Mathematics 308—Fall 1996

Coordinate systems

The aim of this chapter is to understand how PostScript deals with coordinate systems. The user does his drawing in **user coordinates** and PostScript renders them in **page coordinates**. It is important to understand the transformation from one to the other. This part of PostScript is essentially mathematical, and rather elegant.

1. Summary of how to draw in PostScript

When you start a path, you put the command **`newpath`** into your program. Then you set a current point with the **`moveto`** command and put in a number of commands adding pieces to the path you want to draw, such as lines or arcs of circles. Finally, you actually place the path on the current page with one of the three commands **`stroke`**, **`fill`** or **`clip`**.

There are several different parameters involved in how the path is drawn. The most important are **colour** (for us, usually just a shade of grey), the **line details**, the **clipping path** and the **transformation** from the user coordinates to page coordinates.

We already know about most of these. A few extra remarks:

Colour. You always start off with colour black, which is 0 in a scale from 0 to 1. If you are a Canadian nationalist you can put a line

```
/setgrey setgray def
```

and spell things the way you like. If you want to put the current shade on the stack you write

```
currentgray
```

so that if the current colour is either black or white the sequence

```
1 currentgray sub setgray
```

toggles it to the opposite shade.

Clipping. The clipping path sets a path outside which nothing is actually drawn. It gives you a kind of window through which you will see your drawings. It is built like any other path, but replaces **`stroke`** or **`fill`** by **`clip`**. These are the only commands by which paths become part of your picture.

Line details. There are a number of other things you can specify—the width of lines, dash pattern, the way paths are closed up, the way corners are rounded, etc.

Coordinates. In order to render pictures by programming commands, PostScript has to use a coordinate system of some kind. Of course at some point it has to keep track of the actual physical dimensions on the page. These are referred to in **page coordinates**. But it would be a great nuisance if you yourself had to do all the computations in these coordinates, as you would have had to in early printer control languages. Instead, PostScript has as part of its environment a set of data which relate **user coordinates** to page coordinates. This way, you will rarely know exactly what points on the page are involved in your drawing because PostScript will handle the calculations for you internally.

When you start up, you will in fact be using page coordinates. One unit of drawing is equal to $1/72$ of an inch on the page. This is for historical reasons—this unit is almost the same as a printer's point, the unit used since early in the history of printing to specify the size of letters on a printed page. Also, the origin of your coordinate system gets mapped onto the lower left of the page, and the x and y axes are perpendicular, going along the sides of the page.

You can change these by using a number of commands, of varying complexity. The simplest are (1) **`scale`**, which changes the units in the x and y axes; (2) **`translate`**, which moves the origin; and (3) **`rotate`**, which rotates the axes. You can combine these in any order to get more complicated changes, and you can also change the way in which your paths are rendered at a lower level.

Some mathematics is necessary in order to understand how this works. We must first look at how a number of geometrical features are specified in terms of coordinates.

2. Shears

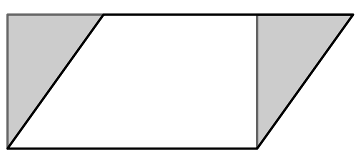
A **shear** is a transformation of a $2D$ figure that has this effect:

It is a bit hard to describe in plain language. Its effect can perhaps be best realized by thinking of the rectangle as a side view of a deck of thick cards:

In other words it slides the components of a figure past each other, and it slides things further if they are higher. From this picture it should be at least intuitively clear that

- *Shears preserve area.*

Roughly speaking this is because sliding a very thin piece of a figure doesn't change its shape. The actual proof that a shear doesn't change area is also very simple:



The idea is that we lop off a triangle from one end and shift it around to the other in order to make a parallelogram into a rectangle. The reason this works is because we can shift that triangle without distorting it. This reasoning also shows that area = base × height. Of course we have to appeal to some more fundamental result to justify this argument. A rigourous proof can be put together by discussing angles cut off by parallel lines. Ultimately it derives from Euclid's parallel postulate, but I won't discuss this further.

There are different directions of shearing possible. The one above is called a shear along the x axis. And there are degrees of shear possible. We can specify a shear along the x axis completely with a single number which tells us how much the unit square is transformed. Its base remains fixed, and the top moves parallel to the x axis. The corner $(0,1)$ is moved to some point $(a,1)$. *Where is the point (x,y) moved to?* The y coordinate remains the same, and the horizontal distance moved (1) depends only the y coordinate; (2) is proportional to the

y coordinate. Therefore (x, y) is shifted by $(ay, 0)$ and gets moved to $(x + ay, y)$. The effect can also be expressed by matrix multiplication:

$$\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} .$$

Therefore

- *Shearing along the x axis is a linear transformation whose matrix is of the form*

$$\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$$

3. Matrices and linear transformations

I recall that a linear transformation is a transformation of points in the plane which takes a point (x, y) to a point (x_*, y_*) whose coordinates are homogeneous linear functions of x and y. This means that

$$\begin{aligned} x_* &= ax + by \\ y_* &= cx + dy \end{aligned}$$

for suitable coefficients a, b, c, d. Equivalently

$$\begin{bmatrix} x_* \\ y_* \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} .$$

Let

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} .$$

Another way of getting the matrix from the transformation is to keep in mind that

- *If T is a linear transformation and A the matrix associated to it then the columns of A are the vectors which T assigns to e_1 and e_2.*

This is a simple calculation. Geometrically, this says that we can reconstruct the matrix of T if we know what T does to the unit square $0 \le x \le 1, 0 \le y \le 1$.

4. Length

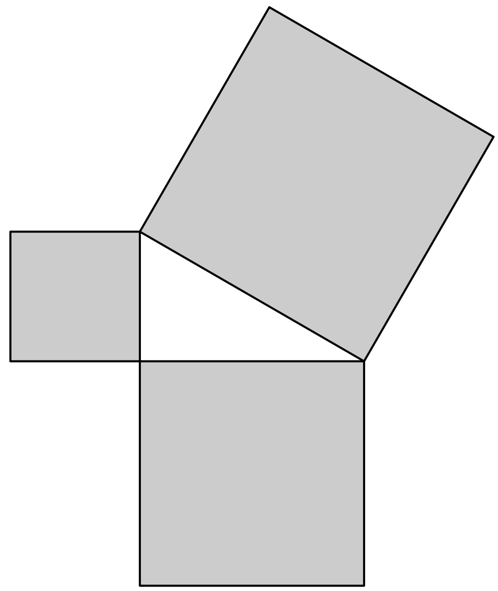
The most important mathematical result we need is that which says that in a coordinate system where x and y are measured uniformly and in which the x and y axes are perpendicular to each other, the distance from the origin to (x, y) is $\sqrt{x^2 + y^2}$. This is just Pythagoras' Theorem. In order to reinforce how important it is, and because readers may enjoy working their way through proofs of it, I shall include a discussion of why it is true. Of course, there are lots of different proofs of this result which been discovered in the more than 3000 years since the formula was discovered. We shall see two here and a few more in the exercises.

We begin with the statement:

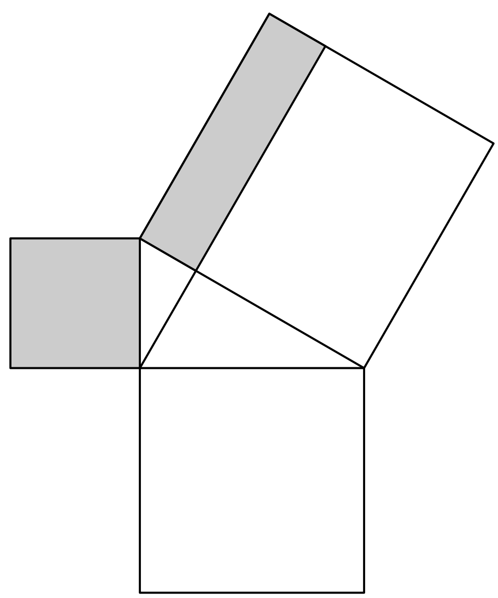
- *For a right triangle with short sides a and b and long side c we have $c^2 = a^2 + b^2$.*

Proof by dissection

How satisfying a proof seems is to a large extent a matter of taste. The proof which I explain first is the one I prefer to all others. It interprets the assertion geometrically in the most direct way possible. We can make Pythagoras' assertion into a geometrical statement by constructing squares on the sides of the original triangle. The area of the large square is then to be the sum of the areas of the smaller squares.

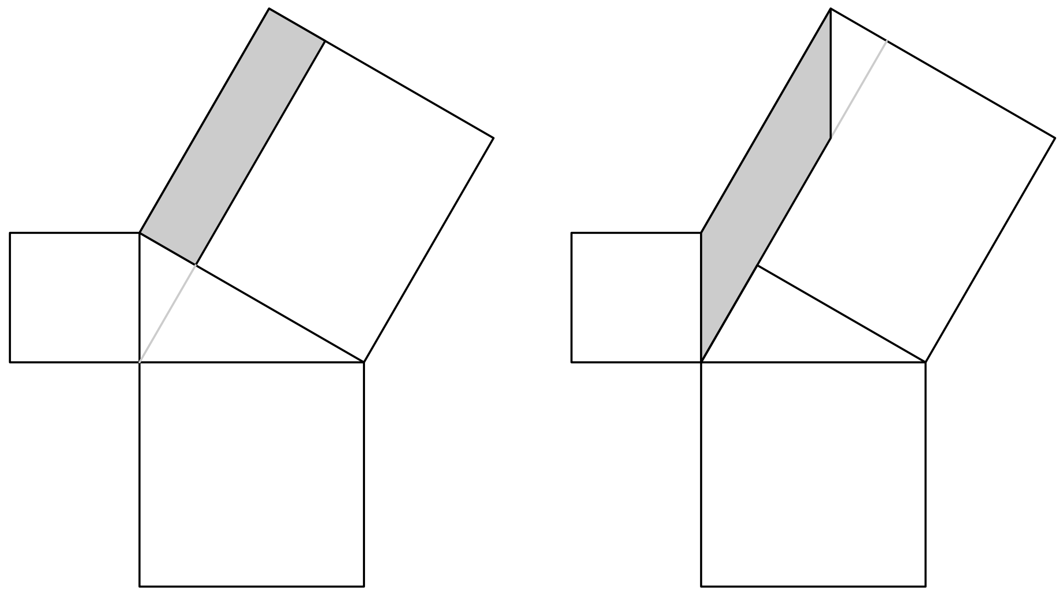


We can make this assertion more precise. Draw a straight line from the corner of the triangle where the right angle occurs across the large square to its opposite side, where it meets that side at a right angle:

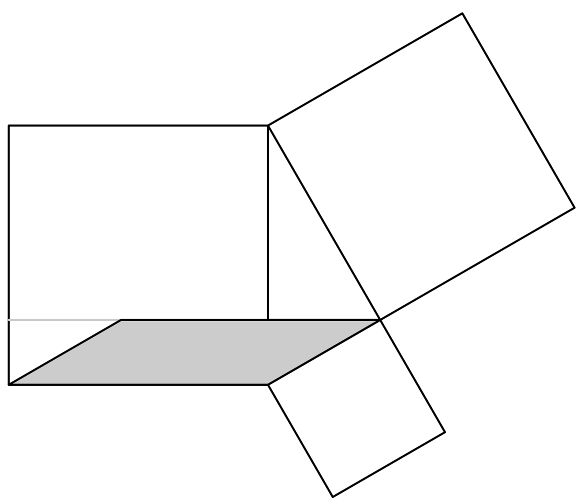


This line divides the large square into two rectangles, and a very explicit version of Pythagoras' Theorem is that the area of each of these is equal to the area of one of the squares. In other words, Pythagoras' Theorem is proven by **dissecting** the square of area c^2 into two smaller pieces of area a^2 and b^2.

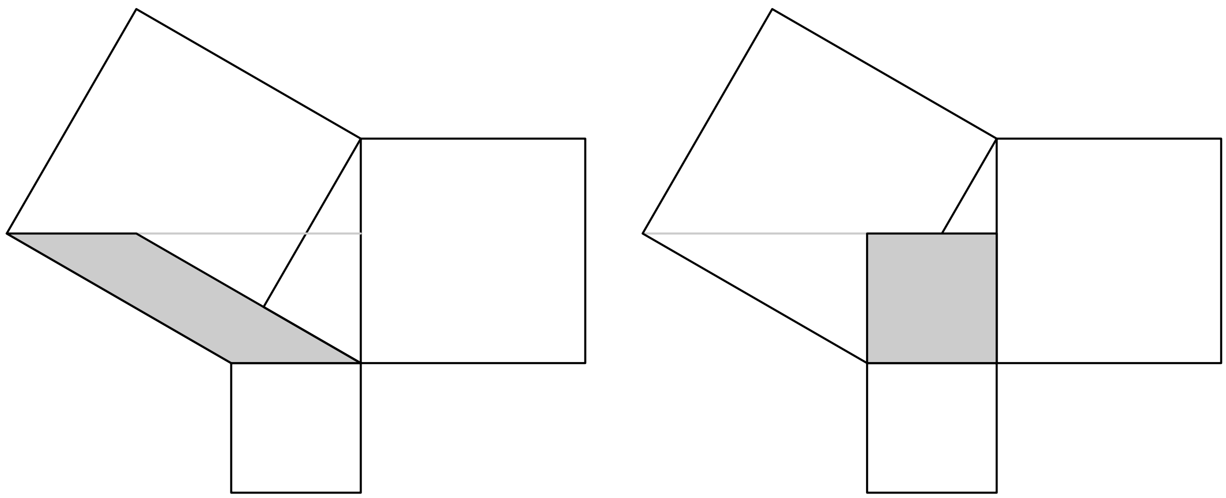
This claim is proven by applying a succession of transformations to a smaller square to turn it into its corresponding rectangle, while preserving its area. In practice we'll go backwards. First we shear the rectangle one way:



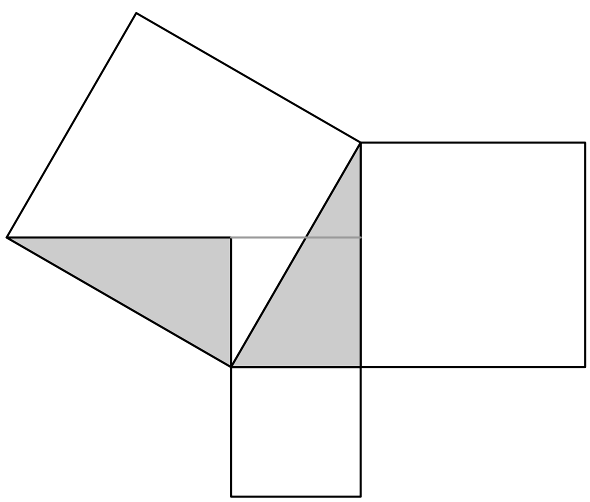
We know that area is preserved. To understand this step better, we can follow it from a different angle:



Then we shear the figure again, rotating the whole picture once more to let us visualize it better:



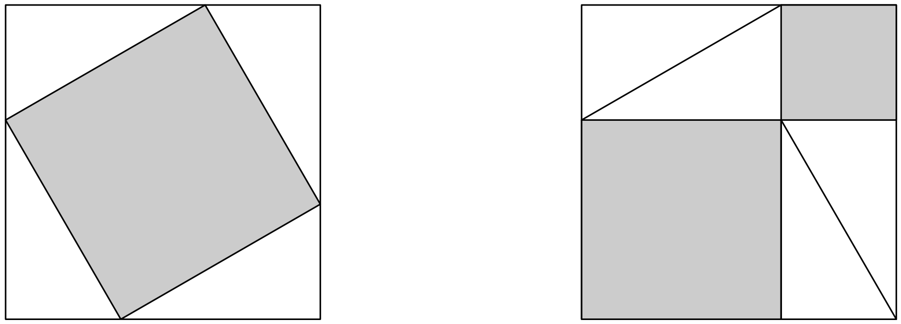
We have sheared the rectangle into a figure which is definitely a rectangle, and which in fact appears to be a copy of the small square. That it is one can be seen by using a rotation:



Of course the same sequence of transformations can be carried out for the other small square.

An algebraic proof

Consider the figure on the left below:

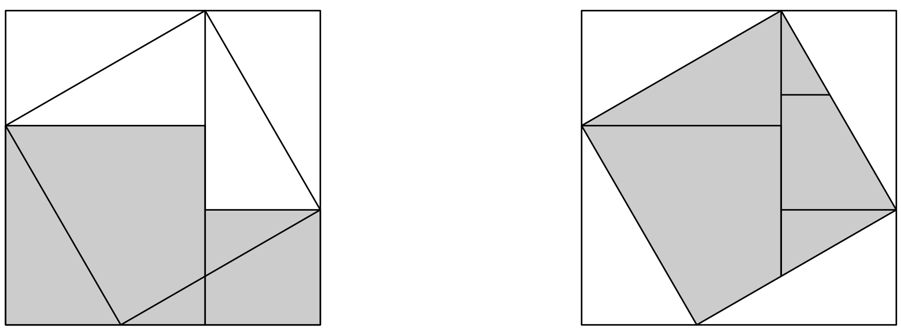


The area of the whole figure is $(a+b)^2$, which by algebra is equal to a^2+b^2+2ab. But it is also

$$c^2 + 4 \cdot \frac{ab}{2} = c^2 + 2ab \ .$$

Therefore $c^2 = a^2 + b^2$. This proof is apparently the oldest known.

This can also be rewritten as a dissection proof. We shift things around a bit, and then lay one of the diagrams over the other. Start with the figures above and continue below.



5. Rotations

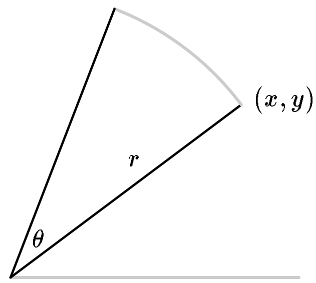
Suppose we rotate the point in the plane with coordinates (x, y) through an angle of θ. What are the coordinates of the point we then get? The answer is

$$(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta) \ .$$

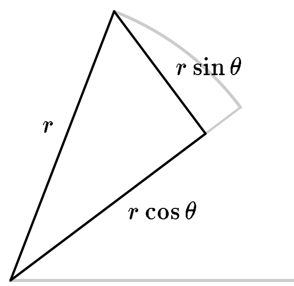
This is a formula used repeatedly, in various guises and in many different circumstances throughout these notes. I will offer three proofs of it.

The direct argument

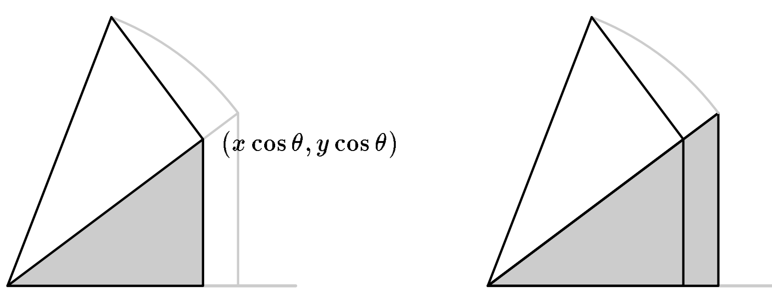
First I will prove it directly from a geometrical argument.



The first step is to drop a perpendicular from the rotated point onto the radius vector of (x, y).

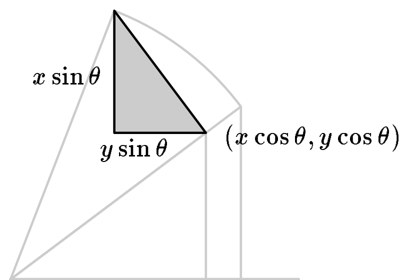


What are the coordinates of the bottom of the perpendicular? Because we have rotated through an angle of θ and rotation preserves distances from the origin, the distance from the origin to the bottom of the perpendicular is $r \cos \theta$, where $r = \sqrt{x^2 + y^2}$ is the length of the original vector (x, y). The length of the perpendicular itself is $r \sin \theta$.



Since the triangle on the left is obtained from the one on the right by a simple scaling operation they are similar. The ratio of the long sides is $r \cos \theta : r$, so the coordinates of the bottom of the perpendicular are $(x \cos \theta, y \cos \theta)$.

We now add a triangle to the picture:



It also is similar to the one of the triangles in the previous figure (the angle at its lower right is obtained by a simple rotation from one of the angles in the smaller of those two), and since its long side is $r\sin\theta$ its bottom has length $x\sin\theta$ and the left side length $y\sin\theta$.

But this tells us immediately that the x-coordinate of the rotated point is $x\cos\theta - y\sin\theta$, and its y-coordinate is $y\cos\theta + x\sin\theta$.

If we take (x, y) to be the point $(\cos\varphi, \sin\varphi)$ we get by rotating $(1, 0)$ through an angle of φ, then on the one hand we get the vector $(\cos(\varphi+\theta), \sin(\varphi+\theta))$ that we would get by rotating $(1, 0)$ through an angle of $\varphi+\theta$, and on the other the formula we have just proven gives a different expression. Therefore

$$(\cos(\varphi+\theta), \sin(\varphi+\theta)) = (\cos\varphi\cos\theta - \sin\varphi\sin\theta, \cos\varphi\sin\theta + \sin\varphi\cos\theta)$$

In fact, the rotation formula is equivalent to the pair of trigonometrical formulas

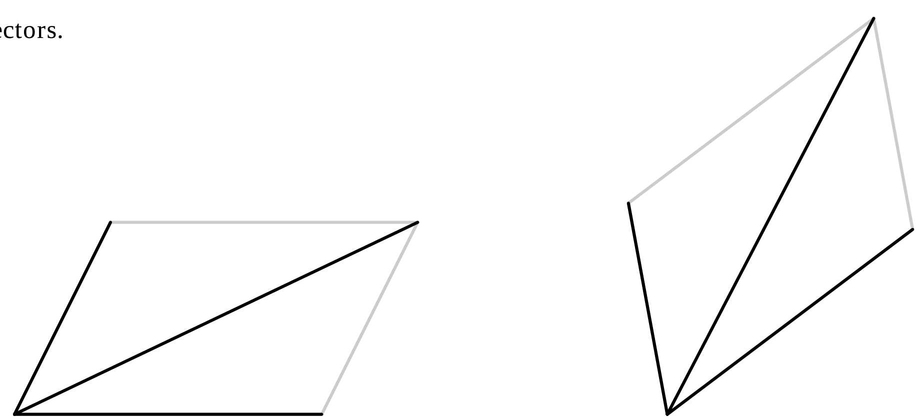
$$\begin{aligned}\cos(\varphi+\theta) &= \cos\varphi\cos\theta - \sin\varphi\sin\theta\\ \sin(\varphi+\theta) &= \sin\varphi\cos\theta + \cos\varphi\sin\theta\end{aligned}$$

Using linearity

Rotation is a **linear** transformation, which in geometrical terms means that if we rotate a parallelogram we obtain a parallelogram. In algebraic terms it means that

$$R(u+v) = R(u) + R(v)$$

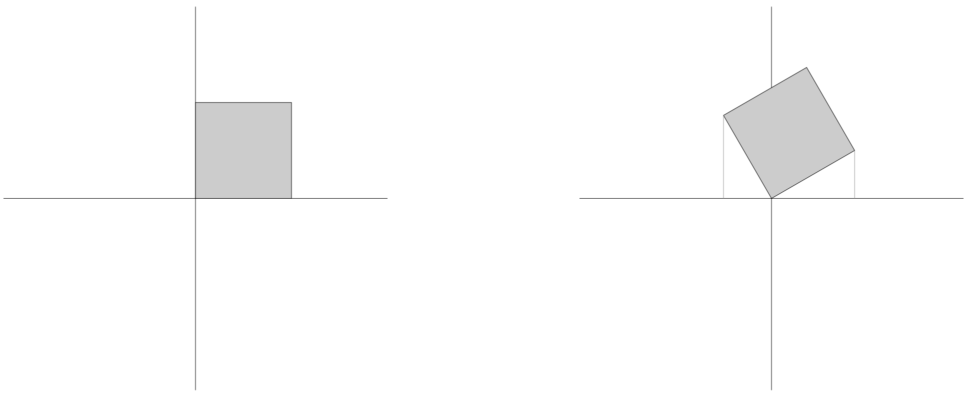
if u and v are vectors.



Since the vector with its head at (x, y) is the vector sum of $(x, 0)$ and $(0, y)$, linearity implies that the rotation of (x, y) is the vector sum of the rotations of $(x, 0)$ and $(0, y)$.

$$
\begin{aligned}
R(x,0) &= (\quad x\cos\theta, x\sin\theta) \\
R(0,y) &= (-y\sin\theta, y\cos\theta) \\
R(x,y) &= (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)
\end{aligned}
$$

Once we realize that rotation is linear, we can read off its effect from its matrix. We can calculate its matrix from its effect on the unit square. This can be easily seen in this picture:



Euler's formula

This one involves on the face of it no geometry at all.

Values of the exponential function can be calculated by the formula

$$e^z = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \frac{z^4}{24} + \cdots + \frac{z^n}{n!} + \cdots$$

The series can be used to check how multiplication and exponents are realted:

$$
\begin{aligned}
e^{w+z} &= \left(1 + w + \frac{w^2}{2} + \frac{w^3}{6} + \cdots\right)\left(1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \cdots\right) \\
&= 1 + (w+z) + \frac{(w^2 + 2wz + z^2)}{2} + \frac{(w^3 + 3w^2z + 3wz^2 + z^3)}{6} + \cdots \\
&= e^{w+z}
\end{aligned}
$$

This series makes sense even if z is a complex number, and we therefore useit to define e^z in this case. If we set $z = ix$ where $i = \sqrt{-1}$ then we get

$$e^{ix} = 1 + ix - \frac{x^2}{2} - \frac{ix^3}{6} + \frac{x^4}{24} + \cdots + \frac{i^n x^n}{n!} + \cdots$$

since

$$i^2 = -1, \quad i^3 = i \cdot i^2 = -i, \quad i^4 = 1, \quad \ldots$$

and therefore

$$
i^n = \begin{cases} 1 & n \text{ is divisible by } 4 \\ i & n = 4m+1 \\ -1 & n = 4m+2 \\ -i & n = 4m+3 \end{cases}
$$

The series for e^{ix} can be rewritten as

$$\begin{aligned} e^{ix} &= 1 + ix - \frac{x^2}{2} - \frac{ix^3}{6} + \frac{x^4}{24} + \cdots + \frac{i^n x^n}{n!} + \cdots \\ &= \left(1 - \frac{x^2}{2} + \frac{x^4}{24} - \cdots\right) + i\left(x - \frac{x^3}{6} + \frac{x^5}{120} - \cdots\right) \end{aligned}$$

The series for the real and imaginary parts are the Taylor series for $\cos x$ and $\sin x$. We therefore get

$$e^{ix} = \cos x + i \sin x$$

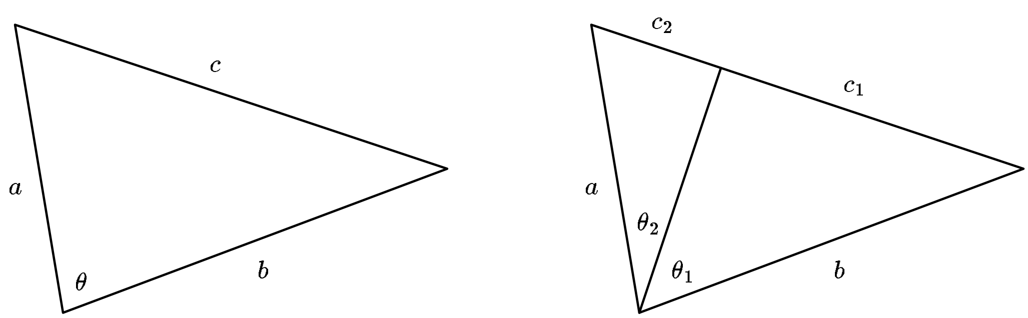
The addition formula remains valid for complex numbers, so

$$\begin{aligned} e^{i(x+y)} &= \cos(x+y) + i\sin(x+y) \\ &= (\cos x + i \sin x)(\cos y + i \sin y) \\ &= (\cos x \cos y - \sin x \sin y) + i(\sin x \cos y + \cos x \sin y) \end{aligned}$$

and we see again the trigonometrical formulas for the sum of angles.

6. Angles

The cosine sum formula has as an immediate consequence the cosine rule for triangles, which is a generalization of Pythagoras' Theorem.



Let the side opposite the origin have length c. By 'dropping' a perpendicular from the origin onto this side we decompose it into two pieces of length, say, c_1 and c_2. Thus

$$c^2 = (c_1 + c_2)^2 = c_1^2 + c_2^2 + 2c_1c_2$$

On the other hand the original angle θ is decomposed into two parts θ_1, θ_2. We know that

$$\cos\theta = \cos\theta_1 \cos\theta_2 - \sin\theta_1 \sin\theta_2$$

Finally, let y be the length of the perpendicular. By Pythagoras' Theorem applied to each of the small triangles and trigonometry

$$\begin{aligned} c_1^2 &= a^2 - y^2 \\ c_2^2 &= b^2 - y^2 \\ c_1 &= a \sin\theta_1 \\ y &= a\cos\theta_1 \\ c_2 &= a\sin\theta_2 \\ y &= a\cos\theta_2 \end{aligned}$$

so that

$$\begin{aligned} c^2 &= (a^2 - y^2) + (b^2 - y^2) + 2ab \sin\theta_1 \sin\theta_2 \\ &= a^2 + b^2 - 2ab\ \cos\theta_1 \cos\theta_2 + 2ab \sin\theta_1 \sin\theta_2 \\ &= a^2 + b^2 - 2ab\cos\theta \end{aligned}$$

The cosine rule in turn relates angles to dot products.

- *For vectors u and v*

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

where θ is the angle between u and v.

If u and v are vectors, then they form two sides of a triangle. On the one hand, the square of the length of the third side is

$$\|u - v\|^2 = \|u - v\| \bullet \|u - v\| = \|u\|^2 + \|v\|^2 - 2u \bullet v$$

and on the other, by the cosine rule, it is

$$\|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\| \cos\theta$$

so by comparison

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

In particular:

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

7. Orthogonal matrices

An **orthogonal matrix** is one whose columns are of unit length and perpendicular to each other. Equivalently, K is orthogonal if

$${}^tK\,K = I = K\,{}^tK$$

or in am equivalent formulation:

- s A matrix K is orthogonal when its inverse K^{-1} isequal to its transpose ${}^tK$.

Because rows of ${}^tK$ are the columns of K, and the entry $x_{i,j}$ in ${}^tK\,K$ is therefore the dot product of the i-th column with the j-th column.

The picture or a calculation shows that rotation matrices are orthogonal. Calculation shows it has determinant 1.

The converse is also true.

- *A matrix is a rotation matrix if and only if it is an orthogonal matrix with positive determinant.*

Suppose K to be an orthogonal matrix. This means that

$${}^tK\,K = I$$

which means no more and no less than that the columns of K have unit length and are perpendicular to each other. If v_1 is the first column, we can write it as $(\cos\theta, \sin\theta)$ for some θ, which is possible for any vector of unit length. The second column v_2 must be v_1 rotated $\pm 90^\circ$. But the condition on the determinant implies that the sign must be positive.

8. Eigenvalues and eigenvalues

We need to recall some simple facts about solving 2×2 systems of equations.

If A is a 2×2 matrix its **transpose adjoint** ${}^tA^*$ is defined by the prescription

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}, \quad {}^tA^* = \begin{bmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{bmatrix} .$$

A simple calculation shows that

$$A\,{}^tA^* = \begin{bmatrix} \det(A) & 0 \\ 0 & \det(A) \end{bmatrix}$$

so that if $\det(A) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1} \neq 0$ then

$$A^{-1} = \frac{1}{\det(A)}\,{}^tA^* = \begin{bmatrix} a_{2,2}/\det & -a_{1,2}/\det \\ -a_{2,1}/\det & a_{2,2}/\det \end{bmatrix} .$$

In this case if we are given an equation $Au = 0$ we can multiply on both sides by A^{-1} to get $u = A^{-1}0 = 0$.

If $\det(A) = 0$ then both of the columns of ${}^tA^*$ are vectors u such that

$$Au = 0$$

If A is not itself the 0 matrix at least one of these columns is not the 0 vector, and otherwise A annihilaites all column vectors.

- *If A is a 2×2 matrix then exactly one of these possibilities occurs:*
  (1) the matrix A is invertible, $\det(A) \neq 0$, and the only vector annihilated by A is the 0 vector itself;
  (2) the matrix A is singular, $\det(A) = 0$, and there exists a non-zero vector u such that $Au = 0$.

An **eigenvector** of a linear transformation T is a vector $v \neq 0$ taken into a multiple of itself by T:

$$Tv = \lambda v$$

for some scalar λ. The scalar is called the associated **eigenvalue**. If M is the matrix of T then to find an eigenvector v we must solve

$$M \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}, \quad v = \begin{bmatrix} x \\ y \end{bmatrix} .$$

This can be rewritten as

$$(M - \lambda I)v = 0$$

where I is the identity matrix

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

Since by definition an eigenvector is not 0, we must therefore have

$$\det(M - \lambda I) = 0 .$$

If

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then

$$M - \lambda I = \begin{bmatrix} a - \lambda & b \\ c & d - \lambda \end{bmatrix}$$

and λ must be a root of the **characteristic polynomial**

$$\lambda^2 - (a + d)\lambda + (ad - bc) = 0 .$$

The system of equations to determine the coordinates of an eigenvector is then

$$\begin{aligned}(a-\lambda)x + by &= 0\\ cx + (d-\lambda)y &= 0\ .\end{aligned}$$

Since the coefficient matrix is singular, one of these equations is a multiple of the other, and we need to consider only one of them. Most of the time we can set

$$x = b, \quad y = (\lambda - a)\ .$$

Often we want eigenvectors of length 1, in which case we divide x and y by $r = \sqrt{x^2 + y^2}$.

Example. Let

$$M = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\ .$$

Its characteristic polynomial is

$$\lambda^2 - 3\lambda + 1 = 0$$

so its eigenvalues are $3/2 \pm \sqrt{5}/2 = \lambda_1, \lambda_2$; numerically $\lambda_1 = 2.618034$, $\lambda_2 = 0.381966$. To find the eigenvectors for λ we solve

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

and get the eigenvector equation

$$(1-\lambda)x + y = 0\ .$$

Explicitly, the eigenvector for λ_1 is

$$\begin{bmatrix} 0.525731 \\ 0.850651 \end{bmatrix}$$

If M is a matrix and v_1, v_2 are eigenvectors for M with eigenvalues λ_1, λ_2 then

$$Mv_1 = \lambda_1 v_1, \quad Mv_2 = \lambda_2 v_2\ .$$

If we make up a matrix V whose columns are the v_i then

$$MV = M\,[\,v_1 \quad v_2\,] = [\,Mv_1 \quad Mv_2\,] = [\,\lambda_1 v_1 \quad \lambda_2 v_2\,] = [\,v_1 \quad v_2\,] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = VD$$

and

$$M = VDV^{-1}\ .$$

In this course, with perhaps a few exceptions, the matrix M will be symmetric. If

$$M = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

then its eigenvalues are

$$\frac{a+c}{2} \pm \frac{\sqrt{(a-c)^2 + 4b^2}}{2}$$

and are both real. To avoid problems with rounding we calculate one root λ by this formula, the one where the sign of the square root is the same as the sign of $a + c$, and calculate the other as $\det/\lambda$.

9. Scaling

A **scale change along the x and y axes** is a transformation which multiplies all x coordinates by one factor and all y coordinates by another. Since it takes (x, y) to (ax, by). it has a matrix of the form

$$\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} .$$

There are other scale changes. The most general possible ones are those which change scale along some pair of orthogonal axes.

How can we find the matrix of a general scale change T? Let v_1 and v_2 be orthogonal vectors of unit length and positive orientation along the axes of the scale change. Then

$$R = [\, v_1 \quad v_2 \,] = \begin{bmatrix} v_{1,x} & v_{2,x} \\ v_{1,y} & v_{2,y} \end{bmatrix}$$

is a rotation matrix, say for angle θ. Suppose T scales along the line through v_1 by a and along that through v_2 by b. To find what the matrix of T is we must find what it does to the vectors e_1 and e_2? Then by definition of coordinates we have

$$\begin{aligned} v_1 &= (\cos\theta)\, e_1 + (\sin\theta)\, e_2 \\ v_2 &= -(\sin\theta)\, e_1 + (\cos\theta)\, e_2 \end{aligned}$$

or

$$\begin{aligned} e_1 &= (\cos\theta)\, v_1 - (\sin\theta)\, v_2 \\ e_2 &= (\sin\theta)\, v_1 + (\cos\theta)\, v_2 \end{aligned}$$

Therefore

$$\begin{aligned} Te_1 &= a(\cos\theta)\, v_1 - b(\sin\theta)\, v_2 \\ Te_2 &= a(\sin\theta)\, v_1 + b(\cos\theta)\, v_2 \end{aligned}$$

We now substitute for v_1 and v_2 in terms of e_1 and e_2. The final result is very simple:

- *If T is a scale change whose axes are the columns of the rotation matrix R with scale factors a and b then the matrix of T is*

$$M_T = RAR^{-1}$$

  where

$$A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} .$$

The matrix M_T is symmetric since

$$\begin{aligned} M_T &= R\,A\,R^{-1} \\ &= R\,A\,{}^tR \\ {}^tM_T &= R\,{}^tA\,{}^tR \\ &= R\,A\,{}^tR \\ &= M_T\ . \end{aligned}$$

Conversely, if M is a symmetric matrix it has real eigenvalues and we can find orthogonal eigenvectors. The linear transformation associated to M will be a scale change along the axes determined by its eigenvectors. Therefore the matrices of scale changing transformations coincide exactly with the symmetric matrices.

Some of them will scale by positive factors. The ones we shall meet will generally arise in the same way.

- *If M is any matrix then the matrix ${}^tMM$ is symmetric and has non-negative eigenvalues.*

The proof hinges on a fundamental equation involving dot products and matrix transposes. If u and v are column vectors of dimension 2 then their dot product can be expressed also as a matrix product:

$$u \bullet v = u_1 v_1 + u_2 v_2 = [\, u_1 \quad u_2 \,] \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = {}^t u \, v \ .$$

If u is an eigenvector of ${}^tM\,M$ with eigenvalue c then ${}^tM\,Mu = cu$. But then

$${}^tM\,Mu \bullet u = cu \bullet u = c\,\|u\|^2$$

on the one hand and

$$Mu \bullet Mu = \|Mu\|^2$$

on the other. But if

$$\|Mu\|^2 = c\,\|u\|^2$$

then since u cannot be 0, c must be non-negative.

10. Factoring linear transformations

The key result in this section is this:

- *Any linear transformation can be written as a composition $R_1 S R_2^{-1}$ where R_1 and R_2 are rotations and S is a scale change along the x and y axes.*

I shall prove this by finding the matrices of all these transformations explicitly. It is a somewhat messy computation.

It depends on being able to find eigenvalues. There is one case which is straightforward, and that is where M is symmetric. In that case

$$M = RSR^{-1}$$

where S is a diagonal matrix with entries are the eigenvalues of M and the columns of R are its eigenvectors. The matrix R can be chosen to be a rotation matrix, in which case R^{-1} is also one.

Let M be an arbitrary invertible matrix. If we could write

$$M = R_1 S R_2^{-1}$$

where the R_i are rotations, then ${}^tS = S$ so

$${}^tM\,M = R_2 S\, {}^tR_1\, R_1 S R_2^{-1} \ .$$

Recall that ${}^t(AB) = {}^tB\,{}^tA$. Since ${}^tR_1\,R_1 = I$, we also have

$${}^tM\,M = R_2 S^2 R_2^{-1} \ .$$

This means that the columns of R_2 are the eigenvectors of ${}^tM\,M$ and that the entries of S^2 are the squares of its eigenvalues.

This suggests how to start with M and find the R_i and S. Let M be an arbitrary matrix. If we write

$${}^tM\,M = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix}$$

we see that the matrix ${}^tM\,M$ is symmetric. We know already that it has non-negative eigenvalues. Therefore we can write

$${}^tM\,M = R_2 D\, R_2^{-1}$$

where the columns of R_2 are the eigenvectors of ${}^tM\,M$. We know that the eigenvalues of M, the diagonal entries of D, are non-negative, so we can write $D = S^2$ with S diagonal. There is a choice of sign—if $\det M < 0$, choose one sign along the diagonal of A negative. This ensures $\det S = \det M$. We now want

$$M = R_1 S R_2^{-1}$$

so we set

$$R_1 = M R_2 S^{-1} \ .$$

I claim that R_1 is a rotation matrix.

We apply the earlier result. The determinant of R_2 is positive because determinants multiply and

$$\det R_2 = \det S \det R_1^{-1} \det M = \det S \det M$$

and because of the choice of signs of S. Furthermore

$$R_1^{-1} R_1 \quad = {}^tR_1\, R_1 = S^{-1} R_2^{-1}\, {}^tM\, M R_2 \quad S S^{-1} R_2^{-1} R_2 \quad S^2 R_2^{-1} R_2 \quad S = I \ .$$

Note that all these steps can be carried out explicitly, if painfully.

Summary. Let

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

(1) Calculate

$${}^tM\, M = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix} = \begin{bmatrix} A & B \\ B & C \end{bmatrix}, \quad D = \det({}^tM\, M) = AC - B^2 \ .$$

(2) Let λ be one of the eigenvalues

$$\lambda = \frac{A + C}{2} + \frac{\sqrt{(A+C)^2 - 4D}}{2} \ .$$

(3) Let

$$x = B, \quad y = \lambda - A, \quad r = \sqrt{x^2 + y^2}, \quad u = \begin{bmatrix} x/r \\ y/r \end{bmatrix}$$

so that u is a normalized eigenvector of ${}^tM\, M$, and let

$$u = \begin{bmatrix} -y/r \\ x/r \end{bmatrix}$$

be v rotated by 90°.

(4) Let

$$R_2 = [\, u \quad v \,], \quad S = \begin{bmatrix} \sqrt{\lambda} & 0 \\ 0 & \det(M)/\sqrt{\lambda} \end{bmatrix} \ .$$

Then calculate

$$R_1 = M R_2 S^{-1} \ .$$

We have $M = R_1 S R_2^{-1}$.

Example. Suppose we are given the matrix

$$M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \ .$$

Then

$$ {}^{t}M\,M = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \; . $$

Its characteristic polynomial is

$$ \lambda^2 - 3\lambda + 1 = 0 $$

so its eigenvalues are $3/2 \pm \sqrt{5}/2 = \lambda_1, \lambda_2$; numerically $\lambda_1 = 2.618034$, $\lambda_2 = 0.381966$. The eigenvector for λ is

$$ \begin{bmatrix} 0.525731 \\ 0.850651 \end{bmatrix} $$

which makes

$$ R_2 = \begin{bmatrix} 0.525731 & -0.850651 \\ 0.850651 & 0.525731 \end{bmatrix} $$

and

$$ S = \begin{bmatrix} 1.618034 & 0.000000 \\ 0.000000 & 0.618034 \end{bmatrix} $$

Finally

$$ R_1 = M R_2 S^{-1} = \begin{bmatrix} 0.850651 & -0.525731 \\ 0.525731 & 0.850651 \end{bmatrix} $$

A consequence of this factorization:

- *A linear transformation changes areas by a factor equal to its determinant. It preserves orientations if and only if its determinant is positive.*

This is because any linear transformation can be factored as above, and determinants and volume change factors multiply under composition.

11. Affine transformations

Suppose given two planes, each with rectangular coordinates. If A is a 2×2 matrix and (τ_x, τ_y) a vector in the second plane, then the map taking

$$ \begin{bmatrix} x \\ y \end{bmatrix} \mapsto A \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix} $$

is called an **affine transformation**. In short, an affine transformation is a linear transformation followed by a vector translation.

Explicitly, if

$$ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} $$

then this map takes

$$ \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix} + \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix} = \begin{bmatrix} ax + by + \tau_x \\ cx + dy + \tau_y \end{bmatrix} \; . $$

- *If A is an invertible matrix, then this transformation is invertible.*

This is because we can solve

$$ \begin{aligned} u &= ax + by + \tau_x \\ v &= cx + dy + \tau_y \end{aligned} $$

or

$$ \begin{aligned} ax + by &= u - \tau_x \\ cx + dy &= v - \tau_y \end{aligned} $$

for (x, y) if we are given (u, v). In terms of matrices if

$$\begin{bmatrix} u \\ v \end{bmatrix} = A \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix}$$

then

$$\begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} u \\ v \end{bmatrix} - A^{-1} \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix}$$

so that the inverse is also an affine transformation.

- *An affine transformation takes lines to lines.*

Let ℓ be a parametrized line of the form

$$t \mapsto \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + t \begin{bmatrix} p \\ q \end{bmatrix} = \begin{bmatrix} x_0 + tp \\ y_0 + tq \end{bmatrix}$$

that is to say going through (x_0, y_0) in the direction (p, q). Then $T\ell$ goes through

$$A \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \begin{bmatrix} \tau_x \\ \tau_y \end{bmatrix}$$

in the direction

$$A \begin{bmatrix} p \\ q \end{bmatrix} .$$

Affine transformations can be characterized in completely geometric terms as maps from one plane to another which take lines to lines. But this is a subtle fact. I explain it an appendix to this section.

12. The transformation matrix in PostScript

Implicit in PostScript drawing are two coordinates, user coordinates and page coordinates. When the programmer draws a line from one point to the other he specifies their coordinates in the user system, and then PostScript renders them into positions on the page. The transformation it applies to go from user coordinates to page coordinates has, in general, no other property than that it takes straight lines to striaght lines. It can be an arbitrary affine transformation.

It is represented in PostScript by an array of six elements

```
[a b c d tx ty]
```

representing the matrix A and translation vector τ. It is called the **Current Transform Matrix** or **CTM** for short. The conventions of interpreting these coefficients are a little different from what we are used to, however, because

- *In PostScript all vectors are row vectors and matrices are applied on the right.*

This is a common problem in computer graphics, where the community seems about evenly divied between row and column interpretation of vectors. Thus the effect of the CTM on coordinates (x, y) is to change them to

$$[\,x \quad y\,] \begin{bmatrix} a & b \\ c & d \end{bmatrix} + [\,\tau_x \quad \tau_y\,]$$

This is something to be careful about, but shouldn't cause serios difficulties. In PostScript it is justifiable because after all it does calculations backwards anyway.

The CTM is modified by the commands **`rotate`**, **`translate`**, and **`scale`** as well as others.

(•) The effect of the sequence

sx sy translate

is to change the CTM to **[a b c d tx+sx ty+sy]**.

(•) The effect of

sx sy scale

is to change the CTM by multiplying its linear component A (the 2×2 matrix) by a diagonal matrix:

$$A \mapsto \begin{bmatrix} sx & 0 \\ 0 & sy \end{bmatrix} A$$

(•) The effect of **rotate** is

$$A \mapsto \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} A$$

To understand this, keep in mind that the matrix of a rotation by θ is

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

If this is different from what you are used to it is because here matrices act on the right.

Any affine transformation can be represented as a linear transformation in three dimensions. We do this by identifying the (x, y) plane with the plane $z = 1$ in three dimensions. The matrix

$$\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ \tau_x & \tau_y & 1 \end{bmatrix}$$

takes this plane to itself, and has the same effect on x and y coordinates as the corresponding affine transformation does. To see this, calculate

$$\begin{bmatrix} x & y & 1 \end{bmatrix} \begin{bmatrix} a & b & 0 \\ c & d & 0 \\ \tau_x & \tau_y & 1 \end{bmatrix}$$

This representation has the virtue that the composition of transformations corresponds to the multiplication of matrices. There is a PostScript command which does exactly that for you.

(•) The sequence

[a b c d tx ty] concat

replaces the CTM by the matrix product

$$\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ \tau_x & \tau_y & 1 \end{bmatrix} \mathrm{CTM}$$

if we think of the CTM as a 3×3 matrix.

It is also possible to set the CTM directly instead of just modifying it, but this is such a terrible idea that I won't tell you how to do it.

In general, it is not a good idea to use **concat** as opposed to applying a succession of translations, scale changes, and rotations. There is no loss of flexibility. It is an easy consequence of the factorization of linear transformations that one can also factor affine transformations:

- *Any affine transformation can be represented as a composition of rotations, scale changes, and translations.*

Still, it is at least conceivable that you might want to apply a shear to your picture, in which case **concat** is what you should use.

13. Ellipses

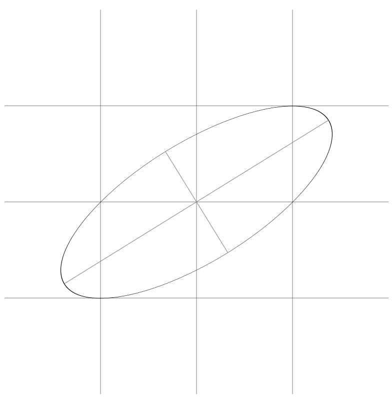
An ellipse is what you get when you scale a unit circle in perpendicular directions, along axes through its centre. Accordng to this definition, a circle is to be considered as a special kind of ellipse.

The data specifying an ellipse comprise the original circle, or equivalently the centre of the circle, together with the scaling factors and axes. Unless the ellipse is a circle, these data are uniquely determined by the ellipse. If we start with the unit circle and scale it, rotate it, and then translate it, we can obtain any given ellipse. Since any affine transformation can be obtained a composition of such operations, its effect on any circle will be to produce an ellipse. This is not quite obvious—for example, if you shear the unit circle along the x-axis it is true that you will get an ellipse, but the axes of this ellipse are not at all easy to calculate, and in fact it is not at all clear *a priori* that the figure you get has any axis of symmetry much less two orthogonal ones.

For example since

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0.850651 & -0.525731 \\ 0.525731 & 0.850651 \end{bmatrix} \begin{bmatrix} 1.618034 & 0 \\ 0 & 0.618034 \end{bmatrix} \begin{bmatrix} 0.525731 & 0.850651 \\ -0.850651 & 0.525731 \end{bmatrix}$$

it takes the unit circle to an ellipse with major half-diameter 1.618034, minor half-diameter 0.618034, and whose major axis goes through $(0.850651, 0.525731)$ and has an angle of about $32°$ with respect to the x-axis.



Incidentally, you draw circles in PostScript with the command **arc**. The sequence

```
1 0 moveto
0 0 10 0 360 arc
stroke
```

draws a circle (an arc from $0°$ to $360°$) of radius 10 with centre $(0,0)$. The command **arc** has what at first sight is a peculiar property—it adds the arc to any current path, so that in order to avoid an extra straight line in your picture you may have to move to the initial point of the arc. We shall see later that this apparent peculiarity is actually a natural way to implement curve drawing.

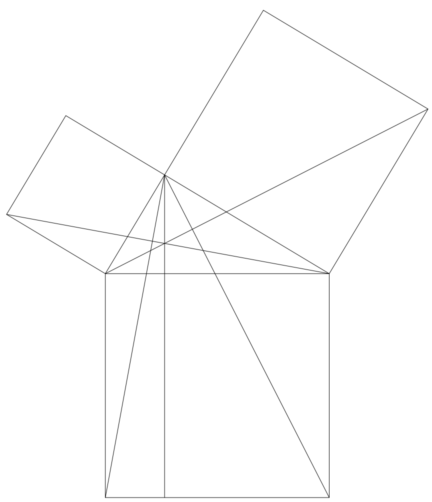
Exercise 13.1. *The ellipse above was made with* **concat** *and then drawing a circle. Write a few lines in* PostScript *which will draw the complete ellipse and its axes.*

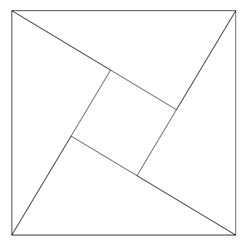
Exercise 13.2. *In the proof of the angle-sum formula using Euler's formula there wasn't any apparent geometry. This is impossible, and there must be some hidden somewhere. Where exactly?*

Exercise 13.3. *If we are given two sets of three points each in the plane, and each set of three is non-collinear, then there exists a unique affine transformation taking one to the other. Write down in your own words a sequence of calculations that will find it.*

Exercise 13.4. *Find the unique affine transformation explicitly when the first three are* $(1, 1)$, $(-1, 2)$, $(1, 3)$ *and the second three are* $(1, -1)$, $(2, 3)$, $(2, -2)$. *Express it in PostScript form (as an array of* 6 *elements, acting on the right). Also write down the* 3×3 *matrix corresponding to it.*

Exercise 13.5. *The following three pictures accompany two proofs of the Pythagoras theorem. (The first is from the proof of Proposition 47 in Euclid.)*





Write proofs to match the pictures. Your proofs should not label points and lines unless absolutely necessary. Instead, I want you to use colouring schemes and lots of drawings to explain what is going on.

There are three diagonal lines which seem to intersect towards the middle of the triangle on the left. Do they in fact intersect, or is it only an artifact of the drawing? Reasons?

Exercise 13.6. *Here is yet another picture for you to translate into a proof of Pythagoras' Theorem. The first trick is to find the triangle!*