

# Perspective Drawing of Simple Shapes with Notes on Calculating Rotations

John Coffey, Cheshire, UK.

2024

Key words : linear perspective, vanishing point, projection, circle, ellipse, rotation, matrix, Euler angle, quaternion, Airy non-linear perspective, computer graphics

This article describes how simple geometrical objects appear when drawn in perspective within a picture, either by hand or with computer graphics. Its emphasise is on the mathematical basis. There are four main sections and supporting appendices:

1. §1 gives an historical overview of perspective and the development of its mathematical basis.
2. §2 examines the maths behind traditional linear perspective, proving how rectangles and circles appear when projected onto the picture.
3. §3 is a lengthy look at rotations of objects in space as performed by today's computer graphics software. This section grew significantly as I realised that there is much more to the calculation of rotations than I had first thought. I have considered hiving this section off into a separate article, but decided to leave it where it is.
4. §4 looks at non-linear perspective which produces pictures as if seen through a wide-angle camera lens and in which straight lines appear bent. The example I take was invented by the Victorian Astronomer Royal, Sir George Airy. Its value is in giving a more realising representation of a wide scene on a flat canvas.
5. Three appendices give supporting details, including proofs of important formulae.

## 1 Perspective in perspective

‘Perspective’ in its broadest sense refers to the methods that artists have used since the renaissance to give an illusion of three dimensional space on the flat surface of a painting, and in particular to depict convincingly objects at different distances from the observer. Earliest artists simply made distant objects smaller than near ones. Even this was a major step towards realistic representational painting, since in the early Middle Ages figures and objects were allocated size according to their importance in the tableau. Thus in Nativity scenes Baby Jesus would be as big as the shepherds. Another device for indicating distance was to

make the farther objects paler and more blue, called ‘aerial perspective’. Landscapes would be added in the background of Biblical scenes and portraits, often depicting a related Bible scene, a walled city and behind them fantastic mountains fading into the clouds.

The scientific study of perspective came in the mid 15th century when men of learning realised that light travels in straight lines. The canvas or drawing paper sheet on which the picture was being made was held vertically in what was called the ‘picture plane’. A distant object was represented by a mark on the canvas in the position at which a ray of light, scattered from the object, appeared to cut through the picture plane on its way to the artist’s eye. This is shown in the 1525 woodcut by Dürer in Figure 1, and similar illustrations were published in early printed books explaining the new science of geometrical perspective. I say ‘artist’s eye’ and not ‘eyes’ because the rays of light were regarded as converging on a single point where the viewer stands; there was no binocular vision. Creating a foreshortened view of objects, people and animals was an interesting challenge to renaissance artists, as in the Dürer example.

As artists sought to draw more complicated architectural features, perspective developed into a branch of geometry. With ruler and compass constructions artists could depict accurately items of considerable complexity such as walls with several arched openings, circular towers with turrets, and spiral staircases. In 1604 the Dutchman Hans Vredeman de Vries published an influential treatise illustrated by immaculate drawings of architectural spaces.

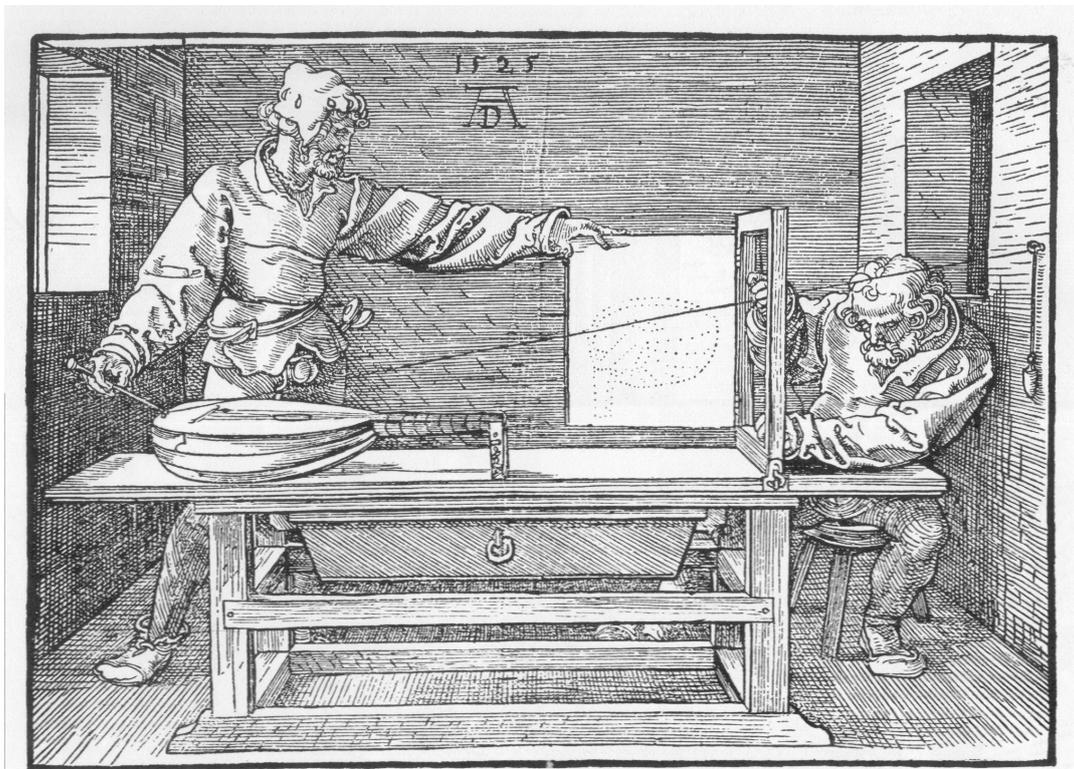


Figure 1: Artists using a taut string to represent a ray of light for drawing a lute in perspective. Woodcut by Albrecht Dürer, 1525.

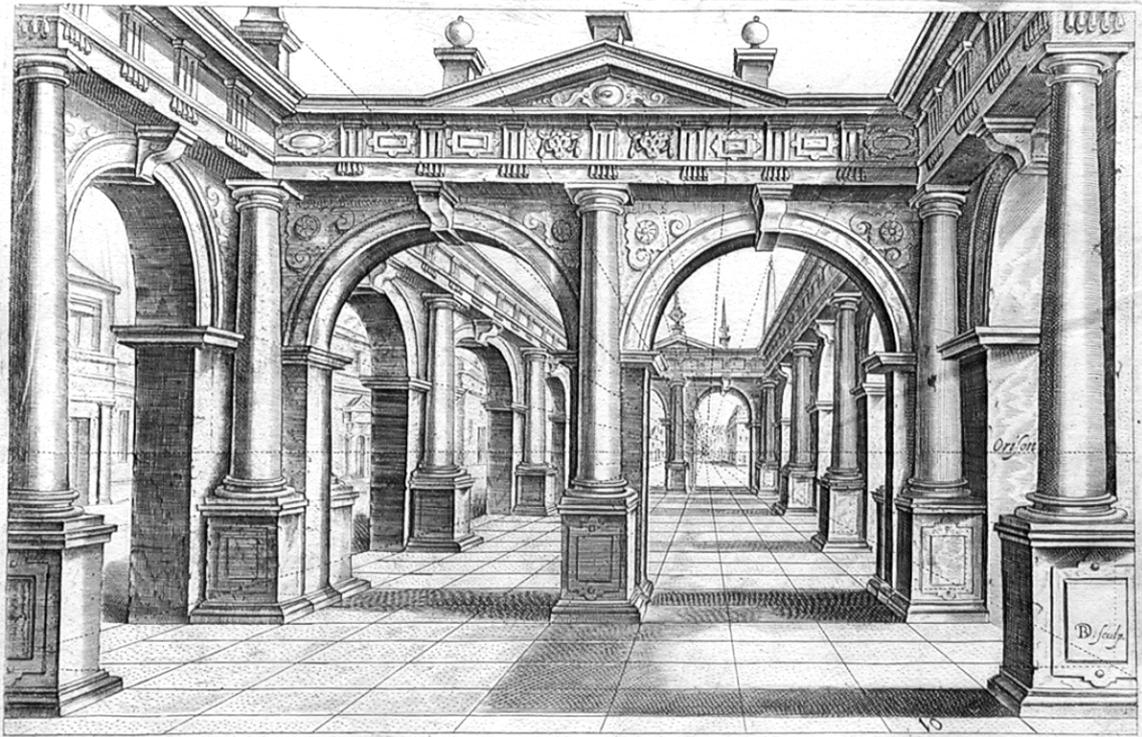


Figure 2: Perspective drawing by de Vries from his 1604 textbook.

Figure 2 is an example. About 1640 the French engineer and architect Girard Desargues published his formulation of perspective into a distinctly new geometry known as ‘projective geometry’. He drew attention to so-called projective properties of geometrical figures, such as the condition for three points to lie on the same straight line, which remain true even when viewed from a different position. The invention of co-ordinate geometry by René Descartes at about the same time allowed curved 3D shapes and their 2D representations to be calculated accurately using algebra. It was shown, for instance, that an ellipse always appears as an ellipse in perspective, and constructions were known for drawing ellipses accurately.

The *camera obscura* (‘dark box’) was also used by artists, probably including Jan Vermeer, to aid the accurate drawing of shapes and, as a bonus, the judgement of colour and tone. This was a hollow box with a ground glass screen at the back and a lens or even just a pin hole at the front through which light from a brightly lit scene could enter the box. The artist could view the image inverted on the screen if he or she covered their head and screen with a black cloth to keep out ambient light. The *camera obscura* was the optical equivalent of the string and viewing frame in Figure 1. This instrument gave the clearest demonstration of light travelling in straight lines. Perspective continued to develop through both mathematical geometry and optical instruments, and in our own time these have become fused in the many 3D computer graphics applications such as in ‘Blender’ ([www.blender.org](http://www.blender.org)) which generate on-screen perspective views in real time.

Architects and engineering draughtsmen, drawing the water-powered mills and steam-driven engines of the industrial revolution, were trained to construct detailed perspective

drawings on a drawing board using simple tools – straight edge, set square and a pair of compasses. In perspective any set of receding parallel lines appears to converge and meet at a ‘vanishing point’. If the lines run in a plane parallel to the ground, the vanishing point will be on the horizon, where sea would meet sky. de Vries’ Figure 2 is an example of ‘1-point perspective’ – as if looking down a tunnel. When buildings are viewed obliquely, the parallel lines of brickwork seen on one side have a different vanishing point from the those in the walls at right-angles, and so the drawing will be in 2-point perspective. If a third set of parallel lines were added in the upwards direction, so as to represent the viewer looking up at a tall building, there would be a third vanishing point in the sky, as in Figure 3. If we look downward, for instance looking at one building from an ever higher one, the third vanishing point will be below eye level.

In the 19th century, over 200 years after de Vries, Desargue and Décartes, the mathematical theory of perspective was revived and reformulated by Jean-Victor Poncelet through his fusion of projective and co-ordinate geometries. Also in the early 19th century other geometries were being invented, particularly the hyperbolic geometry of Nikolai Lobachevsky. These new ways of understanding space were radical abstract departures from the classical geometry of Euclid in which adjacent parallel lines never meet. Poncelet’s projective geometry formally admitted the vanishing point as a new point ‘at infinity’. In one representation of this space the infinite plane of Euclidean geometry is mapped onto the surface a sphere – the Riemann sphere – wrapping the plane around the sphere and gradually shrinking it so there

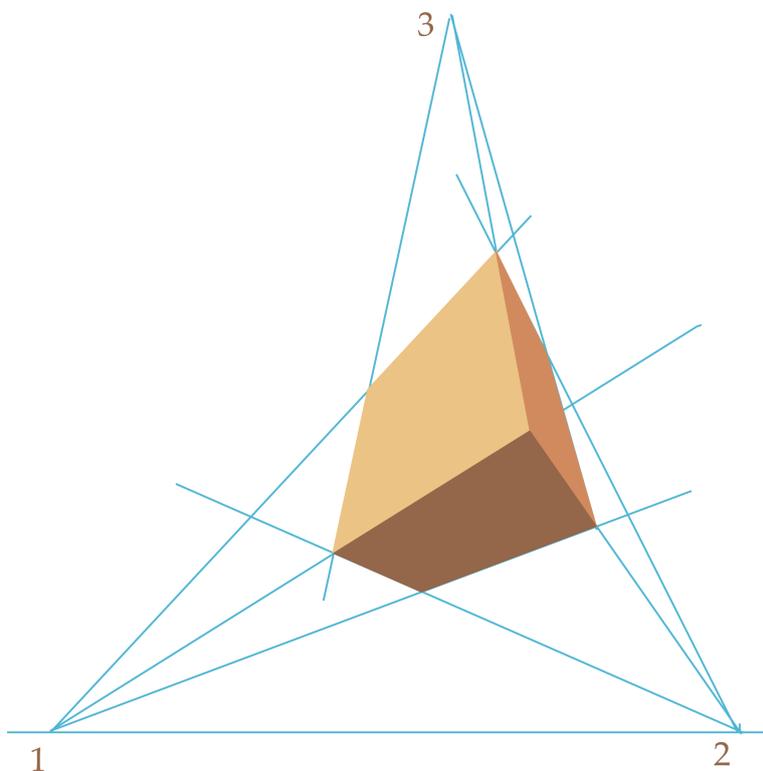


Figure 3: A cuboid seen from below in 3-point perspective. The three vanishing points are numbered. The line 1-2 represents eye level.

are no folds. The centre of the drawing is then at the south pole and the  $360^\circ$  horizon at infinity has shrunk to a single point at the north pole. This has similarities with the geometry of lines drawn on a globe, where lines of longitude meet at the poles.

As an academic subject projective geometry was taught in universities from the 19th century and provided simple proofs of several sophisticated classical theorems including those named after Pascal and Brianchon, both illustrated in Figure 4. In projective geometry all the classic conic sections (circle, ellipse, parabola, hyperbola) are essentially the same object, a remarkable unification. The fact that a consistent geometry could be made without Euclid's postulate about parallel lines never meeting motivated searches for other non-Euclidean geometries, one being the powerful and important hyperbolic geometry noted above. In an influential publication in 1872 Felix Klein redefined geometry as the study of properties of objects which do not vary under displacements, rotations and other transformations in an abstract space, where that space satisfies a set of axioms. A geometry is essentially defined by its invariant quantities.

Even in Euclidean geometry depicting the change in appearance of an object under a change of viewpoint was also aided by mathematics. Matrices provide a means of transforming coordinates under rotations, as do the quaternions invented by Sir William Hamilton in 1843. These are some of the maths behind today's computer graphics, as outlined in §3.

The invention of photography in Victorian times meant that perspective views were produced by an instrument instead of being drawn and painted. Portrait artists became

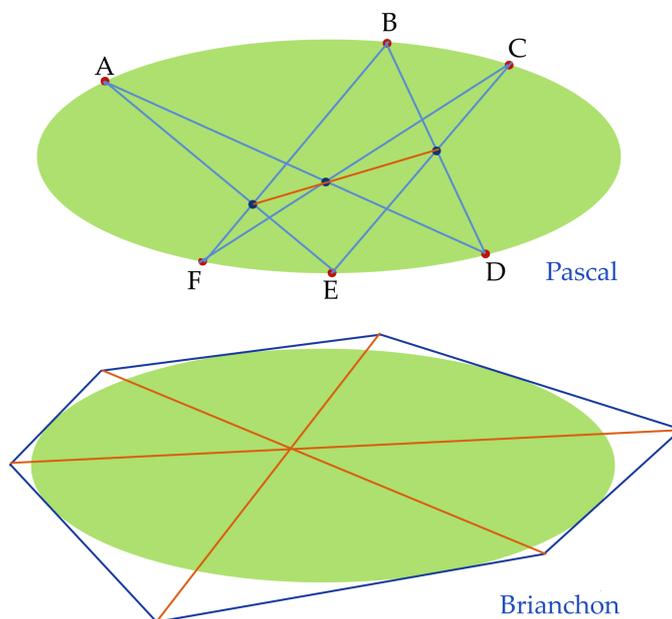


Figure 4: Pascal and Brianchon's theorems of the ellipse. Pascal: A, B, C are arbitrary points on one section of perimeter, joined to their non-corresponding points D+E, D+F, E+F. The points of intersection lie on a straight line. Brianchon: The diagonals of any circumscribed hexagon to an ellipse meet at the same point.

replaced by portrait photographers. A camera's lens projects the scene onto the photographic plate or, today, onto the digital sensor – hence the expression ‘the camera cannot lie’. However, it was recognised early on that every lens has aberrations and so produces an image which is distorted in one way or another; pin-cushion and barrel distortions are two geometrical aberrations. There is a more fundamental principle, however, as to why a single drawing, painting or photograph cannot represent the real world without distortion. This becomes clear when we sit in a room with a geometrical pattern of tiles on the floor, or straight rows of seats to left and right. We do not see the scene before us through a single eye staring resolutely forwards. Instead we have two eyes and turn our heads to look around wherever we are seated. Our viewpoint is always changing. Lines running left to right appear straight from any position of our head, but change their angle as we look along the line. This might lead an artist to draw a straight line as a curve. This is in contrast with the straight left-right lines of strict 1-point perspective seen in the floor tiles in de Vries' Figure 2. 1-point perspective, therefore, is unnatural except when the field of view is narrow.

A more realistic model is to envisage a transparent sphere around your head, like a large goldfish bowl, your eyes at the centre, with yourself looking in all directions, left and right, up and down. Light from the full  $4\pi$  steradians of the external scene enters this sphere. and a true painting would record this over the whole sphere – the ‘picture sphere’ instead of the picture plane. Any flat painting must select only a small solid angle from the total view.

There is a strong analogy with making a map of the world. A globe, being a sphere, can represent the Earth, but it is impossible to depict the whole spherical surface on a flat sheet of paper. In this regard the sphere is different from the cylinder because a cylinder can be opened out into a flat sheet without stretching or cutting. The best that can be done with a sphere is to make an atlas of maps, each centred on a different point of the sphere. There is bound to be distortion around the edges on every page of the atlas because in these places the sphere has been ‘ironed flat’. Map makers have used many different projections of the Earth. British school classrooms used to have a large wall map of the Earth in Mercator's projection. This made Greenland look huge compared with Africa and India. Such maps may have suited the narratives of colonial powers, but other projections such as Lamberts' equal area corrected this bias, though added its own distortions of shape. The Victorian Astronomer Royal, Sir George Airy, used maths to specify a map projection which strikes the balance of showing the largest possible area with the minimum overall distortion. The appearance of a scene in Airy projection is similar to a photograph through a fish-eye lens. It is an example of non-linear perspective.

Paintings and most images are still monocular. Binocular vision requires each eye to be given a different image. Early instruments for this were hand-held viewers into which the user inserted a pair of photographs and then peered at them, seeing a slightly different view of the scene in each eye for the brain to fuse. The photographic pairs had been taken from two camera positions separated by the distance between our eyes. An important development from this was stereoscopic reconnaissance, exploited in World War II using aircraft fitted with two cameras, one on each wing. Expert staff could study these pairs of images through

a stereoscopic viewer and judge the height of terrain and objects on the ground. This in turn developed into photogrammetry, the science of making measurements of position and creating 3D images from overlapping photographs taken from a series of adjacent viewpoints. Meshroom is a free photogrammetry software computer package (<https://alicevision.org>).

In this article I have picked a few of the above topics and given my own account. There is nothing original except my particular way of explaining things. In §2 I discuss how various objects appear when projected onto the picture plane.

## 2 Linear projections onto the picture plane

This section describes the theory behind the traditional way of creating perspective drawings. The geometry is shown in Figure 5. There is an issue over the choice of axes. One would wish a point  $P$  in space to have Cartesian co-ordinates  $(x, y, z)$  with  $x$  horizontal to the right,  $y$  vertically up and  $z$  increasing away from the artist at origin  $O$ . This would be a left-handed system but convention prefers right-handed. In this case, however, right handed axes would involve incorporating  $-$  signs with many quantities, with a risk of confusion. My choice has been to accept the left-handed system and take the scene to lie in the half space  $z > 0$ .

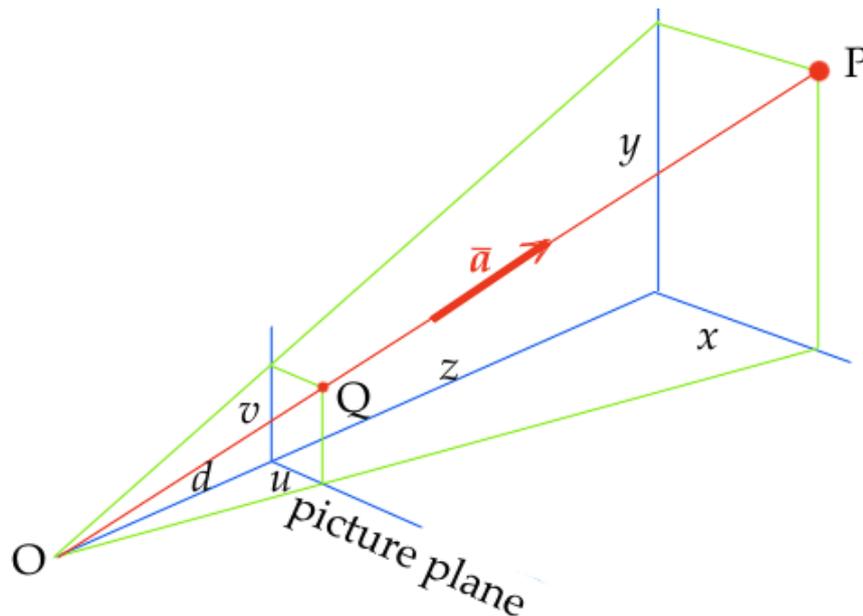


Figure 5: Geometry of linear perspective. The artist's eye is at origin  $O$ . Point  $P$  is projected to  $Q$  on the canvas along ray vector  $\mathbf{a}$ . Note the left-handed axes.

In Figure 5 the artist or viewer is at the origin  $O$ . The line of sight is directly forwards in the positive  $z$  direction. A virtual plane contains the two eyes of the viewer and their line of sight and hence their horizon. I refer to this as the 'viewer's horizon plane'. If the viewer is looking straight forwards, neither up nor down, their horizon will correspond with the horizontal line at which, in a sea-scape, the sea meets the sky. However, if they lift their

head to look upwards, their horizon plane rotates upwards too. The plane of the picture is always at right angles to the viewer's horizon plane at distance  $d > 0$  from their eyes.

## 2.1 Corners and straight edges

**Single point:** Referring to Figure 5, rays of light such as the red line leave an object P in the scene and pass through the picture plane at Q where the artist paints a mark. P is at  $(x, y, z)$  and Q at  $(u, v, d)$ .  $\mathbf{a}$  is a unit vector from P towards O in the direction of the light ray.  $u$  and  $x$ ,  $v$  and  $y$  are in proportion:

$$\frac{u}{d} = \frac{x}{z}, \quad \frac{v}{d} = \frac{y}{z} \quad \text{so } u = x \frac{d}{z}, \quad v = y \frac{d}{z}, \quad z > d. \quad (1)$$

These are the fundamental relations of linear perspective. They are sufficient to map any set of points into their images in the picture. Any point along the infinite line of  $\mathbf{a}$  will be projected to the same single point Q on the canvas. This is an obvious but profound fact – the majority of information about the scene in front of the artist cannot be recorded in the picture. A ray coming along the line of sight maps to the picture's origin,  $(u, v) = (0, 0)$ . If along a receding line there is an array of equally spaced objects with separation  $L$ , such as a line of receding traffic cones, their positions in the picture will change as  $u_n \propto 1/(nL)$ ,  $v_n \propto 1/(nL)$  where  $n$  indexes the objects.

**Straight line:** A general line can be written in vector form as  $t\mathbf{b} + \mathbf{c}$ ; it points in direction  $\mathbf{b}$  and is displaced sideways by  $\mathbf{c}$ .  $t$  is the distance parameter which picks out individual points P along the line. From Eq 1 and with  $d = 1$

$$u = \frac{P_x}{P_z} = \frac{tb_x + c_x}{tb_z + c_z}, \quad v = \frac{P_y}{P_z} = \frac{tb_y + c_y}{tb_z + c_z}.$$

Eliminate  $t$  between  $u$  and  $v$  to obtain the straight line

$$v = mu + c \quad \text{where } m = \frac{b_y c_z - b_z c_y}{b_x c_z - b_z c_x}, \quad c = \frac{b_x c_y - b_y c_x}{b_x c_z - b_z c_x}, \quad (2)$$

the denominators being the same. This proves that a straight line projects as a straight line. While this line in the drawing is unique to the given  $t\mathbf{b} + \mathbf{c}$ , the converse is not the case since any plane can be viewed edge-on and so appear as a line. All lines in that plane will project into the same line in the drawing.

**Parallel lines:** Consider now the projection of two receding parallel lines. Parallel lines all have direction  $\mathbf{b}$  but different offsets  $\mathbf{c}$ . Taking  $d = 1$ ,  $P_1$  on one line projects to

$$u_1 = \frac{tb_x + c_{1x}}{tb_z + c_{1z}}, \quad v_1 = \frac{tb_y + c_{1y}}{tb_z + c_{1z}},$$

while  $P_2$  on a parallel line at similar distance from O projects to

$$u_2 = \frac{tb_x + c_{2x}}{tb_z + c_{2z}}, \quad v_2 = \frac{tb_y + c_{2y}}{tb_z + c_{2z}}.$$

As  $t \rightarrow \infty$  both  $\mathbf{c}$  becomes insignificant and the image  $Q$  converges on

$$(u_b, v_b) = \left( \frac{b_x}{b_z}, \frac{b_y}{b_z} \right). \quad (3)$$

This is the vanishing point, the point at infinity, for all lines parallel to  $\mathbf{b}$ . If  $b_x = 0$ , the vanishing point is centrally placed, straight ahead. If  $b_y = 0$ , it is on the world's horizon, at eye level.

**Lines at right angles:** Consider two sets of parallel lines which set-wise are at right angles to each other but otherwise point in arbitrary directions. One set has lines  $t\mathbf{b} + \mathbf{c}_j$ , all parallel to  $\mathbf{b}$ , and the other set  $\tau\mathbf{g} + \mathbf{h}_k$ , all parallel to  $\mathbf{g}$ . The condition for them to be perpendicular is that  $\mathbf{b} \cdot \mathbf{g} = 0$ . The most frequent case is where  $\mathbf{b}$  and  $\mathbf{g}$  both lie in or parallel to the horizontal  $x - z$  plane and so  $b_y = g_y = 0$ . We then expect two vanishing points on the horizon, one for each set. In this common case

$$\mathbf{b} \cdot \mathbf{g} = b_x g_x + b_z g_z = 0$$

from which

$$\frac{b_x g_x}{b_z g_z} = -1 \quad \text{and hence} \quad u_g = \frac{-1}{u_b}. \quad (4)$$

Here  $u_g$  is the  $u$  co-ordinate in the picture of the vanishing point for the set in direction  $\mathbf{g}$ , and similarly for  $u_b$ . Consider two examples of what this means for drawing a box (a cuboid) in 2-point perspective. First, if the box is turned such that its two visible adjacent vertical faces are at  $\pm 45^\circ$  from the observer, the vanishing point for the right face will be at a distance  $a$  to the right of the drawing's centre, and the vanishing point of the left face at  $a$  to its left. Second, if the cube is turned about its vertical axis so that the face and edges to the right move their vanishing point to  $u = 2a$ , the ones to the left will move to  $u = -1/2a$ . The reciprocal relation ensures that the sides of the cube appear at right angles.

**Example drawings:** The three panels in Figure 6 have been drawn to this specification using  $d = 1$ . The left was calculated and plotted as if by hand from Eq 1. The central and right panels were created in the Blender graphics software, in wireframe and colour rendered modes. Each panel depicts the same three boxes:

1. Green: viewed face-on. Front face extends from  $x = -3$  to  $x = -1$  and  $y$  from 1 to  $-0.5$ . Front face is 4 units from the picture plane at  $z = 5$ , and back face at  $z = 6$ .
2. Blue: top 1 m square,  $y$  dimension 1.5 m, top 1 m below eye level, rotated at  $30^\circ$  and displaced to the right, with nearest vertical edge 3m from the picture plane at  $z = 4$ ,  $x = 1$  m.
3. Red: top and bottom faces 1 m  $\times$  0.5 m, height in  $y$  0.5 m, bottom face 2 m above eye level, rotated at  $30^\circ$  and displaced left with nearest edge 2m from the picture plane ( $z = 3$ ),  $x = -1$  m.

The green box is in 1-point perspective. Its vanishing point is at the origin. The blue and red boxes are examples of 2-point perspective. Table 1 lists the coordinates in 3D space of the corners of the top of the blue box and of their projections in the picture according to Eq 1.

$x$	$y$	$z$	$u$	$v$
1	-1	-4	0.25	-0.25
1.866	-1	4.5	0.415	-0.222
1.366	-1	5.366	0.255	-0.186
0.5	-1	4.866	0.103	-0.206

Table 1: Co-ordinates of the top of the blue cuboid in Figure 6 and their projected values in the picture plane.

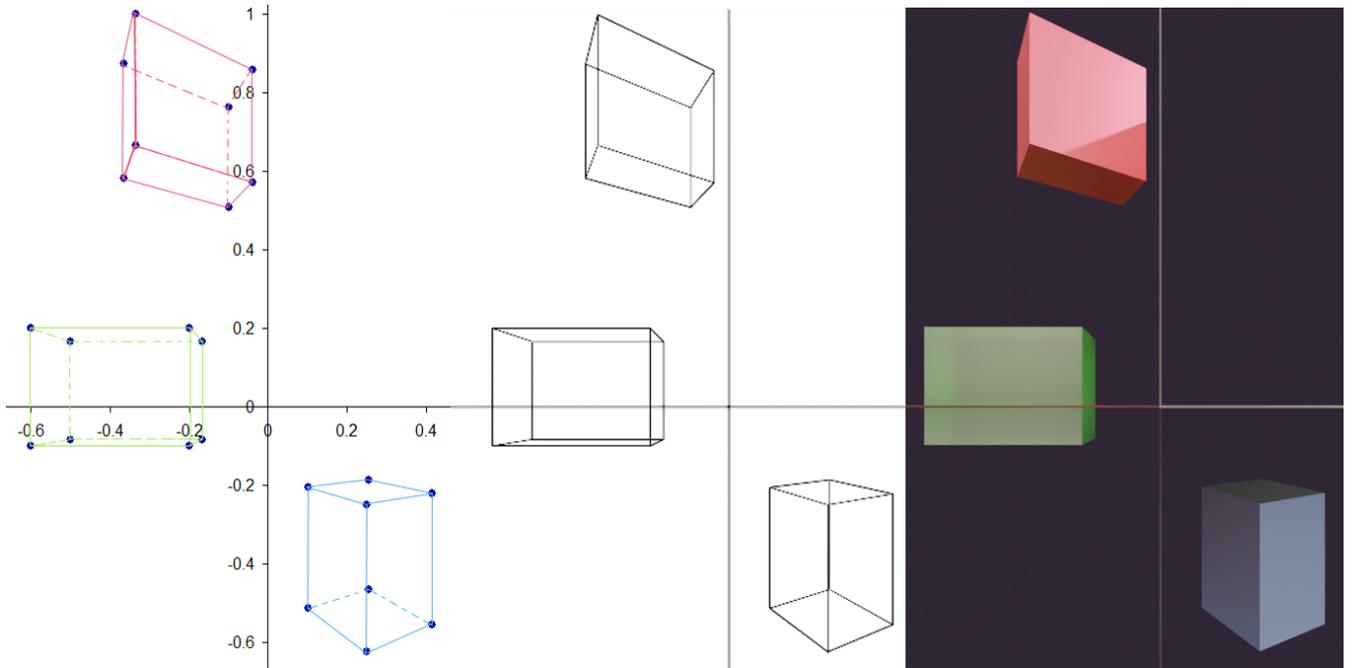


Figure 6: Persepective view of three cuboids. Left: Plotted from Eq 1. Centre: Blender wireframe. Right: Blender rendered in colour.

For an example of 3-point perspective we have to look upwards or downwards so that vertical lines recede from our view. The effect is equivalent to rotating the whole 3D scene in the oppose direction while looking forwards. Figure 7 has been calculated in this way; it is the equivalent of Figure 6, with the viewer moved down  $h$  units but looking upwards at angle  $\theta$ . The  $x$  co-ordinate remains unchanged, but  $y$  and  $z$  transform according to

$$y \rightarrow (y + h) \cos \theta + z \sin \theta, \quad z \rightarrow (y + h) \sin \theta - z \cos \theta. \quad (5)$$

For example, if  $h = 3$  m and  $\theta = 30^\circ$ , the corner of the blue box at  $(1.866, -1, 4.5)$  is at  $(1.866, -0.518, 4.90)$  in the shifted and rotated frame and projects to  $(0.381, -0.106)$  in the picture.

The vanishing points in Figures 6 and 7 are calculated from the vectors  $\mathbf{b}$  through the origin and parallel to the sides of the boxes. Take the edges along the top of the blue box. Two converge to the right in the drawing, and two to the left. The vector  $\mathbf{b}$  for the rightwards edges has components  $(\cos 30^\circ, 0, \sin 30^\circ)$  so by Eq 3 they converge at  $u = 1/\tan 30^\circ = \sqrt{3} = 1.732$ . The leftwards edges have vector  $(-\sin 30^\circ, 0, \cos 30^\circ)$  so converge at  $u = -\tan 30^\circ = -1/\sqrt{3} =$

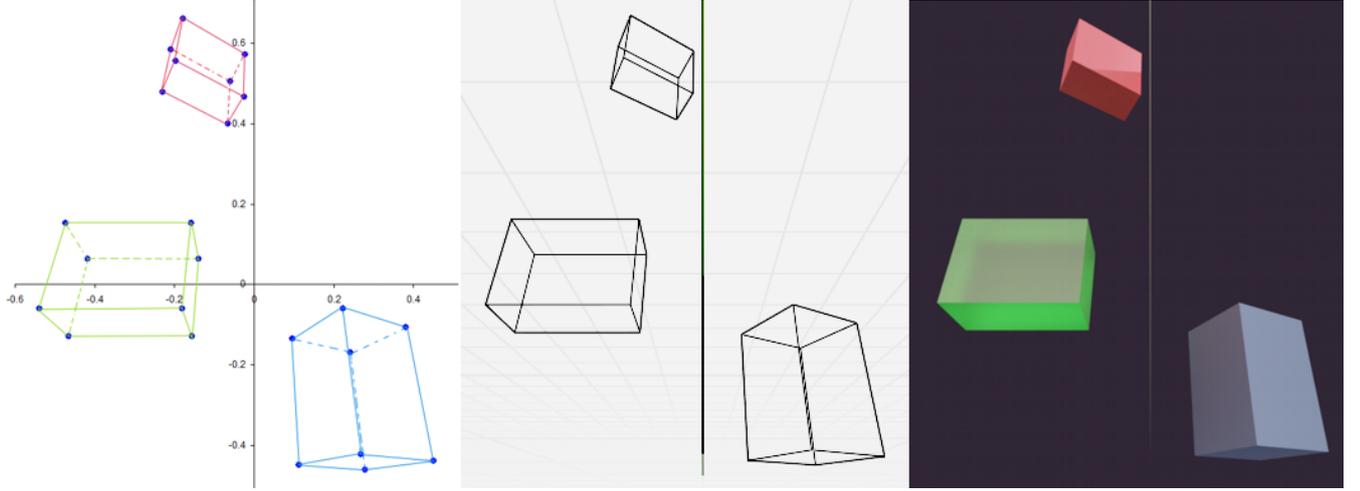


Figure 7: The same three cuboids seen from 3 m below, looking up at  $30^\circ$ .

$-0.577$ . The leftwards vanishing point is shown in Figure 6. In Figure 7 all vertical edges share the vector  $(0, \cos 30^\circ, \sin 30^\circ)$  in the shifted and rotated frame so the third vanishing point is at  $(0, \sqrt{3})$  in the drawing.

## 2.2 Circles and ellipses

**Circle along line of sight:** We start with the simplest case of a circle in a horizontal plane, placed symmetrically in  $x$ , but placed vertically so it is viewed obliquely either from above if  $y_0 < 0$ , or below if  $y_0 > 0$ . Thus the circle is centred at  $C = (0, y_0, z_0)$  with radius  $r$ . Its parametric equation is  $x = r \cos \beta$ ,  $z = r \sin \beta + z_0$  where  $\beta$  is the angle anticlockwise from the  $+x$  direction in the plane of the circle. Taking  $d = 1$ , by Eq 1 a general point on the circle projects to  $(u, v)$  in the picture where

$$u = \frac{r \cos \beta}{r \sin \beta + z_0}, \quad v = \frac{y_0}{r \sin \beta + z_0}, \quad 0 \leq \beta < 2\pi.$$

We wish to see whether the circle projects as an ellipse. One approach is first to find the semi-axes on the assumption that it is an ellipse. At the top and bottom points of curve in the drawing the tangent lines are horizontal. Differentiating with respect to  $\beta$

$$\frac{dv}{d\beta} = \frac{-r y_0 \cos \beta}{z^2} \quad \text{which} = 0 \quad \text{where} \quad \beta = \frac{\pi}{2}, \frac{3\pi}{2};$$

that is, the nearest and farthest points on the circle project to the top and bottom points of its image in the drawing, as we expect. At  $\pi/2$   $v = y_0/(z_0 + r)$  and at  $3\pi/2$   $v = y_0/(z_0 - r)$ . The mean of these is  $y_0 z_0/(z_0^2 - r^2)$  so this must be the  $v$  coordinate of the centre E of the ellipse, assuming it is an ellipse. Call this co-ordinate  $c$ . Similarly the semi-minor axis length,  $b$ , would be half their difference, or  $r y_0/(z_0^2 - r^2)$ .

The centre C of the circle does not project to the centre E of the ellipse. C maps to  $C' = y_0/z_0$  while E is displaced from C by  $\delta$  given by

$$\delta = \frac{y_0}{z_0} - \frac{y_0 z_0}{z_0^2 - r^2} = \frac{-r^2 y_0}{z_0(z_0^2 - r^2)}.$$

If  $y_0 > 0$  such that the circle is seen from below,  $\delta < 0$  meaning that C is slightly closer to O than E.

The semi-major axis  $a$  can be found similarly from

$$\frac{du}{d\beta} = \frac{-r(z_0 \sin \beta + r)}{z^2} \quad \text{which is 0 where} \quad \sin \beta = \frac{-r}{z_0}.$$

Then  $\cos \beta = \sqrt{z_0^2 - r^2}/z_0$ . Substituting into  $u$ , the semi-major axis  $a = r/\sqrt{z_0^2 - r^2}$ . The form of the ellipse would be

$$\frac{u^2}{a^2} + \frac{(v-c)^2}{b^2} = 1$$

which we now test by substituting the above values for  $a$ ,  $b$ ,  $c$ . The algebra is somewhat involved, but the left side does indeed evaluate to 1, so proving that the circle projects to an ellipse in the picture plane.

This algebra points to a recipe for drawing a centrally placed horizontal circle accurately in perspective, using only a straight edge and pair of compasses. The construction is shown in four stages in Figure 8, panels A to D, D being an enlargement of C. The stages are as follows:

1. Panel A. In this supplementary diagram draw the picture plane at distance  $d = 1$  unit from the viewer at O and draw the horizontal circle edge-on as if looking down the  $x$  axis. Construct rays from the closest and furthest points on the circle to cut the picture plane.

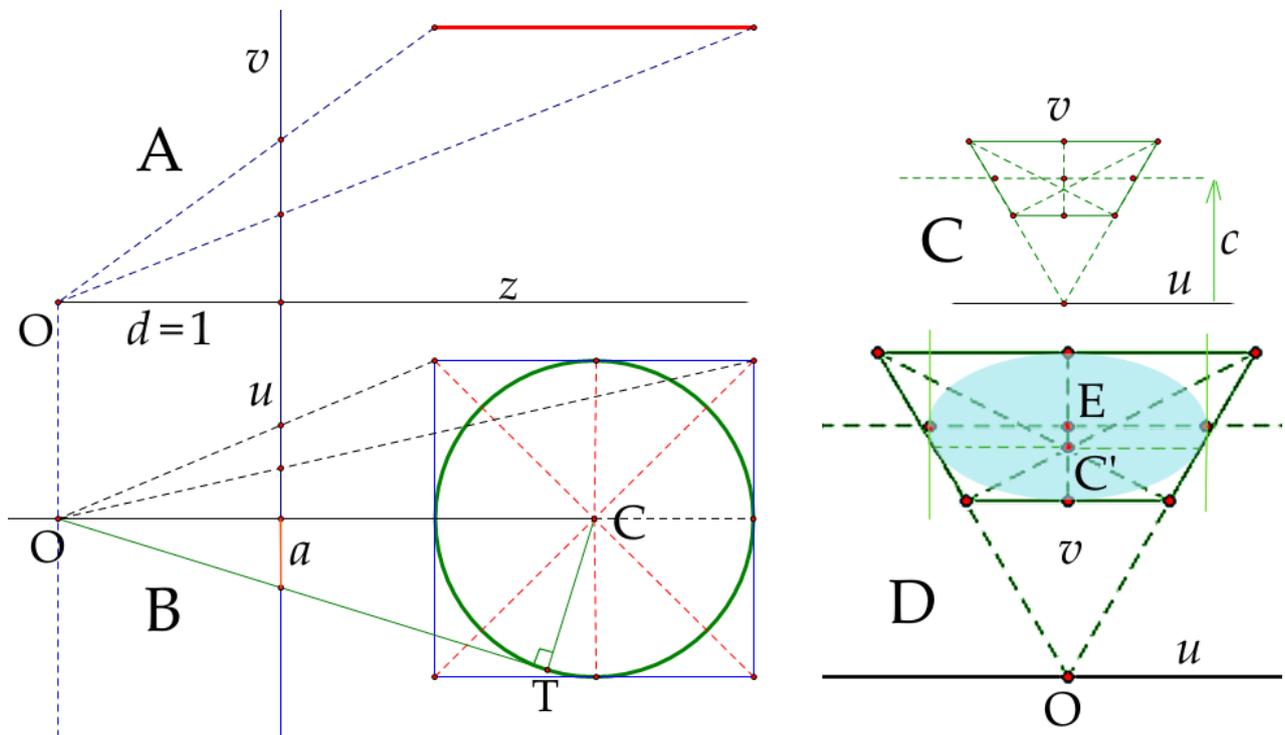


Figure 8: Four panels showing the accurate construction of a centrally placed circle in perspective.

These mark the upper and lower  $v$  co-ordinates of the ellipse. Transfer these dimensions to the main drawing as in panel C and draw lines parallel to the  $u$  axis to represent two sides of the projection of the square which contains the circle. Also construct a parallel line midway between these two. This will be the major axis of the ellipse, whose centre E can now be marked.

2. Panel B. In a second supplementary diagram draw the circle at its correct position along the  $z$  axis and draw the enclosing square. Join the four corners of the square to O and note where these lines cut the picture plane. Transfer these  $u$  distances to panel C and complete the trapezium which is the projection of the square. The trapezium will enclose the ellipse and its sides will be tangents.
3. Panel B again. From O draw the tangent to the circle and mark where it cuts the picture plane. The distance labelled  $a$  is the semi-axis  $a$  of the ellipse. Transfer this to either side of the horizontal centre line through E in panel C. The two points so constructed mark the left and right visible limits of the ellipse, and will lie just inside the trapezium.
4. Panel D. Draw vertical guide lines (parallel to  $v$ ) through these last two points. By hand or otherwise construct the ellipse.
5. If required, construct the diagonals of the trapezium. These represent radial lines in the circle.

A few points are to be noted:

- The distance  $\sqrt{z_0^2 - r^2}$  is the length of the tangent OT in panel B, and  $r/\sqrt{z_0^2 - r^2}$  is  $\tan(\angle COT)$ . Then  $a = d \tan(\angle COT)$ .
- The ellipse can be drawn accurately by at least three constructions:
  1. One uses the parametric equations  $u = a \cos \eta$ ,  $v = b \sin \eta$ . Draw two circles with radii  $a$  and  $b$  both centred on E, then draw any radial line through E. Draw a vertical line through the point at which this radius intersects the circle radius  $a$ , and a horizontal line through the point where it intersects the circle radius  $b$ . The intersection of these two construction lines lies on the ellipse. Build a quadrant of the ellipse point by point and then by symmetry complete the full curve.
  2. The second method is described in a companion article on [www.mathstudio.co.uk](http://www.mathstudio.co.uk) ‘How far can a ladder reach up a wall?’, Figure 7.
  3. Use the fact that the combined distance from one focus to a general point and then to the other focus is a constant, equal to the major axis. The construction can be made with a piece of string and two pins at the foci.

With a suitable rotation of axes the construction also works for vertical circles and those at other angles provided its centre C lies in a plane, normal to the circle, which passes through O. Figure 9 shows receding lines of Romanesque circular arches, with the centre of each circle being at eye-level, each of which could be drawn using the above construction. Moreover, it is clear that the construction will apply equally to ellipses under the same restrictions.



Figure 9: Picture using Blender illustrating 1-point perspective with circular arches and floor inlays.

**Circle displaced sideways:** Figure 10 was produced in Blender. It shows three circles projected as ellipses in 1-point perspective. The rotations of the two ellipses to left and right is clear. You may reasonably point out that a circle does not look like that, but bear in mind that 1-point perspective assumes the viewer is steadfastly looking directly ahead through one eye; he or she does not turn their head to look at the left or right circle for then it would be on their line of sight instead of being seen out of the corner of their eye.



Figure 10: Rotation of ellipses representing circles when displaced sideways.

Suppose the object in the scene is in fact an ellipse lying horizontally, parallel to the  $x - z$  plane, but otherwise in a general orientation and at a general position. How will this project onto the picture plane? I will give only a qualitative explanation as the mathematics is fairly impenetrable, though the Appendix 1 sets out some of it. The Appendix 1 draws attention to the cone of light leaving the object and entering the eye at  $O$ . This has an elliptical cross-section; if it were circular, we would not notice any rotation. When the object lies to the

side,  $x \neq 0$ , the cone is turned around the  $y$ -axis and is cut obliquely by the picture plane as is illustrated in the two panels of Figure 11. Rays from points on the object nearer  $O$  cut the picture higher up – that is, nearer the plane  $y = 0$ . This, together with the elliptical section, accounts for the rotated image.

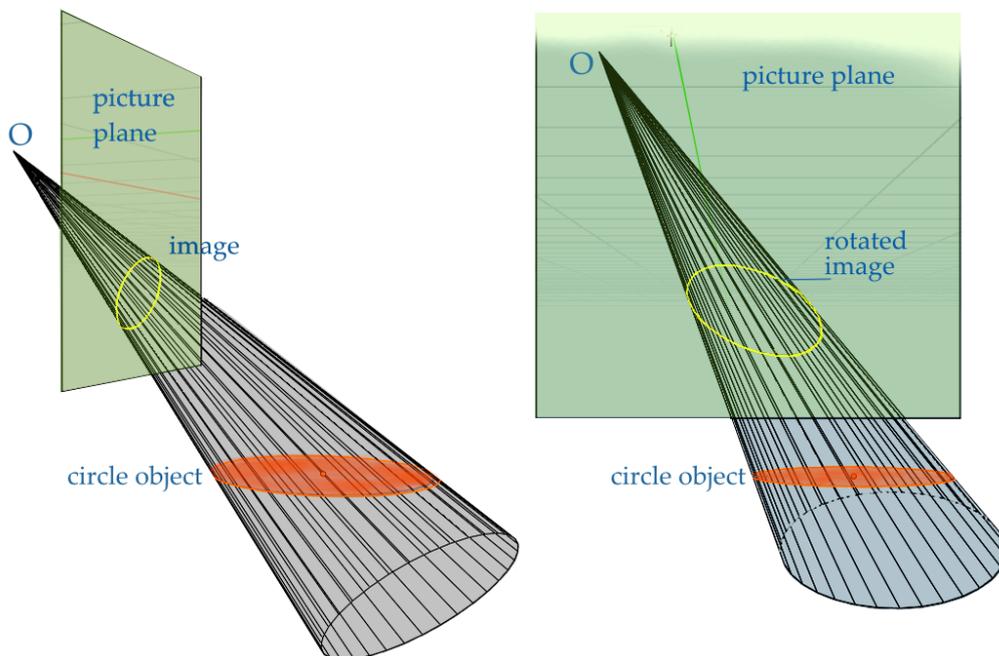


Figure 11: Origin of rotated ellipse image of a circle displaced to the side. Left: view towards  $O$  from near the object. Right: view onto the picture plane.

### 2.3 Wheels and axles

When I was a boy, I had a book called ‘How to Draw Cars’ by the distinguished artist Frank A.A. Wootton (Studio Publications, 1949). His chapter three is about drawing wheels and circular headlights and in it he writes

The most common problem encountered by the student, assuming he can draw an ellipse, is the application of the ellipse to the drawing. I mean the angle of the axis of the ellipse in relation to the axle of the car. This angle is governed by the simple rule – the axis of the ellipse is always at right angles to the axle. ... This applies to any view of the car.

This is probably a rule discovered decades ago by working artists, but does it have mathematical rigour? To gain a feel for the proposition Figures 12 and 13 show a wheel and axle pair generated and rendered in Blender. Each panel in Figure 12 has the object placed in the centre of the picture, with the viewer looking straight at it. Superimposed on each image is a pale yellow rectangle which I have placed so that one edge lies along the centre of the axle and an adjacent edge is along the major axis of the ellipse, being the circle in perspective. Since the corner of each rectangle is a right-angle, this allows the proposition to be judged by eye. Note that the major axis does not pass through the base of the axle; this accords with the theory of §2.2.

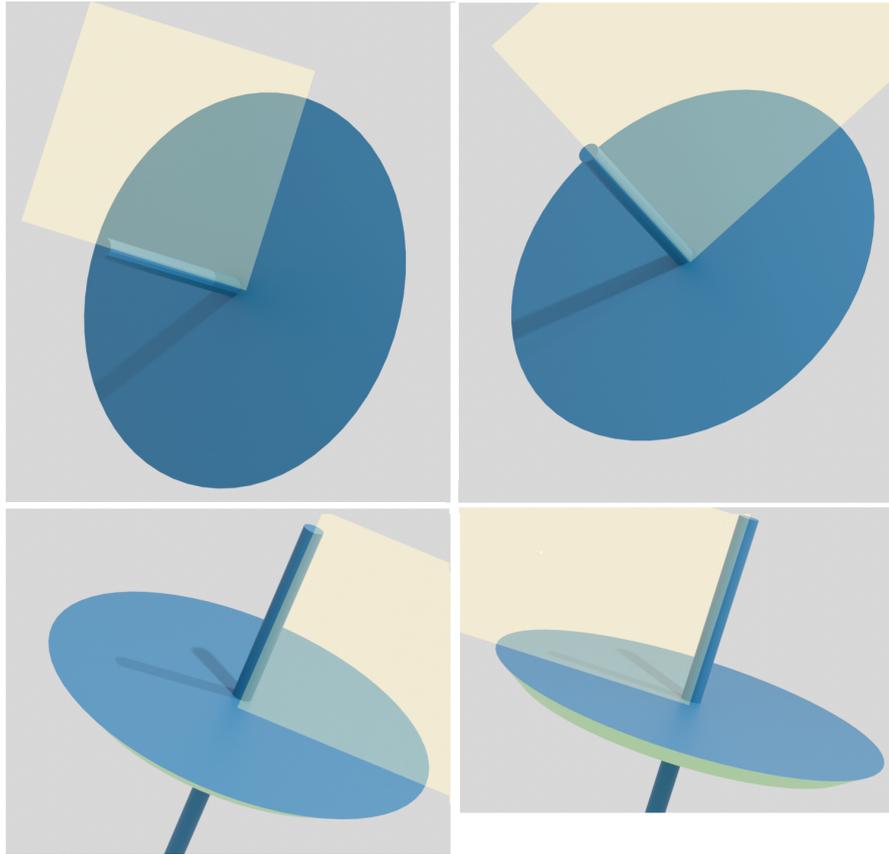


Figure 12: Four examples of a circular wheel and axel drawn in perspective with Blender.

In Figure 13 the wheel-axle pair has been translated along a path at about  $45^\circ$  to the picture plane. In the two cases of distant wheels I have superposed two plane yellow rectangles. They are displaced relative to each other, which shows that the line of the axle is *not* exactly at right-angles to the major axis. This indicates that Wootton's proposition is an approximation, true for wheels close to the centre of the picture, but not exact for ones positioned well to the side. The situation therefore has similarities with the rotated ellipses in Figure 10.

So what is the explanation? Looking at the four examples in Figure 12, it is clear that nothing geometrical is changed if the images are rotated on the page to bring each axle vertically down the centre of the picture. The only difference then between the pictures is the tilt of the parent circle about the  $x$ -axis (horizontal axis) which changes the aspect ratio of the ellipse. However in this rotation, in which the axle tips towards or away from the observer, the axle remain represented in the picture by a vertical line. Moreover this line contains the minor-axis of the projected ellipse. The major axis, always being at right-angles to the minor, is therefore always at right-angles to the axle. Therefore Wootton is correct. He could equally have said that the minor axis of the ellipse which represents the wheel is co-incident with the axle, at least when the wheel is close to the centre of the picture.

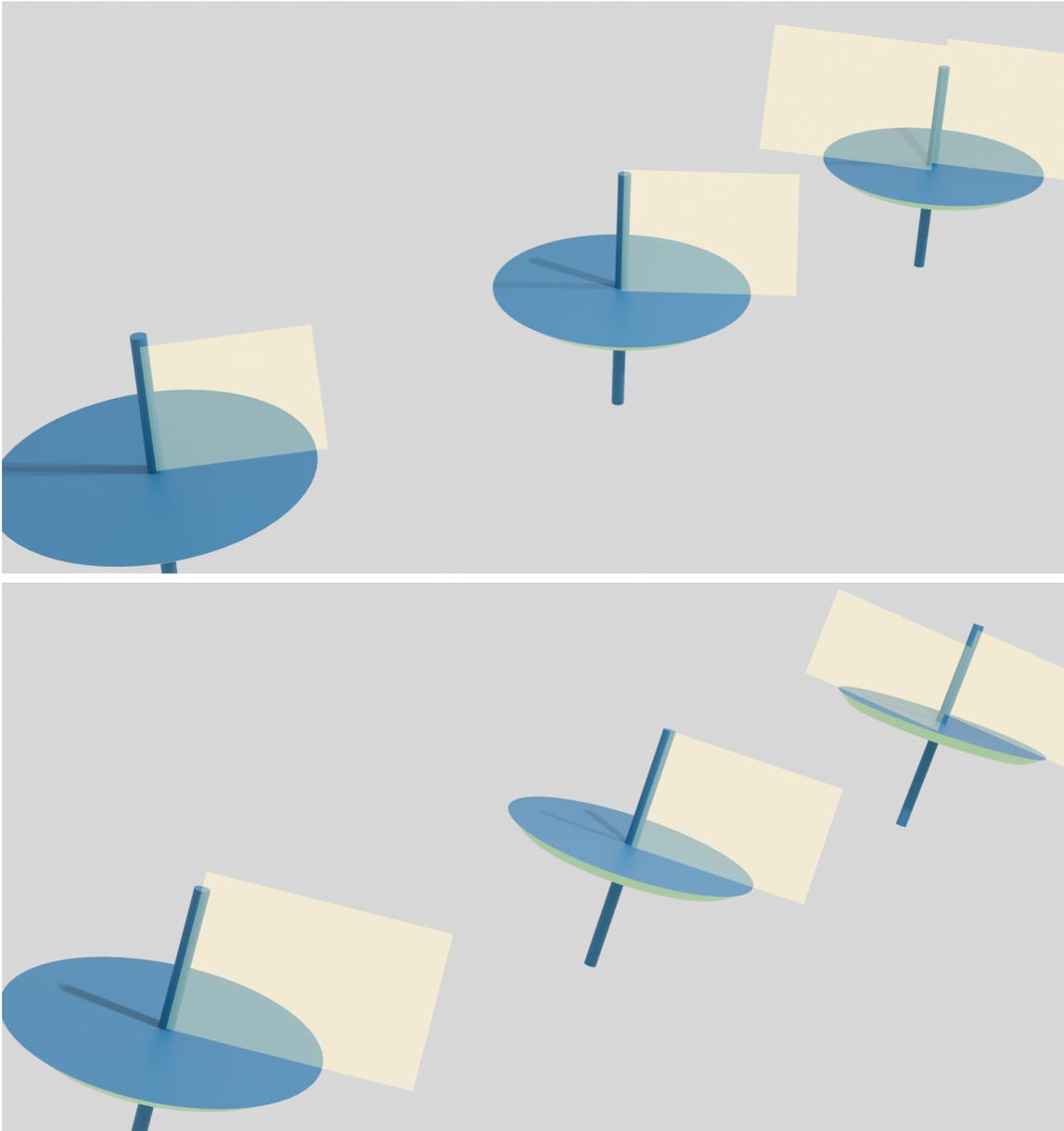


Figure 13: Wheel-axle pairs displaced to left, centre and right.

### 3 Rotating objects within the scene

This is a fairly lengthy discussion of the maths behind the rotation of objects and their images, and has most relevance to computer graphics.

#### 3.1 Main features of rotations

The drawings we are discussing can be sketched or painted by hand, or constructed within a computer graphics program such as Blender or Unity. When a graphics program displays a scene in linear perspective, the artist can use the mouse or other device to change the scale and/or point of view. This requires the image to be redrawn on the screen within a fraction

of a second. Hence not only is rapid computing power highly desirable for a usable computer system, but the algorithms for translation, scale change and rotation about any axis must be fast and reliable. Of these operations, rotation is the most complex and can be achieved with either matrix multiplication or quaternions. Below I outline what is involved. If the reader is disposed to think that rotations are simple, I suggest they study the book by Simon Altmann entitled ‘Rotations, Quaternions and Double Groups’! (OUP, 1986, reprinted by Dover Publications, 2005)

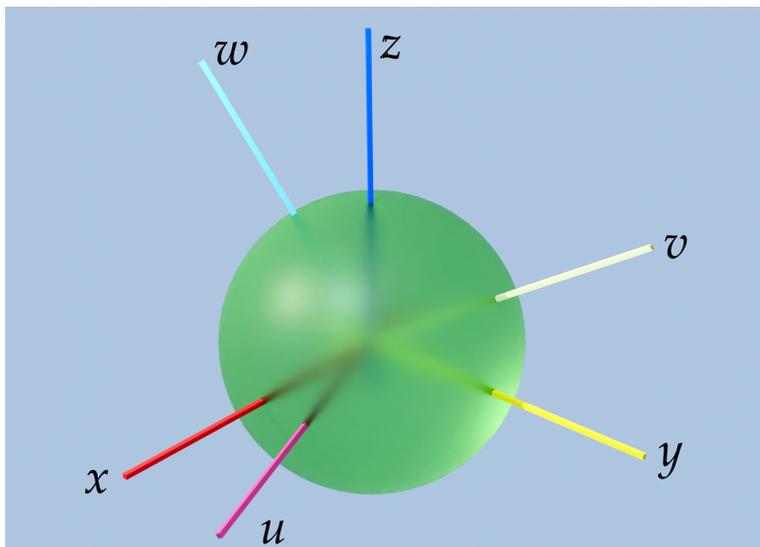


Figure 14: The unit sphere with fixed global axes  $x, y, z$  and local object axes  $u, v, w$  embedded in the rotated sphere.

In using computer graphics software an artist will first create or import the elementary objects into the scene then position them, the lights and the camera in the desired relative positions and orientations. Translation is easy to calculate; simply add to the coordinates of each point on the object the respective  $x, y$  and  $z$  components of the translation vector. Rotations and reflections are more complicated because the components become mixed together. All transformations are vital too to robotics. While it is relatively straightforward to calculate the orientation and position that a robotic hand would arrive at from a given sequence of joint motions, the ‘inverse kinematics’ problem of calculating the joint translations and rotations to achieve a required end configuration is more challenging. It is, of course, something that we humans and all animals do instinctively.

In thinking about rotations it will be helpful to relate them to the unit sphere and the two sets of axes in Figure 14. As noted above, each object is created with its own local or object coordinate frame,  $uvw$ , and placed in a scene which has the fixed ‘global’ coordinate system  $xyz$ . In its initial position and orientation the object will have its local frame aligned with the global one,  $u$  coincident with  $x$  and similarly. The direction of any point in the scene is specified uniquely by a unit vector with end point in the surface of the sphere.

There are several ways by which the an object can be turned into a specified orientation:

1. by ‘axis plus angle’ or ‘skewer and turn’: identify the single axis of rotation and rotate

around this by an angle  $\theta < 2\pi$  or  $-\pi < \theta \leq \pi$ . It is like turning a meat ball or veg burger on a skewer. Only the points along the axis (skewer) do not move. This poses the problem of specifying the orientation of the axis and the angle of turn. Two numbers components are necessary to specify the axis (the third being given by ‘sum of squares equals 1’), and one angle about that axis, making three quantities in all. Appendix 2 supports this topic.

2. the quaternion method of rotation described in §3.3 and 3.4 and Appendix 3 is an efficient implementation of the axis-plus-angle approach.
3. by making a rotation about each of the fixed global axes in turn through what are termed Euler angles. Rotations about global axes are called ‘extrinsic’ rotations.
4. by making a rotation about each of the embedded local axes in turn, called ‘intrinsic’ Euler rotations.

Points 1 and 2 both relate to a rotation as a single turn about one specified axis. Points 3 and 4 relate to three rotations about axes that are either fixed in space (global) or fixed in the object. These approaches are philosophically different. The first is more intuitive, but the second has long been the more widely used. We discuss Euler rotations in §3.2 and the axis-angle approach in §3.3.

### 3.2 Euler rotations using matrices

Rotation is typified by Figure 15. We are looking at the  $x - y$  plane and rotating about the origin. There are two ways in which this rotation can be regarded: either the axes remain fixed and the object rotates about O with respect to them (left panel), or the object remains fixed but the axes frame is rotated to become a new frame in which the co-ordinates are  $(x', y')$ . For example, if a point at  $(1, 0)$  on the  $x$ -axis is rotated by  $90^\circ$  anticlockwise, its co-ordinates in the same frame become  $(0, 1)$ . If the same point stays fixed, but a new frame introduced which is rotated  $90^\circ$  anticlockwise, with the new  $x$  axis pointing along the old  $y$  axis, its co-ordinates are  $(0, -1)$  with respect to the new frame. Looked at another way, referring an object to new axes which are rotated anticlockwise is equivalent to rotating the object clockwise in the old frame.

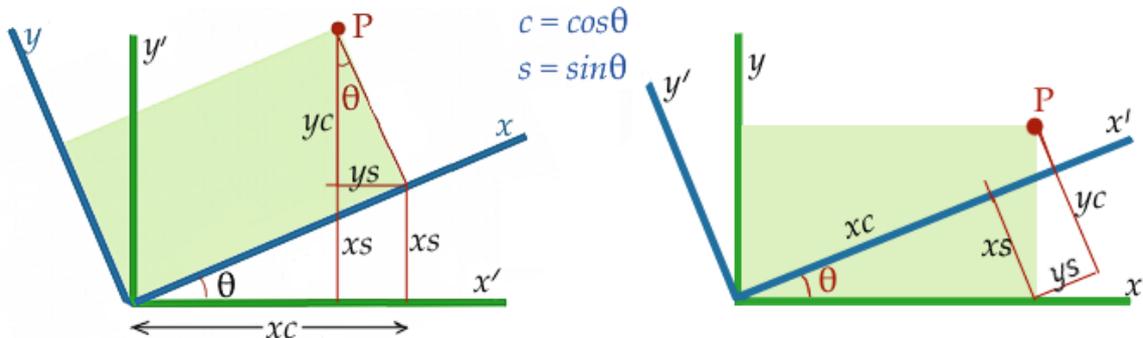


Figure 15: Basis of matrix formula for rotation. Left: Object rotated anticlockwise against fixed axes. Right: axes rotated anticlockwise with respect to fixed object.

Referring to Figure 15 the coordinates of the object after rotation in terms of those before are:

$$\begin{aligned} \text{Fixed axes, object rotated anticlockwise : } & \begin{cases} x' = x \cos \theta - y \sin \theta \\ y' = x \sin \theta + y \cos \theta \end{cases} \\ \text{Object fixed, new axes rotated anticlockwise : } & \begin{cases} x' = x \cos \theta + y \sin \theta \\ y' = -x \sin \theta + y \cos \theta \end{cases} \end{aligned} \quad (6)$$

The first of these in matrix form is  $\mathbf{U} = \mathbf{M}\mathbf{X}$  where  $\mathbf{X} = (x, y)^T$  and

$$\mathbf{M} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

To be clear, in the fixed axes case  $(u, v)$  are the new coordinates of a point on the object with respect to the  $x-y$  frame after the object has rotated  $\theta$  anticlockwise. The object is regarded as having its own ‘local’ or ‘object’ axes embedded within it, and it is this local axis system which has been rotated. In the alternative fixed object case  $(x', y')$  are the coordinates of the object with respect to a new set of axes which have been rotated anticlockwise by  $\theta$  with respect to the previous frame. We will be concerned almost entirely with the first, fixed axes case. If we have the equation of an object in the  $(x, y, z)$  frame, to rotate it anticlockwise about the  $z$ -axis by  $\theta$ , replace  $x$  and  $y$  by the inverse formulae:  $x$  by  $x \cos \theta + y \sin \theta$  and  $y$  by  $-x \sin \theta + y \cos \theta$ . As an example of Eq 6, if the object is initially at  $(a \cos 10^\circ, a \sin 10^\circ)$  and its position is rotated by  $\theta = +60^\circ$ , its new co-ordinates will be

$$x' = a \cos 10^\circ \cos 60^\circ - a \sin 10^\circ \sin 60^\circ = a \cos 70^\circ$$

and similarly for  $y'$ .

Using the identities  $\sin \theta = \cos(\pi/2 - \theta)$ ,  $-\sin \theta = \cos(\pi/2 + \theta)$ , the elements of  $\mathbf{M}$  are seen to be cosines of the angles between the respective old and new co-ordinates.

$$\mathbf{M} = \begin{pmatrix} \cos \theta & \cos(\frac{\pi}{2} + \theta) \\ \cos(\frac{\pi}{2} - \theta) & \cos \theta \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},$$

where, for instance,  $\arccos c_{12}$  is the angle between  $x$  and  $v$ . The  $c_{ij}$  can also be thought of as the components in the  $xyz$  frame of a unit vector pointing in the direction of P.

This is readily generalised to three axes. Figure 15 is now regarded as the view down the  $z$  axis towards the  $x-y$  plane, with anticlockwise rotation about the  $z$  axis. The matrix of direction cosines between old and new axes is

$$\mathbf{M} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}, \quad \begin{array}{c|c} & \text{‘old’, } x, y \\ \hline \text{‘new’, } x' & \\ y' & \end{array} \quad (7)$$

The elements in each row are the components of a unit vector along local/object axes with respect to the fixed global axes. By Pythagoras’ theorem the sum of their squares is 1. The determinant of this matrix is +1, the 1 meaning it is pure rotation with no stretching, and

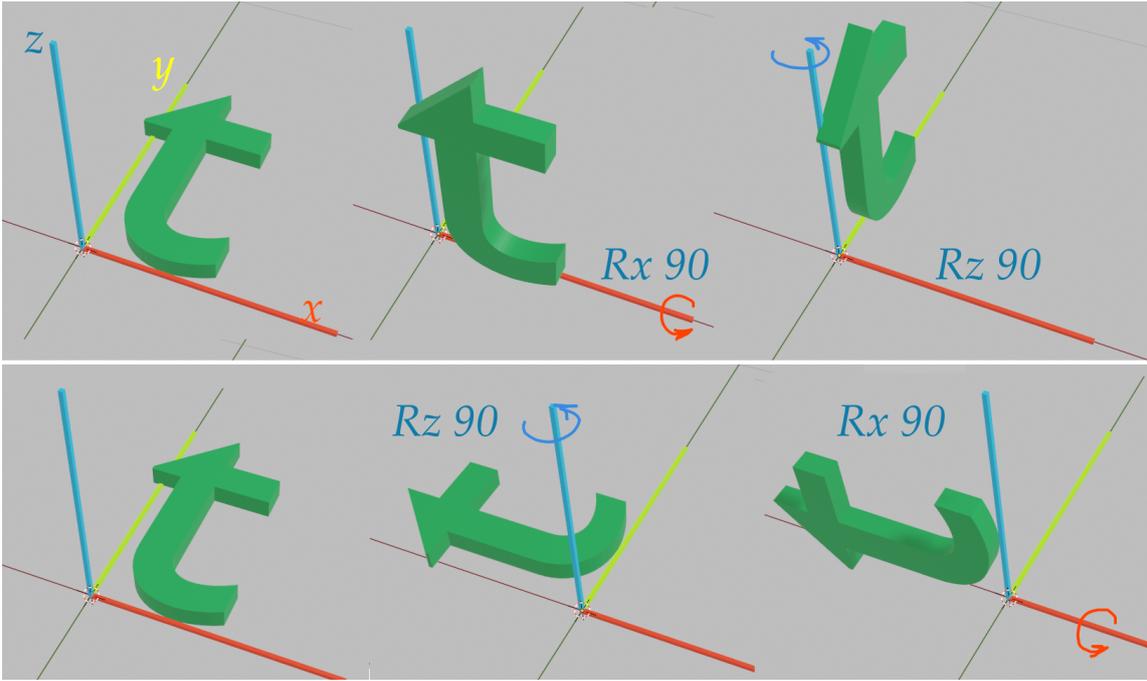


Figure 16: Illustrating that rotations about different axes do not commute. Rx 90 means a rotation about the global  $x$ -axis through  $90^\circ$ . Axes:  $x$  red,  $y$  yellow,  $z$  blue.

the + sign meaning there is no reflection. This is an orthonormal matrix. Only matrices with +1 as an eigenvalue can describe pure rotation since only then has the corresponding unit eigenvector no stretching or shrinking. This eigenvector defines the axis of rotation. The characteristic equation of  $\mathbf{M}$  in Eq 7 is  $(\lambda - 1)(\lambda^2 - 2\lambda \cos \theta + 1)$  and the eigenvalue +1 corresponds to eigenvector  $(0, 0, 1)$ , the  $z$ -axis.

Three rotations about different axes will always suffice to transform the unit sphere, and indeed any object or scene linked to it, into any chosen orientation. This was first proved in the 18th century by the great Leonard Euler. However, there is choice of the axes about which they are made and the order in which they are taken. Order is important since in general rotations about different axes do not commute, a fact illustrated in Figure 16 and reflected in matrix multiplication not being commutative. Rotations about three orthogonal axes are called Euler rotations. Euler rotations about three axes in turn are widely used, though several variants and conventions are in operation. The order of taking the rotations might be  $XYZ$ ,  $ZXZ$ ,  $ZYX$  or another. Blender and other 3D graphics software provide several options.

### 3.2.1 Extrinsic rotations

I now discuss the method of Euler rotations about the global axes – extrinsic rotations – and explain their difference from intrinsic rotations in the next sub-section. Take the order of about the  $z$ -axis first, then about  $y$  and finally about  $x$ . The general matrix of direction

cosines is therefore built up in the stages below. Using the abbreviations

$$sa = \sin \alpha, \quad ca = \cos \alpha, \quad sb = \sin \beta, \quad cb = \cos \beta, \quad sg = \sin \gamma, \quad cg = \cos \gamma.$$

$$\begin{aligned} \mathbf{M}_x &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & ca & -sa \\ 0 & sa & ca \end{pmatrix}, \quad \mathbf{M}_y = \begin{pmatrix} cb & 0 & sb \\ 0 & 1 & 0 \\ -sb & 0 & cb \end{pmatrix}, \quad \mathbf{M}_z = \begin{pmatrix} cg & -sg & 0 \\ sg & cg & 0 \\ 0 & 0 & 1 \end{pmatrix}. \\ \mathbf{M}_y \mathbf{M}_z &= \begin{pmatrix} cb.cg & -cb.sg & sb \\ sg & cg & 0 \\ -sb.cg & sb.sg & cb \end{pmatrix}. \\ \mathbf{M}_{xyz} = \mathbf{M}_x \mathbf{M}_y \mathbf{M}_z &= \begin{pmatrix} cb.cg & -cb.sg & sb \\ sa.sb.cg + ca.sg & -sa.sb.sg + ca.cg & -sa.cb \\ -ca.sb.cg + sa.sg & ca.sb.sg + sa.cg & ca.cb \end{pmatrix}. \end{aligned} \quad (8)$$

This does indeed have one eigenvalue equal to 1 and eigenvector

$$\begin{pmatrix} sa.sb.cg + (ca - cb)sg \\ 1 - cb.cg - ca(cb - cg) + sa.sb.sg \\ ca.sb.sg - sa(cb - cg) \end{pmatrix}$$

which defines the composite axis of rotation. To be clear the order of rotations here is by  $\gamma$  about  $z$ , by  $\beta$  about  $y$  and finally by  $\alpha$  about  $x$ .

In order to work flexibly with rotating objects in a 3D graphics scene, we need to know:

1. to where three Euler rotations will take the object,
2. given a matrix of direction cosines, how to determine the extrinsic Euler rotations to achieve the orientation it describes,
3. how to specify the rotations about  $z$ ,  $y$  then  $x$  which will bring the body into a chosen orientation,

It will be helpful to work with a specific example, so take  $\gamma = \pi/4$ ,  $\beta = \pi/6$  and  $\alpha = \pi/10$ . The matrix is

$$\mathbf{M}_{xyz} = \begin{pmatrix} 0.612 & -0.612 & 0.500 \\ 0.782 & 0.563 & -0.268 \\ -0.118 & 0.555 & 0.824 \end{pmatrix}. \quad (9)$$

This answers point 1 above – just carry out the three rotations in the correct order. The result is a table of direction cosines. This has one eigenvalue of 1 corresponding to normalised eigenvector  $(0.475, 0.357, 0.805)$ , which is the axis of the composite rotation.

At this stage we can also find the angle of rotation about this axis. The trace of the direction cosine matrix (the sum of its diagonal elements) does not change under a similarity transformation which places the matrix in diagonal form. The trace equals the sum of the eigenvalues, which is  $1 + 2\cos\theta$  where  $\theta$  is the angle of rotation about the axis given by the eigenvector whose eigenvalue is 1. This is seen in  $\mathbf{M}_x$ ,  $\mathbf{M}_y$  and  $\mathbf{M}_z$  above. In this example

the trace is  $1.9993$  so  $\cos \theta \approx 1/2$  and  $\theta$  is  $60.03^\circ$ . Interesting though this is, we have not yet a means to move direct to this single axis and make the  $60^\circ$  turn. That will be examined in §3.3. Here the orientation has been arrived at by three Euler rotations about the global axes.

Point 2 involves working backwards from a matrix such as Eq 9 to determine the Euler angles. Reading  $\mathbf{M}_{xyz}$  at Eq 8,  $c_{13} = \sin \beta$ , giving  $\beta = \pi/6 = 30^\circ$  as expected. The ratio  $c_{23}/c_{33} = -\tan \alpha$ . Here  $0.268/0.824 = 0.325 = \tan(\pi/10)$  so  $\alpha$  is  $18^\circ$ . Also  $c_{12}/c_{11} = -\tan \gamma = -1$  here so  $\gamma = \pi/4 = 45^\circ$ . So the three input angles have been recovered.

Point 3 assumes that we can specify the required orientation in terms of angles from the global axes. There are nine direction cosines, but only three are needed to specify the others because the sum of squares in each row and each column is 1, and the dot product of two different rows or two different columns is zero. Having said that, the three independent values cannot be chosen wholly at random since they must satisfy these constraints. In principle it is possible to use these orthogonality relations to solve for the complete matrix and then use the analysis for point 2 above to calculate three Euler angles by which the orientation could be achieved in practice. It is yet to be decided, however, whether this is a well defined, stable process, and whether it will give a unique result. To gain insight into these questions I propose a numerical example.

Suppose, therefore, we want the body  $u$ -axis to be  $70^\circ$  from the global  $x$ -axis, and the  $w$  axis  $80^\circ$  from the  $z$ -axis and  $30^\circ$  from the  $x$ . The direction cosine matrix looks like

$$\begin{pmatrix} \cos 70^\circ & f & g \\ h & j & k \\ \cos 30^\circ & m & \cos 80^\circ \end{pmatrix} = \begin{pmatrix} 0.3420 & f & g \\ h & j & k \\ 0.8660 & m & 0.1736 \end{pmatrix}.$$

Solving for the other elements is rather like solving a sudoku puzzle – you step from cell to cell following the trail of possible values, though there will not be a single route.  $h$  and  $m$  can immediately be found from ‘sum of squares equals 1’, though both can have positive or negative signs:  $h = \pm 0.3647$ ,  $m = \pm 0.4689$ . Taking the dot product of the first two columns gives an equation in  $f$  and  $j$  from which four values of  $j$  can be found corresponding to the  $\pm$  options each for  $h$  and  $m$ . The two  $+$  signs give  $j = -0.9377f - 1.1133$ . Taking now the dot product of the first and third rows gives an equation in  $f$  and  $g$  from which  $f = -0.3703g - 0.6317$ . Now  $g$  can be found from sum of squares in row 1 and checked in column 2. Two possible values are  $g = -0.8898$  or  $g = 0.4783$ . The first of these gives  $k = \pm 0.4221$ . At this stage we have obtained a value for all nine elements of the matrix, but the test is that the determinant be 1. Taking the negative values of  $g$  and  $k$  above gives a matrix with determinant  $-1$ , though this can be reversed by reversing the signs in one row and one column. We may also try the positive values of  $h$ ,  $m$  and  $g$ . Now  $k = \pm 0.8609$  and the  $-$  sign gives a matrix with determinant  $+1$ :

$$\begin{pmatrix} 0.3420 & -0.8089 & 0.4783 \\ 0.3647 & -0.3548 & -0.8609 \\ 0.8660 & 0.4689 & 0.1736 \end{pmatrix}. \quad (11)$$

However other choices of  $h$  and  $m$  also produce matrices with determinant  $+1$ . One is

$$\begin{pmatrix} 0.3420 & 0.3022 & -0.8898 \\ 0.3647 & 0.8300 & 0.4221 \\ 0.8660 & -0.4689 & 0.1736 \end{pmatrix}. \quad (12)$$

This implies that there is not a unique set of Euler angles to achieve a required orientation. The Euler angles, applied in order  $\gamma$  about  $z$ , then  $\beta$  then  $\alpha$ , corresponding to these  $+1$  matrices are

$$\text{Eq 11 : } \alpha = 78.6^\circ, \quad \beta = 28.6^\circ, \quad \gamma = 67.1^\circ.$$

$$\text{Eq 12 : } \alpha = -67.6^\circ, \quad \beta = -62.8^\circ, \quad \gamma = -41.5^\circ.$$

Substituting these values into Eq 8 does indeed recreate the matrices at Eq 11 and 12, further supporting the view that the specified orientation can be reached by two different sets of  $XYZ$  rotations, one set of anticlockwise rotations, the other, as it happens, of clockwise ones.

To test whether this works in practice I used Blender to do a numerical check. I created two sets of axes, each axis modelled a long rod with narrow cross-section. In each set three rods were arranged at mutual right angles and coloured for ready identification as  $x$ ,  $y$ ,  $z$ , or  $u$ ,  $v$ ,  $w$ . The  $xyz$  set remained fixed as the global coordinate frame while the other acted as object and was rotated by the three Euler angles listed above for Eqs 11 and 12. For this Blender's transformation orientation option was set to 'global'. Using the MeasureIt add-on I checked the nine angles between old and new axes, the accuracy being about  $\pm 2^\circ$ . For examples, the angle  $z - w$  was  $81.5^\circ$  instead of  $80^\circ$ . The matrices of measured direction cosines corresponding to Eq 11 and Eq 12 respectively are

$$\begin{pmatrix} 0.331 & -0.825 & 0.471 \\ 0.365 & -0.332 & -0.880 \\ 0.867 & 0.483 & 0.148 \end{pmatrix}, \quad \begin{pmatrix} 0.332 & 0.266 & -0.899 \\ 0.357 & 0.826 & 0.432 \\ 0.855 & -0.471 & 0.184 \end{pmatrix}.$$

These are in fair agreement with the theory. However, as Figure 17 shows, the two orientations are quite different. I have not explored this dichotomy further, but simply presume that the three given directions cosines were not adequate to specify the orientation uniquely. The example serves as a warning to be cautious.

### 3.2.2 Intrinsic rotations

Intrinsic rotations are ones made about the local axes embedded in the object, those labelled  $u$ ,  $v$ ,  $w$  in Figure 14. These object axes start aligned with the global axes  $x$ ,  $y$ ,  $z$  respectively and so the first rotation, which we again take to be about  $z \equiv w$ , is the same in both extrinsic and intrinsic modes of working. The difference starts at the second rotation which is about the  $v$  axis in its new orientation. How are we to determine the matrix to describe this rotation? One way to picture the situation is to see that it is first necessary to unwind the previous rotations to get back to the state where  $v$  was aligned with  $y$ , then make the rotation, and finally reinstate the rotations which were undone. In the simplest case we have to undo the rotation about  $z$ , then make the rotation about  $y \equiv v$ , and then reinstate the rotation about  $z \equiv w$ . So the rotation about  $v$  (reading the order right to left) is  $\mathbf{M}_v = \mathbf{M}_z \mathbf{M}_y \mathbf{M}_z^{-1}$ . The

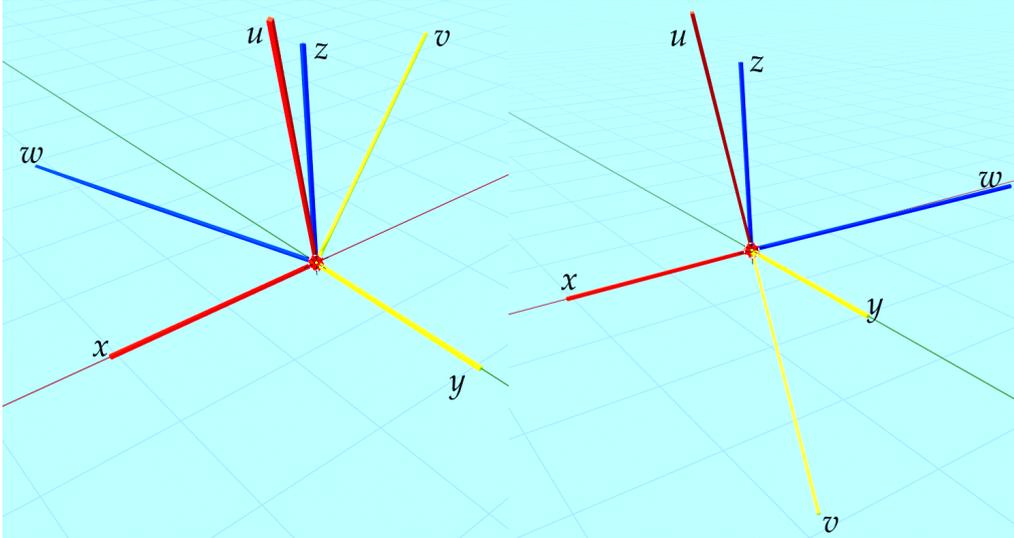


Figure 17:  $uvw$  axes after rotation by three Euler angles. Left: by Eq 11. Right: by Eq 12.

$^{-1}$  denotes the inverse operation which is simply a rotation in the opposite direction. So the combined operation of turning about  $w$  by  $\gamma$  then about  $v$  by  $\beta$  is

$$\mathbf{M}_{vw} = [\mathbf{M}_z \mathbf{M}_y \mathbf{M}_z^{-1}] \cdot \mathbf{M}_z = \mathbf{M}_z \mathbf{M}_y$$

since the forwards and inverse rotations about  $z$  cancelled. This is very similar to the extrinsic rotation except that the order of the matrix multiplication has been reversed. The combined operation for three rotations is

$$\mathbf{M}_{uvw} = [\mathbf{M}_{vw} \mathbf{M}_x \mathbf{M}_{vw}^{-1}] \cdot \mathbf{M}_{vw} = \mathbf{M}_{vw} \mathbf{M}_x = \mathbf{M}_z \mathbf{M}_y \mathbf{M}_x.$$

I find it remarkably elegant that rotations about global and local object axes use the same three basic matrices and differ only in the order, forwards or reverse, in which they are multiplied.

Blender has a comprehensive array of options for rotating objects, including six orders for Euler angles. Intrinsic rotations about local axes operate on a gadget called ‘the gimbal’. This is a virtual apparatus consisting of three mutually perpendicular axes nested one inside the other with the object attached to the innermost. Using the language of aviation, rotations are said to be by ‘roll, pitch and yaw’. The literature warns of a phenomenon called ‘gimbal lock’ which occurs when the second rotation is through  $90^\circ$ . When this happens, two rotation axes have become aligned and lost their independence. In the matrix one row and one column at Eq 8 become  $(1\ 0\ 0)$ ,  $(0\ 1\ 0)$  or  $(0\ 0\ 1)$ , indicating that one axis of rotation has become neutralised. An object in this situation cannot be further manipulated into a general orientation, but rather is locked into rotating in one plane only. The phenomenon is intrinsic to Euler angles so when animating an object, the artist must think of the expected motion at each joint and choose an Euler sequence in which gimbal lock is unlikely to occur. It is reported on the internet that gimbal lock was a concern during the USA Apollo space missions because the spacecraft used a system of mechanical gimbals to stabilize it during its journeys to and from the Moon. NASA engineers were obliged to develop techniques to avoid gimbal lock by using a different set of gimbals for different phases of the mission.

### 3.3 ‘Axis & angle’ rotations

We return to the more fundamental concept of a rotation as a turn  $0 \leq \theta < 2\pi$  or  $\pi < \theta < \pi$  about a single axis. The challenge is to navigate to that axis and make the correct turn. I will start with a geometrical construction which goes a long way to explaining what is happening behind the algebra. This construction compounds two successive rotations into a single rotation. It was discovered by the Frenchman Olinde Rodrigues in 1840, but such is the unfairness of history that it is known as Euler’s construction and Rodrigues is almost forgotten. Like Cayley the commercial lawyer in England, his main occupation for not as a professional mathematician, but as a successful banker and financier.

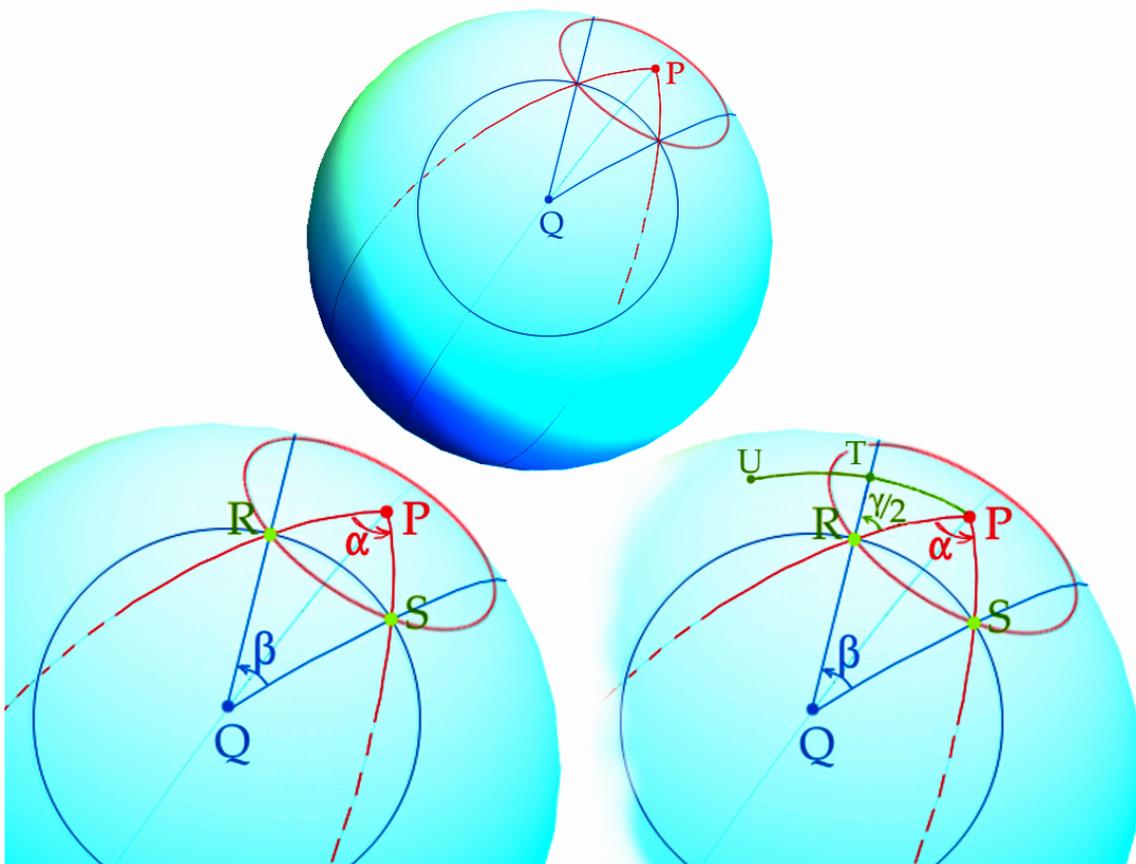


Figure 18: Rodrigues’ geometrical construction for compounding two rotations.

Rodrigue’s argument is illustrated in three stages in Figure 18. The unit sphere is set within a fixed, global coordinate frame  $xyz$  (not shown) to which the axes of rotation are referred. The body being rotated is represented by the surface of the sphere whose centre remains fixed. We will compound a rotation first about an axis through point P and its antipodal point, and then one through Q. P and Q are at arbitrary positions. It is necessary to distinguish between P and Q as poles and points P’ and Q’ fixed in the surface and initially directly under P and Q. P’ and Q’ will turn as the sphere is rotated, but P and Q remain fixed in space. The stages are:

1. Choose two points, P and Q and, using them as centres, draw two small circles on the sphere, letting them intersect at two points as in the top diagram of Figure 18.

2. In the bottom left diagram these intersections are labelled R and S. The great circle through PR is at angle  $\alpha$  to the great circle through S, with S being anticlockwise of R as seen from P. Similarly the great circles through Q as pole which pass through R and S are  $\beta$  apart, with R anticlockwise of Q as seen from Q. The great circle PQ bisects both  $\alpha$  and  $\beta$ .
3. Take the first rotation to be about P through angle  $\alpha$ . This will carry R to the position of S. The second rotation is about Q by  $\beta$ . This will carry S back to R. So the combined effect of the two rotations is to leave R unmoved. Since only the pole (plus internal axis) does not move in a rotation, R must be the pole of the combined rotation  $\beta \cdot \alpha$ ,  $\alpha$  first,  $\beta$  second.
4. Refer to the bottom right panel and consider what happens to the point P' directly underneath P under these combined rotations. Rotation about P has no effect, then rotation about Q through only  $\beta/2$  turns P' to T on the great circle QR. A further turn through  $\beta/2$  brings P' to U. The combined action of  $\alpha$  then  $\beta$  has been to turn points about R through angle  $\gamma$  where  $\gamma/2$  is the angle PRT in the spherical triangle shown.

Altmann reports that this elegant construction went unnoticed. Hamilton was unaware of it when he invented quaternions. Of particular importance is the way the half angles  $\alpha/2$ ,  $\beta/2$ ,  $\gamma/2$  appear in the diagram.

Using spherical trigonometry  $\gamma$  can be calculated in terms of  $\alpha$  and  $\beta$ , and also the unit position vector  $\mathbf{r}$  to R can be found in terms of the unit vectors from O to P and to Q. In a companion article on [www.mathstudio.co.uk](http://www.mathstudio.co.uk) I discuss the basic trigonometry of spherical triangles and prove the results quoted below. In a planar triangle the internal angles add to  $\pi$  radians and the familiar Sine and Cosine formulae give relations between the side lengths,  $a$ ,  $b$ ,  $c$  and their opposite angles,  $A$ ,  $B$ ,  $C$ :

$$\cos C = \frac{a^2 + b^2 - c^2}{2ab}, \quad c^2 = a^2 + b^2 - 2ab \cos C, \quad \frac{\sin A}{a} = \frac{\sin B}{b} = \frac{\sin C}{c}.$$

Spherical geometry has little in common with Euclidean and the relations are distinctly different. A distance on the surface of the unit sphere is the arc length  $d$  along a great circle between its end points, and is numerically equal to the angle in radians subtended at O by the arc. A triangle is made of arcs of three great circles<sup>1</sup>. The sum of the internal angle is greater than  $\pi$ , and the area of the triangle is equal to the 'angular excess' = (sum of internal angles  $-\pi$ ). There are two Cosine rules for angle and side lengths:

$$\cos C = \sin A \sin B \cos c - \cos A \cos B, \quad \cos c = \cos a \cos b + \sin a \sin b \cos C. \quad (13a)$$

The Sine rule is

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C}, \quad (13b)$$

---

<sup>1</sup> This is called a 'strict' or 'authentic' triangle to distinguish it from ones in which one or more sides is a little circle – one which does not have O as its centre.

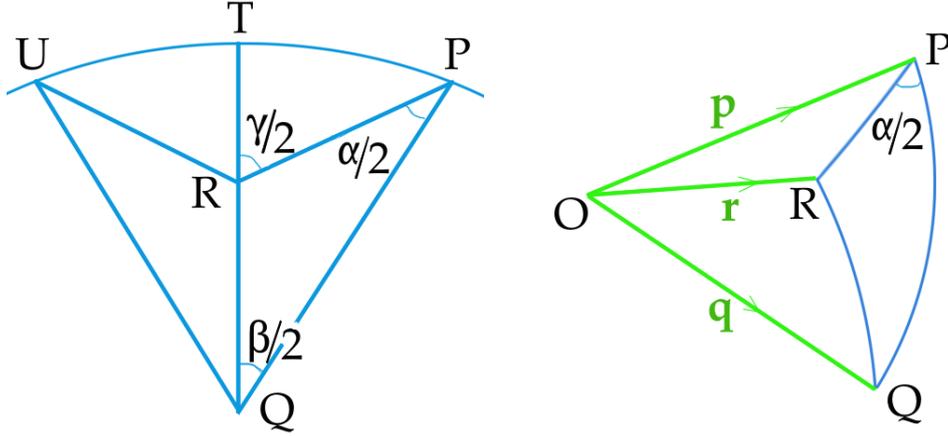


Figure 19: Left: The cosine rule applied to spherical triangle PQR. Right: position vectors of P, Q, R from the centre of the unit sphere.

To determine the rotation we need both the axis and the angle of turn. The spherical Cosine rule above applied to the triangle PQR in Figure 18 will give  $\gamma$ . The left panel in Figure 19 shows triangle PQU. The arc length PQ is given by the dot (scalar) product of unit position vectors  $\mathbf{p}$  and  $\mathbf{q}$ :

$$\cos PQ = \mathbf{p} \cdot \mathbf{q} = p_x q_x + p_y q_y + p_z q_z.$$

Using  $\cos(\pi - \gamma/2) = -\cos(\gamma/2)$ ,

$$\cos \frac{\gamma}{2} = -\sin \frac{\alpha}{2} \cdot \sin \frac{\beta}{2} \cos(PQ) + \cos \frac{\alpha}{2} \cdot \cos \frac{\beta}{2}. \quad (14)$$

Since  $\alpha$ ,  $\beta$  and the distance PQ are all known,  $\gamma$  has been found. Note that  $\alpha$  and  $\beta$  are interchangeable so the order of rotation does not feature.

I will illustrate these calculations with a numerical example based on the two rotations of the letter ‘t’ in Figure 16. In the top panel the first rotation is  $+90^\circ$  about the  $x$ -axis, and the second  $+90^\circ$  about the global  $z$ -axis. So  $\alpha = \pi/2 = \beta$  and arc  $PQ = \pi/2$ . The lower panel in Figure 16 has the order reversed, but in both cases  $\cos(\gamma/2) = \cos^2(\pi/4) = 1/2$  so  $\gamma = +2\pi/3$  radians,  $120^\circ$ .

We now look into determining the unit position vector  $\mathbf{r}$ . This is illustrated with a numerical example in Appendix 2. One approach is to set up and solve two simultaneous equations for components  $r_x$ ,  $r_y$ , and use  $r_x^2 + r_y^2 = 1 - r_z^2$ .  $\mathbf{p} \times \mathbf{r}$  is normal to the plane OPR, and  $\mathbf{p} \times \mathbf{q}$  normal to OPQ, so their scalar product gives angle  $\alpha/2$ . Similarly  $\beta/2$  can be found, and these relations are enough to solve for  $r_x$  and  $r_y$ . Specifically

$$\frac{\mathbf{p} \times \mathbf{r}}{|\mathbf{p} \times \mathbf{r}|} \cdot \frac{\mathbf{p} \times \mathbf{q}}{|\mathbf{p} \times \mathbf{q}|} = \cos \frac{\alpha}{2}, \quad \frac{\mathbf{q} \times \mathbf{r}}{|\mathbf{q} \times \mathbf{r}|} \cdot \frac{\mathbf{p} \times \mathbf{q}}{|\mathbf{p} \times \mathbf{q}|} = \cos \frac{\beta}{2}. \quad (15)$$

More detail is given in Appendix 2. The square roots involved in the norms of these vectors make the equations highly non-linear. Algebraic calculation is cumbersome with nested square

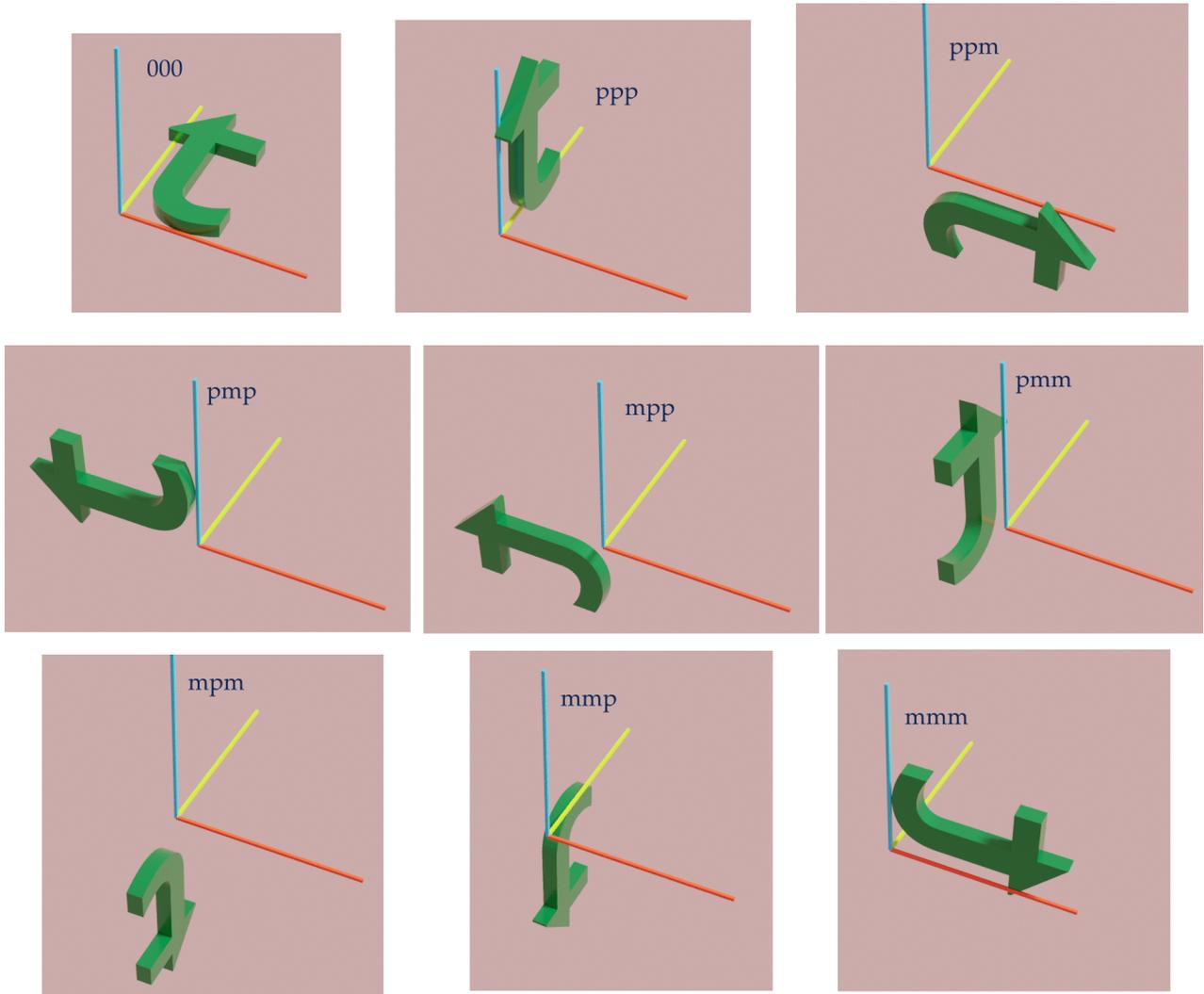


Figure 20: Original figure ‘t’ and eight orientations produced by rotation through  $120^\circ$  about  $(1,1,1)$ -types axes. ‘p’ means + sign and ‘m’ a – sign for the respective component  $r_x, r_y, r_z$ .

roots so numerical solution is usually necessary. In fact this method works almost too well because the + and – signs associated with each square root allow  $r_x, r_y$  and  $r_z$  to be take either sign and so produce  $2^3$  options, all illustrated in Figure 20. In each the rotation is  $+120^\circ$ . The solution for  $\mathbf{r} = (\pm 1/\sqrt{3}, \pm 1/\sqrt{3}, \pm 1/\sqrt{3})$ . The notation in Figure 20 gives the signs of  $r_x, r_y, r_z$  in that order. The top panel of Figure 16 is obtained by  $\text{ppp} \equiv (+1/\sqrt{3}, +1/\sqrt{3}, +1/\sqrt{3})$  and the lower with  $\text{pmp}$ . The intermediate orientations in Figure 16, labelled  $R_x90$  and  $R_z90$ , do not appear in Figure 20 because they involve only a single rotation.

The above method for finding  $\mathbf{r}$ , therefore, involves an additional stage of identifying the correct signs. It is such alternative solutions to the pair of simultaneous equations which correspond to taking Euler rotations in a different order.

There is at least one other method for calculating  $\mathbf{r}$  from  $\mathbf{p}, \mathbf{q}$  and the arc between them, and that it to use the maths of reciprocal vectors to express  $\mathbf{r}$  as a weighted sum of  $\mathbf{p}, \mathbf{q}$

$\mathbf{q}$  and  $\mathbf{p} \times \mathbf{q}$ , a result worked out by Rodrigues. Details of this involved calculation are given in Appendix 2, illustrated by a numerical example. The method leads to a crucial formula for rotations, closely linked to the quaternions. It is the preferred method because it does not have the square root ambiguities of simultaneous equations and is computationally stable and efficient. With the notation in Figures 18 and 19 the formula is

$$\sin R \mathbf{r} = \sin P \cos Q \mathbf{p} + \cos P \sin Q \mathbf{q} + \sin P \sin Q (\mathbf{p} \times \mathbf{q}). \quad (16a)$$

In terms of the rotation angles

$$\sin \frac{\gamma}{2} \mathbf{r} = \sin \frac{\alpha}{2} \cos \frac{\beta}{2} \mathbf{p} + \cos \frac{\alpha}{2} \sin \frac{\beta}{2} \mathbf{q} + \sin \frac{\alpha}{2} \sin \frac{\beta}{2} (\mathbf{p} \times \mathbf{q}). \quad (16b)$$

This and the Cosine rule at Eq 14 are the two formulae closely related to the algebra of quaternions considered at briefly in the next subsection.

We have here a formula for combining two rotations without giving a mathematical expression for either of the single rotations involved beyond stating that the first is about axis  $\mathbf{p}$  through angle  $\alpha$  and the second about axis  $\mathbf{q}$  through angle  $\beta$ . So now introduce the notation that  $\Omega(\psi, \mathbf{u})$  is a rotation by  $+\psi$  about axis with unit vector  $\mathbf{u}$ , and that two rotations combine by

$$\begin{aligned} \Omega(\psi_2, \mathbf{u}_2) \cdot \Omega(\psi_1, \mathbf{u}_1) &= \Omega(\psi_3, \mathbf{u}_3) \\ \text{where } \cos \frac{\psi_3}{2} &= \cos \frac{\psi_1}{2} \cos \frac{\psi_2}{2} - \sin \frac{\psi_1}{2} \sin \frac{\psi_2}{2} (\mathbf{u}_2 \cdot \mathbf{u}_1), \\ \sin \frac{\psi_3}{2} \mathbf{u}_3 &= \cos \frac{\psi_2}{2} \sin \frac{\psi_1}{2} \mathbf{u}_1 + \cos \frac{\psi_1}{2} \sin \frac{\psi_2}{2} \mathbf{u}_2 + \sin \frac{\psi_2}{2} \sin \frac{\psi_1}{2} (\mathbf{u}_2 \times \mathbf{u}_1). \end{aligned} \quad (17)$$

We can discover from this more about the form of a single rotation by letting one of the two be the empty rotation  $(1, 0, 0, 0)$ . Then  $\Omega(\psi, \mathbf{u})$  is the pair  $(\cos(\psi/2), \sin(\psi/2) \mathbf{u})$ . Notice that the form of the product is the same as the form of a single rotation – a scalar part and a 3-vector part. In addition, each must have a type of norm defined such that its value is  $+1$ , so that a pure rotation takes place without scaling or reflection. Having the same form means that rotations can be chained innumerable times.

### 3.4 Rotations using quaternions

A quaternion is a 4-dimensional number generalising the complex numbers and with similar rules of addition and multiplication. They were invented by William Hamilton in 1843 and, after finding little application for a century, have become important for quickly calculating rotations. They have the distinct advantage of not suffering from gimbal lock. Below I note their general form and the rules of arithmetic and give the procedure for using them to rotate a vector and to compound two rotations into one.

I will not go into the mathematics of quaternions deeply though Appendix 3 gives more detail and derives their multiplication formula. Their form is  $q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$  where  $\mathbf{i}, \mathbf{j}, \mathbf{k}$  are a basis of mutually orthogonal unit vectors and  $a, b, c, d$  are real scalar coefficients.  $q$  is often notated simply by its four components  $(a, b, c, d)$ , and it is sometimes useful to

think of  $q$  as the sum of a scalar  $\lambda$  and a 3-vector<sup>2</sup>,  $\lambda + \mathbb{V}$ . Addition is component-wise:  $q_1 + q_2 = (a_1 + a_2) + (b_1 + b_2)\mathbf{i} + (c_1 + c_2)\mathbf{j} + (d_1 + d_2)\mathbf{k}$ . Multiplication is governed by relations analogous to  $i^2 = -1$  in the complex numbers:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1, \quad \mathbf{ijk} = \mathbf{jki} = \mathbf{kij} = -1, \quad \mathbf{ij} = -\mathbf{ji} = -\mathbf{k}, \quad \mathbf{jk} = \mathbf{i} \quad (18)$$

and similar. The anti-commuting property is unfamiliar to many people and may make quaternions seem rather mysterious. Appendix 3 explains their multiplication and shows how the symmetries of  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  can be represented by 2-by-2 matrices or by permutations.

To see how quaternions can be used for rotations, we state that the aim is to obtain an expression where the rotation  $\Omega$  by angle  $+\psi$  about axis  $\mathbf{u}$  can be written as  $\Omega(\psi, \mathbf{u})$  and that two such rotations in succession can be combined as  $\Omega(\psi_2, \mathbf{u}_2).\Omega(\psi_1, \mathbf{u}_1) = \Omega(\psi_3, \mathbf{u}_3)$ . To move towards this objective, cast Eq 14 and 18 into a more compact form by introducing  $\lambda$  and  $\mathbb{V}$

$$\lambda_j = \cos \frac{\psi_j}{2}, \quad \mathbb{V}_j = \sin \frac{\psi_j}{2} \mathbf{u}_j. \quad (19)$$

Then the Cosine rule for  $\psi$  at Eq 14 and the expression at Eq 16b for  $\mathbf{r}$  become

$$\lambda_3 = \lambda_1\lambda_2 - \mathbb{V}_1 \cdot \mathbb{V}_2, \quad \mathbb{V}_3 = \lambda_2\mathbb{V}_1 + \lambda_1\mathbb{V}_2 + \mathbb{V}_2 \times \mathbb{V}_1. \quad (20a)$$

It happens that if  $\Omega(\psi, \mathbf{u})$  is seen as the quaternion  $\lambda + \mathbb{V}$ , quaternion multiplication exactly fits the bill. This is proved by working through the dot and cross vector multiplications in components form, as at the end of Appendix 3. So two rotations are compounded by multiplying the corresponding quaternions in the correct order:

$$\Omega(\lambda_2, \mathbb{V}_2).\Omega(\lambda_1, \mathbb{V}_1) = \Omega(\lambda_1\lambda_2 - \mathbb{V}_1 \cdot \mathbb{V}_2, \lambda_2\mathbb{V}_1 + \lambda_1\mathbb{V}_2 + \mathbb{V}_2 \times \mathbb{V}_1). \quad (20b)$$

It remains to explain how a given general non-unit position vector  $\mathbf{a}$  will be rotated using quaternions. How does the quaternion  $\Omega$  operate on  $\mathbf{a}$ ? Many articles in books and on the internet give the answer to this without deriving the formula or even giving a loose explanation. My own rationale runs as follows. The first step must be to convert  $\mathbf{a}$  into a quaternion so that it and  $\Omega$  can be non-commutatively multiplied. So identify  $\mathbf{a}$  with the quaternion  $(0 + \mathbf{a}) = (0, a_x, a_y, a_z)$ . Now we can try simple multiplication and see if we get another quaternion of the form  $(0 + \mathbf{a}')$  which would allow us to say that the rotated vector is  $\mathbf{a}'$ . It is immediately clear, however, that whether we choose  $(\cos(\psi/2) + \sin(\psi/2)\mathbf{u})(0 + \mathbf{a})$  or  $(0 + \mathbf{a})(\cos(\psi/2) + \sin(\psi/2)\mathbf{u})$ , the leading, scalar component will be  $-\sin(\psi/2)(\mathbf{a} \cdot \mathbf{u})$  which is not in general zero. So something more sophisticated is needed.

Let us stand back from the detail and consider the nature of a rotation. Like all transformations, it amounts to looking at the same object from a different point of view. There is a strong similarity between  $\mathbf{a}$  and its rotated self,  $\mathbf{a}'$ . The two are in the same ‘family’. This concept appears in group theory where structurally similar elements are said to be ‘conjugates’. The term is applied to complex numbers which differ only in the sign of

---

<sup>2</sup>  $\mathbb{V}$  is not an ordinary vector (‘polar vector’) such as a position vector, but an ‘axial vector’ like the cross product. The subtle difference is in the way it transforms under a change of axes. It is really an antisymmetric second rank tensor and is recognised as such because a right-handed sense must also be specified in defining the cross product.

their imaginary parts, and indeed the conjugate of a quaternion has the sign of its vector part swapped. In all these situations if  $\Omega$  is an operation and  $\mathbf{a}$ ,  $\mathbf{a}'$  are two elements, they are conjugates if  $\Omega\mathbf{a} = \mathbf{a}'\Omega$ . To press the point further, look again at Figure 18 in which a rotation by  $\alpha$  about P has carried a point at R to S, then a rotation about Q through  $\beta$  has carried it back to R. We feel that R and S are conjugate points; they lie equally either side of the arc PQ and their roles would be reversed if the rotation about Q had been performed first. Pressing even further, the form  $(0 + \mathbf{a})$  looks like a rotation (by  $\pi$ , and with a stretch if  $|\mathbf{a}| \neq 1$ ), making  $\mathbf{a}$  and  $\Omega$  resemble rotations about axes through P and Q and their conjugate points R and S.

This line of thinking suggests that we examine  $\mathbf{a}'$  formed such that  $\Omega\mathbf{a} = \mathbf{a}'\Omega$ . Applying the inverse quaternion  $\Omega^* = (\cos(\psi/2) - \sin(\psi/2)\mathbf{u})$  isolates  $\mathbf{a}'$ .

$$\mathbf{a}' = \Omega\mathbf{a}\Omega^* = \Omega(\mathbb{V} \cdot \mathbf{a}, \lambda\mathbf{a} - \mathbf{a} \times \mathbb{V}) = (0, (\lambda^2 - |\mathbb{V}|^2)\mathbf{a} + 2\lambda(\mathbb{V} \times \mathbf{a}) + 2(\mathbb{V} \cdot \mathbf{a})\mathbb{V}). \quad (21a)$$

In this the scalar part was  $\lambda(\mathbb{V} \cdot \mathbf{a}) - \mathbb{V} \cdot (\lambda\mathbf{a} - \mathbf{a} \times \mathbb{V})$ , all of which cancels to zero. In the vector part I have used  $\mathbb{V} \times (\mathbf{a} \times \mathbb{V}) = (\mathbb{V} \cdot \mathbb{V})\mathbf{a} - (\mathbb{V} \cdot \mathbf{a})\mathbb{V}$ . Since  $\lambda^2 = \cos^2(\psi/2)$  and  $|\mathbb{V}|^2 = \sin^2(\psi/2)$ ,  $\lambda^2 - |\mathbb{V}|^2 = \cos\psi$ . Therefore

$$\mathbf{a}' = (0, \cos\psi \mathbf{a} + 2(\mathbb{V} \cdot \mathbf{a})\mathbb{V} + 2\lambda(\mathbb{V} \times \mathbf{a})), \quad \mathbb{V} = \sin\frac{\psi}{2} \mathbf{u}. \quad (21b)$$

With zero scalar part this clearly has the correct structure for the vector part to be  $\mathbf{a}'$ .

It will be comforting to check this with a few simple cases, so return to the  $\mathbf{t}$ -shape rotated in Figure 16 and 18. For the top panel in Figure 16 90° rotation about the  $x$ -axis is described by quaternion  $(1/\sqrt{2}, 1/\sqrt{2}, 0, 0)$ . The surface of the  $\mathbf{t}$  initially points upwards along  $z$  so corresponds with quaternion  $(0, 0, 0, 1)$ . The rotation by Eq 21b bring this to  $(0, 0, -1, 0)$  and this is correct because the face now points along the  $-y$  direction. The second rotation about  $z$  requires quaternion  $(1/\sqrt{2}, 0, 0, 1/\sqrt{2})$  and this takes  $(0, 0, -1, 0)$  to  $(0, 1, 0, 0)$ . Again this agrees with the figure as the surface of the  $\mathbf{t}$  is now facing the  $x$  direction. Applying rotations in the opposite order does bring the  $\mathbf{t}$  surface to face the  $-y$  direction, as in the lower panel of Figure 16. Similar checks with other axes add conviction that Eq 21b is correct.

Eq 21b is in fact the most computationally efficient way of rotating a vector. It is unambiguous, with no  $\pm$  signs to choose between. It does not suffer from gimbal lock. It is computationally stable and fast, requiring only 15 multiplications. Compare this with the three matrix multiplications when Euler angles are used. Eq 21b is the most important formula in this section.

## 4 Non-linear projections: Airy's perspective

In §1 I pointed out that making a drawing of a scene is not unlike making a map of a globe, the observer sitting at the centre of the globe and looking radially outwards in all directions. This way of thinking of the art of drawing acknowledges that the artist and viewer both will turn their heads to look to left and right, up and down. As they do so the vanishing point

of 1-point perspective changes with each eye movement, and the artist is faced with trying to represent some aggregated view on the picture plane. Clearly there is no single way to do this, but when computer graphics are programmed with an algorithm for plotting images over a wide solid angle of view, appeal can be made to one of the many map projections used by cartographers. The globe around the artist is likened to the surface of the Earth.

It is impossible to map all of a sphere onto a plane. Moreover, all maps will suffer to some degree from two types of error:

1. unequal areas: two patches of ground on the Earth at different distances from the agreed point of reference (the centre of the map) will in the map be represented by patches which have a different ratio of areas. In other words, area is not faithfully represented across the mapped region.
2. distortion: the angle between two intersecting lines on the Earth is not equal to the angle between their representations on the map. This is equivalent to a rectangle on the earth mapping to a more general quadrilateral, and to a circle on the earth mapping to an ellipse or oval.

These distortions are usually independent. There are dozens of map projections in use<sup>3</sup>. Some achieve equal areas but suffer serious distortion, while ones like the stereographic projection are conformal (map angles between lines correctly), but have large errors in length and area.

Here I discuss the projection devised by the Victorian Astronomer Royal, Sir George Airy, published in 1861<sup>4</sup>. He set about creating an algorithm for mapping a large solid angle of a sphere onto a plane with minimum overall misrepresentation (which he called ‘evil’). He decided on what in cartographic language it is a gnomonic azimuthal projection. Azimuthal means that the Earth is projected onto a plane by rays from a single point. A polar point  $P$  is chosen on the sphere to be the centre of reference. Great circles on the sphere passing through  $P$  then plot as straight lines on the map, all passing radially through the centre of the map,  $P'$ . The error does not vary with azimuthal angle  $\phi$  (as around a clock face). Gnomonic means that the focus of rays from the sphere is at the centre of the sphere, equivalent to where the artist’s eye would be.

Airy gave a mathematical definition to the local error in areas,  $\epsilon_A$  and in distortion,  $\epsilon_D$ , and combined the sum of their squares  $\epsilon_A^2 + \epsilon_D^2$  at each elementary portion of the map. Call this sum of squares  $Q(\theta) > 0$  where  $\theta$  is the arc length in radians around the sphere from its reference centre to the position in question. (The function will not depend on azimuth  $\phi$ .) He then added these elementary  $Q(\theta)$  by integration over the map out to  $\theta_{max}$  to give a number which quantifies the overall quality of the map. The challenge then was to find the form of  $Q(\theta)$  so that this integral is a minimum. This was a problem for the calculus of variations. Airy called his projection the ‘balance of errors’ because it probably achieves the best possible balance between unequal areas and angular distortion. I now outline his derivation and state the resulting mapping function.

---

<sup>3</sup> See for instance [www.map-projections.net](http://www.map-projections.net) and [www.mapmathematics.com](http://www.mapmathematics.com).

<sup>4</sup> Philosophical Magazine, Dec 1861, Vol 22 No 149, p 409.

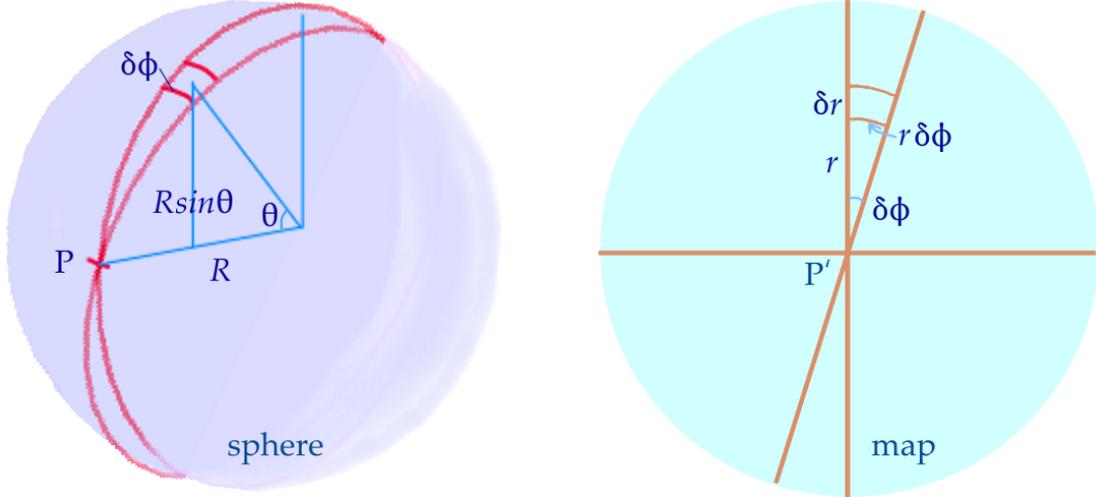


Figure 21: Mapping of element of area from sphere, radius  $R$ , (left) to flat map (right).

The first stage is to determine  $\epsilon_A(\theta)$ ,  $\epsilon_D(\theta)$  and hence  $Q(\theta)$ . Following Airy, we do this symbolically first and later convert to actual distances on the sphere and map. Figure 21 shows the sphere and the map. Consider a curvilinear rectangle on the sphere with length  $a$  along a great circle through the central point of reference, and let  $b$  be the rectangle's azimuthal extent. This element maps to  $a' \times b' = (a + \delta_a)(b + \delta_b)$ . Airy takes the fractional error in area

$$\epsilon_A = \frac{a'b' - ab}{ab} = \frac{(a + \delta_a)(b + \delta_b) - ab}{ab} = \frac{\delta_a}{a} + \frac{\delta_b}{b}.$$

He measures the distortion as

$$\epsilon_D = \frac{a'/b' - a/b}{a/b} = \frac{(a + \delta_a)b - a(b + \delta_b)}{a(b + \delta_b)} \approx \frac{\delta_a}{a} - \frac{\delta_b}{b}.$$

$$\text{So } Q(\theta) = \frac{1}{2}(\epsilon_A^2 + \epsilon_D^2) \rightarrow \left(\frac{\delta_a}{a}\right)^2 + \left(\frac{\delta_b}{b}\right)^2 \quad (22)$$

where an unimportant factor of 2 has been dropped.

$Q$  cannot be used directly in this form, but must be converted to quantities appropriate to the sphere and its map. Figure 21 shows the sphere with radius  $R$ . We decide upon a scale  $s$  for the map, choosing it so that an arc subtended by  $\theta = 1$  radian will map to 1 length unit radially in the map. So  $sR \rightarrow r = 1$ . Now identify the quantities  $a$ ,  $a'$ ,  $\delta a$  etc. used symbolically above.

$$a = sR \delta\theta = \delta\theta, \quad a' = \delta r, \quad \delta a = \delta r - \delta\theta,$$

$$b = sR \sin\theta \delta\phi = \sin\theta \delta\phi, \quad b' = r \delta\phi, \quad \delta b = (r - \sin\theta) \delta\phi.$$

$$\text{Then } \left(\frac{\delta_a}{a}\right) = \frac{\delta r}{\delta\theta} - 1, \quad \left(\frac{\delta_b}{b}\right) = \frac{r}{\sin\theta} - 1$$

$$\text{and } Q(\theta) = \left(\frac{\delta r}{\delta \theta} - 1\right)^2 + \left(\frac{r}{\sin \theta} - 1\right)^2. \quad (23)$$

The is the misrepresentation (‘evil’) in the rectangle with area  $\sin \theta \delta \theta \delta \phi$ , so the overall quality of map is given by

$$2\pi \int_0^{\theta_{max}} Q(\theta) \sin \theta d\theta \quad (24)$$

the  $2\pi$  coming from integration around  $\phi$ . The challenge now – and it is a big challenge – is to find the function  $r(\theta)$  which will minimize this integral.

In principle this can be solved using the calculus of variations. This branch of calculus is a generalisation of the calculus problem of finding the maximum, minimum or other stationary points of a function of a single variable  $f(x)$ . From school maths we know this condition is where  $f'(x) = 0$ . The equivalent of  $f'(x) = 0$  for a ‘functional’  $F$  over a given interval  $(p, q)$  is the Euler-Lagrange differential equation. The relevant theorem states that if the function  $r(\theta)$  makes the integral  $\int_p^q F(r, dr/d\theta, \theta) d\theta$  stationary, then  $r(\theta)$  must satisfy the second order differential equation

$$\frac{d}{d\theta} \left( \frac{\partial F}{\partial R'} \right) - \left( \frac{\partial F}{\partial r} \right) = 0, \quad R' = \frac{dr}{d\theta}. \quad (25)$$

The derivatives are taken as if  $r$ ,  $R'$  and  $\theta$  were independent. In Airy’s case  $F$  is  $Q(r, R', \theta) \sin \theta$  and involves all of  $r$ ,  $R'$  and  $\theta$  explicitly.

In an impressive feat of mathematics running over two pages of his 1861 paper Sir George determines that the formula for his map projection is

$$r(\theta) = \tan \frac{\theta}{2} + 2 \cot \frac{\theta}{2} \ln \sec \frac{\theta}{2} = \tan \frac{\theta}{2} - \frac{2 \ln \cos \frac{\theta}{2}}{\tan \frac{\theta}{2}}. \quad (26)$$

It is plotted in Figure 22 out to  $\theta = 3\pi/4 = 135^\circ$ .

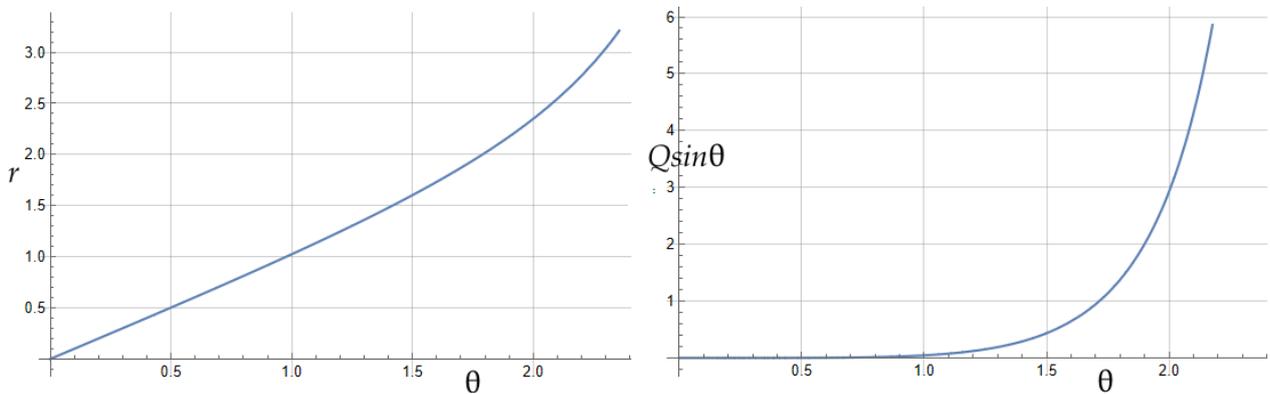


Figure 22: Airy’s Balance of Errors function  $r(\theta)$ .  $\theta$  in radians. Left:  $r(\theta)$ , Right:  $Q(\theta) \sin \theta$ .

Some comments are:

- Being a second order equation, the Euler-Lagrange equation integrates with two disposable constants. Airy chose these so that at  $\theta = 0$  both  $r = \theta$  and  $R'(0) = 0$ , conditions which make the map faithful immediately around the pole  $P$ ,  $P'$ .

- I explicitly stated a scale factor  $s$  so that, with a sphere of radius  $R$ ,  $sR = 1$ , meaning that  $\theta = 1$  radian would map to  $r = 1$  unit. Airy does not state a scale factor explicitly, but he works with the expressions 9 to produce Eq 26. This has  $r(1) = 1.024$  units, and  $r = 1$  unit at  $\theta = 0.977$  radians. I cannot account for this discrepancy in scale and Airy does not mention it. It would not make a practical difference on an actual map.
- Though the integration of  $Q \sin \theta$  is from 0 to an unspecified  $\theta_{max}$ , the two disposable constants have already been allocated so  $\theta_{max}$  does not feature in the expression for  $r$ . This would imply that  $r(\theta)$  in Eq 11 gives the best balance of errors for any  $\theta_{max}$ , and this in turn implies that Eq 26 give the optimum map quality in each finite ring of the map from any  $\theta_1$  to any  $\theta_2$ . This conjecture is open to numerical corroboration, which I now address.

The meaning of “ $r(\theta)$  minimises the integral at Eq 24” is that, if any small function  $h(\theta)$  is added to  $r(\theta)$ , the integral will increase. The only condition is that  $h(\theta)$  be zero at the limits 0 and  $\theta_{max}$ . Therefore, if we add any arbitrary  $h(\theta)$  to Airy’s  $r(\theta)$ , the value of the integral should increase. Suitable test functions could be a parabola or a sine curve with several zeros within the interval  $[0, \theta_{max}]$ . Table 2 gives some examples all of which support  $r(\theta)$  being optimum.

$h(\theta) \setminus \theta_{max}$	1	$\pi/2$	$3\pi/4$
0	0.007260	0.13325	3.0601
$-0.005 \sin \pi\theta$	0.007347		
$+0.01 \sin \pi\theta$	0.007608		
$-0.002 \sin 4\theta$		0.13329	
$+0.005 \sin 4\theta$		0.13349	
$-0.01 \sin 4\theta$			3.0617
$+0.01 \sin 4\theta$			3.0617

Table 2: Values of  $\int_0^{\theta_{max}} Q(\theta) \sin \theta d\theta$  for Airy’s  $r(\theta)$  and  $r(\theta) + h(\theta)$  showing lowest value for  $h = 0$ .

To use Airy non-linear perspective in a wide-angle graphic it is necessary to calculate the angle  $\theta$  by which each point,  $K$ , in space is turned from the  $z$  axis. The  $z$  axis is the forwards line of sight and determines the pole  $P, P'$  of the scene and its map. Let  $K$  have coordinates  $(x, y, z)$ . There are two formulae by which  $\theta$  can be calculated:

1.  $\tan \theta = \sqrt{(x^2 + y^2)}/z$ .
2. using the dot product of vectors,  $\cos \theta = z/\sqrt{(x^2 + y^2 + z^2)}$ .

The plotted point in the picture plane is then  $(r \cos \phi = rx/\sqrt{x^2 + y^2}, r \sin \phi = ry/\sqrt{x^2 + y^2})$ .

I know of no published software which can present a scene in Airy perspective so have written code myself to produce at least a wire-frame picture. Figure 23 shows the parallel rows of six arches and central circular floor insets as in Figure 9. This looks ‘fish-eye’ and unnatural when looked at on the page. However, Figures 9, 10 and 23 were created with the picture plane 1 distance unit from the viewer on a scale by which the circles are 8 units diameter. So if one were to view the image from, say 1 metre way, the image should be magnified to the

size of a wall, such is the solid angle it contains. By design Airy's perspective is less distorted than linear perspective. In the late 1970s I painted a fantasy picture using Airy perspective. It is on the painting page of [www.mathstudio.co.uk](http://www.mathstudio.co.uk) and shows three female figures standing or lying at right-angles to each other in a garden with tiled paving.

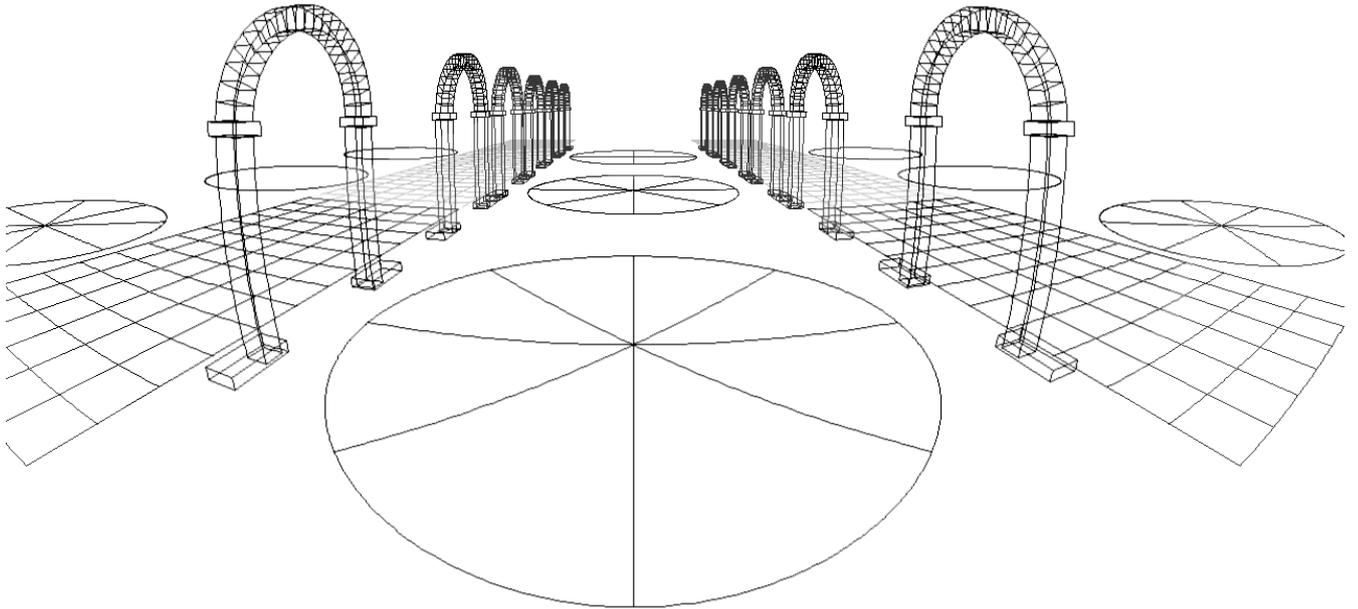


Figure 23: The view in Figure 9 in Airy perspective, with circles to the side similar to Figure 10.

## Appendix 1: An elliptical cone

This appendix looks further at the material in §2.2 about drawing a circle or, more generally, an ellipse in perspective. I will again be using a left-handed axis system so  $z$  is positive in the direction the observer is looking. We take the flat elliptical object to be centrally placed with respect to the viewer at a height  $y_0$ , so we are looking at it from below. Its centre is at  $(0, y_0, z_0)$  and the semi-axes have lengths  $a$  and  $b$ , with  $a$  in the  $x$  direction and  $b$  in the  $z$ . The coordinates of a general point  $P$  on the circumference are  $x = a \cos t$ ,  $z = b \sin t + z_0$  where  $t$  is an angle parameter. The light scattered from or emitted by this object travels within a solid cone with elliptical cross-section as illustrated in Figure 11. We now calculate the equation of this cone.

The construction is similar to that in Figure 8 to which the reader is referred. Take the major axis first. There exists a plane containing the  $y$ -axis and the right-most ray  $R_0$  in the cone which skirts the edge of the ellipse, touching it near  $x = a$ . There is a similar plane which is tangent at the other end of the ellipse near  $x = -a$ . Label the angle of this plane from vertical  $y - z$  plane as  $\alpha$ . This is the semi-angle of the elliptical cone in the  $x$ -direction. A point along the ray  $R_0$  satisfies  $x = z \tan \alpha$ . The ellipse object has equation

$$\frac{x^2}{a^2} + \frac{(z - z_0)^2}{b^2} = 1. \quad (\text{A1.1})$$

Substituting  $x = z \tan \alpha$  determines the intersection of ellipse and plane, but for tangency we require the discriminant ' $b^2 - 4ac$ ' of the resulting quadratic in  $z$  to be zero. This reduces to  $a^2 - (z_0^2 - b^2) \tan^2 \alpha = 0$  so  $\tan \alpha = a / \sqrt{z_0^2 - b^2}$ . In the special case of a circle  $a = b = r$  and we recover  $r / \sqrt{z_0^2 - r^2}$  as in §2.2.

The minor semi-angle of the elliptical cone is found from the rays  $R_1, R_2$  coming to  $O$  respectively from the farthest and nearest points of the ellipse (provided it is not rotated in its plane); these are at the ends of the semi-minor axes. Let the angle between  $R_1$  and the  $z$ -axis be  $\beta_1$ , and that between  $R_2$  and the  $z$ -axis be  $\beta_2 > \beta_1$ . The central axis of the cone will be mid-way between these two angles so we need to determine  $\beta_0 = (\beta_2 + \beta_1)/2$  and the minor semi-angle  $(\beta_2 - \beta_1)/2$ . Clearly  $\tan \beta_1 = y_0 / (z_0 + b)$  and  $\tan \beta_2 = y_0 / (z_0 - b)$ . The combined half angles are found using the trigonometric identities

$$\tan 2u = \frac{2 \tan u}{1 - \tan^2 u} \quad \text{from which} \quad \tan\left(\frac{u}{2}\right) = \frac{1}{\tan u} (\sqrt{1 + \tan^2 u} - 1),$$

$$\tan \beta_0 = \frac{\tan\left(\frac{\beta_2}{2}\right) + \tan\left(\frac{\beta_1}{2}\right)}{1 - \tan\left(\frac{\beta_1}{2}\right) \tan\left(\frac{\beta_2}{2}\right)}. \quad (\text{A1.2})$$

This defines the axis of the elliptical cone. The minor semi-angle is

$$\tan \beta_{semi} = \frac{\tan\left(\frac{\beta_2}{2}\right) - \tan\left(\frac{\beta_1}{2}\right)}{1 + \tan\left(\frac{\beta_1}{2}\right) \tan\left(\frac{\beta_2}{2}\right)}. \quad (\text{A1.3})$$

## Appendix 2: Calculating the rotation axis $\mathbf{r}$

This appendix expands on detail related to §3.3 on how to determine the axis of rotation  $\mathbf{r}$  in an angle-and-axis calculation. In §3.3 two methods were described.

### Simultaneous equations

Let  $\mathbf{p} = (p_x, p_y, p_z)$  and similarly for  $\mathbf{q}$  and  $\mathbf{r}$ . Set  $r_z = \sqrt{1 - r_x^2 - r_y^2}$  and bear in mind the  $\pm$  option.

$$\mathbf{p} \times \mathbf{r} = (-p_z r_y + p_y \sqrt{1 - r_x^2 - r_y^2}, \quad p_z r_x - p_x \sqrt{1 - r_x^2 - r_y^2}, \quad -p_y r_x + p_x r_y).$$

We also need  $\mathbf{p} \times \mathbf{q}$  and  $\mathbf{q} \times \mathbf{r}$  which are similar.  $P \equiv \alpha/2$ , and  $\cos P$  equals the scalar product

$$\frac{\mathbf{p} \times \mathbf{r}}{|\mathbf{p} \times \mathbf{r}|} \cdot \frac{\mathbf{p} \times \mathbf{q}}{|\mathbf{p} \times \mathbf{q}|}.$$

Having worked with this algebraically in Mathematica, subsequent expressions become messy. It is therefore only practical to work numerically with each particular case.

To aid this exposition consider the numerical example illustrated in Figure 24. P has spherical co-ordinates  $\theta = 30^\circ$ ,  $\phi = 60^\circ$  and Q  $\theta = 70^\circ$ ,  $\phi = 80^\circ$ . The left panel of the figure shows the unit sphere with the vertices of the planar triangle PQR just sticking out. The centre panel shows the sphere shrunk to expose the triangle further while the insert on the right shows the spherical triangle itself. The labelling of the opposite sides to the angles is arc QR =  $h_P$ , arc PR =  $h_Q$ , arc PQ =  $h_R$ . In Cartesian co-ordinates P and Q are at

$$P: (0.2500, 0.4330, 0.8660), \quad Q: (0.1632, 0.9254, 0.3420).$$

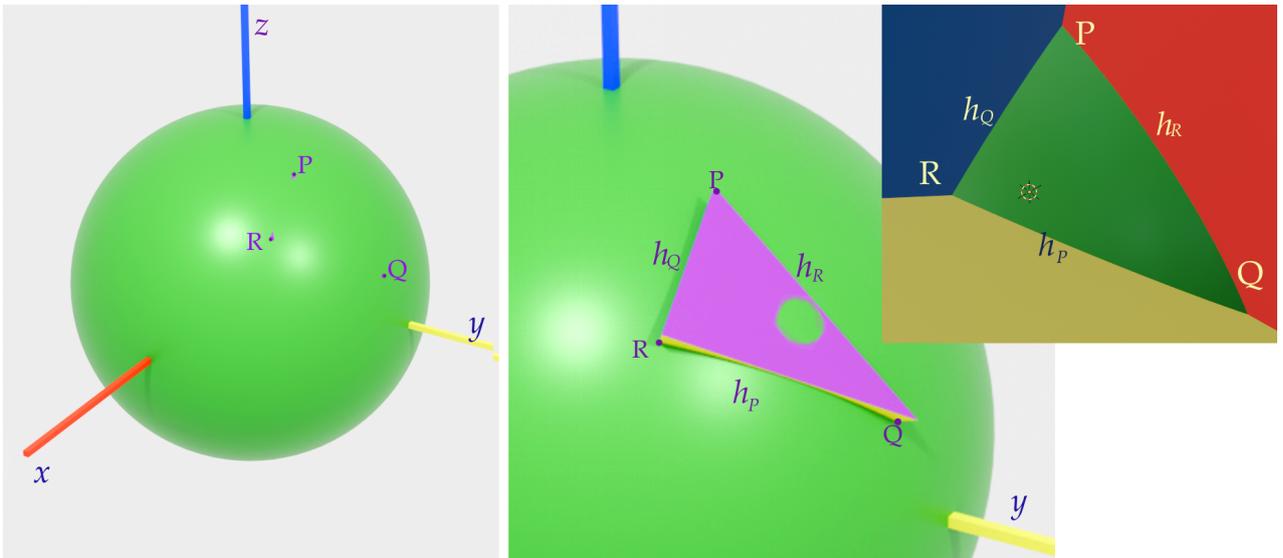


Figure 24: Example triangle. Left: vertices just visible above unit sphere. Centre: enlarged view with planar triangle exposed. Right: spherical triangle formed by blue, red and yellow cutting planes through O.

$$\text{and } \mathbf{s} \equiv \mathbf{p} \times \mathbf{q} = (-0.6533, 0.0558, 0.1607).$$

Choose angle  $P$  to be have absolute value  $70^\circ$  and angle  $Q \equiv \beta/2$  to be  $35^\circ$ . The two messy non-linear simultaneous equations are

$$\cos P = (-0.0315r_x + 0.8976r_y - 0.4397\sqrt{1-r_x^2-r_y^2})/\Delta_P,$$

$$\Delta_P^2 = 0.25 + 0.6875r_x^2 - 0.2165r_xr_y + 0.5625r_y^2 - (0.433r_x + 0.75r_y)\sqrt{1-r_x^2-r_y^2},$$

$$\cos Q = (0.192r_x - 0.3698r_y + 0.909\sqrt{1-r_x^2-r_y^2})/\Delta_Q,$$

$$\Delta_Q^2 = 0.883 + 0.0904r_x^2 - 0.302r_xr_y - 0.7394r_y^2 - (0.1116r_x + 0.633r_y)\sqrt{1-r_x^2-r_y^2}.$$

The two solutions found numerically are

$$\mathbf{r} = (-0.1286, 0.5487, 0.8261) \quad \text{or} \quad (0.5806, 0.4882, 0.6516), \quad (\text{A2.1})$$

giving point R rotated either to the right or left of the arc PQ. We want the second which corresponds to rotation anticlockwise about P by  $-70^\circ$  then anticlockwise about Q by  $+35^\circ$ . The angle  $R = (\pi - \gamma)/2$  is found using the Sine and Cosine Rules at Eqs 13 and 14.

$$-\cos R = -\sin |P| \sin Q \cos h_R + \cos P \cos Q, \quad \cos h_R = \mathbf{p} \cdot \mathbf{q} = 0.7377.$$

Note that  $P$  must be taken as positive since the Cosine rule applies to the magnitude of the angles, not their sense. From this  $R = 83.3^\circ$ ,  $1.453$  radians.

## Reciprocal vectors

This is the method given in Simon Altmann's book on rotations, p 159, though without details. The reciprocal lattice is well recognised in crystallography and optics where it represents the lattice of diffraction spots formed where a wave is diffracted by a rectangular grating. Reciprocal vectors shrink as their direct counterpart increases in length. Three vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  in direct space have three dual reciprocal vectors  $\mathbf{a}'$ ,  $\mathbf{b}'$ ,  $\mathbf{c}'$  in inverse space, with  $\mathbf{a} \cdot \mathbf{a}' = 1$ ,  $\mathbf{a} \cdot \mathbf{b}' = 0$ , and similarly. The reciprocals are calculated using the triple product  $\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$  which equals the volume  $V$  of the parallelepiped spanned by the three vectors.

$$\mathbf{a}' = \frac{\mathbf{b} \times \mathbf{c}}{V}, \quad \mathbf{b}' = \frac{\mathbf{c} \times \mathbf{a}}{V}, \quad \mathbf{c}' = \frac{\mathbf{a} \times \mathbf{b}}{V}. \quad (\text{A2.2})$$

I will outline the theorem that the vector  $\mathbf{r}$  can be uniquely expressed as the weighted sum of three reciprocal vectors which we choose to be  $\mathbf{p}$ ,  $\mathbf{q}$  and  $\mathbf{p} \times \mathbf{q}$ . Start with two vector identities

$$\mathbf{H} \times (\mathbf{C} \times \mathbf{D}) = \mathbf{C}(H \cdot D) - \mathbf{D}(H \cdot C), \quad (\mathbf{A} \times \mathbf{B}) \times \mathbf{H} = \mathbf{B}(A \cdot H) - \mathbf{A}(B \cdot H), \quad (\text{A2.3})$$

the dot products just being scalars. In the first let  $\mathbf{H} = \mathbf{A} \times \mathbf{B}$  and in the second let  $\mathbf{H} = \mathbf{C} \times \mathbf{D}$ , making the two left hand sides equal. Then

$$\mathbf{C}(A \times B \cdot D) - \mathbf{D}(A \times B \cdot C) = \mathbf{B}(A \cdot C \times D) - \mathbf{A}(B \cdot C \times D).$$

The triple products are scalars so write  $A \times B \cdot D$  as  $[ABD]$ . Moreover cyclic permutations are equal, so  $[ABD] = [DAB] = -[DBA]$ . The dot and cross can be placed between either pair of vectors. From this

$$\mathbf{D} = \mathbf{A} \frac{[DBC]}{[ABC]} - \mathbf{B} \frac{[DAC]}{[ABC]} + \mathbf{C} \frac{[DAB]}{[ABC]} = \mathbf{A} \frac{[DBC]}{[ABC]} + \mathbf{B} \frac{[DCA]}{[ABC]} + \mathbf{C} \frac{[DAB]}{[ABC]}.$$

Using Eq A2.2 identify the reciprocal vectors to arrive at the theorem

$$\mathbf{D} = \mathbf{A}(\mathbf{D} \cdot \mathbf{A}') + \mathbf{B}(\mathbf{D} \cdot \mathbf{B}') + \mathbf{C}(\mathbf{D} \cdot \mathbf{C}'). \quad (\text{A2.4})$$

Now let  $\mathbf{D} = \mathbf{r}$ ,  $\mathbf{A} = \mathbf{p}$ ,  $\mathbf{B} = \mathbf{q}$ , and  $\mathbf{C} = \mathbf{s} \equiv \mathbf{p} \times \mathbf{q}$ . Unlike  $\mathbf{p}$  and  $\mathbf{q}$ ,  $\mathbf{s}$  is not a unit vector.

$$\mathbf{r} = \mathbf{p}(\mathbf{r} \cdot \mathbf{p}') + \mathbf{q}(\mathbf{r} \cdot \mathbf{q}') + \mathbf{s}(\mathbf{r} \cdot \mathbf{s}') = \mathbf{p} \frac{[rqs]}{[pqs]} + \mathbf{q} \frac{[rsp]}{[pqs]} + \mathbf{s} \frac{[rpq]}{[pqs]}. \quad (\text{A2.5})$$

This leaves the challenge of finding these four triple products.

The denominator in each reciprocal vector is  $[pqs]$ . Since  $|\mathbf{p} \times \mathbf{q}| = |\mathbf{p}| |\mathbf{q}| \sin h_R = \sin h_R$  gives the area of the base of this solid, and since  $\mathbf{s}$  is normal to the plane OPQ with length  $|\mathbf{p} \times \mathbf{q}|$ , we can conclude that

$$[pqs] = \sin^2 h_R = 1 - \cos^2 h_R. \quad (\text{A2.6})$$

In our example this is  $0 \cdot 4558$  since  $P = -70^\circ$ ,  $Q = +35^\circ$ .

Next take  $[rpq] = [pqr]$ . Start with

$$\sin P = \frac{|(\mathbf{p} \times \mathbf{q}) \times (\mathbf{p} \times \mathbf{r})|}{|(\mathbf{p} \times \mathbf{q})| |\mathbf{p} \times \mathbf{r}|}$$

$$\text{and from Eq A2.2 } (\mathbf{p} \times \mathbf{q}) \times (\mathbf{p} \times \mathbf{r}) = \mathbf{p}[pqr] - \mathbf{r}[pqp] = \mathbf{p}[pqr],$$

the second term being zero. Now  $|\mathbf{p} \times \mathbf{r}| = \sin h_Q$  and  $|\mathbf{p} \times \mathbf{q}| = \sin h_R$ . Using the Sine rule we arrive at two expressions, the first in terms of only given quantities and angle  $R$ :

$$[pqr] = (1 - \cos^2 h_R) \frac{\sin P \sin Q}{\sin R} = \sin h_Q \sin h_R \sin P. \quad (\text{A2.6})$$

In the example this is  $(1 - 0 \cdot 7377^2) \times -0 \cdot 9359 \times 0 \cdot 5879 / 0 \cdot 9931 = -0 \cdot 2474$ .

Now for  $[rsp]$ . Using a similar approach

$$(\mathbf{r} \times \mathbf{s}) \times (\mathbf{s} \times \mathbf{p}) = \mathbf{s}[rsp] - \mathbf{p}[rss] = \mathbf{s}[rsp].$$

Recalling the definition of  $\mathbf{s}$

$$\mathbf{r} \times (\mathbf{p} \times \mathbf{q}) = \mathbf{p}(r \cdot q) - \mathbf{q}(r \cdot p), \quad (\mathbf{p} \times \mathbf{q}) \times \mathbf{p} = \mathbf{q}(p \cdot p) - \mathbf{p}(q \cdot p),$$

$p \cdot p$  being 1. The vector product follows the distributive law so

$$\mathbf{s}[rsp] = \mathbf{s}((r \cdot q) - (r \cdot p)(p \cdot q))$$

$$\text{making } [rsp] = \cos h_P - \cos h_Q \cos h_R. \quad (\text{A2.7})$$

In the example this has value  $0.09003$ .

$[rqs] = [rqs]$  can be obtained in the same way as  $[rsp]$  by working with  $(\mathbf{q} \times \mathbf{s}) \times (\mathbf{s} \times \mathbf{r}) = \mathbf{s}[rqs]$ . The result is

$$[rqs] = \cos h_Q - \cos h_P \cos h_R. \quad (\text{A2.8})$$

In the example its value is  $0.3533$ .

Bringing these results together

$$\mathbf{r} = (0.3533\mathbf{p} + 0.0900\mathbf{q} - 0.2474\mathbf{s}) = (0.5806, 0.4881, 0.6516).$$

This agrees with one of the vectors found by the solution of simultaneous equations.

Altmann reports that Rodrigues found this formula though in a different form. He had

$$\sin R \mathbf{r} = \sin P \cos Q \mathbf{p} + \cos P \sin Q \mathbf{q} + \sin P \sin Q \mathbf{s}. \quad (\text{A2.9})$$

This can be reconciled with my expressions above as follows. Multiply throughout by  $\sin h_R$ . Multiply the term in  $\mathbf{p}$  by  $\sin h_P / \sin h_P$  and the terms in  $\mathbf{q}$  and  $\mathbf{s}$  by  $\sin h_Q / \sin h_Q$ . This gives

$$\begin{aligned} \sin h_R \mathbf{r} &= \left[ \frac{\cos h_Q - \cos h_P \cos h_R}{\sin h_R} \frac{\sin h_P}{\sin h_P} \right] \mathbf{p} \\ &+ \left[ \frac{\cos h_P - \cos h_Q \cos h_R}{\sin h_R} \frac{\sin h_Q}{\sin h_Q} \right] \mathbf{q} + \sin h_Q \sin P \mathbf{s}. \end{aligned}$$

Now a clever step: multiply each term by the ratio in the Sine rule with the corresponding letter, P, Q or R, then cancel the respective  $\sin h_R$ ,  $\sin h_P$ ,  $\sin h_Q$  factors.

$$\begin{aligned} \sin R \mathbf{r} &= \left[ \frac{\cos h_Q - \cos h_P \cos h_R}{\sin h_R} \frac{\sin P}{\sin h_P} \right] \mathbf{p} \\ &+ \left[ \frac{\cos h_P - \cos h_Q \cos h_R}{\sin h_R} \frac{\sin Q}{\sin h_Q} \right] \mathbf{q} + \sin Q \sin P \mathbf{s}. \end{aligned}$$

Now bring in the Cosine rule to replace  $\cos h_P$  by  $\cos h_Q \cos h_R + \sin h_Q \sin h_R \sin P$  and similarly for  $\cos h_Q$ . Several terms and factors cancel from the  $\mathbf{p}$  and  $\mathbf{q}$  terms and we arrive at Rodrigues' expression Eq A2.9.

## Appendix 3: Quaternions and the quaternion group

### Symmetry group properties

The symmetries of  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  give them a group structure. The quaternion group has 8 elements and some similarities with the dihedral group  $D_4$  which are the 8 symmetries of a square. The elements of the quaternion group can be generated from two elements  $A, B$  for which  $\langle A^4 = B^4 = I, A^2 = B^2, BA = A^3B \rangle$ ,  $I$  being the identity. The elements and their inverses are

$$\begin{array}{lcccccccc} \text{element} & I & A & A^2 & A^3 & B & AB & A^2B & A^3B \\ \text{inverse} & I & A^3 & A^2 & A & A^2B & A^3B & B & AB \end{array}$$

There is one element,  $a^2$ , of order 2 and 6 of order 4. A suitable set of three non-commuting order-4 elements can represent  $\mathbf{i}, \mathbf{j}, \mathbf{k}$  and generate the quaternion properties. For example

$$\text{if } \mathbf{i} = A, \mathbf{j} = B, \mathbf{k} = AB, \text{ then } \mathbf{ij} = AB = \mathbf{k}, \mathbf{ji} = BA = A^3B = A^2 \cdot AB = -\mathbf{k}.$$

With  $i^2 = -1$  the elements can also be represented by the two matrices

$$A = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The other elements are then

$$\begin{aligned} A^2 &= \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, & A^3 &= \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, & AB &= \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}. \\ A^2B &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, & A^3B &= \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, & I &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

A general quaternion is therefore a weighted sum of four matrices.

An alternative representation is as permutations. The generating elements can be  $A = (1247)(3685)$ ,  $B = (1348)(2576)$  and then the other non-identity elements are

$$\begin{array}{lcccccc} \text{element} & A^2 & A^3 & AB & A^2B & A^3B \\ \text{permutation} & (14)(27)(38)(56) & (1742)(3586) & (1645)(2378) & (1843)(2675) & (1546)(2873) \end{array}$$

### Algebra of quaternions

Here are some of the properties of more general quaternions.

- as a scalar plus vector  $q = a + \mathbb{V} = a + (b\mathbf{i} + c\mathbf{j} + d\mathbf{k})$ , written  $(a, b, c, d)$ .
- identity  $I = 1 = (1, 0, 0, 0)$ .
- conjugate, analogous to the complex conjugate of complex numbers. Reverse the signs of  $\mathbf{i}, \mathbf{j}$  and  $\mathbf{k}$ .: If  $q = a + \mathbb{V}$ , then  $q^* = a - \mathbb{V}$ ,
- norm. The magnitude of the quaternion.  $|q| = \sqrt{qq^*} = \sqrt{a^2 + b^2 + c^2 + d^2}$ . That is, the square root of the sum of squares of its four coefficients.

- inverse.  $q^{-1}$  is such that  $q^{-1}q = qq^{-1} = (1, 0, 0, 0) = I$ . Thus  $q^{-1} = q^*/|q|$ .

Multiplication is associative meaning that brackets do not matter, so  $ABC = A(BC) = (AB)C$ . The product of  $q = (a, b, c, d) = a + \mathbb{V}$  and  $Q = (\alpha, \beta, \gamma, \delta) = \alpha + \mathbb{W}$  is developed as follows:

$$\begin{aligned} qQ &= a(\alpha, \beta, \gamma, \delta) + b\mathbf{i}(\alpha, \beta, \gamma, \delta) + c\mathbf{j}(\alpha, \beta, \gamma, \delta) + d\mathbf{k}(\alpha, \beta, \gamma, \delta) \\ b\mathbf{i}(\alpha, \beta, \gamma, \delta) &= b\alpha\mathbf{i} - b\beta + b\gamma\mathbf{k} - b\delta\mathbf{j} \\ c\mathbf{j}(\alpha, \beta, \gamma, \delta) &= c\alpha\mathbf{j} - c\beta\mathbf{k} - c\gamma + c\delta\mathbf{i} \\ d\mathbf{k}(\alpha, \beta, \gamma, \delta) &= d\alpha\mathbf{k} + d\beta\mathbf{j} - d\gamma\mathbf{i} - d\delta \end{aligned}$$

If  $qQ = (A, B, C, D) = A + \mathbb{X}$ ,

$$\begin{aligned} A &= a\alpha - b\beta - c\gamma - d\delta, & B &= a\beta + b\alpha + c\delta - d\gamma, \\ C &= a\gamma - b\delta + c\alpha + d\beta, & D &= a\delta + b\gamma - c\beta + d\alpha. \end{aligned}$$

This is sufficient, but if you wish to cast this within the ‘scaler plus vector’ model, the dot and cross products of  $\mathbb{V}$  and  $\mathbb{W}$  can be used to produce a condensed ‘vector’ version of  $qQ$ . First note that  $\mathbf{V} \cdot \mathbf{W} = b\beta + c\gamma + d\delta$  to write  $A$  as  $a\alpha - \mathbf{V} \cdot \mathbf{W}$ . Next note that the cross product  $\mathbb{V} \times \mathbb{W} = (c\delta - d\gamma)\mathbf{i} - (b\delta - d\beta)\mathbf{j} + (b\gamma - c\beta)\mathbf{k}$  and subtract this from  $\mathbb{X}$ . The remainder is  $a\mathbb{W} + \alpha\mathbb{V}$ . We arrive at

$$qQ = (a\alpha - \mathbb{V} \cdot \mathbb{W}) + (a\mathbb{W} + \alpha\mathbb{V} + \mathbb{V} \times \mathbb{W}).$$

This is Eq 18 in alternative notation.

### Some special quaternions

A so-called ‘pure’ quaternion has zero scalar part. Using the notation of §3 in the main article, in  $q = (\lambda, \mathbb{V})$  the  $\lambda = 0$ . If two pure quaternions are non-commutatively multiplied,

$$(0, \mathbb{V}_2).(0, \mathbb{V}_1) = (-\mathbb{V}_2 \cdot \mathbb{V}_1, \mathbb{V}_2 \times \mathbb{V}_1),$$

and if the two vector parts are orthogonal, their dot product is zero and another pure quaternion is the result. A ‘unit quaternion’ is a pure one in which  $\mathbf{V}$  is a unit vector  $\mathbf{u}$ . These follow the same rule as for multiplication of the basic unit quaternions  $\mathbf{i}, \mathbf{j}, \mathbf{k}$  in that

$$(0, \mathbf{u})^2 = (0, \mathbf{u}).(0, \mathbf{u}) = (-\mathbf{u} \cdot \mathbf{u}, \mathbf{u} \times \mathbf{u}) = (-1, 0) \equiv 1.$$

A general quaternion can therefore be split into two components, one containing only a scalar part and the other only the vector part as a scalar-weighted unit vector:

$$(\lambda, \mathbb{V}) = (\lambda, 0) + (0, V\mathbf{u}).$$

This is called the ‘binary form’ of the quaternion in that it has a real part and an ‘imaginary’ part whose square is  $-V^2$ . In terms of rotations in §3.3  $\lambda$  was identified as  $\cos(\psi/2)$  and  $\mathbb{V}$  as  $\sin(\psi/2)\mathbf{u}$ . A pure quaternion, therefore, represents a rotation through  $\psi = \pi$ , a half turn about axis  $\mathbf{u}$ .