Theoretical Stats and Machine Learning (Homework 1)

Student Name: Vu Huyen Trang Pham

UT EID: vp22673

Question 1.

a. ℓ_p norms on $\mathbb{R}^n$.

a1. Comparison and monotonicity. Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $1 \le p \le q \le \infty$.

Prove $\|x\|_q \le \|x\|_p$ for $q < \infty$.

If $x = 0$ the claim is trivial. Otherwise, we define $a_i := \frac{|x_i|}{\|x\|_p} (i = 1, \dots, n)$. Therefore:

$$\sum_{i=1}^n a_i^p = \frac{\sum_{i=1}^n |x_i|^p}{\|x\|_p^p} = 1$$

which implies that $a_i^p \le 1$ and thus $0 \le a_i \le 1$. Because $q \ge p$ and $t \mapsto t^\alpha$ is decreasing in α on $t \in [0, 1]$, we have $a_i^q \le a_i^p$ for all i. We then obtain:

$$\sum_{i=1}^n a_i^q \le \sum_{i=1}^n a_i^p = 1$$

Therefore, $\|x\|_q^q = \sum_{i=1}^n |x_i|^q = \sum_{i=1}^n (\|x\|_p a_i)^q = \|x\|_p^q \sum_{i=1}^n a_i^q \le \|x\|_p^q$, which concludes that $\boxed{\|x\|_q \le \|x\|_p}$.

Prove $\|x\|_p \le n^{\frac{1}{p} - \frac{1}{q}} \|x\|_q$ for $q < \infty$ (Hölder).

By homogeneity, it suffices to assume $\|x\|_q = 1$ ($\sum_{i=1}^n |x_i|^q = 1$). We write:

$$\sum_{i=1}^n |x_i|^p = \sum_{i=1}^n |x_i|^p \cdot 1.$$

Apply Hölder's inequality with conjugate exponents $r = \frac{q}{p} \ge 1, s = \frac{q}{q-p}$ so that $\frac{1}{r} + \frac{1}{s} = 1$, we obtain:

$$\sum_{i=1}^n |x_i|^p \le \Big(\sum_{i=1}^n (|x_i|^p)^r \Big)^{1/r} \Big(\sum_{i=1}^n 1^s \Big)^{1/s} = \Big(\sum_{i=1}^n |x_i|^{pr} \Big)^{1/r} n^{1/s}$$

Because $pr = q$ and $\sum |x_i|^q = 1$, the above result implies:

$$\sum_{i=1}^{n} |x_i|^p \leq n^{1/s}$$
$$\Rightarrow \|x\|_p \leq n^{\frac{1}{p}-\frac{1}{q}} \qquad \text{(take pth roots and use } \frac{1}{s} = 1 - \frac{p}{q})$$

Undoing the normalization (multiplying by $\|x\|_q$), we have: $\boxed{\|x\|_p \leq n^{\frac{1}{p}-\frac{1}{q}} \|x\|_q}$.

The case $q = \infty$.

We have $|x_i| \leq \max_i |x_i| = \|x\|_\infty$, which implies:

$$\|x\|_p^p = \sum_{i=1}^{n} |x_i|^p \leq \sum_{i=1}^{n} \|x\|_\infty^p = n\|x\|_\infty^p$$

so $\boxed{\|x\|_p \leq n^{1/p}\|x\|_\infty = n^{\frac{1}{p}-\frac{1}{\infty}} \|x\|_\infty}$.

Deduction. Fix $x \in \mathbb{R}^n$ and let $1 \leq p \leq q \leq \infty$. From the already proved comparison $\|x\|_q \leq \|x\|_p$, we have that the function $p \mapsto \|x\|_p$ is (weakly) decreasing in p.

Moreover, the unit balls are nested. Indeed, if $x \in B_{\ell_p}^n$, then $\|x\|_p \leq 1$, and therefore $\|x\|_q \leq \|x\|_p \leq 1$, which implies $x \in B_{\ell_q}^n$ ($\boxed{B_{\ell_p}^n \subseteq B_{\ell_q}^n \text{ whenever } p \leq q)}$.

a2. When are the bounds tight?

Let $1 \leq p \leq q \leq \infty$.

Left inequality. The bound $\|x\|_q \leq \|x\|_p$ is tight for one-sparse vectors (vectors with only one nonzero coordinate). For example, let $\boxed{x = (3, 0, \ldots, 0) \in \mathbb{R}^n}$. Therefore:

$$\|x\|_p = (|3|^p)^{1/p} = 3 \qquad \|x\|_q = (|3|^q)^{1/q} = 3$$

which shows that $\|x\|_p = \|x\|_q$.

Right inequality. The bound $\|y\|_p \leq n^{\frac{1}{p}-\frac{1}{q}} \|y\|_q$ is tight for flat vectors (vectors with coordinates all have equal magnitude). For example, let $\boxed{y = (1, 1, \ldots, 1) \in \mathbb{R}^n}$. Therefore:

$$\|y\|_p = \left(\sum_{i=1}^{n} 1^p\right)^{1/p} = n^{1/p} \qquad \|y\|_q = \left(\sum_{i=1}^{n} 1^q\right)^{1/q} = n^{1/q}$$

which shows that $\|y\|_p = n^{\frac{1}{p}-\frac{1}{q}}\|y\|_q$.

a3. [Bonus] The ℓ_∞ limit.

Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ and define $\|x\|_\infty := \max_{1\le i\le n} |x_i|$.

Quantitative bounds. Let $M = \|x\|_\infty$. By definition, there exists j such that $|x_j| = M$. Therefore:

$$\sum_{i=1}^{n} |x_i|^p \ge |x_j|^p = M^p$$

which implies $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p} \ge M = \|x\|_\infty$.

On the other hand, because $|x_i| \le M$ for all i, we have $\sum_{i=1}^n |x_i|^p \le \sum_{i=1}^n M^p = nM^p$. Therefore:

$$\|x\|_p \le (nM^p)^{1/p} = n^{1/p}M = n^{1/p}\|x\|_\infty$$

From the above results, we can conclude: $\boxed{\|x\|_\infty \le \|x\|_p \le n^{1/p}\|x\|_\infty}$.

Convergence as $p \to \infty$. Because $n^{1/p} = e^{(\log n)/p} \longrightarrow 1$ as $p \to \infty$, $n^{1/p}\|x\|_\infty \longrightarrow \|x\|_\infty$. Apply the squeeze theorem for the quantitative bound derived above, we obtain:

$$\boxed{\|x\|_p \longrightarrow \|x\|_\infty \qquad \text{as } p \to \infty}$$

Additionally, if $p \ge \log n$, then $n^{1/p} = e^{(\log n)/p} \le e$. Therefore $\boxed{\|x\|_p \le n^{1/p}\|x\|_\infty \le e\,\|x\|_\infty}$.

b. ℓ_p norms of random variables.

b1. Monotonicity and inclusions.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a real-valued random variable.

Case I: $\le p \le q < \infty$. We consider the function $\varphi(t) = t^{q/p}(t \ge 0)$. Because $q/p \ge 1$, the function φ is convex on $[0, \infty)$. Apply Jensen's inequality, we obtain:

$$(\mathbb{E}|X|^p)^{q/p} \le \mathbb{E}\big(|X|^{p\cdot q/p}\big) = \mathbb{E}|X|^q$$

$$\iff (\mathbb{E}|X|^p)^{1/p} \le (\mathbb{E}|X|^q)^{1/q}$$

$$\iff \boxed{\|X\|_{L_p} \le \|X\|_{L_q}}$$

Case II: $q = \infty$. If $\|X\|_{L_\infty} < \infty$, then by the definition of the essential supremum, we have:

$$
\begin{aligned}
&|X| \le \|X\|_{L_\infty} \quad \text{a.s.} \\
&\Rightarrow |X|^p \le \|X\|_{L_\infty}^p \quad \text{a.s.} \\
&\Rightarrow \mathbb{E}|X|^p \le \|X\|_{L_\infty}^p \\
&\iff \boxed{\|X\|_{L_p} \le \|X\|_{L_\infty}}
\end{aligned}
$$

Set inclusion. If $X \in L_q$, then $\|X\|_{L_q} < \infty$. By the inequality proved above, we obtain $\|X\|_{L_p} \le \|X\|_{L_q} < \infty$, which implies $X \in L_p$. As a result, $\boxed{L_q \subseteq L_p \text{ for } q \ge p}$.

b2. No dimension-free equivalence.

Let $\alpha \in (p, q]$ and consider a random variable with Pareto-type tail:

$$\mathbb{P}(|X| > t) \sim t^{-\alpha} \qquad \text{as } t \to \infty$$

We know that:

$$\mathbb{E}|X|^r < \infty \quad \iff \quad r < \alpha \tag{1}$$

Indeed, recall the tail-integral representation of moments: for any $r > 0$, we have:

$$\mathbb{E}|X|^r = r \int_0^\infty t^{r-1}\, \mathbb{P}(|X| > t)\, dt$$

We now consider the existence of $\mathbb{E}|X|^r$. For simplicity, it is equivalent to consider the existence of $\int_1^\infty t^{r-1}\mathbb{P}(|X| > t)\, dt$ because because the contribution from small t is always finite and cannot affect whether the moment diverges.

By the assumption $\mathbb{P}(|X| > t) \sim t^{-\alpha}$, there exist constants $c_1, c_2 > 0$ and $t_0 > 1$ such that for all $t \ge t_0$,

$$c_1 t^{-\alpha} \le \mathbb{P}(|X| > t) \le c_2 t^{-\alpha}$$

Consequently, for t sufficiently large, we have:

$$c_1 t^{r-\alpha-1} \le t^{r-1}\mathbb{P}(|X| > t) \le c_2 t^{r-\alpha-1}$$

Therefore, $\int_1^\infty t^{r-\alpha-1}\,dt < \infty \iff \int_1^\infty t^{r-1}\mathbb{P}(|X| > t)\,dt < \infty$. Moreover:

$$\int_1^\infty t^{r-\alpha-1}\,dt < \infty \quad \iff \quad r - \alpha - 1 < -1 \quad \iff \quad r < \alpha$$

This implies that $\int_1^\infty t^{r-1}\mathbb{P}(|X| > t)\,dt < \infty \iff r < \alpha$, so 1 is proved.

Apply 1, we have $\mathbb{E}|X|^p < \infty$ but $\mathbb{E}|X|^q = \infty$. Therefore, $X \in L_p$ but $X \notin L_q$, which proves that no universal constant C can satisfy $\|X\|_{L_q} \le C\|X\|_{L_p}$ for all X.

Example. Let $p = 2$ and $q = 4$. Define a random variable X with density:

$$f(x) = 3x^{-4}, \qquad x \ge 1$$

Therefore, X has a Pareto-type tail with parameter $\alpha = 3$, because for $t \ge 1$, $\mathbb{P}(|X| > t) = \mathbb{P}(X > t) = \int_t^\infty 3x^{-4}\,dx = t^{-3}$.

We have:

$$\begin{cases} \mathbb{E}[X^2] = \int_1^\infty x^2 \cdot 3x^{-4}\,dx = 3\int_1^\infty x^{-2}\,dx < \infty, \\ \mathbb{E}[X^4] = \int_1^\infty x^4 \cdot 3x^{-4}\,dx = 3\int_1^\infty 1\,dx = \infty. \end{cases}$$

Therefore, $X \in L_2$ but $X \notin L_4$.

b3. [Bonus] The L_∞ limit.

Assume that $\|X\|_{L_\infty} < \infty$.

Upper bound. By the definition of the essential supremum, we have:

$$|X| \le \|X\|_{L_\infty} \quad \text{a.s.}$$

which implies that:

$$\mathbb{E}|X|^p \le \|X\|_{L_\infty}^p, \qquad \text{and therefore} \qquad \|X\|_{L_p} \le \|X\|_{L_\infty}.$$

Lower bound. By the definition of the essential supremum, for every $\varepsilon > 0$, we have:

$$\mathbb{P}\big(|X| > \|X\|_{L_\infty} - \varepsilon\big) > 0.$$

Let $A_\varepsilon := \big\{|X| > \|X\|_{L_\infty} - \varepsilon\big\}$. On this event, $|X|^p \geq (\|X\|_{L_\infty} - \varepsilon)^p$.

Therefore:

$$\begin{aligned} &\mathbb{E}|X|^p = \mathbb{E}\big(|X|^p \mathbf{1}_{A_\varepsilon}\big) + \mathbb{E}\big(|X|^p \mathbf{1}_{A_\varepsilon^c}\big) \;\geq\; \mathbb{E}\big(|X|^p \mathbf{1}_{A_\varepsilon}\big) \;\geq\; (\|X\|_{L_\infty} - \varepsilon)^p \, \mathbb{P}(A_\varepsilon) \\ \iff &\|X\|_{L_p} \geq (\|X\|_{L_\infty} - \varepsilon) \, \mathbb{P}(A_\varepsilon)^{1/p} \end{aligned}$$

Because $\mathbb{P}(A_\varepsilon)^{1/p} \to 1$ as $p \to \infty$, we obtain $\liminf_{p\to\infty} \|X\|_{L_p} \geq \|X\|_{L_\infty} - \varepsilon$.

As $\varepsilon > 0$ is arbitrary, this implies $\liminf_{p\to\infty} \|X\|_{L_p} \geq \|X\|_{L_\infty}$.

Combine the upper and lower bounds, we can conclude that:

$$\|X\|_{L_p} \longrightarrow \|X\|_{L_\infty} \qquad \text{as } p \to \infty$$

Question 2. Practice with matrix norms.

a. The $p \to q$ operator norm.

a1. Equivalent definitions.

(i) From $\|x\|_p \leq 1$ to $\|x\|_p = 1$. Because $\{x : \|x\|_p = 1\} \subseteq \{x : \|x\|_p \leq 1\}$, we have:

$$\sup_{\|x\|_p = 1} \|Ax\|_q \;\leq\; \sup_{\|x\|_p \leq 1} \|Ax\|_q \tag{2}$$

For the reverse inequality, we fix any x with $0 < \|x\|_p \leq 1$ and define:

$$u := \frac{x}{\|x\|_p}, \qquad \text{so that} \qquad \|u\|_p = 1.$$

By homogeneity of norms, we have:

$$\|Au\|_q = \left\| A\left(\frac{x}{\|x\|_p}\right) \right\|_q = \frac{1}{\|x\|_p} \|Ax\|_q \;\geq\; \|Ax\|_q \qquad \text{(because } \|x\|_p \leq 1\text{)}.$$

Therefore, every value $\|Ax\|_q$ attainable with $\|x\|_p \le 1$ is bounded above by some value $\|Au\|_q$ with $\|u\|_p = 1$. Therefore:

$$\sup_{\|x\|_p \le 1} \|Ax\|_q \;\le\; \sup_{\|u\|_p = 1} \|Au\|_q \tag{3}$$

From 2 and 3, we obtain:

$$\boxed{\sup_{\|x\|_p \le 1} \|Ax\|_q \;=\; \sup_{\|x\|_p = 1} \|Ax\|_q}$$

(ii) From $\|x\|_p = 1$ to the ratio form. First, if $\|x\|_p = 1$, then we have:

$$\|Ax\|_q = \frac{\|Ax\|_q}{\|x\|_p} \le \sup_{z \ne 0} \frac{\|Az\|_q}{\|z\|_p} \qquad \text{(because because the supremum is an upper bound for all values)}$$

Take the supremum over $\|x\|_p = 1$, we obtain:

$$\sup_{\|x\|_p = 1} \|Ax\|_q \;\le\; \sup_{z \ne 0} \frac{\|Az\|_q}{\|z\|_p} \tag{4}$$

Conversely, for any $z \ne 0$ let $u = z/\|z\|_p$ so that $\|u\|_p = 1$. Then

$$\frac{\|Az\|_q}{\|z\|_p} = \left\| A\left(\frac{z}{\|z\|_p}\right)\right\|_q = \|Au\|_q \le \sup_{\|x\|_p = 1} \|Ax\|_q$$

Take the supremum over $z \ne 0$, we obtain the reverse inequality:

$$\sup_{z \ne 0} \frac{\|Az\|_q}{\|z\|_p} \le \sup_{\|x\|_p = 1} \|Ax\|_q \tag{5}$$

From 4 and 5, we have:

$$\boxed{\sup_{z \ne 0} \frac{\|Az\|_q}{\|z\|_p} \;=\; \sup_{\|x\|_p = 1} \|Ax\|_q}$$

Combine (i) and (ii), we obtain the claimed equivalences.

a2. Submultiplicativity. We have the definition of the induced operator norm:

$$\|A\|_{q \to r} = \sup_{\|u\|_q \le 1} \|Au\|_r$$

By homogeneity, this implies that for all $u \in \mathbb{R}^n$,

$$\|Au\|_r \leq \|A\|_{q \to r} \|u\|_q \tag{6}$$

We justify (6) as follows. If $u = 0$, the inequality holds trivially. If $u \neq 0$, we define:

$$v := \frac{u}{\|u\|_q}$$

so that $\|v\|_q = 1$. By the definition of the induced operator norm, we have:

$$\|Av\|_r \leq \sup_{\|w\|_q \leq 1} \|Aw\|_r = \|A\|_{q \to r}.$$

Using the homogeneity of norms, we obtain

$$\|Au\|_r = \|A(\|u\|_q v)\|_r = \|u\|_q \|Av\|_r \leq \|A\|_{q \to r} \|u\|_q,$$

which proves (6).

Now x satisfy $\|x\|_p \leq 1$. Apply (6) with $u = Bx$, we obtain:

$$\|ABx\|_r \leq \|A\|_{q \to r} \|Bx\|_q$$

Using again the definition of the operator norm for B, we obtain:

$$\|Bx\|_q \leq \|B\|_{p \to q} \|x\|_p$$

Combine the two inequalities, we have:

$$\|ABx\|_r \leq \|A\|_{q \to r} \|B\|_{p \to q} \|x\|_p$$

Because $\|x\|_p \le 1$, the above result implies that:

$$\|ABx\|_r \le \|A\|_{q\to r} \|B\|_{p\to q}$$

Finally, take the supremum over all x with $\|x\|_p \le 1$, we conclude that:

$$\boxed{\|AB\|_{p\to r} = \sup_{\|x\|_p \le 1} \|ABx\|_r \le \|A\|_{q\to r} \|B\|_{p\to q}}$$

a3. [Bonus] Duality form and transpose relationship.

Bilinear representation. We have the dual norm identity for ℓ_q:

$$\|u\|_q = \sup_{\|y\|_{q'}=1} y^\top u, \qquad \frac{1}{q} + \frac{1}{q'} = 1$$

for any $u \in \mathbb{R}^m$. Applying this identity to $u = Ax$, we obtain:

$$\|Ax\|_q = \sup_{\|y\|_{q'}=1} y^\top Ax$$

Therefore,

$$\boxed{\|A\|_{p\to q} = \sup_{\|x\|_p=1} \sup_{\|y\|_{q'}=1} y^\top Ax = \sup_{\|x\|_p=1,\ \|y\|_{q'}=1} |y^\top Ax|}$$

The absolute value can be dropped because replacing y by $-y$ does not change the constraint.

Transpose identity. Using the scalar identity, we have:

$$y^\top Ax = x^\top A^\top y$$

We then obtain:

$$\|A\|_{p\to q} = \sup_{\|x\|_p=1,\ \|y\|_{q'}=1} x^\top A^\top y.$$

Fixing y and applying the dual norm identity for ℓ_p, we have:

$$\sup_{\|x\|_p=1} x^\top A^\top y = \|A^\top y\|_{p'}$$

Therefore:

$$\boxed{\|A\|_{p\to q} = \sup_{\|y\|_{q'}=1} \|A^\top y\|_{p'} = \|A^\top\|_{q'\to p'}}$$

In particular, when $p = q = 2$ (so that $p' = q' = 2$), we conclude that:

$$\boxed{\|A^\top\|_{2\to 2} = \|A\|_{2\to 2}}$$

b. Spectral versus Frobenius norms.

b1. Rank-one and diagonal matrices.

Rank-one matrix. Let $A = uv^\top$ with $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$. For any $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$, we have:

$$\begin{aligned} Ax &= u(v^\top x) \\ \Rightarrow \|Ax\|_2 &= |v^\top x|\, \|u\|_2 \qquad \text{(because } v^T x \text{ is a scalar)} \end{aligned}$$

By Cauchy-Schwarz, we have $|v^\top x| \leq \|v\|_2\|x\|_2 = \|v\|_2$ (because $\|x\|_2 = 1$). Therefore:

$$\|Ax\|_2 \leq \|u\|_2\|v\|_2 \tag{7}$$

Because 7 is true for any $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$, we take the supremum over $\|x\|_2 = 1$ to obtain:

$$\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2 \leq \|u\|_2\|v\|_2 \tag{8}$$

Consider the equality of 7, there exists $x = v/\|v\|_2$ (if $v \neq 0$) (for which $|v^\top x| = \|v\|_2$) that leads to $\|Ax\|_2 = \|u\|_2\|v\|_2$ (the case $v = 0$ is trivial). Moreover, $\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2$ is greater than or equal every value of $\|Ax\|_2$, which implies that:

$$\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2 \geq \|u\|_2\|v\|_2 \tag{9}$$

From 8 and 9, we can conclude that:

$$\|A\| = \|uv^\top\| = \|u\|_2\|v\|_2 \tag{10}$$

For the Frobenius norm, we have:

$$\|uv^\top\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n (u_i v_j)^2 = \left(\sum_{i=1}^m u_i^2\right)\left(\sum_{j=1}^n v_j^2\right) = \|u\|_2^2\|v\|_2^2$$

Therefore:

$$\|uv^\top\|_F = \|u\|_2\|v\|_2 \tag{11}$$

Combine 10 and 11, we can conclude: $\boxed{\|uv^\top\| = \|uv^\top\|_F = \|u\|_2\|v\|_2}$.

Diagonal matrix. Let $D = \operatorname{diag}(d_1, \dots, d_n)$ and $d = (d_1, \dots, d_n)^\top$. For any $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$,

$$\|Dx\|_2^2 = \sum_{i=1}^n d_i^2 x_i^2 \leq \left(\max_{1\leq i\leq n} d_i^2\right) \sum_{i=1}^n x_i^2 = \left(\max_{1\leq i\leq n} |d_i|\right)^2 \tag{12}$$

so $\|Dx\|_2 \leq \max_i |d_i|$ and therefore $\|D\| = \sup_{\|x\|_2=1} \|Dx\|_2 \leq \max_i |d_i|$ (because 12 is true for any $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$). Equality holds by taking $x = e_{i^\star}$ where $i^\star \in \arg\max_i |d_i|$, which implies $\|De_{i^\star}\|_2 = |d_{i^\star}|$. Therefore:

$$\boxed{\|D\| = \max_{1\leq i\leq n} |d_i| = \|d\|_\infty}$$

Finally, because D has nonzero entries only on the diagonal, we have:

$$\|D\|_F^2 = \sum_{i,j} D_{ij}^2 = \sum_{i=1}^n d_i^2 = \|d\|_2^2$$

so $\boxed{\|D\|_F = \|d\|_2}$.

b2. General relationship between spectral and Frobenius norms.

Let $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_r(A) > 0$ be the nonzero singular values of A, with $r = \text{rank}(A)$.

We have known that the spectral norm and the Frobenius norm are written as follows:

$$\|A\| = \|A\|_{2\to 2} = \sigma_1(A) \qquad \|A\|_F^2 = \sum_{i=1}^{r} \sigma_i(A)^2$$

Therefore:

$$\|A\|_F^2 = \sum_{i=1}^{r} \sigma_i^2 \;\geq\; \sigma_1^2 \;=\; \|A\|^2 \quad \Longrightarrow \quad \boxed{\|A\| \leq \|A\|_F}$$

In addition, because $\sigma_i \leq \sigma_1$ for all $i \leq r$, we have:

$$\|A\|_F^2 = \sum_{i=1}^{r} \sigma_i^2 \;\leq\; \sum_{i=1}^{r} \sigma_1^2 = r\sigma_1^2 = r\|A\|^2$$

which implies that: $\boxed{\|A\|_F \leq \sqrt{r}\,\|A\|}$.

Tightness of both inequalities. We now show that each inequality can be achieved for any admissible (m, n, r).

(i) Tightness of the lower bound $\|A\| \leq \|A\|_F$. If A has rank one, then it has exactly one nonzero singular value $\sigma_1 > 0$ and therefore we have:

$$\|A\|_F^2 = \sigma_1^2 = \|A\|^2 \quad \Longleftrightarrow \quad \|A\|_F = \|A\|$$

For example, take $A = uv^\top$ with nonzero $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$; then $\text{rank}(A) = 1$ and by part (b1) $\|A\| = \|A\|_F = \|u\|_2\|v\|_2$.

(ii) Tightness of the upper bound $\|A\|_F \leq \sqrt{r}\,\|A\|$. If we choose A so that its r nonzero singular values are all equal ($\sigma_1 = \cdots = \sigma_r = s > 0$). Then we obtain the equality:

$$\|A\| = s \qquad \|A\|_F = \sqrt{\sum_{i=1}^{r} s^2} = \sqrt{r}\,s = \sqrt{r}\,\|A\|$$

For example, let:

$$A = s \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{m \times n},$$

where I_r is the $r \times r$ identity in the top-left corner. This matrix has rank r and exactly r singular values equal to s, therefore it achieves equality in the upper bound.

b3. Frobenius submultiplicativity.

By property of the Frobenius norm, we have $\|AB\|_F^2 = \operatorname{tr}((AB)^\top(AB)) = \operatorname{tr}(B^\top A^\top AB)$.

Moreover, by the definition of the spectral norm, we have $\|A\| = \sup_{\|x\|_2=1} \|Ax\|_2$,, which implies $\|Ax\|_2 \le \|A\| \, \|x\|_2$ for all x (proved in 6).

As a result, for all x, we obtain:

$$x^\top A^\top Ax = \|Ax\|_2^2 \le \|A\|^2 \|x\|_2^2 = x^\top (\|A\|^2 I)x$$

Therefore, $A^\top A \preceq \|A\|^2 I$.

Multiplying on both sides by $B^\top(\cdot)B$ and using monotonicity of the trace for positive semidefinite matrices, we obtain:

$$\operatorname{tr}(B^\top A^\top AB) \le \|A\|^2 \operatorname{tr}(B^\top B).$$

which implies that $\|AB\|_F^2 \le \|A\|^2 \|B\|_F^2$. Take square roots of both side, we conclude that:

$$\boxed{\|AB\|_F \le \|A\| \, \|B\|_F}$$

Apply the above result to $(AB)^T = B^T A^T$ and using the invariance of both norms under transposition, we have:

$$\boxed{\|AB\|_F = \|(AB)^T\|_F = \|B^T A^T\|_F \le \|B^T\| \, \|A^T\|_F = \|A\|_F \, \|B\|}$$

Finally, by part (b2) we know that $\|A\| \le \|A\|_F$ and $\|B\| \le \|B\|_F$. Combining these inequalities with the previous bounds, we have:

$$\boxed{\|AB\|_F \le \|A\| \, \|B\|_F \le \|A\|_F \, \|B\|_F}$$

Therefore, the Frobenius norm is submultiplicative:

$$\boxed{\|AB\|_F \le \|A\|_F\,\|B\|_F}$$

b4. [Bonus] Orthogonal invariance.

Spectral norm. Using the definition $\|M\| = \sup_{\|x\|_2=1} \|Mx\|_2$ and the fact that $\|Uz\|_2 = \|z\|_2$ for all z (because U is an orthogonal matrix), we have:

$$\|UAV\| = \sup_{\|x\|_2=1} \|UAVx\|_2 = \sup_{\|x\|_2=1} \|AVx\|_2$$

Let $y = Vx$. Since V is orthogonal, $\|y\|_2 = \|x\|_2 = 1$, and as x ranges over the unit sphere, so does y. Hence

$$\|UAV\| = \sup_{\|y\|_2=1} \|Ay\|_2 = \|A\|$$

Frobenius norm. We have $\|M\|_F^2 = \operatorname{tr}(M^\top M)$. Apply cyclic property of trace:

$$\|UAV\|_F^2 = \operatorname{tr}\big((UAV)^\top (UAV)\big) = \operatorname{tr}\big(V^\top A^\top U^\top UAV\big) = \operatorname{tr}\big(V^\top A^\top AV\big) = \operatorname{tr}\big(A^\top AVV^\top\big) = \operatorname{tr}(A^\top A) = \|A\|_F^2$$

Therefore $\|UAV\|_F = \|A\|_F$.

In conclusion:

$$\boxed{\|UAV\| = \|A\| \quad \text{and} \quad \|UAV\|_F = \|A\|_F}.$$

Question 3. Practice with variance and covariance identities

a1. Variance via Pythagorean decomposition.

We have:

$$\|Z-m\|_2^2 = \langle Z-m,\, Z-m\rangle = \|Z\|_2^2 - 2\langle Z, m\rangle + \|m\|_2^2$$
$$\Rightarrow \quad \mathbb{E}\|Z-m\|_2^2 = \mathbb{E}\|Z\|_2^2 - 2\,\mathbb{E}\langle Z, m\rangle + \|m\|_2^2.$$

Because m is deterministic, we can write $\mathbb{E}\langle Z, m\rangle$ as follows:

$$\mathbb{E}\langle Z, m\rangle = \langle \mathbb{E}[Z],\, m\rangle = \langle m, m\rangle = \|m\|_2^2.$$

Therefore, we can conclude: $\boxed{\mathbb{E}\|Z-m\|_2^2 = \mathbb{E}\|Z\|_2^2 - \|m\|_2^2}$.

Proof for general case. For any $a \in \mathbb{R}^d$, we can write $Z - a = (Z - m) + (m - a)$.

Therefore, we can expand $\mathbb{E}\|Z-m\|_2^2$ as follows:

$$\begin{aligned}
\|Z-a\|_2^2 &= \|Z-m\|_2^2 + \|m-a\|_2^2 + 2\langle Z-m,\, m-a\rangle \\
\Rightarrow \quad \mathbb{E}\|Z-a\|_2^2 &= \mathbb{E}\|Z-m\|_2^2 + \|m-a\|_2^2 + 2\langle \mathbb{E}[Z-m],\, m-a\rangle \\
&= \boxed{\mathbb{E}\|Z-m\|_2^2 + \|a-m\|_2^2} \qquad \text{(cross term vanishes because } \mathbb{E}[Z-m]=0\text{)}
\end{aligned}$$

We know that $\|a-m\|_2^2 \geq 0$ with equality if and only if $a = m$. Therefore, $\mathbb{E}\|Z-m\|_2^2 = \min_{a \in \mathbb{R}^d} \mathbb{E}\|Z-a\|_2^2$ and the minimizer is unique $a^\star = \mathbb{E}[Z]$.

a2. Variance via symmetrization.

We rewrite $Z - Z'$ as follows:

$$\begin{aligned}
Z - Z' &= (Z-m) - (Z'-m) \\
\Rightarrow \quad \|Z-Z'\|_2^2 &= \|Z-m\|_2^2 + \|Z'-m\|_2^2 - 2\langle Z-m,\ Z'-m\rangle \\
\Rightarrow \quad \mathbb{E}\|Z-Z'\|_2^2 &= \mathbb{E}\|Z-m\|_2^2 + \mathbb{E}\|Z'-m\|_2^2 - 2\,\mathbb{E}\langle Z-m,\ Z'-m\rangle
\end{aligned}$$

Because $Z' \overset{d}{=} Z$, we have $\mathbb{E}\|Z'-m\|_2^2 = \mathbb{E}\|Z-m\|_2^2$. Therefore, we have:

$$\mathbb{E}\|Z-Z'\|_2^2 = 2\mathbb{E}\|Z-m\|_2^2 - 2\,\mathbb{E}\langle Z-m,\ Z'-m\rangle$$

Expand $\langle Z-m,\ Z'-m\rangle$, we have:

$$\begin{aligned}
\langle Z-m,\ Z'-m\rangle &= \langle Z,\ Z'\rangle - \langle Z,\ m\rangle - \langle m,\ Z'\rangle + \langle m,\ m\rangle \\
&= \langle Z,\ Z'\rangle - \langle Z,\ m\rangle - \langle Z',\ m\rangle + \|m\|_2^2.
\end{aligned}$$

We now take the expectation of $\langle Z - m,\ Z' - m\rangle$ as follows:

$$\begin{aligned}
\mathbb{E}\langle Z - m,\ Z' - m\rangle &= \mathbb{E}\langle Z,\ Z'\rangle - \mathbb{E}\langle Z,\ m\rangle - \mathbb{E}\langle Z',\ m\rangle + \|m\|_2^2 \\
&= \langle \mathbb{E}Z,\ \mathbb{E}Z'\rangle - \langle \mathbb{E}Z,\ m\rangle - \langle \mathbb{E}Z',\ m\rangle + \|m\|_2^2 \qquad (m \text{ is deterministic, } Z \text{ and} Z' \text{ independent}) \\
&= \langle m,\ m\rangle - \langle m,\ m\rangle - \langle m,\ m\rangle + \|m\|_2^2 \\
&= 0
\end{aligned}$$

Therefore:

$$\mathbb{E}\|Z - Z'\|_2^2 = 2\,\mathbb{E}\|Z - m\|_2^2,$$

which implies that:

$$\boxed{\mathbb{E}\|Z - \mathbb{E}Z\|_2^2 = \mathbb{E}\|Z - m\|_2^2 = \frac{1}{2}\,\mathbb{E}\|Z - Z'\|_2^2}.$$

a3. [Bonus] Variance for independent sums.

We expand $\left\|\sum_{j=1}^k Z_j\right\|_2^2$ as follows:

$$\begin{aligned}
\left\|\sum_{j=1}^k Z_j\right\|_2^2 &= \left\langle \sum_{j=1}^k Z_j,\ \sum_{\ell=1}^k Z_\ell \right\rangle = \sum_{j=1}^k \sum_{\ell=1}^k \langle Z_j, Z_\ell\rangle \\
&= \sum_{j=1}^k \|Z_j\|_2^2 + 2 \sum_{1\le j<\ell\le k} \langle Z_j, Z_\ell\rangle \\
\Rightarrow \mathbb{E}\left\|\sum_{j=1}^k Z_j\right\|_2^2 &= \sum_{j=1}^k \mathbb{E}\|Z_j\|_2^2 + 2 \sum_{1\le j<\ell\le k} \mathbb{E}\langle Z_j, Z_\ell\rangle
\end{aligned}$$

For $j \neq \ell$, we write $Z_j = (Z_{j1}, \ldots, Z_{jd})^\top$ and $Z_\ell = (Z_{\ell 1}, \ldots, Z_{\ell d})^\top$. We then have:

$$\langle Z_j, Z_\ell\rangle = \sum_{i=1}^d Z_{ji} Z_{\ell i}$$

Because Z_j and Z_ℓ are independent, we have:

$$\mathbb{E}[Z_{ji} Z_{\ell i}] = \mathbb{E}[Z_{ji}]\,\mathbb{E}[Z_{\ell i}].$$

Moreover, $\mathbb{E}[Z_j] = 0$ implies $\mathbb{E}[Z_{ji}] = 0$ for all i, and similarly $\mathbb{E}[Z_{\ell i}] = 0$. Therefore:

$$\mathbb{E}\langle Z_j, Z_\ell \rangle = 0 \qquad \text{for all } j \neq \ell.$$

Because all cross terms vanish, we can conclude that:

$$\boxed{\mathbb{E}\Big\| \sum_{j=1}^{k} Z_j \Big\|_2^2 = \sum_{j=1}^{k} \mathbb{E}\Big\| Z_j \Big\|_2^2}$$

b. Covariance identities.

b1. Expand $(Z - m)(Z - m)^\top$, we have:

$$(Z - m)(Z - m)^\top = ZZ^\top - Zm^\top - mZ^\top + mm^\top$$

Take the expectations, we obtain:

$$\mathrm{Cov}(Z) = \mathbb{E}[ZZ^\top] - \mathbb{E}[Z]m^\top - m\mathbb{E}[Z]^\top + mm^\top = \mathbb{E}[ZZ^\top] - mm^\top$$

b2. For any $v \in \mathbb{R}^d$, we apply the definition of variance to obtain:

$$\mathrm{Var}(v^\top Z) = \mathbb{E}\Big[\big(v^\top Z - \mathbb{E}[v^\top Z]\big)^2\Big] = \mathbb{E}\Big[\big(v^\top (Z - m)\big)^2\Big].$$

Because $v^\top(Z - m)$ is a scalar, we have:

$$\big(v^\top (Z - m)\big)^2 = v^\top (Z - m)(Z - m)^\top v.$$

Therefore $\mathrm{Var}(v^\top Z) = v^\top \mathbb{E}\big[(Z - m)(Z - m)^\top\big]\, v = v^\top \mathrm{Cov}(Z)\, v$.

b3. By linearity of trace and expectation, we have:

$$\mathrm{tr}(\mathrm{Cov}(Z)) = \mathrm{tr}\Big(\mathbb{E}\Big[(Z - m)(Z - m)^\top\Big]\Big) = \mathbb{E}\Big[\mathrm{tr}\Big((Z - m)(Z - m)^\top\Big)\Big]$$

Using the identity $\mathrm{tr}(uu^\top) = \|u\|_2^2$, we conclude that:

$$\mathrm{tr}(\mathrm{Cov}(Z)) = \mathbb{E}\big[\|Z - m\|_2^2\big] = \mathbb{E}\big[\|Z - \mathbb{E}Z\|_2^2\big]$$

Question 4. Practice with moving between moments and tails

a. From tails to moments

a1. Tail integral identities. We first prove the pointwise identity: for every real number $x \geq 0$, we have:

$$x = \int_0^\infty \mathbf{1}\{x > t\}\, dt, \tag{13}$$

because $\mathbf{1}\{x > t\} = 1$ for $t \in [0, x)$ and 0 for $t \geq x$. Therefore, the integral equals $\int_0^x 1\, dt = x$.

Applying (13) with $x = X(\omega)$, we have

$$X = \int_0^\infty \mathbf{1}\{X > t\}\, dt \quad \text{a.s.}$$

Taking expectations and using Tonelli's theorem (because the integrand is nonnegative), we can interchange expectation and integration as follows:

$$\mathbb{E}[X] = \mathbb{E}\left[\int_0^\infty \mathbf{1}\{X > t\}\, dt\right] = \int_0^\infty \mathbb{E}\big[\mathbf{1}\{X > t\}\big]\, dt = \int_0^\infty \mathbb{P}(X > t)\, dt$$

This shows that $\boxed{\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\, dt}$.

More generally, for any $p > 0$ and any $x \geq 0$, we can show that:

$$x^p = \int_0^\infty p t^{p-1} \mathbf{1}\{x > t\}\, dt$$

because the right-hand side equals $\int_0^x p t^{p-1}\, dt = [t^p]_0^x = x^p$. Substituting $x = X(\omega)$, taking expectations, and applying Tonelli again, we obtain:

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

a2. [Bonus] Moment growth from exponential tails. Assume that there exist constants $c, C > 0$ and $\alpha > 0$ such that for all $t \geq 0$,

$$\mathbb{P}(X > t) \leq Ce^{-ct^\alpha}.$$

Using the identity from a1, for any $p > 0$,

$$\begin{aligned}\mathbb{E}[X^p] &= \int_0^\infty pt^{p-1}\mathbb{P}(X > t)\,dt \leq \int_0^\infty pt^{p-1}\,Ce^{-ct^\alpha}\,dt \\ &= Cp\int_0^\infty t^{p-1}e^{-ct^\alpha}\,dt\end{aligned}$$

Let $u = ct^\alpha$, we have $t = (u/c)^{1/\alpha}$ and $dt = \frac{1}{\alpha}c^{-1/\alpha}u^{1/\alpha-1}\,du$. The integral then become:

$$\begin{aligned}\int_0^\infty t^{p-1}e^{-ct^\alpha}\,dt &= \int_0^\infty \left(\frac{u}{c}\right)^{\frac{p-1}{\alpha}} e^{-u}\left(\frac{1}{\alpha}c^{-1/\alpha}u^{1/\alpha-1}\right)du \\ &= \frac{1}{\alpha}c^{-p/\alpha}\int_0^\infty u^{\frac{p}{\alpha}-1}e^{-u}\,du \\ &= \frac{1}{\alpha}c^{-p/\alpha}\,\Gamma\Big(\frac{p}{\alpha}\Big)\end{aligned}$$

Therefore, we obtain:

$$\mathbb{E}[X^p] \leq \frac{C}{\alpha}\,p\,c^{-p/\alpha}\,\Gamma\Big(\frac{p}{\alpha}\Big). \tag{14}$$

From the above result, we can see that $\mathbb{E}[X^p] < \infty$ for all $p < \infty$, which means that $X \in L_p$ for all $p < \infty$.

To obtain a simple growth bound, we use the given estimate: there exists an absolute constant $C_0 > 0$ such that for all $z \geq 1$:

$$\Gamma(z+1) \leq (C_0 z)^z$$

For $p \geq 1$, set $z = p/\alpha$ (so that $z \geq 1$). Using the identity $\Gamma(z) = \Gamma(z+1)/z$, we obtain

$$\Gamma\Big(\frac{p}{\alpha}\Big) = \frac{\Gamma\big(\frac{p}{\alpha}+1\big)}{p/\alpha} \leq \frac{(C_0\,p/\alpha)^{p/\alpha}}{p/\alpha} = \alpha(C_0/\alpha)^{p/\alpha}p^{p/\alpha-1}$$

Because $p^{p/\alpha-1} \leq p^{p/\alpha}$ for all $p \geq 1$, we have: $\Gamma\big(\frac{p}{\alpha}\big) \leq \alpha(C_0/\alpha)^{p/\alpha}\,p^{p/\alpha}$.

Substitute this bound into (14), we obtain:

$$\mathbb{E}[X^p] \le \frac{C}{\alpha}\, p\, c^{-p/\alpha}\, \alpha (C_0/\alpha)^{p/\alpha}\, p^{p/\alpha} = C\, p^{1+p/\alpha} \left(\frac{C_0}{\alpha c}\right)^{p/\alpha}$$

Therefore: $\|X\|_{L_p} = (\mathbb{E}[X^p])^{1/p} \le C^{1/p}\, p^{1/p}\, p^{1/\alpha} \left(\frac{C_0}{\alpha c}\right)^{1/\alpha}$.

For all $p \ge 1$, we have:

$$\begin{cases} C^{1/p} \le \max\{1, C\} & \text{(because } p \ge 1\text{)} \\ p^{1/p} \le e^{1/e} < 2 & \text{(obtained by taking logarithms and differentiating to find the maximum)} \end{cases}$$

so both factors are bounded by absolute constants independent of p. Therefore, there exists a constant $C' := \max\{1, C\}\, e^{1/e} \left(\frac{C_0}{\alpha c}\right)^{1/\alpha}$ depending only on (c, C, α) such that:

$$\|X\|_{L_p} \le C'\, p^{1/\alpha}, \qquad \forall\, p \ge 1$$

Interpretation. When $\alpha = 2$, the tail bound is of the form e^{-ct^2} (sub-Gaussian behavior) and the result gives $\|X\|_{L_p} \lesssim \sqrt{p}$. When $\alpha = 1$, the tail bound is of the form e^{-ct} (sub-exponential behavior) and the result gives $\|X\|_{L_p} \lesssim p$.

b. Classical tail bounds from second moments

b1. Cantelli's inequality (one-sided). Fix $t > 0$. For any $a > 0$, note that on the event $\{Y - \mathbb{E}Y \ge t\}$ we have $Y - \mathbb{E}Y + a \ge t + a$. Therefore:

$$\{Y - \mathbb{E}Y \ge t\} \subseteq \{(Y - \mathbb{E}Y + a)^2 \ge (t+a)^2\}$$

Apply Markov's inequality to the nonnegative random variable $(Y - \mu + a)^2$, we obtain:

$$\mathbb{P}(Y - \mathbb{E}Y \ge t) \le \mathbb{P}\big((Y - \mathbb{E}Y + a)^2 \ge (t+a)^2\big) \le \frac{\mathbb{E}\big[(Y - \mathbb{E}Y + a)^2\big]}{(t+a)^2} \tag{15}$$

Expanding $\mathbb{E}\big[(Y - \mathbb{E}Y + a)^2\big]$, we have:

$$\mathbb{E}\big[(Y - \mathbb{E}Y + a)^2\big] = \mathbb{E}\big[(Y - \mathbb{E}Y)^2\big] + 2a\, \mathbb{E}[Y - \mathbb{E}Y] + a^2 = \operatorname{Var} Y + a^2$$

Substituting this result into (15), we have:

$$\mathbb{P}(Y - \mu \geq t) \leq \frac{\sigma^2 + a^2}{(t+a)^2}, \qquad \forall a > 0.$$

In order to obtain a tighter upper bound for $\mathbb{P}(Y - \mu \geq t)$, we minimize $f(a) := (\operatorname{Var} Y + a^2)/(t+a)^2$ over a. We compute the derivative of $f(a)$ as follows:

$$f'(a) = \frac{2(at - \operatorname{Var} Y)}{(t+a)^3}$$

so the minimum is attained at $a^* = \frac{\operatorname{Var} Y}{t}$. Plugging a^* back, we have:

$$\boxed{\mathbb{P}(Y - \mu \geq t) \leq \frac{\operatorname{Var} Y + (\frac{\operatorname{Var} Y}{t})^2}{(t + \frac{\operatorname{Var} Y}{t})^2} = \frac{\operatorname{Var} Y}{\operatorname{Var} Y + t^2}}$$

b2. Comparison with Chebyshev and a two-sided bound. A naive one-sided bound from Chebyshev is obtained via the inclusion $\{Y - \mathbb{E}Y \geq t\} \subseteq \{|Y - \mathbb{E}Y| \geq t\}$:

$$\mathbb{P}(Y - \mathbb{E}Y \geq t) \leq \mathbb{P}(|Y - \mathbb{E}Y| \geq t) \leq \frac{\operatorname{Var} Y}{t^2}$$

Because $\frac{\operatorname{Var} Y}{\operatorname{Var} Y + t^2} \leq \frac{\operatorname{Var} Y}{t^2} (t > 0)$, we can see that Cantelli is always at least as strong as the naive Chebyshev bound.

For the two-sided bound, we can start as follows:

$$\{|Y - \mathbb{E}Y| \geq t\} = \{Y - \mathbb{E}Y \geq t\} \cup \{\mathbb{E}Y - Y \geq t\}$$

By the union bound, we have:

$$\mathbb{P}(|Y - \mathbb{E}Y| \geq t) \leq \mathbb{P}(Y - \mathbb{E}Y \geq t) + \mathbb{P}(\mathbb{E}Y - Y \geq t)$$

Apply Cantelli to Y, we can bound the first term by $\operatorname{Var} Y/(\operatorname{Var} Y + t^2)$. Apply Cantelli to $-Y$ (we already have $\mathbb{E}[-Y] = -\mathbb{E}Y$ and $\operatorname{Var}(-Y) = \operatorname{Var} Y$) to get

$$\mathbb{P}(\mathbb{E}Y - Y \geq t) = \mathbb{P}(-Y - \mathbb{E}[-Y] \geq t) \leq \frac{\operatorname{Var}(-Y)}{\operatorname{Var}(-Y) + t^2} = \frac{\operatorname{Var}(Y)}{\operatorname{Var}(Y) + t^2}$$

Therefore, we conclude that:

$$\boxed{\mathbb{P}(|Y - \mathbb{E}Y| \geq t) \leq \frac{2\operatorname{Var}(Y)}{\operatorname{Var}(Y) + t^2}}$$

Compare the two-sided Cantelli bound with the standard two-sided Chebyshev inequality $\mathbb{P}(|Y - \mathbb{E}Y| \geq t) \leq \frac{\operatorname{Var} Y}{t^2}$, we can conclude that the Chebyshev bound is tighter for large t.

b3. [Bonus] Paley–Zygmund inequality. We define the event $E := \{X \geq \theta\, \mathbb{E}X\}$. By decomposing this event, we can rewrite the expectation $\mathbb{E}X$ as follows:

$$\mathbb{E}[X] = \mathbb{E}[X\mathbf{1}_{E^c}] + \mathbb{E}[X\mathbf{1}_E]$$

For the complementary event $E^c = \{X < \theta\mathbb{E}X\}$ we can infer that $X \leq \theta\mathbb{E}X$. Therefore:

$$\mathbb{E}[X\mathbf{1}_{E^c}] \leq \theta\mathbb{E}X, \qquad \text{so} \qquad \mathbb{E}[X\mathbf{1}_E] \geq (1-\theta)\mathbb{E}X \tag{16}$$

By Cauchy-Schwarz, we obtain:

$$\mathbb{E}[X\mathbf{1}_E] \leq \left(\mathbb{E}[X^2]\right)^{1/2}\left(\mathbb{E}[\mathbf{1}_E^2]\right)^{1/2} = \left(\mathbb{E}[X^2]\right)^{1/2}\mathbb{P}(E)^{1/2} \tag{17}$$

From 16 and 17, we have:

$$(1-\theta)\mathbb{E}X \leq \mathbb{E}[X\mathbf{1}_E] \leq \left(\mathbb{E}[X^2]\right)^{1/2}\mathbb{P}(E)^{1/2}$$

which implies that:

$$(1-\theta)^2(\mathbb{E}X)^2 \leq \mathbb{E}[X^2]\,\mathbb{P}(E)$$

Finally, we can conclude that:

$$\mathbb{P}(X \geq \theta\,\mathbb{E}X) \geq (1-\theta)^2 \frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}$$

Using $\mathbb{E}[X^2] = \mathrm{Var}(X) + (\mathbb{E}X)^2$, the above result can be rewritten as follows:

$$\mathbb{P}(X \geq \theta\,\mathbb{E}X) \geq (1-\theta)^2 \frac{(\mathbb{E}X)^2}{\mathrm{Var}(X) + (\mathbb{E}X)^2} = \frac{(1-\theta)^2}{1 + \mathrm{Var}(X)/(\mathbb{E}X)^2}$$

In particular, if $\mathrm{Var}(X) \leq c(\mathbb{E}X)^2$ for some $c > 0$, then

$$\boxed{\mathbb{P}(X \geq \theta\,\mathbb{E}X) \geq \frac{(1-\theta)^2}{1+c}}$$

Question 5. Mean versus median.

a. Mean-median closeness (using variance). For this problem, we consider two cases.

Case (I): $M_Z \geq \mathbb{E}Z$. From the definition of median, we have:

$$\mathbb{P}(Z - \mathbb{E}Z \geq M_Z - \mathbb{E}Z) = \mathbb{P}(Z \geq M_Z) \geq \tfrac{1}{2}$$

On the other hand, apply one-sided Cantelli inequality, we have:

$$\mathbb{P}(Z - \mathbb{E}Z \geq M_Z - \mathbb{E}Z) \leq \frac{\mathrm{Var}(Z - \mathbb{E}Z)}{\mathrm{Var}(Z - \mathbb{E}Z) + (M_Z - \mathbb{E}Z)^2} = \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + (M_Z - \mathbb{E}Z)^2}$$

Combining the upper bound and the lower bound of $\mathbb{P}(Z - \mathbb{E}Z \geq M_Z - \mathbb{E}Z)$, we get:

$$\frac{1}{2} \leq \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + (M_Z - \mathbb{E}Z)^2}$$

This implies that $(M_Z - \mathbb{E}Z)^2 \leq \mathrm{Var}(Z)$. Because $M_Z - \mathbb{E}Z \geq 0$, we obtain:

$$M_Z - \mathbb{E}Z \leq \sqrt{\mathrm{Var}(Z)}$$

Case (II): $M_Z \leq \mathbb{E}Z$. Apply the definition of median, we have:

$$\mathbb{P}(\mathbb{E}Z - Z \geq \mathbb{E}Z - M_Z) = \mathbb{P}(-Z \geq -M_Z) = \mathbb{P}(Z \leq M_Z) \geq \frac{1}{2}$$

Moreover, apply the one-sided Cantelli inequality, we get:

$$\mathbb{P}(\mathbb{E}Z - Z \geq \mathbb{E}Z - M_Z) \leq \frac{\mathrm{Var}(\mathbb{E}Z - Z)}{\mathrm{Var}(\mathbb{E}Z - Z) + (\mathbb{E}Z - M_Z)^2} = \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + (\mathbb{E}Z - M_Z)^2}$$

This implies that $\frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z)+(\mathbb{E}Z-M_Z)^2} \geq \frac{1}{2}$, which means $(\mathbb{E}Z - M_Z)^2 \leq \mathrm{Var}(Z)$. Because $\mathbb{E}Z - M_Z \geq 0$, we have:

$$M_Z - \mathbb{E}Z \leq \sqrt{\mathrm{Var}(Z)}.$$

From Case (I) and Case (II), we have:

$$\boxed{|M_Z - \mathbb{E}Z| \leq \sqrt{\mathrm{Var}(Z)}}$$

b. [Bonus] Mean–median closeness using a variance proxy.

Step 1: Prove that $|M_X - \mathbb{E}X| \leq \mathbb{E}|X - M_X|$ (use Jensen inequality).

Because the function $x \mapsto |x|$ is convex, Jensen's inequality gives:

$$|M_X - \mathbb{E}X| = |\mathbb{E}(X - M_X)| \leq \mathbb{E}|X - M_X|$$

Step 2: Prove that $\mathbb{E}|X - M_X| \leq \min\left\{\sqrt{ab},\, a\sqrt{\frac{\pi b}{2}}\right\}$ (use tail integral identity).

Let $Y := |X - M_X| \geq 0$. Apply tail integral identity, we can write $\mathbb{E}Y$ as follows:

$$\mathbb{E}Y = \int_0^\infty \mathbb{P}(Y > t)\, dt$$

Using the assumed tail bound and let $u = \frac{t}{\sqrt{b}}, v = \sqrt{2}u$ (therefore $dt = \sqrt{b}du, du = \frac{1}{\sqrt{2}}dv$), we have:

$$\mathbb{E}|X - M_X| \leq \int_0^\infty a e^{-t^2/b}\, dt = a\sqrt{b} \int_0^\infty e^{-u^2}\, du = \frac{a\sqrt{b}}{\sqrt{2}} \int_0^\infty e^{-v^2/2}\, dv$$

$$= \frac{a\sqrt{b}}{\sqrt{2}} \cdot \frac{1}{2} \int_{-\infty}^{\infty} e^{-v^2/2}\, dv = \frac{a\sqrt{b}}{2\sqrt{2}}\, (\sqrt{2\pi}) \underbrace{\left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-v^2/2}\, dv \right)}_{=1} = a\sqrt{\frac{\pi b}{2}} \tag{18}$$

Using again the tail representation for the second moment, we obtain:

$$\mathbb{E}Y^2 = \int_0^{\infty} 2t\, \mathbb{P}(Y > t)\, dt \leq \int_0^{\infty} 2t\, ae^{-t^2/b}\, dt$$

We have:

$$\int_0^{\infty} 2t\, ae^{-t^2/b}\, dt = -ab \int_0^{\infty} \frac{-2t}{b}\, e^{-t^2/b}\, dt = -ab \left(e^{-t^2/b} \Big|_0^{\infty} \right) = ab$$

Therefore $\mathbb{E}Y^2 \leq ab$.

By Cauchy-Schwarz, we have:

$$\mathbb{E}|X - M_X| = \mathbb{E}Y = \mathbb{E}[Y.1] \leq \sqrt{\mathbb{E}[Y^2]}.\sqrt{\mathbb{E}[1^2]} = \sqrt{\mathbb{E}[Y^2]} \leq \sqrt{ab} \tag{19}$$

From 18 and 19, we can conclude that:

$$\mathbb{E}|X - M_X| \leq \min \left\{ \sqrt{ab},\, a\sqrt{\frac{\pi b}{2}} \right\}$$

From the result of step 1 and step 2, we have:

$$\boxed{|M_X - \mathbb{E}X| \leq \mathbb{E}|X - M_X| \leq \min \left\{ \sqrt{ab},\, a\sqrt{\frac{\pi b}{2}} \right\}}.$$