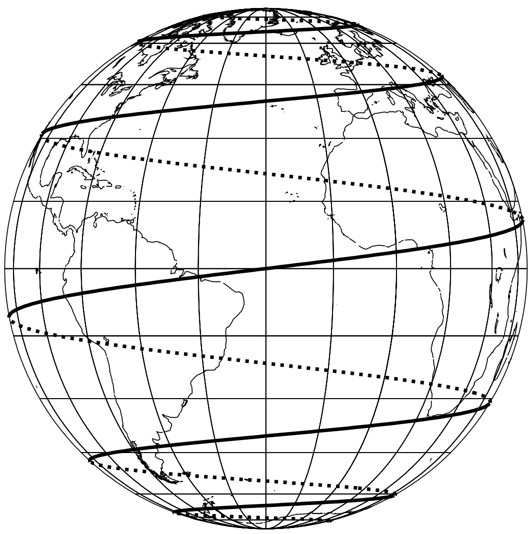
THE MERCATOR PROJECTIONS

THE NORMAL AND TRANSVERSE MERCATOR PROJECTIONS ON THE SPHERE AND THE ELLIPSOID WITH FULL DERIVATIONS OF ALL FORMULAE



PETER OSBORNE

EDINBURGH

2008

The Transverse Mercator projection is the basis of many maps covering individual countries, such as Australia and Great Britain, as well as the set of American UTM projections covering the whole world (other than the polar regions). Such maps are invariably covered by a set of grid lines. It is important to appreciate the following two facts about the Transverse Mercator projection and the grids covering it:

1. Only one grid line runs true north–south. Thus in Britain only the grid line coincident with the meridian at 2°W is true: all others are slightly distorted. The UTM series is a set of 60 projections covering a width of of 6°in latitude: the grid lines run true north–south only on the central meridians at 3°E, 9°E, 15°E, ...

2. The scale on the maps derived from Transverse Mercator projections is not uniform: it is a function of position. For example the Landranger maps of the Ordnance Survey of Great Britain have a nominal scale of 1:50000: this value is only exact on two slightly curved lines almost parallel to the central meridian at 2°W.

The above facts are unknown to the majority of map users. They are the subject of this article together with the presentation of formulae relating latitude and longitude to grid coordinates.

Preface

For many years I had been intrigued by the the statement on the (British) Ordnance Survey maps pointing out that the grid lines are not exactly aligned with meridians and parallels: four precise figures give the magnitude of the deviation at each corner of the map sheets. My first retirement project has been to find out exactly how these figures are calculated and this led to an exploration of all aspects of the Transverse Mercator projection on an ellipsoid of revolution (TME). This projection is also used for the Universal Transverse Mercator series of maps covering the whole of the Earth, except for the polar regions.

The formulae for TME are given in many books and web pages but the full derivations are only to be found in original publications which are not readily accessible: therefore I decided to write a short article explaining the derivation of the formulae. Pedagogical reasons soon made it apparent that it would be necessary to start with the normal and transverse Mercator projection on the sphere before going on to discuss the normal and transverse Mercator projection on the ellipsoid. As a result the length of this document has doubled and redoubled, but I have resisted the temptation to cut out the details which would be straightforward for a professional but daunting for a 'layman'. The mathematics involved is not difficult (depending on your point of view) but it does require the rudiments of complex analysis for the crucial steps. On the other hand the algebra gets fairly heavy at times; Redfearn (see bibliography) talks of a "a particularly tough spot of work" and Hotine talks of reversing series by "brute force and algebra"—so be warned. To make this article as self-contained as possible I have added a number of appendices covering the required mathematics.

My sources for the TME formulae are to be found in Empire Survey Review dating from the nineteen forties to sixties. The actual papers are fairly terse, as is normal for papers by professionals for their peers, and their perusal will certainly not add to the details presented here. Books on mathematical cartography are fairly thin on the ground, moreover they usually try to cover all types of projections whereas we are concerned only with Mercator projections. The few that I found to be of assistance are listed in the bibliography.

I would like to thank Harry Kogon for reading, commenting on and even checking the mathematics outlined in these pages. Any remaining errors (and typographical slips) must be attributed to myself—when you find them please send an email to the address below.

Peter Osborne

Edinburgh, 2008

peter.1@mercator.myzen.co.uk

Contents

Chapter 1

Introduction

Geodesy and the Figure of the Earth

Geodesy is the science concerned with the study of the exact size and shape of the Earth in conjunction with the analysis of the variations of the Earth's gravitational field. This combination of topics is readily appreciated when one realizes that (a) in traditional surveying the instruments were levelled with respect to the gravitational field and (b) in modern satellite techniques we must consider the satellite as an object moving freely in the gravitational field of the Earth. Geodesy is the scientific basis for both traditional triangulation on the actual surface of the earth and modern surveying using GPS methods.

Whichever method we use, traditional or satellite, it is vital to work with well defined reference surfaces to which measurements of latitude and longitude can be referred. Clearly, the actual topographic surface of the Earth is very unsuitable as a reference surface since it has a complicated shape, varying in height by up to twenty kilometres from the deepest oceans to the highest mountains. A much better reference surface is the gravitational equipotential surface which coincides with the mean sea level continued under the continents. This surface is called the **geoid** and its shape is approximately a flattened sphere but with many slight undulations due to the gravitational irregularities arising from the inhomogeneity in the Earth's crust.

However, for the purpose of high precision geodetic surveys, the undulating geoid is not a good enough reference surface and it is convenient to introduce a mathematically exact reference surface which is a good fit to the shape of the geoid. The surface which has been used for the last three hundred years is the oblate **ellipsoid of revolution** formed when an ellipse is rotated about its minor axis. We shall abbreviate 'ellipsoid of revolution' to simply **ellipsoid** in this article, in preference to the term **spheroid** which is used in much of the older literature. (We shall not consider triaxial ellipsoids which do not have an axis of symmetry). The shape and size of the reference ellipsoid which approximates the geoid is usually called the **figure of the Earth**.

The earliest accurate determinations of the figure of the earth were made by comparing two high precision meridian arc surveys, each of which provided a measure of the distance along the meridian per unit degree at a latitude in the middle of each arc. Two such measurements, preferably at very different latitudes, are sufficient to determine two parameters which specify the ellipsoid—the major axis a together with the minor axis b or, more usually, the combination of the major axis with the flattening f (defined below). For example, in the first half of the eighteenth century, French scientists measured a meridian arc of about one degree of latitude in Lapland (crossing the Arctic circle) and a second arc of about three degrees of latitude in Peru (crossing the equator) and confirmed for the first time the oblateness of the ellipsoid. In 1830 Everest calculated an ellipsoid using what he took to be the best two arcs, an earlier Indian Arc surveyed by his predecessor Lambton and once again the arc of Peru. As more and longer arcs were measured the results were combined to give more accurate ellipsoids. For example Airy discussed sixteen arcs before arriving at the result he published in 1830:

$$a = 6377563.4\text{m} \qquad b = 6356256.9\text{m} \qquad f = 1/299.32 \qquad \text{[Airy1830]} \qquad (1.1)$$

where the **flattening** f, defined as $(a - b)/a$, gives a measure of the departure from the sphere. Similarly Clarke used eight arcs to arrive at his 1866 ellipsoid:

$$a = 6378206.4\text{m} \qquad b = 6356583.8\text{m} \qquad f = 1/294.98 \quad \text{[Clarke1866]} \qquad (1.2)$$

Modern satellite methods have introduced global ellipsoid fits to the geoid, that for the World Geodetic System of 1972 being

$$a = 6378135\text{m} \qquad b = 6356750.5\text{m} \qquad f = 1/298.26 \qquad \text{[WGS72]} \qquad (1.3)$$

and that for the Geodetic Reference System of 1980 (GRS80) being

$$a = 6378137\text{m} \qquad b = 6356752.3\text{m} \qquad f = 1/298.26 \qquad \text{[GRS80]} \qquad (1.4)$$

Further satellite models are under development.

There are many ellipsoids in use today and they differ by no more than a kilometre from each other, with an equatorial radius of approximately 6378km (3963 miles) and a polar radius of 6356km (3949 miles) shorter by approximately 22km (14 miles). Note that modern satellite ellipsoids, whilst giving good global fits, are actually poorer fits in *some* regions surveyed on a best-fit ellipsoid derived by traditional (pre-satellite) methods.

The bibliography lists some web surveys of geodesy and also one or two advanced textbooks which give detailed coverage of the subject.

Topographic surveying

The aim of a topographic survey is to provide highly accurate maps of some region referenced to a specific **datum**. By this we mean a choice of a definite reference ellipsoid together with a precise statement as to how the ellipsoid is related to the area under survey.

For example we could specify how the centre of the selected ellipsoid is related to the chosen origin of the survey and also how the orientation of the axes of the ellipsoid are related to the vertical and meridian at the origin. It is very important to realize that the choice of datum for any such survey work is completely arbitrary as long as it is a reasonable fit to the geoid in the region of the survey. The chosen datum is usually stated on the final maps.

As an example, the maps produced by the Ordnance Survey of Great Britain (OSGB) are defined with respect to a datum known as OSGB36 (established for the 1936 re-triangulation) which is still based on the Airy 1830 ellipsoid which was chosen at the start of the original triangulation in the first half of the nineteenth century. This ellipsoid is indeed a good fit to the geoid under Britain but it is a poor fit everywhere else on the globe so it is not used for mapping any other country. The OSGB36 datum defines how the Airy ellipsoid is related to the the ground stations of the survey. Originally, in the nineteenth century, the origin was chosen at Greenwich observatory but, for the 1936 re-triangulation no single origin was chosen, rather the survey was adjusted so that the latitude and longitude of 11 control stations remained as close as possible to their values established in the original nineteenth century triangulation.

Until 1983, the United States, Canada and Mexico used the North American datum established in 1927, namely NAD27. This is based on the Clarke (1866) ellipsoid tied to an origin at Meades Ranch in Kansas where the latitude, longitude, elevation above the ellipsoid and azimuth toward a second station (Waldo) were all fixed. Likewise, much of south east Asia uses the Indian datum, ID1830, which is based on the Everest (1830) ellipsoid tied to an origin at Kalianpur. The modern satellite ellipsoids used in datums such as WGS72, GRS80, WGS84 are defined with respect to the Earth's centre of mass and a defined orientation of axes.

In all, there are two or three hundred datums in use over the world, each with a chosen reference ellipsoid attached to some origin. The ellipsoids used in the datums do not agree in size or position and a major problem for geodesy (and military planners in particular) is how to tie these datums together so that we have an integrated picture of the world's topography. In the past datums were tied together where they overlapped but now we can relate each datum to a single geocentric global datum determined by satellite.

Once the datum for a survey has been chosen we would traditionally have proceeded with a high precision triangulation from which, by using the measured angles and baseline, we can calculate the latitude and longitude of every triangulation station from assumed values of latitude and longitude at the origin. Note that it is the latitude and longitude values on the reference ellipsoid 'beneath' every triangulation station that are calculated and used as input data for the map projections. It is important to realise that once a datum has been chosen for a survey in some region of the Earth (such as Britain or North America) then it should not be altered, otherwise the latitude and longitude of every feature in the survey region would have to be changed (by recalculating the triangulation data). But this has already happened and it will happen again. For example the North American datum NAD27 was replaced by a new datum NAD83 necessitating the recalculation of all coordinates, with resulting changes in position ranging from 10m to 200m. If (when) we use one of the new global datums fitted by satellite technology as the basis for new maps then the latitude and longitude values of every feature will change slightly again.

Cartography

A topographic survey produces a set of geographical locations (latitude and longitude) referenced to some specified ellipsoid. We are then faced with the problem of cartography, the representation of the latitude–longitude data on the ellipsoid by a two-dimensional map. There are an infinite number of projections which address this problem but in this article we consider only the normal (N) and transverse (T) Mercator projections, first on the sphere (S) and then on the ellipsoid (E). We shall abbreviate these projections as NMS, TMS, NME and TME: they are considered in full detail in Chapters 2, 3, 6 and 7 respectively. At a later date Chapter 10 may cover the oblique Mercator projection. Details of all these projections are given, *without derivations*, in the book by Snyder entitled 'Map Projections—A Working Manual'. (See bibliography).

We define a map **projection** by giving two functions $x(\phi, \lambda)$ and $y(\phi, \lambda)$ which specify the plane Cartesian coordinates (x, y) corresponding to the latitude and longitude coordinates $(\phi,\ \lambda)$. For the above projections, other than the oblique Mercator, the fundamental origin is taken as a point O on the equator, the positive x-axis is taken as the eastward direction of the projected equator and the positive y-axis is taken as the northern direction of the projected meridian through O. This convention agrees with that used in Snyder's book but beware other conventions! Many older texts, as well as most current 'continental' sources, adopt a convention with the x-axis as north and the y-axis as sometimes east and sometimes west!

The criteria for a faithful map projection

There are several basic criteria for a faithful map projection but it is important to understand that it is *impossible* to satisfy all these criteria at the same time. This is simply a reflection of the fact that it is impossible to deform a sphere or ellipsoid into a plane without creases or cuts. Thus all *maps* are compromises to some extent and they must fail to meet at least one of the following.

1. *One-to-one correspondence of points*. This will normally be the case for large scale maps of small regions but maps of the whole Earth will usually fail this criterion. Points at which the map fails to be one-to-one are called singular points. For example, in the normal Mercator projection we shall see that the poles are singular because they project into lines.

2. *Uniformity of point (or local) scale*. By point scale we simply mean the ratio of the distance between two nearby points on the map and the corresponding points on the ground. Ideally the point scale factor should have the same value at all points. This criterion is *never* satisfied. In the Mercator projections the scale is 'true' only on two lines at the most.

3. *Isotropy of point scale*. Ideally the scale factor would be isotropic (independent of direction) at any point and as a corollary the shape of any *small* region would be unaltered—such a projection is said to be **orthomorphic** (right shape). By 'small' we mean that, at some level of measurement accuracy, the magnitude of the scale does not vary over the small region. This condition *is* satisfied by the Mercator projections.

4. *Conformal representation*. Consider any two lines on the surface of the Earth which intersect at a point P at an angle θ. Let P' and θ' be the corresponding point and angle on the map projection. The map is said to be conformal if $\theta = \theta'$ at all non-singular points of the map. This has the consequence that the shape of a local feature (such as a short stretch of coastline or a river) is well represented even though there will be distortion over large areas. All of the Mercator projections satisfy this criterion.

5. *Equal area*. We may wish to demand that equal areas on the Earth have equal areas on the projection. This is considered to be 'politically correct' by many proponents of the Peters projection but the downside is that such equal area projections distort shapes in the large. The Mercator projections are not equal area projections and they also distort shapes.

In summary the normal Mercator projection has the properties: (a) there are singular points at the poles, (b) the point scale is isotropic (so the map is orthomorphic) but the magnitude of the scale varies with latitude, being true on two parallels at most, (c) the projection is conformal, (d) the projection does not preserve area. The transverse Mercator projection has the properties: (a) there are singular points on the equator, (b) the scale is isotropic (so the map is orthomorphic) with magnitude varying with *both* latitude and longitude, being true on at most two *curved* lines which cannot be identified with parallels or meridians, (c) the projection is conformal, (d) the projection does not preserve area.

Scale factors and representation factors (RF)

The OSGB produces many series of maps of Great Britain. For example there are over two hundred 'Landranger' map sheets with each 80cm×80cm sheet covering an area of 40km×40km on the ground. This could be taken as implying that the scale of these maps is exactly 2cm to 1km or 1:50000. However, as we pointed out in the last section, this can only be an approximate statement since it is impossible to construct a two-dimensional map of any region of the Earth at a uniform scale. Thus 1:50000 is only a 'nominal' scale for these map sheets, although admittedly the scale variation over any sheet is very small.

Scale variation over map projections is an important topic in the forthcoming chapters. Now it would be clumsy to discuss such variation around a nominal factor of say 1:50000. To this end the projection formulae $x(\phi, \lambda)$ and $y(\phi, \lambda)$ giving the plane Cartesian coordinates will be defined in such a way that the scale will be close to 1:1 where possible. Clearly this means that the projection formulae define a very large hypothetical **super-map**. (My terminology.) For example, the formulae for a normal Mercator projection of the globe generate a mapping coordinate x which will range over an interval of about 24900 miles to give a scale factor of unity on the equator (and increasing as $\sec\phi$ with latitude). Or again, for a projection of Britain, the super-map would be 600km by 1200km and, if we demand that the scale be unity on the central meridian at 2°W, we shall find that the scale factor nowhere exceeds 1.001.

Once we have our mathematical super-map embodying a varying scale (but close to unity) we can construct the *actual* maps for printing by a uniform scaling of the projection coordinates by a constant **representation factor (RF)**, constant that is for a given series of maps. For example the OSGB uses 1:25000, 1:50000, 1:625000 and other values. Thus the RF only arises at the printing stage and we can forget all about it and work with the mathematical super-map for the theoretical analysis of the projections.

Finally, note the usage that a printed map is 'large scale' when the RF, considered as a mathematical fraction, is 'large' and the map covers a small area. The OSGB 1:50000 maps are considered to be in this category and the 1:5000 series are of even larger scale. Conversely small scale maps having a small RF, say 1:1000000 (or simply 1:1M), are used to cover greater regions.

Graticules, grids, azimuths and bearings

The set of meridians and parallels on the reference ellipsoid is called the **graticule**. There is no obligation to show the projection of the graticule on the map but it is usually shown on small scale maps covering large areas and it is usually omitted on large scale maps of small areas. For the OSGB Landranger 1:50000 series there is no graticule but small crosses indicate the intersections of the graticule at $5'$ intervals on the sheet and latitude and longitude values are indicated at the edges.

The projected map is constructed in a plane Cartesian system but once again there is no obligation to show a **reference grid** of lines of constant x and y values. In general small scale maps are not embellished with a grid whereas large scale maps usually do have such a reference grid. The OSGB sheets have a grid at a 2cm intervals on the 1:50000 series so they correspond to a nominal (but not exact) spacing of 1km. Note that *any* kind of grid may be superimposed on a map to meet a user's requirements: it need not be aligned to the Cartesian projection axes.

On the graticule the angle between the meridian at any point A and another short line element AB is called the **azimuth** of that line element. Our convention is that azimuths are measured clockwise from north but other conventions exist. For example, in the past, and occasionally in the present, azimuth has been measured clockwise from south.

On a projection endowed with a grid the angle between the grid line through the projected position of A and the projection of the line AB is called the **grid bearing**. The normal Mercator projections is designed to ensure that the azimuth and grid bearing are equal (if the grid is aligned to the meridians). On the transverse Mercator projections this is not so: the azimuth differs from the grid bearing by a small amount which is termed the **grid convergence**. The difference is tiny but nonetheless it exists: the OSGB 1:50000 map sheets state the value of the grid convergence at each corner of the sheet.

Historical note

Let us state at the outset that **Gerardus Mercator** (1512–1594) did not develop the mathematics that we shall present for "his" projection (NMS); moreover he had nothing at all to do with three other projections that now carry his name—TMS, NME, TME. In 1569 he published his map-chart entitled "Nova et aucta orbis terrae descriptio ad usum navigantium ementate accommadata" which may be translated as "A new and enlarged description of the Earth with corrections for use in navigation". His full explanation is given on the map-chart:

> In this mapping of the world we have [desired] to spread out the surface of the globe into a plane that the places should everywhere be properly located, not only with respect to their true direction and distance from one another, but also in accordance with their true longitude and latitude; and further, that the shape of the lands, as they appear on the globe, shall be preserved as far as possible. For this there was needed a new arrangement and placing of the meridians, so that they shall become parallels, for the maps produced hereto by geographers are, on account of the curving and bending of the meridians, unsuitable for navigation. Taking all this into consideration, we have somewhat increased the degrees of latitude toward each pole, in proportion to the increase of the parallels beyond the ratio they really have to the equator. (Translation from Fite and Freeman—see bibliography).

This is an admirably clear statement and the last two sentences make clear his approach. In order that the meridians should be perpendicular to the equator, and parallel to each other, it is first necessary to increase the map length of a parallel as one moves away from the equator. Now at latitude ϕ the circumference of a parallel is $2\pi a\cos\phi$ and this must be scaled up by a factor of $\sec\phi$ so that the parallel and the equator have the same map length ($2\pi a$). Thus to preserve the shape of say a small rectangle at some latitude, projected from ground to map, it is necessary to increase the meridian scale *at that latitude* by a factor of $\sec\phi$. Exactly how Mercator produced his map is not known. He had had a good mathematical education but in 1559 he would not have had access to tables of the secant function to aid him. So *perhaps* he simply drew rhumb lines on the globe from various points on the equator, and at various azimuths, and took note of which locations on the globe lay on these rhumb lines. He could then adjust the ordinate scale of his projection so that all the locations on any rhumb line on the globe lay on a *straight* line on the projection. Alternatively, it has been suggested that he modified the parallels at ten degree intervals so that projected rhumb lines were approximately straight for each ten degree interval. Whatever method he used it is clear that he had grasped what was required, but his projection may have lacked high accuracy.

Mercator's prime aim was to construct a useful navigator's chart but note that in the above statement he was also concerned that shapes "shall be preserved as far as possible"; he understood the desirability of an orthomorphic projection. He would have noticed the distortion at high latitudes but he was probably satisfied that the appearance at temperate latitudes (Europe) was really quite good. He may have regretted the distortions to landmass shapes but this was far outweighed by the utility of his projection for navigators.

Mercator was extremely secretive about how he had produced his map-chart but this only stimulated others to research their construction. The first to succeed, and publish (albeit unwillingly), was a Cambridge professor of mathematics named **Edward Wright** (1558?–1615). Interestingly his 1599 publication is entitled 'The correction of certain errors in navigation'. He refers to certain errors in Mercator's chart and explains how they could be corrected by using a table of secants. More precisely he realized that if the parallel at latitude ϕ has to move up by a factor proportional to $\sec\phi$ then the net displacement of a parallel from the equator would be given by summing all of the secants from the equator to the parallel in question. In modern parlance this requires an integration of the secant function (see equation 2.28). Of course Wright (like Mercator) was working in pre-calculus and pre-logarithm days and integration of the secant function would not have been possible. Instead, he presented his results in numerical tables of 'cumulative secants' or 'meridional parts' derived by summation from secant tables evaluated at intervals of 1 minute of arc. In other words he carried out a numerical integration with a very fine sub-division.. It is possible that the errors he claims in Mercator's chart were attributable to the fact that Mercator's method was equivalent to a much coarser approximation to the integration. Wright's tables certainly allowed the construction of a *very* accurate chart based on a latitude and longitude values taken from a rather fine globe modelled by his compatriot Emery Molyneux. For many years thereafter the charts were widely described as Wright-Molyneux map projections but the name of Mercator later became the standard appellation.

Wright gives a nice physical construction of the Mercator projection from a sphere.

> Suppose a sphericall superficies with meridians, parallels, rumbes, and the whole hydrographical description drawne thereupon, to be inscribed into a concave cylinder, their axes agreeing in one. Let this sphericall superficies swel like a bladder, (while it is in blowing) equally always in every part thereof (that is, as much in longitude as in latitude) till it apply, and join itself (round about and all alongst, also towards either pole) unto the concave superficies of the cylinder: each parallel on this sphericall superficies increasing successively from the equinoctial [equator] towards either pole, until it come to be of equal diameter with the cylinder, and consequently the meridians still wideening themselves, til they become so far distant every where each from other as they are at the equinoctial. Thus it may most easily be understood, how a sphericall superficies may (by extension) be made cylindrical, . . .

It is easy to see that it works. Mercator's projection is constructed to preserve angles by stretching meridians to compensate exactly for the streching of the parallels. The angle preserving projection is conformal. Now consider Wright's bladder: that it must be infinitely extensible and able to withstand infinite pressure goes without saying. The crucial phrase is "swel . . . equally always in every part thereof". Therefore the tensions over both the initial spherical surface and the final cylindrical surface are uniform, albeit of very different magnitudes. This uniformity guarantees that a crossing of two lines on the sphere will be at at exactly the same angle on the cylinder. Thus we have generated a conformal projection from the sphere to the cylinder. And there is only one such conformal projection.

The logarithm function was invented by **Napier** in 1614 and numerical tables of many logarithmic functions were soon readily available (although analytic Taylor expansions of functions had to wait another hundred years). In the 1640s, another English mathematician called **Henry Bond** (1600–1678) stumbled on the numerical agreement between Wright's tables and those for $\ln[\tan(\theta)]$, as long as θ was identified with $(\phi/2 + \pi/4)$. The mathematical proof of the equivalence immediately became noted as an important problem but it was nearly thirty years before it was solved by **James Gregory** (1638–1675), **Isaac Barrow** (1630–1677) and **Edmond Halley** (1656–1742) acting independently. These proofs eventually coalesced into direct integration of the secant function as presented in Chapter 2. The modification of this integration for the ellipsoid (and NME) would be trivial. (But when? By whom?)

Having given credit to Wright, Bond and others we must now remark that the first to succeed, but not to publish, was almost certainly an English mathematician (and much else besides) called **Thomas Harriot** (1560–1621). He left a large collection of unpublished manuscripts and many years late it became obvious that he had probably duplicated Wright's calculation of meridional parts and moreover appears to have the link to the logarithmic tangent.

The transverse Mercator projection on the sphere was included in a set of seven new projections published in 1772 by a continental (born in Alsace-Lorraine) mathematician and cartographer, **Johann Lambert** (1728–1777). As we shall see in Chapter 3, the derivation of this projection is a straightforward application of spherical trigonometry starting from the normal Mercator result. Apparently Lambert even made some oblique references to the transverse projection on the ellipsoid but it was Carl Friedrich Gauss (1777–1855) who first constructed a conformal projection from the ellipsoid which preserves true scale on one meridian, the projection we shall term TME. (This was in connection with the survey of Hanover commenced in 1818).

Gauss's method involved a double projection, from ellipsoid to sphere and then sphere to plane. The modification of his work to construct a single equivalent projection was developed only as late as 1912 by **L Krüger**. For this reason the transverse Mercator projection on the ellipsoid is often called the Gauss–Krüger projection. This is the method we shall examine.

The transverse Mercator projection was not much used until the middle of the twentieth century when it was advocated for both the new British maps and the proposed world wide system (UTM). In Britain the need for more precise series for the TME projections was met by the papers of **M Hotine**, **L P Lee** and **J C B Redfearn** (see bibliography). The last mentioned produced the most complete form and the solution is often referred to as the Redfearn series. We shall derive these series in full detail.

Outline of following chapters

Chapter 2 starts by describing what we mean by an infinitesimal element on the sphere and goes on to use the planar geometry of such an element (a) to calculate the metric giving the distance between infinitesimally close points, and (b) to define precisely the azimuth angle. We then consider the class of all 'normal' cylindrical projections onto a cylinder tangential to the equator of a sphere and compare and contrast three important examples.

The Mercator projection on the sphere (NMS) is defined as the single member of the class which is such that an azimuth on the sphere and its corresponding grid bearing on the map are equal. This property of conformality is then used to derive the projection formulae. The scale is true only on the equator for basic NMS but we show how it may be modified to give true scale on two parallels instead.

Chapter 3 discusses the transverse Mercator projection on the sphere (TMS). In this case we are considering a projection onto a cylinder which is tangential to the sphere on a pair of meridians which together form a great circle, such as 90°E and 90°W. These projections are rather unusual when applied to the whole globe but in practice we intend to apply them to a narrow strip on either side of the meridian of tangency which is then termed the central meridian of the transverse projection. The crux is that by considering a large number of such projection strips we can cover the whole sphere (except near the poles) with good accuracy. The derivation of the projection formulae is a straightforward exercise in spherical trigonometry. An important new feature is that corresponding azimuths and grid bearings are not equal (even though the transformation remains conformal) and we define their difference as the grid convergence. Finally we present low order series expansions for the projection formulae.

Chapter 4 is the crunch. Our ultimate aim is to derive the projection equations for the transverse Mercator projection on the ellipsoid (TME) in the form of series expansions. The only satisfactory way of obtaining these results is by using a small amount of complex variable theory. This method is complicated by both the geometrical problems of the ellipsoid and also by the fact that we need to carry the series to many terms in order to achieve the required accuracy. Thus, for purely pedagogical reasons, in this chapter we use the complex variable methods to derive the low order series solutions for TMS (derived in Chapter 3) from the standard solution for NMS. That it works is encouragement for proceeding with the major problem of constructing the TME projections from NME.

Chapter 5 derives the properties of the ellipse and ellipsoid that are required in later chapters. In particular we introduce (a) the principal curvatures in the meridian plane and its principal normal plane, (b) the distinction between geocentric, geodetic and reduced latitudes, (c) the distance metric on the ellipsoid and (d) the series expansion which gives the distance along the meridian as a function of latitude.

Chapter 6 derives the normal Mercator projection (NME) on the ellipsoid. The method is a simple generalization of the methods used in Chapter 2 the only difference being in the different form of the infinitesimal distance element on the ellipsoid. The results for the projection equations are obtained in non-trivial closed forms. The inversion of these formulae is not possible in closed form and we must revert to Taylor series expansions. This chapter also contains a digression on 'double' projections and includes a discussion of the transformation of the ellipsoid to the sphere by means of the conformal latitude. (The word 'double' signifying that a second transformation from the sphere to the plane is required to produce a map). As a corollary, the conformal latitude is used to provide a second means of inverting the NME projection formulae.

Chapter 7 uses the techniques developed in Chapter 4 to derive the transverse Mercator projection on the ellipsoid (TME) from that of NME. This derivation requires distinctly heavy algebraic manipulation to achieve our main result, the Redfearn formulae for TME.

Chapter 8 returns to the definitions of point-scale factor and grid convergence presented in Chapter 3 (for TMS) and derives the corresponding results for TME as series expansions. Once again the algebra is fairly heavy.

Chapter 9 applies the general results of Chapters 7 and 8 for the TME projections to two important cases, namely the Universal Transverse Mercator (UTM) and the National Grid of Great Britain (NGGB). The former is actually a set of 60 TME projections each covering 6 degrees of longitude between the latitudes of 80°S and 84°N and the latter is a single projection over approximately 10 degrees of longitude centred on 2°W and covering the latitudes between 50°N and 60°N. We then discuss the variation of scale and grid convergence over the regions of the projection and also assess the accuracy of the TME formulae by examining the terms of the series one by one. We find that for practical purposes some terms may be dropped, as indeed they are in both the UTM and NGGB formulae. Finally the projection formulae are rewritten in the completely different notation used in the OSGB published formulae (see bibliography).

Finally, this work is continuing and two further chapters are intended:

Chapter 10 will cover the derivation of the oblique Mercator projections.

Chapter 11 will not be concerned with projections. It will discuss geodesics on the sphere and ellipsoid and the problem involved in calculating arbitrary distances accurately. In particular we will give the derivation of the Vincenty formulae for long geodesics on the ellipsoid.

This article has tried to be as self-contained as possible and to this end there are seven mathematical appendices. Many of these were developed for other uses so they are more general in nature.

- **A** Curvature in two and three dimensions.
- **B** Inversion of series by Lagrange expansions.
- **C** Plane Trigonometry.
- **D** Spherical Trigonometry.
- **E** Series expansions.
- **F** Calculus of variations.
- **G** Complex variable theory.

Chapter 2

Normal Mercator on the sphere: NMS

Geographical coordinates and Cartesian coordinates. Infinitesimal elements and the metric. General normal cylindrical projection. Angle transformations and scale factors. Three examples of normal cylindrical projections. Derivation of the Mercator projection. The loxodrome or rhumb line. Modified normal cylindrical projections.

2.1 Coordinates and distance on the sphere

Coordinates

The position of a point P on the sphere is denoted by an ordered pair (ϕ, λ) of latitude, longitude values; the meridians (λ constant), the equator ($\phi = 0$) and the small circles (ϕ constant, non-zero) constitute the **graticule** on the sphere. The figure shows a second point Q with coordinates $(\phi + \delta\phi, \lambda + \delta\lambda)$, the meridians through P and Q, arcs of parallels PM, KQ and the geodesic (great circle) through the points P and Q. Such geographical coordinates will be given in degrees but in equations ALL angles will be in radians. The unit **mil**, such that 6400mil=2π radians=360° is sometimes used for small angles, in particular the grid convergence defined in Chapter 3.

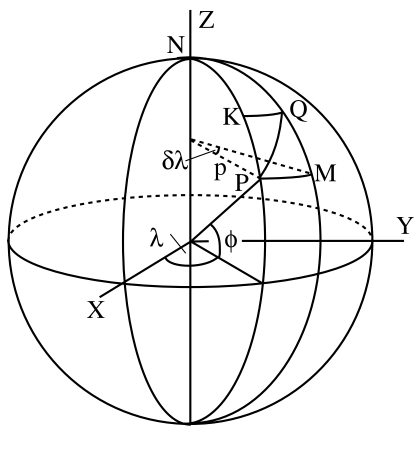


Figure 2.1

The following should be noted:

$$1 \text{ rad} \;=\; 57^\circ.29578 \;=\; 57^\circ\,17'\,44''.8 \;=\; 3437'.75 \;=\; 206264''.8 \;=\; 1018.6\text{mil}$$

$$1^\circ = 0.0174533 \text{ rad}, \quad 1' = 0.000291 \text{ rad} = 3.37 \text{ mil}, \quad 1'' = 0.00000485 \text{ rad}. \qquad (2.1)$$

$$1\text{mil} = 0.9812\text{mrad} = 0^\circ.0049 = 0'.297 = 20''.2.$$

If the distance of P from the axis (*i.e.* the radius of a parallel circle) is denoted by $p(\phi)$, so that $p(\phi) = a\cos\phi$ then the Cartesian coordinates are

$$\begin{aligned} X &= p(\phi)\cos\lambda = a\cos\phi\,\cos\lambda\,, \\ Y &= p(\phi)\sin\lambda\, = a\cos\phi\,\sin\lambda\,, \\ Z &= a\sin\phi\,, \end{aligned} \tag{2.2}$$

with inverse relations

$$\phi = \arctan\left(\frac{Z}{p}\right) = \arctan\left(\frac{Z}{\sqrt{X^2+Y^2}}\right), \qquad \lambda = \arctan\left(\frac{Y}{X}\right). \tag{2.3}$$

For a point at a height h above the surface at P we simply replace a by $a+h$ in the direct transformations: the inverse relations for ϕ and λ are unchanged but they are supplemented with the equation

$$h = \sqrt{X^2+Y^2+Z^2} - a. \tag{2.4}$$

In this chapter we consider a sphere with radius equal to the semi-major axis of the Airy ellipsoid:

$$a = 6377563.396\text{m} \approx 6378\text{km} \approx 3963 \text{ miles}\,.$$

This approximation value will suffice for the moment but more precise results will be needed when we come to consider the large scale maps of TME. For the above radius the circumference of the equator (or any great circle) is approximately 40071km (24900 miles) and the distance between pole and equator is one quarter of that value, 10018km (6225 miles). The closeness of the latter value to 10^7m reflects the original French definition of the metre as 10^{-7} times the pole–equator distance.

Distances on the sphere

In Figure 2.1 the distance PQ in *three* dimensions is unique but the distance *on the surface* of the sphere depends on the path taken between the points. For example, if the points are at the same latitude we can calculate the distance between them by measuring (a) along the parallel circle, (b) along the rhumb line (or loxodrome) which, by definition, intersects meridians at constant azimuth, or (c) along a geodesic which, by definition, gives the shortest distance on the surface.

The only trivially calculated distances are those measured along meridians or parallels: on the meridian in Figure 2.1 we have $PK = a\,\delta\phi$ (where $\delta\phi$ is in radians) and on the parallel $PM = p(\phi)\delta\lambda = a\cos\phi\,\delta\lambda$. For widely separated points these become $PK = a(\phi_2 - \phi_1)$ and $PM = a\cos\phi\,(\lambda_2 - \lambda_1)$. It is useful to have some feel for the distances on meridians and parallels on the sphere:

$$
\begin{array}{lll}
1^\circ & \text{latitude difference on a meridian} & = 111.3\text{km} = 69.16\text{miles}, \\
1' & & = 1.855\text{km} = 1.153\text{miles} \approx 1\text{nml}, \\
1'' & & = 31\text{m} \qquad = 33.8\text{yds}, \\
1^\circ & \text{longitude difference on parallel } 45^\circ\text{N} & = 78.7\text{km} \quad = 48.9\text{miles}, \\
1' & & = 1.312\text{km} = 0.815\text{miles}, \\
1'' & & = 22\text{m} \qquad = 23.9\text{yds}.
\end{array}
\tag{2.5}
$$

One minute of arc on the meridian (of a spherical Earth) was the original definition of the **nautical mile** (nml). On the ellipsoid this definition of the nautical mile would depend on latitude and the choice of ellipsoid so, to avoid discrepancies, the nautical mile is now defined by international treaty as exactly 1852m, corresponding to 1.150779 miles. The original definition remains a good rule of thumb for approximate calculations.

For two points in general position finding the distance is a non-trivial problem. There are two important cases to consider: (a) given the geographic coordinates of two points find the length of the geodesic between them and also the azimuths at the end points of the geodesic joining them; (b) given a starting point and an initial azimuth find the coordinates at a specified distance along the geodesic. These are the two principal geodetic problems and their solution, for both sphere and ellipsoid, is presented in Chapter 11. In the present chapter we consider only the problem of finding the distance between points on the surface which are infinitesimally close. This will suffice for the calculation of scale factors.

Infinitesimal elements

In practical terms an element of area on the sphere can be said to be infinitesimal if, *for a given measurement accuracy*, we cannot distinguish deviations from the plane. To be explicit, consider the spherical element $PMQK$ shown in Figure 2.1, and in enlarged form in Figure 2.2a, where the solid lines PK, MQ, PQ are arcs of great circles, the solid lines

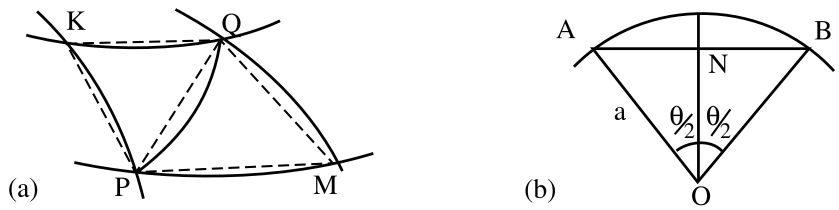


Figure 2.2

PM and KQ are arcs of parallel circles and the dashed lines are straight lines in three dimensions. From Figure 2.2b, for $\theta(\text{rad}) \ll 1$ the arc–chord difference is

$$
\text{arc}\,(AB) - AB = a\theta - 2a\sin\frac{\theta}{2} = a\theta - 2a\left(\frac{\theta}{2} - \frac{1}{3!}\frac{\theta^3}{8} + \cdots\right) = \frac{a\theta^3}{24} + O(a\theta^5). \tag{2.6}
$$

Suppose the accuracy of measurement is 1m. Setting $\theta = \delta\phi$ we see that the difference between the arc and chord PK will be less than 1m, and hence undetectable, if we take

$\delta\phi < (24/a)^{1/3} \approx 0.0155$rad, corresponding to 53′ or a meridian arc length of 99km. Similarly, setting $\theta = \delta\lambda$ and replacing $a \to a\cos\phi$, the difference between the arc and chord PM at a latitude of 45° (where $a\cos\phi = a/\sqrt{2}$) is less than 1m if $\delta\lambda$ is less than 59′, corresponding to an arc length of 78km on the parallel. If we take our limiting accuracy to be 1mm the above values become 9.9km and 7.8km. A surface element of this order or smaller can therefore be well approximated by a planar element.

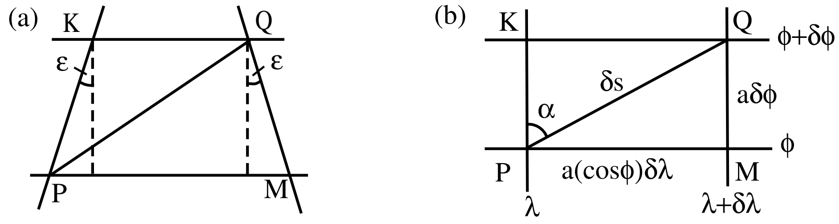


Figure 2.3

We shall now prove that the small surface element $PKQM$ may be well approximated by a *rectangular* element. Figure 2.3a shows the planar trapezium which approximates the surface element. Since $PM = a\cos\phi\,\delta\lambda$ we have

$$KQ - PM = \delta\phi\,\frac{d}{d\phi}\,(a\cos\phi\,\delta\lambda) = -a\sin\phi\,\delta\lambda\,\delta\phi. \tag{2.7}$$

Now the distance $PK \approx a\delta\phi$ so the small angle ϵ is given by

$$\epsilon \approx \sin\epsilon = \frac{PM - KQ}{2}\,\frac{1}{PK} = \frac{1}{2}\sin\phi\,\delta\lambda. \tag{2.8}$$

Clearly ϵ becomes arbitrarily small as Q approaches P and the infinitesimal element is arbitrarily close to the rectangle with sides $a\delta\phi$ and $a\cos\phi\,\delta\lambda$ shown in Figure 2.3b. The planar geometry of the right angled triangle PQM gives two important results for the **azimuth** α and distance:

$$\tan\alpha = \lim_{Q\to P}\frac{a\cos\phi\,\delta\lambda}{a\delta\phi} = \cos\phi\,\frac{d\lambda}{d\phi}\,, \tag{2.9}$$

$$\delta s^2 = PQ^2 = a^2\,\delta\phi^2 + a^2\cos^2\!\phi\,\delta\lambda^2. \tag{2.10}$$

The latter follows more directly from equations (2.2):

$$\begin{aligned}
dX &= -(a\sin\phi\,\cos\lambda)d\phi - (a\cos\phi\,\sin\lambda)d\lambda\\
dY &= -(a\sin\phi\,\sin\lambda)d\phi + (a\cos\phi\,\cos\lambda)d\lambda\\
dZ &= \quad(a\cos\phi)d\phi,
\end{aligned} \tag{2.11}$$

$$ds^2 = \quad dX^2 + dY^2 + dZ^2,$$

$$ds^2 = \quad a^2\,d\phi^2 + a^2\cos^2\!\phi\,d\lambda^2. \tag{2.12}$$

2.2 Normal cylindrical projections

The normal cylindrical projections of a sphere of radius a are defined on a cylinder of radius a which is tangential to the sphere on the equator as shown in Figure (2.4). The axis of the cylinder coincides with the polar diameter NS and the planes through this axis intersect the sphere in its meridians and intersect the cylinder in its generators. The projection

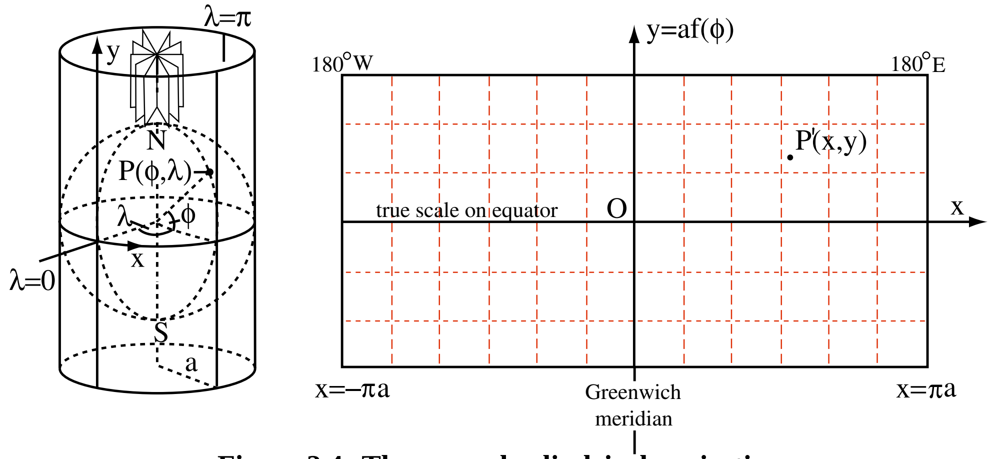


Figure 2.4: The normal cylindrical projection

takes the points of each meridian to points on the corresponding generator of the cylinder according to some formula which is NOT usually a geometric construction—in particular the Mercator projection is not generated by a literal projection from the centre (as stated in some elementary texts). The cylinder is then cut along a generator which has been taken as $\lambda = 180°$ in Figure 2.4 but could have been chosen as any longitude. Finally the cylinder is unrolled to form a flat map, the super-map which we discussed in Chapter 1. Note that the last step of unrolling introduces no further distortions. Axes on the map are chosen with the x-axis along the equator and the y-axis coincident with one particular generator, taken as the Greenwich meridian ($\lambda = 0$) in Figure (2.4). Clearly the meridians on the sphere map into lines of constant x on the projection so the x-equation of the projection is simply $x = a\lambda$ (radians) with this choice of coordinate system. For the y-equation of the projection we admit *any* (sensible) function of ϕ, irrespective of whether or not there is a geometrical interpretation. Thus normal cylindrical projections are defined by[1]

$$x(\lambda, \phi) = a\lambda, \qquad (2.13)$$

$$y(\lambda, \phi) = a\, f(\phi), \qquad (2.14)$$

where λ and ϕ are in radians. With transformations of this form we see that the parallels on the sphere (ϕ constant) project into lines of constant y so that the orthogonal intersections of meridians and parallels of the graticule on the sphere are transformed into orthogonal intersections on the map; we shall see that this is *not* necessarily true for intersections at an arbitrary angle. The spacing of the meridians on the projection is uniform but the spacing of the parallels depends on the choice of the function $f(\phi)$.

[1] In referring to geographic positions it is conventional to use latitude–longitude ordering as in $P(\phi, \lambda)$ but for mathematical functions of these coordinates it is more natural to use the reverse order as in $x(\lambda, \phi)$.

Note that *all* normal cylindrical projections have singular points since the *points* N, S at the poles transform into *lines* given by $y = a\,f(\pm\pi/2)$. On the sphere meridians intersect at the poles but on cylindrical projections meridians do not intersect. All other points of the sphere are non-singular points. Of course there is nothing special about the poles; if we use oblique or transverse projections the geographic poles are regular points and other points become singular—the singularities at the poles are artifacts of the coordinate transformations. For example we shall find that the transverse Mercator projection has singular points on the equator.

The equations (2.13, 2.14) define a projection to a super map of constant width equal to the length of the equator, $2\pi a$ or 40071km. Since the true length of a parallel is $2\pi a\cos\phi$, the scale (map length divided by true length) along a parallel is equal to $\sec\phi$, increasing from 1 on the equator to infinity at the poles. Note that this statement about scale on a parallel applies to *any* normal cylindrical projection but the scale on the meridians, and at other azimuths, will depend on $f(\phi)$. The actual printed projection in Figure 2.4 is about 8cm wide on paper so the RF (representative factor) is 8cm/40071km which is approximately 1 to 500 million or 1:500M.

Angle transformations on normal cylindrical projections

In Figure 2.5 we compare the rectangular infinitesimal element $PMQK$ on the sphere with the corresponding element $P'M'Q'K'$ on the projection. The latter is also a rectangle but without any approximation since the meridians map into lines of constant x and the parallels map into lines of constant y. The angle $K'P'Q'$ is called the **grid bearing** β corresponding

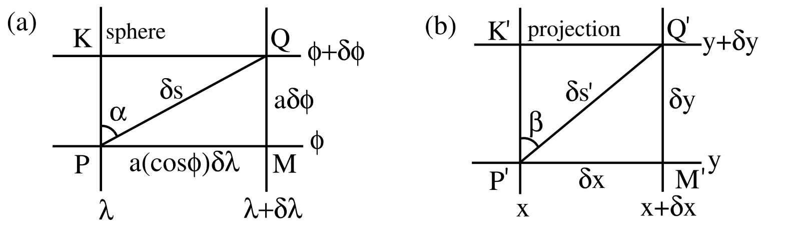


Figure 2.5

to the azimuth α on the globe. The geometry of these rectangular elements gives

$$\text{(a)}\quad \tan\alpha = \frac{a\cos\phi\,\delta\lambda}{a\,\delta\phi} \qquad \text{and} \qquad \text{(b)}\quad \tan\beta = \frac{\delta x}{\delta y} = \frac{\delta\lambda}{f'(\phi)\delta\phi}. \tag{2.15}$$

so that

$$\tan\beta = \frac{\sec\phi}{f'(\phi)}\tan\alpha. \tag{2.16}$$

Note that $\alpha{=}\beta$ on the meridians (both zero) or on the parallels (both $\pi/2$) but in general $\alpha \neq \beta$ unless $f'(\phi) = \sec\phi$. Therefore this is the condition for a normal cylindrical projection to be conformal. It also provides the means of calculating $f(\phi)$ for Mercator projection.

Definition of the point scale factor

If we set $P'Q' = \delta s'$ then for the element on the projection we have

$$\delta s'^2 = \delta x^2 + \delta y^2. \tag{2.17}$$

We now define μ, the point scale *at* P' by

$$\mu = \lim_{Q\to P} \frac{\text{distance } P'Q' \text{ on projection}}{\text{distance } PQ \text{ on sphere}} = \lim_{Q\to P} \frac{\delta s'}{\delta s}, \tag{2.18}$$

or, in terms of the distances squared, (using (2.10))

$$\mu^2 = \lim_{Q\to P} \frac{\delta s'^2}{\delta s^2} = \lim_{Q\to P} \frac{\delta x^2 + \delta y^2}{a^2\,\delta\phi^2 + a^2\cos^2\!\phi\,\delta\lambda^2}. \tag{2.19}$$

Point scale factors on meridians (h) and parallels (k)

When PQ lies along the meridian $\delta\lambda$ and δx are zero and $y = a\,f(\phi)$. The scale factor in this case is conventionally denoted by h. Therefore (2.19) gives

$$\textbf{meridian:} \qquad h = f'(\phi). \tag{2.20}$$

Similarly, when PQ lies along a parallel of latitude, $\delta\phi$ and δy are zero and $x = a\lambda$. The scale factor in this case is conventionally denoted by k. Therefore

$$\textbf{parallel:} \qquad k = \sec\phi. \tag{2.21}$$

Point scale factor in a general direction

Equations (2.15) give $\delta\phi = \cot\alpha\ \cos\phi\,\delta\lambda$ and $\delta y = \cot\beta\,\delta x$. Therefore equation (2.19) gives the scale factor at azimuth α as

$$\mu_\alpha^2 = \lim_{Q\to P} \frac{\delta x^2(1+\cot^2\!\beta)}{a^2\cos^2\!\phi\,\delta\lambda^2(1+\cot^2\!\alpha)}, \tag{2.22}$$

which reduces to

$$\mu_\alpha(\phi) = \sec\phi \left[\frac{\sin\alpha}{\sin\beta}\right], \tag{2.23}$$

where we assume that β has been found in terms of α and ϕ from equation (2.16).

Area scale factor

The area scale is obtained by simply comparing the areas of the two rectangles $PMQK$ and $P'M'Q'K'$. Denoting this scale factor by μ_A and using equations (2.20) and (2.21).

$$\mu_A = \lim_{Q\to P} \frac{\delta x\,\delta y}{(a\cos\phi\,\delta\lambda)(a\,\delta\phi)} = \sec\phi\, f'(\phi) = hk. \tag{2.24}$$

NB. All of these scale factors apply only to the normal cylindrical projections. They are independent of λ, a reflection of the rotational symmetry.

2.3 Three examples of normal cylindrical projections

We shall compare three simple projections of the sphere:

1. The equidistant (Ptolemy, Bonne or Plate Carrée) projection: $f(\phi) = \phi$.
2. Lambert's equal area (sinusoidal) projection: $f(\phi) = \sin\phi$.
3. Mercator's projection: $f(\phi) = \ln[\tan(\phi/2 + \pi/4)]$. (Derived in Section 2.4).

	equidistant	equal-area	Mercator
x transformation	$x = a\lambda$	$x = a\lambda$	$x = a\lambda$
x-range	$(-\pi a,\ \pi a)$	$(-\pi a,\ \pi a)$	$(-\pi a,\ \pi a)$
y transformation	$a\phi$	$a\sin\phi$	$a\ln[\tan(\phi/2{+}\pi/4)]$
y-range	$(-\pi a/2,\ \pi a/2)$	$(-a,\ a)$	$(-\infty, \infty)$
$f(\phi)$	ϕ	$\sin\phi$	$\ln[\tan(\phi/2 + \pi/4)]$
$f'(\phi)$	1	$\cos\phi$	$\sec\phi$
meridian scale (h)	1	$\cos\phi$	$\sec\phi$
parallel scale (k)	$\sec\phi$	$\sec\phi$	$\sec\phi$
scale on equator	1	1	1
area scale (hk)	$\sec\phi$	1	$\sec^2\phi$
angles (eq. (2.16)	$\tan\beta{=}\sec\phi\tan\alpha$	$\tan\beta{=}\sec^2\phi\tan\alpha$	$\tan\beta = \tan\alpha$
aspect ratio	2	π	0
	Figure 2.6	Figure 2.7	Figure 2.8

Table 2.1

The three projections are shown in Figures (2.6–2.8) with every little 'smudge' on the maps indicating a small island or small lake! The maps all have the same x-range of $(-\pi a,\ \pi a)$ (in metres) but varying y-ranges. They all have unit scale on the equator and the RF on these pages is 12cm/40000km or about 1/300M.

On the equator $\phi = 0$ so that (a) all the point scales (h, k, and μ_α) are unity so that the scale is isotropic and small elements retain their shape, (b) the area scale is unity and (c) equation (2.16) shows that $\alpha = \beta$ so that all lines cross the equator on globe and projection at the same angle. Thus on the equator the projections are perfectly behaved and suitable for the *accurate* mapping of countries lying close to the equator.

Away from the equator the divergent parallel scale ($\sec\phi \to \infty$ as $\phi \to \pi/2$) produces gross east-west stretching in high latitudes but the differing meridian stretching leads to different shapes and areas. The acid test is to compare the projection with the outlines shown on any toy or classroom globe. In particular look at the shape of Alaska and the area of Greenland. The latter should be 1/8 that of South America and 1/13 that of Africa.

Finally, each of the projections is annotated on the right with a chequered column corresponding to $5^\circ{\times}5^\circ$ regions on the sphere. The width of these rectangles is the same on all but their height depends on $f(\phi)$.

The equidistant projection: $f(\phi) = \phi$

This projection, whose only merit is the simplicity of construction, has been in use as far back as the time of Ptolemy (83?–161AD): the meridians and parallels of latitude are equidistant parallel lines intersecting at right angles, forming a square grid on the map—the 'Plate Carrée'. Except on the meridians ($\alpha = \beta = 0$) and parallels ($\alpha = \beta = \pi/2$) the map is not conformal ($\alpha \neq \beta$), the scale is not isotropic and the area scale factor is not unity. High latitudes are distorted as expected. The aspect ratio (width:height) is 2.

The very terminology 'equidistant' is misleading, for the scale is true only on the meridians and on the equator. On these lines finite ruler distances measured on the page give accurate true distances simply by dividing by the RF, however separated be the two end points. However the scale is not unity anywhere else. On the parallels it is given

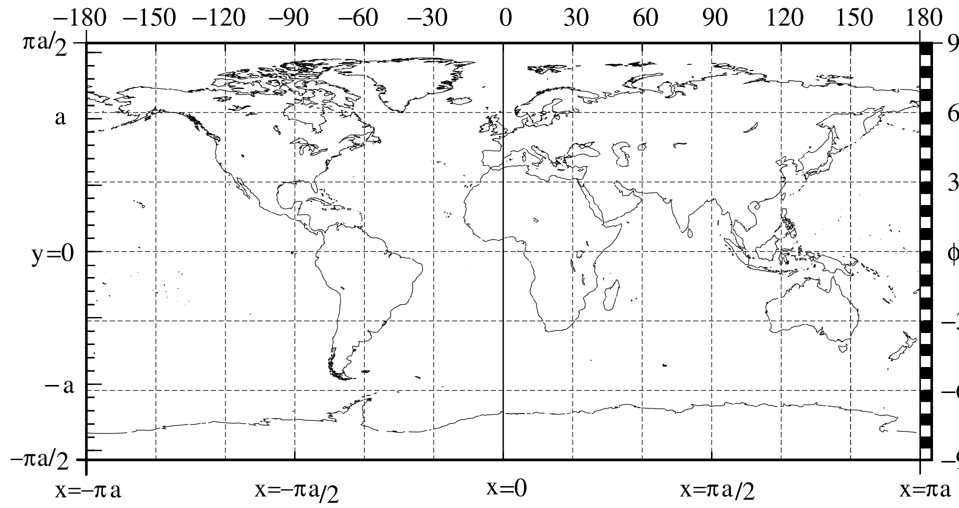


Figure 2.6: The equidistant projection

by $k = \sec\phi$ and the ruler distance must be divided by this factor and the RF to obtain true distances. On a straight line on the projection with a constant bearing β (not equal to zero or $\pi/2$), equation (2.23) shows that the scale is $\mu_\alpha = \sec\phi \sin\alpha(\phi) \operatorname{cosec}\beta$ with $\alpha(\phi) = \arctan(\tan\beta\cos\phi)$ obtained from (2.16). The ϕ dependence of this scale factor means that there is no simple way of measuring along such a line with a ruler in order to determine any kind of distance. A numerical integration of the scale factor along the line on the projection is possible but it would give a distance on that line on the globe which corresponds to β constant; in general such a line is not a parallel, meridian, rhumb line or great circle geodesic, so the utility of such an integration must be queried. The only simple way to obtain the geodesic distance between general points on the projection is to transfer their coordinates to the sphere and then use the standard geodesic formulae presented in Chapter 11.

Lambert's equal area projection: $f(\phi) = \sin\phi$

This projection, which was first presented by Lambert in 1772, has many properties in common with the equidistant projection: it is well behaved at the equator, it is distorted at high latitudes and the scale is unity on the equator only. Its great plus is that it preserves the magnitude of small and large areas since the scale factor on the meridians ($h = \cos\phi$) compensates exactly for the stretching on the parallels ($k = \sec\phi$). The scale factor on the meridians is a compression as compared to the equidistant projection so that the parallels bunch up at high latitudes—witness the box scale on the right hand side. The projection satisfies the Greenland test but still fails the Alaska test.

Once again reading distances from the map is trivial on the equator and parallels (where we must divide by a factor of $\sec\phi$). On meridians it is also simple for, if we know two y-values, y_1 and y_2, not just their separation, we can use $\phi = \arcsin(y/a)$ to find the corresponding latitudes and hence the arc length $a(\phi_2 - \phi_1)$. On general lines of the projection we have $\tan\beta = \sec^2\phi\tan\alpha$ for which we have the same problems as in the previous case.

The projection is neither conformal nor isotropic but on the other hand it is one of the

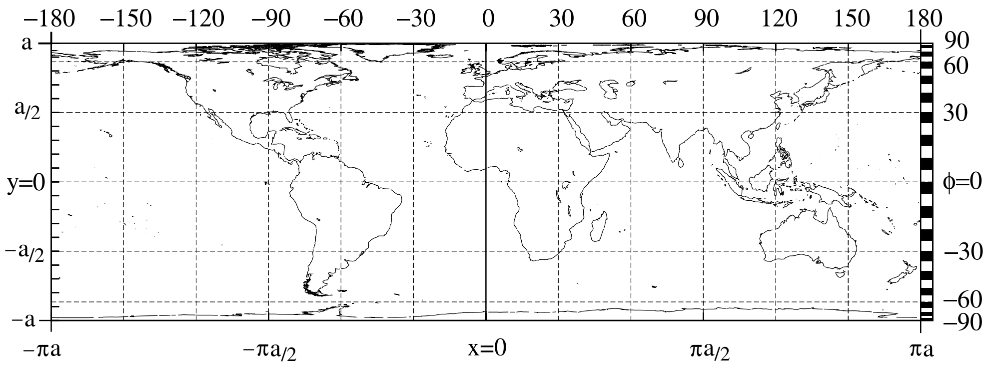


Figure 2.7: Lambert's equal area projection

few which admits of a geometric interpretation because $y = a\sin\phi$ is simply the distance of a point at latitude ϕ above the equatorial plane. Thus points are projected from the sphere to cylinder by lines parallel to the equatorial plane drawn from the axis of the sphere—not projected from the origin. Thus any narrow (in longitude) strip of the map is basically the view of an actual globe from an appropriate distant side position.

The projection is not often used in the form given here: it is usually compressed laterally and extended vertically whilst preserving the equal area property; *i.e.* the transformations become $x = Ka\lambda$ and $y = K^{-1}a\sin\phi$ so that we still have $hk = 1$ but the scale on a parallel is now $K\sec\phi$ and the meridian scale is $K^{-1}\cos\phi$. This means that the true scale is attained only on some parallel other than the equator. For example the equal area projection of Gall (1855), which was republished by Peters in 1973, sets $K = 1/\sqrt{2}$ so that the meridian and parallel scales are unity at latitudes of $\pm 45°$. The distortion of shape in the Peters projection is well known.

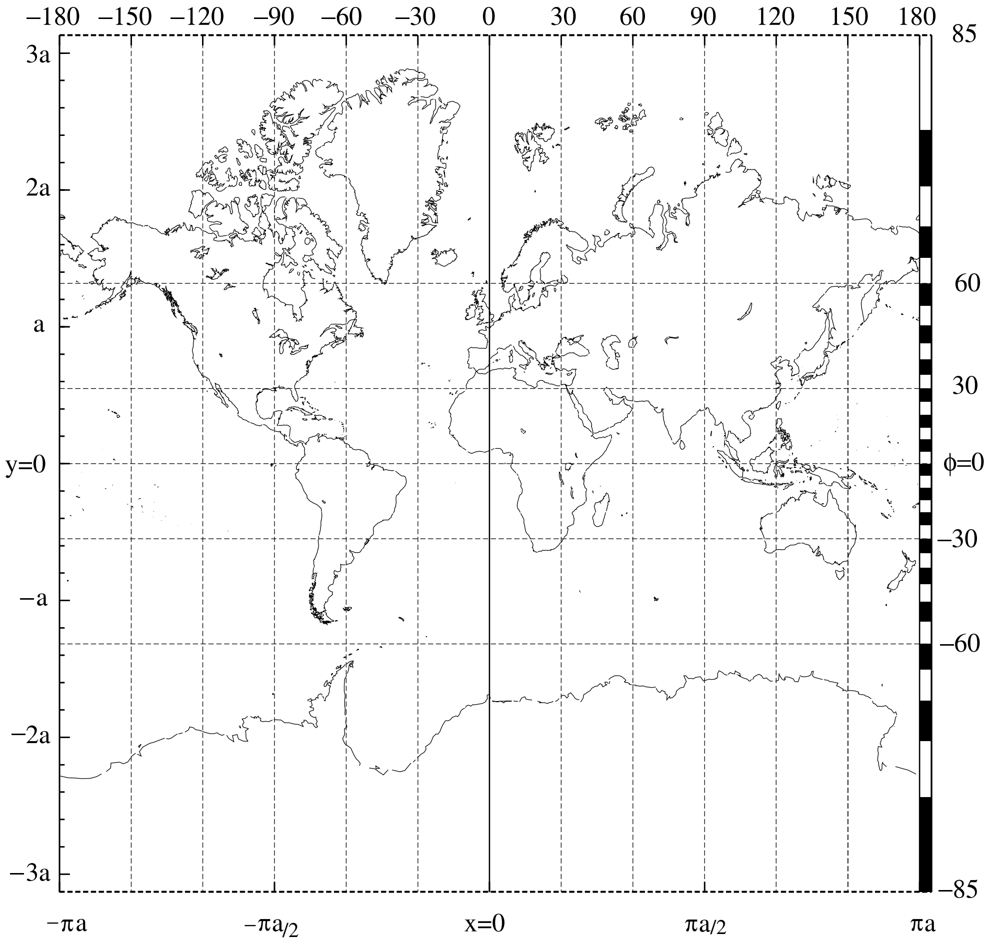


Figure 2.8: The Mercator projection

Mercator's projection: $f(\phi) = \ln[\tan(\phi/2 + \pi/4)]$

Figure 2.8 shows Mercator's projection of 1569. Both scale factors, on meridian and parallel, increase as $\sec\phi$ so that in addition to the stretching along the parallels the meridians are stretched to infinity and the aspect ratio becomes zero. In the above figure we have chosen to truncate the projection at the value $y = \pm 3.13a$ corresponding to $\phi = \pm 85°$ so that the aspect ratio is very close to 1. Truncation at these high latitudes emphasizes the great distortion near the poles—witness the diverging area of Antarctica and a Greenland as big as Africa.

The fundamental property of the Mercator projection is that it is conformal, *i.e.* it is an angle preserving projection; this follows from equation (2.16) since $f'(\phi) = \sec\phi$ implies that $\alpha = \beta$, *i.e.* a curving rhumb line of constant α on the globe projects into a straight line of constant bearing β on the map and vice-versa. For more on rhumb lines see Section 2.5.

The isotropy of scale implies that the shape of *small* elements is well rendered on the projection; the property of orthomorphism. Consider a small square of side L at latitude ϕ where the isotropic scale factor is $\sec\phi$. The projection will map this square into an exact square of side $L\sec\phi$, preserving its shape, only if the variation of $\sec\phi$ over the square can be neglected *at the precision of measurement in use*. Larger shapes will be distorted if $\sec\phi$ varies appreciably over their size making the Mercator projection unsuitable for detailed mapping of large countries.

The only true scale on the Mercator projection is attained at the equator and moreover the area scale approaches unity too. Thus, since we still have conformality, it may well be convenient to use Mercator for accurate large scale mapping near the equator. This is discussed more quantitatively in the next section. Meanwhile we must accept the limitations of using Mercator for small scale maps of the globe—charts excepted!

Using ruler distance to deduce true distance is non-trivial away from the equator. On parallels we can again divide the ruler distance by $k = \sec\phi$ to find the true distance on the parallel (which is not a geodesic difference). On a meridian where the scale is ($h = \sec\phi$) we have to invert $y = f(\phi)$ to find latitudes and hence an arc length: this is discussed in the next section. On lines of constant bearing β the scale factor is also $\sec\phi$. This can be integrated but it gives a distance along a rhumb line—see Section 2.5

Conformal versus orthomorphic

The definition of 'conformal' just given is that the projection preserves the angle between any two lines through any (non-singular) point of the map. This is an exact statement referring to the behaviour of the transformations *at a point*. Many authors use the word 'orthomorphic' (right-shape) as a synonym for conformal but it is important to realise that orthomorphism can only be satisfied approximately in a conformal projection because it is a non-local property. No projection can be completely orthomorphic since even if *small* shapes are preserved then one can find large shapes which are distorted. For this reason we prefer to use the word conformal (exact) rather than orthomorphic (approximate).

2.4 The normal Mercator projection

Derivation of the Mercator projection

The normal Mercator projection is defined to be that normal cylindrical projection which is conformal, so that the azimuth and grid bearing are equal, $\alpha = \beta$. Equation (2.16) shows that this is possible only if

$$f'(\phi) = \frac{df}{d\phi} = \sec\phi, \tag{2.25}$$

and therefore

$$f(\phi) = \int_0^{\phi} \sec\phi\, d\phi, \tag{2.26}$$

with a lower limit such that $y(0) = af(0) = 0$.

Now

$$\begin{aligned}\cos\phi &= \sin(\phi+\pi/2)\\ &= 2\sin(\phi/2{+}\pi/4)\cos(\phi/2{+}\pi/4)\\ &= 2\tan(\phi/2{+}\pi/4)\cos^2(\phi/2{+}\pi/4)\end{aligned}$$

so that

$$y = a\,f(\phi) = \frac{a}{2}\int_0^\phi \frac{\sec^2(\phi/2+\pi/4)}{\tan(\phi/2+\pi/4)}\,d\phi = a\ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right]. \qquad (2.27)$$

There is no need for modulus signs inside the logarithm. For $-\pi/2 \le \phi \le \pi/2$ the argument of the tangent is in the interval $[0,\pi/2]$, therefore the argument of the logarithm is in the range $[0,\infty)$ and the logarithm itself varies from $-\infty$ to ∞.

Mercator parameter and isometric latitude

The function of ϕ which occurs in the expression for the y-coordinate in the Mercator projection is of importance in much that follows. We shall call it the **Mercator parameter** and denote it by $\psi(\phi)$ and write the equations of the projection as

$$x(\lambda,\phi) = a\lambda, \qquad y(\lambda,\phi) = a\psi(\phi),$$

with

$$\psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right] \qquad \frac{d\psi}{d\phi} = \sec\phi. \qquad (2.28)$$

The term Mercator parameter is not standard usage. It is usually called the **isometric latitude**. For example see Snyder(1987). But beware; other authors use the term for a different, but related, function. Note, too, that the symbol ψ is not universal: Lee(1946a) and Redfearn(1948) use X, Maling(1992) uses q and so on.

The origin of the term isometric latitude relates to a re-parameterisation of the sphere in such a way that the isometric latitude (replacing the geodetic latitude) and λ are involved with equal weight in the metric formulae such as (2.12). If we write $\widetilde{\psi}$ for such an isometric latitude then the metric $ds^2 = a^2 d\phi^2 + a^2\cos^2\!\phi\, d\lambda^2$ becomes $ds^2 = a^2\cos^2\!\phi(d\widetilde{\psi}^2 + d\lambda^2)$ if we choose $\cos\phi\, d\widetilde{\psi} = d\phi$. The coefficients in the metric are now both equal to $a^2\cos^2\!\phi$. Since our ψ is defined above (2.28) by $\cos\phi\, d\psi = d\phi$ we see that the functions $\psi(\phi)$ and $\widetilde{\psi}(\phi)$ must be identical. However the two functions are logically distinct and in an elementary treatment we prefer to use different names but allow the same symbol ψ.

Having urged care with notation we must flag a small problem in notation. When we define the Mercator projection on the ellipsoid (NME in Chapter 6) we must define the Mercator parameter ψ in a slightly different way, but such that it reduces to (2.28) as the eccentricity e tends to zero. It would therefore have been natural to define the Mercator parameter on the sphere as ψ_0. We have not *not* done this, assuming that the correct interpretation will be obvious from the context.

Alternative forms of the Mercator parameter

The Mercator parameter can be cast into many forms which may be useful at times; here we present five such. Consider the argument of the logarithm in (2.28):

$$\tan(\phi/2+\pi/4) = \frac{1+\tan(\phi/2)}{1-\tan(\phi/2)} = \frac{\cos(\phi/2)+\sin(\phi/2)}{\cos(\phi/2)-\sin(\phi/2)}$$

$$= \frac{(\cos(\phi/2)+\sin(\phi/2))^2}{\cos^2(\phi/2)-\sin^2(\phi/2)} = \frac{1+\sin\phi}{\cos\phi} = \sec\phi+\tan\phi. \qquad (2.29)$$

Hence

$$\psi(\phi) = \ln[\sec\phi+\tan\phi]. \qquad (2.30)$$

Now rearrange the penultimate term in (2.29):

$$\frac{1+\sin\phi}{\cos\phi} = \left\{\frac{(1+\sin\phi)^2}{1-\sin^2\phi}\right\}^{1/2} = \left\{\frac{1+\sin\phi}{1-\sin\phi}\right\}^{1/2}.$$

Therefore we have

$$\psi(\phi) = \frac{1}{2}\ln\left[\frac{1+\sin\phi}{1-\sin\phi}\right]. \qquad (2.31)$$

Now exponentiate each side of (2.30):

$$\mathrm{e}^{\psi} = \sec\phi+\tan\phi.$$

After a little manipulation we find that

$$2\sinh\psi = \mathrm{e}^{\psi}-\mathrm{e}^{-\psi} = 2\tan\phi,$$

and therefore

$$\text{(a)}\quad \sinh\psi = \tan\phi, \qquad \text{(b)}\quad \cosh\psi = \sec\phi, \qquad \text{(c)}\quad \tanh\psi = \sin\phi, \qquad (2.32)$$

from which we obtain three further variants for $\psi(\phi)$:

$$\psi = \operatorname{arcsinh}(\tan\phi) = \operatorname{arccosh}(\sec\phi) = \operatorname{arctanh}(\sin\phi). \qquad (2.33)$$

The inverse transformation

The inverse transformation of (2.28) is clearly

$$\lambda = \frac{x}{a}, \qquad \psi = \frac{y}{a} \qquad (2.34)$$

where ϕ must be deduced from ψ by using the inverse of any of the four expressions given in (2.28) and (2.32). For example, from (2.28) and (2.32c) we get

$$\text{(a)}\quad \phi = 2\tan^{-1}\left(\mathrm{e}^{y/a}\right)-\frac{\pi}{2}, \qquad \text{(b)}\quad \phi = \sin^{-1}\left[\tanh(y/a)\right]. \qquad (2.35)$$

Comment: the coordinate origin of the transformation may be chosen as any point on the equator so that we write $x = a(\lambda-\lambda_0)$ with a trivial inverse. There is no such freedom of choice for ϕ since its definition is intimately tied to the choice of polar axis and a graticule which assigns $\phi = 0$ on the equator. The range of ϕ is confined to $[-\pi/2, \pi/2]$ and no translation of the ϕ origin is permitted.

2.5 The rhumb line or loxodrome

We have defined the **rhumb line** or **loxodrome** on the sphere as a curve which intersects meridians at a constant azimuth α and is therefore projected to a straight line on the Mercator projection with constant bearing β such that $\alpha = \beta$. Special cases are (a) the equator and any parallel on which $\alpha = \beta = \pi/2$, (b) any meridian, on which $\alpha = \beta = 0$. The utility of a direct link between rhumb lines on the sphere and straight lines on Mercator charts is obvious; it is discussed in many nautical publications and web-sites.

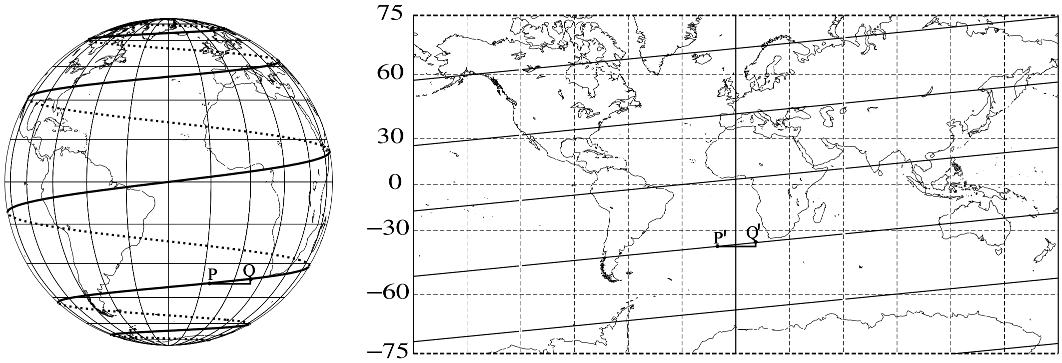


Figure 2.9: Rhumb line on the sphere and the Mercator projection

Figure 2.9 shows such a rhumb line crossing the equator at 30°W and maintaining a constant azimuth of $83°$; it spirals round the sphere from pole to pole whilst, on the projection, it becomes a repeated straight line of infinite total length because the Mercator projection extends to infinity along the y-axis. However, it is easy to show that corresponding rhumb

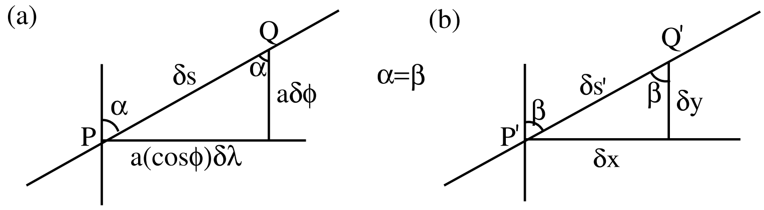


Figure 2.10:

line on the sphere has *finite* length. In Figure 2.10 we show an enlarged view of corresponding elements, PQ on the sphere and $P'Q'$ on the projection. The left hand triangle shows that $\cos\alpha = a\,d\phi/ds$. Since α is a constant this integrates directly to give

$$s_{12} = a\sec\alpha\,(\phi_2 - \phi_1), \qquad \phi_1 \neq \phi_2. \tag{2.36}$$

for a rhumb line from (ϕ_1, λ_1) to (ϕ_2, λ_2). Setting $\phi_1 = -\pi/2$ and $\phi_2 = \pi/2$ we see that the total length from one pole to another is $a\sec\alpha(\pi/2 - (-\pi/2)) = \pi a\sec\alpha$. This becomes πa on a meridian where it does not spiral. Therefore to calculate the distance along a rhumb line we need know only the constant azimuth and the change of latitude. Note that the above result fails on the equator or any parallel since $\phi_1 = \phi_2$. In this case we have

$$s_{12} = a\cos\phi\,(\lambda_2 - \lambda_1) \qquad \phi_1 = \phi_2 = \phi.$$

It is very straightforward to plot the rhumb line on chart and on the sphere. On the chart it is trivial for the equation of a line of gradient $\cot\alpha$ through the point (x_1, y_1) on the projection is simply

$$y - y_1 = \cot\alpha\,(x - x_1). \tag{2.37}$$

The corresponding equation on the sphere follows immediately by using the Mercator projection formulae giving

$$a(\psi - \psi_1) = a\cot\alpha\,(\lambda - \lambda_1), \tag{2.38}$$

which, on substitution for ψ from (2.28) , becomes

$$\lambda(\phi) = \lambda_1 + \tan\alpha\left(\ln\left[\tan\left(\frac{\phi}{2} + \frac{\pi}{4}\right)\right] - \ln\left[\tan\left(\frac{\phi_1}{2} + \frac{\pi}{4}\right)\right]\right). \tag{2.39}$$

To plot the rhumb line on the surface of the sphere we choose a parallel (ϕ) and then evaluate λ at the point where the rhumb line crosses the parallel by using the above result.

Mercator sailing

The above equations solve the basic problem of ‘Mercator sailing’. That is, given a starting point (ϕ_1, λ_1) and a destination (ϕ_2, λ_2) we have to find the azimuth of the rhumb line and the sailing distance. The azimuth follows from (2.38) as long as we have tables of the Mercator parameter or a calculator and the formula (2.28). Notice that in using (2.38) the longitude difference must be expressed in radians by a preliminary calculation. Once we have the azimuth then the sailing distance follows from (2.36) with the latitude in radians.

Of course, even before the advent of technology, there was no need for these calculation as long as one was sailing *short* legs, no more than a few degrees, and as long as one possessed a Mercator chart marked up with scales of latitude and longitude against which we can plot start and final positions. The azimuth is trivial because it equals the bearing marked on the chart as the line joining start and finish points.

The distance is a little more tricky because the chart scales vary non-linearly, but isotropically, as $\sec\phi$. To find the true distance ds, corresponding to ds' we must first compensate by dividing the ruler length of ds' by $\sec\phi$. This will also be the case for finite distances as long as the variation in $\sec\phi$ along the sailing course can be neglected. Now rather than working out $ds'/\sec\phi$ we can simply measure ds' on a ‘stretched rule’, *i.e.* one which has already been stretched by a factor of $\sec\phi$—and there is one on the map already, namely the latitude scale on the vertical edges of the chart. So all one need do is use dividers to transfer ds' to the latitude scale and if this is marked with minutes then we have our final results in traditional nautical miles (on a sphere) without further calculation. For longer voyages, more than a few hundred miles, reading the azimuth from the chart is still trivial but the distance must be calculated using (2.36) because the latitude scale on the sides is too non-linear to use as a measure over such distances.

Meridian parts

Equations (2.36) and (2.38) were the basic equations that navigators could use in the early seventeenth century once they were furnished with Wright's table of Meridian parts. The first edition was based on the division of the meridian into 540 'parts' of $10'$ interval but he soon improved this to 5400 parts at $1'$ interval. For each part he calculated how much it would be stretched by the Mercator scale factor and he then added these successively to obtain his cumulative secants. Nautical texts often write the total of (stretched) Meridian parts from the equator to a latitude ϕ as

$$\mathrm{MP}(\phi) = \sum_{0}^{\phi} \sec\phi_i \qquad \text{at intervals of } 1'. \tag{2.40}$$

Bearing in mind that there are 3437.75 minutes of arc in one radian (Equation 2.1) we see that the above sum is proportional to a good approximation to the integral of $\sec\phi$:

$$\mathrm{MP}(\phi) \approx 3437.75 \int_0^{\phi} \sec\phi \, d\phi = 3437.75\, \psi(\phi). \tag{2.41}$$

Thus the meridional parts are simply proportional to the ordinate of the Mercator projection. Now if we also express $(\lambda_2 - \lambda_1)$ in minutes of arc in equation (2.38) we obtain

$$\tan\alpha = \frac{(\lambda_2 - \lambda_1)'}{\mathrm{MP}(\phi_2) - \mathrm{MP}(\phi_1)} \tag{2.42}$$

where the MP values are obtained from Wright's tables. Thus, as long as we know latitude and longitude values of start and finish, we can work out the required sailing course even if we lack a chart. The sailing distance then follows immediately from equation (2.36) as

$$s_{12} = \sec\alpha\, (\phi_2 - \phi_1)', \tag{2.43}$$

where we have absorbed the factor of a so that if latitudes are expressed in minutes of arc the result is in (traditional) nautical miles for a sphere.

2.6 Scale, distance and accuracy in the Mercator projection

We have already observed that since $\alpha = \beta$ for the Mercator projection the scale factor from equation (2.23) is isotropic and equal to $\sec\phi$. It is customary to use k for the common value of an isotopic scale factors therefore, in terms of the geographical coordinates, we have[2]

$$k(\lambda, \phi) = \sec\phi, \tag{2.44}$$

and in terms of the projection coordinates we use (2.32b) to find

$$k(x, y) = \cosh\psi = \cosh(y/a). \tag{2.45}$$

[2] Note the order. For all functions of two variables defined on the sphere we prefer to write λ, the absciisa-like variable, first. This is unconventional but it makes subsequent chapters more logical.

Note that we should really distinguish these two scale factors, by using distinct functional names such as $k(\lambda, \phi)$ and $\tilde{k}(x, y)$, because replacing ϕ and λ in equation (2.44) by x and y would give $k(x, y) = \sec x$, which is not equation (2.45). This kind of distinction is rarely, if ever, made in the literature of cartography and we shall not do so here; context will make things clear.

Away from the equator the scale factor is $k = \sec\phi > 1$ and an infinitesimal ruler distance on the map will therefore exaggerate distances on the sphere by this amount. It is only on a parallel, where ϕ is constant, that we can measure large distances by dividing the ruler distance by the RF and then dividing again by $\sec\phi$. Unlike the other projections we have considerd, for Mercator we can also easily measure distances on a rhumb line by using (2.36) but geodesic distances must be measured by the methods of Chapter 11.

Our main concern in this article is to investigate the application of the transverse Mercator projections to large scale accurate mapping so it is interesting to see how the normal Mercator fares in this respect when we are close to the equator where $k \approx 1$. It is necessary to define the word 'accurate': we shall use it to mean that the scale variation is within 0.04% of some specified value, corresponding to 4 parts in 10000. We shall call this the **zone of accuracy**.

For normal Mercator the scale varies between $k = 1$ at the equator and 1.0004 as latitudes $\pm\phi_1$ given by $\sec\phi_1 = 1.0004$, or $\cos\phi_1 \approx 0.9996$, corresponding to $\phi_1 = 1.62°$. Thus the zone of accuracy for the Mercator projection is a strip of about 3.24° width centred on the equator—this corresponds to 360km or 200 miles. The projection is certainly suitable for accurate mapping that narrow band of latitudes; in the next section we shall see how the projection may be modified to give a wider zone of accuracy.

We could argue that that the absolute value of the scale is not relevant—only the variation of scale over the mapped region is of interest. Consider, for example, the band of latitudes starting at $10°$N, where $k = 1.015$: we find that the scale has increased by 0.04% when we reach $10.12°$N so that the width of the zone of accuracy starting at $10°$N is only $7'.7$. This is a tiny strip indeed. By $30°$N we have $k = 1.155$ and the zone of accuracy is only $2'.4$. Clearly such narrow zones are unsuitable and the Mercator projection must be limited to a narrow equatorial zone for accurate mapping.

2.7 The modified normal Mercator projection

We have just seen that Mercator is accurate only within a narrow band centred on the equator. We now show how the width of this zone may be enlarged by making the scale on the equator less than 1, but greater than 0.9996 so that we are still with the 0.04% tolerance. We simply modify the projection cylinder by reducing its radius (Figure 2.11). If the cylinder intersects the sphere in the parallels at latitudes $\pm\phi_1$ then the radius of the cylinder must be $a\cos\phi_1$. We then demand that the scale be true on the parallels $\pm\phi_1$ and we can achieve this, and retain conformality, by multiplying both equations of the transformation by the

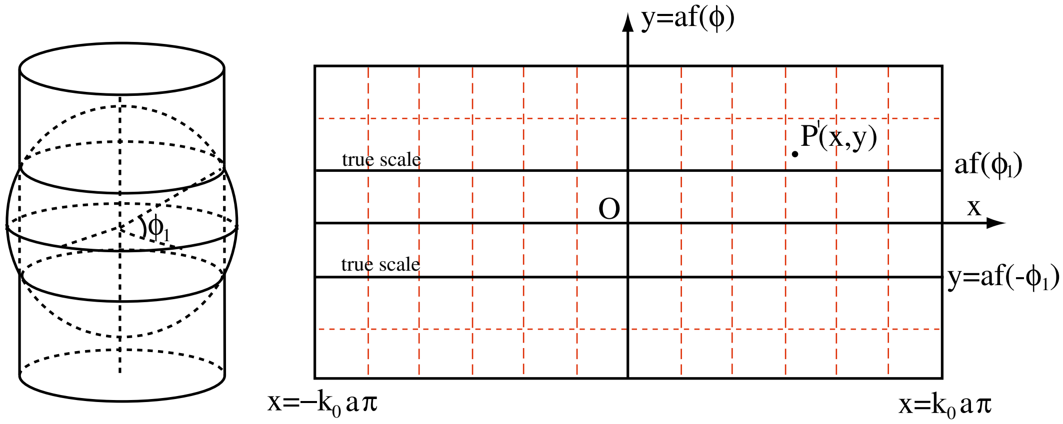


Figure 2.11: The modified normal cylindrical projection

factor $k_0 = \cos\phi_1$ so that they become:

$$x = ak_0\,\lambda, \qquad y = ak_0 \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right], \qquad k_0 = \cos\phi_1 \tag{2.46}$$

The definition of the scale factor in equation (2.19) shows that it remains isotropic with a value

$$k(\lambda,\phi) = k_0\ \sec\phi = \cos\phi_1\ \sec\phi. \tag{2.47}$$

The scale factor is now $k = \cos\phi_1 < 1$ on the equator and increases with latitude. If the lowest acceptable scale factor is $k = 0.9996$ then we must have $\cos\phi_1 = 0.9996$ corresponding to a value of $\phi_1 = 1.62°$. Likewise, if the largest acceptable scale factor is attained at $\phi = \phi_2$ then we must have

$$1.0004 = k(\lambda,\phi_2) = k_0 \sec\phi_2 = \cos\phi_1 \sec\phi_2 = 0.9996 \sec\phi_2. \tag{2.48}$$

This equation gives $\cos\phi_2 \approx 0.9992$ and $\phi_2 = \pm 2.29°$, so that the projection is now reasonably accurate in a strip of total width 4.58° centred on the equator. This corresponds to a north-south distance of about 512km or 284 miles.

Thus for the modified normal Mercator projection the scale on the equator is not unity; the parallels $\pm\phi_1$, on which the scale is true, are called the standard parallels of the modified projection. If we are willing to accept less accuracy then we can take the standard parallels at higher latitudes. For example if we take $\phi_1 = \pm 40°$ then the scale at the equator is $k = 0.76$ and the latitudes at which $k = 1.24$ are $\pm 52°$. Between these latitudes the projection is accurate to within 24%. Similar considerations apply to all normal cylindrical projections.

2.8 Summary of modified NMS

Direct transformation

$$x(\lambda, \phi) = k_0 a(\lambda - \lambda_0),$$

$$y(\lambda, \phi) = k_0 a\psi(\phi), \tag{2.49}$$

Mercator parameter

$$\psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right] = \frac{1}{2}\ln\left[\frac{1+\sin\phi}{1-\sin\phi}\right], \tag{2.50}$$

Inverse transformation

$$\lambda(x, y) = \lambda_0 + \frac{x}{k_0 a}, \qquad \psi(y) = \frac{y}{k_0 a} \tag{2.51}$$

$$\phi(x, y) = 2\tan^{-1}\left[\exp\left(\frac{y}{k_0 a}\right)\right] - \frac{\pi}{2}, \tag{2.52}$$

$$= \sin^{-1}\left[\tanh\left(\frac{y}{k_0 a}\right)\right], \tag{2.53}$$

Conformality

$$\alpha = \beta, \tag{2.54}$$

Scale factors

$$k(\lambda, \phi) = k_0 \sec\phi, \tag{2.55}$$

$$k(x, y) = k_0 \cosh\left(\frac{y}{k_0 a}\right). \tag{2.56}$$

Chapter 3

Transverse Mercator on the sphere: TMS

Abstract

TMS transformations from NMS by rotation of the graticule. Three global TMS maps. Inverse transformations. Meridian distance, footpoint and footpoint latitude. Scale factors. Relation between azimuth and grid bearing. Grid convergence. Conformality, the Cauchy–Riemann conditions and isotropy of scale. Series expansions for the TMS transformation formulae. Modified TMS.

3.1 The derivation of the TMS formulae

In Chapter 2 we constructed the normal Mercator projection (NMS). The strength of NMS is its conformality, preserving local angles exactly and preserving shapes in "small" regions (orthomorphism). Furthermore, meridians project to grid lines and conformality implies that lines of constant azimuth project to constant grid bearings, thereby guaranteeing the continuing usefulness of NMS as an aid to navigation.

As a topographic map of the globe, NMS has shortcomings in that the projection greatly distorts shapes as one approaches the poles—because of the rapid change of scale with latitude. However, the (unmodified) NMS is exactly to scale on the equator and is fairly accurate within a narrow strip of about three degrees centred on the equator (extending to five degrees for modified NMS). It is this accuracy near the equator that we wish to exploit by constructing a projection which takes a complete meridian great circle as a 'kind of equator' and uses 'NMS on its side' to achieve a conformal and accurate projection within a narrow band adjoining the chosen meridian. This is the transverse Mercator projection (TMS) first demonstrated by Johann Lambert in 1772.

The modified versions of the transverse Mercator projection on the ellipsoid (TME, see Chapter 7) are of great importance. They are used for map projections of countries which have a predominantly north-south orientation, such as Great Britain. More importantly they provide a systematic framework for covering the the whole of the globe with conformal and accurate maps. The UTM (Universal Transverse Mercator) series covers the the globe between the latitudes of $84°N$ and $80°S$ with 60 accurate projections of width 6° in longitude centred on meridians at 3°, 9°, 15°, $\ldots$ (The polar regions are always mapped with projections centred at the poles).

Now for NMS the equator has unit scale because we project onto a cylinder tangential to the sphere at the equator. Therefore, for TMS we seek a projection onto a cylinder which is tangential to the sphere on some chosen meridian or strictly, a pair of meridians such as the great circle formed by meridians at Greenwich and 180°E: the geometry is shown in Figure 3.1a. This will guarantee that the scale is unity on the meridian: the problem is to find the functions $x(\lambda, \phi)$ and $y(\lambda, \phi)$ such that the projection is also conformal.

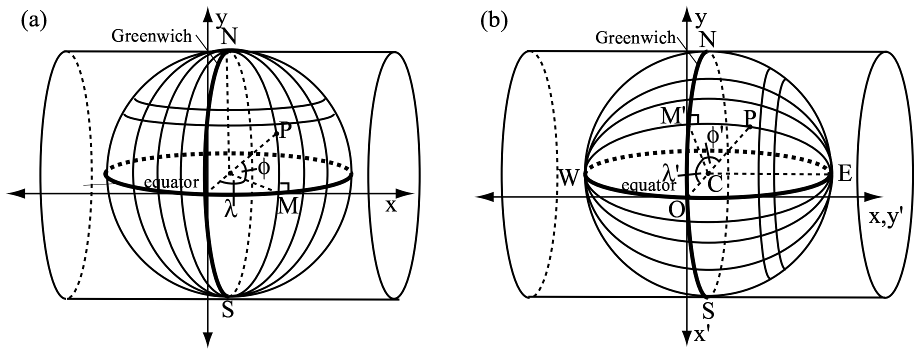


Figure 3.1

The solution is remarkably simple. We first introduce a new graticule which is simply the normal graticule of Figure 3.1a rotated so that its "equator" coincides with the chosen great circle as in Figure 3.1b. Let ϕ' and λ' be the coordinates of P with respect to the new graticule: they are the angles PCM' and OCM' on the figure. Note that, if ϕ' is measured positive from M' to the 'rotated pole' at E, then the sense in which λ' is defined on the rotated graticule is opposite to the sense of λ in the standard graticule of Figure 3.1a. In Figure 3.1b we have also shown x' and y' axes which are related to the rotated graticule in the same way that the axes were assigned for the normal NMS projection in Figure 2.4 so, bearing in mind the sense of λ', the equations (2.28) for NMS with respect to the rotated graticule are

$$x' = -a\lambda', \qquad y' = a\psi(\phi') = a \ln\left[\tan\left(\phi'/2 + \pi/4\right)\right]. \tag{3.1}$$

Now the relation between the actual TMS axes and the primed axes is simply $x = y'$ and $y = -x'$, so that we immediately have the projection formulae with respect to the angles $(\phi', \lambda',)$ of the rotated graticule:

$$x = a\psi(\phi') = a \ln\left[\tan\left(\phi'/2 + \pi/4\right)\right], \qquad y = a\lambda'. \tag{3.2}$$

It will prove more useful to use one of the alternative forms of the Mercator parameter, that in equation (2.31), giving

$$x = \frac{a}{2} \ln\left[\frac{1 + \sin\phi'}{1 - \sin\phi'}\right], \qquad y = a\lambda'. \tag{3.3}$$

All that remains is to derive the relation between (ϕ', λ') and (λ, ϕ) by a straightforward exercise in applying spherical trigonometry to the triangle $NM'P$ defined by the (true) meridians through the origin and an arbitrary point P and by the great circle $WM'PE$

(Figure 3.2a). The right-hand figure shows a similar spherical triangle in standard notation for which the sine and cosine rules (Appendix D) are (for a unit sphere)

$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c}, \tag{3.4}$$

$$\cos a = \cos b \cos c + \sin b \sin c \cos A, \tag{3.5}$$

$$\cos b = \cos c \cos a + \sin c \sin a \cos B, \tag{3.6}$$

$$\cos c = \cos a \cos b + \sin a \sin b \cos C. \tag{3.7}$$

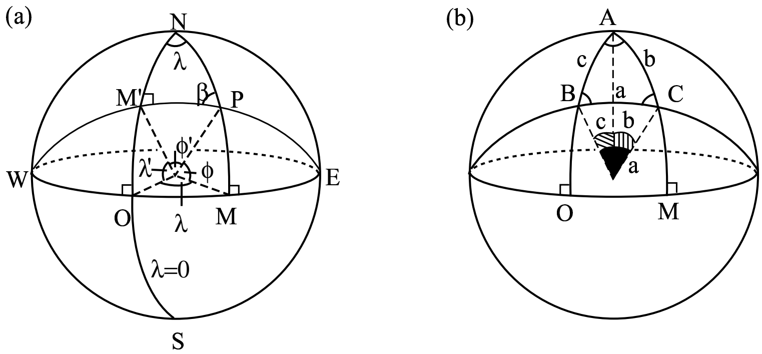


Figure 3.2

With the identifications

$$\begin{aligned} A &\to \lambda, & B &\to \frac{\pi}{2}, & C &\to \beta, \\ a &\to \phi', & b &\to \frac{\pi}{2} - \phi, & c &\to \frac{\pi}{2} - \lambda', \end{aligned} \tag{3.8}$$

the first two terms of the sine rule and the first two cosine rules give

$$\sin\phi' = \sin\lambda\cos\phi, \tag{3.9}$$

$$\cos\phi' = \sin\phi\sin\lambda' + \cos\phi\cos\lambda'\cos\lambda, \tag{3.10}$$

$$\sin\phi = \sin\lambda'\cos\phi' + 0. \tag{3.11}$$

Note the simple expression for $\sin\phi'$ in terms of λ and ϕ; this explains why we chose the alternative form of the NMS transformations in equation (3.3). To obtain the expression for λ' we eliminate $\cos\phi'$ from the last two of these equations. On simplification we find

$$\tan\lambda' = \sec\lambda\tan\phi. \tag{3.12}$$

Therefore our final expressions for TMS centred on the Greenwich meridian are

$$\boxed{\begin{aligned} x(\lambda,\phi) &= \frac{a}{2}\ln\left[\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi}\right] \\ y(\lambda,\phi) &= a\ \arctan\left[\sec\lambda\tan\phi\right] \end{aligned}} \tag{3.13}$$

For a different central meridian we simply replace λ by $\lambda - \lambda_0$.

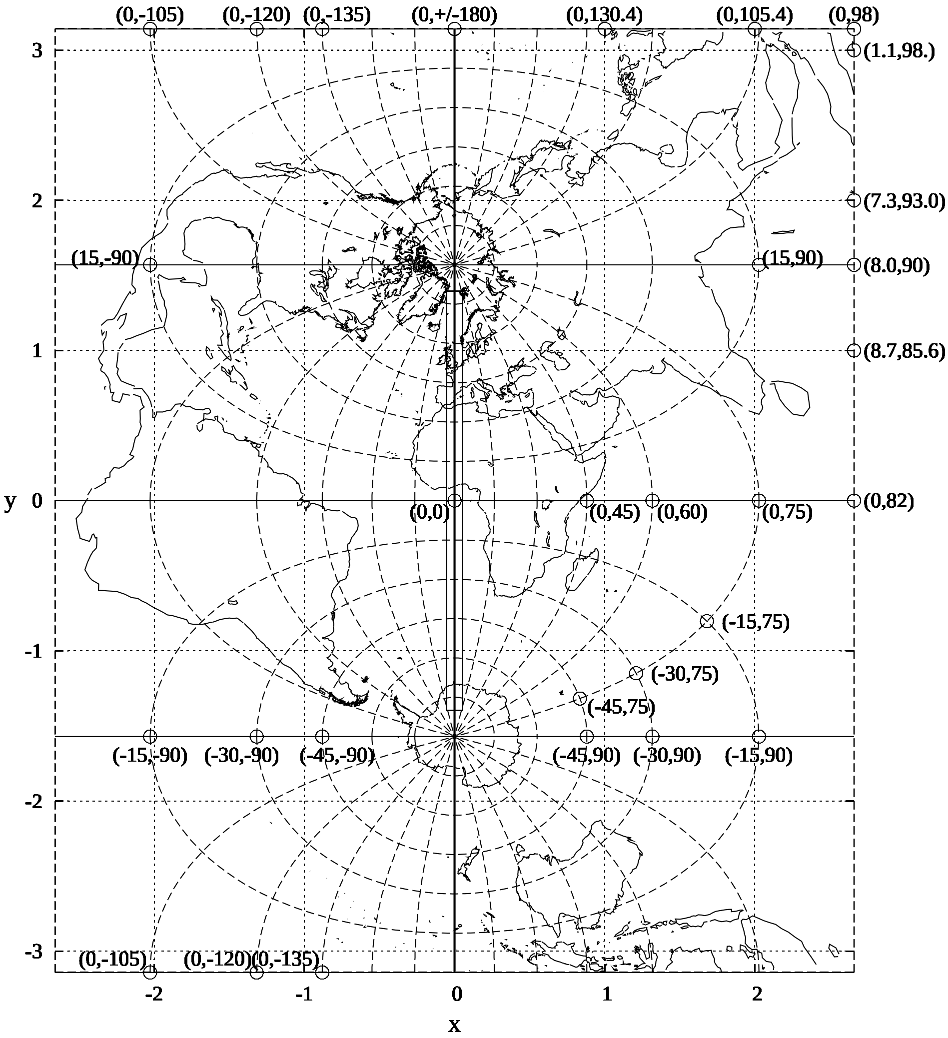


Figure 3.3 : Transverse Mercator centred on Greenwich

3.2 Features of the TMS projection

In Figures 3.3–3.5 we have constructed projections centred on Greenwich ($\lambda_0 = 0$), the Americas ($\lambda_0 = -70°$) and Australasia ($\lambda_0 = 150°$). The axes of these maps are labelled in units of Earth radius. These bizarre TMS projections covering most the Earth have very little utility other than entertainment: it is only the restricted maps near the central meridians that have practical uses for accurate mapping: the thin boxes show how much of the projection is used in one UTM projection (on the ellipsoid).

Many of the features of TMS can be understood by considering its genesis as an NMS projection ‘turned on its side’. For example the central meridian with $\lambda = 0$ projects to a straight line at unit scale defining the y-axis of the projection. It is of finite length, πa between the poles, with the finite sections of length $\pi a/2$ above N and below S corre-

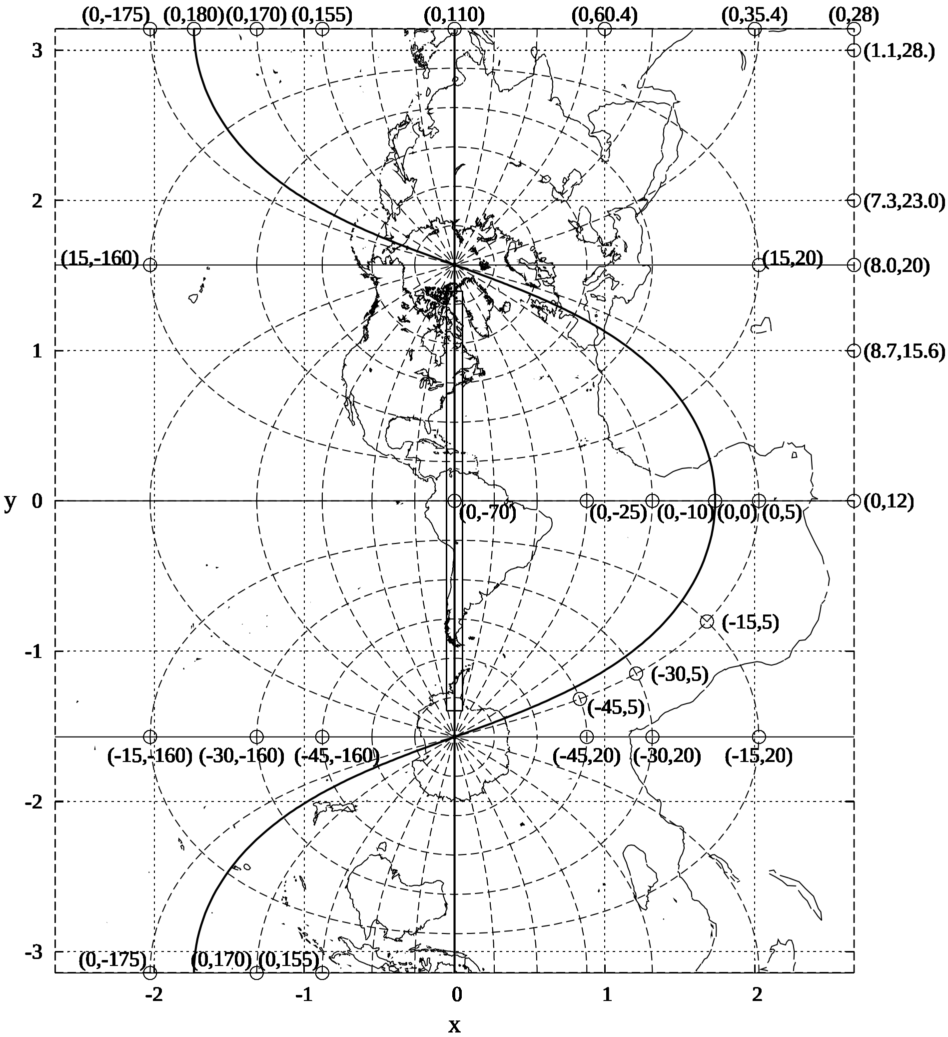


Figure 3.4 : Transverse Mercator centred on 70W

sponding to the inverted meridian of 180°. On the other hand, the x-axis is now of infinite length since E and W, the 'poles' of the rotated graticule, are projected to infinity: we have arbitrarily truncated the x-width at $x = \pm 2.5$ in the figures.

Since all other meridians pass through N and S on the y-axis they are in general complicated curves running from top to bottom of the map. The exceptions are the meridians at $\pm 90°$ which, since they are also great circles through the E and W 'poles', must extend to infinity as horizontal lines on the map. The true parallels, except the equator, map into a set of closed curves around the poles N and S; because of conformality these parallels are orthogonal to the meridians at intersections, some of which are annotated with (latitude,longitude) geographical coordinates.

The equator itself appears on the map in three horizontal lines; the 'front' equator lies

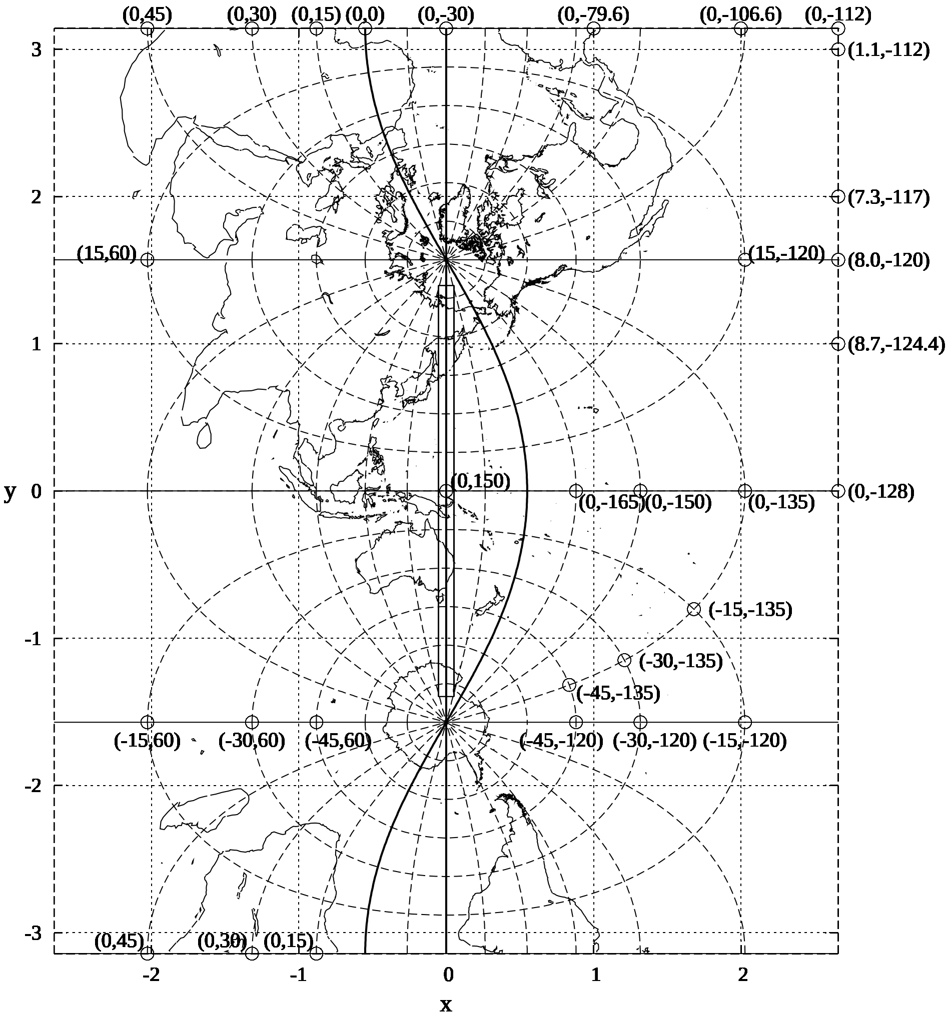


Figure 3.5 : Transverse Mercator centred on 150E

on the x-axis, extends to infinity and reappears as the 'back' equator at the top and bottom of the projection (corresponding to the generator along which the projection cylinder is cut open). Note that longitude increases to the right on the x-axis and to the left at top and bottom.

The grid lines which are parallel to the x-axis correspond to great circles which are 'meridians' of the rotated graticule. On the other hand the grid lines parallel to the y-axis have no readily defined precursors on the sphere. It is important to note that they are parallel to the true meridians only where they cross the equator and nowhere else: the angle between a (north-south) grid line and a curved meridian at a general point is called the grid convergence, discussed in Section 3.6. We have also added some geographical coordinates for grid intersections on the truncated sides and on the equator.

We shall give the formulae for the scale variation of this projection in Section 3.4.

Here we point out that the projection is quite faithful close to the central meridian but there are gross distortions as we move away from the central meridian—look at Africa on the projection centred on the Americas. This is now of no concern for we shall use the projection centred on some λ_0 only near that meridian.

We know from Section 2.5 that a rhumb line, crossing meridians at a constant angle, will reach both poles when extended on the sphere. Therefore, except for the zero azimuth case, a rhumb line cannot be a straight line on the TMS projection. For short distances TM maps are still suitable for navigation, particularly because conformality still guarantees a simple relation (not equality—more anon) between azimuth and grid bearing.

Finally note that we need not have shown the inverted sections of these maps for, by suitable choices of central meridians, what is inverted on one map will be the right way up on some other. Observe that in plotting these maps we cannot get the inverted sections, where either $y > \pi a/2$ or $y < -\pi a/2$, by using equation (3.13) with the arctan function in its principal interval, $(-\pi/2, \pi/2)$. However arctan is a multivalued function, arbitrary to within an additive factor of $N\pi$. To plot the figures we used:

$$y = \left\{ \begin{array}{c} a\pi \\ 0 \\ -a\pi \end{array} \right\} + a \arctan\left[\sec\lambda \tan\phi\right] \qquad \text{for} \left\{ \begin{array}{ll} |\lambda| > \pi/2, & \phi > 0 \\ |\lambda| < \pi/2, & \\ |\lambda| > \pi/2, & \phi < 0 \end{array} \right. . \tag{3.14}$$

3.3 The inverse transformations

Equations (3.13) can be easily inverted to give

$$\sin\lambda\cos\phi = \tanh(x/a), \tag{3.15}$$

$$\sec\lambda\tan\phi = \tan(y/a). \tag{3.16}$$

Eliminating ϕ gives

$$\begin{aligned} \sec^2\!\phi = \sin^2\!\lambda \coth^2(x/a) &= 1 + \cos^2\!\lambda \tan^2(y/a), \\ \tan^2\!\lambda \left(\coth^2(x/a) - 1\right) &= \sec^2(y/a), \\ \tan\lambda &= \sinh(x/a)\sec(y/a), \end{aligned} \tag{3.17}$$

thus determining λ as a function of x and y.

To find ϕ as a function of x and y multiply equations (3.15) and (3.16) to give

$$\tan\lambda\sin\phi = \tanh(x/a)\tan(y/a). \tag{3.18}$$

Using equation (3.17) to eliminate $\tan\lambda$ then gives

$$\sin\phi = \operatorname{sech}(x/a)\sin(y/a). \tag{3.19}$$

Thus the final result for the inverse transformations is

$$\lambda(x,y) = \arctan\left[\sinh(x/a)\sec(y/a)\right], \tag{3.20}$$

$$\phi(x,y) = \arcsin\left[\operatorname{sech}(x/a)\sin(y/a)\right], \tag{3.21}$$

where we take principal values in $[-\pi/2, \pi/2]$ to correspond to the 'front' of the sphere.

The meridian distance, the footpoint and the footpoint latitude

This is an appropriate place for three definitions which, whilst both trivial and superfluous for the sphere, become very important when we study the inverse transformations on the ellipsoid. Let $P'(x,y)$ be a general point on the projection.

- The **meridian distance** $m(\phi) \equiv a\phi$ is the distance on the sphere measured along the central meridian from the origin on the equator to a point at latitude ϕ. On the sphere

$$m(\phi) = a\phi. \tag{3.22}$$

- The **footpoint** of $P'(x,y)$ on the projection is that point P'_1 on the central meridian of the *projection* which has the same ordinate as P'. The coordinates of the footpoint are $P'_1(0,y)$.
- The **footpoint latitude**, ϕ_1, is the latitude of that point P_1 on the central meridian of the *sphere* which projects into the footpoint $P'_1(0,y)$. It is *not* the latitude of the point P which is the inverse of P'.

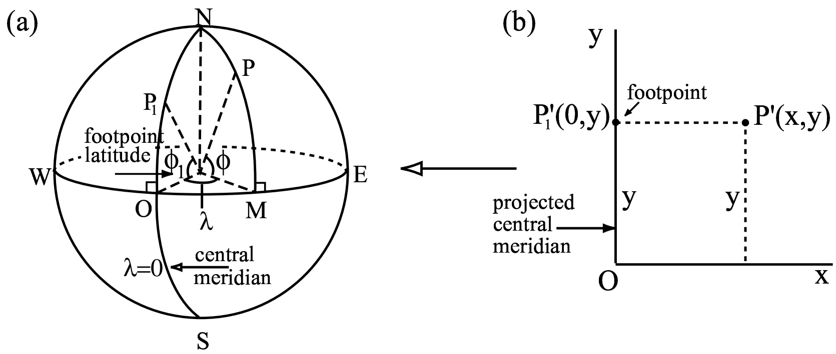


Figure 3.6

From these definitions and equation (3.21) we have

$$\phi_1 = \phi(0,y) = \arcsin\left[\sin(y/a)\right] = y/a. \tag{3.23}$$

This is obvious because, by construction, the scale of the projection is true on the central meridian so that $y = a\phi_1$ and hence $\phi_1 = y/a$. In terms of the meridian distance function, defined in equation (3.22) we see that ϕ_1 satisfies

$$m(\phi_1) = y. \tag{3.24}$$

We now take this equation as the definition of the footpoint latitude since we will find that it continues to hold on the ellipsoid where $m(\phi)$ is a non-trivial function. For future reference we write equations (3.20) and (3.21) as

$$\lambda(x,y) = \arctan\left[\sinh(x/a)\sec\phi_1\right], \tag{3.25}$$

$$\phi(x,y) = \arcsin\left[\operatorname{sech}(x/a)\sin\phi_1\right] \qquad m(\phi_1) = y. \tag{3.26}$$

3.4 The scale factor for the TMS projection

Because of the way in which the TMS was constructed, by applying NMS to a rotated graticule, we *know* that the scale factor for TMS is isotropic and, in terms of the rotated latitude ϕ', its value is $k = \sec\phi'$. Using (3.9) and (3.15) we find the scale factor in terms of either geographical or projection coordinates:

$$k(\lambda,\phi) = \frac{1}{(1-\sin^2\lambda\cos^2\phi)^{1/2}} \tag{3.27}$$

$$k(x,y) = \cosh(x/a). \tag{3.28}$$

[See comments after equation 2.45]. Note that the scale is a complicated function of the geographical coordinates but is simply a function of the x-coordinate on the projection. Both forms show that the scale is unity on the central meridian. ($\lambda = 0$ or $x = 0$).

3.5 Azimuths and grid bearings in TMS

To investigate the relation between azimuths on the sphere and grid bearings on the projection we consider the relation of the infinitesimal elements shown in Figure 3.7. Now strictly, an infinitesimal element on the projection would be a quadrilateral but we have drawn it as curvilinear quadrilateral to emphasize the fact that the meridian MP, the parallel PN and the displacement PQ will in general project to curved lines on the map. The relevant angles must be defined with respect to the tangents of these lines at P'. The angles of concern are

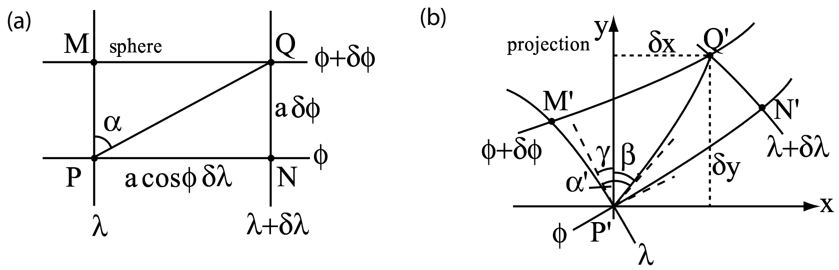


Figure 3.7

- α, the angle between the meridian at P on the sphere and a general displacement PQ.
- α', the angle between the projected meridian and the projected displacement $P'Q'$.
- β, the grid bearing, is the angle between the projected displacement and the y-axis.
- γ, the angle between the projected meridian and the y-axis: this angle is the (grid) **convergence** of the projection at P'; in the next section we show how it is calculated.

- Clearly $\alpha' = \beta + \gamma$.

The construction of TMS guarantees conformality so the corresponding angles α and α' must be equal. Therefore

$$\boxed{\alpha = \beta + \gamma} \tag{3.29}$$

or, in words:

AZIMUTH = GRID BEARING +CONVERGENCE

This equation is to be used in both directions. If we are given an azimuth α at some point on the sphere then the corresponding bearing on the map (chart) can be calculated from $\beta = \alpha - \gamma(\lambda, \phi)$. Likewise, given a bearing on the chart at (x, y) we find the azimuth at the corresponding point on the sphere from $\alpha = \beta + \gamma(x, y)$. Clearly we need to find expressions for the convergence in terms of both geographic and projection coordinates.

Although the convergence can take a wide range of values on small scale TMS projections (such as Figure 3.3), remember that the projection will be applied only in the region very close to the central meridian where the the non-central meridian lines make very small angles with the y-axis. For example, over Great Britain the convergence of the OSGB maps is never greater than $5°$.

3.6 The grid convergence of the TMS projection

The figure shows a section of the $45°$E meridian between the equator and the north pole of the TMS projection of Figure 3.3. Since TMS is conformal the angle between this projected meridian and the y-axis must be $45°$ at the pole. The figure also shows some grid lines and the (y increasing) direction of these lines is defined as **grid north** even though these lines, the y-axis excepted, do not pass through the north pole on the projection. We also define the tangent of the meridian at P' to be the direction of **true north** at that point even though the tangent does not point directly to the pole at N. We can therefore recast the definition of convergence given in the last section. It is the angle between grid north and true north at a point P' on the projection and it is usually specified as so many degrees west or east of grid north. For more general mathematical work we use a signed convergence defined by

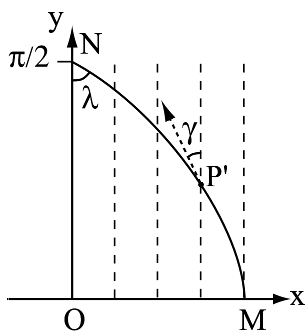


Figure 3.8

$$\tan\gamma = -\left.\frac{dx}{dy}\right|_{P'} \tag{3.30}$$

so that in the quadrant shown in the figure, where $\delta x < 0$ when $\delta y > 0$, the convergence γ is positive. (Thus a positive convergence is to the west of grid north).

Now the increments in $x(\lambda,\phi)$ and $y(\lambda,\phi)$ for arbitrary changes in ϕ and λ are

$$\delta x = \left(\frac{\partial x}{\partial\lambda}\right)\delta\lambda + \left(\frac{\partial x}{\partial\phi}\right)\delta\phi, \tag{3.31}$$

$$\delta y = \left(\frac{\partial y}{\partial\lambda}\right)\delta\lambda + \left(\frac{\partial y}{\partial\phi}\right)\delta\phi, \tag{3.32}$$

but the tangent at P' is along the projection of a meridian on which $\delta\lambda = 0$. Therefore

$$\tan\gamma = -\left.\frac{dx}{dy}\right|_{\delta\lambda=0} = -\left(\frac{\partial x}{\partial\phi}\bigg/\frac{\partial y}{\partial\phi}\right) = -\frac{x_\phi}{y_\phi}, \tag{3.33}$$

$$\boxed{\gamma = -\arctan\left(\frac{x_\phi}{y_\phi}\right)}. \tag{3.34}$$

The partial derivatives must evaluated from equations (3.13); to put them in simpler forms we use equation (3.9) and some equivalent forms

$$\sin\phi' = \sin\lambda\cos\phi, \tag{3.35}$$

$$\begin{aligned}\cos^2\phi' &= 1-\sin^2\lambda\cos^2\phi = \sin^2\phi + \cos^2\lambda\cos^2\phi\\ &= \cos^2\lambda + \sin^2\phi\sin^2\lambda = \cos^2\lambda\cos^2\phi\,(1+\sec^2\lambda\tan^2\phi).\end{aligned} \tag{3.36}$$

Therefore

$$x = \frac{a}{2}\ln\left[\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi}\right], \qquad y = a\ \arctan\left[\sec\lambda\tan\phi\right], \tag{3.37}$$

$$\frac{\partial x}{\partial\lambda} = a\sec^2\phi'\cos\lambda\cos\phi, \qquad \frac{\partial y}{\partial\lambda} = a\sec^2\phi'\sin\lambda\sin\phi\cos\phi, \tag{3.38}$$

$$\frac{\partial x}{\partial\phi} = -a\sec^2\phi'\sin\lambda\sin\phi, \qquad \frac{\partial y}{\partial\phi} = a\sec^2\phi'\cos\lambda. \tag{3.39}$$

The convergence as a function of geographic coordinates follows from equation (3.34):

$$\gamma(\lambda,\phi) = \arctan\left(\tan\lambda\sin\phi\right). \tag{3.40}$$

This result can be written in terms of x and y by using equation (3.18) giving

$$\gamma(x,y) = \arctan\left[\,\tanh(x/a)\tan(y/a)\right]. \tag{3.41}$$

It will prove useful to write this result in terms of x and the footpoint latitude as

$$\gamma(x,y) = \arctan\left[\,\tanh(x/a)\tan\phi_1\right], \qquad m(\phi_1) = y. \tag{3.42}$$

3.7 Conformality of general projections

So far we have claimed, fairly, that TMS is conformal with an isotropic scale factor by virtue of the method we used to derive the projection, *viz* NMS 'on its side'. It is instructive to ask how we may decide that an arbitrary projection from the sphere satisfies these conditions. To this end consider Figure 3.7 where the azimuth angle of the displacement PQ on the sphere is given by

$$\tan\alpha = \lim_{Q\to P} \frac{\cos\phi\,\delta\lambda}{\delta\phi} = \lim_{Q\to P} R\frac{\delta\lambda}{\delta\phi}, \tag{3.43}$$

where (for future developments) it is convenient to set $R = \cos\phi$. Now consider the grid bearing of the corresponding displacement $P'Q'$ for an arbitrary projection. Using equations (3.31, 3.32), with the constraint implied by the above equation, we have

$$\tan\beta = \lim \frac{\delta x}{\delta y} = \lim \left.\frac{x_\lambda\delta\lambda + x_\phi\delta\phi}{y_\lambda\delta\lambda + y_\phi\delta\phi}\right|_{\delta\phi=\frac{R\,\delta\lambda}{\tan\alpha}} = \frac{x_\lambda\tan\alpha + x_\phi R}{y_\lambda\tan\alpha + y_\phi R}. \tag{3.44}$$

We already know $\tan\gamma$ from equation (3.33), therefore we can calculate α', the angle between the projected meridian and parallel, as

$$\tan\alpha' = \tan(\beta+\gamma) = \frac{\tan\beta + \tan\gamma}{1 - \tan\beta\tan\gamma} \tag{3.45}$$

$$= \frac{y_\phi(x_\lambda\tan\alpha + x_\phi R) - x_\phi(y_\lambda\tan\alpha + y_\phi R)}{y_\phi(y_\lambda\tan\alpha + y_\phi R) + x_\phi(x_\lambda\tan\alpha + x_\phi R)}$$

$$= \frac{(x_\lambda y_\phi - x_\phi y_\lambda)\tan\alpha}{R\,(x_\phi^2 + y_\phi^2) + (x_\lambda x_\phi + y_\lambda y_\phi)\tan\alpha}. \tag{3.46}$$

The projection will be conformal if $\tan\alpha' = \tan\alpha$ so that

$$(x_\lambda x_\phi + y_\lambda y_\phi)\tan\alpha + \left[R\,(x_\phi^2 + y_\phi^2) - (x_\lambda y_\phi - x_\phi y_\lambda)\right] \equiv 0. \tag{3.47}$$

This is an identity which must hold for all values of α, therefore the coefficient of $\tan\alpha$ and the constant term must both vanish. This gives two conditions:

$$x_\lambda x_\phi + y_\lambda y_\phi = 0 \tag{3.48}$$

$$R\,(x_\phi^2 + y_\phi^2) = (x_\lambda y_\phi - x_\phi y_\lambda). \tag{3.49}$$

Using (3.48), the second of these equations can be written as

$$R\,y_\phi^2\left(1 + x_\phi^2/y_\phi^2\right) = x_\lambda y_\phi\left(1 + x_\phi^2/y_\phi^2\right), \tag{3.50}$$

so that we must have $x_\lambda = R\,y_\phi$. If we then substitute this back into (3.48) we obtain $y_\lambda = -R\,x_\phi$. Thus, restoring R, the following conditions are necessary (and trivially sufficient) for a conformal transformation from the sphere to the plane.

CAUCHY–RIEMANN $$\boxed{x_\lambda = \cos\phi\; y_\phi, \quad y_\lambda = -\cos\phi\; x_\phi} \tag{3.51}$$

It is trivial to check that these Cauchy–Riemann conditions are satisfied for both NMS and TMS: in the first case have (from 2.28) $x_\lambda{=}a,\ \ y_\phi{=}a\sec\phi$ and $x_\phi{=}y_\lambda{=}0$; TMS follows immediately from equations (3.38, 3.39).

Conformality implies scale isotropy

Consider now the scale factor for an arbitrary transformation. Substituting δx and δy from equations (3.31 , 3.32) into the definition of the scale factor (equation 2.19) we have

$$\mu^2 = \lim_{Q\to P} \frac{\delta s'^2}{\delta s^2} = \lim_{Q\to P} \frac{\delta x^2 + \delta y^2}{a^2\delta\phi^2 + a^2\cos^2\!\phi\,\delta\lambda^2}$$
$$= \lim_{Q\to P} \frac{E\delta\phi^2 + 2F\delta\phi\,\delta\lambda + G\delta\lambda^2}{a^2\delta\phi^2 + a^2\cos^2\!\phi\,\delta\lambda^2}. \tag{3.52}$$

where

$$E(\lambda,\phi) = x_\phi^2 + y_\phi^2, \qquad F(\lambda,\phi) \quad = x_\lambda x_\phi + y_\lambda y_\phi, \qquad G(\lambda,\phi) = x_\lambda^2 + y_\lambda^2. \tag{3.53}$$

An isotropic scale factor must be independent of the azimuth α; in other words (from 3.43) it must be independent of the ratio of $\delta\phi/\delta\lambda$. This is *always* the case when the Cauchy–Riemann equations (3.51) are satisfied, for then we must have $F = 0$ and $G = \cos^2\!\phi\, E$. The isotropic scale factor is then $\sqrt{E}/a$ or

ISOTROPIC SCALE
$$\boxed{k(\lambda,\phi) = \frac{1}{a}\sqrt{x_\phi^2 + y_\phi^2} = \frac{1}{a\cos\phi}\sqrt{x_\lambda^2 + y_\lambda^2}}\,. \tag{3.54}$$

Therefore ALL conformal transformations have isotropic scale factors. It is a simple exercise to show that the above equation reduces to $\sec\phi$ for NMS and to $\sec\phi'$ for TMS; for the latter use equations (3.35–3.39) to confirm the results of Section 3.4.

3.8 Series expansions for the unmodified TMS

In the calculations for the transformations on the ellipsoid we shall have to resort to series solutions. In this section we will derive the corresponding series for TMS. For the direct transformations we hold ϕ constant and expand in terms of λ and for the inverse transformations we hold y constant and expand in terms of x/a. Typically the half-width (at the equator) of a transverse projection is about 3° on the sphere and about 330km on the projection so that $\lambda < .05$ (radians) and $x/a < 0.05$. We shall drop terms involving fifth or higher powers of these small parameters.

The coefficients of the direct series involve trigonometric functions of ϕ, which is not generally a small term: for example $\tan\phi$ is about 1.7 at 60°N. Likewise, the coefficients of the inverse series will be functions of the footpoint latitude ϕ_1 which again is not generally small. It is convenient to introduce the following compact notation for the trigonometric functions of ϕ and ϕ_1:

$$s = \sin\phi \qquad c = \cos\phi \qquad t = \tan\phi \tag{3.55}$$
$$s_1 = \sin\phi_1 \qquad c_1 = \cos\phi_1 \qquad t_1 = \tan\phi_1 \qquad m(\phi_1) = y, \tag{3.56}$$

where $m(\phi) = a\phi$ is the meridian distance and ϕ_1 is the footpoint latitude.

All of the Taylor series that we need for the expansions are collected in Appendix E.

Direct transformation for x

Equation (3.13a) is $\qquad x = \frac{a}{2}\ln\left[\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi}\right] = \frac{a}{2}\ln\left[\frac{1+c\sin\lambda}{1-c\sin\lambda}\right].$

Since $\sin\lambda \ll 1$ we first expand the logarithm with (E.12) and then substitute for $\sin\lambda$ with (E.13) to obtain

$$\begin{aligned} x(\lambda,\phi) &= ac(\lambda - \frac{1}{6}\lambda^3 + \cdots) + \frac{1}{3}ac^3(\lambda - \cdots)^3 + \cdots \\ &= ac\lambda + \frac{1}{6}ac^3(1-t^2)\lambda^3 + \cdots . \end{aligned} \tag{3.57}$$

Direct transformation for y

Equation (3.13b) is $\qquad y(\lambda,\phi) = a\ \arctan\left[\sec\lambda\tan\phi\right] = a\ \arctan\left[t\sec\lambda\right].$

The argument of the arctan is not small but, using (E.16), we have

$$\begin{aligned} t\sec\lambda &= t\left(1 + \frac{1}{2}\lambda^2 + \frac{5}{24}\lambda^4 + \cdots\right) \\ &= t + z, \qquad \text{with} \quad z = t\left(\frac{1}{2}\lambda^2 + \frac{5}{24}\lambda^4 + \cdots\right) \ll 1. \end{aligned}$$

Using (E.9) with $b = t$ we have

$$\begin{aligned} y(\lambda,\phi) &= a\arctan(t+z) \\ &= a\arctan(t) + at\left(\frac{1}{2}\lambda^2 + \frac{5}{24}\lambda^4\right)\frac{1}{1+t^2} + at^2\left(\frac{1}{2}\lambda^2 + \cdots\right)^2\frac{(-t)}{(1+t^2)^2}. \end{aligned}$$

Now $a\arctan(t) = a\arctan(\tan\phi) = a\phi$ so we can write

$$y(\lambda,\phi) = a\phi + \frac{asc}{2}\lambda^2 + \frac{asc^3\lambda^4}{24}\left(5-t^2\right) + \cdots . \tag{3.58}$$

Inverse transformation for λ

Setting $y/a = \phi_1$, the footpoint latitude, equation (3.20) is

$$\lambda(x,y) = \arctan\left[\sinh(x/a)\sec(y/a)\right] = \arctan\left[c_1^{-1}\sinh(x/a)\right].$$

Since $\sinh(x/a) \ll 1$ we can expand with (E.20) and then substitute with (E.21) giving

$$\begin{aligned} \lambda(x,y) &= c_1^{-1}\left(\frac{x}{a} + \frac{1}{6}\frac{x^3}{a^3} + \cdots\right) - c_1^{-3}\frac{1}{3}\left(\frac{x}{a} + \cdots\right)^3 + \cdots \\ &= c_1^{-1}\left(\frac{x}{a}\right) + c_1^{-1}\left(\frac{1}{6} - \frac{1}{3}c_1^{-2}\right)\left(\frac{x}{a}\right)^3 + \cdots \\ &= \frac{1}{c_1}\left(\frac{x}{a}\right) - \frac{(1+2t_1^2)}{6c_1}\left(\frac{x}{a}\right)^3 + \cdots . \qquad \text{where} \quad \phi_1 = y/a. \end{aligned} \tag{3.59}$$

Inverse transformation for ϕ

Setting $y/a = \phi_1$ in equation (3.21) gives

$$\phi(x,y) = \arcsin\left[\operatorname{sech}(x/a)\sin(y/a)\right] = \arcsin\left[s_1 \operatorname{sech}(x/a)\right].$$

We first use (E.24) to write

$$\begin{aligned} s_1 \operatorname{sech}(x/a) &= s_1\left(1 - \frac{1}{2}\left(\frac{x}{a}\right)^2 + \frac{5}{24}\left(\frac{x}{a}\right)^4 + \cdots\right) \\ &= s_1 + z \qquad \text{with} \quad z = s_1\left(-\frac{1}{2}\left(\frac{x}{a}\right)^2 + \frac{5}{24}\left(\frac{x}{a}\right)^4 + \cdots\right) \end{aligned}$$

Using (E.8) with $b = s_1$ we obtain

$$\begin{aligned} \phi(x,y) &= \arcsin s_1 + \frac{s_1}{(1-s_1^2)^{1/2}}\left(-\frac{1}{2}\left(\frac{x}{a}\right)^2 + \frac{5}{24}\left(\frac{x}{a}\right)^4\right) \\ &\qquad + \frac{1}{2}\frac{s_1}{(1-s_1^2)^{3/2}} s_1^2\left(-\frac{1}{2}\left(\frac{x}{a}\right)^2 + \cdots\right)^2 + \cdots \\ &= \phi_1 - \frac{t_1}{2}\left(\frac{x}{a}\right)^2 + \frac{t_1}{24}(5+3t_1^2)\left(\frac{x}{a}\right)^4 + \cdots. \qquad \text{where} \quad \phi_1 = y/a. \end{aligned} \tag{3.60}$$

Series expansion for the scale factor

Using the binomial series (E.29) with $z = -\sin^2\lambda\cos^2\phi = -c^2\sin^2\lambda$ and substituting for $\sin\lambda$ with (E.13), we find that equation (3.27) gives

$$\begin{aligned} k(\lambda,\phi) &= [1 - \sin^2\lambda\cos^2\phi]^{-1/2} \\ &= 1 + \frac{1}{2}c^2\left(\lambda - \frac{1}{6}\lambda^3 + \cdots\right)^2 + \frac{3}{8}c^4\left(\lambda - \cdots\right)^4 \\ &= 1 + \frac{1}{2}c^2\lambda^2 + \frac{1}{24}c^4\lambda^4(5 - 4t^2) + \cdots \end{aligned} \tag{3.61}$$

Similarly equations (3.28) and (E.22) give

$$k(x,y) = \cosh(x/a) = 1 + \frac{1}{2!}\left(\frac{x}{a}\right)^2 + \frac{1}{4!}\left(\frac{x}{a}\right)^4 + \cdots. \tag{3.62}$$

Series expansion for convergence

Equation (3.40) is $\gamma(\lambda,\phi) = \arctan\left[\tan\lambda\sin\phi\right]$. Expanding $\tan\lambda$ with (E.15) and using the expansion for arctan in equation (E.20) gives

$$\begin{aligned} \gamma(\lambda,\phi) &= s(\lambda + (1/3)\lambda^3 + \cdots) - (1/3)s^3(\lambda + \cdots)^3 \\ &= s\lambda + \frac{1}{3}sc^2\lambda^3 + \cdots \end{aligned} \tag{3.63}$$

Equation (3.41) is

$$\gamma(x,y) = \arctan\big[\tanh(x/a)\tan(y/a)\big] = \arctan\big[t_1\tanh(x/a)\big].$$

Expanding $\tanh(x/a)$ for small x with (E.23) and again using (E.20) for arctan gives

$$\begin{aligned}\gamma(x,y) &= t_1\left(\frac{x}{a} - \frac{1}{3}\left(\frac{x}{a}\right)^3 + \cdots\right) - (1/3)t_1^3\left(\frac{x}{a} + \cdots\right)^3 \\ &= t_1\left(\frac{x}{a}\right) - \frac{1}{3}\frac{t_1}{c_1^2}\left(\frac{x}{a}\right)^3 + \cdots \end{aligned} \tag{3.64}$$

3.9 Modified TMS

In Section 2.7 we showed how the NMS was modified to obtain greater accuracy over wider areas by reducing the scale factor on the equator. We do the same for the TMS, reducing the scale on the central meridian by simply multiplying the transformation formulae in equations (3.13) by a factor of k_0. The corresponding equations for the inverses, scale factors and convergence are easily deduced: they are listed below along with the corresponding series solutions. We continue to use the abbreviations for the trig functions of ϕ and ϕ_1 (equations 3.55, 3.56)

Direct transformations

$$x(\lambda,\phi) = \frac{1}{2}k_0 a\ln\left[\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi}\right] = k_0 a\left(c\lambda + \frac{1}{6}c^3\lambda^3(1-t^2) + \cdots\right) \tag{3.65}$$

$$y(\lambda,\phi) = k_0 a\arctan\left[\sec\lambda\tan\phi\right] \quad = k_0 m(\phi) + k_0 a\left(\frac{sc}{2}\lambda^2 + \frac{sc^3\lambda^4}{24}\left(5-t^2\right) + \cdots\right) \tag{3.66}$$

Note that on the central meridian given by $\lambda = 0$ we have $y(\phi,0) = k_0 m(\phi) = k_0 a\phi$. Therefore for the modified projection we must define the footpoint latitude by

$$\phi_1 = \frac{y}{k_0 a}. \tag{3.67}$$

Inverse transformations

$$\lambda(x,y) = \arctan\left[\sinh\frac{x}{k_0 a}\sec\frac{y}{k_0 a}\right] \quad = \frac{1}{c_1}\left(\frac{x}{k_0 a}\right) - \frac{(1+2t_1^2)}{6c_1}\left(\frac{x}{k_0 a}\right)^3 + \cdots \tag{3.68}$$

$$\phi(x,y) = \arcsin\left[\operatorname{sech}\frac{x}{k_0 a}\sin\frac{y}{k_0 a}\right] \quad = \phi_1 - \frac{t_1}{2}\left(\frac{x}{k_0 a}\right)^2 + \frac{t_1}{24}(5+3t_1^2)\left(\frac{x}{k_0 a}\right)^4 + \cdots \tag{3.69}$$

Convergence

$$\gamma(\lambda,\phi) = \arctan(\tan\lambda\sin\phi) \qquad = s\lambda + \frac{1}{3}sc^2\lambda^3 + \cdots \tag{3.70}$$

$$\gamma(x,y) = \arctan\left(\tanh\frac{x}{k_0a}\tan\frac{y}{k_0a}\right) = t_1\left(\frac{x}{a}\right) - \frac{1}{3}\frac{t_1}{c_1^2}\left(\frac{x}{a}\right)^3 + \cdots \tag{3.71}$$

Scale factors

$$k(\lambda,\phi) = \frac{k_0}{(1-\sin^2\lambda\cos^2\phi)^{1/2}} = k_0\left[1 + \frac{1}{2}c^2\lambda^2 + \frac{1}{24}c^4\lambda^4(5-4t^2) + \cdots\right] \tag{3.72}$$

$$k(x,y) = k_0\cosh\left(\frac{x}{k_0a}\right) \qquad = k_0\left[1 + \frac{1}{2!}\left(\frac{x}{k_0a}\right)^2 + \frac{1}{4!}\left(\frac{x}{k_0a}\right)^4 + \cdots\right] \tag{3.73}$$

Consider the scale factor in terms of projection coordinates, that is $k(x,y)$. If we choose $k_0 = 0.9996$ then we easily find that the scale is true when $x/a = \pm 0.0282$ corresponding to $x = \pm 180$km (approximately). Once outside these lines the accuracy decreases as k increases without limit. (The value of k reaches 1.0004 when $x = 255$km so that k increases from 1 to 1.0004 in a distance of 75km. This is less than half of the distance over which the scale changes from $k = 0.9996$ on the central meridian to $k = 1$ at $x = 180$km.)

Thus we see that the modified TMS is reasonably accurate over a width of approximately 510km. We shall see later that this includes most of the area covered by the British grid.

Chapter 4

NMS to TMS by complex variables

Abstract

The methods of complex variable theory are used to derive the TMS from NMS by (a) an explicit closed formula, and (b) a Taylor series expansion.

4.1 Introduction

In Chapter 2 we derived the NMS projection: it takes a point $P(\phi,\lambda)$ on the sphere to a point on a plane defined by projection coordinates (x,y) with

$$x = a\lambda, \qquad y = a\psi = a\ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right]. \tag{4.1}$$

Similarly, in Chapter 3 we derived the TMS projection: it takes a point $P(\phi,\lambda)$ on the sphere to a point on a plane defined by projection coordinates (x,y) with

$$x(\phi,\lambda) = \frac{a}{2}\ln\left[\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi}\right], \tag{4.2}$$

$$y(\phi,\lambda) = a\ \arctan\left[\sec\lambda\tan\phi\right]. \tag{4.3}$$

In addition to these closed forms we also derived series expansions for TMS which neglect terms of order λ^5 and higher:

$$x(\lambda,\phi) = ac\lambda + \frac{1}{6}ac^3(1-t^2)\lambda^3 + \cdots, \tag{4.4}$$

$$y(\lambda,\phi) = a\phi + \frac{asc}{2}\lambda^2 + \frac{asc^3\lambda^4}{24}\left(5-t^2\right) + \cdots. \tag{4.5}$$

The purpose of this chapter is to show how the above projection equations for TMS , both closed forms and series, may be derived directly from the NMS projection equations. To avoid confusion of the two sets of projection coordinates we shall reserve (x,y) for the TMS projection and refer to the NMS projection by coordinates (λ,ψ)—note the order. For the transformation from NMS to TMS we seek functions $x(\lambda,\psi)$ and $y(\lambda,\psi)$: for the inverse transformations we seek two functions $\lambda(x,y)$ and $\psi(x,y)$: the latter then gives $\phi(x,y)$ by inverting one of (2.28, 2.32) or by a Taylor series.

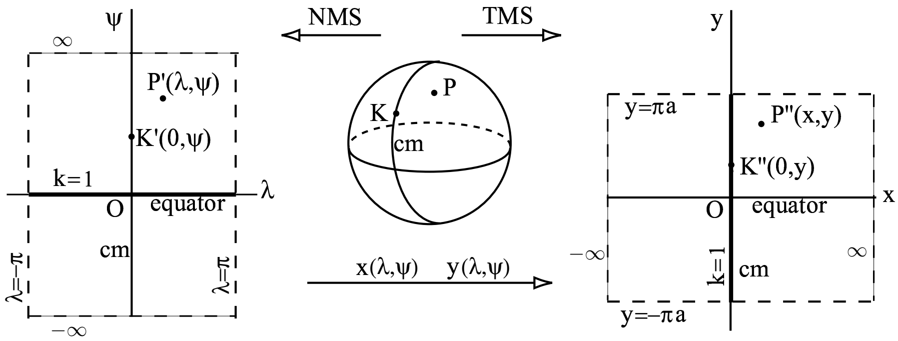


Figure 4.1

Figure 4.1 summarizes the properties of the two projections. NMS is conformal, true to scale on the equator (the λ-axis), with finite extent in λ and infinite extent in ψ. TMS is also conformal, true to scale on the central meridian (the y-axis), with finite extent in y and infinite extent in x. A general point on the sphere $P(\phi, \lambda)$ projects into points $P'(\lambda, \psi)$ and $P''(x, y)$ for NMS, TMS respectively; a general point on the central meridian of the sphere, taken as the Greenwich meridian for simplicity, projects into points $K'(0, \psi)$ and $K''(0, y)$. We shall prove that the following conditions are sufficient to determine the functions $x(\lambda, \psi)$ and $y(\lambda, \psi)$ which define the transformation of NMS to TMS.

- The central meridian of NMS, $\lambda = 0$, transforms to the central meridian of TMS:

$$x(0, \psi) = 0. \tag{4.6}$$

- The scale on the y-axis of TMS is true so that the distance OK'' on TMS is equal to OK on the sphere; therefore $y = m(\phi) = a\phi$.
- The functions $x(\lambda, \psi)$ and $y(\lambda, \psi)$ must describe a conformal transformation: any two lines through P' project into lines intersecting at the same angle at P''.

The first of these conditions is trivial but the others require further discussion.

The meridian distance as a function of ψ

We need to consider the meridian distance as a function M of ψ as well as the usual function m of ϕ; we equate $M(\psi)$ and $m(\phi)$ or, more strictly, we set

$$M\big(\psi\big) = m\big(\phi(\psi)\big). \tag{4.7}$$

Thus the scale condition on the y-axis may be written as

$$y(0, \psi) = M\big(\psi\big). \tag{4.8}$$

On the sphere, where $m(\phi) = a\phi(\psi)$, we can use equation (2.32a) to obtain an explicit expression for $M(\psi)$:

$$M(\psi) = a \arctan\big[\sinh(\psi)\big]. \tag{4.9}$$

In our subsequent calculations we shall need the first four derivatives of $M(\psi)$ with respect to ψ. These are straightforward enough to obtain as functions of ψ from the last equation but, at the end of the day, it will prove more useful to express the derivatives in terms of ϕ. For example we have

$$M'(\psi) \equiv \frac{dM(\psi)}{d\psi} = \frac{dm(\phi)}{d\phi}\frac{d\phi}{d\psi} = a\cos\phi, \tag{4.10}$$

where we have used the equation

$$\frac{d\psi}{d\phi} = \sec\phi \tag{4.11}$$

which is the definition of $\psi(\phi)$ in Section 2.4. Proceeding in this way we can construct the first four derivatives of $M(\psi)$ with respect to ψ but with the results expressed as functions of ϕ (using the compact notation for $\sin\phi$ *etc.* defined in Section 3.8).

$$\begin{aligned}
M' &= a\cos\phi & &= \;\; ac,\\
M'' &= \frac{d(a\cos\phi)}{d\phi}\frac{d\phi}{d\psi} & &= -asc,\\
M''' &= -a(c^2 - s^2)c & &= -ac^3(1-t^2),\\
M'''' &= -a(-3sc^2 - 2sc^2 + s^3)c & &= \;\; asc^3(5-t^2).
\end{aligned} \tag{4.12}$$

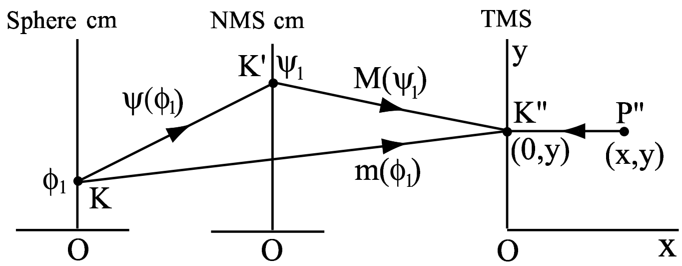


Figure 4.2

The footpoint parameter ψ_1

In Section 3.3, where we discussed the inverse TMS transformations, we introduced the footpoint and the footpoint latitude. In considering the inverse transformations from TMS to NMS it is useful to introduce the **footpoint parameter** ψ_1. All of these parameters are indicated in Figure 4.2 which shows the (x, y) plane of TMS and the central meridians only of NMS and the sphere. Given a point $P''(x, y)$ the footpoint in the TMS plane is $K''(0, y)$ and the footpoint latitude ϕ_1 at K on the sphere is such that $m(\phi_1) = y$. The footpoint parameter in the NMS plane is defined as the point $K'(0, \psi_1)$ such that

$$M(\psi_1) = y = m(\phi_1). \tag{4.13}$$

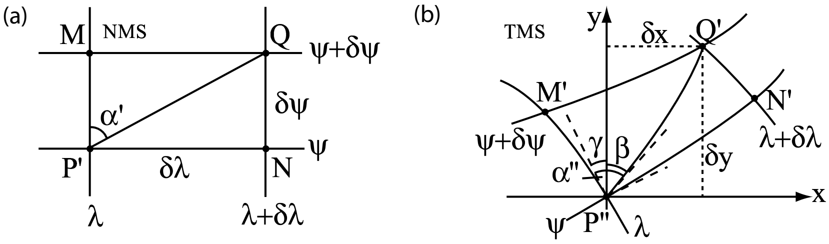


Figure 4.3

The conformality equations and complex functions

It is very straightforward to determine the equations which restrict the functions $x(\lambda,\psi)$ and $y(\lambda,\psi)$ if the transformation $(\lambda,\psi)\to(x,y)$ is to be conformal. Consider infinitesimal elements at P' and P'' as shown in Figure 4.3. Comparing this figure with Figure 3.7 we see that there are only two significant differences. First of all the angle ϕ in Figure 3.7 has been replaced by ψ; secondly the factor of $\cos\phi$ is absent. We can now construct the tangents of all the relevant angles exactly as we did in Section 3.7; imposing conformality by demanding $\alpha'=\alpha''$ we obtain the equations (3.51) with ϕ replaced by ψ and $R=\cos\phi$ replaced by unity.

CAUCHY–RIEMANN $$\boxed{x_\lambda = y_\psi, \quad y_\lambda = -x_\psi} \tag{4.14}$$

Satisfying the Cauchy–Riemann conditions and fitting the scale condition on the y-axis, namely $y(0,\psi)=M(\psi)$, is a non-trivial problem. It becomes much more tractable when we use complex numbers to effect the transformation. The basic idea is to associate with the point $P'(\lambda,\psi)$ of NMS a complex number $\zeta=\lambda+i\psi$. Form a new complex number by constructing a function $z(\zeta)$; the real and imaginary parts of z are then used to define the coordinates of a point $P''(x,y)$ in TMS. Clearly this construction defines two real functions $x(\lambda,\psi)$ and $y(\lambda,\psi)$. The theory of complex numbers tells us that if the function $z(\zeta)$ is differentiable (or analytic, an equivalent term) then x and y *must* satisfy the Cauchy–Riemann conditions and define a conformal transformation.

There is a (very) concise introduction to complex functions in Appendix G. Here we consider a trivial example of a conformal transformation. Consider $z(\zeta)=\zeta+\zeta^2$: this is differentiable, giving $z'(\zeta)=1+2\zeta^2$. Using $i^2=-1$ we have

$$\begin{aligned} z(\zeta) &= \zeta+\zeta^2 = (\lambda+i\psi)+(\lambda+i\psi)^2 = \lambda+i\psi+[\lambda^2+2i\lambda\psi-\psi^2] \\ &= \lambda+\lambda^2-\psi^2+i(\psi+2\lambda\psi). \end{aligned} \tag{4.15}$$

Taking the real and imaginary parts defines the functions $x(\lambda,\psi)=\lambda+\lambda^2-\psi^2$ and $y(\lambda,\psi)=\psi+2\lambda\psi$ which satisfy the Cauchy–Riemann equations $x_\lambda=y_\psi=1+2\lambda$ and $y_\lambda=-x_\psi=2\psi$. On the other hand this example does not satisfy the boundary conditions $x(0,\psi)$=0 and and $y(0,\psi)$=$M(\psi)$ of equations (4.6), (4.8) and (4.9).

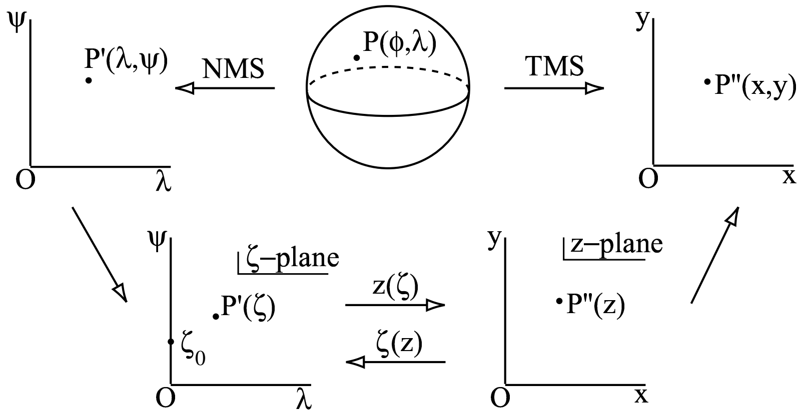


Figure 4.4

4.2 Transformation to the TMS series

The stages of the transformation are summarized in the above figure. We proceed anti-clockwise from the sphere and derive the series solutions for TMS that were presented in Chapter 3, albeit at a very low order which which would be inappropriate for accurate mapping. The same steps will be used when we come to the ellipsoid projections.

- Start at a general point on the sphere with coordinates $P(\phi, \lambda)$ and map to the NMS plane at $P'(\lambda, \psi)$ with ψ the Mercator parameter for the sphere.
- Associate this point with the complex number $\zeta = \lambda + i\psi$ in the complex ζ-plane.
- Use a differentiable function $z(\zeta)$ to construct a conformal map from the complex ζ-plane to the complex z-plane.
- Let x and y be the real and imaginary parts of parts of z so that

$$z(\zeta) = x(\lambda, \psi) + iy(\lambda, \psi). \tag{4.16}$$

- Expand $z(\zeta)$ in a Taylor series about a point ζ_0 on the imaginary axis (λ=0).
- Demand that the central meridians correspond.

$$x(0, \psi) = 0. \tag{4.17}$$

- Demand that the scale be true on the central meridian in the z-plane.

$$y(0, \psi) = M(\psi). \tag{4.18}$$

- The result is a pair of series for x and y agreeing with those derived in Section 3.8.
- Invert the Taylor series to find $\zeta(z)$ and use the real and imaginary parts to find the series for $\lambda(x, y)$ and $\psi(x, y)$.
- Find $\phi(x, y)$ from $\psi(x, y)$.

The direct complex Taylor series

The Taylor series for $z(\zeta)$ about a point $\zeta_0 = i\psi_0$ on the imaginary ($\lambda = 0$) axis of the ζ-plane is

$$z(\zeta) = z_0 + (\zeta - \zeta_0)z'(\zeta_0) + \frac{1}{2!}(\zeta - \zeta_0)^2 z''(\zeta_0) + \frac{1}{3!}(\zeta - \zeta_0)^3 z'''(\zeta_0) + \cdots . \quad (4.19)$$

Equation (4.17) implies that $z_0 = z(\zeta_0)$ is on the imaginary axis of the z-plane so that we must have $z_0 = iy_0$. Then equation (4.18) implies that $y_0 = M(\psi_0)$. Thus the leading term in the expansion may be recast in various forms as and when required:

$$z_0 = z(\zeta_0) = iy_0 = iM(\psi_0) = iM_0 \quad (4.20)$$

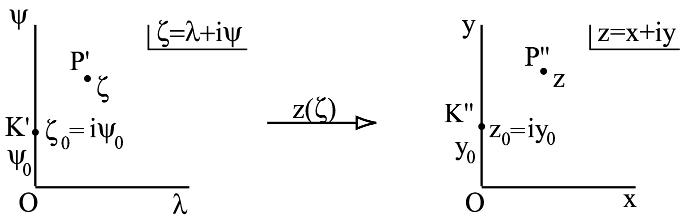


Figure 4.5

It is instructive to consider the derivatives of $z(\zeta)$ from first principles, (as in (G.27):

$$z'(\zeta) = \lim_{\delta\zeta \to 0} \frac{z(\zeta + \delta\zeta) - z(\zeta)}{\delta\zeta}. \quad (4.21)$$

Now because $z(\zeta)$ is analytic we know that this limit is independent of direction and we choose to take it in the ψ direction so that $\delta\lambda = 0$ and $\delta\zeta = i\delta\psi$. Therefore we have

$$z'(\zeta) = \left(\frac{1}{i}\frac{d}{d\psi}\right) z(\zeta). \quad (4.22)$$

Once again, equations (4.17) and (4.18) imply that $z(\zeta)$ reduces to $iM(\psi)$ at an arbitrary point on the imaginary axis. therefore

$$\begin{aligned}
z'(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\big(iM(\psi)\big)\Big|_{\psi_0} &&= \quad M'(\psi_0), \\
z''(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\big(M'(\psi)\big)\Big|_{\psi_0} &&= -iM''(\psi_0), \\
z'''(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\big(-iM''(\psi)\big)\Big|_{\psi_0} &&= -M'''(\psi_0), \\
z''''(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\big(-M'''(\psi)\big)\Big|_{\psi_0} &&= \quad iM''''(\psi_0).
\end{aligned} \quad (4.23)$$

Finally, if we abbreviate $M'(\psi_0)$= M'_0, $M''(\psi_0)$=M''_0 *etc.*, the Taylor series (4.19) may be written as

$$\boxed{z = z_0 + (\zeta - \zeta_0)M'_0 - \frac{i}{2!}(\zeta - \zeta_0)^2 M''_0 - \frac{1}{3!}(\zeta - \zeta_0)^3 M'''_0 + \frac{i}{4!}(\zeta - \zeta_0)^4 M''''_0 + \cdots} \quad (4.24)$$

The direct series for x and y

When we derived the direct series in Section 3.8 we expanded x and y as power series in λ keeping ϕ constant. Now constant ϕ on the sphere corresponds to constant ψ in the ζ-plane. Therefore, if we start from a given point $P'(\lambda, \psi)$ in the ζ plane, that is a given $\zeta = \lambda + i\psi$, we must choose ζ_0 at K' with the *same* ordinate, that is $\zeta_0 = i\psi$: see Figure 4.6. Therefore in the Taylor series (4.24) we must set $\zeta - \zeta_0 = (\lambda + i\psi) - i\psi = \lambda$ so that it becomes a power series in λ. Moreover we must evaluate M and its derivatives with $\psi_0 = \psi$ so that $z_0 = iM_0 \to iM$ and $M_0' \to M'$ *etc.* Thus

$$z = x + iy = iM + \lambda M' - \frac{i}{2!}\lambda^2 M'' - \frac{1}{3!}\lambda^3 M''' + \frac{i}{4!}\lambda^4 M'''' + \cdots \tag{4.25}$$

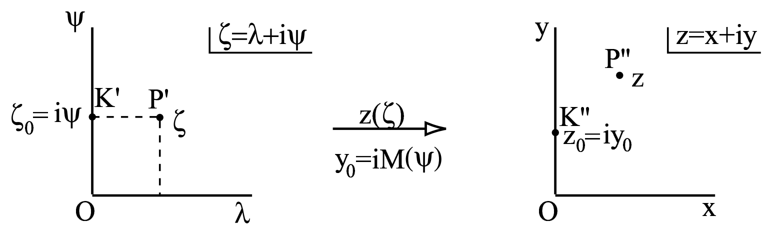


Figure 4.6

The real and imaginary parts of equation (4.25) give x and y as functions of λ and ψ. The derivatives of M are real so that the transformations from NMS→TMS are

$$x(\lambda, \psi) = \lambda M' - \frac{1}{3!}\lambda^3 M''' + \cdots \tag{4.26}$$

$$y(\lambda, \psi) = M - \frac{1}{2!}\lambda^2 M'' + \frac{1}{4!}\lambda^4 M'''' + \cdots . \tag{4.27}$$

On substituting for M and its derivatives using equations (4.12), we obtain the corresponding expressions in terms of λ and ϕ (with $s = \sin\phi$ *etc.*) which define the transformation from sphere to TMS:

$$x(\lambda, \phi) = ac\lambda + \frac{1}{3!}ac^3(1 - t^2)\lambda^3 + \cdots , \tag{4.28}$$

$$y(\lambda, \phi) = a\phi + \frac{1}{2!}asc\,\lambda^2 + \frac{1}{4!}asc^3(5 - t^2)\,\lambda^4 + \cdots . \tag{4.29}$$

These results agree with the expansions obtained in equations (3.57, 3.58).

The Cauchy–Riemann equations

It is instructive to verify that the (plane to plane) Cauchy–Riemann equations (4.14) are indeed satisfied by equations (4.26) and (4.27) (at least if the series are continued to infinity). Evaluating the four partial derivatives we have

$$x_\lambda = \quad y_\psi = M' - \frac{1}{2!}\lambda^2 M''' + \cdots , \tag{4.30}$$

$$x_\psi = -y_\lambda = \lambda M'' - \frac{1}{3!}\lambda^3 M'''' + \cdots . \tag{4.31}$$

Note also that the equations (4.28, 4.29) satisfy the the Cauchy–Riemann equations (3.51) which apply to the transformation from the *sphere* to the TMS plane. Explicitly

$$x_\lambda = \quad \cos\phi\, y_\phi = ac + \frac{1}{2!}ac^3(1-t^2)\lambda^2\cdots, \tag{4.32}$$

$$y_\lambda = -\cos\phi\, x_\phi = asc\,\lambda + \frac{1}{3!}asc^3(5-t^2)\,\lambda^3 + \cdots. \tag{4.33}$$

The inverse complex series: method of Lagrange series inversion

The simplest method of obtaining the inverse series is to use the Lagrange series expansions described in Appendix B; in particular we use the inversion of a fourth order polynomial as described in Section B.4. The beauty of the Lagrange expansions for simple polynomials is that the coefficients can be determined once and for all and applied in various contexts as need arises.

We start by writing the direct Taylor series (4.24) as

$$\frac{z-z_0}{M_0'} = (\zeta-\zeta_0) + \frac{b_2}{2!}(\zeta-\zeta_0)^2 + \frac{b_3}{3!}(\zeta-\zeta_0)^3 + \frac{b_4}{4!}(\zeta-\zeta_0)^4 + \cdots \tag{4.34}$$

where

$$b_2 = -i\frac{M_0''}{M_0'}, \qquad b_3 = -\frac{M_0'''}{M_0'}, \qquad b_4 = i\frac{M_0''''}{M_0'}. \tag{4.35}$$

The series (4.34) and (B.13) are identical if we replace z and ζ in the latter by $(z-z_0)/M_0'$ and $\zeta-\zeta_0$ respectively. Using (B.14) we can immediately find the inverse series of (4.34) as

$$\boxed{\zeta-\zeta_0 = \left(\frac{z-z_0}{M_0'}\right) - \frac{p_2}{2!}\left(\frac{z-z_0}{M_0'}\right)^2 - \frac{p_3}{3!}\left(\frac{z-z_0}{M_0'}\right)^3 - \frac{p_4}{4!}\left(\frac{z-z_0}{M_0'}\right)^4 + \cdots} \tag{4.36}$$

where the p-coefficients follow from (B.12):

$$\begin{aligned} p_2 &= b_2 &&= -\frac{iM_0''}{M_0'} \\ p_3 &= b_3 - 3b_2^2 &&= -\frac{M_0'''}{M_0'} + 3\frac{(M_0'')^2}{(M_0')^2} \\ p_4 &= b_4 - 10b_2b_3 + 15b_2^3 = && \frac{iM_0''''}{M_0'} - 10i\frac{M_0''M_0'''}{(M_0')^2} + 15i\frac{(M_0'')^3}{(M_0')^3}. \end{aligned} \tag{4.37}$$

Using equations (4.12), these become

$$\begin{aligned} p_2 &= \; is_0 \\ p_3 &= \; c_0^2(1+2t_0^2) \\ p_4 &= -is_0c_0^2(5+6t_0^2), \end{aligned} \tag{4.38}$$

with $c_0 = \cos\phi_0$ *etc.* are the usual abbreviations.

The inverse series for ψ and λ

When we derived the inverse series in Section 3.8 we expanded λ and ϕ in power series in x keeping y constant. The corresponding approach for the complex planes is indicated in Figure 4.7. We start with a given point $P''(x, y)$ on the projection corresponding to the point $z = x+iy$ in the complex z-plane. Let $z_0 = iy$ be the point K'' in the z-plane corresponding to the footpoint of P''. Clearly this point projects back to the central meridian of the ζ-plane

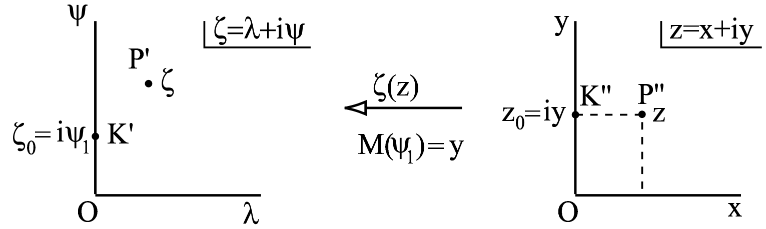


Figure 4.7

at the point $\zeta_0 = i\psi_1$ where $M(\psi_1) = y$, that is ψ_1 is the footpoint parameter defined in Figure 4.2. With these choices we see that we must set $z - z_0 = (x + iy) - iy = x$ in equation (4.36) giving a power series in x as required. M and its derivatives must now be evaluated at ψ_1. Denoting these derivatives by M_1' *etc.* equation (4.36) becomes

$$\lambda + i\psi - i\psi_1 = \frac{x}{M_1'} - \frac{p_2}{2!}\left(\frac{x}{M_1'}\right)^2 - \frac{p_3}{3!}\left(\frac{x}{M_1'}\right)^3 - \frac{p_4}{4!}\left(\frac{x}{M_1'}\right)^4 + \cdots \tag{4.39}$$

where the p coefficients are also evaluated at ψ_1 using equations (4.37) with $M_0' \to M_1'$ *etc*. Taking the real and imaginary parts (noting that p_3 is real whilst p_2 and p_4 are pure imaginary) we find

$$\lambda(x,y) = \frac{x}{M_1'} - \frac{1}{3!}p_3\left(\frac{x}{M_1'}\right)^3 + \cdots \tag{4.40}$$

$$\psi(x,y) = \psi_1 - \frac{1}{2!}\mathrm{Im}\,(p_2)\left(\frac{x}{M_1'}\right)^2 - \frac{1}{4!}\mathrm{Im}\,(p_4)\left(\frac{x}{M_1'}\right)^4 + \cdots . \tag{4.41}$$

Now substitute for the p-coefficients from equations (4.38): these coefficients must now be evaluated at the footpoint latitude $\phi_1 = \phi(\psi_1)$ corresponding to the footpoint parameter ψ_1. Since $M_1' = ac_1$ (equation 4.12) and we find

$$\lambda(x,y) = \frac{1}{c_1}\frac{x}{a} - \frac{1}{3!}\frac{1}{c_1}\left[1 + 2t_1^2\right]\left(\frac{x}{a}\right)^3 + \cdots , \tag{4.42}$$

$$\psi(x,y) = \psi_1 - \frac{1}{2!}\frac{t_1}{c_1}\left(\frac{x}{a}\right)^2 + \frac{1}{4!}\frac{t_1}{c_1}\left[5{+}6t_1^2\right]\left(\frac{x}{a}\right)^4 + \cdots . \tag{4.43}$$

The series for λ is in agreement with equation (3.59) but we must now derive the series for ϕ from that for ψ.

The inverse of the Mercator parameter

The last equation determines $\psi-\psi_1$ as a power series in x with coefficients evaluated at the footpoint latitude ϕ_1. To obtain the corresponding series for ϕ we first construct the Taylor series expansion of $\phi(\psi)$ about the footpoint parameter ψ_1:

$$\phi(\psi) = \phi(\psi_1) + (\psi-\psi_1)\left.\frac{d\phi}{d\psi}\right|_1 + \frac{1}{2!}(\psi-\psi_1)^2\left.\frac{d^2\phi}{d\psi^2}\right|_1 + \cdots\,. \tag{4.44}$$

Once again we could use equation (2.32)a to find the derivatives of $\phi(\psi)$ in terms of the footpoint parameter ψ_1 but we obviously need to express the derivatives in terms of ϕ and evaluate them at ϕ_1. Again we proceed from the definition of $\psi(\phi)$ in Section 2.4:

$$\frac{d\psi}{d\phi} = \sec\phi, \qquad \frac{d\phi}{d\psi} = \left(\frac{d\psi}{d\phi}\right)^{-1} = \cos\phi, \tag{4.45}$$

$$\frac{d^2\phi}{d\psi^2} = \frac{d}{d\psi}(\cos\phi) = -\sin\phi\frac{d\phi}{d\psi} = -\sin\phi\,\cos\phi. \tag{4.46}$$

Substituting these derivatives into the Taylor series, and setting $\phi(\psi_1)=\phi_1$, we have

$$\phi = \phi_1 + (\psi-\psi_1)\cos\phi_1 - \frac{1}{2!}(\psi-\psi_1)^2\sin\phi_1\cos\phi_1 + \cdots\,. \tag{4.47}$$

The inverse series for ϕ

Substituting for $\psi-\psi_1$ from equation (4.43) gives, to order $(x/a)^4$,

$$\begin{aligned}\phi(x,y) = \phi_1 + &\left[-\frac{1}{2!}\left(\frac{x}{a}\right)^2 + \frac{1}{4!}\left(\frac{x}{a}\right)^4\left[5+6t_1^2\right]\right]\frac{t_1}{c_1}c_1\\ &\qquad -\frac{1}{2!}\left[-\frac{1}{2!}\left(\frac{x}{a}\right)^2+\cdots\right]^2\left(\frac{t_1}{c_1}\right)^2 s_1c_1\end{aligned}$$

which simplifies to

$$\phi(x,y) = \phi_1 - \frac{t_1}{2}\left(\frac{x}{a}\right)^2 + \frac{t_1}{24}\left[5+3t_1^2\right]\left(\frac{x}{a}\right)^4 + \cdots\,, \tag{4.48}$$

where $m(\phi_1) = a\phi_1 = y$, in agreement with equation (3.60).

4.3 The inverse complex series: an alternative method

Another way of deriving the inverse series is to take the development given in the first part of Section 4.2 and run it backwards from the z-plane to the ζ-plane. That is we assume the existence of an analytic function $\zeta(z)$ such that (a) the central meridian maps from $x=0$ to $\lambda=0$ and (b) on the ψ-axis the scale is true. Therefore

$$\zeta(z) = \lambda(x,y) + i\psi(x,y), \tag{4.49}$$

$$\lambda(0,y) = 0, \tag{4.50}$$

$$\psi(0,y) = \overline{M}(y), \tag{4.51}$$

where $\overline{M}(y)$ is an inverse to $M(\psi)$ in the sense that $M\big(\overline{M}(y)\big) = y$ and $\overline{M}\big(M(\psi)\big) = \psi$. The Taylor series analogous to (4.24) is then an expansion of $\zeta(z)$ about a point on the $z_0 = iy_0$ on the y-axis of the z-plane:

$$\zeta(z) = \zeta_0 + (z-z_0)\overline{M}'_0 - \frac{i}{2!}(z-z_0)^2\overline{M}''_0 - \frac{1}{3!}(z-z_0)^3\overline{M}'''_0 + \frac{i}{4!}(z-z_0)^4\overline{M}''''_0 + \cdots, \tag{4.52}$$

where $\zeta_0 = i\psi_0 = i\overline{M}(y_0)$ and the derivatives of $\overline{M}$ are with respect to y at y_0.

Now although it is straightforward to construct the function $\overline{M}(y)$ on the sphere we shall construct its derivatives from those of $M(\psi)$. We start by differentiating the identities

$$y = M(\overline{M}(y)) = M(\psi), \tag{4.53}$$

$$\psi = \overline{M}(M(\psi)) = \overline{M}(y), \tag{4.54}$$

to give

$$\frac{dy}{d\psi} = M'(\psi), \qquad \frac{d\psi}{dy} = \overline{M}'(y). \tag{4.55}$$

Therefore, as long as $M'(\psi) \neq 0$, we have

$$\overline{M}'(y) = \frac{d\psi}{dy} = \frac{1}{M'(\psi)}, \tag{4.56}$$

and in general

$$\frac{d(\;)}{dy} = \frac{1}{M'(\psi)}\frac{d(\;)}{d\psi}. \tag{4.57}$$

It is now straightforward to calculate all the derivatives in equation (4.52). For compactness we suppress the argument ψ in $M(\psi)$ and all its derivatives, $M'(\psi)$, $M''(\psi)$ *etc.* Comparing the results with the p-coefficients in equation (4.37) we find

$$\begin{aligned}
\overline{M}'(y) &= \frac{1}{M'} &&= \frac{1}{M'},\\
\overline{M}''(y) &= \frac{1}{M'}\frac{d}{d\psi}\Big[\overline{M}'(y)\Big] = \frac{1}{M'}\frac{d}{d\psi}\left[\frac{1}{M'}\right] = -\frac{M''}{(M')^3} &&= \frac{-ip_2}{M'^2},\\
\overline{M}'''(y) &= \frac{1}{M'}\frac{d}{d\psi}\Big[\overline{M}''(y)\Big] = \frac{1}{M'}\frac{d}{d\psi}\left[\frac{-M''}{(M')^3}\right] = -\frac{M'''}{(M')^4} + 3\frac{(M'')^2}{(M')^5} &&= \frac{p_3}{M'^3},\\
\overline{M}''''(y) &= \frac{1}{M'}\frac{d}{d\psi}\Big[\overline{M}'''(y)\Big] = -\frac{M''''}{(M')^5} + 10\frac{M''M'''}{(M')^6} - 15\frac{(M'')^3}{(M')^7} &&= \frac{ip_4}{M'^4}.
\end{aligned} \tag{4.58}$$

Substituting these derivatives (evaluated at y_0 for $\overline{M}$ and at ϕ_0 for M) into equation (4.52) clearly gives a complex series identical to that of (4.36) and the same results follow. We choose not to follow this method since the calculation of the derivatives to eighth order for the ellipsoid becomes very intricate.

4.4 The closed formula for the transformation

Finally, we present a closed analytic expression whose real and imaginary parts give the TMS transformations of equation (4.2). This section does not readily generalise to the transformations on the ellipsoid: it is included as an interesting digression.

Finding a conformal transformation which satisfies given conditions is not always simple: there is no general technique and we have to rely mainly on our experience—and the fact that there are books which give exhaustive lists of the conformal transformations which have been studied in the last two hundred years. The required transformation is

$$\boxed{z = \frac{i\pi a}{2} - 2ia \operatorname{arccot}\left[\exp(-i\zeta)\right]} \tag{4.59}$$

We now verify that this transformation has the required properties. Substituting $z = x + iy$ and $\zeta = \lambda + i\psi$ in the above gives

$$\cot\left(\frac{x+iy}{-2ia} + \frac{\pi}{4}\right) = \cot\left(\frac{\pi}{4} - \frac{y}{2a} + \frac{ix}{2a}\right) = \exp(-i\lambda + \psi). \tag{4.60}$$

The real and imaginary parts of the cotangent function are given in Appendix G, equation (G.17). We substitute $A = \frac{\pi}{4} - \frac{y}{2a}$ and $B = \frac{x}{2a}$ in that identity but to clarify the algebra we temporarily replace x/a and y/a by x and y respectively. The result is

$$\frac{\cos y - i \sinh x}{\cosh x - \sin y} = \mathrm{e}^{\psi}(\cos\lambda - i\sin\lambda). \tag{4.61}$$

Taking the real and imaginary parts gives

$$\frac{\cos y}{\cosh x - \sin y} = \mathrm{e}^{\psi}\cos\lambda \equiv p \tag{4.62}$$

$$\frac{\sinh x}{\cosh x - \sin y} = \mathrm{e}^{\psi}\sin\lambda \equiv q. \tag{4.63}$$

Re-arranging these equations and also taking their quotient gives

$$\cos y = p\cosh x - p\sin y, \tag{4.64}$$

$$\sinh x = q\cosh x - q\sin y, \tag{4.65}$$

$$\cos y = \frac{p}{q}\sinh x. \tag{4.66}$$

Eliminate y from (4.65) and (4.66) using $\cos^2 y + \sin^2 y = 1$; eliminate x from (4.64) and (4.66) using $\cosh^2 x - \sinh^2 x = 1$. On simplification we find

$$\tanh x = \frac{2q}{p^2 + q^2 + 1} = \frac{2\mathrm{e}^{\psi}\sin\lambda}{\mathrm{e}^{2\psi} + 1} = \sin\lambda\,\operatorname{sech}\psi, \tag{4.67}$$

$$\tan y = \frac{p^2 + q^2 - 1}{2p} = \frac{\mathrm{e}^{2\psi} - 1}{2\mathrm{e}^{\psi}\cos\lambda} = \sec\lambda\sinh\psi. \tag{4.68}$$

This is the final result for the real and imaginary parts of the transformation from the complex ζ-plane to the complex z-plane.

The final step is to transform from ψ to ϕ in (4.67, 4.68) . To do this we use the inverse expressions from equation (2.32), that is

$$\sinh\psi = \tan\phi, \qquad \cosh\psi = \sec\phi, \tag{4.69}$$

and equations (4.67, 4.68) become

$$\tanh x = \sin\lambda\cos\phi, \tag{4.70}$$
$$\tan y = \sec\lambda\ \tan\phi. \tag{4.71}$$

The second equation agrees directly with the y transformation equation (4.2) after restoring $y \to y/a$. Equation (4.3) for the x transformation follows since

$$\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi} = \frac{1+\tanh x}{1-\tanh x} = \frac{\cosh x+\sinh x}{\cosh x-\sinh x} = \frac{2\mathrm{e}^{x}}{2\mathrm{e}^{-x}} = \mathrm{e}^{2x} \to \mathrm{e}^{2x/a}. \tag{4.72}$$

The Cauchy–Riemann conditions

To check the Cauchy–Riemann equations (4.14) we evaluate the partial derivatives of x and y from equations (4.67) and (4.68):

$$(\mathrm{sech}^2 x)\, x_\lambda = \quad \cos\lambda\,\mathrm{sech}\,\psi, \qquad (\sec^2 y)\, y_\lambda = \sin\lambda\sec^2\lambda\sinh\psi, \tag{4.73}$$
$$(\mathrm{sech}^2 x)\, x_\psi = -\sin\lambda\sinh\psi\,\mathrm{sech}^2\psi, \qquad (\sec^2 y)\, y_\psi = \sec\lambda\cosh\psi \tag{4.74}$$

and simplify using

$$\mathrm{sech}^2 x = 1 - \sin^2\lambda\,\mathrm{sech}^2\psi = \mathrm{sech}^2\psi\big[\cosh^2\psi - \sin^2\lambda\big], \tag{4.75}$$
$$\sec^2 y = 1 + \sec^2\lambda\sinh^2\psi = \sec^2\lambda\big[\cosh^2\psi - \sin^2\lambda\big]. \tag{4.76}$$

Chapter 5

The geometry of the ellipsoid

Abstract

Cartesian coordinates. Geodetic and geocentric latitude. Parameters of the ellipsoid. Parameterisation in terms of geodetic latitude. Relation of Cartesian and geographical coordinates. Reduced latitude. Curvature. The metric. Meridian distance and its inverse. Rectifying latitude.

5.1 Coordinates on the ellipsoid

We now model the Earth as an ellipsoid of revolution for which the Cartesian coordinates with respect to its centre satisfy

$$\frac{X^2}{a^2} + \frac{Y^2}{a^2} + \frac{Z^2}{b^2} = 1. \tag{5.1}$$

The definition of longitude λ is exactly the same as on the sphere. The **geodetic latitude** ϕ, which we will simply call the latitude, is the angle at which the normal at P intersects the equatorial plane ($Z = 0$). The crucial new feature is that the normal does not pass through

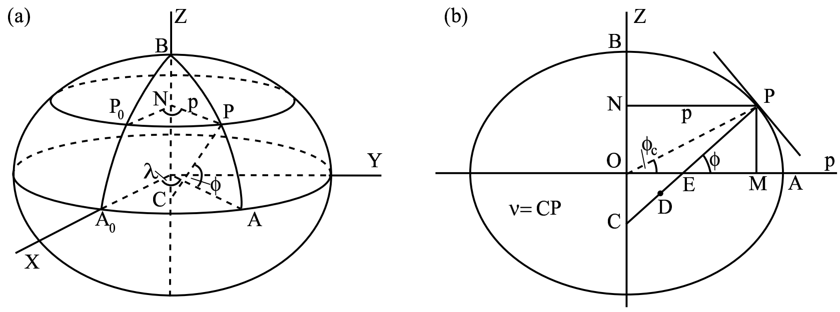


Figure 5.1

the centre of the ellipsoid (except when P is on the equator and at the poles). The line joining P to the centre defines the **geocentric latitude** ϕ_c. We introduce the notation $p(\phi)$ for the distance PN of a point P from the central axis and we also set $\nu(\phi)$ for the length CP of the normal at P to its intersection with the axis. Therefore

$$p(\phi) = \nu(\phi)\cos\phi = \sqrt{X^2 + Y^2}. \tag{5.2}$$

5.2 The parameters of the ellipsoid

Instead of using (a, b) as the basic parameters of the ellipse we can use either (a, e) where e is the **eccentricity**, or (a, f) where f is the **flattening**. These parameters are defined and related by

$$b^2 = a^2(1 - e^2), \qquad f = \frac{a - b}{a}, \qquad e^2 = 2f - f^2. \tag{5.3}$$

For numerical examples we use the values for the Airy (1830) ellipsoid which is used for the OSGB maps:

$$\begin{aligned} a &= 6377563.396\text{m}, & e &= 0.0816733724, & f &= 0.0033408505, \\ b &= 6356256.910\text{m}, & e^2 &= 0.00667053982, & \frac{1}{f} &= 299.3249753. \end{aligned} \tag{5.4}$$

The flattening of the Earth is small. For example, in the figures on the previous page the difference between a sphere of radius a and the ellipsoid should be about the width of one of the lines in the figure. Thus the ellipses shown here, and elsewhere, are greatly exaggerated.

Other parameters used to describe an ellipse

There are several other small parameters which arise naturally in the study of the properties of the ellipse. Two which we shall need are: (a) the second eccentricity, e', and (b) the parameter n (sometimes e_1). They are defined by

$$e'^2 = \frac{a^2 - b^2}{b^2} = \frac{e^2}{1 - e^2}, \qquad n = e_1 = \frac{a - b}{a + b}. \tag{5.5}$$

There are many possible relations between all these parameters. For example we will need the following results:

$$a = b\left(1 - e^2\right)^{-1/2} = b\left(\frac{1 + n}{1 - n}\right) \tag{5.6}$$

$$= b\left(1 + 2n + 2n^2 + 2n^3 + \cdots\right), \tag{5.7}$$

$$e^2 = 1 - \left(\frac{b}{a}\right)^2 = 1 - \left(\frac{1 - n}{1 + n}\right)^2 = \frac{4n}{(1 + n)^2} \tag{5.8}$$

$$= 4n(1 - 2n + 3n^2 - 4n^3 + \cdots). \tag{5.9}$$

5.3 Parameterisation by geodetic latitude

The equation of the cross-section ellipse follows from (5.1) and (5.2):

$$\frac{p^2}{a^2} + \frac{Z^2}{b^2} = 1. \tag{5.10}$$

Differentiating this equation with respect to p gives

$$\frac{dZ}{dp} = -\frac{pb^2}{Za^2}. \tag{5.11}$$

Since the normal and tangent are perpendicular the product of their gradients is -1 and therefore the gradient of the normal is

$$\tan\phi = -\left(\frac{dZ}{dp}\right)^{-1} = \frac{Za^2}{pb^2} = \frac{Z}{p(1-e^2)}. \tag{5.12}$$

Substituting $Z = p(1-e^2)\tan\phi$ in (5.10) gives

$$p^2[1 + (1-e^2)\tan^2\phi] = a^2. \tag{5.13}$$

Thus the required parameterisation is;

$$PN = p(\phi) = \frac{a\cos\phi}{[1-e^2\sin^2\phi]^{1/2}}, \tag{5.14}$$

$$PM = Z(\phi) = \frac{a(1-e^2)\sin\phi}{[1-e^2\sin^2\phi]^{1/2}}. \tag{5.15}$$

Since $CP = PN\sec\phi = p\sec\phi$ we have

$$\boxed{CP = \nu(\phi) = \frac{a}{[1-e^2\sin^2\phi]^{1/2}}} \tag{5.16}$$

in terms of which

$$p(\phi) = \nu(\phi)\ \cos\phi, \tag{5.17}$$

$$Z(\phi) = (1-e^2)\,\nu(\phi)\ \sin\phi. \tag{5.18}$$

The triangle OCE

Later we shall require the sides of the triangle ΔOCE defined by the normal and its intercepts on the axes.

$$\begin{aligned} OE &= OM - EM = p - Z\cot\phi \\ &= \nu\cos\phi - (1-e^2)\nu\cos\phi \\ &= \nu e^2\cos\phi. \\ CE &= \nu e^2 \\ OC &= \nu e^2\sin\phi. \end{aligned} \tag{5.19}$$

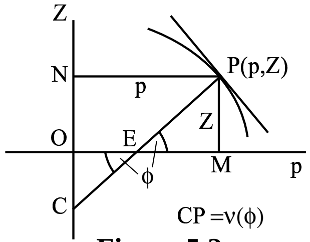


Figure 5.2

The sides of this small triangle are all of order ae^2; for example at latitude $\pm 45°$ the sides OE and OC are about 30km and CE is about 42.5km.

The relation between ϕ and ϕ_c

From Figure 5.1b and equations (5.17) and (5.18) we immediately obtain the relation between ϕ and ϕ_c:

$$\tan\phi_c = \frac{Z}{p} = (1-e^2)\tan\phi. \tag{5.20}$$

Clearly ϕ and ϕ_c are equal only at the equator, $\phi = 0$, or at the poles, $\phi = \pi/2$. Since $e^2 \approx 0.0067$ the difference $\phi - \phi_c$ at any other angle is small (and positive). It is a simple exercise in calculus to find the position and magnitude of the maximum difference. First write

$$\tan(\phi-\phi_c) = \frac{\tan\phi - \tan\phi_c}{1+\tan\phi\tan\phi_c} = \frac{e^2\tan\phi}{1+(1-e^2)\tan^2\phi}. \tag{5.21}$$

Differentiating with respect to ϕ gives

$$\sec^2(\phi-\phi_c)\frac{d(\phi-\phi_c)}{d\phi} = \frac{e^2\sec^2\phi\big[1-(1-e^2)\tan^2\phi\big]}{\big[1+(1-e^2)\tan^2\phi\big]^2}\,. \tag{5.22}$$

Therefore $\phi-\phi_c$ has a turning point, clearly a maximum, when the right hand side vanishes at $\tan\phi = 1/\sqrt{1-e^2}$. Using the value of e for the Airy ellipsoid (equation 5.4) shows that the maximum difference occurs at $\phi \approx 45°.095$, for which $\phi_c \approx 44°.904$ and the latitude difference $\phi-\phi_c \approx 11.5'$. (Note that $e^2 \approx 0.00667$ is the radian measure of 22.9′).

A comment on other latitudes

In addition to the geodetic latitude ϕ and geocentric latitude ϕ_c, we have already discussed the isometric latitude ψ (Section 2.4) and we shall meet three further latitude definitions: the reduced (or parametric) latitude U (Section 5.5), the rectifying latitude μ (Section 5.9) and the conformal latitude χ (Section 6.5). With the exception of the isometric latitude all of these latitudes coincide with the geodetic and geocentric latitudes at the poles and on the equator and the maximum deviations from ϕ are no more than a few minutes of arc. The isometric latitude agrees with the others at the equator only but diverges to infinity at the poles: it is a radically different in character.

5.4 Cartesian and geographic coordinates

Using (5.17) and (5.18) the Cartesian coordinates of a point on the surface are

$$X(\phi) = p(\phi)\cos\lambda = \nu(\phi)\cos\phi\cos\lambda, \tag{5.23}$$

$$Y(\phi) = p(\phi)\sin\lambda = \nu(\phi)\cos\phi\sin\lambda, \tag{5.24}$$

$$Z(\phi) = (1-e^2)\nu(\phi)\sin\phi. \tag{5.25}$$

For given X, Y, Z the inverse relations for ϕ and λ are clearly

$$\text{(a)}\quad \lambda = \arctan\left(\frac{Y}{X}\right), \qquad \text{(b)}\quad \phi = \arctan\left(\frac{Z}{(1-e^2)\sqrt{X^2+Y^2}}\right). \tag{5.26}$$

Now consider a point H at a height h on the normal to the surface at the point P with geographical coordinates ϕ and λ. The distance of this point from the axis is now $p + h\cos\phi$. Also, from (5.20), we have $EP = CP - CE = \nu(1 - e^2)$. The coordinates of H are

$$X(\phi) = \big(\nu(\phi) + h\big)\cos\phi\cos\lambda, \tag{5.27}$$

$$Y(\phi) = \big(\nu(\phi) + h\big)\cos\phi\sin\lambda, \tag{5.28}$$

$$Z(\phi) = \big((1 - e^2)\,\nu(\phi) + h\big)\sin\phi. \tag{5.29}$$

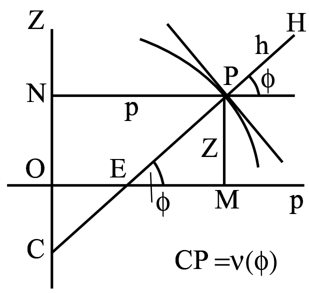


Figure 5.3

For the inverse relations dividing equation (5.28) by (5.27) gives λ explicitly, as in equation (5.26a). To find ϕ and h we can eliminate λ from (5.27) and (5.28) and rewrite equation (5.29) for Z to give

$$\sqrt{X^2 + Y^2} = \big(\nu(\phi) + h\big)\cos\phi, \tag{5.30}$$

$$Z + e^2\nu(\phi)\sin\phi = \big(\nu(\phi) + h\big)\sin\phi. \tag{5.31}$$

Dividing these equations gives an implicit equation for ϕ:

$$\phi = \arctan\left[\frac{Z + e^2\nu(\phi)\sin\phi}{\sqrt{X^2 + Y^2}}\right]. \tag{5.32}$$

There is no closed solution to this equation but we can develop a numerical solution by considering the following **fixed point iteration**:

$$\phi_{n+1} = g(\phi_n) = \arctan\left[\frac{Z + e^2\nu(\phi_n)\sin\phi_n}{\sqrt{X^2 + Y^2}}\right], \qquad n = 0,\, 1,\, 2\ldots. \tag{5.33}$$

Now in most applications we will have $h \ll a$ so that a suitable starting approximation is the value of ϕ obtained by using the $h = 0$ solution, equation (5.26b):

$$\phi_0 = \arctan\left[\frac{Z}{(1 - e^2)\sqrt{X^2 + Y^2}}\right]. \tag{5.34}$$

If the iteration scheme converges so that $\phi_{n+1} \to \phi^*$ and $\phi_n \to \phi^*$ in (5.33) then ϕ^* must be the required solution of equation (5.32). The condition for convergence of this fixed point iteration is that $|g'(\phi)| < 1$: this is true here since $g'(\phi) = O(e^2)$. Once we have found ϕ it is trivial to deduce h from equation (5.30):

$$h = \sec\phi\sqrt{X^2 + Y^2} - \nu(\phi). \tag{5.35}$$

Comment: The formulae of this section are presented without proof in the OSGB web publication mentioned in the Bibliography. We do not use them elsewhere but the method of iteration to a fixed point is important and we will apply it in several places.

5.5 Parameterisation by reduced latitude

There is another important and obvious parameterisation of the ellipse. Construct the **auxiliary circle** of the ellipse: it is concentric and touches the ellipse at the ends of its major axis so that the radius is equal to a. Take a point P on the ellipse and project its ordinate until it meets the auxiliary circle at P' and let angle $P'OA$ be U. The angle U is called the **reduced latitude** (or parametric latitude) of the point P on the ellipse.

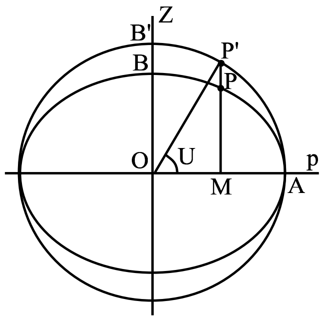


Figure 5.4

The points P and P' clearly have the same abscissa, $p = a\cos U$. Substituting this abscissa into the equation of the ellipse (5.10) we have

$$Z = b\sqrt{1 - p^2/a^2} = b\sin U. \tag{5.36}$$

The pair of equations

$$p = a\cos U, \qquad Z = b\sin U, \tag{5.37}$$

constitutes the required parametric representation of the ellipse. It is clear that that the ellipse is related to the auxiliary circle by scaling in the Z direction by a factor of b/a.

Relations between the reduced and geodetic latitudes

Comparing the parameterisations of p and Z in equations (5.17, 5.18) and (5.37) gives

$$\begin{aligned} p &= \nu(\phi)\ \cos\phi &&= a\cos U, \\ Z &= (1-e^2)\nu(\phi)\ \sin\phi &&= b\sin U. \end{aligned}$$

The basic relation between U and ϕ could be taken as

$$a\cos U = \nu(\phi)\ \cos\phi, \tag{5.38}$$

but it is more useful to divide the expressions for Z and p to find (using $b = a\sqrt{1-e^2}$)

$$\boxed{\tan U = \sqrt{1-e^2}\,\tan\phi} \tag{5.39}$$

It will also be useful to have an expression for ν in terms of U. Using (5.38) and (5.16)

$$\begin{aligned} 1 - e^2\cos^2 U &= 1 - \frac{e^2\nu^2}{a^2}\cos^2\phi = 1 - \frac{e^2\cos^2\phi}{1 - e^2\sin^2\phi} \\ &= \frac{1-e^2}{1-e^2\sin^2\phi}, \end{aligned} \tag{5.40}$$

so that

$$\boxed{\nu = \frac{a}{\left[1 - e^2\sin^2\phi\right]^{1/2}} = \frac{a}{\sqrt{1-e^2}}\left[1 - e^2\cos^2 U\right]^{1/2}} \tag{5.41}$$

We shall need the derivative of U with respect to ϕ. Differentiating (5.39) gives

$$\sec^2 U \frac{dU}{d\phi} = \sqrt{1-e^2}\sec^2\phi = \sqrt{1-e^2}\left[1 + \frac{1}{1-e^2}\tan^2 U\right]. \tag{5.42}$$

$$\frac{dU}{d\phi} = \frac{1-e^2\cos^2 U}{\sqrt{1-e^2}}. \tag{5.43}$$

The difference between reduced and geodetic latitudes

We could use the above derivatives to find the maximum difference between U and ϕ but the result follows by simply comparing equations (5.20) and (5.39). They differ only in that the factor of $(1-e^2)$ in (5.20) is replaced by $\sqrt{1-e^2}$. Since the maximum value of $\phi - \phi_c$ occurred when $\tan\phi = 1/\sqrt{1-e^2}$ we deduce that the maximum value of $\phi - U$ will occur when $\tan\phi = 1/\sqrt[4]{1-e^2}$. This corresponds to $\phi \approx 45^\circ.048$ for which the corresponding value of U is $44^\circ.952$ so that the maximum difference is $\phi - U \approx 5'.7$.

5.6 The curvature of the ellipsoid

We now investigate the properties of the two dimensional curves formed by the intersection of some, but not all, planes with the surface of the ellipsoid: we use the mathematical results established in Appendix A. In particular we investigate two special families of planes. The first family (S) has the normal at P as a common axis and the intersections of its planes with the surface are called the **normal sections** at P. One member of the family is

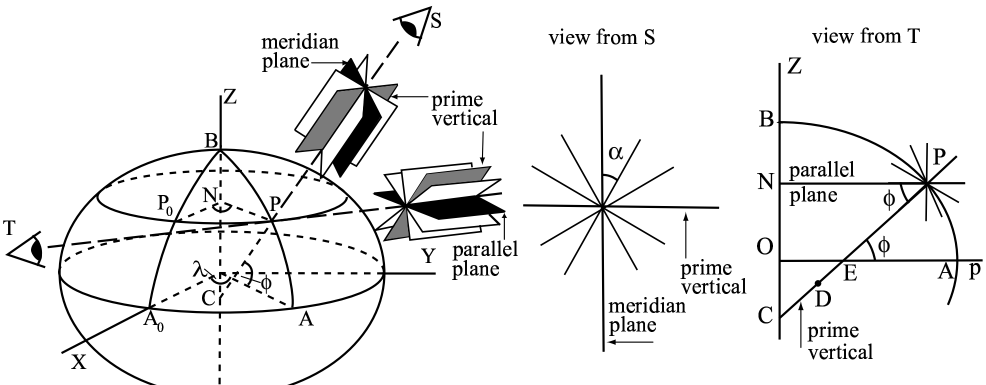


Figure 5.5

the meridian plane (black) containing P and the symmetry axis of the ellipsoid. Another important member of the family is the plane at right angles to the meridian plane: it is called the **prime vertical plane** (shaded grey). Other members of the family are labelled by the angle α between a specific plane and the meridian plane.

The second family of planes (T) has as its axis the tangent to the parallel circle at P: we are interested in just two of its planes. One (black) is the plane of the parallel: its section on the surface is the parallel circle. The other is that which contains the normal at P: this is the prime vertical plane (grey), the only plane common to both families.

Radius of curvature in the meridian plane

The section by the meridian plane is an ellipse whose curvature may be determined from either Cartesian equations or parameterised equations by the well known formulae summarised in Appendix A. The easiest method is to use the parameterisation in terms of the reduced latitude given in (5.36). This has been done as an example in Appendix A: equation (A.12) gives the **meridian curvature** as

$$\kappa = \frac{1}{a}\frac{\sqrt{1-e^2}}{[1-e^2\cos^2 U]^{3/2}}. \tag{5.44}$$

It will be more useful to work with the **meridian radius of curvature** defined by $\rho = 1/\kappa$ and expressed as a function of ϕ. Using equation (5.41) we have

$$\boxed{\rho(\phi) = \frac{a(1-e^2)}{\left(1-e^2\sin^2\phi\right)^{3/2}}} \tag{5.45}$$

Using (5.16) we have the following relation between ρ and ν:

$$\rho = \frac{\nu^3}{a^2}\left(1-e^2\right). \tag{5.46}$$

Furthermore, we have

$$\frac{\rho}{\nu} = \frac{1-e^2}{1-e^2\sin^2\phi}. \tag{5.47}$$

Since (a) the denominator is less than or equal to 1 and (b) the numerator is less than or equal to the denominator, we have

$$(1-e^2)\nu \le \rho \le \nu. \tag{5.48}$$

Now in Figure 5.2 we have $CP = \nu$ and $EP = CP - CE = (1-e^2)\nu$. Therefore the centre of curvature of the meridian is at a point D between C and E, as shown in Figure 5.5.

Radius of curvature in the prime vertical plane

To find the radius of curvature in the prime vertical we consider two planes of the family T: the prime vertical itself (grey) and the parallel plane (black). The radii of curvature in these two planes are related by Meusnier's theorem (Appendix A). This theorem relates the radius of curvature in a normal section to that made by a plane at an oblique angle ϕ:

$$R_{\text{normal}} = \sec\phi\, R_{\text{oblique}} \tag{5.49}$$

We identify the prime vertical plane and parallel plane of the family T with the normal and oblique planes of the theorem. Now the parallel plane intersects the surface in a parallel circle so we know that its radius of curvature is simply $NP = p(\phi)$ in Figure 5.2. But this is just $\nu(\phi)\cos\phi$ and therefore

$$R_{\text{prime vertical}} = \sec\phi\, R_{\text{parallel}} = \sec\phi\; p(\phi) = \nu(\phi). \tag{5.50}$$

Thus we have the important result that the distance $CP = \nu(\phi)$ may be identified as the radius of curvature of the normal section made by the prime vertical plane. The point C where the normal meets the axis is the centre of curvature of this section.

Radius of curvature along a general azimuth

Returning to S, the family of planes on the normal, we now know the curvature of two of the normal sections: ρ^{-1} on the meridian plane and ν^{-1} on the prime vertical. Now consider the curvature, $K(\alpha)$, of the section made by that plane of the family at an angle α, measured clockwise from the meridian plane. Clearly the symmetry of the ellipsoid about any meridian plane implies that $K(-\alpha) = K(\alpha)$ so that $K(\alpha)$ is a symmetric function of α and it must therefore have a turning point at $\alpha = 0$. Therefore the curvature of the meridian section must be either a minimum or maximum and it is therefore one of the principal curvatures at P—see Appendix A.

In the appendix we proved that the planes containing the principal curvatures are orthogonal. Therefore the curvature of the normal section made by the prime vertical plane must be the other principal curvature. Furthermore, equation (5.47) gives $\rho \leq \nu$ and therefore $\rho^{-1} \geq \nu^{-1}$ so that the curvature in the meridian section is the maximum normal section curvature at any point. Introducing the radius of curvature on the general section by $R(\alpha) = 1/K(\alpha)$, we use Euler's formula, equation (A.36), to deduce that

$$\frac{1}{R(\alpha)} = \frac{1}{\rho}\cos^2\alpha + \frac{1}{\nu}\sin^2\alpha. \tag{5.51}$$

Curvatures and their derivatives. The parameter β

In addition to the principal curvatures, ρ and ϕ it is useful to introduce a special notation for their quotient, $\beta = \nu/\rho$:

$$\nu(\phi) = \frac{a}{\left(1 - e^2\sin^2\phi\right)^{1/2}}, \qquad \rho(\phi) = \frac{\nu^3}{a^2}\left(1 - e^2\right), \tag{5.52}$$

$$\beta(\phi) = \frac{\nu}{\rho} = \frac{1 - e^2\sin^2\phi}{1 - e^2}. \qquad \beta - 1 = \frac{e^2\cos^2\phi}{1 - e^2}. \tag{5.53}$$

We shall frequently require the derivatives of the curvatures and their quotient. It is straightforward to show that

$$\frac{d\nu}{d\phi} = (\beta - 1)\rho\tan\phi, \qquad \frac{d\rho}{d\phi} = 3\frac{(\beta - 1)}{\beta}\rho\tan\phi, \qquad \frac{d\beta}{d\phi} = -2(\beta - 1)\tan\phi. \tag{5.54}$$

We need both first and second derivative of ν in the combinations

$$\frac{1}{\nu}\frac{d\nu}{d\phi} = \frac{(\beta - 1)\tan\phi}{\beta}, \qquad \frac{1}{\nu}\frac{d^2\nu}{d\phi^2} = \frac{(\beta - 1)}{\beta} + \frac{1}{\beta^2}\left(2\beta^2 - 5\beta + 3\right)\tan^2\phi. \tag{5.55}$$

Finally we note that the cross-section coordinates (5.17, 5.18) and their derivatives are

$$p(\phi) = \nu(\phi)\ \cos\phi, \qquad Z(\phi) = (1-e^2)\,\nu(\phi)\ \sin\phi, \tag{5.56}$$

$$\frac{dp}{d\phi} = -\rho\sin\phi, \qquad \frac{dZ}{d\phi} = \rho\cos\phi. \tag{5.57}$$

Spherical limit

We shall refer to the limit $e \to 0$ as the spherical limit. Clearly in this limit

$$\nu \to a, \qquad \rho \to a, \qquad \beta \to 1, \qquad \nu',\ \rho',\ \beta'\ \to 0. \tag{5.58}$$

5.7 Distances on the ellipsoid

Derivation of the metric

Starting from the parameterisation of the Cartesian coordinates (Section 5.4):

$$\begin{aligned} X(\phi) &= p(\phi)\cos\lambda = \nu(\phi)\cos\phi\cos\lambda, \\ Y(\phi) &= p(\phi)\sin\lambda = \nu(\phi)\cos\phi\sin\lambda, \\ Z(\phi) &= (1-e^2)\nu(\phi)\sin\phi. \end{aligned} \tag{5.59}$$

we have

$$\begin{aligned} dX &= \dot{p}\cos\lambda\,d\phi - p\sin\lambda\,d\lambda, \quad \text{where DOT} \equiv \frac{d}{d\phi} \\ dY &= \dot{p}\sin\lambda\,d\phi + p\cos\lambda\,d\lambda, \\ dZ &= \dot{Z}\,d\phi. \end{aligned} \tag{5.60}$$

The metric may be written as

$$\begin{aligned} ds^2 &= dX^2 + dY^2 + dZ^2, \\ &= \big(\dot{p}^2 + \dot{Z}^2\big)d\phi^2 + p^2 d\lambda^2. \end{aligned}$$

Using (5.57) and (5.56) we obtain two useful forms:

$$ds^2 = \rho^2\,d\phi^2 + p^2 d\lambda^2, \tag{5.61}$$

$$ds^2 = \rho^2\,d\phi^2 + \nu^2\cos^2\!\phi\,d\lambda^2. \tag{5.62}$$

On the meridian we have $d\lambda = 0$ and on the parallel circle we have $d\phi = 0$ Therefore

$$ds_{\text{meridian}} = \rho\,d\phi, \tag{5.63}$$

$$ds_{\text{parallel}} = \nu\cos\phi\,d\lambda. \tag{5.64}$$

The infinitesimal element on the ellipsoid

The infinitesimal element on the sphere was discussed in Section 2.1 and the same considerations apply on the ellipse. In particular the infinitesimal element may be approximated by a planar rectangular quadrilateral to which we can apply plane trigonometry. Equations (5.63) and (5.64) show that the sides are equal to $\rho\,\delta\phi$ on the meridians and $\nu\cos\phi$

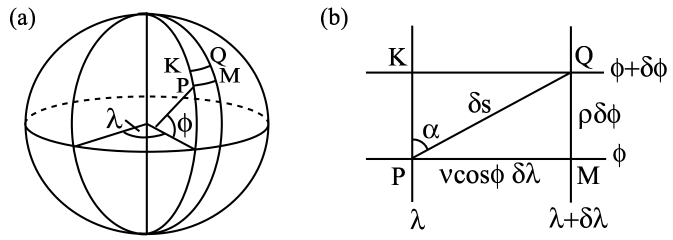


Figure 5.6

on a parallel. The metric (5.62) may be viewed as the application of Pythagoras to the infinitesimal element. The azimuth of an arbitrary displacement is calculated from

$$\tan\alpha = \frac{\nu\cos\phi}{\rho}\frac{d\lambda}{d\phi}. \tag{5.65}$$

Finite distances on the ellipsoid

In general the integration of the metric to find the distance between arbitrary points is non-trivial. The whole of Chapter 11 is devoted to this topic, concluding with the Vincenty series for the geodesic distance. Here we consider only the trivial case of integration along a parallel and the non-trivial case of integration on the meridian where we obtain an elliptic integral. From equations (5.63) and (5.64) we have

$$s_{\text{parallel}} = \int_{\lambda_1}^{\lambda_2} ds_{\text{parallel}} = \int_{\lambda_1}^{\lambda_2} \nu(\phi)\cos\phi\,d\lambda = \nu(\phi)\cos\phi\,\big(\lambda_2-\lambda_1\big), \tag{5.66}$$

$$s_{\text{meridian}} = \int_{\phi_1}^{\phi_2} ds_{\text{meridian}} = \int_{\phi_1}^{\phi_2} \rho(\phi)d\phi = a(1-e^2)\int_{\phi_1}^{\phi_2} \frac{d\phi}{\big(1-e^2\sin^2\phi\big)^{3/2}}. \tag{5.67}$$

5.8 The meridian distance on the ellipsoid

On the sphere we defined the meridian distance $m(\phi)$ as the distance along a meridian from the equator to a point at latitude ϕ; this was trivially $m(\phi) = a\phi$. On the ellipsoid we use the same notation but the definition follows from equation (5.67):

$$m(\phi) = \int_0^{\phi} ds_{\text{meridian}} = \int_0^{\phi} \rho(\phi)d\phi = a(1-e^2)\int_0^{\phi} \frac{d\phi}{\big(1-e^2\sin^2\phi\big)^{3/2}}. \tag{5.68}$$

Series expansion for meridian distance: method I

The above (elliptic) integral cannot be evaluated in closed form but, since $e^2 \approx 0.007$, we expand the denominator as a series and integrate term by term. Setting $s = \sin\phi$ we have

$$m(\phi) = a(1-e^2)\int_0^\phi \left(1 + b_2e^2s^2 + b_4e^4s^4 + b_6e^6s^6 + b_8e^8s^8 + \cdots\right)\, d\phi, \qquad (5.69)$$

where, from (E.30),

$$b_2 = \frac{3}{2}, \qquad b_4 = \frac{15}{8}, \qquad b_6 = \frac{35}{16}, \qquad b_8 = \frac{315}{128}. \qquad (5.70)$$

Using the trigonometric identities (C.32) to (C.38) we can express the $\sin^2\phi, \ldots \sin^8\phi$ in terms $\cos 2\phi, \ldots \cos 8\phi$. Collecting terms with the same cosine factors and integrating will then give a series starting with a ϕ term and followed by terms in $\sin 2\phi, \ldots \sin 8\phi$. The result is:

$$\boxed{m(\phi) = A_0\phi + A_2\sin 2\phi + A_4\sin 4\phi + A_6\sin 6\phi + A_8\sin 8\phi + \cdots\,,} \qquad (5.71)$$

where the coefficients are given by

$$\begin{aligned}
A_0 &= a(1-e^2)\left(1+\frac{b_2e^2}{2}+\frac{3b_4e^4}{8}+\frac{5b_6e^6}{16}+\frac{35b_8e^8}{128}\right) = a\left(1-\frac{e^2}{4}-\frac{3e^4}{64}-\frac{5e^6}{256}-\frac{175e^8}{16{*}1024}\right)\\
A_2 &= \frac{a(1-e^2)}{2}\left(-\frac{b_2e^2}{2}-\frac{b_4e^4}{2}-\frac{15b_6e^6}{32}-\frac{7b_8e^8}{16}\right) = a\left(-\frac{3e^2}{8}-\frac{3e^4}{32}-\frac{45e^6}{1024}-\frac{420e^8}{16{*}1024}\right)\\
A_4 &= \frac{a(1-e^2)}{4}\left(\frac{b_4e^4}{8}+\frac{3b_6e^6}{16}+\frac{7b_8e^8}{32}\right) = a\left(\frac{15e^4}{256}+\frac{45e^6}{1024}+\frac{525e^8}{16{*}1024}\right),\\
A_6 &= \frac{a(1-e^2)}{6}\left(-\frac{b_6e^6}{32}-\frac{b_8e^8}{16}\right) = a\left(-\frac{35e^6}{3072}-\frac{175e^8}{12{*}1024}\right),\\
A_8 &= \frac{a(1-e^2)}{8}\left(\frac{b_8e^8}{128}\right) = a\left(\frac{315e^8}{128{*}1024}\right).
\end{aligned} \qquad (5.72)$$

If we use the numerical values for the Airy ellipsoid (5.4), then, in metres,

$$m(\phi) = 6336914{\cdot}609\phi - 15979{\cdot}859\sin 2\phi + 16{\cdot}711\sin 4\phi - 0{\cdot}022\sin 6\phi + 0{\cdot}00003\sin 8\phi \qquad (5.73)$$

The first four terms have been rounded to the nearest millimetre whilst the last term shows that the $O(e^8)$ terms give rise to sub-millimetre corrections. We shall therefore drop $O(e^8)$ terms in expressions for the meridian distance .

The distance from equator to pole is defined by

$$m_p = m(\pi/2) = \frac{1}{2}\pi A_0 = 10001126{\cdot}081 \text{ metres.} \qquad (5.74)$$

Series expansion for meridian distance: method II

There are other ways of obtaining a series expansion. For example, the OSGB publication uses an expansion in terms of the parameter n. From the relations between a, b, e, n given in Section 5.2 we can write the meridian distance as

$$m(\phi) = b(1-n)(1+n)^2 \int_0^\phi \frac{d\phi}{\left(1+2n\cos 2\phi + n^2\right)^{3/2}}. \tag{5.75}$$

The integral is then evaluated by a change of variable: set $z = \exp(2i\phi)$ for which we have $dz = 2iz\, d\phi$ and $z + z^{-1} = 2\cos 2\phi$. The integrand becomes, to $O(n^3)$,

$$\begin{aligned}
&\left(1+2n\cos 2\phi + n^2\right)^{-3/2} \\
&\quad = (1+nz)^{-3/2}\left(1+nz^{-1}\right)^{-3/2} \\
&\quad = \left(1 + a_1 nz + a_2 n^2 z^2 + a_3 n^3 z^3 +\right)\left(1 + a_1 nz^{-1} + a_2 n^2 z^{-2} + a_3 n^3 z^{-3}\right) \\
&\quad = 1 + a_1^2 n^2 + \left(a_1 n + a_1 a_2 n^3\right)\left[z + \frac{1}{z}\right] + a_2 n^2\left[z^2 + \frac{1}{z^2}\right] + a_3 n^3\left[z^3 + \frac{1}{z^3}\right] + O(n^4)
\end{aligned}$$

where the coefficients are given by (E.30):

$$a_1 = -\frac{3}{2}, \qquad a_2 = \frac{15}{8}, \qquad a_3 = -\frac{35}{16}. \tag{5.76}$$

Apart from the overall constant multiplier the integral becomes, to $O(n^3)$,

$$\begin{aligned}
&\int_1^z \frac{dz}{2iz}\left(1 + a_1^2 n^2 + \left(a_1 n + a_1 a_2 n^3\right)\left[z + \frac{1}{z}\right] + a_2 n^2\left[z^2 + \frac{1}{z^2}\right] + a_3 n^3\left[z^3 + \frac{1}{z^3}\right]\right) \\
&= \frac{1}{2i}\left(\left(1 + a_1^2 n^2\right)\ln z + \left(a_1 n + a_1 a_2 n^3\right)\left[z - \frac{1}{z}\right] + \frac{a_2 n^2}{2}\left[z^2 - \frac{1}{z^2}\right] + \frac{a_3 n^3}{3}\left[z^3 - \frac{1}{z^3}\right]\right)\Bigg|_1^z
\end{aligned}$$

Now the contributions from the lower limit all vanish and at the upper limit we have $\ln z = \ln\left(\exp(2i\phi)\right) = 2i\phi$ and $z - z^{-1} = 2i\sin 2\phi$ *etc.* Therefore the final result is

$$m(\phi) = B_0\phi + B_2\sin 2\phi + B_4\sin 4\phi + B_6\sin 6\phi + \cdots, \tag{5.77}$$

where the coefficients are given to order n^3 by

$$\begin{aligned}
B_0 &= b(1-n)(1+n)^2\left(1 + \frac{9}{4}n^2\right) &&= b\left(1 + n + \frac{5}{4}n^2 + \frac{5}{4}n^3\right), \\
B_2 &= b(1-n)(1+n)^2\left(-\frac{3n}{2} - \frac{45n^3}{16}\right) &&= -b\left(\frac{3}{2}n + \frac{3}{2}n^2 + \frac{21}{16}n^3\right), \\
B_4 &= b(1-n)(1+n)^2\left(\frac{15n^2}{16}\right) &&= b\left(\frac{15}{16}n^2 + \frac{15}{16}n^3\right), \\
B_6 &= b(1-n)(1+n)^2\left(-\frac{35n^3}{48}\right) &&= -b\left(\frac{35}{48}n^3\right).
\end{aligned} \tag{5.78}$$

It is straightforward to show that these coefficients are exactly the same as those obtained in Method I. Simply substitute in the A_n with expressions for e^2 *etc.* obtained from (5.9). (Ignoring terms of $O(e^8)$). Therefore we can write

$$m_p = m(\pi/2) = \frac{1}{2}\pi A_0 = \frac{1}{2}\pi B_0 \tag{5.79}$$

The truncated meridian distance

The results we have just obtained measure the meridian distance from the equator. In practice we often require Δm, the distance from a reference latitude ϕ_0. Using the second form of the series we find

$$\begin{aligned}\Delta m &= m(\phi) - m(\phi_0)\\ &= B_0(\phi-\phi_0) + B_2(\sin 2\phi - \sin 2\phi_0) + B_4(\sin 4\phi - \sin 4\phi_0) + B_6(\sin 6\phi - \sin 6\phi_0)\\ &= B_0(\phi-\phi_0) + 2B_2\sin(\phi-\phi_0)\cos(\phi+\phi_0) + 2B_4\sin 2(\phi-\phi_0)\cos 2(\phi+\phi_0)\\ &\qquad + 2B_6\sin 3(\phi-\phi_0)\cos 3(\phi+\phi_0) + \cdots\,.\end{aligned} \tag{5.80}$$

with the coefficients given by equation (5.78).

5.9 Inverse meridian distance

When we derived the inverse series for TMS in Chapter 3 we expressed the coefficients in terms of the footpoint latitude ϕ_1 which was defined by $m_{\text{sph}}(\phi_1) = a\phi_1 = y$ for a given point (x, y) on the projection. Trivially, $\phi_1 = y/a$. We will have to do the same for the ellipsoid but now we are faced with inverting the series (5.71). There are two methods to choose from. The first is a numerical solution by a fixed point iteration, the second is to apply the Lagrange method to a series for a function closely related to the meridian distance, the rectifying latitude.

Inverse meridian distance by numerical methods

To solve $m(\phi) = y$ when $m(\phi)$ is given by (5.71) or (5.77) we consider the iteration

$$\phi_{n+1} = g(\phi_n) = \phi_n - \frac{(m(\phi_n) - y)}{a}, \qquad n = 0, 1, 2\ldots, \tag{5.81}$$

where the initial value is that for the spherical approximation: namely $\phi_0 = y/a$. Now if the iteration scheme converges so that $\phi_{n+1} \to \phi^*$ and $\phi_n \to \phi^*$ then (5.81) becomes

$$\phi^* = g(\phi^*) = \phi^* - \frac{(m(\phi^*) - y)}{a} \tag{5.82}$$

so that $m(\phi^*) - y = 0$ and ϕ^* is the required solution for the footpoint ϕ_1. Note that since $g'(\phi) \approx 1 - B_0/a = O(e^2) < 1$ the iteration will indeed converge.

Inverse by way of the rectifying latitude and Lagrange series expansion

The **rectifying latitude** is simply a scaled version of the meridian distance. Using (5.79),

$$\mu(\phi) = \frac{\pi}{2}\frac{m(\phi)}{m_p} = \frac{m(\phi)}{A_0} = \frac{m(\phi)}{B_0} \tag{5.83}$$

where $m_p = m(\pi/2)$ is the distance from equator to pole given by (5.71) or (5.77): the constants A_0 and B_0 are given by (5.78) or (5.77) respectively. The rectifying latitude may be used to construct projections from the ellipsoid to the sphere which preserve the meridian distance but here we use it simply as a parameter which facilitates the series inversion. We choose to use the expansion in n, equation (5.77).

$$\mu(\phi) = \frac{1}{B_0}m(\phi) = \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + \cdots \tag{5.84}$$

where $b_2 = B_2/B_0$, $b_4 = B_4/B_0$ *etc*. The coefficients are to be calculated to $O(n^3)$ from equations (5.78):

$$\begin{aligned}
\text{set}\quad \epsilon &= n + \frac{5}{4}n^2 + \frac{5}{4}n^3, \\
B_0^{-1} &= b^{-1}(1+\epsilon)^{-1} = b^{-1}(1 - \epsilon + \epsilon^2 - \epsilon^3 + \cdots) \\
&= b^{-1}\Big(1 - n - \frac{1}{4}n^2 + \frac{1}{4}n^3 + \cdots\Big) \\
b_2 &= -bB_0^{-1}\Big(\frac{3}{2}n + \frac{3}{2}n^2 + \frac{21}{16}n^3 + \cdots\Big) &&= -\frac{3}{2}n + \frac{9}{16}n^3 + \cdots, \\
b_4 &= bB_0^{-1}\Big(\frac{15}{16}n^2 + \frac{15}{16}n^3 + \cdots\Big) &&= \frac{15}{16}n^2 + \cdots \\
b_6 &= -bB_0^{-1}\Big(\frac{35}{48}n^3 + \cdots\Big) &&= -\frac{35}{48}n^3 + \cdots.
\end{aligned} \tag{5.85}$$

Finally, we invert equation (5.84) by the Lagrange expansion of Appendix B, Section B.5:

$$\phi = \mu + D_2 \sin 2\mu + D_4 \sin 4\mu + D_6 \sin 6\mu + \cdots, \qquad \mu = \frac{m}{B_0}, \tag{5.86}$$

where, to $O(n^3)$,

$$\begin{aligned}
D_2 &= -b_2 - b_2 b_4 + \frac{1}{2}b_2^3 &&= \frac{3}{2}n - \frac{27}{32}n^3, \\
D_4 &= -b_4 + b_2^2 &&= \frac{21}{16}n^2, \\
D_6 &= -b_6 + 3b_2 b_4 - \frac{3}{2}b_2^3 &&= \frac{151}{96}n^3.
\end{aligned} \tag{5.87}$$

This completes the derivation of a series for the inverse meridian distance. Note that the definition of the rectifying latitude is such that it is zero on the equator and $\pi/2$ at the pole. From the first term of the above series we see that the maximum of $\phi - \mu$ must occur when $\phi \approx \mu \approx 45^\circ$ and it will be of a magnitude given by D_2, approximately $9'$.

5.10 Ellipsoid: summary

Equation: ellipsoid and cross-section

$$\frac{X^2}{a^2}+\frac{Y^2}{a^2}+\frac{Z^2}{b^2}=1, \qquad \frac{p^2}{a^2}+\frac{Z^2}{b^2}=1. \tag{5.88}$$

Parameters

$$b^2=a^2(1-e^2), \qquad f=\frac{a-b}{a}, \qquad e^2=2f-f^2, \tag{5.89}$$

$$e'^2=\frac{a^2-b^2}{b^2}=\frac{e^2}{1-e^2}, \qquad e_1=n=\frac{a-b}{a+b}. \tag{5.90}$$

Airy ellipsoid

$$a=6377563.396\text{m}, \qquad e=0.081673372415, \qquad f=0.03340850522,$$

$$b=6356256.910\text{m}, \qquad e^2=0.006670539762, \qquad \frac{1}{f}=299.3249753. \tag{5.91}$$

Cartesian coordinates

$$\begin{aligned} X(\phi)&=p(\phi)\cos\lambda=\nu(\phi)\,\cos\phi\cos\lambda,\\ Y(\phi)&=p(\phi)\sin\lambda=\nu(\phi)\,\cos\phi\sin\lambda,\\ Z(\phi)&=(1-e^2)\,\nu(\phi)\,\sin\phi. \end{aligned} \tag{5.92}$$

Coordinate derivatives

$$\frac{dp}{d\phi}=-\rho\sin\phi, \qquad \frac{dZ}{d\phi}=\rho\cos\phi. \tag{5.93}$$

Radii of curvature and their ratio *β*

$$\nu(\phi)=\frac{a}{[1-e^2\sin^2\phi]^{1/2}}, \qquad \rho(\phi)=\frac{\nu^3}{a^2}\left(1-e^2\right), \qquad \beta(\phi)=\frac{\nu}{\rho}=\frac{1-e^2\sin^2\phi}{1-e^2}. \tag{5.94}$$

Curvature derivatives

$$\frac{d\nu}{d\phi}=(\beta-1)\rho\tan\phi, \qquad \frac{d\rho}{d\phi}=3\frac{(\beta-1)}{\beta}\rho\tan\phi, \qquad \frac{d\beta}{d\phi}=-2(\beta-1)\tan\phi. \tag{5.95}$$

Metric

$$ds^2=\rho^2\,d\phi^2+\nu^2\cos^2\phi\,d\lambda^2. \tag{5.96}$$

/continued

Meridian distance

$$m(\phi) = A_0\phi + A_2 \sin 2\phi + A_4 \sin 4\phi + A_6 \sin 6\phi + \cdots , \tag{5.97}$$

$$= B_0\phi + B_2 \sin 2\phi + B_4 \sin 4\phi + B_6 \sin 6\phi + \cdots , \tag{5.98}$$

Rectifying latitude

$$\mu(\phi) = \frac{\pi}{2}\frac{m(\phi)}{m_p}. \tag{5.99}$$

Inverse meridian distance

$$\phi = \mu + D_2 \sin 2\mu + D_4 \sin 4\mu + D_6 \sin 6\mu + \cdots . \tag{5.100}$$

In the above series the A_n, B_n and D_n coefficients are given by (5.72), (5.78) and (5.87).

Chapter 6

Normal Mercator on the ellipsoid (NME)

Abstract

Derivation by analogy with NMS. Inversion of the projection by (a) numerical methods and (b) Taylor series expansion. A digression on double projections through a sphere. The conformal latitude and the application of its series expansion to the inversion problem.

6.1 Normal cylindrical projections on the ellipsoid

The normal Mercator projection on the ellipsoid (NME) is a straightforward generalisation of the normal projection on the sphere (NMS) that we discussed in Chapter 2. It is of course a more accurate projection, the differences between NME and NMS being of order $e^2 \approx 0.0067$. We are not so much interested in NME in itself, but rather as a step on the way to TME, the transverse Mercator projection on the ellipsoid.

NME has the same advantages and disadvantages as NMS. It is constructed to be conformal, preserving angles exactly and mapping rhumb lines on the ellipsoid map into lines of constant bearing on the map. Once again conformality guarantees that the scale at any point is isotropic (independent of direction) so that the projection is locally orthomorphic. As in NMS, the scale does vary with latitude, being exact on the equator and reasonably accurate only within a fairly narrow band centred on the equator. The extent of this region of high accuracy may be increased by modifying the projection so that the scale is exact on a pair of parallels at $\pm\phi_1$. The projection is very distorted at high latitudes and nowhere preserves area.

Before investigating the details of NME we consider an arbitrary normal cylindrical projection on the ellipsoid defined by the equations

$$x(\lambda, \phi) = a\lambda, \tag{6.1}$$

$$y(\lambda, \phi) = a\, f(\phi). \tag{6.2}$$

The geometry of the projection may be illustrated by Figure 2.4 with only one change, the normal at P no longer passes through the centre in general. The meridians on the ellipsoid are projected into lines parallel to the y-axis on the projection and parallel circles are projected into lines parallel to the x-axis. The meridian spacing is equal but the spacing of the projected parallels will of course depend on the nature of the function $f(\phi)$. Note that the orthogonal intersections of meridians and parallels on the graticule are transformed into orthogonal intersections on the map but this is *not* necessarily true for intersections at an arbitrary angle.

The essential difference is that the metric on the ellipsoid is given by (5.62):

$$ds^2 = \rho^2\, d\phi^2 + \nu^2 \cos^2\!\phi\, d\lambda^2, \tag{6.3}$$

where $\nu(\phi)$ and $\rho(\phi)$ are defined by equations (5.16) (5.45) respectively. The corresponding infinitesimal elements on the ellipsoid and the plane of projection are shown in Figure 6.1.

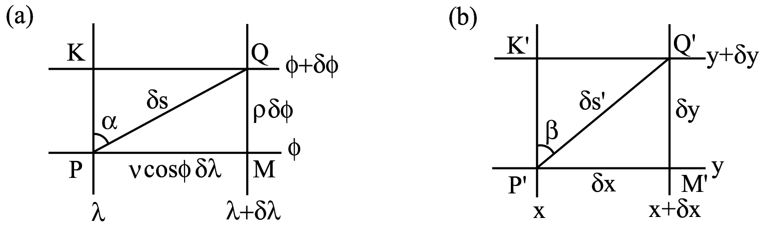


Figure 6.1

Transformation of azimuth to grid bearing

The geometry of the infinitesimal elements gives

$$\text{(a)}\quad \tan\alpha = \frac{\nu\cos\phi\,\delta\lambda}{\rho\,\delta\phi}, \qquad \text{and} \qquad \text{(b)}\quad \tan\beta = \frac{\delta x}{\delta y} = \frac{a\,\delta\lambda}{a\,f'(\phi)\delta\phi}, \tag{6.4}$$

so that

$$\tan\beta = \frac{\rho\sec\phi}{\nu f'(\phi)}\tan\alpha. \tag{6.5}$$

If the projection is conformal, that is $\alpha=\beta$, we must have

$$f'(\phi) = \frac{\rho(\phi)\sec\phi}{\nu(\phi)}. \tag{6.6}$$

We defer the integration to the next section.

The point scale factor

If we denote the distances PQ and $P'Q'$ by δs and $\delta s'$ respectively, the square of the point scale factor is

$$\mu^2 = \lim_{Q\to P}\frac{\delta s'^2}{\delta s^2} = \lim_{Q\to P}\frac{\delta x^2+\delta y^2}{\rho^2\,\delta\phi^2+\nu^2\cos^2\!\phi\,\delta\lambda^2}. \tag{6.7}$$

On PQ and $P'Q'$ equations (6.4) give $\delta\phi = (\nu/\rho)\cot\alpha\,\cos\phi\,\delta\lambda$ and $\delta y = \cot\beta\,\delta x$. Therefore we have

$$\mu_\alpha^2 = \lim_{Q\to P}\frac{\delta x^2(1+\cot^2\!\beta)}{\nu^2\cos^2\!\phi\,\delta\lambda^2(\cot^2\!\alpha+1)}. \tag{6.8}$$

Since $x = a\lambda$ this reduces to

$$\mu_\alpha(\phi) = \frac{a\sec\phi}{\nu(\phi)}\left[\frac{\sin\alpha}{\sin\beta}\right], \tag{6.9}$$

where we assume that β has been found in terms of α and ϕ from equation (6.5). Clearly if the projection is conformal, with $\alpha = \beta$, we have an isotropic scale factor with

$$k(\phi) = \frac{a\sec\phi}{\nu(\phi)}. \tag{6.10}$$

This scale factor differs from that on the sphere by the factor of a/ν, a difference of $O(e^2)$. There is no simple expression for k as a function of y. Given y we must first use one of the methods of inversion discussed in Section 6.3 to find ϕ from y and then we apply (6.10).

6.2 The Mercator parameter on the ellipsoid

The Mercator parameter (or isometric latitude) for NME will be be denoted by $\psi(\phi)$ so that the equations of the projection are still written as

$$x(\lambda,\phi) = a\lambda, \qquad y(\lambda,\phi) = a\psi(\phi). \tag{6.11}$$

Warning. We use the *same* notation for the Mercator parameter on both the sphere and the ellipsoid although they are of course different functions. From this point ψ will always denote the ellipsoidal form which is derived below.

The condition that NME be conformal follows from equation (6.6) when $f(\phi) \to \psi(\phi)$.

$$\frac{d\psi}{d\phi} = \frac{\rho(\phi)\sec\phi}{\nu(\phi)}. \tag{6.12}$$

Substituting for ν and ρ from equations (5.16) and (5.45), splitting into partial fractions and noting that the first term simply gives the same integral as (2.26), we have

$$\begin{aligned}\psi(\phi) &= \int_0^\phi \frac{(1-e^2)}{\cos\phi}\frac{1}{1-e^2\sin^2\phi}\,d\phi \\ &= \int_0^\phi \left[\frac{1}{\cos\phi} - \frac{e^2\cos\phi}{2}\left(\frac{1}{1+e\sin\phi} + \frac{1}{1-e\sin\phi}\right)\right] d\phi \\ &= \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right] - \frac{e}{2}\left[\ln\left(\frac{1+e\sin\phi}{1-e\sin\phi}\right)\right].\end{aligned} \tag{6.13}$$

The first term is just the NMS Mercator parameter for the sphere so we see that the parameters on sphere and ellipse differ by terms of $O(e^2)$. Note that ψ still diverges to $\pm\infty$ at $\phi = \pm\pi/2$. Obviously we recover the parameter for the spherical case when $e \to 0$.

Alternative forms

As in Section 2.4 we can rewrite the Mercator parameter for the ellipsoid in many different forms. The most useful are (a) a simple rearrangement of (6.13), (b) likewise but with $\tan(\phi/2+\pi/4)$ replaced by $\cot(\pi/4-\phi/2)$ and (c) a replacement of the tangent term as in equation (2.31):

$$\psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right], \tag{6.14}$$

$$\psi(\phi) = -\ln\left[\tan\left(\frac{\pi}{4}-\frac{\phi}{2}\right)\left(\frac{1+e\sin\phi}{1-e\sin\phi}\right)^{e/2}\right], \tag{6.15}$$

$$\psi(\phi) = \frac{1}{2}\ln\left[\left(\frac{1+\sin\phi}{1-\sin\phi}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e}\right]. \tag{6.16}$$

6.3 The inverse NME transformation

Inverting the equations $x = a\lambda$ and $y = a\psi$ to find λ and ψ is trivial but finding a value of ϕ from $\psi = y/a$ is anything but trivial. There is no way in which we can manipulate any of the expressions for $\psi(\phi)$ to give $\phi(\psi)$ in a closed form. We have to resort to one of the following methods.

- Attempt to expand $\psi(\phi)$ as a series in ϕ and then invert the series by a Lagrange expansion. (As we did for $m(\phi)$).
- Construct an iterative scheme from which we can find a numerical value for ϕ for any given value of $\psi = y/a$.
- Use a Taylor series expansion of $\phi(\psi)$. In particular we shall need an expansion about the footpoint parameter ψ_1 when we come to the transverse Mercator on the ellipse (TME) in the next chapter.

It is clear that the first method is problematic. The range of ϕ is finite, $[-\pi/2, \pi/2]$, but we have $\psi \to \pm\infty$ as $\phi \to \pm\pi/2$; there is obviously no hope of obtaining a series of the form $\psi = \phi +$ terms of order e, $e^2, \ldots$ which would be amenable to inversion by a Lagrange expansion. However, in Sections 6.4–6.6, we shall see that this is possible for another parameter which is closely related to ψ. In the remainder of this section we consider the numerical solution and the Taylor series.

Numerical inversion by a fixed point iteration

For positive values of ψ and the corresponding ϕ it is convenient to write (6.15) as an implicit equation

$$\phi = \frac{\pi}{2} - 2\arctan\left[\exp(-\psi)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right], \tag{6.17}$$

from which we construct an iteration

$$\phi_{n+1} = \frac{\pi}{2} - 2\arctan\left[\exp(-\psi)\left(\frac{1-e\sin\phi_n}{1+e\sin\phi_n}\right)^{e/2}\right], \tag{6.18}$$

where $n = 0, 1, 2\ldots$ For the initial value ϕ_0 we use the spherical approximation. Setting $e = 0$ in equation (6.17) we have

$$\phi_0 = \frac{\pi}{2} - 2\arctan[\exp(-\psi)]. \tag{6.19}$$

The choice of (6.15) rather than (6.14) introduces the factor of $\exp(-\psi)$ which clearly facilitates the convergence when ψ is large and positive. If the iteration does converge to a value ϕ^*, then $\phi_n \to \phi^*$ and $\phi_{n+1} \to \phi^*$ in (6.18) so that

$$\phi^* = \frac{\pi}{2} - 2\arctan\left[\exp(-\psi)\left(\frac{1-e\sin\phi^*}{1+e\sin\phi^*}\right)^{e/2}\right]. \tag{6.20}$$

Thus ϕ^* is the required solution of (6.17).

Inverse by Taylor series

In the next chapter we shall require the Taylor series of $\phi(\psi)$ about the footpoint parameter $\psi = \psi_1$. To order $(\psi - \psi_1)^4$ we have

$$\phi(\psi) = \phi_1 + (\psi-\psi_1)\left.\frac{d\phi}{d\psi}\right|_1 + \frac{(\psi-\psi_1)^2}{2!}\left.\frac{d^2\phi}{d\psi^2}\right|_1 + \frac{(\psi-\psi_1)^3}{3!}\left.\frac{d^3\phi}{d\psi^3}\right|_1 + \frac{(\psi-\psi_1)^4}{4!}\left.\frac{d^4\phi}{d\psi^4}\right|_1, \tag{6.21}$$

where $\phi_1 = \phi(\psi_1)$ is the footpoint latitude. The definitions of the footpoint parameter and footpoint latitude are as given in equation (4.13) but with the meridian distance given by that on the ellipsoid by equation (5.71) or (5.77).

Now although we do not know the function $\phi(\psi)$ we do know its first derivative *as a function of* ϕ. Equation (6.12) gives

$$\frac{d\phi}{d\psi} = \frac{\nu(\phi)\cos\phi}{\rho(\phi)} = \beta(\phi)\cos\phi = \beta c, \tag{6.22}$$

where we have introduced the usual compact notation

$$s = \sin\phi \qquad c = \cos\phi \qquad t = \tan\phi \qquad \beta = \beta(\phi) = \nu/\rho\,. \tag{6.23}$$

Note that an expression for $d\phi/d\psi$ evaluated at the footpoint latitude ϕ_1 is exactly what we shall need in the application of this series.

We now construct expressions for all the derivatives in the Taylor series as functions of ϕ. We need the derivative of $\beta(\phi)$ given in equation (5.54) as $\beta' = (2-2\beta)t$ where $t = \tan\phi$ (and $dt/d\phi = 1+t^2$).

$$\frac{d\phi}{d\psi} = \beta c,$$

$$\begin{aligned}\frac{d^2\phi}{d\psi^2} &= \frac{d}{d\psi}[\beta c] = \frac{d}{d\phi}[\beta c]\frac{d\phi}{d\psi} = \left[\beta' c - \beta s\right](\beta c)\\ &= \left[(2-2\beta)tc - \beta s\right](\beta c) = c^2 t\left[-3\beta^2+2\beta\right],\end{aligned}$$

$$\begin{aligned}\frac{d^3\phi}{d\psi^3} &= \frac{d}{d\phi}\left\{c^2t\left[-3\beta^2+2\beta\right]\right\}\frac{d\phi}{d\psi}\\ &= (\beta c)\Big\{\big(-2cst + c^2(1+t^2)\big)\left[-3\beta^2+2\beta\right] + c^2t\left[-6\beta+2\right](2-2\beta)t\Big\}\\ &= c^3\left[\beta^3(-3+15t^2)+\beta^2(2-18t^2)+\beta(4t^2)\right],\end{aligned}$$

$$\begin{aligned}\frac{d^4\phi}{d\psi^4} &= \frac{d}{d\phi}\left\{c^3\left[\beta^3(-3+15t^2)+\beta^2(2-18t^2)+\beta(4t^2)\right]\right\}\frac{d\phi}{d\psi}\\ &= c^4t\left[\beta^4(57-105t^2)+\beta^3(-68+180t^2)+\beta^2(16-84t^2)+\beta(8t^2)\right]\end{aligned}\qquad(6.24)$$

These derivatives must be evaluated at ϕ_1 and substituted into the Taylor series which we now write as

$$\phi-\phi_1 = (\psi-\psi_1)\,\beta_1c_1 + \frac{(\psi-\psi_1)^2}{2!}\beta_1c_1^2t_1D_2 + \frac{(\psi-\psi_1)^3}{3!}\beta_1c_1^3D_3 + \frac{(\psi-\psi_1)^4}{4!}\beta_1c_1^4t_1D_4,\qquad(6.25)$$

where $\beta_1 = \nu(\phi_1)/\rho(\phi_1)$, $c_1 = \cos\phi_1$, $t_1 = \tan\phi_1$ and

$$\begin{aligned}D_2 &= -3\beta_1+2\\ D_3 &= \beta_1^2(-3+15t_1^2)+\beta_1(2-18t_1^2)+4t_1^2\\ D_4 &= \beta_1^3(57-105t_1^2)+\beta_1^2(-68+180t_1^2)+\beta_1(16-84t_1^2)+8t_1^2.\end{aligned}\qquad(6.26)$$

We shall also need these coefficients in the spherical limit ($e\to 0,\ \beta\to 1$):

$$\begin{aligned}\overline{D}_2 &= -1\\ \overline{D}_3 &= -1+t_1^2\\ \overline{D}_4 &= 5-t_1^2.\end{aligned}\qquad(6.27)$$

Comment. The following three sections are a digression. We shall not need any of these results in the derivation of TME. They do, however, provide another approach to finding $\phi(\psi)$.

6.4 Double projections via the sphere

Projections can be defined from the ellipsoid to any surface of 'reasonable' shape, not just to the plane. Here we shall consider only the case of projections from the ellipsoid to a sphere of radius R. Such a projection can then be followed by one of the many known projections from the sphere to the plane to produce a double projection with desirable properties.

If we denote the latitude and longitude coordinate on the sphere by (Φ, Λ) then the projection to the sphere is defined by two (well-behaved) functions, $\Phi(\phi, \lambda)$ and $\Lambda(\phi, \lambda)$, where ϕ and λ are the usual geodetic longitude and latitude on the ellipsoid. To complete the double projection we define coordinates (x, y) on the plane by specifying a further two functions $x(\Phi, \Lambda)$ and $y(\Phi, \Lambda)$.

The basic properties of such a projection from the ellipsoid to the sphere can be investigated by comparing the infinitesimal elements shown in Figure 6.2.

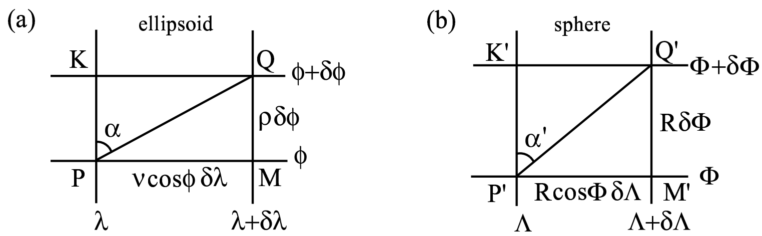


Figure 6.2

The geometry of the infinitesimal elements gives

$$\text{(a)} \quad \tan\alpha = \frac{\nu\cos\phi\,\delta\lambda}{\rho\,\delta\phi} \qquad \text{and} \qquad \text{(b)} \quad \tan\alpha' = \frac{R\cos\Phi\,\delta\Lambda}{R\delta\Phi}. \tag{6.28}$$

Restricted projections to the sphere

The restricted projections are those in which $\Lambda = \lambda$ and Φ is a function of ϕ only. In this case the relation between α and α' then becomes

$$\tan\alpha' = \frac{\cos\Phi}{\Phi'(\phi)}\,\frac{\rho}{\nu\cos\phi}\tan\alpha. \tag{6.29}$$

It is straightforward to calculate the scale factor for any azimuth but we shall consider only the scale factor on the meridian of the sphere; denoting this by $h(\phi)$, as in Section 2.2, we have

$$h(\phi) = \frac{R\delta\,\Phi}{\rho\,\delta\phi} = \frac{R\Phi'(\phi)}{\rho}. \tag{6.30}$$

Note that we have not specified how the radius R is to be chosen. There are many possible choices which, whilst not affecting the angle transformations, will of course influence the scale factor. All of the following have been used for R.

- The semi-major axis.
- An arithmetic or geometric mean of the semi-axes.
- The meridian radius of curvature, ρ, at a latitude where we seek the best fit.
- The Gaussian radius of curvature, $\sqrt{\rho\nu}$, at a latitude where we seek the best fit.
- R such that the ellipsoid and sphere have the same volume.
- R such that the ellipsoid and sphere have the same surface area.
- R such that the ellipsoid and sphere have the same equator–pole distance.

A conformal projection to the sphere: the conformal latitude

The function of $\Phi(\phi)$ which generates a conformal restricted projection to the sphere is called the **conformal latitude**, for which we use the notation $\chi(\phi)$ (in agreement with Snyder—see Bibliography). There are many other nomenclatures in the literature. Beware also that many older books apply 'conformal latitude' to that function of ϕ which we have already defined as the Mercator parameter.

Equation (6.29) shows that the projection is conformal, that is $\alpha{=}\alpha'$, if $\Phi = \chi$ satisfies the condition

$$\sec\chi\;\chi'(\phi) = \frac{\rho(\phi)\sec\phi}{\nu(\phi)}, \tag{6.31}$$

which integrates to

$$\int_0^{\chi(\phi)} \sec\chi\,d\chi = \int_0^{\phi} \frac{\rho(\phi)\sec\phi}{\nu(\phi)} d\phi. \tag{6.32}$$

The integral on the left is the same as that for the Mercator parameter on the sphere, equation (2.26), whilst the integral on the right is that which gives the Mercator parameter on the ellipse, equation (6.13). Therefore

$$\ln\left[\tan\left(\frac{\chi(\phi)}{2}+\frac{\pi}{4}\right)\right] = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right] \equiv \psi(\phi), \tag{6.33}$$

$$\chi(\phi) = 2\arctan\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right] - \frac{\pi}{2}. \tag{6.34}$$

This rather complicated transformation (along with $\Lambda = \lambda$) has been constructed to guarantee a conformal projection from ellipsoid to sphere. But note that equations (6.30) and (6.31) clearly show that the scale cannot be uniform on any meridian of the sphere. Therefore following this projection with TMS from sphere to plane will produce a conformal projection of the ellipsoid to the plane but with a non-uniform scale on the central meridian. When we meet TME (next Chapter) we shall find that it is a conformal projection of the ellipsoid to the plane with a uniform scale on the central meridian.

A rectifying projection from ellipse to the sphere

As a second example of a projection to the sphere consider that defined by setting $\Phi{=}\mu(\phi)$, where $\mu(\phi)$ is the rectifying latitude defined in Section 5.9, equation (5.83):

$$\Phi(\phi) = \mu(\phi) = \frac{\pi}{2}\frac{m(\phi)}{m_p} \tag{6.35}$$

where $m_p = m(\pi/2)$ is the meridian distance between equator and pole on the ellipsoid. The corresponding distance on the sphere is $(\pi/2)R$ and the two are clearly equal if we set $R = 2m_p/\pi$. In terms of the notation developed in Section 5.8 we have $R = A_0 = B_0$ so that we have $a > R > b$ as expected.

The scale factor on the meridian is then

$$h = \frac{R\Phi'(\phi)}{\rho} = \frac{R}{\rho}\frac{\pi}{2}\frac{m'(\phi)}{m_p} = 1, \tag{6.36}$$

since $m'(\phi) = \rho$ from equation (5.68). Thus this projection to the sphere conserves the scale factor and total length on every meridian but on the other hand it is not a conformal projection since $\mu(\phi)$ and $\chi(\phi)$ are different functions.

6.5 A series expansion for the conformal latitude

Equation (6.34) shows that χ and ϕ are equal at the equator and at the pole and elsewhere they differ by $O(e)$ terms, (actually by $O(e^2)$ as we shall see), so there is every reason to expect them to be related by a series which can be inverted by the Lagrange method. In this section we will find coefficients such that

$$\chi(\phi) = \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots\,, \tag{6.37}$$

$$\phi(\chi) = \chi + d_2 \sin 2\chi + d_4 \sin 4\chi + d_6 \sin 6\chi + d_8 \sin 8\chi + \cdots\,. \tag{6.38}$$

To develop the direct series for $\chi(\phi)$ we first introduce two new (small) parameters, A and η, such that

$$\left(\frac{1 - e\sin\phi}{1 + e\sin\phi}\right)^{e/2} = \exp A = 1 + \eta. \tag{6.39}$$

Taking logarithms

$$A = \frac{e}{2}\ln\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right) = \frac{e}{2}\ln\left(\frac{1-es}{1+es}\right).$$

Using the series (E.12) we obtain A and its powers to order $O(e^8)$:

$$\begin{aligned}
A &= -e^2 s\left[1+\frac{1}{3}e^2s^2+\frac{1}{5}e^4s^4+\frac{1}{7}e^6s^6\right],\\
A^2 &= \quad e^4s^2\left[1+\frac{2}{3}e^2s^2+\frac{23}{45}e^4s^4\right],\\
A^3 &= -e^6s^3\left[1+e^2s^2\right],\\
A^4 &= \quad e^8s^4.
\end{aligned}$$

Therefore

$$\begin{aligned}
\eta &= \exp A\ -1 = A+\frac{1}{2!}A^2+\frac{1}{3!}A^3+\frac{1}{4!}A^4+\cdots\\
&= p_2e^2+p_4e^4+p_6e^6+p_8e^8+\cdots,\\
\eta^2 &= p_2^2e^4+2p_2p_4e^6+(p_4^2+2p_2p_6)e^8+\cdots,\\
\eta^3 &= p_2^3e^6+3p_2^2p_4e^8+\cdots,\\
\eta^4 &= p_2^4e^8+\cdots,
\end{aligned}\tag{6.40}$$

where

$$\begin{aligned}
p_2 &= -s\\
p_4 &= \quad (s^2/6)(3-2s),\\
p_6 &= -(s^3/30)(5-10s+6s^2),\\
p_8 &= \quad (s^4/2520)(105-420s+644s^2-360s^3).
\end{aligned}\tag{6.41}$$

We now introduce the abbreviation, (and use equation 2.29),

$$b=\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)=\sec\phi+\tan\phi=\frac{1}{c}(1+s),\tag{6.42}$$

and note for future reference that

$$\frac{b}{1+b^2}=\frac{\tan(\phi/2+\pi/4)}{\sec^2(\phi/2+\pi/4)}=\frac{1}{2}\sin\left(\phi+\frac{\pi}{2}\right)=\frac{1}{2}\cos\phi=\frac{c}{2}.\tag{6.43}$$

With this notation equation (6.34) becomes

$$\frac{\chi}{2}+\frac{\pi}{4}=\arctan[b(1+\eta)] \;=\; \arctan(b+\eta b).\tag{6.44}$$

For the inverse tangent we use the series (E.9) with $z=\eta b$. Therefore

$$\begin{aligned}
\frac{\chi}{2}+\frac{\pi}{4} &= \frac{\phi}{2}+\frac{\pi}{4}+\frac{\eta b}{1!}\frac{1}{1+b^2}-\frac{(\eta b)^2}{2!}\frac{2b}{(1+b^2)^2}\\
&+\frac{(\eta b)^3}{3!}\left[\frac{-2}{(1+b^2)^2}+\frac{8b^2}{(1+b^2)^3}\right]-\frac{(\eta b)^4}{4!}\left[\frac{-24b}{(1+b^2)^3}+\frac{48b^3}{(1+b^2)^4}\right]+\cdots,
\end{aligned}\tag{6.45}$$

Substitute first for the powers of $[b/(1+b^2)]$ from equation (6.43) and then substitute for the remaining powers of b from equation (6.42). The result is

$$\chi = \phi + c\eta - \frac{c}{2}(1+s)\eta^2 + \frac{c}{6}(1+3s+2s^2)\eta^3 - \frac{c}{4}(s+2s^2+s^3)\eta^4 + \cdots\,. \tag{6.46}$$

Substitute for the powers of η in terms of the p-coefficients (equation 6.40) and then substitute for the p-coefficients using equation 6.41. This gives a series of the form

$$\chi = \phi + q_2 e^2 + q_4 e^4 + q_6 e^6 + q_8 e^8 + \cdots\,, \tag{6.47}$$

where the coefficients are given by

$$\begin{aligned}
q_2 &= cp_2 = -sc,\\
q_4 &= cp_4 - \frac{c}{2}(1+s)(p_2^2) = 0.s^2c - \frac{5}{6}s^3c,\\
q_6 &= cp_6 - \frac{c}{2}(1+s)(2p_2p_4) + \frac{c}{6}(1+3s+2s^2)(p_2^3) = \frac{1}{6}s^3c + 0.s^4c - \frac{13}{15}s^5c,\\
q_8 &= cp_8 - \frac{c}{2}(1+s)(p_4^2+2p_2p_6) + \frac{c}{6}(1+3s+2s^2)(3p_2^2p_4) - \frac{c}{4}(s+2s^2+s^3)(p_2^4)\\
&= 0.s^4c + \frac{9}{24}s^5c + 0.s^6c - \frac{1237}{1260}s^7c.
\end{aligned} \tag{6.48}$$

Note the cancellation of all the terms involving s^2c, s^4c and s^6c. This suggests that there must be a smarter way of carrying out this expansion. Finally, using the identities for sc, s^3c, s^5c, s^7c given in Appendix C, equations (C.39) *etc.*

$$\boxed{\chi = \phi + b_2\sin 2\phi + b_4\sin 4\phi + b_6\sin 6\phi + b_8\sin 8\phi + \cdots} \tag{6.49}$$

where

$$\begin{aligned}
b_2 &= -\frac{e^2}{2} - \frac{5e^4}{24} - \frac{3e^6}{32} - \frac{281e^8}{5760} - \cdots\,,\\
b_4 &= \frac{5e^4}{48} + \frac{7e^6}{80} + \frac{697e^8}{11520} + \cdots\,,\\
b_6 &= -\frac{13e^6}{480} - \frac{461e^8}{13440} - \cdots\,,\\
b_8 &= \frac{1237e^8}{161280} + \cdots\,.
\end{aligned} \tag{6.50}$$

The leading correction term (b_2) gives a maximum value of $\chi - \phi \approx 12'$ at $\phi \approx 45°$.

The inverse series

The inverse of the above series for χ may be calculated as a Lagrange expansion as in Appendix B, equation (B.17). The result is

$$\boxed{\phi = \chi + d_2\sin 2\chi + d_4\sin 4\chi + d_6\sin 6\chi + d_8\sin 8\chi + \cdots} \tag{6.51}$$

where

$$\begin{aligned}
d_2 &= -b_2 - b_2 b_4 + \frac{1}{2}b_2^3 &&= \frac{e^2}{2} + \frac{5e^4}{24} + \frac{e^6}{12} + \frac{13e^8}{360} + \cdots\,,\\
d_4 &= -b_4 + b_2^2 - 2b_2 b_6 + 4b_2^2 b_4 - \frac{4}{3}b_2^4 &&= \frac{7e^4}{48} + \frac{29e^6}{240} + \frac{811e^8}{11520} + \cdots\,,\\
d_6 &= -b_6 + 3b_2 b_4 - \frac{3}{2}b_2^3 &&= \frac{7e^6}{120} + \frac{81e^8}{1120} + \cdots\,,\\
d_8 &= -b_8 + 2b_4^2 + 4b_2 b_6 - 8b_2^2 b_4 + \frac{8}{3}b_2^4 &&= \frac{4279e^8}{161280} + \cdots\,. \qquad (6.52)
\end{aligned}$$

6.6 The inverse of the Mercator parameter

Returning to equation (6.33) we see that the relation between the Mercator parameter and the conformal latitude is

$$\psi(\chi) = \ln\left[\tan\left(\frac{\chi}{2} + \frac{\pi}{4}\right)\right] = \ln\left[\cot\left(\frac{\pi}{4} - \frac{\chi}{2}\right)\right]. \qquad (6.53)$$

Inverting the second of these gives

$$\chi(\psi) = \frac{\pi}{2} - 2\arctan\left[\exp(-\psi)\right]. \qquad (6.54)$$

This result, alongwith the last section, provides another inversion of the Mercator parameter. Given $\psi = y/a$ we use equation (6.54) to calculate χ and then use the series (6.51) to find ϕ. This solves the problem of the inverse transformation and also allows us to find the scale factor for any y using (6.10).

6.7 Summary of modified NME

NME can be modified exactly as NMS to provide a slightly wider domain near the equator in which the scale is accurate to within a given tolerance. For a tolerance of 1 in 2500 the range of latitude will differ from that for NMS (Section 2.7) by terms of order e^2. The technique is the same, we simply introduce a factor of k_0 into the transformations.

Direct transformation

$$x = k_0 a\lambda, \qquad y = k_0 a\psi(\phi), \qquad \psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2} + \frac{\pi}{4}\right)\left(\frac{1 - e\sin\phi}{1 + e\sin\phi}\right)^{e/2}\right] \qquad (6.55)$$

Inverse transformation

$$\lambda = x/k_0 a, \qquad \phi = \phi(\psi) \quad \text{with} \quad \psi = y/k_0 a \tag{6.56}$$

where $\phi(\psi)$ is calculated by either of the following two methods:

Inverse of the Mercator parameter I

$\phi(\psi)$ may be evaluated numerically by the iteration of

$$\phi_{n+1} = \frac{\pi}{2} - 2\arctan\left[\exp(-\psi)\left(\frac{1 - e\sin\phi_n}{1 + e\sin\phi_n}\right)^{e/2}\right], \tag{6.57}$$

with a starting value

$$\phi_0 = \frac{\pi}{2} - 2\arctan\left[\exp(-\psi).\right]$$

Inverse of the Mercator parameter II

$\phi(\psi)$ may also be calculated by using the series

$$\phi(\chi) = \chi + d_2\sin 2\chi + d_4\sin 4\chi + d_6\sin 6\chi + d_8\sin 8\chi + \cdots, \tag{6.58}$$

where the coefficients are given in equation (6.52) and χ (the conformal latitude) is defined in terms of ψ by

$$\chi(\psi) = \frac{\pi}{2} - 2\arctan\left[\exp(-\psi)\right]. \tag{6.59}$$

Scale factor

$$k(\phi) = \frac{k_0 a \sec\phi}{\nu(\phi)} \tag{6.60}$$

To find the scale for a given y on the projection we use the above results to first find ϕ when given $\psi = y/k_0 a$.

Chapter 7

Transverse Mercator on the ellipsoid (TME)

Abstract

TME is derived as a series by a complex transformation from the NME projection. The method parallels that used in Chapter 4 for the derivation of the TMS series from NMS.

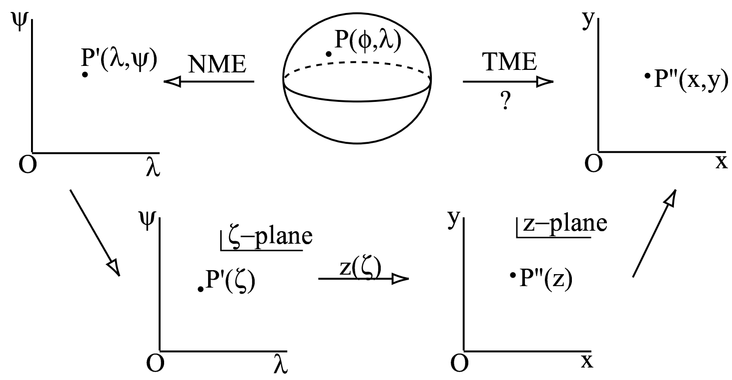


Figure 7.1

7.1 Introduction

In Chapter 6 we derived the NME projection: it can be considered as a conformal transformation from a point $P(\phi, \lambda)$ on the ellipsoid to a point on the complex ζ-plane defined by $\zeta = \lambda + i\psi$ where $\psi(\phi)$ is the Mercator parameter for the ellipsoid given in (6.14).

Let (x, y) be the coordinates of the required TME projection and let $z = x + iy$ be a general point on the associated complex plane. The aim of this chapter is to find a conformal transformation

$$\zeta \to z(\zeta) \equiv x(\lambda, \psi) + iy(\lambda, \psi), \tag{7.1}$$

such that (a) the central meridians, $\lambda = 0$ and $x = 0$, map into each other, and (b) the scale is true on the y-axis. Our method parallels that of Chapter 4 with the appropriate definitions of the Mercator parameter and meridian distance for the ellipsoid. The resulting series are those first given by Lee (to sixth order) and Redfearn (to eighth order)—see bibliography.

The meridian distance

The meridian distance on the ellipsoid was obtained as a series in Section 5.8: two possible forms are given in equations (5.71) or (5.77). The first of these is (to sufficient accuracy)

$$m(\phi) = A_0\phi + A_2 \sin 2\phi + A_4 \sin 4\phi + A_6 \sin 6\phi, \tag{7.2}$$

where the A-coefficients are given in equations (5.72). In considering the transformation from the complex ζ-plane to the complex z-plane it is useful to express the meridian distance as a function of ψ and write it as $M(\psi)$, where

$$M\big(\psi(\phi)\big) = m(\phi). \tag{7.3}$$

There is no closed expression for $M(\psi)$ analogous to (4.6); this is of no import since we only need its derivatives. (See next page).

Footpoint latitude and parameter

Given a point P'' with projection coordinates (x, y) then the projection coordinates of the footpoint are $(0, y)$. The definition of the footpoint latitude ϕ_1 and the footpoint parameter ψ_1 are unchanged from those of Sections 3.3 and 4.1: they are the solutions of

$$m(\phi_1) = y, \qquad M(\psi_1) = y. \tag{7.4}$$

We shall need to calculate the footpoint latitude (but not the footpoint parameter) for a given y. One method of finding the solution of $m(\phi)=y$ is to use the fixed point iteration given in equation (5.81),

$$\phi_{n+1} = g(\phi_n) = \phi_n - \frac{\big(m(\phi_n) - y\big)}{a}, \qquad n = 0, 1, 2 \ldots, \tag{7.5}$$

starting with the spherical approximation $\phi_0 = y/a$.

Alternatively, we can use the series given in (5.86) with $m(\phi) = y$,

$$\phi = \mu + d_2 \sin 2\mu + d_4 \sin 4\mu + d_6 \sin 6\mu + \cdots, \qquad \mu = \frac{y}{B_0}, \tag{7.6}$$

where the d-coefficients are given in (5.87).

The Mercator parameter: derivative and inverse

The Mercator parameter on the ellipsoid is given in equation (6.14) as

$$\psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2} + \frac{\pi}{4}\right)\left(\frac{1 - e\sin\phi}{1 + e\sin\phi}\right)^{e/2}\right]. \tag{7.7}$$

We shall not need this explicit form but we shall require require its derivatives. From (6.12)

$$\frac{d\psi}{d\phi} = \frac{\rho(\phi)}{\nu(\phi)\cos\phi}, \qquad \frac{d\phi}{d\psi} = \frac{\nu(\phi)\cos\phi}{\rho(\phi)}. \tag{7.8}$$

We shall also need $\phi(\psi)$, the inverse of the Mercator parameter, as a fourth order Taylor series about the footpoint parameter ψ_1. This is given in equation (6.25).

$$\phi-\phi_1 = (\psi-\psi_1)\,\beta_1 c_1 + \frac{(\psi-\psi_1)^2}{2!}\beta_1 c_1^2 t_1 D_2 + \frac{(\psi-\psi_1)^3}{3!}\beta_1 c_1^3 D_3 + \frac{(\psi-\psi_1)^4}{4!}\beta_1 c_1^4 t_1 D_4 \tag{7.9}$$

where the D-coefficients are given in (6.26). The '1' suffix of course denotes a term calculated at the footpoint latitude.

The derivatives of the meridian distance

We shall need the derivative of the $M(\psi)$ as functions of ϕ. From (5.68) we have

$$\frac{dm(\phi)}{d\phi} = \rho(\phi), \tag{7.10}$$

and using (7.8) we obtain

$$M'(\psi) \equiv \frac{dM(\psi)}{d\psi} = \frac{dM(\psi(\phi))}{d\phi}\frac{d\phi}{d\psi} = \frac{dm(\phi)}{d\phi}\frac{\nu\cos\phi}{\rho} = \nu(\phi)\cos\phi. \tag{7.11}$$

Proceeding in this way we can construct all the derivatives of $M(\psi)$ with respect to ψ but with the results expressed as functions of ϕ. Denoting the n-th derivative of M with respect to ψ by $M^{(n)}$ the exact results for the first six derivatives are given below. We use the usual compact notation for $\sin\phi$ *etc.* and also make frequent use of the derivatives of $\nu(\phi)$ and $\beta(\phi)$ given in equation (5.54):

$$\frac{d\nu}{d\phi} = (\beta-1)\rho\tan\phi, \qquad \frac{d\beta}{d\phi} = -2(\beta-1)\tan\phi. \tag{7.12}$$

$$M^{(1)} = \frac{dM}{d\psi} = \nu c. \tag{7.13}$$

$$\begin{aligned} M^{(2)} = \frac{d^2M}{d\psi^2} &= \frac{d}{d\phi}\left(M^{(1)}\right)\frac{d\phi}{d\psi} = \frac{d}{d\phi}(\nu c)\frac{d\phi}{d\psi} = [(\beta-1)\rho tc - \nu s]\,\frac{\nu c}{\rho} \\ &= -\nu sc. \end{aligned} \tag{7.14}$$

$$\begin{aligned} M^{(3)} = \frac{d^3M}{d\psi^3} &= -\left[(\beta-1)\rho tsc + \nu(c^2-s^2)\right]\,\frac{\nu c}{\rho} = -\nu c^3\left(\beta - t^2\right) \\ &\equiv -\nu c^3 W_3. \end{aligned} \tag{7.15}$$

/cont.

$$
\begin{aligned}
M^{(4)} = \frac{d^4M}{d\psi^4} &= -\Big[\, \big\{(\beta{-}1)\rho tc^3 - 3\nu c^2 s\big\}\,(\beta{-}t^2) + \nu c^3\,\big\{-2(\beta{-}1)t - 2t(1{+}t^2)\big\}\,\Big]\,\frac{\nu c}{\rho} \\
&= \nu sc^3\left[4\beta^2+\beta-t^2\right] \qquad (7.16)\\
&\equiv \nu sc^3 W_4.
\end{aligned}
$$

$$
\begin{aligned}
M^{(5)} = \frac{d^5M}{d\psi^5} &= \Big[\,\big\{(\beta-1)\rho tsc^3 + \nu(c^4-3s^2c^2)\big\}\,(4\beta^2+\beta-t^2) \\
&\qquad + \nu sc^3\,\big\{(8\beta+1)(-2\beta+2)t - 2t(1+t^2)\big\}\,\Big]\,\frac{\nu c}{\rho} \\
&= \nu c^5\left[4\beta^3(1-6t^2)+\beta^2(1+8t^2)-2\beta t^2+t^4\right] \qquad (7.17)\\
1ex &\equiv \nu c^5 W_5.
\end{aligned}
$$

$$
\begin{aligned}
M^{(6)} = \frac{d^6M}{d\psi^6} &= \Big[\,\big\{(\beta-1)\rho tc^5 + \nu(-5sc^4)\big\}\,W_5 + \nu c^5 W_5'\Big]\,\frac{\nu c}{\rho} \\
&= \nu sc^5\Big[(-4\beta-1)\,\big\{4\beta^3(1-6t^2)+\beta^2(1+8t^2)-2\beta t^2+t^4\big\} \\
&\qquad + \beta t^{-1}\,\big\{\big(12\beta^2(1-6t^2)+2\beta(1+8t^2)-2t^2\big)\,(-2\beta+2)\,t \\
&\qquad\qquad - \big(24\beta^3-8\beta^2+2\beta\big)\,\big(2t(1+t^2)\big) + 4t^3(1+t^2)\big\}\,\Big] \\
&= -\nu sc^5\left[8\beta^4(11-24t^2)-28\beta^3(1-6t^2)+\beta^2(1-32t^2)-2\beta t^2+t^4\right] \\
&\equiv -\nu sc^5 W_6 \qquad (7.18)
\end{aligned}
$$

We shall find that the derivatives $M^{(7)}$ and $M^{(8)}$ multiply λ^7 and λ^8 terms respectively and we shall later justify the neglect of terms of order $e^2\lambda^7$ and $e^2\lambda^8$. Accordingly, we evaluate these derivatives in the spherical limit in which $e \to 0$ and $\beta \to 1$, (except that the overall multiplicative factors of ν are not set equal to a for the sake of visual conformity with the lower order derivatives, not to improve accuracy). Noting that $\nu' = \beta' = 0$ in this limit we find

$$
\begin{aligned}
M^{(7)} = \frac{d^7M}{d\psi^7} &= -\left[\nu\left(c^6-5s^2c^4\right)\,W_6\big|_{\beta=1} + \nu sc^5\;W_6'\big|_{\beta=1}\right]\frac{\nu c}{\rho} \\
&= -\nu c^7\left[\left(1-5t^2\right)\left(61-58t^2+t^4\right)+t\left(-116t+4t^3\right)\left(1+t^2\right)\right] \\
&= -\nu c^7\left(61-479t^2+179t^4-t^6\right). \\
&\equiv -\nu c^7\overline{W}_7. \qquad (7.19)
\end{aligned}
$$

$$
\begin{aligned}
M^{(8)} = \frac{d^8M}{d\psi^8} &= \left[7\nu c^6 s\overline{W}_7 - \nu c^7\overline{W}_7{}'\right]\frac{\nu c}{\rho} \\
&= \nu sc^7\left[7\left(61{-}479t^2{+}179t^4{-}t^6\right) - \frac{1}{t}\left(-958t{+}716t^3{-}6t^5\right)\left(1{+}t^2\right)\right] \\
&= \nu sc^7\left(1385-3111t^2+543t^4-t^6\right). \\
&\equiv \nu sc^7\overline{W}_8. \qquad (7.20)
\end{aligned}
$$

Note the minus signs introduced in the definitions of W_3, W_6 and $\overline{W}_7$.

Summary of derivatives

$$
\begin{aligned}
M^{(1)} &= \quad \nu c \\
M^{(2)} &= -\nu sc \\
M^{(3)} &= -\nu c^3\ W_3 \qquad W_3(\phi) = \beta - t^2 \\
M^{(4)} &= \quad \nu sc^3 W_4 \qquad W_4(\phi) = 4\beta^2 + \beta - t^2 \\
M^{(5)} &= \quad \nu c^5\ W_5 \qquad W_5(\phi) = 4\beta^3(1{-}6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4 \\
M^{(6)} &= -\nu sc^5 W_6 \qquad W_6(\phi) = 8\beta^4(11{-}24t^2){-}28\beta^3(1{-}6t^2){+}\beta^2(1{-}32t^2){-}2\beta t^2 + t^4 \\
M^{(7)} &= -\nu c^7\ \overline{W}_7 \qquad \overline{W}_7(\phi) = 61 - 479t^2 + 179t^4 - t^6 + O(e^2) \\
M^{(8)} &= \quad \nu sc^7 \overline{W}_8 \qquad \overline{W}_8(\phi) = 1385 - 3111t^2 + 543t^4 - t^6 + O(e^2).
\end{aligned} \tag{7.21}
$$

The bar on $\overline{W}_7$ and $\overline{W}_8$ denotes that the term is evaluated in the spherical limit. This notation will be standard from here on. Later we will need the expressions for $W_3, \ldots W_6$ in the spherical approximation: setting $\beta = 1$ gives

$$
\begin{aligned}
W_3(\phi) &\to \overline{W}_3(\phi) = 1 - t^2, \\
W_4(\phi) &\to \overline{W}_4(\phi) = 5 - t^2, \\
W_5(\phi) &\to \overline{W}_5(\phi) = 5 - 18t^2 + t^4, \\
W_6(\phi) &\to \overline{W}_6(\phi) = 61 - 58t^2 + t^4.
\end{aligned} \tag{7.22}
$$

7.2 Derivation of the Redfearn series

The direct complex series

Following Section 4.2, the complex Taylor series of $z(\zeta)$ about ζ_0 on the central meridian is

$$
\begin{aligned}
z = z_0 &+ (\zeta{-}\zeta_0)M_0^{(1)} - \frac{i}{2!}(\zeta{-}\zeta_0)^2 M_0^{(2)} - \frac{1}{3!}(\zeta{-}\zeta_0)^3 M_0^{(3)} + \frac{i}{4!}(\zeta{-}\zeta_0)^4 M_0^{(4)} \\
&+ \frac{1}{5!}(\zeta{-}\zeta_0)^5 M_0^{(5)} - \frac{i}{6!}(\zeta{-}\zeta_0)^6 M_0^{(6)} - \frac{1}{7!}(\zeta{-}\zeta_0)^7 M_0^{(7)} + \frac{i}{8!}(\zeta{-}\zeta_0)^8 M_0^{(8)} + \cdots,
\end{aligned} \tag{7.23}
$$

where $M_0^{(n)} = M^{(n)}(\psi_0)$, the n-th derivative of $M(\psi)$ with respect to ψ evaluated at ψ_0. The leading term in the expansion will be recast in various forms when required:

$$
z_0 = z(\zeta_0) = iy_0 = iM(\psi_0) = iM_0 \tag{7.24}
$$

The direct series for x and y

For the direct series we start from a given (arbitrary) point P' at $\zeta = \lambda + i\psi$ and choose $\zeta_0 = i\psi$ with the *same* ordinate in the ζ-plane. Therefore in the Taylor series (7.23) we set

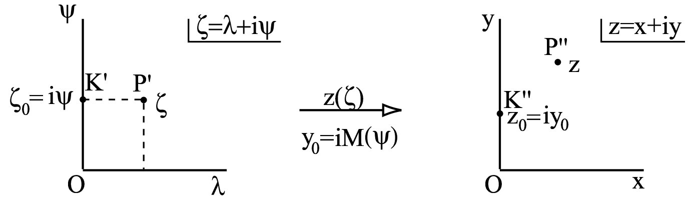


Figure 7.2

$\zeta - \zeta_0 = \lambda$ and evaluate the derivatives at $\psi_0 = \psi$. Writing $M^{(n)}(\psi)$ as $M^{(n)}$ and using $z_0 = iM_0 \to iM$ the generalisation of equation (4.25) is

$$\begin{aligned} z = x + iy = iM + \lambda M^{(1)} - \frac{i}{2!}\lambda^2 M^{(2)} - \frac{1}{3!}\lambda^3 M^{(3)} + \frac{i}{4!}\lambda^4 M^{(4)} \\ + \frac{1}{5!}\lambda^5 M^{(5)} - \frac{i}{6!}\lambda^6 M^{(6)} - \frac{1}{7!}\lambda^7 M^{(7)} + \frac{i}{8!}\lambda^8 M^{(8)} + \cdots \end{aligned} \tag{7.25}$$

The real and imaginary parts of equation (7.25) give x and y as functions of λ and ψ:

$$x(\lambda, \psi) = \lambda M^{(1)} - \frac{1}{3!}\lambda^3 M^{(3)} + \frac{1}{5!}\lambda^5 M^{(5)} - \frac{1}{7!}\lambda^7 M^{(7)} + \cdots \tag{7.26}$$

$$y(\lambda, \psi) = M - \frac{1}{2!}\lambda^2 M^{(2)} + \frac{1}{4!}\lambda^4 M^{(4)} - \frac{1}{6!}\lambda^6 M^{(6)} + \frac{1}{8!}\lambda^8 M^{(8)} + \cdots . \tag{7.27}$$

Writing M and its derivatives as functions of ϕ from (7.3) and (7.21) gives the Redfearn formulae for the direct transformation as power series in λ (radians):

$$x(\lambda, \phi) = \lambda\nu c + \frac{\lambda^3 \nu c^3}{3!} W_3 + \frac{\lambda^5 \nu c^5}{5!} W_5 + \frac{\lambda^7 \nu c^7}{7!} W_7, \tag{7.28}$$

$$y(\lambda, \phi) = m(\phi) + \frac{\lambda^2 \nu s c}{2} + \frac{\lambda^4 \nu s c^3}{4!} W_4 + \frac{\lambda^6 \nu s c^5}{6!} W_6 + \frac{\lambda^8 \nu s c^7}{8!} W_8. \tag{7.29}$$

Since all the coefficients on the right hand sides are now expressed in terms of ϕ, we have replaced $x(\lambda, \psi)$ and $y(\lambda, \psi)$ on the left hand side by $x(\lambda, \phi)$ and $y(\lambda, \phi)$ respectively.

Conformality and the Cauchy–Riemann equations

The conformality of the above transformations may be confirmed by evaluating the Cauchy–Riemann equations (4.14)

$$x_\lambda = y_\psi = M^{(1)} - \frac{1}{2!}\lambda^2 M^{(3)} + \frac{1}{4!}\lambda^4 M^{(5)} - \frac{1}{6!}\lambda^6 M^{(7)} + \cdots , \tag{7.30}$$

$$x_\psi = -y_\lambda = \lambda M^{(2)} - \frac{1}{3!}\lambda^3 M^{(4)} + \frac{1}{5!}\lambda^5 M^{(6)} - \frac{1}{7!}\lambda^7 M^{(8)} + \cdots . \tag{7.31}$$

The inverse complex series

We start by dividing the direct Taylor series (7.23) by a factor of $M_0^{(1)}$ which, from (7.21), is equal to $\nu_0 c_0$. Therefore

$$\frac{z-z_0}{\nu_0 c_0} = (\zeta-\zeta_0) + \frac{b_2}{2!}(\zeta-\zeta_0)^2 + \frac{b_3}{3!}(\zeta-\zeta_0)^3 + \cdots + \frac{b_8}{8!}(\zeta-\zeta_0)^8 + \cdots \qquad (7.32)$$

where we have set $z_0 = iy_0 = iM_0$. The b-coefficients are

$$\begin{aligned}
b_2 &= \frac{-iM_0^{(2)}}{\nu_0 c_0} = is_0\\
b_3 &= \frac{-M_0^{(3)}}{\nu_0 c_0} = c_0^2 W_3(\phi_0)\\
b_4 &= \frac{iM_0^{(4)}}{\nu_0 c_0} = is_0 c_0^2 W_4(\phi_0)\\
b_5 &= \frac{M_0^{(5)}}{\nu_0 c_0} = c_0^4 W_5(\phi_0)\\
b_6 &= \frac{-iM_0^{(6)}}{\nu_0 c_0} = is_0 c_0^4 W_6(\phi_0)\\
b_7 &= \frac{-M_0^{(7)}}{\nu_0 c_0} = c_0^6 W_7(\phi_0)\\
b_8 &= \frac{iM_0^{(8)}}{\nu_0 c_0} = is_0 c_0^6 W_8(\phi_0)
\end{aligned} \qquad (7.33)$$

where the functions on the right hand sides are evaluated at ϕ_0 such that $\psi_0 = \psi(\phi_0)$.

The Lagrange inversion of an eighth order series is developed in Appendix B, Sections B.6–B.8. If we identify the series (7.32) with (B.23) by replacing $(z-z_0)/\nu_0 c_0$ and $(\zeta-\zeta_0)$ by w and z respectively we can use (B.24) to deduce that the inverse of (7.32) is

$$\zeta-\zeta_0 = \left(\frac{z-z_0}{\nu_0 c_0}\right) - \frac{p_2}{2!}\left(\frac{z-z_0}{\nu_0 c_0}\right)^2 - \frac{p_3}{3!}\left(\frac{z-z_0}{\nu_0 c_0}\right)^3 - \cdots - \frac{\bar{p}_8}{8!}\left(\frac{z-z_0}{\nu_0 c_0}\right)^8, \qquad (7.34)$$

where the p-coefficients are given by equations (B.25) and (B.30). We shall actually need these coefficients at the footpoint latitude ϕ_1 and we choose to write them as

$$\begin{aligned}
p_2 &= ic_1 t_1, &&\\
p_3 &= c_1^2\, V_3 && V_3 = \beta_1 + 2t_1^2,\\
p_4 &= ic_1^3 t_1\, V_4 && V_4 = 4\beta_1^2 - 9\beta_1 - 6t_1^2,\\
p_5 &= c_1^4\, V_5 && V_5 = 4\beta_1^3(1-6t_1^2) - \beta_1^2(9-68t_1^2) - 72\beta_1 t_1^2 - 24t_1^4,\\
p_6 &= ic_1^5 t_1\, V_6 && V_6 = 8\beta_1^4(11-24t_1^2) - 84\beta_1^3(3-8t_1^2) + 225\beta_1^2(1-4t_1^2) + 600\beta_1 t_1^2 + 120t_1^4,\\
\bar{p}_7 &= c_1^6\, \bar{V}_7 && \bar{V}_7 = 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6,\\
\bar{p}_8 &= ic_1^7 t_1\, \bar{V}_8 && \bar{V}_8 = -1385 - 7266t_1^2 - 10920t_1^4 - 5040t_1^6.
\end{aligned} \qquad (7.35)$$

The inverse series for ψ and λ

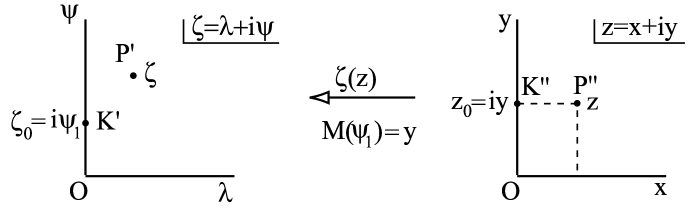


Figure 7.3

For the inverse we start from an arbitrary point with projection coordinates $P''(x,y)$ and move to the footpoint at $K''(0,y)$ so that we set $z-z_0=(x+iy)-iy=x$ in equation (7.34). We must then set $\zeta_0=i\psi_1$ where ψ_1 is the footpoint parameter such that $M(\psi_1)=y$. Let ϕ_1 be the corresponding footpoint latitude such that $m(\phi_1)=y$. Therefore (7.34) becomes

$$\lambda+i\psi-i\psi_1=\frac{x}{\nu_1c_1}-\frac{p_2}{2!}\left(\frac{x}{\nu_1c_1}\right)^2-\frac{p_3}{3!}\left(\frac{x}{\nu_1c_1}\right)^3-\cdots-\frac{p_8}{8!}\left(\frac{x}{\nu_1c_1}\right)^8, \tag{7.36}$$

where the p coefficients at the footpoint latitude ϕ_1 have already been given in (7.35). Taking the real and imaginary parts we have

$$\lambda(x,y)=\frac{x}{\nu_1c_1}-\frac{x^3}{3!\,\nu_1^3c_1}V_3-\frac{x^5}{5!\,\nu_1^5c_1}V_5-\frac{x^7}{7!\,\nu_1^7c_1}\overline{V}_7,\quad \text{where } m(\phi_1)=y, \tag{7.37}$$

$$\psi-\psi_1=-\frac{x^2t_1}{2!\,\nu_1^2c_1}-\frac{x^4t_1}{4!\,\nu_1^4c_1}V_4-\frac{x^6t_1}{6!\,\nu_1^6c_1}V_6-\frac{x^8t_1}{8!\,\nu_1^8c_1}\overline{V}_8. \tag{7.38}$$

We see that the spherical approximation has been used only in the last term of each series. Therefore it is equivalent to neglecting terms of order $e^2(x/a)^7$ and $e^2(x/a)^8$. This will be justified when we look at the typical magnitude of such terms.

Finally, we note for future reference the spherical limits of the terms $V_3,\ldots V_6$: since $\beta_1\to 1$ as $e\to 1$ we have

$$\begin{aligned}
V_3\to\overline{V}_3&=\quad 1+2t_1^2,\\
V_4\to\overline{V}_4&=-5-6t_1^2,\\
V_5\to\overline{V}_5&=-5-28t_1^2-24t_1^4,\\
V_6\to\overline{V}_6&=\quad 61+180t_1^2+120t_1^4.
\end{aligned} \tag{7.39}$$

The inverse series for ϕ

In Chapter 6, equation (6.21) we derived the following fourth order Taylor series for the inverse of the Mercator parameter on the ellipsoid:

$$\phi-\phi_1=(\psi-\psi_1)\,\beta_1c_1+\frac{(\psi-\psi_1)^2}{2!}\beta_1c_1^2t_1D_2+\frac{(\psi-\psi_1)^3}{3!}\beta_1c_1^3D_3+\frac{(\psi-\psi_1)^4}{4!}\beta_1c_1^4t_1D_4, \tag{7.40}$$

where the D-coefficients are given in (6.26). All that remains is to substitute for $(\psi - \psi_1)$ using (7.38). It is convenient to use a temporary abbreviation, setting $\widetilde{x} = x/\nu_1$.

$$\begin{aligned}
\psi - \psi_1 &= -\frac{t_1}{c_1}\left[\frac{1}{2!}\widetilde{x}^2 + \frac{1}{4!}V_4\widetilde{x}^4 + \frac{1}{6!}V_6\widetilde{x}^6 + \frac{1}{8!}\overline{V}_8\widetilde{x}^8\right],\\
(\psi - \psi_1)^2 &= \frac{t_1^2}{c_1^2}\left[\frac{1}{2!2!}\widetilde{x}^4 + \frac{2}{2!4!}V_4\widetilde{x}^6 + \frac{2}{2!6!}V_6\widetilde{x}^8 + \frac{1}{4!4!}V_4^2\widetilde{x}^8\right],\\
(\psi - \psi_1)^3 &= -\frac{t_1^3}{c_1^3}\left[\frac{1}{2!2!2!}\widetilde{x}^6 + \frac{3}{2!2!4!}V_4\widetilde{x}^8\right],\\
(\psi - \psi_1)^4 &= \frac{t_1^4}{c_1^4}\left[\frac{1}{2!2!2!2!}\widetilde{x}^8\right] \qquad \text{where } \widetilde{x} = \frac{x}{\nu_1}.
\end{aligned} \tag{7.41}$$

Substituting these expressions into the Taylor series (7.40) gives

$$\begin{aligned}
\phi - \phi_1 = &- \frac{1}{2!}\widetilde{x}^2\beta_1 t_1\,[\,1\,]\\
&- \frac{1}{4!}\widetilde{x}^4\beta_1 t_1\left[V_4 - 3t_1^2 D_2\right]\\
&- \frac{1}{6!}\widetilde{x}^6\beta_1 t_1\left[V_6 - 15t_1^2 D_2 V_4 + 15t_1^2 D_3\right]\\
&- \frac{1}{8!}\widetilde{x}^8\beta_1 t_1\left[\overline{V}_8 - 28t_1^2\overline{D}_2\overline{V}_6 - 35t_1^2\overline{D}_2\overline{V}_4^{\,2} + 210t_1^2\overline{D}_3\overline{V}_4 - 105t_1^4\overline{D}_4\right],
\end{aligned} \tag{7.42}$$

where we use the spherical approximation in evaluating the eighth order term. Substituting for the D-coefficients from equations (6.26, 6.27) and the V-coefficients from equations (7.35,7.39) our final result for ϕ is

$$\phi(x,y) = \quad \phi_1 - \frac{x^2\beta_1 t_1}{2\nu_1^2} - \frac{x^4\beta_1 t_1}{4!\nu_1^4}U_4 - \frac{x^6\beta_1 t_1}{6!\nu_1^6}U_6 - \frac{x^8\beta_1 t_1}{8!\nu_1^8}\overline{U}_8, \tag{7.43}$$

where

$$\begin{aligned}
U_4 &= 4\beta_1^2 - 9\beta_1(1 - t_1^2) - 12t_1^2,\\
U_6 &= 8\beta_1^4(11 - 24t_1^2) - 12\beta_1^3(21 - 71t_1^2) + 15\beta_1^2(15 - 98t_1^2 + 15t_1^4)\\
&\qquad + 180\beta_1(5t_1^2 - 3t_1^4) + 360t_1^4\\
\overline{U}_8 &= -1385 - 3633t_1^2 - 4095t_1^4 - 1575t_1^6.
\end{aligned} \tag{7.44}$$

Later we shall require U_4 and U_6 in the spherical approximation:

$$\begin{aligned}
U_4 \to \overline{U}_4 &= -5 - 3t_1^2,\\
U_6 \to \overline{U}_6 &= \quad 61 + 90t_1^2 + 45t_1^4,
\end{aligned} \tag{7.45}$$

7.3 Summary of Redfearn's transformation formulae

Direct $$x(\lambda,\phi) = \lambda\nu c + \frac{\lambda^3\nu c^3}{3!}W_3 + \frac{\lambda^5\nu c^5}{5!}W_5 + \frac{\lambda^7\nu c^7}{7!}\overline{W}_7 \tag{7.46}$$

$$y(\lambda,\phi) = m(\phi) + \frac{\lambda^2\nu sc}{2} + \frac{\lambda^4\nu sc^3}{4!}W_4 + \frac{\lambda^6\nu sc^5}{6!}W_6 + \frac{\lambda^8\nu sc^7}{8!}\overline{W}_8 \tag{7.47}$$

Inverse $$\lambda(x,y) = \frac{x}{\nu_1 c_1} - \frac{x^3}{3!\,\nu_1^3 c_1}V_3 - \frac{x^5}{5!\,\nu_1^5 c_1}V_5 - \frac{x^7}{7!\,\nu_1^7 c_1}\overline{V}_7 \tag{7.48}$$

$$\phi(x,y) = \phi_1 - \frac{x^2\beta_1 t_1}{2\nu_1^2} - \frac{x^4\beta_1 t_1}{4!\nu_1^4}U_4 - \frac{x^6\beta_1 t_1}{6!\nu_1^6}U_6 - \frac{x^8\beta_1 t_1}{8!\nu_1^8}\overline{U}_8 \tag{7.49}$$

$$\psi(x,y) = \psi_1 - \frac{x^2 t_1}{2!\,\nu_1^2 c_1} - \frac{x^4 t_1}{4!\,\nu_1^4 c_1}V_4 - \frac{x^6 t_1}{6!\,\nu_1^6 c_1}V_6 - \frac{x^8 t_1}{8!\,\nu_1^8 c_1}\overline{V}_8 \tag{7.50}$$

where

$$\begin{aligned}
W_3 &= \beta - t^2\\
W_4 &= 4\beta^2 + \beta - t^2\\
W_5 &= 4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4\\
W_6 &= 8\beta^4(11-24t^2) - 28\beta^3(1-6t^2) + \beta^2(1-32t^2) - 2\beta t^2 + t^4\\
\overline{W}_7 &= 61 - 479t^2 + 179t^4 - t^6 + O(e^2)\\
\overline{W}_8 &= 1385 - 3111t^2 + 543t^4 - t^6 + O(e^2)
\end{aligned} \tag{7.51}$$

$$\begin{aligned}
V_3 &= \beta_1 + 2t_1^2\\
V_4 &= 4\beta_1^2 - 9\beta_1 - 6t_1^2\\
V_5 &= 4\beta_1^3(1-6t_1^2) - \beta_1^2(9-68t_1^2) - 72\beta_1 t_1^2 - 24t_1^4\\
V_6 &= 8\beta_1^4(11-24t_1^2) - 84\beta_1^3(3-8t_1^2) + 225\beta_1^2(1-4t_1^2) + 600\beta_1 t_1^2 + 120t_1^4\\
\overline{V}_7 &= 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6\\
\overline{V}_8 &= -1385 - 7266t_1^2 - 10920t_1^4 - 5040t_1^6
\end{aligned} \tag{7.52}$$

$$\begin{aligned}
U_4 &= 4\beta_1^2 - 9\beta_1(1-t_1^2) - 12t_1^2\\
U_6 &= 8\beta_1^4(11-24t_1^2) - 12\beta_1^3(21-71t_1^2) + 15\beta_1^2(15-98t_1^2+15t_1^4)\\
&\qquad + 180\beta_1(5t_1^2 - 3t_1^4) + 360t_1^4\\
\overline{U}_8 &= -1385 - 3633t_1^2 - 4095t_1^4 - 1575t_1^6
\end{aligned} \tag{7.53}$$

and (a) $m(\phi)$ is the meridian distance which may be calculated from the series (5.71) or (5.77); (b) s, c, t denote $\sin\phi$, $\cos\phi$, $\tan\phi$; (c) the functions $\nu(\phi)$ and $\beta(\phi)$ are defined in equation (5.52, 5.53); (d) λ is measured in radians from the central meridian; (e) the subscripted terms in the inverse series are to be evaluated at the footpoint latitude ϕ_1 such that $m(\phi_1) = y$.

Comments

1. The series given on the previous page are in full accordance with those printed in Redfearn's article in the (Empire) Survey Review. They differ in format since Redfearn writes the series in terms of $\rho(\phi)$ and $\nu(\phi)$ rather than $\beta(\phi)\ \big(=\nu/\rho\big)$ and $\nu(\phi)$. In the original paper Redfearn's formulae a few '1' subscripts are omitted in the inverse series.

2. We shall defer an analysis of the accuracy of the above series until we present the results for scale and convergence in the next chapter. Then in Chapter 9 we discuss the results in the context of two important applications, UTM and NGGB.

3. The TME projection may be modified so that the scale on the central meridian is less than unity, equal to or approximately equal to 0.9996 for UTM and NGGB respectively. Unit scale is achieved on two lines but for TME these are neither meridians nor grid lines. This modification reduces the range of scale variation.

4. For computational purposes the series should be written in a 'nested' form. For example equation (7.46) can be written as

$$x = \nu\widetilde{\lambda}\left[1 + \widetilde{\lambda}^2\left\{\frac{W_3}{3!} + \widetilde{\lambda}^2\left(\frac{W_5}{5!} + \widetilde{\lambda}^2\frac{\overline{W}_7}{7!}\right)\right\}\right], \qquad \widetilde{\lambda} = \lambda c. \tag{7.54}$$

Chapter 8

Scale and convergence in TME

Abstract

Cauchy–Riemann conditions for the inverse. Grid convergence in geographical coordinates. Azimuths and bearings. Grid convergence in projection coordinates. Scale factors. Redfearn series for convergence and scale. Modified TME.

8.1 Cauchy–Riemann conditions for NME to TME

In Chapter 7 we proved that the transformation NME→TME was conformal by checking that the series (7.28, 7.29) satisfied the Cauchy–Riemann equations (4.14). That is,

$$\zeta \to z(\zeta) \equiv x(\lambda,\psi) + iy(\lambda,\psi) : \qquad x_\lambda = y_\psi, \qquad x_\psi = -y_\lambda. \tag{8.1}$$

Since a conformal transformation is angle-preserving then the inverse transformation must also be angle-preserving and conformal, satisfying the following conditions:

$$z \to \zeta(z) \equiv \lambda(x,y) + i\psi(x,y) \quad : \qquad \lambda_x = \psi_y, \qquad \lambda_y = -\psi_x. \tag{8.2}$$

It is straightforward, and useful, to show that the Cauchy–Riemann equations for the inverse transformation follow from (8.1). Consider the identities

$$\begin{aligned} x &= x\big(\lambda(x,y),\, \psi(x,y)\big), \\ y &= y\big(\lambda(x,y),\, \psi(x,y)\big). \end{aligned} \tag{8.3}$$

Differentiate both these identities by x and then both by y, giving four equations:

$$\begin{aligned} 1 &= x_\lambda\lambda_x + x_\psi\psi_x, & 0 &= x_\lambda\lambda_y + x_\psi\psi_y, \\ 0 &= y_\lambda\lambda_x + y_\psi\psi_x, & 1 &= y_\lambda\lambda_y + y_\psi\psi_y. \end{aligned} \tag{8.4}$$

Eliminating ψ_x then λ_x from the left pair and then ψ_y and λ_y from the right pair:

$$y_\psi = J\,\lambda_x, \qquad y_\lambda = -J\,\psi_x, \qquad x_\psi = -J\,\lambda_y, \qquad x_\lambda = J\,\psi_y, \tag{8.5}$$

where

$$\begin{aligned} J &= x_\lambda y_\psi - y_\lambda x_\psi = x_\lambda^2 + y_\lambda^2 = |x_\lambda + iy_\lambda|^2 = |z'(\zeta)|^2 \\ &= \frac{1}{|\zeta'(z)|^2} = \frac{1}{|\lambda_x + i\psi_x|^2} = \frac{1}{\lambda_x^2 + \psi_x^2}. \end{aligned} \tag{8.6}$$

Now neither $z'(\zeta)$ nor $\zeta'(z)$ can vanish, for otherwise the transformation and its inverse would degenerate into z=constant or ζ=constant respectively (so that every $P(\phi, \lambda)$ would map to the same point). Therefore J must be non-zero and bounded and equations (8.5) can be used to show that the equations (8.2) follow from (8.1).

8.2 Azimuths and grid bearings in TME

Following Section 3.5 we consider the geometry of infinitesimal elements on the ellipsoid and corresponding elements on the NME and TME complex planes. Both steps of the projection are conformal and the azimuth α on the ellipsoid is equal to the angle between projected meridian and the projected displacement on both NME and TME planes. The

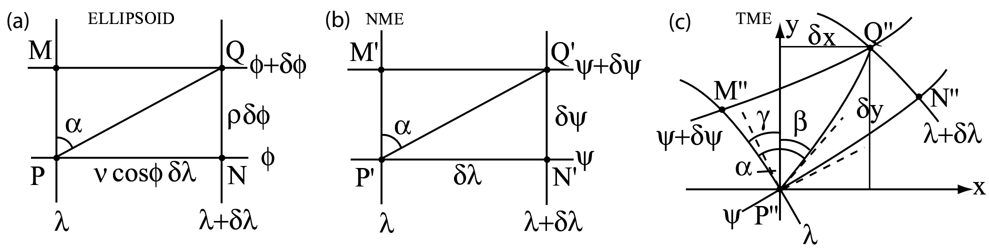


Figure 8.1

grid bearing β on the TME projection is defined as the angle between grid north (on which $\delta x = 0$) and the tangent to the projected displacement $P''Q''$. The TME projection convergence, γ, is defined as the angle between TME grid north and the projected meridian—and because of conformality we see that the displacents $P'M'$, $P'Q'$ and $P'N'$ are all rotated by γ to the corresponding displacements $P''M''$, $P''Q''$ and $P''N''$. With these definitions we see that just as in Section 3.5 we have

$$\boxed{\alpha = \beta + \gamma} \tag{8.7}$$

This equation can be used to find bearings from azimuths or vice versa if we know $\gamma(\lambda, \phi)$ and $\gamma(x, y)$ respectively. That is α=β+$\gamma(x, y)$ or β=α−$\gamma(\phi, \lambda)$.

8.3 Grid convergence in TME

We have defined grid (or projection) convergence, γ, as the angle between true north and grid north on the projection at P'', where true north is defined by the tangent to the projected meridian at P''. Since $\delta\lambda = 0$ on the projected meridian, and remembering that we are now considering x and y as functions of λ and ψ, we have

$$\tan\gamma = -\left.\frac{dx}{dy}\right|_{\delta\lambda=0} = -\lim \left.\frac{x_\lambda\,\delta\lambda + x_\psi\,\delta\psi}{y_\lambda\,\delta\lambda + y_\psi\delta\psi}\right|_{\delta\lambda=0} = -\frac{x_\psi}{y_\psi} = \frac{y_\lambda}{x_\lambda}, \tag{8.8}$$

where we use the Cauchy-Riemann conditions (8.1) in the last step. This last form is to be preferred since the Redfearn formulae (7.46, 7.47) give x and y as power series in λ with coefficients as functions of ϕ. Therefore we have

$$\gamma(\lambda,\phi) = \quad \arctan\left(\frac{y_\lambda}{x_\lambda}\right). \tag{8.9}$$

The calculation of the convergence in projection coordinates follows immediately if we use equations (8.5) to set $y_\lambda = -J\,\psi_x$ and $x_\lambda = J\,\psi_y$:

$$\gamma(x,y) = -\arctan\left(\frac{\psi_x}{\lambda_x}\right), \tag{8.10}$$

where ψ and λ are power series in x (7.50, 7.48) with coefficients evaluated at the footpoint ϕ_1.

8.4 Point scale factors in TME

If we work from first principles (compare equation 2.19) the square of the point scale factor is given by the ratio of the metric distances on the TME projection and the ellipsoid:

$$\mu^2 = \frac{\delta x^2 + \delta y^2}{\rho^2\,\delta\phi^2 + \nu^2\cos^2\!\phi\,\delta\lambda^2}. \tag{8.11}$$

From the derivative of the Mercator parameter on the ellipsoid, equation (6.12), we have that $\rho\,\delta\phi = \nu\cos\phi\,\delta\psi$. Therefore we can write the above as

$$\mu^2 = \frac{a^2}{\nu^2\cos^2\!\phi}\;\frac{\delta x^2 + \delta y^2}{a^2(\delta\psi^2 + \delta\lambda^2)} \;\equiv\; k^2_{\mathrm{NME}} m^2 \tag{8.12}$$

The first of these factors is just the square of the scale for the transformation from the ellipsoid to NME—see equation (6.10). The second factor, which we have defined as m^2, is the square of the scale factor between the NME and TME projections (or the magnification of the conformal transformation between the NME, TME complex planes). Using the Cauchy–Riemann equations (8.1) we can rewrite the numerator as follows:

$$\begin{aligned}\delta x^2 + \delta y^2 &= (x_\psi\delta\psi + x_\lambda\delta\lambda)^2 + (y_\psi\delta\psi + y_\lambda\delta\lambda)^2 && (8.13)\\ &= (x_\psi^2 + y_\psi^2)\delta\psi^2 + 2(x_\psi x_\lambda + y_\psi y_\lambda)\delta\psi\,\delta\lambda + (x_\lambda^2 + y_\lambda^2)\delta\lambda^2 \\ &= (x_\lambda^2 + y_\lambda^2)(\delta\psi^2 + \delta\lambda^2). && (8.14)\end{aligned}$$

Therefore the scale factor can be written as

$$k(\lambda,\phi) = \frac{1}{\nu(\phi)\cos\!\phi}\,\left\{x_\lambda^2 + y_\lambda^2\right\}^{1/2} = \frac{x_\lambda\sec\gamma(\lambda,\phi)}{\nu(\phi)\cos\!\phi}, \tag{8.15}$$

where we use (8.9) in the second step. We have also replaced μ by k, the usual notation for an isotropic scale factor, since x_λ and y_λ are independent of α (depending only on position). Using equations (8.6, 8.10) this result can be transformed to

$$k(x,y) = \frac{1}{\nu(\phi)\cos\!\phi}\,\left\{\lambda_x^2 + \psi_x^2\right\}^{-1/2} = \frac{1}{\nu(\phi)\cos\!\phi}\;\frac{1}{\lambda_x\sec\gamma(x,y)}, \tag{8.16}$$

where it is assumed that we have first calculated $\phi(x,y)$.

Alternative derivations of scale and convergence

In Section 8.2 we remarked that all linear elements in the complex NME plane are rotated by the angle of convergence γ when we transform to the complex TME plane. Therefore if we denote the element $P'Q'$ in the NME complex plane by $\delta\zeta = r\exp[i\theta]$ then the element $P''Q''$ will be represented by $\delta z = R\exp\big[i(\theta+\gamma)\big]$ where $m = R/r$ is the magnification of the element. Since the limit of the ratio of these elements is the value of $z'(\zeta)$ at P we must have

$$x_\lambda + iy_\lambda = z'(\zeta) = \lim \frac{\delta z}{\delta\zeta} = \frac{R}{r}e^{i\gamma} = me^{i\gamma}. \tag{8.17}$$

Therefore

$$\gamma = \arctan\left(\frac{y_\lambda}{x_\lambda}\right), \qquad m = \{x_\lambda^2 + y_\lambda^2\}^{1/2}, \tag{8.18}$$

reproducing the previous definitions in equations (8.9) and (8.15), the latter after multiplication by the scale factor k_{NME} for the transformation from the ellipsoid to NME and after dividing by a factor of a to allow for the fact that we chose (Chapter 4) to label the NME complex plane by (λ,ψ) rather than $(a\lambda, a\psi)$.

8.5 Series for partial derivatives

The expressions for the convergence and scale factors in equations (8.9), (8.10), (8.15) and (8.16) depend on the partial derivatives x_λ, y_λ, λ_x, ψ_x of the Redfearn series (7.46–7.50). Setting $\widetilde\lambda = \lambda c$ and $\widetilde x = x/\nu_1$ we have

$$x_\lambda = \frac{\partial x}{\partial\lambda} = \nu c\left[1 + \frac{1}{2}\widetilde\lambda^2 W_3 + \frac{1}{24}\widetilde\lambda^4 W_5 + \frac{1}{720}\widetilde\lambda^6 \overline{W}_7,\right], \qquad \widetilde\lambda = \lambda c \tag{8.19}$$

$$y_\lambda = \frac{\partial y}{\partial\lambda} = \nu s\,\widetilde\lambda\left[1 + \frac{1}{6}\widetilde\lambda^2 W_4 + \frac{1}{120}\widetilde\lambda^4 W_6 + \frac{1}{5040}\widetilde\lambda^6 \overline{W}_8\right], \tag{8.20}$$

$$\lambda_x = \frac{\partial\lambda}{\partial x} = \frac{1}{\nu_1 c_1}\left[1 - \frac{1}{2}\widetilde x^2 V_3 - \frac{1}{24}\widetilde x^4 V_5 - \frac{1}{720}\widetilde x^6 \overline{V}_7\right], \qquad \widetilde x = \frac{x}{\nu_1} \tag{8.21}$$

$$\psi_x = \frac{\partial\psi}{\partial x} = -\frac{t_1\widetilde x}{\nu_1 c_1}\left[1 + \frac{1}{6}\widetilde x^2 V_4 + \frac{1}{120}\widetilde x^4 V_6 + \frac{1}{5040}\widetilde x^6 \overline{V}_8\right]. \tag{8.22}$$

Note that, apart from the overall multiplicative terms, the series for λ_x is obtained from that for x_λ by the replacements $\widetilde\lambda \to \widetilde x$ and $W_n \to -V_n$ (n odd); the series for ψ_x is obtained from that for y_λ by the replacements $\widetilde\lambda \to \widetilde x$ and $W_n \to V_n$ (n even).

NB. In constructing the Redfearn series we discarded terms of order λ^9 and $(x/a)^9$ so we must discard terms of order λ^8 and $(x/a)^8$ (and higher order) in the derivatives and in any expressions obtained by manipulation of the above series. Moreover the coefficients of $\widetilde\lambda^7$, $\widetilde x^7$, $\widetilde\lambda^8$ and $\widetilde x^8$ terms of the Redfearn series were evaluated in the spherical approximation ($e = 0$, $\beta = 1$); therefore, for consistency, we must use the spherical approximation in coefficients of terms of the order $\widetilde\lambda^6$, $\widetilde x^6$, $\widetilde\lambda^7$ and $\widetilde x^7$ wherever they arise in the manipulation of the series for the derivatives.

The quotient of y_λ and x_λ

Using (E.31) we find the inverse of x_λ in (8.19) to be such that

$$\frac{\nu c}{x_\lambda} = 1 - \widetilde{\lambda}^2 \left(\frac{W_3}{2}\right) - \widetilde{\lambda}^4 \left(\frac{W_5}{24} - \frac{W_3^2}{4}\right) - \widetilde{\lambda}^6 \left(\frac{\overline{W}_7}{720} - \frac{\overline{W}_3\overline{W}_5}{24} + \frac{\overline{W}_3^3}{8}\right). \tag{8.23}$$

Note that the W_3 and W_5 which the inversion casts into the last term have been replaced by their spherical limits. The product of the above with (8.20) gives

$$\frac{y_\lambda}{x_\lambda} = t\,\widetilde{\lambda}\left[1 + a_2\widetilde{\lambda}^2 + a_4\widetilde{\lambda}^4 + \bar{a}_6\widetilde{\lambda}^6\right], \tag{8.24}$$

where the coefficients (and required spherical limits) are calculated using (7.21, 7.22)

$$a_2 = \frac{W_4}{6} - \frac{W_3}{2} \quad = \frac{1}{3}\left[2\beta^2 - \beta + t^2\right], \qquad \bar{a}_2 = \frac{1}{3}\left(1 + t^2\right)$$

$$a_4 = \frac{W_6}{120} - \frac{W_3 W_4}{12} + \frac{W_3^2}{4} - \frac{W_5}{24},$$

$$= \frac{1}{15}\left[\beta^4(11-24t^2) - \beta^3(11-36t^2) + \beta^2(2-4t^2) - 4\beta t^2 + 2t^4\right], \quad \bar{a}_4 = \frac{1}{15}\left(2+4t^2+2t^4\right)$$

$$\bar{a}_6 = \frac{\overline{W}_8}{5040} - \frac{\overline{W}_3\overline{W}_6}{240} + \frac{\overline{W}_3^{\,2}\overline{W}_4}{24} - \frac{\overline{W}_4\overline{W}_5}{144} - \frac{\overline{W}_3^{\,3}}{8} + \frac{\overline{W}_3\overline{W}_5}{24} - \frac{\overline{W}_7}{720}$$

$$= \frac{1}{315}\left[17 + 51t^2 + 51t^4 + 17t^6\right]. \tag{8.25}$$

The quotient of ψ_x and λ_x

Bearing in mind the comment made immediately after equation (8.22) we see that if we define

$$\frac{\psi_x}{\lambda_x} = -t_1\,\widetilde{x}\,\left[1 + r_2\widetilde{x}^2 + r_4\widetilde{x}^4 + \bar{r}_6\widetilde{x}^6\right] \tag{8.26}$$

then the coefficients follow by analogy with equations (8.25). Using (7.35, 7.39) we find that the r-coefficients and their spherical limits are

$$r_2 = \frac{V_4}{6} + \frac{V_3}{2} \quad = \frac{1}{3}\left[2\beta_1^2 - 3\beta_1\right], \qquad \bar{r}_2 = -\frac{1}{3}$$

$$r_4 = \frac{V_6}{120} + \frac{V_3 V_4}{12} + \frac{V_3^2}{4} + \frac{V_5}{24},$$

$$= \frac{1}{15}\left[\beta_1^4(11-24t_1^2) - 3\beta_1^3(8-23t_1^2) + 15\beta_1^2(1-4t_1^2) + 15\beta_1 t_1^2\right], \qquad \bar{r}_4 = \frac{2}{15}$$

$$\bar{r}_6 = \frac{\overline{V}_8}{5040} + \frac{\overline{V}_3\overline{V}_6}{240} + \frac{\overline{V}_3^{\,2}\overline{V}_4}{24} + \frac{\overline{V}_4\overline{V}_5}{144} + \frac{\overline{V}_3^{\,3}}{8} + \frac{\overline{V}_3\overline{V}_5}{24} + \frac{\overline{V}_7}{720} = -\frac{17}{315}. \tag{8.27}$$

Note the absence of terms in t^2, t^4 or t^6 in the coefficients $\bar{r}_2$, $\bar{r}_4$ or $\bar{r}_6$.

8.6 Convergence in geographical coordinates

From (8.9) and (8.24) we have

$$\tan\gamma(\phi,\lambda) = \widetilde{\lambda} t\left[1 + a_2\widetilde{\lambda}^2 + a_4\widetilde{\lambda}^4 + \bar{a}_6\widetilde{\lambda}^6\right] \tag{8.28}$$

and we calculate γ as $\arctan(\tan\gamma)$ by using the series (E.20):

$$\gamma = \tan\gamma - \frac{1}{3}\tan^3\!\gamma + \frac{1}{5}\tan^5\!\gamma - \frac{1}{7}\tan^7\!\gamma + \cdots . \tag{8.29}$$

To order $\widetilde{\lambda}^7$ the higher powers of $\tan\gamma$ are given by

$$\begin{aligned}
&\tan^2\!\gamma = \widetilde{\lambda}^2 t^2\left[1 + 2a_2\widetilde{\lambda}^2 + (2\bar{a}_4 + \bar{a}_2^2)\widetilde{\lambda}^4\right] && \tan^5\!\gamma = \widetilde{\lambda}^5 t^5\left[1 + 5\bar{a}_2\widetilde{\lambda}^2\right]\\
&\tan^3\!\gamma = \widetilde{\lambda}^3 t^3\left[1 + 3a_2\widetilde{\lambda}^2 + 3(\bar{a}_4 + \bar{a}_2^2)\widetilde{\lambda}^4\right] && \tan^6\!\gamma = \widetilde{\lambda}^6 t^6\,[\,1\,]\\
&\tan^4\!\gamma = \widetilde{\lambda}^4 t^4\left[1 + 4\bar{a}_2\widetilde{\lambda}^2\right] && \tan^7\!\gamma = \widetilde{\lambda}^7 t^7\,[\,1\,]
\end{aligned} \tag{8.30}$$

so that

$$\gamma = \widetilde{\lambda} t + \frac{\widetilde{\lambda}^3 t}{3}\left[3a_2 - t^2\right] + \frac{\widetilde{\lambda}^5 t}{5}\left[5a_4 - 5a_2 t^2 + t^4\right] + \frac{\widetilde{\lambda}^7 t}{7}\left[7\bar{a}_6 - 7\big(\bar{a}_4 + \bar{a}_2^2\big)t^2 + 7\bar{a}_2 t^4 - t^6\right]. \tag{8.31}$$

Substituting the a-coefficients from (8.25) gives our final result

$$\gamma(\phi,\lambda) = \widetilde{\lambda} t + \frac{1}{3}\widetilde{\lambda}^3 t\,H_3 + \frac{1}{15}\widetilde{\lambda}^5 t\,H_5 + \frac{1}{315}\widetilde{\lambda}^7 t\,\bar{H}_7, \qquad \widetilde{\lambda} = \lambda c. \tag{8.32}$$

$$\begin{aligned}
H_3 &= 2\beta^2 - \beta,\\
H_5 &= \beta^4(11 - 24t^2) - \beta^3(11 - 36t^2) + \beta^2(2 - 14t^2) + \beta t^2,\\
\bar{H}_7 &= 17 - 26t^2 + 2t^4.
\end{aligned} \tag{8.33}$$

8.7 Convergence in projection coordinates

From (8.10) and (8.26) and

$$\tan\gamma(x,y) = t_1\widetilde{x}\left[1 + r_2\widehat{x}^2 + r_4\widehat{x}^4 + \bar{r}_6\widehat{x}^6\right]. \tag{8.34}$$

Comparing this equation with (8.28) we see from equation (8.31) that

$$\gamma = \widetilde{x} t_1 + \frac{\widetilde{x}^3 t_1}{3}\left[3r_2 - t_1^2\right] + \frac{\widetilde{x}^5 t_1}{5}\left[5r_4 - 5r_2 t_1^2 + t_1^4\right] + \frac{\widetilde{x}^7 t_1}{7}\left[7\bar{r}_6 - 7\big(\bar{r}_4 + \bar{r}_2^2\big)t_1^2 + 7\bar{r}_2 t_1^4 - t_1^6\right]. \tag{8.35}$$

Substituting the r-coefficients from (8.27) gives our final result

$$\gamma(x,y) = \widetilde{x}t_1 + \frac{1}{3}\widetilde{x}^3 t_1 K_3 + \frac{1}{15}\widetilde{x}^5 t_1 K_5 + \frac{1}{315}\widetilde{x}^7 t_1 K_7, \qquad \widetilde{x} = \frac{x}{\nu_1}, \tag{8.36}$$

where

$$\begin{aligned} K_3 &= 2\beta_1^2 - 3\beta_1 - t_1^2, \\ K_5 &= \beta_1^4(11 - 24t_1^2) - 3\beta_1^3(8 - 23t_1^2) + 5\beta_1^2(3 - 14t_1^2) + 30\beta_1 t_1^2 + 3t_1^4, \\ \bar{K}_7 &= -17 - 77t_1^2 - 105t_1^4 - 45t_1^6. \end{aligned} \tag{8.37}$$

Setting $\beta_1 = \nu_1/\rho_1$ gives the Redfearn series except that he writes ν rather than a more correct ν_1 *etc.*

8.8 Scale factor in geographical coordinates

From (8.15) we have

$$k(\lambda,\phi) = \frac{x_\lambda \sec\gamma(\lambda,\phi)}{\nu\cos\phi}\,, \tag{8.38}$$

where x_λ is given in equation (8.19) and we evaluate $\sec\gamma$ by using the binomial series (E.28) and the expressions for $\tan^n\gamma$ given in (8.30):

$$\frac{x_\lambda}{\nu c} = 1 + \frac{1}{2}\widetilde{\lambda}^2 W_3 + \frac{1}{24}\widetilde{\lambda}^4 W_5 + \frac{1}{720}\widetilde{\lambda}^6 \overline{W}_7, \tag{8.39}$$

$$\begin{aligned} \sec\gamma &= \left\{1 + \tan^2\gamma\right\}^{1/2} = 1 + \frac{1}{2}\tan^2\gamma - \frac{1}{8}\tan^4\gamma + \frac{1}{16}\tan^6\gamma + \cdots \\ &= 1 + \frac{1}{2}\widetilde{\lambda}^2 t^2 + \frac{1}{8}\widetilde{\lambda}^4\left[8a_2t^2 - t^4\right] + \frac{1}{16}\widetilde{\lambda}^6\left[16\bar{a}_4t^2 + 8\bar{a}_2^2t^2 - 8\bar{a}_2t^4 + t^6\right] \end{aligned} \tag{8.40}$$

$$\begin{aligned} k(\phi,\lambda) &= 1 + \frac{1}{2}\widetilde{\lambda}^2\big(W_3 + t^2\big) + \frac{1}{24}\widetilde{\lambda}^4\big(W_5 + 6t^2W_3 + 24a_2t^2 - 3t^4\big) \\ &\quad + \frac{\widetilde{\lambda}^6}{720}\big(\overline{W}_7 + 15t^2\overline{W}_5 + 360\bar{a}_2t^2\overline{W}_3 - 45t^4\overline{W}_3 + 720\bar{a}_4t^2 + 360\bar{a}_2^2t^2 - 360\bar{a}_2t^4 + 45t^6\big) \end{aligned} \tag{8.41}$$

Finally, using the W-coefficients from (7.21 7.22) and the a-coefficients from (8.25),

$$k(\lambda,\phi) = 1 + \frac{1}{2}\widetilde{\lambda}^2 H_2 + \frac{1}{24}\widetilde{\lambda}^4 H_4 + \frac{1}{720}\widetilde{\lambda}^6 \bar{H}_6, \qquad \widetilde{\lambda} = \lambda c, \tag{8.42}$$

$$\begin{aligned} H_2 &= \beta \\ H_4 &= 4\beta^3(1 - 6t^2) + \beta^2(1 + 24t^2) - 4\beta t^2 \\ \bar{H}_6 &= 61 - 148t^2 + 16t^4. \end{aligned} \tag{8.43}$$

8.9 Scale factor in projection coordinates

From (8.16) we have

$$k(x,y) = \frac{1}{\nu\cos\phi}\,\frac{1}{\lambda_x \sec\gamma}. \tag{8.44}$$

Ignoring the factor of $\nu\cos\phi$ for the moment we follow the same steps as in the previous section but calculate $\sec\gamma$ using the expression for $\tan\gamma$ given in equation (8.34):

$$\lambda_x = \frac{1}{\nu_1 c_1}\left[1 - \frac{1}{2}\widetilde{x}^2 V_3 - \frac{1}{24}\widetilde{x}^4 V_5 - \frac{1}{720}\widetilde{x}^6 \overline{V}_7,\right], \tag{8.45}$$

$$\begin{aligned}\sec\gamma &= \left\{1+\tan^2\gamma\right\}^{1/2} = 1 + \frac{1}{2}\tan^2\gamma - \frac{1}{8}\tan^4\gamma + \frac{1}{16}\tan^6\gamma + \cdots \\ &= 1 + \frac{1}{2}\widetilde{x}^2 t_1^2 + \frac{1}{8}\widetilde{x}^4\left[8r_2 t_1^2 - t_1^4\right] + \frac{1}{16}\widetilde{x}^6\left[16\bar{r}_4 t_1^2 + 8\bar{r}_2^2 t_1^2 - 8\bar{r}_2 t_1^4 + t_1^6\right],\end{aligned} \tag{8.46}$$

$$\begin{aligned}\lambda_x\sec\gamma &= \frac{1}{\nu_1 c_1}\Big[1 + \frac{1}{2}\widetilde{x}^2\big(-V_3 + t_1^2\big) + \frac{1}{24}\widetilde{x}^4\big(-V_5 - 6t_1^2 V_3 + 24 r_2 t_1^2 - 3t_1^4\big) \\ &\quad + \frac{\widetilde{x}^6}{720}\big(-\overline{V}_7 - 15t_1^2\overline{V}_5 - 360\bar{r}_2 t_1^2\overline{V}_3 + 45t_1^4\overline{V}_3 + 720\bar{r}_4 t_1^2 + 360\bar{r}_2^2 t_1^2 - 360\bar{r}_2 t_1^4 + 45t_1^6\big)\Big] \\ &= \frac{1}{\nu_1 c_1}\left[1 + \frac{1}{2}\widetilde{x}^2 p_2 + \frac{1}{24}\widetilde{x}^4 p_4 + \frac{1}{720}\widetilde{x}^6\bar{p}_6\right],\end{aligned} \tag{8.47}$$

where the coefficients and their spherical limits are evaluated using (7.35, 7.39) and (8.27):

$$\begin{aligned} p_2 &= -\beta_1 - t_1^2 & \bar{p}_2 &= -1 - t_1^2 \\ p_4 &= -4\beta_1^3(1-6t_1^2) + \beta_1^2(9-52t_1^2) + 42\beta_1 t_1^2 + 9t_1^4 & \bar{p}_4 &= 5 + 14t_1^2 + 9t_1^4 \\ \bar{p}_6 &= -61 - 331t_1^2 - 495t_1^4 - 225t_1^6. \end{aligned} \tag{8.48}$$

The factor $\nu\cos\phi$

We expand $f(\phi) = \nu(\phi)\cos\phi$ in a Taylor series about the footpoint latitude ϕ_1:

$$f(\phi) = f(\phi_1) + (\phi-\phi_1)f'(\phi_1) + \frac{1}{2!}(\phi-\phi_1)^2 f''(\phi_1) + \frac{1}{3!}(\phi-\phi_1)^3 f'''(\phi_1). \tag{8.49}$$

The series terminates with the third order term since equation (7.49) shows that $(\phi-\phi_1)^4$ is of order $\widetilde{x}^8$ and is therefore neglected; further $(\phi-\phi_1)^3$ is of order $\widetilde{x}^6$ so the third derivative must be evaluated in the spherical limit, ($\nu = a,\ \nu' = \nu'' = 0$). Therefore

$$f(\phi) = \nu c, \quad f'(\phi) = \nu' c - \nu s, \quad f''(\phi) = \nu'' c - 2\nu' s - \nu c, \quad f'''(\phi) = as,$$

and

$$\nu\cos\phi = \nu_1 c_1 + [\nu_1' c_1 - \nu_1 s_1](\phi-\phi_1) + \frac{1}{2!}[\nu_1'' c_1 - 2\nu_1' s_1 - \nu_1 c_1](\phi-\phi_1)^2 + \frac{1}{3!}[a s_1](\phi-\phi_1)^3. \tag{8.50}$$

Divide by $\nu_1 c_1$ and using the expressions for ν'/ν and ν''/ν given in (5.55):

$$\frac{\nu\cos\phi}{\nu_1 c_1} = 1 - \frac{t_1}{\beta_1}(\phi-\phi_1) - \frac{1}{2\beta_1^2}(\beta_1 + 3\beta_1 t_1^2 - 3t_1^2)(\phi-\phi_1)^2 + \frac{t_1}{6}(\phi-\phi_1)^3. \quad (8.51)$$

Substituting for $(\phi-\phi_1)$ from (7.49) we have

$$\frac{\nu\cos\phi}{\nu_1 c_1} = 1 - \frac{t_1}{\beta_1}\left[-\frac{1}{2}\beta_1 t_1 \widetilde{x}^2 - \frac{1}{24}\beta_1 t_1 \widetilde{x}^4 U_4 - \frac{1}{720} t_1 \widetilde{x}^6 \overline{U}_6\right]$$
$$- \frac{1}{2\beta_1^2}(\beta_1 + 3\beta_1 t_1^2 - 3t_1^2)\left[\frac{1}{4}\beta_1^2 t_1^2 \widetilde{x}^4 + \frac{1}{24} t_1^2 \widetilde{x}^6 \overline{U}_4\right] + \frac{t_1}{6}\left[-\frac{1}{8} t_1^3 \widetilde{x}^6\right]. \quad (8.52)$$

Using the U-coefficients in (7.44, 7.45) we find

$$\frac{\nu\cos\phi}{\nu_1 c_1} = 1 + \frac{1}{2}\widetilde{x}^2 q_2 + \frac{1}{24}\widetilde{x}^4 q_4 + \frac{1}{720}\widetilde{x}^6 \bar{q}_6, \quad (8.53)$$

$$\begin{aligned}
q_2 &= t_1^2 & \bar{q}_2 &= t_1^2,\\
q_4 &= t_1^2 U_4 - 3t_1^2(\beta_1 + 3\beta_1 t_1^2 - 3t_1^2) = 4\beta_1^2 t_1^2 - 12\beta_1 t_1^2 - 3t_1^4 & \bar{q}_4 &= -8t_1^2 - 3t_1^4,\\
\bar{q}_6 &= t_1^2 \overline{U}_6 - 15 t_1^2 \overline{U}_4 - 15 t_1^4 = 136 t_1^2 + 120 t_1^4 + 45 t_1^6, & & \quad (8.54)
\end{aligned}$$

The scale factor

Multiplying the series (8.47) and (8.53) gives the product

$$\nu\cos\phi\,\lambda_x \sec\gamma = 1 + \frac{1}{2}\widetilde{x}^2 g_2 + \frac{1}{24}\widetilde{x}^4 g_4 + \frac{1}{720}\widetilde{x}^6 \bar{g}_6, \quad (8.55)$$

$$\begin{aligned}
g_2 &= p_2 + q_2 = -\beta_1 & \bar{g}_2 &= -1\\
g_4 &= p_4 + q_4 + 6p_2 q_2 = -4\beta_1^3(1-6t_1^2) + \beta_1^2(9-48t_1^2) + 42\beta_1 t_1^2 & \bar{g}_4 &= 5\\
\bar{g}_6 &= \bar{p}_6 + \bar{q}_6 + 15(\bar{p}_2\bar{q}_4 + \bar{q}_2\bar{p}_4) = -61. & & \quad (8.56)
\end{aligned}$$

The actual scale factor (8.44) is the inverse of (8.55); using (E.32) we have,

$$k(x,y) = 1 + \frac{1}{2}\widetilde{x}^2 K_2 + \frac{1}{24}\widetilde{x}^4 K_4 + \frac{1}{720}\widetilde{x}^6 \bar{K}_6, \qquad \widetilde{x} = \frac{x}{\nu_1} \quad (8.57)$$

where

$$\begin{aligned}
K_2 &= -g_2 &&= \beta_1\,,\\
K_4 &= -g_4 + 6g_2^2 &&= 4\beta_1^3(1-6t_1^2) - 3\beta_1^2(1-16t_1^2) - 24\beta_1 t_1^2,\\
\bar{K}_6 &= -\bar{g}_6 + 30\bar{g}_2\bar{g}_4 - 90\bar{g}_2^3 = 1. && \quad (8.58)
\end{aligned}$$

Setting $\beta_1 = \nu_1/\rho_1$ gives the Redfearn series except that in the sixth order term he has a denominator of $\nu_1^3\rho_1^3$ as against ν_1^6 here. Since we assume the spherical limit for this term both could be replaced by a^6 and there is no inconsistency.

8.10 Final results: Redfearn's modified TME series

Direct series

As for NMS, TMS and NME simply multiply (7.46) and (7.47) by a factor of k_0.

$$x(\lambda,\phi) = k_0\nu\left[\widetilde{\lambda} + \frac{\widetilde{\lambda}^3}{3!}W_3 + \frac{\widetilde{\lambda}^5}{5!}W_5 + \frac{\widetilde{\lambda}^7}{7!}\overline{W}_7\right], \qquad \widetilde{\lambda} = \lambda c \tag{8.59}$$

$$y(\lambda,\phi) = k_0\left[m(\phi) + \frac{\widetilde{\lambda}^2\nu t}{2} + \frac{\widetilde{\lambda}^4\nu t}{4!}W_4 + \frac{\widetilde{\lambda}^6\nu t}{6!}W_6 + \frac{\widetilde{\lambda}^8\nu t}{8!}\overline{W}_8\right]. \tag{8.60}$$

Inverse series:

Set $x{\to}x/k_0$ and replace $\widetilde{x} = x/\nu_1$ by $\widehat{x} = x/k_0\nu_1$. The **footpoint latitude**, ϕ_1, must be found from (7.4) or (7.5) with $y{\to}y/k_0$: it is the solution of $m(\phi_1) = y/k_0$. Equations 7.48, 7.50 and 7.49 become

$$\lambda(x,y) = \frac{\widehat{x}}{c_1} - \frac{\widehat{x}^3}{3!\,c_1}V_3 - \frac{\widehat{x}^5}{5!\,c_1}V_5 - \frac{\widehat{x}^7}{7!\,c_1}\overline{V}_7, \qquad \widehat{x} = \frac{x}{k_0\nu_1} \tag{8.61}$$

$$\psi(x,y) = \psi_1 - \frac{\widehat{x}^2 t_1}{2\,c_1} - \frac{\widehat{x}^4 t_1}{4!\,c_1}V_4 - \frac{\widehat{x}^6 t_1}{6!\,c_1}V_6 - \frac{\widehat{x}^8 t_1}{8!\,c_1}\overline{V}_8. \qquad m(\phi_1) = \frac{y}{k_0} \tag{8.62}$$

$$\phi(x,y) = \phi_1 - \frac{\widehat{x}^2\beta_1 t_1}{2} - \frac{\widehat{x}^4\beta_1 t_1}{4!}U_4 - \frac{\widehat{x}^6\beta_1 t_1}{6!}U_6 - \frac{\widehat{x}^8\beta_1 t_1}{8!}\overline{U}_8, \tag{8.63}$$

Scale and convergence.

The calculations of the present chapter may be applied to the modified series above. Clearly the derivatives x_λ, y_λ pickup a factor of k_0 and the derivatives λ_x, ψ_x pickup a factor of $1/k_0$. The modified forms of 8.42, 8.32, 8.57 and 8.36 are

$$k(\lambda,\phi) = k_0\left[1 + \frac{1}{2}\widetilde{\lambda}^2 H_2 + \frac{1}{24}\widetilde{\lambda}^4 H_4 + \frac{1}{720}\widetilde{\lambda}^6\overline{H}_6\right], \qquad \widetilde{\lambda} = \lambda c, \tag{8.64}$$

$$\gamma(\lambda,\phi) = \widetilde{\lambda}t + \frac{1}{3}\widetilde{\lambda}^3 t\,H_3 + \frac{1}{15}\widetilde{\lambda}^5 t\,H_5 + \frac{1}{315}\widetilde{\lambda}^7 t\,\overline{H}_7, \tag{8.65}$$

$$k(x,y) = k_0\left[1 + \frac{1}{2}\widehat{x}^2 K_2 + \frac{1}{24}\widehat{x}^4 K_4 + \frac{1}{720}\widehat{x}^6\overline{K}_6\right], \qquad m(\phi_1) = \frac{y}{k_0}, \tag{8.66}$$

$$\gamma(x,y) = \widehat{x}t_1 + \frac{1}{3}\widehat{x}^3 t_1 K_3 + \frac{1}{15}\widehat{x}^5 t_1 K_5 + \frac{1}{315}\widehat{x}^7 t_1\overline{K}_7, \qquad \widehat{x} = \frac{x}{k_0\nu_1}. \tag{8.67}$$

Note 1: both series for the scale factor give a value of k_0 on the central meridian.

Note 2: as usual $c{=}\cos\phi$, $t{=}\tan\phi$, $\beta{=}\nu(\phi)/\rho(\phi)$ from (5.53) and the '1' subscript denotes a function evaluated at the footpoint latitude. For convenience all of the required coefficients are collected on the following page.

All coefficients

$$
\begin{aligned}
&x(\lambda,\phi) && W_3 = \beta - t^2\\
& && W_5 = 4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4\\
& && \overline{W}_7 = 61 - 479t^2 + 179t^4 - t^6 + O(e^2)\\[6pt]
&y(\lambda,\phi) && W_4 = 4\beta^2 + \beta - t^2\\
& && W_6 = 8\beta^4(11{-}24t^2) - 28\beta^3(1{-}6t^2) + \beta^2(1{-}32t^2) - 2\beta t^2 + t^4\\
& && \overline{W}_8 = 1385 - 3111t^2 + 543t^4 - t^6 + O(e^2)\\[6pt]
&\lambda(x,y) && V_3 = \beta_1 + 2t_1^2\\
& && V_5 = 4\beta_1^3(1-6t_1^2) - \beta_1^2(9-68t_1^2) - 72\beta_1 t_1^2 - 24t_1^4\\
& && \overline{V}_7 = 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6\\[6pt]
&\psi(x,y) && V_4 = 4\beta_1^2 - 9\beta_1 - 6t_1^2\\
& && V_6 = 8\beta_1^4(11{-}24t_1^2){-}84\beta_1^3(3{-}8t_1^2){+}225\beta_1^2(1{-}4t_1^2){+}600\beta_1 t_1^2{+}120t_1^4\\
& && \overline{V}_8 = -1385 - 7266t_1^2 - 10920t_1^4 - 5040t_1^6\\[6pt]
&\phi(x,y) && U_4 = 4\beta_1^2 - 9\beta_1(1-t_1^2) - 12t_1^2\\
& && U_6 = 8\beta_1^4(11-24t_1^2) - 12\beta_1^3(21-71t_1^2) + 15\beta_1^2(15-98t_1^2+15t_1^4)\\
& && \qquad\qquad + 180\beta_1(5t_1^2 - 3t_1^4) + 360t_1^4\\
& && \overline{U}_8 = -1385 - 3633t_1^2 - 4095t_1^4 - 1575t_1^6\\[6pt]
&k(\lambda,\phi) && H_2 = \beta\\
& && H_4 = 4\beta^3(1-6t^2) + \beta^2(1+24t^2) - 4\beta t^2\\
& && \bar{H}_6 = 61 - 148t^2 + 16t^4\\[6pt]
&\gamma(\lambda,\phi) && H_3 = 2\beta^2 - \beta\\
& && H_5 = \beta^4(11-24t^2) - \beta^3(11-36t^2) + \beta^2(2-14t^2) + \beta t^2\\
& && \bar{H}_7 = 17 - 26t^2 + 2t^4\\[6pt]
&k(x,y) && K_2 = \beta_1\\
& && K_4 = 4\beta_1^3(1-6t_1^2) - 3\beta_1^2(1-16t_1^2) - 24\beta_1 t_1^2\\
& && \bar{K}_6 = 1\\[6pt]
&\gamma(x,y) && K_3 = 2\beta_1^2 - 3\beta_1 - t_1^2\\
& && K_5 = \beta_1^4(11-24t_1^2) - 3\beta_1^3(8-23t_1^2) + 5\beta_1^2(3-14t_1^2) + 30\beta_1 t_1^2 + 3t_1^4\\
& && \bar{K}_7 = -17 - 77t_1^2 - 105t_1^4 - 45t_1^6
\end{aligned}
\tag{8.68}
$$

Chapter 9

The UTM and NGGB projections

Abstract

A review of projections, grids and origins. UTM and NGGB projections. Numerical discussion of the variation of scale and convergence. The accuracy of the Redfearn series. Approximations to the series. The OSGB series.

9.1 Projections, grids and origins

So far we have rather casually mixed the terms ‘projection coordinates’ and ‘grid coordinates’. We must now be a little more precise for the two are logically distinct and rarely equal. In fact we must consider four reference systems:

- Geographic coordinates on an ellipsoid (of revolution).
- The projection coordinates in the plane.
- A set of grid coordinates relative to true and false origins.
- An alpha-numeric grid reference system.

We will discuss each of these points in turn with examples relating to the National Grid of Great Britain (NGGB) and the Universal Transverse Mercator (UTM).

Geographic coordinates

The basic data of any survey are the latitude and longitude of all features of interest. It is important to remember that in each survey the (ϕ, λ) coordinates are defined with respect to one particular ellipsoid of revolution. For example the NGGB projection is based on the Airy 1830 ellipsoid and UTM is based on the International 1924 (aka Hayford 1909) ellipsoid. Remember that there is no such thing as an ‘absolute’ value of latitude or longitude—the same location has different geographic coordinates with respect to different ellipsoids.

Projection coordinates

We consider here only the modified Transverse Mercator projection coordinates derived from the latitude and longitude values and the parameters of the chosen ellipsoid by the Redfearn formula for $x(\lambda, \phi)$ and $y(\lambda, \phi)$ in equations (8.59, 8.60). These formulae give Cartesian coordinates with a **projection origin** where the central meridian crosses the equator. Our convention (in agreement with Snyder) is that the positive y-axis is directed northwards along the projection of the central meridian and the positive x-axis is eastwards along the projection of the equator. (Beware of alternative conventions. For example the original paper of Redfearn, and most 'continental' texts, have x and y interchanged).

In the introduction we called the Cartesian representation defined by equations (8.59) and (8.60) a 'super-map'. From (8.59) we see that if we restrict λ to be less than say 5°, or about 0.1 radians, then the x-range of this super-map will of the order 10^6 metres and from (8.60) we see that the distance between the equator and pole on the central meridian is $k_0 m(\pi/2) \approx 10^7$ metres. Actual printed maps are then obtained by scaling the super-map by the representation factor (RF). For example the Landranger maps produced by the Ordnance Survey of Great Britain (OSGB) have an RF of 1:50000 so that a sheet which measure 80cm square represents an area approximately 40km square on the super-map or on the ground. The caveat 'approximately' is necessary since we know that the scale in the TME projection is the complicated function given by equation (8.64), approximately unity when close to the central meridian.

Note that the maps produced in this way need not be embellished by lines of any kind. There is no obligation to show the x- and y-axes or the origin of the projection coordinates. There is also no obligation to show any lines marking constant values of x and y. There is no obligation to show lines (curves) of constant latitude or longitude.

Grid coordinates: true and false origins

It is of course *useful* to to overlay the map with some kind of reference grid system so that we can refer precisely to the locations of features of interest but there is no necessity that the grid should coincide with the Cartesian system used in the construction of the map. For example we might wish to superimpose a polar grid centred on some point so that positions were related to that point by a distance and a direction. Or a surveyor might construct a large scale map overlain with a grid aligned to some prominent feature such as a river or highway. Thus we see that grids are an arbitrary addition to the map but, if present, we must know how to relate them to the underlying projection and geographic coordinates. In practice, most grids *are* taken as Cartesian coordinate systems aligned to the projection coordinates. This is certainly the case for both UTM and NGGB but the origins of these grids are chosen differently. For UTM the grid origin is chosen to be coincident with the projection origin but, for NGGB, it is taken at latitude 49°N on the central meridian. These are examples of the choice of the **true origin** for a grid.

Now TM projection coordinates are not positive everywhere; they are negative both west of the central meridian and south of the projection origin. It follows that the grid coordinates referred to a true origin on the central meridian will also take positive and negative values. This is unsatisfactory for many practical applications and we therefore introduce a **false origin** of the grid at a position such that the grid coordinates relative to that point are positive throughout the region of interest. These positive coordinates, rounded to the nearest metre, are called **Eastings** (E) and **Northings** (N). The choice of false origin for UTM and NGGB is described in the next sections.

9.2 The UTM projections

UTM is a collection of 60 distinct modified TME projections based on the International 1924 (aka Hayford 1909) ellipsoid; each is of 6° width in longitude and their central meridians are at $\lambda_0 = -177, -171, -165\ldots$. The figure, which is not to scale, shows zone number 30 which includes my home city of Edinburgh at H: the projection is centred on 3°W and it covers the region between parallels at 84°N and 80°S and between the meridians at 6°W and Greenwich. The actual aspect ratio of the grid overlying the projection can be appreciated from the box shown in Figure 3.5; the difference between TMS and TME is not visible at such a small scale.

The true origin T of the UTM grid coincides with the origin of the projection O where the central meridian meets the equator at 3°W. We treat the hemispheres differently and introduce *two* false origins, both are 500km west of the true origin but one, for the northern hemisphere, is at F_1 on the equator and the other, for the southern hemisphere, is at a point F_2 10000km below the equator on the projection. Therefore the eastings and northings for points in the northern and southern hemispheres are

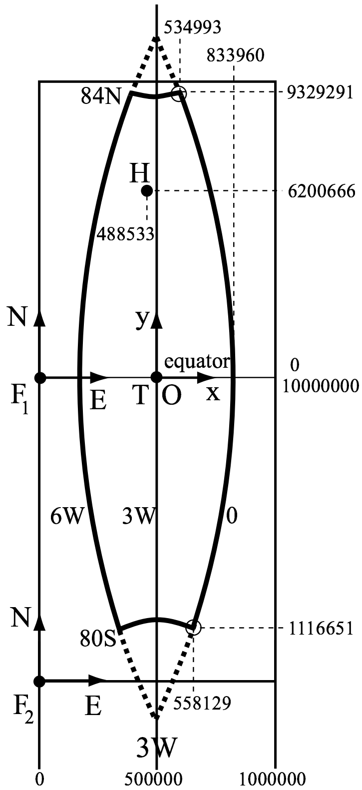


Figure 9.1

$$E = E0 + x(\lambda, \phi), \qquad E0 = 500000 \tag{9.1}$$

$$N = N0 + y(\lambda, \phi) \qquad N0 = \begin{cases} 0 & \text{(N)} \\ 10000000 & \text{(S)} \end{cases} \tag{9.2}$$

where x, y are calculated (to the nearest metre) using the modified TME series (8.59, 8.60) taking λ relative to 3°W (in radians) and k_0 =0.9996. Note that $(E0, N0)$ are the coordinates of the true origin (at $x = y = 0$) relative to the false origin on each grid. The figure shows the eastings and northings of several points on the perimeter of the zone: (a) at $\lambda = 0$ on the equator we have $E = 833960$, $N = 0$; (b) at $\lambda = 0$ on the parallel at 84°N we have $E = 534993$, $N = 9329291$. Thus the width of the UTM zone at the equator is approximately (recall $k_0 \neq 1$) 668km and at latitude 84°N it is approximately 70km.

The inverse relations are the series (8.61, 8.63) with the replacement of x by $E - E0$ and coefficients evaluated at a footpoint latitude such that

$$m(\phi_1) = \frac{y}{k_0} = \frac{N - N0}{k_0}. \tag{9.3}$$

Grid reference systems for UTM

Eastings and Northings are measured in metres and provide a purely numeric grid reference system. For example the grid reference of my armchair is E=488533, N=6200666 (in zone 30 north). This is a rather unwieldy system because we have to give large numbers: in the northern hemisphere E lies between 166040 and 833960 and N lies between 0 and 9329291. To avoid such large numbers the grid reference system has been modified to an alpha-numeric referencing system with the following components:

- The zone number: in my case 30.
- Each zone is split into twenty latitude sub-zones, nineteen of extent 8° starting from 80°S and one of 12° finishing at 84°N. Each of these twenty latitude bands is designated a zone letter from C to X, with I and O excepted (to avoid ambiguity with digits 1 and 0). My home (approximately $55°57'$N, $3°11'$W) is just within the northern limit of zone 30U, centred on 3°W and running from 48° to 56°N. The extent of this sub-zone in projection coordinates is approximately 890km north-south whilst the width varies from 447km to 373km as one goes north.
- Within each of the sub-zones the 100km squares are labelled with row and column letters; for example Edinburgh is in a 100km square labelled UG. Thus 30UUG fixes my home to within 100km. (For a full description of the labelling of the 100km squares see the DMA manual listed in the bibliography.)
- Within a 100km square the Eastings and Northings range from 0 to 99999m so that 1m accuracy is given by two five digit numbers. Thus the full UTM grid reference of my armchair is 30UUG 88533 00666.
- Such precision is often superfluous and a pair of numbers with 4, 3, 2, 1 digits may be used for accuracy to within 10m, 100m, 1km, 10km (of the left hand and bottom edges of a box of that size). Thus
  - 30UUG 8853 0066 fixes my home to within 10m
  - 30UUG 885 006 fixes my home to within 100m
  - 30UUG 88 00 fixes the centre of Edinburgh within 1km
  - 30UUG 8 0 fixes Edinburgh within 10km

9.3 The British national grid: NGGB

The British national grid (NGGB) is a grid overlain on the TME projection centred on longitude 2°W and based on the Airy 1830 ellipsoid, for which the parameters are given in Section 5.2. It is a modified projection with a value of $k_0 = 0.9996012717$ on the central meridian. This value arose when the 1936 re-survey was constrained to be as close as possible to the previous survey at a number of selected points. For most practical purposes, and in the remainder of this chapter, we will take $k_0 \approx 0.9996$.

Clearly only a small area of the projection is needed, just the small box shown in Figure 9.2, so we take the true origin at latitude $\phi_0 = 49°$, or 0.855rad; this parallel is slightly south of the area mapped by the OSGB. The false origin is then chosen west and north with E0=400000m and N0=-100000m so that mapped area is completely to its east and north and all E and N values are positive. Recall that E0 and N0 give the position of the true origin relative to the false origin.

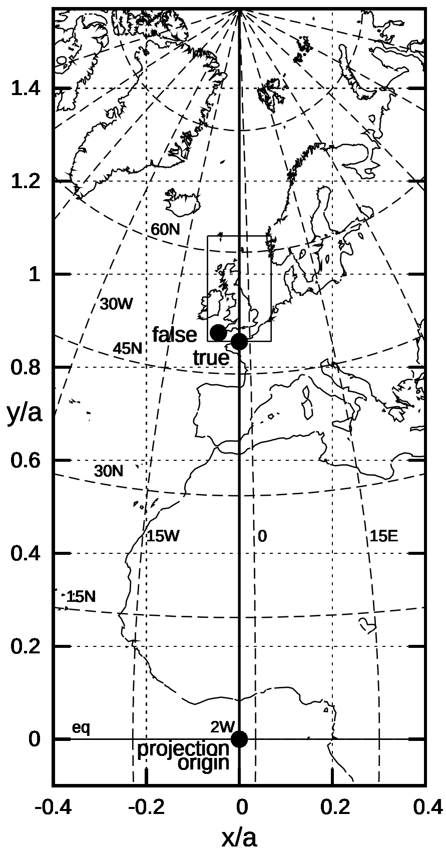


Figure 9.2

Figure 9.3 shows the grid in greater detail. There are several features to be noted. (a) Eastings and northings are normally used and quoted in metres but in this figure we have shown their values in kilometres; (b) the region over which NGGB is extant is defined in terms of ranges of eastings (0–700km) and northings (0–1300km) relative to the false origin; this is unlike the UTM zone which is defined on the ellipsoid in terms of meridians and parallels; (c) the 100km squares are annotated by with the letter pairs which are used in the alpha-numeric grid reference scheme; (d) meridians and parallels are the *curved* lines—compare with the TMS projections such as Figure 3.3; (e) the lines on which k is equal to 0.9996, 1 and 1.0004 (bounding the region of 'acceptable' scale accuracy) are shown as lines which appear to be straight—but see Section 9.4 for a truer statement.

The territory covered by NGGB requires a much wider longitude range than UTM; it must extend 2°E to 7°W, an interval of 9 degrees compared to the 6 degrees of UTM. Moreover this longitude range is not symmetric about the central meridian— 4 degrees to the east but 5 degrees to the west. This means that NGGB requires use of the transformation formulae at up to 5 degrees from the central meridian.

[Aside. Although the grid clearly covers Northern Ireland (and part of Eire) it is not used as a reference system in that province. The excluded region is shown in Figure 9.3 (heavy dotted line). In both Northern Ireland and Eire one uses a TME projection based

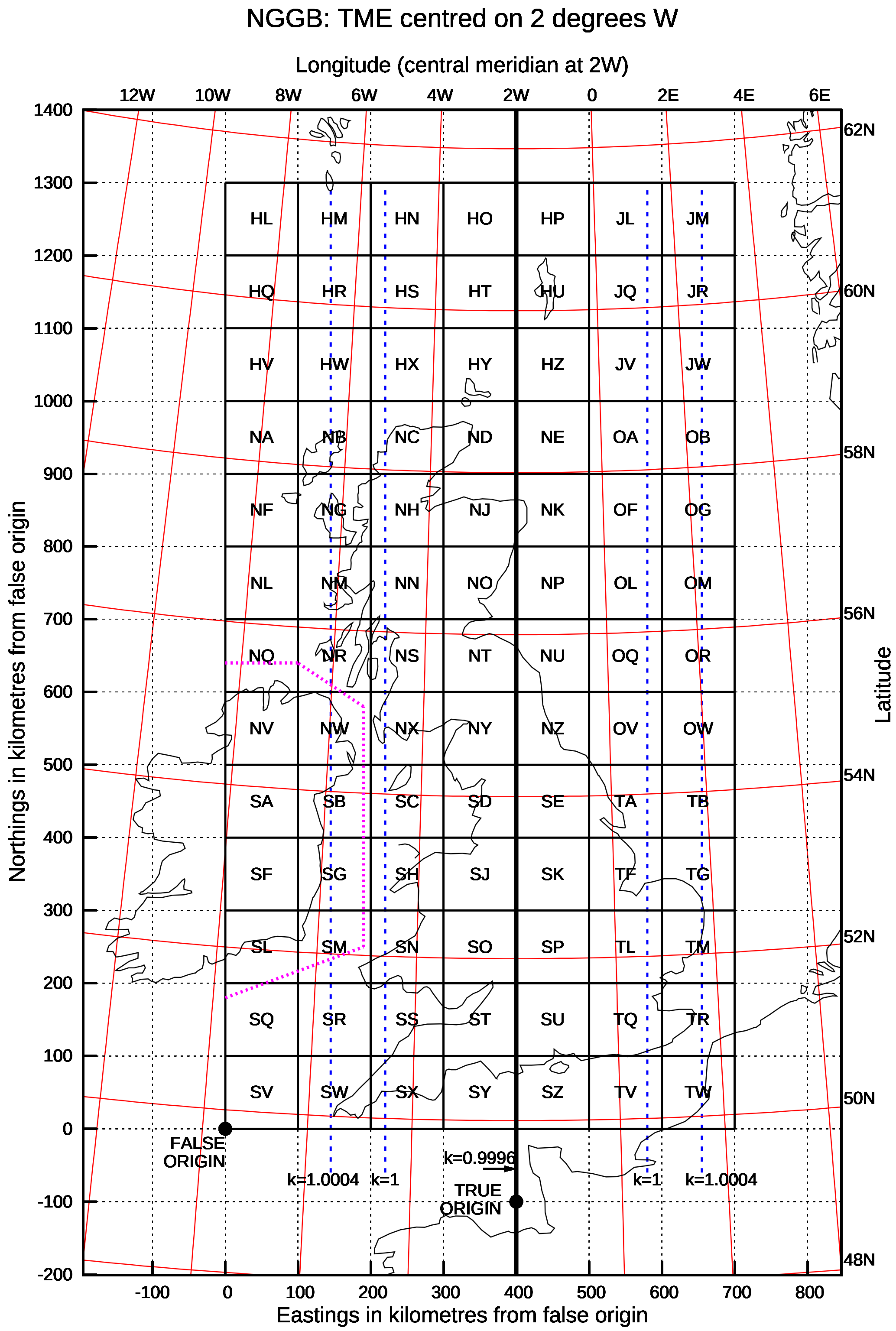

NGGB: TME centred on 2 degrees W
Longitude (central meridian at 2W)
12W 10W 8W 6W 4W 2W 0 2E 4E 6E
62N 60N 58N 56N 54N 52N 50N 48N
Latitude
1400 1300 1200 1100 1000 900 800 700 600 500 400 300 200 100 0 -100 -200
Northings in kilometres from false origin
-100 0 100 200 300 400 500 600 700 800
Eastings in kilometres from false origin
HL HM HN HO HP JL JM
HQ HR HS HT HU JQ JR
HV HW HX HY HZ JV JW
NA NB NC ND NE OA OB
NF NG NH NJ NK OF OG
NL NM NN NO NP OL OM
NQ NR NS NT NU OQ OR
NV NW NX NY NZ OV OW
SA SB SC SD SE TA TB
SF SG SH SJ SK TF TG
SL SM SN SO SP TL TM
SQ SR SS ST SU TQ TR
SV SW SX SY SZ TV TW
FALSE ORIGIN
TRUE ORIGIN
k=0.9996
k=1.0004
k=1
k=1
k=1.0004

Figure 9.3

on a slightly modified Airy ellipsoid centred on the meridian at 8° W; the grid having a true origin on this meridian at latitude $53^\circ 30'$N and a false origin such that E0=200000m and N0=250000m. The scale modification is $k_0 = 1.000035$. (See the Bibliography for a reference to a discussion of this unusual factor).

Returning to NGGB we see that the equations for E and N must take into account the shift along the meridian to the true origin as well as the translation to the false origin:

$$E = E0 + x(\lambda, \phi) \qquad\qquad = x + 400000, \tag{9.4}$$

$$N = N0 + [y(\lambda, \phi) - k_0 m(\phi_0)] = y - 5542868, \tag{9.5}$$

where we have calculated $m(\phi_0)$ with $\phi_0 = 49^\circ = 0.8552$ radians using (5.71). For x and y we use the Redfearn series (8.59, 8.60) with the understanding that λ (in radians) is measured from the central meridian at 2°W. The inverse series are again calculated from (8.61, 8.63) with x replaced by $E - E0$ and the footpoint calculated from

$$m(\phi_1) = \frac{y}{k_0} = \frac{N - N0}{k_0} + m(\phi_0) \tag{9.6}$$

Using the above formulae I calculate that, to the nearest metre, I am sitting at E=325701, N=673642. This is in the 100km square labelled NT so the alpha-numeric grid reference is NT 25701 73642 to within 1m. As for UTM we can use shorter grid references such as NT25707364, NT257736, NT2573, NT27 for accuracy to within 10m, 100m, 1km, 10km.

9.4 Scale variation in TME projections

Scale variation in TMS

Since the differences between TMS and TME are small (of order e^2) it is instructive to review the properties of the scale factor in TMS as a function of the projection coordinates. If we use the notation $k(x,y)\big|_{\text{TMS}}$ for the TMS scale factor we have the exact result from equation (3.73),

$$k(x,y)\Big|_{\text{TMS}} = k_0 \cosh\left(\frac{x}{k_0 a}\right) \tag{9.7}$$

and the series to sixth order in x/a is

$$k(x,y)\Big|_{\text{TMS}} = k_0 \left[1 + \frac{1}{2}\widehat{x}^2 + \frac{1}{24}\widehat{x}^4 + \frac{1}{720}\widehat{x}^6\right] \qquad \widehat{x} = \frac{x}{k_0 a}. \tag{9.8}$$

This scale factor is a function of x only and it is straightforward to calculate (using 9.7) the values at which $k = 1$ and $k = 1.0004$. Taking $k_0 = 0.9996$ and the radius of the sphere as $a = 6378$km, (approximately equal to the semi-major axis of the Airy ellipsoid), we find that these **isoscale** lines are at approximately x=180km and $x = 255$km respectively. Figure 9.4 shows a plot of the scale factor as a function of x. We shall prove that the differences between the scale factors of TMS and TME are so small that the corresponding plot for TME at any fixed y value is indistinguishable from Figure 9.4.

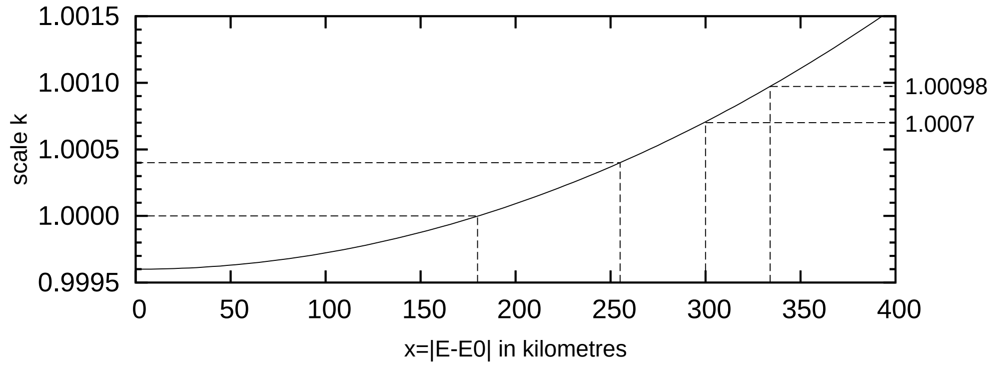


Figure 9.4

Scale variation in TME

The corresponding result for TME is given by equation (8.66):

$$k(x,y) = k_0\left[1+\frac{1}{2}\widehat{x}^2K_2+\frac{1}{24}\widehat{x}^4K_4+\frac{1}{720}\widehat{x}^6\bar{K}_6\right] \qquad \widehat{x}=\frac{x}{k_0\nu_1}, \tag{9.9}$$

where the K coefficients are given in (8.68) as

$$\begin{aligned} K_2 &= \beta_1\,,\\ K_4 &= 4\beta_1^3(1-6t_1^2)-3\beta_1^2(1-16t_1^2)-24\beta_1 t_1^2,\\ \bar{K}_6 &= 1, \end{aligned} \tag{9.10}$$

where β_1 and t_1 are evaluated at the footpoint latitude such that $m(\phi_1) = y/k_0$ and y is related to the northing coordinate by equation (9.2) or (9.5) for UTM or NGGB respectively.

To simplify the comparison of the TMS and TME scale factors it is instructive, and useful, to write K_2 and K_4 in terms of the following parameter of order e^2:

$$\eta^2 = \beta - 1 = \frac{\nu}{\rho} - 1 = \frac{e^2\cos^2\phi}{1-e^2} = e'^2\cos^2\phi. \tag{9.11}$$

The importance of this parameterisation is that it allows us to identify the way in which the K-coefficients depend on the eccentricity of the ellipse. It was not introduced at an earlier stage because the parameter β is much better suited to the derivations of all the coefficients. For the non-trivial coefficients in the TME scale factor we have

$$\begin{aligned} K_2 &= 1+\eta_1^2\\ K_4 &= 1+6\eta_1^2+\eta_1^4(9-24t_1^2)+\eta_1^6(4-24t_1^2) \end{aligned} \tag{9.12}$$

Clearly as $e^2 \to 0$ the TME scale factor reduces to the TMS scale factor. Their difference is given by terms of order $\eta^2\widehat{x}^2$ and $\eta^2\widehat{x}^4$: these are typically no more than 10^{-5} and 10^{-8} respectively so that Figure 9.4 is an adequate representation of the TME scale variation with x for any value of η_1 calculated for the footpoint value corresponding to y. (Note that in working with the Redfearn series we must not assume that the t_1 terms are small: the footpoint latitude may be of the order $80°$ for which $t_1 = \tan\phi_1 \approx 5.7$ so that $t_1^2 \approx 32$).

Figure 9.4 is annotated with the scale factors for the extreme cases which may arise in TME. For UTM the greatest extent in projection coordinate is on the equator where $x \approx 334$km and k reaches its worst value of almost 1.001—still perfectly acceptable in all practical applications. For NGGB the worst cases are the extremes of the East Anglia coast and the Outer Hebrides where $|x|$ is about 300km and $k \approx 1.0007$. The bulk of the NGGB grid is within approximately 255km of the central meridian where $|k-1| < 0.0004$. Note that a typical map sheet corresponds to a small section of the plot. For example on my 'local' sheet of the OSGB, bounded by E=316km and E=356km and close to the central meridian on which E=400km, the scale varies from $k = 0.999686$ on the west to $k = 0.999624$ on the east.

Isoscale lines of TME

We will now investigate the lines on which k is a constant and show how little they differ from the straight lines x=constant which we found for the TMS projection. To do this we make a Lagrange inversion of the series (9.9) for the TME scale factor. Start by writing this equation as

$$w \equiv \frac{2}{K_2}\left(\frac{k}{k_0} - 1\right) = \widehat{x}^2 + \frac{1}{12}\widehat{x}^4\frac{K_4}{K_2} + \frac{1}{360}\widehat{x}^6\frac{\bar{K}_6}{K_2} \qquad \widehat{x} = \frac{x}{k_0\nu_1}. \tag{9.13}$$

From equations (B.10–B.11), with $z = \widehat{x}^2$ and $b_4 = 0$, we immediately obtain the inverse

$$\widehat{x}^2 = w - \left(\frac{K_4}{12K_2}\right)w^2 - \left(\frac{K_2\bar{K}_6 - 5K_4^2}{360K_2^2}\right)w^3 \qquad w = \frac{2}{K_2}\left(\frac{k}{k_0} - 1\right), \tag{9.14}$$

from which we can find $x = k_0\nu_1\widehat{x}$, and hence $(E-E0)$, as a function of k for a given value of N. We have used this result to calculate the eastings at which $k = 1$ and $k = 1.0004$ for the UTM projection at four particular values of N for each of which we have first calculated the footpoint from equation (9.3). From the fourth column we see that the deviation in

N	footpoint	k=0.9996	k=1.0	k=1.0004
9000000	81°.05846848	0	180946	255887
6000000	54°.14694587	0	180556	255336
3000000	27°.12209299	0	180010	254564
0	0°.00000000	0	179759	254208

UTM: value of $|E-E0|$ at which k=0.9996, 1, 1.0004

eastings on the $k = 1$ isoscale is just approximately 1km over in a range of 9000km from the equator to just over 80°N. The deviation of the $k = 1.004$ isoscale is still under 2km. Thus the isoscale lines are essentially parallel to the central meridian.

Similar results can be calculated for the British grid:

N	footpoint	k=0.9996	k=1.0	k=1.0004
1200000	60°.68382350	0	180369	255275
800000	57°.09116995	0	180302	255180
400000	53°.49645197	0	180231	255080
0	49°.89956809	0	180157	254976

NGGB: value of $|E - E0|$ at which k=0.9996, 1, 1.0004

Clearly the lines on which $k = 1$ and $k = 1.0004$, which are shown in Figure 9.3, are indistinguishable from straight lines parallel to the central meridian. For the $k = 1$ locus the change in eastings is only 212m over the full northings extent of 1200km.

To end this discussion of the scale factor in TME note that much of mainland Britain is just within the scale variation of 0.9996 to 1.0004. This *may* be why these numbers were adopted as suitable criteria for mapping in the first place. This is of course speculation and we would be interested to hear of any evidence either way.

9.5 Convergence in the TME projection

When we first discussed convergence in Section 3.6 we observed that on any particular meridian (on the TMS projection) the convergence, defined as the angle between grid north and true north (the tangent to the meridian), must increase from zero at the equator to λ at the pole. The same must be true for TME although we require only values up to the northerly limit of UTM or NGGB. The following table shows the convergence for these two projections for several latitude values—on a bounding meridian for UTM (at λ_0+3°) and on the extreme meridian intersecting the land area covered by NGGB.

Convergence along a projected meridian			
UTM at λ_0+3°		NGGB at 7°W	
84°N	2°59′ 1″W	60°N	4°19′ 58″E
80°N	2°57′ 16″W	58°N	4°14′ 36″E
60°N	2°35′ 55″W	56°N	4° 8′ 55″E
40°N	1°55′ 46″W	54°N	4° 2′ 55″E
20°N	1° 1′ 37″W	52°N	3°56′ 38″E
0°N	0°	50°N	3°50′ 3″E

For UTM the values clearly approach the limiting value of 3° which would be attained at the pole. For NGGB the specified meridian is 5° west of the central meridian and the convergence clearly approaches this value at the northern extremity of the grid.

The convergence varies only slightly over any one of the map sheets. For example exact calculations give the convergence at the corners of the NGGB Edinburgh sheet bounded by E=316km, E=356km, N=650km and N=690km as

Boundary of Edinburgh sheet							
	E	N	γ	γ	E	N	
NW	316	690	$1°7'\,14.94''$E	$35'\,13.82''$E	356	690	NE
SW	316	650	$1°6'\,20.85''$E	$34'\,45.48''$E	356	650	SE

Note that the variation of convergence from top to bottom is much less than the variation from east to west. Similar figures are given at the corners of every map in the OSGB series: the values at other points may be approximated by interpolation.

It is important to observe that convergence (γ) values are not vanishingly small and they must be taken into account in relating an azimuth (α) to a grid bearing (β) by the relation $\alpha = \beta + \gamma$ discussed in Section 8.2. This correction is important in high accuracy applications.

9.6 The accuracy of the TME transformations

One obvious test of the accuracy of the TME transformations is to start from given geographical coordinates (ϕ, λ), transform to projection coordinates with the direct Redfearn series (8.59, 8.60) and then reverse the transformation with the inverse series (8.61, 8.63) We should then be back where we started. Before doing so we must decide on our standard of accuracy. We shall work to within 1mm in the projection coordinates and to within 0.0001$''$ in geographical coordinates. These accuracies are approximately equivalent, for we see from (2.5) that 0.0001$''$ is equivalent to 3mm along the meridian and less than 2mm along a parallel for examples calculated within the region of the NGGB (at latitudes from 50°N–60°N). To cope with rounding errors we compute to two extra places of decimals. These accuracies are purely to assess the mathematical consistency of our transformations for, in practice, no survey claims accuracies better than 10cm. The following example shows that this first test is satisfied with flying colours. Note that we use eastings and nothings rather than x and y, the projection coordinates. For NGGB they are related by (9.4, 9.5) and the footpoint is to be calculated from (9.6).

	Lat	Lon	E	N
Redfearn–direct	$52°39'\,27.2531''$	$1°43'\,4.5177''$	651409.903	313177.270
	E	N	Lat	Lon
Redfearn-Inverse	651409.903	313177.270	$52°39'\,27.2531''$	$1°43'\,4.5177''$

Another test is to assess the outcome of small changes in the inputs to the direct and inverse series. Sticking to the same coordinates as above we perturb the geographical coordinates by 0.0001$''$ in latitude, longitude separately and together: for the inverse we perturb the projection coordinates by 0.001mm. The results are shown overleaf.

	Lat	Lon	E	N
NGGB-direct	$52°39'\,27.2531''$	$1°43'\,4.5177''$	651409.903	313177.270
Lat + $0.0001''$	$52°39'\,27.2532''$	$1°43'\,4.5177''$	651409.903	313177.273
Lon + $0.0001''$	$52°39'\,27.2531''$	$1°43'\,4.5178''$	651409.905	313177.270
Both together	$52°39'\,27.2532''$	$1°43'\,4.5178''$	651409.905	313177.274

	E	N	Lat	Lon
NGGB-Inverse	651409.903	313177.270	$52°39'\,27.2531''$	$1°43'\,4.5177''$
E + 1mm	651409.904	313177.270	$52°39'\,27.2531''$	$1°43'\,4.5178''$
N + 1mm	651409.903	313177.271	$52°39'\,27.2531''$	$1°43'\,4.5177''$
Both together	651409.904	313177.271	$52°39'\,27.2531''$	$1°43'\,4.5178''$

Thus we see that $0.0001''$ changes induce a maximum change on the projection of no more than 4mm: for the inverse transformation 1mm changes in the projection coordinates change the geographical coordinates by no more than $0.0001''$.

Now when Redfearn published his series he was 'simply' extending the series that had been published earlier by Lee who had discarded terms smaller that $\lambda^4 e^2$, $\lambda^5 e^2$, $(x/a)^4 e^2$ and $(x/a)^5 e^2$ in the series for x, y, ϕ and λ respectively. Redfearn observed that the coefficients in Lee's series were increasing rapidly, particularly at larger latitudes where t and t_1 are not small, and consequently it seemed possible that some omitted terms might actually be larger than the smallest ones retained. This in fact proved to be the case for two of the terms omitted by Lee. Redfearn's analysis to higher order makes clear which terms can be safely omitted, as is done in many published expressions for the transformations.

To compare the size of the terms we again introduce the parameter η^2, which is $O(e^2)$, defined in equation (9.11). The transformation equations are those of Section 8.10 but we now write them in terms of eastings and northings using equations (9.4–9.6) and we also replace λ by $\lambda - \lambda_0$.

$$E - E0 = k_0\nu\left[\widetilde{\lambda} + \frac{\widetilde{\lambda}^3}{3!}W_3 + \frac{\widetilde{\lambda}^5}{5!}W_5 + \frac{\widetilde{\lambda}^7}{7!}\overline{W}_7\right], \tag{9.15}$$

$$N - N0 + k_0 m(\phi_0) = k_0\left[m(\phi) + \frac{\widetilde{\lambda}^2\nu t}{2} + \frac{\widetilde{\lambda}^4\nu t}{4!}W_4 + \frac{\widetilde{\lambda}^6\nu t}{6!}W_6 + \frac{\widetilde{\lambda}^8\nu t}{8!}\overline{W}_8\right], \tag{9.16}$$

$$\lambda(E,N) = \frac{\widehat{x}}{c_1} - \frac{\widehat{x}^3}{3!\,c_1}V_3 - \frac{\widehat{x}^5}{5!\,c_1}V_5 - \frac{\widehat{x}^7}{7!\,c_1}\overline{V}_7, \tag{9.17}$$

$$\phi(E,N) = \phi_1 - \frac{\widehat{x}^2\beta_1 t_1}{2} - \frac{\widehat{x}^4\beta_1 t_1}{4!}U_4 - \frac{\widehat{x}^6\beta_1 t_1}{6!}U_6 - \frac{\widehat{x}^8\beta_1 t_1}{8!}\overline{U}_8, \tag{9.18}$$

where

$$\widetilde{\lambda} = (\lambda - \lambda_0)c, \qquad \widehat{x} = \frac{E - E0}{k_0\nu_1}, \qquad m(\phi_1) = \frac{N - N0}{k_0} + m(\phi_0) \tag{9.19}$$

and the coefficients, given in equations (7.51–7.53), are now written in terms of η as shown overleaf.

The significance of the vertical rules in the rewritten coefficients will be discussed shortly.

$$\begin{aligned}
W_3 &= 1 - t^2 + \eta^2 \,\Big| \\
W_4 &= 5 - t^2 + 9\eta^2 \,\Big|\, + 4\eta^4 \\
W_5 &= 5 - 18t^2 + t^4 + \eta^2(14 - 58t^2) \,\Big|\, + \eta^4(13 - 64t^2) + \eta^6(4 - 24t^2) \\
W_6 &= 61 - 58t^2 + t^4 \,\Big|\, + \eta^2(270 - 330t^2) + \eta^4(445 - 680t^2) \\
&\qquad\qquad\qquad + \eta^6(324 - 600t^2) + \eta^8(88 - 192t^2) \\
\overline{W}_7 &= \Big|\, 61 - 479t^2 + 179t^4 - t^6 \\
\overline{W}_8 &= \Big|\, 1385 - 3111t^2 + 543t^4 - t^6
\end{aligned} \tag{9.20}$$

$$\begin{aligned}
V_3 &= 1 + 2t_1^2 + \eta_1^2 \,\Big| \\
V_5 &= -5 - 28t_1^2 - 24t_1^4 \,\Big|\, - \eta_1^2(6 + 8t_1^2) + \eta_1^4(3 - 4t_1^2) + \eta_1^6(4 - 24t_1^2) \\
\overline{V}_7 &= 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6 \,\Big|
\end{aligned} \tag{9.21}$$

$$\begin{aligned}
U_4 &= -5 - 3t_1^2 - \eta_1^2(1 - 9t_1^2) \,\Big|\, + 4\eta_1^4 \\
U_6 &= 61 + 90t_1^2 + 45t_1^4 \,\Big|\, + \eta_1^2(46 - 252t_1^2 - 90t_1^4) + \eta_1^4(-3 - 66t_1^2 + 225t_1^4) \\
&\qquad\qquad\qquad + \eta_1^6(100 + 84t_1^2) + \eta_1^8(88 - 192t_1^2) \\
\overline{U}_8 &= \Big|\, -1385 - 3633t_1^2 - 4095t_1^4 - 1575t_1^6
\end{aligned} \tag{9.22}$$

As an example we consider the direct and inverse transformations for a location in the Outer Hebrides with $\phi = 58°$N and $\lambda = 7°$W and projection coordinates given by E =104647.323m and N =912106.244m. This point has about the greatest value of $\lambda - \lambda_0$ that we can get in the NGGB and moreover it is about as far north as we can get so the value of $\tan\phi$ in the coefficients is not small (t=1.6). In the tables shown overleaf all the sub-terms have been displayed according to their power of η^2, essentially e^2, and their power of either $\widetilde{\lambda}$ or $\widehat{x}$ for the direct and inverse series respectively.

In these two tables the upright rules correspond to those in the expressions for the coefficients in (9.20). We now see that they demarcate the significant terms: all terms to the right of them are negligible and hence we can drop these terms from the series with impunity. Since the chosen point was the most extreme for NGGB we can obviously neglect these terms for all applications of NGGB.

Thus, at the end of the day, after much hard graft, we have thrown away almost all of the higher order terms except for the seventh order term in x/a in the inverse series for λ. Clearly terms of order $O\big(e^2(x/a)^7\big)$ would also be negligible so at last we have justified the use of the spherical approximation in calculating the higher order terms. Note that we could not have assessed the size of the higher order terms without working them out!

One can of course use the Redfearn series as they stand for they are simple to encode on any computer. Remember, however, that when these series were first developed it was imperative to simplify the working as much as possible for hand(-machine) calculations.

In: $\phi = 58^\circ$ $\lambda = -7^\circ$					Out: E =104647.323m
	η^0	η^2	η^4	η^6	η^8
$\widetilde{\lambda}^1$	104482.705				
$\widetilde{\lambda}^3$	164.425	-0.199			
$\widetilde{\lambda}^5$	0.389	0.003	6.0E-06	4.3E-09	
$\widetilde{\lambda}^7$	4.9E-06				

In: $\phi = 58^\circ$ $\lambda = -7^\circ$					Out: N =912106.244m
	η^0	η^2	η^4	η^6	η^8
$\widetilde{\lambda}^0$	901166.420				
$\widetilde{\lambda}^2$	10935.050				
$\widetilde{\lambda}^4$	4.753	0.032	2.8E-05		
$\widetilde{\lambda}^6$	-0.011	-1.5E-04	-6.4E-07	-1.1E-09	-7.1E-13
$\widetilde{\lambda}^8$	-1.6E-05				

In: E =104647.323, N =912106.244					Out: ϕ =58° $0'$ $0.0000''$
	η_1^0	η_1^2	η_1^4	η_1^6	η_1^8
$\widehat{x}^0$	58° $5'\,53.7728''$				
$\widehat{x}^2$	$-5'\,54.5718''$				
$\widehat{x}^4$	$0.8042''$	$-$ $0.0025''$	-8.9E-07$''$		
$\widehat{x}^6$	$-$ $0.0027''$	1.0E-05$''$	-2.1E-08$''$	-9.4E-12$''$	2.3E-14$''$
$\widehat{x}^8$	1.1E-05$''$				

In: E =104647.323, N =912106.244					Out: λ =-7° $0'$ $0.0000''$
	η_1^0	η_1^2	η_1^4	η_1^6	η_1^8
$\widehat{x}^1$	-7° $0'\,39.4213''$				
$\widehat{x}^3$	$39.5711''$	$0.0120''$			
$\widehat{x}^5$	$-$ $0.1626''$	-3.4E-05$''$	-1.8E-08$''$	-2.6E-10$''$	
$\widehat{x}^7$	$0.0008''$				

Lee would no doubt have had this in mind when he dropped the sixth order terms in his calculations. To be exact, of the required terms he dropped the term in λ^6 in the direct series for y, the term in $(x/a)^6$ in the inverse series for ϕ and the term in $(x/a)^7$ in the inverse series for λ; at the same time he included the term in $(x/a)^5 e^2$ in the inverse series for λ although we now see that it is negligible.

The approximations used in Snyder are similar but he retains the terms $\eta_1^2(46-252t_1^2)$ in U_6; we have shown this is negligible for NGGB but it is required at the higher latitudes that we meet in UTM. (Comment: Snyder uses e'^2 explicitly in some coefficients. To compare with the expressions above first set $e'^2 = \eta^2 \sec^2\phi = \eta^2[1 + t^2]$).

9.7 The truncated TME series

We have stressed that although no penalty is incurred by using the full Redfearn series as summarised in Section 8.10, the coefficients truncated at the vertical rules in equations (9.20–9.22) will be perfectly adequate. Dropping the small terms equations (9.15–9.19) become

$$E(\phi,\lambda) = E0 + k_0\nu\left[\widetilde{\lambda} + \frac{\widetilde{\lambda}^3}{3!}W_3^T + \frac{\widetilde{\lambda}^5}{5!}W_5^T\right], \tag{9.23}$$

$$N(\phi,\lambda) = N0 + k_0\left[m(\phi) - m(\phi_0)\right] + k_0\left[\frac{\widetilde{\lambda}^2\nu t}{2} + \frac{\widetilde{\lambda}^4\nu t}{4!}W_4^T + \frac{\widetilde{\lambda}^6\nu t}{6!}W_6^T\right], \tag{9.24}$$

$$\lambda(E,N) = \frac{\widehat{x}}{c_1} - \frac{\widehat{x}^3}{3!\,c_1}V_3^T - \frac{\widehat{x}^5}{5!\,c_1}V_5^T - \frac{\widehat{x}^7}{7!\,c_1}\overline{V}_7^T, \tag{9.25}$$

$$\phi(E,N) = \phi_1 - \frac{\widehat{x}^2\beta_1 t_1}{2} - \frac{\widehat{x}^4\beta_1 t_1}{4!}U_4^T - \frac{\widehat{x}^6\beta_1 t_1}{6!}U_6^T, \tag{9.26}$$

where

$$\widetilde{\lambda} = (\lambda - \lambda_0)c, \qquad \widehat{x} = \frac{E - E0}{k_0\nu_1}, \qquad m(\phi_1) = \frac{N - N0}{k_0} + m(\phi_0), \tag{9.27}$$

and, as usual, $c{=}\cos\phi$, $t{=}\tan\phi$, $\beta{=}\nu(\phi)/\rho(\phi)$ from (5.53) and the '1' subscript denotes a function evaluated at the footpoint latitude.

The truncated coefficients follow from equations (9.20–9.22)

$$\begin{aligned}
W_3^T &= 1 - t^2 + \eta^2\\
W_4^T &= 5 - t^2 + 9\eta^2\\
W_5^T &= 5 - 18t^2 + t^4 + \eta^2(14 - 58t^2)\\
W_6^T &= 61 - 58t^2 + t^4
\end{aligned} \tag{9.28}$$

$$\begin{aligned}
V_3^T &= 1 + 2t_1^2 + \eta_1^2\\
V_5^T &= -5 - 28t_1^2 - 24t_1^4\\
\overline{V}_7^T &= 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6
\end{aligned} \tag{9.29}$$

$$\begin{aligned}
U_4^T &= -5 - 3t_1^2 - \eta_1^2(1 - 9t_1^2)\\
U_6^T &= 61 + 90t_1^2 + 45t_1^4
\end{aligned} \tag{9.30}$$

where

$$\eta^2 = \beta - 1 = \frac{\nu}{\rho} - 1 = \frac{e^2\cos^2\phi}{1 - e^2} = e'^2\cos^2\phi. \tag{9.31}$$

9.8 The OSGB series

The published form of the series used by the OSGB uses a different notation for the truncated series of the previous section. First use equations (5.78, 5.80) to set

$$
\begin{aligned}
M &= k_0\,[m(\phi) - m(\phi_0)] \qquad (9.32)\\
&= k_0 b\Bigg[\left(1+n+\frac{5}{4}n^2+\frac{5}{4}n^3\right)(\phi-\phi_0) - \left(3n+3n^2+\frac{21}{8}n^3\right)\sin(\phi-\phi_0)\cos(\phi+\phi_0)\\
&\quad + \left(\frac{15}{8}n^2+\frac{15}{8}n^3\right)\sin 2(\phi-\phi_0)\cos 2(\phi+\phi_0) - \left(\frac{35}{24}n^3\right)\sin 3(\phi-\phi_0)\cos 3(\phi+\phi_0)\Bigg]
\end{aligned}
$$

where n is defined in equation (5.5) as $(a-b)/(a+b)$.

Equations (9.24), (9.23), (9.26) and (9.25) may be written as

$$N = \mathrm{I} + \mathrm{II}(\lambda-\lambda_0)^2 + \mathrm{III}(\lambda-\lambda_0)^4 + \mathrm{IIIA}(\lambda-\lambda_0)^6 \qquad (9.33)$$

$$E = E0 + \mathrm{IV}(\lambda-\lambda_0) + \mathrm{V}(\lambda-\lambda_0)^3 + \mathrm{VI}(\lambda-\lambda_0)^5, \qquad (9.34)$$

$$\phi = \phi_1 - \mathrm{VII}(E-E0)^2 + \mathrm{VIII}(E-E0)^4 - \mathrm{IX}(E-E0)^6, \qquad (9.35)$$

$$\lambda = \lambda_0 + \mathrm{X}(E-E0) - \mathrm{XI}(E-E0)^3 + \mathrm{XII}(E-E0)^5 - \mathrm{XIIA}(E-E0)^7, \qquad (9.36)$$

where ϕ_1 must be calculated from equation

$$m(\phi_1) = \frac{y}{k_0} = \frac{N-N0}{k_0} + m(\phi_0) \qquad (9.37)$$

by the methods of Section 5.9.

If we introduce $\widetilde{\nu} = k_0\nu$ and $\widetilde{\rho} = k_0\rho$ the coefficients I–XIIA may be written in terms of the truncated coefficients (9.28–9.30) as

$$\mathrm{I} = M + N0 \qquad \mathrm{II} = \frac{\widetilde{\nu}sc}{2} \qquad \mathrm{III} = \frac{\widetilde{\nu}sc^3W_4^T}{4!} \qquad \mathrm{IIIA} = \frac{\widetilde{\nu}sc^5W_6^T}{6!}$$

$$\mathrm{IV} = \widetilde{\nu}c \qquad \mathrm{V} = \frac{\widetilde{\nu}c^3W_3^T}{3!} \qquad \mathrm{VI} = \frac{\widetilde{\nu}c^5W_5^T}{5!}$$

$$\mathrm{VII} = \frac{t_1}{2\widetilde{\rho}_1\widetilde{\nu}_1} \qquad \mathrm{VIII} = \frac{-t_1U_4^T}{4!\widetilde{\rho}_1\widetilde{\nu}_1^3} \qquad \mathrm{IX} = \frac{t_1U_6^T}{6!\widetilde{\rho}_1\widetilde{\nu}_1^5}$$

$$\mathrm{X} = \frac{1}{\widetilde{\nu}_1c_1} \qquad \mathrm{XI} = \frac{V_3^T}{3!\widetilde{\nu}_1^3c_1} \qquad \mathrm{XII} = \frac{-V_5^T}{5!\widetilde{\nu}_1^5c_1} \qquad \mathrm{XIIA} = \frac{\overline{V}_7^T}{7!\widetilde{\nu}_1^7c_1}$$

and, as usual, $c{=}\cos\phi$, $t{=}\tan\phi$, $\beta{=}\nu(\phi)/\rho(\phi)$ from (5.53) and the '1' subscript denotes a function evaluated at the footpoint latitude. The OSGB publication exhibits the above formulae without the tildes on ν and ρ because the definitions of ν and ρ in that publication already absorb a factor of k_0.

Chapter 10

The oblique Mercator projection

10.1 Introduction and summary

IN PREPARATION

Chapter 11

Geodesics on the sphere and ellipsoid

11.1 Introduction and summary

This chapter is a not directly concerned with the Mercator (or any) transformation, rather it tackles the problem of finding geodesics, curves of shortest length, between points on a sphere and on an ellipsoid. To be precise we solve the following two problems on both the sphere and the ellipsoid.

1. **The direct problem**.

   Define a geodesic by an initial azimuth α_1 at a starting point P_1 with geographic coordinates (ϕ_1, λ_1). Find the coordinates (ϕ_2, λ_2) of a point P_2 at a distance s measured along the geodesic; find also the azimuth α_2 of the geodesic at P_2.

2. **The inverse problem**.

   Given the coordinates (ϕ_1, λ_1) and (ϕ_2, λ_2) of two points P_1 and P_2 find s, the geodesic distance between them, and also α_1, α_2, the azimuths of the geodesic at the points.

Note that we need only relative longitudes so we calculate only $\lambda = \lambda_2 - \lambda_1$ in the direct problem; for the inverse problem only λ is given.

For the case of the sphere we shall see that we can derive the results (a) by spherical trigonometry or (b) by integrating in closed form simple first order differential equations describing the geodesics. For the ellipsoid there is no equivalent to spherical trigonometry and we have no option but to integrate the differential equations; moreover, since the equations cannot be integrated in closed form we must resort to series expansions.

The results for the sphere have been long established and approximations to the ellipsoid problems have been presented by many authors over the last 150 years. The results we present here were given in the Survey Review by Vincenty. (See bibliography)

THIS CHAPTER IS IN PREPARATION

Appendix A

Curvature in 2 and 3 dimensions

A.1 Planar curves

A straight line has zero curvature. The curvature, κ, of a general curve in the plane is defined as the rate of change of the direction of its tangent with respect to the distance travelled along the line:

$$\kappa = \frac{d\theta}{ds}. \tag{A.1}$$

If we are given the equation of the curve as $y = f(x)$ with respect to Cartesian axes then it is natural to choose the x-axis as the reference for the direction of the tangent.

The geometry of the small inset in the figure shows that

$$\tan\theta = \frac{dy}{dx} = y'(x), \qquad \cos\theta = \frac{dx}{ds} \tag{A.2}$$

Differentiating the first of these statements by s and using the second gives

$$\sec^2\theta \frac{d\theta}{ds} = \frac{d[y'(x)]}{ds} = y''(x)\frac{dx}{ds}$$
$$= y''(x)\cos\theta. \tag{A.3}$$

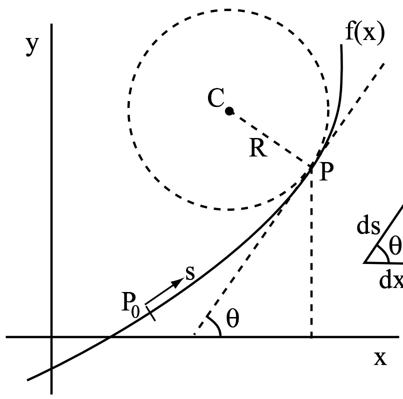


Figure A.1

Now $\sec\theta = \pm\sqrt{1+\tan^2\theta} = \pm\sqrt{1+y'^2}$ so we obtain $d\theta/ds$ and

$$\boxed{\kappa \ = \ \pm\frac{y''(x)}{[1+y'^2]^{3/2}}. \qquad \text{DASH} \equiv \frac{d}{dx}} \tag{A.4}$$

The choice of sign is a matter for convention in every case. We shall illustrate this point immediately.

The curvature of a circle

For a general circle of radius a at the origin we have $x^2 + y^2 = a^2$ so that on the two semicircles $y > 0$ and $y < 0$,

$$y(x) = \pm\sqrt{a^2 - x^2}, \qquad y'(x) = \frac{\mp x}{\sqrt{a^2 - x^2}}, \qquad y''(x) = \frac{\mp a^2}{(a^2 - x^2)^{3/2}}. \tag{A.5}$$

Substituting these values in equation (A.4) we see that the curvature of the upper semicircle is $\kappa = \pm(-1/a)$ whilst for the lower semicircle $\kappa = \pm(1/a)$. Now it is conventional to define the curvature of a circle to be *positive* so we must choose the negative sign in the definition for the case of the upper semicircle and the positive sign for the lower; with these choices of sign we have a constant curvature $\kappa = 1/a$. Therefore the curvature of a circle is the inverse of the radius and vice-versa.

The osculating circle and the radius of curvature

The particular circle which touches the curve at P (Figure A.1) and also shares the same curvature at that point is called the 'osculating circle' (osculating=kissing) or the 'circle of curvature'. The radius of this circle defines R, the radius of curvature of the curve at that point. Clearly

$$R = \frac{1}{\kappa}. \tag{A.6}$$

Curves in parametric form

The previous results related to a curve in two dimensions described by a single function $y(x)$ in Cartesian coordinates. We now consider the situation where these Cartesian coordinates are parameterised by two functions of u; that is the position of a point on the curve is written as $\mathbf{r}(u) = \big(x(u),\, y(u)\big)$. We shall investigate three types of parameterisation: (1) the parameter is assumed to be perfectly general, *not* necessarily the distance along the path; (2) the parameter *is* taken as the arc length s; (3) the parameter is taken as the angle between the tangent and the x-axis.

Case 1: Arbitrary parameterisation: $x(u),\ y(u)$. Set DOT $\equiv \dfrac{d}{du}$

$$y'(x) = \frac{dy}{dx} = \frac{\dot{y}}{\dot{x}},$$

$$y''(x) = \frac{d}{du}\left(\frac{\dot{y}}{\dot{x}}\right)\frac{du}{dx} = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{\dot{x}^3}$$

$$\boxed{\frac{1}{R} \;=\; \kappa \;=\; \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{[\dot{x}^2 + \dot{y}^2]^{3/2}} \qquad \text{DOT} \equiv \frac{d}{du}}. \tag{A.7}$$

Case 2: Special parameterisation: $u \to s$. Given $x(s),\ y(s)$. Set DASH $\equiv \dfrac{d}{ds}$

Since s is the arc length we have $dx^2 + dy^2 = ds^2$ and consequently $x'^2 + y'^2 = 1$. Therefore (A.7) becomes

$$\boxed{\frac{1}{R} \;=\; \kappa \;=\; x'y'' - y'x'' \qquad \text{DASH} \equiv \frac{d}{ds}} \tag{A.8}$$

Case 3: Special parameterisation: $u \to \theta$. Given $x(\theta),\ y(\theta)$. Set DOT $\equiv \dfrac{d}{d\theta}$

Since θ is the angle between the tangent and the x-axis we have $\tan\theta = \dfrac{dy}{dx} = \dot{y}/\dot{x}$

Differentiating with respect to θ gives $\sec^2\theta = \dfrac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{\dot{x}^2}$.

But we also have $\sec^2\theta = 1 + \tan^2\theta = \dfrac{\dot{x}^2 + \dot{y}^2}{\dot{x}^2}$

Therefore we must have $\dot{x}\ddot{y} - \ddot{x}\dot{y} = \dot{x}^2 + \dot{y}^2$ so that (A.7) becomes

$$\boxed{\frac{1}{R} \;=\; \kappa \;=\; \frac{1}{[\dot{x}^2 + \dot{y}^2]^{1/2}} \qquad \text{DOT} \equiv \frac{d}{d\theta}} \tag{A.9}$$

Note that this result follows directly from equation (A.1) since $\frac{1}{\kappa}\, d\theta = ds = [dx^2 + dy^2]^{1/2}$.

Curvature of the ellipse

The Cartesian equation of an ellipse is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \tag{A.10}$$

where the semi-axes are related to the eccentricity by $b = a\sqrt{1-e^2}$. Now the ellipse may be obtained by scaling the auxiliary circle by a factor of b/a in the y direction. Since an arbitrary point P' on the circle is $(a\cos U,\ a\sin U)$ the corresponding point on the ellipse is $P(a\cos U,\ b\sin U)$. We call U the 'parametric' or 'reduced' latitude in cartography and the 'eccentric anomaly' in astronomy.)

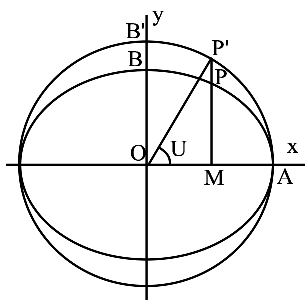


Figure A.2

We calculate the curvature from equation (A.7) setting:

$$\begin{aligned} x &= a\cos U, & y &= b\sin U, \\ \dot{x} &= -a\sin U, & \dot{y} &= b\cos U, \\ \ddot{x} &= -a\cos U, & \ddot{y} &= -b\sin U, \end{aligned} \tag{A.11}$$

giving

$$\kappa = \frac{+ab}{[a^2 - (a^2 - b^2)\cos^2 U]^{3/2}} = \frac{1}{a}\frac{\sqrt{1-e^2}}{[1 - e^2\cos^2 U]^{3/2}}. \tag{A.12}$$

A.2 Curves in three dimensions

First consider two neighbouring points, $P(s)$ and $Q(s+\delta s)$, on a curve parameterised by its arc length s (Figure A.3a). The chord length between these points is given by $\delta s^2 = \delta\mathbf{r}\cdot\delta\mathbf{r}$. The tangent vector at P is the limit of $\delta\mathbf{r}/\delta s$ and therefore has the properties

$$\mathbf{t} = \mathbf{r}', \qquad \mathbf{t}\cdot\mathbf{t} = 1 \qquad \mathbf{t}\cdot\mathbf{t}' = 0, \tag{A.13}$$

where the last follows by differentiation of the second.

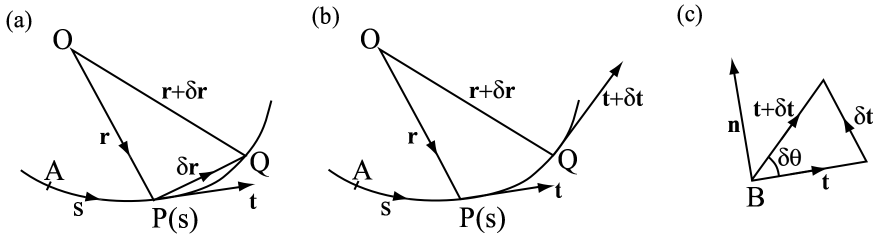


Figure A.3

Consider the tangents at neighbouring points (Figure A.3b); $\mathbf{t}(s)$ and $\mathbf{t}(s+\delta s) = \mathbf{t}+\delta\mathbf{t}$ are compared in the third figure by bringing them together at some point B. Tangent vectors are unit vectors so that $|\mathbf{t}| = |\mathbf{t}+\delta\mathbf{t}| = 1$; therefore in the limit of $\delta s \to 0$ we see that $\delta\mathbf{t}$ is in the direction of $\mathbf{n}$, a unit vector normal to $\mathbf{t}$ and in the 'osculating plane' defined by the two vectors $\mathbf{t}(s)$ and $\mathbf{t}(s+\delta s)$. Furthermore, if the angle between the unit tangent vectors is $\delta\theta$ then as $\delta s \to 0$ we must have $|\delta\mathbf{t}| = \delta\theta$. Consequently

$$\mathbf{t}' = \lim_{\delta s\to 0}\frac{\delta\mathbf{t}}{\delta s} = \frac{d\theta}{ds}\mathbf{n} = \kappa\mathbf{n}. \tag{A.14}$$

The vector $\mathbf{n}$ is called the principal normal to the curve and the curvature κ, is defined as for planar curves. We can invert this relation and write

$$\mathbf{n} = \frac{1}{\kappa}\mathbf{t}' = \frac{1}{\kappa}\mathbf{r}''. \tag{A.15}$$

Note that $\mathbf{n}$ is defined to be a unit vector; on the other hand $\mathbf{t}'$ is not a unit vector, its length is equal to the curvature κ. Since $\mathbf{t}\cdot\mathbf{t}' = 0$ we must have $\mathbf{n}\cdot\mathbf{t} = 0$.

Now introduce the unit 'binormal' vector $\mathbf{b}$, defined so that it forms a right-handed orthogonal triad with $\mathbf{t}$ and $\mathbf{n}$. Therefore

$$\mathbf{b} = \mathbf{t}\times\mathbf{n}, \tag{A.16}$$

$$\begin{aligned} &\mathbf{t}\cdot\mathbf{n} = 0 \qquad &&\mathbf{n}\cdot\mathbf{b} = 0 \qquad &&\mathbf{b}\cdot\mathbf{t} = 0,\\ &\mathbf{t}\times\mathbf{n} = \mathbf{b}, \qquad &&\mathbf{n}\times\mathbf{b} = \mathbf{t}, \qquad &&\mathbf{b}\times\mathbf{t} = \mathbf{n}. \end{aligned} \tag{A.17}$$

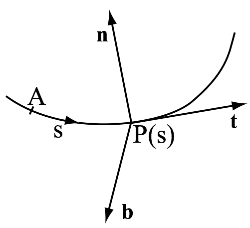


Figure A.4

Torsion and the Frenet–Serret formulae

Since $\mathbf{b}$ is a unit vector, differentiation gives $\mathbf{b}\cdot\mathbf{b}'=0$. Furthermore if we differentiate the relation $\mathbf{t}\cdot\mathbf{b}=0$ we get

$$\mathbf{t}'\cdot\mathbf{b}+\mathbf{t}\cdot\mathbf{b}'=0. \tag{A.18}$$

Now since $\mathbf{t}'=\kappa\mathbf{n}$ the first of these terms must vanish so we must have $\mathbf{t}\cdot\mathbf{b}'=0$. Consequently $\mathbf{b}'$ is perpendicular to both $\mathbf{t}$ and $\mathbf{b}$ and it is therefore a vector in the direction of $\mathbf{n}$. The vector $\mathbf{b}'$ is not a unit vector and its magnitude is defined to be τ, the 'torsion' of the curve. We choose to set

$$\mathbf{b}'=-\tau\mathbf{n}. \tag{A.19}$$

The torsion of the curve is a measure of the rate of rotation of the vectors $\mathbf{b}$, and hence $\mathbf{n}$, about the tangent vector as s increases. The negative sign associates a 'right-handed' rule as part of the definition.

We have expressions for the derivatives of $\mathbf{t}$ and $\mathbf{b}$ in equations (A.15) and (A.19). We now evaluate the derivative of $\mathbf{n}$ from $\mathbf{b}\times\mathbf{t}$:

$$\mathbf{n}'=\mathbf{b}'\times\mathbf{t}+\mathbf{b}\times\mathbf{t}'=-\tau\mathbf{n}\times\mathbf{t}+\mathbf{b}\times(\kappa\mathbf{n})=\tau\mathbf{b}-\kappa\mathbf{t}. \tag{A.20}$$

This equation together with the derivatives of $\mathbf{t}$ and $\mathbf{b}$ constitute the set of Frenet-Serret relations:

$$\begin{aligned}\mathbf{t}'&=\kappa\mathbf{n},\\ \mathbf{n}'&=\tau\mathbf{b}-\kappa\mathbf{t},\\ \mathbf{b}'&=-\tau\mathbf{n}.\end{aligned} \tag{A.21}$$

These equations show that the form of a curve in three dimensions is essentially determined by the two functions $\kappa(s)$ and $\tau(s)$ and an initial orthonormal triad.

A.3 Curvature of surfaces

At any point on a surface we can define the curvature of a line on the surface passing through that point. Rather than build up a large part of differential geometry we shall give elementary proofs of two important results that we need.

First consider those curves which are defined by the intersection of a plane with the surface. The most important case is a plane which contains the normal **N** at the point P; such a plane defines a 'normal section' of the surface. We shall consider all the normal sections at a given point and investigate the curvature at P of their intersections. The principal result is that the maximum and minimum curvatures arise on two normal sections at right-angles to each other; these are the 'principal' curvatures which we will denote by κ_1 and κ_2. Euler's formula gives the curvature on any other normal section in terms of the principal curvatures.

The other main result that we need is Meusnier's theorem. This relates the curvature on a normal section to the curvature of the sections made by planes oblique to the chosen normal plane, *i.e.* sharing the same tangent at P. It is convenient to prove this theorem first.

A.4 Meusnier's theorem

Without loss of generality we choose axes with the origin O at an arbitrary point on a surface and such that the xy-plane is tangential at the point. Consider the family of planes which contain the tangent directed along the x-axis. Each plane intersects the the surface in a plane curve; let $g(x)$ be the curve on the normal plane and $w(x)$ on an oblique plane inclined at an angle ϕ to the normal plane. If κ, R denote the curvature and radius of curvature at the origin of $g(x)$ and $w(x)$ on the normal and oblique planes repectively, then

$$\boxed{\kappa_{\text{oblique}} = \sec\phi\,\kappa_{\text{normal}}, \qquad R_{\text{oblique}} = \cos\phi\,R_{\text{normal}}.} \tag{A.22}$$

The choice of coordinates implies that $z = f(x,y)$, the 'height' of the surface above the xy-plane, is such that $f(0,0) = 0$ and has partial derivatives $f_x(0,0){=}0$ and $f_y(0,0){=}0$. The Taylor series at the origin is then

$$z = f(x,y) = \frac{1}{2}Ax^2 + Bxy + \frac{1}{2}Cy^2 + \cdots, \tag{A.23}$$

with $A = f_{xx}(0,0)$, $B = f_{xy}(0,0)$ and $C = f_{yy}(0,0)$.

The intersection of the xz-plane ($y = 0$) and the surface is the curve $g(x)$ which, near P, is given by

$$g(x) = f(x,0) = \frac{1}{2}Ax^2 + O(x^3). \tag{A.24}$$

The curvature of $g(x)$ at P follows from (A.4)

$$\kappa_{\text{normal}} = \left.\frac{g''(x)}{[1+(g')^2]^{3/2}}\right|_{x=0} = A. \tag{A.25}$$

On the oblique plane at P we have $z = f(x,y)$ with $z = w\cos\phi$ and $y = w\sin\phi$. Therefore

$$w\cos\phi = f(x, w\sin\phi) = \frac{Ax^2}{2} + Bxw\sin\phi + \frac{Cw^2}{2}\sin^2\phi$$

It is clear from this equation that for small x and arbitrary ϕ we must have $w(x) = O(x^2)$. (For suppose on the contrary that $w(x) = O(x)$, then the LHS of the previous equation would be $O(x)$ and the RHS would be $O(x^2)$).

$$w(x) = \sec\phi\left(\frac{1}{2}Ax^2 + O(x^3)\right). \tag{A.26}$$

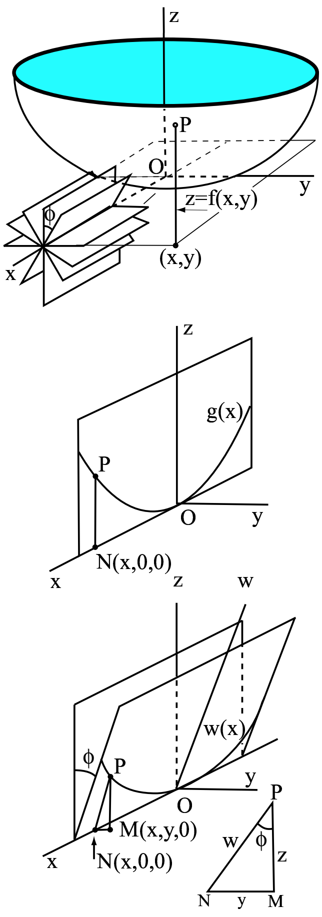


Figure A.5

Equations (A.26) and (A.24) give Meusnier's theorem

$$\kappa_{\text{oblique}} = \left.\frac{w''}{[1+(w')^2]^{3/2}}\right|_{x=0} = A\sec\phi = \sec\phi\,\kappa_{\text{normal}}. \tag{A.27}$$

A.5 Curvature of normal sections

We now consider the set of planes which have as axis the normal to the surface at a given point. The intersections of these planes with the surface define the 'normal sections' at that point. Once again we take the given point as the origin of our coordinate systems and define the tangent plane at the origin to be the xy-plane. Therefore the equation of the surface may be taken as in the last section:

$$z = f(x, y) = \frac{1}{2}Ax^2 + Bxy + \frac{1}{2}Cy^2 + \cdots . \quad \text{(A.28)}$$

Now one of the planes in the normal set is the xz-plane. This plane was also the first we considered in the proof of Meusnier's theorem. It intersects the surface in the curve $g(x)$ which, in the neighbourhood of the origin, is given by

$$g(x) = f(x, 0) = \frac{1}{2}Ax^2 + O(x^3), \quad \text{(A.29)}$$

and, from equation (A.25), we know that its curvature at the origin is equal to A. Similarly the curvature of the section by the yz-plane is equal to C.

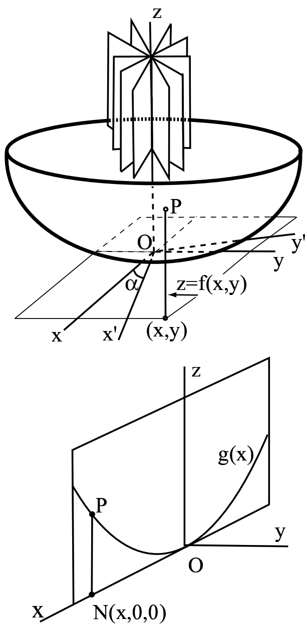


Figure A.6

Principal Axes

We exploit the freedom to choose any pair of orthogonal lines as axes in the xy-plane. If new $x'y'$-axes are rotated from the original by an angle α then we must set

$$x = \cos\alpha\, x' - \sin\alpha\, y', \quad \text{(A.30)}$$

$$y = \sin\alpha\, x' + \cos\alpha\, y'. \quad \text{(A.31)}$$

Abbreviate $c = \cos\alpha$ and $s = \sin\alpha$ and set $x = cx' - sy'$ and $y = sx' + cy'$ in the equation of the surface (A.28). In terms of these new coordinates

$$z = h(x', y') = \frac{1}{2}A(cx'-sy')^2 + B(cx'-sy')(sx'+cy') + \frac{1}{2}C(sx'+cy')^2 + \cdots . \quad \text{(A.32)}$$

Now the coefficient of the $x'y'$ cross term is equal to $[-Asc + B(c^2 - s^2) + Csc]$ and this will vanish if $(A - C)\sin 2\alpha = 2B\cos 2\alpha$ or $\tan 2\alpha = 2B/(A - C)$. There is always a solution for α and therefore we can always rotate axes so that the equation of the surface may be taken without a cross term. This defines the **principal axes** in the tangent plane at the given point.

Curvature in an arbitrary normal plane: Euler's formula

We now re-interpret Figure A.6 by assuming that the principal axes have been found and they have been chosen as the x-axis and y-axis. Therefore the equation of the surface with respect to the principal axes in the tangent plane is of the form

$$z = f(x, y) = \frac{1}{2}\kappa_1 x^2 + \frac{1}{2}\kappa_2 y^2 + \cdots\,, \tag{A.33}$$

where κ_1, κ_2 could be related to $A,\ B,\ C$. On the normal plane which includes the x principal axis we have $y = 0$ and $z = (1/2)\kappa_1 x^2$ so that the curvature of the section is κ_1. Similarly κ_2 is the curvature of the normal section which includes the y principal axis. κ_1 and κ_2 are called the **principal curvatures** of the surface at P.

We shall now find the curvature of the section made by a normal plane which makes an arbitrary angle α to one of the principal axes, say the x-axis. Once again we rotate axes, away from the principal axes, so that the new x' axis lies in the chosen plane and y' is orthogonal to it. This is achieved by exactly the same rotation as before, namely $x = cx' - sy'$ and $y = sx' + cy'$. The equation of the surface now takes the form:

$$z = h(x', y') = \frac{1}{2}\kappa_1(cx' - sy')^2 + \frac{1}{2}\kappa_2(sx' + cy')^2 + \cdots\,. \tag{A.34}$$

Now the chosen plane at the angle α is the plane with $y' = 0$ in the new coordinates so its section with the surface is given by

$$\begin{aligned} p(x') = h(x', 0) &= \frac{1}{2}\kappa_1(cx')^2 + \frac{1}{2}\kappa_2(sx')^2 + \cdots \\ &= \frac{1}{2}x'^2(\kappa_1\cos^2\!\alpha + \kappa_2\sin^2\!\alpha) + \cdots\,. \end{aligned} \tag{A.35}$$

We now evaluate its curvature at the origin using equation (A.4), giving:

EULER'S FORMULA

$$\boxed{\kappa(\alpha) = \kappa_1\cos^2\alpha + \kappa_2\sin^2\alpha} \tag{A.36}$$

for the curvature of the normal section made by a plane making an angle α with one of the principal normal planes.

Without loss of generality let us take $\kappa_1 > \kappa_2$, then we have

$$\kappa_1 - \kappa(\alpha) = (\kappa_1 - \kappa_2)\sin^2\alpha \geq 0, \tag{A.37}$$

$$\kappa(\alpha) - \kappa_2 = (\kappa_1 - \kappa_2)\cos^2\alpha \geq 0. \tag{A.38}$$

$$\kappa_1 \geq \kappa(\alpha) \geq \kappa_2. \tag{A.39}$$

Thus we have proved that the curvatures of normal sections at a point are such that the minimum and maximum values, the principal curvatures, are associated with orthogonal planes and the curvature on any other plane is given by the Euler formula. Note that we have not provided any way of calculating the curvature for an arbitrary surface for in general we do not have equations for the surface in the form of (A.28). The general study requires the machinery of differential geometry (see Bibliography) but for surfaces of revolution such as the ellipsoid we shall find that this not required.

Two definitions of average curvature

The mean curvature at a point on a surface is $$H = \frac{1}{2}(\kappa_1 + \kappa_2). \tag{A.40}$$

The Gaussian curvature at a point on a surface is $$G = \sqrt{\kappa_1 \kappa_2}. \tag{A.41}$$

These definitions are useful in various ways—for example, when we seek to approximate the surface of small part of the ellipsoid by a sphere.

Appendix B

Lagrange expansions

B.1 Introduction

We wish to investigate the inversion of a finite series such as

$$w = z + a_2z^2 + a_3z^3 + a_4z^4 + a_5z^5 \cdots \tag{B.1}$$

where both z and w are assumed to be small, less than 1, whilst the coefficients are of of order unity. The series we shall meet in the cartographic applications will typically be Taylor series truncated after a few terms. Now since $z^n \ll z$ for $z < 1$ and $n > 1$ we must have $z \approx w$ and we might expect it to be represented by a series of the form

$$z = b_1w + b_2w^2 + b_3w^3 + b_4w^4 + b_5w^5 + \cdots\,. \tag{B.2}$$

One way of finding the coefficients is to substitute the series for z into every term on the right hand side of (B.1) and compare coefficients of w^n on both sides. This is demonstrated explicitly in the next section. Fortunately a more general method exists, namely the Lagrange expansions defined in Section B.3. This is essential for the inversion of the eighth order series that we shall encounter.

A second category of problem is illustrated by a series of the form

$$w = z + c_2 \sin 2z + c_4 \sin 4z + c_6 \sin 6z + \cdots\,, \tag{B.3}$$

where we might have z and w as $O(1)$ whilst the coefficients c_n are small. The method of Lagrange expansions will show that there is an inverse given by

$$z = w + d_2 \sin 2w + d_4 \sin 4w + d_6 \sin 6w + \cdots\,. \tag{B.4}$$

B.2 Direct inversion of power series

The power series may be solved simply by back substitution, *i.e.* we substitute z from (B.2) into the terms on the right hand side of (B.1) and compare coefficients of w. If we retain only terms up to $O(w^4)$ we have

$$\begin{aligned} w &= (b_1 w + b_2 w^2 + b_3 w^3 + b_4 w^4) + a_2 w^2 (b_1 + b_2 w + b_3 w^2 + \cdots)^2 \\ &\quad + a_3 w^3 (b_1 + b_2 w + \cdots)^3 + a_4 w^4 (b_1 + \cdots)^4, \\ &= (b_1 w + b_2 w^2 + b_3 w^3 + b_4 w^4) + a_2 w^2 (b_1^2 + 2b_1 b_2 w + b_2^2 w^2 + 2b_1 b_3 w^2 + \cdots) \\ &\quad + a_3 w^3 (b_1^3 + 3b_1^2 b_2 w + \cdots) + a_4 w^4 b_1^4 + O(w^5). \end{aligned}$$

Comparing coefficients gives

$$\begin{aligned} w^1 &: \quad 1 = b_1, \\ w^2 &: \quad 0 = b_2 + a_2 b_1^2, \\ w^3 &: \quad 0 = b_3 + 2a_2 b_1 b_2 + a_3 b_1^3, \\ w^4 &: \quad 0 = b_4 + a_2(b_2^2 + 2b_1 b_3) + 3a_3 b_1^2 b_2 + a_4 b_1^4. \end{aligned}$$

These equations are solved in turn to give

$$b_1 = 1, \qquad b_2 = -a_2, \qquad b_3 = -a_3 + 2a_2^2, \qquad b_4 = -a_4 + 5a_2 a_3 - 5a_2^3. \tag{B.5}$$

This method is straightforward but becomes progressively harder as we go up to $O(z^8)$ terms. Moreover the method is inapplicable for the trigonometric series. Fortunately there is a more general and elegant approach.

B.3 Lagrange's theorem

The general form of the series (B.1, B.3) is

$$w = z + f(z), \tag{B.6}$$

with $|f(z)| \ll |z|$ and $w \approx z$. The theorem of Lagrange states that in a suitable domain the solution of this equation is

$$\boxed{z = w + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \left(\frac{d}{dw}\right)^{(k-1)} [f(w)]^k} \tag{B.7}$$

The proof of this theorem will be given in the last section of this appendix.

B.4 Application to a fourth order polynomial

Consider the finite polynomial

$$w = z + a_2 z^2 + a_3 z^3 + a_4 z^4, \tag{B.8}$$

which is a case of equation (B.6) with

$$f(z) = a_2 z^2 + a_3 z^3 + a_4 z^4.$$

We now apply the theorem with

$$f(w) = a_2 w^2 + a_3 w^3 + a_4 w^4.$$

In evaluating the inverse we shall only retain terms up to w^4 in the series for z (although the Lagrange expansion is infinite). Therefore in evaluating the power $[f(w)]^k$ we need retain only those powers of w which give terms no higher than w^4 after differentiating $(k-1)$ times. For example $[f(w)]^3$ has terms of order w^6, w^7, $\ldots w^{12}$ but only the first of these terms contributes after differentiating 2 times. No terms of order w^4 arise from $[f(w)]^4$ and higher powers. Therefore we keep only

$$\begin{aligned} f(w) &= a_2 w^2 + a_3 w^3 + a_4 w^4, \\ [f(w)]^2 &= w^4(a_2^2) + w^5(2a_2 a_3) + \cdots, \\ [f(w)]^3 &= w^6(a_2^3) + \cdots . \end{aligned}$$

Calculate the derivatives

$$\begin{aligned} f(w) &= a_2 w^2 + a_3 w^3 + a_4 w^4, \\ \frac{1}{2!} D[f(w)]^2 &= 2w^3(a_2^2) + 5w^4(a_2 a_3) + \cdots , \\ \frac{1}{3!} D^2[f(w)]^3 &= 5w^4(a_2^3) + \cdots . \end{aligned}$$

Substitute in Lagrange's expansion:

$$z = w - f(w) + \frac{1}{2!} D[f(w)]^2 - \frac{1}{3!} D^2[f(w)]^3 + \cdots .$$

The final result is

$$z = w - w^2 \left[a_2\right] - w^3 \left[a_3 - 2a_2^2\right] - w^4 \left[a_4 - 5a_2 a_3 + 5a_2^3\right] - \cdots . \tag{B.9}$$

The coefficients are in agreement with equation (B.5).

Modified fourth order polynomial

It will be convenient to consider a modified version of equation (B.8) with coefficients are of the form $a_n = b_n/n!$. In this case the above equations become

$$w = z + \frac{b_2}{2!}z^2 + \frac{b_3}{3!}z^3 + \frac{b_4}{4!}z^4, \tag{B.10}$$

$$z = w - \frac{p_2}{2!}w^2 - \frac{p_3}{3!}w^3 - \frac{p_4}{4!}w^4 + \cdots \tag{B.11}$$

where the p-coefficients are given by

$$p_2 = b_2, \qquad p_3 = b_3 - 3b_2^2, \qquad p_4 = b_4 - 10b_2b_3 + 15b_2^3. \tag{B.12}$$

Alternative notation

For the applications to cartography it is convenient to use the following notation for the direct and inverse series:

$$z = \zeta + \frac{b_2}{2!}\zeta^2 + \frac{b_3}{3!}\zeta^3 + \frac{b_4}{4!}\zeta^4, \tag{B.13}$$

$$\zeta = z - \frac{p_2}{2!}z^2 - \frac{p_3}{3!}z^3 - \frac{p_4}{4!}z^4 + \cdots \tag{B.14}$$

B.5 Application to a trigonometric series

Consider equation (B.6), that is

$$w(z) = z + f(z), \tag{B.15}$$

with $f(z)$ defined by the following finite trigonometric series:

$$f(z) = b_2 \sin 2z + b_4 \sin 4z + b_6 \sin 6z + b_8 \sin 8z, \tag{B.16}$$

where the coefficients b_n are small enough for the condition $|f(z)| \ll z,\, w$ to be valid; note that we are assuming that w and z are of order unity. For the applications we have in mind we shall have $|b_n| = O(e^n)$ where e is the eccentricity of the ellipsoid. In deriving the inversion we shall truncate the infinite Lagrange expansion at terms of order e^8; thus we retain terms proportional to $b_2,\, b_4,\, b_2^2,\, b_6,\, b_2b_4,\, b_2^3,\, b_8, b_2b_6,\, b_2^2b_4,\, b_4^2,\, b_2^4$ and drop higher powers.

In the following steps we make use of several trigonometric identities from Appendix C.

$$
\begin{aligned}
f(w) &= b_2 \sin 2w + b_4 \sin 4w + b_6 \sin 6w + b_8 \sin 8w, \\
[f(w)]^2 &= b_2^2 \sin^2 2w + 2b_2 b_4 \sin 2w \sin 4w + b_4^2 \sin^2 4w + 2b_2 b_6 \sin 2w \sin 6w + \cdots \\
&= \frac{1}{2} b_2^2 (1 - \cos 4w) + b_2 b_4 (\cos 2w - \cos 6w) \\
&\quad + \frac{1}{2} b_4^2 (1 - \cos 8w) + b_2 b_6 (\cos 4w - \cos 8w) + \cdots \\
[f(w)]^3 &= b_2^3 \sin^3 2w + 3b_2^2 b_4 \sin^2 2w \sin 4w + \cdots \\
&= \frac{1}{4} b_2^3 (3 \sin 2w - \sin 6w) + \frac{3}{4} b_2^2 b_4 (2 \sin 4w - \sin 8w) + \cdots \\
[f(w)]^4 &= b_2^4 \sin^4 2w + \cdots = \frac{1}{8} b_2^4 (3 - 4 \cos 4w + \cos 8w) + \cdots .
\end{aligned}
$$

Calculate the derivatives

$$
\begin{aligned}
f(w) &= b_2 \sin 2w + b_4 \sin 4w + b_6 \sin 6w + b_8 \sin 8w, \\
\frac{1}{2!} D[f(w)]^2 &= b_2^2 \sin 4w + b_2 b_4 (-\sin 2w + 3 \sin 6w) \\
&\quad + 2b_4^2 \sin 8w + 2b_2 b_6 (-\sin 4w + 2 \sin 8w) + \cdots \\
\frac{1}{3!} D^2[f(w)]^3 &= \frac{1}{2} b_2^3 (-\sin 2w + 3 \sin 6w) + 4b_2^2 b_4 (-\sin 4w + 2 \sin 8w) + \cdots \\
\frac{1}{4!} D^3[f(w)]^4 &= \frac{4}{3} b_2^4 (-\sin 4w + 2 \sin 8w) + \cdots .
\end{aligned}
$$

Finally, substituting into

$$
z = w - f(w) + \frac{1}{2!} D[f(w)]^2 - \frac{1}{3!} D^2[f(w)]^3 + \frac{1}{4!} D^3[f(w)]^4 + \cdots
$$

and grouping terms according to the trigonometric functions gives

$$
\boxed{z = w + d_2 \sin 2w + d_4 \sin 4w + d_6 \sin 6w + d_8 \sin 8w + \cdots ,} \tag{B.17}
$$

where

$$
\begin{aligned}
d_2 &= -b_2 - b_2 b_4 + \frac{1}{2} b_2^3, \\
d_4 &= -b_4 + b_2^2 - 2b_2 b_6 + 4b_2^2 b_4 - \frac{4}{3} b_2^4, \\
d_6 &= -b_6 + 3b_2 b_4 - \frac{3}{2} b_2^3, \\
d_8 &= -b_8 + 2b_4^2 + 4b_2 b_6 - 8b_2^2 b_4 + \frac{8}{3} b_2^4.
\end{aligned} \tag{B.18}
$$

B.6 Application to an eighth order polynomial

We now invert a series of the form

$$w = z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 + a_6 z^6 + a_7 z^7 + a_8 z^8. \tag{B.19}$$

retaining only the terms up to w^8 in the series for z. This problem is a trivial generalisation of the derivation for the fourth order series developed in Section B.4, only the algebra is a little more involved. We set $f(w) = a_2 w^2 + a_3 w^3 \cdots$ in the Lagrange expansion and evaluate the powers of $[f(w)]^k$; recall that we need retain only those powers of w which give terms no higher than w^8 after differentiating $k-1$ times.

$$\begin{aligned}
f(w) &= a_2 w^2 + a_3 w^3 + a_4 w^4 + a_5 w^5 + a_6 w^6 + a_7 w^7 + a_8 w^8, \\
[f(w)]^2 &= w^4(a_2^2) \\
&\quad + w^5(2a_2a_3) \\
&\quad + w^6(2a_2a_4 + a_3^2) \\
&\quad + w^7(2a_2a_5 + 2a_3a_4) \\
&\quad + w^8(2a_2a_6 + 2a_3a_5 + a_4^2) \\
&\quad + w^9(2a_2a_7 + 2a_3a_6 + 2a_4a_5) + O(w^{10}), \\
[f(w)]^3 &= w^6(a_2^3) \\
&\quad + w^7(3a_2^2a_3) \\
&\quad + w^8(3a_2^2a_4 + 3a_2a_3^2) \\
&\quad + w^9(3a_2^2a_5 + 6a_2a_3a_4 + a_3^3) \\
&\quad + w^{10}(3a_2^2a_6 + 6a_2a_3a_5 + 3a_2a_4^2 + 3a_3^2a_4) + O(w^{11}), \\
[f(w)]^4 &= w^8(a_2^4) \\
&\quad + w^9(4a_2^3a_3) \\
&\quad + w^{10}(4a_2^3a_4 + 6a_2^2a_3^2) \\
&\quad + w^{11}(4a_2^3a_5 + 12a_2^2a_3a_4 + 4a_2a_3^3) + O(w^{12}), \\
[f(w)]^5 &= w^{10}(a_2^5) \\
&\quad + w^{11}(5a_2^4a_3) \\
&\quad + w^{12}(5a_2^4a_4 + 10a_2^3a_3^2) + O(w^{13}), \\
[f(w)]^6 &= w^{12}(a_2^6) \\
&\quad + w^{13}(6a_2^5a_3) + O(w^{14}), \\
[f(w)]^7 &= w^{14}(a_2^7) + O(w^{15}), \\
[f(w)]^8 &= O(w^{16}).
\end{aligned}$$

Evaluate the derivatives, (writing D for d/dw):

$$
\begin{aligned}
f(w) &= \ a_2w^2 + a_3w^3 + a_4w^4 + a_5w^5 + a_6w^6 + a_7w^7 + a_8w^8, \\
\frac{1}{2!}D[f(w)]^2 &= +\,2w^3(a_2^2) \\
&\quad + \frac{5}{2}w^4(2a_2a_3) \\
&\quad + 3w^5(2a_2a_4 + a_3^2) \\
&\quad + \frac{7}{2}w^6(2a_2a_5 + 2a_3a_4) \\
&\quad + 4w^7(2a_2a_6 + 2a_3a_5 + a_4^2) \\
&\quad + \frac{9}{2}w^8(2a_2a_7 + 2a_3a_6 + 2a_4a_5) + O(w^9), \\
\frac{1}{3!}D^2[f(w)]^3 &= +\,5w^4(a_2^3) \\
&\quad + 7w^5(3a_2^2a_3) \\
&\quad + \frac{28}{3}w^6(3a_2^2a_4 + 3a_2a_3^2) \\
&\quad + 12w^7(3a_2^2a_5 + 6a_2a_3a_4 + a_3^3) \\
&\quad + 15w^8(3a_2^2a_6 + 6a_2a_3a_5 + 3a_2a_4^2 + 3a_3^2a_4) + O(w^9), \\
\frac{1}{4!}D^3[f(w)]^4 &= +\,14w^5(a_2^4) \\
&\quad + 21w^6(4a_2^3a_3) \\
&\quad + 30w^7(4a_2^3a_4 + 6a_2^2a_3^2) \\
&\quad + \frac{165}{4}w^8(4a_2^3a_5 + 12a_2^2a_3a_4 + 4a_2a_3^3) + O(w^9), \\
\frac{1}{5!}D^4[f(w)]^5 &= +\,42w^6(a_2^5) \\
&\quad + 66w^7(5a_2^4a_3) \\
&\quad + 99w^8(5a_2^4a_4 + 10a_2^3a_3^2) + O(w^9), \\
\frac{1}{6!}D^5[f(w)]^6 &= +\,132w^7(a_2^6) \\
&\quad + \frac{429}{2}w^8(6a_2^5a_3) + O(w^9), \\
\frac{1}{7!}D^6[f(w)]^7 &= +\,429w^8(a_2^7) + O(w^9), \\
\frac{1}{8!}D^7[f(w)]^8 &= \ O(w^9).
\end{aligned}
$$

Substitute the above in the Lagrange expansion:

$$
\begin{aligned}
z = w - f(w) &+ \frac{1}{2!}D[f(w)]^2 - \frac{1}{3!}D^2[f(w)]^3 + \frac{1}{4!}D^3[f(w)]^4 - \frac{1}{5!}D^4[f(w)]^5 \\
&+ \frac{1}{6!}D^5[f(w)]^6 - \frac{1}{7!}D^6[f(w)]^7 + \frac{1}{8!}D^7[f(w)]^8. \qquad \text{(B.20)}
\end{aligned}
$$

Final result for basic eighth order series

The inverse of the series

$$w = z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 + a_6 z^6 + a_7 z^7 + a_8 z^8, \tag{B.21}$$

is

$$\begin{aligned} z = w &- w^2[a_2] \\ &- w^3\left[a_3 - 2a_2^2\right] \\ &- w^4\left[a_4 - 5a_2a_3 + 5a_2^3\right] \\ &- w^5\left[a_5 - 3(2a_2a_4 + a_3^2) + 21a_2^2a_3 - 14a_2^4\right] \\ &- w^6\left[a_6 - 7(a_2a_5 + a_3a_4) + 28(a_2^2a_4 + a_2a_3^2) - 84a_2^3a_3 + 42a_2^5\right] \\ &- w^7\left[a_7 - 4(2a_2a_6 + 2a_3a_5 + a_4^2) + 12(3a_2^2a_5 + 6a_2a_3a_4 + a_3^3)\right. \\ &\qquad \left. - 30(4a_2^3a_4 + 6a_2^2a_3^2) + 330a_2^4a_3 - 132a_2^6\right] \\ &- w^8\left[a_8 - 9(a_2a_7 + a_3a_6 + a_4a_5) + 15(3a_2^2a_6 + 6a_2a_3a_5 + 3a_2a_4^2 + 3a_3^2a_4)\right. \\ &\qquad - 165(a_2^3a_5 + 3a_2^2a_3a_4 + a_2a_3^3) + 99(5a_2^4a_4 + 10a_2^3a_3^2) \\ &\qquad \left. - 1287a_2^5a_3 + 429a_2^7\right] \end{aligned} \tag{B.22}$$

B.7 Application to a modified eighth order series

Replacing a_n by $b_n/n!$ gives the following pair of inverse series:

$$w = z + \frac{b_2}{2!}z^2 + \frac{b_3}{3!}z^3 + \frac{b_4}{4!}z^4 + \frac{b_5}{5!}z^5 + \frac{b_6}{6!}z^6 + \frac{b_7}{7!}z^7 + \frac{b_8}{8!}z^8. \tag{B.23}$$

$$z = w - \frac{p_2}{2!}w^2 - \frac{p_3}{3!}w^3 - \frac{p_4}{4!}w^4 - \frac{p_5}{5!}w^5 - \frac{p_6}{6!}w^6 - \frac{p_7}{7!}w^7 - \frac{p_8}{8!}w^8. \tag{B.24}$$

where

$$\begin{aligned} p_2 &= [b_2] \\ p_3 &= \left[b_3 - 3b_2^2\right] \\ p_4 &= \left[b_4 - 10b_2b_3 + 15b_2^3\right] \\ p_5 &= \left[b_5 - (15b_2b_4 + 10b_3^2) + 105b_2^2b_3 - 105b_2^4\right] \\ p_6 &= \left[b_6 - (21b_2b_5 + 35b_3b_4) + (210b_2^2b_4 + 280b_2b_3^2) - 1260b_2^3b_3 + 945b_2^5\right] \\ p_7 &= \left[b_7 - (28b_2b_6 + 56b_3b_5 + 35b_4^2) + (378b_2^2b_5 + 1260b_2b_3b_4 + 280b_3^3)\right. \\ &\qquad \left. - (3150b_2^3b_4 + 6300b_2^2b_3^2) + 17325b_2^4b_3 - 10395b_2^6\right] \\ p_8 &= [b_8 - (36b_2b_7 + 84b_3b_6 + 126b_4b_5) \\ &\qquad + (630b_2^2b_6 + 2520b_2b_3b_5 + 1575b_2b_4^2 + 2100b_3^2b_4) \\ &\qquad - (6930b_2^3b_5 + 34650b_2^2b_3b_4 + 15400b_2b_3^3) \\ &\qquad + (51975b_2^4b_4 + 138600b_2^3b_3^2) - 270270b_2^5b_3 + 135135b_2^7] \end{aligned} \tag{B.25}$$

Comment: these results are extended to 12th order series in papers by (a) W E Bleieck and (b) W G Bickley and J C P Miller. See bibliography.

B.8 Application to series for TME

In evaluating the inverse of the complex series that arises in the derivation of the transverse Mercator projection on the ellipsoid (TME) we have the following coefficients

$$\begin{array}{llll} b_2 = i\,s & b_3 = \; c^2 W_3 & b_4 = \; isc^2 W_4 & b_5 = \; c^4 W_5 \\ b_6 = i\,sc^4 W_6 & b_7 = \; c^6 \overline{W}_7 & b_8 = i\,sc^6 \overline{W}_8 & \end{array} \tag{B.26}$$

where $i = \sqrt{-1}$, $s = \sin\phi,\; c = \cos\phi,\; t = \tan\phi$ and the W functions are of the form

$$\begin{aligned} W_3 &= \beta - t^2 \\ W_4 &= 4\beta^2 + \beta - t^2 \\ W_5 &= 4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4 \\ W_6 &= 8\beta^4(11-24t^2) - 28\beta^3(1-6t^2) + \beta^2(1-32t^2) - 2\beta t^2 + t^4 \\ \overline{W}_7 &= 61 - 479t^2 + 179t^4 - t^6 \\ \overline{W}_8 &= 1385 - 3111t^2 + 543t^4 - t^6, \end{aligned} \tag{B.27}$$

where β is defined in equation (5.53). Substituting for the b-coefficients in (B.25) gives

$$\begin{aligned} p_2 &= ict\;\;[\,1\,] \\ p_3 &= \;\; c^2 \left[W_3 + 3t^2\right] \\ p_4 &= ic^3 t \left[W_4 - 10W_3 - 15t^2\right] \\ p_5 &= \;\; c^4 \left[W_5 + (15t^2 W_4 - 10W_3^2) - 105t^2 W_3 - 105t^4\right] \\ p_6 &= ic^5 t \left[W_6 - (21W_5 + 35W_3W_4) - (210t^2 W_4 - 280W_3^2) \;+1260t^2 W_3 + 945t^4\right] \\ p_7 &= \;\; c^6 \Big[\overline{W}_7 + (28t^2 W_6 - 56W_3W_5 + 35t^2 W_4^2) \;-(378t^2 W_5 + 1260t^2 W_3 W_4 - 280W_3^3) \\ &\qquad -(3150t^4 W_4 - 6300t^2 W_3^2) + 17325t^4 W_3 + 10395t^6\Big] \\ p_8 &= ic^7 t \Big[\overline{W}_8 - (36\overline{W}_7 + 84W_3W_6 + 126W_4W_5) \\ &\qquad -(630t^2 W_6 - 2520W_3W_5 + 1575t^2 W_4^2 - 2100W_3^2 W_4) \\ &\qquad +(6930t^2 W_5 + 34650t^2 W_3 W_4 - 15400W_3^3) \\ &\qquad +(51975t^4 W_4 - 138600t^2 W_3^2) \;-270270t^4 W_3 - 135135t^6\Big] \end{aligned} \tag{B.28}$$

Now substitute for the W. For $p_2, \ldots p_6$ we use the expressions given in (B.27). For p_7, p_8 we use the spherical approximation (5.58) for all the terms on the right hand sides. That is we set $\beta = 1$ in $W_3, \ldots W_6$ on the right hand sides using the approximations

$$\begin{aligned} W_3 &\to \overline{W}_3 = 1 - t^2, \\ W_4 &\to \overline{W}_4 = 5 - t^2, \\ W_5 &\to \overline{W}_5 = 5 - 18t^2 + t^4, \\ W_6 &\to \overline{W}_6 = 61 - 58t^2 + t^4. \end{aligned} \tag{B.29}$$

The p-coefficients now become

$$
\begin{aligned}
p_2 &= ict\ \ [\,1\,] \\
p_3 &= \quad c^2\left[\beta - t^2 + 3t^2\right] \\
p_4 &= ic^3 t\left[4\beta^2 + \beta - t^2 - 10(\beta - t^2) - 15t^2\right] \\
p_5 &= \quad c^4\left[4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4\right. \\
&\qquad + 15t^2(4\beta^2 + \beta - t^2) \\
&\qquad \left.- 10(\beta - t^2)^2 - 105t^2(\beta - t^2) - 105t^4\right] \\
p_6 &= ic^5 t\left[8\beta^4(11 - 24t^2) - 28\beta^3(1-6t^2) + \beta^2(1-32t^2) - 2\beta t^2 + t^4\right. \\
&\qquad - 21\left\{4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4\right\} \\
&\qquad - 35(\beta - t^2)(4\beta^2 + \beta - t^2) - 210t^2(4\beta^2 + \beta - t^2) \\
&\qquad \left.+ 280(\beta - t^2)^2 + 1260t^2(\beta - t^2) + 945t^4\right] \\
\bar{p}_7 &= \quad c^6\left[61 - 479t^2 + 179t^4 - t^6 + 28t^2(61 - 58t^2 + t^4) - 56(1-t^2)(5 - 18t^2 + t^4)\right. \\
&\qquad + 35t^2(5-t^2)^2 - 378t^2(5 - 18t^2 + t^4) - 1260t^2(1-t^2)(5-t^2) \\
&\qquad + 280(1-t^2)^3 - 3150t^4(5-t^2) + 6300t^2(1-t^2)^2 + 17325t^4(1-t^2) \\
&\qquad \left.+ 10395t^6\right] \\
\bar{p}_8 &= ic^7 t\left[1385 - 3111t^2 + 543t^4 - t^6 - 36(61 - 479t^2 + 179t^4 - t^6)\right. \\
&\qquad - 84(1-t^2)(61 - 58t^2 + t^4) - 126(5-t^2)(5 - 18t^2 + t^4) \\
&\qquad - 630t^2(61 - 58t^2 + t^4) + 2520(1-t^2)(5 - 18t^2 + t^4) \\
&\qquad - 1575t^2(5-t^2)^2 + 2100(1-t^2)^2(5-t^2) + 6930t^2(5 - 18t^2 + t^4) \\
&\qquad + 34650t^2(1-t^2)(5-t^2) - 15400(1-t^2)^3 + 51975t^4(5-t^2) \\
&\qquad \left.- 138600t^2(1-t^2)^2 - 270270t^4(1-t^2) - 135135t^6\right]
\end{aligned}
$$

Note that we have changed p_7, p_8 to $\bar{p}_7$, $\bar{p}_8$ to show that these coefficients have been evaluated in the spherical approximation. Finally, simplifying these expressions gives

$$
\begin{aligned}
p_2 &= ict\ \ [\,1\,] \\
p_3 &= \quad c^2\left[\beta + 2t^2\right] \\
p_4 &= ic^3 t\left[4\beta^2 - 9\beta - 6t^2\right] \\
p_5 &= \quad c^4\left[4\beta^3(1-6t^2) - \beta^2(9 - 68t^2) - 72\beta t^2 - 24t^4\right] \\
p_6 &= ic^5 t\left[8\beta^4(11 - 24t^2) - 84\beta^3(3 - 8t^2) + 225\beta^2(1 - 4t^2) + 600\beta t^2 + 120t^4\right] \\
\bar{p}_7 &= \quad c^6\left[61 + 662t^2 + 1320t^4 + 720t^6\right] \\
\bar{p}_8 &= ic^7 t\left[-1385 - 7266t^2 - 10920t^4 - 5040t^6\right]
\end{aligned}
\tag{B.30}
$$

B.9 Proof of the Lagrange expansion

This derivation of the Lagrange expansion is included since it is to be found only in older textbooks—see bibliography. This account is based on a simplified version of that in Whittaker's Modern Analysis (1902!!) where there is a more general statement of the theorem. The derivation requires an excursion into complex analysis. In particular we require three results which follow from Cauchy's theorem. Since these results can be found in most texts on complex analysis we quote them without proof.

Definition: a function $f(z)$ is **analytic** in some domain D if it is single valued and differentiable within D, except possibly at a finite number of points, the **singularities** of $f(z)$. If no point of D is a singularity then we say that $f(z)$ is **regular**.

- **Cauchy's integral formula**: let $f(z)$ be an analytic function, regular within a closed contour C and continuous within and on C, and let a be a point within C. If in addition $f(z)$ has derivatives of all orders, then the n-th derivative at a is

$$f^{(n)}(a) = \frac{n!}{2\pi i}\oint_C \frac{f(z)}{(z-a)^{n+1}}\,dz. \tag{B.31}$$

- The following result is usually found as a corollary to the proof of **the principle of the argument**. If $f(z)$ and $g(z)$ are regular within and on a closed contour C and $f(z)$ has a simple zero at $z = a$ then

$$g(a) = \frac{1}{2\pi i}\oint_C \frac{g(z)f'(z)}{f(z)}\,dz. \tag{B.32}$$

- **Rouché's theorem**: if $f(z)$ and $g(z)$ are two functions regular within and on a closed contour C, on which $f(z)$ does not vanish and also $|g(z)| < |f(z)|$, then $f(z)$ and $f(z) + g(z)$ have the same number of zeroes within C.

Let $p(z)$ be regular within and on a closed contour C and let there be a *single* simple zero at the point $z = w$ inside C. Consider the equation

$$p(z) = t, \tag{B.33}$$

where t is a constant such that

$$|p(z)| > |t| \quad \text{at all points of } C. \tag{B.34}$$

By Rouché's theorem (with $f \to p$ and $g \to -t$) we see that $p(z)$ and $p(z)-t$ have the same number of zeroes inside C, namely one. The zero of $p(z)$ is of course $z = w$: let the zero of $p(z) - t$ be $z = a$. Therefore setting $f(z) = p(z) - t$ and $g(z) = z$ in equation (B.32), noting that t is a constant, we find the solution $z = a$ of (B.33) is

$$z = a = \frac{1}{2\pi i}\oint_C \frac{zp'(z)}{p(z)-t}dz. \tag{B.35}$$

Expanding the integrand

$$a = \frac{1}{2\pi i}\oint_C \frac{zp'(z)}{p(z)}\left[1+\sum_1^\infty\left(\frac{t}{p(z)}\right)^n\right]dz. \tag{B.36}$$

Since $|t| < |p(z)|$ on C the series is convergent and we can integrate term by term to find

$$a = \sum_0^\infty A_n t^n, \tag{B.37}$$

where

$$A_0 = \frac{1}{2\pi i}\oint_C \frac{zp'(z)}{p(z)}dz, \qquad A_n = \frac{1}{2\pi i}\oint_C \frac{zp'(z)}{[p(z)]^{n+1}}dz \quad (n \geq 1). \tag{B.38}$$

Now since $p(z)$ has a simple zero at $z = w$ the first integral may be integrated by setting $g(z) = z$ in equation (B.32).

$$A_0 = w. \tag{B.39}$$

For the second integral we integrate by parts. The integral of the total derivative is zero because the change in a single valued function around a closed curve is zero. Therefore

$$A_n = \frac{1}{2\pi i}\frac{1}{n}\oint_C \frac{1}{[p(z)]^n}dz, \quad (n \geq 1). \tag{B.40}$$

Now set

$$p(z) = (z-w)q(z) = \frac{z-w}{r(z)}. \tag{B.41}$$

so that

$$A_n = \frac{1}{2\pi i}\frac{1}{n}\oint_C \frac{[r(z)]^n}{(z-w)^n}dz. \tag{B.42}$$

$p(z)$ has one zero inside C, at $z = w$, so $q(z)$ will have no zeroes within C and $r(z)$ will have no poles within C. Using the Cauchy integral formula (B.31) A_n becomes

$$\begin{aligned} A_n &= \frac{1}{n!}D_z^{(n-1)}[r(z)]^n\Big|_{z=w} \\ &= \frac{1}{n!}D_w^{(n-1)}[r(w)]^n \qquad (n \geq 1). \end{aligned} \tag{B.43}$$

Therefore the solution of

$$\frac{z-w}{r(z)} = t \tag{B.44}$$

is given by

$$z = a = w + \sum_1^\infty \frac{t^n}{n!}D_w^{(n-1)}[r(w)]^n. \tag{B.45}$$

Finally we set

$$f(z) = -t\, r(z), \tag{B.46}$$

so that equation (B.44) becomes

$$w = z + f(z), \tag{B.47}$$

with the solution

$$z = w + \sum_{1}^{\infty} \frac{(-1)^n}{n!} D_w^{(n-1)} [f(w)]^n. \tag{B.48}$$

This is the form of the expansion given in Section B.3. The domain of validity is discussed in the textbooks. In the current applications we start from convergent series for $w(z)$ and find that the above series for $z(w)$ is also convergent.

Appendix C

Plane Trigonometry

A brief reminder of some identities from plane trigonometry which are required at various points in the main text.

basic definition	$\exp ix = \cos x + i \sin x$	(C.1)
	$\exp i(x+y) = (\cos x + i \sin x)(\cos y + i \sin y)$	(C.2)
real part of (C.2)	$\cos(x+y) = \cos x\ \cos y - \sin x\ \sin y$	(C.3)
$y \to -y$	$\cos(x-y) = \cos x\ \cos y + \sin x\ \sin y$	(C.4)
imag. part of (C.2)	$\sin(x+y) = \sin x\ \cos y + \cos x\ \sin y$	(C.5)
$y \to -y$	$\sin(x-y) = \sin x\ \cos y - \cos x\ \sin y$	(C.6)
(C.5) / (C.3)	$\tan(x+y) = \dfrac{\tan x + \tan y}{1 - \tan x \tan y}$	(C.7)
(C.6) / (C.4)	$\tan(x-y) = \dfrac{\tan x - \tan y}{1 + \tan x \tan y}$	(C.8)
(C.5) + (C.6)	$\sin x \cos y = \dfrac{1}{2}\left[\sin(x+y) + \sin(x-y)\right]$	(C.9)
$x \leftrightarrow y$	$\cos x \sin y = \dfrac{1}{2}\left[\sin(x+y) - \sin(x-y)\right]$	(C.10)
(C.3) + (C.4)	$\cos x \cos y = \dfrac{1}{2}\left[\cos(x+y) + \cos(x-y)\right]$	(C.11)
(C.4) – (C.3)	$\sin x \sin y = \dfrac{1}{2}\left[\cos(x-y) - \cos(x+y)\right]$	(C.12)
$x \pm y \to x,\ y$ in (C.9)	$\sin x + \sin y = 2 \sin \dfrac{x+y}{2} \cos \dfrac{x-y}{2}$	(C.13)
$x \pm y \to x,\ y$ in (C.10)	$\sin x - \sin y = 2 \cos \dfrac{x+y}{2} \sin \dfrac{x-y}{2}$	(C.14)
$x \pm y \to x,\ y$ in (C.11)	$\cos x + \cos y = 2 \cos \dfrac{x+y}{2} \cos \dfrac{x-y}{2}$	(C.15)
$x \pm y \to x,\ y$ in (C.12)	$\cos x - \cos y = -2 \sin \dfrac{x+y}{2} \sin \dfrac{x-y}{2}$	(C.16)

$y = x$ in (C.4) $$1 = \cos^2 x + \sin^2 x \tag{C.17}$$

$$\sec^2 x = 1 + \tan^2 x \tag{C.18}$$

$$\csc^2 x = \operatorname{cosec}^2 x = 1 + \cot^2 x \tag{C.19}$$

$y = x$ in (C.3) $$\cos 2x = \cos^2 x - \sin^2 x \tag{C.20}$$

use (C.17) $$= 1 - 2\sin^2 x \tag{C.21}$$

use (C.17) $$= 2\cos^2 x - 1 \tag{C.22}$$

$y = x$ in (C.5) $$\sin 2x = 2\sin x \ \cos x \tag{C.23}$$

from (C.21) $$\sin^2 x = \frac{1}{2}\left[1 - \cos 2x\right] \tag{C.24}$$

from (C.22) $$\cos^2 x = \frac{1}{2}\left[1 + \cos 2x\right] \tag{C.25}$$

use (C.21) $$\sin^3 x = \frac{1}{2}\sin x\left[1 - \cos 2x\right]$$

use (C.9) $$= \frac{1}{2}\sin x - \frac{1}{4}\left[\sin 3x - \sin x\right]$$

$$= \frac{3}{4}\sin x - \frac{1}{4}\sin 3x \tag{C.26}$$

from (C.26) $$\sin^3 x \ \cos x = \frac{3}{4}\sin x\cos x - \frac{1}{4}\sin 3x\cos x$$

use (C.9) $$= \frac{3}{8}\sin 2x - \frac{1}{8}[\sin 4x + \sin 2x]$$

$$= \frac{1}{8}\left[2\sin 2x - \sin 4x\right] \tag{C.27}$$

from (C.24) $$\sin^4 x = \frac{1}{4}\left[1 - \cos 2x\right]^2$$

use (C.25) $$= \frac{1}{4}\left[1 - 2\cos 2x + \frac{1}{2} + \frac{1}{2}\cos 4x\right]$$

$$= \frac{1}{8}\left[3 - 4\cos 2x + \cos 4x\right] \tag{C.28}$$

use (C.25) $$\cos^3 x = \frac{1}{2}\cos x\left[1 + \cos 2x\right]$$

use (C.11) $$= \frac{1}{2}\cos x + \frac{1}{4}\left[\cos 3x + \cos x\right]$$

$$= \frac{3}{4}\cos x + \frac{1}{4}\cos 3x \tag{C.29}$$

NOTATION
$$S_k \equiv \sin kx \tag{C.30}$$
$$C_k \equiv \cos kx \tag{C.31}$$

Hence
$$\sin^2 x = \frac{1}{2}\left[1 - C_2\right] \tag{C.32}$$
$$\sin^3 x = \frac{1}{4}\left[3S - S_3\right] \tag{C.33}$$
$$\sin^4 x = \frac{1}{8}\left[3 - 4C_2 + C_4\right] \tag{C.34}$$
$$\sin^5 x = \frac{1}{16}\left[10S - 5S_3 + S_5\right] \tag{C.35}$$
$$\sin^6 x = \frac{1}{32}\left[10 - 15C_2 + 6C_4 - C_6\right] \tag{C.36}$$
$$\sin^7 x = \frac{1}{64}\left[35S - 21S_3 + 7S_5 - S_7\right] \tag{C.37}$$
$$\sin^8 x = \frac{1}{128}\left[35 - 56C_2 + 28C_4 - 8C_6 + C_8\right] \tag{C.38}$$

$$\sin x \, \cos x = \frac{1}{2}\left[S_2\right] \tag{C.39}$$
$$\sin^3 x \, \cos x = \frac{1}{8}\left[2S_2 - S_4\right] \tag{C.40}$$
$$\sin^5 x \, \cos x = \frac{1}{32}\left[5S_2 - 4S_4 + S_6\right] \tag{C.41}$$
$$\sin^7 x \, \cos x = \frac{1}{128}\left[14S_2 - 14S_4 + 6S_6 - S_8\right] \tag{C.42}$$

Hyperbolic functions

The basic definitions are

$$\cosh x = \frac{e^x + e^{-x}}{2}, \qquad \sinh x = \frac{e^x - e^{-x}}{2}. \tag{C.43}$$

From equation (C.1) the corresponding equations are

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}, \qquad \sin x = \frac{e^{ix} - e^{-ix}}{2i}, \tag{C.44}$$

so we can immediately deduce that

$$\cos ix = \cosh x, \qquad \sin ix = i \sinh x, \qquad \tan ix = i \tanh x, \tag{C.45}$$
$$\cosh ix = \cos x, \qquad \sinh ix = i \sin x, \qquad \tanh ix = i \tan x. \tag{C.46}$$

These identities can be used to derive all the hyperbolic formulae from the trigonometric identities simply by replacing x and y by ix and iy. This effectively changes all cosine terms to cosh. Each sine term becomes $i\sinh$ and where there is a single sinh in each term of the identity an overall factor of i will cancel. Terms which have a product of two sines will become a product of two $i\sinh$ terms giving an overall sign change. Likewise for the tangent terms. We list only the identities corresponding to (C.3–C.8) and (C.17–C.23).

$$\cosh(x \pm y) = \cosh x\ \cosh y \pm \sinh x\ \sinh y \tag{C.47}$$

$$\sinh(x \pm y) = \sinh x\ \cosh y \pm \cosh x\ \sinh y \tag{C.48}$$

$$\tanh(x \pm y) = \frac{\tanh x \pm \tanh y}{1 \pm \tanh x \tanh y} \tag{C.49}$$

$$1 = \cosh^2 x - \sinh^2 x \tag{C.50}$$

$$\mathrm{sech}^2 x = 1 - \tanh^2 x \tag{C.51}$$

$$\mathrm{cosech}^2 x = \coth^2 x - 1 \tag{C.52}$$

$$\begin{aligned} \cosh 2x &= \cosh^2 x + \sinh^2 x \\ &= 1 + 2\sinh^2 x \\ &= 2\cosh^2 x - 1 \end{aligned} \tag{C.53}$$

$$\sinh 2x = 2\sinh x \cosh x \tag{C.54}$$

Appendix D

Spherical trigonometry

D.1 Introduction

A great circle on a sphere is defined by the intersection of any plane through the centre of the sphere with the surface of the sphere. Any two points on the sphere must lie on some great circle and the shorter part of that great circle is also the shortest distance between the points. In general three great circles define a spherical triangle (Figure D.1) and this appendix develops the trigonometry of such triangles. There are many (old) text books and we recommend Todhunter's book on Spherical Trigonometry—see Bibliography.

Consider the three great circles defining the triangle ABC: they meet again in the points A_1, B_1 and C_1 defining the triangle $A_1B_1C_1$. In fact they define eight triangles since each pair of geodesics bounds four triangles but ABC and $A_1B_1C_1$ are counted three times. (Think of slicing an apple into eight pieces with three diametral cuts). Note that we do not consider the 'improper' triangles such as that formed by the *interior* arcs BA , BC together with the *exterior* arc AC_1A_1C. Such improper triangles have one angle greater than π. Their solution presents no difficulty but we refer to Todhunter's book for details.

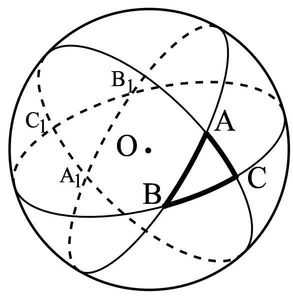


Figure D.1

We now restrict attention to triangles in which the angles are less than or equal to π. The case of all equal to π is degenerate for the 'triangle' must then be three points on one great circle with the sum of the angles equal to 3π and the sum of the sides equal to 2π (on the unit sphere). The rigorous proof of this last statement is to be found in Euclid: Book 11, Proposition 21—see Bibliography. We shall see that it has as a corollary that the sum of the angles of a spherical triangle is greater than π. This lower bound is approached by small triangles (sides much less than the radius) that are almost planar.

Figure D.2 shows the spherical triangle in more detail: A, B, C label the vertices and also give the measure (in radians) of the angles of the spherical triangle and the angles between planes OAC, OAB and OBC. The sides of the spherical triangle are a, b, c; these give the distances along the great circle arcs joining the vertices. The angles subtended by the sides at the centre are α, β and γ so that $a = \alpha R$ etc. The aim of this appendix is to prove the principal relations between the six elements of a spherical triangle.The fundamental relation is the spherical cosine rule. ALL OTHER RULES, and there are many, can be derived from the cosine rule.

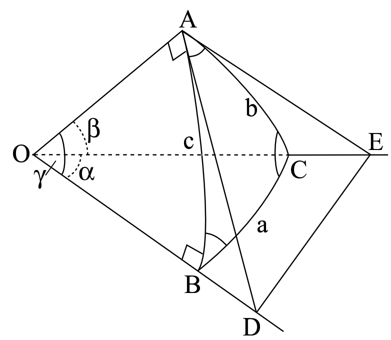


Figure D.2

D.2 Spherical cosine rule

Geometric proof

In Figure D.2 AD and AE are the tangents to the sides of the spherical triangle at A. As long as the angles β and γ are strictly less than $\pi/2$ the tangent to the side AB meets the radius OB extended to D and the tangent to the side AC meets the radius OC extended to E. Since any tangent to the sphere is normal to the radius at the point of contact we have that the triangles OAD and OAE are right angled.

We apply the planar cosine rule to the triangles ODE and ADE:

$$DE^2 = OD^2 + OE^2 - 2OD.OE\cos\alpha,$$
$$DE^2 = AD^2 + AE^2 - 2AD.AE\cos A.$$

Subtracting these equations and using Pythagoras' theorem to set $OD^2 - AD^2 = OA^2$ and $OE^2 - AE^2 = OA^2$ we obtain

$$0 = 2OA^2 + 2AD.AE\cos A - 2OD.OE\cos\alpha$$

Dividing each term by the product $OD.OE$ and using $OA/OD = \cos\gamma$ *etc.* gives

$$\cos\alpha = \cos\gamma\cos\beta + \sin\gamma\sin\beta\cos A.$$

It is conventional to express these identities in terms of the actual sides so that we should set $\alpha = a/R$ *etc.* If we assume that the lengths have been scaled to a *unit* sphere then the above, alongwith the two relations obtained by cyclic permutations, becomes

$$\boxed{\begin{aligned}\cos a &= \cos b\,\cos c + \sin b\,\sin c\,\cos A,\\ \cos b &= \cos c\,\cos a + \sin c\,\sin a\,\cos B,\\ \cos c &= \cos a\,\cos b + \sin a\,\sin b\,\cos C.\end{aligned}} \tag{D.1}$$

For a small triangle with a, b, $c \ll 1$ on the unit sphere, the spherical cosine rules reduce to the planar cosine rule if we neglect cubic terms. For example the first becomes

$$1 - \frac{a^2}{2} = \left(1 - \frac{b^2}{2}\right)\left(1 - \frac{c^2}{2}\right) + bc\cos A,$$

which simplifies to

$$a^2 = b^2 + c^2 - 2bc\cos A.$$

The above proof assumes that the angles β and γ are less than ninety degrees for the constructions as drawn. This restriction may be removed; it is discussed in detail in Todhunter's book (pages 16 to 19). The following alternative proof does not rely on these assumptions.

Cosine rule: vector proof

For any given spherical triangle we can introduce Cartesian axes with the z-axis along OA and the xz-plane defined by the plane OAB. Take the radius of the sphere as unity and define vectors $\mathbf{B}$ and $\mathbf{C}$ along the radii OB and OC respectively. The angle between the planes AOM and AON is given by $\angle MON = A$, so the components of these unit vectors are

$$\begin{aligned} \mathbf{B} &= (\sin c,\ 0\,,\ \cos c), \\ \mathbf{C} &= (\sin b\ \cos A, \sin b\ \sin A,\ \cos b) \end{aligned} \qquad \text{(D.2)}$$

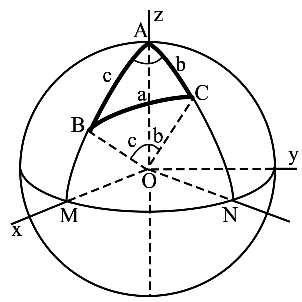


Figure D.3

Now the angle between the unit vectors is simply a, the angle subtended at the centre by the arc BC. Therefore

$$\mathbf{B}\cdot\mathbf{C} = \cos a = \sin b\ \sin c\ \cos A + 0 + \cos b\ \cos c, \qquad \text{(D.3)}$$

in agreement with our previous result for the cosine rule. This is the simplest proof of the cosine rule: it needs no restrictions on the angles.

D.3 Spherical sine rule

Derivation from the cosine rule

From equation (D.1) we have

$$\begin{aligned} \cos A &= \frac{\cos a - \cos b\ \cos c}{\sin b\ \sin c} \\ \sin^2 A &= 1 - \left(\frac{\cos a - \cos b\ \cos c}{\sin b\ \sin c}\right)^2 \\ &= \frac{(1 - \cos^2 b)(1 - \cos^2 c) - (\cos a - \cos b\ \cos c)^2}{\sin^2 b\ \sin^2 c}. \end{aligned}$$

Therefore

$$\frac{\sin A}{\sin a} = \Delta(a,b,c), \tag{D.4}$$

where

$$\Delta(a,b,c) = \frac{[1-\cos^2 a - \cos^2 b - \cos^2 c + 2\cos a \cos b \cos c]^{1/2}}{\sin a \sin b \sin c}. \tag{D.5}$$

Since Δ is invariant under a cyclic permutation of $a,\ b,\ c$ we deduce the spherical sine rule

$$\boxed{\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c} = \Delta(a,b,c)} \tag{D.6}$$

The three separate rules are

$$\begin{aligned} \sin b\ \sin A &= \sin a\ \sin B, \\ \sin c\ \sin A &= \sin a\ \sin C, \\ \sin b\ \sin C &= \sin c\ \sin B. \end{aligned} \tag{D.7}$$

Spherical sine rule: geometric proof

Consider the following construction. Take any point P on the line OA and drop a perpendicular to the point N in the plane OBC. Draw the perpendicular from N to the line OB at the point M. Therefore the three triangles PMN, PON and ONM are all right angled triangles and we can therefore use Pythagoras' theorem to deduce that

$$\begin{aligned} PM^2 &= MN^2 + PN^2, \\ OP^2 &= ON^2 + PN^2, \\ ON^2 &= OM^2 + MN^2. \end{aligned}$$

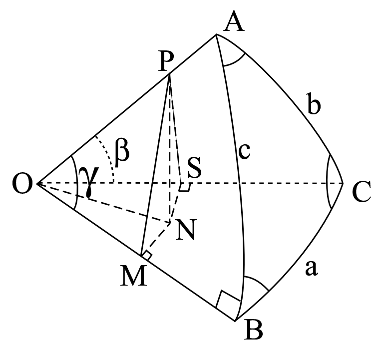


Figure D.4

Therefore we must have

$$PM^2 = (ON^2 - OM^2) + (OP^2 - ON^2) = OP^2 - OM^2,$$

so that the triangle OPM must have a right angle at M. From this we first deduce that $PM = OP \sin\gamma$. Secondly we note that since PM and NM are both normal to OB then the angle PMN is the angle between the planes OAB and OBC; this is the angle B so that we must have

$$PN = PM \sin B = OP \sin\gamma \sin B.$$

We now repeat the argument with the construction of NS perpendicular to OC and prove that triangle OPS is right angled and the angle PSN is equal to C. (M, N and S are not collinear). Therefore we find

$$PN = PS \sin C = OP \sin\beta \sin C.$$

Comparing the two expressions for PN we deduce that

$$\sin\gamma\sin B = \sin\beta\sin C.$$

This whole process can be repeated with P an arbitrary point on OB or OC and dropping perpendiculars onto the face OAC and OAB respectively. Clearly this will give

$$\sin\gamma\sin A = \sin\alpha\sin C,$$
$$\sin\beta\sin A = \sin\alpha\sin B.$$

On the unit sphere the angles α, η, γ will be replaced by a, b, c giving equations (D.6). Note that the construction and proof will need slight modifications if either of the angles B or C exceeds $\pi/2$. This is discussed in Todhunter.

D.4 Solution of spherical triangles I

In general, if we know three elements of a triangle then we might expect to find the other three elements by direct application of the spherical sine and cosine rules. This is NOT possible: to complete the solution in many cases we shall need further rules developed in the ensuing sections.

The six distinct ways in which three elements may be given are shown in Figure D.5 along with a seventh case involving four given elements. In each figure the given elements are shown below and the given angles are marked with a small arc and the given sides are marked with a cross bar; each figure has variations given by cyclic permutations. The solution of such spherical triangles is harder than in the planar case because we do not know the sum of the angles: given two angles we do not know the third.

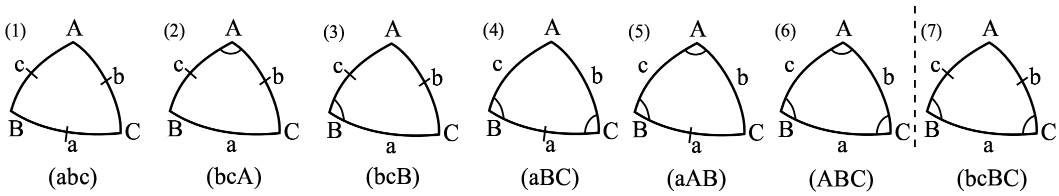


Figure D.5

- **Case 1**: this can be solved by using the cosine rule.
- **Case 2**: cosine rule gives a and then we are back to Case 1.
- **Case 3**: sine rule gives C and then we are in Case 7.
- **Case 4**: no progress possible with only sine and cosine rules.
- **Case 5**: sine rule gives b and then we are in Case 7.
- **Case 6**: no progress possible. This case doesn't arise in plane geometry.
- **Case 7**: no progress possible with only sine and cosine rules.

This is an appropriate point to mention that any determination of an angle or side from its sine will generally lead to ambiguities since $\sin x = \sin(\pi - x)$. However the angles and sides on the unit sphere are in the interval $(0, \pi)$ so their determination from cosines, secants, tangents or cotangents will be unambiguous. Likewise the sine, cosine or tangent of any half-angle (or side) is positive and its inverse is also unambiguous. Many of the formulae that we will derive were established to avoid the sine ambiguity.

D.5 Polar triangles and the supplemental cosine rules

Figure D.6a below shows the three great circles which intersect to form the spherical triangle ABC. In addition we show the normals to the plane of each great circle; each intersects the sphere in two points each of which is a 'pole' when a specific great circle is identified as an equator. As shown some of the poles (small solid circles) are visible and some(open circles) are on the hidden face. Three of these six poles may be used to define the polar triangle. The convention is that A and its pole A' lie on the same side of the diametral plane containing BC; likewise for the others. We shall now prove the following statements.

- The sides of the polar triangle $A'B'C'$ are the supplements of the angles of the original triangle ABC. (We assume a unit sphere on which the lengths of the sides are equal to the radian measure of the angles they subtend at the centre).
- The angles of the polar triangle $A'B'C'$ are the supplements of the sides of the original triangle ABC.

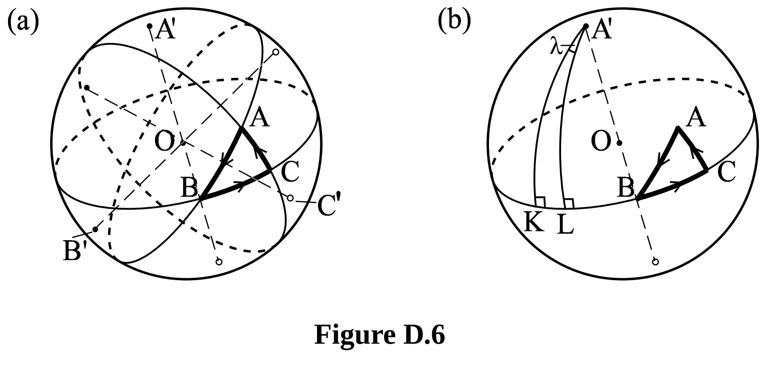


Figure D.6

Figure D.6b shows the triangle ABC, the pole A' of A and the corresponding 'equator' formed by extending the side BC. Note three properties:

1. Any great circle through the pole A' to its equator BC is a quadrant arc of length $\pi/2$ (on the unit sphere), *i.e.* $A'K = A'L = \pi/2$.
2. Any great circle through A' intersects its equator BC at a right angle, as at K and L.
3. The angle λ (in radians) between two such quadrant arcs is equal to the length of the segment cut on the equator by the arcs, *i.e.* $\lambda = \angle KA'L = KL$.

Figure D.7, which is neither an elevation nor a perspective view, shows the *schematic* relation between the triangle ABC and its polar triangle $A'B'C'$. The sides of ABC are extended along their great circles to meet the sides of $A'B'C'$ at the points shown. From the the three properties discussed in the previous paragraph we can deduce the following results.

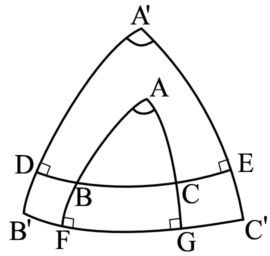


Figure D.7

- The great circles $A'B'$ and $A'C'$ through the pole A' intersect the equator corresponding to A', that is BC extended, at points D and E. The intersections are right angles and the distance DE is equal to the angle A' expressed in radians. Therefore $A' = DE$.

- $B'G$ is a great circle through the pole B' meeting its corresponding equator CA at G. The intersection at G is at right angles and the length $B'G = \pi/2$. Similarly $C'F$ is a great circle through the pole C' meeting its corresponding equator AB at F: the intersection at F is at right angles and the length $C'F = \pi/2$.

- Now consider the intersections of the great circle $B'C'$ with the great circles defined by AB and AC. Since the angles at F and G are right angles we deduce that A must be the pole to the equator $B'C'$. Similarly B, C must be the poles of the equators $C'A'$ and $A'B'$ repectively. We conclude that the polar triangle of the polar triangle $A'B'C'$ must be the original triangle ABC. Consequently (1) since C is the pole of $A'B'$ we must have $CD = \pi/2$; (2) since B is the pole of $C'A'$ we must have $BE = \pi/2$; (3) since A is the pole of $B'C'$ we must have $FG = A$.

We now have all the information we need to deduce

$$A' = DE = DC + BE - BC = \frac{\pi}{2} + \frac{\pi}{2} - a = \pi - a,$$
$$a' = B'C' = B'G + FC' - FG = \frac{\pi}{2} + \frac{\pi}{2} - A = \pi - A.$$

Similar results follow for the other angles and sides of the polar triangle so that:

$$\begin{aligned} A' &= \pi - a \qquad & B' &= \pi - b \qquad & C' &= \pi - c, \\ a' &= \pi - A \qquad & b' &= \pi - B \qquad & c' &= \pi - C. \end{aligned} \tag{D.8}$$

An important corollary follows from the existence of the polar triangle. We have already stated that Euclid proves that the sum of the sides of a spherical triangle on the unit sphere satisfies $\sigma = a+b+c < 2\pi$. Applying this to the polar triangle gives $3\pi - A - B - C < 2\pi$ so that Σ, the sum of the angles, is greater than π. Since we conventionally take the angles to be less than π then we must have $\pi < \Sigma < 3\pi$. (The restriction to angles and sides less than π may be lifted; the so-called improper triangles so formed are discussed in Todhunter. We have no need to consider them here.)

Supplemental cosine rules

As an example of using the polar triangle let us apply the cosine rules of (D.1) to $A'B'C'$:

$$\begin{aligned}\cos a' &= \cos b' \, \cos c' + \sin b' \, \sin c' \, \cos A',\\ \cos b' &= \cos c' \, \cos a' + \sin c' \, \sin a' \, \cos B',\\ \cos c' &= \cos a' \, \cos b' + \sin a' \, \sin b' \, \cos C'.\end{aligned} \tag{D.9}$$

Now substitute for angles and sides usung equation (D.8) noting that $\cos(\pi - \theta) = -\cos\theta$ and $\sin(\pi - \theta) = \sin\theta$:

$$\boxed{\begin{aligned}\cos A + \cos B \, \cos C &= \sin B \, \sin C \, \cos a,\\ \cos B + \cos C \, \cos A &= \sin C \, \sin A \, \cos b,\\ \cos C + \cos A \, \cos B &= \sin A \, \sin B \, \cos c.\end{aligned}} \tag{D.10}$$

Now these equations, obtained by applying a known rule to the polar triangle, are obviously *new* relations between the elements of the original triangle; they are called the supplemental cosine rules. This is an example of a powerful method of generating a new formula from any that we have already found.

The supplemental cosine rules clearly provide a way of solving a spherical triangle when all three angles are given. This is Case 6 in Figure D.5.

Note that a new rule does not always arise. For example, applying the sine rule to $A'B'C'$ gives

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

On substituting (D.8) we have the usual rules simply inverted:

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

Alternative derivation of the supplemental cosine rules

It is possible to derive the supplemental cosine rules directly without appealing to the polar triangle. For example, in the first formula of (D.10) substitute for the terms on the left-hand side using the normal cosine rules:

$$\begin{aligned}\cos A + \cos B \cos C &= \frac{(\cos a - \cos b \cos c)\sin^2 a + (\cos b - \cos c \cos a)(\cos c - \cos a \cos b)}{\sin^2 a \sin b \sin c}\\ &= \frac{\cos a[1 - \cos^2 a - \cos^2 b - \cos^2 c + 2\cos a \cos b \cos c]}{\sin^2 a \sin b \sin c}\\ &= \cos a \, \Delta^2 \sin b \, \sin c\\ &= \cos a \, \sin B \, \sin C,\end{aligned}$$

where we have used the definition of Δ in (D.5) and also the sine rule (D.6). Thus we could have proceeded in this way and then deduced the existence of the polar triangle as a corollary without the geometrical proof that we presented earlier.

D.6 The cotangent four-part formulae

The six elements of a triangle may be written in an anti-clockwise order as $(aCbAcB)$. The cotangent, or four-part, formulae relate two sides and two angles forming four *consecutive* elements around the triangle, for example $(aCbA)$ or $BaCb$. The six distinct formulae that we shall prove are

$$\boxed{\begin{array}{lll} (a) & \cos b\,\cos C = \cot a\,\sin b - \cot A\,\sin C, & (aCbA) \\ (b) & \cos b\,\cos A = \cot c\,\sin b - \cot C\,\sin A, & (CbAc) \\ (c) & \cos c\,\cos A = \cot b\,\sin c - \cot B\,\sin A, & (bAcB) \\ (d) & \cos c\,\cos B = \cot a\,\sin c - \cot A\,\sin B, & (AcBa) \\ (e) & \cos a\,\cos B = \cot c\,\sin a - \cot C\,\sin B, & (cBaC) \\ (f) & \cos a\,\cos C = \cot b\,\sin a - \cot B\,\sin C, & (BaCb), \end{array}} \tag{D.11}$$

where the subset of elements involved is shown to the right of every equation. In the first equation, for the set $aCbA$, we term a and A the outer elements and C and b the inner elements. With this notation the general form of the equations is

$$\boxed{\begin{aligned} \cos(\text{inner side}).\cos(\text{inner angle}) \quad = \quad & \cot(\text{outer side}).\sin(\text{inner side}) \\ & -\cot(\text{outer angle}).\sin(\text{inner angle}) \end{aligned}} \tag{D.12}$$

Note that the 'inner' elements of each set formula occur twice and cannot be deduced from the other elements; only the 'outer' elements of each set may be derived in terms of the other three. For example in the first equation involving the set $aCbA$ we can only determine the outer side a in terms of CbA or the outer angle A in terms of aCb. Note also that the outer angle or side is determined from its cotangent so that there is no ambiguity.

To prove the first formula start from the cosine rule (D.1a) and on the right-hand side substitute for $\cos c$ from (D.1c) and for $\sin c$ from (D.6):

$$\begin{aligned} \cos a &= \cos b \cos c + \sin b \sin c \cos A \\ &= \cos b\,(\cos a \cos b + \sin a \sin b \cos C) + \sin b \sin C \sin a \cot A \\ \cos a \sin^2 b &= \cos b \sin a \sin b \cos C + \sin b \sin C \sin a \cot A. \end{aligned}$$

The result follows on dividing by $\sin a \sin b$. Similar techniques with the other two cosine rules give D.11c,e. Equations D.11b,d,f follow by applying D.11e,a,c to the polar triangle.

Solution of spherical triangles II

The four-part formulae may be used to give solutions to two of the cases discussed in Section D.4. In Case 2 in Figure D.5, where we are given (bAc), we can use equations D.11b,c to find the angles C, B from their cotangents: we can then find a from D.11a without any sine ambiguity. We can now solve Case 4, where we are given (BaC), by using equations D.11e,f to give the sides c, b and we can then find A from D.11a. We are still left with the problem solving Case 7 (since Cases 3, 5 can also be reduced to Case 7).

D.7 Half-angle and half-side formulae

If $2s = (a + b + c)$ is the sum of the sides and $2S = (A + B + C)$ is the sum of the angles, then we can easily prove the following formulae:

$$\begin{array}{ll}
\sin\dfrac{A}{2} = \left[\dfrac{\sin(s{-}b)\sin(s{-}c)}{\sin b \sin c}\right]^{1/2} & \sin\dfrac{a}{2} = \left[\dfrac{-\cos S\cos(S{-}A)}{\sin B\sin C}\right]^{1/2} \\
\cos\dfrac{A}{2} = \left[\dfrac{\sin s\sin(s{-}a)}{\sin b\sin c}\right]^{1/2} & \cos\dfrac{a}{2} = \left[\dfrac{\cos(S{-}B)\cos(S{-}C)}{\sin B\sin C}\right]^{1/2} \\
\tan\dfrac{A}{2} = \left[\dfrac{\sin(s{-}b)\sin(s{-}c)}{\sin s\sin(s{-}a)}\right]^{1/2} & \tan\dfrac{a}{2} = \left[\dfrac{-\cos S\cos(S{-}A)}{\cos(S{-}B)\cos(S{-}C)}\right]^{1/2}
\end{array} \tag{D.13}$$

To prove the first formula use $\cos A = 1 - 2\sin^2(A/2)$ and the cosine rule (D.1).

$$\begin{aligned}
\sin^2\frac{A}{2} &= \frac{1-\cos A}{2} \\
&= \frac{1}{2} - \frac{\cos a - \cos b\cos c}{2\sin b\sin c} = \frac{\cos(b{-}c) - \cos a}{2\sin b\sin c} \\
&= \frac{1}{\sin b\sin c}\sin\left(\frac{a+b-c}{2}\right)\sin\left(\frac{a-b+c}{2}\right).
\end{aligned}$$

Since $2(s{-}b) = (a + b + c) - 2b = a - b + c$ *etc.* we obtain the first result. The second follows from $1{+}\cos A = 2\cos^2(A/2)$ and the third from their quotient. The results in the right hand column follow by applying the first column formulae to the polar triangle. They also follow from (D.10) and by starting with $\cos a = 1 - 2\sin^2(a/2)$ *etc.* .

It is worth commenting on the negative signs under some radicals. Take the expression for $\sin a/2$ as an example. Since $\pi < A + B + C < 3\pi$ we have $\pi/2 < S < 3\pi/2$ so that $\cos S < 0$. Now in any spherical triangle the side BC is the shortest distance between B and C so we must have $BC < BA + AC$, or $a < b + c$; *i.e.* any side is less than the sum of the others. Applying this to the polar triangle we have $\pi{-}A < (\pi{-}B) + (\pi{-}C)$; therefore $2(S{-}A) = B{+}C{-}A < \pi$ or $(S{-}A) < \pi/2$. Furthermore, since $A < \pi$ we have $B{+}C{-}A > -\pi$ and consequently $2(S{-}A) > -\pi$. Therefore $-\pi/2 < (S{-}A) < \pi/2$ and $\cos(S{-}A) > 0$. These results guarantee that the expressions under the radical are positive.

Solution of spherical triangles III

The above formulae are clearly applicable to the cases where we know either three sides or three angles, cases which we have solved by either the normal or supplemental cosine rules. The expressions given here involving tangents of half angles are to be preferred whenever the angle or side to be found is very small or nearly π.

Delambre (or Gauss) analogies.

$$\frac{\sin\frac{1}{2}(A+B)}{\cos\frac{1}{2}C} = \frac{\cos\frac{1}{2}(a-b)}{\cos\frac{1}{2}c} \qquad \frac{\sin\frac{1}{2}(A-B)}{\cos\frac{1}{2}C} = \frac{\sin\frac{1}{2}(a-b)}{\sin\frac{1}{2}c}$$

$$\frac{\cos\frac{1}{2}(A+B)}{\sin\frac{1}{2}C} = \frac{\cos\frac{1}{2}(a+b)}{\cos\frac{1}{2}c} \qquad \frac{\cos\frac{1}{2}(A-B)}{\sin\frac{1}{2}C} = \frac{\sin\frac{1}{2}(a+b)}{\sin\frac{1}{2}c} \tag{D.14}$$

These are proved by expanding the numerator on the left hand side and using the half angle formulae. For example, using equations C.5, C.13 and C.23

$$\begin{aligned}
\sin\frac{1}{2}(A+B) &= \sin\frac{A}{2}\cos\frac{B}{2} + \cos\frac{A}{2}\sin\frac{B}{2} \\
&= \left[\frac{\sin s\,\sin^2(s-b)\,\sin(s-c)}{\sin a\ \sin b\,\sin^2 c}\right]^{1/2} + \left[\frac{\sin s\,\sin^2(s-a)\,\sin(s-c)}{\sin a\ \sin b\,\sin^2 c}\right]^{1/2} \\
&= \frac{\sin(s-b)+\sin(s-a)}{\sin c}\left[\frac{\sin s\,\sin(s-c)}{\sin a\,\sin b}\right]^{1/2} \\
&= \frac{\sin\frac{1}{2}c\ \cos\frac{1}{2}(a-b)}{\sin\frac{1}{2}c\ \cos\frac{1}{2}c}\cos\frac{1}{2}C,
\end{aligned}$$

and hence the required result.

Napier's analogies

Published by Napier in 1614. His methods were purely geometric but we obtain them by dividing the Delambre formulae.

$$\tan\tfrac{1}{2}(A+B) = \frac{\cos\frac{1}{2}(a-b)}{\cos\frac{1}{2}(a+b)}\cot\tfrac{1}{2}C \qquad \tan\tfrac{1}{2}(a+b) = \frac{\cos\frac{1}{2}(A-B)}{\cos\frac{1}{2}(A+B)}\tan\tfrac{1}{2}c$$

$$\tan\tfrac{1}{2}(A-B) = \frac{\sin\frac{1}{2}(a-b)}{\sin\frac{1}{2}(a+b)}\cot\tfrac{1}{2}C \qquad \tan\tfrac{1}{2}(a-b) = \frac{\sin\frac{1}{2}(A-B)}{\sin\frac{1}{2}(A+B)}\tan\tfrac{1}{2}c \tag{D.15}$$

Solution of spherical triangles IV

We now have all we need to solve all the possible configurations shown in Figure D.5. Napier's analogies clearly provide the means of solving Case 7, and hence Cases 3, 5. They also provide a means of progressing without the trouble of ambiguities arising from the use of the sine rule. For example in Case 4 given a, B, C, we can use the Napier analogies to find $b \pm c$ and then again to find A.

D.8 Right-angled triangles

There are many problems in which one of the angles, say C, is equal to $\pi/2$. In this case there are only 5 elements and in general two will suffice to solve the triangle. We shall show that the solution of such a triangle can be presented as a set of 10 equations involving 3 elements so that every element can be expressed in terms of any pair of the other elements.

The required 10 equations involving C are found from the third cosine rule (D.1), two sine rules (D.7), four cotangent formulae (D.11) and all three of the supplemental cosine rules (D.10). Setting $C = \pi/2$ we obtain (from the equations indicated)

$$
\begin{array}{llll}
\text{(D.1c)} & \cos c = \cos a\, \cos b, & \text{(D.11b)} & \tan b = \cos A\, \tan c, \\
\text{(D.7b)} & \sin a = \sin A\, \sin c, & \text{(D.11e)} & \tan a = \cos B\, \tan c, \\
\text{(D.7c)} & \sin b = \sin B\, \sin c, & \text{(D.10a)} & \cos A = \sin B\, \cos a, \\
\text{(D.11a)} & \tan a = \tan A\, \sin b, & \text{(D.10b)} & \cos B = \sin A\, \cos b, \\
\text{(D.11f)} & \tan b = \tan B\, \sin a, & \text{(D.10c)} & \cos c = \cot A\, \cot B.
\end{array} \quad \text{(D.16)}
$$

As an example suppose we are given a and c (and $C = \pi/2$). Then we can find c, A, B from the first, fourth and fifth equations.

Napier's rules for right-angled triangles

Napier showed that the ten equations which give all possible relations in a right-angled triangle can be summarised by two simple rules along with a simple picture. We define the 'circular parts' of the triangle to be a, b, $\frac{1}{2}\pi - A$, $\frac{1}{2}\pi - c$, and $\frac{1}{2}\pi - B$. These are arranged around the circle in the natural order of the triangle, C omitted between a and b. Choose any of the five sectors and call it the middle part. The sectors next to it are called the 'adjacent' parts and the remaining two parts are the 'opposite' parts. Napiers rules are:

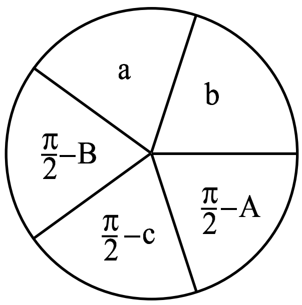


Figure D.8

$$
\boxed{\begin{array}{lcl}
\text{sine of middle part} & = & \text{product of tangents of adjacent parts,} \\
\text{sine of middle part} & = & \text{product of cosines of opposite parts.}
\end{array}} \quad \text{(D.17)}
$$

For example if we take $\frac{1}{2}\pi - c$ as the middle part the first rule gives $\sin(\pi/2 - c) = \tan(\pi/2 - A)\tan(\pi/2 - B)$ which gives the last of the equations in (D.16); if we apply the second rule we get $\sin(\pi/2 - c) = \cos a \cos b$ which is the first of the equations in (D.16).

D.9 Quadrantal triangles

The triangle ABC is quadrantal if at least one side subtends an angle of $\pi/2$ at the centre of the sphere. Without loss of generality take $c = \pi/2$. Therefore the angle $C' = \pi - c$ of the polar triangle is equal to $\pi/2$. Now apply Napier's rules to the polar triangle with $C' = \pi/2$ and

$$a' = \pi - A, \quad b' = \pi - B, \quad A' = \pi - a, \quad B' = \pi - b.$$

The circular parts of the polar triangle

$$a', \qquad b', \qquad \frac{\pi}{2} - A', \qquad \frac{\pi}{2} - c', \qquad \frac{\pi}{2} - B',$$

must be replaced by

$$\pi - A, \qquad \pi - B, \qquad a - \frac{\pi}{2}, \qquad C - \frac{\pi}{2}, \qquad b - \frac{\pi}{2},$$

Noting that $\sin(x - \pi/2) = -\cos x$, $\cos(x - \pi/2) = \sin x$ and $\tan(x - \pi/2) = -\cot x$ we have the following equations:

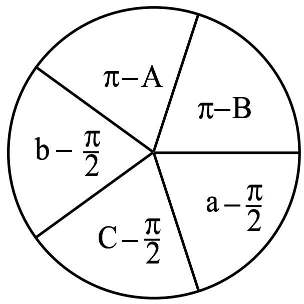


Figure D.9

$$\begin{aligned} \cos C &= -\cos A \, \cos B, & \tan B &= -\cos a \, \tan C, \\ \sin A &= \sin a \, \sin C, & \tan A &= -\cos b \, \tan C, \\ \sin B &= \sin b \, \sin C, & \cos a &= \sin b \, \cos A, \\ \tan A &= \tan a \, \sin B, & \cos b &= \sin a \, \cos B, \\ \tan B &= \tan b \, \sin A, & \cos C &= -\cot a \, \cot b. \end{aligned} \tag{D.18}$$

Example

As an example of a quadrantal triangle we consider a problem arising in the discussion of geodesics on a sphere in Chapter 11. With the following identifications

$$a = s, \qquad b = \frac{\pi}{2} - \phi, \qquad c = \frac{\pi}{2},$$
$$A = \lambda, \qquad B = \alpha_0, \qquad C = \pi - \alpha. \tag{D.19}$$

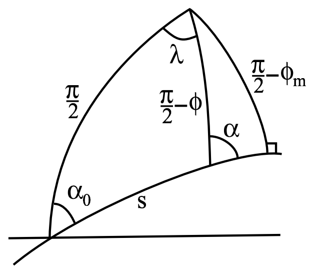


Figure D.10

the equations (D.18) become

$$\begin{aligned} \cos\alpha &= \cos\lambda \, \cos\alpha_0, & \tan\alpha_0 &= \cos s \, \tan\alpha, \\ \sin\lambda &= \sin s \, \sin\alpha, & \tan\lambda &= \sin\phi \, \tan\alpha, \\ \sin\alpha_0 &= \cos\phi \, \sin\alpha, & \cos s &= \cos\phi \, \cos\lambda, \\ \tan\lambda &= \tan s \, \sin\alpha_0, & \sin\phi &= \sin s \, \cos\alpha_0, \\ \tan\alpha_0 &= \cot\phi \, \sin\lambda, & \cos\alpha &= \cot s \, \tan\phi. \end{aligned} \tag{D.20}$$

The practical problems are (a) given α_0 and s find λ, ϕ and α; (b) given λ, ϕ find α_0 and s. For the first we use the fourth, ninth, and then the first equation. For the second we use the fifth and eighth equations.

Solution of spherical triangles V

The rules for right angled triangles provide another method for the solution of spherical triangles in general. Consider the triangle ABC shown in the figure; $b,\ c,\ B$ are assumed given. Draw the great circle through A which meets BC at right angles at the point D. We first solve the triangle ABD using c and B to find AD, BD and $\angle BAD$. Then in triangle ACD we use AD and b to find CD and the angles $\angle CAD$ and C. The difficulty with this method, apart from the increased number of steps, is to find the most appropriate construction.

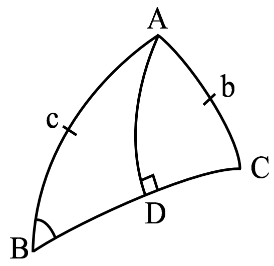


Figure D.11

Appendix E

Power series expansions

E.1 General form of the Taylor and Maclaurin series

Taylor's theorem may be written in the form:

$$f(z) = f(b) + \frac{(z-b)}{1!}f'(b) + \frac{(z-b)^2}{2!}f''(b) + \frac{(z-b)^3}{3!}f'''(b) + \cdots . \tag{E.1}$$

or, alternatively,

$$f(b+z) = f(b) + \frac{z}{1!}f'(b) + \frac{z^2}{2!}f''(b) + \frac{z^3}{3!}f'''(b) + \cdots . \tag{E.2}$$

When $b = 0$ we obtain Maclaurin's series.

$$f(z) = f(0) + \frac{z}{1!}f'(0) + \frac{z^2}{2!}f''(0) + \frac{z^3}{3!}f'''(0) + \cdots . \tag{E.3}$$

E.2 Miscellaneous Taylor series

$$\sin(b+z) = \sin b + z\cos b - \frac{z^2}{2!}\sin b - \frac{z^3}{3!}\cos b + \frac{z^4}{4!}\sin b + \cdots \tag{E.4}$$

$$\cos(b+z) = \cos b - z\sin b - \frac{z^2}{2!}\cos b + \frac{z^3}{3!}\sin b + \frac{z^4}{4!}\cos b + \cdots \tag{E.5}$$

$$\tan(b+z) = \tan b + z\sec^2 b + z^2\tan b\sec^2 b + \frac{z^3}{3}(1+3\tan^2 b)\sec^2 b + \cdots \tag{E.6}$$

$$\tan\left(\frac{\pi}{4}+z\right) = 1 + 2z + 2z^2 + \frac{8}{3}z^3 + \cdots \tag{E.7}$$

$$\arcsin(b+z) = \arcsin b + z\frac{1}{(1-b^2)^{1/2}} + \frac{z^2}{2}\frac{b}{(1-b^2)^{3/2}} + \cdots \tag{E.8}$$

$$\arctan(b+z) = \arctan b + z\frac{1}{1+b^2} - \frac{z^2}{2!}\frac{2b}{(1+b^2)^2} + \frac{z^3}{3!}\left[\frac{-2}{(1+b^2)^2} + \frac{8b^2}{(1+b^2)^3}\right]$$
$$- \frac{z^4}{4!}\left[\frac{-24b}{(1+b^2)^3} + \frac{48b^3}{(1+b^2)^4}\right] + \cdots \tag{E.9}$$

/over

E.3 Miscellaneous Maclaurin series

- Logarithms

$$\ln(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \frac{1}{4}z^4 + \cdots \qquad -1 < z \le 1 \quad \text{(E.10)}$$

$$\ln(1-z) = -z - \frac{1}{2}z^2 - \frac{1}{3}z^3 - \frac{1}{4}z^4 + \cdots \qquad -1 \le z < 1 \quad \text{(E.11)}$$

$$\ln\left(\frac{1+z}{1-z}\right) = 2z + \frac{2}{3}z^3 + \frac{2}{5}z^5 + \frac{2}{7}z^7 + \cdots \qquad |z| < 1 \quad \text{(E.12)}$$

- Trigonometric

$$\sin z = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \frac{1}{7!}z^7 + \cdots \qquad |z| < \infty \quad \text{(E.13)}$$

$$\cos z = 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \frac{1}{6!}z^6 + \cdots \qquad |z| < \infty \quad \text{(E.14)}$$

$$\tan z = z + \frac{1}{3}z^3 + \frac{2}{15}z^5 + \frac{17}{315}z^7 + \cdots \qquad |z| < \frac{\pi}{2} \quad \text{(E.15)}$$

$$\sec z = 1 + \frac{1}{2}z^2 + \frac{5}{24}z^4 + \frac{61}{720}z^6 + \cdots \qquad |z| < \frac{\pi}{2} \quad \text{(E.16)}$$

$$\csc z = \frac{1}{z} + \frac{1}{6}z + \frac{7}{360}z^3 + \frac{31}{15120}z^5 + \cdots \qquad 0 < |z| < \pi \quad \text{(E.17)}$$

$$\cot z = \frac{1}{z} - \frac{1}{3}z - \frac{1}{45}z^3 - \frac{2}{945}z^5 - \cdots \qquad 0 < |z| < \pi \quad \text{(E.18)}$$

- Inverse trig

$$\arcsin z = z + \frac{1}{6}z^3 + \frac{3}{40}z^5 + \frac{5}{112}z^7 + \cdots \qquad |z| < 1 \quad \text{(E.19)}$$

$$\arctan z = z - \frac{1}{3}z^3 + \frac{1}{5}z^5 - \frac{1}{7}z^7 + \frac{1}{9}z^9 - \cdots \qquad |z| < 1 \quad \text{(E.20)}$$

- Hyperbolic

$$\sinh z = z + \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \frac{1}{7!}z^7 + \cdots \qquad |z| < \infty \quad \text{(E.21)}$$

$$\cosh z = 1 + \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \frac{1}{6!}z^6 + \cdots \qquad |z| < \infty \quad \text{(E.22)}$$

$$\tanh z = z - \frac{1}{3}z^3 + \frac{2}{15}z^5 - \frac{17}{315}z^7 + \cdots \qquad |z| < \frac{\pi}{2} \quad \text{(E.23)}$$

$$\operatorname{sech} z = 1 - \frac{1}{2}z^2 + \frac{5}{24}z^4 - \frac{61}{720}z^6 \cdots \qquad |z| < \frac{\pi}{2} \quad \text{(E.24)}$$

/over

E.4 Miscellaneous Binomial series

Setting $f(z) = (1+z)^n$, n an integer, in the Maclaurin series gives the standard binomial series:

$$(1+z)^n = 1 + nz + \frac{n(n-1)}{2!}z^2 + \cdots + \frac{n!}{(n-r)!\,r!}z^r + \cdots\,, \tag{E.25}$$

$$(1+z)^{-1} = 1 - z + z^2 - z^3 + z^4 - \cdots\,, \tag{E.26}$$

$$(1+z)^{-2} = 1 - 2z + 3z^2 - 4z^3 + 5z^4 \cdots\,. \tag{E.27}$$

When n is a half-integer we obtain

$$(1+z)^{1/2} = 1 + \frac{1}{2}z - \frac{1}{8}z^2 + \frac{1}{16}z^3 - \frac{5}{128}z^4 + \cdots\,, \tag{E.28}$$

$$(1+z)^{-1/2} = 1 - \frac{1}{2}z + \frac{3}{8}z^2 - \frac{5}{16}z^3 + \frac{35}{128}z^4 - \cdots\,, \tag{E.29}$$

$$(1+z)^{-3/2} = 1 - \frac{3}{2}z + \frac{15}{8}z^2 - \frac{35}{16}z^3 + \frac{315}{128}z^4 - \cdots\,, \tag{E.30}$$

We will also need the the inverse of $(1 + a_2z^2 + a_4z^4 + a_6z^6)$. Therefore replacing z by $(a_2z^2 + a_4z^4 + a_6z^6)$ in (E.26) gives

$$\begin{aligned}(1{+}a_2z^2{+}a_4z^4{+}a_6z^6)^{-1} &= 1 - (a_2z^2 + a_4z^4 + a_6z^6)\\ &\quad + (a_2^2z^4 + 2a_2a_4z^6 + \cdots)^2 - (a_2^3z^6 + \cdots)^3 + O(z^8)\\ &= 1 - (a_2)z^2 - (a_4{-}a_2^2)z^4 - (a_6{-}2a_2a_4{+}a_2^3)z^6{+}O(z^8).\end{aligned} \tag{E.31}$$

Furthermore

$$\left(1{+}\frac{a_2z^2}{2}{+}\frac{a_4z^4}{24}{+}\frac{a_6z^6}{720}\right)^{-1} = 1 - \frac{z^2}{2}(a_2) - \frac{z^4}{24}\big(a_4{-}6a_2^2\big) - \frac{z^6}{720}\big(a_6{-}30a_2a_4{+}90a_2^3\big). \tag{E.32}$$

Appendix F

Calculus of variations

The simplest problem in the calculus of variations is as follows. Let $F(x, y, y')$ be a function of x and some *unspecified* function $y(x)$ and also its derivative. For every $y(x)$ we construct the following integral between *fixed* points A and B at which $x = a$ and $x = b$:

$$J[y] = \int_a^b F(x, y, y')dx. \tag{F.1}$$

The problem is to find the particular function $y(x)$ which, for a given function $F(x, y, y')$, minimises or maximises $J[y]$. In general we will not be able to say that we have a maximum or a minimum solution but the context of any particular problem will usually decide the matter. The following method only guarantees that $J[y]$ will be extremal. The solution here is valid for twice continuously differentiable functions.

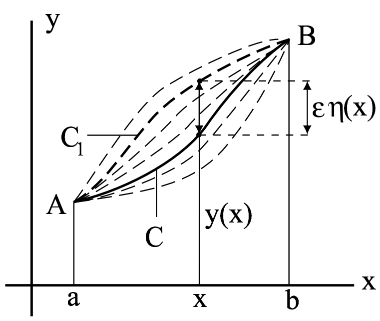


Figure F.1

We first tighten our notation a little. We assume that an extremal function can be found and that it is denoted by $y(x)$, the heavy path C in the figure; $J[y]$ then refers to the value of the integral on the extremal path. We consider the set of all paths AB defined by functions $\overline{y}$ where

$$\overline{y}(x) = y(x) + \epsilon\eta(x), \tag{F.2}$$

where $\eta(x)$ is an arbitrary function such that $\eta(a) = \eta(b) = 0$, thus guaranteeing that the end points of all paths are the same. One of these paths is denoted C_1 in the figure. The set of integrals for one given $\eta(x)$ and varying ϵ may be considered as generating a function $\Phi(\epsilon)$ such that

$$\Phi(\epsilon) = J[\overline{y}] = J[y + \epsilon\eta] = \int_a^b F(x,\, y + \epsilon\eta,\, y' + \epsilon\eta')dx. \tag{F.3}$$

In this notation the value of the integral on the extremal path is $\Phi(0)$ and the condition that it is an extremum is

$$\left.\frac{d\Phi}{d\epsilon}\right|_{\epsilon=0} = 0. \tag{F.4}$$

The Taylor series for the integrand is

$$F(x,\, y + \epsilon\eta,\, y' + \epsilon\eta') = F + F_y\epsilon\eta + F_{y'}\epsilon\eta' + O(\epsilon^2) \tag{F.5}$$

where F_y and $F_{y'}$ denote partial derivatives of F with respect to y and y' respectively. Substituting this series into the integral and differentiating with respect to ϵ gives

$$\left.\frac{d\Phi}{d\epsilon}\right|_{\epsilon=0} = \int_a^b \left[F_y\eta + F_{y'}\eta'\right] dx = 0. \tag{F.6}$$

The second term may be integrated by parts to give

$$\int_a^b F_{y'}\eta' dx = \left.\left[F_{y'}\eta\right]\right|_a^b - \int_a^b \eta\frac{d}{dx}\left[F_{y'}\right]\, dx. \tag{F.7}$$

Since the first term vanishes we have proved that for an extremal

$$\int_a^b \eta(x)\, H(x)\, dx = 0, \tag{F.8}$$

$$\text{where} \qquad H(x) = \frac{d}{dx}\left[F_{y'}\right] - F_y. \tag{F.9}$$

We now show that equation (F.8) implies that $H(x) = 0$. This result rejoices under the grand name of ‘the fundamental lemma’ of the calculus of variations. The proof is by contradiction: first suppose that $H(x) \neq 0$, say positive, at some point x_0 in (a, b). Then there must be an interval (x_1, x_2) surrounding x_0 in which $H(x) > 0$. Since $\eta(x)$ can be any suitably differentiable function we take $\eta = (x_2 - x)^4(x - x_1)^4$ in $[x_1, x_2]$ and zero elsewhere. Clearly, for such a function we must have $\int_a^b \eta H\, dx > 0$, in contradiction to (F.8). Therefore our hypothesis that $H \neq 0$ is not valid. Therefore we must have $H = 0$, giving the Euler–Lagrange equations:

EULER–LAGRANGE

$$\boxed{\frac{d}{dx}\left[\frac{\partial F}{\partial y'}\right] - \frac{\partial F}{\partial y} = 0.} \tag{F.10}$$

In this equation the partial derivatives indicate merely the formal operations of differentiating $F(x, y, y')$ with respect to y and y' as if they were independent variables. On the other hand the operator d/dx is a regular derivative and the above equation expands to

$$\frac{\partial^2 F}{\partial x \partial y'} + \frac{\partial^2 F}{\partial y \partial y'}y' + \frac{\partial^2 F}{\partial y'^2}y'' - \frac{\partial F}{\partial y} = 0. \tag{F.11}$$

This is a second order ordinary differential equation for $y(x)$: it has a solution with two arbitrary constants which must be fitted at the end points.

An alternative form of the Euler–Lagrange equations

Using the Euler equation (F.10) we have

$$\frac{d}{dx}\left[y' F_{y'} - F(x, y, y')\right] = y'' F_{y'} + y' \frac{d}{dx}\left[F_{y'}\right] - F_x - F_y y' - F_{y'} y'' = -F_x, \quad \text{(F.12)}$$

giving an alternative equation

$$\frac{d}{dx}\left[y' F_{y'} - F\right] + F_x = 0 \quad \text{(F.13)}$$

Functions of the form $F(y,\, y')$

Equation (F.13) shows that if F is independent of x then the equations integrate immediately since $F_x = 0$:

$$y' F_{y'} - F = \text{constant}. \quad \text{(F.14)}$$

Functions of the form $F(x,\, y')$

If F is independent of y then equation (F.10) can be integrated directly since $F_y = 0$:

$$F_{y'} = \text{constant}. \quad \text{(F.15)}$$

Extensions

There are many variants of the above results:

1. F has two (or more) dependent functions: $F\big(x,\, u(x),\, u'(x),\, v(x),\, v'(x)\big)$
2. F has two (or more) independent variables: $F\big(x,\, y,\, u(x, y),\, u'(x, y)\big)$
3. Both of above: $F\big(x,\, y,\, u(x, y),\, u'(x, y),\, v(x, y),\, v'(x, y)\big)$
4. F involves higher derivatives: $F(x,\, y,\, y',\, y'', \ldots)$.
5. The end points are not held fixed.

Only the first of the above concerns us here. The proof is along the same lines as above but we need to make two independent variations and set

$$\bar{u}(x) = u(x) + \epsilon \eta^{(u)}(x),$$
$$\bar{v}(x) = v(x) + \epsilon \eta^{(v)}(x).$$

Equation (F.8) now becomes of the form

$$\int_a^b \left[\eta^{(u)}(x)\, H^{(u)}(x) + \eta^{(v)}(x)\, H^{(v)}(x)\right] dx = 0. \quad \text{(F.16)}$$

Since $\eta^{(u)}$ and $\eta^{(v)}$ are arbitrary independent functions we obtain $H^{(u)} = H^{(v)} = 0$, *i.e.*

$$\frac{d}{dx}\left[\frac{\partial F}{\partial u'}\right] - \frac{\partial F}{\partial u} = 0, \tag{F.17}$$

$$\frac{d}{dx}\left[\frac{\partial F}{\partial v'}\right] - \frac{\partial F}{\partial v} = 0. \tag{F.18}$$

Sufficiency

The Euler–Lagrange equations have been shown to be a necessary conditions for the existence of an extremal integral. The proof of sufficiency is non-trivial and is discussed in advanced texts.

Appendix G

Complex variable theory

G.1 Complex numbers and functions

Complex numbers

A complex number z is a pair of real numbers, x, y combined with the basic 'imaginary' number 'i' in the expression $z = x + iy$. Such complex numbers may be manipulated just as real numbers with the proviso that $i^2 = -1$. We say that x is the real part of the complex number, $x = \text{Re}\,(z)$, and y is the imaginary part, $y = \text{Im}\,(z)$. From $z = x + iy$ we form its complex conjugate $z^* = x - iy$. Note that $zz^* = (x + iy)(x - iy) = x^2 + y^2$. The complex number $z = x + iy$ may be be represented as a point (x, y) in a plane which is called the complex z-plane. It is also useful to introduce polar coordinates in the plane and write

$$z = x + iy = r(\cos\theta + i\sin\theta). \tag{G.1}$$

In this context we say that r is the 'modulus' of z and θ is the 'argument' of z and write

$$r = |z| = \left[x^2 + y^2\right]^{1/2}, \qquad \theta = \arg(z) = \arctan\left(\frac{y}{x}\right). \tag{G.2}$$

Note that we can also write $r = |z| = \sqrt{zz^*}$.

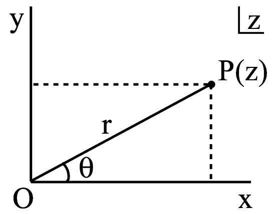


Figure G.1

Complex functions: examples

- The simplest complex function we can consider is a finite polynomial such as:

$$w(z) = 3 + z + z^2. \tag{G.3}$$

  If we substitute $z = x + iy$ in this expression, using $i^2 = -1$, we obtain

$$w(z) = u(x, y) + iv(x, y) \qquad \text{where} \quad \begin{cases} u(x, y) = 3 + x + x^2 - y^2, \\ v(x, y) = y + 2xy. \end{cases} \tag{G.4}$$

  Here we have written $w(z)$ in terms of two real functions of two variables. All complex functions can be split up in this way.

- Complex functions may be defined by convergent infinite series of the form

$$w(z) = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 + \cdots , \tag{G.5}$$

where the coefficients will, in general, be complex numbers. The real and imaginary parts of $w(z)$ will be infinite series.

- The complex exponential function is defined by the series

$$\exp z = 1 + \frac{1}{1!}z + \frac{1}{2!}z^2 + \frac{1}{3!}z^3 + \frac{1}{4!}z^4 + \frac{1}{5!}z^5 + \cdots . \tag{G.6}$$

It can be proved that this series is convergent for all values of z. Note that when z is purely real, $z = x$, the series reduces to the usual real definition of $\exp(x)$. When z is purely imaginary, $z = i\theta$ say, we get a very interesting result:

$$\exp i\theta = 1 + \frac{i}{1!}\theta - \frac{1}{2!}\theta^2 - \frac{i}{3!}\theta^3 + \frac{1}{4!}\theta^4 - \frac{i}{5!}\theta^5 + \cdots . \tag{G.7}$$

Now the real terms in this expansion are simply those in the expansion of $\cos\theta$, whilst the imaginary terms are those that arise in the expansion of $\sin\theta$. Therefore we can write the polar coordinate expression of z in (G.1) as

$$z = r(\cos\theta + i\sin\theta) = r\exp(i\theta) = r\mathrm{e}^{i\theta}. \tag{G.8}$$

If we raise this result to the n-th power we obtain De Moivre's theorem:

$$z^n = r^n \exp(in\theta) = r^n(\cos n\theta + i\sin n\theta). \tag{G.9}$$

- We can also define the sine and cosine functions of a complex number by the series

$$\cos z = 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \cdots , \tag{G.10}$$

$$\sin z = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \cdots . \tag{G.11}$$

Once again, when z is purely real, $z = x$, these series reduce to the usual real definition of $\cos(x)$ and $\sin(x)$. When z is purely imaginary, $z = i\theta$ say, we get the following important results using (E.21,E.22):

$$\cos i\theta = \cosh\theta, \tag{G.12}$$

$$\sin i\theta = i\sinh\theta, \tag{G.13}$$

$$\tan i\theta = i\tanh\theta. \tag{G.14}$$

For a general complex number we can use compound angle formulae (C.3,C.4) to determine the real and imaginary parts of the sine and cosine functions:

$$\cos(x+iy) = \cos x\cos iy - \sin x\sin iy = \cos x\cosh y - i\sin x\sinh y, \tag{G.15}$$

$$\sin(x+iy) = \sin x\cos iy + \cos x\sin iy = \sin x\cosh y + i\cos x\sinh y. \tag{G.16}$$

- In Chapter 4 we require the real and imaginary parts of $\cot z$; these follow from the quotient of the last two equations. Simply multiply top and bottom by the complex conjugate of the denominator and use the identities for $\cos 2x$, $\cosh 2x$ *etc.* given in Appendix C. It will be convenient to use different notation for this example.

$$\begin{aligned}\cot(A+iB) &= \frac{\cos A\cosh B - i\sin A\sinh B}{\sin A\cosh B + i\cos A\sinh B}\\ &= \frac{\sin A\cos A - i\sinh B\cosh B}{\sin^2 A\cosh^2 B + \cos^2 A\sinh^2 B}\\ &= \frac{\sin 2A - i\sinh 2B}{\cosh 2B - \cos 2A} \qquad \text{(G.17)}\end{aligned}$$

- Similarly, we define the hyperbolic sine and cosine functions of a complex numbers. When z is purely real, $z = x$, these series reduce to the usual real definition of $\cosh x$ and $\sinh x$. When z is purely imaginary, $z = i\theta$ say, we get the counterparts of (G.12–G.14)

$$\cosh z = 1 + \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \cdots\,, \qquad \text{(G.18)}$$

$$\sinh z = z + \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \cdot, \qquad \text{(G.19)}$$

$$\cosh i\theta = \cos\theta, \qquad \text{(G.20)}$$

$$\sinh i\theta = i\sin\theta, \qquad \text{(G.21)}$$

$$\tanh i\theta = i\tan\theta. \qquad \text{(G.22)}$$

G.2 Differentiation of complex functions

Before presenting the definition of differentiation of a complex function we examine two aspects of real differentiation.

Real differentiation in one dimension

The usual definition of the derivative of a real function $f(x)$ is

$$f'(x) = \lim_{\delta x\to 0}\frac{f(x+\delta x) - f(x)}{\delta x}. \qquad \text{(G.23)}$$

The 'small print' of the definition is that the limit when x tends to zero from above ($\delta x \to 0+$) should be equal to the limit when x tends to zero from below ($\delta x \to 0-$). In principle these limits could be different and we would then have to define two different derivatives, say $f'_+(x)$ and $f'_-(x)$. A simple example where the limits differ is the function $f(x) = |x|$, for which $f'_+ = +1$ and $f'_- = -1$ at the origin. The only point we wish to make is that even in one dimension we must be careful about directions when defining derivatives.

Real differentiation in two dimensions

In two dimensions we can define two partial derivatives: that with respect to x being the derivative of $f(x, y)$ when y is held constant, and that with respect to y being the derivative of $f(x, y)$ when x is held constant. The notation and definitions of the partial derivatives is

$$\left(\frac{\partial f}{\partial x}\right)_y = \lim_{\delta x \to 0} \frac{f(x+\delta x, y) - f(x, y)}{\delta x}, \qquad \left(\frac{\partial f}{\partial y}\right)_x = \lim_{\delta y \to 0} \frac{f(x, y+\delta y) - f(x, y)}{\delta y}. \tag{G.24}$$

The brackets and subscripts are usually dropped if there is no ambiguity introduced thereby. There is no reason why the two derivatives should be equal, or even related in any particlar way.

The above derivatives are along the directions of the coordinate axes but it is perfectly reasonable to seek a derivative of $f(x, y)$ along any specified direction. To do this we use Taylor's theorem (in two dimensions), keeping only the first order terms, so that for a general displacement with components δx and δy,

$$\delta f = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y. \tag{G.25}$$

If the direction of the displacement is taken in the direction of the unit vector $\mathbf{n}$ and the magnitude of the displacement is δs, then we can set $\delta x = n_x \delta s$ and $\delta y = n_y \delta s$. We can then define a directional derivative in two dimensions as

$$\left(\frac{df}{ds}\right)_{\mathbf{n}} = n_x \frac{\partial f}{\partial x} + n_y \frac{\partial f}{\partial y}. \tag{G.26}$$

The point we wish to stress is that the derivatives of functions of two variables are essentially dependent on direction.

Differentiation of complex functions

We have seen that a complex function can always be split into two functions of two variables as in (G.4) and therefore the differentiation of a complex function $w(z) = u + iv$ may be expected to parallel the partial differentiation of $f(x, y)$ given above. This would mean that complex functions were no more that a combination of two real functions. Instead, we define the derivative of $w(z)$ in a way that parallels the the definition of the derivative of a real function in one dimension, namely

$$w'(z) = \lim_{\delta z \to 0} \frac{w(z+\delta z) - w(z)}{\delta z}. \tag{G.27}$$

The crucial step is that we demand that this limit should exist *independent of the direction* in which δz tends to zero. If such a limit exists in all points of some region of the complex plane then we say that $w(z)$ is an *analytic* (or *regular*) function of z (in that region). This restriction on differentiation is very strong and as a result analytic functions are very special, with many interesting properties.

The Cauchy–Riemann conditions

The Cauchy–Riemann conditions are a pair of equations which are *necessarily* satisfied if $w(z)$ is differentiable. To derive them first write out the definition of the derivative in terms of the functions $u(x,y)$ and $v(x,y)$ and set $\delta z = \delta x + i\delta y$.

$$w'(z) = \lim_{\delta(x+iy)\to 0} \left(\frac{u(x+\delta x, y+\delta y) + iv(x+\delta x, y+\delta y) - u(x,y) - iv(x,y)}{\delta x + i\delta y} \right). \tag{G.28}$$

Consider two special cases. In the first we let δz tend to zero along the real x-axis. Therefore we set $\delta y = 0$, so that the limits reduce to partial derivatives with respect to x:

$$w'(z) = u_x + iv_x. \tag{G.29}$$

Repeating with limit taken along the y-axis, so that $\delta x = 0$, we have

$$w'(z) = \frac{1}{i}(u_y + iv_y) = v_y - iu_y. \tag{G.30}$$

If we now demand that these two derivatives $w'(z)$ are the same we have the Cauchy–Riemann equations:

$$\boxed{u_x = v_y, \qquad u_y = -v_x} \tag{G.31}$$

It can also be shown that if the partial derivatives u_x *etc.* are continuous, then the Cauchy–Riemann conditions are sufficient for the derivative $w'(z)$ to exist.

Simple examples of differentiation

- As an example of differentiation and the Cauchy–Riemann conditions consider the function $w(z) = z^3$. Since

$$\begin{aligned} w = u + iv &= (x+iy)^3 \\ &= x^3 + 3x^2(iy) + 3x(iy)^2 + (iy)^3 \end{aligned} \tag{G.32}$$

the real and imaginary parts and their partial derivatives are

$$\begin{aligned} u &= x^3 - 3xy^2, & v &= 3x^2y - y^3. \\ u_x &= 3x^2 - 3y^2, & v_x &= 6xy, \\ u_y &= -6xy, & v_y &= 3x^2 - 3y^2. \end{aligned}$$

These equations show that the Cauchy–Riemann equations (G.31) are indeed satisfied and we can use either (G.29) or (G.30) to identify the derivative as

$$\begin{aligned} w'(z) &= u_x + iv_x = v_y - iu_y \\ &= 3x^2 - 3y^2 + i6xy = 3(x+iy)^2 \\ &= 3z^2. \end{aligned} \tag{G.33}$$

- Similarly we can prove that $w(z) = z^n$ is analytic with a derivative given by $w'(z) = nz^{n-1}$. (The proof is easier if the Cauchy–Riemann conditions are written in terms of polar coordinates and z^n is written as $r^n \mathrm{e}^{in\theta}$.

- Consider the function $w(z) = \sin z$. The real and imaginary parts of the sine function were determined in equation (G.16) so that $w(z) = u + iv$ where

$$\begin{aligned} u &= \sin x \cosh y, & v &= \cos x \sinh y. \\ u_x &= \cos x \cosh y, & v_x &= -\sin x \sinh y, \\ u_y &= \sin x \sinh y, & v_y &= \cos x \cosh y. \end{aligned}$$

  Once again the Cauchy–Riemann equations are indeed satisfied and we can identify the derivative from equation (G.15):

$$\begin{aligned} w'(z) &= u_x + iv_x = v_y - iu_y \\ &= \cos x \cosh y - i \sin x \sinh y \\ &= \cos z. \end{aligned} \tag{G.34}$$

- In similar ways we can show that all the derivatives of ‘standard’ functions parallel those that arise for functions of one real variable.

Taylor’s theorem

We state without proof or qualification that under ‘reasonable’ conditions an analytic function may be represented by a convergent Taylor’s series. In the following development we shall use the theorem in the following form.

$$w(z) = w(z_0) + \frac{1}{1!}(z-z_0)w'(z_0) + \frac{1}{2!}(z-z_0)^2 w''(z_0) + \frac{1}{3!}(z-z_0)^3 w'''(z_0) + \cdots . \tag{G.35}$$

It is also true that any convergent power series defines an analytic function. Proofs of these statements are to be found in the standard texts on complex functions.

/continued overleaf

G.3 Functions and maps

Mathematicians and geographers both use the term map in essentially the same way. A complex function $w(z)$ may be viewed as simply a pair of real functions which define a map in the sense that it takes a point (x, y) in the complex z-plane into a point (u, v) in the complex w-plane by virtue of the two functions $u(x, y)$ and $v(x, y)$. Points go to points, regions go to regions, curves through a point go to curves through the image point, (Figure G.2). The important result is that if $w(z)$ is an analytic function then γ, the angle of intersection of two curves C_1 and C_2 at P, is equal to γ' the angle of intersection of the image curves at the image point. Such maps are said to be be conformal. We proceed immediately to the proof of this statement.

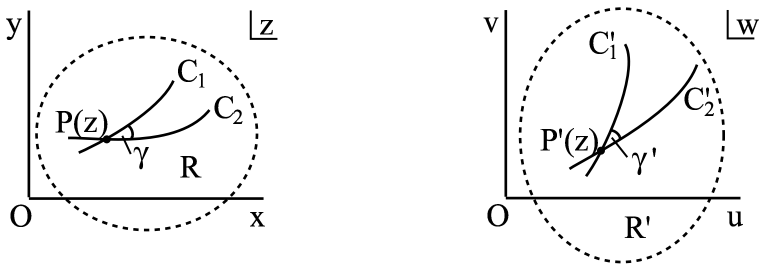


Figure G.2

Proof of the conformality property

Let z_0 be a fixed point on the curve C in the z-plane. Let z be a nearby point on C and write $z - z_0 = r\mathrm{e}^{i\theta}$. Note that θ is the angle between real axis and the chord; in the limit $z \to z_0$, this angle will approach the angle between the real axis and the tangent to C at z_0. Let w_0

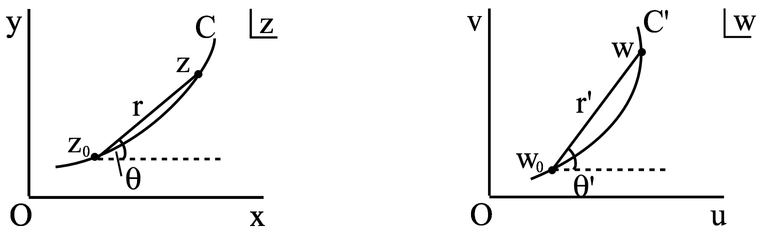


Figure G.3

and w be the corresponding image points and set $w - w_0 = r'\mathrm{e}^{i\theta'}$. Taylor's theorem tells us that

$$w(z) = w(z_0) + \frac{1}{1!}(z - z_0)w'(z_0) + \frac{1}{2!}(z - z_0)^2 w''(z_0) + \cdots , \tag{G.36}$$

so that we can write

$$\frac{w - w_0}{z - z_0} = w'(z_0) + \frac{1}{2!}(z - z_0)w''(z_0) + \cdots . \tag{G.37}$$

The derivative of the function $w(z)$ at z_0 is a unique complex number which depends only on the position z_0 and we can write it as $A(z_0)\exp(i\alpha(z_0))$ where $A(z_0)$ and α are both real.

Therefore in the limit as $z \to z_0$ equation (G.37) becomes

$$\lim_{z\to z_0}\left(\frac{r'}{r}\exp\left[i(\theta'-\theta)\right]\right) = w'(z_0) = A(z_0)\exp(i\alpha(z_0)), \tag{G.38}$$

since the remaining terms on the RHS vanish in the limit. We deduce that

$$\lim_{z\to z_0}\left(\frac{r'}{r}\right) = A(z_0), \tag{G.39}$$

$$\exp i(\theta_0' - \theta_0) = \exp[i\alpha(z_0)], \tag{G.40}$$

where θ_0 and θ_0' are the angles between the tangents and the real axes. Note that the second of these equations can be derived only when $A \neq 0$. The value of $\theta - \theta'$ becomes indeterminate if $A = 0$ so we must therefore demand that $w'(z_0) \neq 0$.

The second of the above limits, when it exists, shows that $\theta_0' = \theta_0 + \alpha(z_0)$, that is the tangent at P is rotated by an angle α when it is mapped to the w-plane. This will be true of all curves through P and consequently the angle of intersection of any two curves will be preserved under the mapping. This is the definition of a conformal mapping.

For a given measurement accuracy we can always find an infinitesimal region around P in which the variation of A and α is imperceptible. Equations (G.39,G.40) then imply that the small region is scaled and rigidly rotated, so preserving its shape. This is the property of orthomorphism.

References and Bibliography

This short bibliography lists some of the papers and books that I found useful in preparing this article. A more extensive list with hypertext links will be published in due course. In addition a Wikipedia search on ‘geodesy’, ‘cartography’, ‘Mercator’, ‘NGGB’, ‘UTM’ *etc.* will generate a great many relevant web pages. On such subjects these Wikpedia pages are faily reliable (but not infallible) and most topics can be verified by following up the many references.

Geodesy

In addition to the Wiki pages we select only two.

- The American Practical Navigator/Chapter 2 at
  http://www.answers.com/topic/the-american-practical-navigator-chapter-2
  This is a good elementary discussion in practical terms.

- The Ordnance Survey of Great Britain (OSGB) have an excellent web site on which can be found *A guide to coordinate systems in Great Britain*. This article discusses modern approaches to reference systems. It can be found at:
  http://www.ordnancesurvey.co.uk/oswebsite/gps/information/index.html

This article mentions geodesy only briefly in the introduction. Those who wish to go (much) deeper can consult the following texts:

- CLARKE, A R (1880), *Geodesy*, Clarendon Press, Oxford.

  A classic which is old enough to be very clear! Clarke discusses techniques which are now very outmoded but this is still a fascinating insight to the work of a nineteenth geodesist. Particularly interesting chapters show how survey results are combined to define the figure of the Earth. Clarke’s ellipsoid of 1866 was the basis of the United States map projections until very recently.

- BOMFORD G, (1971 and later), *Geodesy*, Clarendon Press, Oxford.

  A more modern classic: comprehensive but heavy going. The earlier editions cover traditional (land-based) methods but the later editions have a fair amount on satellite techniques. Expensive to get hold of this book.

- TORGE W, (1980), *Geodesy*, de Gruyter, ISBN-13: 978-3110170726. (3rd edition)

  Clear but fairly advanced survey based on modern satellite methods. The latest edition is available in a reasonably priced paperback format.

General cartography and projections

There are not many textbooks which cover the more mathematical aspects of projections but the following is a good text at an intermediate level covering the general features of all projections and details of some. The mathematics is not too demanding but does not extend to a derivation of the full Redfearn formulae. Lots of interesting material.

- MALING D H,
  (1992), *Coordinate Systems and Map Projections*,
  Pergamon, ISBN: 0-08-037234-1.

The following is a comprehensive summary of just about all projections in use, including of course *all* the Mercator projections: normal, transverse and oblique. The introductions to the projections are very readable and moreover each is complemented with an historical survey. BUT there are no derivations of any projection formulae.

- SNYDER J P, (1987),
  Map Projections: a Working Manual,
  US Geological Survey, Professional Paper 1395
  Published by US Government Printing Office but also available on the web at:
  http://pubs.er.usgs.gov/usgspubs/pp/pp1395

For a more advanced treatment covering all projections (but with many gaps in the mathematical development requiring large amounts of work to fill in):

- BUGAYEVSKIY L M AND SNYDER J P, (1995),
  Map Projections: A Reference Manual,
  CRC Press, ISBN: ISBN 0748403043.

The article referred to in the geodesy section, *A guide to coordinate systems in Great Britain*, also includes a statement of the transverse Mercator projection formulae (without derivations) with examples of transforming between grid and geographical coordinates. There is also an excellent (short) Wikipedia page entitled ‘British National Grid Reference System’ which gives a direct link to the above article.

Survey Review (SR)

The Survey Review was the principal British source of papers on surveying and cartography at the time (mid twentieth century) when the first modern British maps based on the transverse Mercator projection were being prepared by the OSGB. Note that until 1962 the journal was entitled the Empire Survey Review. Note also that this journal is published in parts (from four to eight a year) and bound in two year periods so both volume and part number are specified. Sadly this journal is not readily available in general libraries.

Two important papers are listed below. The first, by Lee, is the first article in the journal to present a correct derivation of the Transverse Mercator projection formulae. The second paper, by Redfearn, presents a derivation of the series to high enough order to be applicable to all practical problems. By one of those quirks of fate it is Redfearn's name that has entered the literature. The set of papers by Hotine present a much less elegant derivation which refuses to countenance the existence of complex variable methods: they are not discussed here.

- LEE L P, (1946), Survey Review, Vol **8**, Part 58 pp 142–152.
  The transverse Mercator projection of the spheroid. (Errata and comments in Vol **8**, Part 61 pp 277–278).
- REDFEARN J C B, (1948), Survey Review, Vol **9**, Part 69 pp 318–322.
  Transverse Mercator formulae.
- HOTINE, M (1946, 1948), Survey Review, Vol **8**, Part 62 pp 301–311 and Vol **9**, Part 63 pp 29–35, Part 64 pp 52–70, Part 65 pp 112–123, Part 66 pp 157–166.
  The orthomorphic projection of the spheroid, parts I–V.

It should be said at once that these papers are fairly terse when it comes to the derivations of the projection formulae. The present article errs in the other direction and it is fair to say that no extra details will be found in the original papers. Note also that this article uses a different convention for the names of axes; basically x and y are exchanged.

The discussion of geodesic problems in Chapter 11 includes a greatly expanded version of the following paper:

- VINCENTY, T (1976), Survey Review, Vol **23**, Part 176 pp 88–93.
  Direct and inverse solutions of geodesics on the ellipsoid with applications of nested tables.

Mathematics

The appendices include all the mathematics we require and a little more besides. They are derived from first principles and should hopefully not require further background reading. In their preparation I found that modern texts were not helpful on the whole because they were too distant from application. The older books were much more useful. A few texts are listed here.

Spherical Trigonometry

- SMART W M, (1962), *Textbook on spherical astronomy*, Cambridge.

  First chapter is a compact survey of Spherical Trigonometry.

- TODHUNTER I, (1859 . . . 1901), *Spherical Trigonometry*, Macmillan (London).

  A splendid traditional account of the subject. Many editions. Final edition (1901) revised by J G Leathem is best. An earlier edition is on the web at the Cornell centre for Historical Mathematical Monographs:

  http://historical.library.cornell.edu/math/

- EUCLID, *The elements*.

  Book 11 contains results required by Todhunter. There are several good references in the Wikipedia page for Euclid. One of the nicest is a website by David Joyce at Clark university:
  http://aleph0.clarku.edu/ djoyce/java/elements/toc.html

Differential Geometry

- WEATHERBURN C E, (1939), *Differential Geometry*, Cambridge.

  An old, but good, straightforward account in approachable notation. Modern texts tend to set up much more 'elaborate' machinery before encountering reality.

Lagrange Expansions

The derivation of the Lagrange expansions has essentially disappeared from modern texts. The proof in Appendix B is a combination of

- WHITTAKER C E, (1902), *Modern Analysis*, Cambridge.
- COPSON E T, (1935), *Theory of Functions of a Complex Variable*, Oxford

The twelfth order Lagrange series are published in the Philosophical Magazine.

- BICKLEY W G, MILLER J C P, (1949) *Phil Mag* **3** 35-36.
  Notes on the reversion of a series.

Index X

Index