Great Britain, for this had been derived from astro-geodetic distances obtained during the original Primary Triangulation of the country. In the days before digital computing, once a national survey had been computed using a particular reference figure it would have been extremely inconvenient and costly to convert the positions of many hundreds or even thousands of control points to another spheroid. It was done in the USSR when the decision was taken in 1942 to transform the entire control network from the Bessel spheroid to the newly described Krassovska figure, but that was a practically unique example. It follows that usually a national survey continued to be based upon a particular figure long after the original reasons for its choice had ceased to be valid.

This argument carries less weight today than before digital computing became commonplace. It is interesting to note in this context that probably the first major use of digital computing in geodesy and surveying was the work undertaken by the US Army Map Service shortly after World War II, when they accomplished the formidable task of reducing the national surveys of western Europe to a common datum on the International Spheroid. This is known as the European Datum, 1950, or ED50. This network had hitherto been based upon a multiplicity of different points of origin, reference spheroids, units of measure and projections. We shall also refer to the change in the North American Datum from NAD 27 into NAD 83 during the 1980s, which amongst other changes includes that from the Clarke 1866 figure to GRS 80.

Nevertheless the use of different figures still remains. It arises partly from historical accident, partly from inertia and partly for reasons of national prestige. Sometimes it also happens that the chosen spheroid fits the shape of the geoid in that country better than any of the others.

Finally the continuity of use is important. Indeed Chovitz (1981) has argued that this continuity is at least as important as the formal accuracy of recording the length of the major semi-axis and flattening. Some of the better-known figures, such as Airy, Everest and the three useful Clarke determinations, have been slightly modified on many occasions for use in different places or for different purposes. Typical examples include retaining the original value for the semi-axis, a, but using it with a slightly different (rounded) value for f. Other changes have been enforced by the discrepancies introduced to the dimensions of the semi-axes through converting from British Standard into metric units or vice-versa. For example, Strasser (1975) has shown how US legislation concerning the definition of the metre has created numerous difficulties in reconciling different versions of the Clarke 1866 figure. Sometimes we know enough about the history of a survey to understand where discrepancies have arisen. More often it may be extremely difficult to reconcile these so that mistakes are sometimes made in choosing the correct version of Everest or Airy.

CHAPTER 2

Coordinate reference systems on the plane

It is impossible not to feel stirred at the thought of the emotions of men at certain historic moments of adventure and discovery—Columbus when he first saw the Western shore, Pizarro when he stared at the Pacific Ocean, Franklin when the electric spark came from the string of his kite, Galileo when he first turned his telescope to the heavens. Such moments are also granted to students in the abstract region of thought, and high among them must be placed the morning when Descartes lay in bed and invented the method of co-ordinate geometry.

A. N. Whitehead

Introduction

In this chapter we review some of the fundamental ideas about the plane coordinate systems which are used in surveying and mapping, both from the viewpoint of studying the mathematics of map projections and the practical tasks which arise in cartography.

Coordinates are a convenient method of recording position in space. They may be used to locate position in two dimensions, such as a point on a graph. An extension of this method to map use allows the location of a place by its grid reference. Definition of coordinate position on the surface of a three-dimensional body such as a sphere or spheroid is rather more difficult. However, the reader should already be aware of the method of describing location by means of latitude and longitude, which are geographical coordinates. These are defined in Chapters 3 and 4, where the differences between defining latitude on a sphere and on a spheroid are introduced. In addition to providing a means of reference, coordinates can also be used as a convenient way of solving certain geometrical problems. The branch of mathematics known as coordinate geometry analyses problems through the relationship between points as defined by their coordinates. By these means, for example, it is possible to derive algebraic expressions defining different kinds of curve which cannot be done by Euclidean geometry. Coordinate geometry is an exceptionally powerful tool in the study of the theory of map projections, and without its help it is practically impossible to pass beyond the elementary descriptive stage. Plane coordinate geometry is usually studied first through the
medium of the conic sections or the definition of the different kinds of curve formed by the surface of a cone where this has been intersected by a plane. Two of the resulting sections, the ellipse and the circle, are of fundamental importance to the theory of distortions in map projections.

There are an infinite number of ways in which one point on a plane surface may be referred to another point on the same plane. Every map projection creates a unique reference system which satisfies this requirement and an infinite of different map projections could theoretically be described. However it is desirable to use some kind of coordinate system to describe, analyse and construct each of these projections. Any system to be used for such purposes ought to be easy to understand and simple to express algebraically. For plane representation the choice lies between plane cartesian coordinates and polar coordinates.

**Plane cartesian coordinates**

The reader will already be familiar with graphs as a method of plotting two variables on specially ruled paper and with the National Grid on Ordnance Survey maps. The graph and the National Grid are simple, but special, examples of plane cartesian coordinates. In the general case, any plane coordinate system which makes use of linear measurements in two directions from a pair of fixed axes can be regarded as a cartesian system. The coordinate system comprises sets or families of lines which intersect one another to form a network when plotted. The only necessary conditions which must be fulfilled are:

- that the two families of lines are distinct from one another;
- that every line of one family should intersect every line of the other family at one point only;
- that no two lines of the same family should intersect one another.

Thus a cartesian coordinate system can comprise families of straight lines or curves which may intersect at any angle. However, it is a distinct advantage if the special case is chosen in which both families of lines are straight and that they are orthogonal, or intersect at right angles. This special case, characterised by ordinary graph paper and by the National Grid on Ordnance Survey maps, may be called a plane rectangular cartesian coordinate system, or, in short, rectangular coordinates.

In Fig. 2.01 the origin of the rectangular coordinate system is the point \( O \), through which two orthogonal axes, \( OX \) and \( OY \), have been plotted. These axes define the directions of the two families of lines. Since the axes are straight lines and perpendicular to one another, it follows that all the lines composing one family will be parallel to one another and that all the points of intersection within the network are made from lines which are perpendicular to one another. The position of a point \( A \) is defined by

![Diagram of coordinate reference systems on the plane](image)

**Fig. 2.01** Plane rectangular cartesian coordinates.

the two linear measurements \( OM \) and \( ON \) made from the origin to the points \( M \) and \( N \) on the two axes, which are drawn perpendicular from \( A \) to the axes. Clearly \( AM \) is parallel to \( OY \) and \( AN \) is parallel to \( OX \). The mathematical convention is to refer to the horizontal axis \( OX \) as the \( X \)-axis or abscissa. The vertical line \( OY \) is called the \( Y \)-axis or ordinate. However, the convention is not always observed in the study of geodesy, surveying and map projections. In some books the notation is reversed and \( OX \) is the axis pointing upwards on the page. There are cogent reasons for this change in notation, to do with the direction in which angles are measured, as described on p. 34, but the change in axes is extremely confusing to the beginner. We shall use the standard mathematical, or graph, convention throughout most of this book and refer to the coordinates of the point \( A \) as being \((x, y)\) according to the axes illustrated in Fig. 2.01. It is not until Chapter 15 that we have to change the notation for particular purposes. Even then we use it sparingly.

The units into which the axes are subdivided for the purposes of linear measurement are quite arbitrary. For example, graph paper is available with both millimetre and inch ruling, with various combinations of multiples and fractions of these. The National Grid is measured in metres. We shall make considerable use of units of earth radius, \( R \), in which coordinates are expressed in multiples or decimals of \( R \) without having to convert into units suitable for plotting on a sheet of paper.

There is a sign convention to be observed in the use of rectangular coordinates. This states that the \( X \)-axis is reckoned positive towards the right and the \( Y \)-axis is positive towards the top of the page. In other words, a point in the top right-hand quarter of a graph illustrated by Fig.
Coordinate Systems and Map Projections

2.01 is defined by positive values of x and y, whereas a point in the bottom left-hand quarter has negative values for x and y. The quarters are termed quadrants and these are numbered 1–4 in a clockwise direction commencing with the top right quadrant. Hence the sign convention is:

1st quadrant  
2nd quadrant  
3rd quadrant  
4th quadrant  

The map grid as an example of plane rectangular coordinates

A grid has been defined in the Glossary of Technical Terms in Cartography (Royal Society, 1966) as *'a cartesian reference system using distances measured on a chosen projection'*.† In the first edition of this book the author disagreed with the last seven words in this definition, but as a major contributor to the Glossary felt a certain loyalty to the deliberations of the working group, limiting himself to making only a mild criticism of this particular definition. Professor E. H. Thompson (1973) was not restrained by such inhibitions, and in his important review of the first edition of this book made the following characteristically forthright statement:

It is sad to see an author, who has clearly thought out so much of the problem for himself, committing old faults because his courage fails him at the last minute. He says 'For the moment it will suffice to regard a grid as a system of rectangular coordinates superimposed upon a plane corresponding to the ground'. Why ‘For the moment’? Grids are simply sets of squares and to paraphrase Gertrude Stein, a square is a square is a square. It is indeed a pity that we are also given a definition from the Royal Society Glossary of Technical Terms in Cartography... Whatever has a projection to do with a grid? The sin is Dr Maing's only in so far as he perpetuates it and he barely does that for he says, about the above definition, ‘...the last seven words...are probably necessary but tend to confuse the issue. They are not necessary...’ and the issue by being quite wrong.

One family of lines is orientated approximately north–south and the other family, by definition, is perpendicular to them. Measurements along the axes are made in some units used for ground measurement. Nowadays the metric system is used almost everywhere, but formerly some grids used feet or yards as the unit. By virtue of the approximate orientation of a grid, the abscissa of a point is usually called its Easting and the ordinate is its Northing. Thus E corresponds to x and N corresponds to y in the mathematical and graph conventions. We will introduce this substitution without further comment where it is appropriate to refer to a point by its (E, N) coordinates rather than by (x, y). The order in which the grid coordinates are recorded is often confusing to the beginner, who has probably only just learnt to describe geographical position in the order ‘latitude-followed-by-longitude’. If it is remembered that a grid is like a graph, then the logic of using ‘Easting-followed-by-Northing’ matching the ‘x-followed-by-y’ graph convention is apparent.

We do not attempt to describe in detail how a grid reference may be obtained from a map, for it is assumed that the reader can do this already. Military manuals, such as Ministry of Defence (1973, 1978) are always painstaking in describing this aspect of map use, for it is vital to military communications. The practices adopted by the Ordnance Survey for use with the National Grid are described in Ordnance Survey (1951) and Harley (1975). This distinguishes the slightly different procedures to be adopted at different map scales. Moreover many Ordnance Survey and other national survey maps have the appropriate instructions, with a worked example, printed in the margin.

Because a grid is a form of graph it must have an origin. Moreover if the grid is to satisfy its purpose to serve as a national or international standard of reference, the point of origin must be explicitly stated, together with the orientation of the axes at this point. It is this aspect of a grid which introduces the confusing ideas in the second part of the definition given on p. 30. For example, the National Grid (Fig. 2.02) has its origin at the point with latitude 49°N, longitude 2°W. This is situated in the Golf de St Malo, about 20 km south-east of St Helier in Jersey. The same point is also taken as the origin of the map projection used by the Ordnance Survey for all topographical maps of England, Scotland and Wales. We defer the projection part of the problem to a later chapter. Here it is desirable to consider two properties of the grid, its orientation and the system of numbering along the axes.

The ordinate of the system is orientated so that it coincides with the meridian 2°W. It follows that since all meridians point towards true north (see Chapter 3 for justification of this statement), the ordinate of the National Grid also points towards true north. Since the grid is composed of families of straight lines, it follows that all other vertical grid lines point in the constant direction defined by the ordinate. This constant direction may be called grid north. On the other hand all meridians converge towards the geographical poles, therefore a meridian through a point lying east or west of longitude 2°W does not coincide everywhere with a grid line through the same point. This gives rise to the angular discrepancy between the meridians and grid lines which is illustrated

†Frequent reference will be made in this book to the labours of the United Kingdom Working Group on Terminology and to the preparation of the Glossary of Technical Terms in Cartography, published by the Royal Society in 1966. The definitions in that work were subsequently combined with other national contributions to the Multilingual Dictionary of Technical Terms in Cartography, published by ICA in 1973. The preferred terms relating to map projections which appear in those works are used throughout the book. Definitions which are those used in the Glossary are prefixed with the symbol *.
in a much exaggerated form in Fig. 2.03. The angle is known as grid convergence. Within the range in longitude occupied by southern England, the amount of convergence is small, for example it is 2° 54' near Lands' End and nearly 3° on the Norfolk coast.

The choice of the meridian 2°W as the longitude for the origin is simply because this lies near the middle of the part of the British Isles covered by the National Grid. It is a line which passes through the Isle of Purbeck in Dorset, through Birmingham, Berwick and Fraserburgh. From the sign convention used with graphs this means that everywhere in Britain lying to the west of the Birmingham-Berwick-Fraserburgh line, i.e. all Wales, most of Scotland and much of England, would be assigned negative Easting coordinates and referred to in this inconvenient way. The method of overcoming likely confusion is to imagine that the origin of the National Grid has been shifted westwards until the whole country lies in the first quadrant of the graph. In the example of the British National Grid the shift in origin is 400 km to the west and 100 km to the north of the point near the Channel Islands, so that zero on the National Grid lies at a point located about 80 km west of the Scilly Isles. This is equivalent to assigning the arbitrary coordinate values $E = 400\,000\, m$, $N = 100\,000\, m$ to the true origin and renumbering the grid lines. The point $E = 0\, m$, $N = 0\, m$ is referred to as the false origin of the grid to distinguish it from the point in latitude 49°N, 2°W which is the true origin. They way in which the shift has been applied may be imagined mathematically as the parallel shift of each axis through the defined distances. This is called translation of the axes.

**Plane polar coordinates**

Polar coordinates define position by means of one linear measurement and one angular measurement. The pair of orthogonal axes passing through the origin is replaced by a single line $OQ$, in Fig. 2.04, passing through the origin $O$, or pole of the system. The position of any point $A$ may be defined with reference to this pole and the polar axis or initial line, $OQ$ by means of the distance $OA = r$ and the angle $QOA = \theta$. The line $OA$ is known as the radius vector and the angle $\theta$ is the vectorial angle.
which the radius vector makes with the initial line. Hence the position of
A may be defined by the coordinates \((r, \theta)\). The order of referring to
the radius vector followed by the vectorial angle is standard to all branches
of pure and applied mathematics. The vectorial angle may be expressed
in sexagesimal (degree) or centesimal (grad) units to plot or locate a point
instrumentally. 

In the theoretical derivation of map projections, where \(\theta\) enters directly
into an equation and is not introduced as some trigonometric function
of the angle, it is necessary to express this angle in absolute angular units,
or radians. This is because both elements of the coordinate system must
have the character of length.

The direction in which the vectorial angle is measured depends upon the
purpose for which polar coordinates are used. Usually the mathematician
regards \(+\theta\) as the anticlockwise angle measured from the initial line. This
is the sign convention which is used, for example, in vector algebra. On
the other hand, the navigator, surveyor and cartographer are accustomed
to measure a positive angle in the clockwise direction. This is because
direction on the earth's surface is conventionally measured clockwise
from north or clockwise from a reference object. In many practical applica-
tions, formal recognition of the sign of an angle is unimportant because
the user can visualise the relationship between angles measured on the
360\(^\circ\) circle. However, difficulties arise in automatic data processing
because the standard subroutines, for example those to convert from
rectangular into polar coordinates, invariably use the mathematical con-
vention. This kind of calculation, which is described in the next section,
is extremely common in surveying and cartography. Consequently
the user of a computer or calculator must be aware of the difference in
convention, how the instrument deals with such data and write suitable
program steps which overcome the difficulty. Similarly in writing programs
for digital processing it is frequently necessary to introduce a series of
tests and conditional statements to allow uninterrupted processing of
data which have been collected according to the clockwise convention.
The simplest way of overcoming the difficulty is to interchange the axes,
so that the \(x\)-axis points towards the north. This is equivalent to a rotation

\*One right angle is represented by 90\(^\circ\) in sexagesimal notation, 100\(^\circ\) in centesimal units
or \(\pi/2\) radians. Many pocket calculators can operate in all three modes.

\[
\begin{align*}
\text{Fig. 2.04} & \quad \text{Plane polar coordinates.} \\
\text{Fig. 2.05} & \quad \text{The relationship between plane rectangular and plane polar co-
 ordinates with common origin and one common axis.}
\end{align*}
\]

plus a reflection of Fig. 2.05, which may be verified by tracing this diagram
on a piece of transparent plastic.

**Transformation from polar into rectangular coordinates and vice-versa**

Figure 2.05 illustrates the relationship between the rectangular and polar
coordinates of a point \(A\). The rectangular coordinates of the point are
\((x, y)\) referred to the origin \(O\) and the axes \(OX\) and \(OY\). Superimposed
upon this is a system of polar coordinates in which the pole also lies at
\(O\) and the initial line coincides with \(OY\). Then the polar coordinates of
\(A\) are \((r, \theta)\) where \(r = OA\) and \(\theta = \text{angle } YOA\), \(AN = x\) and
\(AM = NO = y\). It is evident from the right-angled triangle \(AON\) that

\[
\begin{align*}
x & = r \cdot \sin \theta \\
y & = r \cdot \cos \theta
\end{align*}
\]

(2.01)  \hspace{1cm} (2.02)

The inverse transformation from rectangular to polar coordinates can be
accomplished using a variety of different formulae. For example

\[
\begin{align*}
\tan \theta & = x/y \\
r & = y \cdot \sec \theta \\
r & = x \cdot \cosec \theta \\
r^2 & = x^2 + y^2 \\
\sin \theta & = x/r \\
\cos \theta & = y/r
\end{align*}
\]

(2.03)  \hspace{1cm} (2.04)  \hspace{1cm} (2.05)  \hspace{1cm} (2.06)  \hspace{1cm} (2.07)  \hspace{1cm} (2.08)
Note that these expressions are based upon the assumption that the angle $\theta$ has been measured 'clockwise-from-grid-north'. The coordinate expressions corresponding to these in most mathematical textbooks are derived from the complement of the vectorial angle, i.e. $AOX = 90^\circ - \theta$.

From the expressions which may be used to transform from rectangular to polar coordinates, the formulae (2.03) and either (2.04) or (2.05) used to be the most convenient in numerical work, and the reader would be warned against using Pythagoras' Theorem (2.06) to find the length of the radius vector because this was slow and inconvenient to calculate by logarithms. Nowadays most pocket calculators can be used to obtain square roots directly, so this caveat no longer applies.

**Two-dimensional coordinate transformations**

A series of numerical procedures which are commonly required in the mapping sciences are the two-dimensional linear transformations from one cartesian coordinate system into another. We provide here five examples of applications, and this list is by no means exhaustive. It includes:

1. Determination of the positions of intersections of a grid to be plotted on a map manuscript which has been compiled from and shows a different grid. This is necessary for mapping the zone of overlap between two grid systems and both of them have to be shown on the map.

2. Determination of the positions of intersections of a new grid to be plotted on a map manuscript originally compiled on a different grid which has been superseded. Now that most national surveys are based upon either the Universal Transverse Mercator (UTM) projection or the similar Soviet Unified Reference System (SURS), the need for this conversion is much less than it was in the early postwar decades, when many separate, or local, grid systems were still in use.

3. Conversion of the coordinate output of some other mapping process so that the results can be used with a particular grid. A typical example of this kind of work is when aerial triangulation has been carried out in an analogue photogrammetric plotter. The output from this includes a stream of (X, Y) *model coordinates* for control points which have been observed in the plotter and whose positions are recorded with respect to the axial movements of the plotter. These now have to be transferred to the same system as the map grid in order to fit the photogrammetric control to ground surveys. The concept of the *analytical plotter* which has more or less replaced the older analogue instruments is based upon continuous transformation from the plane of the aerial photograph to that of the map by digital methods.

4. Perhaps the most important application of all now arises in digital mapping, in the use of *vector* digitised map information to refer digitiser coordinates to the map grid. The majority of instruments functioning in

---

The vector mode comprise a special table containing the electronic hardware which converts the positions of a measuring mark mounted in a special cursor into rectangular coordinates defined by the manufacturer of the table. Information about position is obtained by pointing to or tracing the map detail (called *line-following*) with the measuring mark. The coordinates of a single point or points along a line are recorded and stored in digital form on tape or disc according to the (x, y) coordinate system built into the instrument. Hence the (E, N) grid of a topographical map is converted into the (x, y) coordinates of the digitiser and the precise relationship between the two depends upon the way in which the map sheet was placed upon and attached to the table. In order to reproduce any of the map detail in a desired form it is necessary to convert back from the (x, y) system of digitised coordinates into the (E, N) system of the map grid. This is usually done by digitising the four corners of the map and using these control points to determine the translation, rotation and scale change components of the transformation.

5. A second stage of this kind of digital mapping is contained in the need to change from one map projection to another, from a source map on one map projection to a new map which is compiled upon another. We consider this particular application of the two-dimensional transformations in detail in Chapter 19. Here we confine our attention to the two simplest methods:
The linear conformal, similarity or Helmert transformation, expressed in the general form:

\[
\begin{align*}
X &= A + Cx + Dy \\
Y &= B - Dx + Cy
\end{align*}
\]  

(2.09)

- The affine transformation:

\[
\begin{align*}
X &= A + Cx + Dy \\
Y &= B - Ex + Fy
\end{align*}
\]  

(2.10)

In these equations the known \((x, y)\) coordinates of a point in one system are transformed into the \((X, Y)\) coordinates of a second system, through the use of four or six coefficients \(A-F\). In the first we see that the \(C\) and \(D\) coefficients are common to both the equations for \(X\) and \(Y\), but in affine transformation it is necessary to introduce separate corrections for each direction. The risk of confusion of the coefficient \(E\) in equation (2.10) with the abbreviation for Easting should be noted.

**Linear conformal, similarity or Helmert transformation**

Both transformations may be resolved into three components:

- translation of the axes or change of origin, corresponding to the coefficients \(A\) and \(B\) in both equations (2.09 and 2.10);
- change in scale from one grid system to the other;
- rotation of the axes of one grid system with respect to their directions in the other.

The difference between the Helmert and affine transformations comes in the treatment of scale changes and rotations of the axes.

**Translation of the axes or change of origin**

We have already described this transformation for it has been used to introduce a false origin to a grid. This is simplest if the axes of the original system and those of the final system are parallel to one another as illustrated in Fig. 2.07. In this figure the point \(A\) has \((x, y)\) coordinates in the original system which has its origin at \(O\). We wish to refer the point to the second system in \((x', y')\) coordinates which have their origin at \(O'\). The differences between \(O\) and \(O'\) are the coordinate displacements \(x''\) and \(y''\). It follows that the new coordinates of \(A\) may be written

\[
\begin{align*}
x' &= x + x'' \\
y' &= y + y''
\end{align*}
\]  

(2.11)  

(2.12)

![Fig. 2.07 Translation of the axes of a plane rectangular coordinate system.](image)

The signs of \(x''\) and \(y''\) depend upon the direction in which the shift has been made. However, in dealing with grids of topographical maps, the false origin has usually been assigned to a position which lies to the south and west of any point likely to be referred to the grid, thereby avoiding the inconvenience of having negative grid references. It follows that normally \(x' > x\) and that \(y' > y\) so that \(x''\) and \(y''\) are both positive corrections. We may express the pair of equations (2.11) and (2.12) in the form of matrix addition,

\[
\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x'' \\ y'' \end{pmatrix}
\]  

(2.13)

**Change in scale from one coordinate system to another**

Consider two points, \(A\) and \(B\), which are common to two coordinate systems. In the first system the straight line \(AB\) joins the pair of points and in the second system the corresponding line is \(ab\). If \(AB \neq ab\), a scale factor must be introduced to convert coordinates in the first system into coordinates within the second system. This scale factor is

\[
m = ab/AB
\]  

(2.14)

from which it follows that

\[
\begin{align*}
x' &= m \cdot x \\
y' &= m \cdot y
\end{align*}
\]  

(2.15)  

(2.16)

In matrix notation this has the form

\[
\begin{pmatrix} x' \\ y' \end{pmatrix} = m \cdot \begin{pmatrix} x \\ y \end{pmatrix}
\]  

(2.17)

where the term \(m\) is appropriately called a *scalar*. A typical application
of this part of the transformation is the conversion of \((x, y)\) projection coordinates, which are given in units of earth radius, into the \((x', y')\) system of master grid coordinates which are needed to plot points on a master grid in millimetres. We shall see in Chapter 8 that this is the customary method of constructing a map to a required scale.

**Rotation of the axes about the origin**

We assume that the origin of each system is the same point, \(O\), but the axes have been rotated through the angle \(\alpha\). Thus \(OX\) becomes \(OX'\) and \(OY\) becomes \(OY'\), as illustrated in Figs 2.08 and 2.09. These two figures illustrate the difference between the clockwise and anticlockwise rotations of the axes. We shall study the effect of a clockwise rotation of the axes in detail.

If \(A = (x, y)\) in the first system it is required to determine its \((x', y')\) coordinates after rotation of the axes to form the second system. From equations (2.01) and (2.02) we know that \(x = r \cdot \sin \theta\) and \(y = r \cdot \cos \theta\), where \(\theta\) is the angle \(AOY\). Moreover the angle \(AOY' = \theta - \alpha\). Therefore

\[
\begin{align*}
  x' &= r \cdot \sin(\theta - \alpha) \\
  y' &= r \cdot \cos(\theta - \alpha)
\end{align*}
\]  

(2.18)  
(2.19)

The sine and cosine of the difference between two angles are well-known formulae from plane trigonometry. Here

\[
\begin{align*}
  \sin(\theta - \alpha) &= \sin \theta \cdot \cos \alpha - \cos \theta \cdot \sin \alpha \\
  \cos(\theta - \alpha) &= \cos \theta \cdot \cos \alpha + \sin \theta \cdot \sin \alpha
\end{align*}
\]  

(2.20)  
(2.21)

Substituting these expressions in equations (2.18) and (2.19)

\[
\begin{align*}
  x' &= r \cdot \sin \theta \cdot \cos \alpha - r \cdot \cos \theta \cdot \sin \alpha \\
  y' &= r \cdot \cos \theta \cdot \cos \alpha + r \cdot \sin \theta \cdot \sin \alpha
\end{align*}
\]  

(2.22)  
(2.23)

From equations (2.01) and (2.02) we may now substitute \(x\) and \(y\) for \(r \cdot \sin \theta\) and \(r \cdot \cos \theta\) respectively. Thus

\[
\begin{align*}
  x' &= x \cdot \cos \alpha - y \cdot \sin \alpha \\
  y' &= x \cdot \sin \alpha + y \cdot \cos \alpha
\end{align*}
\]  

(2.24)  
(2.25)

Note the order in which the terms for \(x\) and \(y\) are written. This corresponds to the rules governing the order in which terms and coefficients are written in matrices, so that these two equations have the matrix notation

\[
\begin{pmatrix}
  x' \\
  y'
\end{pmatrix} =
\begin{pmatrix}
  \cos \alpha & -\sin \alpha \\
  \sin \alpha & \cos \alpha
\end{pmatrix}
\begin{pmatrix}
  x \\
  y
\end{pmatrix}
\]  

(2.26)

The \(2 \times 2\) matrix containing the trigonometric coefficients is known as the **rotation matrix**. We turn now to the anticlockwise rotation of the axes illustrated by Fig 2.09, where the angle \(Y'O\overrightarrow{A} = \theta + \alpha\). Using the same arguments with the trigonometric expressions defining the sine and cosine of the sum of two angles, the final equations are:

\[
\begin{align*}
  x' &= x \cdot \cos \alpha + y \cdot \sin \alpha \\
  y' &= -x \cdot \sin \alpha + y \cdot \cos \alpha
\end{align*}
\]  

(2.27)  
(2.28)
which means that the rotation matrix is now

\[ R = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \] (2.29)

We observe that the two elements in \( \sin x \) have different signs and the position of \( -\sin \alpha \) has changed between (2.26) which refers to the clockwise rotation and (2.29) describing the anticlockwise rotation.

**Coordinate transformations involving all three displacements**

We may now combine the effects of all three displacements to produce the pair of equations

\[ x' = (m \cdot x \cdot \cos \alpha + m \cdot y \cdot \sin \alpha) + x'' \] (2.30)

\[ y' = (m \cdot x \cdot \sin \alpha + m \cdot y \cdot \cos \alpha) + y'' \] (2.31)

Several different versions may be used to express the result in matrix form. The simplest is to write

\[ \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} m \cdot \cos \alpha & m \cdot \sin \alpha \\ -m \cdot \sin \alpha & m \cdot \cos \alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x'' \\ y'' \end{pmatrix} \] (2.32)

In many survey applications there is a convention of writing \( P = m \cdot \sin \alpha \) and \( Q = m \cdot \cos \alpha \). Consequently the expression (2.32) may be written

\[ \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} Q & P \\ -P & Q \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x'' \\ y'' \end{pmatrix} \] (2.33)

The inverse transformation is that of determining the \((x, y)\) coordinates whose \((x', y')\) coordinates are already known. It may be required in converting from one map projection to another, because this is often a two-way process, as shown in Chapter 9. It can be shown that the inverse transformation corresponding to (2.33) is

\[ \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Q' & -P' \\ P' & Q' \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} \] (2.34)

where \( Q' = \cos \alpha/m \) and \( P' = \sin \alpha/m \).

**Affine transformation**

The assumption which is made in the Helmert transformation is that the scalar, \( m \), is a single, unique value. In other words the ratio \( ab/AB \) is the same whatever the directions of these lines. This is a reasonable assumption to make in some work, but it may not be justified for other jobs. For example in photogrammetry the location of image points on a film may be affected by deformation of the film base by stretching and shrinking, and this is not usually the same in all directions. In the extraction of positional information by digitising a paper map, the influence of differential stretching or shrinking of the paper must be considered. This may be large and unpredictable, as described by Maling (1989). For these applications it is desirable to use the affine transformation because this allows for different scales in the directions of the two axes, \( m_x \) and \( m_y \). This may also be combined with small departures of the coordinate axes from the perpendicular, as illustrated in Fig. 2.10. Here we see that the \((x, y)\) axes intersect at an angle \( \gamma \neq 90^\circ \). We need to determine six coefficients to solve equation (2.10).

**Grid-on-Grid Calculations**

The linear conformal transformation from one cartesian system to another is, as already stated, commonly used in cartography. From the nature of the first problem, all these transformations may be called Grid-on-Grid Calculations.

Although equation (2.33), with appropriate changes in notation from \( x \) to \( E \) and \( y \) to \( N \), specify the final equations need to convert from the known \((E', N')\) coordinates into the required \((E, N)\) values, it is still necessary to determine suitable numerical values for \( P, Q, E' \) and \( N' \).

Provided that there are at least two points which are common to both systems, these terms can be calculated and used to convert as many additional points as are required. The method of solving the unknowns
may be carried out as below:

Fig. 2.11 The Grid-on-Grid problem. Stage 1, defining the relationship of two points A and B, whose coordinates on both grids are already known. E' and N' denote the initial grid; E and N denote the second grid to which other points are to be transformed.

In Fig. 2.11 the two points A and B are common to both grids. We use the following notation to described each point:

<table>
<thead>
<tr>
<th>Point</th>
<th>1st grid</th>
<th>2nd grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>E', N'</td>
<td>E_a, N_a</td>
</tr>
<tr>
<td>B</td>
<td>E_b, N_b</td>
<td>E_b, N_b</td>
</tr>
</tbody>
</table>

The coordinate differences between the two points may be expressed as follows, using the convention that the Greek letter \( \delta \) signifies the difference between two coordinate values.

\[
\begin{align*}
&\text{1st grid} & & \text{2nd grid} \\
E'_a - E_b = \delta E' & & E_a - E_b = \delta E \\
N'_a - N_b = \delta N' & & N_a - N_b = \delta N
\end{align*}
\]

These terms have the geometrical significance which is illustrated in Fig. 2.11. Using arguments similar to those already used to determine the effects of rotation and scale change upon the coordinates, it can be shown that

\[
\begin{align*}
Q &= \frac{[\delta E \cdot \delta N' - \delta N \cdot \delta E']}{[\delta E^2 + \delta N'^2]} \\
P &= \frac{[\delta N \cdot \delta N' + \delta E \cdot \delta E']}{[\delta E^2 + \delta N'^2]}
\end{align*}
\]  

Fig. 2.12 The Grid-on-Grid problem. Stage 2, indicating the relationship of any point, \( P \), whose coordinates on the initial grid (\( E_a, N_a \)) are known, to the second grid upon which it must be plotted.

The translation terms \( E'', N'' \), corresponding to \( x'' \) and \( y'' \) in equations (2.33) etc. may be found from

\[
\begin{align*}
E'' &= E_a - P \cdot E'_a - Q \cdot N'_a \\
&= E_a - P \cdot E_b - Q \cdot N_b \\
N'' &= N_a + Q \cdot E'_a - P \cdot N'_a \\
&= N_a + Q \cdot E_b - P \cdot N_b
\end{align*}
\]

Hence the required equations to transform the (\( E', N' \)) coordinates of any other point, \( P \) (Fig. 2.12) to the (E, N) system are

\[
\begin{align*}
E &= Q \cdot E' + P \cdot N' + E'' \\
N &= -Q \cdot E' + P \cdot N' + N''
\end{align*}
\]

which, converted into matrix notation provides an expression like (2.33).

The equations (2.35)–(2.38) have been given here without proof, but their derivation can be found, for example, in Ministry of Defence (1978). In Admiralty (1965) there is also described the method of solving the coefficients when there are three points common to both systems. If there are more than three common points, such as occurs in vector digitising and in the adjustment of aerial triangulation to many ground control points, the determination of the coefficients from only two or three of them is inadequate because the coordinates of any of those points may contain small errors and the use of them will introduce error into the transformation of all other points. Under these circumstances all of the data which are available for the determination of \( P \) and \( Q \) ought to be
Coordinate Systems and Map Projections

taken into consideration. This involves a solution of the coefficients by
the methods of least squares, which is a more sophisticated numerical
solution based upon statistical error analysis.

The best procedure is to translate the axes of both system to a common
origin at the centroid of the n points, obtained simply by determining the
mean value of each coordinate. Thus for n points, labelled \( i = 1 \ldots n \),
\[
E_G = \frac{\sum E_i}{n} \quad \quad \quad \quad (2.43)
\]
\[
N_G = \frac{\sum N_i}{n} \quad \quad \quad \quad (2.44)
\]
with similar determinations for \( E_0 \) and \( N_0 \).

The individual coordinates, \( E_i, N_i, E', N' \) are now referred to these
centroids as origin and the analysis of the most probable values for \( P \)
and \( Q \) derived by standard routines. Modern textbooks on survey
adjustments and computations, e.g. Hirvonen (1971), Cooper (1974), Mikhail (1976),
Mikhail and Gracie (1981), and Methley (1986) all deal with the subject,
and this book deals later (Chapter 19) with polynomial transformations,
of which these are elementary examples.

The reader who is particularly concerned with the adjustment of vector
digitised coordinates measured from paper maps which may also have
been folded is referred specifically to the important paper by Sprinsky

CHAPTER 3

Coordinate reference systems on the sphere

"What's the good of Mercator's North Poles and Equators,
Tropics, Zones and Meridian lines?"
So the Bellman would cry: and the crew would reply
"They are merely conventional signs."

Lewis Carroll, The Hunting of the Snark

Introduction

It has been assumed in Chapters 1 and 2 that the reader already knows
something about the terms which are used to describe planes, arcs and
angles on the earth. For example, the idea of latitude and longitude;
parallels and meridians and the convergence of the meridians have been
introduced without formal definition. However, it is desirable to consider
these definitions and develop further our knowledge about the geometry
of the earth. There are two reasons for this. First, we need to introduce
a standardised system of algebraic notation for the different quantities
which will be used throughout this book. Secondly it is necessary to
demonstrate certain important geometrical differences between the sphere
and the spheroid. In order to appreciate the distinctions to be made
between these bodies it is essential to know precisely what is represented
by planes, arcs and angles on each of them.

Some of the properties of a sphere have already been described
in Chapter 1. These may be summarised as a preliminary to further
definitions:

- A sphere is a solid body whose curved surface is everywhere equi-
distant from its centre.
- It follows that any sphere has constant radius.
- If a tangent plane meets any point on the curved surface, a line
  normal to this plane at the point of tangency is a radius to the centre
  of the sphere.
- The distance between two points on the sphere can be defined and
  measured either as the angular distance or the arc distance. There is