The Hölder and Minkowski inequalities

Brent Baccala
September 10, 2008

Lecture 1 was the first time that I had seen the Minkowski inequality in its "double integral" formulation, so I did some digging and came up with a proof. I use the Lebesgue integral throughout, so it can get rather "baroque" as John said, but there are some interesting intermediate results, including a converse of the Hölder theorem and a theorem of Abel that shows how to adjust a divergent series to make it convergent.

Throughout this paper, p and q will be assumed to be real positive numbers related by $\frac{1}{p} + \frac{1}{q} = 1$. I'll exclude ∞ throughout, so this implies that $p > 1$ and $q > 1$. Integrals, unless specified otherwise, are always over some measure space X with measure μ.

Hölder's inequality

Theorem A

If $u, v \in \mathbb{R}$, $u \geq 0$ and $v \geq 0$, then

$$uv \leq \frac{u^p}{p} + \frac{v^q}{q}$$

and equality holds if $u^p = v^q$.

Proof

First note that:

$$\begin{aligned}
\frac{1}{p} + \frac{1}{q} &= 1 \\
p + q &= pq \\
p &= pq - q \\
p &= q(p-1) \qquad (1)
\end{aligned}$$

If u or v is zero, the inequality obviously holds, so assume that both numbers are positive and consider

$$f(u,v) = \frac{u^p}{p} + \frac{v^q}{q} - uv$$

Holding v fixed, we vary u:

$$\frac{df}{du} = u^{p-1} - v \qquad \frac{d^2 f}{du^2} = (p-1)u^{p-2}$$

$p > 1$ and u is positive, so $\frac{d^2 f}{du^2}$ is always positive, so the function has a minimum at $v = u^{p-1}$, where (1) shows that

$$v^q = u^{q(p-1)} = u^p$$

At the minimum,

$$f = \frac{u^p}{p} + \frac{u^p}{q} - uu^{p-1} = u^p(\frac{1}{p} + \frac{1}{q}) - u^p = 0$$

so $f(u,v) \geq 0$ for all u and v. The theorem follows. $\square$

Theorem B

Let f and g be non-negative real functions on X. If $\int f^p\, d\mu = \int g^q\, d\mu = 1$, then $\int fg\; d\mu \le 1$, and equality holds if $f^p = g^q$ a.e.

Proof

$$\begin{aligned} \int fg\, d\mu &\le \int \Big(\frac{f^p}{p} + \frac{g^q}{q}\Big)\, d\mu && \text{Theorem A} \\ &= \frac{1}{p}\int f^p\, d\mu + \frac{1}{q}\int g^q\, d\mu \\ &= \frac{1}{p} + \frac{1}{q} = 1 \end{aligned}$$

Equality holds if $f^p = g^q$, but since we can modify the functions on any set of measure zero without affecting the integrals, we only need $f^p = g^q$ a.e.

□

Theorem C (Hölder's inequality)

Let f and g be complex functions on X. Then

$$\int |fg|\, d\mu \le \Big\{\int |f|^p\, d\mu\Big\}^{\frac{1}{p}} \Big\{\int |g|^q\, d\mu\Big\}^{\frac{1}{q}}$$

Equality holds if f^p and g^q are *effectively proportional*: there exist complex numbers α and β, not both zero, such that $\alpha f^p = \beta g^q$ a.e. The second number β is largely extraneous; it is present only to handle the zero cases correctly.

Proof

Set $F = \int |f|^p d\mu$ and $G = \int |g|^q d\mu$. We assume that these integrals exist; if they diverge, then the inequality is trivially satisfied. Now let

$$\bar{f} = \frac{|f|}{F^{1/p}} \qquad \bar{g} = \frac{|g|}{G^{1/q}}$$

so that $\int \bar{f}^p\, d\mu = \int \bar{g}^q\, d\mu = 1$, and now,

$$\begin{aligned} \int |fg|\, d\mu &= F^{1/p} G^{1/q} \int \bar{f}\bar{g}\, d\mu \\ &\le F^{1/p} G^{1/q} && \text{Theorem B} \\ &= \Big\{\int |f|^p\, d\mu\Big\}^{\frac{1}{p}} \Big\{\int |g|^q\, d\mu\Big\}^{\frac{1}{q}} \end{aligned}$$

If f^p and g^q are effectively proportional, we can assume that β is 1 (divide through by β if it's not zero; swap f and g if it is). So $\alpha f^p = g^q$ a.e, and taking the modulus of both sides yields $|\alpha||f|^p = |g|^q$ a.e. $G = \int |\alpha||f|^p d\mu = |\alpha| F$, and

$$\bar{g}^q = \frac{1}{G}\,|g|^q = \frac{1}{|\alpha| F}\,|\alpha|\,|f|^p = \frac{1}{F}\,|f|^p = \bar{f}^p \quad \text{a.e.}$$

□

Minkowski's inequality

This presentation is adapted from Hardy, Littlewood, and Pólya, *Inequalities* (Cambridge, 1934), and I use their theorem numbers throughout, though I prove the theorems in logical, not numerical, order.

Theorem 42

If x and r are real numbers, $x > 0$ and $x \neq 1$, then

$$x^r - 1 > r(x-1) \qquad (r > 1)$$

$$x^r - 1 < r(x-1) \qquad (0 < r < 1)$$

Proof (My proof, not theirs)

Consider $f(x) = x^r - 1 - r(x-1)$:

$$f'(x) = rx^{r-1} - r = r(x^{r-1} - 1)$$
$$f''(x) = r(r-1)x^{r-2}$$

The only critical point is at $x = 1$ ($x > 0$, so we don't have to worry about $x = -1$), and $f(1) = 0$. If $r > 1$, then $f''(x)$ is always positive, $x = 1$ is thus a minimum, and therefore:

$$f(x) = x^r - 1 - r(x-1) > 0 \qquad r > 1, \quad x \neq 1$$

$$x^r - 1 > r(x-1) \qquad r > 1, \quad x \neq 1$$

The $0 < r < 1$ case is similar, except that $x = 1$ is now a maximum.

$\square$

Theorem 41

If x and y are positive real numbers, $x \neq y$, and r is a real number, then

$$rx^{r-1}(x-y) > x^r - y^r > ry^{r-1}(x-y) \qquad r < 0 \text{ or } r > 1$$

$$rx^{r-1}(x-y) < x^r - y^r < ry^{r-1}(x-y) \qquad 0 < r < 1$$

Proof

Use Theorem 42 with $\frac{x}{y}$ or $\frac{y}{x}$ to establish the $r > 1$ and $0 < r < 1$ cases, i.e:

$$(\frac{x}{y})^r - 1 > r(\frac{x}{y} - 1) \qquad r > 1$$

$$x^r - y^r > ry^{r-1}(x-y) \qquad r > 1$$

For the $r < 0$ cases, use $s = -r$, so $s + 1 > 1$

$$x^r - y^r = x^{-s} - y^{-s} = x^{-s}y^{-s-1}(y^{s+1} - x^s y)$$
$$= x^{-s}y^{-s-1}\big(y^{s+1} - x^{s+1} - x^s(y-x)\big)$$

$$> x^{-s}y^{-s-1}\big\{(s+1)x^s(y-x) - x^s(y-x)\big\} \qquad \text{use } r > 1 \text{ case of this theorem}$$

$$= x^{-s}y^{-s-1}sx^s(y-x) = ry^{r-1}(x-y) \qquad \square$$

Theorem 162 (Abel)

Given a real, non-negative, divergent series $\sum a_i$, let $A_n = \sum_{i=1}^{n} a_i$. Then:

i) $\sum \dfrac{a_n}{A_n}$ is divergent, and

ii) $\sum \dfrac{a_n}{{A_n}^{1+\delta}}$ is convergent for $\delta > 0$

Hardy, et. al, stated this theorem without the "non-negative" requirement, but the theorem isn't true without this extra condition, because some A_n could be zero. Consider the series $1 - 1 + 1 - 1 + \cdots$ which obviously fails the theorem's construction.

Proof

Note that $\forall n, r \in \mathbb{Z}^+$, $A_{n+r} \geq A_n$, so

$$\frac{a_{n+1}}{A_{n+1}} + \frac{a_{n+2}}{A_{n+2}} + \cdots + \frac{a_{n+r}}{A_{n+r}} \geq \frac{a_{n+1} + a_{n+2} + \cdots + a_{n+r}}{A_{n+r}} = \frac{A_{n+r} - A_n}{A_{n+r}} = 1 - \frac{A_n}{A_{n+r}}$$

Now fix n and let $r \to \infty$. A_n is finite and $A_{n+r} \to \infty$, so $(1 - \frac{A_n}{A_{n+r}}) \to 1$, and for any n, the tail of series (i) will be greater than or equal to 1, which precludes convergence.

Now assume $0 < \delta < 1$. The series:

$$\sum \frac{A_n^\delta - A_{n-1}^\delta}{A_{n-1}^\delta A_n^\delta} = \sum \left(\frac{1}{A_{n-1}^\delta} - \frac{1}{A_n^\delta}\right)$$

is convergent (since $A_n \to \infty$). We use it in a comparison test to establish convergence of the next series, since Theorem 41 tells us that $A_n^\delta - A_{n-1}^\delta > \delta A_n^{\delta-1}(A_n - A_{n-1}) = \delta A_n^{\delta-1} a_n$, so

$$\sum \frac{\delta A_n^{\delta-1} a_n}{A_{n-1}^\delta A_n^\delta} = \sum \frac{\delta a_n}{A_{n-1}^\delta A_n}$$

is convergent. Since $A_n > A_{n-1}$, we now see that series (ii) is convergent.

Since increasing δ shrinks the terms of series (ii), establishing the result for $0 < \delta < 1$ established it for all $\delta > 0$

$\square$

Theorem 161

Given a real, non-negative sequence a_i, if for all real sequences b_i such that $\sum b_i^p$ is convergent, $\sum a_i b_i$ is also convergent, then $\sum a_i^q$ is convergent, i.e:

$$\left\{\forall b \Big(\sum b_i^p \text{ converges} \Rightarrow \sum a_i b_i \text{ converges}\Big)\right\} \Rightarrow \sum a_i^q \text{ converges}$$

Again, the book's statement of this theorem omitted "non-negative". Also note that I'm using $\frac{1}{p} + \frac{1}{q} = 1$ again.

Proof

By proving the contraposition; we'll establish that…

Theorem 161a

Given a real, non-negative, divergent series $\sum a_i^q$, we can construct a real sequence b_i such that $\sum b_i^p$ converges and $\sum a_i b_i$ diverges, i.e:

$$\sum a_i^q \text{ diverges} \Rightarrow \left\{\exists b \Big(\sum b_i^p \text{ converges } \wedge \sum a_i b_i \text{ diverges}\Big)\right\}$$

Proof

Note first that the theorem holds even if the series "diverges" because one (or more) of its terms is actually infinite. Just pick b_i to be 1 at a single term where a_i is ∞ and 0 elsewhere. So we can assume that all the a_i's are finite.

Let $u_i = a_i^q$, so $a_i = u_i^{1/q}$ and $\sum u_i$ diverges.

Let $U_n = \sum_{i=1}^n u_i$ and take $v_i = \frac{1}{U_i}$.

Theorem 162 shows that $\sum u_i v_i = \sum \frac{u_i}{U_i}$ diverges, while $\sum u_i v_i^p = \sum \frac{u_i}{U_i^p}$ converges (since $p > 1$).

Now take $b_i = u_i^{1/p} v_i$. So $\sum b_i^p = \sum u_i v_i^p$ converges, while $\sum a_i b_i = \sum u_i^{1/p + 1/q} v_i = \sum u_i v_i$ diverges.

Thus, b_i is the desired sequence.

$\square$

Theorem 190 (Hölder converse)

Let X be a σ-finite measure space with measure μ.

Given a measurable function $f : X \to \mathbb{C}$, if $\forall\, g \in L^p, fg \in L^1$, then $f \in L^q$.

Proof

We just proved this (Theorem 161) for counting measures, now we have to extend that result. Hardy, et. al. only proved this for the real line. I've tried to come up with a more general version.

A measure space is σ-*finite* if it can be written as a countable union of subspaces, each of finite measure.

Again, we proceed by proving the theorem's contraposition. Assume $f \notin L^q$. We'll construct a function $g \in L^p$ such that $fg \notin L^1$. First we construct a function s with three properties: $0 \leq s \leq |f|$, $s \notin L^q$, and s's range consists of at most countably many values.

Decompose X into a countable union of subspaces of finite measure: $X = \cup X_i$. Now take $\sum \epsilon_i$ to be an infinite non-negative series that converges to a finite value (say $\sum 1/2^n$), and assign to each subspace a term of the series. Take $\delta_i = \epsilon_i/\mu(X_i)$, a finite value. Define $s(x)$, for $x \in X_i$, to be the real q^{th} root of the largest integer multiple of δ_i less than $|f|^q(x)$.

Now for $x \in X_i$, $|f|^q(x) - s^q(x) \leq \delta_i$, so $\int_{X_i} |f|^q - s^q \, d\mu \leq \delta_i \mu(X_i) = \epsilon_i$. So $\int_X |f|^q - s^q \, d\mu \leq \sum \epsilon_i$, which is finite, so $\int_X s^q \, d\mu = \infty$ since $\int_X |f|^q \, d\mu = \infty$.

Since s's range in each X_i consists of at most countably many values (the roots of the multiples of δ_i), and there are at most countably many X_i's, a Cantor-type diagonalization argument shows that s's range for all of X is at most countable. Enumerate s's values as c_j and let $X_j = \{x | s(x) = c_j\}$ (this is a different decomposition than the X_i's). So s can be written:

$$s = \sum c_j \, \mathbb{1}_{X_j}$$

and

$$\int s^q = \sum c_j^q \, \mu(X_j)$$

and this last series is divergent because the integral is divergent. Take $a_j = c_j \, \mu(X_j)^{1/q}$, use Theorem 161a to construct b_j (since $\sum a_j^q$ diverges), and now construct:

$$g = \sum b_j \, \mu(X_j)^{-1/p} \, \mathbb{1}_{X_j}$$

$\int g^p = \sum b_j^p$ converges, so $g \in L^p$, and

$$\int |fg| \geq \int sg = \sum c_j b_j \, \mu(X_j)^{1-1/p} = \sum c_j b_j \, \mu(X_j)^{1/q} = \sum a_j b_j$$

This series diverges, so $fg \notin L^1$.

$\square$

Theorem 191

For any σ-finite measure space, a necessary and sufficient condition that $\|f\|_p \leq F$ is that for any G, $\|fg\|_1 \leq FG$ for all g such that $\|g\|_q \leq G$, i.e:

$$\|f\|_p \leq F \Longleftrightarrow \forall\, g \left(\|g\|_q \leq G \Longrightarrow \|fg\|_1 \leq FG\right)$$

Proof

Hölder's inequality establishes the $\Rightarrow$ direction.

For $\Leftarrow$, Theorem 190 shows that $f \in L^p$, since $\forall \alpha \in \mathbb{R}$, $\|\alpha g\| = |\alpha|\,\|g\|$, so any $g \in L^q$ can be scaled down to make its q-norm meet the conditions of the theorem. Of course, this will also scale down the 1-norm by α, but the original 1-norm would have had to be finite, so $fg \in L^1$ and Theorem 190 establishes that $f \in L^p$.

Now we want to construct g so that f^p and g^q are effectively proportional. Define $\bar{g} = f^{p/q}$ where f is finite, which is a.e, and leave $\bar{g}$ undefined where f is infinite. Since $\int |\bar{g}|^q = \int |f|^p$, $f \in L^p$ implies $\bar{g} \in L^q$. Now let $g = \alpha\bar{g}$ and set α so that $\|g\|_q = G$.

Effective proportionality means equality holds in $\|fg\|_1 = \|f\|_p\,\|g\|_q$. So if $\|f\|_p > F$, then $\|fg\|_1 > FG$, contrary to hypothesis.

□

Theorem 202 (Minkowski's inequality)

Let X and Y be σ-finite measure spaces with measures t and u, respectively, and let f be a complex function on $X \times Y$. Then

$$\left(\int |\textstyle\int f\,dt|^p\,du\right)^{1/p} \leq \int (\textstyle\int |f|^p\,du)^{1/p}\,dt$$

or, more compactly,

$$\|\textstyle\int f\,dt\|_p \leq \displaystyle\int \|f\|_p\,dt$$

where the norms (both here and in the proof) are understood to be with respect to (Y, u).

Proof

Let $J(u) = \int f\,dt$. For $\|J\|_p$ to be bound by a constant, say, $\|J\|_p \leq M$, Theorem 191 says that $\forall g$, $\|g\|_q \leq 1 \Rightarrow \|Jg\|_1 \leq M$.

So assume $\|g\|_q \leq 1$, and compute:

$$\|Jg\|_1 = \int |Jg|\,du = \int |\textstyle\int f\,dt|\,|g|\,du$$

$$\leq \int (\textstyle\int |f|\,dt)\,|g|\,du \qquad \text{Baby Rudin, Theorem 6.25}$$

$$= \int (\textstyle\int |fg|\,du)\,dt \qquad \text{Fubini}$$

$$= \int \|fg\|_1\,dt \leq \int \|f\|_p\,dt \qquad \text{Hölder, and } \|g\|_q \leq 1$$

Take $M = \int \|f\|_p\,dt$ and the theorem follows. Note that while we used q during the proof, it disappeared from the final result, so the theorem is valid for any $p \geq 1$. Note also that we needed the σ-finiteness assumption not only to apply Theorem 191, but also to apply Fubini's theorem. □