A BRIEF INTRODUCTION TO THE CAUCHY-SCHWARZ AND HÖLDER INEQUALITIES

LEO GOLDMAKHER

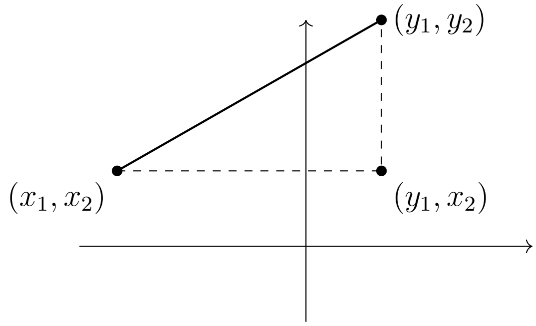
ABSTRACT. The Cauchy-Schwarz inequality is fundamental to many areas of mathematics, physics, engineering, and computer science. We introduce and motivate this inequality, show some applications, and indicate some generalizations, including a simpler form of Hölder's inequality than is usually presented.

1. MOTIVATING CAUCHY-SCHWARZ

Recall that the standard distance (aka "Euclidean metric") between two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ is defined

$$d(\mathbf{x}, \mathbf{y}) := \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2},$$

a simple consequence of the pythagorean theorem.



Calculating the distance between $\mathbf{x} = (x_1, x_2)$ *and* $\mathbf{y} = (y_1, y_2)$

The Euclidean metric famously satisfies the triangle inequality

$$d(\mathbf{x}, \mathbf{y}) \le d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}), \tag{1.1}$$

which asserts that the shortest path from one point of the plane to another is via a straight line. This is intuitively clear, but the proof isn't immediately obvious. We warm up by thinking about a special case:

Proposition 1.1. $d(\mathbf{x}, \mathbf{y}) \le d(\mathbf{x}, 0) + d(0, \mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$.

Exercise 1. Show that the proposition implies (1.1).

Scratchwork for proof of the Proposition:

Writing out the claim in terms of the formal definition of the Euclidean distance, we see that we're trying to prove

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \le \sqrt{x_1^2 + x_2^2} + \sqrt{y_1^2 + y_2^2}.$$

Squaring both sides and simplifying shows this is equivalent to proving

$$-2x_1y_1 - 2x_2y_2 \le 2\sqrt{x_1^2 + x_2^2}\sqrt{y_1^2 + y_2^2}.$$

Dividing by 2 and squaring both sides, we see that our claim is equivalent to

$$(x_1y_1 + x_2y_2)^2 \le (x_1^2 + x_2^2)(y_1^2 + y_2^2). \tag{†}$$

Thus the triangle inequality for the Euclidean metric in $\mathbb{R}^2$ is equivalent to the inequality (†). A similar argument shows that the triangle inequality for the Euclidean metric in $\mathbb{R}^n$ is equivalent to the following:

Theorem 1.2 (Cauchy-Schwarz)**.** *For any real numbers x_i, y_i we have*

$$(x_1y_1 + x_2y_2 + \cdots + x_ny_n)^2 \leq (x_1^2 + x_2^2 + \cdots + x_n^2)(y_1^2 + y_2^2 + \cdots + y_n^2).$$

We'll prove this below, but for now let's explore some easy consequences of this inequality. Taking all the $y_i = 1$, we immediately deduce

$$(x_1 + x_2 + \cdots + x_n)^2 \leq (x_1^2 + x_2^2 + \cdots + x_n^2)n. \tag{*}$$

Rearranging yields

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \leq \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}},$$

or in words, that the mean is bounded above by the root mean square. This already hints at a connection to probability and statistics. Inspired by this, let's impose the additional condition $x_1 + x_2 + \cdots + x_n = 1$ (the sum of probabilities of all possible events is 1); if this is the case, $(*)$ implies

$$x_1^2 + x_2^2 + \cdots + x_n^2 \geq \frac{1}{n}.$$

For example, this shows that if I roll a 6-sided die twice, the probability that both rolls produce the same number is at least $1/6$ – *no matter whether or not the die is fairly weighted.*

Let's reconsider the original Cauchy-Schwarz inequality from a different perspective. What does the quantity $x_1y_1 + x_2y_2 + \cdots + x_ny_n$ remind you of? The dot product of $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$! Thus we can rewrite Cauchy-Schwarz in the more compact form

$$(\mathbf{x} \cdot \mathbf{y})^2 \leq (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y}).$$

This change of perspective is not merely notationally convenient, but also suggests a short proof. Recall that for any two vectors, $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}|\ |\mathbf{b}| \cos\theta$ where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$ (this is a consequence of the law of cosines).

Proof of Cauchy-Schwarz. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have

$$\begin{aligned}(\mathbf{x} \cdot \mathbf{y})^2 &= (|\mathbf{x}|\ |\mathbf{y}| \cos\theta)^2 = |\mathbf{x}|^2\ |\mathbf{y}|^2 \cos^2\theta \\ &\leq |\mathbf{x}|^2\ |\mathbf{y}|^2 = (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y}).\end{aligned}$$

□

In fact, examining this proof we see that equality holds in Cauchy-Schwarz iff the angle between $\mathbf{x}$ and $\mathbf{y}$ is a multiple of π, or in other words, iff $\mathbf{x}$ is a rescaling of $\mathbf{y}$. Thus, we can write the theorem in a stronger form:

Theorem 1.3 (Cauchy-Schwarz, v2.0)**.** *Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have*

$$(\mathbf{x} \cdot \mathbf{y})^2 \leq (\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})$$

with equality if and only if $\mathbf{x}$ is a rescaling of $\mathbf{y}$.

This refinement of Cauchy-Schwarz is quite useful, as the following example demonstrates.

Example 1. What is the maximum of $2x + 3y + 5z$ over all the points (x, y, z) on the unit sphere?

By Cauchy-Schwarz we have

$$(2x + 3y + 5z)^2 \leq (2^2 + 3^2 + 5^2)(x^2 + y^2 + z^2) = 38,$$

with equality iff (x, y, z) is a rescaling of $(2, 3, 5)$. Thus the function $2x + 3y + 5z$ takes on a maximal value of $\sqrt{38}$, and this happens precisely at the points $\pm\left(\frac{2}{\sqrt{38}}, \frac{3}{\sqrt{38}}, \frac{5}{\sqrt{38}}\right)$.

We've now seen applications of Cauchy-Schwarz to problems in geometry, probability, and optimization. This is the tip of the iceberg; Cauchy-Schwarz is extremely useful throughout mathematics, physics, engineering, and computer science.

2. Generalizing Cauchy-Schwarz

Returning to Cauchy-Schwarz in the form presented in Theorem 1.2, observe that the right hand side of the inequality doesn't change if we replace any of the x_i by $-x_i$. Right away we deduce the stronger bound

$$(|x_1y_1| + |x_2y_2| + \cdots + |x_ny_n|)^2 \leq (x_1^2 + x_2^2 + \cdots + x_n^2)(y_1^2 + y_2^2 + \cdots + y_n^2).$$

The take-home message here is that it suffices to state the Cauchy-Schwarz inequality (as well as other inequalities) for the special case of *non-negative* real numbers; inserting absolute values often produces a stronger version for free!

As a student I found Cauchy-Schwarz difficult to remember, and came up with the following mnemonic:

$$\boxed{\textit{The square of the sum of products} \leq \textit{the product of the sum of squares.}}$$

To be useful one must remember the direction of the inequality, but this doesn't need to be memorized: a single random example suffices to determine this, e.g. $\mathbf{x} = (1,0)$ and $\mathbf{y} = (0,1)$. As with any mnemonic, however, there's a possibility of misinterpretation. In this case, the word *product* is potentially misleading; how many variables are we taking the product of? This is important, since taking the trivial product of a single variable leads to the false conclusion $(a+b)^2 \leq a^2 + b^2$. What we actually mean by "product" is the product of *two* variables. Thus a more precise version of the mnemonic might read *the square of the sum of products (of two variables)* $\leq$ *the product of the sum of the squares*, but this is clunkier so in practice I suppress the parenthetical remark.

One nice feature of the above mnemonic is that it inspires generalizations. For example, one might guess

$$\boxed{\text{The cube of the sum of products (of three variables)} \leq \text{the product of the sum of cubes.}}$$

This turns out to be true, with one caveat: all the variables have to be non-negative. (As above, one can remove this restriction by inserting absolute values into the inequality.) Taking this idea and running with it, we might be led to conjecture the following:

Theorem 2.1 (Hölder's inequality)**.** *For any positive integer m, we have*

$$\Big(\sum_i \underbrace{\alpha_i\beta_i\cdots\omega_i}_{m\ \textit{variables}}\Big)^m \leq \underbrace{\left(\sum_i \alpha_i^m\right)\left(\sum_i \beta_i^m\right)\cdots\left(\sum_i \omega_i^m\right)}_{m\ \textit{factors}} \tag{$\heartsuit$}$$

assuming all of the $\alpha_i, \beta_i, \ldots, \omega_i$ are non-negative real numbers.

The proof of this is outlined in the exercises. Just as Cauchy-Schwarz is the natural tool for proving the triangle inequality in $\mathbb{R}^n$ with respect to the Euclidean metric, Hölder's inequality is useful for proving the triangle inequality in some other spaces that arise in analysis (called L^p spaces).

Recasting Cauchy-Schwarz in the language of vectors (as in Theorem 1.3) provides a different avenue for generalization, to arbitrary vector spaces endowed with an *inner product* (a generalization of the dot product). We won't discuss this here, but you will encounter this in any course on functional analysis or Hilbert spaces. An example of this is a version of Cauchy-Schwarz for integrals rather than sums; see exercise **1.d** below.

3. Exercises

1. Practice with Cauchy-Schwarz:

(a) Prove that $|x_1y_1 + \cdots + x_ny_n|^2 \leq (|x_1|^2 + \cdots + |x_n|^2)(|y_1|^2 + \cdots + |y_n|^2)$ for any $x_i, y_i \in \mathbb{C}$.

(b) Prove that for any $a, b, c > 0$,

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} \geq \frac{3}{2}.$$

(c) Given $\theta_1, \theta_2, \cdots, \theta_n \in [0, \frac{\pi}{2})$, prove that

$$\frac{1}{\cos^2\theta_1} + \cdots + \frac{1}{\cos^2\theta_n} \geq n + \frac{(\sin\theta_1 + \cdots + \sin\theta_n)^2}{\cos^2\theta_1 + \cdots + \cos^2\theta_n}.$$

(d) Prove that $\left|\int_a^b f(x)g(x)\,dx\right|^2 \leq \Big(\int_a^b |f(x)|^2\,dx\Big)\Big(\int_a^b |g(x)|^2\,dx\Big)$.

2. The purpose of this problem is to discover a few more proofs of Cauchy-Schwarz.

(a) Prove directly that $(x_1y_1+x_2y_2)^2 \leq (x_1^2+x_2^2)(y_1^2+y_2^2)$. Then deduce Theorem 1.2 by induction.

(b) Given real numbers a_i and b_i, consider the function

$$F(x) := \sum_i (a_ix - b_i)^2.$$

Observe that $F(x)$ is quadratic in x, i.e. can be written in the form Ax^2+Bx+C. What can you deduce about the discriminant B^2-4AC? Use this to prove Theorem 1.3.

(c) Prove that

$$\left(\sum_i x_iy_i\right)^2 = \left(\sum_i x_i^2\right)\left(\sum_i y_i^2\right) - \frac{1}{2}\sum_{i,j}(x_iy_j - x_jy_i)^2,$$

and deduce Theorem 1.3.

3. The goal of this problem is to prove Theorem 2.1, i.e. that the inequality ($\heartsuit$) holds for all positive integers m.

(a) Prove that if ($\heartsuit$) holds when $m=a$ and when $m=b$, then it also must hold for $m=ab$.

(b) Prove that ($\heartsuit$) holds whenever m is a power of 2.

(c) Prove that for any non-negative real numbers a_n, b_n, c_n we have

$$\sum a_nb_nc_n \leq \Big(\sum a_nb_nc_n\Big)^{1/4}\Big(\sum a_n^3\Big)^{1/4}\Big(\sum b_n^3\Big)^{1/4}\Big(\sum c_n^3\Big)^{1/4}.$$

(d) Prove ($\heartsuit$) for $m=3$. [*Hint: use the previous part!*]

(e) Prove Theorem 2.1. [*Hint: use strong induction. m is either even or odd...*]

4. Frequently in the literature, *Hölder's inequality* refers to the bound

$$a_1b_1+\cdots+a_nb_n \leq (a_1^p+\cdots+a_n^p)^{1/p}(b_1^q+\cdots+b_n^q)^{1/q} \tag{$\ddagger$}$$

for any positive real numbers p, q satisfying $1/p+1/q=1$ and any non-negative $a_i, b_i \in \mathbb{R}$. The goal of this exercise is to prove that ($\ddagger$) is equivalent to Theorem 2.1.

(a) Prove that ($\ddagger$) implies Theorem 2.1.

(b) Show that Theorem 2.1 implies ($\ddagger$) for all $p \in \mathbb{Q}$. [*Hint: Warm up with the case $p=3$.*]

(c) Deduce that ($\ddagger$) holds for all real numbers $p>1$. [*Hint: Think of the numbers a_i, b_i being fixed and of p as the variable.*]

5. In this exercise we outline a beautiful argument (due to Orr Shalit) that deduces ($\ddagger$) directly from Cauchy-Schwarz.[1] Throughout, let

$$S := \{1/p \in (0,1) : \text{the bound } (\ddagger) \text{ holds for all positive real choices of } a_i, b_i\}.$$

(a) Prove that if $\frac{1}{p} \in S$, then $\frac{1}{2p} \in S$. [*Hint: Write $a_kb_k = (a_kb_k^\sigma)(b_k^{1-\sigma})$, and apply Cauchy-Schwarz to $\sum a_kb_k$. Then choose σ appropriately.*]

(b) Deduce that $\frac{a}{2^k} \in S$ for all positive integers k and all $a \in \{1,2,\ldots,2^k-1\}$. [*Hint: Do you happen to know any elements of S a priori?*]

(c) Conclude that $S=(0,1)$.

DEPT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA, USA 01267
Email address: Leo.Goldmakher@williams.edu

[1]For a completely different proof of ($\ddagger$), see my article *Hölder's inequality as a convexity result*, available online.