

# A Primer on Inequalities

## Introduction: Basic Inequalities

When working in the set of real numbers, we have a law of trichotomy. Given  $x, y \in \mathbb{R}$  exactly one of these is true:  $x < y$ ,  $y < x$ , or  $x = y$ . This (almost) defining characteristic of the real line means that inequalities are a valuable tool in calculus and calculus' big brother, real analysis. At our level it's hard to give authentic examples of why inequalities are important, but here's a couple of examples.

- Perhaps we want to know whether or not the infinite sum  $\sum a_k$  is finite, but  $a_k$  is very hard to understand. If we could find a simple series  $\sum b_k$  so that either  $\sum a_k \leq \sum b_k$  or conversely, our problem could be circumvented.
- The definite integral on  $[0, 1]$  can be viewed as a function from the set of continuous functions  $C[0, 1]$  to the real numbers:

$$f \mapsto \int_0^1 f(x) dx.$$

Is this operation a continuous one? That is, if two functions  $f, g$  are 'close together' are their integrals 'close' as well? In functional analysis we find that for any linear operation  $T$  on a function space,  $T$  is continuous if we can find an inequality of the form

$$\|Tf\| \leq C\|f\|$$

that holds for all  $f$  in the domain. Continuous operations are extremely important in the study of PDE, quantum mechanics, dynamical systems, and even image reconstruction.

Clearly these applications are beyond the scope of this course. How goal is to work with inequalities in an authentic way; that is, we don't want to solve inequalities (Find all reals  $x$  so that  $3x + 2 > 5$ ), but rather prove inequality statements that are generally true. To illuminate this point, consider the simplest inequality: for all real  $x$ ,

$$x^2 \geq 0.$$

This statement is always true, and we can use it in clever ways to prove more general facts.

*Example.* For all  $a, b, c \in \mathbb{R}$  prove that

$$a^2 + b^2 + c^2 \geq ab + bc + ca.$$

Let  $a, b, c \in \mathbb{R}$ . We begin with a simple observation, albeit one that might seem unmotivated:

$$(a - b)^2 + (b - c)^2 + (c - a)^2 \geq 0.$$

Expanding the binomials gives the result. □

Note well that we always start our proofs with a statement we know to be true! We cannot begin our proof of the previous result by writing "Since

$$a^2 + b^2 + c^2 \geq ab + bc + ca$$

we can rearrange..." Let's see one more example before moving to more interesting problems.

*Example.* Given positive numbers  $a, b$  prove that

$$\frac{a + b}{2} \geq \sqrt{ab}.$$

In this problem we've restricted ourselves to positive numbers. The advantage of this is that  $\sqrt{a}$  and  $\sqrt{b}$  are real numbers as well (you might wonder why a complex number such as  $\sqrt{-1}$  would bother us—it turns out that complex numbers cannot be ordered in an algebraically-meaningful way, so we must avoid them for now).

Now for the proof. Let  $a, b \in \mathbb{R}$  be positive and note that

$$(\sqrt{a} - \sqrt{b})^2 \geq 0.$$

Expanding the binomial gives the result. □

This inequality we've just proved is far more important than one might think. We'll continue this line of thought in the next section.

## Arithmetic and Geometric Means

Given positive numbers  $a_1, a_2, \dots, a_n$  we can form the arithmetic mean

$$A = \frac{a_1 + a_2 + \dots + a_n}{n}$$

or the geometric mean

$$G = \sqrt[n]{a_1 a_2 \dots a_n}.$$

Perhaps a surprising fact is that  $A \geq G$  no matter what. We'll examine two major proofs of this fact. The first is elementary—meaning that it doesn't use high-powered mathematical tools, not that it's easy. The second we'll see in a later section below, once we have those tools. We call this theorem the arithmetic mean–geometric mean inequality, or AM–GM for short.

**Theorem (AM–GM).** *Given positive real numbers  $a_1, a_2, \dots, a_n$  we have that*

$$\frac{a_1 + a_2 + \dots + a_n}{n} \geq \sqrt[n]{a_1 a_2 \dots a_n}.$$

*Proof.* The case  $n = 1$  is trivial, so assume  $n \geq 2$ . The following proof is due to Cauchy, and it essentially is an induction argument on  $n$ . That said, it's very unusual in its structure. The base case  $n = 2$  was discussed in an example in the previous section. Next we prove that whenever the result holds for  $n$ , it holds for  $2n$  as well. That is, we'll first prove the result for powers of 2:  $n = 2, 4, 8, 16, \dots$ . Assume we know the result holds for some  $n$ . Now consider  $2n$  positive real numbers  $a_1, \dots, a_n$  and  $b_1, \dots, b_n$ . We use the induction hypothesis and the base case to find

$$\begin{aligned} (a_1 + \dots + a_n) + (b_1 + \dots + b_n) &\geq n(a_1 \dots a_n)^{1/n} + n(b_1 \dots b_n)^{1/n} \\ &\geq 2n \sqrt{(a_1 \dots a_n)^{1/n} (b_1 \dots b_n)^{1/n}} \\ &= 2n \sqrt[2n]{a_1 \dots a_n b_1 \dots b_n}, \end{aligned}$$

which is what we wanted. We've finished our first induction, and we know the theorem to be true for infinitely many  $n$ . Next we prove that whenever the result is true for  $n$ , it's also true for  $n - 1$ . That is, we'll use a backwards induction; this will catch all those values of  $n$  between powers of 2, finishing the proof.

Let  $n \geq 4$  and assume the result holds for  $n$ . Consider the  $n - 1$  positive numbers  $a_1, a_2, \dots, a_{n-1}$ . Define  $a_n$  to be the geometric mean of the other  $n - 1$  numbers. Then we have

$$a_1 + a_2 + \dots + a_{n-1} + a_n \geq n \sqrt[n]{a_1 a_2 \dots a_{n-1} a_n} = n a_n.$$

To see why the last equality holds, consider this anecdote: A student has a 70% going into a final exam. He scores a 70% on the final, so his final average is 70%. That is, adding the geometric mean to the list did not change the geometric mean (manipulating exponents proves this fact). Finally, algebra gives us

$$a_1 + a_2 + \dots + a_{n-1} \geq n a_n - a_n,$$

so the result follows upon division by  $n - 1$ . Whew! □

It's time to get our hands dirty and prove some inequalities using our new tool.

*Example.* Let  $\theta \in (0, \pi/2)$ . Then

$$\tan \theta + \cot \theta = \tan \theta + \frac{1}{\tan \theta} \geq 2 \sqrt{\tan \theta \cdot \frac{1}{\tan \theta}} = 2.$$

If  $\theta = \pi/4$  equality occurs; we've minimized a function without calculus, hooray. □

*Example.* Let  $a, b, c \in \mathbb{R}$  be positive. Prove that

$$\frac{ab}{c} + \frac{bc}{a} + \frac{ca}{b} \geq a + b + c.$$

We need to use the AM–GM cleverly here. Note that

$$\frac{ab}{c} + \frac{bc}{a} \geq 2 \sqrt{\frac{ab}{c} \cdot \frac{bc}{a}} = 2b.$$

Similarly  $bc/a + ca/b \geq 2c$  and  $ca/b + ab/c \geq 2a$ . Adding these 3 inequalities gives the result. □

*Example.* Prove

$$\binom{n}{1} \binom{n}{2} \cdots \binom{n}{n-1} \leq \left( \frac{2^n - 2}{n-1} \right)^{n-1}$$

when  $n \geq 2$ . This follows from the AM–GM once we remember that

$$\sum_{k=0}^n \binom{n}{k} = 2^n$$

and that  $\binom{n}{0} = \binom{n}{n} = 1$ . □

*Example.* Given a positive integer  $n$ , let  $\tau(n)$  denote the number of positive integers dividing  $n$  and let  $\sigma(n)$  denote the sum of these positive divisors. We claim that for any  $n \geq 1$

$$\frac{\sigma(n)}{\tau(n)} \geq \sqrt{n}.$$

Denote the divisors of  $n$  as  $1 = d_1 < d_2 < \dots < d_k = n$ . Note that the arithmetic mean of these numbers is exactly  $\sigma(n)/\tau(n)$ . By the AM–GM we have

$$\frac{\sigma(n)}{\tau(n)} \geq \sqrt[k]{d_1 d_2 \cdots d_k}.$$

Now for a trick. Notice that  $d_j \cdot d_{k-j} = n$  for each  $j$ ; for example the divisors of 12 are 1, 2, 3, 4, 6, 12 and sure enough  $1 \cdot 12 = 2 \cdot 6 = 3 \cdot 4 = 12$ . Thus we can write

$$\left( \frac{\sigma(n)}{\tau(n)} \right)^2 \geq \sqrt[k]{d_1 d_2 \cdots d_k} \cdot \sqrt[k]{d_k d_{k-1} \cdots d_1} = \sqrt[k]{n^k} = n,$$

which proves the result. □

## The Cauchy–Schwarz Inequality

It’s time to meet the most useful inequality in mathematics. Commonly called the Cauchy–Schwarz inequality, this statement was actually proven in different forms by Augustin-Louis Cauchy and Hermann Schwarz. In fact, the mathematician Viktor Bunyakovsky (a student of Cauchy) proved a third version; to this day many Russian mathematicians refer to the inequality as that of Bunyakovsky. We’ll use the earliest version, due to Cauchy.

**Theorem** (Cauchy–Schwarz Inequality, Cauchy’s version, 1821). *Let  $a = (a_1, a_2, \dots, a_n)$  and  $b = (b_1, b_2, \dots, b_n)$  be vectors in  $\mathbb{R}^n$ . It follows that*

$$(a_1 b_1 + a_2 b_2 + \cdots + a_n b_n)^2 \leq (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2). \quad (\text{CS})$$

*In vector notation (which is often less helpful) we have  $a \cdot b \leq \|a\| \cdot \|b\|$ .*

For completeness and cultural awareness, we present the other 2 forms of the inequality, which we won’t use.

**Theorem** (Cauchy–Schwarz, Bunyakovsky’s version, 1859). *Let  $f, g : X \rightarrow \mathbb{C}$  be measurable functions on a measure space  $X$ . (In lay terms,  $f$  and  $g$  are functions for which integrals can be defined, and  $X$  is a space we can integrate over, such as an interval). Then we have*

$$\int_X |fg| \leq \left( \int_X |f|^2 \right)^{1/2} \left( \int_X |g|^2 \right)^{1/2}.$$

*Note that these are definite integrals, written in the style of measure theory.*

**Theorem** (Cauchy–Schwarz, Schwarz’ version, 1888). *Let  $V$  be an inner product space (that is, a vector space with a complex-valued operation like the dot product, where we denote  $\mathbf{u} \cdot \mathbf{v}$  instead by  $\langle u, v \rangle$ ). Then for any vectors  $u, v \in V$  we have*

$$|\langle u, v \rangle| \leq \|u\| \cdot \|v\|.$$

Without further ado, let’s prove Cauchy’s inequality. The idea is simple—when a real quadratic function is always nonnegative, its discriminant is also.

*Proof.* Let  $a, b \in \mathbb{R}^n$  be arbitrary. Given any real number  $t \in \mathbb{R}$  note that

$$\|a + tb\|^2 \geq 0.$$

Recall that the dot product satisfies  $\|u\|^2 = u \cdot u$  for any vector  $u$ . Thus

$$(a + tb) \cdot (a + tb) \geq 0,$$

which expands to give

$$(b \cdot b)t^2 + 2(a \cdot b)t + (a \cdot a) \geq 0.$$

This quadratic polynomial in  $t$  is always nonnegative, so the discriminant must be nonnegative:

$$(2a \cdot b)^2 - 4(b \cdot b)(a \cdot a) \geq 0,$$

which rearranges into the result. □

Cauchy's inequality is powerful. Let's solve a number of problems with it.

*Example.* Let  $a_1, a_2, \dots, a_n$  be positive reals. Prove that

$$(a_1 + a_2 + \dots + a_n)(a_1^{-1} + a_2^{-1} + \dots + a_n^{-1}) \geq n^2.$$

Since each  $a_k$  is positive, the square root  $\sqrt{a_k}$  is real. Using Cauchy's inequality on the vectors  $(\sqrt{a_k})$  and  $(1/\sqrt{a_k})$  gives

$$(\sqrt{a_1}^2 + \dots + \sqrt{a_n}^2)(1/\sqrt{a_1}^2 + \dots + 1/\sqrt{a_n}^2) \geq \left( \sqrt{a_1} \cdot \frac{1}{\sqrt{a_1}} + \dots + \sqrt{a_n} \cdot \frac{1}{\sqrt{a_n}} \right)^2,$$

which gives the result. □

*Example.* Let  $(a_k)$  be an infinite sequence of real numbers. Prove

$$\sum_{k=1}^{\infty} a_k x^k \leq \frac{1}{\sqrt{1-x^2}} \left( \sum_{k=1}^{\infty} a_k^2 \right)^{1/2}$$

whenever  $0 \leq x < 1$ . This follows immediately from Cauchy's inequality if we recall that

$$1 + x^2 + x^4 + x^6 + \dots = \frac{1}{1-x^2}$$

for all  $x \in (-1, 1)$ . □

*Example.* Given positive real numbers  $a, b, c$  which sum to 6, prove that

$$\left(a + \frac{1}{b}\right)^2 + \left(b + \frac{1}{c}\right)^2 + \left(c + \frac{1}{a}\right)^2 \geq \frac{75}{4}.$$

Using Cauchy with the vectors  $(1, 1, 1)$  and  $(a + 1/b, b + 1/c, c + 1/a)$  gives

$$(1^2 + 1^2 + 1^2) \left[ \left(a + \frac{1}{b}\right)^2 + \left(b + \frac{1}{c}\right)^2 + \left(c + \frac{1}{a}\right)^2 \right] \geq \left(a + b + c + \frac{1}{a} + \frac{1}{b} + \frac{1}{c}\right)^2.$$

Notice that  $a + b + c = 6$  and  $1/a + 1/b + 1/c \geq 3/2$  by a previous example. Thus

$$3 \left[ \left(a + \frac{1}{b}\right)^2 + \left(b + \frac{1}{c}\right)^2 + \left(c + \frac{1}{a}\right)^2 \right] \geq \left(6 + \frac{3}{2}\right)^2,$$

which is equivalent to the given inequality. □

*Example.* Given a regular  $n$ -gon with side length 1 pick an arbitrary point  $P$  in the interior. Draw line segments perpendicular to each of the  $n$  sides and let  $x_1, x_2, \dots, x_n$  be their lengths. Prove that

$$\frac{1}{x_1} + \dots + \frac{1}{x_n} > 2\pi.$$

Note that we need to prove strict inequality (equality not allowed). Our proof should indicate why this occurs. If we draw line segments from  $P$  to each vertex of the polygon, we obtain a subdivision into  $n$  triangles; the  $k$ -th triangle has height  $x_k$  and base 1. Thus the area of the  $k$ -th triangle is  $x_k/2$ , causing the total area of the polygon to be  $(x_1 + \dots + x_n)/2$ . On the other hand, a regular polygon with  $n$  sides and side length 1 has area  $(n/4) \cot(n/\pi)$ . Combining these facts gives us

$$x_1 + x_2 + \dots + x_n = \frac{n}{2} \cot\left(\frac{\pi}{n}\right). \quad (\star)$$

Finally, recall an earlier example wherein we proved

$$(x_1 + \dots + x_n) \left( \frac{1}{x_1} + \dots + \frac{1}{x_n} \right) \geq n^2.$$

Using equation  $(\star)$  in this inequality gives us

$$\frac{1}{x_1} + \dots + \frac{1}{x_n} \geq 2n \tan\left(\frac{\pi}{n}\right) > 2n \left(\frac{\pi}{n}\right) = 2\pi,$$

where we have used the fact that  $\tan x > x$  for  $0 < x < \pi/2$ . □

## Convexity and Jensen's Inequality

So far every inequality we've studied has followed from the nonnegativity of a square. Now we turn to a different 'source' of inequality, convexity.

**Definition.** A function  $f : (a, b) \rightarrow \mathbb{R}$  is called convex iff whenever  $x, y \in (a, b)$  and  $0 < t < 1$  we have

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

For  $t = 1/2$ , this reduces to the memorable form

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x) + f(y)}{2}.$$

When  $-f$  is convex, we call  $f$  concave.

Graphically, the graph of a convex function lies underneath its secant line (see figure 1). Before we proceed, we make a few remarks about convex functions.

- Convex functions are always continuous. A proof of this is beyond the scope of this course, but it's nice to know that convex functions are reasonably well-behaved.
- Convex functions need not be differentiable. Just consider  $f(x) = |x|$  at  $x = 0$ .
- If a function  $f : (a, b) \rightarrow \mathbb{R}$  is twice differentiable *everywhere* in its domain, then  $f$  is convex if and only if  $f''(x) \geq 0$  throughout  $(a, b)$ . This is convexity as most people remember it from calculus. We won't prove this either; nonetheless, feel free to use this fact to prove functions are convex!
- The statement

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x) + f(y)}{2}$$

is *not* enough to guarantee that  $f$  is convex, unless we know a priori that  $f$  is continuous. Again, this is beyond our scope.

Convexity is a very powerful tool. The following inequality, due to Johann Jensen, shows that convexity applies to averages of more than 2 terms. First we state a simple version, which is typically all we need.

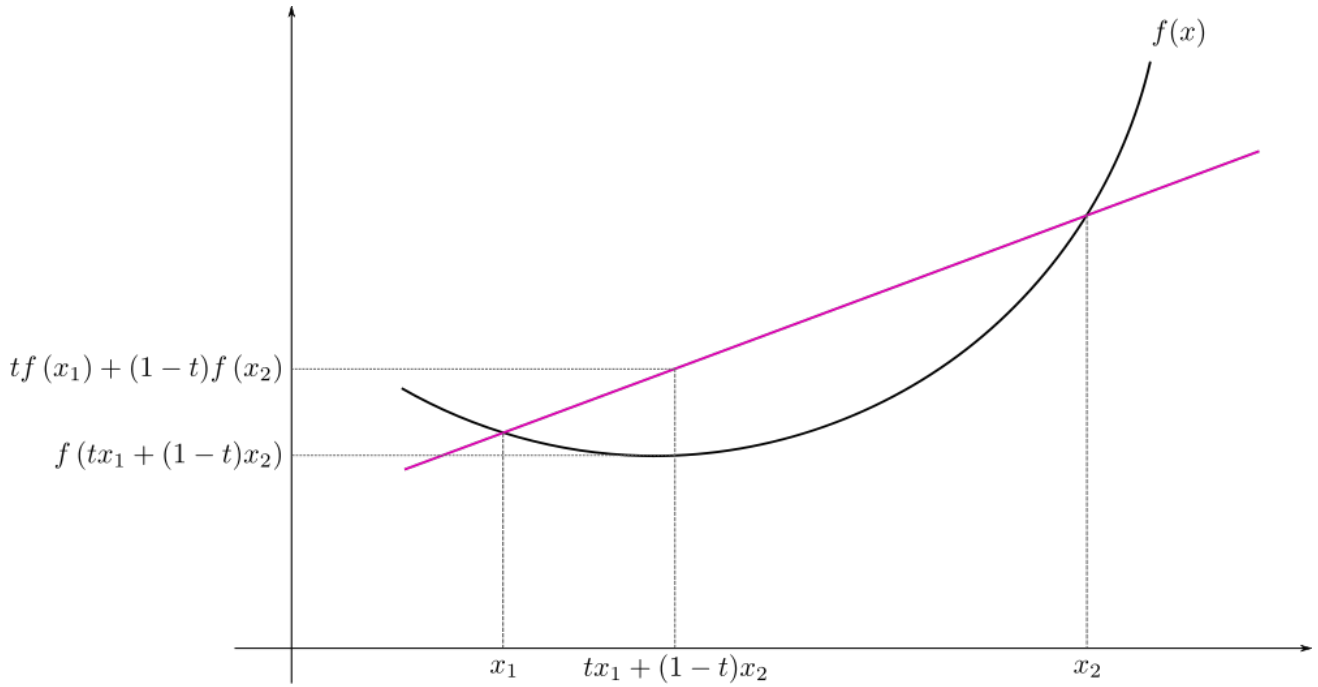


Figure 1: The graph of a convex function lies below the line segment joining its endpoints.

**Theorem** (Jensen's Inequality, simple version, 1906). *Let  $f : (a, b) \rightarrow \mathbb{R}$  be convex. Given  $x_1, x_2, \dots, x_n \in (a, b)$  we have that*

$$f\left(\frac{x_1 + x_2 + \dots + x_n}{n}\right) \leq \frac{f(x_1) + f(x_2) + \dots + f(x_n)}{n}.$$

*If  $f$  is concave, the inequality is reversed.*

**Theorem** (Jensen's Inequality, general version, 1906). *Let  $f : (a, b) \rightarrow \mathbb{R}$  be convex. Given  $x_1, x_2, \dots, x_n \in (a, b)$  and positive numbers  $t_1, \dots, t_n$  which sum to 1, we have that*

$$f(t_1x_1 + t_2x_2 + \dots + t_nx_n) \leq t_1f(x_1) + t_2f(x_2) + \dots + t_nf(x_n).$$

*The simple version follows when  $t_1 = t_2 = \dots = t_n = 1/n$ . Again, a concave function simply reverses the inequality.*

Of course, there is a more general setting in which Jensen's inequality is used. For cultural awareness (and nothing more), we'll state it here.

**Theorem** (Jensen's inequality for integration on probability spaces). *Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be convex and  $X$  be a measure space with total measure (e.g., length, area, etc.) equal to 1. Given an integrable function  $g : X \rightarrow \mathbb{R}$ , we have that*

$$f\left(\int_X g\right) \leq \int_X f \circ g.$$

We'll prove the general Jensen inequality for sums, which implies the simple one.

*Proof.* We induct on  $n$ , the number of terms in the sum. When  $n = 2$  Jensen's inequality is simply the definition of convexity. Now assume that whenever  $y_1, \dots, y_n \in (a, b)$  and  $\lambda_1, \dots, \lambda_n \in (0, 1)$  are such that  $\sum \lambda_k = 1$ , we have

$$f(\lambda_1y_1 + \dots + \lambda_ny_n) \leq \lambda_1f(y_1) + \dots + \lambda_nf(y_n).$$

That is, assume the result for some integer  $n \geq 2$ . Given  $x_1, \dots, x_n, x_{n+1} \in (a, b)$  and positive reals  $t_k \in (0, 1)$  which sum to 1, we use clever algebra to write

$$\begin{aligned} f\left(\sum_{k=1}^{n+1} t_k x_k\right) &= f\left(t_{n+1}x_{n+1} + (1-t_{n+1})\sum_{k=1}^n \frac{t_k}{1-t_{n+1}}x_k\right) \\ &\leq t_{n+1}f(x_{n+1}) + (1-t_{n+1})f\left(\sum_{k=1}^n \frac{t_k}{1-t_{n+1}}x_k\right). \end{aligned}$$

Notice that  $\sum_{k=1}^n t_k/(1-t_{n+1}) = 1$ , so the induction hypothesis gives us

$$f\left(\sum_{k=1}^{n+1} t_k x_k\right) \leq t_{n+1} f(x_{n+1}) + (1-t_{n+1}) \sum_{k=1}^n \frac{t_k}{1-t_{n+1}} f(x_k) = \sum_{k=1}^{n+1} t_k f(x_k),$$

and the result holds for all  $n$  by induction.  $\square$

Convexity is a great tool for proving inequalities about functions you understand well, such as the tangent, logarithm and squaring functions. Let's look at some examples.

*Example.* Given a triangle  $ABC$  prove that

$$\sin A + \sin B + \sin C \leq \frac{3\sqrt{3}}{2}.$$

Since  $ABC$  is a triangle,  $A+B+C = \pi$  and each angle lies within  $(0, \pi)$ . Furthermore,  $\sin$  is a concave function on  $(0, \pi)$ . By Jensen's inequality we have

$$\frac{\sin A + \sin B + \sin C}{3} \leq \sin\left(\frac{A+B+C}{3}\right) = \sin(\pi/3) = \frac{\sqrt{3}}{2},$$

which rearranges into the result.  $\square$

*Example.* Suppose the positive numbers  $a, b, c$  satisfy  $a+b+c = abc$ . Prove that

$$\frac{1}{\sqrt{1+a^2}} + \frac{1}{\sqrt{1+b^2}} + \frac{1}{\sqrt{1+c^2}} \leq \frac{3}{2}. \quad (\star\star)$$

Inspired by trigonometry, we might try writing  $a = \tan A$ ,  $b = \tan B$ , and  $c = \tan C$  with  $A, B, C \in (0, \pi/2)$ . What relationship holds between the angles if we make this substitution? Exercising my omnipotence, expanding  $\tan(A+B+C)$  gives us

$$\tan(A+B+C) = \frac{\tan A + \tan B + \tan C - \tan A \tan B \tan C}{1 - \tan A \tan B - \tan B \tan C - \tan C \tan A} = 0,$$

which implies that  $A+B+C = \pi$ . That is,  $ABC$  is a triangle. I'm being a bit disingenuous here; I'm not clever, just showing a well-known fact about triangles. Anyway, the inequality  $(\star\star)$  that we want to prove can be written simply as

$$\cos A + \cos B + \cos C \leq \frac{3}{2}.$$

On the interval  $(0, \pi/2)$  cosine is concave, so Jensen's inequality gives the result:

$$\cos A + \cos B + \cos C \leq 3 \cos\left(\frac{A+B+C}{3}\right) = 3 \cos\left(\frac{\pi}{3}\right) = \frac{3}{2},$$

as desired.  $\square$

*Example.* Let's give a more conceptual proof of the AM-GM using convexity. Given positive reals  $a_1, \dots, a_n$  let's prove once again that

$$\frac{a_1 + \dots + a_n}{n} \geq \sqrt[n]{a_1 \dots a_n}.$$

This time we'll use positivity in a different way; rewrite the  $n$  numbers as  $a_k = e^{x_k}$  with each  $x_k \in \mathbb{R}$ . The exponential function  $x \mapsto e^x$  is convex, so Jensen's inequality gives

$$\frac{1}{n} \sum_{k=1}^n e^{x_k} \geq \exp\left(\frac{x_1 + \dots + x_n}{n}\right) = [e^{x_1} e^{x_2} \dots e^{x_n}]^{1/n}.$$

Rewriting this in terms of the numbers  $a_k$  gives the result. Much easier than forward-backward induction!  $\square$

## More Examples and Applications

In this section we attack various problems and discuss uses of inequalities in science. Try to decide a fruitful method for each before reading the solution.

*Example.* Let  $x_1, \dots, x_n$  be positive reals summing to 1. Prove that

$$\left(1 + \frac{1}{x_1}\right) \cdots \left(1 + \frac{1}{x_n}\right) \geq (n+1)^n.$$

A product of  $n$  terms might invoke feelings of an AM–GM proof, but the AM–GM shows that products are *smaller* than other things, not larger. We need a different approach. Taking a logarithm produces a different-looking kind of problem.

Consider the function  $f : (0, \infty) \rightarrow \mathbb{R}$  given by  $f(x) = \ln(1 + 1/x)$ . Then  $f''(x) = 1/x^2 - 1/(x+1)^2 > 0$ , so  $f$  is convex. By Jensen's inequality we have

$$\begin{aligned} \ln \prod_{k=1}^n \left(1 + \frac{1}{x_k}\right) &= \sum_{k=1}^n \ln \left(1 + \frac{1}{x_k}\right) \\ &\geq n \ln \left(1 + \frac{n}{x_1 + \cdots + x_n}\right) \\ &= \ln [(n+1)^n]. \end{aligned}$$

Exponentiation preserves inequalities, so applying exp cancels the logarithm to obtain the result.  $\square$

*Example.* Let  $a_1, \dots, a_n$  be positive reals with sum  $s$ . Prove both

$$\sum_{k=1}^n \frac{a_k}{s - a_k} \geq \frac{n}{n-1} \quad \text{and} \quad \sum_{k=1}^n \frac{s - a_k}{a_k} \geq n(n-1).$$

There are many ways to prove each of these statements. We'll use Jensen for the first and Cauchy for the second. Define  $f : (0, s) \rightarrow \mathbb{R}$  via  $f(x) = x/(s-x)$ . Then  $f''(x) = 2s/(s-x)^3 > 0$ , so  $f$  is convex. Jensen gives

$$\sum_{k=1}^n f(a_k) \geq n f(s/n) = \frac{s}{s - s/n} = \frac{n}{n-1}.$$

For the other inequality note that  $\sum a_k/s = 1$  and

$$\sum_{k=1}^n \frac{s}{a_k} = \left(\sum_{k=1}^n \frac{a_k}{s}\right) \left(\sum_{k=1}^n \frac{s}{a_k}\right) \geq n^2.$$

Therefore

$$\sum_{k=1}^n \frac{s - a_k}{a_k} = \sum_{k=1}^n \left(\frac{s}{a_k} - 1\right) = \sum_{k=1}^n \left(\frac{s}{a_k}\right) - n \geq n^2 - n = n(n-1). \quad \square$$

*Example.* Let  $a > b > 0$  and consider the ellipse in  $\mathbb{R}^2$  with equation  $x^2/a^2 + y^2/b^2 = 1$ . While the area of the ellipse is easy to compute (simply  $\pi ab$ ), the perimeter is notoriously difficult. The perimeter  $\ell$  is given by

$$\ell = 4a \int_0^{\pi/2} \sqrt{1 - \epsilon^2 \sin^2 t} dt,$$

where  $\epsilon$  is the eccentricity, defined as

$$\epsilon = \sqrt{\frac{a^2 - b^2}{b^2}}.$$

Recall that as  $\epsilon \rightarrow 0$ , the ellipse becomes a circle. This arc length integral is impossible to evaluate analytically (that is, there is no technique to find the answer without numerical approximation). We can use an inequality to estimate it, however. Using the Cauchy-Schwarz inequality for integrals, we can write

$$\ell = 4a \int_0^{\pi/2} 1 \cdot \sqrt{1 - \epsilon^2 \sin^2 t} dt \leq 4a \left(\int_0^{\pi/2} 1^2 dt\right)^{1/2} \left(\int_0^{\pi/2} (1 - \epsilon^2 \sin^2 t) dt\right)^{1/2} = 2a\pi \sqrt{1 - \frac{\epsilon^2}{2}}.$$

This upper bound is very close to the true value  $\ell$  when  $\epsilon$  is near 0. Specifically, the error is  $O(\epsilon^4)$  as  $\epsilon \rightarrow 0$ .  $\square$



*Example.* Let's take a moment to talk economics. Consider a real-valued function  $U$  that measures your economic happiness—we call this function utility. In reality, everything is basically a function of everything else, so I can't write a concise  $U : \star \rightarrow \mathbb{R}$ . Instead, let's focus on wealth. Alice is going to make some money in the stock market this year, or so she hopes. She doesn't know how much wealth  $w$  she'll make, so she wonders about her expected happiness. Which is larger: the expected happiness she derives from her earnings  $E[U(w)]$ , or the happiness she'd derive from her expected earnings  $U(E[w])$ ? Despite how it sounds, these are in fact different!

Now, I'm assuming a few things about Alice. As Alice earns more money, a value of a little bit more money (marginal utility) decreases. The utility function is concave! Expected value is computed with an integral against a probability measure, so thanks to Jensen's inequality we find

$$U(E[w]) = U\left(\int_X w\right) \geq \int_X U \circ w = E[U(w)],$$

so we conclude Alice would rather be guaranteed some money than have variance around that value. In other words, Alice would rather have \$100 than a 50–50 chance of either \$50 or \$150.  $\square$

*Example.* Let's talk some physics. It is fairly well-known that quantum mechanics claims the workings of the universe are inherently random. Measuring a property of a particle will give a random result with some probability distribution. After our measurement, we have changed the particle (collapsed the wavefunction). How do we know that when we measure the position of an electron and find it, that it wasn't just "there all along"? Einstein himself posed this sort of question in the early days of modern physics—he didn't like where this stuff was leading. People who disagreed with the postulates of quantum mechanics proposed "hidden variable" theories. In short, physics isn't random, it's just determined by some processes we don't see. How could one ever refute such a statement? Or prove it?

The answer, surprisingly enough, is with inequalities. No, really. In 1964 John Bell published a paper addressing the issue. He proved a collection of inequalities that any local hidden variable theory must satisfy. For example, the first one states that (under some hypotheses) for 3 different probabilistic events  $a, b, c$  we have

$$1 + C(b, c) \geq |C(a, b) - C(a, c)|,$$

where  $C(\cdot, \cdot)$  denotes the correlation between events. It becomes a simple matter of testing in the lab to find a situation that breaks Bell's inequalities. Sure enough, electron spin can be shown to disobey Bell's theorem to within an accuracy that guarantees the inequalities are wrong (this was done in 1981). Yep, physics is weird.  $\square$

We conclude with 2 tangentially related applications. One easy, one less so.

*Example.* A marching band stands in rectangular formation. The band director notices that each row has a tallest person; among those tall people Alice is the shortest. The director also notices each column has a shortest person. Among these short people, Bob is the tallest. Decide who is taller, Alice or Bob, assuming no two people have the same height.

The idea here is simple. If Alice is in Bob's row or column, Alice is clearly the taller one. Otherwise some person, say Carl, is in the same row as Alice and column as Bob. Alice is taller than Carl, but Bob is shorter than Carl. So again Alice is the tallest one.  $\square$

**Theorem** (Isoperimetric Inequality). *Suppose a smooth curve  $C$  in the plane has length  $2\pi$  and bounds a region with area  $A$ . Then  $A \leq \pi$  with equality if and only if  $C$  is a circle.*

We'll need a lemma before we begin. This result is sometimes called a Poincaré inequality.

**Lemma.** *Let  $f : [0, 2\pi] \rightarrow \mathbb{R}$  be smooth (that is, infinitely differentiable). Then*

$$\int_0^{2\pi} |f(t) - a|^2 dt \leq \int_0^{2\pi} |f'(t)|^2 dt,$$

where  $a$  is the average value of  $f$ :

$$a = \frac{1}{2\pi} \int_0^{2\pi} f(t) dt.$$

*Proof.* Write the complex-valued Fourier series of  $f$ :

$$f(t) \sim \sum_{n=-\infty}^{\infty} c_n e^{int}$$

so that  $a = c_0$  and

$$\int_0^{2\pi} |f(t) - a|^2 dt = \sum_{n \neq 0} |c_n|^2,$$

$$\int_0^{2\pi} |f'(t)|^2 dt = \sum_{n \neq 0} n^2 |c_n|^2.$$

Putting these together gives

$$\int_0^{2\pi} (|f'(t)|^2 - |f(t) - a|^2) dt = \sum_{n \neq 0} (n^2 - 1) |c_n|^2 \geq 0.$$

Equality holds if and only if  $a_n = 0$  for all  $n \neq -1, 0, 1$ ; that is,  $f(t) = a + b \cos t$  for some constants  $a, b$ .  $\square$

*Proof of Theorem.* This is an important result with a rich history. The problem is subtle because we cannot assume in advance that there is a ‘best’ curve—that is, one which bounds the most area. In only using smooth curves (those with no spikes or cusps) we’re restricting ourselves slightly, but the problem can be reduced to this case. There’s a proof using Fourier series, but we’ll instead use Green’s theorem and some basic inequalities.

Let the curve  $C$  be parametrized by via  $\gamma(s) = (x(s), y(s))$  with  $0 \leq s \leq 2\pi$ . We can do this in a way so that  $\gamma$  is unit-speed; that is,  $\|\gamma'(s)\| = 1$  for all  $s$ . In lay terms, we are traversing the curve a constant speed of 1. First we note that

$$\int_0^{2\pi} y'(s) ds = y(2\pi) - y(0) = 0,$$

since the end of the curve is the beginning. Meanwhile define the average value of  $x(s)$

$$\frac{1}{2\pi} \int_0^{2\pi} x(s) ds = a.$$

Using Green’s theorem, we can produce a well-known area formula for the region inside  $C$ :

$$A = \int_C x dy = \int_0^{2\pi} x(s)y'(s) ds = \int_0^{2\pi} (x - a)y'(s) ds + a \int_0^{2\pi} y'(s) ds = \int_0^{2\pi} (x - a)y'(s) ds.$$

Next use the AM–GM to write

$$(x - a)y' \leq \frac{(x - a)^2 + y'^2}{2}.$$

This gives us

$$A \leq \frac{1}{2} \int_0^{2\pi} [(x(s) - a)^2 + y'(s)^2] ds \leq \frac{1}{2} \int_0^{2\pi} (x'(s)^2 + y'(s)^2) ds,$$

where the last step follows from the lemma. Finally, we’ve assumed that  $\|\gamma'(s)\|^2 = 1$ , so that

$$A \leq \frac{1}{2} \int_0^{2\pi} 1 ds = \pi.$$

From our proof of the lemma, equality happens if and only if  $x(s) = a + b \cos s$  for some constants  $a, b$ . Since  $x'^2 + y'^2 = 1$ , this forces  $y(s) = c + b \sin s$  for some constant  $c$ . The curve parametrized by  $\gamma(s) = (a + b \cos s, c + b \sin s)$  is a circle.  $\square$