Hölder, Minkowski, Riesz, Helly

Bill Casselman
University of British Columbia
cass@math.ubc.ca

This essay contains proofs of Hölder's and Minkowski's inequalities, as well as some examples of how they are used in the theory of Banach spaces. There is nothing new, but my reliance on original works seems to be unusual.

Contents

1. Hölder's Lemma [Minkowski.tex]

Otto Hölder was one of the first mathematicians to be interested in convexity as a separate topic in mathematics. As far as I know, the following elementary result was first formulated in [Hölder:1889]:

[holders] **1.1. Lemma.** *Suppose φ to be a real function on an interval in $\mathbb{R}$ with continuous and weakly increasing derivative. For (x_i) in the interval and all $a_i \ge 0$ with $\sum a_i = 1$*

$$\varphi\Big(\sum a_i x_i\Big) \le \sum a_i \varphi(x_i)\,.$$

As we'll see more clearly in a moment, the hypothesis amounts to the assertion that the graph of $y = \varphi(x)$ is convex. The Lemma can be summarized informally as

$$\varphi(\text{mean}) \le \text{mean}(\varphi)\,.$$

In practice, the hypothesis is verified by showing $\varphi'' > 0$.

Proof. The assertion is equivalent to the claim that

$$\varphi\left(\frac{\sum a_i x_i}{\sum a_i}\right) \le \frac{\sum a_i \varphi(x_i)}{\sum a_i}\,.$$

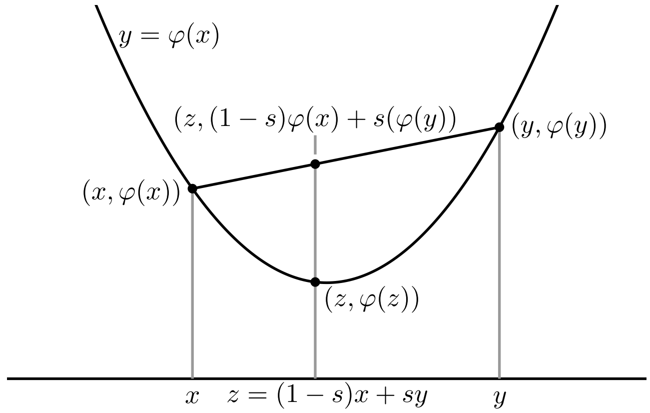
if all $a_i > 0$. Since

$$\frac{\sum_1^n a_i X_i)}{\sum a_i} = \frac{\sum_1^{n-1} a_i}{\sum_1^n a_i} \cdot \left(\frac{\sum_1^{n-1} a_i X_i}{\sum_1^{n-1} a_i}\right) + \frac{a_n X_n}{\sum_1^n a_i}\,,$$

induction reduces the claim to the case $n = 2$: if $z = (1-s)x + sy$ with $0 \le s \le 1$, then

$$\varphi(z) \le (1-s)\varphi(x) + s\varphi(y)\,.$$

In other words, in between x and y the graph of φ lies below or on the straight line from $(x, \varphi(x))$ to $(y, \varphi(y))$:



images/convex-phi.eps

For the proof:

$$\begin{aligned}\big((1-s)\varphi(x)+s\varphi(y)\big)-\varphi(z) &= \frac{(y-z)\varphi(x)+(z-x)\varphi(y)}{y-x}-\varphi(z)\\ &= \frac{(y-z)(\varphi(x)-\varphi(z))+(z-x)(\varphi(y)-\varphi(z))}{y-x}\\ &= \frac{(z-x)(\varphi(y)-\varphi(z))-(y-z)(\varphi(z)-\varphi(x))}{y-x}\\ &= \frac{(z-x)(y-z)\varphi'(\theta_y)-(y-z)(z-x)\varphi'(\theta_x)}{y-x}\\ &= \frac{(z-x)(y-z)(\varphi'(\theta_y)-\varphi'(\theta_x))}{y-x}\geq 0\,.\end{aligned}$$

Here, according to one version of the mean value theorem, $x \leq \theta_x \leq z \leq \theta_y \leq y$.

Replacing φ by its negative:

[holders-cor] **1.2. Corollary.** *Suppose φ to be a real function on an interval in $\mathbb{R}$ with continuous and weakly decreasing derivative. For (x_i) in the interval and all $a_i \geq 0$ with $\sum a_i = 1$*

$$\varphi\Big(\sum a_i x_i\Big) \geq \sum a_i \varphi(x_i)\,.$$

2. Hölder's inequality [Minkowski.tex]

If one sets $\varphi(x) = x^p$ in Hölder's Lemma, with $p > 1$, one deduces the result commonly known as Hölder's inequality.

[holders-ineq] **2.1. Proposition.** *For $p > 1$, $1/q + 1/p = 1$*

$$\sum x_i y_i \leq \Big(\sum |x_i|^q\Big)^{1/q} \cdot \Big(\sum |y_i|^p\Big)^{1/p}.$$

Proof. It suffices to prove this for $x_i, y_i > 0$, since

$$\sum x_i y_i \leq \sum |x_i y_i|\,.$$

♡ **[holders]** If we set $\varphi(x) = x^p$ in Lemma 1.1, we get

$$\left(\frac{\sum a_i b^i}{\sum a_i}\right)^p \le \frac{\sum a_i b_i^p}{\sum a_i},$$

and then

$$\sum a_i b_i \le \left(\sum a_i\right)^{1-1/p} \left(\sum a_i b_i^p\right)^{1/p}.$$

But now is we set $1/q = 1 - 1/p$, $a_i = x_i^q$, $a_i b_i^p = y_i^p$ we deduce

$$\begin{aligned} a_i^{1/q} &= x_i \\ a_i^{1/p} b_i &= y_i \\ a^{1/p+1/q} b_i &= a_i b_i = x_i y_i\,. \end{aligned}$$

3. Minkowski's inequality [Minkowski.tex]

Minkowski's inequality is found on pp. 115–117 of [Minkowski:1896/1910], which put convexity into the main stream of mathematics:

[minkowskis] **3.1. Proposition.** *For* $p > 1$

$$\left(\sum (x_i + y_i)^p\right)^{1/p} \le \left(\sum |x_i|^p\right)^{1/p} + \left(\sum |y_i|^p\right)^{1/p}.$$

Minkowski's proof is an elementary calculus exercise, and has the rare virtue of being at once clever and straightforward. I reproduce it here.

Proof. It is to be shown that

$$\psi = \left(\sum a_i^p\right)^{1/p} + \left(\sum b_i^p\right)^{1/p} - \left(\sum (a_i + b_i)^p\right)^{1/p} \ge 0\,.$$

Fix the array (a_i) and consider this a function of (b_i) alone. Note that

$$\frac{\partial \psi}{\partial b_i} = \left(\frac{b_i}{\|b\|_p}\right)^{p-1} \left(\frac{(a_i + b_i)}{\|a+b\|_p}\right)^{p-1} = \text{say } \beta_i^{p-1} - \gamma_i^{p-1}$$

if

$$\beta = \frac{b}{\|b\|_p}, \quad \gamma = \frac{a+b}{\|a+b\|_p}\,.$$

We then have

[test] **(3.2)**
$$\begin{aligned} 1 = \|\beta\|_p &= (\beta_1^p + \cdots + \beta_n^p)^{1/p} \\ &= \|\gamma\|_p = (\gamma_1^p + \cdots + \gamma_n^p)^{1/p}\,. \end{aligned}$$

I'll say that b is exceptional if it is proportional to a, or equal to 0. For exceptional points, $\psi = 0$. Suppose we are given a point b that is not exceptional. Then some $\beta(k) \ne \gamma(k)$. We cannot then have $\beta(k) \le \gamma(k)$ ♡ **[test]** for all k because (3.2) would not hold. So there exists some $\beta(k) > \gamma(k) \ge 0$, and since $p > 1$ we must have at the point $\partial\psi/\partial b(k) > 0$. This implies some point $b - h\varepsilon(k)$ has a smaller value of ψ for very small h.

Now fix $c > 0$ and consider the region $b_i \le c$ for all i. It is compact, and there must exist a point b where ψ is a minimum. If this minimum is not 0, then b is not exceptional and there exists a point in the region where ψ is less. Therefore b must be exceptional, and the minimum is 0.

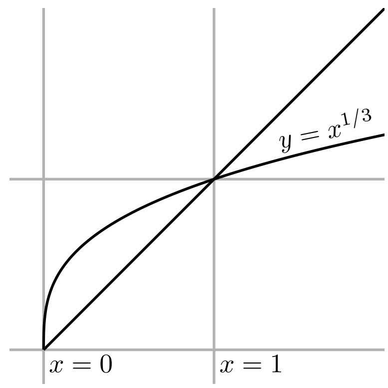
Remarks. This inequality is also a consequence of the corollary of Hölder's Lemma (the version with $\varphi'' < 0$) for $\varphi(x) = (1 + x^{1/p})^p$. This was observed with a note of surprise in [Bourbaki:1981].

4. Helly's version [Minkowski.tex]

I'll offer here geometric proofs of Minkowski's and Hölder's inequalities, suggested in [Riesz:1913] and [Helly:1921].

I'll use an elementary observation: *if* $s \geq 0$ *and* $x > 0$ *then* $x \leq 1$ *if and only if* $x^s \leq 1$.

In a picture:

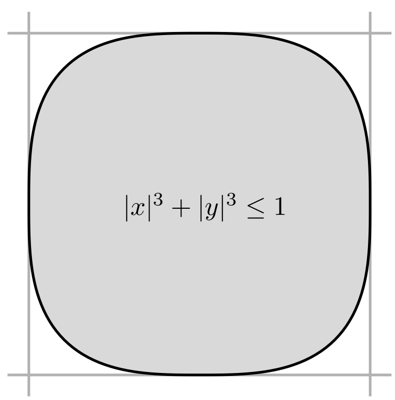


-images/minkowskisa.eps

It is Riesz who seems to have first observed in print that the basic idea of the inequalities we are looking at is contained in the following, although Minkowoski must certainly have known it:

[convex] **4.1. Proposition.** *The region* $\|x\|_p \leq 1$ *is bounded and convex.*

For example, in two dimensions (given the first observation above):



-images/minkowskisc.eps

Proof. That it is bounded should be clear, since $\sup |x_i| \leq \sum |x_i|^p$. Given the observation, it then suffices to show the region $\sum |x_i|^p \leq 1$ is convex. So suppose $\sum |x_i|^p \leq 1$, $\sum |y_i|^p \leq 1$. It is to be proved that $\sum |(1-s)x_i + sy_i|^p| \leq 1$ for all s in $[0,1]$.

♡ [holders] But it follows from Lemma 1.1 that

$$\begin{aligned}\sum \big|(1-s)x_i + sy_i\big|^p &\leq \sum (1-s)|x_i|^p + s|y_i|^p \\ &= (1-s)\Big(\sum |x_i|^p\Big) + s\Big(\sum |y_i|^p\Big) \\ &= (1-s)\|x\|_p + s\|y\|_p \\ &\leq (1-s) + s = 1\end{aligned}$$

Suppose for the moment that Ω is any bounded, closed, convex region in $\mathbb{R}^n$. Each array (u_i) defines a linear function $\langle u, x\rangle = \sum u_i x_i$ on $\mathbb{R}^n$. Define

$$\|u\|_* = \sup_{x \in \Omega} |\langle u, x\rangle| \,.$$

Since the region Ω is closed and convex, it is the intersection of all regions $\langle u, x\rangle \le \|u\|_*$.

I recall that a **norm** on $\mathbb{R}^n$ is a function $\|x\|$ taking non-negative values such that

(a) $\|x+y\| \le \|x\| + \|y\|$;
(b) $\|cx\| = |c|\|x\|$;
(c) $\|x\| = 0$ if and only if $x = 0$.

[convex-dual] **4.2. Proposition.** *Given the closed, bounded, convex set Ω, $\|u\|_*$ is a norm.*

Proof.

$$\|u+v\|_* = \sup_{x\in\Omega}\langle u+v, x\rangle = \langle u, x\rangle + \langle v, x\rangle \le \|u\|_* + \|v\|_* .$$

[holders-convex] **4.3. Corollary.** *In this situation*

$$|\langle u, x\rangle| \le \|u\|_* \cdot \|x\| .$$

As we shall see in a moment, this is a generalization of Hölder's inequality.

What is $\|u\|_*$ if $\|x\| = \|x\|_p$ with $1 < p < \infty$? The answer is a straightforward application of Lagrange multipliers. Thsi tells you that to find the critical values of $f(x)$ on $g(x) = 0$, you find the points of $g(x) = 0$ where ∇f is proportional to ∇g. In our case that gives the equation

[hold] **(4.4)**
$$(a_i) = \lambda(x_i^{p-1})$$

for some constant λ, subject to the condition $\sum x_i^p = 1$. The equation $a_i = \lambda x_i^{p-1}$ gives $x_i = (a_i/\lambda)^{1/p-1}$, and if $q = p/p-1$ this leads to

$$\lambda = \left(\sum a_i^q\right)^{1/q} = \|a\|_q .$$

But then at the maximum

$$\sum a_i x_i = \lambda = \|a\|_q$$

and then to Hölder's inequality.

♡ [convex-dual] Minkowski's inequality is now an immediate consequence of Proposition 4.2.

5. Applications [Minkowski.tex]

• Suppose $1 \le p < \infty$. Minkowski's inequality implies that

$$\|c\|_p = \left(\sum |c(i)|^p\right)^{1/p}$$

is a norm on the vector space of sequences $(c(i))$ of complex numbers. Let ℓ^p be the space of all sequences for which the norm is finite.

[lp] **5.1. Proposition.** *The vector space ℓ^p is complete.*

Proof. Replacing f_n by a subsequence if necessary, we may assume that $\|f_{n+1} - f_n\|_p \le 1/2^n$. Since $|f(n)| \le \|f\|_p$, the point values of f_n converge. The limit will be in ℓ^p.

For a while, suppose $p > 1$, $1/p + 1/q = 1$.

In the Hilbert space ℓ^2, we know that $\langle f, \overline{f}\rangle = \|f\|_2^2$. There is a similar phenomenon in ℓ^p:

[innerp] **5.2. Lemma.** *Suppose $(g(k))$ to be a sequence of complex numbers. If $f(k) = \overline{g(k)}\,|g(k)|^{q-2}$ then*

$$\|f\|_p^p = \sum f(k)g(k) = \|g\|_q^q .$$

Proof. Because $|f(k)|^p = f(k)g(k) = |g(k)|^q$.

For f in ℓ^p and g in ℓ^q, Hölder's inequality implies that

$$\langle g, f\rangle = \sum g(i)f(i) \le \|g\|_q \cdot \|f\|_p < \infty\,.$$

If we define

$$T_g\colon\ \ell^p \longrightarrow \mathbb{C}, \quad f \longmapsto \langle g, f\rangle$$

then T_g is in the continuous dual of ℓ^p and its norm is

$$|||T_g||| = \sup_{f\neq 0} \frac{\langle g, f\rangle}{\|f\|_p} \le \|g\|_q\,.$$

[p-duals] **5.3. Proposition.** *For $1 < p < \infty$ the map $g \mapsto T_g$ is an isometry of ℓ^q with the continuous dual of ℓ^p.*

Proof. I first show that $|||T_g||| = \|g\|_q$. Since

$$\langle g, f\rangle \le |||T_g||| \cdot \|f\|_p \le \|g\|_q \cdot \|f\|_p\,,$$

it suffices to find $f \neq 0$ in ℓ^p such that

$$\langle g, f\rangle = \|g\|_q \cdot \|f\|_p\,.$$

For this to happen, it is plausible to restrict the search to f such that each $|f(k)|^p$ is a scalar multiple of $|g(k)|^q$, or even $|f(k)|^p = |g(k)|^q$. For this, set

$$f(k) = \overline{g(k)}\,|g(k)|^{q-2}\,,$$

♡ [innerp] and apply Lemma 5.2.

It must now be shown that every continuous linear map γ from ℓ^p to $\mathbb{C}$ is equal to some T_g. It is easy to define the candidate. Let ε_m be the vector in ℓ^p with

$$\varepsilon_m(k) = \begin{cases} 1 & \text{if } k = m \\ 0 & \text{otherwise.} \end{cases}$$

Then set

$$g(m) = \langle \gamma, \varepsilon(m)\rangle = g(m)\,.$$

It must be shown that g is in ℓ^q and that $T_g = \gamma$. The second is easy, since the finite sums of the ε_k are dense in ℓ^p.

As for the first: since γ is continuous, we know that $|\langle\gamma, f\rangle| \le C \cdot \|f\|_p$ for some $C > 0$ and all f in ℓ^p. Suppose $g^{[m]}$ to be g truncated to the first m terms. Then Then

$$|\langle\gamma, f\rangle| = \|g^{[m]}\|_q^q \le C \cdot \|f^{[m]}\|_p = C \cdot \|g^{[m]}\|_q^{q/p}\,,$$

leading to

$$\|g^{[m]}\|_q^{q-q/p} = \|g^{[m]}\|_q \le C\,.$$

Let m go to ∞.

So the dual of ℓ^p is ℓ^q and the dual of ℓ^q is ℓ^p. In other words, these spaces ℓ^p are **reflexive**.

• Define ℓ^∞ to be the vector space of all bounded sequences with norm

$$\|c\|_\infty = \sup |c_i|\,,$$

and ℓ_∞ to be the subspace of sequences with $\lim c_i = 0$.

[linfty] **5.4. Lemma.** *The space ℓ^∞ is complete, and ℓ_∞ is a closed subspace.* The inner product

$$\langle g, f\rangle = \sum f_i g_i$$

defines a pairing of ℓ^∞ and ℓ^1.

[infty-dual] **5.5. Proposition.** *This pairing identifies ℓ^1 with the continuous dual of ℓ_∞, and ℓ^∞ with the continuous dual of ℓ^1.*

In other words, the spaces ℓ_∞ is *not* reflexive. The space ℓ_∞ is closed in ℓ^∞. According to Hahn-Banach, there are many continuous lineare forms on ℓ^∞ vanishing on ℓ_∞. Fopr example, one of them is

$$(x_n) \longrightarrow \lim_{m\to\infty} \frac{x_{m+1} + x_{m+2} + \cdots + x_{2m}}{m} \, .$$

As Helly remarks, the spaces ℓ_∞ and ℓ^∞ differ in one important way—the space ℓ_∞ possesses a countable dense subset, whereas ℓ^∞ does not.

6. References [Minkowski.tex]

1. Nicolas Bourbaki, Chapitres 1–4 of **Integration**, Springer, 1981.

2. Édouard Helly, 'Ueber Systeme linearer Gleichungen mit unendlichen vielen Unbekannten', *Monatsheft für Mathematik und Physik*, **XXXI** (1921), 60–91.

3. Otto Hölder, 'Über eine Mittelwertssatz', *Nachrichten von der Königliche Gesellschaft der Wissenschaften und der Georg-Augusts-Universiät zu Göttingen*, 1889.

4. Hermann Minkowski, **Geometrie der Zahlen**, Teubner, 1896/1910.

5. Fréderique Riesz, **Les systemes d'équations linéaires à une infinité d'inconnues**, Gauthier-Villars, 1913.