Useful Inequalities from Jensen to Young to Hölder to Minkowski

Frank R. Kschischang
Department of Electrical & Computer Engineering
University of Toronto

September 25, 2019

Abstract

The aim of this note is to establish the triangle inequality for p-norms in $\mathbb{C}^n$, a result known as Minkowski's inequality. On the way to this result, we will establish a number of other famous inequalities.

1 The Triangle Inequality for Complex Numbers

We will start with a basic inequality for complex numbers. Throughout these notes, if $z = a + b\,\mathrm{i}$ is any complex number with $a, b \in \mathbb{R}$, we will write z^* to denote its complex conjugate $a - b\,\mathrm{i}$. Recall that for $z \in \mathbb{C}$, we have $\mathrm{Re}(z) \leq |z|$, with equality if and only if z is real-valued and non-negative.

Theorem 1 (Triangle Inequality for Complex Numbers). *Every pair z_1, z_2 of complex numbers satisfies*

$$|z_1 + z_2| \leq |z_1| + |z_2|,$$

with equality achieved if and only if $z_1 = z_2 = 0$ or (if $z_2 \neq 0$) if $z_1 = az_2$ for some non-negative real number a or (if $z_1 \neq 0$) if $z_2 = az_1$ for some non-negative real number a.

Proof. We write

$$\begin{aligned}
|z_1 + z_2|^2 &= (z_1 + z_2)(z_1 + z_2)^* \\
&= z_1 z_1^* + z_1 z_2^* + z_2 z_1^* + z_2 z_2^* \\
&= |z_1|^2 + 2\mathrm{Re}[z_1 z_2^*] + |z_2|^2 \\
&\leq |z_1|^2 + 2|z_1 z_2^*| + |z_2|^2 \\
&= |z_1|^2 + 2|z_1| \cdot |z_2| + |z_2|^2 \\
&= (|z_1| + |z_2|)^2 .
\end{aligned}$$

The result follows by taking the square-root of both sides, an operation that preserves the inequality since the square-root function is monotonically increasing. Equality is certainly achieved if $z_1 = z_2 = 0$. Otherwise, we need $\mathrm{Re}[z_1 z_2^*] = |z_1 z_2^*|$, which arises if and only if $z_1 z_2^* = b$ for some non-negative real value b. Assuming $z_2 \neq 0$, and multiplying both sides by $z_2/|z_2|^2$, we find that equality is achieved if and only if $z_1 = \frac{b}{|z_2|^2} z_2$, i.e., if and only if $z_1 = a z_2$ for some non-negative real number a. The case when $z_1 \neq 0$ is similar. $\square$

The condition for equality in this triangle inequality can be concisely expressed in terms of the *phase* or *argument* of the two complex numbers in question. Recall that any nonzero complex number z can be written in the form $re^{\mathrm{i}\theta}$ where $r = |z|$ is the *magnitude* of z and $\theta \in [0, 2\pi)$ is the *phase* (or *argument*) of z. We have $|z_1 + z_2| = |z_1| + |z_2|$ if and only if $z_1 = 0$ or $z_2 = 0$ (or both), or if z_1 and z_2 are both nonzero and have the same phase.

The triangle inequality extends to any finite number of complex numbers. In general we have, for any $z_1, z_2, \ldots, z_n \in \mathbb{C}$, that

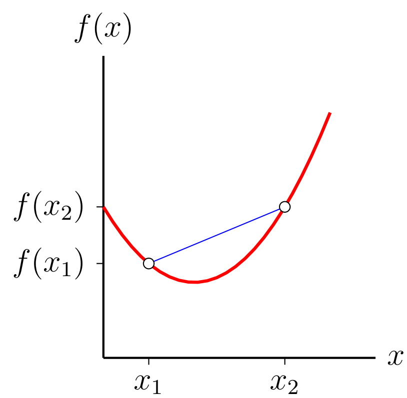
$$|z_1 + z_2 + \cdots + z_n| \leq |z_1| + |z_2| + \cdots + |z_n|$$

with equality achieved if and only if all nonzero z_i have the same phase or all z_i are zero.

2 Convex Functions and Convex Combinations

Let I be an interval of $\mathbb{R}$. A function $f : I \to \mathbb{R}$ is said to be *convex* over I if for every pair of elements $x_1, x_2 \in I$, and every $a \in [0, 1]$, we have

$$f(ax_1 + (1-a)x_2) \leq af(x_1) + (1-a)f(x_2). \qquad (1)$$



In other words, a convex function lies below the line joining two points on its graph, as illustrated in the figure.

For example, if f is twice-differentiable on I and if $f''(x) \geq 0$ for all $x \in I$, then $f(x)$ is convex. A function is said to be *strictly convex* if the inequality (1) holds strictly (i.e., without equality) whenever $x_1 \neq x_2$ and $a \notin \{0, 1\}$. Thus for a strictly convex function, equality in (1) can be achieved if and only if $x_1 = x_2$ or $a \in \{0, 1\}$.

A function f for which $-f$ is (strictly) convex is called (strictly) *concave*. For example, $\ln(x)$ is strictly concave over $(0, \infty)$, since $\frac{\mathrm{d}^2}{\mathrm{d}x^2} \ln(x) = -1/x^2 < 0$.

Note that if $0 \leq a \leq 1$, then $\min(x_1, x_2) \leq ax_1 + (1-a)x_2 \leq \max(x_1, x_2)$; thus the point $ax_1 + (1-a)x_2$ is indeed an element of I (and between x_1 and x_2). More generally, for any non-negative real numbers $p_1, \ldots, p_m$ summing to one, i.e., satisfying $p_1 + \cdots + p_m = 1$, and for any points $x_1, \dots, x_m \in I$, the point $p_1 x_1 + \cdots + p_m x_m$ is called a *convex combination* of $x_1, \dots, x_m$. Since

$$\min(x_1, \dots, x_m) \leq p_1 x_1 + p_2 x_2 + \cdots + p_m x_m \leq \max(x_1, \dots, x_m),$$

every convex combination of any finite number of points of I is again a point of I.

3 Jensen's Inequality

Jensen's inequality, named after the Danish engineer Johan Jensen (1859–1925), can be stated as follows.

Theorem 2 (Jensen's Inequality)**.** *Let m be a positive integer and let $f : I \to \mathbb{R}$ be convex over the interval $I \subseteq \mathbb{R}$. For any (not necessarily distinct) points $x_1, \dots, x_m \in I$ and any non-negative real numbers $p_1, \dots, p_m$ summing to one,*

$$f(p_1 x_1 + p_2 x_2 + \cdots + p_m x_m) \leq p_1 f(x_1) + p_2 f(x_2) + \cdots + p_m f(x_m).$$

Proof. We proceed by induction on m, denoting the induction hypothesis as $P(m)$. The truth of $P(1)$ is a triviality, and $P(2)$ is true by definition of convexity. Suppose $P(m)$ is true for some $m \geq 2$, let $x_1, \dots, x_{m+1}$ be any $m+1$ points of I, and let $p_1, \dots, p_{m+1}$ be any $m+1$ non-negative real numbers summing to one. If $p_1 = 1$, then $p_j = 0$ for all $j > 1$, and it is trivially true that $f(p_1 x_1 + \cdots + p_{m+1} x_{m+1}) \leq p_1 f(x_1) + \cdots + p_{m+1} f(x_{m+1})$. Otherwise $p_1 < 1$, and we have

$$p_1 x_1 + \cdots + p_{m+1} x_{m+1} = p_1 x_1 + (1 - p_1) z$$

where

$$z = \frac{p_2}{1 - p_1} x_2 + \cdots + \frac{p_{m+1}}{1 - p_1} x_{m+1}.$$

Note that z is a convex combination of $x_2, \dots, x_{m+1}$, and hence $z \in I$. We then have

$$\begin{aligned} f(p_1 x_1 + (1 - p_1) z) &\leq p_1 f(x_1) + (1 - p_1) f(z) \\ &\leq p_1 f(x_1) + (1 - p_1) \left(\frac{p_2}{1 - p_1} f(x_2) + \cdots + \frac{p_{m+1}}{1 - p_1} f(x_{m+1}) \right) \\ &= p_1 f(x_1) + \cdots + p_{m+1} f(x_{m+1}), \end{aligned}$$

where the first inequality follows from the convexity of f, and the second inequality follows from the induction hypothesis. Thus $P(m)$ implies $P(m+1)$. Since $P(1)$ and $P(2)$ are true, and $P(m)$ implies $P(m+1)$ for all $m \geq 2$, by induction it follows that $P(m)$ is true for all positive integers m. □

It can be shown that in the case when $f : I \to \mathbb{R}$ is strictly convex, equality in Jensen's inequality can be achieved for $x_1, \dots, x_m \in I$ and $p_1, \dots, p_m \geq 0$ summing to one if and only if, for some $x \in I$, $x_i = x$ for all i satisfying $p_i > 0$. In other words, to achieve equality in

Jensen's inequality when f is strictly convex, all x_i's contributing (with positive coefficient) to the convex combination must be equal. (Equality is also achieved when $f(x)$ is an affine function; however such a function is not strictly convex.)

When f is concave, the sense of Jensen's inequality is reversed, i.e.,

$$p_1 f(x_1) + \cdots + p_{m+1} f(x_{m+1}) \leq f(p_1 x_1 + \cdots + p_{m+1} x_{m+1}).$$

For example, taking $I = (0, \infty)$, and $f(x) = \ln(x)$, we have

$$p_1 \ln x_1 + \cdots + p_m \ln x_m \leq \ln(p_1 x_1 + \cdots + p_m x_m)$$

or

$$\ln(x_1^{p_1}) + \cdots + \ln(x_m^{p_m}) \leq \ln(p_1 x_1 + \cdots + p_m x_m).$$

Exponentiating both sides, since $\exp(x)$ is monotonically increasing, we obtain the following theorem, which we term the "Generalized AM-GM Inequality," where AM stands for "arithmetic mean" and GM stands for "geometric mean." This terminology is not standard; however, see Theorem 4 below.

Theorem 3 (Generalized AM-GM Inequality)**.** *For every $x_1, \ldots, x_m > 0$,*

$$x_1^{p_1} \cdots x_m^{p_m} \leq p_1 x_1 + \cdots + p_m x_m$$

for any non-negative real numbers $p_1, \ldots, p_m$ summing to one. Equality is achieved if and only if $x_i = c$ for all i satisfying $p_i > 0$, for some positive constant c.

A special case of Theorem 3 is the following.

Theorem 4 (Inequality of Arithmetic and Geometric Means)**.** *For any positive real numbers $x_1, \ldots, x_m$,*

$$\left(\prod_{i=1}^{m} x_i \right)^{1/m} \leq \frac{1}{m} \sum_{i=1}^{m} x_i,$$

with equality achieved if and only if $x_i = c$ for some constant c, for all $i \in \{1, \ldots, m\}$.

Proof. This is Theorem 3 in the special case when $p_1 = \cdots = p_m = 1/m$. The left-hand side is the geometric mean of the given set of numbers and the right-hand side is the arithmetic mean. □

4 Young's Inequality

Young's inequality, named after the English mathematician William Henry Young (1863–1942), can be stated as follows.

Theorem 5 (Young's Inequality)**.** *For any non-negative real numbers a and b and any positive real numbers p and q such that $\frac{1}{p} + \frac{1}{q} = 1$,*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

with equality achieved if and only if $a^p = b^q$.

Proof. When a and b are positive, this is Theorem 3 in the special case when $m = 2$, $x_1 = a^p$, $x_2 = b^q$, $p_1 = \frac{1}{p}$ and $p_2 = \frac{1}{q}$. Indeed, we then have

$$ab = x_1^{p_1} x_2^{p_2} \leq p_1 x_1 + p_2 x_2 = \frac{a^p}{p} + \frac{b^q}{q}.$$

If one (or both) of a or b is zero, the inequality also holds. □

5 Hölder's Inequality

We can use Young's inequality to prove Hölder's inequality, named after the German mathematician Otto Ludwig Hölder (1859–1937).

Theorem 6 (Hölder's Inequality)**.** *For any pair of vectors x and y in $\mathbb{C}^n$, and for any positive real numbers p and q satisfying $\frac{1}{p} + \frac{1}{q} = 1$, we have*

$$\sum_{i=1}^{n} |x_i y_i| \leq \|x\|_p \cdot \|y\|_q,$$

where

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \quad \|y\|_q = \left(\sum_{i=1}^{n} |x_i|^q \right)^{\frac{1}{q}}$$

are the p- and q-norms of x and y, respectively. If one (or both) of x or y is zero, then equality is achieved. If both x and y are nonzero, equality is achieved if and only for each $i \in \{1, \ldots, n\}$ we have

$$\left(\frac{|x_i|}{\|x\|_p} \right)^p = \left(\frac{|y_i|}{\|y\|_q} \right)^q.$$

Proof. If one (or both) of x or y is zero, the inequality certainly holds with equality. Otherwise, assume x and y are both nonzero, and let $u = \frac{x}{\|x\|_p}$ and let $v = \frac{y}{\|y\|_q}$, and note that

$\|u\|_p = \|v\|_q = 1$. Then

$$\begin{aligned}
\sum_{i=1}^{n} |u_i v_i| &= \sum_{i=1}^{n} |u_i||v_i| \\
&\leq \sum_{i=1}^{n} \left(\frac{|u_i|^p}{p} + \frac{|v_i|^q}{q} \right) \qquad \text{(by Young's inequality)} \\
&= \frac{(\|u\|_p)^p}{p} + \frac{(\|v\|_q)^q}{q} \\
&= \frac{1}{p} + \frac{1}{q} \\
&= 1.
\end{aligned}$$

Now multiply both sides by the positive quantity $\|x\|_p \cdot \|y\|_q$ to obtain the statement of the theorem. To achieve equality, each term in the sum must achieve equality in Young's inequality, i.e., for all $i \in \{1, \ldots, n\}$, $|u_i|^p = |v_i|^q$, which translates to the statement in the theorem since $|u_i| = |x_i|/\|x\|_p$ and $|v_i| = |y_i|/\|y\|_q$. □

The Cauchy-Schwarz inequality, named after the French mathematician, engineer, and physicist Augustin-Louis Cauchy (1789–1857) and the German mathematician Karl Hermann Amandus Schwarz[1] (1843–1921), is closely related to Hölder's inequality in the case $p = q = 2$.

Theorem 7 (Cauchy-Schwarz Inequality)**.** *For any pair of vectors x and y in $\mathbb{C}^n$,*

$$\left| \sum_{i=1}^{n} x_i y_i^* \right| \leq \|x\|_2 \cdot \|y\|_2,$$

where equality is achieved if and only $y = \lambda x$ for some scalar $\lambda \in \mathbb{C}$.

Proof. We write

$$\begin{aligned}
\left| \sum_{i=1}^{n} x_i y_i^* \right| &\leq \sum_{i=1}^{n} |x_i y_i^*| \\
&= \sum_{i=1}^{n} |x_i y_i| \\
&\leq \|x\|_2 \cdot \|y\|_2,
\end{aligned}$$

where the first inequality is the triangle inequality for complex numbers and the second inequality is Hölder's inequality in the case $p = q = 2$. For equality to hold, it must hold in both inequalities, which certainly occurs if one (or both) of x or y is zero. If both are nonzero, the first equality holds with equality if and only if every nonzero $x_i y_i^*$ has the same

[1] Caution: some authors are tempted to misspell this as Schwartz!

phase, i.e., $x_i y_i^* = r_i e^{\mathrm{i}\theta}$ for some fixed θ. The second inequality holds with equality if and only if

$$\frac{|x_i|}{\|x\|_2} = \frac{|y_i|}{\|y\|_2}$$

for every i, which implies that $|y_i| = c|x_i|$ for some positive constant c. Putting these two conditions together, if $x_i = a_i e^{\mathrm{i}\phi_i}$, then $y_i = c a_i e^{\mathrm{i}(\phi_i - \theta)}$, or, in other words, $y = \lambda x$ for $\lambda = c e^{-\mathrm{i}\theta}$. □

6 Minkowski's Inequality

We can use Hölder's inequality to prove Minkowski's inequality, named after the German mathematician Hermann Minkowski (1864–1909). Minkowski's inequality is the triangle inequality for p-norms.

Theorem 8 (Minkowski's Inequality)**.** *For any pair of vectors u and v in $\mathbb{C}^n$, and for any $p > 1$, we have*

$$\|u + v\|_p \leq \|u\|_p + \|v\|_p.$$

Equality holds if and only if $au = bv$ for some non-negative real constants a and b, not both zero.

Proof. The theorem is clearly true if u and v are both zero, and it holds if $u + v$ is zero. Otherwise, we write

$$\begin{aligned}
(\|u + v\|_p)^p &= \sum_{i=1}^{n} |u_i + v_i|^p \\
&= \sum_{i=1}^{n} |u_i + v_i| \cdot |u_i + v_i|^{p-1} \\
&\leq \sum_{i=1}^{n} (|u_i| + |v_i|) \cdot |u_i + v_i|^{p-1} \\
&= \sum_{i=1}^{n} |u_i||u_i + v_i|^{p-1} + \sum_{i=1}^{n} |v_i||u_i + v_i|^{p-1} \\
&\leq \left(\sum_{i=1}^{n} |u_i|^p \right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n} \left(|u_i + v_i|^{p-1}\right)^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} \\
&\quad + \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n} \left(|u_i + v_i|^{p-1}\right)^{\frac{p}{p-1}} \right)^{\frac{p-1}{p}} \\
&= (\|u\|_p + \|v\|_p) \cdot (\|u + v\|_p)^{p-1}.
\end{aligned}$$

The first inequality is an application of the triangle inequality for complex numbers, and the second inequality is two applications of Hölder's inequality, taking $q = p/(p-1)$ so that

$\frac{1}{p}+\frac{1}{q}=1$. The theorem follows by dividing both sides by the positive quantity $(\|u+v\|_p)^{p-1}$. To achieve equality it is necessary that the triangle inequality for complex numbers holds with equality for each term, i.e., that u_i and v_i, if nonzero, have the same phase. We also require equality in each application of Hölder's inequality. Let $w = (w_1, \ldots, w_n)$ where $w_i = |u_i + v_i|^{p-1}$. For equality to hold in the first application of Hölder's inequality, we need, for each $i \in \{1, \ldots, n\}$, that

$$\left(\frac{|u_i|}{\|u\|_p}\right)^p = \left(\frac{w_i}{\|w\|_q}\right)^q,$$

and similarly for the second application of Hölder's inequality. Now

$$\begin{aligned}
\|w\|_q &= \left(\sum_{i=1}^{n} w_i^q\right)^{1/q} \\
&= \left(\sum_{i=1}^{n} (|u_i + v_i|^{p-1})^{p/(p-1)}\right)^{(p-1)/p} \\
&= \left(\sum_{i=1}^{n} (|u_i + v_i|)^p\right)^{(p-1)/p} \\
&= \|u+v\|_p^{p-1};
\end{aligned}$$

thus $\|w\|_q^q = \|u+v\|_p^p$. Also $w_i^q = |u_i + v_i|^p$. Thus, taking pth roots, the conditions for equality in Hölder's inequality become

$$\frac{|u_i|}{\|u\|_p} = \frac{|u_i + v_i|}{\|u+v\|_p} = \frac{|v_i|}{\|v\|_p}.$$

Thus we need $|u_i|$ and $|v_i|$ to be proportional. Summarizing, we find that equality in Minkowski's inequality is achieved if and only if $au = bv$ for some non-negative real scalars a and b, not both zero. $\square$

The alert reader will note that Minkowski's inequality also holds in the case $p = 1$. This must be proved separately, and is left as an exercise.

Acknowledgement

The author is grateful for useful comments and suggestions from Dr. Qun Zhang.