

1 A quick proof of the Cauchy-Schwarz inequality

(a) Given vectors $u, v \in \mathbb{R}^n$, show that

$$\|u\|_2^2 + \|v\|_2^2 \geq 2u^\top v.$$

Hint: Start by expanding $\|u - v\|_2^2$.

Solution:

Note that $\|u - v\|_2^2$ is nonnegative since it is the square of a real quantity. Thus,

$$0 \leq \|u - v\|_2^2 = (u - v)^\top (u - v) = u^\top u + v^\top v - 2u^\top v.$$

Rearranging, we get

$$\|u\|_2^2 + \|v\|_2^2 \geq 2u^\top v. \quad (1)$$

(b) Assume u and v are nonzero. Apply the result from part (a) of this problem to the normalized vectors $u' := u/\|u\|_2$ and $v' := v/\|v\|_2$ to arrive at the conclusion that

$$u^\top v \leq \|u\|_2 \|v\|_2.$$

Note that this also holds if one or both of the vectors u and v is the zero vector, but it is obvious in these cases.

Solution:

Note that $\|u'\|_2 = \|v'\|_2 = 1$. Applying the inequality from part (a) of this problem, we get

$$2 = 1 + 1 = \|u'\|_2^2 + \|v'\|_2^2 \geq 2u'^\top v' = 2u^\top v / (\|u\|_2 \|v\|_2).$$

Rearranging, we get

$$\|u\|_2 \|v\|_2 \geq u^\top v.$$

(c) The standard statement of the Cauchy-Schwarz inequality, which follows immediately from the result in part (b), is

$$|u^\top v| \leq \|u\|_2 \|v\|_2 \quad \forall u, v \in \mathbb{R}^n.$$

In fact one can even strengthen this further to the sometimes strictly stronger statement that $\sum_{i=1}^n |u_i| |v_i| \leq \|u\|_2 \|v\|_2$ by simply replacing u by the vector with coordinates $|u_i|$, which has the same ℓ_2 norm as u , and v by the vector with coordinates $|v_i|$, which has the same ℓ_2 norm as v .

Now, consider a fixed, nonzero $u_0 \in \mathbb{R}^n$. For what v is the inequality an equality? Your answer should contain u_0 .

Solution:

Obviously, if $v = 0$, the LHS and RHS are both zero, so that's one such v . Now, suppose $v \neq 0$, so u_0 and v are both nonzero. Then they have nonzero norms, so we can rearrange the inequality to

$$\frac{|u_0^\top v|}{\|u_0\|_2 \|v\|_2} \leq 1 \iff |\cos(\theta)| \leq 1,$$

where θ is the angle between u_0 and v .

For the inequality to be an equality, we need $|\cos(\theta)| = 1$, which happens precisely when $\theta = 0$ or $\theta = \pi$. Geometrically, this means that the vectors u_0 and v are collinear, either pointing in the same direction ($\theta = 0$) or opposite directions ($\theta = \pi$). Algebraically, we can parameterize this set as

$$\{v \in \mathbb{R}^n \mid v = \alpha u_0, \alpha \in \mathbb{R}\}.$$

To check our work, we can plug back into the original inequality:

$$\begin{aligned} |u_0^\top v| &= |u_0^\top (\alpha u_0)| \\ &= |\alpha| |u_0^\top u_0| \\ &= |\alpha| \|u_0\|_2^2 \\ &= |\alpha| \|u_0\|_2 \|u_0\|_2 \\ &= \|u_0\|_2 \|\alpha u_0\|_2 \\ &= \|u_0\|_2 \|v\|_2 \end{aligned}$$

So the inequality is indeed an equality when the vectors are collinear.

2 Monotonicity of ℓ_p norms.

(a) First, we'll show that for $a_1, \dots, a_n \in \mathbb{R}_{>0}$ (i.e., each a_i is a positive real number) and $0 < m < 1$,

$$\left(\sum_{i=1}^n a_i \right)^m \leq \sum_{i=1}^n a_i^m,$$

where the inequality is strict if $n \geq 2$.

i. Define

$$f(a_1, \dots, a_n) := \left(\sum_{i=1}^n a_i \right)^m - \sum_{i=1}^n a_i^m.$$

We'll show $f(a_1, \dots, a_n) \leq 0$ to prove the result. To start, compute $\partial f(a_1, \dots, a_n) / \partial a_j$, where $j \in \{1, \dots, n\}$.

Solution:

Applying the chain rule, we have for each $j \in \{1, \dots, n\}$ that

$$\frac{\partial f}{\partial a_j}(a_1, \dots, a_n) = m \left(\left(\sum_{i=1}^n a_i \right)^{m-1} - a_j^{m-1} \right).$$

ii. Using the result of the previous part, show that $\partial f(a_1, \dots, a_n) / \partial a_j \leq 0$ for all j , with strict inequality when $n \geq 2$. Remember that we assume $a_1, \dots, a_n \in \mathbb{R}_{>0}$.

Solution:

When $n = 1$ we have $\partial f(a_1, \dots, a_n) / \partial a_j = 0$ for all j , so we may assume that $n \geq 2$ and attempt to show the claimed strict inequality. We'll establish this through a chain of logical equivalences. Using the partial derivative we computed in the previous part, we have:

$$\begin{aligned} & \frac{\partial f}{\partial a_j}(a_1, \dots, a_n) < 0 \\ \iff & m \left(\left(\sum_{i=1}^n a_i \right)^{m-1} - a_j^{m-1} \right) < 0 \\ \iff & \left(\sum_{i=1}^n a_i \right)^{m-1} - a_j^{m-1} < 0 & (m > 0) \\ \iff & \left(\sum_{i=1}^n a_i \right)^{m-1} < a_j^{m-1} \\ \iff & \sum_{i=1}^n a_i > a_j & (m-1^{\text{th}} \text{ root, and } m-1 < 0 \text{ so inequality flips}) \\ \iff & a_j + \sum_{i=1, i \neq j}^n a_i > a_j \\ \iff & \sum_{i=1, i \neq j}^n a_i > 0. \end{aligned}$$

This final inequality is clearly true: we assume all a_i are positive, so their sum (with one element removed) must also be positive. So the original inequality must also be true, proving that the derivative is negative wherever the arguments are all positive.

- iii. Use the result of the previous part to conclude that $f(a_1, \dots, a_n) \leq 0$ when $a_1, \dots, a_n \in \mathbb{R}_{>0}$, with strict inequality when $n \geq 2$, thus proving the original claim.

Hint: To start, it may help to compute $f(0, \dots, 0)$.

Solution:

Clearly, $f(0, \dots, 0) = 0$ since all terms in both summations vanish when all the a_i are zero. When $n = 1$ we have $f(a_1) = 0$ for all a_1 , and in particular for $a_1 \in \mathbb{R}_{>0}$, so we may assume that $n \geq 2$.

We know from the previous part of the problem that all the partial derivatives of f are strictly negative when $a_i > 0 \forall i$. If we follow the path from $(0, \dots, 0)$ to (a_1, \dots, a_n) given by

$$t \rightarrow f(ta_1, \dots, ta_n), \quad t \in [0, 1],$$

and observe that

$$\frac{d}{dt}f(ta_1, \dots, ta_n) = \sum_{j=1}^n a_j \frac{\partial f}{\partial a_j}(ta_1, \dots, ta_n) < 0 \text{ for all } t > 0,$$

we have

$$\begin{aligned} f(a_1, \dots, a_n) &= f(0, \dots, 0) + \int_0^1 \frac{d}{dt}f(ta_1, \dots, ta_n) dt \\ &< f(0, \dots, 0) \\ &= 0, \end{aligned}$$

which proves the original result. The second line follows because the integrand is negative for all $t > 0$, so the integral over $t \in (0, 1)$ is negative.

- (b) For $x \in \mathbb{R}^n$, recall the definition of the ℓ_p norm:

$$\|x\|_p := \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

Using the result of part (a), conclude that for $1 \leq q < p < \infty$ and nonzero $x \in \mathbb{R}^n$ having at least two nonzero coordinates, we have

$$\|x\|_p < \|x\|_q.$$

Solution:

Let $[n] = \{1, \dots, n\}$. Setting $m = q/p$ and $a_i = |x_i|^p$, we have from part (a) of the problem that

$$\left(\sum_{i \in [n]: |x_i| > 0} |x_i|^p \right)^{q/p} < \sum_{i \in [n]: |x_i| > 0} (|x_i|^p)^{q/p} = \sum_{i \in [n]: |x_i| > 0} |x_i|^q,$$

where we have used the assumption that x has at least two nonzero coordinates. The desired inequality follows from taking the $1/q$ -th power of both sides.

Note that if e_i denotes the unit vector in direction i for any $i \in [n]$ we have $\|e_i\|_\infty = \|e_i\|_q$ for all $1 \leq q < \infty$. Thus the strict inequality proved in part (b) of the problem does not hold when x has exactly one nonzero coordinate, and it is also obvious that the inequality cannot be strict for the zero vector. When x has at least two nonzero coordinates it is also straightforward to see that the inequality extends to $p = \infty$ in the sense that $\|x\|_q > \|x\|_\infty$ for all $1 \leq q < \infty$. To see this, simply pick some p such that $1 \leq q < p < \infty$. It should be obvious that $\|x\|_p \geq \|x\|_\infty$ for all $x \in \mathbb{R}^n$, and since we have proved that $\|x\|_p < \|x\|_q$ when x has at least two nonzero coordinates, we conclude that $\|x\|_\infty < \|x\|_q$ when x has at least two nonzero coordinates.

3 Gram-Schmidt process

Let \mathcal{P} be the subspace in \mathbb{R}^3 spanned by the vectors $x_1 := [2 \ 2 \ 1]^\top$ and $x_2 := [2 \ 0 \ -1]^\top$. In the first two parts of the problem we will establish that \mathcal{P} is a plane and then we will proceed with the rest of the problem.

- (a) Find the angle between the vectors x_1 and x_2 .

Solution:

We have $\|x_1\|_2 = 3$, $\|x_2\|_2 = \sqrt{5}$, and $x_1^\top x_2 = 3$. The angle θ between the vectors x_1 and x_2 therefore satisfies

$$\cos(\theta) = \frac{x_1^\top x_2}{\|x_1\|_2 \|x_2\|_2} = \frac{1}{\sqrt{5}}.$$

We can write $\theta = \arccos(\frac{1}{\sqrt{5}})$, but of course what the angle is depends on how one chooses to define the arc cosine.

- (b) Show that x_1 and x_2 are linearly independent.

Hint: Think about the situation geometrically – what does the angle between the vectors tell us about their linear (in)dependence?

Solution:

Since x_1 and x_2 are both nonzero, they would be linearly dependent iff the cosine of the angle between them were either 1 or -1 . This is not the case here, so they are linearly independent.

- (c) Find an orthonormal basis $B_{\mathcal{P}}$ for the plane \mathcal{P} using the Gram-Schmidt process.

Solution:

Using the Gram-Schmidt process we first find an orthogonal basis for \mathcal{P} as follows:

$$\begin{aligned} v_1 &:= x_1 = [2 \ 2 \ 1]^\top \\ v_2 &:= x_2 - \frac{\langle x_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 = [2 \ 0 \ -1]^\top - \frac{3}{9} [2 \ 2 \ 1]^\top \\ &= [4/3 \ -2/3 \ -4/3]^\top. \end{aligned}$$

From this we can find orthonormal basis $B_{\mathcal{P}}$ as,

$$\begin{aligned} w_1 &= v_1 / \|v_1\|_2 = \frac{1}{3} [2 \ 2 \ 1]^\top \\ w_2 &= v_2 / \|v_2\|_2 = \frac{1}{3} [2 \ -1 \ -2]^\top. \end{aligned}$$

- (d) Extend $B_{\mathcal{P}}$ to B , an orthonormal basis for \mathbb{R}^3 .

Solution:

To extend $B_{\mathcal{P}}$ to an orthonormal basis for \mathbb{R}^3 , we'll first find x_3 such that x_1, x_2 and x_3 are linearly independent. For instance, we can check that $x_3 := [1 \ 0 \ 0]^\top$ satisfies this condition, by checking that the matrix

$$\begin{bmatrix} 1 & 2 & 2 \\ 0 & 2 & -1 \\ 0 & 1 & -2 \end{bmatrix}$$

is nonsingular (e.g., by showing its determinant is nonzero).

Now, we can apply the Gram-Schmidt process again and find v_3 as the vector orthogonal to $\text{span}(x_1, x_2)$ (or equivalently $\text{span}(v_1, v_2)$) which when added to the projection of x_3 on $\text{span}(x_1, x_2)$ returns x_3 . Namely, we find v_3 as follows:

$$\begin{aligned}v_3 &:= x_3 - \frac{\langle x_3, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \frac{\langle x_3, v_2 \rangle}{\langle v_2, v_2 \rangle} v_2 \\ &= [1 \ 0 \ 0]^\top - \frac{2}{9} [2 \ 2 \ 1]^\top - \frac{4/3}{4} [4/3 \ -2/3 \ -4/3]^\top \\ &= [1/9 \ -2/9 \ 2/9]^\top.\end{aligned}$$

Then we can find the third orthonormal basis vector w_3 in the orthonormal basis B as,

$$w_3 := v_3 / \|v_3\| = \frac{1}{3} [1 \ -2 \ 2]^\top.$$

(e) Use B to find the distance of the vector $[1 \ 2 \ 3]^\top$ from the plane \mathcal{P} .

Solution:

The distance of vector $z := [1 \ 2 \ 3]^\top$ from the plane \mathcal{P} can be computed using basis B . The distance from z to \mathcal{P} is the magnitude of the component of z along the basis vector w_3 when z is expressed in the orthonormal basis B . This is because the other two components of the orthonormal basis B lie in the plane \mathcal{P} . This distance is therefore

$$d = |\langle z, w_3 \rangle| = 1.$$

4 Extrema of the inner product over an Euclidean ball

Let $y \in \mathbb{R}^n$ be a given nonzero vector, and let $\mathcal{X} := \{x \in \mathbb{R}^n : \|x\|_2 \leq r\}$, where r is some given strictly positive number.

Solution:

Let $V \in \mathbb{R}^{n,n-1}$ be a matrix whose columns form an orthonormal basis for the subspace of \mathbb{R}^n orthogonal to y .

- (a) Determine the optimal value p_1^* and the optimal set (i.e., the set of *all* optimal solutions) of the problem $\min_{x \in \mathcal{X}} |y^\top x|$.

Solution:

Clearly $\min_{x \in \mathcal{X}} |y^\top x| \geq 0$, i.e. p_1^* , cannot be strictly negative. In fact, $p_1^* = 0$, since this value is attained by $x = 0$. Further, this value is attained by and only by any vector $x \in \mathcal{X}$ orthogonal to y . The optimal set can therefore be written as

$$\mathcal{X}_{opt} = \{x : x = Vz, \|z\|_2 \leq r\}.$$

Here we used the fact that $\|x\|_2 = \|Vz\|_2 = \|z\|_2$ because V has orthonormal columns.

- (b) Determine the optimal value p_2^* and the optimal set of the problem $\max_{x \in \mathcal{X}} |y^\top x|$.

Solution:

The Cauchy-Schwarz inequality tells us that for any x with $\|x\|_2 \leq r$ we have $|y^\top x| \leq r\|y\|_2$, with equality iff x has $\|x\|_2 = r$ and is proportional to y . The optimal value $\max_{x \in \mathcal{X}} |y^\top x|$, i.e. p_2^* , is therefore attained by and only by those x with $\|x\|_2 = r$ that are proportional to y , i.e of the form $x = \alpha y$ for some α . We must then have $|\alpha| = r/\|y\|_2$, which gives $p_2^* = r\|y\|_2$. The optimal set consists of two vectors

$$\mathcal{X}_{opt} = \left\{x : x = \alpha y, \alpha = \pm \frac{r}{\|y\|_2}\right\}.$$

- (c) Determine the optimal value p_3^* and the optimal set of the problem $\min_{x \in \mathcal{X}} y^\top x$.

Solution:

For $p_3^* := \min_{x \in \mathcal{X}} y^\top x$ we have $p_3^* = -r\|y\|_2$, which is attained at the unique optimal point $x^* = -\frac{r}{\|y\|_2}y$. This follows from the Cauchy-Schwarz inequality, as in part 2. of the problem.

- (d) Determine the optimal value p_4^* and the optimal set of the problem $\max_{x \in \mathcal{X}} y^\top x$.

Solution:

For $p_4^* := \max_{x \in \mathcal{X}} y^\top x$ we have $p_4^* = r\|y\|_2$, which is attained at the unique optimal point $x^* = \frac{r}{\|y\|_2}y$. This also follows from the Cauchy-Schwarz inequality, as in part 2. of the problem.