1
Starting with Cauchy

Cauchy's inequality for real numbers tells us that

$$a_1b_1 + a_2b_2 + \cdots + a_nb_n \le \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}\sqrt{b_1^2 + b_2^2 + \cdots + b_n^2},$$

and there is no doubt that this is one of the most widely used and most important inequalities in all of mathematics. A central aim of this course — or master class — is to suggest a path to mastery of this inequality, its many extensions, and its many applications — from the most basic to the most sublime.

The Typical Plan

The typical chapter in this course is built around the solution of a small set of *challenge problems.* Sometimes a challenge problem is drawn from one of the world's famous mathematical competitions, but more often a problem is chosen because it illustrates a mathematical technique of wide applicability.

Ironically, our first challenge problem is an exception. To be sure, the problem hopes to offer honest coaching in techniques of importance, but it is unusual in that it asks you to solve a problem that you are likely to have seen before. Nevertheless, the challenge is sincere; almost everyone finds some difficulty directing fresh thoughts toward a familiar problem.

Problem 1.1 *Prove Cauchy's inequality. Moreover, if you already know a proof of Cauchy's inequality, find another one!*

Coaching for a Place to Start

How does one solve a problem in a fresh way? Obviously there cannot be any universal method, but there are some hints that almost always help. One of the best of these is to try to solve the problem by means of a *specific principle* or *specific technique.*

Here, for example, one might insist on proving Cauchy's inequality just by algebra — or just by geometry, by trigonometry, or by calculus. Miraculously enough, Cauchy's inequality is wonderfully provable, and each of these approaches can be brought to a successful conclusion.

A Principled Beginning

If one takes a dispassionate look at Cauchy's inequality, there is another principle that suggests itself. Any time one faces a valid proposition that depends on an integer n, there is a reasonable chance that mathematical induction will lead to a proof. Since none of the standard texts in algebra or analysis gives such a proof of Cauchy's inequality, this principle also has the benefit of offering us a path to an "original" proof — provided, of course, that we find any proof at all.

When we look at Cauchy's inequality for $n = 1$, we see that the inequality is trivially true. This is all we need to start our induction, but it does not offer us any insight. If we hope to find a serious idea, we need to consider $n = 2$ and, in this second case, Cauchy's inequality just says

$$(a_1b_1 + a_2b_2)^2 \leq (a_1^2 + a_2^2)(b_1^2 + b_2^2). \tag{1.1}$$

This is a simple assertion, and you may see at a glance why it is true. Still, for the sake of argument, let us suppose that this inequality is not so obvious. How then might one search systematically for a proof?

Plainly, there is nothing more systematic than simply expanding both sides to find the equivalent inequality

$$a_1^2b_1^2 + 2a_1b_1a_2b_2 + a_2^2b_2^2 \leq a_1^2b_1^2 + a_1^2b_2^2 + a_2^2b_1^2 + a_2^2b_2^2,$$

then, after we make the natural cancellations and collect terms to one side, we see that inequality (1.1) is also equivalent to the assertion that

$$0 \leq (a_1b_2)^2 - 2(a_1b_2)(a_2b_1) + (a_2b_1)^2. \tag{1.2}$$

This equivalent inequality actually puts the solution of our problem within reach. From the well-known factorization $x^2 - 2xy + y^2 = (x - y)^2$ one finds

$$(a_1b_2)^2 - 2(a_1b_2)(a_2b_1) + (a_2b_1)^2 = (a_1b_2 - a_2b_1)^2, \tag{1.3}$$

and the nonnegativity of this term confirms the truth of inequality (1.2). By our chain of equivalences, we find that inequality (1.1) is also true, and thus we have proved Cauchy's inequality for $n = 2$.

The Induction Step

Now that we have proved a nontrivial case of Cauchy's inequality, we are ready to look at the induction step. If we let $H(n)$ stand for the hypothesis that Cauchy's inequality is valid for n, we need to show that $H(2)$ and $H(n)$ imply $H(n+1)$. With this plan in mind, we do not need long to think of first applying the hypothesis $H(n)$ and then using $H(2)$ to stitch together the two remaining pieces. Specifically, we have

$$\begin{aligned} &a_1b_1 + a_2b_2 + \cdots + a_nb_n + a_{n+1}b_{n+1} \\ &\qquad = (a_1b_1 + a_2b_2 + \cdots + a_nb_n) + a_{n+1}b_{n+1} \\ &\qquad \le (a_1^2 + a_2^2 + \cdots + a_n^2)^{\frac{1}{2}} (b_1^2 + b_2^2 + \cdots + b_n^2)^{\frac{1}{2}} + a_{n+1}b_{n+1} \\ &\qquad \le (a_1^2 + a_2^2 + \cdots + a_n^2 + a_{n+1}^2)^{\frac{1}{2}} (b_1^2 + b_2^2 + \cdots + b_n^2 + b_{n+1}^2)^{\frac{1}{2}}, \end{aligned}$$

where in the first inequality we used the induction hypothesis $H(n)$, and in the second inequality we used $H(2)$ in the form

$$\alpha\beta + a_{n+1}b_{n+1} \le (\alpha^2 + a_{n+1}^2)^{\frac{1}{2}} (\beta^2 + b_{n+1}^2)^{\frac{1}{2}}$$

with the new variables

$$\alpha = (a_1^2 + a_2^2 + \cdots + a_n^2)^{\frac{1}{2}} \quad \text{and} \quad \beta = (b_1^2 + b_2^2 + \cdots + b_n^2)^{\frac{1}{2}}.$$

The only difficulty one might have finding this proof comes in the last step where we needed to see how to use $H(2)$. In this case the difficulty was quite modest, yet it anticipates the nature of the challenge one finds in more sophisticated problems. The actual application of Cauchy's inequality is never difficult; the challenge always comes from seeing *where* Cauchy's inequality should be applied and *what* one gains from the application.

The Principle of Qualitative Inferences

Mathematical progress depends on the existence of a continuous stream of new problems, yet the processes that generate such problems may seem mysterious. To be sure, there is genuine mystery in any deeply original problem, but most new problems evolve quite simply from well-established principles. One of the most productive of these principles calls on us to expand our understanding of a *quantitative* result by first focusing on its *qualitative* inferences.

Almost any significant quantitative result will have some immediate qualitative corollaries and, in many cases, these corollaries can be derived independently, without recourse to the result that first brought them to light. The alternative derivations we obtain this way often help us to see the fundamental nature of our problem more clearly. Also, much more often than one might guess, the qualitative approach even yields new quantitative results. The next challenge problem illustrates how these vague principles can work in practice.

Problem 1.2 *One of the most immediate qualitative inferences from Cauchy's inequality is the simple fact that*

$$\sum_{k=1}^{\infty} a_k^2 < \infty \text{ and } \sum_{k=1}^{\infty} b_k^2 < \infty \text{ imply that } \sum_{k=1}^{\infty} |a_k b_k| < \infty. \tag{1.4}$$

Give a proof of this assertion that does not call on Cauchy's inequality.

When we consider this challenge, we are quickly drawn to the realization that we need to show that the product $a_k b_k$ is small when a_k^2 and b_k^2 are small. We could be sure of this inference if we could prove the existence of a constant C such that

$$xy \le C(x^2 + y^2) \qquad \text{for all real } x, y.$$

Fortunately, as soon as one writes down this inequality, there is a good chance of recognizing why it is true. In particular, one might draw the link to the familiar factorization

$$0 \le (x - y)^2 = x^2 - 2xy + y^2,$$

and this observation is all one needs to obtain the bound

$$xy \le \frac{1}{2}x^2 + \frac{1}{2}y^2 \qquad \text{for all real } x, y. \tag{1.5}$$

Now, when we apply this inequality to $x = |a_k|$ and $y = |b_k|$ and then sum over all k, we find the interesting *additive* inequality

$$\sum_{k=1}^{\infty} |a_k b_k| \le \frac{1}{2}\sum_{k=1}^{\infty} a_k^2 + \frac{1}{2}\sum_{k=1}^{\infty} b_k^2. \tag{1.6}$$

This bound gives us another way to see the truth of the qualitative assertion (1.4) and, thus, it passes one important test. Still, there are other tests to come.

A Test of Strength

Any time one meets a new inequality, one is almost duty bound to test the strength of that inequality. Here that obligation boils down to asking how close the new additive inequality comes to matching the quantitative estimates that one finds from Cauchy's inequality.

The additive bound (1.6) has two terms on the right-hand side, and Cauchy's inequality has just one. Thus, as a first step, we might look for a way to combine the two terms of the additive bound (1.6), and a natural way to implement this idea is to normalize the sequences $\{a_k\}$ and $\{b_k\}$ so that each of the right-hand sums is equal to one.

Thus, if neither of the sequences is made up of all zeros, we can introduce new variables

$$\hat{a}_k = a_k \Big/ \Big(\sum_j a_j^2 \Big)^{\frac{1}{2}} \quad \text{and} \quad \hat{b}_k = b_k \Big/ \Big(\sum_j b_j^2 \Big)^{\frac{1}{2}},$$

which are *normalized* in the sense that

$$\sum_{k=1}^{\infty} \hat{a}_k^2 = \sum_{k=1}^{\infty} \Big\{ a_k^2 \Big/ \Big(\sum_j a_j^2 \Big) \Big\} = 1$$

and

$$\sum_{k=1}^{\infty} \hat{b}_k^2 = \sum_{k=1}^{\infty} \Big\{ b_k^2 \Big/ \Big(\sum_j b_j^2 \Big) \Big\} = 1.$$

Now, when we apply inequality (1.6) to the sequences $\{\hat{a}_k\}$ and $\{\hat{b}_k\}$, we obtain the simple-looking bound

$$\sum_{k=1}^{\infty} \hat{a}_k \hat{b}_k \le \frac{1}{2} \sum_{k=1}^{\infty} \hat{a}_k^2 + \frac{1}{2} \sum_{k=1}^{\infty} \hat{b}_k^2 = 1$$

and, in terms of the original sequences $\{a_k\}$ and $\{b_k\}$, we have

$$\sum_{k=1}^{\infty} \Big\{ a_k \Big/ \Big(\sum_j a_j^2 \Big)^{\frac{1}{2}} \Big\} \Big\{ b_k \Big/ \Big(\sum_j b_j^2 \Big)^{\frac{1}{2}} \Big\} \le 1.$$

Finally, when we clear the denominators, we find our old friend Cauchy's inequality — though this time it also covers the case of possibly infinite sequences:

$$\sum_{k=1}^{\infty} a_k b_k \le \Big(\sum_{j=1}^{\infty} a_j^2 \Big)^{\frac{1}{2}} \Big(\sum_{j=1}^{\infty} b_j^2 \Big)^{\frac{1}{2}}. \tag{1.7}$$

The additive bound (1.6) led us to a proof of Cauchy's inequality which is quick, easy, and modestly entertaining, but it also connects to a larger theme. Normalization gives us a systematic way to pass from an additive inequality to a multiplicative inequality, and this is a trip we will often need to make in the pages that follow.

Item in the Dock: The Case of Equality

One of the enduring principles that emerges from an examination of the ways that inequalities are developed and applied is that many benefits flow from understanding when an inequality is sharp, or nearly sharp. In most cases, this understanding pivots on the discovery of the circumstances where equality can hold.

For Cauchy's inequality this principle suggests that we should ask ourselves about the relationship that must exist between the sequences $\{a_k\}$ and $\{b_k\}$ in order for us to have

$$\sum_{k=1}^{\infty} a_k b_k = \left(\sum_{k=1}^{\infty} a_k^2\right)^{\frac{1}{2}} \left(\sum_{k=1}^{\infty} b_k^2\right)^{\frac{1}{2}}. \tag{1.8}$$

If we focus our attention on the nontrivial case where neither of the sequences is identically zero and where both of the sums on the right-hand side of the identity (1.8) are finite, then we see that each of the steps we used in the derivation of the bound (1.7) can be reversed. Thus one finds that the identity (1.8) implies the identity

$$\sum_{k=1}^{\infty} \hat{a}_k \hat{b}_k = \frac{1}{2}\sum_{k=1}^{\infty} \hat{a}_k^2 + \frac{1}{2}\sum_{k=1}^{\infty} \hat{b}_k^2 = 1. \tag{1.9}$$

By the two-term bound $xy \le (x^2 + y^2)/2$, we also know that

$$\hat{a}_k \hat{b}_k \le \frac{1}{2}\hat{a}_k^2 + \frac{1}{2}\hat{b}_k^2 \qquad \text{for all } k = 1, 2, \ldots, \tag{1.10}$$

and from these we see that if strict inequality were to hold for even one value of k then we could not have the equality (1.9). This observation tells us in turn that the case of equality (1.8) can hold for nonzero series only when we have $\hat{a}_k = \hat{b}_k$ for all $k = 1, 2, \ldots$. By the definition of these normalized values, we then see that

$$a_k = \lambda b_k \qquad \text{for all } k = 1, 2, \ldots, \tag{1.11}$$

where the constant λ is given by the ratio

$$\lambda = \left(\sum_{j=1}^{\infty} a_j^2\right)^{\frac{1}{2}} \Big/ \left(\sum_{j=1}^{\infty} b_j^2\right)^{\frac{1}{2}}.$$

Here one should note that our argument was brutally straightforward, and thus, our problem was not much of a challenge. Nevertheless, the result still expresses a minor miracle; the *one* identity (1.8) has the strength to imply an *infinite* number of identities, one for each value of $k = 1, 2, \ldots$ in equation (1.11).

Benefits of Good Notation

Sums such as those appearing in Cauchy's inequality are just barely manageable typographically and, as one starts to add further features, they can become unwieldy. Thus, we often benefit from the introduction of shorthand notation such as

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{j=1}^{n} a_j b_j \tag{1.12}$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$. This shorthand now permits us to write Cauchy's inequality quite succinctly as

$$\langle \mathbf{a}, \mathbf{b} \rangle \le \langle \mathbf{a}, \mathbf{a} \rangle^{\frac{1}{2}} \langle \mathbf{b}, \mathbf{b} \rangle^{\frac{1}{2}}. \tag{1.13}$$

Parsimony is fine, but there are even deeper benefits to this notation if one provides it with a more abstract interpretation. Specifically, if V is a real vector space (such as $\mathbb{R}^d$), then we say that a function on $V \times V$ defined by the mapping $(\mathbf{a}, \mathbf{b}) \mapsto \langle \mathbf{a}, \mathbf{b} \rangle$ is an *inner product* and we say that $(V, \langle \cdot, \cdot \rangle)$ is a *real inner product space* provided that the pair $(V, \langle \cdot, \cdot \rangle)$ has the following five properties:

(i) $\langle \mathbf{v}, \mathbf{v} \rangle \ge 0$ for all $\mathbf{v} \in V$,
(ii) $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = 0$,
(iii) $\langle \alpha \mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\alpha \in \mathbb{R}$ and all $\mathbf{v}, \mathbf{w} \in V$,
(iv) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, and finally,
(v) $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ for all $\mathbf{v}, \mathbf{w} \in V$.

One can easily check that the shorthand introduced by the sum (1.12) has each of these properties, but there are many further examples of useful inner products. For example, if we fix a set of positive real numbers $\{w_j : j = 1, 2, \ldots, n\}$ then we can just as easily define an inner product on $\mathbb{R}^n$ with the weighted sums

$$\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{j=1}^{n} a_j b_j w_j \tag{1.14}$$

and, with this definition, one can check just as before that $\langle \mathbf{a}, \mathbf{b} \rangle$ satisfies all of the properties that one requires of an inner product. Moreover, this example only reveals the tip of an iceberg; there are many useful inner products, and they occur in a great variety of mathematical contexts.

An especially useful example of an inner product can be given by considering the set $V = C[a,b]$ of real-valued continuous functions on the bounded interval $[a,b]$ and by defining $\langle \cdot, \cdot \rangle$ on V by setting

$$\langle f, g \rangle = \int_a^b f(x)g(x)\,dx, \qquad (1.15)$$

or more generally, if $w : [a,b] \to \mathbb{R}$ is a continuous function such that $w(x) > 0$ for all $x \in [a,b]$, then one can define an inner product on $C[a,b]$ by setting

$$\langle f, g \rangle = \int_a^b f(x)g(x)w(x)\,dx.$$

We will return to these examples shortly, but first there is an opportunity that must be seized.

An Opportunistic Challenge

We now face one of those pleasing moments when good notation suggests a good theorem. We introduced the idea of an inner product in order to state the basic form (1.7) of Cauchy's inequality in a simple way, and now we find that our notation pulls us toward an interesting conjecture: Can it be true that in every inner product space one has the inequality $\langle \mathbf{v}, \mathbf{w} \rangle \le \langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} \langle \mathbf{w}, \mathbf{w} \rangle^{\frac{1}{2}}$? This conjecture is indeed true, and when framed more precisely, it provides our next challenge problem.

Problem 1.3 *For any real inner product space $(V, \langle \cdot, \cdot \rangle)$, one has for all $\mathbf{v}$ and $\mathbf{w}$ in V that*

$$\langle \mathbf{v}, \mathbf{w} \rangle \le \langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} \langle \mathbf{w}, \mathbf{w} \rangle^{\frac{1}{2}}; \qquad (1.16)$$

moreover, for nonzero vectors $\mathbf{v}$ and $\mathbf{w}$, one has

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} \langle \mathbf{w}, \mathbf{w} \rangle^{\frac{1}{2}} \qquad \text{if and only if } \mathbf{v} = \lambda \mathbf{w}$$

for a nonzero constant λ.

As before, one may be tempted to respond to this challenge by just rattling off a previously mastered textbook proof, but that temptation should still be resisted. The challenge offered by Problem 1.3 is important, and it deserves a fresh response — or, at least, a relatively fresh response.

For example, it seems appropriate to ask if one might be able to use some variation on the additive method which helped us prove the plain vanilla version of Cauchy's inequality. The argument began with the observation that $(x-y)^2 \geq 0$ implies $xy \leq x^2/2 + y^2/2$, and one might guess that an analogous idea could work again in the abstract case.

Here, of course, we need to use the defining properties of the inner product, and, as we go down the list looking for an analog to $(x-y)^2 \geq 0$, we are quite likely to hit on the idea of using property (i) in the form

$$\langle \mathbf{v}-\mathbf{w}, \mathbf{v}-\mathbf{w} \rangle \geq 0.$$

Now, when we expand this inequality with the help of the other properties of the inner product $\langle \cdot, \cdot \rangle$, we find that

$$\langle \mathbf{v}, \mathbf{w} \rangle \leq \frac{1}{2}\langle \mathbf{v}, \mathbf{v} \rangle + \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle. \tag{1.17}$$

This is a perfect analog of the additive inequality that gave us our second proof of the basic Cauchy inequality, and we face a classic situation where all that remains is a "matter of technique."

A Retraced Passage — Conversion of an Additive Bound

Here we are oddly lucky since we have developed only one technique that is even remotely relevant — the normalization method for converting an additive inequality into one that is multiplicative. Normalization means different things in different places, but, if we take our earlier analysis as our guide, what we want here is to replace $\mathbf{v}$ and $\mathbf{w}$ with related terms that reduce the right side of the bound (1.17) to 1.

Since the inequality (1.16) holds trivially if either $\mathbf{v}$ or $\mathbf{w}$ is equal to zero, we may assume without loss of generality that $\langle \mathbf{v}, \mathbf{v} \rangle$ and $\langle \mathbf{w}, \mathbf{w} \rangle$ are both nonzero, so the normalized variables

$$\hat{\mathbf{v}} = \mathbf{v}/\langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} \quad \text{and} \quad \hat{\mathbf{w}} = \mathbf{w}/\langle \mathbf{w}, \mathbf{w} \rangle^{\frac{1}{2}} \tag{1.18}$$

are well defined. When we substitute these values for $\mathbf{v}$ and $\mathbf{w}$ in the bound (1.17), we then find $\langle \hat{\mathbf{v}}, \hat{\mathbf{w}} \rangle \leq 1$. In terms of the original variables $\mathbf{v}$ and $\mathbf{w}$, this tells us $\langle \mathbf{v}, \mathbf{w} \rangle \leq \langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} \langle \mathbf{w}, \mathbf{w} \rangle^{\frac{1}{2}}$, just as we wanted to show.

Finally, to resolve the condition for equality, we only need to examine our reasoning in reverse. If equality holds in the abstract Cauchy inequality (1.16) for nonzero vectors $\mathbf{v}$ and $\mathbf{w}$, then the normalized variables $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$ are well defined. In terms of the normalized variables, the equality of $\langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} \langle \mathbf{w}, \mathbf{w} \rangle^{\frac{1}{2}}$ tells us that $\langle \hat{\mathbf{v}}, \hat{\mathbf{w}} \rangle = 1$, and this tells us in turn that $\langle \hat{\mathbf{v}} - \hat{\mathbf{w}}, \hat{\mathbf{v}} - \hat{\mathbf{w}} \rangle = 0$ simply by expansion of the inner product. From this we deduce that $\hat{\mathbf{v}} - \hat{\mathbf{w}} = 0$; or, in other words, $\mathbf{v} = \lambda \mathbf{w}$ where we set $\lambda = \langle \mathbf{v}, \mathbf{v} \rangle^{\frac{1}{2}} / \langle \mathbf{w}, \mathbf{w} \rangle^{\frac{1}{2}}$.

The Pace of Science — The Development of Extensions

Augustin-Louis Cauchy (1789–1857) published his famous inequality in 1821 in the second of two notes on the theory of inequalities that formed the final part of his book *Cours d'Analyse Algébrique*, a volume which was perhaps the world's first rigorous calculus text. Oddly enough, Cauchy did not use his inequality in his text, except in some illustrative exercises. The first time Cauchy's inequality was applied in earnest by anyone was in 1829, when Cauchy used his inequality in an investigation of Newton's method for the calculation of the roots of algebraic and transcendental equations. This eight-year gap provides an interesting gauge of the pace of science; now, each month, there are hundreds — perhaps thousands — of new scientific publications where Cauchy's inequality is applied in one way or another.

A great many of those applications depend on a natural analog of Cauchy's inequality where sums are replaced by integrals,

$$\int_a^b f(x)g(x)\,dx \le \left(\int_a^b f^2(x)\,dx\right)^{\frac{1}{2}} \left(\int_a^b g^2(x)\,dx\right)^{\frac{1}{2}}. \tag{1.19}$$

This bound first appeared in print in a *Mémoire* by Victor Yacovlevich Bunyakovsky which was published by the Imperial Academy of Sciences of St. Petersburg in 1859. Bunyakovsky (1804–1889) had studied in Paris with Cauchy, and he was quite familiar with Cauchy's work on inequalities; so much so that by the time he came to write his *Mémoire*, Bunyakovsky was content to refer to the classical form of Cauchy's inequality for finite sums simply as *well-known.* Moreover, Bunyakovsky did not dawdle over the limiting process; he took only a single line to pass from Cauchy's inequality for finite sums to his continuous analog (1.19). By ironic coincidence, one finds that this analog is labelled as inequality **(C)** in Bunyakovsky's *Mémoire*, almost as though Bunyakovsky had Cauchy in mind.

Bunyakovsky's *Mémoire* was written in French, but it does not seem to have circulated widely in Western Europe. In particular, it does not seem to have been known in Göttingen in 1885 when Hermann Amandus Schwarz (1843–1921) was engaged in his fundamental work on the theory of minimal surfaces.

In the course of this work, Schwarz had the need for a two-dimensional integral analog of Cauchy's inequality. In particular, he needed to show that if $S \subset \mathbb{R}^2$ and $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$, then the double integrals

$$A = \iint_S f^2 \, dxdy, \quad B = \iint_S fg \, dxdy, \quad C = \iint_S g^2 \, dxdy$$

must satisfy the inequality

$$|B| \le \sqrt{A} \cdot \sqrt{C}, \tag{1.20}$$

and Schwarz also needed to know that the inequality is strict unless the functions f and g are proportional.

An approach to this result via Cauchy's inequality would have been problematical for several reasons, including the fact that the strictness of a discrete inequality can be lost in the limiting passage to integrals. Thus, Schwarz had to look for an alternative path, and, faced with necessity, he discovered a proof whose charm has stood the test of time.

Schwarz based his proof on one striking observation. Specifically, he noted that the real polynomial

$$p(t) = \iint_S \Big(tf(x,y) + g(x,y) \Big)^2 \, dxdy = At^2 + 2Bt + C$$

is always nonnegative, and, moreover, $p(t)$ is strictly positive unless f and g are proportional. The binomial formula then tells us that the coefficients must satisfy $B^2 \le AC$, and unless f and g are proportional, one actually has the strict inequality $B^2 < AC$. Thus, from a single algebraic insight, Schwarz found everything that he needed to know.

Schwarz's proof requires the wisdom to consider the polynomial $p(t)$, but, granted that step, the proof is lightning quick. Moreover, as one finds from Exercise 1.11, Schwarz's argument can be used almost without change to prove the inner product form (1.16) of Cauchy's inequality, and even there Schwarz's argument provides one with a quick understanding of the case of equality. Thus, there is little reason to wonder why Schwarz's argument has become a textbook favorite, even though it does require one to pull a rabbit — or at least a polynomial — out of a hat.

The Naming of Things — Especially Inequalities

In light of the clear historical precedence of Bunyakovsky's work over that of Schwarz, the common practice of referring to the bound (1.19) as Schwarz's inequality may seem unjust. Nevertheless, by modern standards, both Bunyakovsky and Schwarz might count themselves lucky to have their names so closely associated with such a fundamental tool of mathematical analysis. Except in unusual circumstances, one garners little credit nowadays for crafting a continuous analog to a discrete inequality, or vice versa. In fact, many modern problem solvers favor a method of investigation where one rocks back and forth between discrete and continuous analogs in search of the easiest approach to the phenomena of interest.

Ultimately, one sees that inequalities get their names in a great variety of ways. Sometimes the name is purely descriptive, such as one finds with the triangle inequality which we will meet shortly. Perhaps more often, an inequality is associated with the name of a mathematician, but even then there is no hard-and-fast rule to govern that association. Sometimes the inequality is named after the first finder, but other principles may apply — such as the framer of the final form, or the provider of the best known application.

If one were to insist on the consistent use of the rule of first finder, then Hölder's inequality would become Rogers's inequality, Jensen's inequality would become Hölder's inequality, and only riotous confusion would result. The most practical rule — and the one used here — is simply to use the traditional names. Nevertheless, from time to time, it may be scientifically informative to examine the roots of those traditions.

Exercises

Exercise 1.1 (The 1-Trick and the Splitting Trick)

Show that for each real sequence $a_1, a_2, \ldots, a_n$ one has

$$a_1 + a_2 + \cdots + a_n \le \sqrt{n}(a_1^2 + a_2^2 + \cdots + a_n^2)^{\frac{1}{2}} \tag{a}$$

and show that one also has

$$\sum_{k=1}^{n} a_k \le \left(\sum_{k=1}^{n} |a_k|^{2/3} \right)^{\frac{1}{2}} \left(\sum_{k=1}^{n} |a_k|^{4/3} \right)^{\frac{1}{2}}. \tag{b}$$

The two tricks illustrated by this simple exercise will be our constant companions throughout the course. We will meet them in almost countless variations, and sometimes their implications are remarkably subtle.

Exercise 1.2 (Products of Averages and Averages of Products)

Suppose that $p_j \ge 0$ for all $j = 1, 2, \ldots, n$ and $p_1 + p_2 + \cdots + p_n = 1$. Show that if a_j and b_j are nonnegative real numbers that satisfy the termwise bound $1 \le a_j b_j$ for all $j = 1, 2, \ldots, n$, then one also has the aggregate bound for the averages,

$$1 \le \left\{ \sum_{j=1}^{n} p_j a_j \right\} \left\{ \sum_{j=1}^{n} p_j b_j \right\}. \tag{1.21}$$

This graceful bound is often applied with $b_j = 1/a_j$. It also has a subtle complement which is developed much later in Exercise 5.8.

Exercise 1.3 (Why Not Three or More?)

Cauchy's inequality provides an upper bound for a sum of pairwise products, and a natural sense of confidence is all one needs to guess that there are also upper bounds for the sums of products of three or more terms. In this exercise you are invited to justify two prototypical extensions. The first of these is definitely easy, and the second is not much harder, provided that you do not give it more respect than it deserves:

$$\left(\sum_{k=1}^{n} a_k b_k c_k \right)^4 \le \left(\sum_{k=1}^{n} a_k^2 \right)^2 \sum_{k=1}^{n} b_k^4 \sum_{k=1}^{n} c_k^4, \tag{a}$$

$$\left(\sum_{k=1}^{n} a_k b_k c_k \right)^2 \le \sum_{k=1}^{n} a_k^2 \sum_{k=1}^{n} b_k^2 \sum_{k=1}^{n} c_k^2. \tag{b}$$

Exercise 1.4 (Some Help From Symmetry)

There are many situations where Cauchy's inequality conspires with symmetry to provide results that are visually stunning. Here are two examples from a multitude of graceful possibilities.

(a) Show that for all positive x, y, z one has

$$S = \left(\frac{x+y}{x+y+z} \right)^{1/2} + \left(\frac{x+z}{x+y+z} \right)^{1/2} + \left(\frac{y+z}{x+y+z} \right)^{1/2} \le 6^{1/2}.$$

(b) Show that for all positive x, y, z one has

$$x + y + z \le 2 \left\{ \frac{x^2}{y+z} + \frac{y^2}{x+z} + \frac{z^2}{x+y} \right\}.$$

Exercise 1.5 (A Crystallographic Inequality with a Message)

Recall that $f(x) = \cos(\beta x)$ satisfies the identity $f^2(x) = \frac{1}{2}(1 + f(2x))$, and show that if $p_k \ge 0$ for $1 \le k \le n$ and $p_1 + p_2 + \cdots + p_n = 1$ then

$$g(x) = \sum_{k=1}^{n} p_k \cos(\beta_k x) \quad \text{satisfies} \quad g^2(x) \le \frac{1}{2} \big\{ 1 + g(2x) \big\}.$$

This is known as the Harker–Kasper inequality, and it has far-reaching consequences in crystallography. For the theory of inequalities, there is an additional message of importance; given any functional *identity* one should at least consider the possibility of an analogous *inequality* for a more extensive class of related functions, such as the class of mixtures used here.

Exercise 1.6 (A Sum of Inversion Preserving Summands)

Suppose that $p_k > 0$ for $1 \le k \le n$ and $p_1 + p_2 + \cdots + p_n = 1$. Show that one has the bound

$$\sum_{k=1}^{n} \left(p_k + \frac{1}{p_k} \right)^2 \ge n^3 + 2n + 1/n,$$

and determine necessary and sufficient conditions for equality to hold here. We will see later (Exercise 13.6, p. 206), that there are analogous results for powers other than 2.

Exercise 1.7 (Flexibility of Form)

Prove that for all real x, y, α and β one has

$$(5\alpha x + \alpha y + \beta x + 3\beta y)^2 \le (5\alpha^2 + 2\alpha\beta + 3\beta^2)(5x^2 + 2xy + 3y^2). \tag{1.22}$$

More precisely, show that the bound (1.22) is an immediate corollary of the Cauchy–Schwarz inequality (1.16) provided that one designs a special inner product $\langle \cdot, \cdot \rangle$ for the job.

Exercise 1.8 (Doing the Sums)

The effective use of Cauchy's inequality often depends on knowing a convenient estimate for one of the bounding sums. Verify the four following classic bounds for real sequences:

$$\sum_{k=0}^{\infty} a_k x^k \le \frac{1}{\sqrt{1 - x^2}} \left(\sum_{k=0}^{\infty} a_k^2 \right)^{\frac{1}{2}} \qquad \text{for } 0 \le x < 1, \tag{a}$$

$$\sum_{k=1}^{n} \frac{a_k}{k} < \sqrt{2} \left(\sum_{k=1}^{n} a_k^2 \right)^{\frac{1}{2}}, \tag{b}$$

$$\sum_{k=1}^{n} \frac{a_k}{\sqrt{n + k}} < (\log 2)^{\frac{1}{2}} \left(\sum_{k=1}^{n} a_k^2 \right)^{\frac{1}{2}}, \quad \text{and} \tag{c}$$

$$\sum_{k=0}^{n} \binom{n}{k} a_k \leq \binom{2n}{n}^{\frac{1}{2}} \left(\sum_{k=0}^{n} a_k^2 \right)^{\frac{1}{2}}. \tag{d}$$

Exercise 1.9 (Beating the Obvious Bounds)

Many problems of mathematical analysis depend on the discovery of bounds which are stronger than those one finds with the direct application of Cauchy's inequality. To illustrate the kind of opportunity one might miss, show that for any real numbers a_j, $j = 1, 2 \ldots, n$, one has the bound

$$\left| \sum_{j=1}^{n} a_j \right|^2 + \left| \sum_{j=1}^{n} (-1)^j a_j \right|^2 \leq (n+2) \sum_{j=1}^{n} a_j^2.$$

Here the direct application of Cauchy's inequality gives a bound with $2n$ instead of the value $n+2$, so for large n one does better by a factor of nearly two.

Exercise 1.10 (Schur's Lemma — The R and C Bound)

Show that for each rectangular array $\{c_{jk} : 1 \leq j \leq m, 1 \leq k \leq n\}$, and each pair of sequences $\{x_j : 1 \leq j \leq m\}$ and $\{y_k : 1 \leq k \leq n\}$, we have the bound

$$\left| \sum_{j=1}^{m} \sum_{k=1}^{n} c_{jk} x_j y_k \right| \leq \sqrt{RC} \left(\sum_{j=1}^{m} |x_j|^2 \right)^{1/2} \left(\sum_{k=1}^{n} |y_k|^2 \right)^{1/2} \tag{1.23}$$

where R and C are the *row sum* and *column sum* maxima defined by

$$R = \max_j \sum_{k=1}^{n} |c_{jk}| \quad \text{and} \quad C = \max_k \sum_{j=1}^{m} |c_{jk}|.$$

This bound is known as *Schur's Lemma*, but, ironically, it may be the second most famous result with that name. Nevertheless, this inequality is surely the single most commonly used tool for bounding a quadratic form. One should note in the extreme case when $n = m$, $c_{jk} = 0$ $j \neq k$, and $c_{jj} = 1$ for $1 \leq j \leq n$, Schur's Lemma simply recovers Cauchy's inequality.

Exercise 1.11 (Schwarz's Argument in an Inner Product Space)

Let v and w be elements of the inner product space $(V, \langle \cdot, \cdot \rangle)$ and consider the quadratic polynomial defined for $t \in \mathbb{R}$ by

$$p(t) = \langle \mathbf{v} + t\mathbf{w}, \mathbf{v} + t\mathbf{w} \rangle.$$

Observe that this polynomial is nonnegative and use what you know about the solution of the quadratic equation to prove the inner product version (1.16) of Cauchy's inequality. Also, examine the steps of your proof to establish the conditions under which the case of equality can apply. Thus, confirm that Schwarz's argument (page 11) applies almost without change to prove Cauchy's inequality for a general inner product.

Exercise 1.12 (Example of a Self-generalization)

Let $\langle \cdot, \cdot \rangle$ denote an inner product on the vector space V and suppose that $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ and $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ are sequences of elements of V. Prove that one has the following vector analog of Cauchy's inequality:

$$\sum_{j=1}^{n} \langle \mathbf{x}_j, \mathbf{y}_j \rangle \leq \left(\sum_{j=1}^{n} \langle \mathbf{x}_j, \mathbf{x}_j \rangle \right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} \langle \mathbf{y}_j, \mathbf{y}_j \rangle \right)^{\frac{1}{2}}. \tag{1.24}$$

Note that if one takes $n = 1$, then this bound simply recaptures the Cauchy–Schwarz inequality for an inner product space, while, if one keeps n general but specializes the vector space V to be $\mathbb{R}$ with the trivial inner product $\langle \mathbf{x}, \mathbf{y} \rangle = xy$, then the bound (1.24) simply recaptures the plain vanilla Cauchy inequality.

Exercise 1.13 (Application of Cauchy's Inequality to an Array)

Show that if $\{a_{jk} : 1 \leq j \leq m,\ 1 \leq k \leq n\}$ is an array of real numbers then one has

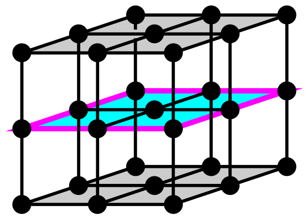
$$m \sum_{j=1}^{m} \left(\sum_{k=1}^{n} a_{jk} \right)^2 + n \sum_{k=1}^{n} \left(\sum_{j=1}^{m} a_{jk} \right)^2 \leq \left(\sum_{j=1}^{m} \sum_{k=1}^{n} a_{jk} \right)^2 + mn \sum_{j=1}^{m} \sum_{k=1}^{n} a_{jk}^2.$$

Moreover, show that equality holds here if and only if there exist α_j and β_k such that $a_{jk} = \alpha_j + \beta_k$ for all $1 \leq j \leq m$ and $1 \leq k \leq n$.

Exercise 1.14 (A Cauchy Triple and Loomis–Whitney)

Here is a generalization of Cauchy's inequality that has as a corollary a discrete version of the Loomis–Whitney inequality, a result which in the continuous case provides a bound on the volume of a set in terms of the volumes of the projections of that set onto lower dimensional subspaces. The discrete Loomis–Whitney inequality (1.26) was only recently developed, and it has applications in information theory and the theory of algorithms.

(a) Show that for any nonnegative a_{ij}, b_{jk}, c_{ki} with $1 \leq i, j, k \leq n$ one



Here we have a set A with cardinality $|A| = 27$ with projections that satisfy $|A_x| = |A_y| = |A_z| = 9$.

Fig. 1.1. The discrete Loomis–Whitney inequality says that for any collection A of points in $\mathbb{R}^3$ one has $|A| \le |A_x|^{\frac{1}{2}}|A_y|^{\frac{1}{2}}|A_z|^{\frac{1}{2}}$. The cubic arrangement indicated here suggests the canonical situation where one finds the case of equality in the bound.

has the triple product inequality

$$\sum_{i,j,k=1}^{n} a_{ij}^{\frac{1}{2}}\, b_{jk}^{\frac{1}{2}}\, c_{ki}^{\frac{1}{2}} \le \left\{ \sum_{i,j=1}^{n} a_{ij} \right\}^{\frac{1}{2}} \left\{ \sum_{j,k=1}^{n} b_{jk} \right\}^{\frac{1}{2}} \left\{ \sum_{k,i=1}^{n} c_{ki} \right\}^{\frac{1}{2}}. \tag{1.25}$$

(b) Let A denote a finite set of points in $\mathbb{Z}^3$ and let A_x, A_y, A_z denote the projections of A onto the corresponding coordinate planes that are orthogonal to the x, y, or z-axes. Let $|B|$ denote the cardinality of a set $B \subset \mathbb{Z}^3$ and show that the projections provide an upper bound on the cardinality of A:

$$|A| \le |A_x|^{\frac{1}{2}}|A_y|^{\frac{1}{2}}|A_z|^{\frac{1}{2}}. \tag{1.26}$$

Exercise 1.15 (An Application to Statistical Theory)

If $p(k;\theta) \ge 0$ for all $k \in D$ and $\theta \in \Theta$ and if

$$\sum_{k \in D} p(k;\theta) = 1 \qquad \text{for all } \theta \in \Theta, \tag{1.27}$$

then for each $\theta \in \Theta$ one can think of $\mathcal{M}_\theta = \{p(k;\theta) : k \in D\}$ as specifying a *probability model* where $p(k;\theta)$ represents the probability that we "observe k" when the parameter θ is the true "state of nature." If the function $g : D \to \mathbb{R}$ satisfies

$$\sum_{k \in D} g(k)p(k;\theta) = \theta \qquad \text{for all } \theta \in \Theta, \tag{1.28}$$

then g is called an *unbiased estimator* of the parameter θ. Assuming that D is finite and $p(k;\theta)$ is a differentiable function of θ, show that one has the lower bound

$$\sum_{k\in D}(g(k)-\theta)^2 p(k;\theta) \geq 1/I(\theta) \tag{1.29}$$

where $I:\Theta\to\mathbb{R}$ is defined by the sum

$$I(\theta) = \sum_{k\in D}\left\{p_\theta(k;\theta)/p(k;\theta)\right\}^2 p(k;\theta), \tag{1.30}$$

where $p_\theta(k;\theta) = \partial p(k;\theta)/\partial\theta$. The quantity defined by the left side of the bound (1.29) is called the *variance* of the unbiased estimator g, and the quantity $I(\theta)$ is known as the *Fisher information* at θ of the model $\mathcal{M}_\theta$. The inequality (1.29) is known as the Cramér–Rao lower bound, and it has extensive applications in mathematical statistics.