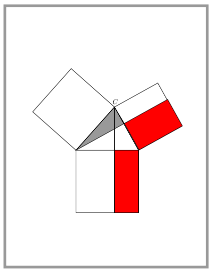


CHAPTER 2

Elementary coordinate geometry

The page on which you draw may, for all practical purposes, be considered as a window onto a plane extending uniformly to infinity. We shall not look too closely at the assumptions made in this statement, but instead rely strongly on intuition depending on visual experience to deduce important facts about this plane.

Using computers to draw requires translating from geometry to numbers—i. e. to a coordinate system—and back again. There are a few basic formulas that are used over and over again. It is best to memorize them. Calculating the distance between points whose coordinates are given requires Pythagoras' Theorem, which we recall almost at the beginning of this chapter. Before that, however, comes a discussion of the areas of parallelograms; and even before that comes a short note about distinguishing points from vectors. Towards the end of the chapter we shall look at a number of results related principally to projections.

For many readers, the results presented in this chapter will be well known. Even for them, however, the use of visual reasoning might be interesting and, in some aspects, novel.

1. Points and vectors

It is important to distinguish **points** from **vectors**, even though a coordinate system assigns a pair of numbers to either a point or a vector. Points are . . . well, points. They possess no attribute other than **position**, and in particular they are (in spite of how they are drawn!) without dimension or size or color or smell or . . . anything other than position. Vectors, on the other hand, have magnitude and direction. They measure **relative position**. It is very important to keep in mind that *both points and vectors are objects independent of which coordinate system is being used*.

Vectors can be added to each other, and they can be multiplied by constants. There is also a kind of limited algebra involving points. If $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$ are two points then there is a unique vector with tail at P and head at Q whose coordinates are $x_Q - x_P$ and $y_Q - y_P$, describing the relative position of the two points. It is written as $Q - P$. One reason that it is common to confuse points with vectors is that to each point P corresponds the vector $P - O$ from the origin to P. However, if the coordinate system changes, the origin may change. The points themselves won't change, but the vectors they correspond to will likely do so.

If we are given a coordinate system, the vector with coordinates x, y will be $[x, y]$ and the point with those coordinates will be (x, y). The (x, y) corresponds to the vector $[x, y]$ from the origin to itself—but I repeat that this point and this vector are not the same geometrical object.

If P is a point and v a vector it makes sense to consider $P + v$ as a point—it is the point Q such that $Q - P = v$. It is the point P **displaced** or **translated** by v. If t is a real number between 0 and 1 then the point t of the way from P to Q is equal to $P + t(Q - P)$ with coordinates $(1 - t)x_P + tx_Q, (1 - t)y_P + ty_Q)$. I write it as $(1-t)P + tQ$. It is a kind of weighted average of P and Q. For example, the point point midway between P and Q is $(1/2)P + (1/2)Q$. As we shall see later, we can also take weighted averages of collections of several points.

In summary: we can subtract two points to get a vector; or calculate a weighted average to obtain another point; but the sum of two points or a scalar multiple of a point makes no intrinsic sense.

2. Areas of parallelograms

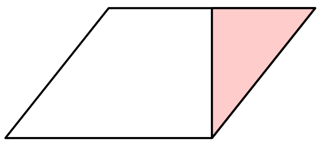
Area is a somewhat sophisticated concept, not easily analyzed in complete rigour. We are used to thinking of it as a number, but of course the number involved depends on the units involved—it is really a ratio of the area of a region to that of a unit square. So area seems to be a fundamental, geometrical characteristic of a region. It is interesting that Euclid starts off Book I of the *Elements* with properties of area that are encapsulated in a few particularly simple axioms. One of these is that *congruent regions have the same area*. Recall that one region is said to be **congruent** to another if it is obtained from it by translation, rotation, or reflection, without altering the relative distances between points of the region. Another basic principle is the **additive principle of areas**: *If two regions have the same area and congruent regions are added to each, then the new regions also have the same area*. This leads to the following more general criterion, which is very close to the one used implicitly by Euclid in his treatment of area:

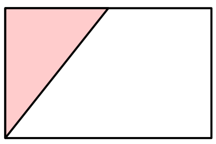
- *(Euclid's first criterion for areas) Two regions have the same area if they can be chopped into smaller pieces which are congruent.*

This does not allow for a treatment of areas with curved boundaries, but it does allow us to see that

- *A parallelogram has the same area as a rectangle with the same base and height.*

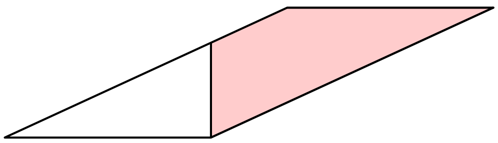
Why is this true? A proof according to Euclid's criterion must show how to decompose the parallelogram and the rectangle into congruent pieces. In some circumstances, this is simple. The complexity of the decomposition involved depends on how skewed the parallelogram is, or how far removed it is from the rectangle it is to be compared to. If it not too skewed, then we can lop off a triangle at one end of the parallelogram and paste it in at the other to make a rectangle.



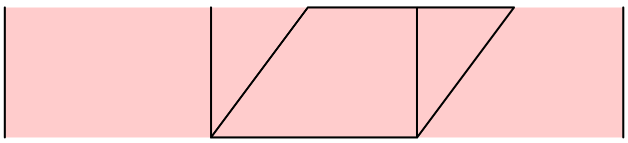


But this means exactly that in these circumstances we can decompose the rectangle and the parallelogram into congruent pieces.

If the parallelogram is very skewed, however, then what we lop off at one end is not a triangle, and this argument fails.



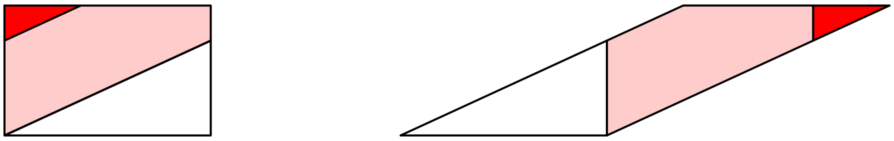
The first, simple argument works when the parallelogram is **mildly skewed**—i. e. when the piece chopped off one end is indeed a triangle. This happens when the entire parallelogram fits into the region shown in this figure:



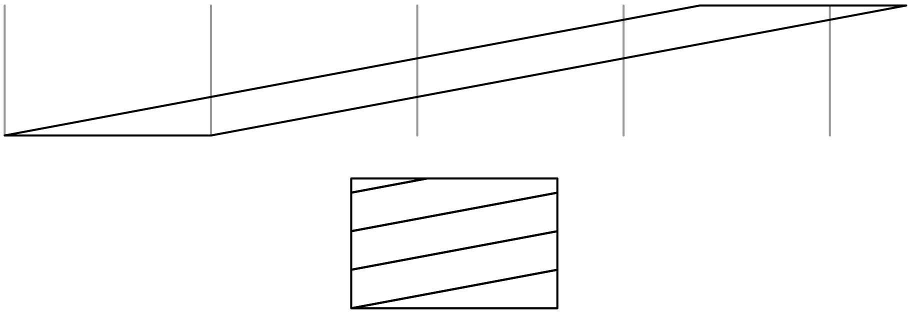
Just about all proofs of the result are the same for mildly skewed parallelograms. There are lots of different ways to proceed for the rest. Here are a few:

Proof 1. We can get an idea of a possible way to proceed if we again translate the lopped off region to the left and glue it on, just as if it were a triangle.

The natural thing to do now is to lop off the bit of triangle at the far right and shift it back again to fill in a rectangle. This gives us finally a way to chop up both the rectangles and the original parallelogram into congruent regions.



As the parallelogram gets more and more skewed, the number of pieces the parallelogram gets chopped up into increases, but there is a definite pattern to the way things go. Here are a couple of pictures to show what happens:



Exercise 1. *Define the* **skew** *of a parallelogram to be the length of the perpendicular projection of its upper left corner onto its base line, divided by the length of the base. Count negatively to the left. A parallelogram is a rectangle if and only if its skew is* 0. *The argument above shows that if the skew* s *satisfies* $0 < s \le 1$, *then the simple decomposition will prove the claim. Explain by a picture what happens if* $-1 \le s < 0$.

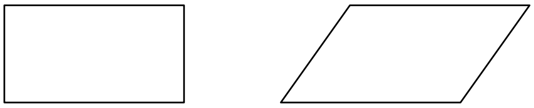
Exercise 2. *Explain the argument in the previous exercise by producing figures in PostScript.*

Exercise 3. *The second group of pictures shows what happens if* $1 < s \le 2$. *What about* $-2 \le s < -1$? $2 < s \le 3$?

Exercise 4. *If the skew* s *satisfies* $n < s \le n+1$ *(*n *positive), what is the least number of pieces in the decomposition of the parallelogram and rectangle into congruent pieces suggested by the above reasoning?*

Exercise 5. *The reasoning above has just shown how the decomposition of rectangle and parallelogram works in a few cases, and the exercises above have shown how to include a few more cases. Write out in detail a recipe for making congruent decompositions of rectangle and parallelogram that will prove the claim when the skew* s *satisfies* $0 < n < s \le n+1$.

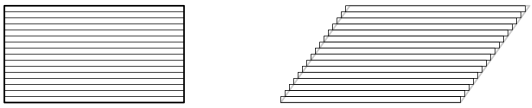
Proof 2. The transformation of a rectangle into a parallelogram with the same base and height is called a **shear**.



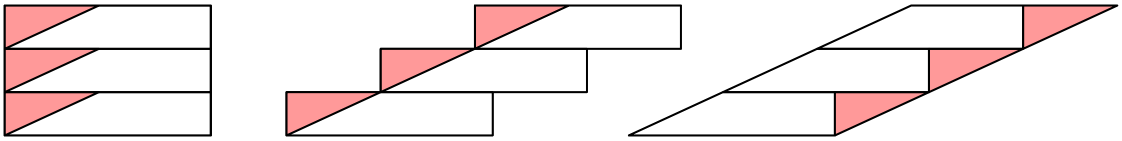
The result we are proving amounts to this:

- *Shears preserve area.*

A shear can be visualized as a continuous sequence of sliding motions, if you think of the original rectangle as made up of very thin strips piled on top of each other. Like a sliding deck of cards.

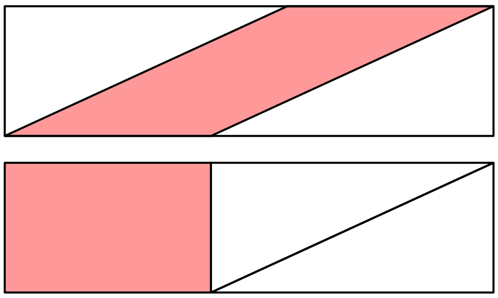


In this way, preservation of area under shear becomes intuitive—you can think of the rectangle as an infinite number of horizontal strips piled on top of one another. Shearing it just translates each of these, not changing its area, hence not changing the area of the total figure as it is sheared. This sort of reasoning is not always dependable, but it is valuable nonetheless. Historically, it played an important role in the development of calculus long before the nature of limits was understood clearly. Here, however, it suggests an entirely valid and perhaps the best motivated proof of the result. We don't have to chop up the rectangle into an infinite number of horizontal strips, but just enough strips so that each one becomes only a mildly skewed parallelogram when it is sheared.



Proof 3. That two parallelograms with the same base and height have the same area is Proposition I.35 in Euclid's *Elements*. But his proof of it depends on the **subtractive principle of areas**: *If congruent regions are taken away from two regions of equal area, then the remaining regions have equal area.*

The simplest of all proofs depends on this principle, but it is not the same as Euclid's. It can be explained in a single pair of diagrams:



Exercise 6. *Analyze Euclid's own proof of I.35 by breaking it up into a sequence of pictures.*

Exercise 7. *Neither the result nor any of these proofs depends on interpreting area as a number, nor even how to compare the area of two distinct rectangles. Make a first step in this direction by explaining in your own words how to construct geometrically, for any rectangle with base b and height a, a square of the same area. (This is II.14 of Euclid's Elements).*

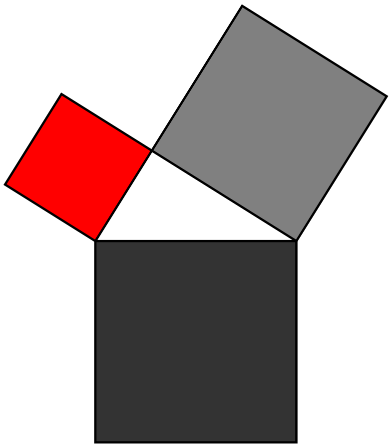
3. Lengths

The principal result concerning lengths is Pythagoras' Theorem.

- *For a right-angled triangle with short sides a and b and long side c, $c^2 = a^2 + b^2$.*

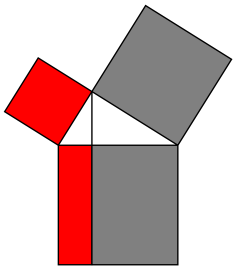
This result, as also the one in the previous section, can be phrased in terms of equality of areas. We erect squares on each of the sides of the given triangle. The Theorem asserts neither more nor less than that

- *The area of the largest square is the sum of the areas of the other two.*

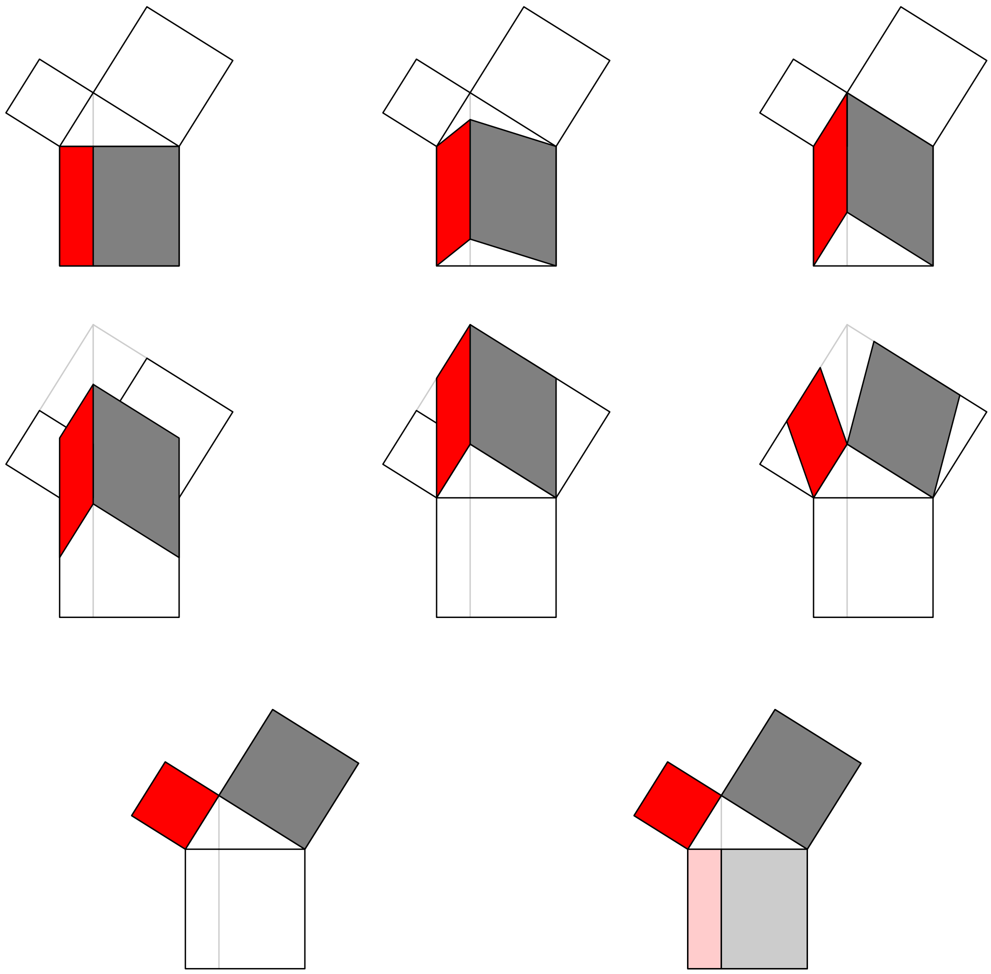


There are many ways to prove Pythagoras' Theorem. There is even a book which purports to contain 365 different proofs, one for each day of the year (and includes a few extra). The proof given here is close to Euclid's own (Pythagoras' Theorem is I.47 in the *Elements*). I first saw it in a book by Howard Eves, but it probably derives originally from the proof of a generalization of Pythagoras' Theorem due to the later Alexandrian mathematician Pappus.

It exhibits a decomposition of the larger square (the 'square on the hypotenuse') into rectangles whose areas match the smaller squares (the 'squares on the sides').



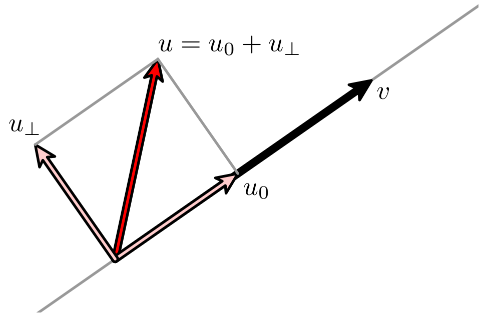
The proof proceeds by a sequence of shears and translations, which we know to preserve areas, transforming the rectangles in the large square into the squares on the sides.



Exercise 8. *This is very elegant, but if looked at closely there appear to be a few gaps. Find them and fill them in.*

4. Vector projections

If v is a vector in the plane, then any other vector u can be expressed as the sum of a vector u_0 parallel to v and a vector $u_\perp$ perpendicular to it.



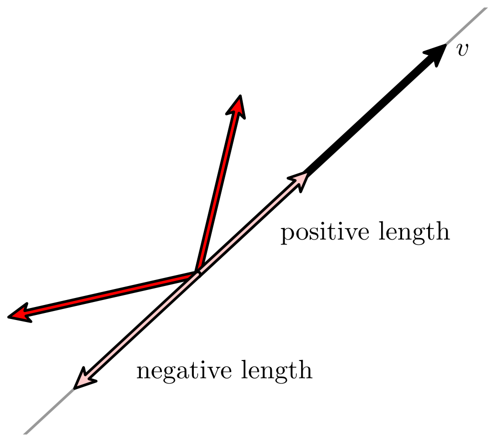
We ask the following question:

- *If $v = [a, b]$ and $u = [x, y]$, how do we calculate u_0?*

The projection u_0 will be a scalar multiple of v, say $u_0 = cv$, and our problem is to calculate c. The length of the projection will be

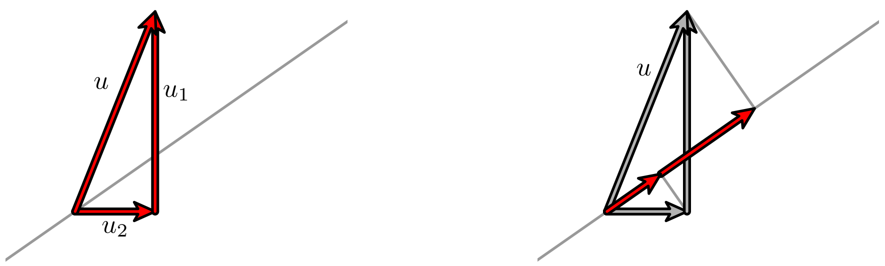
$$\|u_0\| = |c|\|v\| \ .$$

So if we know the length $\|u_0\|$, we can calculate $|c| = \|u_0\|/\|v\|$. In order to get the sign of c, we introduce the notion of **signed length**. If the ordinary length of the projection is s, then its signed length (relative to the vector v) is just s if the projection is in the same direction as v, but $-s$ if in the opposite direction.



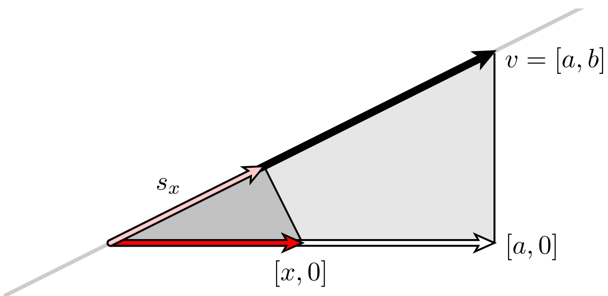
We now need to find a formula for the signed length of u_0.

The first observation is that the parallel projection is an additive function of u, which means that if $u = u_1 + u_2$ then the projection of u is the sum of the projections of u_1 and u_2.



Since u is equal to the sum of its projections onto the x and y axes, it is only necessary to find the signed lengths of the projections of $[x, 0]$ and $[0, y]$ and add them together.

Let's look at the projection of $[x, 0]$.

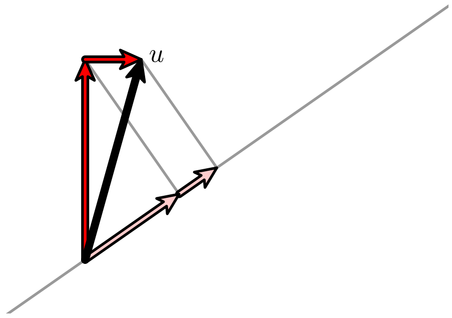


Let s_x be the signed length of the projection of $[x, 0]$, and let $v = [a, b]$. The two triangles in the figure are similar, so we see that

$$\frac{s_x}{x} = \frac{a}{\|v\|}, \quad s_x = \frac{ax}{\|v\|} .$$

Similarly if s_y is that of $[0, y]$ then

$$s_y = \frac{by}{\|v\|} .$$



The figure deals with positive lengths, but the final result remains valid for negative ones as well. Hence the signed length of $u = [x, y]$ is

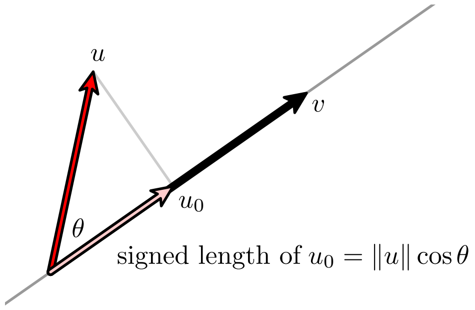
$$s = s_x + s_y = \frac{ax + by}{\|v\|} .$$

- *If $u = [x, y]$ and $v = [a, b]$ are vectors, the projection of u onto a line parallel to v is*

$$\frac{ax + by}{\sqrt{a^2 + b^2}} \, \frac{[a, b]}{\sqrt{a^2 + b^2}} = \left(\frac{ax + by}{a^2 + b^2} \right) [a, b] .$$

There is another formula for the signed length. If θ is the angle between u and v then

$$s = \|u\| \cos\theta .$$



If we compare the two formulas we see that the angle θ between the vectors $v = [a, b]$ and $u = [x, y]$ can be found from this identity:

$$\cos\theta = \frac{ax + by}{\|u\| \, \|v\|} .$$

If $u = [a, b]$ and $v = [x, y]$ then the numerator of the formula above is called their **dot product**:

$$u \bullet v = ax + by .$$

Thus the formula for the cosine of the angle between them becomes

$$\cos\theta = \frac{u \bullet v}{\|u\|\,\|v\|} \ .$$

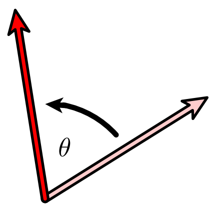
The dot product satisfies a number of simple formal algebraic rules:

$$\begin{aligned} cx \bullet y &= c(x \bullet y) \\ x \bullet cy &= c(x \bullet y) \\ (x+y) \bullet z &= x \bullet z + y \bullet z \\ x \bullet x &= \|x\|^2 \end{aligned}$$

where $\|x\|$ is the length of the vector x, the distance of its head from its tail.

5. Rotations

We now look at a new problem: *Suppose we start with the vector $v = [x, y]$ and rotate it around the origin by angle θ. What are the coordinates of the new vector?*

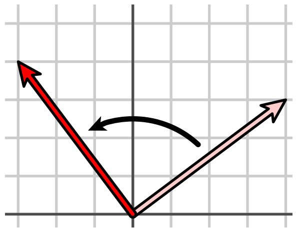


Rotation by θ

The answer to this depends directly on the answer to the simplest case, that where the angle is 90°.

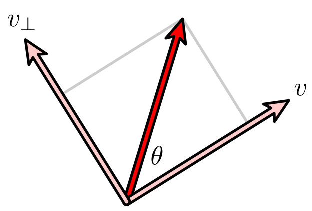
- *If $v = [x, y]$ then rotating v counter-clockwise by 90° gives us $v_\perp = [-y, x]$.*

This can be seen easily in this picture:



Rotation by 90°

But now we are in good shape, since from this figure



v rotated by θ is $(\cos\theta)v + (\sin\theta)v^\perp$

we can deduce:

- *If $v = [x, y]$ then rotating v by θ gives us*

$$(\cos\theta)\, v + (\sin\theta)\, v_\perp = [x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta] \ .$$

This expression can be calculated by a matrix. Rotation by θ takes the vector $[\, x \quad y \,]$ to

$$[\, x \quad y \,] \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

In this book, vectors will usually be row vectors, and matrices will multiply them on the right. This is a common convention in computer graphics, as opposed to that in mathematics, and makes especially good sense in dealing with PostScript calculations, as we shall see.

If v is the unit vector $[\cos\alpha, \sin\alpha]$ and V_* is v rotated by β, we obtain on the one hand the vector corresponding to angle $\alpha + \beta$, and on the other, according to the formula for rotations I have just derived, the vector

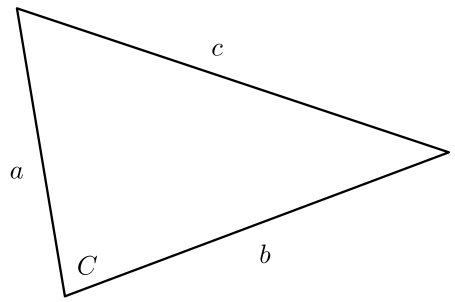
$$[\cos(\alpha+\beta), \sin(\alpha+\beta)] = [\cos\alpha, \sin\alpha] \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \ .$$

This gives us the **cosine and sine sum rules**:

$$\begin{aligned} \cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta \\ \sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta \end{aligned}$$

6. The cosine rule

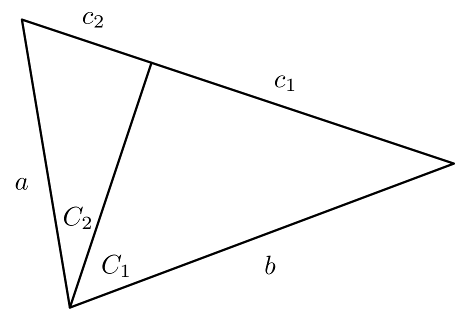
The **cosine rule** is a generalization of Pythagoras' Theorem that applies to triangles which are not necessarily right-angled.



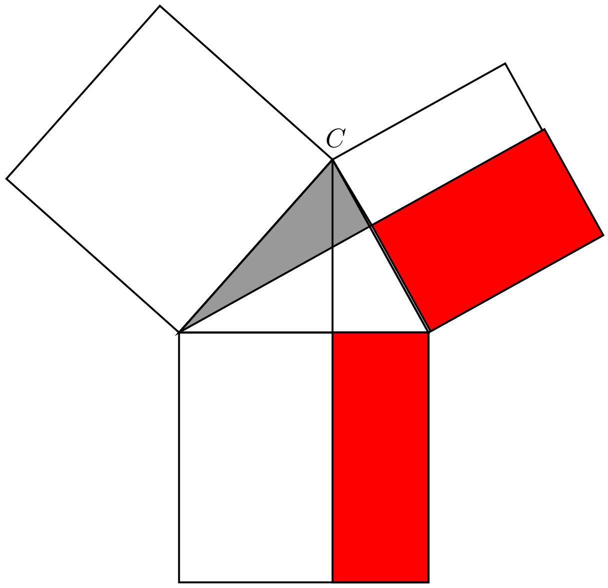
- (Cosine rule) *In a triangle with sides a, b, c, and angle C opposite c*

$$c^2 = a^2 + b^2 - 2ab\cos C \ .$$

I sketch three proofs, the first a mixture of algebra and geometry, the second almost purely geometric, and the third almost entirely algebraic.



The first one applies the cosine sum formula and Pythagoras' Theorem to the diagram above.



The second generalizes Euclid's proof of Pythagoras' Theorem. It begins by showing that the two rectangular areas in the diagram above have equal areas, and then finally applies the definition of cosine.

As for the third, it uses the dot-product in 2D. Recall:

- *If the angle between two vectors u and v is θ then*

$$\cos\theta = \frac{u \bullet v}{\|u\|\,\|v\|}\ .$$

Now start by writing the cosine rule in terms of vectors. We want to show that

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

where u and v are vectors along the sides of the triangle with lengths a and b, and therefore the third side of the triangle is $u - v$. Following the equation above we reduce

$$\begin{aligned}\frac{(u-v)\bullet(u-v) - u\bullet u - v\bullet v}{2\|u\|\,\|v\|} &= \frac{u\bullet u - 2\,u\bullet v + v\bullet v - u\bullet u - v\bullet v}{2\|u\|\,\|v\|} \\ &= \frac{u\bullet v}{\|u\|\,\|v\|} \\ &= \cos\theta \quad \text{sure enough.}\end{aligned}$$

Exercise 9. *Finish both the first two proofs. For the second, add several diagrams illustrating the argument.*

7. Dot products in higher dimensions

The dot product of two vectors in any number n of dimensions greater than two or three is **by definition** the sum of the products of their coordinates:

$$(x_1, x_2, \ldots, x_n)\bullet(y_1, y_2, \ldots, y_n) = x_1y_1 + x_2y_2 + \cdots + x_ny_n \;.$$

The formal rules we have seen to be true in two dimensions hold also in three dimensions and more.

- *For vectors u and v in 2 or 3 dimensions*

$$u\bullet v = \|u\|\,\|v\|\,\cos\theta$$

  where θ is the angle between them.

The proof of this is just the reverse of the third argument in the last section.

In particular:

- *The dot product of two vectors is 0 precisely when they are perpendicular to each other.*

Of course it is only in 2 or 3 dimensions that we have a geometric definition of the angle between two vectors. In higher dimensions, this formula is used to define that angle algebraically.

8. Lines

One way of representing lines in the plane is by means of an equation

$$y = mx + b \;.$$

But this cannot represent vertical lines, those parallel to the y-axis, which have an equation $x = a$. The uniform way to represent all lines is by means of an equation

$$Ax + By + C = 0 \;.$$

For example, $y - mx - b = 0$ or $x - a = 0$. Lines which are not vertical are those with $B \neq 0$, in which case we can solve for y. The problem with this scheme is that if $Ax + By + C = 0$ is the equation of a line and c is a non-zero constant then $cAx + cBy + cC = (cA)x + (cB)y + (cC) = 0$ is also the equation of the same line. This means that the coordinates $[A, B, C]$ of a line are **homogeneous**—only determined up to multiplication by a non-zero scalar. This is the first place in which homogeneous coordinates occur in this book. They will play an extremely important role later on, especially when we come to 3D graphics, and also in understanding how PostScript handles coordinates in 2D.

In the equation $y = mx + b$ both m and b have a geometrical interpretation—m is the slope and b the y-intercept. What is the geometrical significance of A, B, and C in the equation $Ax + By + C = 0$?

Suppose $C = 0$. The equation is $Ax + By = 0$, which can be rewritten

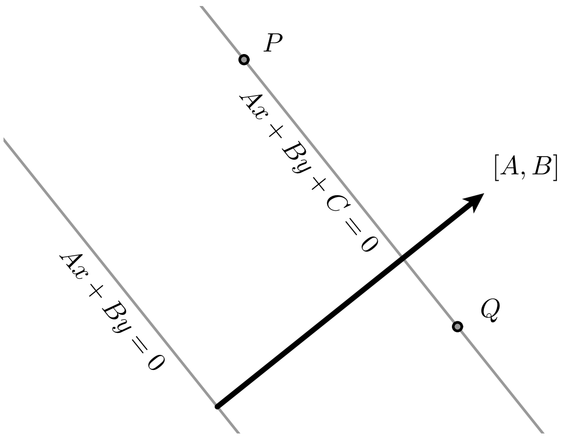
$$[A, B] \bullet [x, y] = 0 \ .$$

But this is the condition that $[x, y]$ be perpendicular to $[A, B]$. In other words, $[A, B]$ is the direction perpendicular to the line $Ax + By = 0$. In other words, the line $Ax + By = 0$ is the unique line that is (1) perpendicular to the vector $[A, B]$ and (2) passing through the origin.

Now look at the general case $Ax + By + C = 0$. If $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$ are two points on this line, then

$$\begin{aligned} Ax_P + By_P + C &= 0 \\ Ax_Q + By_Q + C &= 0 \\ A(x_Q - x_P) + B(y_Q - y_P) &= 0 \end{aligned}$$

which says that the vector $Q - P$ is perpendicular to $[A, B]$. In other words, we have the following picture:



- *The vector $[A, B]$ is perpendicular to the line $Ax + By + C = 0$.*

What is the meaning of C? If $C = 0$ the line passes through the origin. This makes plausible:

- *The quantity*

$$\frac{-C}{\sqrt{A^2 + B^2}}$$

is the signed distance of the line $Ax + By + C = 0$ from the line $Ax + By = 0$.

Exercise 10. *Explain why this is true. (Hint: use projections.)*

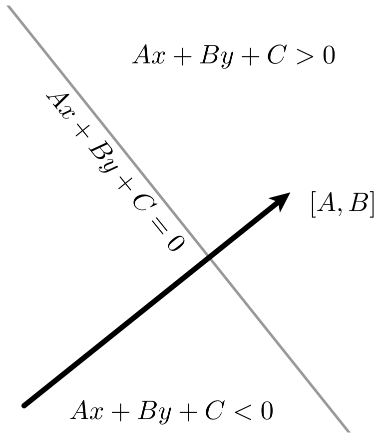
Exercise 11. *Given a line $Ax + By + C = 0$ and a point P, find a formula for the perpendicular projection of P onto the line.*

Exercise 12. *Given two lines $A_1x + B_1y + C_1 = 0$ and $A_2x + B_2y + C_2 = 0$, find a formula for the point of intersection.*

Exercise 13. *Given two points P and Q, find a formula for the line containing them.*

Exercise 14. *Given a line $Ax + By + C = 0$, find a formula for the line obtained by rotating it by $90°$ around the origin.*

The line $Ax + By + C = 0$ separates the plane into two halves, one where $Ax + By + C > 0$ and the other where it is negative. Which side is which?



As we cross the line in the direction of $[A, B]$ *the values of* $Ax + By + C$ *change from negative to positive.* This is easy to see indirectly. If $(x, y) = (tA, tB)$ then $Ax + By + C = t(A^2 + B^2) + C$ and for $t \gg 0$ will definitely be positive.

The function $f(x, y) = Ax + By + C$ is called an **affine function**. One useful property of affine functions is this:

- *If* P *and* Q *are two points in the plane,* t *a real number, and* f *an affine function then*

$$f\big((1-t)P + tQ\big) = (1-t)f(P) + tf(Q) \ .$$

Recall that $(1 - t)P + tQ$ is the weighted average of P and Q. As t varies over all real numbers this expression produces all points on the line through P and Q. Proving this property is an easy calculation.

Exercise 15. *An affine function* $f(x) = Ax + By + C$ *is equal to* -4 *at* $(0, 0)$ *and* 7 *at* $(1, 2)$. *Where on the line segment between these two points is* $f(x) = 0$*?*

9. Code

The file `eves-animation.eps` is a page-turning animation of Eves' proof of Pythagoras' Theorem.

References

1. Euclid, **The Elements**, translated by T. L. Heath. This is available in a commonly found Dover reprint. The part due to Euclid, but not Heath's very valuable comments, is also available on the 'Net at

`http://aleph0.clarku.edu/ djoyce/java/elements/elements.html`

The commentary used to be found at

`http://www.perseus.tufts.edu/`

but at the moment I write this (December, 2003) all links to Heath's comments, other than the initial chapters, are unfortunately dead.

2. Elisha S. Loomis, **The Pythagorean proposition: its demonstrations analyzed and classified, and bibliography of sources for data of the four kinds of proofs**, National Council of Teachers, Washington, 1968. Not wildly exciting. Not as much variety as you might hope for, either. But still curious.

3. H. Eves, **In mathematical circles**, 1969. The idea for the sequence of pictures for Pythagoras' Theorem is taken from p. 75.