1 ℓ_p and L_p norms

(a) **ℓ_p norms on $\mathbb{R}^n$.** **ℓ_p norms on $\mathbb{R}^n$.** Let $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. For $1 \leqslant p < \infty$, define

$$\|x\|_{\ell_p} := \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, \qquad \|x\|_{\ell_\infty} := \max_{1\leqslant i\leqslant n} |x_i|.$$

Also, define the unit ball $B^n_{\ell_p} := \{x \in \mathbb{R}^n : \|x\|_{\ell_p} \leqslant 1\}$.

(a1) (Comparison and monotonicity) Prove that if $1 \leqslant p \leqslant q \leqslant \infty$, then

$$\|x\|_{\ell_q} \leqslant \|x\|_{\ell_p} \leqslant n^{\frac{1}{p}-\frac{1}{q}} \|x\|_{\ell_q}.$$

Deduce that $\|x\|_{\ell_p}$ is (weakly) decreasing in p and that the unit balls are nested:

$$B^n_{\ell_p} \subseteq B^n_{\ell_q} \text{ for } p \leqslant q.$$

Solution:
First we prove $\|x\|_{l_q} \leqslant \|x\|_{l_p}$:

i. Case 1: $q < \infty$.
If $x \neq 0$ normalize x, and let $y = x/\|x\|_{\ell_p}$. Then, $\|y\|_{\ell_p} = 1$, i.e., $\sum_{i=1}^n |y_i|^p = 1$. Since $\sum_i |y_i|^p = 1$, each $|y_i|^p \leqslant 1 \implies |y_i| \leqslant 1$. So, for $q \geq p$, $|y_i|^q \leqslant |y_i|^p$. Summing, we get,

$$\sum_{i=1}^n |y_i|^q \leqslant \sum_{i=1}^n |y_i|^p = 1,$$

which implies that $\|y\|_{\ell_q} \leqslant 1$. Hence,

$$\|x\|_{\ell_q} = \|x\|_{\ell_p} \|y\|_{\ell_q} \leqslant \|x\|_{\ell_p}. \quad \square$$

ii. Case 2: $q = \infty$.
For each i, $|x_i|^p \leqslant \sum_j |x_j|^p = \|x\|^p_{\ell_p}$, so $|x_i| \leqslant \|x\|_{\ell_p}$. Taking the maximum,

$$\max\{|x_i|\} \leqslant \|x\|_{\ell_p} \implies \|x\|_{\ell_\infty} \leqslant \|x\|_{\ell_p}.$$

Hence, in all cases,

$$\boxed{\|x\|_{\ell_q} \leqslant \|x\|_{\ell_p}.}$$

Now, we prove $\|x\|_{\ell_p} \leqslant n^{\frac{1}{p}-\frac{1}{q}} \|x\|_{\ell_q}$:

i. Case 1: $q < \infty$.
For $x = 0$, the inequality holds trivially. For $x \neq 0$, normalize x and let $y = x/\|x\|_{\ell_q}$ such that $\|y\|_{\ell_q} = 1$ i.e., $\sum_i |y_i|^q = 1$. Let $r = \dfrac{q}{p} \geqslant 1$ and $s = \dfrac{r}{r-1} = \dfrac{q}{q-p}$ so that $\dfrac{1}{r} + \dfrac{1}{s} = 1$. Now, we apply Holder's inequality to $\sum_i |y_i|^p \cdot 1$, which gives

$$\sum_{i=1}^n |y_i|^p \leqslant \left(\sum_{i=1}^n (|y_i|^p)^r\right)^{1/r} \left(\sum_{i=1}^n 1^s\right)^{1/s} = \left(\sum_{i=1}^n |y_i|^{pr}\right)^{1/r} n^{1/s}.$$

Here, $pr = q$ and $\sum_i |y_i|^q = 1$, so

$$\sum_{i=1}^{n} |y_i|^p \leqslant n^{1/s} \implies \|y\|_{l_p} = \left(\sum_{i=1}^{n} |y_i|^p\right)^{1/p} \leqslant n^{1/ps}.$$

Here, $\dfrac{1}{ps} = \dfrac{q-p}{pq} = \dfrac{1}{p} - \dfrac{1}{q}$. Hence, going back to x from y,

$$\|y\|_{\ell_p} = \left\| \frac{x}{\|x\|_{\ell_q}} \right\|_{\ell_p} \leqslant n^{\frac{1}{p}-\frac{1}{q}} \implies \|x\|_{\ell_p} \leqslant n^{\frac{1}{p}-\frac{1}{q}} \|x\|_{\ell_q}. \quad \square$$

ii. Case 2: $q = \infty$.

Here, we need to prove $\|x\|_{\ell_p} \leqslant n^{\frac{1}{p}} \|x\|_{\ell_q}$ as $1/q = 0$ for $q = \infty$. Then,

$$\|x\|_{\ell_p} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \leqslant \left(\sum_{i=1}^{n} (\max |x_i|)^p\right)^{1/p} = n^{1/p} \max |x_i| = n^{1/p} \|x\|_{\ell_\infty}.$$

Hence, in all cases,

$$\boxed{\|x\|_{\ell_p} \leqslant n^{\frac{1}{p}-\frac{1}{q}} \|x\|_{\ell_q}.}$$

Finally, combining all the results, if $1 \leqslant p \leqslant q \leqslant \infty$, then

$$\boxed{\|x\|_{\ell_q} \leqslant \|x\|_{\ell_p} \leqslant n^{\frac{1}{p}-\frac{1}{q}} \|x\|_{\ell_q}.}$$

As we proved $\|x\|_{\ell_q} \leqslant \|x\|_{\ell_p}$ for $q \geqslant p$, $\|x\|_{\ell_p}$ is non-increasing in p, i.e., it is (weakly) decreasing in p. And, for $q \geqslant p$, if $\|x\|_{\ell_p} \leq 1$, then using above result $\|x\|_{\ell_q} \leqslant \|x\|_{\ell_p} \leqslant 1$. Hence, if $x \in B^n_{l_p}$, then $x \in B^n_{l_q}$, i.e., $B^n_{l_p} \subseteq B^n_{l_q}$ for $p \leqslant q$. $\square$

(a2) (When are the bounds tight?) Give examples of nonzero vectors $x, y \in \mathbb{R}^n$ such that

$$\|x\|_{\ell_p} = \|x\|_{\ell_q} \qquad \text{and} \qquad \|y\|_{\ell_p} = n^{\frac{1}{p}-\frac{1}{q}} \|y\|_{\ell_q}.$$

Solution:

Examples for tighter bounds:

i. For $x = (1, 0, \cdots, 0)^\top \in \mathbb{R}^n$, $\|x\|_{\ell_p} = \|x\|_{\ell_p} = 1$.

ii. For $y = (1, 1, \cdots, 1)^\top \in \mathbb{R}^n$, $\|y\|_{\ell_p} = n^{1/p}$ and $\|y\|_{\ell_q} = n^{1/q}$.
Dividing $\|y\|_{\ell_p}$ by $\|y\|_{\ell_q}$, we get $\|y\|_{\ell_p} = n^{\frac{1}{p}-\frac{1}{q}} \|y\|_{\ell_q}$.

(a3) [Bonus] (The ℓ_∞ limit) Show that $\|x\|_{\ell_p} \to \|x\|_{\ell_\infty}$ as $p \to \infty$. Also, show the quantitative bound

$$\|x\|_{\ell_\infty} \leqslant \|x\|_{\ell_p} \leqslant n^{1/p} \|x\|_{\ell_\infty}.$$

Conclude that if $p \geqslant \log n$ then $\|x\|_{\ell_p} \leqslant e \|x\|_{\ell_\infty}$. (Here, log denotes the natural logarithm.)

[Bonus] Solution:

Substituting $q = \infty$ in the inequality we prove in (a1), we get,

$$\|x\|_{\ell_\infty} \leqslant \|x\|_{\ell_p} \leqslant n^{1/p} \|x\|_{\ell_\infty}.$$

We know $n^{1/p} \to 1$ as $p \to \infty$. So, taking the limit of $p \to \infty$,

$$\lim_{p\to\infty} \|x\|_{\ell_\infty} \leqslant \lim_{p\to\infty} \|x\|_{\ell_p} \leqslant \lim_{p\to\infty} n^{1/p} \|x\|_{\ell_\infty} \implies \|x\|_{\ell_\infty} \leqslant \lim_{p\to\infty} \|x\|_{\ell_p} \leqslant \|x\|_{\ell_\infty}.$$

Hence, using the sandwich theorem,

$$\boxed{\lim_{p\to\infty} \|x\|_{\ell_p} = \|x\|_{\ell_\infty} \text{ i.e., } \|x\|_{\ell_p} \to \|x\|_{\ell_\infty} \text{ as } p \to \infty.}$$

Now, if $p \geq \log n$, then $\log n^{1/p} \leqslant 1$, which implies $n^{1/p} \leqslant e$. Hence, $\|x\|_{\ell_p} \leqslant e\|x\|_{\ell_\infty}$. □

(b) L_p **norms of random variables.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a real-valued random variable. For $1 \leqslant p < \infty$, define

$$\|X\|_{L_p} := (\mathbb{E}|X|^p)^{1/p}, \qquad \|X\|_{L_\infty} := \operatorname{ess\,sup}|X|.$$

Write $L_p := \{X : \|X\|_{L_p} < \infty\}$.

(b1) (Monotonicity and inclusions) Prove that, for $1 \leqslant p \leqslant q \leqslant \infty$,

$$\|X\|_{L_p} \leqslant \|X\|_{L_q}.$$

Deduce the set inclusion $L_q \subseteq L_p$ for $q \geqslant p$.

Solution:

i. Case 1: $q = \infty$

By the definition of the essential supremum, $|X| \leqslant \|X\|_{L_\infty}$ almost everywhere. So, $(|X|)^p \leqslant (\|X\|_{L_\infty})^p$ and taking expectation gives,

$$\mathbb{E}[(|X|)^p] \leqslant \mathbb{E}[(\|X\|_{L_\infty})^p] = \|X\|^p_{L_\infty}.$$

So,

$$\|X\|_{L_p} \leqslant \|X\|_{L_\infty}.$$

ii. Case 2: $q < \infty$

Consider $\phi(t) = t^{q/p}$ on $[0, \infty)$. Since $q \geqslant p$, ϕ is convex. Then,

$$\begin{aligned}
\|X\|_{L_p} = (\mathbb{E}[|X|^p])^{1/p} &= \left[(\mathbb{E}[|X|^p])^{q/p}\right]^{1/q} \\
&\leqslant \left(\mathbb{E}\left[(|X|^p)^{q/p}\right]\right)^{1/q} \quad \text{[Using Jensen's inequality on } \phi(\cdot)] \\
&= (\mathbb{E}[|X|^q])^{1/q} \\
&= \|X\|_{L_q}
\end{aligned}$$

So, for $1 \leqslant p \leqslant q < \infty$,

$$\|X\|_{L_p} \leqslant \|X\|_{L_q}.$$

Hence, combining all cases, for $1 \leqslant p \leqslant q \leqslant \infty$,

$$\boxed{\|X\|_{L_p} \leqslant \|X\|_{L_q}.}$$

Now, let $X \in L_q$, then $\|X\|_{L_q} < \infty$. By the inequality above,

$$\|X\|_{L_p} \leqslant \|X\|_{L_q} < \infty,$$

so $X \in L_p$. Therefore,

$$\boxed{L_q \subseteq L_p \text{ for } q \geqslant p.}$$

(b2) (No dimension-free equivalence) Explain why there cannot exist a universal constant C such that $\|X\|_{L_q} \leqslant C\|X\|_{L_p}$ holds for all random variables on general probability spaces whenever $p < q$. Give an explicit example of X such that $X \in L_p$ but $X \notin L_q$.

Solution:

There cannot exist a universal constant C such that

$$(\mathbb{E}[|X|^q])^{1/q} \leqslant C\,(\mathbb{E}[|X|^p])^{1/p}$$

for all random variables on general probability spaces whenever $p < q$ because the higher order moment for the random variable might not exist. If the higher order moment (q-th moment) of random variable X doesn't exist then $\mathbb{E}[|X|^q] = \infty$, which cannot be bounded above by a finite p-th moment ($\mathbb{E}[|X|^p] < \infty$) fo X. For example, let $p \in \mathbb{N}$. Take $\Omega = (0,1)$ with Lebesgue measure, and define

$$X(\omega) = \sum_{n=1}^{\infty} a_n \mathbf{1}_{A_n}\{\omega\},$$

with disjoint sets A_n of probability $\mathbb{P}\{A_n\} = 2^{-n}$ and amplitudes $a_n = 2^{n/p} n^{-2/p}$. Then,

$$\mathbb{E}[|X|^p] = \sum_{n=1}^{\infty} a_n^p \mathbb{P}\{A_n\} = \sum_{n=1}^{\infty} (2^n n^{-2})2^{-n} = \sum_{n=1}^{\infty} n^{-2} < \infty.$$

But for $q > p$,

$$\mathbb{E}[|X|^q] = \sum_{n=1}^{\infty} a_n^q \mathbb{P}\{A_n\} = \sum_{n=1}^{\infty} (2^{nq/p} n^{-2q/p})2^{-n} = \sum_{n=1}^{\infty} 2^{n(q/p-1)} n^{-2q/p} = \infty$$

because $q/p - 1 > 0$ makes the $2^{n(q/p-1)}$ explode. So $X \in L_p$ but $X \notin L_q$.

(b3) [Bonus] (The L_∞ limit) Assume $\|X\|_{L_\infty} < \infty$. Show that $\|X\|_{L_p} \to \|X\|_{L_\infty}$ as $p \to \infty$.

[Bonus] Solution:

By the definition of the essential supremum, $|X| \le \|X\|_\infty$ almost everywhere. So, using the result from part (b1), we get

$$\limsup_{p\to\infty} \|X\|_{L_p} \le \|X\|_{L_\infty}.$$

Now, let $M = \|X\|_{L_\infty}$. For any $\varepsilon > 0$, let $A_\varepsilon = \{\omega : |X(\omega)| > M - \varepsilon\}$. By the definition of the essential supremum, $\mathbb{P}\{A_\varepsilon\} > 0$. On A_ε, $|X|^p > (M-\varepsilon)^p$, so using the indicator function $\mathbf{1}_{A_\varepsilon}$:

$$\mathbb{E}[|X|^p] \ge \mathbb{E}[|X|^p \mathbf{1}_{A_\varepsilon}] \ge (M-\varepsilon)^p \mathbb{P}\{A_\varepsilon\}.$$

Taking the p-th root:

$$\|X\|_{L_p} \ge (M-\varepsilon)\mathbb{P}\{A_\varepsilon\}^{1/p}.$$

As $p \to \infty$, $\mathbb{P}\{A_\varepsilon\}^{1/p} \to 1$. Thus,

$$\liminf_{p\to\infty} \|X\|_{L_p} \ge M - \varepsilon.$$

Since ε is arbitrary and holds for every $\varepsilon > 0$, $\liminf_{p\to\infty} \|X\|_{L_p} \ge M = \|X\|_{L_\infty}$.

Hence, using the sandwich theorem,

$$\boxed{\lim_{p\to\infty} \|X\|_{L_p} = \|X\|_{L_\infty}.}$$

2 Matrix norms

Throughout, let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$ be conformable matrices, and let $1 \leqslant p, q, r \leqslant \infty$. Recall $\|x\|_p$ denotes the ℓ_p norm on Euclidean spaces. For any $s \in [1,\infty]$, let s' denote the conjugate exponent: $\frac{1}{s} + \frac{1}{s'} = 1$ (with the conventions $1/\infty = 0$ and $1/0 = \infty$).

(a) **The $p \to q$ operator norm.** Define the induced (operator) norm

$$\|A\|_{p\to q} := \sup_{\|x\|_p \leqslant 1} \|Ax\|_q.$$

(a1) (Equivalent definitions) Show that

$$\|A\|_{p\to q} = \sup_{\|x\|_p = 1} \|Ax\|_q = \sup_{x\neq 0} \frac{\|Ax\|_q}{\|x\|_p}.$$

Solution:
First we prove $\sup_{\|x\|_p\leqslant 1} \|Ax\|_q = \sup_{\|x\|_p=1} \|Ax\|_q$. Clearly $\{\|x\|_p = 1\} \subseteq \{\|x\|_p \leqslant 1\}$, so

$$\sup_{\|x\|_p=1} \|Ax\|_q \leqslant \sup_{\|x\|_p\leqslant 1} \|Ax\|_q.$$

Now, for the other direction of the inequality, take any x with $\|x\|_p < 1$. If $x = 0$, $\|Ax\|_q = 0$, so it doesn't affect the inequality. For $x \neq 0$, define $y = x/\|x\|_p$ so $\|y\|_p = 1$. Then, by homogeneity,

$$\|Ax\|_q = \|x\|_p \|Ay\|_q \leqslant \|Ay\|_q.$$

Hence, every value attained inside the ball is less than some value on the sphere, i.e,

$$\sup_{\|x\|_p\leqslant 1} \|Ax\|_q \leqslant \sup_{\|x\|_p=1} \|Ax\|_q.$$

Combining both directions, we get,

$$\boxed{\sup_{\|x\|_p\leqslant 1} \|Ax\|_q = \sup_{\|x\|_p=1} \|Ax\|_q.}$$

Now, we prove $\sup_{\|x\|_p=1} \|Ax\|_q = \sup_{x\neq 0} \frac{\|Ax\|_q}{\|x\|_p}$. For any $x \neq 0$, as before, define $y = x/\|x\|_p$ so $\|y\|_p = 1$, because the equality holds trivially for $x = 0$. Then,

$$\frac{\|Ax\|_q}{\|x\|_p} = \frac{\|x\|_p\|Ay\|_q}{\|x\|_p} = \|Ay\|_q$$

Here, as $x \neq 0$ ranges over $\mathbb{R}^n$, the ratio $\|Ax\|_q/\|x\|_p$ takes exactly the same set of values as $\|Ay\|_q$ with $\|y\|_p = 1$. So, their supremums are equal, i.e.,

$$\boxed{\sup_{x\neq 0} \frac{\|Ax\|_q}{\|x\|_p} = \sup_{\|y\|_p=1} \|Ay\|_q.}$$

Putting both results together,

$$\boxed{\|A\|_{p\to q} = \sup_{\|x\|_p\leqslant 1}\|Ax\|_q = \sup_{\|x\|_p=1}\|Ax\|_q = \sup_{x\neq 0}\frac{\|Ax\|_q}{\|x\|_p}.}$$

(a2) (Submultiplicativity) Prove the composition bound

$$\|AB\|_{p\to r} \leqslant \|A\|_{q\to r}\|B\|_{p\to q}.$$

Solution:
We know,

$$\|A\|_{q\to r} = \sup_{u\neq 0}\frac{\|Au\|_r}{\|u\|_q} \implies \|Au\|_r \leqslant \|A\|_{q\to r}\|u\|_q \quad \forall u.$$

Substituting $u = Bx$, $\|ABx\|_r \leqslant \|A\|_{q\to r}\|Bx\|_q$. Taking the supremum over $\|x\|_p \leqslant 1$,

$$\begin{aligned}\sup_{\|x\|_p\leqslant 1}\|ABx\|_r \leqslant \sup_{\|x\|_p\leqslant 1}\|A\|_{q\to r}\|Bx\|_q &\iff \|AB\|_{p\to r} \leqslant \|A\|_{q\to r}\sup_{\|x\|_p\leqslant 1}\|Bx\|_q \\ &\iff \|AB\|_{p\to r} \leqslant \|A\|_{q\to r}\|B\|_{p\to q}. \quad \square\end{aligned}$$

(a3) [Bonus] (Duality form and transpose relationship) Show the bilinear representation

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

Deduce the transpose identity

$$\|A^\top\|_{q'\to p'} = \|A\|_{p\to q}.$$

In particular, conclude that for the $2\to 2$ operator norm (spectral norm),

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

[Bonus] Solution:
First of all, for a given x, we know that,

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

because if $y^\top Ax < 0$, we can we can flip $y \to -y$ without changing $\|y\|_{q'}$. Now, using the dual norm identity of l_q on $u = Ax$:

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

Taking supremum over $\|x\|_p = 1$ gives,

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

Since the feasible sets are products of the compact sets, we can write it as a joint supremum, i.e.,

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

Hence, combining with the result for the absolute value, we get,

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

Here $y^\top Ax$ is a scalar, so rewrite as $y^\top Ax = x^\top A^\top y$. So, the bilinear form becomes:

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

Using the dual norm identity of $l_{p'}$ on $u = A^\top y$:

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

and taking the supremum over $\|y\|_{q'} = 1$ gives

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

where we write the joint supremum as the sets are compact. Hence,

$$\boxed{\|A^\top\|_{q'\to p'} = \|A\|_{p\to q}.}$$

For $p = q = 2$, $p' = q' = 2$, so for the $2 \to 2$ operator norm (spectral norm):

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

i.e., the singular values of a matrix A is invariant of the transpose. □

(b) **Spectral versus Frobenius norms.** In this part, write $\|A\|$ for the spectral norm $\|A\|_{2\to 2}$. Recall the Frobenius norm $\|A\|_F = \big(\sum_{i,j} A_{ij}^2\big)^{1/2}$.

(b1) (Rank-one and diagonal matrices) Let $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$. Show that

$$\|uv^\top\| = \|uv^\top\|_F = \|u\|_2\|v\|_2.$$

Let $D \in \mathbb{R}^{n\times n}$ be a diagonal matrix with entries $(d_1, \dots, d_n)$. Write $d = (d_1, \dots, d_n)^\top \in \mathbb{R}^n$. Show that

$$\|D\| = \max_{1\leqslant i\leqslant n} |d_i| = \|d\|_\infty \quad \text{and} \quad \|D\|_F = \|d\|_2.$$

Solution:
For the matrix $uv^\top$, $(uv^\top)_i j = u_i v_j$. So,

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

Now, for any $x \in \mathbb{R}^n$, $v^\top x$ is a scalar, so

$$\|uv^\top x\|_2 = |v^\top x|\, \|u\|_2 \leqslant \|v\|_2\|x\|_2\|u\|_2,$$

where the inequality comes from using Cauchy-Schwarz inequality.

This gives, for any x,

$$\frac{\|uv^\top x\|_2}{\|x\|_2} \leqslant \|u\|_2\|v\|_2 \implies \sup_{x\neq 0}\frac{\|uv^\top x\|_2}{\|x\|_2} \leqslant \|u\|_2\|v\|_2 \implies \|uv^\top\| \leqslant \|u\|_2\|v\|_2.$$

For $x = v$,

$$\frac{\|uv^\top x\|_2}{\|x\|_2} = \frac{\|uv^\top v\|_2}{\|v\|_2} = \frac{\|u\|_2\|v\|_2^2}{\|v\|_2} = \|u\|_2\|v\|_2,$$

i.e., the supremum is attained at $x = v$ and so $\|uv^\top\| = \|u\|_2\|v\|_2$. Hence, we have

$$\boxed{\|uv^\top\| = \|uv^\top\|_F = \|u\|_2\|v\|_2.}$$

In matrix D, only the diagonal entries are non-zero, so

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

We know, for a diagonal matrix D, D^2 is also diagonal and its largest eigenvalue is $\max_i d_i^2$. So, using the definition of spectral norm:

$$\|D\| = \sigma_1(D) = \sqrt{\max_i d_i^2} = \max_i |d_i| \implies \|D\| = \max_{1\leqslant i\leqslant n}|d_i| = \|d\|_\infty. \quad \square$$

(b2) (General relationship) If $\operatorname{rank}(A) = r$, show that

$$\|A\| \leqslant \|A\|_F \leqslant \sqrt{r}\|A\|.$$

Show that both inequalities can be tight (achieved) for any admissible (m, n, r).

Solution:

Since A has $\operatorname{rank}(A) = r$, let $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r > 0$ are the nonzero singular values of A. Then, by the definition of Frobenius and Spectral norms:

$$\|A\|_F^2 = \sum_{i=1}^{r}\sigma_i^2 \leqslant \sum_{i=1}^{r}\sigma_1^2 \leqslant r\sigma_1^2 \implies \|A\|_F \leqslant \sqrt{r}\sigma_1 = \sqrt{r}\|A\|.$$

And, we know $\sigma_1^2 \leqslant \sum_{i=1}^{r}\sigma_i^2$, which gives $\|A\| \leqslant \|A\|_F$. Hence, combining both:

$$\boxed{\|A\| \leqslant \|A\|_F \leqslant \sqrt{r}\|A\|.}$$

i. Tightness of $\|A\| \leqslant \|A\|_F$:
If $\operatorname{rank}(A) = 1$ i.e., $\sigma_1 > 0$ and $\sigma_2 = \sigma_3 = \cdots = \sigma_r = 0$, then $\sum_i \sigma_i^2 = \sigma_1^2 \implies \|A\| = \|A\|_F$. For example, let $A = uv^\top$ for any $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$, with $m, n \geqslant 1$. Here, $\operatorname{rank}(A) = 1$, and as we proved in (b1):

$$\|uv^\top\| = \|uv^\top\|_F \text{ i.e., } \|A\| = \|A\|_F.$$

ii. Tightness of $\|A\|_F \leqslant \sqrt{r}\|A\|$:
For any admissible (m, n, r) i.e., $r \leqslant \min\{m, n\}$, take:

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

where $s > 0$ is any value of choice. Here, $\operatorname{rank}(A) = r$ and $\sum_i \sigma_i^2 = rs^2$. So, $\|A\| = s$ and $\|A\|_F = \sqrt{r}s$. Hence, if all nonzero singular values of A are equal then $\|A\|_F = \sqrt{r}\|A\|$.

(b3) (Frobenius submultiplicativity) Show that for conformable matrices A, B,

$$\|AB\|_F \leqslant \|A\|\|B\|_F \qquad \text{and} \qquad \|AB\|_F \leqslant \|A\|_F\|B\|.$$

Conclude (by combining an inequality from (b2)) that $\|\cdot\|_F$ is submultiplicative, i.e.,

$$\|AB\|_F \leqslant \|A\|_F\|B\|_F.$$

Solution:
Using the definition of the Frobenius norm:

$$\|AB\|_F = \operatorname{tr}((AB)^\top AB) = \operatorname{tr}(B^\top A^\top AB).$$

We know, for any matrix A and any vector x, using the Spectral norm:

$$\|Ax\|_2^2 \leqslant \|A\|^2\|x\|_2^2 \implies x^\top A^\top Ax \leqslant \|A\|^2 x^\top x = x^\top(\|A\|^2 I)x.$$

So, we get the positive semi-definite ordering:

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

Using this ordering and monotonicity of the trace:

$$\|AB\|_F = \operatorname{tr}(B^\top A^\top AB) \leqslant \operatorname{tr}(B^\top\|A\|IB) = \|A\|\operatorname{tr}(B^\top B) = \|A\|\|B\|_F.$$

We know, the Frobenius norm is invariant of the transpose, so $\|AB\|_F = \|(AB)^\top\|_F = \|B^\top A^\top\|_F$. Using the result just proved,

$$\|AB\|_F = \|B^\top A^\top\|_F \leqslant \|B^\top\|\|A^\top\|_F.$$

But again, $\|A^\top\|_F = \|A\|_F$ and $\|B^\top\| = \|B\|$ using the result proved in (a3). So, we have

$$\|AB\|_F \leqslant \|B^\top\|\|A^\top\|_F = \|A\|_F\|B\|.$$

Hence, combining both we get,

$$\boxed{\|AB\|_F \leqslant \|A\|\|B\|_F \qquad \text{and} \qquad \|AB\|_F \leqslant \|A\|_F\|B\|.}$$

We proved in (b2), that for any matrix A, $\|A\| \leqslant \|A\|_F$. This gives:

$$\boxed{\|AB\|_F \leqslant \|A\|_F\|B\|_F.}$$

(b4) [Bonus] (Orthogonal invariance) Let $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ be orthogonal. Show that

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

[Bonus] Solution:
Here, $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal, so

$$U^\top U = I_m \qquad \text{and} \qquad V^\top V = V^\top V = I_n.$$

Using the definition of the Spectral norm:

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

For any orthogonal matrix U, $(Ux)^\top(Ux) = x^\top U^\top Ux = x^\top x = \|x\|_2$, i.e., orthogonal matrices preserve l_2 (Euclidean) norm. So, $\|UAVx\|_2 = \|AVx\|_2$. This gives:

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

Now, let $y = Vx$. Again, as orthogonal matrices preserve the Euclidean norm, $\|y\|_2 = \|x\|_2 = 1$. So,

$$\|UAV\| = \sup_{\|x\|_2=1} \|AVx\|_2 = \sup_{\|y\|_2=1} \|Ay\|_2 = \|A\|. \quad \square$$

Next, using the definition of the Frobenius norm:

$$\begin{aligned}
\|UAV\|_F &= \mathrm{tr}((UAV)^\top UAV) \\
&= \mathrm{tr}(V^\top A^\top U^\top UAV) \\
&= \mathrm{tr}(V^\top A^\top AV) && [\text{because } U^\top U = I_m] \\
&= \mathrm{tr}(A^\top AVV^\top) && [\text{cyclic property of trace}] \\
&= \mathrm{tr}(A^\top A) && [\text{because } VV^\top = I_n] \\
&= \|A\|_F. \quad \square
\end{aligned}$$

Hence, for orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$,

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

3 Variance and covariance identities

Unless specified otherwise, assume all random variables/vectors below have finite second moments. Throughout, for $u, v \in \mathbb{R}^d$ we write $\langle u, v\rangle := u^\top v$ and $\|u\|_2^2 = \langle u, u\rangle$.

(a) **Variance identities.** Let $Z \in \mathbb{R}^d$ be a random vector with mean $m := \mathbb{E}[Z] \in \mathbb{R}^d$ (defined coordinatewise). The quantity $\mathbb{E}\|Z - \mathbb{E}Z\|_2^2$ is sometimes called the "total variance".

(a1) (Variance via Pythagorean decomposition) Assume $\mathbb{E}\|Z\|_2^2 < \infty$ and let $m = \mathbb{E}Z$. Show that

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

More generally, show that for every $a \in \mathbb{R}^d$,

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

and conclude that

$$\mathbb{E}\|Z - m\|_2^2 = \min_{a\in\mathbb{R}^d} \mathbb{E}\|Z - a\|_2^2, \qquad \text{with unique minimizer } a^\star = \mathbb{E}[Z].$$

Solution:
For any $a \in \mathbb{R}^d$,

$$\|Z - a\|_2^2 = \|Z - m + m - a\|_2^2 = \|Z - m\|_2^2 + 2\langle Z - m, m - a\rangle + \|m - a\|_2^2.$$

Taking expectations:

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

Here, $m - a$ is deterministic and $\mathbb{E}[Z - m] = 0$, so

$$\mathbb{E}\langle Z - m, m - a\rangle = \langle \mathbb{E}[Z - m], m - a\rangle = \langle 0, m - a\rangle = 0.$$

Hence, for every $a \in \mathbb{R}^d$,

$$\boxed{\mathbb{E}\|Z - a\|_2^2 = \mathbb{E}\|Z - m\|_2^2 + \|a - m\|_2^2}.$$

If we let $a = m$, then $\|a - m\|_2^2 = 0$. And,

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

Hence, for $m = \mathbb{E}Z$,

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

Here, $\mathbb{E}\|Z - a\|_2^2 \geqslant \mathbb{E}\|Z - m\|_2^2$ and the equality holds iff $\mathbb{E}\|m - a\|_2^2 = 0 \Leftrightarrow m = a$.
Hence,

$$\boxed{\mathbb{E}\|Z - m\|_2^2 = \min_{a\in\mathbb{R}^d} \mathbb{E}\|Z - a\|_2^2, \qquad \text{with unique minimizer } a^\star = \mathbb{E}[Z].}$$

(a2) (Variance via symmetrization) Let Z' be an independent copy of Z (i.e., $Z' \perp\!\!\!\perp Z$ and $Z' \overset{d}{=} Z$). Show that

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

Solution:
Here, Z' is an independent copy of Z, so $\mathbb{E}Z = \mathbb{E}Z'$ and $\mathbb{E}\|Z - \mathbb{E}Z\|_2^2 = \mathbb{E}\|Z' - \mathbb{E}Z'\|_2^2$.

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

Now, using the linearity of the inner product,

$$\langle Z - \mathbb{E}Z, \mathbb{E}Z - Z'\rangle = \langle Z, \mathbb{E}Z - Z'\rangle - \langle \mathbb{E}Z, \mathbb{E}Z - Z'\rangle = \langle Z, \mathbb{E}Z\rangle - \langle Z, Z'\rangle - \langle \mathbb{E}Z, \mathbb{E}Z - Z'\rangle.$$

Taking expectation:

$$\begin{aligned}
\mathbb{E}\langle Z - \mathbb{E}Z, \mathbb{E}Z - Z'\rangle &= \langle \mathbb{E}Z, \mathbb{E}Z\rangle - \langle \mathbb{E}Z, \mathbb{E}Z'\rangle - \langle \mathbb{E}Z, \mathbb{E}Z - \mathbb{E}Z'\rangle \\
&= \langle \mathbb{E}Z, \mathbb{E}Z\rangle - \langle \mathbb{E}Z, \mathbb{E}Z\rangle - \langle \mathbb{E}Z, 0\rangle \\
&= 0
\end{aligned}$$

So,

$$\mathbb{E}\|Z - Z'\|_2^2 = \mathbb{E}\|Z - \mathbb{E}Z\|_2^2 + \mathbb{E}\|Z' - \mathbb{E}Z'\|_2^2 = 2\mathbb{E}\|Z - \mathbb{E}Z\|_2^2.$$

Rearranging, we get

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

(a3) [Bonus] (Variance for independent sums) Let $Z_1, \dots, Z_k \in \mathbb{R}^d$ be independent random vectors with $\mathbb{E}[Z_j] = 0$ and $\mathbb{E}\|Z_j\|_2^2 < \infty$. Show that

$$\mathbb{E}\left\|\sum_{j=1}^k Z_j\right\|_2^2 = \sum_{j=1}^k \mathbb{E}\|Z_j\|_2^2.$$

[Bonus] Solution:
Here, using the linearity of the inner product,

$$\mathbb{E}\left\|\sum_{j=1}^k Z_j\right\|_2^2 = \mathbb{E}\left\langle \sum_{j=1}^k Z_j, \sum_{j=1}^k Z_j\right\rangle = \mathbb{E}\left[\sum_{j=1}^k \sum_{l=1}^k \langle Z_j, Z_l\rangle\right].$$

Using linearity of expectations,

$$\mathbb{E}\left[\sum_{j=1}^k \sum_{l=1}^k \langle Z_j, Z_l\rangle\right] = \sum_{j=1}^k \sum_{l=1}^k \mathbb{E}\langle Z_j, Z_l\rangle = \sum_{j=1}^k \mathbb{E}\|Z_j\|_2^2 + \sum_{j\neq l} \mathbb{E}\langle Z_j, Z_l\rangle.$$

Since all the Z_i are independent, for $j \neq l$,

$$\mathbb{E}\langle Z_j, Z_l\rangle = \langle \mathbb{E}Z_j, \mathbb{E}Z_l\rangle = \langle 0, 0\rangle = 0.$$

Hence,

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

(b) **Covariance identities.** For a random vector $Z \in \mathbb{R}^d$ with $m = \mathbb{E}[Z]$, define the covariance matrix

$$\mathrm{Cov}(Z) := \mathbb{E}[(Z-m)(Z-m)^\top] \in \mathbb{R}^{d\times d}.$$

Show that:

(b1) $\mathrm{Cov}(Z) = \mathbb{E}[ZZ^\top] - mm^\top$.
Solution:

$$\begin{aligned}
\mathrm{Cov}(Z) &= \mathbb{E}[(Z-m)(Z-m)^\top] \\
&= \mathbb{E}[ZZ^\top - Zm^\top - mZ^\top - mm^\top] \\
&= \mathbb{E}[ZZ^\top] - \mathbb{E}[Zm^\top] - \mathbb{E}[mZ^\top] - mm^\top \\
&= \mathbb{E}[ZZ^\top] - \mathbb{E}[Z]m^\top - m\mathbb{E}[Z^\top] - mm^\top \\
&= \mathbb{E}[ZZ^\top] - mm^\top. \quad \square
\end{aligned}$$

(b2) For every $v \in \mathbb{R}^d$, $\mathrm{Var}(v^\top Z) = v^\top \mathrm{Cov}(Z) v$.
Solution:
Here, for any $v \in \mathbb{R}^d$, $v^\top Z \in \mathbb{R}$ (scalar) and $\mathbb{E}[v^\top Z] = v^\top \mathbb{E} Z = v^\top m$. Now,

$$\begin{aligned}
\mathrm{Var}(v^\top Z) &= \mathbb{E}\left[(v^\top Z - v^\top m)^2\right] \\
&= \mathbb{E}\left[(v^\top (Z-m))^2\right] \\
&= \mathbb{E}[v^\top (Z-m)(Z-m)^\top v] \\
&= v^\top \mathbb{E}[(Z-m)(Z-m)^\top] v \\
&= v^\top \mathrm{Cov}(Z) v. \quad \square
\end{aligned}$$

(b3) $\mathrm{tr}(\mathrm{Cov}(Z)) = \mathbb{E}\|Z - \mathbb{E}Z\|_2^2$.
Solution:
Using the linearity of the trace and expectation:

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

We know, $\mathrm{tr}(uu^\top) = \|u\|_2^2$. So,

$$\mathrm{tr}((Z-m)(Z-m)^\top) = \|Z-m\|_2^2 = \|Z - \mathbb{E}Z\|_2^2.$$

Hence,

$$\mathrm{tr}(\mathrm{Cov}(Z)) = \mathbb{E}\|Z - \mathbb{E}Z\|_2^2. \quad \square$$

4 Moving between moments and tails

Throughout, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Assume all random variables below are real-valued and measurable, and that any expectations that appear are finite.

(a) **From tails to moments.** Let $X \geqslant 0$ be a nonnegative random variable.

(a1) (Moment identities using tail integrals) Show that

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}\{X > t\}\, dt.$$

More generally, show that for any $p > 0$,

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

Solution:
Here, X is a nonnegative random variable, so

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

using the pth-moment identity. Now, $g : X \times t \to [0, \infty)$ defined by $g(x,t) = pt^{p-1}\mathbf{1}\{x > t\} f_X(x)$ is a nonnegative measurable function, so using Tonelli's theorem,

$$\begin{aligned}
\mathbb{E}[X^p] &= \int_0^\infty \int_0^\infty p t^{p-1} \mathbf{1}\{x > t\} f_X(x)\, d(x,t) \\
&= \int_0^\infty p t^{p-1} \left(\int_0^\infty \mathbf{1}\{x > t\} f_X(x)\, dx \right) dt \\
&= \int_0^\infty p t^{p-1} \left(\int_t^\infty f_X(x)\, dx \right) dt \\
&= \int_0^\infty p t^{p-1} \mathbb{P}\{X > t\}\, dt.
\end{aligned}$$

Hence,

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

And, for $p = 1$, the expression becomes

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

(a2) [Bonus] (Moment growth bounds from exponential tail behavior) Assume there exist constants $c, C > 0$ and $\alpha > 0$ such that for all $t \geqslant 0$,

$$\mathbb{P}\{X > t\} \leqslant C \exp(-ct^\alpha).$$

Use part (a1) to show that $X \in L_p$ for all $p < \infty$ and that there exists a constant $C' = C'(c, C, \alpha)$ such that

$$\|X\|_{L_p} \leqslant C' p^{1/\alpha} \qquad \text{for all } p \geqslant 1.$$

Finally, interpret the cases $\alpha = 2$ and $\alpha = 1$ in terms of the heuristic moment growth for sub-Gaussian and sub-exponential random variables.

[Bonus] Solution:
Using the moment identity proved in part (a1), for $p \geqslant 1$:

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

Plugging in the given tail bound:

$$\mathbb{E}[X^p] \leqslant Cp\int_0^\infty t^{p-1}e^{-ct^\alpha}\, dt.$$

Let $u = ct^\alpha \implies t = (u/c)^{1/\alpha}$ and $dt = (1/\alpha)c^{-1/\alpha}u^{1/\alpha-1}\, du$. Then,

$$\int_0^\infty t^{p-1}e^{-ct^\alpha} = \int_0^\infty \left(\frac{u}{c}\right)^{(p-1)/\alpha} e^{-u}\cdot\frac{1}{\alpha}c^{-1/\alpha}u^{1/\alpha-1}\, du = \frac{c^{-p/\alpha}}{\alpha}\int_0^\infty u^{p/\alpha-1}e^{-u}\, du,$$

where the last integral is a Gamma function and integrates to $\Gamma\left(\frac{p}{\alpha}\right)$. Hence,

$$\boxed{\mathbb{E}[X^p] \leqslant \frac{Cp}{\alpha}c^{-p/\alpha}\Gamma\left(\frac{p}{\alpha}\right) < \infty, \quad \forall p < \infty \implies X \in L_p, \quad \forall p < \infty.}$$

Now, taking the p–th roots:

$$\|X\|_{L_p} = (\mathbb{E}[X^p])^{1/p} \leqslant \left(\frac{C}{\alpha}\right)^{1/p} p^{1/p}c^{-1/\alpha}\Gamma\left(\frac{p}{\alpha}\right)^{1/p}.$$

i. If $p \geqslant \alpha$, then $p/\alpha \geqslant 1$. So, let $z = p/\alpha$ and using the given bound for $\Gamma(z+1)$:

$$\Gamma\left(\frac{p}{\alpha}\right) = \frac{\Gamma\left(\frac{p}{\alpha}+1\right)}{p/\alpha} \leqslant \Gamma\left(\frac{p}{\alpha}+1\right) \leqslant \left(C_0\frac{p}{\alpha}\right)^{p/\alpha} \implies \Gamma\left(\frac{p}{\alpha}\right)^{1/p} \leqslant \left(C_0\frac{p}{\alpha}\right)^{1/\alpha}.$$

Plugging it back, we get,

$$\|X\|_{L_p} \leqslant \left(\frac{C}{\alpha}\right)^{1/p} p^{1/p}c^{-1/\alpha}\left(C_0\frac{p}{\alpha}\right)^{1/\alpha}.$$

We know, $\arg\max_x x^{1/x} = e$, so $p^{1/p} \leqslant e^{1/e} < e$ and $(C/\alpha)^{1/p} \leqslant \max\{1, C/\alpha\}$ for $p \geqslant 1$. So,

$$\|X\|_{L_p} \leqslant e\left(\frac{C_0}{c\alpha}\right)^{1/\alpha}\max\left\{1, \frac{C}{\alpha}\right\}p^{1/\alpha} \implies \|X\|_{L_p} \leqslant C'p^{1/\alpha},$$

for $C' = C'(c, C, \alpha) = eC_0^{1/\alpha}(c\alpha)^{-1/\alpha}\max\{1, C/\alpha\}$.

ii. If $p \leqslant \alpha$, then using the monotonicity proved in Problem 1 (b1):

$$\|X\|_{L_p} \leqslant \|X\|_{L_\alpha} \leqslant \left[C\alpha\int_0^\infty t^{\alpha-1}e^{-ct^\alpha}\, dt\right]^{1/\alpha},$$

using the given tail bound. Let $u = ct^\alpha \implies t = (u/c)^{1/\alpha}$ and $dt = (1/\alpha)c^{-1/\alpha}u^{1/\alpha-1}\, du$. Then,

$$\int_0^\infty t^{\alpha-1}e^{-ct^\alpha} = \int_0^\infty \left(\frac{u}{c}\right)^{(\alpha-1)/\alpha} e^{-u}\cdot\frac{1}{\alpha}c^{-1/\alpha}u^{1/\alpha-1}\, du = \frac{1}{c\alpha}\int_0^\infty e^{-u}\, du = \frac{1}{c\alpha}.$$

So,

$$\|X\|_{L_p} \leqslant \left(\frac{C\alpha}{c\alpha}\right)^{1/\alpha} = \left(\frac{C}{c}\right)^{1/\alpha} \leqslant \left(\frac{C}{c}\right)^{1/\alpha} p^{1/\alpha},$$

because for $p \geqslant 1$, $p^{1/\alpha} \geqslant 1$. Thus, $\|X\|_{L_p} \leqslant C' p^{1/\alpha}$ with $C' = C'(c, C, \alpha) = (C/c)^{1/\alpha}$.

Hence, combining both cases,

$$\boxed{\exists C' = C'(c, C, \alpha) \qquad \text{such that} \qquad \|X\|_{L_p} \leqslant C' p^{1/\alpha} \qquad \text{for all } p \geqslant 1.}$$

Now,

i. For $\alpha = 2$, $\mathbb{P}\{X > t\} \leqslant Ce^{-ct^2}$ i.e., Gaussian-type tail. Then,

$$\|X\|_{L_p} \leqslant C' \sqrt{p},$$

i.e., moments grow like $\sqrt{p}$.

ii. For $\alpha = 1$, $\mathbb{P}\{X > t\} \leqslant Ce^{-ct}$ i.e., Exponential tail. Then,

$$\|X\|_{L_p} \leqslant C' p,$$

i.e., moments grow linearly in p.

(b) **Classical tail bounds from second moments.** Let X and Y be real-valued random variables.

(b1) (Upper bound) Fix $t > 0$. Show the one-sided *Cantelli's inequality*

$$\mathbb{P}\{Y - \mathbb{E}Y \geqslant t\} \leqslant \frac{\mathrm{Var}(Y)}{\mathrm{Var}(Y) + t^2}.$$

Solution:

Here, let $\mathrm{Var}(Y) = \sigma^2$. For all $a > 0$,

$$\begin{aligned}
\mathbb{P}\{Y - \mathbb{E}Y \geqslant t\} &= \mathbb{P}\{Y - \mathbb{E}Y + a \geqslant t + a\} \\
&\leqslant \mathbb{P}\{(Y - \mathbb{E}Y + a)^2 \geqslant (t + a)^2\} \\
&\leqslant \frac{\mathbb{E}[(Y - \mathbb{E}Y + a)^2]}{(t + a)^2} \qquad \text{[Using Markov's inequality]} \\
&\leqslant \frac{\mathrm{Var}(Y) + a^2}{(t + a)^2} = \frac{\sigma^2 + a^2}{(t + a)^2}.
\end{aligned}$$

Now, we will optimize the upper bound to find a tighter bound. We will find the value of a that minimizes $f(a) = \dfrac{\sigma^2 + a^2}{(t + a)^2}$. We have both $a, t > 0$, so

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

equals to 0 only when $at - \sigma^2 = 0$, which gives

$$\arg_a \{f(a) = 0\} = a^* = \frac{\sigma^2}{t}.$$

Now, the second derivative is

$$f''(a) = \frac{2(t^2 - 2at + 3\sigma^2)}{(t+a)^4}.$$

$f''(a^*) = \dfrac{2(t^2+\sigma^2}{(t+\sigma^2/t)^4} > 0$ so a^* is the global minimum of $f(a)$, and $f(a^*) = \dfrac{\sigma^2 + \sigma^4/t^2}{(t+\sigma^2/t)^2} = \dfrac{\sigma^2}{\sigma^2+t^2}$. So, the inequality becomes:

$$\mathbb{P}\{Y - \mathbb{E}Y \geqslant t\} \leqslant \inf_a f(a) = \frac{\sigma^2}{\sigma^2+t^2}.$$

Hence,

$$\boxed{\mathbb{P}\{Y - \mathbb{E}Y \geqslant t\} \leqslant \frac{\operatorname{Var}(Y)}{\operatorname{Var}(Y)+t^2}.}$$

(b2) (Comparison with Chebyshev's inequality) Use Chebyshev's inequality to give a "naive" one-sided bound on $\mathbb{P}\{Y - \mathbb{E}Y \geqslant t\}$ and compare it to Cantelli's inequality from (b1). Now apply Cantelli to Y and to $-Y$ and deduce a two-sided tail bound

$$\mathbb{P}\{|Y - \mathbb{E}Y| \geqslant t\} \leqslant \frac{2\operatorname{Var}(Y)}{\operatorname{Var}(Y)+t^2}.$$

Compare this to the standard two-sided Chebyshev inequality.

Solution:

Chebyshev's inequality gives

$$\mathbb{P}\left\{|Y - \mathbb{E}Y| \geqslant t\right\} \leqslant \frac{\operatorname{Var}(Y)}{t^2}.$$

Since $\{Y - \mathbb{E}Y \geqslant t\} \subseteq \{|Y - \mathbb{E}Y| \geqslant t\}$, $\mathbb{P}\left\{Y - \mathbb{E}Y \geqslant t\right\} \leqslant \mathbb{P}\left\{|Y - \mathbb{E}Y| \geqslant t\right\}$. This gives a "naive" one-sided bound:

$$\mathbb{P}\left\{Y - \mathbb{E}Y \geqslant t\right\} \leqslant \frac{\sigma^2}{t^2}.$$

Because $\sigma^2 \geqslant 0$, $\sigma^2 + t^2 \geqslant t^2$, and $\sigma^2/(\sigma^2+t^2) \leqslant \sigma^2/t^2$. Hence, Cantelli's inequality is always at least as strong as the naive one-sided Chebyshev bound. Let's apply Cantelli's inequality in (b1) to $-Y$. $\mathbb{E}[-Y] = -\mathbb{E}Y$ and $\operatorname{Var}(-Y) = \sigma^2$. Then, using (b1),

$$\mathbb{P}\left\{-Y + \mathbb{E}Y \geqslant t\right\} \leqslant \frac{\sigma^2}{\sigma^2+t^2} \implies \mathbb{P}\left\{Y - \mathbb{E}Y \leqslant -t\right\} \leqslant \frac{\sigma^2}{\sigma^2+t^2}.$$

Now, using the union bound,

$$\mathbb{P}\left\{|Y - \mathbb{E}Y| \geqslant t\right\} \leqslant \mathbb{P}\left\{Y - \mathbb{E}Y \geqslant t\right\} + \mathbb{P}\left\{Y - \mathbb{E}Y \leqslant -t\right\} = \frac{2\sigma^2}{\sigma^2+t^2}.$$

Hence, the two-sided Cantelli bound is:

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

For $t^2 \leqslant \sigma^2$, $\frac{2\sigma^2}{\sigma^2+t^2} \leqslant \frac{\sigma^2}{t^2}$, so the two-sided Cantelli bound is stronger than the standard two-sided Chebyshev bound for small t. However, for large t, i.e., $t^2 \geqslant \sigma^2$, $\frac{\sigma^2}{t^2} \leqslant \frac{2\sigma^2}{\sigma^2+t^2}$. So, the standard two-sided Chebyshev bound is stronger than the two-sided Cantelli bound for large t. And for much larger t, $t^2 >> \sigma^2$, $\frac{2\sigma^2}{\sigma^2+t^2} \approx \frac{2\sigma^2}{t^2}$. So, the Chebyshev bound is almost half of the two-sided Cantelli bound.

(b3) [Bonus] (Lower bound) Let $X \geqslant 0$ with $\mathbb{E}[X^2] < \infty$ and $\mathbb{E}[X] > 0$. Fix $\theta \in (0,1)$. Prove the *Paley–Zygmund inequality*

$$\mathbb{P}\{X \geqslant \theta\mathbb{E}X\} \geqslant (1-\theta)^2\frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]} = (1-\theta)^2\frac{(\mathbb{E}X)^2}{\mathrm{Var}(X)+(\mathbb{E}X)^2} = \frac{(1-\theta)^2}{1+\mathrm{Var}(X)/(\mathbb{E}X)^2}.$$

[Bonus] Solution:
We can write the expectation of X as:

$$\begin{aligned}\mathbb{E}X &= \mathbb{E}[X\mathbf{1}\{X < \theta\mathbb{E}X\}] + \mathbb{E}[X\mathbf{1}\{X \geqslant \theta\mathbb{E}X\}] \\ &\leqslant \theta\mathbb{E}X + \mathbb{E}[X\mathbf{1}\{X \geqslant \theta\mathbb{E}X\}].\end{aligned}$$

This gives $(1-\theta)\mathbb{E}X \leqslant \mathbb{E}[X\mathbf{1}\{X \geqslant \theta\mathbb{E}X\}] \implies (1-\theta)^2(\mathbb{E}X)^2 \leqslant (\mathbb{E}[X\mathbf{1}\{X \geqslant \theta\mathbb{E}X\}])^2$. We know, using Cauchy-Schwarz inequality,

$$(\mathbb{E}[X\mathbf{1}\{X \geqslant \theta\mathbb{E}X\}])^2 \leqslant \mathbb{E}\left[X^2\right]\mathbb{E}[\mathbf{1}\{X \geqslant \theta\mathbb{E}X\}] = \mathbb{E}\left[X^2\right]\mathbb{P}\left\{X \geqslant \theta\mathbb{E}X\right\}.$$

Putting it together, we get,

$$(1-\theta)^2(\mathbb{E}X)^2 \leqslant \mathbb{E}\left[X^2\right]\mathbb{P}\left\{X \geqslant \theta\mathbb{E}X\right\}.$$

Finally, rearranging, we get the Paley-Zygmund inequality:

$$\boxed{\mathbb{P}\{X \geqslant \theta\mathbb{E}X\} \geqslant (1-\theta)^2\frac{(\mathbb{E}X)^2}{\mathbb{E}[X^2]}.}$$

We know, $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$, so the inequality can be written as:

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

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

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

5 Mean versus median

(a) **Mean–median closeness (using variance).** Let Z be a real-valued random variable with $\mathbb{E}[Z^2] < \infty$, and let M_Z be a median of Z, meaning

$$\mathbb{P}(Z \geqslant M_Z) \geqslant \frac{1}{2} \qquad \text{and} \qquad \mathbb{P}(Z \leqslant M_Z) \geqslant \frac{1}{2}.$$

Show that

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

Solution:
If $M_Z > \mathbb{E}Z$, then $\mathbb{P}\{Z \geqslant M_Z\} = \mathbb{P}\{Z - \mathbb{E}Z \geqslant M_Z - \mathbb{E}Z\} \geqslant 1/2$. So, using one-sided Cantelli's inequality,

$$\frac{1}{2} \leqslant \mathbb{P}\{Z - \mathbb{E}Z \geqslant M_Z - \mathbb{E}Z\} \leqslant \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + (M_Z - \mathbb{E}Z)^2}.$$

Taking the extreme bounds,

$$\frac{1}{2} \leqslant \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + (M_Z - \mathbb{E}Z)^2} \implies (M_Z - \mathbb{E}Z)^2 \leqslant \mathrm{Var}(Z) \implies (M_Z - \mathbb{E}Z) \leqslant \sqrt{\mathrm{Var}(Z)}.$$

Similarly, if $M_Z < \mathbb{E}Z$, then $\mathbb{P}\{Z \leqslant M_Z\} = \mathbb{P}\{-Z \geqslant -M_Z\} = \mathbb{P}\{\mathbb{E}Z - Z \geqslant \mathbb{E}Z - M_Z\} \geqslant 1/2$. Using one-sided Cantelli's inequality,

$$\frac{1}{2} \leqslant \mathbb{P}\{\mathbb{E}Z - Z \geqslant \mathbb{E}Z - M_Z\} \leqslant \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + (\mathbb{E}Z - M_Z)^2}.$$

Taking the extreme bounds,

$$\frac{1}{2} \leqslant \frac{\mathrm{Var}(Z)}{\mathrm{Var}(Z) + (\mathbb{E}Z - M_Z)^2} \implies (\mathbb{E}Z - M_Z)^2 \leqslant \mathrm{Var}(Z) \implies (\mathbb{E}Z - M_Z) \leqslant \sqrt{\mathrm{Var}(Z)}.$$

If $M_Z = \mathbb{E}Z$, then the inequality is trivially true because variance is non-negative. Hence,

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

(b) [Bonus] **Mean–median closeness (using variance proxy).** Let X be a real-valued random variable with median M_X. Assume there exist constants $a > 0$ and $b > 0$ such that for all $t > 0$,

$$\mathbb{P}(|X - M_X| > t) \leqslant a e^{-t^2/b}.$$

Show that

$$|M_X - \mathbb{E}X| \leqslant \mathbb{E}|X - M_X| \leqslant \min\left\{\sqrt{ab},\, a\sqrt{b\pi/4}\right\}.$$

[Bonus] Solution:
Absolute value is a convex function, so using Jensen's inequality:

$$|\mathbb{E}[M_X - X]| \leqslant \mathbb{E}\,|M_X - X| \implies |M_X - \mathbb{E}X| \leqslant \mathbb{E}|X - M_X|.$$

Now,

$$\mathbb{E}\,|X - M_X| = \int_0^\infty \mathbb{P}\left\{|X - M_X| \geqslant t\right\}\,dt \leqslant \int_0^\infty ae^{-t^2/b}\,dt = \frac{a\sqrt{\pi b}}{2}\int_{-\infty}^\infty \frac{e^{-t^2/b}}{\sqrt{\pi b}}\,dt,$$

where the final integrand is the kernel of a Normal$(0, b/2)$ density, which integrates to 1. So,

$$\mathbb{E}\,|X - M_X| \leqslant a\sqrt{b\pi/4}.$$

Similarly,

$$\begin{aligned}
\mathbb{E}\big[(X - M_X)^2\big] &= \int_0^\infty 2t\mathbb{P}\left\{|X - M_X| \geqslant t\right\}\,dt \\
&\leqslant \int_0^\infty 2tae^{-t^2/b}\,dt \\
&= ab\int_0^\infty be^{-u}\,du \qquad [\text{By substituting } u = t^2/b] \\
&= ab.
\end{aligned}$$

Using Jensen's inequality,

$$(\mathbb{E}\,|X - M_X|)^2 \leqslant \mathbb{E}\big[(X - M_X)^2\big] \implies \mathbb{E}\,|X - M_X| \leqslant \sqrt{\mathbb{E}[(X - M_X)^2]} = \sqrt{ab}.$$

Hence,

$$\boxed{|M_X - \mathbb{E}X| \leqslant \mathbb{E}|X - M_X| \leqslant \min\left\{\sqrt{ab},\, a\sqrt{b\pi/4}\right\}.}$$

* * * End of Solutions * * *