Inequalities

Youming Zhao
Email: youming0.zhao@gmail.com

First draft: January 1, 2022 Last update: October 14, 2023

Contents

1 Definitions

1.1 Norm

A function $f : \mathbf{R}^n \to \mathbf{R}$ with $\mathrm{dom} f = \mathbf{R}^n$ is called a norm if

- f is nonnegative: $f(x) \geq 0$ for all $x \in \mathbf{R}^n$
- f is definite: $f(x) = 0$ if and only if $x = 0$
- f is homogeneous: $f(tx) = |t|f(x)$, for all $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$
- f satisfies the triangle inequality: $f(x + y) \leq f(x) + f(y)$, for all $x, y \in \mathbf{R}^n$.

1.2 Vector p-norms

A vector p-norm denoted $\|\cdot\|_p$ is defined as,

$$\|\mathbf{x}\|_p := (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{\frac{1}{p}} = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}, \quad p \geq 1, \mathbf{x} \in \mathbf{R}^n.$$

The most commonly used vector p-norms are ℓ_1-norm, ℓ_2-norm, and ℓ_∞-norm.

$$\|\mathbf{x}\|_1 := |x_1| + |x_2| + \cdots + |x_n| = \sum_{i=1}^{n} |x_i|$$

$$\|\mathbf{x}\|_2 := \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{x^T x}$$

$$\|\mathbf{x}\|_\infty := \max_{1 \geq i \geq n} |x_i|.$$

where the $|\cdot|$ sign is not omitted in the definition of $|\cdot|_2$ to emphasize the importance of the $|\cdot|$ operation in the calculations of all p-norms. $\|\mathbf{x}\|_2$ is also known as Euclidean norm.

1.3 Dual norm

Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$. The associated dual norm, denoted $\|\cdot\|_*$, is defined as

$$\|z\|_* = \sup\{z^T x \mid \|x\| \leq 1\}.$$

It is easy to show that the dual norm satisfies all properties of a norm, so the dual norm is a norm.

Proposition 1 (Inner product, norm, and dual norm). *Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$. The associated dual norm $\|z\|_*$ satisfies*

$$z^T x \leq \|x\| \|z\|_*,$$

where $x, z \in \mathbf{R}^n$.

Proof. From the definition of dual norm, we have

$$z^T x \leq \|z\|_*$$

with all x satisfying $\|x\| \leq 1$. So the inequality also holds for $\|x\| = 1$,

$$z^T x \leq \|z\|_* = \|x\| \|z\|_*.$$

Let $x = ty$ with $t > 0$. Then we get

$$z^T (ty) \leq \|ty\| \|z\|_* = t\|y\| \|z\|_* \Longleftrightarrow z^T y \leq \|y\| \|z\|_*,$$

as desired. Note that there is no requirement on the value of $\|y\|$. □

Proposition 2 (Dual norm of Euclidean norm is Euclidean norm). *The dual norm of Euclidean norm is Euclidean norm, i.e.,*

$$\|z\|_2 = \sup\{z^T x \mid \|x\|_2 \leq 1\}.$$

Proof. According to Cauchy-Schwarz inequality, we get

$$z^T y \leq \|y\|_2 \|z\|_2.$$

Given z, the equality holds if and only if $y = z$. If $\|y\|_2$ is required to be not greater than 1, then $y = \frac{z}{\|z\|_2}$ which maximizes $z^T y$ and the maximum is $z^T \frac{z}{\|z\|_2} = \|z\|_2$. Thus, $\|z\|_2 = \sup\{z^T x \mid \|x\|_2 \leq 1\}$, which is exactly the definition of dual norm. □

Proposition 3 (Dual norm of ℓ_∞-norm is ℓ_1-norm). *The dual norm of the ℓ_∞-norm is the ℓ_1-norm.*

Proof. Since $\|x\|_\infty \leq 1$, $x_i = \mathrm{sgn}(z_i)$ for each $i \in \{1, \ldots, n\}$ can maximize $z^T x$, where sgn(t) is the sign function which outputs 1 for positive inputs, -1 for negative inputs, and 0 for zero inputs, respectively. Thus,

$$\sup\{z^T x \mid \|x\|_\infty \leq 1\} = \sum_{i=1}^{n} |z_i| = \|z\|_1,$$

as desired. □

Proposition 4 (Dual norm of ℓ_1-norm is ℓ_∞-norm). *The dual norm of the ℓ_1-norm is the ℓ_∞-norm.*

Proof. We find the maximum of all $|z_i|, \forall i \in \{1, \ldots, n\}$ and denote it by $|z_j|$. Then we let $x_j = 1$ and $x_i = 0$ with $i \neq j$, which satisfies the requirement of $\|x\|_1 \leq 1$. Then,

$$\sup\{z^T x \mid \|x\|_1 \leq 1\} = |z_j| = \|z\|_\infty,$$

as desired. □

Proposition 5 (Dual norm of ℓ_p-norm is ℓ_q-norm). *The dual norm of the ℓ_p-norm is the ℓ_q-norm, where $p, q \geq 1$ and $1/p + 1/q = 1$. That is,*

$$z^T x \leq \|x\|_p \|z\|_q.$$

This proposition is similar to Hölder inequality.

2 Quadratic mean, average mean, geometric mean, and harmonic mean

The content in this section is largely taken from Xicheng Peng et al. Exploring Inequalities. 2016. In this section, we give the relationships between quadratic mean (QM), average mean (AM), geometric mean (GM), and harmonic mean (HM). The result is QM≥AM≥GM≥HM.

Theorem 6. *For any positive real number $a_1, a_2, \cdots, a_n$, we have the following inequalities*

$$\overbrace{\frac{n}{\frac{1}{a_1} + \frac{1}{a_1} + \cdots + \frac{1}{a_n}}}^{\text{Harmonic Mean}} \leq \overbrace{\sqrt[n]{a_1 a_2 \cdots a_n}}^{\text{Geometric Mean}} \leq \overbrace{\frac{a_1 + a_2 + \cdots + a_n}{n}}^{\text{Arithmetic Mean}} \leq \overbrace{\sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}}}^{\text{Quadratic Mean}} \tag{1}$$

where the equalities hold if and only if $a_1 = a_2 = \cdots = a_n$. This inequality chain is also denoted as $H(n) \leq G(n) \leq A(n) \leq Q(n)$.

We will provide two proofs. The first proof employs Jensen's inequality and is simpler compared with the second proof.

Proof. Let $f(x) = -\ln x$, then $f''(x) = \frac{1}{x^2} > 0$ which implies that f is convex. For all $a_i > 0$, by Jensen's inequality,

$$
\begin{aligned}
f(\frac{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}{n}) &\leq \frac{1}{n}\left(f(\frac{1}{a_1}) + f(\frac{1}{a_2}) + \cdots + f(\frac{1}{a_n})\right) \\
-\ln(\frac{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}{n}) &\leq \frac{1}{n}\left(-\ln(\frac{1}{a_1}) - \ln(\frac{1}{a_2}) + \cdots - \ln(\frac{1}{a_n})\right) \\
\ln(\frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}) &\leq \frac{1}{n}(\ln a_1 + \ln a_2 + \cdots + \ln a_n) \\
\ln(\frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}}) &\leq \ln \sqrt[n]{a_1 a_2 \cdots a_n}
\end{aligned}
$$

Since ln is increasing,

$$\frac{n}{\frac{1}{a_1} + \frac{1}{a_2} + \cdots + \frac{1}{a_n}} \leq \sqrt[n]{a_1 a_2 \cdots a_n}$$

where the equality holds if and only if $a_1 = a_2 = \cdots = a_n$. The "if" part is obvious. For the "only if" part, suppose the equality holds with $a_1 \neq a_2 = a_3 = \cdots = a_n$, then let $a_1 = ka_2 (k > 0)$ we have

$$\frac{n}{(n - 1 + \frac{1}{k})\frac{1}{a_2}} = a_2 \sqrt[n]{k} \iff \frac{n}{(n - 1 + \frac{1}{k})} = \sqrt[n]{k} \iff n = \sqrt[n]{k}(n - 1 + \frac{1}{k})$$

Let $x = \sqrt[n]{k}$ and it is clear that $x > 0$. Then

$$nx^{n-1} = (n-1)x^n + 1 \iff (n-1)x^n - nx^{n-1} + 1 = 0$$

$$
\begin{aligned}
(n-1)x^n - nx^{n-1} + 1 &= (n-1)x^n - (n-1)x^{n-1} - x^{n-1} + 1 \\
&= (n-1)x^{n-1}(x-1) - (x^{n-1} - 1) \\
&= (n-1)x^{n-1}(x-1) - (x-1)(x^{n-2} + x^{n-3} + \cdots + 1) \\
&= (x-1)[(n-1)x^{n-1} - (x^{n-2} + x^{n-3} + \cdots + 1)] \\
&= (x-1)[(n-1)x^{n-1} - (n-1)x^{n-2} + (n-2)x^{n-2} - x^{n-3} - \cdots - 1] \\
&= (x-1)[(n-1)x^{n-2}(x-1) + (n-2)x^{n-2} - x^{n-3} - \cdots - 1] \\
&= (x-1)[(n-1)x^{n-2}(x-1) + (n-2)x^{n-2} - (n-2)x^{n-3} + (n-3)x^{n-3} - \cdots - 1] \\
&= (x-1)[(n-1)x^{n-2}(x-1) + (n-2)x^{n-3}(x-1) + (n-3)x^{n-3} - \cdots - 1] \\
&= (x-1)[(n-1)x^{n-2}(x-1) + (n-2)x^{n-3}(x-1) + (n-3)x^{n-4}(x-1) + \cdots + (x-1)] \\
&= (x-1)^2[(n-1)x^{n-2} + (n-2)x^{n-3} + (n-3)x^{n-4} + \cdots + 1]
\end{aligned}
$$

Due to the fact that any polynomial that has positive coefficients cannot have roots on the nonnegative real axis[1], 1 is the only roots which contradicts the supposition that $a_1 \neq a_2$. Hence, $a_1 = a_2 = \cdots = a_n$ is the necessary and sufficient condition for the equality to hold.

Now we show GM≤AM,

$$
\begin{aligned}
f(\frac{a_1 + a_2 + \cdots + a_n}{n}) &\leq \frac{1}{n}\left(f(a_1) + f(a_2) + \cdots + f(a_n)\right) \\
-\ln(\frac{a_1 + a_2 + \cdots + a_n}{n}) &\leq \frac{1}{n}\left(-\ln(a_1) - \ln(a_1) + \cdots - \ln(a_1)\right)
\end{aligned}
$$

[1] `https://mtns2018.hkust.edu.hk/media/files/0073.pdf`. Besides that paper, we can also prove this fact via contradiction. Suppose this kind of polynomial $P(x)$ has some positive real roots, say $P(x_0) = 0$, then each term of $P(x_0)$ is positive which leads to $P(x_0) > 0$, a contradiction. Thus, x_0 does not exist.

$$\ln(\frac{a_1+a_2+\cdots+a_n}{n}) \geq \frac{1}{n}(\ln a_1+\ln a_2+\cdots+\ln a_n)$$
$$\ln(\frac{a_1+a_2+\cdots+a_n}{n}) \geq \ln \sqrt[n]{a_1a_2\cdots a_n}$$
$$\frac{a_1+a_2+\cdots+a_n}{n} \geq \sqrt[n]{a_1a_2\cdots a_n}$$

Let $g(x)=x^2$, then $g''(x)=2>$ which indicates that g is convex.

Now we show AM≤QM. By Jensen's inequality, we have

$$g(\frac{a_1+a_2+\cdots+a_n}{n}) \leq \frac{1}{n}\left(g(a_1)+g(a_2)+\cdots+g(a_n)\right)$$
$$(\frac{a_1+a_2+\cdots+a_n}{n})^2 \leq \frac{1}{n}(a_1^2+a_2^2+\cdots+a_n^2)$$
$$\frac{a_1+a_2+\cdots+a_n}{n} \leq \sqrt{\frac{a_1^2+a_2^2+\cdots+a_n^2}{n}}$$

This completes our proof. □

We first introduce a lemma for the second proof.

Lemma 7. *If $a_i>0, i=1,2,\cdots,n$, and $a_1a_2\cdots a_n=1$, then $a_1+a_2+\cdots+a_n \geq n$ where the equality holds iff $a_1=a_2=\cdots=a_n=1$.*

Proof. Let $x_i=\ln a_i, i=1,2,\cdots,n$.

$$a_1a_2\cdots a_n=1 \iff \ln a_1+\ln a_2+\cdots+\ln a_n=0 \iff x_1+x_2+\cdots+x_n=0$$

Since $e^x \geq x+1$,

$$a_1+a_2+\cdots+a_n=e^{x_1}+e^{x_2}+\cdots+e^{x_n} \geq (x_1+1)+(x_2+1)+\cdots+(x_n+1)=n$$

where the equality holds iff $x_i=0, \forall i=1,2,\cdots,n$, i.e., $a_i=1, \forall i=1,2,\cdots,n$ due to the fact that $e^x=x+1$ iff $x=0$. Thus, $a_1+a_2+\cdots+a_n=n$ iff $a_1=a_2=\cdots=a_n=1$. □

Next, we prove this inequality chain using Lemma 7.

Proof. 1. $H(n) \leq G(n)$

$$\frac{n}{\frac{1}{a_1}+\frac{1}{a_2}+\cdots+\frac{1}{a_n}} \leq \sqrt[n]{a_1a_2\cdots a_n} \iff \frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_1}+\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_2}+\cdots+\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_n} \geq n$$

We observe that $\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_1}\cdot\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_2}\cdot\cdots\cdot\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_n}=1$. By Lemma 7, the above holds. The equality holds iff $\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_1}=\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_2}=\cdots=\frac{\sqrt[n]{a_1a_2\cdots a_n}}{a_n}$, i.e., $a_1=a_2=\cdots=a_n$.

2. $G(n) \leq A(n)$

$$\sqrt[n]{a_1a_2\cdots a_n} \leq \frac{a_1+a_2+\cdots+a_n}{n} \iff \frac{a_1}{\sqrt[n]{a_1a_2\cdots a_n}}+\frac{a_2}{\sqrt[n]{a_1a_2\cdots a_n}}+\cdots+\frac{a_n}{\sqrt[n]{a_1a_2\cdots a_n}} \geq n$$

We observe that $\frac{a_1}{\sqrt[n]{a_1a_2\cdots a_n}}\cdot\frac{a_2}{\sqrt[n]{a_1a_2\cdots a_n}}\cdot\cdots\cdot\frac{a_n}{\sqrt[n]{a_1a_2\cdots a_n}}=1$. By Lemma 7, the above holds. The equality holds iff $\frac{a_1}{\sqrt[n]{a_1a_2\cdots a_n}}=\frac{a_2}{\sqrt[n]{a_1a_2\cdots a_n}}=\cdots=\frac{a_n}{\sqrt[n]{a_1a_2\cdots a_n}}$, i.e., $a_1=a_2=\cdots=a_n$.

3. $A(n) \leq Q(n)$. Let $c = \frac{a_1+a_2+\cdots+a_n}{n}$ and $a_i = c + \alpha_i, \forall i = 1$. Then

$$\begin{aligned} a_1 + a_2 + \cdots + a_n &= nc + \alpha_1 + \alpha_2 + \cdots + \alpha_n \\ &= a_1 + a_2 + \cdots + a_n + (\alpha_1 + \alpha_2 + \cdots + \alpha_n) \end{aligned}$$

So, $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 0$.

$$\begin{aligned} \sqrt{\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n}} &= \sqrt{\frac{(c+\alpha_1)^2 + (c+\alpha_2)^2 + \cdots + (c+\alpha_n)^2}{n}} \\ &= \sqrt{\frac{nc^2 + 2c(\alpha_1 + \alpha_2 + \cdots + \alpha_n) + \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2}{n}} \\ &= \sqrt{\frac{nc^2 + \alpha_2^2 + \cdots + \alpha_n^2}{n}} \\ &= \sqrt{c^2 + \frac{\alpha_2^2 + \cdots + \alpha_n^2}{n}} \\ &\geq \sqrt{c^2} = c \end{aligned}$$

where the third equality follows from $\alpha_1 + \alpha_2 + \cdots + \alpha_n = 0$ and $c = \frac{a_1+a_2+\cdots+a_n}{n}$.

Lemma 7 is not involved with the proof of $A(n) \leq Q(n)$. □

Note. AM≥GM≥HM can be proved using induction; see Theorem 1.2.2 of Jixiu Chen et al. Mathematical Analysis, third edition.

2.1 An inequality about the difference between AM and GM

Proposition 8. *Given $a, b, c \in \mathbf{R}_+$, show* $3(\frac{a+b+c}{3} - \sqrt[3]{abc}) \geq 2(\frac{a+b}{2} - \sqrt{ab})$.

Proof. After simple algebra, it suffices to show

$$c + 2\sqrt{ab} \geq 3\sqrt[3]{abc}$$

If we consider $2\sqrt{ab}$ as $\sqrt{ab} + \sqrt{ab}$ on the LHS, we can use AM-GM to get

$$c + \sqrt{ab} + \sqrt{ab} \geq 3\sqrt[3]{\sqrt{ab} \cdot \sqrt{ab} \cdot c} = 3\sqrt[3]{abc}.$$

This completes our proof. □

This result can be easily generalized to the general case, namely,

$$(n+1)(\frac{a_1^2 + a_2^2 + \cdots + a_{n+1}^2}{n+1} - \sqrt[n+1]{a_1 a_2 \cdots a_{n+1}}) \geq n(\frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n} - \sqrt[n]{a_1 a_2 \cdots a_n}).$$

2.2 Power mean inequality

Refer to Page 106 of Xicheng Peng et al. Exploring Inequalities. 2016. This will be done later.

2.3 Weighted power mean inequality

Refer to Page 108 of Xicheng Peng et al. Exploring Inequalities. 2016. This will be done later.

2.4 Applications

Proposition 9. *Given positive reals a, b, c and $a+b+c=1$, show that $ab+bc+ca \leq \frac{1}{3}$.*

Proof. Since the geometric mean is not less than the average mean, we have

$$\sqrt{\frac{a^2+b^2+c^2}{3}} \geq \frac{a+b+c}{3} = \frac{1}{3} \Longleftrightarrow a^2+b^2+c^2 \geq \frac{1}{3}.$$

Furthermore, we have

$$ab+bc+ca = \frac{(a+b+c)^2-(a^2+b^2+c^2)}{2} = \frac{1-(a^2+b^2+c^2)}{2} \leq \frac{1-1/3}{2} = \frac{1}{3},$$

as desired. □

3 Young inequality

Theorem 10. *Given $x, y \geq 0$, $p, q \geq 1$, and $1/p + 1/q = 1$, the inequality*

$$xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q,$$

where the equality holds if and only if $x^p = y^q$.

Proof. The claim is obvious for the case when either $x=0$ or $y=0$. When x and y are positive reals, we let $f(t) = e^x$, then $f'' > 0$. So f is convex. By the Jensen's inequality,

$$f(\frac{1}{p}\ln x^p + \frac{1}{q}\ln y^q) \leq \frac{1}{p}f(\ln x^p) + \frac{1}{q}f(\ln y^q)$$
$$\Downarrow$$
$$e^{\frac{1}{p}\ln x^p + \frac{1}{q}\ln y^q} \leq \frac{1}{p}e^{\ln x^p} + \frac{1}{q}e^{\ln y^q} \Longleftrightarrow e^{\frac{1}{p}\ln x^p}e^{\frac{1}{q}\ln y^q} \leq \frac{1}{p}x^p + \frac{1}{q}y^q \Longleftrightarrow xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q,$$

as desired. The equality follows from the condition for the equality of Jensen's inequality to hold with any convex function. □

4 Hölder inequality

A classic result concerning p-norms is the **Hölder inequality in inner-product form**:

Theorem 11 (Hölder inequality). *For any $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ and $p, q > 1$ satisfying $1/p + 1/q = 1$, the following*

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$$

holds. The equality holds if and only if $\frac{|x_1|^p}{|y_1|^q} = \frac{|x_2|^p}{|y_2|^q} = \cdots = \frac{|x_n|^p}{|y_n|^q}$ holds.

We provide two proofs here. The first proof employs Young inequality and the second proof makes use of the fact that the dual norm of p-norm is q-norm with $p, q \geq 0$ and $1/p + 1/q = 1$.

Proof. The claim is trivial either $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} = \mathbf{0}$. Suppose that $\mathbf{x} \neq \mathbf{0}$ or $\mathbf{y} \neq \mathbf{0}$. For any $i \in \{1, 2, \ldots, n\}$, letting $s = \frac{|x_i|}{(\sum_{i=1}^n |x_i|^p)^{1/p}}$ and $t = \frac{|y_i|}{(\sum_{i=1}^n |y_i|^q)^{1/q}}$, by Young's inequality, we have

$$st = \frac{|x_i|}{(\sum_{i=1}^n |x_i|^p)^{1/p}} \frac{|y_i|}{(\sum_{i=1}^n |y_i|^q)^{1/q}} \leq \frac{1}{p}\left(\frac{|x_i|}{(\sum_{i=1}^n |x_i|^p)^{1/p}}\right)^p + \frac{1}{q}\left(\frac{|y_i|}{(\sum_{i=1}^n |y_i|^q)^{1/q}}\right)^q$$

$$\Updownarrow$$

$$\frac{|x_i y_i|}{(\sum_{i=1}^n |x_i|^p)^{1/p} (\sum_{i=1}^n |y_i|^q)^{1/q}} \le \frac{1}{p} \frac{|x_i|^p}{\sum_{i=1}^n |x_i|^p} + \frac{1}{q} \frac{|y_i|^p}{\sum_{i=1}^n |y_i|^q}$$

Summing up over i,

$$\frac{\sum_{i=1}^n |x_i y_i|}{(\sum_{i=1}^n |x_i|^p)^{1/p} (\sum_{i=1}^n |y_i|^q)^{1/q}} \le \frac{1}{p} \frac{\sum_{i=1}^n |x_i|^p}{\sum_{i=1}^n |x_i|^p} + \frac{1}{q} \frac{\sum_{i=1}^n |y_i|^p}{\sum_{i=1}^n |y_i|^q}$$

$$\Updownarrow$$

$$\frac{\sum_{i=1}^n |x_i y_i|}{(\sum_{i=1}^n |x_i|^p)^{1/p} (\sum_{i=1}^n |y_i|^q)^{1/q}} \le \frac{1}{p} + \frac{1}{q} = 1$$

$$\Updownarrow$$

$$|\mathbf{x}^T \mathbf{y}| \le \sum_{i=1}^n |x_i y_i| \le (\sum_{i=1}^n |x_i|^p)^{1/p} (\sum_{i=1}^n |y_i|^q)^{1/q}, \tag{2}$$

as desired. □

Now we present the second proof.

Proof. Recall the definition of dual norm,

$$\|x\|_p = \max_{\|z\|_q \le 1} x^T z.$$

Thus, $x^T z \le \|x\|_p$ holds for any z satisfying $\|z\|_q \le 1$, including $\|z\|_q = 1$. When $\|z\|_q = 1$, we have

$$x^T z \le \|x\|_p \|z\|_q$$

Now let $z = ty$ with $t > 0$. Thus, we have

$$x^T (ty) \le \|x\|_p \|ty\|_q \Longleftrightarrow x^T y \le \|x\|_p \|y\|_q$$

where $\|y\|_q = \|z/t\|_q = 1/t \|z\|_q = 1/t > 0$. This completes the proof. □

The key part of the above proof is using the dual representation of the ℓ_p norm, namely, $\|x\|_p = \max_{\|z\|_q \le 1} x^T z$.

4.1 Generalized Hölder inequality

Setting $\lambda_1 = 1/p$ and $\lambda_2 = 1/q$ in (2) yields

$$(\sum_{i=1}^n |x_i|^{1/\lambda_1})^{\lambda_1} (\sum_{i=1}^n |y_i|^{1/\lambda_2})^{\lambda_2} \ge \sum_{i=1}^n (|x_i|^{1/\lambda_1})^{\lambda_1} (|y_i|^{1/\lambda_1})^{\lambda_1},$$

where $0 < \lambda_1, \lambda_2 < 1$ and $\lambda_1 + \lambda_2 = 1$. Continuing on this notational trick by setting $a_{i1} = |x_i|^{1/\lambda_1}$ and $a_{i2} = |x_i|^{1/\lambda_2}$ for any $i \in \{1, 2, \ldots, n\}$ gives

$$(\sum_{i=1}^n a_{i1})^{\lambda_1} (\sum_{i=1}^n a_{i2})^{\lambda_2} \ge \sum_{i=1}^n (a_{i1})^{\lambda_1} (a_{i2})^{\lambda_2}.$$

where $a_{i1}, a_{i2} \ge 0$ for any $i \in \{1, 2, \ldots, n\}$. This form is beautiful. If it can be generalized to λ_m ($m > 2$), it will be wonderful. Yes, it is. We formalize it into the following theorem.

Theorem 12 (Generalized Hölder inequality). *Given a matrix $A \in \mathbf{R}^{n\times m}$ with all nonnegative entries a_{ij} and $0 < \lambda_j < 1$ satisfying $\lambda_1 + \lambda_2 + \cdots + \lambda_m = 1$, it holds that*

$$\begin{aligned}&(\sum_{i=1}^{n} a_{i1})^{\lambda_1}(\sum_{i=1}^{n} a_{i2})^{\lambda_2}\cdots(\sum_{i=1}^{n} a_{im})^{\lambda_m}\\ \geq& {a_{11}}^{\lambda_1}{a_{12}}^{\lambda_2}\cdots {a_{1m}}^{\lambda_m} + {a_{21}}^{\lambda_1}{a_{22}}^{\lambda_2}\cdots {a_{2m}}^{\lambda_m} + \cdots + {a_{n1}}^{\lambda_1}{a_{n2}}^{\lambda_2}\cdots {a_{nm}}^{\lambda_m}\\ =&\sum_{i=1}^{n} {a_{i1}}^{\lambda_1}{a_{i2}}^{\lambda_2}\cdots {a_{im}}^{\lambda_m}\end{aligned}$$

Compactly,

$$\prod_{j=1}^{m}(\sum_{i=1}^{n} a_{ij})^{\lambda_j} \geq \sum_{i=1}^{n}\prod_{j=1}^{m} {a_{ij}}^{\lambda_j}.$$

Proof. We prove this generalized Hölder inequality by mathematical induction. When $m = 2$, it is exactly Theorem 11, i.e., the canonical Hölder inequality. Suppose the claim holds when $m = k$. When $m = k + 1$, let $\lambda_1 + \lambda_2 + \cdots + \lambda_k = s$ and denote $t_i = \lambda_i/s, i = 1, 2, \ldots, k$. Then

$$\begin{aligned}\sum_{i=1}^{n} {a_{i1}}^{\lambda_1}{a_{i2}}^{\lambda_2}\cdots {a_{im}}^{\lambda_m} &= \sum_{i=1}^{n}({a_{i1}}^{t_1}{a_{i2}}^{t_2}\cdots {a_{ik}}^{t_k})^s {a_{i,k+1}}^{\lambda_{k+1}}\\ &\leq \left(\sum_{i=1}^{n}\left(({a_{i1}}^{t_1}{a_{i2}}^{t_2}\cdots {a_{ik}}^{t_k})^s\right)^{\frac{1}{s}}\right)^s \left(\sum_{i=1}^{n}({a_{i,k+1}}^{\lambda_{k+1}})^{\frac{1}{\lambda_{k+1}}}\right)^{\lambda_{k+1}}\\ &= \left(\sum_{i=1}^{n} {a_{i1}}^{t_1}{a_{i2}}^{t_2}\cdots {a_{ik}}^{t_k}\right)^s \left(\sum_{i=1}^{n} a_{i,k+1}\right)^{\lambda_{k+1}}\\ &\leq \left(\left(\sum_{i=1}^{n} a_{i1}\right)^{t_1}\left(\sum_{i=1}^{n} a_{i2}\right)^{t_2}\cdots\left(\sum_{i=1}^{n} a_{ik}\right)^{t_k}\right)^s \left(\sum_{i=1}^{n} a_{i,k+1}\right)^{\lambda_{k+1}}\\ &= \left(\sum_{i=1}^{n} a_{i1}\right)^{\lambda_1}\left(\sum_{i=1}^{n} a_{i2}\right)^{\lambda_2}\cdots\left(\sum_{i=1}^{n} a_{ik}\right)^{\lambda_k}\left(\sum_{i=1}^{n} a_{i,k+1}\right)^{\lambda_{k+1}}\end{aligned}$$

where the first inequality and the second inequality follow from the classic Hölder inequality and the induction assumption, respectively. This completes the proof. □

The generalized Hölder inequality can also be applied to the case where $\lambda_1 + \lambda_2 + \cdots + \lambda_m < 1$. For this, we have the following corollary.

Corollary 13. *Given a matrix $A \in \mathbf{R}^{n\times m}$ with all nonnegative entries a_{ij} and $0 < \lambda_j < 1$ satisfying $\lambda_1 + \lambda_2 + \cdots + \lambda_m = r < 1$, it holds that*

$$\begin{aligned}&(\sum_{i=1}^{n} a_{i1})^{\lambda_1}(\sum_{i=1}^{n} a_{i2})^{\lambda_2}\cdots(\sum_{i=1}^{n} a_{im})^{\lambda_m}\\ \geq& n^{r-1}{a_{11}}^{\lambda_1}{a_{12}}^{\lambda_2}\cdots {a_{1m}}^{\lambda_m} + n^{r-1}{a_{21}}^{\lambda_1}{a_{22}}^{\lambda_2}\cdots n^{r-1}{a_{2m}}^{\lambda_m} + \cdots + n^{r-1}{a_{n1}}^{\lambda_1}{a_{n2}}^{\lambda_2}\cdots {a_{nm}}^{\lambda_m}\\ =& n^{r-1}\sum_{i=1}^{n} {a_{i1}}^{\lambda_1}{a_{i2}}^{\lambda_2}\cdots {a_{im}}^{\lambda_m}\end{aligned}$$

Compactly,

$$\prod_{j=1}^{m}(\sum_{i=1}^{n} a_{ij})^{\lambda_j} \geq n^{r-1}\sum_{i=1}^{n}\prod_{j=1}^{m} {a_{ij}}^{\lambda_j}.$$

Proof. By the generalized Hölder inequality, letting $\alpha = 1 - r$ yields

$$\begin{aligned}
&(1+1+\cdots+1)^{\alpha}(\sum_{i=1}^{n} a_{i1})^{\lambda_1}(\sum_{i=1}^{n} a_{i2})^{\lambda_2}\cdots(\sum_{i=1}^{n} a_{im})^{\lambda_m} \\
\geq& {a_{11}}^{\lambda_1}{a_{12}}^{\lambda_2}\cdots{a_{1m}}^{\lambda_m} + {a_{21}}^{\lambda_1}{a_{22}}^{\lambda_2}\cdots{a_{2m}}^{\lambda_m} + \cdots + {a_{n1}}^{\lambda_1}{a_{n2}}^{\lambda_2}\cdots{a_{nm}}^{\lambda_m} \\
=& \sum_{i=1}^{n} {a_{i1}}^{\lambda_1}{a_{i2}}^{\lambda_2}\cdots{a_{im}}^{\lambda_m}.
\end{aligned}$$

Equivalently,

$$(\sum_{i=1}^{n} a_{i1})^{\lambda_1}(\sum_{i=1}^{n} a_{i2})^{\lambda_2}\cdots(\sum_{i=1}^{n} a_{im})^{\lambda_m} \geq n^{r-1}\sum_{i=1}^{n} {a_{i1}}^{\lambda_1}{a_{i2}}^{\lambda_2}\cdots{a_{im}}^{\lambda_m}$$

□

4.2 Cauchy-Schwarz inequality

A very important special case of Hölder inequality is **Cauchy-Schwarz inequality**:

$$|x^T y| \leq \|x\|_2\|y\|_2.$$

which can also be expressed as $(x^T y)^2 \leq \|x\|_2^2\|y\|_2^2$. Here, $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$. Thus,

$$(x_1y_1 + x_2y_2 + \ldots + x_ny_n)^2 \leq (x_1^2 + x_2^2 + \ldots + x_n^2)(y_1^2 + y_2^2 + \ldots + y_n^2)$$
$$(\sum_{i=1}^{n} x_iy_i)^2 = \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2$$

4.3 Two variants of Cauchy-Schwarz inequality

Given $a_i \in \mathbf{R}$ and $b_i > 0 (i = 1, 2, \cdots, n)$, let $y_i = \sqrt{b_i}$ and $x_i = \frac{a_i}{\sqrt{b_i}}$. Then we have an important **variant** of Cauchy-Schwarz inequality.

$$(\sum_{i=1}^{n} \frac{a_i}{\sqrt{b_i}}\sqrt{b_i})^2 = (\sum_{i=1}^{n} a_i)^2 \leq (\sum_{i=1}^{n} \frac{a_i^2}{b_i})\sum_{i=1}^{n} b_i \Longleftrightarrow \sum_{i=1}^{n} \frac{a_i^2}{b_i} \geq \frac{(\sum_{i=1}^{n} a_i)^2}{\sum_{i=1}^{n} b_i}.$$

For $a_i > 0$ and $b_i > 0 (i = 1, 2, \cdots, n)$, letting $x_i = \sqrt{\frac{a_i}{b_i}}$ and $y_i = \sqrt{a_ib_i}$ gives

$$\sum_{i=1}^{n} x_i^2 \cdot \sum_{i=1}^{n} y_i^2 \geq (\sum_{i=1}^{n} x_iy_i)^2 \Longleftrightarrow \sum_{i=1}^{n} \frac{a_i}{b_i} \cdot \sum_{i=1}^{n} a_ib_i \geq (\sum_{i=1}^{n} a_i)^2 \Longleftrightarrow \sum_{i=1}^{n} \frac{a_i}{b_i} \geq \frac{(\sum_{i=1}^{n} a_i)^2}{\sum_{i=1}^{n} a_ib_i}.$$

5 Minkowski inequality

Theorem 14 (Minkowski inequality). *Given $x_i, y_i \geq 0, i = 1, 2, \ldots, n$ and $p \geq 1$, the following*

$$(\sum_{i=1}^{n} |x_i + y_i|^p)^{\frac{1}{p}} \leq (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^{n} |y_i|^p)^{\frac{1}{p}}$$

holds. By the definition of p-norm (see section 1.2), the equivalent vector form

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$$

holds.

Since the following proof employs Hölder inequality in which $1/p + 1/q = 1$ is required and $q \to +\infty$ as $p \to 1$, we need to consider the case of $p = 1$ separately. The triangle inequality is enough to show the case of $p = 1$.

Proof. We observe that when either $x_i = 0$ or $y_i = 0$, or $x_i + y_i = 0, i = 1, 2, \ldots, n$, the result is trivial. Now we show the case when $\sum_{i=1}^n x_i + y_i \neq 0$.

We first prove the case of $p = 1$. By the triangle inequality, we have

$$\sum_{i=1}^n |x_i + y_i| \leq \sum_{i=1}^n (|x_i| + |y_i|) = \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| \Longleftrightarrow \|\mathbf{x} + \mathbf{y}\|_1 \leq \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1.$$

Now we turn to the case of $p > 1$.

$$\sum_{i=1}^n |x_i + y_i|^p = \sum_{i=1}^n |x_i + y_i||x_i + y_i|^{p-1} \leq \sum_{i=1}^n |x_i||x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i||x_i + y_i|^{p-1},$$

where the last inequality follows from the triangle inequality.

Applying Hölder inequality on both terms of the RHS gives

$$\sum_{i=1}^n |x_i||x_i + y_i|^{p-1} \leq (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |x_i + y_i|^{q(p-1)})^{\frac{1}{q}} = (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |x_i + y_i|^p)^{\frac{1}{q}}$$
$$\sum_{i=1}^n y_i |x_i + y_i|^{p-1} \leq (\sum_{i=1}^n |y_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |x_i + y_i|^{q(p-1)})^{\frac{1}{q}} = (\sum_{i=1}^n |y_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |x_i + y_i|^p)^{\frac{1}{q}}$$

where the equalities follow from the fact that $q(p-1) = p$ due to Hölder inequality's condition, i.e., $1/p + 1/q = 1$. Adding the above two inequalities up yields,

$$\sum_{i=1}^n |x_i + y_i|^p \leq (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |x_i + y_i|^p)^{\frac{1}{q}} + (\sum_{i=1}^n |y_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |x_i + y_i|^p)^{\frac{1}{q}}$$
$$\Updownarrow$$
$$\frac{\sum_{i=1}^n |x_i + y_i|^p}{(\sum_{i=1}^n |x_i + y_i|^p)^{\frac{1}{q}}} \leq (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^n |y_i|^p)^{\frac{1}{p}}$$
$$\Updownarrow$$
$$(\sum_{i=1}^n |x_i + y_i|^p)^{1-\frac{1}{q}} \leq (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^n |y_i|^p)^{\frac{1}{p}} \Longleftrightarrow (\sum_{i=1}^n |x_i + y_i|^p)^{\frac{1}{p}} \leq (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} + (\sum_{i=1}^n |y_i|^p)^{\frac{1}{p}},$$

as desired. □

6 Rearrangement inequalities

For two sequences $a_1 \leq a_2 \leq \cdots \leq a_n$ and $b_1 \leq b_2 \leq \cdots \leq b_n$ with reals a_i and b_i, the following inequalities

$$\underbrace{a_1 b_n + a_2 b_{n-1} + \cdots + a_n b_1}_{\text{reverse sum}} \leq \underbrace{a_1 b_{\pi(1)} + a_2 b_{\pi(2)} + \cdots + a_n b_{\pi(n)}}_{\text{disordered sum}} \leq \underbrace{a_1 b_1 + a_2 b_2 + \cdots + a_n b_n}_{\text{sequential sum}}$$

hold, where $\pi(1), \pi(2), \ldots, \pi(n)$ is any permutation of $1, 2, \ldots, n$. Simply speaking, **reverse sum $\leq$ disordered sum $\leq$ sequential sum**.

Proof. [2] We want to prove that the identity permutation maximizes $a_1 b_{\pi(1)} + a_2 b_{\pi(2)} + \cdots + a_n b_{\pi(n)}$. Suppose for the sake of contradiction $\pi(i)$ is the smallest integer such that $\pi(i) \neq i$, then $\pi(i) = j (j > i)$ (since $1, \ldots, i-1$ have been assigned). Meanwhile, there exists a number $k > i$ such that $\pi(k) = i$ as some number must be assigned to i.

Now, since $i < j$, it follows that $b_i \leq b_j$. Likewise, since $i < k$, it follows that $a_i \leq a_k$. Thus,

$$(a_k - a_i)(b_j - b_i) \geq 0 \Longrightarrow a_k b_j + a_i b_i \geq a_i b_j + a_k b_i,$$

which demonstrates that the sum $a_1 b_{\pi(1)} + a_2 b_{\pi(2)} + \cdots + a_n b_{\pi(n)}$ is not decreased by changing $\pi(j) = i$ and $\pi(k) = i$ to $\pi(i) = i$ and $\pi(k) = j$. This implies the identity permutation gives the maximum possible value of the sum $a_1 b_{\pi(1)} + a_2 b_{\pi(2)} + \cdots + a_n b_{\pi(n)}$, as desired. Moreover, the reverse identity permutation gives the minimum possible value of the sum $a_1 b_{\pi(1)} + a_2 b_{\pi(2)} + \cdots + a_n b_{\pi(n)}$. $\square$

6.1 Chebyshev's inequality

For two sequences $a_1 \leq a_2 \leq \cdots \leq a_n$ and $b_1 \leq b_2 \leq \cdots \leq b_n$ with reals a_i and b_i, by the rearrangement inequalities, we always have the following inequalities

$$\begin{gathered}
x_1 y_n + x_2 y_{n-1} + \cdots + x_n y_1 \leq x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \leq x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\
x_1 y_n + x_2 y_{n-1} + \cdots + x_n y_1 \leq x_1 y_2 + x_2 y_3 + \cdots + x_n y_1 \leq x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\
x_1 y_n + x_2 y_{n-1} + \cdots + x_n y_1 \leq x_1 y_3 + x_2 y_4 + \cdots + x_n y_2 \leq x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \\
\cdots \\
x_1 y_n + x_2 y_{n-1} + \cdots + x_n y_1 \leq x_1 y_n + x_2 y_1 + \cdots + x_n y_{n-1} \leq x_1 y_1 + x_2 y_2 + \cdots + x_n y_n
\end{gathered}$$

Summing up the above inequalities gives

$$\begin{gathered}
n(x_1 y_n + x_2 y_{n-1} + \cdots + x_n y_1) \leq (x_1 + x_2 + \cdots + x_n)(y_1 + y_2 + \cdots + y_n) \leq n(x_1 y_1 + x_2 y_2 + \cdots + x_n y_n) \\
\frac{1}{n}(x_1 y_n + x_2 y_{n-1} + \cdots + x_n y_1) \leq \frac{1}{n^2}(x_1 + x_2 + \cdots + x_n)(y_1 + y_2 + \cdots + y_n) \leq \frac{1}{n}(x_1 y_1 + x_2 y_2 + \cdots + x_n y_n) \\
\frac{1}{n}\sum_{i=1}^{n} x_i y_{n+1-i} \leq \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \frac{1}{n}\sum_{i=1}^{n} y_i \leq \frac{1}{n}\sum_{i=1}^{n} x_i y_i
\end{gathered}$$

which is called Chebyshev's inequality.

[2] https://brilliant.org/wiki/rearrangement-inequality/