Math 79SI Notes

Mason Rogers

8 Choosing Strategies

8.1 Background

The technical aspects of last section focused on what *not* to do when coming up with a proof; by contrast, this section will focus on how to come up with good arguments. Of course, as with writing, the only way to improve is to practice solving problems and reading other people's work. We have developed a number of different strategies for writing proofs, but a typical math problem will not fall naturally within a single one of the techniques we have introduced. Frequently, problems require complex, multi-step arguments that employ several different techniques. In this section, we will focus on breaking down difficult problems into manageable pieces, then choosing the appropriate mathematical tools for finishing each piece. We will also discuss how the conceptual *insight* provided by good proofs enables us to generalize familiar intuition to new settings, developing certain ideas like 'projection' and 'angle' to situations far beyond $\mathbb{R}^2$ and $\mathbb{R}^n$ where one might not expect geometric intuition to be relevant.

8.2 Examples

We will begin this section with a discussion of angles between vectors. Most of the geometric questions we have about vectors can be answered in terms of lengths of vectors and angles formed between pairs of vectors. In $\mathbb{R}^2$, these notions are easy to define in a way that matches our intuition for 'length' and 'angle'.

From the Pythagorean Theorem, we know that we can describe the length of a vector $v = (v_1, v_2) \in \mathbb{R}^2$ by

$$\|v\| = \sqrt{v_1^2 + v_2^2} = \sqrt{v \cdot v}.$$

Using the law of cosines, we can also describe angles between nonzero vectors. Choose $u = (u_1, u_2)$ and $v = (v_1, v_2)$ in $\mathbb{R}^2$, and let θ be the non-reflex (i.e. less than 180 degrees) angle between them, so $0 \le \theta \le \pi$. Then by the Law of Cosines and properties of the dot product,

$$\cos(\theta) = \frac{\|u\|^2 + \|v\|^2 - \|u+v\|^2}{2\|u\|\|v\|} = \frac{u \cdot u + v \cdot v - (u+v)\cdot(u+v)}{2\|u\|\|v\|} = \frac{u \cdot v}{\|u\|\|v\|}.$$

In $\mathbb{R}^2$, it is in some sense easy to measure the angle between two nonzero vectors because of the existence of polar coordinates as follows. For two nonzero vectors u and v, we can write $u = (r_u, \theta_u)$ and $v = (r_v, \theta_v)$ with $r_u, r_v > 0$ and $\theta_u, \theta_v \in [0, 2\pi)$ uniquely. We then have the angle between them given by

$$\theta = \min\{|\theta_u - \theta_v|,\ 2\pi - |\theta_u - \theta_v|\}.$$

In other words, we can define angles *between* nonzero vectors in terms of the *absolute* angles the vectors make with the positive x-axis, which themselves are defined by inverse trigonometric functions of the vectors' Cartesian coordinates.

In $\mathbb{R}^n$ for $n > 2$, the picture is less clear, for we do not have polar coordinates we can use to assign to any nonzero vector an 'absolute' angles with respect to some fixed half-line. But perhaps we can adapt the *algebraic formulas* for length and angle in $\mathbb{R}^2$, namely

$$\|v\| = \sqrt{v \cdot v}, \cos(\theta) = u \cdot v/(\|u\|\|v\|),$$

to *define* angles between nonzero vectors in $\mathbb{R}^n$! If we can do this, we can adapt much of our geometric intuition for vectors in the plane to try to say things about vectors in n-dimensional space.

Definition 8.1. For nonzero vectors $u, v \in \mathbb{R}^n$, we define the *angle* between them to be the unique $\theta_{uv} \in [0, \pi]$ satisfying

$$\cos(\theta_{uv}) = \frac{u \cdot v}{\|u\|\|v\|}.$$

But there is one problem: in order to define θ_{uv} in this way, we need to be sure that the right side lies in the interval $[-1, 1]$, so it *is* the cosine of some (unique) $\theta \in [0, \pi]$. In other words, we need to prove:

Proposition 8.2. For $u, v \in \mathbb{R}^n$, $|u \cdot v| \leq \|u\|\|v\|$.

Idea. It is useful to split into two cases: u, v are linearly independent and u, v are linearly dependent. The first case should follow quickly because if $v \neq 0$ then there is a concrete relationship $u = \lambda v$ for some $\lambda \in \mathbb{R}$. In the second case, we can restrict the problem to $\text{span}(u, v)$, which is a *2-dimensional* subspace of $\mathbb{R}^n$ on which we can try to apply our intuition from the Euclidean plane!

Proof attempt. First, suppose u and v are linearly dependent. In the case that u or v is 0, the desired inequality becomes $0 \leq 0$, which is true. Thus, we reduce to the case $u, v \neq 0$, so $u = \lambda v$ for some $\lambda \in \mathbb{R}$. Then

$$|u \cdot v| = |\lambda v \cdot v| = |\lambda|\|v\|^2 = \|\lambda v\|\|v\| = \|u\|\|v\|,$$

so the inequality holds (and is actually an equality).

Now suppose u and v are linearly independent, so $\text{span}(u, v)$ is a 2-dimensional subspace of $\mathbb{R}^n$, which is a plane. Let $\Pi = \text{span}(u, v)$. If there is an orthonormal (in $\mathbb{R}^n$) basis (i.e. a basis of orthogonal unit vectors) $\{e_1, e_2\}$ of Π then in coordinates relative to such a basis the dot product (and so length) come out *exactly* as in Euclidean geometry:

$$(ae_1 + be_2) \cdot (a'e_1 + b'e_2) = aa'e_1 \cdot e_1 + (ab' + a'b)e_1 \cdot e_2 + bb'e_2 \cdot e_2 = aa' + bb'.$$

We have then reduced to the known 2-dimensional case! $\square$

Our above attempt did not actually finish the job: to reduce to the 2-dimensional case, we need to find an orthonormal basis for Π. Why should one exist for such general Π? Especially for $n > 3$, the existence of such a basis is not obvious.

However, our attempt gave us two crucial insights: we proved equality (and therefore also the inequality) in the case where u and v are linearly dependent, and we learned that if we want to reduce to a 2-dimensional picture, we just need to find in Π something like the 'standard basis' for $\mathbb{R}^2$, to make Π with the dot product *from* $\mathbb{R}^n$ look like $\mathbb{R}^2$ with *its* usual dot product. We can combine these two findings to develop a new approach:

Idea. We already finished the case where u and v are linearly dependent, so we now assume u and v are linearly independent, and in particular both are nonzero. We will try to break u into components along v and orthogonal to v under the dot product in $\mathbb{R}^n$. It will be the orthogonal component that is responsible for the inequality, since the parallel component gave equality in the linearly dependent case.

Proof. In 2 dimensions, the projection formula for the component of u along $v \neq 0$ is given by

$$\operatorname{proj}_v(u) = \frac{u \cdot v}{v \cdot v} v. \tag{1}$$

We may hope that this formula extended to $\mathbb{R}^n$ *as a definition* satisfies the *properties* we know in $\mathbb{R}^2$, such as the fact that $u - \operatorname{proj}_v(u)$ and v are orthogonal. Indeed, taking equation (1) as the *definition* of the left side, we compute

$$\begin{aligned}(u - \operatorname{proj}_v(u)) \cdot v = u \cdot v - \operatorname{proj}_v(u) \cdot v &= u \cdot v - \frac{(u \cdot v)(v \cdot v)}{v \cdot v} \\ &= u \cdot v - u \cdot v \\ &= 0.\end{aligned}$$

We now have the tools we need to finish in two ways.

One way is to observe that $w := u - \operatorname{proj}_v(u)$ is nonzero (since $\operatorname{proj}_v(u) \in \operatorname{span}(v)$ by definition yet $u \notin \operatorname{span}(v)$ by the linear independence of u and v) and orthogonal to v by the calculation above. Thus, $\{w/\|w\|, v/\|v\|\}$ is orthonormal in the 2-dimensional plane Π, hence an orthonormal basis of Π. This allows us to identify Π with $\mathbb{R}^2$ *respecting dot products*,[1] so we are done.

Alternatively, let $z = \operatorname{proj}_v(u)$ (and note $z \neq u$ by the above argument that $w \neq 0$), so $(u - z) \cdot v = 0$ and $u - z \neq 0$. We then have via the *definition* of z that

$$\begin{aligned}0 < \|u - z\|^2 = \left\|u - \frac{u \cdot v}{v \cdot v} v\right\|^2 &= u \cdot u - 2\frac{(u \cdot v)^2}{v \cdot v} + \frac{(u \cdot v)^2 (v \cdot v)}{v \cdot v} \\ &= \|u\|^2 - 2\frac{(u \cdot v)^2}{v \cdot v} + \frac{(u \cdot v)^2 (v \cdot v)}{v \cdot v} \\ &= \|u\|^2 - \frac{(u \cdot v)^2}{v \cdot v}\end{aligned}$$

[1] In mathematics, we say an identification or map between objects 'respects' a property if the property is unchanged by the map.

$$= \|u\|^2 - \frac{|u \cdot v|^2}{\|v\|^2},$$

from which we conclude $|(u \cdot v)| < \|u\|\|v\|$ as desired. $\square$

We thus have two complete proofs of Proposition 8.2 (the second perhaps less intuitive), plus a little bit more: the inequality is equality *if and only if* u and v are linearly dependent. We therefore have the Cauchy-Schwarz Inequality:

Theorem 8.3. (Cauchy-Schwarz Inequality in $\mathbb{R}^n$) For any vectors $u, v \in \mathbb{R}^n$, $|u \cdot v| \leq \|u\|\|v\|$, with equality if and only if u and v are linearly dependent.

The first upshot of this result is that our definition of angles between vectors is well-posed. From the equality case of the Cauchy-Schwarz Inequality (i.e. the case of linear independence), we also can see that our *definition* of angle matches our geometric intuition: for nonzero u and v, $\cos(\theta_{uv})$ equals $+1$ and -1 exactly when u and v are parallel or antiparallel (i.e. $u = \lambda v$ for $\lambda > 0$ or $\lambda < 0$, respectively). Moreover, $\cos(\theta_{uv}) = 0$ exactly when $u \cdot v = 0$ in $\mathbb{R}^n$, as expected.

Another upshot of our result comes into focus upon recognition that our proof *never required us to write u and v in terms of coordinates in $\mathbb{R}^n$*. All we needed was certain abstract properties of $\mathbb{R}^n$ and the dot product: a vector space over $\mathbb{R}$ (a concept to be defined fully in the next section) with "length" given by a "dot product".[2] In fact, we only used the bilinearity,[3] symmetry, and nonnegativity properties of the dot product on $\mathbb{R}^n$ in our proof (e.g. $v \cdot v \geq 0$ for all v, with equality if and only if $v = 0$). We may therefore ask if there are any other mathematical structures on which we can define a natural 'dot product' as well as what the significance of the resulting 'Cauhcy-Schwarz Inequality' is in those settings. The following examples will all be vector spaces over $\mathbb{R}$, i.e. sets equipped with the operation of addition and scalar multiplication satisfying properties akin to $\mathbb{R}^n$, on which we can *define* a suitable 'dot product' to make sense of the Cauchy-Schwarz Inequality.

Example 8.4. Let X and Y be random variables (i.e. variables whose values are the result of random processes) and let $E[-]$ denote the expected value. One can check that the assignment $(X, Y) \mapsto E[XY]$ is symmetric and bilinear and that $E[X^2] \geq 0$ with equality if and only if $X = 0$. We can thus use the second proof of the Cauchy-Schwarz Inequality in this setting to obtain

$$|E[XY]|^2 \leq E[X^2]E[Y^2].$$

More meaningfully, applying the inequality to the variables $X - E[X]$ and $Y - E[Y]$ gives the *covariance inequality*

$$(\mathrm{Cov}(X, Y))^2 \leq \mathrm{Var}(X)\mathrm{Var}(Y),$$

which bounds the amount that two random variables can covary (i.e. vary interdependently) by how much they each vary individually.

[2]This is hinting at the notion of an inner product space, a topic within functional analysis.

[3]A function $f : V \times V \to \mathbb{R}$ is called *bilinear* if it is linear in both arguments, so $u \mapsto f(u, v)$ is linear for all $v \in V$ and $v \mapsto f(u, v)$ is linear for every $u \in V$.

Example 8.5. Let $\{a_n\}_{n\in\mathbb{N}}$ and $\{b_n\}_{n\in\mathbb{N}}$ be sequences of real numbers which satisfy $\sum_n a_n^2 < \infty$ and $\sum_n b_n^2 < \infty$. It can be checked that the assignment $(\{a_n\},\{b_n\}) \mapsto \sum_n a_n b_n$ is symmetric and bilinear, and clearly $\sum_n a_n^2 \geq 0$ with equality if and only if $a_n = 0$ for all n. The Cauchy-Schwarz Inequality therefore applies in this setting as

$$\left|\sum_n a_n b_n\right|^2 \leq \left(\sum_n a_n^2\right)\left(\sum_n b_n^2\right).$$

This is useful to prove that certain series are convergent.

Example 8.6. Let $f : [0,1] \to \mathbb{R}$ and $g : [0,1] \to \mathbb{R}$ be continuous functions. One can check that the assignment $(f,g) \mapsto \int_0^1 fg\ dx$ is symmetric and bilinear, and it is true that $\int_0^1 f^2\ dx \geq 0$ with equality if and only if $f = 0$. This gives an application of the Cauchy-Schwarz Inequality to vector spaces of *functions*, specifically

$$\left(\int_0^1 fg\ dx\right)^2 \leq \left(\int_0^1 f^2\ dx\right)\left(\int_0^1 g^2\ dx\right).$$

The notion of treating functions as 'points' in a vector space (such as the set of all polynomials, the set of all continuous functions on $[0,1]$, the set of all infinitely differentiable functions on $\mathbb{R}$, etc.) may seem alien at first, for functions do not match the mental picture we have of vectors as arrows with a length and direction. However, we can *expand* our mental picture, and with a little practice, treating function spaces as vector spaces is highly useful. For instance, the set of solutions to a linear ordinary differential equation of order n is an n-dimensional vector space (which we will explore in the next section). We can also adapt the Gram-Schmidt process on $\mathbb{R}^n$ to apply to the space of polynomials with degree less than n equipped with the 'dot product' $(f,g) \mapsto \int_0^1 f(x)g(x)\ dx$ to construct sets of 'orthonormal polynomials' which are useful in solving differential equations and in physics. (See Exercise 1b.)

This is the value of *appropriate generalization* in mathematics: we start with a mental picture and use it to prove something specific, but when we step back and look at the *essential properties* of concepts used in the proof, we sometimes find them in another structure that initially may seem quite different! Ultimately, we could recast this entire preceding discussion in the language of inner product spaces (i.e. vector spaces over $\mathbb{R}$ with a suitable 'dot product'), the broad setting in which the Cauchy-Schwarz Inequality holds.

We can also leverage our geometric intuition for vectors in $\mathbb{R}^n$ in these new (sometimes 'infinite-dimensional'!) settings. Focusing specifically on Example 8.6, we may be curious if we can find an infinite 'orthonormal basis' $\{e_n\}_{n\in\mathbb{Z}}$ for the space of continuous functions on $[0,1]$ and talk about 'orthogonally projecting' functions onto the elements of this basis. It turns out (after a bit of work) that the family of functions

$$\{1,\ \{\sqrt{2}\sin(2\pi kx)\}_{k\in\mathbb{N}},\ \{\sqrt{2}\cos(2\pi \ell x)\}_{\ell\in\mathbb{N}}\}$$

does the trick, and the coefficients given by projecting a function f onto each of these basis vectors have important physical and mathematical significance. These are the beginnings of the topic of *Fourier analysis*. The 'Fourier coefficients' of a function $f : [0, 1] \to \mathbb{R}$ given by $c_n = \int_0^1 f e^{2\pi i n x}\, dx$ are helpful to solve certain partial differential equations and to study f itself. And while this formula for a 'projection' looks different from the orthogonal projection formula onto a line in $\mathbb{R}^n$, the appropriate notion of 'dot product' makes the two formulas instances of the *same* concept!

The process of generalization is central to studying mathematics, and it is both useful and rewarding to discover a broad setting that isolates the essential properties of a mathematical structure.

8.3 Exercises

1. This problem will consider what the Gram-Schmidt process looks like for vector spaces of functions. On $\mathbb{R}^n$, the Gram-Schmidt process takes a list of $k \leq n$ linearly independent vectors $\{v_1, \ldots, v_k\}$ and returns a list of k orthogonal vectors $\{e_1, \ldots, e_k\}$ such that $\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k)$ via the following procedure:

   Set $e_1 = v_1$. We will recursively define e_j for $j > 1$ by subtracting off the 'parallel component' of v_j which lies in the subspace $\operatorname{span}(e_1, \ldots, e_{j-1})$ to get a vector e_j orthogonal to each element of $\{e_1, \ldots, e_{j-1}\}$. Precisely, for $1 < j \leq k$, define $e_j = v_j - \sum_{i<j} \operatorname{proj}_{e_i}(v_j)$. Afterwards, we can divide each e_i by its length if we want to get an orthonormal set.

   As an example, consider the vectors $(1, 1, 0), (0, 2, 1), (3, 1, 2) \in \mathbb{R}^3$. We begin with $v_1 = (1, 1, 0)$ and set $e_1 = v_1 = (1, 1, 0)$. Since $v_2 = (0, 2, 1)$, we have

   $$e_2 = v_2 - \operatorname{proj}_{e_1}(v_2) = (0, 2, 1) - (1, 1, 0) = (-1, 1, 1).$$

   Lastly, we have $v_3 = (3, 1, 2)$, so

   $$e_3 = v_3 - \operatorname{proj}_{e_1}(v_3) - \operatorname{proj}_{e_2}(v_3) = (3, 1, 2) - (2, 2, 0) - (0, 0, 0) = (1, -1, 2).$$

   (a) What basis does the Gram-Schmidt process give for $\mathbb{R}^3$ given the inputs

   $$\{(7, 0, 24), (1, 0, 1), (3, 4, 5)\}?$$

   (b) Using the notion of 'dot product' given in Example 8.6, what orthonormal set does the Gram-Schmidt process return given the inputs $\{1, x, x^2, x^3\}$? The resulting set (up to rescaling factors) is the first few terms of what are known as the *Legendre polynomials*, which have applications in math and physics.

2. Let $\{e_1, \ldots, e_k\}$ be a sequence of vectors in $\mathbb{R}^n$ given by performing the Gram-Schmidt procedure on linearly independent vectors $\{v_1, \ldots, v_k\}$. Prove that the e_i are nonzero and orthogonal: $e_i \cdot e_j = 0$ if and only if $i \neq j$. (*Hint: Induct on k.*)

3. Let $\{e_1, \ldots, e_k\}$ be a sequence of vectors in $\mathbb{R}^n$ given by performing the Gram-Schmidt procedure to linearly independent vectors $\{v_1, \ldots, v_k\}$. Prove that

   $$\operatorname{span}(e_1, \ldots, e_k) = \operatorname{span}(v_1, \ldots, v_k).$$

   (*Hint: Induct on k.*)

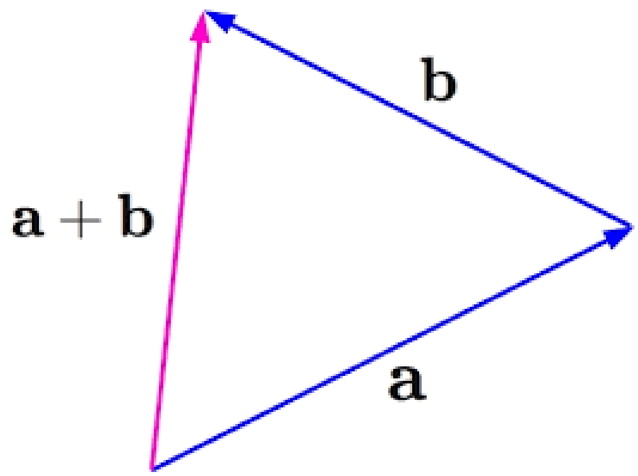

Figure 1: Illustration of the triangle inequality, which states that the length of a sum of vectors is never greater than the sum of their lengths.

4. Use the Cauchy-Schwarz Inequality for $\mathbb{R}^n$ to prove the 'triangle inequality' for $\mathbb{R}^n$, which states that $\|u+v\| \leq \|u\| + \|v\|$ for all $u, v \in \mathbb{R}^n$, with equality if and only if u is a nonnegative scalar multiple of v. (*See Figure 1.*)

5. Use the Cauchy-Schwarz Inequality for continuous functions from $[0,1]$ to $\mathbb{R}$ and the previous exercise to formulate and prove an analog of triangle inequality for such functions. What does this say in terms of definite integrals?

6. Read another proof of the Cauchy-Schwarz Inequality online. In what setting does the proof apply? What approach does the proof take? Compare and contrast with the proof given in this section.

7. Prove the Pythagorean Theorem for $\mathbb{R}^n$, which states for $u, v \in \mathbb{R}^n$ that if $u \cdot v = 0$ then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$. Observe that the proof uses the *definition* $\|x\|^2 = x \cdot x = x_1^2 + \ldots + x_n^2$, which seems to assume exactly the Pythagorean Theorem (at least when $n = 2$)! Reflect: in your opinion, is the Pythagorean Theorem still a theorem within the language of dot products, or a definition? What logical status does it hold?

8.4 Further Reading

"Various Proofs of the Cauchy-Schwarz Inequality", Hui-Hua Wu and Shanhe Wu	More proofs of the Cauchy-Schwarz Inequality framed in the setting of $\mathbb{R}^n$. It is worth noting which proofs use techniques specific to $\mathbb{R}^n$ and which can be extended to more general vector spaces with 'dot products'. rgmia.org/papers/v12e/Cauchy-Schwarzinequality.pdf
"Greek Versus Babylonian Mathematics", Richard Feynman	In this short video (10:20), physicist Richard Feynman talks about the relationship between physics and math as well as the differences between what is axiomatic and what is worthy to prove from different perspectives. youtube.com/watch?v=YaUlqXRPMmY