

High-Dimensional Probability

Lectured by [Hao Ge](#)

L^AT_EXed by [Chengxin Gong](#)

January 8, 2024

Contents

0	Appetizer	2
1	Preliminaries on random variables	2
2	Concentration of sums of independent random variables	2
3	Random vectors in high dimensions	4
4	Random matrices	5
5	Concentration without independence	7
6	Quadratic forms, symmetrization, contraction	9
7	Random processes	10
8	Chaining	12

[Reference](#) [High-Dimensional Probability: An Introduction with Applications in Data Science \(Roman Vershynin\)](#)

0 Appetizer

- Convex combination: For $z_1, z_2, \dots, z_m \in \mathbb{R}^n$, the form of $\sum_{i=1}^m \lambda_i z_i$ with $\lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1$. Convex hull of $T \subset \mathbb{R}^n$: $\text{conv}(T) = \{\text{convex combinations of } z_1, \dots, z_m \in T, m \in \mathbb{N}\}$.
- Caratheodory's theorem: Every point in the convex hull of a set $T \subset \mathbb{R}^n$ can be expressed as a convex combination of at most $n + 1$ points from T .
- Approximate Caratheodory's theorem: Consider $T \subset \mathbb{R}^n$, $\text{diam}(T) = \sup\{\|s - t\|_2, s, t \in T\} < 1$. Then for any $x \in \text{conv}(T)$ and any k , one can find points $x_1, x_2, \dots, x_k \in T$ such that $\|x - \frac{1}{k} \sum_{i=1}^k x_i\|_2 \leq \frac{1}{\sqrt{k}}$ (repetition is allowed).

Proof WLOG assume $\|t\|_2 \leq 1, \forall t \in T$. Fix $x \in \text{conv}(T), x = \sum_{i=1}^m \lambda_i z_i, z_i \in T$. Define $Z, \mathbb{P}(Z = z_i) = \lambda_i, \mathbb{E}Z = x$. Consider i.i.d. Z_1, Z_2, \dots of $Z, \frac{1}{n} \sum_{j=1}^n Z_j \rightarrow x$ a.s. $n \rightarrow +\infty$. $\mathbb{E}\|x - \frac{1}{k} \sum_{j=1}^k Z_j\|_2^2 = \frac{1}{k^2} \sum_{j=1}^k \mathbb{E}\|Z_j - x\|_2^2 = \frac{1}{k^2} \sum_{j=1}^k (\mathbb{E}\|Z_j\|_2^2 - \|\mathbb{E}Z_j\|_2^2) \leq \frac{1}{k} \Rightarrow \exists$ a realization of Z_1, \dots, Z_k such that $\|x - \frac{1}{k} \sum_{j=1}^k Z_j\|_2 \leq \frac{1}{\sqrt{k}}$. \square

- Corollary (Covering polytopes by balls): P is a polytope in \mathbb{R}^n with N vertices, $\text{diam}(P) \leq 1$. Then P can be covered by at most $N^{\lceil 1/\epsilon^2 \rceil}$ Euclidean balls of radii $\epsilon > 0$.

1 Preliminaries on random variables

- Jensen's inequality: convex $\phi, \phi(\mathbb{E}X) \leq \mathbb{E}\phi(X). \Rightarrow \|X\|_{L^p} \leq \|X\|_{L^q}$ for $p \leq q$.
- Minkowski inequality: $p \geq 1, \|X + Y\|_{L^p} \leq \|X\|_{L^p} + \|Y\|_{L^p}$.
- Cauchy-Schwarz inequality: $\mathbb{E}|XY| \leq \|X\|_{L^2} \|Y\|_{L^2}$.
- Holder inequality: $p, q \in (1, +\infty), \frac{1}{p} + \frac{1}{q} = 1$ or $p = 1, q = \infty, \mathbb{E}\|XY\| \leq \|X\|_{L^p} \|Y\|_{L^q}$.
- $X \geq 0$, then $\mathbb{E}X = \int_0^{+\infty} \mathbb{P}(X > t) dt$.
- Markov inequality: $X \geq 0, t > 0, \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$.
- LLN: X_1, \dots, X_n, \dots i.i.d., $\mathbb{E}X_i = \mu, \text{Var}(X_i) = \sigma^2, S_N = X_1 + X_2 + \dots + X_N$. Then: (WLLN) $\mathbb{P}(|\frac{S_N}{N} - \mu| > \epsilon) \rightarrow 0, \forall \epsilon > 0$; (SLLN) $\mathbb{P}(\frac{S_N}{N} \rightarrow \mu, N \rightarrow +\infty) = 1$.
- CLT: $Z_N = \frac{S_N - \mathbb{E}S_N}{\sqrt{\text{Var}(S_N)}} \xrightarrow{d} \mathcal{N}(0, 1)$.
- $X_{N,i}, 1 \leq i \leq N$ independent $\text{Ber}(p_{N,i}), \max_{i \leq N} p_{N,i} \rightarrow 0, S_N = \sum_{i=1}^N X_{N,i}, \mathbb{E}S_N \rightarrow \lambda < +\infty$. Then $S_N \xrightarrow{d} \text{Poisson}(\lambda)$.

2 Concentration of sums of independent random variables

- Question: N times, $\mathbb{P}(\text{head} \geq \frac{3}{4}N) = ?$ Let S_N be the number of heads, $\mathbb{E}S_N = \frac{N}{2}, \text{Var}(S_N) = \frac{N}{4}$. (1) Chebyshev's inequality: $\mathbb{P}(S_N \geq \frac{3}{4}N) \leq \mathbb{P}(|S_N - \frac{N}{2}| \geq \frac{N}{4}) \leq \frac{4}{N}$; (2) $Z_N = \frac{S_N - \frac{N}{2}}{\sqrt{N/4}}$, expect: $\mathbb{P}(Z_N \geq \sqrt{N/4}) \approx \mathbb{P}(g \geq \sqrt{N/4}) \leq \frac{1}{\sqrt{2\pi}} e^{-N/8}$ where $g \sim \mathcal{N}(0, 1)$.
- For all $t > 0, (\frac{1}{t} - \frac{1}{t^3}) \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \leq \mathbb{P}(g \sim \mathcal{N}(0, 1) \geq t) \leq \frac{1}{t} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$.
- Berry-Esseen bound: $|\mathbb{P}(Z_N \geq t) - \mathbb{P}(g \geq t)| \leq \frac{\rho}{\sqrt{N}}$ where $\rho = \mathbb{E}|X_1 - \mu|^3 / \sigma^3$. And in general, no improvement since $\mathbb{P}(S_N = \frac{N}{2}) \sim \frac{1}{\sqrt{N}} \Rightarrow \mathbb{P}(Z_N = 0) \sim \frac{1}{\sqrt{N}}$ but $\mathbb{P}(g = 0) = 0$.
- Hoeffding's inequality: X_1, \dots, X_N i.i.d. symmetric Bernoulli ($\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$), $a = (a_1, \dots, a_N) \in \mathbb{R}^N$. Then $\forall t \geq 0, \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) \leq e^{-t^2/2\|a\|_2^2}, \mathbb{P}(|\sum_{i=1}^N a_i X_i| \geq t) \leq 2e^{-t^2/2\|a\|_2^2}$.

Proof WLOG, $\|a\|_2^2 = 1$. For $\lambda > 0, \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) = \mathbb{P}(e^{\lambda \sum_{i=1}^N a_i X_i} \geq e^{\lambda t}) \leq e^{-\lambda t} \mathbb{E}e^{\lambda \sum_{i=1}^N a_i X_i} = e^{-\lambda t} \prod_{i=1}^N \mathbb{E}e^{\lambda a_i X_i} = e^{-\lambda t} \prod_{i=1}^N \cosh(\lambda a_i) \leq e^{-\lambda t} \prod_{i=1}^N e^{\lambda^2 a_i^2 / 2} = e^{-\lambda t + \frac{\lambda^2}{2}} \Rightarrow \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) \leq \inf_{\lambda \geq 0} e^{-\lambda t + \frac{\lambda^2}{2}} = e^{-\frac{t^2}{2}} (\lambda = t). \quad \square$

- Bounded r.v.s: X_1, \dots, X_N independent, $X_i \in [m_i, M_i]$. Then $\forall t \geq 0, \mathbb{P}(\sum_{i=1}^N (X_i - \mathbb{E}X_i) \geq t) \leq e^{-\frac{2t^2}{\sum_{i=1}^N (M_i - m_i)^2}}$.
- Chernoff's inequality: $X_i \sim \text{Ber}(p_i)$ independent, $S_N = \sum_{i=1}^N X_i, \mu = \mathbb{E}S_N \Rightarrow \forall t > \mu, \mathbb{P}(S_N \geq t) \leq e^{-\mu(\frac{t}{\mu})^t}$.
Proof $\mathbb{P}(S_N \geq t) \leq e^{-\lambda t} \prod_{i=1}^N \mathbb{E}e^{\lambda X_i}$. $\mathbb{E}e^{\lambda X_i} = e^{\lambda p_i} + (1 - p_i) = 1 + (e^\lambda - 1)p_i \leq e^{(e^\lambda - 1)p_i} \Rightarrow \mathbb{P}(S_N \geq t) \leq e^{-\lambda t} e^{(e^\lambda - 1)\mu}$. Take $\lambda^* = \log(t/\mu)$. \square
- Chernoff bound: $\mathbb{P}(|S_N - \mu| \geq \delta\mu) \leq 2e^{-c\mu\delta^2}$.
- $d = (n - 1)p$ is the expected degree. There is an absolute constant C s.t. for $G(n, p)$, $d \geq C \log n$. Then with high prob (for example 0.9), all vertices of G have degrees between $0.9d$ and $1.1d$.
Proof Ex 2.3.5 $\Rightarrow \mathbb{P}(|d_i - d| \geq \delta d) \leq 2e^{-c\delta^2 d}$. Union bound: $\mathbb{P}(\exists i, |d_i - d| \geq \delta d) \leq n \cdot 2e^{-c\delta^2 d} \leq n \cdot 2 \dots n^{-C c \delta^2} = 2n^{1 - C c \delta^2} \leq 1 - p^*$ (let $C c \delta^2 > 1$). \square
- Sub-gaussian properties: The following are equivalent: (i) $\mathbb{P}(|X| \geq t) \leq 2e^{-t^2/k_1^2}$ for all $t \geq 0$; (ii) $\|X\|_{L^p} = (\mathbb{E}|X|^p)^{1/p} \leq k_2\sqrt{p}$ for all $p \geq 1$; (iii) $\mathbb{E}e^{\lambda^2 X^2} \leq e^{k_3^2 \lambda^2}$ for all λ s.t. $|\lambda| \leq \frac{1}{k_3}$; (iv) $\mathbb{E}e^{X^2/k_4^2} \leq 2$; (v) $\mathbb{E}e^{\lambda X} \leq e^{k_5^2 \lambda^2}$, for all $\lambda \in \mathbb{R}$ (if $\mathbb{E}X = 0$).
Proof (i) \Rightarrow (ii): WLOG $k_1 = 1$. $\mathbb{E}|X|^p = \int_0^{+\infty} \mathbb{P}(|X| \geq t) p t^{p-1} dt \leq \int_0^{+\infty} 2e^{-t^2} p t^{p-1} dt = p \Gamma(\frac{p}{2}) \stackrel{\Gamma(x) \leq 3x^x \text{ for } x \geq \frac{1}{2}}{\leq} 3p(\frac{p}{2})^{p/2} \Rightarrow \|X\|_p \leq \frac{1}{\sqrt{2}}(3p)^{1/p} p^{1/2} \leq 3\sqrt{p}$.
(ii) \Rightarrow (iii): WLOG $k_2 = 1$. $\mathbb{E}e^{\lambda^2 X^2} = \mathbb{E}[1 + \sum_{p=1}^{+\infty} \frac{(\lambda^2 X^2)^p}{p!}]$. $\mathbb{E}|X|^{2p} \leq (2p)^p, p! \geq (\frac{p}{e})^p \Rightarrow \mathbb{E}e^{\lambda^2 X^2} \leq 1 + \sum_{p=1}^{+\infty} \frac{(2\lambda^2 p)^p}{(\frac{p}{e})^p} = \frac{1}{1 - 2e\lambda^2}$ (if $2e\lambda^2 < 1$) $\stackrel{\frac{1}{1-x} \leq e^{2x} \text{ for } x \in [0, \frac{1}{2}]}{\leq} e^{4e\lambda^2}$ (if $2e\lambda^2 \leq 1/2 \Leftrightarrow \lambda \leq \frac{1}{2\sqrt{e}}$).
(iii) \Rightarrow (iv): trivial.
(iv) \Rightarrow (i): $\mathbb{P}(|X| \geq t) = \mathbb{P}(e^{X^2} \leq e^{t^2}) \leq e^{-t^2} \mathbb{E}e^{X^2} \leq 2e^{-t^2}$.
(iii) \Rightarrow (v): WLOG $k_3 = 1$. If $|\lambda| \leq 1$, then $\mathbb{E}e^{\lambda X} \leq \mathbb{E}(\lambda X + e^{\lambda^2 X^2}) = \mathbb{E}e^{\lambda^2 X^2} \leq e^{\lambda^2}$. If $|\lambda| \geq 1$, then $\mathbb{E}e^{\lambda X} \leq e^{\frac{\lambda^2}{2}} \mathbb{E}e^{\frac{X^2}{2}} e^{\frac{\lambda^2}{2}} e^{\frac{1}{2}} \leq e^{\lambda^2}$.
(v) \Rightarrow (i): mimic the proof of (iv) \Rightarrow (i). \square
- Sub-gaussian r.v.: satisfy the above sub-gaussian properties. $\|X\|_{\psi_2} = \inf\{t > 0 : \mathbb{E}e^{X^2/t^2} \leq 2\}$. Thus $\mathbb{P}(|X| \geq t) \leq 2e^{-ct^2/\|X\|_{\psi_2}^2}$; $\|X\|_{L^p} \leq C\|X\|_{\psi_2}\sqrt{p}$; if $\mathbb{E}X = 0$ then $\mathbb{E}e^{\lambda X} \leq e^{C\lambda^2\|X\|_{\psi_2}^2}$.
- Maximum of sub-gaussians: $K = \max_{i \leq N} \|X_i\|_{\psi_2}$. Then $\mathbb{E}\max_{i \leq N} X_i \leq CK\sqrt{\log N}$.
- Let X_1, \dots, X_N be independent and mean zero sub-gaussian, then $\|\sum_{i=1}^N X_i\|_{\psi_2}^2 \leq C \cdot \sum_{i=1}^N \|X_i\|_{\psi_2}^2$.
Proof $\forall \lambda \in \mathbb{R}, \mathbb{E}e^{\lambda \sum_{i=1}^N X_i} = \prod_{i=1}^N \mathbb{E}e^{\lambda X_i} \leq \prod_{i=1}^N e^{c\lambda^2 \|X_i\|_{\psi_2}^2} = e^{c\lambda^2 \sum_{i=1}^N \|X_i\|_{\psi_2}^2}$ \square
- Centering: X is sub-gaussian $\Rightarrow X - \mathbb{E}X$ is sub-gaussian and $\|X - \mathbb{E}X\|_{\psi_2} \leq C\|X\|_{\psi_2}$.
Proof $\|\mathbb{E}X\|_{\psi_2} \leq C_1|\mathbb{E}X| \leq C_1\mathbb{E}|X| = C_1\|X\|_{L^1} \leq C_1C_2\|X\|_{\psi_2}$. \square
- Sub-exponential properties: The following are equivalent: (1) $\mathbb{P}(|X| \geq t) \leq 2e^{-t/k_1}, \forall t \geq 0$; (2) $\|X\|_{L^p} \leq k_2 p, p \geq 1$; (3) $\mathbb{E}e^{\lambda|X|} \leq e^{k_3\lambda}$ for all $0 \leq \lambda \leq \frac{1}{k_3}$; (4) $\mathbb{E}e^{|X|/k_4} \leq 2$; (5) if $\mathbb{E}X = 0, \mathbb{E}e^{\lambda X} \leq e^{k_5^2 \lambda^2}$ for $|\lambda| \leq \frac{1}{k_5}$.
Proof (2) \Rightarrow (5): $k_2 = 1, \mathbb{E}e^{\lambda X} = 1 + \sum_{p=2}^{+\infty} \frac{\lambda^p \mathbb{E}X^p}{p!} \leq 1 + \sum_{p=2}^{+\infty} \frac{\lambda^p p^p}{(p/e)^p} = 1 + \frac{(e\lambda)^2}{1 - e\lambda} (|e\lambda| < 1)$. If $|e\lambda| \leq \frac{1}{2}, 1 + \frac{(e\lambda)^2}{1 - e\lambda} \leq 1 + 2e^2\lambda^2 \leq e^{2e^2\lambda^2} \leq e^{4e^2\lambda^2}$, i.e. $k_5 = 2e$.
(5) \Rightarrow (1): $k_5 = 1, |x|^p \leq p^p(e^x + e^{-x}) \Rightarrow \mathbb{E}|X|^p \leq p^p(\mathbb{E}e^X + \mathbb{E}e^{-X}) \leq 2ep^p$. \square
- $\|X\|_{\psi_1} = \inf\{t > 0 : \mathbb{E}e^{|X|/t} \leq 2\}$. X is sub-gaussian $\Leftrightarrow X^2$ is sub-exponential. $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$.
- X, Y are sub-gaussian $\Rightarrow XY$ is sub-exponential and $\|XY\|_{\psi_1} \leq \|X\|_{\psi_2}\|Y\|_{\psi_2}$.
Proof WLOG $\|X\|_{\psi_2} = \|Y\|_{\psi_2} = 1$. $\mathbb{E}e^{XY} \leq \mathbb{E}e^{\frac{X^2+Y^2}{2}} = \mathbb{E}[e^{\frac{X^2}{2} + \frac{Y^2}{2}}] \leq \frac{1}{2}(\mathbb{E}e^{X^2} + \mathbb{E}e^{Y^2}) = 2$. \square
- Orlicz function/space: $\psi : [0, +\infty) \rightarrow [0, +\infty)$, convex, increasing, $\psi(0) = 0, \psi(x) \rightarrow +\infty, x \rightarrow +\infty$. $\|X\|_{\psi} := \inf\{t > 0 : \mathbb{E}\psi(|X|/t) \leq 1\}$. $L_{\psi} := \{X : \|X\|_{\psi} < +\infty\}$ is Banach space. Examples: (1) $L_p : \psi(x) = x^p, p \geq 1$; (2) $L_{\psi_2} : \psi_2(x) = e^{x^2} - 1, L_{\infty} \subset L_{\psi_2} \subset L_p$.

- Bernstein's inequality: X_1, \dots, X_N independent, mean zero and sub-exponential. Then for $t \geq 0$, $\mathbb{P}(|\sum_{i=1}^N X_i| \geq t) \leq 2e^{-c \min(\frac{t^2}{\sum_{i=1}^N \|X_i\|_{\psi_1}^2}, \frac{t}{\max_i \|X_i\|_{\psi_1}})}$.
Proof $S = \sum_{i=1}^N X_i$. $\mathbb{P}(S \geq t) \leq e^{-\lambda t} \prod_{i=1}^N \mathbb{E}e^{\lambda X_i}$. $\mathbb{E}e^{\lambda X_i} \leq e^{c\lambda^2 \|X_i\|_{\psi_1}^2}$ if $|\lambda| \leq \frac{c}{\max \|X_i\|_{\psi_1}}$. Then $\mathbb{P}(S \geq t) \leq e^{-\lambda t + c\lambda^2 \sigma^2}$ where $\sigma^2 := \sum_{i=1}^N \|X_i\|_{\psi_1}^2$. The following is to find the minimum of a quadratic function with the restriction $|\lambda| \leq \frac{c}{\max \|X_i\|_{\psi_1}}$. \square
- Corollary 1: $\mathbb{P}(|\sum_{i=1}^N a_i X_i| \geq t) \leq 2e^{-c \min(\frac{t^2}{K^2 \|a\|_2^2}, \frac{t}{K \|a\|_\infty})}$ where $K = \max_i \|X_i\|_{\psi_1}$.
- Corollary 2: $|X_i| \leq K$, then $\mathbb{P}(|\sum_{i=1}^N X_i| \geq t) \leq 2 \exp(-\frac{t^2/2}{\sigma^2 + Kt/3})$ where $\sigma^2 = \sum_{i=1}^N \mathbb{E}X_i^2$.

3 Random vectors in high dimensions

- $X \in \mathbb{R}^n$, independent sub-gaussian coordinate X_i , $\mathbb{E}X_i^2 = 1$. Then $\| \|X\|_2 - \sqrt{n} \|_{\psi_2} \leq CK^2$, $K = \max_i \|X_i\|_{\psi_2}$.
Proof $\mathbb{E}X_i^2 = 1 \Rightarrow K \geq 1$. $\|X_i^2 - 1\|_{\psi_1} \leq C\|X_i^2\|_{\psi_1} = C\|X_i\|_{\psi_2}^2 \leq CK^2$. Bernstein's inequality: $\mathbb{P}(|\frac{1}{n}\|X\|_2^2 - 1| \geq u) = \mathbb{P}(|\frac{1}{n}\sum_{i=1}^n (X_i^2 - 1)| \geq u) \leq 2e^{-cn \min(\frac{u^2}{K^4}, \frac{u}{K^2})} \leq 2e^{-\frac{cn}{K^4} \min(u^2, u)}$. For any $\delta > 0$, $\mathbb{P}(|\frac{1}{\sqrt{n}}\|X\|_2 - 1| \geq \delta) \leq \mathbb{P}(|\frac{1}{n}\|X\|_2^2 - 1| \geq \max(\delta, \delta^2)) \leq 2e^{-\frac{cn}{K^4} \delta^2} \Rightarrow \mathbb{P}(\| \|X\|_2 - \sqrt{n} \| \geq t) \leq 2e^{-ct^2/K^4}$. \square
- Isotropy: $\Sigma(X) = \mathbb{E}XX^T = I$. If $\Sigma \neq I_n$, then let $Z = \Sigma^{-1/2}X$. X is isotropic $\Leftrightarrow \mathbb{E}\langle X, x \rangle^2 = \|x\|_2^2$ for any $x \in \mathbb{R}^n$.
Proof $\mathbb{E}\langle X, x \rangle^2 = \mathbb{E}(x^T X X^T x) = x^T (\mathbb{E}X X^T) x$. $\|x\|_2^2 = x^T I_n x$. $\Rightarrow \mathbb{E}X X^T = I_n$. \square
- X is isotropic $\Rightarrow \mathbb{E}\|X\|_2^2 = n$. If X, Y are independent and isotropic $\Rightarrow \mathbb{E}\langle X, Y \rangle^2 = n$.
Proof $\mathbb{E}\|X\|_2^2 = \mathbb{E}(X^T X) = \mathbb{E}(\text{tr}(X^T X)) = \text{tr}(\mathbb{E}X X^T) = n$.
 $\mathbb{E}\langle X, Y \rangle^2 = \mathbb{E}(X^T Y Y^T X) = \mathbb{E}(\text{tr}(X^T Y Y^T X)) = \mathbb{E}(\text{tr}(X X^T Y Y^T)) = \text{tr}(\mathbb{E}X X^T) (\mathbb{E}Y Y^T) = n$. \square
- Examples: $X \sim U(\sqrt{n}\mathbb{S}^{n-1})$, $X \sim U(\{-1, 1\}^n)$, $X = (X_1, \dots, X_n)$ i.i.d., $\mathbb{E}X_i = 0$, $\text{Var}(X_i) = 1$ are all isotropic.
- $g \sim \mathcal{N}(0, I_n)$, then $\mathbb{P}(\|g\|_2 - \sqrt{n} \geq t) \leq 2e^{-ct^2}$.
- Frame: $\{u_i\}_{i=1}^N \subset \mathbb{R}^n$, Approximate Parseval's identity: $A\|x\|_2^2 \leq \sum_{i=1}^N \langle u_i, x \rangle^2 \leq B\|x\|_2^2$. A, B : frame bounds. $A = B$: tight frame ($\Leftrightarrow \sum_{i=1}^N u_i u_i^T = AI_n$) and in this case, $\sum_{i=1}^N \langle u_i, x \rangle u_i = Ax$.
- (a) Tight frame $\{u_i\}_{i=1}^N, A = B, X = \text{Unif}\{u_i, i = 1, 2, \dots, N\}$, then $(\frac{N}{A})^{1/2}X$ is isotropic. (b) X is isotropic, $\mathbb{P}(X = x_i) = p_i, i = 1, 2, \dots, N$. Then $u_i = \sqrt{p_i}x_i$ form a tight frame with $A = B = 1$.
- Isotropic convex sets: $X \sim \text{Unif}(K), K \subset \mathbb{R}^n$ convex, bounded, non-empty interior (convex body). Assume $\mathbb{E}X = 0, \Sigma = \text{Cov}(X)$. Then $Z = \Sigma^{-1/2}X$ is isotropic and $Z \sim \text{Unif}(\Sigma^{-1/2}K)$.
- $X \in \mathbb{R}^n$ is sub-gaussian $\Leftrightarrow \forall X \in \mathbb{R}^n, \langle X, x \rangle$ are sub-gaussian. $\|X\|_{\psi_2} = \sup_{x \in \mathbb{S}^{n-1}} \|\langle X, x \rangle\|_{\psi_2}$.
- $X = (X_1, \dots, X_n)$ independent, mean zero, sub-gaussian coordinate. Then X is sub-gaussian with $\|X\|_{\psi_2} \leq C \max_{i \leq n} \|X_i\|_{\psi_2}$.
Proof $\|\langle X, x \rangle\|_{\psi_2}^2 = \|\sum X_i x_i\|_{\psi_2}^2 \leq C \sum_{i=1}^n x_i^2 \|X_i\|_{\psi_2}^2 \leq C \max_{i \leq n} \|X_i\|_{\psi_2}^2$. \square
- Gaussian dist: $X \sim \mathcal{N}(0, I_n), \|X\|_{\psi_2} \leq C$.
- Discrete dist: $X \sim \text{Unif}\{\sqrt{n}e_i, i = 1, 2, \dots, n\}, \|X\|_{\psi_2} \simeq \sqrt{\frac{n}{\log n}}$.
- Uniform dist: $X \sim \text{Unif}\{\sqrt{n}\mathbb{S}^{n-1}\}, \|X\|_{\psi_2} \leq C$.
Proof $g \sim \mathcal{N}(0, I_n), X \stackrel{d}{=} \frac{\sqrt{ng}}{\|g\|_2}$. $p(t) := \mathbb{P}(|X_1| \geq t) = \mathbb{P}(\frac{|g_1|}{\|g\|_2} \geq \frac{t}{\sqrt{n}})$. $\| \|g\|_2 - \sqrt{n} \|_{\psi_2} \leq C \Rightarrow \mathbb{P}(\|g\|_2 \leq \frac{\sqrt{n}}{2}) \leq 2e^{-cn}$. Need to show that all one-dimensional marginals $\langle X, x \rangle$ are sub-gaussian. By rotation invariance, we may assume that $x = (1, 0, \dots, 0)$. Let $\mathcal{E} = \{\|g\|_2 \geq \frac{\sqrt{n}}{2}\} \Rightarrow p(t) \leq \mathbb{P}(\frac{|g_1|}{\|g\|_2} \geq \frac{t}{\sqrt{n}}, \mathcal{E}) + \mathbb{P}(\mathcal{E}) \leq \mathbb{P}(|g_1| \geq \frac{t}{2}, \mathcal{E}) + 2e^{-cn} \leq 2e^{-t^2/8} + 2e^{-cn} \stackrel{t \leq \sqrt{n}}{\leq} 4e^{-ct^2}$. \square
- Grothendieck's inequality: $A = \{a_{ij}\}_{m \times n}$ of real numbers. Assume $\forall x_i, y_i \in \{-1, 1\}$, we have $|\sum_{i,j} a_{ij} x_i y_j| \leq 1$. Then for any Hilbert space \mathcal{H} , any $u_i, v_j \in \mathcal{H}$ satisfying $\|u_i\| = \|v_j\| = 1$, we have $|\sum_{i,j} a_{ij} \langle u_i, v_j \rangle| \leq K$ with $K \leq 1.783$.

Proof (1) Reduction. For any $u_i, v_j \in \mathbb{R}^N$ s.t. $\|u_i\|_2 = \|v_j\|_2 = 1, K_{u,v} := \sum_{i,j} a_{ij} \langle u_i, v_j \rangle, K = \sup_{\|u\|_2=\|v\|_2=1} K_{u,v}$.

(2) Introduce randomness. $g \sim \mathcal{N}(0, I_N), U_i = \langle g, u_i \rangle, V_j = \langle g, v_j \rangle, \mathbb{E} U_i V_j = \langle u_i, v_j \rangle. K_{u,v} = \mathbb{E}(\sum_{i,j} a_{ij} U_i V_j) \Rightarrow K_{u,v} \leq R^2$ if $|U_i| \leq R, |V_j| \leq R$.

(3) Truncation. Given $R \geq 1, U_i = U_i^- + U_i^+, U_i^- = U_i 1_{\{|U_i| \leq R\}}, V_j = V_j^- + V_j^+, |U_i^-| \leq R, |V_j^-| \leq R. \|U_i^+\|_{L^2}^2 \leq 2(R + \frac{1}{R}) \frac{1}{\sqrt{2\pi}} e^{-R^2/2} < \frac{4}{R^2} (R > 1)$.

(4) Breaking up the sum. $\mathbb{E} \sum a_{ij} U_i V_j = \mathbb{E} \sum a_{ij} U_i^- V_j^- + \mathbb{E} \sum a_{ij} U_i^+ V_j^- + \mathbb{E} \sum a_{ij} U_i^- V_j^+ + \mathbb{E} \sum a_{ij} U_i^+ V_j^+ := S_1 + S_2 + S_3 + S_4. S_1 \leq R^2, S_2 \leq K \max \|U_i^+\|_2 \max \|V_j^-\|_2 \leq K \frac{2}{R}, S_3 \leq K \frac{2}{R}, S_4 \leq K \frac{4}{R^2}$.

(5) Putting everything together. $K_{u,v} \leq R^2 + \frac{4K}{R} + \frac{4K}{R^2} \Rightarrow K \leq R^2 + \frac{4K}{R} + \frac{4K}{R^2} \Rightarrow K \leq \frac{R^2}{1 - \frac{4}{R} - \frac{4}{R^2}}$. \square

• Remark: The assumption can be equivalently stated as $|\sum_{i,j} a_{ij} x_i y_j| \leq \max_i |x_i| \max_j |y_j|$. The conclusion can be equivalently stated as $|\sum_{i,j} a_{ij} \langle u_i, v_j \rangle| \leq K \max_i \|u_i\| \max_j \|v_j\|$.

• Semidefinite programming: $\max \langle A, X \rangle$ s.t. $X \succeq 0, \langle B_i, X \rangle = b_i, A, B_i n \times n, b_i$ real number, $\langle A, X \rangle = \text{tr}(A^T X) = \sum_{i,j=1}^n A_{ij} X_{ij}$.

• Semidefinite relaxation: $\max \sum_{i,j=1}^n A_{ij} x_i x_j, x_i = \pm 1, A n \times n$ symmetric matrix. Relax to $\max \sum_{i,j=1}^n A_{ij} \langle X_i, X_j \rangle, \|X_i\|_2 = 1, X_i \in \mathbb{R}^n$.

• A positive semidefinite, $\text{INT}(A) := \max \sum_{i,j=1}^n A_{ij} x_i x_j, x_i = \pm 1, \text{SDP}(A) := \max \sum_{i,j=1}^n A_{ij} \langle X_i, X_j \rangle, \|X_i\|_2 = 1$. Then $\text{INT}(A) \leq \text{SDP}(A) \leq 2K \cdot \text{INT}(A)$.

• Maximum cut: $G = (V, E)$ finite simple, $V \rightarrow V_1 + V_2$, cut number of edges crossing between V_1 and V_2 . MAX-CUT(G): NP-hard. Adjacency matrix $A = \{A_{ij}\}_{n \times n}, A_{ij} = \begin{cases} 1, & i \leftrightarrow j \\ 0, & \text{otherwise} \end{cases}$. Partition: $X = (x_i)_{n \times 1}, x_i = \pm 1. \text{CUT}(G, X) = \frac{1}{4} \sum_{i,j=1}^n A_{ij} (1 - x_i x_j). \text{MAX-CUT}(G) = \frac{1}{4} \max \{ \sum_{i,j} A_{ij} (1 - x_i x_j) : x_i = \pm 1 \}$.

• 0.5-approximation algorithm: Partition at random, $\mathbb{E} \text{CUT}(G, X) = 0.5|E| \geq 0.5 \text{MAX-CUT}(G)$.

• 0.878-approximation algorithm: $\text{SDP}(G) = \frac{1}{4} \max \{ \sum_{i,j=1}^n A_{ij} (1 - \langle X_i, X_j \rangle), x_i \in \mathbb{R}^n, \|X_i\|_2 = 1 \}$. $X_1, \dots, X_n \rightarrow x_1, \dots, x_n : g \sim \mathcal{N}(0, I_n), x_i := \text{sgn}(\langle X_i, g \rangle). \mathbb{E} \text{CUT}(G, X) \geq 0.878 \text{SDP}(G) \geq 0.878 \text{MAX-CUT}(G)$.

Proof $\mathbb{E} \text{CUT}(G, X) = \frac{1}{4} \sum_{i,j=1}^n A_{ij} (1 - \mathbb{E} x_i x_j)$ and $1 - \mathbb{E} x_i x_j = 1 - \mathbb{E} \text{sgn} \langle g, X_i \rangle \text{sgn} \langle g, X_j \rangle = 1 - \frac{2}{\pi} \arcsin \langle X_i, X_j \rangle \geq 0.878 (1 - \langle X_i, X_j \rangle)$. \square

• $u, v \in \mathbb{S}^{n-1}, \mathbb{E} \text{sgn}(\langle g, u \rangle) \text{sgn}(\langle g, v \rangle) = \frac{2}{\pi} \arcsin \langle u, v \rangle$.

• There exists a Hilbert space \mathcal{H} and $\phi, \psi : \mathbb{S}^{n-1} \rightarrow \mathbb{S}(\mathcal{H})$ s.t. $\frac{2}{\pi} \arcsin \langle \phi(u), \psi(v) \rangle = \beta \langle u, v \rangle$ for all $u, v \in \mathbb{S}^{n-1}$ and $\beta = \frac{2}{\pi} \log(1 + \sqrt{2})$.

Proof $\langle \phi(u), \psi(v) \rangle = \sin(\frac{\beta\pi}{2} \langle u, v \rangle)$. Ex 3.7.6 $\Rightarrow \exists \mathcal{H}, \phi, \psi. \sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} + \dots, \sinh t = t + \frac{t^3}{3!} + \frac{t^5}{5!}, \|\phi(u)\|_2^2 = \|\psi(u)\|_2^2 = \sinh(\frac{\beta\pi}{2}) = 1$ for all $u \in \mathbb{S}^{n-1} \Rightarrow \beta = \frac{2}{\pi} \log(1 + \sqrt{2})$.

Proof of Grothendieck's inequality with $K \leq \frac{1}{\beta} \approx 1.783$ WLOG $u_i, v_j \in \mathbb{S}^{N-1}$, then $\frac{2}{\pi} \arcsin \langle u'_i, v'_j \rangle = \beta \langle u_i, v_j \rangle, \mathcal{H} = \mathbb{R}^M, g \sim \mathcal{N}(0, I_M). \beta \sum a_{ij} \langle u_i, v_j \rangle = \sum_{i,j} a_{ij} \frac{2}{\pi} \arcsin \langle u'_i, v'_j \rangle = \sum_{i,j} a_{ij} \mathbb{E} \text{sgn} \langle g, u'_i \rangle \text{sgn} \langle g, v'_j \rangle \leq 1$. \square

4 Random matrices

• Singular vector decomposition: $A_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^T = \sum_{i=1}^n s_i U_i V_i^T, \Sigma = \text{diag}(s_1, s_2, \dots, s_r), s_i \geq 0$ singular values. $s_i = \sqrt{\lambda_i(AA^T)} = \sqrt{\lambda_i(A^T A)}$. If A is symmetric, $s_i = |\lambda_i(A)|$.

• Courant-Fisher's min-max theorem: $\lambda_i(A) = \max_{\dim E=i} \min_{x \in \mathbb{S}(E)} \langle Ax, x \rangle, s_i(A) = \max_{\dim E=i} \min_{x \in \mathbb{S}(E)} \|Ax\|_2$.

• Operator norm/spectral norm: $\|A\| = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x \in \mathbb{S}^{n-1}} \|Ax\|_2 = s_1(A)$. Or equivalently, $\|A\| = \max_{x \in \mathbb{S}^{n-1}, y \in \mathbb{S}^{m-1}} \langle Ax, y \rangle$.

• $s_n(A) > 0 \Leftrightarrow m \geq n = \text{rank}(A), s_n(A) = \frac{1}{\|A^+\|}$ where A^+ is pseudo-inverse (the norm of A^{-1} restriction to the image of A).

- Frobenius norm: $\|A\|_F = (\sum_{i=1}^n \sum_{j=1}^n A_{ij}^2)^{\frac{1}{2}} = (\sum_{i=1}^n s_i^2(A))^{\frac{1}{2}}$.
- Low-rank approximation: $\text{rank}(A) = r, k < r, A_k := \sum_{i=1}^k s_i u_i v_i^T, \|A - A_k\| = \min_{\text{rank}(A') \leq k} \|A - A'\|$ (holds for $\|\cdot\|, \|\cdot\|_F$).
- Approximate isometries: $m\|x\|_2 \leq \|Ax\|_2 \leq n\|x\|_2$ where $m = s_n(A), n = s_1(A)$, or $s_n\|x - y\|_2 \leq \|Ax - Ay\|_2 \leq s_1\|x - y\|_2$.
- $A_{m \times n}, \delta > 0$. If $\|A^T A - I_n\| \leq \max(\delta, \delta^2)$, then $(1 - \delta)\|x\|_2 \leq \|Ax\|_2 \leq (1 + \delta)\|x\|_2$ for all x .
Proof WLOG $\|x\|_2 = 1$. $|\|Ax\|_2^2 - 1| = |\langle (A^T A - I_n)x, x \rangle| \leq \max(\delta, \delta^2) \Rightarrow \max(\|Ax\|_2 - 1, (\|Ax\|_2 - 1)^2) \leq \max(\delta, \delta^2) \Rightarrow \|\|Ax\|_2 - 1\| \leq \delta$. \square
- $Q_{n \times m}, QQ^T = I_n \Leftrightarrow P = Q^T Q$ is an orthogonal proj in \mathbb{R}^m onto a subspace with $\dim n$.
- ϵ -net: (T, d) a metric space, $K \subset T, \epsilon > 0$. $\mathcal{N} \subset K$ is an ϵ -net of K if $\forall x \in K, \exists x_0 \in \mathcal{N}$ s.t. $d(x, x_0) \leq \epsilon$. Covering number: smallest $|\mathcal{N}| = |\mathcal{N}(K, d, \epsilon)|$.
- Compactness: $\mathcal{N}(K, d, \epsilon) < +\infty$ for all $\epsilon > 0$.
- ϵ -separated: $\mathcal{P} \subset T$ is ϵ -separated if $d(x, y) > \epsilon$ for all $x, y \in \mathcal{P}$. Packing number: largest $|\mathcal{P}| = |\mathcal{P}(K, d, \epsilon)|$.
- \mathcal{P} is a maximal ϵ -separated subset $\Rightarrow \mathcal{P}$ is a ϵ -net of K .
- $\mathcal{P}(K, d, 2\epsilon) \leq \mathcal{N}(K, d, \epsilon) \leq \mathcal{P}(K, d, \epsilon)$.

Proof The upper bound follows from the previous lemma. For the lower bound, choose an 2ϵ -separated subset $\mathcal{P} = \{x_i\}$ in K and an ϵ -net $\mathcal{N} = \{y_j\}$ of K . $\forall x_i, \exists y_j \in \mathcal{N}$, s.t. $|x_i - y_j| < \epsilon$. $\forall y_j$, there exists at most a $x_j \in \mathcal{P}$ s.t. $|x_i - y_j| < \epsilon$. \square

- Minkowski sum: $A, B \in \mathbb{R}^n, A + B := \{a + b, a \in A, b \in B\}$.
- $K \subset \mathbb{R}^n, \epsilon > 0, \frac{|K|}{|\epsilon B_2^n|} \leq \mathcal{N}(K, \epsilon) \leq \mathcal{P}(K, \epsilon) \leq \frac{|K + \frac{\epsilon}{2} B_2^n|}{|\frac{\epsilon}{2} B_2^n|}$ where $|\cdot|$ denotes the volume in \mathbb{R}^n , B_2^n denotes the unit Euclidean ball in \mathbb{R}^n .
- Corollary: Let $K = B_2^n$. $|\epsilon B_2^n| = \epsilon^n |K|, |K + \frac{\epsilon}{2} B_2^n| = (1 + \frac{\epsilon}{2})^n |K| \Rightarrow (\frac{1}{\epsilon})^n \leq \mathcal{N}(B_2^n, \epsilon) \leq (1 + \frac{2}{\epsilon})^n$. $\epsilon \in (0, 1] \Rightarrow (\frac{1}{\epsilon})^n \leq \mathcal{N}(B_2^n, \epsilon) \leq (\frac{3}{\epsilon})^n$.
- Hamming cube: $x, y \in \{0, 1\}^n, d_H(x, y) := \#\{i : x(i) \neq y(i)\}$.
- (T, d) a metric space, $K \subset T, \mathcal{C}(K, d, \epsilon)$ the smallest number of bits sufficient specify every points $x \in K$ with accuracy ϵ in the metric d . Then $\log_2 \mathcal{N}(K, d, \epsilon) \leq \mathcal{C}(K, d, \epsilon) \leq \log_2 \mathcal{N}(K, d, \frac{\epsilon}{2})$. $\log_2 \mathcal{N}(K, \epsilon)$ is often called the metric entropy of K .

Proof Lower bound. Assume $\mathcal{C}(K, d, \epsilon) \leq N$. There exists a transformation of $x \in K$ into bit strings of length N . A partition of K into at most 2^N subsets.

Upper bound. Assume $\log_2 \mathcal{N}(K, d, \frac{\epsilon}{2}) \leq N$. There exists an $\frac{\epsilon}{2}$ -net \mathcal{N} with $|\mathcal{N}| \leq 2^N$. To every point $x \in K$, assign a point $x_0 \in \mathcal{N}$ that is closest to x . The encoding $x \mapsto x_0$ represents points in K with accuracy ϵ . \square

- Error correcting code: Fix integers k, n and r . Encoder $\{0, 1\}^k \rightarrow \{0, 1\}^n$, Decoder $\{0, 1\}^n \rightarrow \{0, 1\}^k, D(y) = x$ if $x \in \{0, 1\}^k, y \in \{0, 1\}^n$ and $d_H(E(x), y) \leq r$.
- If $\log_2 \mathcal{P}(\{0, 1\}^n, d_H, 2r) \geq k$, then there exists an error correcting code, k bits $\rightarrow n$ bits, correct r error.
Proof $\exists \mathcal{P} \in \{0, 1\}^n, |\mathcal{P}| = 2^k$ s.t closed balls centered at \mathcal{P} with radii r are disjoint. $E : \{0, 1\}^k \rightarrow \mathcal{N}$ one to one; $D : \{0, 1\}^n \rightarrow \{0, 1\}^k$ nearest-neighbor decodes. \square
- If $n \geq k + 2r \log_2(\frac{en}{2r})$, then there exists an error correcting code that encodes k -bit strings into n -bit strings and can correct r errors.

Proof $\mathcal{P}(\{0, 1\}^n, d_H, 2r) \geq \mathcal{N}(\{0, 1\}^n, d_H, 2r) \geq \frac{2^n}{\sum_{k=0}^{2r} \binom{n}{k}} \geq 2^n (\frac{2r}{en})^{2r} \geq 2^k$. \square

- $A_{m \times n}$, $\epsilon \in [0, 1)$. Then for any ϵ -set \mathcal{N} of \mathbb{S}^{n-1} , $\sup_{x \in \mathcal{N}} \|Ax\|_2 \leq \|A\| \leq \frac{1}{1-\epsilon} \sup_{x \in \mathcal{N}} \|Ax\|_2$.

Proof Fix $x \in \mathbb{S}^{n-1}$, $\|A\| = \|Ax\|_2$. $\exists x_0 \in \mathcal{N}$, $\|x - x_0\|_2 \leq \epsilon$, $\|Ax - Ax_0\|_2 \leq \|A\| \|x - x_0\|_2 \leq \epsilon \|A\| \Rightarrow \|Ax_0\|_2 \geq \|Ax\|_2 - \epsilon \|A\|$.
 $\|A(x - x_0)\|_2 \geq \|A\| - \epsilon \|A\|$. \square

- $A_{m \times n} = \{A_{ij}\}$, A_{ij} independent mean zero sub-gaussian, $K = \max_{i,j} \|A_{ij}\|_{\psi_2}$. Then for any $t > 0$, $\mathbb{P}(\|A\| \leq CK(\sqrt{m} + \sqrt{n} + t)) \geq 1 - 2e^{-t^2}$.

Proof Step 1: Approximation. Choose $\epsilon = 1/4$ and ϵ -net \mathcal{N} of \mathbb{S}^{n-1} , ϵ -net \mathcal{M} of \mathbb{S}^{m-1} with $|\mathcal{N}| \leq 9^n, |\mathcal{M}| \leq 9^m$. Ex 4.4.3 $\Rightarrow \|A\| \leq 2 \max_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle$.

Step 2: Concentration. $\langle Ax, y \rangle = \sum_{i,j} A_{ij} x_i y_j$, $\|\langle Ax, y \rangle\|_{\psi_2}^2 \leq C \sum_{i,j} \|A_{ij}\|_{\psi_2}^2 x_i^2 y_j^2 \leq CK^2 \Rightarrow \mathbb{P}(\langle Ax, y \rangle \geq u) \leq 2e^{-cu^2/K^2}$.

Step 3: Union bound. $\mathbb{P}(\max_{x \in \mathcal{N}, y \in \mathcal{M}} \langle Ax, y \rangle \geq u) \leq \sum_{x \in \mathcal{N}, y \in \mathcal{M}} \mathbb{P}(\langle Ax, y \rangle \geq u) \leq 9^{n+m} 2e^{-cu^2/K^2}$. Take $u = CK(\sqrt{m} + \sqrt{n} + t)$, $u^2 \geq C^2 K^2 (m + n + t^2)$. C sufficiently large s.t. $cu^2/K^2 \geq 3(n + m + t^2)$. \square

- $A_{n \times n}$ symmetric, $A_{ij}, i \leq j$ independent mean zero sub-gaussian. Then for $t \geq 0$, $\mathbb{P}(\|A\| \leq CK(\sqrt{n} + t)) \geq 1 - 4e^{-t^2}$.

Proof $A = \underbrace{A^+ + A^-}_{\text{upper + lower triangular matrix}}$, $\mathbb{P}(\|A^+\| \leq CK(\sqrt{n} + t)) \geq 1 - 2e^{-4t^2} \geq 1 - 2e^{-t^2}$, $\mathbb{P}(\|A^-\| \leq CK(\sqrt{n} + t)) \geq 1 - 2e^{-4t^2} \geq 1 - 2e^{-t^2}$, $\mathbb{P}(\|A\| \leq CK(\sqrt{n} + t)) \geq \mathbb{P}(\|A^+\| \leq \frac{C}{2}K(\sqrt{n} + t), \|A^-\| \leq \frac{C}{2}K(\sqrt{n} + t)) \geq 2(1 - 2e^{-t^2}) - 1 = 1 - 4e^{-t^2}$. \square

- Stochastic block model (SBM): $G(n, p, q)$, $p > q$, n vertices, two community of size $n/2$, $x, y \in$ same community $\Rightarrow \mathbb{P}(x \sim y) = p$, otherwise $\mathbb{P}(x \sim y) = q$. $A = \{A_{ij}\}$, $A_{ij} = 1$ if $i \sim j$ otherwise 0. $A = \mathbb{E}A + R := D + R$, $\|D\| = \frac{p+q}{2} \cdot n$, $\mathbb{P}(\|R\| \leq C\sqrt{n}) \geq 1 - 4e^{-n}$.

- Weyl's inequality: Symmetric matrices S and T with same dim, $\max_i |\lambda_i(S) - \lambda_i(T)| \leq \|S - T\|$.

- Davis-Kahan: Fix i , $\min_{j \neq i} |\lambda_i(S) - \lambda_j(S)| = \delta > 0$. Then $\sin \angle(v_i(S), v_i(T)) \leq \frac{2\|S-T\|}{\delta} \Rightarrow \exists \theta \in \{-1, 1\}$, $\|v_i(S) - \theta v_i(T)\|_2 \leq \frac{\|S-T\|}{\delta} \cdot 2^{3/2}$.

- Spectral clustering: Recall SBM $A = D + R$ and let $S = D, T = A = D + R$ in Davis-Kahan. $\delta = \min(\lambda_2, \lambda_2 - \lambda_1) = \min(\frac{p-q}{2}, q)n := \mu n$. $\mathbb{P}(\|R\| = \|T - S\| \leq C\sqrt{n}) \geq 1 - 4e^{-n} \Rightarrow \exists \theta \in \{\pm 1\}$, $\|v_2(D) - \theta v_2(A)\| \leq \frac{C}{\mu\sqrt{n}}$. Let $u_2(D) = (1, 1, \dots, 1, -1, -1, \dots, -1) \Rightarrow \|u_2(D) - \theta u_2(A)\| \leq \frac{C}{\mu} \Rightarrow \sum_{j=1}^n |u_2(D)_j - \theta u_2(A)_j|^2 \leq \frac{C}{\mu^2}$. Thus the number of disagreeing signs between $u_2(D)$ and $u_2(A)$ must be bounded by $\frac{C}{\mu^2}$.

- $A_{m \times n}$, rows A_i independent mean zero sub-gaussian, isotropic. Then for any $t \geq 0$, $\sqrt{m} - CK^2(\sqrt{n} + t) \leq s_n(A) \leq s_1(A) \leq \sqrt{m} + CK^2(\sqrt{n} + t)$ with prob $\geq 1 - 2e^{-t^2}$. Here $K = \max_i \|A_i\|_{\psi_2}$.

Proof Only need to prove $\|\frac{1}{m}A^T A - I_n\| \leq \epsilon := K^2 \max\{\delta, \delta^2\}$, $\delta = C(\frac{\sqrt{n}}{\sqrt{m}} + \frac{t}{\sqrt{m}})$.

Step 1: Approximation. Find an $\frac{1}{4}$ -net \mathcal{N} of the unit space \mathbb{S}^{n-1} , $|\mathcal{N}| \leq 9^n$. $\|\frac{1}{m}A^T A - I_n\| \leq 2 \max_{x \in \mathcal{N}} |\langle (\frac{1}{m}A^T A - I_n)x, x \rangle| = 2 \max_{x \in \mathcal{N}} |\frac{1}{m}\|Ax\|_2^2 - 1|$.

Step 2: Concentration. $X_i := \langle A_i, x \rangle$ independent, mean zero, $\|X_i\|_{\psi_2} \leq K$, $\mathbb{E}X_i^2 = 1$. $\mathbb{P}(|\frac{1}{m}\|Ax\|_2^2 - 1| \geq \frac{\epsilon}{2}) \leq 2e^{-c_1 \delta^2 m} \leq 2e^{-c_1 C^2 (n+t^2)}$.

Step 3: Union bound. $\mathbb{P}(\max_{x \in \mathcal{N}} |\frac{1}{m}\|Ax\|_2^2 - 1| \geq \frac{\epsilon}{2}) \leq 9^n \cdot 2e^{-c_1 C^2 (n+t^2)} \leq 2e^{-t^2}$. \square

- $X \in \mathbb{R}^n$ sub-gaussian. $\mathbb{E}X = 0, \Sigma = \mathbb{E}X X^T, X_i \stackrel{d}{=} X$ i.i.d., $\Sigma_m = \frac{1}{m} \sum_{i=1}^m X_i X_i^T$. Assume there exists $K \geq 1$ s.t. $\|\langle X, x \rangle\|_{\psi_2}^2 \leq K^2 \|\langle X, x \rangle\|_{L^2}^2$. Then for m , $\mathbb{E}\|\Sigma_m - \Sigma\| \leq CK^2(\sqrt{\frac{n}{m}} + \frac{n}{m})\|\Sigma\|$.

Proof $Z_i = \Sigma^{-1/2} X_i, Z = \Sigma^{-1/2} X, \mathbb{E}Z_i Z_i^T = I_n, \|Z\|_{\psi_2} \leq K, \|Z_i\|_{\psi_2} \leq K$. Then $\|\Sigma_m - \Sigma\| = \|\Sigma^{1/2} R_m \Sigma^{1/2}\| \leq \|R_m\| \|\Sigma\|$ where $R_m = \frac{1}{m} \sum_{i=1}^m Z_i Z_i^T - I$. Consider an $m \times n$ random matrix A whose rows are Z_i^T . $\mathbb{E}\|R_m\| = \mathbb{E}\|\frac{1}{m}A^T A - I\| \leq CK^2(\sqrt{\frac{n}{m}} + \frac{n}{m})$. \square

5 Concentration without independence

- $(X, d_X) \xrightarrow{f} (Y, d_Y), d_Y(f(u), f(v)) \leq L \cdot d_X(u, v), \forall u, v \in X$. The infimum of all L in this definition is called the Lipschitz norm of f and is denoted $\|f\|_{\text{Lip}}$.

- $\epsilon > 0, A_\epsilon = A + \epsilon B_2^n, A \subset \mathbb{R}^n$, \min_A volume of A_ϵ with colume A fixed is achieved when A is a ball.

- $\sigma_{n-1}(A)$ normalized area on \mathbb{S}^{n-1} , $\epsilon > 0$. With given $\sigma_{n-1}(A)$, $\min_A \sigma_{n-1}(A_\epsilon)$ is achieved when A is a spherical cap.
- $A \subset \sqrt{n}\mathbb{S}^{n-1}$. If $\sigma(A) \geq \frac{1}{2}$, then $\forall t \geq 0, \sigma(A_t) \geq 1 - 2e^{-ct^2}$.
Proof Let $H = \{x \in \sqrt{n}\mathbb{S}^{n-1}, x_1 \leq 0\}$, $\sigma(A) \geq \sigma(H)$. The latest thm $\Rightarrow \sigma(A_t) \geq \sigma(H_t) = \mathbb{P}(X \in H_t)$. $H_t \supset \{x \in \sqrt{n}\mathbb{S}^{n-1}, x_1 \leq \frac{t}{\sqrt{2}}\} \Rightarrow \sigma(H_t) \geq \mathbb{P}(X_1 \leq \frac{t}{\sqrt{2}})$. $\|X_1\|_{\psi_2} \leq C \Rightarrow \mathbb{P}(X_1 \leq \frac{t}{\sqrt{2}}) \geq 1 - 2e^{-ct^2}$. \square
- $X \sim \text{Unif}(\sqrt{n}\mathbb{S}^{n-1})$, $f : \sqrt{n}\mathbb{S}^{n-1} \rightarrow \mathbb{R}$. Then $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq C\|f\|_{\text{Lip}}$.
Proof WLOG $\|f\|_{\text{Lip}} = 1$, $\mathbb{P}(f(X) \geq M) \geq \frac{1}{2}, \mathbb{P}(f(X) \leq -M) \geq \frac{1}{2}$. $A := \{x \in \sqrt{n}\mathbb{S}^{n-1} : f(x) \leq M\}$. $\mathbb{P}(X \in A) \geq \frac{1}{2} \Rightarrow \mathbb{P}(A_t) \geq 1 - 2e^{-ct^2} \Rightarrow \mathbb{P}(f(X) \leq M+t) \geq 1 - 2e^{-ct^2}$. By centering, $f(X) - \mathbb{E}f(X) = f(X) - M - (\mathbb{E}f(X) - M)$ is sub-gaussian. \square
- $X \sim \mathcal{N}(0, I_n)$, $\gamma_n(A) = \mathbb{P}(X \in A)$, $\epsilon > 0$, $\gamma_n(A)$ given, half spaces minimize $\gamma_n(A_\epsilon)$.
- $X \sim \mathcal{N}(0, I_n)$, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\|f\|_{\text{Lip}} < \infty$. Then $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq C\|f\|_{\text{Lip}}$.
- Hamming cube, $d(x, y) = \frac{1}{n}|\{i : x_i \neq y_i\}|$, $\mathbb{P}(A) = \frac{|A|}{2^n}$. $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq \frac{C\|f\|_{\text{Lip}}}{\sqrt{n}}$.
- $S_n : n!$ permutation of n symbols. $d(\pi, \rho) = \frac{1}{n}|\{i : \pi(i) \neq \rho(i)\}|$, $\mathbb{P}(A) = \frac{|A|}{n!}$. $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq \frac{C\|f\|_{\text{Lip}}}{\sqrt{n}}$.
- Special orthogonal group $\text{SO}(n)$, determinant = 1, $d = \|\cdot\|_F$, \mathbb{P} is uniform measure. $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq \frac{C\|f\|_{\text{Lip}}}{\sqrt{n}}$.
- $G_{n,m}$ all m -dim subspaces of \mathbb{R}^n ($\simeq \mathcal{P}_{G_{n,m}}$ orthogonal projections), $d(E, F) = \|\mathcal{P}_E - \mathcal{P}_F\|$, \mathbb{P} is uniform measure. A random subspace E can be constructed by computing the column span (i.e. the image) of a random $n \times m$ Gaussian random matrix G with i.i.d. $\mathcal{N}(0, 1)$ entries. $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq \frac{C\|f\|_{\text{Lip}}}{\sqrt{n}}$.
- A random vector X in \mathbb{R}^n with density $p(x) = e^{-U(x)}$, Hess $U(x) \succeq \kappa I_n$. $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq \frac{C\|f\|_{\text{Lip}}}{\sqrt{\kappa}}$.
- $X = (X_1, \dots, X_n)$ independent coordinates, $|X_i| \leq 1$ a.s., f convex and Lipschitz. $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq C\|f\|_{\text{Lip}}$.
- $E \sim \text{Unif}(G_{n,m})$, $z \in \mathbb{R}^n$, $\epsilon > 0$. Then (a) $(\mathbb{E}\|P_E z\|_2^2)^{\frac{1}{2}} = \sqrt{\frac{m}{n}}\|z\|_2$; (b) $\mathbb{P}(|\|P_E z\|_2 - \sqrt{\frac{m}{n}}\|z\|_2| \leq \epsilon\sqrt{\frac{m}{n}}\|z\|_2) \geq 1 - 2e^{-c\epsilon^2 m}$.
Proof (a): WLOG $\|z\|_2 = 1$. Rotational invariance: $\mathbb{P}(E \in A) = \mathbb{P}(U(E) \in A)$ where U is $n \times n$ orthogonal \Rightarrow The dist. of $P_E z$ is the same if we fix E , $z \in \text{Unif}(\mathbb{S}^{n-1})$. WLOG $Pz = (z_1, \dots, z_m, 0, \dots, 0)$. $\mathbb{E}\|Pz\|_2^2 = m\mathbb{E}z_1^2 = \frac{m}{n}$.
 (b): $f : z \rightarrow \|Pz\|_2, \|f\|_{\text{Lip}} = 1 \Rightarrow \mathbb{P}(|\|Pz\|_2 - \mathbb{E}\|Pz\|_2| \leq \frac{C}{\sqrt{n}}) \geq 1 - 2e^{-c\epsilon^2 m}$. \square
- Johnson-Lindenstrauss lemma: \mathcal{X} a set of N points in \mathbb{R}^n , $\epsilon > 0$, $m \geq \frac{C}{\epsilon^2} \log N$, $E \sim \text{Unif}(G_{n,m})$, $Q = \sqrt{\frac{n}{m}}\mathcal{P}_E$. Then $\mathbb{P}(|\|Qx - Qy\|_2 - \|x - y\|_2| \leq \epsilon\|x - y\|_2 \text{ for any } x, y \in \mathcal{X}) \geq 1 - 2e^{-c\epsilon^2 m}$.
Proof Let $\mathcal{X} - \mathcal{X} := \{x - y : x, y \in \mathcal{X}\}$. The latest lemma $\Rightarrow \forall z, \mathbb{P}((1 - \epsilon)\sqrt{\frac{m}{n}}\|z\|_2 \leq \|Pz\|_2 \leq (1 + \epsilon)\sqrt{\frac{m}{n}}\|z\|_2) \geq 1 - 2e^{-c\epsilon^2 m}$. Union bound: $\mathbb{P}(\dots, \text{ for any } z \in \mathcal{X} - \mathcal{X}) \geq 1 - N^2 \cdot 2e^{-c\epsilon^2 m} \geq 1 - 2e^{-c'\epsilon^2 m}$. \square
- $f : \mathbb{R} \rightarrow \mathbb{R}$, $X = \sum_{i=1}^n \lambda_i u_i u_i^T$, define $f(X) = \sum_{i=1}^n f(\lambda_i) u_i u_i^T$.
- P.S.D. order: $X \succeq 0, X \succeq Y$ if $X - Y \succeq 0$.
- Golden-Thompson inequality: $\text{tr}(e^{A+B}) \leq \text{tr}(e^A e^B)$.
- Lieb's inequality: $H : n \times n$ symmetric matrix, X P.D., $f(X) = \text{tr}(e^{H+\log X})$. Then f is concave.
- X is a random P.D. matrix $\Rightarrow \mathbb{E}f(X) \leq f(\mathbb{E}X)$. $X = e^Z$, Z symmetric. Then $\mathbb{E}\text{tr}(e^{H+Z}) \leq \text{tr}(e^{H+\log \mathbb{E}e^Z})$.
- X_1, \dots, X_N independent mean zero $n \times n$ symmetric random matrices, $\|X_i\| \leq K$ a.s. for all i . Then for $\forall t \geq 0$, $\mathbb{P}(\|\sum_{i=1}^N X_i\| \geq t) \leq 2ne^{-\frac{t^2/2}{\sigma^2 + Kt/3}}$ where $\sigma^2 = \|\sum_{i=1}^N \mathbb{E}X_i^2\|$.
Proof Step 1: Reduction to MGF. $S := \sum_{i=1}^N X_i$. $\|S\| = \max_i |\lambda_i(S)| = \max(\lambda_{\max}(S), \lambda_{\max}(-S))$. $\mathbb{P}(\lambda_{\max}(S) \geq t) \leq e^{-\lambda t} \mathbb{E}e^{\lambda \lambda_{\max}(S)}$. $E := \mathbb{E}e^{\lambda \lambda_{\max}(S)} = \mathbb{E}\lambda_{\max}(e^{\lambda S}) \Rightarrow E \leq \mathbb{E}\text{tr}(e^{\lambda S})$.
 Step 2: Apply Lieb's inequality. $\mathbb{E}\text{tr}(e^{\lambda S}) = \mathbb{E}\text{tr}(e^{\sum_{i=1}^{N-1} \lambda X_i + \lambda X_N}) \leq \mathbb{E}\text{tr}(e^{\sum_{i=1}^{N-1} \lambda X_i + \log \mathbb{E}e^{\lambda X_N}}) \leq \text{tr}(e^{\sum_{i=1}^N \log \mathbb{E}e^{\lambda X_i}})$.
 Step 3: Lemma: X is an $n \times n$ symmetric mean zero random matrix, $\|X\| \leq K$ a.s. Then $\mathbb{E}e^{\lambda X} \preceq e^{g(\lambda)\mathbb{E}X^2}$ where $g(\lambda) = \frac{\lambda^2/2}{1 - |\lambda|K/3}, |\lambda| < 3/K$.

Proof $e^z \leq 1 + z + \frac{1}{1-|z|/3} \frac{z^2}{2}$ if $|z| < 3$. Let $z = \lambda x$. If $|x| \leq K, |\lambda| < \frac{3}{K}$, then $e^{\lambda x} \leq 1 + \lambda x + g(\lambda)x^2$. (b) of Ex. 5.4.5 \Rightarrow If $\|X\| \leq K, |\lambda| < 3/K, \mathbb{E}e^{\lambda X} \preceq I + g(\lambda)\mathbb{E}X^2$ (since $\mathbb{E}X = 0$) $\preceq e^{g(\lambda)\mathbb{E}X^2}$. \square

Step 4: $E \leq \text{tr}(e^{\sum_{i=1}^N \log \mathbb{E}e^{\lambda X_i}})$. The latest lemma + (g) of Ex.5.4.5 $\Rightarrow \log \mathbb{E}e^{\lambda X_i} \preceq g(\lambda)\mathbb{E}X_i^2 \Rightarrow \sum_{i=1}^N \log \mathbb{E}e^{\lambda X_i} \preceq g(\lambda) \cdot Z$ where $Z := \sum_{i=1}^N \mathbb{E}X_i^2$ and $\sigma^2 = \|Z\|$. (e) of Ex.5.4.5 $\Rightarrow \text{tr}(e^{\sum_{i=1}^N \log \mathbb{E}e^{\lambda X_i}}) \leq \text{tr}(e^{g(\lambda)Z}) \Rightarrow E \leq \text{tr}(e^{g(\lambda)Z}) \leq n\lambda_{\max}(e^{g(\lambda)Z}) = ne^{g(\lambda)\|Z\|} = ne^{g(\lambda)\sigma^2}$. Minimize for λ as a function of t with $0 < \lambda < 3/K$. \square

- $X \in \mathbb{R}^n, \Sigma = \mathbb{E}XX^T, \Sigma_m = \frac{1}{m} \sum_{i=1}^m X_i X_i^T, X_i \stackrel{\text{i.i.d.}}{\sim} X, \|X\|_2 \leq K(\mathbb{E}\|X\|_2^2)^{\frac{1}{2}}$ a.s.. Then $\mathbb{E}\|\Sigma_m - \Sigma\| \leq C(\sqrt{\frac{K^2 n \log n}{m}} + \frac{K^2 n \log n}{m})\|\Sigma\|$.

Proof $\mathbb{E}\|X\|_2^2 = \mathbb{E}X^T X = \mathbb{E}\text{tr}(X^T X) = \mathbb{E}\text{tr}(XX^T) = \text{tr}(\Sigma) \Rightarrow \|X\|_2^2 \leq K^2 \text{tr}(\Sigma)$ a.s.. Ex 5.4.11 $\Rightarrow \mathbb{E}\|\Sigma_m - \Sigma\| = \frac{1}{m}\mathbb{E}\|\sum_{i=1}^m (X_i X_i^T - \Sigma)\| \leq \frac{1}{m}(\sigma\sqrt{\log n} + M \log n)$ where $\sigma^2 = \|\sum_{i=1}^m \mathbb{E}(X_i X_i^T - \Sigma)^2\| = m\|\mathbb{E}(XX^T - \Sigma)^2\|$ and M is chosen s.t. $\|XX^T - \Sigma\| \leq M$ a.s.. Then $\mathbb{E}(XX^T - \Sigma)^2 = \mathbb{E}(XX^T)^2 - \Sigma^2 \preceq \mathbb{E}(XX^T)^2 = \mathbb{E}(\|X\|_2^2 XX^T) \preceq K^2 \text{tr}(\Sigma)\Sigma \Rightarrow \sigma^2 \leq K^2 m \text{tr}(\Sigma)\|\Sigma\|$. $\|XX^T - \Sigma\| \leq \|X\|_2^2 + \|\Sigma\| \leq K^2 \text{tr}(\Sigma) + \|\Sigma\| \leq 2K^2 \text{tr}(\Sigma) := M$ (since $K \geq 1$ and $\|\Sigma\| \leq \text{tr}(\Sigma)$). Substitute our bounds for σ^2 and M into the previous bound $\frac{1}{m}(\sigma\sqrt{\log n} + M \log n)$. \square

6 Quadratic forms, symmetrization, contraction

- $Y \perp Z, \mathbb{E}Z = 0, F$ convex, then $\mathbb{E}F(Y) \leq \mathbb{E}F(Y + Z)$.

Proof $F(y) = F(\mathbb{E}(y + Z)) \leq \mathbb{E}F(y + Z) \Rightarrow \mathbb{E}F(Y) = \mathbb{E}(\mathbb{E}(F(Y + \mathbb{E}Z)|Y)) = \mathbb{E}(\mathbb{E}(F(\mathbb{E}(Y + Z)|Y))) \leq \mathbb{E}(\mathbb{E}(F(Y + Z)|Y)) = \mathbb{E}F(Y + Z)$. \square

- Decoupling: $A_{n \times n}$ diagonal-free (i.e. the diagonal entries of A equal zero), $X = (X_1, \dots, X_n)$ independent mean zero. Then for every convex function $F: \mathbb{R} \rightarrow \mathbb{R}, \mathbb{E}F(X^T A X) \leq \mathbb{E}F(4X^T A X')$ where $X' \stackrel{\text{i.i.d.}}{\sim} X, X' \perp X$.

Proof $\delta_1, \dots, \delta_n \stackrel{\text{i.i.d.}}{\sim} \text{Ber}(1, \frac{1}{2}), I = \{i : \delta_i = 1\}, \mathbb{E}\delta_i(1 - \delta_j) = \frac{1}{4}, X^T A X = \sum_{i \neq j} a_{ij} X_i X_j = 4\mathbb{E}\delta \sum_{i,j} \delta_i(1 - \delta_j) a_{ij} X_i X_j = 4\mathbb{E}I \sum_{(i,j) \in I \times I^c} a_{ij} X_i X_j \Rightarrow \mathbb{E}_X F(X^T A X) = \mathbb{E}_X F(4\mathbb{E}I \sum_{(i,j) \in I \times I^c} a_{ij} X_i X_j) \leq \mathbb{E}_I \mathbb{E}_X F(4 \sum_{(i,j) \in I \times I^c} a_{ij} X_i X_j)$. There exists an I s.t. $\mathbb{E}_X F(4 \sum_{(i,j) \in I \times I^c} a_{ij} X_i X_j) = \mathbb{E}_X F(4 \sum_{(i,j) \in I \times I^c} a_{ij} X_i X_j') \geq \mathbb{E}F(X^T A X)$. LHS $\leq \mathbb{E}_X F(4 \sum_{i,j} a_{ij} X_i X_j')$ by the latest lemma since $\mathbb{E}[(\sum_{(i,j) \in I \times I} + \sum_{(i,j) \in I^c \times I^c} + \sum_{(i,j) \in I^c \times I}) a_{ij} X_i X_j'] | \{X_i, i \in I\}, \{X_j', j \in I^c\}] = 0$. \square

- $X, X' \sim \mathcal{N}(0, I_n), X \perp X'$, then $\mathbb{E}e^{\lambda X^T A X'} \leq e^{C\lambda^2 \|A\|_F^2}, |\lambda| \leq \frac{c}{\|A\|}$.

Proof $A = \sum_i s_i u_i v_i^T, X^T A X' = \sum_i s_i \underbrace{\langle u_i, X \rangle}_{:=g_i} \underbrace{\langle v_i, X' \rangle}_{:=g_i'}$. $(g_1, \dots, g_n) \perp (g_1', \dots, g_n') \sim \mathcal{N}(0, I_n) \Rightarrow \mathbb{E}e^{\lambda X^T A X'} = \prod_{i=1}^n \mathbb{E}e^{\lambda s_i g_i g_i'} = \prod_{i=1}^n \mathbb{E}e^{\lambda^2 s_i^2 g_i'^2 / 2} \leq \prod_{i=1}^n e^{C\lambda^2 s_i^2 (\lambda^2 s_i^2 \leq c)} \leq e^{C\lambda^2 \|A\|_F^2} (\lambda^2 \leq \frac{c}{\max_i s_i^2} = \frac{c}{\|A\|^2})$. \square

- X, X' independent sub-gaussian mean zero, $\|X\|_{\psi_2} \leq K, \|X'\|_{\psi_2} \leq K. g, g' \sim \mathcal{N}(0, I_n), g \perp g'$. Then $\mathbb{E}e^{\lambda X^T A X'} \leq \mathbb{E}e^{CK^2 \lambda g^T A g'}$.

Proof Conditioned on $X', \mathbb{E}_X e^{\lambda X^T A X'} \leq e^{C\lambda^2 K^2 \|A X'\|_2^2}, \mathbb{E}_g e^{\mu g^T A X'} = e^{\frac{\mu^2 \|A X'\|_2^2}{2}}. \mu = \sqrt{2c}K\lambda \Rightarrow \mathbb{E}_X e^{\lambda X^T A X'} \leq \mathbb{E}_g e^{\sqrt{2c}K\lambda g^T A X'} \Rightarrow \mathbb{E}e^{\lambda X^T A X'} \leq \mathbb{E}e^{\sqrt{2c}K\lambda g^T A g'} \leq \mathbb{E}e^{2cK^2 \lambda g^T A g'}$. \square

- Hanson-Wright inequality: $X = (X_1, \dots, X_n)$ independent mean zero sub-gaussian, then $\mathbb{P}(|X^T A X - \mathbb{E}X^T A X| \geq t) \leq 2e^{-c \min(\frac{t^2}{K^4 \|A\|_F^2}, \frac{t}{K^2 \|A\|})}$.

Proof WLOG $K = 1. X^T A X = \sum_{i,j} a_{ij} X_i X_j, \mathbb{E}X^T A X = \sum_i a_{ii} \mathbb{E}X_i^2, X^T A X - \mathbb{E}X^T A X = \sum_i a_{ii} (X_i^2 - \mathbb{E}X_i^2) + \sum_{i \neq j} a_{ij} X_i X_j$. $p := \mathbb{P}(X^T A X - \mathbb{E}X^T A X \geq t) \leq \mathbb{P}(\sum_i a_{ii} (X_i^2 - \mathbb{E}X_i^2) \geq \frac{t}{2}) + \mathbb{P}(\sum_{i \neq j} a_{ij} X_i X_j \geq \frac{t}{2}) := p_1 + p_2$.

Step 1: $\|X_i^2 - \mathbb{E}X_i^2\|_{\psi_1} \lesssim 1$. Bernstein $\Rightarrow p_1 \leq e^{-c \min(\frac{t^2}{\sum a_{ii}^2}, \frac{t}{\max_i |a_{ii}|})} \leq e^{-c \min(\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|})}$.

Step 2: $S := \sum_{i \neq j} a_{ij} X_i X_j. p_2 \leq e^{-\frac{\lambda t}{2}} \mathbb{E}e^{\lambda S}, \mathbb{E}e^{\lambda S} \leq \mathbb{E}e^{4\lambda X^T A X'} \leq \mathbb{E}e^{c_1 \lambda g^T A g'} \leq e^{C\lambda^2 \|A\|_F^2}$ (with $\lambda \leq \frac{c}{\|A\|}$). \square

- $B_{m \times n}, X \in \mathbb{R}^n, \{X_i\}$ independent mean-zero, unit-variance, sub-gaussian. Then $|\|BX\|_2 - \|B\|_F|_{\psi_2} \leq CK^2 \|B\|, K = \max_i \|X_i\|_{\psi_2}$.

Proof $A = B^T B, X^T A X = \|BX\|_2^2, \mathbb{E}X^T A X = \|B\|_F^2, \|A\| = \|B\|^2, \|A\|_F = \|B^T B\|_F \leq \|B^T\| \|B\|_F = \|B\| \|B\|_F$. Thus $\forall u \geq 0, \mathbb{P}(|\|BX\|_2^2 - \|B\|_F^2| \geq u) \leq e^{-\frac{c}{K^4} \min(\frac{u^2}{\|B\|^2 \|B\|_F^2}, \frac{u}{\|B\|^2})}$. Let $u = \epsilon \|B\|_F^2, \mathbb{P}(|\|BX\|_2^2 - \|B\|_F^2| \geq \epsilon \|B\|_F^2) \leq 2e^{-c \min(\epsilon^2, \epsilon) \frac{\|B\|_F^2}{K^4 \|B\|^2}}$. Let $\delta^2 = \min(\epsilon^2, \epsilon)$, then $\epsilon = \max(\delta, \delta^2), |\|BX\| - \|B\|_F| \geq \delta \|B\|_F \Rightarrow |\|BX\|_2^2 - \|B\|_F^2| \geq \epsilon \|B\|_F^2 \Rightarrow \mathbb{P}(|\|BX\|_2 - \|B\|_F| \geq \delta \|B\|_F) \leq 2e^{-c\delta^2 \frac{\|B\|_F^2}{K^4 \|B\|^2}}$. \square

- Symmetrization: X_1, X_2, \dots, X_N independent, mean zero in a normed space, $\epsilon_1, \epsilon_2, \dots, \epsilon_N$ a sequence of independent symmetric Bernoulli random variables. Then $\frac{1}{2}\mathbb{E}\|\sum_{i=1}^N \epsilon_i X_i\| \leq \mathbb{E}\|\sum_{i=1}^N X_i\| \leq 2\mathbb{E}\|\sum_{i=1}^N \epsilon_i X_i\|$.

Proof Upper bound. $X' \perp X, X' \stackrel{d}{=} X$. $p = \mathbb{E}\|\sum_{i=1}^N X_i\| \leq \mathbb{E}\|\sum_{i=1}^N X_i - \sum_{i=1}^N X'_i\| = \mathbb{E}\|\sum_{i=1}^N \epsilon_i(X_i - X'_i)\| \leq \mathbb{E}\|\sum_{i=1}^N \epsilon_i X_i\| + \mathbb{E}\|\sum_{i=1}^N \epsilon_i X'_i\| = 2\mathbb{E}\|\sum_{i=1}^N \epsilon_i X_i\|$. \square

- $A_{n \times n}$ symmetric independent mean zero. Then $\mathbb{E}\|A\| \leq C\sqrt{\log n}\mathbb{E}\max\|A_i\|_2$ where A_i is i -th row of A .

Proof $A = \sum_{i \leq j} Z_{ij}$ independent mean zero symmetric where $Z_{ij} = \begin{cases} A_{ij}(e_i e_j^T + e_j e_i^T), & i \leq j \\ A_{ii} e_i e_i^T & i = j \end{cases} \Rightarrow \mathbb{E}\|A\| \leq 2\mathbb{E}\|\sum_{i \leq j} \epsilon_{ij} Z_{ij}\|$.

Ex 5.4.13(a) \Rightarrow Conditioned on $\{Z_{ij}\}$, $\mathbb{E}_\epsilon\|\sum_{i \leq j} \epsilon_{ij} Z_{ij}\| \leq C\sqrt{\log n}\|\sum_{i \leq j} Z_{ij}^2\|^{1/2} \Rightarrow \mathbb{E}\|\sum_{i \leq j} \epsilon_{ij} Z_{ij}\| \leq C\sqrt{\log n}\mathbb{E}\|\sum_{i \leq j} Z_{ij}^2\|^{1/2}$, $\sum_{i \leq j} Z_{ij}^2 = \sum_{i=1}^n (\sum_{j=1}^n A_{ij}^2) e_i e_i^T = \sum_{i=1}^n \|A_i\|_2^2 e_i e_i^T \Rightarrow \|\sum_{i \leq j} Z_{ij}^2\|^{1/2} = \max\|A_i\|_2$. \square

- Matrix completion: $X_{n \times n}$, $\text{rank}(X) = r \ll n, Y_{ij} = \delta_{ij} X_{ij}, \delta_{ij} \sim \text{Ber}(p), p = \frac{m}{n^2}, \hat{X} = \arg \min_{\text{rank}(A') \leq r} \|p^{-1}Y - A'\|$. Then $\mathbb{E}\frac{1}{n}\|\hat{X} - X\|_F \leq C\sqrt{\frac{rn \log n}{m}}\|X\|_\infty$.

Proof Step 1. $\|\hat{X} - X\| \leq \|\hat{X} - p^{-1}Y\| + \|p^{-1}Y - X\| \leq 2\|p^{-1}Y - X\| = \frac{2}{p}\|Y - pX\|$. $(Y - pX)_{ij} = (\delta_{ij} - p)X_{ij}$ independent mean zero, Ex 6.5.2 $\Rightarrow \mathbb{E}\|Y - pX\| \leq C\sqrt{\log n}(\mathbb{E}\max_i\|(Y - pX)_i\|_2 + \mathbb{E}\max_j\|(Y - pX)_j\|_2)$. $\|(Y - pX)_i\|_2^2 = \sum_{j=1}^n (\delta_{ij} - p)^2 X_{ij}^2 \leq \sum_{j=1}^n (\delta_{ij} - p)^2 \|X\|_\infty^2$. Ex 6.6.2 $\Rightarrow \mathbb{E}\max_i \sum_{j=1}^n (\delta_{ij} - p)^2 \leq Cpn \Rightarrow \frac{2}{p}\|Y - pX\| \leq C\sqrt{\frac{n \log n}{p}}\|X\|_\infty$.

Step 2. $\text{rank}(X) \leq r, \text{rank}(\hat{X}) \leq r, \text{rank}(\hat{X} - X) \leq 2r$. $\|\hat{X} - X\|_F \leq \sqrt{2r}\|\hat{X} - X\| \Rightarrow \mathbb{E}\|\hat{X} - X\|_F \leq C\sqrt{\frac{rn \log n}{p}}\|X\|_\infty$. \square

- Contraction principle: X_1, \dots, X_N vectors in some normed space, $a = (a_1, \dots, a_N) \in \mathbb{R}^N$. Then $\mathbb{E}\|\sum_{i=1}^N a_i \epsilon_i X_i\| \leq \|a\|_\infty \mathbb{E}\|\sum_{i=1}^N \epsilon_i X_i\|$.

Proof WLOG $\|a\|_\infty \leq 1, f(a) = \mathbb{E}\|\sum_{i=1}^N a_i \epsilon_i X_i\|$ is convex, which implies the maximum of f is attained at the boundary. Thus $f(a) \leq f(a^*) = \mathbb{E}\|\sum_{i=1}^N \epsilon_i X_i\|$ with $a_i^* = 1$ or -1 . \square

- Symmetrization with Gaussians: X_1, \dots, X_N independent mean zero, $g_1, \dots, g_N \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), \frac{C}{\sqrt{\log N}}\mathbb{E}\|\sum_{i=1}^N g_i X_i\| \leq \mathbb{E}\|\sum_{i=1}^N X_i\| \leq 3\mathbb{E}\|\sum_{i=1}^N g_i X_i\|$.

Proof Upper: $\mathbb{E}\|\sum_{i=1}^N X_i\| \leq 2\mathbb{E}\|\sum_{i=1}^N \epsilon_i X_i\| = 2\sqrt{\frac{\pi}{2}}\mathbb{E}_{X, \epsilon}\|\sum_{i=1}^N \epsilon_i \mathbb{E}_g |g_i| X_i\| \leq 2\sqrt{\frac{\pi}{2}}\mathbb{E}\|\sum_{i=1}^N \epsilon_i |g_i| X_i\| = 2\sqrt{\frac{\pi}{2}}\mathbb{E}\|\sum_{i=1}^N g_i X_i\|$.

Lower: $\mathbb{E}\|\sum_{i=1}^N g_i X_i\| = \mathbb{E}\|\sum_{i=1}^N \epsilon_i g_i X_i\| \leq \mathbb{E}_g \mathbb{E}_X (\|g\|_\infty \mathbb{E}_\epsilon \|\sum_{i=1}^N \epsilon_i X_i\|) = \mathbb{E}_g \|g\|_\infty \mathbb{E}_{X, \epsilon} \|\sum_{i=1}^N \epsilon_i X_i\| \leq 2\mathbb{E}_g \|g\|_\infty \mathbb{E}_X \|\sum_{i=1}^N X_i\| \leq C\sqrt{\log N} \mathbb{E}_X \|\sum_{i=1}^N X_i\|$. \square

7 Random processes

- Basic concepts: $\{X_t\}_{t \in T} \subset \mathbb{R}^n, \mathbb{E}X_t = 0, \forall t \in T, \Sigma(t, s) = \text{Cov}(X_t, X_s) = \mathbb{E}X_t X_s, d(t, s) = \|X_t - X_s\|_{L^2} = (\mathbb{E}(X_t - X_s)^2)^{1/2}$ (increments).

- Gaussian process: $T_0 \subset T, |T_0| < \infty, \{X_t\}_{t \in T_0}$ has normal distribution.

- Y is a mean zero Gaussian r.v. in \mathbb{R}^n . Then there exists $t_1, \dots, t_n \in \mathbb{R}^n$ s.t. $Y \stackrel{d}{=} (\langle g, t_i \rangle)_{i=1}^n, g \sim \mathcal{N}(0, I_n)$.

- Gaussian integration by parts: $X \sim \mathcal{N}(0, 1)$. Then for f differentiable, $\mathbb{E}f'(X) = \mathbb{E}Xf(X)$.

Proof f has bounded support: $\mathbb{E}f(X) = \int_{\mathbb{R}} f'(x)\phi(x)dx = -\int_{\mathbb{R}} f(x)\phi'(x)dx$. General $f: f_n \rightarrow f$. \square

- $X \sim \mathcal{N}(0, \Sigma), f: \mathbb{R}^n \rightarrow \mathbb{R}$. Then $\mathbb{E}Xf(X) = \Sigma \cdot \mathbb{E}\nabla f(X)$.

- $X \sim \mathcal{N}(0, \Sigma^X), Y \sim \mathcal{N}(0, \Sigma^Y), X \perp Y, Z(u) = \sqrt{u}X + \sqrt{1-u}Y, u \in [0, 1]$. Then $f: \mathbb{R}^n \rightarrow \mathbb{R}$ twice-differentiable, $\frac{d}{du}\mathbb{E}f(Z(u)) = \frac{1}{2}\sum_{i,j=1}^n (\Sigma_{ij}^X - \Sigma_{ij}^Y)\mathbb{E}[\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(u))]$.

Proof $\frac{d}{du}\mathbb{E}f(Z(u)) = \frac{1}{2}\sum_{i=1}^n \mathbb{E}\frac{\partial f}{\partial x_i}(Z(u))(\frac{X_i}{\sqrt{u}} - \frac{Y_i}{\sqrt{1-u}})$. $\sum_{i=1}^n \frac{1}{\sqrt{u}}\mathbb{E}X_i \frac{\partial f}{\partial x_i}(Z(u)) := \sum_{i=1}^n \frac{1}{\sqrt{u}}\mathbb{E}X_i g_i(X)$ (conditioned on Y) where $g_i(X) := \frac{\partial f}{\partial x_i}(\sqrt{u}X + \sqrt{1-u}Y)$. $\mathbb{E}X_i g_i(X) = \sum_{j=1}^n \Sigma_{ij}^X \mathbb{E}\frac{\partial g_i}{\partial x_j}(X) = \sum_{j=1}^n \Sigma_{ij}^X \mathbb{E}\frac{\partial^2 f}{\partial x_i \partial x_j}(\sqrt{u}X + \sqrt{1-u}Y)\sqrt{u}$. \square

- $\mathbb{E}X_i^2 = \mathbb{E}Y_i^2, \mathbb{E}(X_i - X_j)^2 \leq \mathbb{E}(Y_i - Y_j)^2, \frac{\partial^2 f}{\partial x_i \partial x_j} \geq 0, \forall i \neq j$. Then $\mathbb{E}f(X) \geq \mathbb{E}f(Y)$.

- Slepian's inequality: Let $\{X_t\}_{t \in T}$ and $\{Y_t\}_{t \in T}$ be two mean zero Gaussian processes. Assume $\mathbb{E}X_t^2 = \mathbb{E}Y_t^2, \mathbb{E}(X_t - X_s)^2 \leq \mathbb{E}(Y_t - Y_s)^2$. Then for every $t \in \mathbb{R}, \mathbb{P}(\sup_{t \in T} X_t \geq t) \leq \mathbb{P}(\sup_{t \in T} Y_t \geq t)$ and $\mathbb{E}\sup_{t \in T} X_t \leq \mathbb{E}\sup_{t \in T} Y_t$.

Proof Let $f(x) \approx 1_{\{\max x_i < t\}} = \prod_{i=1}^n 1_{\{x_i < t\}}$ and use the latest lemma. \square

- Sudakov-Fernique's inequality: Let $\{X_t\}_{t \in T}$ and $\{Y_t\}_{t \in T}$ be two mean zero Gaussian processes. Assume $\mathbb{E}(X_t - X_s)^2 \leq \mathbb{E}(Y_t - Y_s)^2$. Then $\mathbb{E} \sup_{t \in T} X_t \leq \mathbb{E} \sup_{t \in T} Y_t$.

Proof Let $f(x) = \frac{1}{\beta} \log \sum_{i=1}^n e^{\beta x_i}$. $f(x) \rightarrow \max x_i$ as $\beta \rightarrow \infty$. $\frac{\partial^2 f}{\partial x_i \partial x_j} \leq 0$. \square

- $A_{m \times n}$ independent $\mathcal{N}(0, 1)$ entries. Then $\mathbb{E} \|A\| \leq \sqrt{m} + \sqrt{n}$.

Proof $\max_{u \in \mathbb{S}^{n-1}, v \in \mathbb{S}^{m-1}} \langle Au, v \rangle := \max_{(u,v) \in T} X_{uv}$ where $T = \mathbb{S}^{n-1} \times \mathbb{S}^{m-1}$ and $X_{uv} \sim \mathcal{N}(0, 1)$. $\mathbb{E}(X_{uv} - X_{wz})^2 = \mathbb{E}(\langle Au, v \rangle - \langle Aw, z \rangle)^2 = \mathbb{E}(\sum_{i,j} A_{ij}(u_j v_i - w_j z_i))^2 = \sum_{i,j} (u_j v_i - w_j z_i)^2 = \|uv^T - wz^T\|_F^2 \leq \|u - w\|_2^2 + \|v - z\|_2^2$. Define $Y_{uv} = \langle g, u \rangle + \langle h, v \rangle$, $g \sim \mathcal{N}(0, I_n)$, $h \sim \mathcal{N}(0, I_m)$, $g \perp h$. $\mathbb{E}(Y_{uv} - Y_{wz})^2 = \|u - w\|_2^2 + \|v - z\|_2^2$. Then $\mathbb{E} \|A\| = \mathbb{E} \sup_{(u,v) \in T} X_{uv} \leq \mathbb{E} \sup_{(u,v) \in T} Y_{uv} = \mathbb{E} \sup_{u \in \mathbb{S}^{n-1}} \langle g, u \rangle + \mathbb{E} \sup_{v \in \mathbb{S}^{m-1}} \langle h, v \rangle = \mathbb{E} \|g\|_2 + \mathbb{E} \|h\|_2 \leq \sqrt{n} + \sqrt{m}$. \square

- $\mathbb{P}(\|A\| \geq \sqrt{m} + \sqrt{n} + t) \leq 2e^{-ct^2}$.

Proof $A \sim \mathcal{N}(0, I_{nm})$, $f(A) = \|A\| \leq \|A\|_2 \Rightarrow \|f\|_{\text{Lip}} \leq 1 \Rightarrow \|f(A) - \mathbb{E}f(A)\|_{\psi_2} \leq C$. \square

- Sudakov's minoration inequality: $\{X_t\}_{t \in T}$ mean zero Gaussian process. $\forall \epsilon > 0$, $\mathbb{E} \sup_{t \in T} X_t \geq C\epsilon \sqrt{\log \mathcal{N}(T, d, \epsilon)}$.

Proof Assume $\mathcal{N}(T, d, \epsilon) = N < \infty$. Let \mathcal{N} be a maximal ϵ -separated subset of T . $|\mathcal{N}| \geq N$. It suffices to show $\mathbb{E} \sup_{t \in \mathcal{N}} X_t \geq C\epsilon \sqrt{\log N}$. $Y_t := \frac{\epsilon}{\sqrt{2}} g_t$, $g_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. $\mathbb{E}(X_t - X_s)^2 = d(t, s)^2 \geq \epsilon^2 = \mathbb{E}(Y_t - Y_s)^2 \Rightarrow \mathbb{E} \sup_{t \in \mathcal{N}} X_t \geq \mathbb{E} \sup_{t \in \mathcal{N}} Y_t = C\epsilon \sqrt{\log N}$. \square

- $X_t = \langle g, t \rangle$, $g \sim \mathcal{N}(0, I_n)$, $d(s, t) = \|t - s\|_2$, $\mathbb{E} \sup_{t \in T} \langle g, t \rangle \geq C\epsilon \sqrt{\log \mathcal{N}(T, \epsilon)}$.

- P a polytope in \mathbb{R}^n with N vertices, diameter is bounded by 1. Then for $\epsilon > 0$, $\mathcal{N}(P, \epsilon) \leq N^{c/\epsilon^2}$.

Proof x_1, x_2, \dots, x_N vertices of P , $\mathbb{E} \sup_{t \in P} \langle g, t \rangle = \mathbb{E} \sup_{i \leq N} \langle g, x_i \rangle \leq C\sqrt{\log N}$. \square

- Gaussian width: $w(T) := \mathbb{E} \sup_{x \in T} \langle g, x \rangle$, $g \sim \mathcal{N}(0, I_n)$.

- Properties of Gaussian width: (a) $w(T) < \infty \Leftrightarrow T$ is bounded; (b) For every orthogonal matrix U and vector y , $w(UT + y) = w(T)$; (c) $w(\text{conv}(T)) = w(T)$; (d) $w(T + S) = w(T) + w(S)$, $w(aT) = |a|w(T)$; (e) $w(T) = \frac{1}{2}w(T - T) = \frac{1}{2}\mathbb{E} \sup_{x, y \in T} \langle g, x - y \rangle$; (f) $\frac{1}{\sqrt{2\pi}} \text{diam}(T) \leq w(T) \leq \frac{\sqrt{n}}{2} \text{diam}(T)$.

Proof (e): $w(T) = \frac{1}{2}(w(T) + w(T)) = \frac{1}{2}(w(T) + w(-T)) \stackrel{\text{(d)}}{=} \frac{1}{2}w(T - T)$.

(f): Lower bound. Fix $x, y \in T$, $x - y, y - x \in T - T$, $w(T) \geq \frac{1}{2}\mathbb{E} \max(\langle x - y, g \rangle, \langle y - x, g \rangle) = \frac{1}{2}\mathbb{E} | \langle x - y, g \rangle | = \sqrt{\frac{1}{2\pi}} \|x - y\|_2$.

Upper bound. $w(T) = \frac{1}{2}\mathbb{E} \sup_{x, y \in T} \langle g, x - y \rangle \leq \frac{1}{2}\mathbb{E} \sup_{x, y \in T} \|g\|_2 \|x - y\|_2 = \frac{1}{2}\mathbb{E} \|g\|_2 \text{diam}(T)$ and $\mathbb{E} \|g\|_2 \leq (\mathbb{E} \|g\|_2^2)^{\frac{1}{2}} = \sqrt{n}$. \square

- Spherical width: $w_{\mathbb{S}}(T) = \mathbb{E} \sup_{x \in T} \langle \theta, x \rangle$, $\theta \sim \text{Unif}(\mathbb{S}^{n-1})$.

- $(\sqrt{n} - C)w_{\mathbb{S}}(T) \leq w(T) \leq (\sqrt{n} + C)w_{\mathbb{S}}(T)$.

Proof $g = \|g\|_2 \cdot \frac{g}{\|g\|_2} := r \cdot \theta$, $r \perp \theta$. $w(T) = \mathbb{E} \sup_{x \in T} \langle r\theta, x \rangle = \mathbb{E} r \mathbb{E} \sup_{x \in T} \langle \theta, x \rangle = \mathbb{E} \|g\|_2 w_{\mathbb{S}}(T)$. Ex 3.1.4 $\Rightarrow |\mathbb{E} \|g\|_2 - \sqrt{n}| \leq C$. \square

- Squared version of the Gaussian width: $h(T)^2 = \mathbb{E} \sup_{t \in T} \langle g, t \rangle^2$, $g \sim \mathcal{N}(0, I_n)$.

- Stable dimension: bounded $T \subset \mathbb{R}^n$, $d(T) := \frac{h(T-T)^2}{\text{diam}^2(T)} \asymp \frac{w^2(T)}{\text{diam}^2(T)}$.

- $d(T) \leq \text{dim}(T)$.

Proof Let $\text{dim}(T) = k$ and $T \subset \mathbb{R}^k$. $h(T - T)^2 = \mathbb{E} \sup_{x, y \in T} \langle g, x - y \rangle^2$. $x - y = \text{diam}(T) \cdot z$ for some $z \in B_2^k$. $\therefore h(T - T)^2 \leq \text{diam}^2(T) \mathbb{E} \sup_{z \in B_2^k} \langle g, z \rangle^2 = \text{diam}^2(T) \mathbb{E} \|g\|_2^2 = \text{diam}^2(T) \cdot k$. \square

- Stable rank: $A_{m \times n}$, $r(A) := \frac{\|A\|_F^2}{\|A\|_2^2} = d(AB_2^n) \leq \text{rank}(A) = \text{dim}(AB_2^n)$.

- Gaussian complexity: $\gamma(T) := \mathbb{E} \sup_{x \in T} |\langle g, x \rangle|$, $g \sim \mathcal{N}(0, I_n)$.

- $T \subset \mathbb{R}^n$, \mathcal{P} projection onto $E \sim \text{Unif}(G_{n,m})$. $\forall m \leq n$, with probability at least $1 - 2e^{-m}$, $\text{diam}(\mathcal{P}T) \leq C[w_{\mathbb{S}}(T) + \sqrt{\frac{m}{n}} \text{diam}(T)]$.

Proof Step 1: Approximation. WLOG $\text{diam}(T) \leq 1$. $Q_{n \times n}$: choosing the first m rows of $U_{n \times n} \sim \text{Unif}(O(n))$. Then $\|\mathcal{P}x\|_2 \stackrel{d}{=} \|Qx\|_2, \forall x \in \mathbb{R}^n$. $Q^T z \sim \text{Unif}(\mathbb{S}^{n-1}), \forall z \in \mathbb{S}^{n-1}$. $\text{diam}(\mathcal{P}T) \stackrel{d}{=} \text{diam}(QT) = \sup_{x \in T-T} \|Qx\|_2 = \sup_{x \in T-T} \max_{z \in \mathbb{S}^{m-1}} \langle Qx, z \rangle = \sup_{x \in T-T} \max_{z \in \mathbb{S}^{m-1}} \langle x, Q^T z \rangle$. Choose an $\frac{1}{2}$ -net \mathcal{N} of $\mathbb{S}^{m-1}, |\mathcal{N}| \leq 5^m \Rightarrow \text{diam}(QT) \leq 2 \max_{z \in \mathcal{N}} \sup_{x \in T-T} \langle x, Q^T z \rangle$.

Step 2: Concentration. For $z \in \mathcal{N}$, $\mathbb{E} \sup_{x \in T-T} \langle Q^T z, x \rangle = w_{\mathbb{S}}(T-T) = 2w_{\mathbb{S}}(T)$. The function $f : Q^T z \rightarrow \sup_{x \in T-T} \langle Q^T z, x \rangle$ is Lipschitz on $\mathbb{S}^{n-1} \Rightarrow \langle Q^T z, x \rangle$ is sub-gaussian $\Rightarrow \mathbb{P}(\sup_{x \in T-T} \langle Q^T z, x \rangle \geq 2w_{\mathbb{S}}(T) + t) \leq 2e^{-cnt^2}$.

Step 3: Union bound. $\mathbb{P}(\max_{z \in \mathcal{N}} \sup_{x \in T-T} \langle Q^T z, x \rangle \geq 2w_{\mathbb{S}}(T) + t) \leq 5^m 2e^{-cnt^2} \leq 2e^{-m}$ ($t = C\sqrt{\frac{m}{n}}$ and C large enough). \square

- Phase transition: Equivalently write it as $\text{diam}(\mathcal{P}T) \leq C \max(w_{\mathbb{S}}(T), \sqrt{\frac{m}{n}} \text{diam}(T))$. Set $w_{\mathbb{S}}(T) = \sqrt{\frac{m}{n}} \text{diam}(T) \Rightarrow m = \frac{(\sqrt{n} w_{\mathbb{S}}(T))^2}{\text{diam}^2(T)} \asymp \frac{w^2(T)}{\text{diam}^2(T)} \asymp d(T)$. That is, if $m > d(T)$, $\text{diam}(\mathcal{P}T) \leq C \sqrt{\frac{m}{n}} \text{diam}(T)$; if $m < d(T)$, $\text{diam}(\mathcal{P}T) \leq C w_{\mathbb{S}}(T)$.
- Random matrix $G_{m \times n}$ with independent $\mathcal{N}(0, 1)$ entries. $\forall m \leq n$, with probability at least $1 - 2e^{-m}$, $\text{diam}(GT) \leq C[w(T) + \sqrt{m} \text{diam}(T)]$.

8 Chaining

- Sub-gaussian increments: $\{X_t\}_{t \in T}, (T, d)$. Exist $K \geq 0$, s.t. $\|X_t - X_s\|_{\psi_2} \leq Kd(t, s)$ for $t, s \in T$.
- $\{X_t\}_{t \in T}, \mathbb{E}X_t = 0, (T, d)$, sub-gaussian increments. Then $\mathbb{E} \sup_{t \in T} X_t \leq CK \sum_{k \in \mathbb{Z}} 2^{-k} \sqrt{\log \mathcal{N}(T, d, 2^{-k})}$.

Proof Step 1: Chaining setup. WLOG assume $K = 1$ and T is finite. $\epsilon_k = 2^{-k}, k \in \mathbb{Z}$. Choose an ϵ_k -net T_k of T so that $|T_k| = \mathcal{N}(T, d, \epsilon_k)$. T is finite $\Rightarrow \exists$ small enough $\kappa \in \mathbb{Z}$ and large enough $K \in \mathbb{Z}$ s.t. $T_\kappa = \{t_0\}, T_\kappa = T$. For a point $t \in T$, $\pi_k(t)$: a closest point in T_k . Then $d(t, \pi_k(t)) \leq \epsilon_k$. $\mathbb{E}X_{t_0} = 0 \Rightarrow \mathbb{E} \sup_{t \in T} X_t = \mathbb{E} \sup_{t \in T} (X_t - X_{t_0})$. Since $X_t - X_{t_0} = \sum_{k=\kappa+1}^K (X_{\pi_k(t)} - X_{\pi_{k-1}(t)})$, $\mathbb{E} \sup_{t \in T} (X_t - X_{t_0}) \leq \sum_{k=\kappa+1}^K \mathbb{E} \sup_{t \in T} (X_{\pi_k(t)} - X_{\pi_{k-1}(t)})$.

Step 2: Control the increments. $\|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}\|_{\psi_2} \leq d(\pi_k(t), \pi_{k-1}(t)) \leq \epsilon_k + \epsilon_{k-1} \leq 2\epsilon_{k-1}$. Ex 2.5.10 $\Rightarrow \mathbb{E} \sup_{t \in T} (X_{\pi_k(t)} - X_{\pi_{k-1}(t)}) \leq C \cdot 2\epsilon_{k-1} \cdot \sqrt{\log(|T_k| \cdot |T_{k-1}|)} \leq C \cdot 2\epsilon_{k-1} \sqrt{2 \log |T_k|}$.

Step 3: Summing up the increments. $\mathbb{E} \sup_{t \in T} (X_t - X_{t_0}) \leq C \sum_{k=\kappa+1}^K \epsilon_{k-1} \sqrt{\log |T_k|}$. \square

- Dudley's integral inequality: $\mathbb{E} \sup_{t \in T} X_t \leq CK \int_0^\infty \sqrt{\log \mathcal{N}(T, d, \epsilon)} d\epsilon$.
- *Proof* $2^{-k} = 2 \int_{2^{-k-1}}^{2^{-k}} d\epsilon$. $\sum_{k \in \mathbb{Z}} 2^{-k} \sqrt{\log \mathcal{N}(T, d, 2^{-k})} \leq 2 \sum_{k \in \mathbb{Z}} \int_{2^{-k-1}}^{2^{-k}} \sqrt{\log \mathcal{N}(T, k, \epsilon)} d\epsilon$. \square
- $\mathbb{E} \sup_{t, s \in T} |X_t - X_s| \leq CK \int_0^\infty \sqrt{\log \mathcal{N}(T, d, \epsilon)} d\epsilon$.
- $T \subset \mathbb{R}^n, w(T) \leq C \int_0^{+\infty} \sqrt{\log \mathcal{N}(T, \epsilon)} d\epsilon$.
- $T \subset \mathbb{R}^n, s(T) := \sup_{\epsilon \geq 0} \epsilon \sqrt{\log \mathcal{N}(T, \epsilon)}$. Then $cs(T) \leq w(T) \leq Cs(T) \log n$.
- Empirical process: $f \in \mathcal{F}, f : w \rightarrow \mathbb{R}, (\Omega, \Sigma, \mu)$. X is a random point in Ω . $X \sim \mu$. X_1, \dots, X_n i.i.d. copies of X . $X_f := \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}f(X)$ empirical process indexed by \mathcal{F} .
- $X_i \in [0, 1], \mathcal{F} := \{f : [0, 1] \rightarrow \mathbb{R}, \|f\|_{\text{Lip}} \leq L\}$. Then $\mathbb{E} \sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}f(X)| \leq \frac{CL}{\sqrt{n}}$.

Proof WLOG $\mathcal{F} = \{f : [0, 1] \rightarrow [0, 1], \|f\|_{\text{Lip}} \leq 1\}$.

Step 1: Check sub-gaussian increments. Let $Z_i := (f-g)(X_i) - \mathbb{E}(f-g)(X)$. Then $\forall f, g \in \mathcal{F}, \|X_f - X_g\|_{\psi_2} = \frac{1}{n} \|\sum_{i=1}^n Z_i\|_{\psi_2} \lesssim \frac{1}{n} (\sum_{i=1}^n \|Z_i\|_{\psi_2}^2)^{\frac{1}{2}}$. $\|Z_i\|_{\psi_2} \lesssim \|(f-g)(X_i)\|_{\psi_2} \lesssim \|f-g\|_\infty$.

Step 2: Apply Dudley's inequality. $\mathbb{E} \sup_{f \in \mathcal{F}} |X_f| \leq \frac{1}{\sqrt{n}} \int_0^1 \sqrt{\log \mathcal{N}(\mathcal{F}, \|\cdot\|_\infty, \epsilon)} d\epsilon \stackrel{\text{Ex 8.2.6}}{\leq} \frac{1}{\sqrt{n}} \int_0^1 \sqrt{\frac{C}{\epsilon} \log \frac{C}{\epsilon}} d\epsilon \lesssim \frac{1}{\sqrt{n}}$. \square

- VC dimension: $\mathcal{F} = \{f : \Omega \rightarrow \{0, 1\}\}$. Shattered $\Lambda \subset \Omega$, any $g : \Lambda \rightarrow \{0, 1\}$ can be obtained by $f \in \mathcal{F}$ restricted on Λ . $\text{VC}(\mathcal{F}) = \max_\Lambda |\Lambda|$.
- $|\mathcal{F}| \leq |\{\Lambda \subset \Omega, \Lambda \text{ is shattered by } \mathcal{F}\}|$.

Proof $|\Omega| = 1$, trivial. If $|\Omega| = n+1, \Omega = \Omega_0 \cup \{x_0\}, |\Omega_0| = n$. $\mathcal{F}_0 = \{f \in \mathcal{F} : f(x_0) = 0\}, \mathcal{F}_1 = \{f \in \mathcal{F} : f(x_0) = 1\}, S(\mathcal{F}) = |\{\Lambda \subset \Omega : \Lambda \text{ is shattered by } \mathcal{F}\}|$. Then $S(\mathcal{F}_0) \geq |\mathcal{F}_0|, S(\mathcal{F}_1) \geq |\mathcal{F}_1|, |\mathcal{F}| = |\mathcal{F}_0| + |\mathcal{F}_1|$.

(1) Λ is shattered by $\mathcal{F}_0(\mathcal{F}_1)$ but not by $\mathcal{F}_1(\mathcal{F}_0)$. Then $\Lambda_0 \in \mathcal{F}_0(\mathcal{F}_1), \notin \mathcal{F}_1(\mathcal{F}_0)$.

(2) Λ is shattered by \mathcal{F}_0 and \mathcal{F}_1 . Replace it with $\Lambda \cup \{x_0\}$. \square

- $|\Omega| = n$. Then $|\mathcal{F}| \leq \sum_{k=0}^d C_n^k \leq (\frac{en}{d})^d$ where $d = \text{VC}(\mathcal{F})$.
- Dimension reduction: $|\mathcal{F}| = N$, a class of boolean functions. Assume $\|f - g\|_{L^2(\mu)} > \epsilon$ for $f, g \in \mathcal{F}$. Then there exists $n \leq C\epsilon^{-4} \log N$ and $\Omega_n \subset \Omega, |\Omega_n| = n$ s.t. μ_n is uniform probability mass on $\Omega_n, \|f - g\|_{L^2(\mu_n)} \geq \frac{\epsilon}{2}$ for all $f, g \in \mathcal{F}$.

Proof $X_1, X_2, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$. Denote $h := (f-g)^2, \Omega_n = \{X_1, \dots, X_n\}$. $\|f-g\|_{L^2(\mu_n)}^2 - \|f-g\|_{L^2(\mu)}^2 = \frac{1}{n} \sum_{i=1}^n (h(X_i) - \mathbb{E}h(X))$. $\|h(X_i) - \mathbb{E}h(X)\|_{\psi_2} \lesssim \|h(X_i)\|_{\psi_2} \lesssim \|h(X_i)\|_{\infty} \leq 1 \Rightarrow \mathbb{P}(\|f-g\|_{L^2(\mu_n)}^2 - \|f-g\|_{L^2(\mu)}^2 > \frac{\epsilon^2}{4}) \leq 2e^{-cn\epsilon^4}$. Therefore, $\|f-g\|_{L^2(\mu_n)} \geq \frac{3}{4}\epsilon^2$ hold for all $f, g \in \mathcal{F}$ with prob $\geq 1 - 2N^2e^{-cn\epsilon^4}$. $n = C\epsilon^{-4} \log N$ sufficiently large. \square

- $\forall \epsilon \in (0, 1), \mathcal{N}(\mathcal{F}, L^2(\mu), \epsilon) \leq (\frac{2}{\epsilon})^{Cd}$.

Proof Choose $N \geq \mathcal{N}(\mathcal{F}, L^2(\mu), \epsilon)$ ϵ -separated functions in \mathcal{F} . $|\Omega_n| = n \leq C\epsilon^{-4} \log N$ s.t. $\mathcal{F}|_{\Omega_n} := \mathcal{F}_n$ is still $\frac{\epsilon}{2}$ -separated in $L^2(\mu_n)$. $N \leq (\frac{en}{dn})^{dn} \leq (\frac{C\epsilon^{-4} \log N}{dn})^{dn}$ where $d_n = \text{VC}(\mathcal{F}_n) \Rightarrow N \leq (C\epsilon^{-4})^{2d_n}$. \square

- $\text{VC}(\mathcal{F}) \geq 1$. Let $X_1, \dots, X_n \in \Omega \sim \mu$. Then $\mathbb{E} \sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}f(X)| \leq C\sqrt{\frac{\text{VC}(\mathcal{F})}{n}}$.

Proof Ex 8.3.24 $\Rightarrow \mathbb{E} \sup_{f \in \mathcal{F}} |\frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbb{E}f(X)| \leq \frac{2}{\sqrt{n}} \mathbb{E} \sup_{f \in \mathcal{F}} |Z_f|$ where $Z_f = \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i f(X_i)$ and $\epsilon_i \stackrel{\text{i.i.d.}}{\sim}$ symmetric Bernoulli. Conditioned on $(X_i), \|Z_f - Z_g\|_{\psi_2} = \frac{1}{\sqrt{n}} \|\sum_{i=1}^n \epsilon_i (f-g)(X_i)\|_{\psi_2} \lesssim [\frac{1}{n} \sum_{i=1}^n (f-g)^2(X_i)]^{\frac{1}{2}} = \|f-g\|_{L^2(\mu_n)}$. Applying Dudley's inequality, $\frac{2}{\sqrt{n}} \mathbb{E} \sup_{f \in \mathcal{F}} Z_f \lesssim \frac{1}{\sqrt{n}} \mathbb{E} X \int_0^1 \sqrt{\log \mathcal{N}(\mathcal{F}, L^2(\mu_n), \epsilon)} d\epsilon \lesssim \frac{1}{\sqrt{n}} \mathbb{E} X \int_0^1 \sqrt{\text{VC}(\mathcal{F}) \log \frac{2}{\epsilon}} d\epsilon \lesssim \sqrt{\frac{\text{VC}(\mathcal{F})}{n}}$. \square

- $\mathcal{R}(f_n^*) - \mathcal{R}(f^*) \leq 2 \sup_{f \in \mathcal{F}} |\mathcal{R}_n(f) - \mathcal{R}(f)|$.

Proof $\epsilon = \sup_{f \in \mathcal{F}} |\mathcal{R}_n(f) - \mathcal{R}(f)|$. $\mathcal{R}(f_n^*) \leq \mathcal{R}_n(f_n^*) + \epsilon \leq \mathcal{R}_n(f^*) + \epsilon \leq \mathcal{R}(f^*) + 2\epsilon$. \square

- For two-class classification, $\text{VC}(\mathcal{F}) \geq 1$. Then $\mathbb{E} \mathcal{R}(f_n^*) \leq \mathcal{R}(f^*) + C\sqrt{\frac{\text{VC}(\mathcal{F})}{n}}$ where $\mathcal{R}(\cdot)$ is the MSE risk.

Proof Only to show $\mathbb{E} \sup_{f \in \mathcal{F}} |\mathcal{R}_n(f) - \mathcal{R}(f)| \lesssim \sqrt{\frac{\text{VC}(\mathcal{F})}{n}}$. LHS = $\frac{1}{n} \sum_{i=1}^n [l(x_i) - \mathbb{E}l(x)]$ where $l = (f-T)^2$ is Boolean. Let $\mathcal{L} = \{(f-T)^2 : f \in \mathcal{F}\}$. Dudley's inequality \Rightarrow LHS $\lesssim \frac{1}{\sqrt{n}} \mathbb{E} X \int_0^1 \sqrt{\log \mathcal{N}(\mathcal{L}, L^2(\mu_n), \epsilon)} d\epsilon \stackrel{\text{Ex 8.4.6}}{\leq} \frac{1}{\sqrt{n}} \mathbb{E} X \int_0^1 \sqrt{\log \mathcal{N}(\mathcal{F}, L^2(\mu_n), \epsilon)} d\epsilon$. \square

- Talagrand's γ_2 functional: (T, d) metric space, $(T_k)_{k=1}^{\infty} (T_k \subset T)$ admissible sequence iff $|T_0| = 1, |T_k| \leq 2^{2^k}, \forall k$. $\gamma_2(T, d) := \inf_{(T_k)} \sup_{t \in T} \sum_{k=0}^{\infty} 2^{k/2} d(t, T_k)$.

- $\{X_t\}_{t \in T}$ mean zero sub-gaussian increments. Then $\mathbb{E} \sup_{t \in T} X_t \leq CK\gamma_2(T, d)$.

Proof Step 1: Chaining setup. WLOG $K = 1, |T| < \infty$. Let (T_k) be an admissible sequence, $T_0 = \{t_0\}, t_0 = \pi_0(t) \rightarrow \pi_1(t) \rightarrow \dots \rightarrow \pi_k(t) = t, d(t, \pi_k(t)) = d(t, T_k)$. Then $X_t - X_{t_0} = \sum_{k=1}^K (X_{\pi_k(t)} - X_{\pi_{k-1}(t)})$.

Step 2: Controlling the increments. Fix k and t , for $u \geq 0, \mathbb{P}(|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| \leq Cu2^{k/2} d(\pi_k(t), \pi_{k-1}(t))) \geq 1 - 2e^{-8u^2 2^k}$. Unfix t and $k, |T_k| |T_{k-1}| \leq |T_k|^2 \leq 2^{2^{k+1}}$. Let $A = \{|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| \leq Cu2^{k/2} d(\pi_k(t), \pi_{k-1}(t)) \text{ for } \forall k, t\}$. Then $\mathbb{P}(A) \geq 1 - \sum_{k=1}^{\infty} 2^{2^{k+1}} 2e^{-8u^2 2^k} \geq 1 - 2e^{-u^2}$ if $u > C'$.

Step 3: Summing up the increments. In event $A, \sup_{t \in T} |X_t - X_{t_0}| \leq C_1 u \gamma_2(T, d) \Rightarrow \|\sup_{t \in T} |X_t - X_{t_0}|\|_{\psi_2} \leq C_2 \gamma_2(T, d)$. \square

- $\{X_t\}_{t \in T}$ mean zero Gaussian process on $T, d(t, s) = \|X_t - X_s\|_{L^2}$. Then $c\gamma_2(T, d) \leq \mathbb{E} \sup_{t \in T} X_t \leq C\gamma_2(T, d)$.
- Talagrand's comparison inequality: $\{X_t\}_{t \in T}$ mean zero, $\{Y_t\}_{t \in T}$ mean zero Gaussian, $\forall t, s \in T, \|X_t - X_s\|_{\psi_2} \leq K \|Y_t - Y_s\|_{L^2} \Rightarrow \mathbb{E} \sup_{t \in T} X_t \leq CK \mathbb{E} \sup_{t \in T} Y_t$.
- $A_{m \times n}, A_{ij}$ independent mean zero sub-gaussian, $T \subset \mathbb{R}^n, S \subset \mathbb{R}^m$. Then $\mathbb{E} \sup_{x \in T, y \in S} \langle Ax, y \rangle \leq CK[w(T)\text{rad}(S) + w(S)\text{rad}(T)]$ where $K = \max_{ij} \|A_{ij}\|_{\psi_2}, \text{rad}(T) := \sup_{x \in T} \|x\|_2$.

Proof WLOG $K = 1$. $X_{uv} := \langle Au, v \rangle, u \in T, v \in S$. Then $\|X_{uv} - X_{wz}\|_{\psi_2} = \|\sum_{i,j} A_{ij}(u_i v_j - w_i z_j)\|_{\psi_2} \leq (\sum_{i,j} \|A_{ij}(u_i v_j - w_i z_j)\|_{\psi_2}^2)^{\frac{1}{2}} \leq (\sum_{i,j} \|u_i v_j - w_i z_j\|_2^2)^{\frac{1}{2}} = \|uv^T - wz^T\|_F \leq \|(u-w)v^T\|_F + \|w(v-z)^T\|_F = \|u-w\|_2 \|v\|_2 + \|w\|_2 \|v-z\|_2 \leq \|u-w\|_2 \text{rad}(S) + \|v-z\|_2 \text{rad}(T)$. Let $Y_{uv} = \langle g, u \rangle \text{rad}(S) + \langle h, v \rangle \text{rad}(T)$ where $g \sim \mathcal{N}(0, I_n), h \sim \mathcal{N}(0, I_m)$. $\|Y_{uv} - Y_{wz}\|_2^2 = \|u-w\|_2^2 \text{rad}(S)^2 + \|v-z\|_2^2 \text{rad}(T)^2 \Rightarrow \|X_{uv} - X_{wz}\|_{\psi_2} \lesssim \|Y_{uv} - Y_{wz}\|_2$. Applying the comparison inequality, $\mathbb{E} \sup_{u \in T, v \in S} X_{uv} \lesssim \mathbb{E} \sup_{u \in T, v \in S} Y_{uv} = \mathbb{E} \sup_{u \in T} \langle g, u \rangle \text{rad}(S) + \mathbb{E} \sup_{v \in S} \langle h, v \rangle \text{rad}(T) = w(T)\text{rad}(S) + w(S)\text{rad}(T)$. \square