Theoretical Stats and Machine Learning (Homework 3)

Student Name: Vu Huyen Trang Pham

UT EID: vp22673

Question 1. More sub-Gaussian characterizations

(a)(i) Assume $\|X\|_{\psi_2} \le L$. By Homework 2, Exercise 5 (tail characterization of sub-Gaussianity), there exists an absolute constant $c_0 > 0$ such that for all $t \ge 0$,

$$\mathbb{P}(|X| \ge t) \le 2\exp\Big(-c_0\,\frac{t^2}{L^2}\Big).$$

Let $g \sim \mathcal{N}(0,1)$. We will show that there is an absolute constant $C > 0$ such that with $K := CL$ we have

$$\mathbb{P}(|X| \ge t) \le 2\,\mathbb{P}(|Kg| \ge t) \qquad \forall\, t \ge 0,$$

i.e. $X \preceq Kg$.

To lower bound Gaussian tails, use the Mills ratio: for all $u > 0$,

$$\mathbb{P}(g \ge u) \ge \frac{\varphi(u)}{u + u^{-1}}, \qquad \varphi(u) := \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

Therefore for $u > 0$,

$$\mathbb{P}(|g| \ge u) = 2\mathbb{P}(g \ge u) \ge \frac{2}{\sqrt{2\pi}} \cdot \frac{e^{-u^2/2}}{u + u^{-1}}.$$

In particular, for $u > 1/2$ we have $u^{-1} \le 2$, so $u + u^{-1} \le u + 2 \le 3(u+1)$, and therefore there exists an absolute constant $c_1 > 0$ such that

$$\mathbb{P}(|g| \ge u) \ge c_1\,\frac{1}{u+1}\,e^{-u^2/2} \qquad (u > 1/2).$$

For $u \in [0, 1/2]$, we also have a crude lower bound $\mathbb{P}(|g| \ge u) \ge \mathbb{P}(|g| \ge 1/2) =: c_2 > 0$. Thus, after decreasing constants if needed, we may write for all $u \ge 0$,

$$\mathbb{P}(|g| \ge u) \ge c_3\,\frac{1}{u+1}\,e^{-u^2/2}.$$

Now fix $t \geq 0$ and set $u := t/K$. Then

$$\mathbb{P}(|Kg| \geq t) = \mathbb{P}(|g| \geq t/K) = \mathbb{P}(|g| \geq u) \geq c_3 \frac{1}{u+1} e^{-u^2/2}.$$

Using $\frac{1}{u+1} = e^{-\log(u+1)}$ and the crude inequality

$$\log(u+1) \leq \frac{u^2}{4} + 1 \qquad (u \geq 0),$$

we obtain

$$\frac{1}{u+1} \geq e^{-u^2/4-1},$$

and therefore

$$\mathbb{P}(|Kg| \geq t) \geq c_3 e^{-1} \exp\Big(-\frac{3}{4}u^2\Big) = c_4 \exp\Big(-c_5 \frac{t^2}{K^2}\Big),$$

for absolute constants $c_4, c_5 > 0$.

Choose $K := CL$ with $C > 0$ large enough so that $c_5/K^2 \geq c_0/L^2$, i.e. $C^2 \geq c_5/c_0$, and also so that $2c_4 \geq 1$ (increase C if necessary; constants are absolute). Then for all $t \geq 0$,

$$2e^{-c_0 t^2/L^2} \leq 2c_4 e^{-c_5 t^2/K^2} \leq 2\,\mathbb{P}(|Kg| \geq t).$$

Combining with the upper bound on $\mathbb{P}(|X| \geq t)$ yields

$$\mathbb{P}(|X| \geq t) \leq 2e^{-c_0 t^2/L^2} \leq 2\,\mathbb{P}(|Kg| \geq t) \qquad \forall\, t \geq 0,$$

so $X \preceq Kg$ with $K = CL$, as desired.

(a)(ii) Conversely, suppose that (1) holds for some $K > 0$, i.e.

$$\mathbb{P}(|X| \geq t) \leq 2\,\mathbb{P}(|Kg| \geq t) \qquad \forall\, t \geq 0,$$

where $g \sim \mathcal{N}(0,1)$. Then for all $t \geq 0$,

$$\mathbb{P}(|X| \geq t) \leq 2\,\mathbb{P}(|g| \geq t/K).$$

Using the standard Gaussian upper tail bound

$$\mathbb{P}(|g| \geq u) \leq 2e^{-u^2/2} \qquad (u \geq 0),$$

we obtain

$$\mathbb{P}(|X| \geq t) \leq 2\,\mathbb{P}(|g| \geq t/K) \leq 4\exp\Big(-\frac{t^2}{2K^2}\Big) \qquad \forall\, t \geq 0.$$

We now show that this Gaussian-type tail bound implies a finite ψ_2 norm. Using the tail-integral identity

$$\mathbb{E}e^{X^2/\alpha^2} = 1 + \int_0^\infty \mathbb{P}\big(e^{X^2/\alpha^2} \geq s\big)\,ds = 1 + \int_0^\infty \mathbb{P}\big(|X| \geq \alpha\sqrt{\log s}\big)\,ds,$$

and substituting the tail bound, we get

$$\mathbb{E}e^{X^2/\alpha^2} \leq 1 + 4\int_0^\infty \exp\Big(-\frac{\alpha^2}{2K^2}\log s\Big)\,ds = 1 + 4\int_0^\infty s^{-\alpha^2/(2K^2)}\,ds.$$

Choose $\alpha = cK$ with $c > \sqrt{2}$ so that $\alpha^2/(2K^2) > 1$. Then the integral converges and equals

$$\int_0^\infty s^{-p}\,ds < \infty \qquad (p > 1),$$

therefore $\mathbb{E}e^{X^2/\alpha^2} \leq 2$ for a suitable absolute choice of c. Therefore,

$$\|X\|_{\psi_2} \leq \alpha = cK,$$

which implies $\|X\|_{\psi_2} \lesssim K$, as desired.

Question 2: Sub-exponential characterizations

(a1) (i) $\Rightarrow$ (ii). Assume that for some $K > 0$,

$$\mathbb{P}(|Z| \geq t) \leq 2\exp(-t/K) \qquad \forall t \geq 0.$$

Let $X := |Z| \geq 0$. By the tail-integral identity (HW1, Ex. 4), for any $p > 0$,

$$\mathbb{E}[X^p] = \int_0^\infty p\, t^{p-1}\, \mathbb{P}(X > t)\, dt.$$

Therefore, for any $p \geq 1$,

$$\begin{aligned}\mathbb{E}|Z|^p &= \int_0^\infty p\, t^{p-1}\, \mathbb{P}(|Z| > t)\, dt \\ &\leq \int_0^\infty p\, t^{p-1} \cdot 2e^{-t/K}\, dt = 2p \int_0^\infty t^{p-1} e^{-t/K}\, dt.\end{aligned}$$

With the change of variables $u = t/K$ (so $t = Ku$, $dt = K\, du$), we get

$$\mathbb{E}|Z|^p \leq 2pK^p \int_0^\infty u^{p-1} e^{-u}\, du = 2pK^p\, \Gamma(p) = 2pK^p\, (p-1)! = 2K^p\, p!.$$

Finally, using the elementary bound $p! \leq p^p$, we have $(p!)^{1/p} \leq p$. therefore

$$(\mathbb{E}|Z|^p)^{1/p} \leq (2K^p p!)^{1/p} = K\, 2^{1/p} (p!)^{1/p} \leq K \cdot 2 \cdot p.$$

Thus (ii) holds with an absolute constant $C = 2$:

$$(\mathbb{E}|Z|^p)^{1/p} \leq CKp \qquad \forall p \geq 1.$$

(a2) (ii) $\Rightarrow$ (iii). Assume (ii): there exists an absolute constant $C_0 > 0$ such that for all $p \geq 1$,

$$(\mathbb{E}|Z|^p)^{1/p} \leq C_0 K\, p, \qquad \text{therefore} \qquad \mathbb{E}|Z|^p \leq (C_0 K p)^p.$$

Let $\alpha > 0$ be a constant to be chosen. Using the power series expansion of the exponential and

Tonelli's theorem (all terms are nonnegative),

$$\mathbb{E}\exp\left(\frac{|Z|}{\alpha K}\right) = \sum_{p=0}^{\infty} \frac{\mathbb{E}|Z|^p}{(\alpha K)^p\, p!} \le \sum_{p=0}^{\infty} \frac{(C_0 K p)^p}{(\alpha K)^p\, p!} = \sum_{p=0}^{\infty} \left(\frac{C_0}{\alpha}\right)^p \frac{p^p}{p!}.$$

By Stirling's lower bound $p! \ge (p/e)^p$, we have $\frac{p^p}{p!} \le e^p$, so

$$\mathbb{E}\exp\left(\frac{|Z|}{\alpha K}\right) \le \sum_{p=0}^{\infty} \left(\frac{eC_0}{\alpha}\right)^p.$$

Choose $\alpha = 2eC_0$. Then $(eC_0)/\alpha = 1/2$ and the series is geometric:

$$\mathbb{E}\exp\left(\frac{|Z|}{2eC_0 K}\right) \le \sum_{p=0}^{\infty} 2^{-p} = 2.$$

Thus (iii) holds after adjusting constants, i.e., replacing K by $K' := 2eC_0 K$.

(a3) (iii) $\Rightarrow$ (iv). Assume (iii): $\mathbb{E}\exp(|Z|/K) \le 2$ and $\mathbb{E}Z = 0$. Fix $\lambda \in \mathbb{R}$ with $|\lambda| \le \frac{1}{2K}$. We use the elementary inequality

$$e^x \le 1 + x + \frac{x^2}{2} e^{|x|} \qquad (\forall x \in \mathbb{R}).$$

With $x = \lambda Z$ this gives

$$e^{\lambda Z} \le 1 + \lambda Z + \frac{\lambda^2 Z^2}{2} e^{|\lambda Z|}.$$

Taking expectations and using $\mathbb{E}Z = 0$,

$$\mathbb{E}e^{\lambda Z} \le 1 + \frac{\lambda^2}{2}\, \mathbb{E}\big[Z^2 e^{|\lambda Z|}\big]. \tag{1}$$

Step 1: Bound $\mathbb{E}[Z^2 e^{|\lambda Z|}]$. First note that for all $t \ge 0$,

$$t^2 \le C_1 K^2 e^{t/K}, \tag{2}$$

for an absolute constant $C_1 > 0$ (e.g. take $C_1 = 1$, because $x^2 \le e^x$ for all $x \ge 0$ after rescaling). Applying (2) with $t = |Z|$ yields

$$Z^2 \le C_1 K^2 e^{|Z|/K}.$$

Moreover, because $|\lambda| \le \frac{1}{2K}$, we have $e^{|\lambda Z|} \le e^{|Z|/(2K)}$. Therefore,

$$Z^2 e^{|\lambda Z|} \le C_1 K^2 \exp\!\left(\frac{|Z|}{K}\right) \exp\!\left(\frac{|Z|}{2K}\right) = C_1 K^2 \exp\!\left(\frac{3|Z|}{2K}\right).$$

To control the last expectation, use the tail bound implied by (iii): for $t \ge 0$,

$$\mathbb{P}(|Z| \ge t) \le \mathbb{E}e^{|Z|/K}\, e^{-t/K} \le 2e^{-t/K}.$$

Then for any $\beta \in (0,1)$,

$$\mathbb{E}e^{\beta|Z|/K} = 1 + \int_0^\infty \frac{\beta}{K} e^{\beta t/K} \mathbb{P}(|Z| \ge t)\, dt \le 1 + \int_0^\infty \frac{\beta}{K} e^{\beta t/K} \cdot 2e^{-t/K}\, dt = 1 + \frac{2\beta}{1-\beta}.$$

Taking $\beta = \frac{1}{2}$ gives $\mathbb{E}e^{|Z|/(2K)} \le 3$. therefore, by Hölder,

$$\mathbb{E}e^{\frac{3|Z|}{2K}} = \mathbb{E}\Big[e^{\frac{|Z|}{K}} e^{\frac{|Z|}{2K}}\Big] \le \left(\mathbb{E}e^{|Z|/K}\right)\left(\mathbb{E}e^{|Z|/(2K)}\right) \le 2 \cdot 3 = 6,$$

and consequently,

$$\mathbb{E}\big[Z^2 e^{|\lambda Z|}\big] \le 6C_1K^2.$$

Step 2: Local quadratic log-MGF. Plugging this bound into (1) yields

$$\mathbb{E}e^{\lambda Z} \le 1 + C_2\lambda^2K^2, \qquad C_2 := 3C_1.$$

Using $1 + u \le e^u$ for all $u \in \mathbb{R}$, we obtain

$$\mathbb{E}e^{\lambda Z} \le \exp\!\big(C_2\lambda^2K^2\big),$$

and therefore

$$\psi_Z(\lambda) = \log \mathbb{E}e^{\lambda Z} \le C_2\lambda^2K^2 \qquad \text{for all } |\lambda| \le \frac{1}{2K}.$$

This is (iv) with absolute constants $c = \frac{1}{2}$ and $C = C_2$.

(a4) (iv) $\Rightarrow$ (i). Assume (iv): there exist absolute constants $c_0, C_0 > 0$ such that for all $|\lambda| \le c_0/K$,

$$\psi_Z(\lambda) = \log \mathbb{E} e^{\lambda Z} \le C_0 \lambda^2 K^2.$$

Fix $t \ge 0$. For any $\lambda \in [0, c_0/K]$, by Chernoff's bound,

$$\mathbb{P}(Z \ge t) = \mathbb{P}\left(e^{\lambda Z} \ge e^{\lambda t}\right) \le e^{-\lambda t} \mathbb{E} e^{\lambda Z} \le \exp\bigl(-\lambda t + \psi_Z(\lambda)\bigr) \le \exp\bigl(-\lambda t + C_0 \lambda^2 K^2\bigr).$$

We minimize the RHS over $\lambda \in [0, c_0/K]$. The unconstrained minimizer of $-\lambda t + C_0 \lambda^2 K^2$ is

$$\lambda^* = \frac{t}{2 C_0 K^2}.$$

We consider two cases.

Case 1: $t \le 2 C_0 c_0 K$. Then $\lambda^* \le c_0/K$, so we may plug $\lambda = \lambda^*$:

$$\mathbb{P}(Z \ge t) \le \exp\left(-\frac{t^2}{4 C_0 K^2}\right).$$

Case 2: $t > 2 C_0 c_0 K$. Then $\lambda^* > c_0/K$, so we take the endpoint $\lambda = c_0/K$:

$$\mathbb{P}(Z \ge t) \le \exp\left(-\frac{c_0}{K} t + C_0 \frac{c_0^2}{K^2} K^2\right) = \exp\left(-\frac{c_0}{K} t + C_0 c_0^2\right) \le \exp\left(-\frac{c_0}{2K} t\right),$$

where the last inequality uses $t > 2 C_0 c_0 K$ to absorb the additive constant $C_0 c_0^2$ into the linear term.

Combining the two cases, there exists an absolute constant $c > 0$ such that for all $t \ge 0$,

$$\mathbb{P}(Z \ge t) \le \exp\left(-c \min\left(\frac{t^2}{K^2}, \frac{t}{K}\right)\right).$$

Applying the same argument to $-Z$ gives the two-sided bound

$$\mathbb{P}(|Z| \ge t) \le 2 \exp\left(-c \min\left(\frac{t^2}{K^2}, \frac{t}{K}\right)\right).$$

In particular, because $\min(\frac{t^2}{K^2}, \frac{t}{K}) \geq \frac{t}{K}$ for all $t \geq 0$, we obtain (i) after adjusting constants:

$$\mathbb{P}(|Z| \geq t) \leq 2\exp\left(-\frac{t}{K'}\right) \quad \text{for some } K' \asymp K. \quad \square$$

Question 3: Bounding ℓ_p norm of random vector

a. Expected ℓ_p norm (finite p). Fix $p \in [1, \infty)$. Recall

$$\|X\|_p = \Big(\sum_{i=1}^{n} |X_i|^p\Big)^{1/p}.$$

Let $S := \sum_{i=1}^{n} |X_i|^p$. because the map $\phi(u) = u^{1/p}$ is concave on $\mathbb{R}_+$ for $p \geq 1$, by Jensen's inequality we have

$$\mathbb{E}\|X\|_p = \mathbb{E}\, S^{1/p} \;\leq\; \big(\mathbb{E}S\big)^{1/p} = \Big(\sum_{i=1}^{n} \mathbb{E}|X_i|^p\Big)^{1/p}.$$

Using the given sub-Gaussian moment bound (for $q = p \geq 1$),

$$\big(\mathbb{E}|X_i|^p\big)^{1/p} \leq CK\sqrt{p} \quad \Longrightarrow \quad \mathbb{E}|X_i|^p \leq (CK\sqrt{p})^p.$$

Therefore,

$$\mathbb{E}\|X\|_p \leq \Big(n\,(CK\sqrt{p})^p\Big)^{1/p} = CK\sqrt{p}\,n^{1/p}.$$

This proves that for every $p \in [1, \infty)$,

$$\boxed{\mathbb{E}\|X\|_p \;\leq\; CK\sqrt{p}\,n^{1/p}}.$$

(b) Expected ℓ_∞ norm.

Let $Y := \|X\|_\infty = \max_{1\leq i\leq n} |X_i| \geq 0$.

Step 1: Union bound + sub-Gaussian tail. For any $u \geq 0$,

$$\{Y \geq u\} = \Big\{\max_{1\leq i\leq n} |X_i| \geq u\Big\} = \bigcup_{i=1}^{n} \{|X_i| \geq u\}.$$

By the union bound,

$$\mathbb{P}(Y \geq u) \leq \sum_{i=1}^{n} \mathbb{P}(|X_i| \geq u).$$

Using the sub-Gaussian tail bound,

$$\mathbb{P}(|X_i| \geq u) \leq 2\exp\Big(-c\,\frac{u^2}{K^2}\Big),$$

therefore

$$\mathbb{P}(Y \geq u) \leq 2n\exp\Big(-c\,\frac{u^2}{K^2}\Big). \tag{3}$$

Step 2: Choose $u = CK(\sqrt{\log n} + t)$. Fix $t \geq 0$ and set

$$u = AK(\sqrt{\log n} + t),$$

where $A > 0$ is a universal constant. because $(a+b)^2 \geq a^2 + b^2$ for $a, b \geq 0$,

$$\frac{u^2}{K^2} = A^2(\sqrt{\log n} + t)^2 \geq A^2(\log n + t^2).$$

Substituting into (3) gives

$$\mathbb{P}\big(Y \geq AK(\sqrt{\log n} + t)\big) \leq 2ne^{-cA^2 \log n}e^{-cA^2t^2} = 2\,n^{1-cA^2}e^{-cA^2t^2}.$$

Choosing A large enough so that $cA^2 \geq 1$ yields $n^{1-cA^2} \leq 1$, therefore there exist universal constants $C, c' > 0$ such that for all $t \geq 0$,

$$\boxed{\mathbb{P}\Big(\|X\|_\infty \geq CK(\sqrt{\log n} + t)\Big) \leq 2e^{-c't^2}}.$$

Step 3: Deduce the expectation bound.

For any nonnegative random variable,

$$\mathbb{E}Y = \int_0^\infty \mathbb{P}(Y \geq s)\,ds.$$

Let $s_0 := CK\sqrt{\log n}$. Split the integral:

$$\mathbb{E}Y = \int_0^{s_0} \mathbb{P}(Y \geq s)\, ds + \int_{s_0}^{\infty} \mathbb{P}(Y \geq s)\, ds.$$

First term. because probabilities are always bounded by 1,

$$\mathbb{P}(Y \geq s) \leq 1,$$

we obtain

$$\int_0^{s_0} \mathbb{P}(Y \geq s)\, ds \leq \int_0^{s_0} 1\, ds = s_0.$$

Second term. For $s \geq s_0$, write

$$s = s_0 + CK\, t, \quad \text{so that} \quad t = \frac{s - s_0}{CK} \geq 0.$$

The tail bound then implies

$$\mathbb{P}(Y \geq s) \leq 2 \exp\Big(-c' \Big(\tfrac{s-s_0}{CK}\Big)^2 \Big).$$

therefore

$$\int_{s_0}^{\infty} \mathbb{P}(Y \geq s)\, ds \leq \int_{s_0}^{\infty} 2 \exp\Big(-c' \Big(\tfrac{s-s_0}{CK}\Big)^2 \Big)\, ds.$$

Perform the change of variables

$$r = \frac{s - s_0}{CK}, \quad ds = CK\, dr,$$

which yields

$$= 2CK \int_0^{\infty} e^{-c' r^2}\, dr.$$

The integral $\int_0^\infty e^{-c'r^2}\, dr$ is a Gaussian integral, which is finite and in fact equals

$$\int_0^{\infty} e^{-c' r^2}\, dr = \frac{1}{2}\sqrt{\frac{\pi}{c'}}.$$

Therefore the second term is bounded by a constant multiple of K:

$$\int_{s_0}^{\infty} \mathbb{P}(Y \geq s)\, ds \leq C_1 K.$$

Combining both parts, we conclude:

$$\mathbb{E}\|X\|_\infty = \mathbb{E}Y \leq s_0 + C_1 K = CK\sqrt{\log n} + C_1 K \leq C_2 K\sqrt{\log n},$$

for a universal constant $C_2 > 0$ (absorbing the additive K term into $K\sqrt{\log n}$ for $n \geq e$).

$$\boxed{\mathbb{E}\|X\|_\infty \leq CK\sqrt{\log n}}.$$

For $1 \leq n < e$, we have $\sqrt{\log n} \leq 1$, so the bound

$$\mathbb{E}\|X\|_\infty \leq CK\sqrt{\log n} + C_1 K \leq (C + C_1)K$$

holds; by enlarging the universal constant C_2 if necessary. Therefore, this implies:

$$\mathbb{E}\|X\|_\infty \leq C_2 K\sqrt{\log n} \quad \text{for all } n \geq 1.$$

c. Two regimes of expected ℓ_p norms.

We distinguish two cases.

Case 1: $1 \leq p \leq \log n$. From part (a), we directly obtain:

$$\mathbb{E}\|X\|_p \leq CK\sqrt{p}\, n^{1/p}.$$

Case 2: $p \geq \log n$. For any $x \in \mathbb{R}^n$, we have the deterministic inequality

$$\|x\|_p = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p} \leq \Big(n \max_i |x_i|^p\Big)^{1/p} = n^{1/p}\|x\|_\infty.$$

Applying this to X and taking expectations, we get:

$$\mathbb{E}\|X\|_p \le n^{1/p}\,\mathbb{E}\|X\|_\infty.$$

By part (b), we have:

$$\mathbb{E}\|X\|_\infty \le CK\sqrt{\log n},$$

Therefore

$$\mathbb{E}\|X\|_p \le CK\,n^{1/p}\sqrt{\log n}.$$

because $p \ge \log n$,

$$n^{1/p} = e^{\frac{\log n}{p}} \le e,$$

so this factor is bounded by a universal constant. Therefore

$$\mathbb{E}\|X\|_p \le CK\sqrt{\log n}.$$

Combining both cases, we can conclude:

$$\mathbb{E}\|X\|_p \lesssim \begin{cases} CK\sqrt{p}\,n^{1/p}, & 1 \le p \le \log n, \\ \\ CK\sqrt{\log n}, & p \ge \log n, \end{cases}$$

up to universal constants.

Question 4: Bound L_p norms of linear forms

a. Sub-Gaussian Khintchine L_p bound ($p \ge 2$).

Let

$$S(a) = \sum_{i=1}^{N} a_i X_i, \qquad \|a\|_2 = \Big(\sum_{i=1}^{N} a_i^2\Big)^{1/2}.$$

Assume the X_i are independent with $\mathbb{E}X_i = 0$, $\mathbb{E}X_i^2 = 1$, and

$$\|X_i\|_{\psi_2} \le K_2 \quad \text{for all } i.$$

Lower bound. For $p \geq 2$, $\|Y\|_{L_p} \geq \|Y\|_{L_2}$. Moreover,

$$\begin{aligned}\|S(a)\|_{L_2}^2 &= \mathbb{E}\Big(\sum_{i=1}^N a_i X_i\Big)^2 \\ &= \sum_{i=1}^N a_i^2 \mathbb{E}X_i^2 + \sum_{i\neq j} a_i a_j \mathbb{E}[X_i X_j].\end{aligned}$$

By independence and $\mathbb{E}X_i = 0$, we have $\mathbb{E}[X_i X_j] = 0$ for $i \neq j$, and $\mathbb{E}X_i^2 = 1$. therefore

$$\|S(a)\|_{L_2}^2 = \sum_{i=1}^N a_i^2 = \|a\|_2^2,$$

so

$$\|S(a)\|_{L_p} \geq \|a\|_2.$$

Upper bound.

because $\|X_i\|_{\psi_2} \leq K_2$, by definition of the ψ_2 norm there exists $\widetilde{K} \leq 2K_2$ such that

$$\mathbb{E}\exp(X_i^2/\widetilde{K}^2) \leq 2.$$

For any $\lambda \in \mathbb{R}$, the inequality

$$\lambda x \leq \frac{x^2}{2\widetilde{K}^2} + \frac{\lambda^2\widetilde{K}^2}{2} \quad \text{(because } (x/\widetilde{K} - \lambda\widetilde{K})^2 \geq 0\text{)}$$

implies

$$e^{\lambda X_i} \leq \exp\Big(\frac{X_i^2}{2\widetilde{K}^2}\Big)\exp\Big(\frac{\lambda^2\widetilde{K}^2}{2}\Big).$$

Taking expectations and using Cauchy–Schwarz,

$$\mathbb{E}e^{\lambda X_i} \leq \exp\Big(\frac{\lambda^2\widetilde{K}^2}{2}\Big)\Big(\mathbb{E}\exp(X_i^2/\widetilde{K}^2)\Big)^{1/2} \leq \sqrt{2}\, e^{\lambda^2\widetilde{K}^2/2} \leq e^{C\lambda^2 K_2^2}.$$

Replacing λ by λa_i gives

$$\mathbb{E}e^{\lambda a_i X_i} \leq \exp(C\lambda^2 a_i^2 K_2^2).$$

because the X_i are independent,

$$\begin{aligned}\mathbb{E}e^{\lambda S(a)} &= \prod_{i=1}^{N} \mathbb{E}e^{\lambda a_i X_i} \\ &\leq \exp\Big(C\lambda^2 K_2^2 \sum_{i=1}^{N} a_i^2\Big) = \exp\big(C\lambda^2 K_2^2 \|a\|_2^2\big).\end{aligned}$$

By Chernoff's bound, for any $t > 0$,

$$\mathbb{P}(S(a) \geq t) \leq \exp\Big(-\lambda t + C\lambda^2 K_2^2 \|a\|_2^2\Big).$$

Optimizing over λ yields

$$\mathbb{P}(|S(a)| \geq t) \leq 2\exp\Big(-c\,\frac{t^2}{K_2^2\|a\|_2^2}\Big).$$

Finally, using

$$\mathbb{E}|Y|^p = \int_0^\infty pt^{p-1}\,\mathbb{P}(|Y| \geq t)\,dt,$$

the Gaussian tail above implies for all $p \geq 1$,

$$\|S(a)\|_{L_p} \leq CK_2\sqrt{p}\,\|a\|_2.$$

Combining both bounds, we obtain:

$$\|a\|_2 \leq \|S(a)\|_{L_p} \leq CK_2\sqrt{p}\,\|a\|_2, \qquad p \geq 2.$$

b. Sub-exponential Bernstein–Khintchine tail bound. Assume $\mathbb{E}X_i = 0$ and $\|X_i\|_{\psi_1} \leq K_1$ for all i. Let

$$S(a) = \sum_{i=1}^{N} a_i X_i, \qquad \|a\|_2 = \Big(\sum_{i=1}^{N} a_i^2\Big)^{1/2}, \qquad \|a\|_\infty = \max_{1\leq i\leq N} |a_i|.$$

Define $Y_i := a_i X_i$. because the ψ_1-norm scales linearly, we have:

$$\|Y_i\|_{\psi_1} = \|a_i X_i\|_{\psi_1} = |a_i|\,\|X_i\|_{\psi_1} \leq |a_i| K_1.$$

therefore each Y_i is mean-zero sub-exponential and the Y_i are independent. Set

$$K_i := \|Y_i\|_{\psi_1}, \qquad V := \sum_{i=1}^{N} K_i^2, \qquad B := \max_{1\le i\le N} K_i.$$

Then

$$B = \max_i \|Y_i\|_{\psi_1} \le K_1 \max_i |a_i| = K_1\|a\|_\infty, \qquad V = \sum_i \|Y_i\|_{\psi_1}^2 \le K_1^2 \sum_i a_i^2 = K_1^2\|a\|_2^2.$$

By Bernstein's inequality for sums of independent mean-zero sub-exponential random variables, for all $t \ge 0$, we have:

$$\mathbb{P}\Big\{\sum_{i=1}^{N} Y_i \ge t\Big\} \le 2\exp\Big(-c\,\min\Big(\frac{t^2}{V}, \frac{t}{B}\Big)\Big),$$

where $c > 0$ is an absolute constant. Applying the same bound to $-\sum_i Y_i$ and using $\sum_i Y_i = S(a)$ gives

$$\mathbb{P}\{|S(a)| \ge t\} \le 2\exp\Big(-c\,\min\Big(\frac{t^2}{V}, \frac{t}{B}\Big)\Big).$$

Finally, substitute the bounds on V and B:

$$\mathbb{P}\{|S(a)| \ge t\} \le 2\exp\Bigg(-c\,\min\Big(\frac{t^2}{K_1^2\|a\|_2^2}, \frac{t}{K_1\|a\|_\infty}\Big)\Bigg), \qquad t \ge 0,$$

which is the desired result.

c. Sub-exponential Bernstein–Khintchine L_p bound ($p \ge 2$).

From part (b), there exists an absolute constant $c > 0$ such that for all $t \ge 0$, we have:

$$\mathbb{P}\{|S(a)| \ge t\} \le 2\exp\Bigg(-c\,\min\Big(\frac{t^2}{K_1^2\|a\|_2^2}, \frac{t}{K_1\|a\|_\infty}\Big)\Bigg). \tag{4}$$

Set $\sigma := K_1\|a\|_2$ and $b := K_1\|a\|_\infty$. Then (4) can be rewritten as follows:

$$\mathbb{P}\{|S(a)| \ge t\} \le 2\exp\Big(-c\min(t^2/\sigma^2,\ t/b)\Big).$$

Using the tail integral identity:

$$\mathbb{E}|S(a)|^p = \int_0^\infty p t^{p-1} \mathbb{P}(|S(a)| \ge t)\, dt,$$

and the inequality $e^{-c\min(u,v)} \le e^{-cu} + e^{-cv}$ for $u, v \ge 0$, we obtain:

$$\begin{aligned}
\mathbb{E}|S(a)|^p &\le 2p \int_0^\infty t^{p-1} \exp\Big(- c\, \min(t^2/\sigma^2,\ t/b)\Big)\, dt \\
&\le 2p \int_0^\infty t^{p-1} e^{-ct^2/\sigma^2}\, dt + 2p \int_0^\infty t^{p-1} e^{-ct/b}\, dt.
\end{aligned}$$

For the "Gaussian part", apply the Gamma-type estimate

$$\int_0^\infty t^{p-1} e^{-\alpha t^2}\, dt = \frac{1}{2}\alpha^{-p/2}\Gamma(p/2), \qquad \alpha > 0,$$

with $\alpha = c/\sigma^2$:

$$2p \int_0^\infty t^{p-1} e^{-ct^2/\sigma^2}\, dt = 2p \cdot \frac{1}{2}\Big(\frac{\sigma^2}{c}\Big)^{p/2} \Gamma(p/2) \le C^p\, p\, \sigma^p\, \Gamma(p/2).$$

Using the crude bound $\Gamma(p/2) \le (Cp)^{p/2}$ for $p \ge 2$, we obtain:

$$2p \int_0^\infty t^{p-1} e^{-ct^2/\sigma^2}\, dt \le (C\sigma\sqrt{p})^p.$$

For the "exponential part", use

$$\int_0^\infty t^{p-1} e^{-\alpha t}\, dt = \alpha^{-p}\Gamma(p), \qquad \alpha > 0,$$

with $\alpha = c/b$ to get:

$$2p \int_0^\infty t^{p-1} e^{-ct/b}\, dt = 2p\Big(\frac{b}{c}\Big)^p \Gamma(p) \le C^p\, p\, b^p\, \Gamma(p).$$

Using the crude bound $\Gamma(p) \le (Cp)^p$ for $p \ge 2$ gives

$$2p \int_0^\infty t^{p-1} e^{-ct/b}\, dt \le (Cbp)^p.$$

Combining both parts, we can conclude:

$$\mathbb{E}|S(a)|^p \le (C\sigma\sqrt{p})^p + (Cbp)^p \le \big(C(\sigma\sqrt{p} + bp)\big)^p,$$

where we used $(x^p + y^p)^{1/p} \le x + y$ for $x, y \ge 0$. Taking pth roots and recalling $\sigma = K_1\|a\|_2$ and $b = K_1\|a\|_\infty$, we obtain for all $p \ge 2$

$$\|S(a)\|_{L_p} \le C\big(\sigma\sqrt{p} + bp\big) = CK_1\big(\sqrt{p}\,\|a\|_2 + p\,\|a\|_\infty\big)$$

which is the claim.

Question 5:

a. Mean and variance.

We have: $Z = g^\top A g = \sum_{i,j=1}^n a_{ij} g_i g_j = 2\sum_{1\le i<j\le n} a_{ij} g_i g_j \qquad (A \text{ symmetric}, a_{ii} = 0).$

a1. Compute $\mathbb{E}[Z]$. We have:

$$\mathbb{E}Z = \mathbb{E}\sum_{i,j} a_{ij} g_i g_j = \sum_{i,j} a_{ij}\, \mathbb{E}[g_i g_j] \qquad \text{(linearity)}$$

We have: $\mathbb{E}[g_i g_j] = \mathbb{E}[g_i]\mathbb{E}[g_j] = 0$ when $i \ne j$ because independence and $\mathbb{E}[g_i g_j] = E[g_i^2] = \mathrm{Var}[g_i] = 1$ when $i = j$. Denote $\delta_{ij} = 0$ when $i \ne j$ and $\delta_{ij} = 1$ when $i = j$, we obtain:

$$\begin{aligned}\mathbb{E}[Z] &= \sum_{i,j} a_{ij}\,\delta_{ij} \\ &= \sum_{i=1}^n a_{ii} = \mathrm{tr}(A) = 0 \qquad (\mathrm{diag}(A) = 0).\end{aligned}$$

a2. Compute $\mathrm{Var}[Z]$. We have: $\mathrm{Var}[Z] = \mathbb{E}[Z^2]$ because $\mathbb{E}[Z] = 0$.

We have:

$$Z = 2\sum_{i<j} a_{ij} g_i g_j.$$

Therefore:

$$Z^2 = 4\sum_{i<j}\sum_{k<\ell} a_{ij}a_{k\ell}\, g_i g_j g_k g_\ell$$

$$\Rightarrow \mathbb{E}[Z^2] = 4\sum_{i<j}\sum_{k<\ell} a_{ij}a_{k\ell}\, \mathbb{E}[g_i g_j g_k g_\ell]$$

Moreover:

$$\mathbb{E}[g_i g_j g_k g_\ell] = \delta_{ij}\delta_{k\ell} + \delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk} \qquad \text{(by Wick's formula)}$$

We also have:

$$\delta_{ij}\delta_{k\ell} = 0 \qquad (\text{if } i \neq j,\ k \neq \ell),$$

$$\delta_{ik}\delta_{j\ell} \neq 0 \iff i = k,\ j = \ell,$$

$$\delta_{i\ell}\delta_{jk} \neq 0 \iff i = \ell,\ j = k \qquad (\text{impossible because } i < j,\ k < \ell).$$

Therefore:

$$\mathbb{E}[g_i g_j g_k g_\ell] = \begin{cases} \mathbb{E}[g_i^2 g_j^2] = 1, & (i,j) = (k,\ell) \\ 0, & \text{otherwise} \end{cases}$$

This implies $\mathbb{E}[Z^2] = 4\sum_{i<j} a_{ij}^2$. Moreover $\|A\|_F^2 = \sum_{i,j} a_{ij}^2 = 2\sum_{i<j} a_{ij}^2$(because A is symmetric, diag$= 0$).

Therefore, we conclude that: $\boxed{\mathrm{Var}(Z) = 2\|A\|_F^2.}$

b. Variance bound via Gaussian Poincaré.

Let $f(x) = x^\top A$, and $Z = f(g) = g^\top A g$ with $g \sim N(0, I_n)$.

Apply Gaussian Poincaré inequality, we obtain:

$$\mathrm{Var}(Z) = \mathrm{Var}(f(g)) \leq \mathbb{E}\|\nabla f(g)\|_2^2$$

We have: if A is symmetric then $\nabla f(x) = 2Ax$. Indeed:

$$\begin{aligned} f(x) &= \sum_{i,j=1}^{n} a_{ij} x_i x_j, \\ \Rightarrow \frac{\partial f}{\partial x_k}(x) &= \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i \qquad \text{(differentiate termwise)} \\ &= 2 \sum_{j=1}^{n} a_{kj} x_j \qquad \text{(because } A \text{ is symmetric)} \\ &\Rightarrow \nabla f(x) = 2Ax \end{aligned}$$

Apply the above result, we obtain:

$$\begin{aligned} \mathrm{Var}(Z) &\le \mathbb{E}\|2Ag\|_2^2 = 4\,\mathbb{E}\|Ag\|_2^2 \\ &= 4\,\mathbb{E}\big[g^\top A^\top A g\big] \qquad \text{(because } \|Ag\|_2^2 = g^\top A^\top A g\text{)} \\ &= 4\,\mathrm{tr}(A^\top A) \qquad \text{(because } \mathbb{E}[g^\top M g] = \mathbb{E}[\sum_{i,j=1}^{n} m_{ij} g_i g_j] = \sum_{i1}^{n} m_{ii} = \mathrm{tr}(M) \text{ for } g \sim N(0, I)\text{)} \\ &= 4\|A\|_F^2. \end{aligned}$$

Conclusion: $\boxed{\mathrm{Var}(Z) \le 4\|A\|_F^2.}$

From part a, we have $\mathrm{Var}(Z) = 2\|A\|_F^2$, this shows that Gaussian Poincaré is not tight here; the extra factor comes from using an inequality.

(c) Diagonalization and distributional representation.

Diagonalize A, we have:

$$A = U^\top \Lambda U, \qquad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.

Then $Z = g^\top A g = g^\top U^\top \Lambda U g = (Ug)^\top \Lambda (Ug)$.

Let $y := Ug$. Because U is orthogonal and $g \sim N(0, I_n)$, the Gaussian distribution is rotationally invariant, we have:

$$y \sim N(0, I_n)$$

Therefore:

$$Z = y^\top \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2.$$

Because $\mathrm{tr}(A) = 0$ ($\mathrm{diag}(A) = 0$), the sum of eigenvalues is zero:

$$\sum_{i=1}^{n} \lambda_i = 0.$$

Therefore, can write Z as:

$$Z = \sum_{i=1}^{n} \lambda_i (y_i^2 - 1).$$

Finally, because the coordinates of y are i.i.d. $N(0,1)$ (same distribution as g_i), we can conclude:

$$\boxed{Z \overset{d}{=} \sum_{i=1}^{n} \lambda_i (g_i^2 - 1).}$$

d. CGF of a centered χ_1^2.

Let $G \sim N(0,1)$ and define $Y := G^2 - 1$. We compute the moment generating function of Y.

First, we have:

$$\mathbb{E}e^{\theta Y} = \mathbb{E}e^{\theta(G^2-1)} = e^{-\theta}\, \mathbb{E}e^{\theta G^2}.$$

To evaluate $\mathbb{E}e^{\theta G^2}$, use the density of the standard normal:

$$\begin{aligned} \mathbb{E}e^{\theta G^2} &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\Big(\theta x^2 - \frac{x^2}{2}\Big)\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\Big(-\frac{1-2\theta}{2} x^2\Big)\, dx \end{aligned}$$

This is a Gaussian integral, which equals $1/\sqrt{1-2\theta}$ if $1 - 2\theta > 0$, i.e., $\theta < \frac{1}{2}$. Therefore:

$$\mathbb{E}e^{\theta G^2} = \frac{1}{\sqrt{1-2\theta}}.$$

Thus $\mathbb{E}e^{\theta Y} = e^{-\theta}(1-2\theta)^{-1/2}$. Take taking logarithms, we have:

$$\boxed{\log \mathbb{E}e^{\theta Y} = -\theta - \tfrac{1}{2}\log(1-2\theta), \quad \theta < \tfrac{1}{2}.}$$

Next we prove the bound $\log \mathbb{E}e^{\theta Y} \leq \frac{\theta^2}{1-(2\theta)_+}, \qquad \theta < \frac{1}{2}$, where $(u)_+ := \max\{u, 0\}$.

Case 1: $\theta \leq 0$.

Then $(2\theta)_+ = 0$, so the right-hand side is θ^2. Let $x = -2\theta \geq 0$. Then

$$-\theta - \tfrac{1}{2}\log(1 - 2\theta) = \tfrac{x}{2} - \tfrac{1}{2}\log(1 + x).$$

Using the inequality $\log(1 + x) \geq x - x^2/2$ for $x \geq 0$, we obtain

$$\tfrac{x}{2} - \tfrac{1}{2}\log(1 + x) \leq \tfrac{x^2}{4} = \theta^2.$$

Case 2: $0 \leq \theta < \frac{1}{2}$.

Then $(2\theta)_+ = 2\theta$ and the right-hand side equals $\theta^2/(1 - 2\theta)$. Let $x = 2\theta \in [0, 1)$. The inequality we need to prove becomes

$$-x - \log(1 - x) \leq \frac{x^2}{2(1 - x)}.$$

Define $\phi(x) := -x - \log(1 - x) - \frac{x^2}{2(1-x)}$. We have $\phi'(x) = \frac{x^2}{2(1-x)^2} \geq 0$, so ϕ is increasing. Because $\phi(0) = 0$, we conclude $\phi(x) \leq 0$ for all $x \in [0, 1)$, proving the claim. Therefore:

$$\boxed{\log \mathbb{E}e^{\theta Y} \leq \frac{\theta^2}{1 - (2\theta)_+}, \quad \theta < \tfrac{1}{2}.}$$

e. [Bonus] A Bernstein-type CGF bound

From part (c), we have the distributional representation

$$Z \stackrel{d}{=} \sum_{i=1}^{n} \lambda_i(g_i^2 - 1),$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, and $g_i \sim \mathcal{N}(0, 1)$ are independent.

because the g_i are independent, we obtain

$$\begin{aligned}\log \mathbb{E} e^{\lambda Z} &= \log \mathbb{E} \exp\left(\lambda \sum_{i=1}^{n} \lambda_i (g_i^2 - 1)\right) \\ &= \sum_{i=1}^{n} \log \mathbb{E} \exp\bigl(\lambda \lambda_i (g_i^2 - 1)\bigr) .\end{aligned}$$

For each i, apply part (d) with

$$\theta = \lambda \lambda_i.$$

The bound in part (d) requires $\theta < \frac{1}{2}$. because $|\lambda_i| \leq \|A\|$, the condition

$$|\lambda| < \frac{1}{2\|A\|}$$

ensures

$$|\lambda \lambda_i| \leq |\lambda| \|A\| < \frac{1}{2},$$

so the bound from part (d) is valid for every i.

therefore,

$$\log \mathbb{E} e^{\lambda \lambda_i (g_i^2 - 1)} \leq \frac{\lambda^2 \lambda_i^2}{1 - 2(\lambda \lambda_i)_+}.$$

because $(x)_+ \leq |x|$, we have

$$(\lambda \lambda_i)_+ \leq |\lambda| |\lambda_i| \leq |\lambda| \|A\|.$$

Therefore,

$$1 - 2(\lambda \lambda_i)_+ \geq 1 - 2|\lambda| \|A\|.$$

Summing over i gives

$$\begin{aligned}\log \mathbb{E} e^{\lambda Z} &\leq \sum_{i=1}^{n} \frac{\lambda^2 \lambda_i^2}{1 - 2|\lambda| \|A\|} \\ &= \frac{\lambda^2}{1 - 2|\lambda| \|A\|} \sum_{i=1}^{n} \lambda_i^2.\end{aligned}$$

Finally, because

$$\sum_{i=1}^{n} \lambda_i^2 = \|A\|_F^2,$$

we conclude that for all

$$|\lambda| < \frac{1}{2\|A\|},$$

$$\boxed{\log \mathbb{E} e^{\lambda Z} \leq \frac{\lambda^2 \|A\|_F^2}{1 - 2|\lambda|\|A\|}}.$$