High-Dimensional Probability

Lectured by Hao Ge   LaTeXed by Chengxin Gong

October 27, 2023

Contents

0 Appetizer

- Convex combination: For $z_1, z_2, \cdots, z_m \in \mathbb{R}^n$, 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 $T \subset \mathbb{R}^n$: $\mathrm{conv}(T) = \{\text{convex combinations of } z_1, \cdots, z_m \in T, m \in \mathbb{N}\}$.

- 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 $T \subset \mathbb{R}^n$, $\mathrm{diam}(T) = \sup\{||s-t||_2, s, t \in T\} < 1$. Then for any $x \in \mathrm{conv}(T)$ and any k, one can find points $x_1, x_2, \cdots, x_k \in T$ 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 $x \in \mathrm{conv}(T), x = \sum_{i=1}^m \lambda_i z_i, z_i \in T$. Define $Z, \mathbb{P}(Z = z_i) = \lambda_i, \mathbb{E}Z = x$. Consider i.i.d. $Z_1, Z_2, \cdots$ of Z, $\frac{1}{n}\sum_{j=1}^n Z_j \to x$ a.s. $n \to +\infty$. $\mathbb{E}||x - \frac{1}{k}\sum_{j=1}^k Z_j||_2^2 = \frac{1}{k^2}\sum_{j=1}^k \mathbb{E}||Z_j - x||_2^2 = \frac{1}{k^2}\sum_{j=1}^k (\mathbb{E}||Z_j||^2 - ||\mathbb{E}Z_j||_2^2) \leq \frac{1}{k} \Rightarrow \exists$ a realization of $Z_1, \cdots, Z_k$ such that $||x - \frac{1}{k}\sum_{j=1}^k Z_j||_2 \leq \frac{1}{\sqrt{k}}$. □

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

1 Preliminaries on random variables

- Jensen's inequality: convex ϕ, $\phi(\mathbb{E}X) \leq \mathbb{E}\phi(X)$. $\Rightarrow ||X||_{L^p} \leq ||X||_{L^q}$ 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 $\mathbb{E}X = \int_0^{+\infty} \mathbb{P}(X > t)\mathrm{d}t$.

- Markov inequality: $X \geq 0, t > 0, \mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$.

- LLN: $X_1, \cdots, X_n, \cdots$ i.i.d., $\mathbb{E}X_i = \mu, \mathrm{Var}(X_i) = \sigma^2, S_N = X_1 + X_2 + \cdots + X_N$. Then: (WLLN) $\mathbb{P}(|\frac{S_N}{N} - \mu| > \epsilon) \to 0, \forall \epsilon > 0$; (SLLN) $\mathbb{P}(\frac{S_N}{N} \to \mu, N \to +\infty) = 1$.

- CLT: $Z_N = \frac{S_N - \mathbb{E}S_N}{\sqrt{\mathrm{Var}(S_N)}} \xrightarrow{d} \mathcal{N}(0,1)$.

- $X_{N,i}, 1 \leq i \leq N$ independent $\mathrm{Ber}(p_{N,i}), \max_{i \leq N} p_{N,i} \to 0, S_N = \sum_{i=1}^N X_{N,i}, \mathbb{E}S_N \to \lambda < +\infty$. Then $S_N \xrightarrow{d}$ $\mathrm{Poisson}(\lambda)$.

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, $\mathbb{E}S_N = \frac{N}{2}, \mathrm{Var}(S_N) = \frac{N}{4}$. (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: $X_1, \cdots, X_N$ i.i.d. symmetric Bernoulli ($\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$), $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$. 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 $\lambda > 0, \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) = \mathbb{P}(e^{\lambda \sum a_i X_i} \geq e^{\lambda t}) \leq e^{-\lambda t}\mathbb{E}e^{\lambda \sum_{i=1}^N a_i X_i} = e^{-\lambda t}\prod_{i=1}^N \mathbb{E}e^{\lambda a_i X_i} = e^{-\lambda t}\prod_{i=1}^N \cosh(\lambda a_i) \leq e^{-\lambda t}\prod_{i=1}^N e^{\lambda^2 a_i^2/2} = e^{-\lambda t + \frac{\lambda^2}{2}} \Rightarrow \mathbb{P}(\sum_{i=1}^N a_i X_i \geq t) \leq \inf_{\lambda \geq 0} e^{-\lambda t + \frac{\lambda^2}{2}} = e^{-\frac{t^2}{2}} (\lambda = t)$. □

- Bounded r.v.s: $X_1, \cdots, X_N$ 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, $S_N = \sum_{i=1}^N X_i, \mu = \mathbb{E}S_N \Rightarrow \forall t > \mu, \mathbb{P}(S_N \geq t) \leq e^{-\mu}(\frac{e\mu}{t})^t$.

  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 $\lambda^* = \log(t/\mu)$. □

- $d = (n-1)p$ 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: $\mathbb{P}(\exists i, |d_i - d| \geq \delta d) \leq n \cdot 2e^{-c\delta^2 d} \leq n \cdot 2 \cdots n^{-Cc\delta^2} = 2n^{1 - Cc\delta^2} \leq 1 - p^*$ (let $Cc\delta^2 > 1$). □

- 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 $k_1 = 1$. $\mathbb{E}|X|^p = \int_0^{+\infty} \mathbb{P}(|X| \geq t) p t^{p-1} \mathrm{d}t \leq \int_0^{+\infty} 2e^{-t^2} p t^{p-1} \mathrm{d}t = p\Gamma(\frac{p}{2}) \overset{\Gamma(x) \leq 3x^x \text{ for } x \geq \frac{1}{2}}{\leq} 3p(\frac{p}{2})^{p/2} \Rightarrow ||X||_p \leq \frac{1}{\sqrt{2}}(3p)^{1/p} p^{1/2} \leq 3\sqrt{p}$.

  (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 $2e\lambda^2 < 1$) $\overset{\frac{1}{1-x} \leq e^{2x} \text{ for } x \in [0, \frac{1}{2}]}{\leq} e^{4e\lambda^2}$ (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). □

- 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 $\mathbb{E}e^{\lambda X} \leq e^{C\lambda^2 ||X||_{\psi_2}^2}$.

- Let $X_1, \cdots, X_N$ be i.i.d. 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 $\forall \lambda \in \mathbb{R}, \mathbb{E}e^{\lambda \sum_{i=1}^N X_i} = \prod_{i=1}^N \mathbb{E}e^{\lambda X_i} \leq \prod_{i=1}^N e^{c\lambda^2 ||X_i||_{\psi_2}^2} = e^{c\lambda^2 \sum_{i=1}^n ||X_i||_{\psi_2}^2}$ □

- 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 $||\mathbb{E}X||_{\psi_2} \leq C_1|\mathbb{E}X| \leq C_1\mathbb{E}|X| = C_1||X||_{L^1} \leq C_1C_2||X||_{\psi_2}$. □

- 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): $k_5 = 1, |x|^p \leq p^p(e^x + e^{-x}) \Rightarrow \mathbb{E}|X|^p \leq p^p(\mathbb{E}e^X + \mathbb{E}e^{-X}) \leq 2ep^p$. □

- $||X||_{\psi_1} = \inf\{t > 0 : \mathbb{E}e^{|X|/t} \leq 2\}$. 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 $||X||_{\psi_2} = ||Y||_{\psi_2} = 1$. $\mathbb{E}e^{XY} \leq \mathbb{E}e^{\frac{X^2 + Y^2}{2}} = \mathbb{E}[e^{\frac{X^2}{2} + \frac{Y^2}{2}}] \leq \frac{1}{2}(\mathbb{E}e^{X^2} + \mathbb{E}e^{Y^2}) = 2$. □

- Orlicz function/space: $\psi : [0, +\infty) \to [0, +\infty)$, convex, increasing, $\psi(0) = 0$, $\psi(x) \to +\infty, x \to +\infty$. $||X||_\psi := \inf\{t > 0 : \mathbb{E}\psi(|X|/t) \leq 1\}$. $L_\psi := \{X : ||X||_\psi < +\infty\}$ 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: $X_1, \cdots, X_N$ i.i.d., 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 $S=\sum_{i=1}^N X_i$. $\mathbb{P}(S\ge t)\le e^{-\lambda t}\prod_{i=1}^N \mathbb{E}e^{\lambda X_i}$. $\mathbb{E}e^{\lambda X_i}\le e^{c\lambda^2||X_i||_{\psi_1}^2}$ if $|\lambda|\le \frac{c}{\max||X_i||_{\psi_1}}$. Then $\mathbb{P}(S\ge t)\le e^{-\lambda t+c\lambda^2\sigma^2}$ 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 $|\lambda|\le\frac{c}{\max||X_i||_{\psi_1}}$. □

- Corollary 1: $\mathbb{P}(|\sum_{i=1}^N a_iX_i|\ge t)\le 2e^{-c\min(\frac{t^2}{K^2||a||_2^2},\frac{t}{K||a||_\infty})}$ where $K=\max_i||X_i||_{\psi_1}$.
- Corollary 2: $|X_i|\le K$, then $\mathbb{P}(|\sum_{i=1}^N X_i|\ge t)\le 2\exp(-\frac{t^2/2}{\sigma^2+Kt/3})$ 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 $\|\|X\|_2-\sqrt{n}\|_{\psi_2}\le CK^2, K=\max_i||X_i||_{\psi_2}$.

  Proof $\mathbb{E}X_i^2=1\Rightarrow K\ge 1$. $||X_i^2-1||_{\psi_1}\le C||X_i^2||_{\psi_1}=C||X_i||_{\psi_2}^2\le CK^2$. Bernstein's inequality: $\mathbb{P}(|\frac{1}{n}||X||_2^2-1|\ge u)=\mathbb{P}(|\frac{1}{n}\sum_{i=1}^n(X_i^2-1)|\ge u)\le 2e^{-cn\min(\frac{u^2}{K^4},\frac{u}{K^2})}\le 2e^{-\frac{cn}{K^4}\min(u^2,u)}$. For any $\delta>0$, $\mathbb{P}(|\frac{1}{\sqrt{n}}||X||_2-1|\ge\delta)\le\mathbb{P}(|\frac{1}{n}||X||_2^2-1|\ge\max(\delta,\delta^2))\le 2e^{-\frac{cn}{K^4}\delta^2}\Rightarrow\mathbb{P}(|\,\|X\|_2-\sqrt{n}|\ge t)\le 2e^{-ct^2/K^4}$. □

- Isotropy: $\Sigma(X)=\mathbb{E}XX^T=I$. If $\Sigma\ne I_n$, 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 $\mathbb{E}\langle X,x\rangle^2=\mathbb{E}(x^TXX^Tx)=x^T(\mathbb{E}XX^T)x$. $||x||_2^2=x^TI_nx$. $\Rightarrow\mathbb{E}XX^T=I_n$. □

- 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 $\mathbb{E}||X||_2^2=\mathbb{E}(X^TX)=\mathbb{E}(\mathrm{tr}(X^TX))=\mathrm{tr}(\mathbb{E}XX^T)=n$.

  $\mathbb{E}\langle X,Y\rangle^2=\mathbb{E}(X^TYY^TX)=\mathbb{E}(\mathrm{tr}(X^TYY^TX))=\mathbb{E}(\mathrm{tr}(XX^TYY^T))=\mathrm{tr}((\mathbb{E}XX^T)(\mathbb{E}YY^T))=n$. □

- Examples: $X\sim U(\sqrt{n}\mathbb{S}^{n-1}), X\sim U(\{-1,1\}^n), X=(X_1,\cdots,X_n)$ i.i.d., $\mathbb{E}X_i=0,\mathrm{Var}(X_i)=1$ are all isotropic.
- $g\sim\mathcal{N}(0,I_n)$, then $\mathbb{P}(|\,\|g\|_2-\sqrt{n}|\ge t)\le 2e^{-ct^2}$.
- Frame: $\{u_i\}_{i=1}^N\subset\mathbb{R}^n$, Approximate Parseval's identity: $A||x||_2^2\le\sum_{i=1}^N\langle u_i,x\rangle^2\le B||x||_2^2$. A,B: frame bounds. $A=B$: tight frame($\Leftrightarrow\sum_{i=1}^N u_iu_i^T=AI_n$) and in this case, $\sum_{i=1}^N\langle u_i,x\rangle u_i=Ax$.
- (a) Tight frame $\{u_i\}_{i=1}^N, A=B, X=\mathrm{Unif}\{u_i,i=1,2,\cdots,N\}$, then $(\frac{N}{A})^{1/2}X$ is isotropic. (b) X is isotropic, $\mathbb{P}(X=x_i)=p_i, i=1,2,\cdots,N$. Then $u_i=\sqrt{p_i}x_i$ form a tight frame with $A=B=1$.
- Isotropic convex sets: $X\sim\mathrm{Unif}(K), K\subset\mathbb{R}^n$ convex, bounded, non-empty interior (convex body). Assume $\mathbb{E}X=0,\Sigma=\mathrm{Cov}(X)$. Then $Z=\Sigma^{-1/2}X$ is isotropic and $Z\sim\mathrm{Unif}(\Sigma^{-1/2}K)$.
- $X\in\mathbb{R}^n$ is sub-gaussian $\Leftrightarrow\forall X\in\mathbb{R}^n,\langle X,x\rangle$ are sub-gaussian. $||X||_{\psi_2}=\sup_{x\in\mathbb{S}^{n-1}}||\langle X,x\rangle||_{\psi_2}$.
- $X=(X_1,\cdots,X_n)$ independent, mean zero, sub-gaussian coordinate. Then X is sub-gaussian with $||X||_{\psi_2}\le C\max_{i\le n}||X_i||_{\psi_2}$.

  Proof $||\langle X,x\rangle||_{\psi_2}^2=||\sum X_ix_i||_{\psi_2}^2\le C\sum_{i=1}^n x_i^2||X_i||_{\psi_2}^2\le C\max_{i\le n}||X_i||_{\psi_2}^2$. □

- Gaussian dist: $X\sim\mathcal{N}(0,I_n),||X||_{\psi_2}\le C$.
- Discrete dist: $X\sim\mathrm{Unif}\{\sqrt{n}e_i,i=1,2,\cdots,n\},||X||_{\psi_2}\simeq\sqrt{\frac{n}{\log n}}$.
- Uniform dist: $X\sim\mathrm{Unif}\{\sqrt{n}\mathbb{S}^{n-1}\},||X||_{\psi_2}\le C$.

  Proof $g\sim\mathcal{N}(0,I_n),X\overset{d}{=}\frac{\sqrt{n}g}{||g||_2}$. $p(t):=\mathbb{P}(|X_1|\ge t)=\mathbb{P}(\frac{|g|}{||g||_2}\ge\frac{t}{\sqrt{n}})$. $\|\|g\|_2-\sqrt{n}\|_{\psi_2}\le C\Rightarrow\mathbb{P}(\|g\|_2\le\frac{\sqrt{n}}{2})\le 2e^{-cn}$. Need to show that all one-dimensional marginals $\langle X,x\rangle$ are sub-gaussian. By rotation invariance, we may assume that $x=(1,0,\cdots,0)$. Let $\mathcal{E}=\{\|g\|_2\ge\frac{\sqrt{n}}{2}\}\Rightarrow p(t)\le\mathbb{P}(\frac{|g|}{\|g\|_2}\ge\frac{t}{\sqrt{n}},\mathcal{E})+\mathbb{P}(\mathcal{E})\le\mathbb{P}(|g_1|\ge\frac{t}{2},\mathcal{E})+2e^{-cn}\le 2e^{-t^2/8}+2e^{-cn}\overset{t\le\sqrt{n}}{\le}4e^{-ct^2}$. □

- Grothendieck's inequality: $A=\{a_{ij}\}_{m\times n}$ of real numbers. Assume $\forall x_i,y_i\in\{-1,1\}$, we have $|\sum_{i,j}a_{ij}x_iy_j|\le 1$. Then for any Hilbert space $\mathscr{H}$, any $u_i,v_j\in\mathscr{H}$ satisfying $||u_i||=||v_j||=1$, we have $|\sum_{i,j}a_{ij}\langle u_i,v_j\rangle|\le K$ with $K\le 1.783$.

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

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

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

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

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

- 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, $\langle A, X\rangle = \mathrm{tr}(A^TX) = \sum_{i,j=1}^n A_{ij}X_{ij}$.

- Semidefinite relaxation: $\max \sum_{i,j=1}^n A_{ij}x_ix_j, x_i = \pm 1, A$ $n \times n$ 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, $\mathrm{INT}(A) := \max \sum_{i,j=1}^n A_{ij}x_ix_j, x_i = \pm 1, \mathrm{SDP}(A) := \max \sum_{i,j=1}^n A_{ij}\langle X_i, X_j\rangle, ||X_i||_2 = 1$. Then $\mathrm{INT}(A) \le \mathrm{SDP}(A) \le 2K \cdot \mathrm{INT}(A)$.

- Maximum cut: $G = (V, E)$ finite simple, $V \to V_1 + V_2$, 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: $X = (x_i)_{n\times 1}, x_i = \pm 1$. $\mathrm{CUT}(G, X) = \frac{1}{4}\sum_{i,j=1}^n A_{ij}(1 - x_ix_j)$. $\text{MAX-CUT}(G) = \frac{1}{4}\max\{\sum_{i,j} A_{ij}(1 - x_ix_j) : x_i = \pm 1\}$.

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

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

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

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

- There exists a Hilbert space $\mathcal{H}$ and $\phi, \psi : \mathbb{S}^{n-1} \to \mathbb{S}(\mathcal{H})$ 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 $\Rightarrow \exists\mathcal{H}, \phi, \psi$. $\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} + \cdots, \sinh t = t + \frac{t^3}{3!} + \frac{t^5}{5!}, ||\phi(u)||_2^2 = ||\psi(u)||_2^2 = \sinh(\frac{\beta\pi}{2}) = 1$ for all $u \in \mathbb{S}^{n-1} \Rightarrow \beta = \frac{2}{\pi}\log(1 + \sqrt{2})$.

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

4 Random matrices

- Singular vector decomposition: $A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V_{n\times n}^T = \sum_{i=1}^n s_iU_iV_i^T, \Sigma = \mathrm{diag}(s_1, s_2, \cdots, s_r), s_i \ge 0$ sigular values. $s_i = \sqrt{\lambda_i(AA^T)} = \sqrt{\lambda_i(A^TA)}$. 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, $||A|| = \max_{x\in\mathbb{S}^{n-1}, y\in\mathbb{S}^{m-1}}\langle Ax, y\rangle$.

- $s_n(A) > 0 \Leftrightarrow m \ge n = \mathrm{rank}(A), s_n(A) = \frac{1}{||A^+||}$ 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: $\mathrm{rank}(A) = r, k < r, A_k := \sum_{i=1}^k s_i u_i v_i^T, ||A - A_k|| = \min_{\mathrm{rank}(A')\leq k} ||A - A'||$ (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 $||A^TA - I_n|| \leq \max(\delta, \delta^2)$, then $(1-\delta)||x||_2 \leq ||Ax||_2 \leq (1+\delta)||x||_2$ for all x.

  Proof WLOG $||x||_2 = 1$. $|\|Ax\|_2^2 - 1| = |\langle (A^TA - I_n)x, x\rangle| \leq \max(\delta, \delta^2) \Rightarrow \max(||Ax||_2 - 1, (||Ax||_2 - 1)^2) \leq \max(\delta, \delta^2) \Rightarrow |\|Ax\|_2 - 1| \leq \delta$. □

- $Q_{n\times m}, QQ^T = I_n \Leftrightarrow P = Q^TQ$ 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. $|x_i - y_j| < \epsilon$. □

- 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 cubde: $x, y \in \{0, 1\}^n, d_H(x, y) := \#\{i : x(i) \neq y(i)\}$.

- (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. □

- Error correcting code: Fix integers k, n and r. Encoder $\{0, 1\}^k \to \{0, 1\}^n$, Decoder $\{0, 1\}^n \to \{0, 1\}^k$, $D(y) = x$ 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 $\to n$ 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. $E : \{0, 1\}^k \to \mathcal{N}$ one to one; $D : \{0, 1\}^n \to \{0, 1\}^k$ nearest-neighbor decodes. □

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

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

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

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