1

Basic formulae

1.1 Introduction

Before the development of radar, there was no direct means of measuring the distance of any astronomical object. Even today the application of radar is severely limited. It is possible to measure the radar distances of some objects within the solar system, but this is not technically possible for more distant objects, like the stars, and is unlikely to be in the foreseeable future. The distances of the stars can only be inferred from slight periodic variations in their positions, variations that are caused by parallax. When the term position is used here, we mean the apparent direction in which the star is located. This is something that can be measured directly with great accuracy, and is neatly expressed as two angular coordinates. For positional purposes it is often convenient to ignore the distance of the star, since it is inaccessible to direct measurement, and to treat all the stars as if they were at the same distance, that is, as if they were situated on the surface of a sphere centred on the observer. This is what is meant by the 'celestial sphere'. Its radius is quite arbitrary, though presumably very large by any terrestrial standard. There is no loss of generality, however, if the radius of the celestial sphere is adopted as the unit of length in positional astronomy; indeed, this leads to considerable simplification.

1.2 Spherical geometry – great circle arcs

A sphere is defined as the surface whose points are all equidistant from a fixed point, the centre. The sphere is a two-dimensional surface which is finite but unbounded. Spherical geometry, meaning geometry performed on the *surface* of a sphere, is therefore, a two-dimensional geometry, but it differs significantly from the ordinary two-dimensional plane geometry of Euclid. In particular, there are no straight lines on the

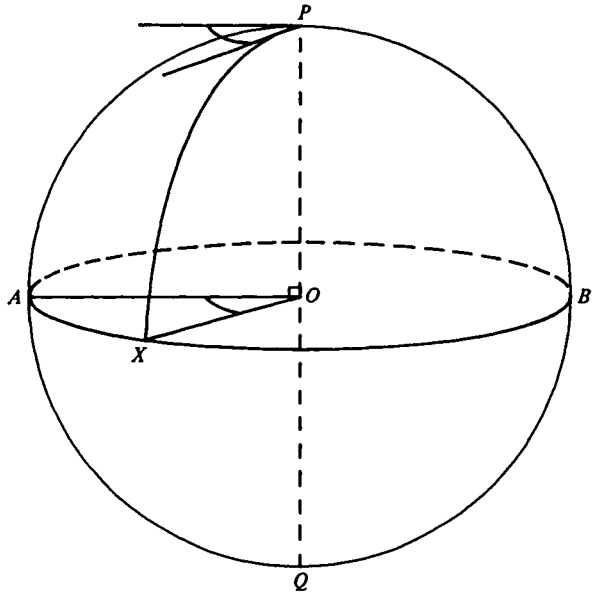
surface of a sphere. The equivalent curves are the arcs of great circles, which are defined as follows:

Definition. Any plane through the centre of the sphere intersects the sphere in a great circle. The *poles* of the great circle are the two extremities of the diameter of the sphere drawn perpendicular to that great circle.

It is immediately clear that a great circle is indeed a circle whose radius is the radius of the sphere, which we have taken as unity. A great circle AXB is shown in Fig. 1.1 and also the poles of this great circle, P and Q. These two points are said to be *diametrically opposite*. Another great circle is also drawn in Fig. 1.1, namely $PAQB$. It is easily seen, by rotating the plane of this great circle, that *any* great circle through P must also pass through the diametrically opposite point Q. This is a special property, however, of diametrically opposite points.

Consider, on the other hand, two general points of the sphere, such as A and X. The great circle on which they lie is uniquely determined. For the points A and X and the centre of the sphere O together define a unique plane, and this plane cuts the sphere in a great circle. The great circle arc AX is in fact the shortest curve that can be drawn on the sphere's surface to link these two points. This important property of great circle arcs means that

Figure 1.1



they are the geodesics of the sphere, analogous to straight lines in Euclidean geometry.

Since the sphere has unit radius, the following important conclusion is readily made: the length of a great circle arc is equal to the angle, in radians, that it subtends at the centre of the sphere. Although, strictly, radians should always be used, there is no ambiguity in practice in specifying the length of a great circle arc in degrees. For example, if we say that the great circle arc AX in Fig. 1.1 is 45°, we mean that it subtends this angle at O and that its length is, therefore, $\frac{1}{4}\pi$. To be precise, there are two great circle arcs joining A and X, namely AX already discussed and the arc ABX measured in the opposite direction and of length $2\pi - AX$. When we speak of the great circle arc joining two points, however, we shall always intend the shorter of the two arcs, which is always less than π in length.

Now join PX (i.e. construct the great circle arc PX). Then PX and PA are two great circle arcs intersecting at P, and their intersection produces a spherical angle APX, which can be defined in several equivalent ways. For example:

Definition. The *spherical angle* between two intersecting great circle arcs is the angle between their planes.

Alternatively, the spherical angle may be defined as the angle between the tangents to the two intersecting great circle arcs at their point of intersection. With either definition, it is readily seen, from Fig. 1.1, that

$$\begin{aligned}\text{Spherical angle } APX &= A\hat{O}X,\\ &= \text{great circle arc } AX. \qquad (1.1)\end{aligned}$$

Three points on a sphere which lie on the same great circle are analogous to three collinear points in plane geometry. Consider, however, three general points A, B, C which do not lie on the same great circle. Then, as shown in Fig. 1.2, these points may be joined in pairs by great circle arcs BC, CA, AB, each of which is less than π in length. The resulting figure is called a *spherical triangle*. Its parts consist of the three sides, which are the great circle arcs already mentioned, and the three included spherical angles. It will be convenient to denote each angle by the capital letter of its vertex and each side by the small letter corresponding to the opposite angle. Thus $BC = a$, $CA = b$, $AB = c$, as indicated in Fig. 1.2.

The three points A, B, C define a plane, and, since the three points do not lie on a single great circle, this plane does not pass through the centre of the sphere. We can construct a plane, however, passing through the sphere's centre parallel to the plane ABC. This latter plane divides the sphere into two hemispheres, and it is clear that the spherical triangle is confined to one

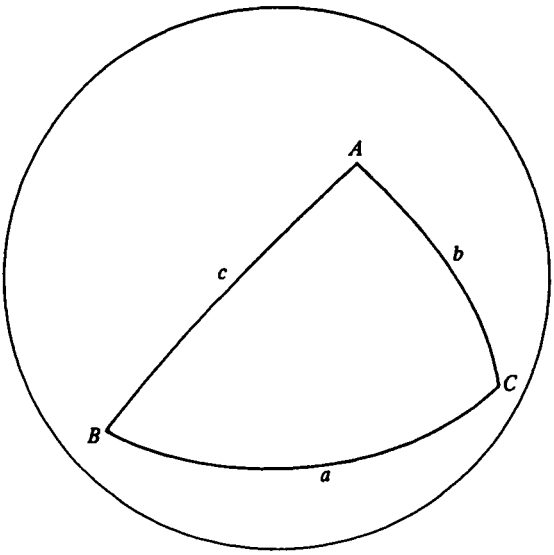
of these hemispheres. It is then intuitively obvious that each of the angles of the spherical triangle is less than 180°.

Spherical triangles have certain properties in common with plane triangles. For example, any side is less than the sum of the other two. There are also important differences, however. In plane geometry the sum of the angles of a triangle is 180°. In a spherical triangle the sum of the angles is not fixed but always exceeds this value. A plane triangle may have one right angle, but only one. A spherical triangle may have one, two, or even three right angles. For example, referring back to Fig. 1.1, PAX is a spherical triangle in which angles A and X are both right angles. Trigonometrical formulae may be used to relate the parts of a spherical triangle. These will be developed in a later section, but it is useful to note one point at the outset. All parts of a spherical triangle lie between 0 and 180°; they are confined to the first two quadrants. The inverse cosine is single-valued in this range; the inverse sine is not. It is desirable, therefore, to use formulae that give the cosine rather than the sine of the part required, since, otherwise, an ambiguity inevitably exists.

Till now we have been concerned only with the geodesics of the sphere – great circles. Another curve of importance is defined as follows:

Definition. A plane that does not pass through the centre of the sphere intersects the sphere, if at all, in a *small circle*. The poles of the small circle

Figure 1.2



are the extremities of the diameter of the sphere that is perpendicular to the small circle's plane.

It is evident that a small circle is, indeed, a circle whose radius is less than that of the sphere, justifying the name. In Fig. 1.3, the small circle AB is shown, and also the parallel great circle CD. Both have the points P and Q as their poles. Let the radius of the small circle be r and let the great circle arc $AP = \theta$. Then from the plane triangle AOS, it is readily seen that

$$AS = AO \sin A\hat{O}S,$$

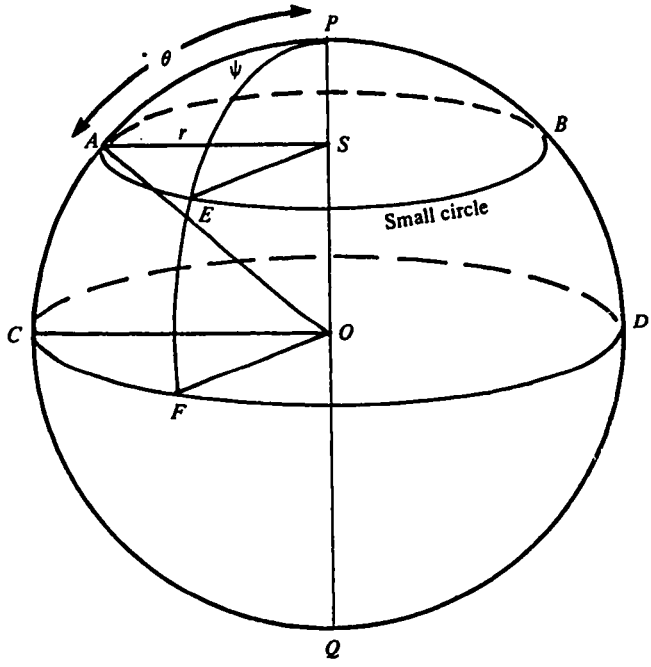
i.e.

$$r = \sin \theta. \tag{1.2}$$

Let E be any point on the small circle AB. Join the great circle arc PE and produce it to meet CD in F. Clearly $PE = \theta$. In fact, since all points on the small circle have the same separation on the sphere from the pole P, the small circle is the analogy of the circle in plane geometry. Let spherical angle $APE = \psi$. Then, equivalently, $C\hat{O}F = \psi$. Now, since AS and ES are, respectively, parallel to CO and FO, it follows that $A\hat{S}E = \psi$. The length of the small circle arc AE can now be derived as

$$\text{small circle arc } AE = rA\hat{S}E = \psi \sin \theta. \tag{1.3}$$

Figure 1.3



Notice that θ is the separation of the points on the small circle from their pole, and ψ is the spherical angle at the pole subtended by the arc in question.

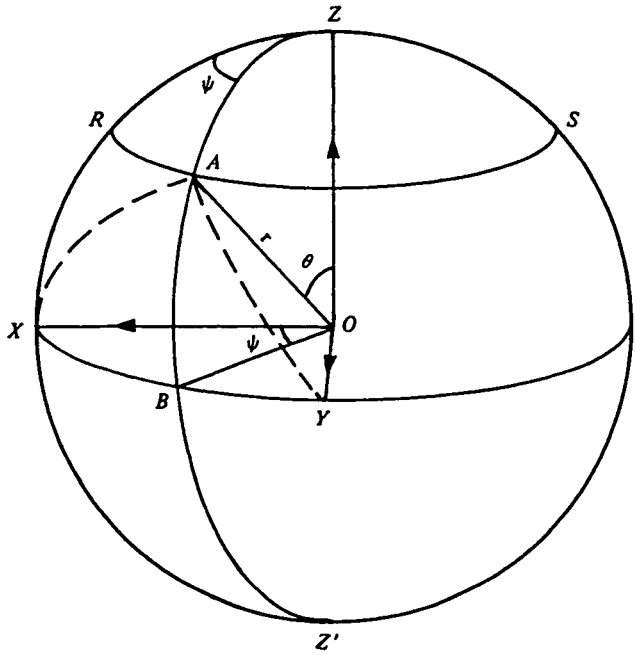
The distance given in equation (1.3) is not, of course, the shortest distance on the sphere from A to E. The length of the *great* circle arc AE will be less than this. It is important to recognize that a small circle arc cannot be part of a spherical triangle. Certainly there is a spherical triangle APE, but it is not shown in Fig. 1.3. In particular, small circle arc AE must be replaced by the great circle arc, and, as a result, the angles at A and E will not be right angles as they appear in the diagram.

1.3 Spherical polar coordinates

A number of different coordinate systems can be set up on the celestial sphere. The systems used in practice are all basically similar and are just different forms of spherical polar coordinates.

Suppose a set of right-handed rectangular Cartesian axes $Oxyz$ are set up at the centre O of a unit sphere. Further, let the positive directions of these axes intersect the sphere in the points X, Y and Z, as shown in Fig. 1.4. Then the great circles XY and ZX represent the x–y and z–x planes respectively.

Figure 1.4



Let A be any point of the sphere with Cartesian coordinates (x, y, z). Then

$$x^2+y^2+z^2=1, \tag{1.4}$$

and so one of the coordinates is redundant. It is often more convenient, therefore, to employ spherical polar coordinates, (r, θ, ψ) say. With their usual definitions, the radial coordinate r of the point A is OA, the polar coordinate θ is $Z\hat{O}A$, and the azimuthal coordinate ψ is the angle between the plane ZOA and the z–x plane.

Since the point A is on the unit sphere, its radial coordinate $r=1$. Moreover, its angular coordinates are simply related to the position of A on the sphere. From the definitions and results of the last section, it is immediately seen that the polar angle θ is the great circle arc length ZA, and the azimuthal angle ψ is the spherical angle XZA. To span the entire sphere, the coordinates θ and ψ must be in the ranges:

$$\begin{aligned} 0 \leqslant \theta \leqslant \pi, \\ 0 \leqslant \psi < 2\pi. \end{aligned} \tag{1.5}$$

To set up a coordinate system on the celestial sphere, therefore, it is necessary to select a pole of the coordinate system Z, from which the polar angle θ is measured, and a reference great circle ZX from which the azimuthal angle ψ is measured. All the coordinate systems used in spherical astronomy are essentially of this form; their differences arise principally from a different choice of pole. Occasionally, it is true, a left-handed system may be used, and quite often the complement $(\frac{1}{2}\pi-\theta)$ is used rather than θ itself. Apart from these slight differences, however, the coordinate systems of spherical astronomy are all different examples of the spherical polar coordinates (θ, ψ) defined above. The coordinate grid is as follows: the curves $\theta=$constant are small circles with pole Z, like RAS in Fig. 1.4, and the curves $\psi=$constant are the semi-great circles like $ZABZ'$.

Despite the inherent redundancy, it can still be advantageous to use Cartesian coordinates. This often leads to a form of equations suitable for computation and allows the elegance of vector methods. Thus, if $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors in the positive x-, y- and z-directions, then the location of A on the celestial sphere may be given by its position vector, $\mathbf{r}_A$, say, where

$$\mathbf{r}_A = x\mathbf{i}+y\mathbf{j}+z\mathbf{k}. \tag{1.6}$$

This is, of course, a unit vector, and (x, y, z) are the direction cosines of the line OA. So, in terms of great circle arcs,

$$\begin{aligned} x &= \cos XA, \\ y &= \cos YA, \\ z &= \cos ZA. \end{aligned} \tag{1.7}$$

Moreover, in terms of the spherical polar coordinates, we have the well-known transformations

$$x = \sin\theta\cos\psi,$$
$$y = \sin\theta\sin\psi, \tag{1.8}$$
$$z = \cos\theta.$$

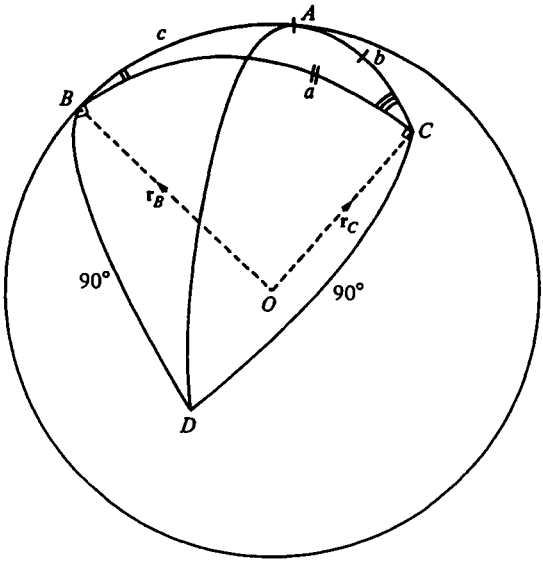
Any problem in spherical astronomy may be treated by the methods of spherical trigonometry; alternatively, a three-dimensional vector approach may be adopted. The method chosen is largely a matter of personal preference. It has been assumed that the reader is already acquainted with basic vector methods, although, for reference purposes, essential vector formulae are summarized in Appendix A. The basic formulae of spherical trigonometry, on the other hand, are developed *ab initio* in the next section.

1.4 Spherical trigonometry – basic formulae

Let ABC be the spherical triangle shown in Fig. 1.5. If we adopt a spherical polar coordinate system (θ, ψ) with the point A as pole and the arc AB as the reference great circle, then the point B is given by $\theta = c$, $\psi = 0$, and the point C by $\theta = b$, $\psi = A$. Let $\mathbf{r}_B$, $\mathbf{r}_C$ be the position vectors of the points B and C respectively. Then, by equation (1.8), it follows that

$$\mathbf{r}_B = (\sin c, 0, \cos c), \tag{1.9}$$
$$\mathbf{r}_C = (\sin b\cos A, \sin b\sin A, \cos b). \tag{1.10}$$

Figure 1.5



Now the angle between these two vectors is equal to the side BC of the spherical triangle. So, taking the scalar product, it follows that, since both are unit vectors, $\mathbf{r}_B \cdot \mathbf{r}_C = \cos a$. This scalar product can also be formed from equations (1.9) and (1.10) to yield the important result

$$\cos a = \cos b \cos c + \sin b \sin c \cos A. \tag{1.11}$$

This is the most fundamental formula of spherical trigonometry and will be referred to simply as the *cosine formula*. It is similar to the cosine formula for a plane triangle, and, like it, expresses one side of the triangle in terms of the other two sides and the included angle. There are two ready applications of equation (1.11). If the sides b and c and the included angle are known, it can be used as it stands to give the third side a. Alternatively, if all three sides are known, it may be used to derive the angle A. In addition to equation (1.11), companion formulae may be obtained by cyclical permutation of the symbols.

All the formulae of spherical trigonometry can be derived by successive applications of the cosine formula and subsequent trigonometric and algebraic manipulation. Two of these formulae can, however, be more directly obtained by considering the vector $\mathbf{r}_C \times \mathbf{r}_B$. Since the angle between these two unit vectors is equal to the arc BC, the vector product has a magnitude $\sin a$. Moreover, it will be directed towards a point on the sphere that is 90° from both C and B. This is the point D, shown in Fig. 1.5, which is the pole of the side BC of the original spherical triangle. If $\mathbf{r}_D$ is the position vector of the point D, then

$$\mathbf{r}_C \times \mathbf{r}_B = \sin a\, \mathbf{r}_D. \tag{1.12}$$

Now the left-hand side of this vector equation can be obtained from equations (1.9) and (1.10), giving the result

$$\mathbf{r}_C \times \mathbf{r}_B = (\sin b \cos c \sin A,\ \cos b \sin c - \sin b \cos c \cos A,\ -\sin b \sin c \sin A). \tag{1.13}$$

Further, by equation (1.8), we may write the right-hand side of equation (1.12) as

$$\sin a\, \mathbf{r}_D = \sin a(\sin AD \cos BAD,\ \sin AD \sin BAD,\ \cos AD). \tag{1.14}$$

By equating the components of these last two equations, important new results are derived.

Consider spherical triangle BAD. Since D is the pole of BC, the arc $BD = 90°$ and it is perpendicular to BC. Therefore, the angle $ABD = 90° + B$. Applying the cosine formula to spherical triangle BAD then yields

$$\cos AD = \cos 90° \cos c + \sin 90° \sin c \cos (90° + B),$$

i.e.

$$\cos AD = -\sin c \sin B.$$

Substitute this result into the z-component of equation (1.14). Then equating z-components of equations (1.13) and (1.14) will yield

$$\sin b \sin c \sin A = \sin a \sin c \sin B,$$

or

$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b}.$$

By symmetry, this last equation can be written in more complete form as

$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c}. \tag{1.15}$$

This result will be known as the *sine formula*. Its similarity to the sine formula in a plane triangle is evident and it suffers from the same inherent defect of ambiguity.

If the sine formula is now applied to the spherical triangle BAD, the following result is obtained

$$\sin AD \sin BAD = \sin BD \sin ABD = \cos B.$$

Now substitute this result into the y-component of equation (1.14), then equating y-components of (1.13) and (1.14) will give the important result

$$\sin a \cos B = \cos b \sin c - \sin b \cos c \cos A. \tag{1.16}$$

This will be known as the *analogue formula*.

It will be found that the applications of the analogue formula are generally similar to those of the sine formula. It is, however, patently more complicated and perhaps slightly difficult to remember. On close inspection readers will probably discover for themselves some symmetry in the structure of the formula, but a visual aid to memory is given in Fig. 1.6. This

Figure 1.6