sup_{t \in M} Y_t$$

HW1

$$= \frac{\varepsilon}{\sqrt{2}}\mathbb{E}\max_{t \in M} g_t \geq C\varepsilon\sqrt{\log M}$$

We next study a particular type of a Gaussian process, which is an analogue of the Rademacher complexity.

<u>Defn</u>: The <u>Gaussian width</u> of a set $T \subseteq \mathbb{R}^n$ is $w(T) := \mathbb{E}\sup_{x \in T}\langle g, x\rangle$

where $g \sim N(0, I)$.

Properties:

(1) $w(T) < \infty$ iff T is bounded.

(2) Affine Invariance:

$$w(UT+y) = w(T)$$

for any orthogonal $U \in \mathbb{R}^{n \times n}$ and $y \in \mathbb{R}^n$.

(3) $w(\mathrm{conv}(T)) = w(T)$

(4) $w(T+S) = w(T) + w(S)$

$$w(aT) = |a| w(T) \qquad \forall a \in \mathbb{R}.$$

(5) $w(T) = \frac{1}{2} w(T-T) = \frac{1}{2} \mathbb{E} \sup_{x,y \in T} \langle g, x-y \rangle$

(6) $\frac{1}{\sqrt{2\pi}} \mathrm{diam}(T) \leq w(T) \leq \frac{\sqrt{n}}{2} \mathrm{diam}(T)$

(7) For any 1-Lipschitz $\varphi_i : \mathbb{R} \to \mathbb{R}$ it holds

$$w(\varphi \circ T) \leq w(T)$$

pf: (1)–(4) follow from defn of support functions. To see (5) observe

$$w(T) = \frac{1}{2}(w(T) + w(T)) = \frac{1}{2}(w(T) + w(-T)) \overset{(4)}{=} \frac{1}{2} w(T-T).$$

To see (6), fix $x, y \in T$. Then

$$\omega(T) \geq \frac{1}{2} \mathbb{E} \max(\langle x-y, g\rangle, \langle y-x, g\rangle)$$

$$= \frac{1}{2} \mathbb{E} |\langle x-y, g\rangle| = \frac{1}{2}\sqrt{\frac{2}{\pi}} \|x-y\|$$

Take $\sup_{x,y \in T}$.

Next

$$\omega(T) = \frac{1}{2} \mathbb{E}_g \sup_{x,y \in T} \langle g, x-y\rangle$$

$$\leq \frac{1}{2} \sup_{x,y\in T} \|x-y\|_2 \cdot \mathbb{E}_g \|g\| \leq \frac{\sqrt{n}}{2} \operatorname{diam}(T)$$

(7) is the contraction inequality we already proved.

<u>Lemma</u>: For any $A \in \mathbb{R}^{m \times n}$, it holds

$$\omega(AT) \leq \|A\|_2\, \omega(T)$$

[Hint: singular value decomposition + concraction]

<u>Thm:</u> (Gaussian vs Rademacher Complexity)
For any set $T \subseteq \mathbb{R}^d$, it holds:

$$\frac{\omega(T)}{2\sqrt{\log d}} \leq \mathcal{R}(T) \leq \sqrt{\frac{\pi}{2}}\,\omega(T)$$

<u>pf:</u>
$$\mathcal{R}(T) = \mathbb{E}_{\varepsilon} \sup_{x \in T} \sum_{i=1}^{d} \varepsilon_i x_i$$
$$= \sqrt{\frac{\pi}{2}} \cdot \mathbb{E}_{\varepsilon} \sup_{x \in T} \sum_{i=1}^{d} \varepsilon_i \mathbb{E}|g_i| x_i$$
$$\leq \sqrt{\frac{\pi}{2}} \cdot \mathbb{E}_{\varepsilon, g} \sup_{x \in T} \sum_{i=1}^{d} \varepsilon_i |g_i| x_i$$

$\varepsilon_i |g_i| \overset{\mathcal{L}}{=} g_i$
$$= \sqrt{\frac{\pi}{2}} \cdot \mathbb{E}_{g} \sup_{x \in T} \sum_{i=1}^{d} g_i x_i$$
$$= \sqrt{\frac{\pi}{2}}\,\omega(T)$$

Conversely
$$\omega(T) = \mathbb{E}_g \sup_{x \in T} \sum_{i=1}^{d} g_i x_i$$
$$= \mathbb{E}_{g,\varepsilon} \sup_{x \in T} \sum_{i=1}^{d} \varepsilon_i g_i x_i$$

$$= \mathbb{E}_g \mathcal{R}(g \circ T) \overset{\text{contraction}}{\leq} \mathbb{E}_g \max_{i=1,\dots,d} |g_i| \mathcal{R}(T)$$

$$\leq 2\sqrt{\log d}\, \mathcal{R}(T) \quad \boxtimes$$

There is a close cousin of Gaussian width called spherical width.

<u>Defn</u>: The <u>spherical width</u> of $T \in \mathbb{R}^d$ is
$$s(T) = \mathbb{E}_\theta \sup_{x \in T} \langle \theta, x \rangle$$
where $\theta \sim \mathrm{Unif}(S^{n-1})$

<u>Lemma</u>: $(\sqrt{n} - C)\, s(T) \leq w(T) \leq (\sqrt{n} + C)\, s(T)$

<u>pf</u>: $w(T) = \mathbb{E}_g \sup_{x \in T} \langle g, x \rangle = \mathbb{E}_g \sup_{x \in T} \langle \|g\| \cdot \frac{g}{\|g\|}, x \rangle$
$$= \mathbb{E}_g \|g\|_2 \cdot s(T)$$

Let's compare the sizes of $\mathbb{B}_2^d, \mathbb{B}_1^d, \mathbb{B}_\infty^d$.

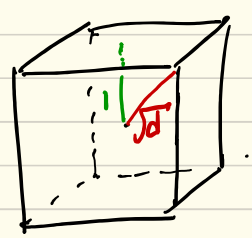
	$\log N(\cdot, \delta)$	$\omega(\cdot)$	R_I	R_C	# vertices
$\mathbb{B}_2^d$	$d \log(\frac{1}{\delta})$	$\sqrt{d}$	—	—	—
$\mathbb{B}_1^d$	$\lesssim \frac{\log d}{\delta^2}$ ②	$\sqrt{\log d}$ ①	$\frac{1}{\sqrt{d}}$	1	d
$\mathbb{B}_\infty^d$	$\lesssim \frac{d^3}{\delta^2}$ ④	d ③	1	$\sqrt{d}$	2^d

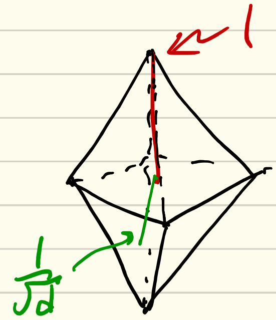
pf: ① $\mathbb{E}_g \sup_{x \in \mathbb{B}_1} \langle g, x \rangle = \mathbb{E}_g \|g\|_\infty \simeq \sqrt{\log d}$

② Sudakov $\Rightarrow \omega(\mathbb{B}_1^d) \gtrsim \max_\delta C \delta \sqrt{\log N(\mathbb{B}_1^d, \delta)}$

③ $\omega(\mathbb{B}_\infty^d) = \mathbb{E}_g \|g\|_1 = \sqrt{\frac{2}{\pi}}\, d$

④ Sudakov





$\omega(\mathbb{B}_\infty) \approx \omega(\text{circumscribed ball})$ $\quad \omega(\mathbb{B}_1) \approx \omega(\text{inscribed ball})$

Algebraic Dimension is highly unstable

Defn: Define the stable dimension of T by

$$d(T) = \frac{h(T-T)^2}{\mathrm{diam}(T)^2} \sim \frac{w(T)^2}{\mathrm{diam}(T)^2}$$

where $h(S) = \sqrt{\mathbb{E}_g \sup_{x \in S} \langle g, x\rangle^2}$

Lemma: (HW)

$$2w(T) \leq h(T-T) \leq 2C\,w(T)$$

Lemma: $d(T) \leq \dim(\mathrm{span}\, T)$

pf: We can assume $T \subseteq \mathrm{Span}\{e_1, \dots, e_k\}$ where $k = \dim(T)$. Then

$$h(T-T)^2 = \mathbb{E}_g \sup_{x \in T-T} \langle g, x\rangle^2 \leq \mathbb{E}_{g \sim N(0, I_k)} \|g\|^2 \cdot \mathrm{diam}(T)^2.$$

$$= k\, \mathrm{diam}(T)^k \quad \blacksquare$$

<u>Exercise</u> (HW): Let $A \in \mathbb{R}^{m \times n}$. Then

$$d(AB_2^d) = \frac{\|A\|_F^2}{\|A\|_2^2}$$

- Stable rank of matrix
- changes gradually with perturbations

<u>Random Projections of Sets</u>

<u>Thm:</u> Let $T \subset \mathbb{R}^n$ and let P be a projection onto a uniformly random m-dimensional subspace. Then w.p. $1-2e^{-m}$, we have

$$\operatorname{diam}(PT) \leq C\left(\delta(T) + \sqrt{\frac{m}{n}}\operatorname{diam}(T)\right)$$

Lemma (HW): Let P be a projection in $\mathbb{R}^n$ onto a uniformly random m-dimensional subspace. Let Q be an $m \times n$ orthogonal matrix drawn uniformly. Then

(a) For any $x \in \mathbb{R}^n$, it holds:

$\|Px\|_2$ and $\|Qx\|_2$ have the same distribution.

(b) For any z with $\|z\|=1$, it holds

$$Q^T z \sim \mathrm{Unif}(S^{n-1})$$

pf of thm: Covering argument. Can replace P by Q by Lemma. Can assume $\mathrm{Diam}(T) \le 1$

Approximation: Let $\mathcal{N}$ be $\frac{1}{2}$-net of S^{m-1}.
Then $|\mathcal{N}| \le 5^m$.

Then

$$\mathrm{diam}(QT) \le \sup_{x \in T-T} \|Qx\|_2$$

$$= \sup_{x \in T-T} \sup_{z \in S^{m-1}} \langle Qx, z \rangle$$

$$\le 2 \sup_{x \in T-T} \left[\sup_{z \in \mathcal{N}} \langle x, Q^T z \rangle \right]$$

$$\le 2 \sup_{z \in \mathcal{N}} \left[\sup_{x \in T-T} \langle x, Q^T z \rangle \right]$$

Fix $z \in \mathcal{N}$. Then $Q^T z \sim \mathrm{Unif}(S^{n-1})$.

Note: $\mathbb{E}_Q \sup_{x \in T-T} \langle x, Q^T z \rangle = s(T-T) = 2s(T)$

Concentration: The function $\theta \mapsto \sup_{x \in T-T} \langle \theta, x \rangle$ is 1-Lipschitz on S^{n-1} $\Rightarrow$ Concentration

$$\mathbb{P}\left[\sup_{x \in T-T} \langle x, Q^T z \rangle \ge 2s(T) + t \right] \le 2\exp(-cnt^2)$$

Union bound:

$$\mathbb{P}\left[\max_{z \in \mathcal{N}} \sup_{x \in T-T} \langle Q^T z, x\rangle \geq 2s(T) + t\right]$$

$$\leq 5^m \cdot 2\exp(-cnt^2)$$

$$= 2\exp(m\log(5) - cnt^2)$$

set $t = c'\sqrt{\frac{m}{n}}$

$$\leq e^{-m}.$$

$\blacksquare$

Exercise (HW) For all bounded sets:

$$\mathbb{E}\,\mathrm{diam}(PT) \geq c\left(s(T) + \sqrt{\frac{m}{n}}\,\mathrm{diam}(T)\right)$$

The phase transition:

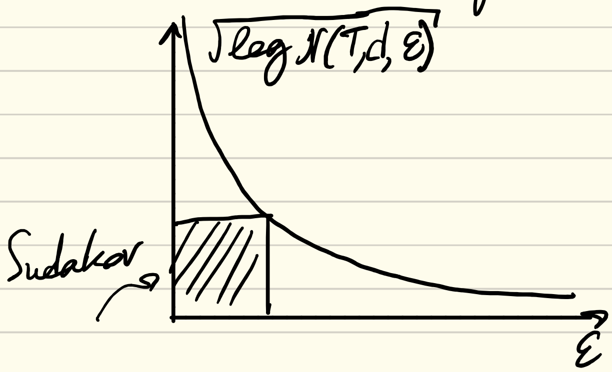
$$\mathrm{diam}(PT) \leq C\begin{cases} \sqrt{\frac{m}{n}}\,\mathrm{diam}(T), & \text{if } m \geq d(T) \\ s(T), & \text{if } m \leq d(T) \end{cases}$$

Chaining

We now prove the following upper bound

$$\mathbb{E} \sup_{t \in T} X_t \le CK \int_0^\infty \sqrt{\log \mathcal{N}(T,d,\varepsilon)}\, d\varepsilon$$

on subGaussian processes.



Defn: Consider a random process $\{X_t\}_{t \in T}$ on a metric space (T,d) if $\exists K \ge 0$ s.t.

$$\|X_t - X_s\|_{\psi_2} \le K\, d(t,s) \quad \forall t,s$$

subGaussian norm

Ex: If X_t is a canonical Gaussian process and $d(t,s) = \|t-s\|_2$, then $K=1$.

Thm: (Discrete version)

Let $\{X_t\}$ be a mean-zero subGaussian process on (T,d). Then

$$\mathbb{E}\sup_{t\in T} X_t \le CK \sum_{k\in\mathbb{Z}} 2^{-k}\sqrt{\log N(T,d,2^{-k})}$$

[Idea: Let $\pi(\cdot)$ be project on ε-net repeat

$\Rightarrow \mathbb{E}\sup_{t\in T} X_t \le \mathbb{E}\sup_{t\in T} X_{\pi(t)} + \mathbb{E}\sup_{t\in T}\left[X_t - X_{\pi_t}\right]$]

pf: WLOG assume $K=1$, T finite.

Set $\varepsilon_k = 2^{-k}$, $k\in\mathbb{Z}$

and let T_k be an ε_k-net of T.

Since T is finite $\exists\, \kappa_0, \kappa_\infty$ s.t.

$T_{\kappa_0} = \{t_0\}$ for some $t_0\in T$

$T_{\kappa_\infty} = T$

For $t\in T$, let $\pi_k(t)\in T_k$ satisfy

$$d(t,\pi_k(t)) \le \varepsilon_k .$$

Since $\mathbb{E} X_{t_0} = 0$, we can write

$$\mathbb{E} \sup_{t \in T} X_t = \mathbb{E} \sup_{t \in T} \left[X_t - X_{t_0} \right]$$

Write

$$X_t - X_{t_0} = \sum_{k=k_0+1}^{k_\infty} \left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \right)$$

So

$$\mathbb{E} \sup_{t \in T} (X_t - X_0) \le \sum_{k=k_0+1}^{k_\infty} \mathbb{E} \sup_{t \in T} \left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \right)$$

supremum is over $|T_k| \cdot |T_{k-1}| \le |T_k|^2$

For fixed t, have

$$\begin{aligned} \| X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \|_{\psi_2} &\le d(\pi_k(t), \pi_{k-1}(t)) \\ &\le d(\pi_k(t), t) + d(t, \pi_{k-1}(t)) \\ &\le \varepsilon_k + \varepsilon_{k-1} \\ &\le 2 \varepsilon_{k-1} \end{aligned}$$

So $\mathbb{E}\sup_{t\in T}\left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\right) = C\,\varepsilon_{k-1}\sqrt{\log|T_k|}$

Exercise 2.5.10. (HW)

So

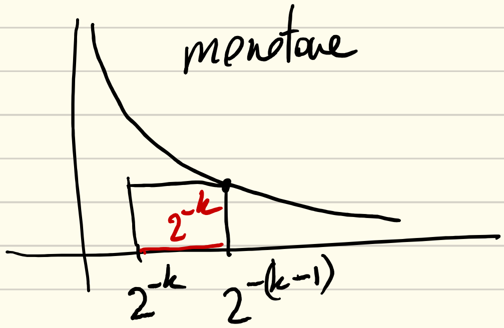
$$\mathbb{E}\sup_{t\in T} X_t - X_0 \le C_1 \sum_{k=k_0}^{k_\infty - 1} 2^{-k}\sqrt{\log N(T,d,2^{-k})}$$

$\boxed{T_k}$

Thm: (Dudley's Integral)
Let $\{X_t\}_{t\in T}$ be a mean zero random process on (T,d) with subGaussian increments. Then

$$\mathbb{E}\sup_{t\in T} X_t \le CK\int_0^\infty \sqrt{\log N(T,d,\varepsilon)}\,d\varepsilon .$$

pf:



$\boxed{2}$

Thm (HW)

Let $\{X_t\}_{t \in T}$ be a subGaussian process on (T,d), possibly not centered. Then w.p. $1-2\exp(-\delta^2)$ have:

$$\sup_{t,s \in T} |X_t - X_s| \le CK\left(\int_0^\infty \sqrt{\log N(T,d,\varepsilon)}\, d\varepsilon + \delta \cdot \mathrm{diam}(T) \right)$$

Thm: (Two-sided Sudakov)

Let $\{X_t\}_{t \ge 0}$ be a canonical Gaussian process on a set T. Define

$$\varphi(T) = \sup_{\varepsilon \ge 0} \varepsilon \sqrt{\log N(T,\varepsilon)}$$

Then

$$c_2 \varphi(T) \le W(T) \le c_1 (\log n)\, \varphi(T)$$

[Read in Vershynin]

Improved Uniform Laws

Let $\mathcal{F}$ be a class of indicator functions with VC-dimension v.
We showed

$$\mathcal{R}_n(\mathcal{F}) \leq 2\sqrt{\frac{v \log(n+1)}{n}}$$

Let's get a better bound on

$$\mathcal{R}_n(\mathcal{F}) = \mathbb{E}_x \mathbb{E}_\varepsilon \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \varepsilon_i f(x_i)$$

Define $Z_f := \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(x_i)$ and consider the process $\{Z_f\}_{f \in \mathcal{F}}$

Let's compute

$$\|Z_f - Z_g\|_{\psi_2} = \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^n \varepsilon_i \left(f(x_i) - g(x_i)\right) \right\|_{\psi_2}$$

$$\leq \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^n \left(f(x_i) - g(x_i)\right)^2} \leq \frac{1}{\sqrt{n}} \|f-g\|_2$$

So Dudley

$$\Rightarrow \mathbb{E}_{\varepsilon} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_i \varepsilon_i f(x_i) \right| \leq \frac{c}{\sqrt{n}} \int_0^2 \sqrt{\log N(\mathcal{F}, \delta, \|\cdot\|_2)}$$

<u>Thm:</u> $N(\mathcal{F}, \delta, \|\cdot\|_2) \leq \left(\frac{2}{\delta}\right)^{cv}$

[Vershynin Thm 8.3.18]

So continuing

$$\leq \frac{c}{\sqrt{n}} \int_0^2 \sqrt{v \ln\left(\frac{2}{\varepsilon}\right)}\, d\varepsilon \leq c\sqrt{\frac{v}{n}}$$

So $\boxed{\mathcal{R}_n(\mathcal{F}) \leq c\sqrt{\frac{v}{n}}}$

<u>Thm:</u> (Improved Glivenko-Cantelli)

Let $X_1, \ldots, X_n$ be iid random variables with CDF F. Define $F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \leq t\}}$

Then $\mathbb{E}\|F_n - F\|_\infty \leq \frac{c}{\sqrt{n}}$

<u>pf:</u> Let $\mathcal{F} = \{\mathbb{1}_{(-\infty, t]} : t \in \mathbb{R}\}$

$\Rightarrow vc(\mathcal{F}) = 1.$ ☒

Another Example:

Define
$$\mathcal{F} := \{ f : [0,1] \to [0,1] \mid \|f\|_{Lip} \le L \}$$

Thm: Let $X_1, \ldots, X_n$ be iid random variables in $[0,1]$. Then

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E} f(X) \right| \le \frac{CL}{\sqrt{n}}$$

pf: WLOG assume $L = 1$.

Again, let's upper bound
$$R_n(F) = \mathbb{E}_X \underbrace{\mathbb{E}_\varepsilon \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i)}$$

Define $z_f = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i)$

Then
$$\|z_f - z_g\|_{\psi_2} = \frac{1}{n} \left\| \sum_{i=1}^{n} \varepsilon_i (f(x_i) - g(x_i)) \right\|_{\psi}$$
$$\le \frac{1}{n} \sqrt{\sum_{i=1}^{n} (f(x_i) - g(x_i))^2} \le \frac{1}{\sqrt{n}} \|f - g\|_\infty$$

Dudley $\Rightarrow$

$$\mathbb{E} \sup_{f \in \mathcal{F}} |Z_f| \leq \frac{c}{\sqrt{n}} \int_0^{\overset{\text{diameter}}{1}} \sqrt{\log N(\mathcal{F}, \|\cdot\|_\infty, \varepsilon)}\, d\varepsilon.$$

Lemma (HW): $N(\mathcal{F}, \|\cdot\|_\infty, \varepsilon) \leq \left(\frac{2}{\varepsilon}\right)^{\frac{2}{\varepsilon}}$

$$\Rightarrow \mathbb{E} \sup_{f \in \mathcal{F}} |Z_f| \leq \frac{c}{\sqrt{n}} \int_0^1 \sqrt{\frac{1}{\varepsilon} \log\left(\frac{k}{\varepsilon}\right)}\, d\varepsilon$$

$$\leq \frac{\hat{c}}{\sqrt{n}}$$

$\boxtimes$

So far we know that

$$\sup_{\varepsilon > 0} \varepsilon \sqrt{\log N(T, \varepsilon, \|\cdot\|_2)} \leq \mathbb{E} \sup_{t \in T} X_t \leq K \int_0^1 \sqrt{\log N(T, \varepsilon, d)}\, d\varepsilon$$

Gaussian SubGaussian

<u>Question</u>: Is there a tighter bound that separates ① the subGaussian parameter? ② size of T.

Yes!

Recall the key bound from chaining

$$(*)\quad \mathbb{E} \sup_{t \in T} X_t \le c \sum_{k=k_0}^{k_\infty} \varepsilon_{k-1} \sqrt{\log |T_k|}$$

where T_k are ε_k-nets of T.

$\Rightarrow$ We chose ε_k followed by T_k.
Let's switch the order.

Fix subset $T_1, T_2, \dots, T_k, \dots \subset T$
such that

$$|T_0| = 1 \quad \text{and} \quad |T_k| \le 2^{2^k} \text{ for } k \ge 1$$

The sequence $(T_k)_{k=0}^{\infty}$ is called <u>admissable</u>

Define $\varepsilon_k = \sup_{t \in T} d(t, T_k)$

Then T_k is an ε_k-net of T.

$(*)$ becomes

$$\mathbb{E} \sup_{t \in T} X_t \le c \sum_{k=0}^{\infty} 2^{k/2} \sup_{t \in T} d(t, T_k)$$

__Defn__ (Talagrand's γ_2 functional)

Let (T,d) be a metric space. Define

$$\gamma_2(T,d) = \inf_{(T_k)} \sup_{t \in T} \sum_{k=0}^{\infty} 2^{k/2} d(t, T_k)$$

where $\inf_{(T_k)}$ is taken over all admissable sequences.

__Thm__ (Generic chaining bound)

Let $\{X_t\}_{t \in T}$ be a mean zero subGaussian process satisfying

$$\| X_t - X_s \|_{\psi_2} \leq K d(t,s).$$

Then

$$\mathbb{E} \sup_{t \in T} X_t \leq C K \gamma_2(T,d)$$

pf: WLOG assume $K=1$.

We do the chaining as before

$$t_0 \to \pi_0(t) \to \pi_1(t) \to \cdots \to \pi_N(t) = t$$

$$\Rightarrow X_t - X_0 = \sum_{k=1}^{N} \left(X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \right)$$

Fix k, t. We know

$$\| X_{\pi_k(t)} - X_{\pi_{k-1}(t)} \|_{\psi_2} \le d(\pi_k(t), \pi_{k-1}(t))$$

$$\Rightarrow \mathbb{P}\left[| X_{\pi_k(t)} - X_{\pi_{k-1}(t)} | \le C u 2^{k/2} d(\pi_k(t), \pi_{k-1}(t)) \right]$$

$$\ge 1 - 2\exp(-8u^2 2^k)$$

Take union bound over

$$|T_k|\,|T_{k-1}| \le |T_k|^2 = 2^{2^{k+1}} \qquad k \in \mathbb{N}$$

So get uniform bound w.p.

$$1 - \sum_{k=1}^{\infty} 2^{2^{k+1}} \cdot 2\exp(-8u^2 2^k) \overset{?}{\ge} 1 - 2\exp(-cu^2)$$

check

for $u > \hat{c}$.

In this event, get

$$|X_t - X_0| \le Cu \sum_{k=1}^{\infty} 2^{k/2} d(\pi_k(t), \pi_{k-1}(t))$$

$$\le C u \underbrace{\sum_{k=1}^{\infty} 2^{k/2} d(t, T_k)}_{\gamma_2^{\text{"}}(T,d)}$$

$$\Rightarrow \sup_{t \in T} |X_t - X_0| \le C u \, \gamma_2(T,d)$$

Now integrate the tail $\square$

Remarkably $\gamma_2(T,d)$ is also a lower bound

<u>Thm</u> (Talagrand) Let $\{X_t\}_{t \in T}$ be a <u>Gaussian Process</u> on T. Define $d(t,s) = \sqrt{\mathbb{E}\|X_t - X_s\|_2^2}$. Then

$$c\,\gamma_2(T,d) \le \mathbb{E} \sup_{t \in T} X_t \le C \gamma_2(T,d)$$

For us the most important consequence is the following generalization of Sudakov-Fernique comparison inequality.

Thm: (Talagrand Comparison Inequality) Let $\{X_t\}_{t\in T}$ be a mean-zero random process on T and let $\{Y_t\}_{t\in T}$ be a Gaussian process. Assume

$$\|X_t - X_s\|_{\psi_2} \le K \|Y_t - Y_s\|_2$$

Then

$$\mathbb{E} \sup_{t\in T} X_t \le CK\, \mathbb{E} \sup_{t\in T} Y_t.$$

pf: Define $d(t,s) = \|Y_t - Y_s\|_2$. Then

$$\mathbb{E} \sup_{t\in T} X_t \underset{\substack{\uparrow \\ \text{upper bound} \\ \text{in Talagrand}}}{\le} CK\, \gamma_2(T,d) \underset{\substack{\uparrow \\ \text{lower bound} \\ \text{in Talagrand}}}{\le} \hat{C} K\, \mathbb{E} \sup_{t\in T} Y_t. \quad \boxtimes$$

Cor: Let $\{X_t\}_{t\in T}$ be a mean-zero random process on $T \subseteq \mathbb{R}^n$. Assume

$$\|X_t - X_s\|_{\psi_2} \le K \|t-s\|_2$$

Then
$$\mathbb{E} \sup_{t\in T} X_t \le CK\, w(T)$$

pf: Define $Y_t = \langle g, t\rangle$ where $g \in N(0, I)$. Apply previous thm. $\boxtimes$

Cor: (HW) Let $\{X_t\}_{t\in T}$ be a random process on a set $T \subseteq \mathbb{R}^n$. Assume

$$\|X_t - X_s\|_{\psi_2} \le K \|t-s\|_2 \qquad \forall t, s \in T$$

Then

$$\mathbb{E} \sup_{t\in T} |X_t| \le CK\, \mathbb{E} \sup_{t\in T} |\langle g, t\rangle|$$

Thm: (SubGaussian Chevet Inequality)
Let $A \in \mathbb{R}^{m \times n}$ be a random matrix with independent mean zero K-subGaussian parameters. Let $T \in \mathbb{R}^n$, $S \subset \mathbb{R}^m$ be arbitrary bounded sets. Then

$$\mathbb{E} \sup_{x \in T, y \in S} \langle Ax, y \rangle \leq CK(w(T)\,\mathrm{rad}(S) + \mathrm{rad}(T)\, w(S))$$

pf: WLOG assume $K=1$. Define
$X_{u,v} := \langle Au, v \rangle$ for $u \in T$, $v \in S$.
Compute

$$\|X_{uv} - X_{wz}\|_{\psi_2} = \Big\| \sum_{i,j} A_{ij}(u_i v_j - w_i z_j) \Big\|_{\psi_2}$$

$$\leq \Big(\sum_{i,j} |u_i v_j - w_i z_j|^2 \Big)^{1/2}$$

$$= \| uv^T - wz^T \|_F$$

$$= \| u(v-z)^T + (u-w)z^T \|_F$$

$$\leq \|u(v-z)^T\|_F + \|(u-w)z^T\|_F$$

$$= \|v-z\|_2 \|u\|_2 + \|u-w\|_2 \|z\|_2$$

$$\leq \|v-z\|_2 \cdot rad(T) + \|u-w\|_2 \cdot rad(S)$$

Define $Y_{uv} := \langle g,u\rangle rad(T) + \langle h,v\rangle rad\, S$

where $g \sim N(0,I)$, $h \sim N(0,I)$

Then $\|Y_{uv} - Y_{wz}\|^2 = \|u-w\|_2^2 \, rad(T)^2 + \|v-z\|_2^2 \, rad(S)^2$

So

$$\|X_{uv} - X_{wz}\|_{\psi_2} \leq C \|Y_{uv} - Y_{wz}\|_2$$

Talagrand

$$\Rightarrow \mathbb{E} \sup_{u\in T, v\in S} X_{uv} \leq C \, \mathbb{E} \sup_{u\in T, v\in S} Y_{uv}$$

$$= \mathbb{E} \sup_{u\in T} \langle g,u\rangle rad(S) + \mathbb{E} \sup_{v\in S} \langle h,v\rangle rad(T)$$

$$= w(T)\, rad(S) + w(S)\, rad(T)$$

$\boxtimes$

$\boxed{\text{Chapter 7}}$

Matrix deviation inequality and geometric consequences.

- Matrix Deviation Inequality
- M^*-bound and escape theorem.

We would like to prove the following.

<u>Thm</u>: Let $A \in \mathbb{R}^{m \times n}$ have independent, isotropic, and σ-subGaussian rows. Then for any $T \subseteq \mathbb{R}^n$, we have:

$$(\star) \quad \mathbb{E} \sup_{x \in T} \Big| \|Ax\|_2 - \sqrt{m}\, \|x\|_2 \Big| \leq C \sigma^2 \gamma(T)$$

where $\gamma(T) = \mathbb{E} \sup_{x \in T} |\langle g, x \rangle|$ with $g \sim N(0, I)$

<u>Remark</u>:

- Clearly $\omega(T) \leq \gamma(T)$
- If $T = -T$, then $\omega(T) = \gamma(T)$
- $\omega(T) = \frac{1}{2}\, \omega(T-T) = \frac{1}{2}\, \gamma(T-T)$
- (HW) For any $T \subseteq \mathbb{R}^n$ and $y \in T$ it holds:

$$\frac{1}{3}\left[\omega(T) + \|y\|_2\right] \leq \gamma(T) \leq 2\left[\omega(T) + \|y\|_2\right]$$

Observe

$$\mathbb{E}\frac{\|Ax\|_2}{\|x\|} = \mathbb{E}\sqrt{\sum_{i=1}^{m}\frac{\langle A_i, x\rangle^2}{\|x\|^2}}$$

independent σ-subGaussian

and $\mathbb{E}\langle A_i, x\rangle^2 = \langle xx^T, \mathbb{E}A_iA_i^T\rangle$

$= \|x\|_2^2$

concentration of norm

$$\Rightarrow (\sqrt{m} - C\sigma^2) \leq \mathbb{E}\frac{\|Ax\|_2}{\|x\|_2} \leq (\sqrt{m} + C\sigma^2)$$

$$\Rightarrow \quad -C\sigma^2\|x\|_2 \leq \mathbb{E}\|Ax\|_2 - \sqrt{m}\|x\|_2 \leq C\sigma^2\|x\|_2$$

Notice

$$\gamma(T) = \mathbb{E}\sup_{y\in T}|\langle g, y\rangle| \geq \mathbb{E}|\langle g, x\rangle| = \hat{C}\|x\|_2$$

So A is amounts to

$$\mathbb{E}\sup_{x\in T}\left|\|Ax\|_2 - \mathbb{E}\|Ax\|_2\right| \leq C\sigma^2\gamma(T)$$

We will use Talagrand's comparison inequality. Define

$$W_x := \|Ax\|_2 - \sqrt{m}\,\|x\|_2$$

We must show

$$\|W_x - W_y\|_{\psi_2} \leq C\sigma^2 \|x-y\|_2 \quad \forall x, y \in T.$$

We prove a few special cases and then prove the general case.

<u>Setting 1</u>: $\|x\|_2 = 1$, $y = 0$

We need to show

$$\big\| \|Ax\|_2 - \sqrt{m} \big\|_{\psi_2} \leq C\sigma^2.$$

Observe $\|Ax\|_2 = \|w\|_2$ where $w \in \mathbb{R}^m$ has independent σ-SubGaussian coordinates with $\mathbb{E}\, w_i^2 = 1$. So we already proved this.

Setting 2: $\|x\|_2 = \|y\|_2 = 1$

We must show

$$\big\| \|Ax\|_2 - \|Ay\|_2 \big\|_{\psi_2} \leq C\sigma^2 \|x-y\|_2$$

Let's analyze the squares first

$$Z := \frac{\|Ax\|_2^2 - \|Ay\|_2^2}{\|x-y\|_2} = \frac{\langle A(x-y), A(x+y)\rangle}{\|x-y\|_2} = \langle Au, Av\rangle$$

where $u = \frac{x-y}{\|x-y\|_2}$, $v = x+y$.

Claim: $\mathbb{P}[|Z| \geq s\sqrt{m}] \leq 2\exp\left(-\frac{cs^2}{\sigma^4}\right)$

for any $0 < s \leq \sqrt{m}$

pf: Observe $Z = \sum_{i=1}^{m} \underbrace{\langle A_i, u\rangle \langle A_i, v\rangle}_{\text{independent}}$

and

$$\mathbb{E}\langle A_i, x-y\rangle\langle A_i, x+y\rangle = \mathbb{E}\left(\langle A_i, x\rangle^2 - \langle A_i, y\rangle^2\right) = 0$$

Notice $\langle A_i, u\rangle\langle A_i, v\rangle$ is subexponential

with parameter $\|\langle A_i, u\rangle\|_{\psi_2} \cdot \|\langle A_i, v\rangle\|_{\psi_2} \leq 2\sigma^2$ ☒

Getting rid of the squares

Lemma: Let $x, y \in \mathbb{R}^n$ with $\|x\|_2 = \|y\|_2 = 1$.
Then

$$\big\| \|Ax\|_2 - \|Ay\|_2 \big\|_{\psi_2} \le C\sigma^2 \|x-y\|_2.$$

pf: We want to prove

$$p(s) := \mathbb{P}\left[\frac{\big| \|Ax\|_2 - \|Ay\|_2 \big|}{\|x-y\|_2} \ge s \right] \le 4\exp\left(-\frac{cs^2}{4\sigma^2}\right)$$

Case 1: $s \le 2\sqrt{m}$. Observe

$$p(s) = \mathbb{P}\left[\frac{\big| \|Ax\|_2^2 - \|Ay\|_2^2 \big|}{\|x-y\|_2} \ge s\left(\|Ax\|_2 + \|Ay\|_2\right) \right]$$

$$= \mathbb{P}\left[|Z| \ge s\left(\|Ax\|_2 + \|Ay\|_2\right) \right]$$

$$\le \mathbb{P}\left[|Z| \ge s\|Ax\|_2 \right]$$

$$\le \underbrace{\mathbb{P}\left[|Z| \ge \frac{s\sqrt{m}}{2} \right]}_{= p_1(s)} + \underbrace{\mathbb{P}\left[\|Ax\|_2 < \frac{\sqrt{m}}{2} \right]}_{= p_2(s)}$$

We already prove $p_1(s) \leq 2\exp\left(-\frac{cs^3}{\sigma^4}\right)$
[setting 2]

Moreover

$$p_2(s) = \mathbb{P}\left[\|Ax\|_2 < \frac{\sqrt{m}}{2}\right]$$

$$\leq \mathbb{P}\left[\left|\|Ax\|_2 - \sqrt{m}\right| \geq \frac{\sqrt{m}}{2}\right]$$

$$\overset{\text{setting 1}}{\leq} 2\exp\left(-\frac{cs^2}{\sigma^4}\right).$$

So $p(s) \leq 4\exp\left(\frac{-cs^2}{\sigma^4}\right)$

<u>Case 2</u>: $s > 2\sqrt{m}$.

Recall $p(s) = \mathbb{P}\left[\frac{\left|\|Ax\|_2 - \|Ay\|_2\right|}{\|x-y\|_2} \geq s\right]$

Reverse Triangle Inequality

$$\Rightarrow \left|\|Ax\|_2 - \|Ay\|_2\right| \leq \|A(x-y)\|_2$$

So

$$p(s) \leq \mathbb{P}\left[\|Au\|_2 \geq s\right]$$

$$\leq \mathbb{P}\left[\|Au\|_2 - \sqrt{m} \geq \frac{s}{2}\right]$$

$$\leq 2\exp\left(-\frac{cs^2}{\sigma^4}\right)$$

setting, $\boxtimes$

Now let's prove the general case. Fix $x, y \in \mathbb{R}^n$. Can rescale and assume

$$\|x\|_2 = 1 \quad \text{and} \quad \|y\|_2 \geq 1$$

Define $\bar{y} = \frac{y}{\|y\|_2}$.

$$\Rightarrow \|W_x - W_y\|_{\psi_2} \leq \underbrace{\|W_x - W_{\bar{y}}\|_{\psi_2}}_{\leq C\sigma^2\|x-\bar{y}\|_2} + \|W_{\bar{y}} - W_y\|_{\psi_2}$$

Notice

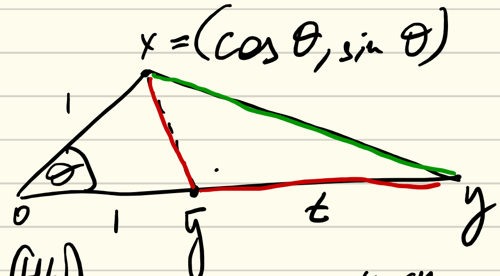
$$\|W_{\bar{y}} - W_y\|_{\psi_2} = \underbrace{\|\bar{y} - y\|_2}_{= \|y\| - 1} \|W_{\bar{y}}\|_{\psi_2}$$

<u>We know:</u> (Setting 1)

$$\|W_{\bar{y}}\|_{\psi_2} \le C\sigma^2$$

So $\|W_x - W_y\|_{\psi_2} \le C\sigma^2 \left(\|x-\bar{y}\|_2 + \|\bar{y}-y\|_2\right)$

$$\le C\sqrt{2}\,\sigma^2 \|x-y\|_2$$



(HW)

<u>Cor:</u> Let $A \in \mathbb{R}^{m\times n}$ with A_i independent, isotropic, σ^2-subGaussian. Then for any $T \subseteq \mathbb{R}^n$ and $u > 0$, we have

$$\sup_{x\in T} \left| \|Ax\|_2 - \sqrt{m}\,\|x\|_2 \right| \le C\sigma^2 \left(w(T) + u\,\mathrm{rad}(T)\right)$$

w.p. $1 - 2\exp(-u^2)$.

Some Consequences:

① Bounds on singular values

Let $T = S^{n-1}$ ← sphere. Then we learn

$$\sqrt{m} - C\sigma^2(\sqrt{n}+u) \le \|Ax\|_2 \le \sqrt{m} + C\sigma^2(\sqrt{n}+u)$$

$\forall x \in S^{n-1}$ w.p. $1-2\exp(-u^2)$

② Diameter of random projections

Define $P = \frac{1}{\sqrt{n}} A$.

Then

$$\mathbb{E}\,\mathrm{diam}(PT) \le \sqrt{\frac{m}{n}}\,\mathrm{diam}(T) + C\sigma^2 s(T)$$

pf:

$$\mathbb{E}\sup_{x\in T} \|Ax\|_2 \le \sqrt{m}\sup_{x\in T}\|x\|_2 + C\sigma^2\gamma(T)$$

replace T by $T-T$ and divide by $\sqrt{n}$. ☒

③ Johson-Lindenstrauss

Let $S = \{x_i\}_{i=1}^N \subseteq \mathbb{R}^n$.

Set $T := \left\{ \frac{x-y}{\|x-y\|} : x, y \in S \right\}$

Recall $\gamma(T) \leq C\sqrt{\log n}$. So

$$\sup_{x,y \in S} \left| \frac{\|Ax - Ay\|_2}{\|x-y\|_2} - \sqrt{m} \right| \leq C\sqrt{\log N}$$

w.h.p.

$\Rightarrow$

$\forall x, y \in S$

$$\left(1 - C\sqrt{\frac{\log N}{m}}\right)\|x-y\|_2 \leq \frac{1}{\sqrt{m}}\|Ax - Ay\| \leq \left(1 + C\sqrt{\frac{\log N}{m}}\right)\|x-y\|_2$$

Accuracy ε requires $m \geq \frac{\log N}{\varepsilon^2}$

Prop: Fix a set $T \subseteq \mathbb{R}^n$. Let $A \in \mathbb{R}^{m \times n}$ have independent, isotropic, σ-subGaussian rows A_i. Then with probability 0.99 we have

$$\|x-y\|_2 - \delta \le \frac{1}{\sqrt{m}} \|Ax - Ay\|_2 \le \|x-y\|_2 + \delta \quad \forall x,y \in T$$

where
$$\delta = \frac{C\sigma^2 w(T)}{\sqrt{m}}$$

pf: Apply matrix deviation to $T-T$ $\square$

Remark: Choose $m \geq \frac{C\sigma^4}{\varepsilon^2} d(T)$

$$d(T) = \frac{w(T)^2}{\mathrm{diam}(T)^2}$$

Then
$$\delta = C\varepsilon\, \mathrm{diam}(T).$$

Random Sections:

The following two consequences of the matrix deviation inequality will be the main tools for sparse recovery.

Thm (M^*-bound) Consider $T \subseteq \mathbb{R}^n$ and let $A \in \mathbb{R}^{m \times n}$ have independent, isotropic, σ-subGaussian rows. Then $E = \ker(A)$ satisfies

$$\mathbb{E}\, \mathrm{diam}(T \cap E) \leq \frac{C\sigma^2 w(T)}{\sqrt{m}}$$

pf: We'll actually prove

$$\mathbb{E} \sup_z \mathrm{diam}(T \cap (z+E)) \leq \frac{C\sigma^2 w(T)}{\sqrt{m}}$$

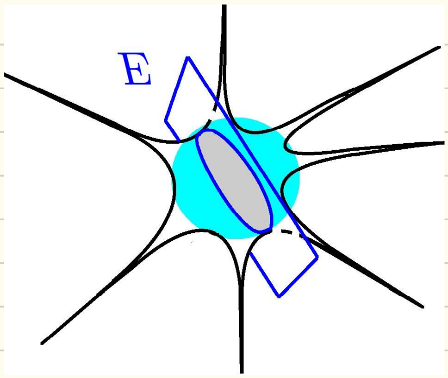
Matrix deviation with $T-T$ gives

$$\mathbb{E} \sup_{x,y \in T} \big| \|Ax - Ay\| - \sqrt{m}\|x-y\| \big| \leq C\sigma^2 \underbrace{\gamma(T-T)}_{=2w(T)}$$

$$\geq$$

$$\mathbb{E} \sup_z \sup_{x,y \in (z+E) \cap T} \sqrt{m}\, \|x-y\| = \sqrt{m}\, \mathbb{E} \sup_z \mathrm{diam}((z+E) \cap T)$$

$\square$



Ex: Let $T = B_1^n$. Then

$$\mathbb{E}\, \mathrm{diam}(T \cap E) \le C \sqrt{\frac{\log n}{m}}$$

so if $m = \delta n$, then $\dim E = (1-\delta)n$

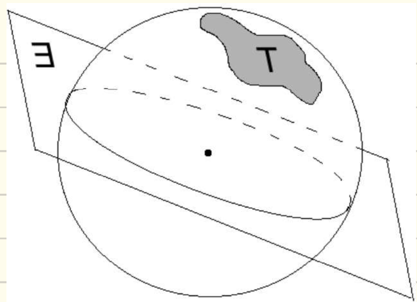
but $\mathbb{E}\, \mathrm{diam}(T \cap E) \le C \sqrt{\frac{\log n}{\delta n}}$

Remark: How big must m be to guarantee

$$\mathbb{E}\, \mathrm{diam}(T \cap E) \le \varepsilon\, \mathrm{diam}(T)$$

Answer: $m \ge C \frac{\sigma^4}{\varepsilon^2} \cdot d(T)$

$$\underbrace{\qquad}_{\frac{w(T)^2}{\mathrm{diam}\, T}}$$



Thm (Escape Thm) Consider $T \subseteq S^{n-1}$ and let $A \in \mathbb{R}^{m \times n}$ have independent, isotropic, σ-subGaussian rows. If

$$m \geq C\sigma^4 w(T)^2$$

then $E = \ker(A)$ satisfies $T \cap E = \emptyset$ w.p. $1 - 2\exp\left(\frac{-cm}{\sigma^4}\right)$.

pf: High Probability Matrix Deviation gives $\sup_{x \in T} \left| \|Ax\|_2 - \sqrt{m}\|x\|_2 \right| \leq C\sigma^2(w(T)+u)$ w.p. $1-\exp(-u^2)$. In this event, for and $x \in T \cap E$, it holds

$$\sqrt{m} \leq C\sigma^2(w(T)+u)$$

Set $u = \frac{\sqrt{m}}{2C\sigma^2}$ and get a contradiction $\boxtimes$