On the Cauchy-Schwarz Inequality

by Sambuddha Roy

Goal:

To prove the Cauchy-Schwarz inequality i.e.

$$\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \cdot \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2} \geqslant (x_1 y_1 + x_2 y_2 + \cdots + x_n y_n)$$

. $x = (x_1, x_2, \cdots, x_n)$ and $y = (y_1, y_2, \cdots, y_n)$.

A **restatement** of this is as follows: Let x denote the vector $(x_1, x_2, \cdots, x_n)$ and y denote the vector $(y_1, y_2, \cdots, y_n)$; then $\|x\|_2 \|y\|_2 \geqslant x.y$, where $\|x\|_2$ is the $2-$ norm of a vector x, and $x.y$ denotes the usual *inner product* of two vectors x and y.

Special Cases:

As is usual, when we approach a problem, we try to tease out the building blocks of the problem so to say - i.e. the *real reasons* that make the problem hard (if it is indeed hard). A natural way is to understand some special cases, explore the "neighborhood" (so to say) of the problem.

For this problem, there are essentially two (independent) dimensions: one is p (note that q is related to p by the relation $\frac{1}{p} + \frac{1}{q} = 1$, and the other dimension is n.

We proceed to enumerate some special cases here.

The case $p = q = 2$

We notice that this special case i.e.

$$\|x\|_2 \|y\|_2 \geqslant x.y$$

is nothing but the famous **Cauchy-Schwarz** inequality. We will assume this as proven; for instance, see the blog article on Cauchy Schwarz.

The next couple of special cases deal with the other independent parameter n:

The case $n=1$

This is tantamount to proving

$$x^p/p+y^q/q \geqslant x.y$$

where p and q satisfy $1/p+1/q=1$.

Let us call this the **basic inequality.**

For $p=q=2$, this is just the simple AM-GM for two quantities. We take motivation from that, and remove pesky fractions first by letting N be an integer such that $N/p, N/q$ are both integers; denote $P_1=N/p$ and $Q_1=N/q$, so that $pP_1=qQ_1=N$.

The inequality to be proven then translates to

$$(P_1x^p+Q_1y^q)/N \geqslant x.y$$

of course note that $N=P_1 + Q_1$. Now it should be clear that this follows from AM-GM. Take P_1 quantities equal to x^p and Q_1 quantities equal to y^q. Following through with the $\mathrm{AM}-\mathrm{GM}$ inequality yields the inequality above.

The case of equality: is achieved when $x^p=y^q$.

The case $n=2$

Here, the desired inequality is

$$(x_1^p+x_2^p)^{1/p}\cdot(y_1^q+y_2^q)^{1/q} \geqslant x.y$$

We can write down two $n=1$ inequalities (one for x_1 and y_1 and the other for x_2, y_2) as specified by the basic inequality, and add the two inequalities together to get:

$$\frac{x_1^p+x_2^p}{p}+\frac{y_1^q+y_2^q}{q} \geqslant x.y$$

Where $x = (x_1, x_2)$ and $y = (y_1, y_2)$. Call this the **combo inequality**. How is this inequality faring with regard to what we want to prove? It is easy to see that the LHS of this last inequality is actually larger than the LHS desired for the case $n = 2$ of Holder's Inequality (How? This is by another application of the *basic inequality*.) So this inequality is faring *worse* than the inequality we need to prove.

It seems that we are stumped, right?

Another Try:

Terry Tao and others suggest that in order to milk the most out of an inequality try to apply it to the scenario where *the case of equality can potentially hold*. Thus we realise that we would not be able to get from the *basic inequality* to Holder as above, because the case of equality can be widely flouted ($x_1^p + x_2^p$ may be widely different from $y_1^q + y_2^q$). Of course, if both $x_1^p \approx y_1^q$, and $x_2^p \approx y_2^q$, then the foregoing cannot happen. What if $x_1^p \gg y_1^q$ but $x_2^p \ll y_2^q$? In this case almost anything can happen: either $x_1^p + x_2^p \approx y_1^q + y_2^q$ or $x_1^p + x_2^p \gg y_1^q + y_2^q$ or $x_1^p + x_2^p \ll y_1^q + y_2^q$. How do we then make the claim in Holder's inequality?

Enter Lagrangian Multipliers:

This is a situation commonly encountered in optimization where two *disparate* quantities enter some objective function to be optimized. How do such components play off against each other? Suppose you had an objective function consisting of the *sum of the prices of pencils + sum of prices of houses in India.* It is clear that the components in this (albeit artificial) objective function is grossly disparate; thus one of the components (in this case, price of houses) have overarching *influence* over the objective function. We need something that would equate two disparate components of such an objective function (if we do mean the price of pencils to play a role in our objective function at all). How do we accomplish that? One easy way out is to multiply one of the components in such objective functions by a number; such considerations are the underpnnings of what are called *Lagrangian multipliers*. Such multipliers can be thought of as "manual knobs", or an "exit-for-domain-knowledge" in order to tweak the *relative importances* of the components of the objective function.

Rather surprisingly, this simple idea underscores various intricate algorithms in optimization as well as machine learning.

Armed with this knowledge, let's proceed to apply this to our situation. Take a Lagrangian multiplier λ as follows:

$$\frac{x_1^p+x_2^p}{p}+\frac{(\lambda y_1)^q+(\lambda y_2)^q}{q}\geqslant x.(\lambda y)$$

and simplifying we get that

$$\frac{x_1^p+x_2^p}{p}+\lambda^q\cdot\frac{y_1^q+y_2^q}{q}\geqslant \lambda x.y$$

Note that we derive this inequality by applying the **combo inequality** to $x_1, \lambda y_1$ and $x_2, \lambda y_2$. Also note that we really should be calling λ^q as the Lagrangian multiplier.

Now we can at least hope to achieve the case of equality, and we set λ accordingly:

$$\lambda^q=\frac{x_1^p+x_2^p}{y_1^q+y_2^q}$$

This expression looks quite formidable, but we are almost there! From the above, the expression on the LHS now reads simply as $(x_1^p+x_2^p)$ while the RHS reads as $\lambda x.y$. So then we have that:

$$\frac{x_1^p+x_2^p}{\lambda}\geqslant x.y$$

and now we simplify this to:

$$\frac{x_1^p+x_2^p}{(x_1^p+x_2^p)^{1/q}}\cdot(y_1^q+y_2^q)^{1/q}\geqslant x.y$$

and noting that $1-1/q=1/p$, we are done:

$$(x_1^p+x_2^p)^{1/p}\cdot(y_1^q+y_2^q)^{1/q}\geqslant x.y$$

The case of n quantities throws up no new surprises, and so we have proven Holder's inequality for ourselves.

The Bigger Picture

The basic inequality can be interpreted as the *Fenchel-Young inequality* for the norm function x^p/p. The inequality then just says that the Fenchel dual of this function is y^q/q where $1/p + 1/q = 1$. From the Fenchel Young Inequality, we are deriving Holder's inequality.

The bigger question is whether the Fenchel Young inequality is in a sense stronger than Holder's inequality. Holder's inequality is essentially a statement about *dual norms*. Given a norm $\|\cdot\|$ and its dual $\|\cdot\|_*$, Holder's inequality states:

$$\|x\| \cdot \|y\|_* \geqslant x.y$$

Fenchel Young Inequality on the other hand deals with the *Fenchel dual* of functions; given a function f and its Fenchel dual f^*, it holds that

$$f(x) + f^*(y) \geqslant x.y$$

For every norm $\|\cdot\|$ is there a corresponding function f, such that the following chain holds?

$$f(x) + f^*(y) \geqslant \|x\| \cdot \|y\|_* \geqslant x.y$$