Cauchy-Schwartz inequality and geometric incidence problems

Misha Rudnev

Abstract

In this lecture, I will discuss a couple of examples showing what a nice way of bookkeeping the Cauchy-Schwartz inequality is. They are the box inequality and the geometric incidence inequality. There are no prerequisites to be able to follow the exposition, except familiarity with the sum notations.

Notation

Recall that a shortcut to denote the sum of $n > 1$ numbers $a_1, a_2, \ldots, a_n$ is

$$a_1 + a_2 + \ldots + a_n = \sum_{i=1}^{n} a_i. \tag{1}$$

The index of summation is a dummy variable. It runs from 1 through n, marking the identity of the numbers $a_1, \ldots, a_n$ and can be tinkered with, renamed, etc., similar to how this is done with definite integrals. I.e.

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j = \sum_{\alpha=5}^{n+4} a_{\alpha-4} = a_1 + \ldots + a_n.$$

The only thing that matters is the individuality of each of the numbers $a_1, \ldots, a_n$.

From basic rules of addition and multiplication it follows that if c is a constant,

$$\begin{array}{rcccl} c\sum_{i=1}^n a_i & = & \sum_{i=1}^n (ca_i), & & \\ \left(\sum_{i=1}^n a_i\right)\cdot\left(\sum_{j=1}^m b_j\right) & = & \sum_{i=1}^n a_i\left(\sum_{j=1}^m b_j\right) & = & \sum_{i=1}^n\sum_{j=1}^m a_ib_j, \\ \left(\sum_{i=1}^n a_i\right)^2 & = & \sum_{i=1}^n\sum_{j=1}^n a_ia_j, & & \end{array} \tag{2}$$

an this is all that's going to be needed as far as algebra is concerned. *Warning:* above and throughout, the symbol $\cdot$ is used most of the time to denote the usual multiplication of real numbers. It is only in the ensuing Proof 2 that it stands for the dot product of vectors, which are printed in boldface.

A separate geometric issue is that if E is a set in $\mathbb{R}^d$, i.e. a collection of points on the real line for $d = 1$, in the plane for $d = 2$, in space for $d = 3$, etc., then rather than dealing with the set itself, it is convenient to work with the *characteristic function* $\chi_E(\mathbf{x})$ of E, defined as follows:

$$\chi_E(\mathbf{x}) = \begin{cases} 1 & \text{for} \quad \mathbf{x} \in E, \\ 0 & \text{for} \quad \mathbf{x} \notin E. \end{cases} \tag{3}$$

Observe that generally $\mathbf{x}$ is meant as a vector here: if $\mathbf{x} \in \mathbb{R}^d$, then $\mathbf{x} = (x_1, \ldots, x_d)$. So the characteristic function tests each point $\mathbf{x}$ on whether or not it is a member of E: if yes it returns one, otherwise zero. It's quite convenient for intersections for instance: clearly $\chi_{E_1 \cap E_2}(\mathbf{x}) = \chi_{E_1}(\mathbf{x})\chi_{E_2}(\mathbf{x})$ for two sets E_1 and E_2.

A nice property of any characteristic function, which immediately follows from definition (3) is that

$$\chi_E^2 = \chi_E, \quad \text{for any set} \ \ E. \tag{4}$$

Cauchy-Schwartz inequality

The Cauchy-Schwartz is implied by convexity of the parabola, or the fact that the derivative of $y = x^2$, equal to 2 is positive, for all x. Here come two rather similar ways to derive it.

Theorem 1 (Cauchy-Schwartz inequality). *If $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are nonzero real numbers, then*

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^n b_i^2\right)^{\frac{1}{2}}, \tag{5}$$

with the equality only if $a_i = cb_i$, for all $i = 1, \ldots, n$ and some fixed constant c.

The requirement that the a's and b's be nonzero is only because otherwise they certainly do not contribute in either sum.

Proof 1: If a, b are two real numbers, then $(a-b)^2 \geq 0$, with the equality only if $a = b$. Opening brackets results in

$$ab \leq \frac{a^2 + b^2}{2}. \tag{6}$$

Returning to (5), let us denote

$$A = \left(\sum_{i=1}^n a_i^2\right)^{\frac{1}{2}}, \quad B = \left(\sum_{i=1}^n b_i^2\right)^{\frac{1}{2}}. \tag{7}$$

They are both non-zero. So we can write a truism

$$\sum_{i=1}^n a_i b_i = AB \sum_{i=1}^n \frac{a_i}{A} \cdot \frac{b_i}{B},$$

and now apply (6) n times, for each i:

$$\frac{a_i}{A} \cdot \frac{b_i}{B} \leq \frac{\frac{a_i^2}{A^2} + \frac{b_i^2}{B^2}}{2}.$$

Hence

$$\sum_{i=1}^n a_i b_i \leq AB \frac{\frac{\sum_{i=1}^n a_i^2}{A^2} + \frac{\sum_{i=1}^n b_i^2}{B^2}}{2} = AB,$$

which proves the inequality (5). Indeed, by (7) each of the summands on top of the big fraction above equals 1. The claim about the constant c can also be recovered from above, but it's seen even easier in the second proof that follows.

Proof 2: Consider two n-dimensional vectors, i.e arrays of reals, $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$. They can be added component-wise as well as multiplied by real numbers, also component-wise. We can also define the dot product:

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + \ldots + a_n b_n = \sum_{i=1}^{n} a_i b_i,$$

this is the left-hand side of (5).

Defined as it is, for any $\mathbf{a}$, one can see that $\mathbf{a} \cdot \mathbf{a} \geq 0$, with the equality only if every single component of $\mathbf{a}$ is zero. Note that $\mathbf{a} \cdot \mathbf{a} = A^2$, where A has been defined in (7) and appears in the right-hand side of (5).

Introduce a real variable c, consider a vector $\mathbf{a} - c\mathbf{b}$. Then $(\mathbf{a} - c\mathbf{b}) \cdot (\mathbf{a} - c\mathbf{b}) \geq 0$, with the equality only if $a_i = cb_i$, for all i. Now open the parentheses and get for all c:

$$(\mathbf{b} \cdot \mathbf{b})\, c^2 - 2(\mathbf{a} \cdot \mathbf{b})\, c + (\mathbf{a} \cdot \mathbf{a}) \; > \; 0,$$

as long as $\mathbf{a} \neq c\mathbf{b}$.

This is a quadratic inequality in c, and the fact that is greater than zero for all c means that the corresponding quadratic equation has no roots. Thus we must have

$$4(\mathbf{a} \cdot \mathbf{b})^2 - 4(\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) < 0,$$

or equal to zero provided that $a_i = cb_i$, for all i. After cancelling 4, we get precisely the Cauchy-Schwartz inequality (5), in which squares have been taken of both sides.

Box inequality

How many mirrors does a hippo need to see that it's fat? The answer is, no more than three.

To formalise the question, consider a set E of N points in space $\mathbb{R}^3$. Write $|E| = N$ to say that E has N points. The bigger N, the fatter the set. Now consider the projections of the set E to the coordinate planes. Let $\pi_z(E)$ denote the projection of E onto the coordinate plane (x, y), along the z-axis. Similarly, define the projections $\pi_x(E)$ and $\pi_y(E)$ along the x and y axes. Note that the elements of the set $\pi_z(E)$ are points on the xy coordinate plane.

Denote $|\pi_z(E)|$ the number of elements, or size, of the projection along the z-axis, same for the two other projections.

Theorem 2 (Box inequality). *At least one of the projection sizes $|\pi_z(E)|$, $|\pi_x(E)|$, $|\pi_y(E)|$ is as large as $N^{\frac{2}{3}}$. I.e.*

$$\max_{x,y,z}(|\pi_z(E)|,\, |\pi_x(E)|,\, |\pi_y(E)|) \;\geq\; |E|^{\frac{2}{3}}. \tag{8}$$

Before proving this theorem, let us see what happens in two dimensions. If E is a set of N points with coordinates (x,y), let $\pi_x(E)$ be its projection along the x-axis (i.e. onto the y-axis) and $\pi_y(E)$ be its projection along the y-axis (i.e. onto the x-axis). Let us show that in this case

$$\max_{x,y}(|\pi_x(E)|,\,|\pi_y(E)|)\;\geq\;\sqrt{N}. \tag{9}$$

Indeed, suppose one of the projections, say $\pi_y(E)$ has the size smaller than $\sqrt{N}$. I.e. $|\pi_y(E)| = c\sqrt{N}$, with $c\leq 1$. The prototype of $x\in\pi_y(E)$ in the set E is a vertical column of points of E over x that project onto x. If we have the total of N points in the plane arranged into $c\sqrt{N}$ distinct columns, then some column has at least $\frac{N}{c\sqrt{N}}\geq\sqrt{N}$ points. This proves(9).

However, the same argument in three dimensions is harder (yet not impossible) to pull through – try it!

To prove (8), let's use characteristic functions, see the Notation section, and Cauchy-Schwartz. For simplicity, let us denote $E_1 = \pi_z(E)$, $E_2 = \pi_x(E)$, $E_3 = \pi_y(E)$. To reduce the number of indices further, let us denote $\chi(x,y,z)$ the characteristic function of E and $\chi_1(x,y)$, $\chi_2(y,z)$, $\chi_3(z,x)$ the characteristic functions of the projections E_1, E_2, E_3 of E, respectively.

First off, note the characteristic functions inequality:

$$\chi(x,y,z)\;\leq\;\chi_1(x,y)\chi_2(y,z)\chi_3(z,x). \tag{10}$$

It simply says that if the point $(x,y,z)\in E$, then it *has* projections, i.e. $(x,y)\in E_1$, $(y,z)\in E_2$, and $(z,x)\in E_3$. Note that (10) is a bona fide inequality. *Exercise:* give an example of E, such that the right hand side is strictly greater than the left-hand side for some point (x,y,z).

And now observe that

$$N = |E| = \sum_{x,y,z}\chi(x,y,z). \tag{11}$$

Namely every time $(x,y,z)\in E$, the sum above picks up 1. Then, using (10), we have

$$N\leq\sum_{x,y,z}\chi_1(x,y)\chi_2(y,z)\chi_3(z,x) = \sum_{x,y}\chi_1(x,y)\sum_z\chi_2(y,z)\chi_3(z,x).$$

Now let us use Cauchy-Schwartz (5), plus the fact that in the proof of this formula we did not care how many terms there were under the sum, or how they were indexed. I.e. (5) is true for *any* summations, no matter how many indices there are. Then

$$\sum_{x,y}\chi_1(x,y)\left(\sum_z\chi_2(y,z)\chi_3(z,x)\right)\leq\left(\sum_{x,y}\chi_1^2(x,y)\right)^{\frac12}\left(\sum_{x,y}\left[\sum_z\chi_2(y,z)\chi_3(z,x)\right]^2\right)^{\frac12}. \tag{12}$$

The first multiplier simply equals $\sqrt{|E_1|}$. Indeed, $\chi_1^2=\chi_1$ as in (4), and the sum just does the accounting for the projection onto the xy plane, similar to (11) for E itself.

In the second sum, let us use Cauchy-Schwartz again regarding the summation in z. For each (x,y) we have

$$\left[\sum_z\chi_2(y,z)\chi_3(z,x)\right]^2\leq\left(\sum_z\chi_2^2(y,z)\right)\cdot\left(\sum_z\chi_3^2(z,x)\right).$$

Once again, the squares in the right-hand side can be removed by (4). So, as the summations in the right hand side above are independent, now we've got

$$\sum_{x,y}\left[\sum_z \chi_2(y,z)\chi_3(z,x)\right]^2 \le \left(\sum_{y,z}\chi_2(y,z)\right)\cdot\left(\sum_{z,x}\chi_3(z,x)\right) = |E_2|\,|E_3|.$$

Returning now to (12) and further back, we have

$$|E| \le \sqrt{|E_1|}\sqrt{|E_2|}\sqrt{|E_3|}. \tag{13}$$

This does the job, because it means that if $\max(\sqrt{|E_1|},\ \sqrt{|E_2|},\ \sqrt{|E_3|})$ is the largest of the three terms in the right-hand side, then

$$|E| \le [\max(\sqrt{|E_1|},\ \sqrt{|E_2|},\ \sqrt{|E_3|})]^3.$$

This is equivalent to (8).

Problem 3. Try to argue that if E is now a "continuous" set of volume V in space, the same argument goes through, and one of the three projections – or the mirrors – cast by the set E – or the hippo – has the area of at least $V^{\frac{2}{3}}$.

Problem 4. Give an example of a set E, showing that the box inequality is sharp. I.e. when the largest projection size equals precisely $|E|^{\frac{2}{3}}$.

Problem 5. The inequalities (9) and (8) suggest that if the set E now lives in $\mathbb{R}^d$, in $d \ge 2$ dimensions, then the largest projection of E on one of the $d-1$ dimensional coordinate hyperplanes has the size of at least $|E|^{\frac{d-1}{d}}$. Try to copy the argument in the proof for $d = 3$ to show this. You will need a generalisation of the Cauchy-Schwartz inequality, called the Hölder inequality. It says that if $p > 1$ and q is such that $\frac{1}{p} + \frac{1}{q} = 1$ (in particular $p = 2$ implies $q = 2$), then

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

Incidences between lines and points in the plane

Suppose, you have n distinct points in the plane and m distinct straight lines. An *incidence counting function* δ_{pl} equals 1 when the point p sits on the line l and is zero otherwise. So the total number of incidences

$$I = \#\{(p,l) : p \in l\} \ = \sum_{p=1}^{n}\sum_{l=1}^{m}\delta_{pl}. \tag{14}$$

What's the best upper bound for I, no matter how the points and lines are situated? The obvious one is mn, when each point sits on each line, but this is hardly possible to arrange for.

Theorem 6. *The total number of incidences*

$$I \ \le\ 2\max(m\sqrt{n},\ n\sqrt{m}). \tag{15}$$

To prove it, we shall apply Cauchy-Schwartz to (14) and use the following geometric principles: (i) two distinct lines intersect at most at one point, and (ii) no more than one line passes through a given pair of distinct points.

Now, by Cauchy-Schwartz

$$I = \sum_{p=1}^{n} 1 \cdot \left(\sum_{l=1}^{m} \delta_{pl} \right) \leq \sqrt{n} \sqrt{\sum_{p=1}^{n} \left(\sum_{l=1}^{m} \delta_{pl} \right)^2} = \sqrt{n} \sqrt{\sum_{p=1}^{n} \sum_{l=1}^{m} \sum_{\lambda=1}^{m} \delta_{pl} \delta_{p\lambda}}, \tag{16}$$

see (2) for the last step. Now consider the double sum in all pairs of lines (l, λ):

$$\sum_{l=1}^{m} \sum_{\lambda=1}^{m} \delta_{pl} \delta_{p\lambda} = \sum_{l=1}^{m} \delta_{pl} \delta_{pl} + \sum_{l,\lambda=1,\ldots,m;\, l\neq\lambda} \delta_{pl} \delta_{p\lambda}.$$

The first term simply corresponds to the case $l = \lambda$, in which case $\delta_{pl}\delta_{p\lambda} = \delta_{pl}\delta_{pl} = \delta_{pl}$, similar to (4). So

$$\sum_{p=1}^{n} \sum_{l=1}^{m} \delta_{pl} \delta_{pl} = \sum_{p=1}^{n} \sum_{l=1}^{m} \delta_{pl} = I.$$

Hence, changing the order of summation in (16), we have

$$I \leq \sqrt{n} \sqrt{I + \sum_{l,\lambda=1,\ldots,m;\, l\neq\lambda} \left(\sum_{p=1}^{n} \delta_{pl} \delta_{p\lambda} \right)}.$$

Given a pair of distinct lines (l, λ), there is at most one point p, belonging to both, i.e. when $\delta_{pl}\delta_{p\lambda} = 1$. So, the sum in brackets is at most 1, and we proceed as

$$I \leq \sqrt{n} \sqrt{I + \sum_{l,\lambda=1,\ldots,m;\, l\neq\lambda} 1} \leq \sqrt{n}\sqrt{I + m^2}.$$

Now, if $I \geq m^2$, this means $I \leq \sqrt{n}\sqrt{2I}$, i.e. $I \leq 2n$. If $I < m^2$, then we have $I \leq \sqrt{n}\sqrt{2m^2}$. So

$$I \leq 2 \max(n,\, m\sqrt{n}). \tag{17}$$

We can now repeat the procedure, only initially applying Cauchy-Schwartz to the summation in lines l in (14) and (16) rather than in points p. Effectively, m and n now will swap their rôles. The geometric principle to be used now is that there is at most one line that can pass through a given pair of distinct points. This will give us

$$I \leq 2 \max(m,\, n\sqrt{m}). \tag{18}$$

Together with (17), this implies (15). The proof is complete.

Now notice that the geometric principles (i) and (ii) we used were quite general and did not require that the lines be exactly straight. Basically, if we take any curves – let us still refer to them as "lines" – such that (i) there is at most a finite – i.e. independent of "large" numbers m, n – number of points allowed to sit at the intersection of any two distinct lines, and (ii) there is at most a finite number of lines that may pass through any pair of distinct points, the bound (15) still remains true, possibly with a larger constant C substituting 2 in it. In particular, the bound (15) is easily seen to be true as it is, if all the lines are circles of *the same radius* – in this case the "finite number" in clauses (i), (ii) equals 2. If the circles have different radii, the clause (ii) can be violated: one can draw as many circles of different radii passing through a given pair of points as they wish.

Theorem 6 for points and straight lines can be improved, but this is not easy and has been done only in the early 1980s by Szemerédi and Trotter, [2]. The generally unimprovable bound for the number of incidences between m lines and n points, satisfying the geometric conditions (i) and (ii) is

$$I \leq C[m + n + (mn)^{\frac{2}{3}}], \tag{19}$$

for some constant C.

Theorem 6 as it is, for the case of circles of the same radius, already has interesting corollaries.

Corollary 7. *For any collection of $n \geq 2$ distinct points in the plane, the distance 1 between a pair of points cannot occur more than $2n^{\frac{3}{2}}$ times.*

Indeed, draw a unit circle around each point. The number of times the unit distance occurs is the number of incidences in this arrangement of points and unit circles. Now apply (15) with $m = n$.

Corollary 8. *For any collection of $n \geq 2$ distinct points in the plane, there are at least $\frac{1}{8}\sqrt{n}$ distinct distances.*

Indeed, there are $\frac{n^2-n}{2}$ distinct pairs of points, and a single distance, be it 1 or anything else, cannot occur more than $2n^{\frac{3}{2}}$ times, by the previous corollary. So the number of distinct distances is at least $\frac{(n^2-n)/2}{2n\sqrt{n}} \geq \frac{1}{8}\sqrt{n}$.

These results were obtained in 1946 by Erdös, [1], who conjectured in particular that in fact the number of distinct distances should be at least $C\frac{n}{\sqrt{\log n}}$ for some constant C and n large enough. No one has succeeded in proving this so far. The bound (19) plugged into Corollaries 7 and 8, results immediately in some $Cn^{\frac{2}{3}}$ distinct distances. The best bound known today is that the number of distinct distances between a large number n points in the plane is at least $Cn^{.86}$, for some constant C. This result has already been obtained in the XXI century, by Katz and Tardos, [3].

References

[1] P. Erdös. *On a set of distances of n points.* American Math. Monthly **53** (1946) pp. 248–250.

[2] E. Szemerédi and W. Trotter. *Extremal problems in discrete geometry.* Combinatorica **3** (1983) pp. 381–392.

[3] N. Katz and G. Tardos. *A new entropy inequality for the Erdös distance problem.* Comtemp. Maths. **342**. Towards a theory of geometric graphs, pp. 119–126. AMS 2004.