

High-Dimensional Probability

Lectured by [Hao Ge](#)

L^AT_EXed by [Chengxin Gong](#)

November 24, 2023

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

[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
- $d = (n - 1)p$ 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: $\mathbb{P}(\exists i, |d_i - d| \geq \delta d) \leq n \cdot 2e^{-c\delta^2 d} \leq n \cdot 2 \dots n^{-C\delta^2} = 2n^{1-C\delta^2} \leq 1 - p^*$ (let $C\delta^2 > 1$). \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 $k_1 = 1$. $\mathbb{E}|X|^p = \int_0^{+\infty} \mathbb{P}(|X| \geq t) p t^{p-1} dt \leq \int_0^{+\infty} 2e^{-t^2} p t^{p-1} dt = p \Gamma(\frac{p}{2}) \leq \frac{\Gamma(x) \leq 3x^x \text{ for } x \geq \frac{1}{2}}{3p(\frac{p}{2})^{p/2}} \Rightarrow \|X\|_p \leq \frac{1}{\sqrt{2}}(3p)^{1/p} p^{1/2} \leq 3\sqrt{p}$.
(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 $2e\lambda^2 < 1$) $\leq e^{4e\lambda^2}$ (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 $\mathbb{E}e^{\lambda X} \leq e^{C\lambda^2\|X\|_{\psi_2}^2}$.
- 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 $|e\lambda| \leq \frac{1}{2}, 1 + \frac{(e\lambda)^2}{1 - e\lambda} \leq 1 + 2e^2\lambda^2 \leq e^{4e^2\lambda^2}$, 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 $t \geq 0, \mathbb{P}(|\sum_{i=1}^N X_i| \geq t) \leq 2e^{-c \min(\frac{t^2}{\sum_{i=1}^N \|X_i\|_{\psi_1}^2}, \max_i \frac{t}{\|X_i\|_{\psi_1}})}$.

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 $\|X\|_2 - \sqrt{n} \leq CK^2$, $K = \max_i \|X_i\|_{\psi_2}$.
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: $\Sigma(X) = \mathbb{E} X X^T = I$. 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 ($\Leftrightarrow \sum_{i=1}^N u_i u_i^T = A I_n$) 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 $g \sim \mathcal{N}(0, I_n)$, $X \stackrel{d}{=} \frac{\sqrt{ng}}{\|g\|_2}$. $p(t) := \mathbb{P}(|X_1| \geq t) = \mathbb{P}(\frac{|g_1|}{\|g\|_2} \geq \frac{t}{\sqrt{n}})$. $\|g\|_2 - \sqrt{n} \leq C \Rightarrow \mathbb{P}(\|g\|_2 \leq \frac{\sqrt{n}}{2}) \leq 2e^{-cn}$. 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 $\Rightarrow \exists \mathcal{H}, \phi, \psi$. $\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \dots, \sinh t = t + \frac{t^3}{3!} + \frac{t^5}{5!} + \dots, \|\phi(u)\|_2^2 = \|\psi(u)\|_2^2 = \sinh(\frac{\beta\pi}{2}) = 1$ 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: $A, B \in \mathbb{R}^n, A + B := \{a + b, a \in A, b \in B\}$.
- $K \subset \mathbb{R}^n, \epsilon > 0, \frac{|K|}{|B_2^n|} \leq \mathcal{N}(K, \epsilon) \leq \mathcal{P}(K, \epsilon) \leq \frac{|K + \frac{\epsilon}{2} B_2^n|}{|\frac{\epsilon}{2} B_2^n|}$ where $|\cdot|$ denotes the volume in \mathbb{R}^n , B_2^n denotes the unit Euclidean ball in \mathbb{R}^n .
- Corollary: Let $K = B_2^n$. $|B_2^n| = \epsilon^n |K|, |K + \frac{\epsilon}{2} B_2^n| = (1 + \frac{\epsilon}{2})^n |K| \Rightarrow (\frac{1}{\epsilon})^n \leq \mathcal{N}(B_2^n, \epsilon) \leq (1 + \frac{2}{\epsilon})^n$. $\epsilon \in (0, 1] \Rightarrow (\frac{1}{\epsilon})^n \leq \mathcal{N}(B_2^n, \epsilon) \leq (\frac{3}{\epsilon})^n$.
- 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 $\mathcal{P}(\{0, 1\}^n, d_H, 2r) \geq \mathcal{N}(\{0, 1\}^n, d_H, 2r) \geq \frac{2^n}{\sum_{k=0}^{2r} C_n^k} \geq 2^n (\frac{2r}{en})^{2r} \geq 2^k$. \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 $A = \underbrace{A^+ + A^-}_{\text{upper} + \text{lower triangular matrix}}, \mathbb{P}(\|A^+\| \leq CK(\sqrt{n} + t)) \geq 1 - 2e^{-4t^2} \geq 1 - 2e^{-t^2}, \mathbb{P}(\|A^-\| \leq CK(\sqrt{n} + t)) \geq 1 - 2e^{-4t^2} \geq 1 - 2e^{-t^2}$.
 $1 - 2e^{-t^2}, \mathbb{P}(\|A\| \leq CK(\sqrt{n} + t)) \geq \mathbb{P}(\|A^+\| \leq \frac{C}{2}K(\sqrt{n} + t), \|A^-\| \leq \frac{C}{2}K(\sqrt{n} + t)) \geq 2(1 - 2e^{-t^2}) - 1 = 1 - 4e^{-t^2}$. \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 $\sin \angle(v_i(S), v_i(T)) \leq \frac{\|S - T\|}{\delta} \Rightarrow \exists \theta \in \{-1, 1\}, \|v_i(S) - \theta v_i(T)\|_2 \leq \frac{\|S - T\|}{\delta} \cdot 2^{3/2}$.

- 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 $\mathbb{S}^{n-1}, |\mathcal{N}| \leq 9^n$. $\|\frac{1}{m} A^T A - I_n\| \leq 2 \max_{x \in \mathcal{N}} |\langle \frac{1}{m} A^T A - I_n, x \rangle| = 2 \max_{x \in \mathcal{N}} |\frac{1}{m} \|Ax\|_2^2 - 1|$.

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. $\mathbb{P}(|\frac{1}{m} \|Ax\|_2^2 - 1| \geq \frac{\epsilon}{2}) \leq 9^n \cdot 2e^{-c_1 C^2 (n+t^2)} \leq 2e^{-t^2}$. \square

- $X \in \mathbb{R}^n$ sub-gaussian. $\mathbb{E}X = 0, \Sigma = \mathbb{E}XX^T, X_i \stackrel{d}{=} X$ 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 volume 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 $H = \{x \in \sqrt{n}\mathbb{S}^{n-1}, x_1 \leq 0\}$. $\sigma(A) \geq \frac{1}{2}\sigma(H)$. Thm 5.1.6 $\Rightarrow \sigma(A_t) \geq \sigma(H_t) = \mathbb{P}(X \in H_t)$. $H_t \supset \{x \in \sqrt{n}\mathbb{S}^{n-1}, x_1 \leq \frac{t}{\sqrt{2}}\} \Rightarrow \sigma(H_t) \geq \mathbb{P}(X_1 \leq \frac{t}{\sqrt{2}})$. $\|X_1\|_{\psi_2} \leq C \Rightarrow \mathbb{P}(X_1 \leq \frac{t}{\sqrt{2}}) \geq 1 - 2e^{-ct^2}$. \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 ($\simeq P_{G_{n,m}}$ orthogonal projections), $d(E, F) = \|P_E - P_F\|$, \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 $p(x) = e^{-U(x)}$, $\text{Hess } U(x) \succeq \kappa I_n$. $\|f(X) - \mathbb{E}f(X)\|_{\psi_2} \leq \frac{C\|f\|_{\text{Lip}}}{\sqrt{\kappa}}$.
- $X = (X_1, \dots, X_n)$ 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 $Pz = (z_1, \dots, z_m, 0, \dots, 0)$. $\mathbb{E}\|Pz\|_2^2 = m\mathbb{E}z_i^2 = \frac{m}{n}$.

(b): $f : z \rightarrow \|Pz\|_2$, $\|f\|_{\text{Lip}} = 1 \Rightarrow \|\|Pz\|_2 - \mathbb{E}\|Pz\|_2\|_{\psi_2} \leq \frac{C}{\sqrt{n}}$ \square

- \mathcal{X} a set of N points in \mathbb{R}^n , $\epsilon > 0$, $m \geq \frac{C}{\epsilon^2} \log N$, $E \sim \text{Unif}(G_{n,m})$, $Q = \sqrt{\frac{n}{m}}P_E$. 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. $\mathbb{E}\text{tr}(e^{\lambda S}) = \mathbb{E}\text{tr}(e^{\sum_{i=1}^{N-1} \lambda X_i + \lambda X_N}) \leq \mathbb{E}\text{tr}(e^{\sum_{i=1}^{N-1} \lambda X_i + \log \mathbb{E}e^{\lambda X_N}}) \leq \text{tr}(e^{\sum_{i=1}^{N-1} \log \mathbb{E}e^{\lambda X_i}})$.

Step 3: Lemma: X is an $n \times n$ symmetric mean zero random matrix, $\|X\| \leq K$ a.s. Then $e^{\lambda X} \preceq e^{g(\lambda)\mathbb{E}X^2}$ 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 $|z| < 3$. $z = \lambda x$. If $|x| \leq K, |\lambda| < \frac{3}{K}$, $e^{\lambda x} \leq 1 + \lambda x + g(\lambda)x^2$. (b) of Ex. 5.4.5 \Rightarrow If $\|X\| \leq K, |\lambda| < 3/K$, $\mathbb{E}e^{\lambda X} \leq I + g(\lambda)\mathbb{E}X^2$ (since $\mathbb{E}X = 0$) $\leq e^{g(\lambda)\mathbb{E}X^2}$. \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 $\Rightarrow \log \mathbb{E}e^{\lambda X_i} \leq g(\lambda)\mathbb{E}X_i^2 \Rightarrow \sum_{i=1}^N \log \mathbb{E}e^{\lambda X_i} \leq g(\lambda) \cdot Z$ 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 $\mathbb{E}\|X\|_2^2 = \mathbb{E}XX^T = \mathbb{E}\text{tr}(X^T X) = \mathbb{E}\text{tr}(XX^T) = \text{tr}(\Sigma) \Rightarrow \|X\|_2^2 \leq K^2 \text{tr}(\Sigma)$ a.s.. Ex 5.4.11 $\Rightarrow \mathbb{E}\|\Sigma_m - \Sigma\| = \frac{1}{m} \mathbb{E}\|\sum_{i=1}^m (X_i X_i^T - \Sigma)\| \lesssim \frac{1}{m}(\sigma\sqrt{\log n} + M \log n)$ 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 $\mathbb{E}(XX^T - \Sigma)^2 = \mathbb{E}(XX^T)^2 - \Sigma^2 \leq \mathbb{E}(XX^T)^2 = \mathbb{E}(\|X\|_2^2 XX^T) \leq K^2 \text{tr}(\Sigma)\Sigma \Rightarrow \sigma^2 \leq K^2 m \text{tr}(\Sigma)\|\Sigma\|$. $\|XX^T - \Sigma\| \leq \|X\|_2^2 + \|\Sigma\| \leq K^2 \text{tr}(\Sigma) + \|\Sigma\| \leq 2K^2 \text{tr}(\Sigma) := M$ (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

- $Y \perp Z, \mathbb{E}Z = 0$, 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 $X' \stackrel{d}{=} X, X' \perp X$.

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 $A = \sum_i s_i u_i u_i^T, X^T A X' = \sum_i s_i \underbrace{\langle u_i, X \rangle}_{:=g_i} \underbrace{\langle u_i, X' \rangle}_{:=g'_i}$. $(g_1, \dots, g_n) \perp (g'_1, \dots, g'_n) \sim \mathcal{N}(0, I_n) \Rightarrow \mathbb{E}e^{\lambda X^T A X'} = \prod_{i=1}^n \mathbb{E}e^{\lambda s_i g_i g'_i} = \prod_{i=1}^n \mathbb{E}e^{\lambda^2 s_i^2 g_i^2 / 2} \leq \prod_{i=1}^n e^{C\lambda^2 s_i^2 (\lambda^2 s_i^2 \leq c)} \leq e^{C\lambda^2 \|A\|_F^2} (\lambda^2 \leq \frac{c}{\max_i s_i^2} = \frac{c}{\|A\|^2})$. \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' , $\mathbb{E}_X e^{\lambda X^T A X'} \leq e^{C\lambda^2 K^2 \|A X'\|_2^2}, \mathbb{E}_g e^{\mu g^T A X'} = e^{\frac{\mu^2 \|A X'\|_2^2}{2}}$. $\mu = \sqrt{2c}K\lambda \Rightarrow \mathbb{E}_X e^{\lambda X^T A X'} \leq \mathbb{E}_g e^{\sqrt{2c}K\lambda g^T A X'} \Rightarrow \mathbb{E}e^{\lambda X^T A X'} \leq \mathbb{E}e^{\sqrt{2c}K\lambda g^T A X'} \leq \mathbb{E}e^{2cK^2 \lambda g^T A g'}$. \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: $\|X_i^2 - \mathbb{E}X_i^2\|_{\psi_2} \lesssim 1$. Bernstein $\Rightarrow p_1 \leq e^{-c \min(\frac{t^2}{\sum_i a_{ii}^2}, \frac{t}{\max_i |a_{ii}|})} \leq e^{-c \min(\frac{t^2}{\|A\|_F^2}, \frac{t}{\|A\|})}$.

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 $|\|BX\|_2 - \|B\|_F|_{\psi_2} \leq CK^2 \|B\|, K = \max_i \|X\|_{\psi_2}$.