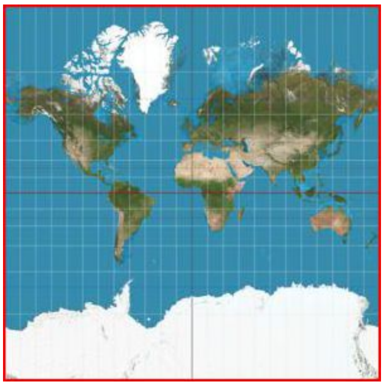
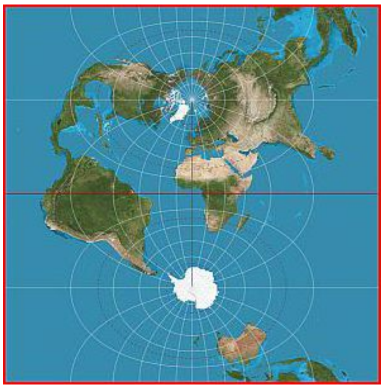
THE MERCATOR PROJECTIONS

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





PETER OSBORNE

EDINBURGH

2013

This article describes the mathematics of the normal and transverse Mercator projections on the sphere and the ellipsoid with full derivations of all formulae.

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 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 central meridian at 2°W is true: all other meridians deviate from grid lines. The UTM series is a set of 60 distinct Transverse Mercator projections each covering a width 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 and distant approximately 180km east and west of it. The scale on the central meridian is constant but it is slightly less than the nominal value.

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 has 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 (NMS and TMS) before going on to discuss the normal and transverse Mercator projection on the ellipsoid (NME and TME). As a result, the length of this document has doubled and redoubled but I have resisted the temptation to cut out 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 (1948) talks of a "a particularly tough spot of work" and Hotine (1946) talks of reversing series by "brute force and algebra"—so be warned. Repeating this may be seen as a perverse undertaking on my part. 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 also 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 books that I found to be of assistance are listed in the Literature (L) or References R but they are supplemented with research papers and web material.

This second edition (2013) adds further material and enters the world of hyperref.

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, 2013

Source files

peter.1@mercator.myzen.co.uk

Contents

Chapter 1

Introduction

1.1 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 (from 1734–1749) , French scientists measured a meridian arc of about one degree of latitude in Scandanavia crossing the Arctic circle (French Geodesic Mission to Finland) and a second arc of about three degrees crossing the equator in Peru (French Geodesic Mission to Peru) and confirmed for the first time the oblateness of the ellipsoid. (See Clarke (1880), pp 4–13) More accurate measurements of a French arc, supervised by Delambre from 1792, determined the meridian quadrant from equator to pole through Paris, as 5130766 toises (of Peru), the standard of length used in the measurement of the Peruvian arc. The standard toise bar was held in Paris. The French then defined the meridian quadrant to be 10,000,000 metres and the first standard metre bar was constructed at a length of 0.5130766 toise. (Clarke (1880), pp 18–22).

In 1830 George Everest calculated an ellipsoid using what he took to be the best two arcs, an earlier Indian Arc surveyed by his predecessor William 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 George Biddell Airy discussed sixteen arcs before arriving at the result he published in 1830:

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

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

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

Modern satellite methods have introduced global ellipsoid fits to the geoid, that for the Geodetic Reference System of 1980 (GRS80) being

$$a = 6378137\text{m} \qquad b = 6356752.3\text{m} \qquad f = 1/298.26 \qquad \text{[GRS80/WGS84]} \tag{1.3}$$

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 semi-axis 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.

1.2 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, 1999) are defined with respect to a datum OSGB36 (established for the 1936 re-survey) 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 the World Geodetic System suchs WGS72, GRS80, WGS84 are defined with respect to the Earth's centre of mass and a defined orientation of axes. See Global Positioning System.

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.

1.3 Cartography

A topographic survey produces a set of geographical locations (latitude and longitude) referenced to some specified datum. The problem of cartography, the representation of the latitude–longitude data on thedatum 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. Formulae (without derivations) may also be found in in *Map Projections—A Working Manual*, (Snyder, 1987).

We define a map **projection** by two functions $x(\phi,\lambda)$ and $y(\phi,\lambda)$ which specify the plane Cartesian coordinates (x,y) corresponding to the latitude and longitude coordinates (ϕ, λ). For the above projections 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 convention x-north and y-east is also useful when complex mathematics is used, for example Karney (2011),

1.4 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. (This follows from the Theorema Egregium of Gauss. See Gauss (1827)) Thus all *maps* are compromises to some extent and they must fail to meet at least one of the following five properties

1. *One-to-one correspondence of points*. This will normally be the case for large scale maps of small regions but global maps 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 the poles are singular because they project into lines.

2. *Uniformity of point (or local) scale*. By point scale we 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 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 Gall-Peters projection but the downside is that such equal area projections distort shapes in the large. The Mercator projections do not preserve area .

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.

1.5 The representative fraction (RF) and the scale factor

The OSGB (1999) produces many series of maps of Great Britain. For example there are over two hundred 'Landranger' map sheets which are endorsed with the phrase '1:50,000 scale', implying that each 80cm×80cm sheet covers an area of 40km×40km on the ground. This statement is misleading. To clarify the issue we distinguish two concepts: the **representative fraction** and the **scale factor**.

There are four conceptual steps involved in making a map: (1) a survey produces latitude, longitude data on a spherical or ellipsoidal datum; (2) the datum is reduced to a small model, the reduction factor being the representative fraction; (3) the position locations on the small model are projected (by specified formulae, not simply literally) onto a cylindrical or conical sheet 'wrapped' about the model; (4) the sheet is cut and opened out to give a planar map projection. If, in this construction, the cylinder was tangential to the model at the equator, the map distances on the equator will equal these on the reduced model and we say that the map scale factor is unity on the equator. The scale factor at other points will vary in a way which is determined by the projection formulae and maintaining a small variation of the scale factor over the map is an important criterion in the choice of projection. Note that the second and third of the conceptual steps may be interchanged. Of course one doesn't construct physical models, neither does one wrap them in sheets of paper: only one step is needed from a data-base of locations straight to a printer by way of a computer program.

Returning to the example of the 1:50,000 map series produced by the OSGB we now interpret that figure as the *constant* representative fraction. For the details of the projection we have to consult the OSGB (1999) literature where we find that the map scale factor, the ratio of *nearby* distances on the map divided by the corresponding distance on the reduced datum model, is fixed as 0.9996 on the meridian at 2°W and elsewhere varies according to precise formulae which we shall display later. This implies that 2cm on the map, the spacing between the grid lines, represents a true distance on the ground of between 0.9996km and 1.0007km. Thus each 80cm*x*80cm map sheet covers only approximately 40km×40km. The precise variation of scale factor with position will be calculated in later chapters.

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.

This is an appropriate point to mention the concept of a zero dimension for a map projection. This is the smallest size that can be printed on the map and remain visible to the naked eye. Before the age of digital maps this was often taken as 0.2mm, corresponding to 10m on a 1:50000 map. Thus narrow streams or roads cannot be shown to scale on such a map. Even wide roads, such as motorways, are often shown at exaggerated scales. Modern digital systems are more powerful since they can show more and more detail as the map is zoomed.

1.6 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 projection but it is usually shown on small scale maps covering large areas, such as world maps. and it is usually omitted on large scale maps of small areas. For the OSGB 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 of the sheet.

The projected map is usually constructed in a plane Cartesian coordinate 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 1:50000 map sheets have a grid at a 2cm intervals corresponding 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, nor need it be a Cartesian grid.

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. (Occasionally 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**. This clear distinction in terminology shall be adhered to in this work but it is by no means universal.

On normal Mercator projections the projected graticule is aligned to the underlying Cartesian system so the constant-x grid lines correspond to meridians running north-south. The projection is also constructed to ensure that the azimuth and grid bearing are equal. Therefore a **rhumb line**, a course of constant azimuth on the sphere, becomes a straight line on the projection.

On the transverse Mercator projections the situation is more complicated. The projection of the graticule is a set of complex curves which, in general, are not aligned to the underlying Cartesian reference grid: the only exceptions are the equator and the central meridian. As a result the constant-x grid lines do not run north-south and the azimuth is not equal to the grid-bearing, instead it is equal to the angle between the projected meridian and the projected line segment AB. The angle between the projected meridian and the constant-x grid line is called the **grid convergence**. On large scale maps of restricted regions it is a small angle but nonetheless important for high accuracy work. The OSGB 1:50000 map sheets state the value of the grid convergence at each corner of the sheet.

1.7 Historical outline

Gerardus Mercator (1512–1594) did not develop the mathematics that we shall present for “his” projection (NMS) in Chapter 2; 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 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. (Full text is available on Wikipedia at Mercator 1569 world map, Legend 3).

This is an admirably clear statement of 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 length of a parallel on the projection as one moves away from the equator. Since the circumference of a parallel at latitude ϕ is $2\pi R\cos\phi$ it must be scaled up by a factor of $\sec\phi$ so that it has the same length as the equator on the projection ($2\pi R$). Thus, to guarantee that an azimuth is equal to its corresponding grid bearing, or equivalently rhumb lines project to straight lines, it is necessary to increase the meridian scale at latitude ϕ by a factor of $\sec\phi$. This leads to a gradual increase in the spacing of the parallels on the projection as against the uniform spacing of the equirectangular projection.

Exactly how Mercator produced his map is not known: he left no account of his method. He was familiar with the writing of Pedro Nunes (Randles, 2000) who had showed that rhumb lines are spirals from pole to pole on the sphere, and he had found means of marking such rhumbs on his terrestial globe of 1541. It is possible that he employed mechanical means, using templates, one for each of the principal rhumbs. He *could* have measured the coordinates of points on a rhumb and transferred them to a plane chart with the parallels adjusted so that the rhhumbs became straight lines. On the other hand his latitude scale is fairly accurate and most writers assume that he had some means of calculating the spacings. Ten such methods are discussed by Hollander (2005).

The first to publish an account of the construction of a Mercator chart was a Cambridge professor of mathematics named Edward Wright 1558?–1615). His publication entitled 'The correction of certain errors in navigation' (Wright, 1599) discusses the errors of the equirectangular projection and shows how the angles, at least, are correct in Mercator's chart and goes on to explain the construction of such a chart by using a table of secants. He published a very fine chart based on accurate positions taken from a globe modelled by his compatriot Emery Molyneux. For many years thereafter the charts were widely described as Wright-Molyneux map projection.

In addition to his mathematical derivation of the projection Wright imagined a physical construction:

> 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 how this 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: it must be infinitely extensible and able to withstand infinite pressure as it slides over the perfectly smooth cylinder. 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 John 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. See Halley (1696) 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) is trivial.

Having given credit to Wright, Bond and others it is now believed the English mathematician Thomas Harriot (1560–1621) was possibly the first to calculate the spacings of the Mercator projection. His unpublished works have had to await study until very recently. They contain evidence of a method equivalent to that of Edward Wright and moreover he seems to have devised a formula equivalent to the logarithmic tangent formula derived from calculus almost one hundred years later. (Pepper, 1967; Lohne, 1965, 1979; Taylor and Sadler, 1953; Stedall, 2000)

The transverse Mercator projection on the sphere was included in a set of seven new projections published (Lambert, 1772) by the Swiss mathematician and cartographer, Johann Heinrich Lambert. As we shall see in Chapter 3, the derivation of this projection on the sphere is a straightforward application of spherical trigonometry starting from the normal Mercator result. The generalisation to the ellipsoid was carried out by Carl Friedrich Gauss (1777–1855) in connection with the survey of Hanover commenced in 1818. His projection is conformal and preserves true scale on one meridian: this is the projection we shall term TME.

Gauss left few details of his work and the most accessible account is that of Krüger (1912). For this reason the transverse Mercator projection on the ellipsoid is often called the Gauss–Krüger projection. Kruger developed two expressions for the TME projection, one a power series in the longitude difference from a central meridian and a second using a power series in the parameter which describes the flattening of the ellipsoid. The first of these methods was extended to higher order by Lee (1945), and Redfearn (1948) in Britain and by Thomas (1952) in the USA. Their results are used for the OSGB map series and the UTM series respectively. (There are only sub-millimetre differences). It is this method which is described in this article.

It should be noted that in addition to the representation of the projection by power series approximations an *exact* solution was devised by E. H. Thompson and published by Lee (1976). That solution and the second solution of Krüger are available in Karney (2011).

Finally, I give the abstract of the 1948 paper by Redfearn. The actual paper is highly condensed (6 pages!) and although it is available (at a price) it will add nothing to the work presented here.

> The Transverse Mercator Projection, now in use for the new O.S. triangulation and mapping of Great Britain, has been the subject of several recent articles in the Empire Surpey Review. The formulae of the projection itself have been given by various writers, from Gauss, Schreiber and Jordan to Hristow, Tardi, Lee, Hotine and other—not, it is to be regretted, with complete agreement, in all cases. For the purpose for which these formulae have hitherto been employed, in zones of restricted width and in relatively low latitudes, the completeness with which they were given was adequate, and the omission of certain smaller terms, in the fourth and higher powers of the eccentricity, was of no practical importance. In the case of the British grid, however, we have to cover a zone which must be considered as having a total width of some ten to twelve degrees of longitude at least, and extending to latitude 61N. This means, firstly, that terms which have as their initial co-efficients the fourth and sixth powers of the longitude will be of greater magnitude than usual, and secondly that the powers of tan are likewise greatly increased. Lastly, an inspection of the formulae (as hitherto available) shows a definite tendency for the numerical co-efficients of terms to increase as the terms themselves decrease.

1.8 Chapter outlines

Chapter 2 starts by describing by discussing angles and distances on a sphere of radius equal to the mean radius of the WGA84 ellipsoid. We consider the class of all ‘normal’ (equatorial) cylindric projections onto a cylinder tangential to the equator of a sphere and compare and contrast four 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 by a comparison of infinitesimal elements on the sphere and the plane. Rhumb lines and their properties are defined in detail and contrasted with great-circles. Secant projections are introduced to control scale variation.

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 great circle formed by a meridian and its continuation, *e.g.* the meridian at 21°E and 159°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) at high 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. 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, (e) auxiliary latitudes and double projections.

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.

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 applies the general results of Chapter 7 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).

Appendices There are eight mathematical appendices. Some of these were developed for teaching purposes 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.
- **H** Maxima code.

Blank page. A contradiction.

Chapter 2

Normal Mercator on the sphere: NMS

Coordinates and distance on the sphere. Infinitesimal elements and the metric. Normal cylindrical projection. Angle transformations and scale factors. Four examples of normal cylindrical projections. Derivation of the Mercator projection. Rhumb lines and loxodromes. Distances on the Mercator projection. Secant (modified) normal cylindrical projections.

2.1 Coordinates and distance on the sphere

Basic definitions

The intersection of a plane through the centre of the (spherical) Earth with its surface is a **great circle**: other planes intersect the surface in **small circles**. The intersections of the rotation axis of the Earth with its surface define the poles N and S. The **meridians** are those lines joining the poles which are defined by the intersection of planes through the rotation axis with the surface: the meridians are great circles. The **parallels** are defined by the intersections of planes normal to the rotation axis with the surface and the equator is the special case when the plane is through the centre: the equator is a great circle and other parallels are small circles.

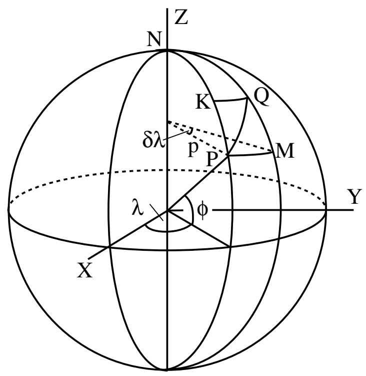


Figure 2.1

The position of a point P on the sphere is denoted by an ordered pair (ϕ, λ) of latitude and longitude values. **Latitude** is the angle between the normal at P and the equatorial plane: it is constrained to the interval of $[-90°, 90°]$ or $[-\pi/2, \pi/2]$ radians. On the sphere any normal to the surface passes through the centre. **Longitude** is the angle between the meridian through P and an arbitrarily chosen reference meridian (established at Greenwich by the Prime Meridian Conference (1884)): it may be defined on either of the intervals $[-180°, 180°]$ or $[0°, 360°]$ with radian equivalents $[-\pi, \pi]$ and $[0, 2\pi]$ radians respectively. 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 great circle through the points P and Q.

Geographical coordinates are normally given as degrees, minutes and decimal seconds or in degrees and decimal minutes or simply in decimal degrees. In equations, however, all angles must be in radians—except where explicitly stated. The unit **mil**, such that 6400mil=2π radians=360°, is sometimes used for small angles, in particular the grid convergence defined in Section 3.6. The relations between these units are as follows:

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

$$1^\circ = 0.0174533\text{ rad},\quad 1' = 0.000291\text{ rad} = 0.296\text{ mil},\quad 1'' = 0.00000485\text{ rad}. \qquad (2.1)$$

$$1\text{mil} = 0.000982\text{ rad} = 0^\circ.0563 = 3'.37 = 202''.$$

Radius of the sphere

The Earth is more accurately represented by an ellipsoid (Chapter 5) with semi-major axis (equatorial radius), a, and semi-minor axis b (often mistakenly called the polar radius) and there are several choices for the radius, R, of a sphere approximating such an ellipsoid. (The notation R will be applied to both the full size approximation or to the sphere reduced by the representative fraction, RF). The most important possibilities are:

- the major axis of the ellipsoid, a,
- the tri-axial arithmetic mean, $(a+a+b)/3)$,
- radius of equal area sphere, $(a/\sqrt{2})\left[1+e^{-1}(1-e^2)\tanh^{-1}e\right]^{1/2}$, (Section 5.13),
- radius of equal volume sphere, $(a^2b)^{1/3}$.

For other possible choices see Maling 1992 (page 76) and Wikipedia Earth radius For WGS (1984) the values to within 10 cm. are

- $a = 6,378,137.0$m, $\quad b = 6,356,752.3$m,
- triaxial mean radius: 6,371,008.8m,
- equal volume radius: 6,371,000.8m,
- equal area radius: 6,371,007.2m.

In practical calculations with a spherical model it is acceptable to take the mean radius as $R = 6371$km (3958 miles). For example this is the value taken by the FAI (International Air Federation). No higher accuracy is required since we must use an ellipsoidal model for more precise calculations. For this radius the circumference is 40,030km (24,868 miles) and the meridional quadrant (pole to equator) is 10,007km (6217 miles). One degree of latitude corresponds to 111.2km. and one minute of latitude corresponds to 1853m.

Cartesian coordinates

If the radius of a parallel circle is $p(\phi) = R\cos\phi$ the Cartesian coordinates of P are

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

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 R by $R+h$ in the direct transformations: the inverse relations for ϕ and λ are unchanged but they are supplemented with the equation

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

The unit vector, **n**, from the centre of the sphere toward a point on the surface is

$$\mathbf{n} = (\cos\phi\,\cos\lambda,\ \cos\phi\,\sin\lambda,\ \sin\phi). \tag{2.5}$$

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 two points in general position the important distances are along a great circle, which is the shortest distance, or along a rhumb line which, by definition, intersects meridians at constant azimuth. (Rhumb line distances are discussed in Section 2.5). For example, if the points are at the same latitude we can calculate the rhumb distance between them by measuring along the parallel circle; the shorter great circle distance deviates north (south) in the northern (southern) hemisphere.

The only trivially calculated distances are those measured along meridians or parallels: on the meridian in Figure 2.1 we have $PK = R\,\delta\phi$ and on the parallel $PM = p(\phi)\delta\lambda = R\cos\phi\,\delta\lambda$ (where $\delta\phi$ and $\lambda\phi$ are in radians). For widely separated points these become $PK = R(\phi_2 - \phi_1)$ and $PM = R\cos\phi\,(\lambda_2 - \lambda_1)$. It is useful to have some feel for the distances on meridians and parallels on the sphere (radius 6371km). To the nearest metre:

	radius (m)	circumference	1°	1′	1″
meridian	6,371,008	40,030,173	111,195	1853	31
equator	6,371,008	40,030,173	111,195	1853	31
parallel at 15°	6,153,921	38,666,178	107,406	1790	30
parallel at 30°	5,517,454	34,667,147	96,297	1604	27
parallel at 45°	4,504,983	28,305,607	78,626	1310	22
parallel at 60°	3,185,504	20,015,096	55,597	927	15
parallel at 75°	1,648,938	10,360,571	28,779	480	8

Table 2.1

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 therefore, to avoid discrepancies, the nautical mile is now defined by international treaty as *exactly* 1852m (1.151 miles). The original definition remains a good rule of thumb for approximate calculations but note that it corresponds to a spherical Earth model of radius equal to 6366.7km rather than the value of 6371km which we used for the previous table.

The **great circle distance**, g_{12} between two points in general position is R times the angle (in radians) which the circular arc between them subtends at the centre. That angle is defined by the two unit vectors giving the positions of the end points.

$$\begin{aligned}
\mathbf{n}_1 &= (\cos\phi_1\,\cos\lambda_1,\ \cos\phi_1\,\sin\lambda_1,\ \sin\phi_1),\\
\mathbf{n}_2 &= (\cos\phi_2\,\cos\lambda_2,\ \cos\phi_2\,\sin\lambda_2,\ \sin\phi_2),\\
\mathbf{n}_1\cdot\mathbf{n}_2 &= \cos\phi_1\,\cos\phi_2\,\cos(\lambda_2-\lambda_1)+\sin\phi_1\sin\phi_2\\
g_{12} &= R\cos^{-1}\left[\cos\phi_1\,\cos\phi_2\,\cos(\lambda_2-\lambda_1)+\sin\phi_1\sin\phi_2\right]. \qquad (2.6)
\end{aligned}$$

This formula is not well conditioned at short distances and alternative forms are preferable. (See Wikipedia Great-circle distance). There are a number of great circle distance calculators available on the web. The FAI use the mean sphere (and WGS84). The Ed Williams' Aviation page has a more comprehensive list of Earth models. The csgnetwork uses a model a model with radius 6366.7km: it may also be used to calculate way point values.

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 specific, 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

Figure 2.2

lines 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

$$\text{arc}\,(AB) - AB = R\theta - 2R\sin\frac{\theta}{2} = R\theta - 2R\left(\frac{\theta}{2} - \frac{1}{3!}\frac{\theta^3}{8} + \cdots\right) = \frac{R\theta^3}{24} + O(R\theta^5). \qquad (2.7)$$

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 by measurement, if we take $\delta\phi < (24/R)^{1/3} \approx 0.0155$rad, corresponding to $53'$ or a meridian arc length of 99km. Similarly, setting $\theta = \delta\lambda$ and replacing R by $p = R\cos\phi$, the difference between the arc and chord PM at a latitude of 45° (where $\cos\phi = 1/\sqrt{2}$) is less than 1m if $\delta\lambda$ is less than $59'$, corresponding to an arc length of 78km on that 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 which can be mapped without need for projection.

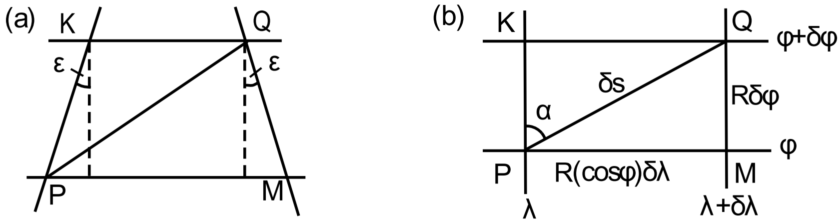


Figure 2.3

We 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 $PM = R\cos\phi\,\delta\lambda$ we have

$$KQ - PM = \delta\phi\,\frac{d}{d\phi}\,(R\cos\phi\,\delta\lambda) = -R\sin\phi\,\delta\lambda\,\delta\phi. \tag{2.8}$$

Now the distance $PK \approx R\delta\phi$ so the small angle ε, called spherical convergence (note **[1]**) is given by

$$\varepsilon \approx \sin\varepsilon = \frac{PM - KQ}{2}\frac{1}{PK} = \frac{1}{2}\sin\phi\,\delta\lambda. \tag{2.9}$$

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

$$\tan\alpha = \lim_{Q\to P}\frac{R\cos\phi\,\delta\lambda}{R\delta\phi} = \cos\phi\,\frac{d\lambda}{d\phi}\,, \tag{2.10}$$

$$\delta s^2 = PQ^2 = R^2\,\delta\phi^2 + R^2\cos^2\!\phi\,\delta\lambda^2. \tag{2.11}$$

The latter result also follows directly from equations (2.2):

$$\begin{aligned}
\delta X &= -(R\sin\phi\,\cos\lambda)\delta\phi - (R\cos\phi\,\sin\lambda)\delta\lambda \\
\delta Y &= -(R\sin\phi\,\sin\lambda)\delta\phi + (R\cos\phi\,\cos\lambda)\delta\lambda \\
\delta Z &= (R\cos\phi)\delta\phi, \qquad\qquad (2.12)\\
\delta s^2 &= \delta X^2 + \delta Y^2 + \delta Z^2, \\
\delta s^2 &= R^2\cos^2\!\phi\,\delta\lambda^2 + R^2\,\delta\phi^2. \qquad\qquad (2.13)
\end{aligned}$$

This expression defines the metric on the surface of the sphere.

2.2 Normal (equatorial) cylindrical projections

The **normal**, or equatorial, **aspect** of cylindrical projections of a (reduced) sphere of radius R are defined on a cylinder of radius R which is tangential to the sphere on the equator as shown in Figure (2.4). (When the cylinder is tangential to a meridian the aspect is said to be **transverse**: for other orientations the aspect is **oblique**.) 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.

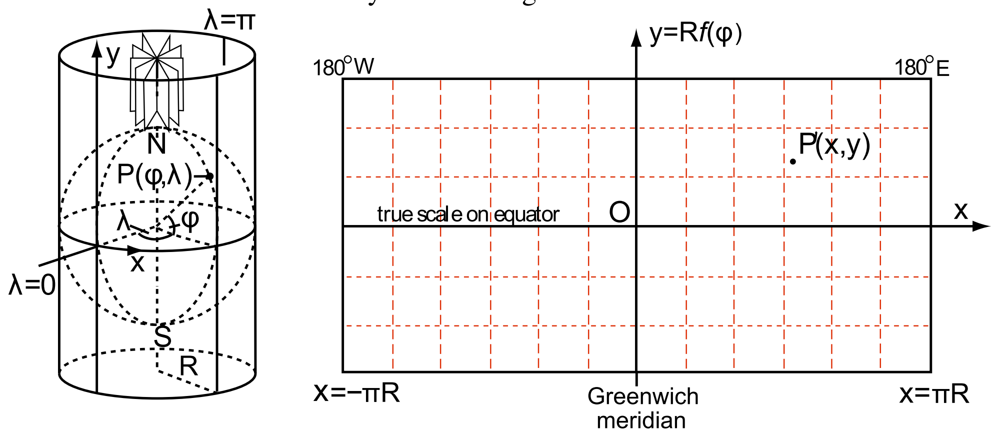


Figure 2.4: The normal cylindrical projection

The projection takes points on the 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 on many web sites): see the next section. The cylinder is then cut along a generator which has been taken at $\lambda = 180°$ in Figure 2.4 but could have been chosen at any longitude. Finally the cylinder is unrolled to form the flat map. 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 $x = R\lambda$ (radians). For the y-equation of the projection we admit *any* "sensible" function of ϕ, irrespective of whether or not there is a geometrical interpretation. Therefore normal cylindrical projections are defined by (note **[2]**)

$$x(\lambda,\phi) = R\lambda, \tag{2.14}$$
$$y(\lambda,\phi) = Rf(\phi), \tag{2.15}$$

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 projection. The spacing of the meridians on the projection is uniform but the spacing of the parallels depends on the choice of the function $f(\phi)$.

Note that *all* normal cylindrical projections have singular points: the *points* N, S at the poles transform into *lines* given by $y = Rf(\pm\pi/2)$. On the sphere meridians intersect at the poles but on normal 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 aspects 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 (2.14, 2.15) define a projection to a map of constant width, W, equal to the length of the equator, $2\pi R$. Since the true length of a parallel is $2\pi R\cos\phi$, the scale factor, the map projection length divided by the corresponding true length (on the reduced sphere), along a parallel is equal to $\sec\phi$: this factor increases 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 other lines, will depend on $f(\phi)$.

The actual printed projection in Figure 2.4 has a value of W approximately equal to 8cm, corresponding to $R = 1.27$cm and an RF (representative fraction) of approximately 1 over 500 million or 1:500M.

Angle transformations on normal cylindrical projections: conformality

In Figure 2.5 we compare the rectangular infinitesimal element $PMQK$ on the sphere with the corresponding rectangular element $P'M'Q'K'$ on the projection. We define the angle α on the sphere to be the **azimuth** and the corresponding angle β on the projection to be the **grid bearing**. This distinction in terminology is not widespread.

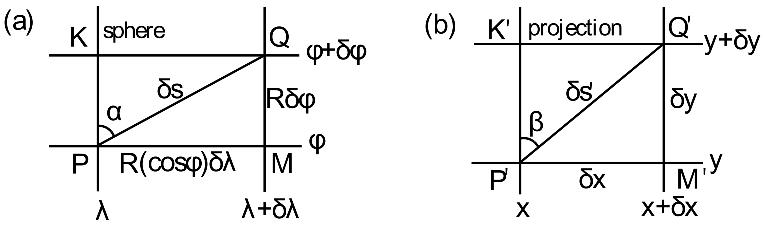


Figure 2.5

The geometry of the rectangular elements gives

$$\text{(a)}\quad \tan\alpha = \frac{R\cos\phi\,\delta\lambda}{R\,\delta\phi} \qquad \text{and} \qquad \text{(b)}\quad \tan\beta = \frac{\delta x}{\delta y} = \frac{\delta\lambda}{f'(\phi)\,\delta\phi}, \tag{2.16}$$

so that

$$\tan\beta = \frac{\sec\phi}{f'(\phi)}\tan\alpha. \tag{2.17}$$

If $\alpha=\beta$ it follows that the angle between *any* two azimuths is equal to the angle between the corresponding grid bearings. In this case we say that the projection is **conformal**. The condition for this is that $f'(\phi) = \sec\phi$: this is the basis for the derivation of the Mercator projection given in Section 2.4.

The point scale factor

Define μ, the point scale *at the point* P' on the projection by

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

$$= \lim_{Q\to P} \frac{\sqrt{\delta x^2 + \delta y^2}}{\sqrt{R^2\cos^2\phi\,\delta\lambda^2 + R^2\,\delta\phi^2}}. \tag{2.19}$$

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

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

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

On a parallel $\delta\phi$ and δy are zero and $\delta x = R\delta\lambda$. The scale factor in this case is conventionally denoted by k.

parallel scale: $$k(\phi) = \sec\phi. \tag{2.21}$$

The parallel scale factor, plotted alongside, is the same for all normal cylindrical projections.

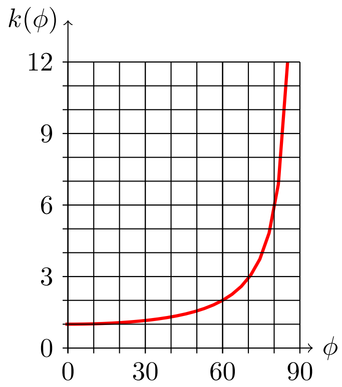


Figure 2.6

Point scale factor in a general direction: isotropy of scale

Equations (2.16) 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

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

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

where $\beta(\alpha,\phi)$ can be found from Equation (2.17). For a conformal projection with $\alpha = \beta$ the general scale factor is equal to $\sec\phi$, independent of α: it is **isotropic**.

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

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

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 Four examples of normal cylindrical projections

Consider the following projections of the unit sphere, $R = 1$:

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

The following table summarizes the properties of these projections.

	equirectangular	equal-area	Mercator	central
x	λ	λ	λ	λ
x-range	$(-\pi, \pi)$	$(-\pi, \pi)$	$(-\pi, \pi)$	$(-\pi, \pi)$
$y = f(\phi)$	ϕ	$\sin\phi$	$\ln[\tan(\phi/2 + \pi/4)]$	$\tan\phi$
y-range	$(-\pi/2, \pi/2)$	$(-1, 1)$	$(-\infty, \infty)$	$(-\infty, \infty)$
$f'(\phi)$	1	$\cos\phi$	$\sec\phi$	$\sec^2\phi$
meridian h	1	$\cos\phi$	$\sec\phi$	$\sec^2\phi$
parallel k	$\sec\phi$	$\sec\phi$	$\sec\phi$	$\sec\phi$
equator k	1	1	1	1
area (hk)	$\sec\phi$	1	$\sec^2\phi$	$\sec^3\phi$
(2.17): $\tan\beta =$	$\sec\phi\tan\alpha$	$\sec^2\phi\tan\alpha$	$\tan\alpha$	$\cos\phi\tan\alpha$
aspect ratio	2	π	0	0
	Figure 2.7	Figure 2.8	Figure 2.10	Figure 2.11

Table 2.2

The four projections are shown in Figures (2.7–2.11). The maps all have the same x-range of $(-\pi, \pi)$ (on the unit sphere) but varying y-ranges. They are portrayed on the printed page with a width of 12cm corresponding to an RF of approximately 1/300M. 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 for all projections but their height depends on $f(\phi)$.

On the equator $\phi = 0$ so that both $\cos\phi$ and $\sec\phi$ are equal to unity. All the scale factors are unity and equation (2.17) shows that $\alpha = \beta$ where any line segment crosses the equator. The projections are perfectly well behaved and quite suitable for accurate large scale mapping close to the equator: both relative distance and shape are well preserved in a comparison with their actual representations on the sphere. Look at an actual globe and compare all four projections in locations such as Africa, the Caribbean and Indonesia.

Away from the equator all the projections have a parallel scale factor equal to $\sec\phi$: a necessary consequence of the attempt to project the spherical surface onto a rectangular domain. This factor increases to infinity as $\phi \to \pm\pi/2$ so that the poles of the sphere are stretched out to lines across the full width of the projection, at finite or infinite values of the y-coordinate. The poles are singular points of the projection where the one-to-one correspondence between sphere and projection breaks down. The horizontal stretching at high latitudes is leads to distortions in all four projections when they are compared with an actual globe. The shape of Alaska is good measure of this distortion. Only the Mercator projection preserves good local shape.

Relative area is another good criterion in assessing the projections. On the globe the area of Greenland is 1/8 that of South America and 1/13 that of Africa. Only the Lambert equal-area projection preserves these values.

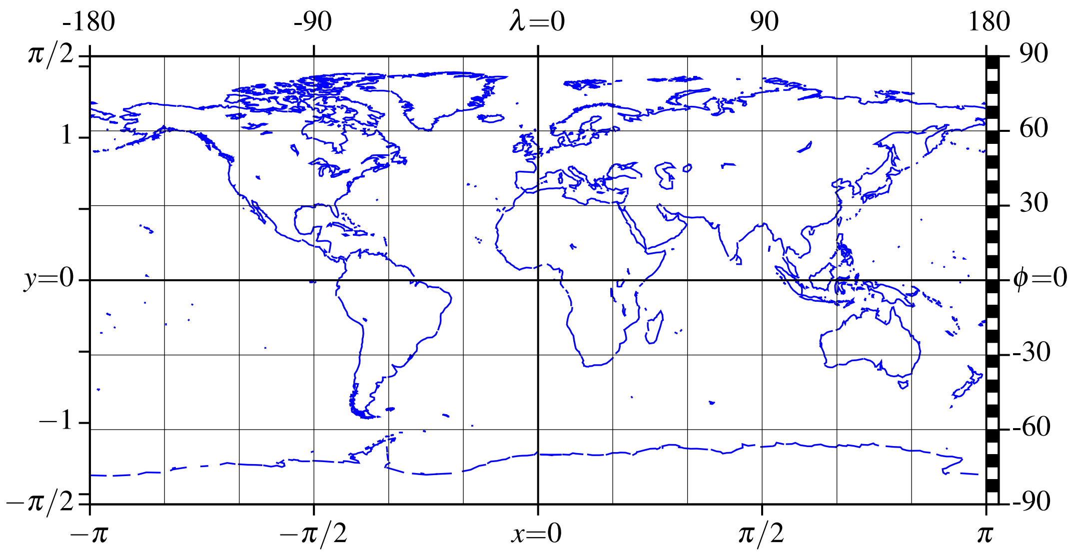


Figure 2.7: Equirectangular projection (R=1)

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

This projection, the simplest of all, has been in use since the time of Ptolemy (83?–161AD) who attributes its first use to Marinus (Snyder, 1993, page 6). The meridians and parallels of latitude are equidistant parallel lines intersecting at right angles. (It becomes a rectangular grid when applied to a secant projection; see Section 2.7). The overall **aspect ratio** (width:height) is 2. The projection is also known as the 'Plate Carrée', the Plane Chart or the equidistant projection. The projection was very important in the sixteenth century because it was used to underpin the new portolans and charts which were being extended to cover ocean sailing as against the local European sailing of earlier times. The projection is still useful for many applications where the metric properties are irrelevant, for example index map sheets for all sorts of topics.

Like all normal cylindrical projections the equirectangular is well behaved near the equator but as we move away from the equator the failings of its simplicity become more pronounced. Note first Equation 2.17 reduces to $\tan\alpha = \cos\phi \tan\beta$ so that $\alpha \neq \beta$ unless $\alpha = \beta = 0$ or $\pi/2$. The projection is certainly not conformal and the attractive compass roses shown on early charts give *incorrect* values for the corresponding azimuths on the sphere. This may be of no concern for short journeys in equatorial regions but it is very significant on long oceanic voyages and at extreme latitudes, even if the sailors of the sixteenth century were aware of this deficiency and had evolved rules of thumb to compensate. Note that the errors are not small: if we take $\beta = 45°$ and a latitude $\phi = 30°$ then $\alpha = \arctan\sqrt{3}/2 = 40.9°$ but at latitude $\phi = 60°$ and $\alpha = \arctan 0.5 = 26.6°$

The alternative name "equidistant" is also misleading because the scale factor is uniform, equal to 1, only on the equator and the meridians where the true distance is equal to the ruler distance divided by the RF. On a parallel we must first divide by a factor of $\sec\phi$. Along any other line in the projection the scale factor depends on both the grid bearing and the latitude: from Equation (2.23) we have $\mu = \sec\phi\sin\alpha(\beta,\phi)\operatorname{cosec}\beta$ where, from Equation (2.17), $\alpha = \arctan[\tan\beta\cos\phi]$. It is then possible to relate elements of length, ds' on the projection and ds on the sphere, and even construct an integral for finite segments but there is basically no point in doing so since the path on the sphere corresponding to this straight line on the projection is neither a rhumb nor a great circle. It can't be a rhumb because the azimuth, α is not constant. It is, moreover, a curve on the sphere which, if we assume that it starts at $\phi = \lambda = 0$, attains the pole. For example if the line on the projection is $\phi = 3\lambda$ it reaches the pole on the sphere when $\lambda = \pi/6$. But the only great circles through the pole are the meridians so the path on the sphere cannot be a great circle. For the true great circle distance between general points on the equirectangular projection we must use the standard geodesic formulae of equation 2.6.

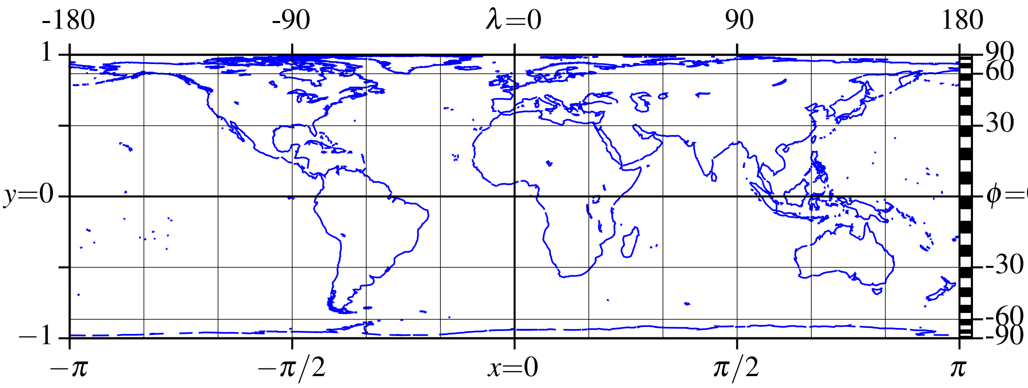


Figure 2.8: Lambert equal area projection (R=1)

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

This projection (Lambert, 1772) is constructed to guarantee the equality of corresponding area elements on the (reduced) sphere and the projection. Since the parallel scale factor is still as $k = \sec\phi$ we must have a scale factor on the meridian given by $h = \cos\phi$. It follows from Equation eq:02nms:20 that $f(\phi) = \sin\phi$. Thus the ratio of the area of Greenland to that of Africa is correctly portrayed (as 1/13). In common with all cylindrical projections it is well behaved with good shapes and distances near the equator but it is distorted at high latitudes: look at look at Alaska for example. The projection is not conformal.

There are problems with angles, exactly as for the equirectangular projection. We now have $\tan\alpha = \cos^2\phi \tan\beta$ so that if we again take $\beta = 45°$ and a latitude $\phi = 30°$ then $\alpha = \arctan 0.75 = 36.9°$ and at latitude $\phi = 60°$ we have $\alpha = \arctan 0.25 = 14.0°$. The grid bearings are completely unreliable.

Once again, interpreting ruler distances on the projection is trivial only on the equator and parallels (where we must first divide by a factor of $\sec\phi$). This time a ruler distance on the meridian has no simple interpretation because of the of the varying scale factor $h = \cos\phi$. However if we measure the ruler distances of two points on a meridian from the equator, say y_1 and y_2, not just their separation, we can then use $y = \sin\phi$ to find the corresponding latitudes, ϕ_1 and ϕ_2: the length $y_2 - y_1$ then corresponds to a distance on the sphere equal to $R(\phi_2 - \phi_1)$. On oblique lines of the projection we have exactly the same difficulties that we encountered with the equirectangular projection.

The projection is one of the few which admits of a geometric interpretation because $y = R\sin\phi$ is simply the distance NP' of a point P at latitude ϕ above the equatorial plane. P is projected from the sphere to cylinder along the line KPP' 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 a classroom globe from a distant side position.

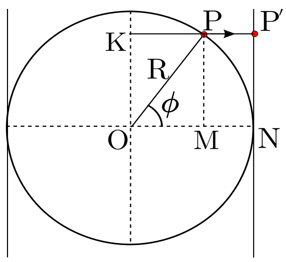


Figure 2.9

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

Figure 2.10 shows Mercator's projection. Since $f(\pi/2) = \infty$ the projection extends to infinity in the y direction and the map must be truncated at an arbitrarily chosen latitude. Here we have chosen $\phi = \pm 85.051°$ so that the aspect ratio is equal to 1. Truncation at these high latitudes emphasizes the great distortion near the poles—witness the diverging area of Antarctica and Greenland as big as Africa. Note that the original (Mercator, 1569) projection was truncated asymmetrically and as a result Europe, already at a larger scale than Africa, moved nearer to the centre: a source of controversy in the twentieth century (Monmonier, 2004).

The fundamental property of the Mercator projection is that it is conformal, *i.e.* it is an angle preserving projection. This follows from equation (2.17) since $f'(\phi) = \sec\phi$ implies that $\alpha = \beta$. One important corollary is that a rhumb line on the sphere, which crosses the converging meridians on the sphere at a constant angle α, projects into a straight line of constant grid bearing β on the projection where the meridians are parallel verticals. This is discussed more fully in Section 2.5.

Conformality implies isotropy of scale: meridian scale (h), parallel scale (k) and general scale (μ_α) are all equal to $\sec\phi$ in the Mercator projection. Therefore a small region of the sphere is projected with very little change of shape, hence the use of the term **orthomorphic** projection, (greek: right shape). Witness the realistic shape of small islands far from the equator, for example Iceland or Great Britain: larger regions such as Greenland or Antarctica are distorted because the scale factor changes markedly over their extent.

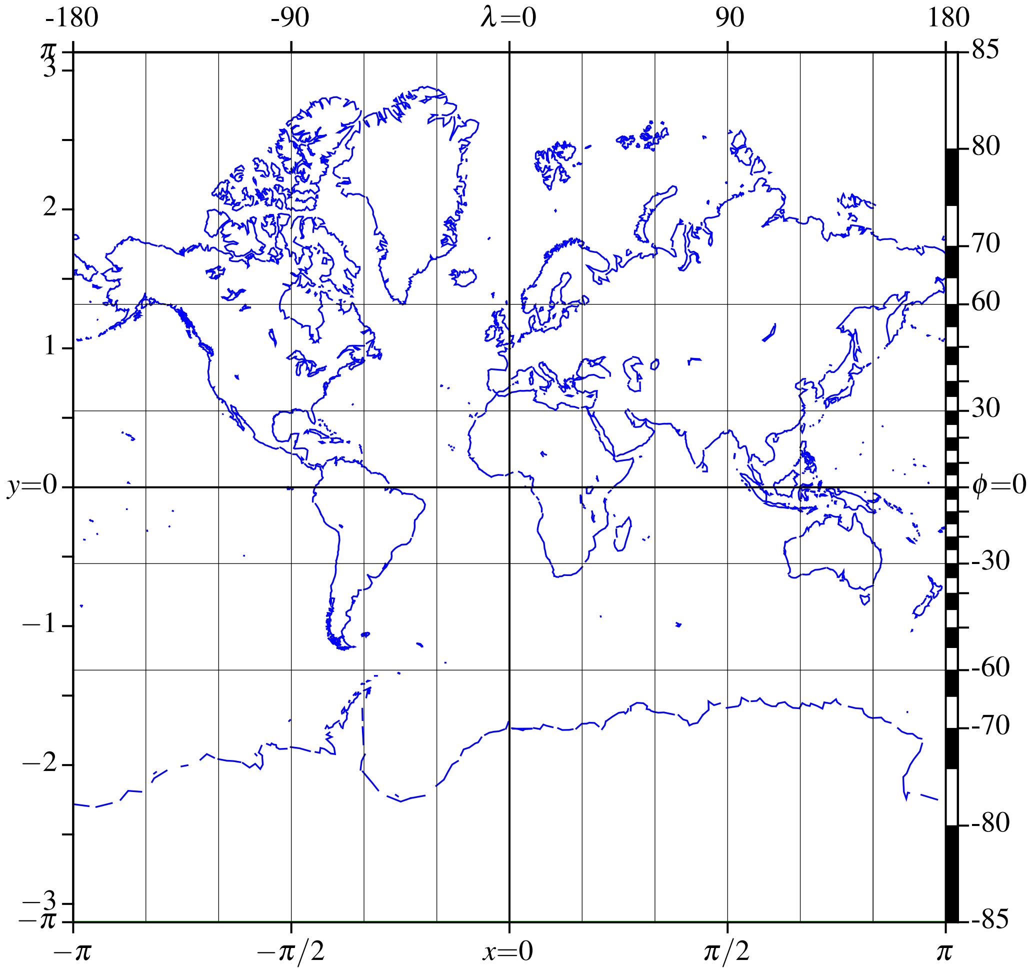


Figure 2.10: Mercator projection truncated at 83.05°(R=1)

The words conformality and orthomorphism are often used interchangeably: this is misleading. Conformality is an *exact* local (point) property: the angle between two lines intersecting at a point on the sphere is the same on the projection. Orthomorphism is an *approximate* non-local property because the shape of small elements is preserved only to the extent that the latitude variation of scale is undetectable: this depends on the accuracy of measurement. Conformality and low scale distortion near the equator, where $\sec\phi$ is approximately unity, means that the Mercator projection *is* suitable for accurate large scale mapping near the equator. This is discussed more quantitatively in Section 2.7. It is quite inappropriate for small scale projections of the world, or large regions of the world, oceanic charts excepted.

Ruler distances on the equator and parallels can be simply related to true distances, as in the previous projections, but there are no simple interpretations for ruler distances of other lines on the projection. Since straight lines on the projection correspond to rhumb lines these distances are important and we shall return to their calculation in Section 2.6.

Central cylindric projection: $f(\phi) = \tan\phi$

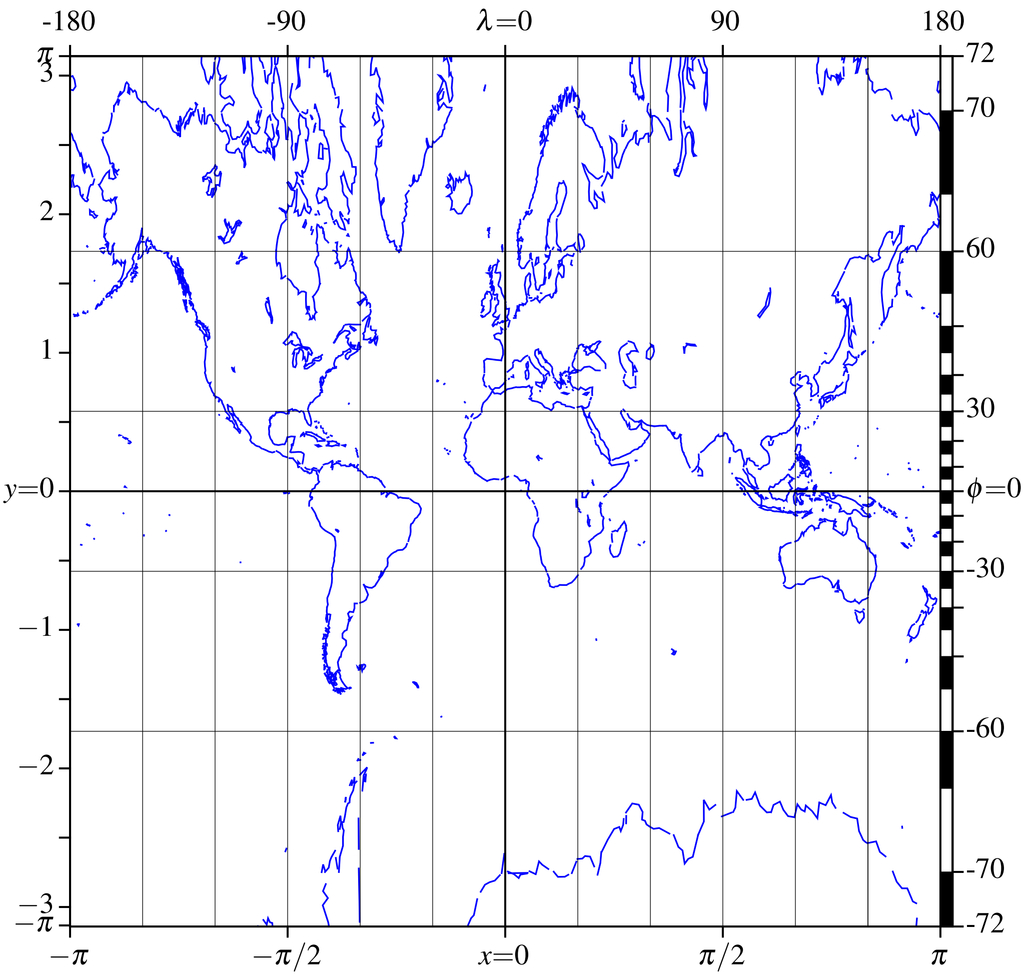


Figure 2.11: Central cylindrical projection truncated at 72°(R=1)

The only reason for including the central cylindric projection, a direct projection from the centre of the sphere with $y = NP' = R\tan\phi$, is that it is often often claimed to show the construction of the Mercator projection. This is of course completely wrong: the Mercator projection is NOT constructed in this way. The projection is shown truncated at 72.34° to give unit aspect ratio. The central projection is completely lacking in any virtues and it has never been used for any practical mapping. We shall not consider it further.

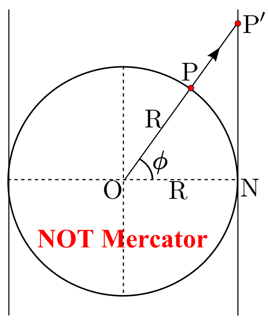


Figure 2.12

2.4 The normal Mercator projection

Derivation of the Mercator projection

The generic function $f(\phi)$ will be replaced by $\psi(\phi)$ for the normal Mercator projection. The condition that the projection is conformal, $\alpha = \beta$, follows from Equation (2.17):

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

and therefore

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

choosing a lower limit such that $y(0) = R\,\psi(0) = 0$. The integrand may be rewritten using

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

so that (note [3])

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

Therefore the Mercator projection in normal (equatorial) aspect is

$$\boxed{x = R\lambda \qquad y = R\psi(\phi) \qquad \psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right]} \tag{2.28}$$

The following table gives the value of $\psi(\phi)$ at some selected latitudes. Note that since $\tan(-\phi/2+\pi/4) = \cot(\phi/2+\pi/4)$ we must have $\psi(-\phi) = -\psi(\phi)$.

ϕ	$\psi(\phi)$	ϕ	$\psi(\phi)$	ϕ	$\psi(\phi)$
0°	0.00	40°	0.76	74.6°	2
5°	0.09	45°	0.88	75°	2.03
10°	0.18	49.6°	1	80°	2.44
15°	0.26	50°	1.01	84.3°	3
20°	0.36	55°	1.15	85°	3.13
25°	0.45	60°	1.32	85.05°	π
30°	0.55	65°	1.51	89°	4.74
35°	0.65	70°	1.73	90°	∞

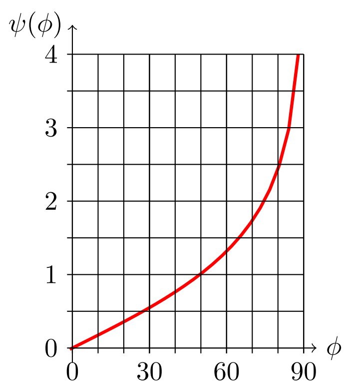


Figure 2.13

Mercator parameter and isometric latitude

The function $\psi(\phi)$ 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**: this is not standard usage although the term **Mercator latitude** was suggested by Lee (1946a). In advanced texts, such Snyder (1987), it is called the **isometric latitude**. But beware; other authors (see Adams, 1921) use the term for a different, but related, function. Note, too, that the symbol ψ is not universal: Lee (1945) and Redfearn (1948) use χ, Maling (1992) use q and so on.

The term isometric latitude arises because the metric can be written in terms of λ and ψ as $ds^2 = R^2\cos^2\phi(d\lambda^2 + d\psi^2)$. The coefficients in the metric are now both equal to $R^2\cos^2\phi$ so that equal increments of λ and ψ correspond to the same linear displacements (at the latitude concerned).

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* 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):

$$\tan(\phi/2+\pi/4) = \frac{1+\tan(\phi/2)}{1-\tan(\phi/2)} = \frac{\cos(\phi/2)+\sin(\phi/2)}{\cos(\phi/2)-\sin(\phi/2)}$$
$$= \frac{(\cos(\phi/2)+\sin(\phi/2))^2}{\cos^2(\phi/2)-\sin^2(\phi/2)} = \frac{1+\sin\phi}{\cos\phi} = \sec\phi+\tan\phi. \tag{2.29}$$

Hence

$$\psi(\phi) = \ln[\sec\phi+\tan\phi]. \tag{2.30}$$

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

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

Exponentiate each side of (2.30) and then invert:

$$e^{\psi} = \sec\phi+\tan\phi,$$
$$e^{-\psi} = \sec\phi-\tan\phi,$$

so that

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

and therefore

$$\text{(a)}\quad \sinh\psi = \tan\phi, \qquad \text{(b)}\quad \operatorname{sech}\psi = \cos\phi, \qquad \text{(c)}\quad \tanh\psi = \sin\phi, \tag{2.32}$$

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

$$\psi = \sinh^{-1}\tan\phi = \operatorname{sech}^{-1}\cos\phi = \cosh^{-1}\sec\phi = \tanh^{-1}\sin\phi. \tag{2.33}$$

Inverse transformations and inverse scale factor

The inverse transformation for λ is trivial: $\lambda = x/R$ (if $\lambda_0 = 0$). To find ϕ first set $\psi = y/R$ and use the inverse of any of the expressions for ψ given above. For example:

$$\phi = 2\tan^{-1}e^{\psi} - \frac{\pi}{2} = \sin^{-1}\tanh\psi = \tan^{-1}\sinh\psi = \operatorname{gd}\psi, \qquad \psi = \frac{y}{R} \tag{2.34}$$

where we have introduced the **gudermannian** function **gd** defined below. (It may be recast into many forms; see appendices at C.59, G.24 and web Weisstein, 2012):

$$\operatorname{gd}(x) = \int_0^x \operatorname{sech}\theta\, d\theta = \tan^{-1}\sinh x, \qquad \operatorname{gd}^{-1}(x) = \int_0^x \sec\theta\, d\theta = \sinh^{-1}\tan x. \tag{2.35}$$

The scale factor can also be considered as a function of the coordinates on the projection. Using Equation (2.32b) we have

$$k(x,y) = \cosh\psi = \cosh(y/a). \tag{2.36}$$

2.5 Rhumb lines and loxodromes

Rhumb lines were first discussed in the sixteenth century by Pedro Nunes as curves of constant azimuth which spiralled from pole to pole. The more academic word **loxodrome** (Greek loxos: oblique + dromos: running) appeared early in the seventeenth century. At that time both of these terms excluded simple parallel or meridian sailing but modern usage includes these cases. This is unfortunate. We shall permit rhumb lines to include all possible directions but we shall restrict the use of loxodrome to azimuths which are neither parallels nor meridians. This is consistent with the definition of the loxodrome in mathematics as a spherical helix: see Weisstein (2012). The distinction is important for there are two topologically distinct classes of rhumbs: (a) closed parallels; (b) open lines, loxodromes, running from pole to pole with meridians as degenerate cases. The importance of rhumb lines follows from the conformality property: $\alpha = \beta$ implies that a rhumb line with constant α is projected to a straight line on the Mercator projection.

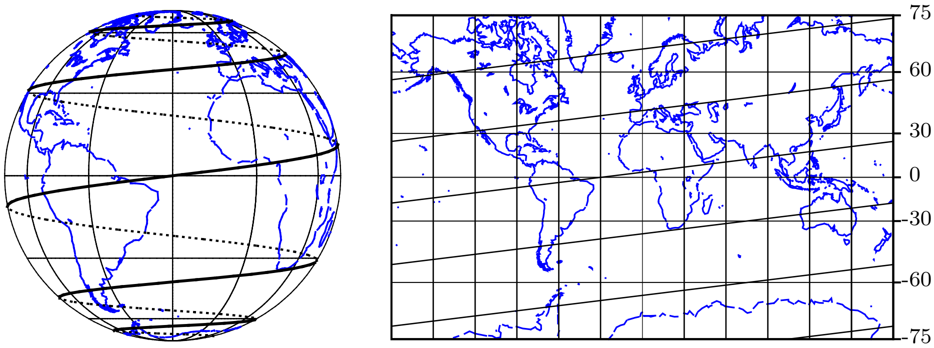


Figure 2.14: Rhumb line on the sphere and the Mercator projection

Figure 2.14 shows a loxodrome crossing the equator at 30°W and maintaining a constant azimuth of 83°: it spirals round the sphere covering a *finite* distance from pole to pole even though it makes an infinite number of turns about the axis. (These statements are proved below). On the projection, it is a repeated straight line of *infinite* total length as $|y| \to \infty$. The intercepts with the Greenwich meridian are calculated later using Equation 2.43.

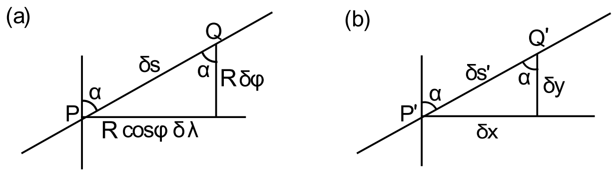


Figure 2.15: Infinitesimal elements of a rhumb on the sphere and the projection

Consider the rhumb distance, r_{12}, on the sphere between $P(\phi_1, \lambda_1)$ and $Q(\phi_2, \lambda_2)$. If the rhumb is a parallel the distance is the radius of the parallel circle times the change of longitude. On a meridian the distance is simply the radius of the sphere times the change of latitude. For an infinitesimal element of a loxodrome on the sphere, PQ in Figure 2.15a we have $\cos\alpha = R\,d\phi/ds$. Since α is a constant this integrates trivially. In summary:

$$r_{12} = R\cos\phi\,(\lambda_2 - \lambda_1), \qquad \text{parallel}, \tag{2.37}$$

$$r_{12} = R(\phi_2 - \phi_1), \qquad \text{meridian}, \tag{2.38}$$

$$r_{12} = R\sec\alpha\,(\phi_2 - \phi_1), \qquad \text{loxodrome}. \tag{2.39}$$

Therefore, to calculate the distance along a loxodrome, we need know only the constant azimuth and the change of latitude. This is an important result. Note (a) the meridian result can be deduced as the $\alpha \to 0$ limit of the loxodrome result (although the figure is inappropriate); (b) the parallel result is not related to the loxodrome result by any limiting process. The latter is a reflection of the different topological nature of parallels and loxodromes.

The above equations show that the length of a loxodrome is finite. Setting $\phi_1 = -\pi/2$ and $\phi_2 = \pi/2$ in Equation 2.39 we obtain the total length from one pole to another as $\pi R\sec\alpha$. This reduces to πR on a meridian.

Equations of the loxodrome

To find the equation, on the sphere, of the loxodrome which starts at the point (ϕ_1, λ_1) at an azimuth α note that on the projection it is the straight line

$$y - y_1 = (x - x_1)\cot\alpha. \tag{2.40}$$

where $y_1 = R\psi(\phi_1)$ and $x_1 = R\lambda_1$. Using Equations 2.33 and 2.34 gives

$$\psi(\phi) = \psi(\phi_1) + (\lambda - \lambda_1)\cot\alpha, \tag{2.41}$$

$$\lambda(\phi) = \lambda_1 + \tan\alpha\left[\tanh^{-1}\sin\phi - \tanh^{-1}\sin\phi_1\right], \tag{2.42}$$

$$\phi(\lambda) = \sin^{-1}\tanh\left[\tanh^{-1}\sin\phi_1 + (\lambda - \lambda_1)\cot\alpha\right]. \tag{2.43}$$

As an example take $\lambda_1 = \phi_1 = -10^\circ$ and $\lambda = \lambda_2 = \phi_2 = 40^\circ$. Transforming λ to radian measure, Equation 2.41 gives

$$\tan\alpha = \frac{\pi}{180}\frac{50}{[\psi(40) - \psi(-10)]} = 0.929$$

and therefore $\alpha = 42.9^\circ$. Note that this is just the *shortest* rhumb line through the two points. If the rhumb makes one complete revolution before getting to the second point, then replacing $\Delta\lambda = 50^\circ$ by $\Delta\lambda = 410^\circ$ we find $\alpha = 82.5^\circ$ and so on.

Once α has been determined further points on the rhumb are found from Equation 2.43. For example, for the loxodrome with $\alpha = 83^\circ$, $\phi_1 = 0$ and $\lambda_1 = -30^\circ$ (Figure 2.14), we calculate the intercepts on the Greenwich meridian ($\lambda = 0,\ 360,\ 720,\ \ldots$) as

3.7°, 43.1°, 67.3°, 79.4°, 85.0°, 87.7°, 88.9°, 89.5°, 89.8° ....

Equation 2.42 shows that λ becomes infinite at the poles so that the loxodrome must encircle the pole an infinite number of times as it approaches, even though it is of finite length. This is a geometrical version of Zeno's paradox.

Mercator sailing

The above results solve the two basic problem of Mercator sailing, by which we mean **loxodromic sailing**; *i.e.* the trivial cases of sailing along parallels or meridians are excluded from the discussion. The two problems are

1. Given a starting point $P(\phi_1, \lambda_1)$ and a destination $Q(\phi_2, \lambda_2)$ find the azimuth α of the loxodrome line and the sailing distance, d. (The inverse problem).

2. Given an initial point $P(\phi_1, \lambda_1)$, a loxodrome of azimuth α and a sailing distance d, find the destination $Q(\phi_2, \lambda_2)$. (The direct problem).

We now outline the solution of these problems

For the inverse problem we are given $\phi_1, \lambda_1, \phi_2, \lambda_2$. First calculate ψ_1, ψ_2 from 2.33 (or simply use a table of meridian parts as described below): α then follows from 2.41 and the distance from 2.39 (in nautical miles if the latitudes are given in minutes of arc). On a chart the angle is measured from the slope of PQ and and the distance, if moderate, can be approximated without calculation by using dividers to transfer the interval PQ to the latitude scale on the chart, taking care to place the dividers symmetrically with respect to the mid-latitude point so that the latitude scale is approximately uniform for that interval. The latitude difference in minutes gives the distance in nautical miles.

For the direct problem we are given $\phi_1, \lambda_1, r_{12}, \alpha$. Equation 2.39 gives $\delta\phi$; hence ϕ_2. Calculate (or use tables for) ψ_1, ψ_2 and then $\delta\lambda$ is found from 2.41; hence λ_2. Again, without calculation, open the dividers to a distance d (given in nautical miles measured as minutes on the latitude scale and transfer the dividers to the line through P with azimuth α, fixing Q; ϕ_2, λ_2 can then be read from the chart.

Meridian parts

At the end of the sixteenth century Edward Wright published *Certaine Errors in Navigation* (Wright, 1599) in which he criticised the Plane Chart, *i.e.* the equirectangular chart. He stressed that over large regions the plane chart was unreliable in respect of both distance and direction. (See Section 2.3). He advocated the use of Mercator's chart and constructed his own version, stressing that he had no assistance in devising his method. He did exploit Mercator's statement that conformality, not that he used the word, is achieved by compensating the stretching of parallels to the same width as the equator by an equal amount of stretching in the meridian direction. He implemented this by dividing the plane chart he divides it into one minute (of latitude) strips and stretches each by a factor of the $\sec\phi$ evaluated at the upper edge of the strip; hence giving a slight overestimate. The amount any point moves up is the sum of the increments of all the strips beneath. His results are summarised in a table of cumulative secants at intervals of one second of arc and beginning

$\sec 1'$	1.000000042
$\sec 1' + \sec 2'$	2.000000211
$\sec 1' + \sec 2' + \sec 3'$	3.000000592
$\sec 1' + \sec 2' + \sec 3' + \sec 4'$	4.000001269

In the first edition of the book he listed only rounded values at intervals of $10'$ but the full table appears in later editions (Monmonier, 2004, Chapter5). We consider only the first table. The cumulative secants, multiplied by a factor of 10, later became called **meridional parts** and the number of such parts in the interval from zero up to an angle ϕ is denoted by

$$\mathrm{MP}(\phi) = 10\sum_{0}^{\phi} \sec\phi_i \qquad \text{at intervals of } 1'. \tag{2.44}$$

N.B. Modern tables and web calculators usually omit this factor of 10.

Some values selected from the published table are shown in the following table: Wright truncated his calculated values for simplicity of use. He also chose the unit of the meridional part (MP) to be one tenth of the length of an equatorial minute of longitude. Therefore if the width of a given chart is W cm. the unit of the meridional part is $W/216000$ cm. An extra line has been interpolated in the table at $85°3'$ corresponding to the latitude at which Wright's figure in *Certaine Errors* was truncated at 108000MP so that the aspect ratio (of the northern hemisphere only) was exactly 2. Note that the first entries, up to $2°30'$, show the increased spacing was undetectable up to that point in the rounded figures of his table.

Lat.	MP	Lat.	MP	Lat.	MP
$10'$	100	$10°$	6030	$70°$	59667
$20'$	200	$20°$	12251	$80°$	83773
$30'$	300	$30°$	18884	$85°3'$	108000
...	...	$40°$	26228	$89°$	163176
$2°30'$	1500	$50°$	34746	$89°50'$	226223
$2°40'$	1601	$60°$	45277	$90°$	∞

Table 2.3

Wright asserted that the MP values gave the correct spacing of the Mercator parallels. This is obvious since his construction is just a numerical integration of $\sec\phi$ replacing Equation 2.27. Of course the calculus hadn't been invented in his day and the 'log-tan' formula was derived only one hundred years later, (although Wright's contemporary and compatriot, Thomas Harriot, seems to have arrived at the formula by his own original method, Lohne (1965)). The numerical integration may be written as

$$\psi(\phi) = \int_0^\phi \sec\phi\, d\phi \approx \sum_0^\phi \sec\phi_i \delta\phi_i \qquad \delta\phi \text{ in radians.} \tag{2.45}$$

An interval of $1'$ corresponds to $\delta\phi = 0.000291\text{rad} = 1/3437.75\text{rad}$. Therefore

$$\text{MP}(\phi) = 10\sum_0^\phi \sec\phi_i = 34377.5\,\psi(\phi). \tag{2.46}$$

A table of meridional parts may be used to solve the Mercator sailing problems by calculation. For example, in the inverse problem the direction of the azimuth may be evaluated from Equation 2.40:

$$\cot\alpha = \frac{\Delta y}{\Delta x} = \frac{\Delta\psi}{\Delta\lambda} = \frac{3437.75\Delta\psi}{\Delta\lambda'} = \frac{\text{MP}(\phi_2) - \text{MP}(\phi_1)}{10(\lambda_2' - \lambda_1')} \tag{2.47}$$

and the distance (in nautical miles) from Equation 2.39: $d = R\sec\alpha\,\Delta\phi'$ follows . In the direct problem $\Delta\phi = (d/R)\cos\alpha$ is known and hence we have ϕ_2. Then λ_2 follows from the last equation.

Wright's tables of rhumbs

In addition to the table of Meridian parts Wright published tables of rhumb line coordinates, meaning by 'rhumb' the seven loxodromes of the first quadrant making angles of $11.25°$, $22.5°$, $33.75°$, $45°$, $56.25°$, $67.50°$, $78.75°$ with the equator at the point of crossing: these angles are the complements of their azimuths. With these values he was able to plot rhumb lines on globes and also map projections other than mercator. (He mentions stereographic projections particularly.)

Without loss of generality we assume the rhumbs cross the equator at the zero of longitude. He observes that if we take a step of $1°$ of longitude (600 MP) along the equator the ordinate of the next point on the first rhumb is $600\tan(11.25°) = 117.8097$ MP. Therefore, since the rhumb is a straight line on the projection, the ordinates of successive points are simple multiples: 0, 117.81, 235.62, 353.43, $\cdots$. He then used the (full) table of meridian parts to invert these MP values to give latitudes at steps of one degree of longitude along the rhumb. Some selected values for the first rhumb are given in the following table:

Long.	Lat.	Long.	Lat.	Long.	Lat.
$0°$	$0°$	$60°$	$11°'50$	$360°$	$58°1'$
$10°$	$1°59'$	$70°$	$13°47'$	$2{\times}360°$	$80°36'$
$20°$	$3°58'$	$80°$	$15°42'$	$3{\times}360°$	$87°18'$
$30°$	$5°57'$	$90°$	$17°37'$	$4{\times}360°$	$89°13'$
$40°$	$7°55'$	$180°$	$33°40'$	$5{\times}360°$	$89°46'$
$50°$	$9°53'$	$270°$	$47°13'$	$6{\times}360°$	$89°59'$

Table 2.4

He gives similar tables for each of the seven rhumbs, all in steps of one degree of longitude from zero up to a value at which the latitude is equal to $89°59'$. For example the table for the fourth rhumb terminates at $540°$ and that for the seventh rhumb at $114°$.

2.6 Distances on rhumbs and great circles

The distinction between rhumb distance and great circle distance was clearly understood by Mercator but he stressed that the rhumb line distance is an acceptable approximation for true great circle distance for courses of short or moderate distance, particularly at lower latitudes. (See Mercator, 1569, Legend 12 on the 1569 map.) He even quantifies his statement: "When the great circle distances which are to be measured in the vicinity of the equator do not exceed 20 degrees of a great circle, or 15 degrees near Spain and France, or 8 and even 10 degrees in northern parts it is convenient to use rhumb line distances". Moreover he explains how the rhumb distance could be calculated directly from his projection, essentially by the method we have outlined in Section 2.5, and he would have known how to derive Equation 2.6 for it is the just the cosine rule of spherical trigonometry: see Appendix D.

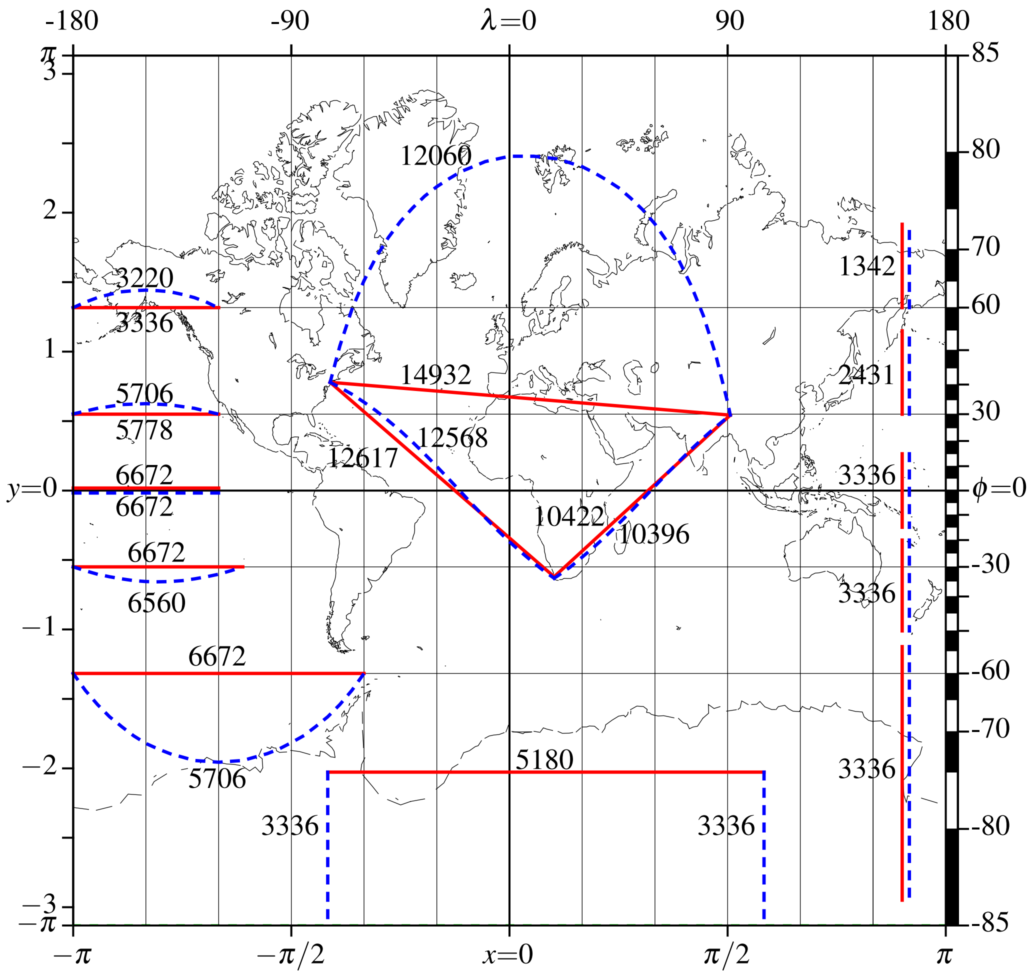


Figure 2.16: Distances on rhumb lines (solid red) and great circles (dashed blue.)

In Figure 2.16 we illustrate the differences between ruler distances d, rhumb line distances r (solid red in the figure) and true great circle distances g (dashed blue in the figure). The great circle distances are always less than or equal to the rhumb line distances.

On the equator

This case is trivial:

$$\frac{d}{\mathrm{RF}} = g = r. \tag{2.48}$$

With radius and great circle circumference equal to 6,371km and 40,030km respectively the 60 degree segment shown has length 40030/6 = 6672km. On a map printed with an RF of 1/300M, for which $R = 2.12$cm and $W = 13.34$cm, the ruler measurement would be 2.22cm.

On other parallels

Parallels are rhumb lines but, apart from the equator, they are not great circles. In the northern (southern) hemisphere the great circle joining two points on a parallel runs north (south) of the parallel. The rhumb distances are calculated from Equation 2.37 and the great circle distances are calculated from Equation 2.6, both using the mean radius of 6371km.

$$\frac{d\cos\phi}{\mathrm{RF}} = r > g = R\left(\cos^2\phi\,\cos\Delta\lambda + \sin^2\phi\right). \tag{2.49}$$

Figure 2.16 shows two segments of parallels above the equator which have the same ruler length as that on the equator. There are two segments below the equator which have the same rhumb lengths as that on the equator.

Note the special case shown at latitude 75°S. The separation in longitude is 180° so the great circle route is the meridian over the pole: it is considerably shorter even though it has infinite distance on un-truncated Mercator projection. The length of the great circle is equal to 40030/12 = 3336km. Note that this distance, and twice its value, appears in several places in the figure.

On meridians

A meridian on the projection is a great circle but the continuous scale variation means that ruler distance is not simply related to true distance. If the projection is marked with an accurate latitude scale, as is usually the case for nautical charts, the meridian distance may be obtained directly as $R\Delta\phi$ (Equation 2.38). Otherwise it is necessary to measure the ruler distances of each point from the equator. If these are y_1 and y_2, with $y_2 > y_1$, first find the latitudes using $y = R\,\mathrm{gd}\,\phi$ and evaluate $R(\phi_2 - \phi_1)$:

$$r = g = R\,\mathrm{gd}^{-1}\left(\frac{y_2}{R}\right) - R\,\mathrm{gd}^{-1}\left(\frac{y_1}{R}\right) = R\tan^{-1}\left[\sinh\left(\frac{y_2}{R}\right)\right] - R\tan^{-1}\left[\sinh\left(\frac{y_1}{R}\right)\right]. \tag{2.50}$$

General distances

The figure shows a large triangle joining the cities of New York (40.7°N, 74.0°W), Cape Town (33.0°S, 18.4°E) and Lhasa (29.7°N, 91.1°E). Once again the ruler distance measurements are not simply related to either rhumb distances or great circle distances which were calculated from Equations 2.39 and 2.6 respectively. Note the great difference for the two routes between New York and Lhasa. Where the rhumb joins two points across the equator it is close to the great circle and the differences are small.

Short distances

Over a small region centred on latitude ϕ, where the scale factor does not vary too much, the ruler distance d corresponds to a true distance of $(d\cos\phi)/RF$ and moreover this is equal to the rhumb distance, r and the great circle distance, g.

2.7 The secant normal Mercator projection

Zones of high accuracy in scale

When the Earth is modelled by a sphere reduced by the representative fraction (RF) and then projected to the plane the scale factor is unity on the equator. Near the equator the Mercator projection is perfectly suitable for high accuracy conformal mapping. We can make this statement more quantitative by demanding that the scale variation must remain within (say) 0.1% of the exact scale on the equator. This value is illustrative only.

With this choice, the scale factor varies from 1 on the equator to 1.001, a value which is attained at latitudes given by $k = \sec\phi = 1.001$, or $\phi = \pm 2.56°$. Therefore the accuracy of the projection is within 0.1% in a zone of width of 5.12°, corresponding to a north south distance of 570km centred on the equator.

It could be argued 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, a band of latitudes starting at 10°N, where $k = 1.0154$. We find that the scale has increased by 0.1% to $k = 1.0164$ when we reach 10.32°N so that the width of the zone of accuracy starting at 10°N is only 19′ (or 35km) . Clearly such narrow zones are unsuitable for accurate mapping.

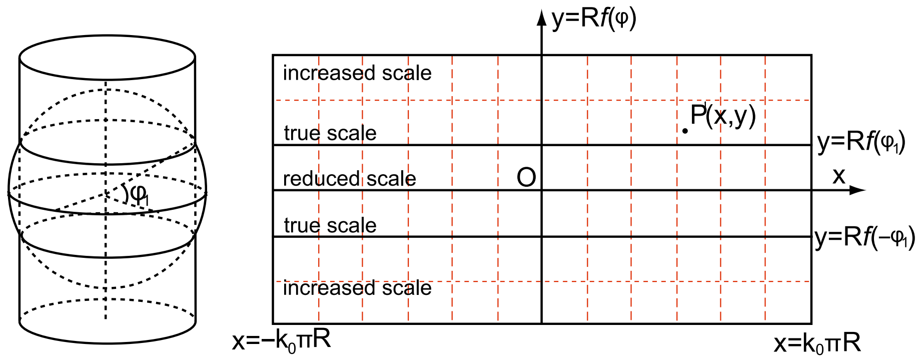


Figure 2.17: The modified normal cylindrical projection

Secant projections

Secant (or 'modified') projections provide a means of extending the zone of accuracy of a map projection. The basic idea is that the scale factor is allowed to be less than 1 and it is only the modulus of the difference between the scale factor and unity that is confined to a given range. This is achieved for cylindrical projections by projecting the sphere (according to a defined mathematical function, not just literally) to a cylinder which cuts the sphere in two parallels at $\pm\phi_1$. Hence the terminology 'secant' meaning cutting. The parallels at $\pm\phi_1$ are called the **standard parallels** of the secant projection.

Clearly the points on the parallel at ϕ_1 lie on both the sphere and on the cylinder and therefore the scale factor at that latitude must be $k = 1$ and the width of the projection must be $2\pi\cos\phi_1$. Parallels on the projection between latitudes $\pm\phi_1$ must be contracted and outside that interval they must be stretched. In general the scale factor on a parallel is

$$k(\phi) = \frac{\text{projected length}}{\text{true length}} = \frac{2\pi\cos\phi_1}{2\pi\cos\phi} = \cos\phi_1\,\sec\phi. \tag{2.51}$$

The usual parallel scale factor on the normal cylindrical, $\sec\phi$, follows from Equation 2.19 using $x = R\lambda$. To obtain the modified scale factor we must augment the x-transformation equation by a factor k_0 equal to $\cos\phi_1$. To preserve conformality we must also multiply the y-transformation equation by the same factor (so that the ratio of $\delta x/\delta y$ in Equation 2.16 is unchanged). The equations of the secant equatorial Mercator projection are therefore:

$$x = k_0 R\lambda, \quad y = k_0 R \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right], \qquad k_0 = \cos\phi_1, \quad k(\phi) = k_0\sec\phi. \tag{2.52}$$

If, as before, we allow a 0.1% variation in the scale factor we have $k(0) = k_0 = 0.999$ on the equator and therefore $\phi_1 = 2.56°$. The scale factor will then be equal to its maximum permitted value at a latitude ϕ_2 such that $1.001 = 0.999\cos\phi_2$: this gives $\phi_2 = 3.62°$ and the width of the zone of high accuracy is now 7.24 degrees corresponding to 805km. This is to be compared with the previous zone width of 570km. The secant projection has achieved a 50% greater zone at the same accuracy.

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%.

For the highly accurate large scale maps (of narrow zones) of the value of k_0 is taken as 0.9996. (Web links: UTM, OSGB). See also UTM (1989) and OSGB (1999).

Other secant projections

All cylindrical projections have a secant form. There are two standard parallels at $\pm\phi_1$ on which parallel scale is unity and the x-transformation equation is as above for Mercator with $k_0 = \cos\phi_1$. The y-transformation equation depends on the projection.

For the secant **equirectangular** projection the equations are

$$x = k_0 R\lambda, \qquad y = R\phi, \qquad k(\phi) = k_0\sec\phi, \qquad h = 1. \tag{2.53}$$

The very earliest (Greek) example of this projection used a value of $k_0 = 0.81$ corresponding to the latitude of Rhodes at $\phi_1 = 36°N$. For this value the spacing of the meridians is about 4/5 that of the parallels so that the graticule takes on a rectangular appearance (Snyder, 1993).

For the secant **equal area** projection the equations are

$$x = k_0 R\lambda, \qquad y = k_0^{-1} R\sin\phi, \qquad k(\phi) = k_0\sec\phi, \qquad h = k_0^{-1}\cos\phi. \tag{2.54}$$

The product fo the scale factors remains unchanged at unity. With standard parallels at $\phi_1 = \pm 45°$ the value of k_0 is $/\sqrt{2} = 0.707$ so that the equator is contracted by 30%. This form of the equal area projection was presented by Gall in 1855 and later republished by Peters in 1973. The distortion of shape in the Gall-Peters projection is well known: central Africa becomes much too narrow compared with its Mediterranean coastline.

2.8 Summary of NMS

Mercator parameter

$$\begin{aligned}
\psi(\phi) &= \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right] = \frac{1}{2}\ln\left[\frac{1+\sin\phi}{1-\sin\phi}\right], \\
&= \ln(\sec\phi+\tan\phi) = \sinh^{-1}\tan\phi\,, \\
&= \operatorname{sech}^{-1}\cos\phi = \cosh^{-1}\sec\phi\,, \\
&= \tanh^{-1}\sin\phi = \operatorname{gd}^{-1}(\phi)\,. \qquad (2.55)
\end{aligned}$$

Inverse parameter

$$\begin{aligned}
\phi(\psi) &= 2\tan^{-1}\exp\psi-\frac{\pi}{2} = \sin^{-1}\tanh\psi, \\
&= \tan^{-1}\sinh\psi = \cos^{-1}\operatorname{sech}\psi\,, \\
&= \sec^{-1}\cosh\psi = \operatorname{gd}\psi\,. \qquad (2.56)
\end{aligned}$$

Direct projection

$$x(\lambda,\phi) = k_0R(\lambda-\lambda_0), \qquad (2.57)$$

$$y(\lambda,\phi) = k_0R\psi(\phi), \qquad (2.58)$$

Inverse projection

$$\lambda(x,y) = \lambda_0+\frac{x}{k_0R}, \qquad (2.59)$$

$$\phi(x,y) = \operatorname{gd}\psi(y), \qquad \psi(y)=\frac{y}{k_0R} \qquad (2.60)$$

Scale factors

$$k(\lambda,\phi) = k_0\sec\phi \qquad (2.61)$$

$$k(x,y) = k_0\cosh\psi(y), \qquad \psi(y)=\frac{y}{k_0R} \qquad (2.62)$$

Notes

[1] The angle $2\varepsilon \approx \sin\phi\,\delta\lambda$ in Figure 2.3 may be taken as a measure of the **spherical convergence** of the meridians on the sphere; it must not be confused with the projection (grid) convergence defined in Section 3.6. There is no accepted definition of spherical convergence (Lee, 1946b). That it *is* a suitable measure of meridian convergence on the sphere may be seen from the fact that it sensibly interpolates between zero on the equator, where meridians are parallel, and the value $\delta\lambda$ at the pole, the angle at which the meridians *PK* and *MQ* intersect there.

[2] 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 $x(\lambda,\phi)$.

[3] Note that 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 ∞.

Blank page. A contradiction.

Chapter 3

Transverse Mercator on the sphere: TMS

Abstract

TMS transformations from NMS by rotation of the graticule. Four global TMS projections. 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. Secant 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 rhumb lines 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 very accurate within a narrow strip of about three degrees centred on the equator (extending to five degrees for the secant 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 Lambert (1772).

The crucial point is that if we have a projection which is very accurate close to one meridian then a set of such projections will provide accurate coverage of the whole sphere.

The secant versions of the transverse Mercator projection on the ellipsoid (TME), are of great importance. One such projection may be used for map projections of countries which have a predominantly north-south orientation, for example the Ordnance Survey of Great Britain; see also OSGB (1999). The Universal Transverse Mercator set of projections cover the entire sphere (between the latitudes of 84°N and 80°S) using 60 zones of width 6° in longitude centred on meridians at 3°, 9°, 15°, . . . (UTM, 1989).

In unmodified NMS the equator has unit scale because we project onto a cylinder tangential to the sphere at the equator, (Figure 2.3) 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 construct the functions $x(\lambda,\phi)$ and $y(\lambda,\phi)$ such that the projection is also conformal.

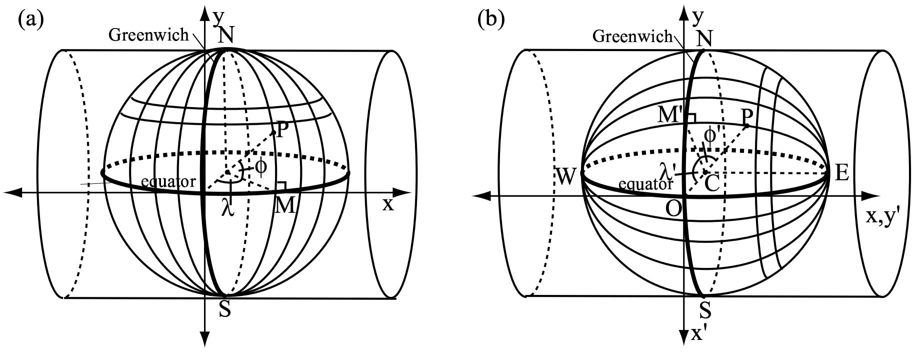


Figure 3.1

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 meridian 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 λ' is measured positive from the origin O towards M': this 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 3.1a. Therefore, bearing in mind the sense of λ', the equations (2.28) for NMS with respect to the rotated graticule are

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

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:

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

It is more useful to adopt one of the alternative forms of the Mercator parameter given in Equation (2.55). (Each would give a different expression for our final result.) We choose

$$x = R\tanh^{-1}\sin\phi', \qquad y = R\lambda'. \tag{3.3}$$

All that remains is to derive the relation between (ϕ',λ') and (λ,ϕ) by 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) which is a "meridian" of the rotated graticule. 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)

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

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

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

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

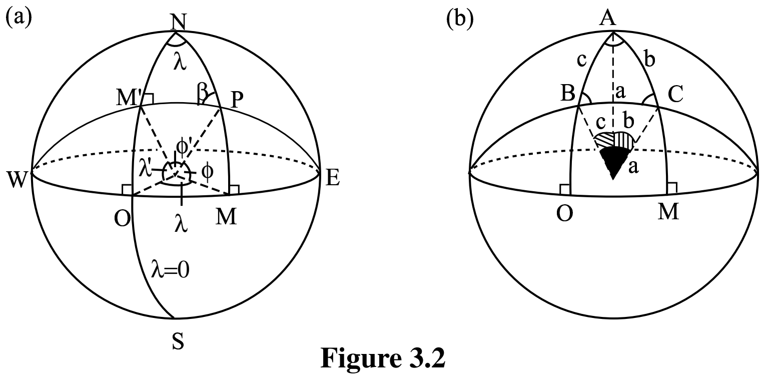


Figure 3.2

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

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

Our choice (3.3) gives the equations for TMS centred on the Greenwich meridian as

$$\boxed{\begin{aligned} x(\lambda,\phi) &= R\tanh^{-1}[\sin\lambda\cos\phi] \\ y(\lambda,\phi) &= R\tan^{-1}[\sec\lambda\tan\phi] \end{aligned}} \tag{3.13}$$

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

Alternative notation

Using Equation 2.31 instead of Equation 2.33 we can replace Equation 3.3 by

$$x = \frac{R}{2} \ln \left[\frac{1 + \sin\phi'}{1 - \sin\phi'} \right], \qquad y = R\lambda'. \tag{3.14}$$

The corresponding equations are:

$$\begin{aligned} x(\lambda, \phi) &= \frac{R}{2} \ln \left[\frac{1 + \sin\lambda \cos\phi}{1 - \sin\lambda \cos\phi} \right] \\ y(\lambda, \phi) &= R \tan^{-1} \left[\sec\lambda \tan\phi \right] \end{aligned} \tag{3.15}$$

Central meridian

On the central meridian, where λ=0, the TMS equations become x=0 and y=$R\phi$ so the scale on the central meridian is constant and the range of y is $[-\pi R/2, \pi R/2]$.

Equator

On the equator, where ϕ=0, we have y=0 but the equation for x diverges to infinity as $\lambda \to \pm\pi/2$. Thus it would appear that only the front hemisphere, between longitudes $-\pi/2$ and $\pi/2$ (or $-90°$ to $90°$), can be mapped by the projection.

Extended transformation

The rear hemisphere, remote from the central meridian, can be mapped in the same projection by modifying the equation for y noting that the inverse tangent is arbitrary to within an additive factor of $N\pi$ (integer N). If we restrict the value of $\tan^{-1}[\sec\lambda \tan\phi]$ to the interval $[-\pi/2, \pi/2]$ then we can set

$$y = \left\{ \begin{array}{c} \pi R \\ 0 \\ -\pi R \end{array} \right\} + R \tan^{-1} \left[\sec\lambda \tan\phi \right] \quad \text{for} \left\{ \begin{array}{ll} |\lambda| > \pi/2, & 0 \le \phi \le \pi/2 \\ |\lambda| \le \pi/2, & |\phi| \le \pi/2 \\ |\lambda| > \pi/2, & 0 \ge \phi \ge -\pi/2 \end{array} \right. . \tag{3.16}$$

For example, where $|\lambda| > \pi/2$ and $\phi > 0$ we have $\sec\lambda \tan\phi < 0$ and as ϕ increases from 0 to $\pi/2$ the value of y decreases from πR to $\pi R/2$. Therefore the rear northern hemisphere appears inverted above the north pole. Similarly the rear southern hemisphere appears inverted below the south pole. This is how Figures 3.3–3.6, were constructed. Note that the equator now includes the top and bottom edges of the projection ($y = \pm\pi R$); latitude increases from these lines towards the poles. Note that above and below the poles longitudes increase from right to left. The meridians at $\pm\pi/2$ ($\pm 90°$) from the central meridian project to lines $\pi R/2$.

The inverse transformations

Invert Equations 3.13:

$$\sin\lambda\cos\phi = \tanh(x/R), \tag{3.17}$$

$$\sec\lambda\tan\phi = \tan(y/R). \tag{3.18}$$

Eliminating ϕ gives

$$\sec^2\phi = \sin^2\lambda\coth^2(x/R) = 1+\cos^2\lambda\tan^2(y/R),$$

$$\tan^2\lambda\left(\coth^2(x/R)-1\right) = \sec^2(y/R),$$

$$\tan\lambda = \sinh(x/R)\sec(y/R), \tag{3.19}$$

thus determining λ as a function of x and y. To find ϕ as a function of x and y first multiply equations (3.17) and (3.18) to give

$$\tan\lambda\sin\phi = \tanh(x/R)\tan(y/R). \tag{3.20}$$

Eliminating $\tan\lambda$ from the last two equations gives

$$\sin\phi = \operatorname{sech}(x/R)\sin(y/R). \tag{3.21}$$

Thus the the inverse transformations are

$$\lambda(x,y) = \tan^{-1}\left[\sinh(x/R)\sec(y/R)\right], \tag{3.22}$$

$$\phi(x,y) = \sin^{-1}\left[\operatorname{sech}(x/R)\sin(y/R)\right]. \tag{3.23}$$

If $|y|\le\pi/2$ the inverse sine and tangent are restricted to the interval $[-\pi/2,\pi/2]$ thus giving points on the front of the sphere. If $|y|\ge\pi/2$ the inverse tangent must be taken outwith the interval $[-\pi/2,\pi/2]$ to give points on the rear of the sphere.

3.2 Features of the TMS projection

Figures 3.3–3.6, show the TMS projections centred on (a) Greenwich (λ_0=0), (b) Africa (λ_0=21°), (c) the Americas ($\lambda_0=-87°$); (c) Australasia and Japan (λ_0=135°). The axes of these projections are labelled in units of Earth radius, equivalent to setting R=1. They have been truncated at $x=\pm\pi$ to give a unit aspect ratio. The figures are overlain by a 15° graticule (relative to the equator and central meridian) and annotated with (latitude,longitude) coordinates of selected interior intersections and boundary points. The Greenwich meridian and its great circle continuation is shown in green (Figures 3.4–3.6 only).

These slightly bizarre (x-truncated) TMS projections covering most the Earth are not of practical use. On the other hand, when these projections are generalised to the ellipsoid (TME), the small regions within the narrow (red) rectangles on the central meridian (Figures 3.4–3.6 only) are the areas covered by the highly accurate large scale projections for UTM zones 34, 16 and 53. In general the central part of the projection is suitable for large scale maps of predominantly north-south land masses, such as Great Britain.

The following table compares and contrasts the features of NMS and TMS.

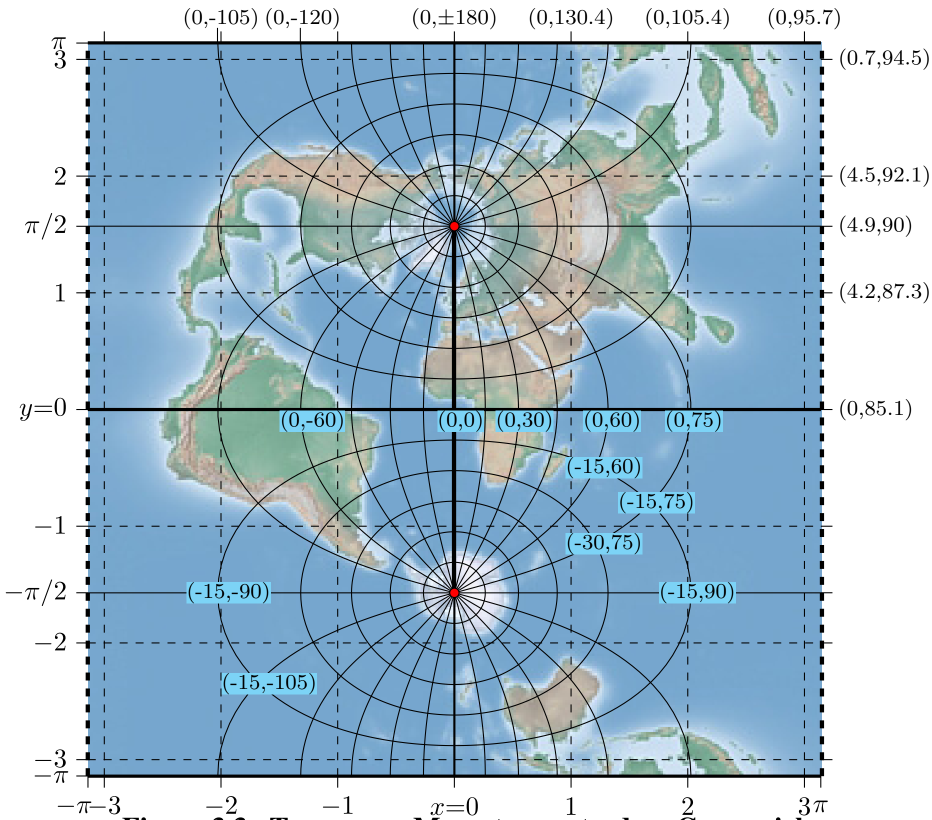


Figure 3.3 : Transverse Mercator centred on Greenwich

NMS	TMS
• Meridians project to lines of constant x. Central meridian is x=0.	• Central meridian is x=0. Meridians at ±90° are lines of constant y through poles. Other meridians project to complicated curves.
• The equator projects to the straight line y=0.	• The equator projects to three straight lines: y=0 and the top and bottom edges. At top and bottom longitude increases to the left.
• Parallel circles project to straight lines of constant y.	• Parallel circles project to closed curves around poles.
• The poles project to lines, $y = \pm\pi$ for R=1.	• The poles project to points.
• Meridians and parallels intersect at right angles.	• Meridians and parallels intersect at right angles.

(continued)

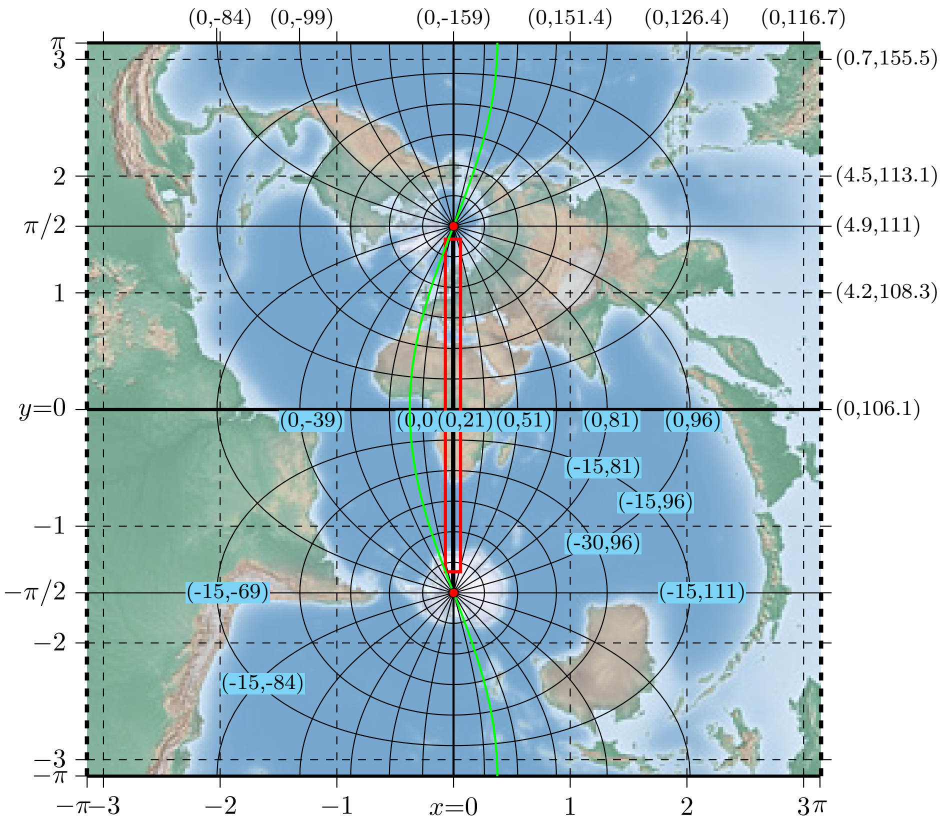


Figure 3.4 : Transverse Mercator centred on 21E

NMS	TMS
• The projection is unbounded in y. The poles lie at infinity. Figure 2.8 is truncated at latitude $\pm 85°$ to give unit aspect ratio.	• The projection is unbounded in x. The points on the equator at $\lambda_0 \pm 90°$ are projected to infinity. The TMS projections are truncated at the x-value corresponding to longitudes $\lambda_0 \pm 85°$ on the equator.
• The projection is conformal. The shapes of small elements are well preserved.	• The projection is conformal. The shapes of small elements are well preserved.
• Distortion increases with y. The projection is not suited for world maps.	• Distortion increases with x. The projection is not suited for world maps.
• Distortion is small near the equator: the projection is suitable for accurate mapping of equatorial regions	• Distortion is small near the central meridian: the projection is suitable for accurate mapping near the central meridian.

(continued)

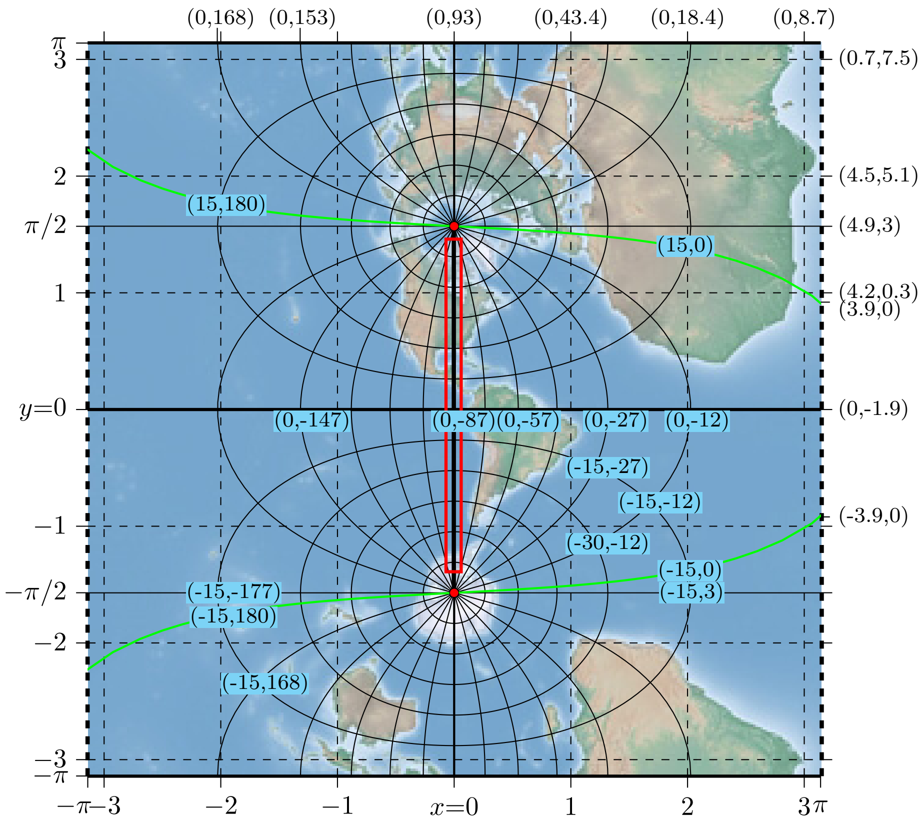


Figure 3.5 : Transverse Mercator centred on 87W

NMS	TMS
• The point scale factor is independent of direction (isotropic). It is a function of *y* on the projection so that it depends on latitude only. The scale is true on the equator.	• The point scale factor is independent of direction (isotropic). It is a function of *x* on the projection but depends on both latitude and longitude. The scale is true on the central meridian.
• The projection is reasonably accurate near the equator. The scale factorincreses by 1% (10%) at latitudes of 8° (25°).	• The projection is reasonably accurate near the central meridian. The scale factor increases by 1% (10%) at when x=0.14R, (0.44R). A strip x<const. is not readily related to regions defined in terms of geographical coordinates.
• In the secant version the scale is reduced on the equator and it is true on two 'standard' parallels.	• In the secant version the scale is reduced on the central meridian: it is true on two lines parallel to the central meridian on the projection.

(continued)

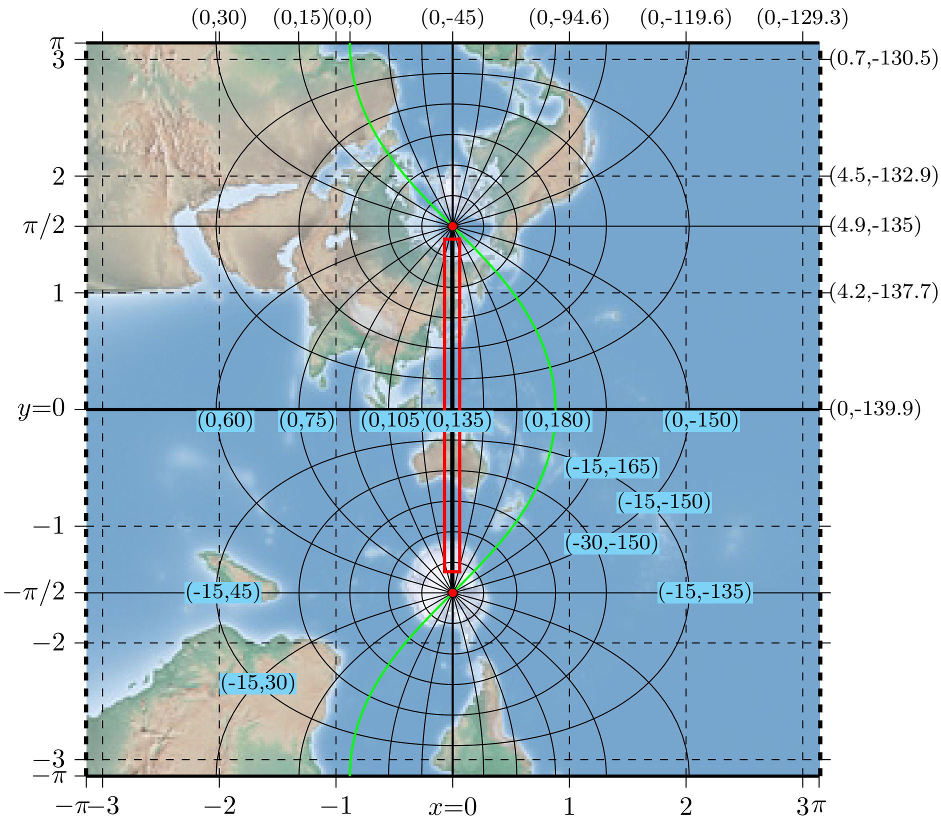


Figure 3.6 : Transverse Mercator centred on 135E

NMS	TMS
• Convergence (the angle between projected meridians and grid lines) is identically zero. Grid north and true north coincide.	• Convergence is zero on the equator and increases towards the poles. Grid north and true north do not coincide.
• Loxodromes running from pole to pole on the sphere project to straight lines.	• Loxodromes running from pole to pole on the sphere cannot project to straight lines.
• Area is not conserved but is approximately so near the equator. Greenland is much larger than Australia: it should be about one third of the size.	• Area is not conserved but is approximately so near the central meridian. For λ=135°E the shapes and relative sizes of Greenland and Australia are both good. Areas far from the central meridian are enlarged, *e.g.* Ethiopia.

Comment The following pages contain versions of Figures 3.3–3.6 which are suitable for printing.

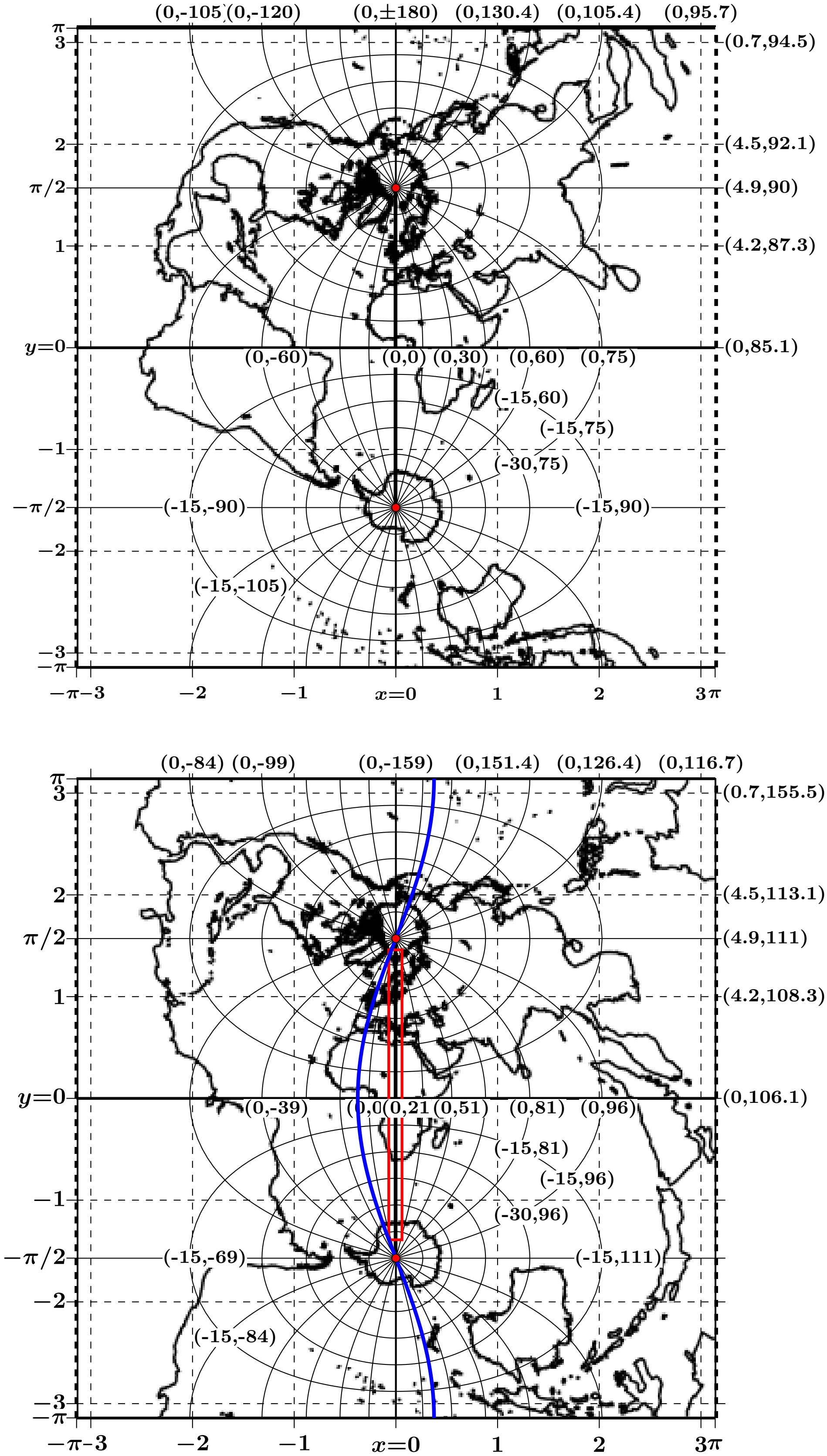


Figure 3.7 : Transverse Mercator centred on Greenwich (top) and 21E (bottom)

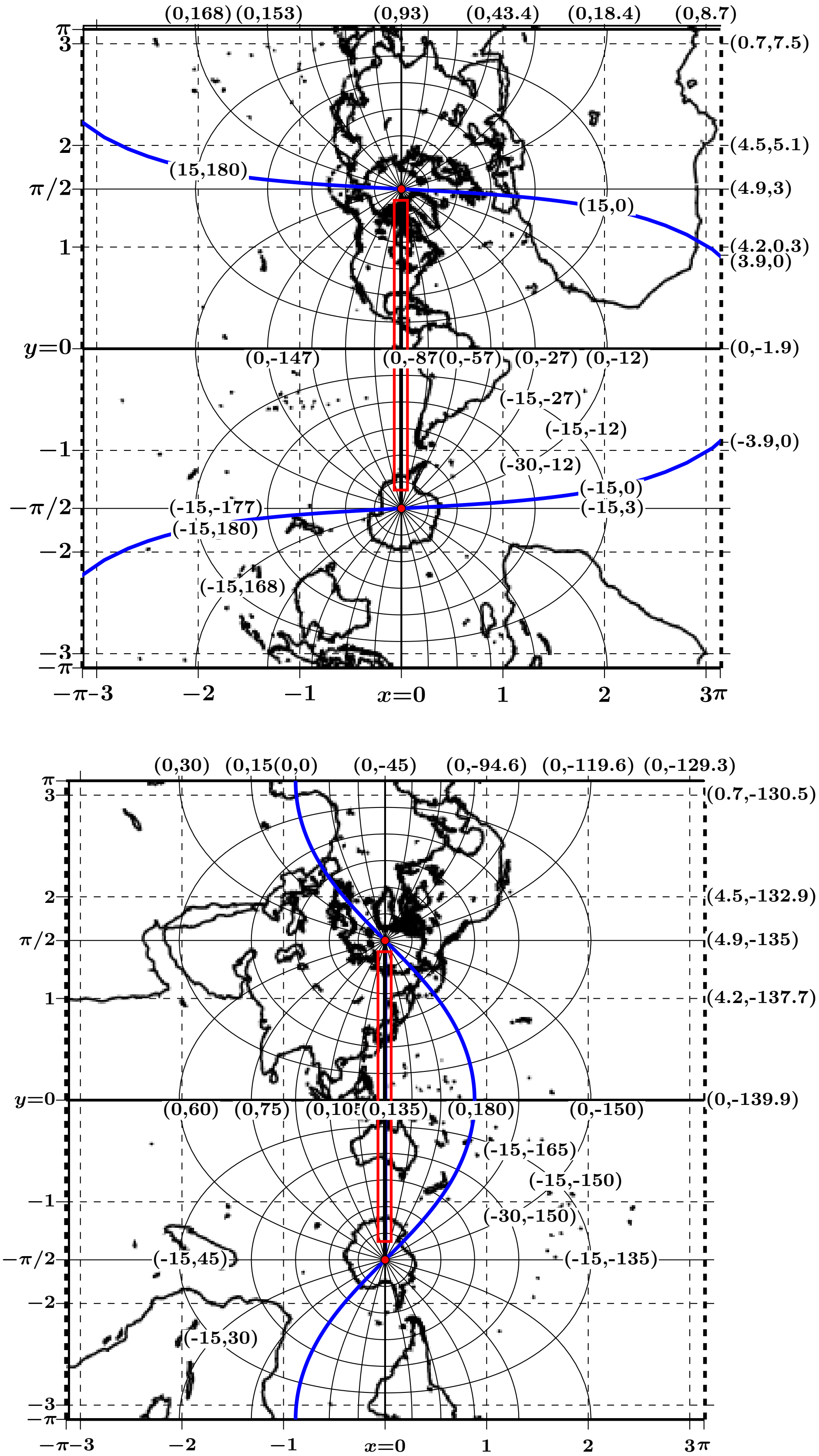


Figure 3.8 : Transverse Mercator centred on 87W (top) and 135E (bottom)

3.3 Meridian distance, footpoint and footpoint latitude

We introduce three definitions which, whilst both trivial and superfluous for the sphere, become very important when we study the transformations on the ellipsoid (TME). Let $P(\lambda,\phi)$ be a general point on the sphere which projects to $P'(x,y)$.

- The **meridian distance**, $m(\phi)$, is the distance measured along a meridian from the equator to a point at latitude ϕ:

$$m(\phi) = R\phi. \tag{3.24}$$

  On the central meridian of TMS we have $y(0,\phi) = m(\phi)$.

- The **footpoint** associated with any point $P'(x,y)$ 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 the point P_1 on the central meridian of the *sphere* which projects into the footpoint $P_1'(0,y)$. In general this latitude is *not* equal to that at the point $P(\lambda,\phi)$ which is the inverse of $P'(x,y)$.

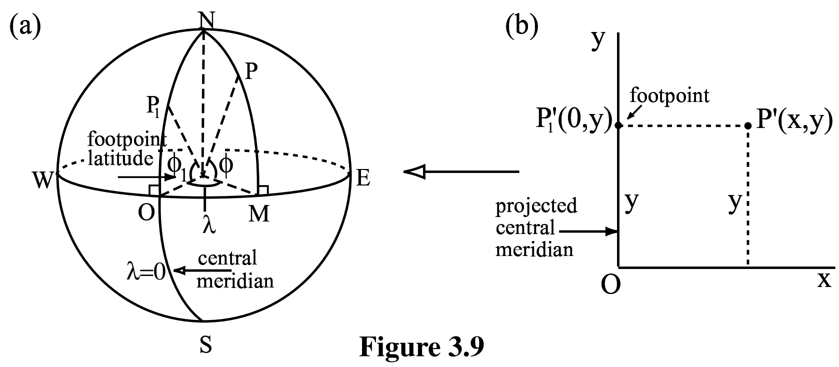


Figure 3.9

From these definitions and equation (3.23) we have

$$\phi_1 = \phi(0,y) = \sin^{-1}[\sin(y/R)] = y/R. \tag{3.25}$$

This is obvious because, by construction, the scale of the projection is true on the central meridian so that $y = R\phi_1$ and hence $\phi_1 = y/R$. Equation (3.24) implies

$$m(\phi_1) = y. \tag{3.26}$$

We now take this equation as a definition of the footpoint latitude which will continue to hold on the ellipsoid where $m(\phi)$ is a non-trivial function. For future reference we write equations (3.22) and (3.23) in terms of the footpoint latitude as

$$\lambda(x,y) = \tan^{-1}[\sinh(x/R)\sec\phi_1], \tag{3.27}$$

$$\phi(x,y) = \sin^{-1}[\operatorname{sech}(x/R)\sin\phi_1] \qquad m(\phi_1) = y. \tag{3.28}$$

The reason for stressing the role of the footpoint is that the series solutions for the inverse transformation of TME *must* be expressed as Taylor series about the footpoint latitude where $m(\phi)$ is a non-trivial relation, unlike the simple $m(\phi)$=$R\phi$ in the present case.

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 (3.17) we find the scale factor in terms of either geographical or projection coordinates:

$$k(\lambda,\phi) = \frac{1}{(1-\sin^2\lambda\cos^2\phi)^{1/2}} \tag{3.29}$$

$$k(x,y) = \cosh(x/R). \tag{3.30}$$

Scale variation is complicated in terms of geographical coordinates but is very simple in terms of projection coordinates. For example $\cosh 0.14 = 1.01$ so that within the strip $x < 0.14R$ ($x < 0.44R$) the scale variation is less than 1% (10%).

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 infinitesimal elements shown in Figure 3.10. 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 projection. The relevant angles must be defined with respect to the tangents of these lines at P'. The angles of interest are

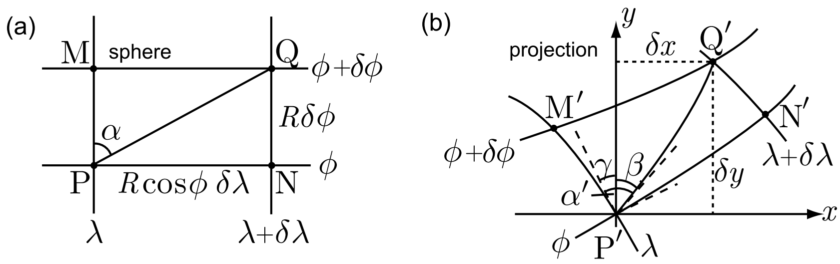


Figure 3.10

- α, the **azimuth** at P between the meridian PM and a displacement PQ on the sphere.
- α', the angle at P' between the projected meridian $P'M'$ and the displacement $P'Q'$.
- β, the **grid bearing** at P' between the projected displacement $P'Q'$ and the y-axis.
- γ, the (grid) **convergence** between the projected meridian and the y-axis.
- Clearly $\alpha' = \beta + \gamma$.

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

$$\boxed{\alpha = \beta + \gamma} \tag{3.31}$$

or, in words:

AZIMUTH = GRID BEARING +CONVERGENCE

This equation is to be used in both directions. If we are given an azimuth α at $P(\lambda,\phi)$ on the sphere then the corresponding bearing at P' on the map projection can be calculated from $\beta = \alpha - \gamma(\lambda,\phi)$. Likewise, given a bearing at a point $P'(x,y)$ on the projection we find the azimuth at the corresponding point at P 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 (generalised to the ellipsoid) 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° and in a UTM zone it is never greater than 3%.

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.

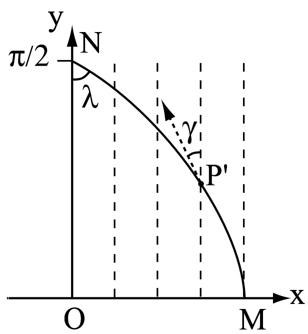


Figure 3.11

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

$$\tan\gamma = -\left.\frac{dx}{dy}\right|_{P'} \tag{3.32}$$

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).

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

$$\delta x = \left(\frac{\partial x}{\partial\lambda}\right)\delta\lambda + \left(\frac{\partial x}{\partial\phi}\right)\delta\phi = x_\lambda\delta\lambda + x_\phi\delta\phi, \tag{3.33}$$

$$\delta y = \left(\frac{\partial y}{\partial\lambda}\right)\delta\lambda + \left(\frac{\partial y}{\partial\phi}\right)\delta\phi = y_\lambda\delta\lambda + y_\phi\delta\phi, \tag{3.34}$$

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

$$\tan\gamma = -\lim\left.\frac{\delta x}{\delta y}\right|_{\delta\lambda=0} = -\frac{x_\phi}{y_\phi}, \tag{3.35}$$

$$\boxed{\gamma = -\tan^{-1}\left(\frac{x_\phi}{y_\phi}\right)}. \tag{3.36}$$

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

$$\sin\phi' = \sin\lambda\cos\phi, \tag{3.37}$$

$$\begin{aligned}\cos^2\phi' &= 1-\sin^2\lambda\cos^2\phi = \sin^2\phi + \cos^2\lambda\cos^2\phi\\ &= \cos^2\lambda + \sin^2\phi\sin^2\lambda = \cos^2\lambda\cos^2\phi\,(1+\sec^2\lambda\tan^2\phi).\end{aligned} \tag{3.38}$$

Therefore

$$x = \frac{R}{2}\tanh^{-1}[\sin\lambda\cos\phi] \qquad y = R\tan^{-1}[\sec\lambda\tan\phi], \tag{3.39}$$

$$x_\lambda = R\sec^2\phi'\cos\lambda\cos\phi, \qquad y_\lambda = R\sec^2\phi'\sin\lambda\sin\phi\cos\phi, \tag{3.40}$$

$$x_\phi = -R\sec^2\phi'\sin\lambda\sin\phi, \qquad y_\phi = R\sec^2\phi'\cos\lambda. \tag{3.41}$$

The convergence as a function of geographic coordinates follows from equation (3.36):

$$\gamma(\lambda,\phi) = \tan^{-1}(\tan\lambda\sin\phi). \tag{3.42}$$

This result can be written in terms of x and y by using equation (3.20) giving

$$\gamma(x,y) = \tan^{-1}[\tanh(x/R)\tan(y/R)] \tag{3.43}$$

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

$$\gamma(x,y) = \tan^{-1}\left[\tanh(x/R)\tan\phi_1\right], \qquad m(\phi_1) = y. \tag{3.44}$$

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 is conformal. To this end consider Figure 3.10 where the azimuth angle of the displacement PQ on the sphere is given by

$$\tan\alpha = \lim_{Q\to P} \frac{\cos\phi\,\delta\lambda}{\delta\phi}. \tag{3.45}$$

Consider the grid bearing of the corresponding displacement $P'Q'$ for an arbitrary projection. Using equations (3.33, 3.34), with the constraint implied by the above equation, we have

$$\tan\beta = \lim \frac{\delta x}{\delta y} = \lim \left.\frac{x_\lambda\,\delta\lambda + x_\phi\,\delta\phi}{y_\lambda\,\delta\lambda + y_\phi\,\delta\phi}\right|_{\delta\phi=\frac{\cos\phi\,\delta\lambda}{\tan\alpha}} = \frac{x_\lambda\tan\alpha + x_\phi\cos\phi}{y_\lambda\tan\alpha + y_\phi\cos\phi}. \tag{3.46}$$

We already know $\tan\gamma$ from equation (3.35), therefore we can calculate α', the angle between the projected meridian and parallel, as

$$\tan\alpha' = \tan(\beta+\gamma) = \frac{\tan\beta + \tan\gamma}{1-\tan\beta\tan\gamma} \tag{3.47}$$

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

$$= \frac{(x_\lambda y_\phi - x_\phi y_\lambda)\tan\alpha}{\cos\phi\,(x_\phi^2 + y_\phi^2) + (x_\lambda x_\phi + y_\lambda y_\phi)\tan\alpha}. \tag{3.48}$$

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

$$(x_\lambda x_\phi + y_\lambda y_\phi)\tan\alpha + \left[\cos\phi\,(x_\phi^2+y_\phi^2) - (x_\lambda y_\phi - x_\phi y_\lambda)\right] \equiv 0. \tag{3.49}$$

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:

$$x_\lambda x_\phi + y_\lambda y_\phi = 0 \tag{3.50}$$

$$\cos\phi\,(x_\phi^2 + y_\phi^2) = (x_\lambda y_\phi - x_\phi y_\lambda). \tag{3.51}$$

Using (3.50), the second of these equations can be written as

$$\cos\phi\, y_\phi^2\left(1 + x_\phi^2/y_\phi^2\right) = x_\lambda y_\phi\left(1 + x_\phi^2/y_\phi^2\right), \tag{3.52}$$

so that we must have $x_\lambda = \cos\phi\, y_\phi$. If we then substitute this back into (3.50) we obtain $y_\lambda = -\cos\phi\, x_\phi$. Thus the following conditions are necessary (and trivially sufficient) for a conformal transformation from the sphere to the plane.

CAUCHY–RIEMANN

$$\boxed{x_\lambda = \cos\phi\, y_\phi, \qquad y_\lambda = -\cos\phi\, x_\phi} \tag{3.53}$$

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 (3.40, 3.41).

Conformality implies scale isotropy

Consider now the scale factor for an arbitrary transformation. Substituting δx and δy from equations (3.33 , 3.34) into the definition of the scale factor (equation 2.18) we have

$$\mu^2 = \lim_{Q\to P}\frac{\delta s'^2}{\delta s^2} = \lim_{Q\to P}\frac{\delta x^2+\delta y^2}{R^2\delta\phi^2+R^2\cos^2\phi\,\delta\lambda^2}$$
$$= \lim_{Q\to P}\frac{E\delta\phi^2+2F\delta\phi\,\delta\lambda+G\delta\lambda^2}{R^2\delta\phi^2+R^2\cos^2\phi\,\delta\lambda^2}. \qquad (3.54)$$

where

$$E(\lambda,\phi)=x_\phi^2+y_\phi^2, \qquad F(\lambda,\phi) = x_\lambda x_\phi+y_\lambda y_\phi, \qquad G(\lambda,\phi)=x_\lambda^2+y_\lambda^2. \qquad (3.55)$$

The Cauchy–Riemann equations (3.53) imply that $F=0$ and $G=\cos^2\phi\,E$ so the above limit reduces to E/R^2 independent of α. Therefore the scale factor is isotropic with a value given by

ISOTROPIC SCALE
$$\boxed{k(\lambda,\phi)=\frac{1}{R}\sqrt{x_\phi^2+y_\phi^2}=\frac{1}{R\cos\phi}\sqrt{x_\lambda^2+y_\lambda^2}}\,. \qquad (3.56)$$

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 (3.37–3.41) to confirm the results of Section 3.4.

3.8 Series expansions for the unmodified TMS

The transverse Mercator on the ellipsoid (TME) will be determined as power series. Here we derive the corresponding series for TMS; thy will provide a check for the TME series in the limit of the eccentricity of the ellipsoid tending to zero. 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 $x/a < 0.05$. In this section we shall drop terms involving fifth or higher powers of these small parameters but when we construct the series for TME we will retain higher orders

The coefficients of the direct series involve trigonometric functions of ϕ which not not necessarily small: 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:

$$s=\sin\phi \qquad c=\cos\phi \qquad t=\tan\phi \qquad (3.57)$$
$$s_1=\sin\phi_1 \qquad c_1=\cos\phi_1 \qquad t_1=\tan\phi_1 \qquad m(\phi_1)=y, \qquad (3.58)$$

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 $$x = \frac{R}{2}\ln\left[\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi}\right] = \frac{R}{2}\ln\left[\frac{1+c\sin\lambda}{1-c\sin\lambda}\right].$$

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

$$\begin{aligned} x(\lambda,\phi) &= Rc(\lambda - \frac{1}{6}\lambda^3 + \cdots) + \frac{1}{3}Rc^3(\lambda - \cdots)^3 + \cdots \\ &= Rc\lambda + \frac{1}{6}Rc^3(1-t^2)\lambda^3 + \cdots . \end{aligned} \tag{3.59}$$

Direct transformation for *y*

Equation (3.13b) is $$y(\lambda,\phi) = R\tan^{-1}\left[\sec\lambda\tan\phi\right] = R\tan^{-1}\left[t\sec\lambda\right].$$

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

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

Now $R\tan^{-1}(t) = R\tan^{-1}(\tan\phi) = R\phi = m(\phi)$ so we can write

$$y(\lambda,\phi) = m(\phi) + \frac{Rsc}{2}\lambda^2 + \frac{Rsc^3\lambda^4}{24}(5-t^2) + \cdots . \tag{3.60}$$

Inverse transformation for λ

Setting $y/R = \phi_1$, the footpoint latitude, equation (3.22) is

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

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

$$\begin{aligned} \lambda(x,y) &= \frac{1}{c_1}\left(\frac{x}{R} + \frac{1}{6}\frac{x^3}{R^3} + \cdots\right) - \frac{1}{3c_1^3}\left(\frac{x}{R} + \cdots\right)^3 + \cdots \\ &= \frac{1}{c_1}\left(\frac{x}{R}\right) + \frac{1}{c_1}\left(\frac{1}{6} - \frac{1}{3}c_1^{-2}\right)\left(\frac{x}{R}\right)^3 + \cdots \\ &= \frac{1}{c_1}\left(\frac{x}{R}\right) - \frac{(1+2t_1^2)}{6c_1}\left(\frac{x}{R}\right)^3 + \cdots \qquad \text{where} \quad m(\phi_1) = y. \end{aligned} \tag{3.61}$$

Inverse transformation for ϕ

Setting $y/R = \phi_1$ in equation (3.23) gives

$$\phi(x,y) = \sin^{-1}[\operatorname{sech}(x/R)\sin(y/R)] = \sin^{-1}[s_1 \operatorname{sech}(x/R)].$$

Use (E.24) to write

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

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

$$\begin{aligned} \phi(x,y) &= \sin^{-1} s_1 + \frac{s_1}{(1-s_1^2)^{1/2}}\left(-\frac{1}{2}\left(\frac{x}{R}\right)^2 + \frac{5}{24}\left(\frac{x}{R}\right)^4\right) \\ &\qquad + \frac{1}{2}\frac{s_1}{(1-s_1^2)^{3/2}} s_1^2 \left(-\frac{1}{2}\left(\frac{x}{R}\right)^2 + \cdots\right)^2 + \cdots \\ &= \phi_1 - \frac{t_1}{2}\left(\frac{x}{R}\right)^2 + \frac{t_1}{24}(5+3t_1^2)\left(\frac{x}{R}\right)^4 + \cdots. \qquad \text{where} \quad m(\phi_1) = y. \end{aligned} \tag{3.62}$$

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 (3.29) gives

$$\begin{aligned} k(\lambda,\phi) &= [1 - \sin^2\lambda\cos^2\phi]^{-1/2} \\ &= 1 + \frac{1}{2}c^2\left(\lambda - \frac{1}{6}\lambda^3 + \cdots\right)^2 + \frac{3}{8}c^4(\lambda - \cdots)^4 \\ &= 1 + \frac{1}{2}c^2\lambda^2 + \frac{1}{24}c^4\lambda^4(5 - 4t^2) + \cdots \end{aligned} \tag{3.63}$$

Similarly equations (3.30) and (E.22) give

$$k(x,y) = \cosh\left(\frac{x}{R}\right) = 1 + \frac{1}{2!}\left(\frac{x}{R}\right)^2 + \frac{1}{4!}\left(\frac{x}{R}\right)^4 + \cdots. \tag{3.64}$$

Series expansion for convergence

Equation (3.42) is $\gamma(\lambda,\phi) = \tan^{-1}\left[\tan\lambda\sin\phi\right]$. Expanding $\tan\lambda$ with (E.15) and using the expansion for arctan in equation (E.20) gives

$$\begin{aligned} \gamma(\lambda,\phi) &= s(\lambda + (1/3)\lambda^3 + \cdots) - (1/3)s^3(\lambda + \cdots)^3 \\ &= s\lambda + \frac{1}{3}sc^2\lambda^3 + \cdots \end{aligned} \tag{3.65}$$

Equation (3.43) is

$$\gamma(x,y)=\tan^{-1}\left[\tanh(x/R)\tan(y/R)\right]=\tan^{-1}\left[t_1\tanh(x/R)\right].$$

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

$$\begin{aligned}\gamma(x,y) &= t_1\left(\frac{x}{R}-\frac{1}{3}\left(\frac{x}{R}\right)^3+\cdots\right)-\frac{1}{3}t_1^3\left(\frac{x}{R}+\cdots\right)^3\\ &= t_1\left(\frac{x}{R}\right)-\frac{1}{3}t_1(1+t_1^2)\left(\frac{x}{R}\right)^3+\cdots \qquad \text{where} \quad m(\phi_1)=y. \end{aligned} \tag{3.66}$$

3.9 Secant 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 multiplying the transformation formulae in equations (3.13) by a factor of k_0. The corresponding modifications for the inverses, scale factors and convergence are easily deduced: they are listed below along with the corresponding series solutions. On the central meridian we now have $y(0,\phi)=k_0 m(\phi)$ where we still have $m(\phi)=R\phi$. The corresponding definition of the footpoint latitude becomes

$$m(\phi_1)=\frac{y}{k_0}. \tag{3.67}$$

We continue to use the abbreviations for the trig functions of ϕ and ϕ_1 (equations 3.57, 3.58)

Direct transformations

$$x(\lambda,\phi)=\frac{1}{2}k_0R\ln\left[\frac{1+\sin\lambda\cos\phi}{1-\sin\lambda\cos\phi}\right]=k_0R\left(c\lambda+\frac{1}{6}c^3\lambda^3(1-t^2)+\cdots\right) \tag{3.68}$$

$$y(\lambda,\phi)=k_0R\tan^{-1}\left[\sec\lambda\tan\phi\right] \qquad =k_0m(\phi)+k_0R\left(\frac{sc}{2}\lambda^2+\frac{sc^3\lambda^4}{24}(5-t^2)+\cdots\right) \tag{3.69}$$

Inverse transformations

$$\lambda(x,y)=\tan^{-1}\left[\sinh\frac{x}{k_0R}\sec\frac{y}{k_0R}\right] \quad =\frac{1}{c_1}\left(\frac{x}{k_0R}\right)-\frac{(1+2t_1^2)}{6c_1}\left(\frac{x}{k_0R}\right)^3+\cdots \tag{3.70}$$

$$\phi(x,y)=\sin^{-1}\left[\operatorname{sech}\frac{x}{k_0R}\sin\frac{y}{k_0R}\right] \quad =\phi_1-\frac{t_1}{2}\left(\frac{x}{k_0R}\right)^2+\frac{t_1}{24}(5+3t_1^2)\left(\frac{x}{k_0R}\right)^4+\cdots \tag{3.71}$$

Convergence

$$\gamma(\lambda,\phi)=\tan^{-1}(\tan\lambda\sin\phi) \qquad =s\lambda+\frac{1}{3}sc^2\lambda^3+\cdots \tag{3.72}$$

$$\gamma(x,y)=\tan^{-1}\left(\tanh\frac{x}{k_0R}\tan\frac{y}{k_0R}\right)=t_1\left(\frac{x}{k_0R}\right)-\frac{1}{3}t_1(1+t_1^2)\left(\frac{x}{k_0R}\right)^3+\cdots \tag{3.73}$$

Scale factors

$$k(\lambda,\phi) = \frac{k_0}{(1-\sin^2\lambda\cos^2\phi)^{1/2}} = k_0\left[1+\frac{1}{2}c^2\lambda^2+\frac{1}{24}c^4\lambda^4(5-4t^2)+\cdots\right] \tag{3.74}$$

$$k(x,y) = k_0\cosh\left(\frac{x}{k_0R}\right) = k_0\left[1+\frac{1}{2}\left(\frac{x}{k_0R}\right)^2+\frac{1}{24}\left(\frac{x}{k_0R}\right)^4+\cdots\right] \tag{3.75}$$

Consider the scale factor in terms of projection coordinates, that is $k(x,y)$. If we choose $k_0 = 0.9996$ then we find that the scale is true (k=1) when $x/R = \pm 0.0282$ corresponding to $x = \pm 180$km (approximately). Once outside these lines the accuracy decreases as k increases. 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 very accurate, within 0.04% over a width of approximately 510km. We shall see later that this includes most of the area covered by the British grid. These values are only slightly altered on the ellipsoid (TME) but the lines of unit scale are no longer straight on the projection.

Blank page. A contradiction.

Chapter 4

NMS to TMS by complex variables

Abstract

Complex variable theory is used to derive TMS from NMS by closed formulae and Taylor series expansions. Scale and convergence are written in terms of the derivative of the complex function describing the transformation.

4.1 Introduction

In Chapter 2 we derived the NMS projection: it takes a point $P(\phi,\lambda)$ on the sphere to a point P' on a plane with projection coordinates (λ,ψ) where (from 2.55)

$$\psi = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right] = \tanh^{-1}(\sin\phi) = \mathrm{gd}^{-1}\phi. \tag{4.1}$$

In Chapter 3 we derived the TMS projection: it takes a point $P(\phi,\lambda)$ on the sphere to a point P'' on a plane with projection coordinates (x,y) where (from 3.13)

$$x(\phi,\lambda) = R\tanh^{-1}[\sin\lambda\cos\phi], \qquad , \tag{4.2}$$

$$y(\phi,\lambda) = R\tan^{-1}[\sec\lambda\tan\phi]. \tag{4.3}$$

In addition to these exact closed forms we obtained low order series such as (3.59–3.66):

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

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

We shall show how the above equations for TMS, both exact closed forms and low order series, may be derived from NMS by using the theory of complex variables. This should be possible since both NMS and TMS are conformal transformations from the sphere and the theory of analytic complex functions incorporates conformality in a natural way.

Let (λ,ψ) be the real and imaginary axes of a complex ζ-plane with $\zeta=\lambda+i\psi$. Likewise, (x,y) are the real and imaginary axes of a complex z-plane with $z=x+iy$. For the direct transformation we seek functions $x(\lambda,\psi)$ and $y(\lambda,\psi)$ and for the inverse transformations functions $\lambda(x,y)$ and $\psi(x,y)$. The transformations to and from the sphere are accomplished through the known relations between ψ and ϕ. Almost any function $z(\zeta)$ defines a conformal transformation between the two complex planes but we shall find that imposition of boundary conditions on the central meridian fixes the function uniquely.

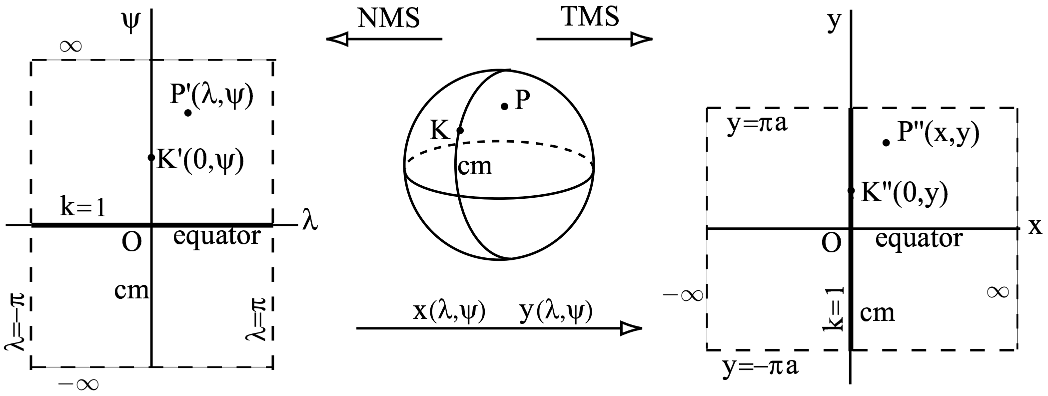


Figure 4.1

Figure 4.1 (with $a{\to}R$ twice) summarizes the properties of the two projections. A general point P on the sphere with coordinates (λ,ϕ) projects into $P'(\lambda,\psi)$ for NMS and $P''(x,y)$ for TMS A general point $K(0,\phi)$ on the central meridian of the sphere, taken as the Greenwich meridian for simplicity, projects into points $K'(0,\psi)$ and $K''(0,y)$.

NMS is conformal, of finite extent in λ and infinite extent in ψ. It is true to scale on the equator: $k=1$ when $\phi=0$ (equation 2.21).

TMS is also conformal, of finite extent in y and infinite extent in x. It is true to scale on the central meridian: $k=1$ when $x=0$, (equation 3.30). Equations 3.13 also imply that where $\lambda=0$ we must have $y(0,\phi)=R\phi=m(\phi)$. We shall show that if we demand that TMS has these properties then the functions $x(\lambda,\psi)$ and $y(\lambda,\psi)$ are completely determined.

The meridian distance as a function of ψ

The condition that λ=0 maps to x=0 is trivial but the second, true scale on the central meridian, implies that $y(0,\psi)$ must equal the meridian distance which we have written as a function of ϕ, namely $m(\phi)=R\phi$. It is convenient to introduce $M(\psi)$, the meridian distance as a function of ψ, by

$$M(\psi)=m\,(\phi(\psi))=R\,\phi(\psi), \tag{4.6}$$

and write the scale condition in two equivalent ways.

$$y(0,\psi)=M\big(\psi\big), \tag{4.7}$$

$$y(0,\psi(\phi))=m(\phi)=R\phi. \tag{4.8}$$

On the sphere we can use any one of the equations (2.56) to obtain an explicit expression for $M(\psi)$: we choose

$$M(\psi)=R\tan^{-1}\big[\sinh\psi\big]. \tag{4.9}$$

In subsequent calculations we need the first four derivatives of $M(\psi)$ with respect to ψ. These are straightforward enough to obtain as functions of ψ, For example

$$M'(\psi)=\frac{R\cosh\psi}{1+\sinh^2\psi}=R\,\mathrm{sech}\,\psi. \tag{4.10}$$

However, it will prove more useful to express the derivatives of M in terms of ϕ. For example we have

$$M'(\psi(\phi)) \equiv \frac{dM(\psi(\phi))}{d\psi} = \frac{dm(\phi)}{d\phi}\frac{d\phi}{d\psi} = R\cos\phi, \tag{4.11}$$

where we have used the defining equation for NMS (2.25)

$$\frac{d\psi}{d\phi} = \sec\phi \tag{4.12}$$

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 of Section 3.8, ($s = \sin\phi$ *etc.*).

$$\begin{aligned}
M' &= R\cos\phi &&= \quad Rc,\\
M'' &= \frac{d(Rc)}{d\phi}\frac{d\phi}{d\psi} &&= -Rsc,\\
M''' &= -R(c^2 - s^2)c &&= -Rc^3(1-t^2),\\
M'''' &= -R(-3sc^2 - 2sc^2 + s^3)c &&= \quad Rsc^3(5-t^2).
\end{aligned} \tag{4.13}$$

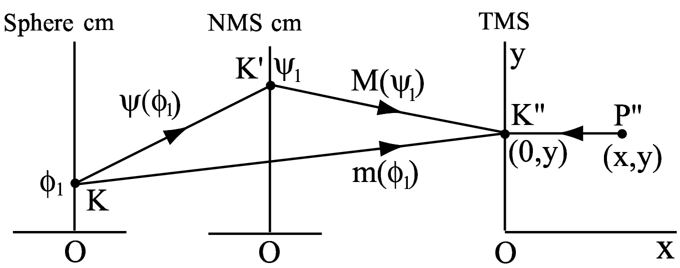


Figure 4.2

The footpoint parameter ψ_1

In Section 3.1, where we discussed the inverse TMS transformations, we introduced the footpoint and the footpoint latitude. We shall drive the inverse transformations from TMS to NMS as Taylor series expansions about the **footpoint parameter** ψ_1. Figure 4.2 shows the (x,y) plane of TMS and the central meridians only of NMS and the sphere. For any point $P''(x,y)$ its footpoint is at $K''(0,y)$. The footpoint latitude is that of the point $K(0,\phi_1)$ on the sphere such that $m(\phi_1) = y$: the footpoint parameter is the ordinate of that point $K'(0,\psi_1)$ in the NMS plane such that $M(\psi_1) = y$. We have

$$y = m(\phi_1) = R\phi_1 = M(\psi_1) = R\tan^{-1}\sinh\psi_1. \tag{4.14}$$

Conformality and complex functions

There is a very brief introduction to the theory of complex functions in Appendix G (and in the first chapter of any book on complex functions). The central point is that a complex function, $w(z)$, of a complex variable z, has a unique derivative independent of direction if, and only if, the Cauchy–Riemann conditions are satisfied: in the notation of this section the partial derivatives of the real and imaginary parts of $z(\zeta)$, namely $x(\lambda,\psi)$ and $y(\lambda,\psi)$, must satisfy

CAUCHY–RIEMANN $$\boxed{x_\lambda = y_\psi, \quad y_\lambda = -x_\psi} \tag{4.15}$$

In Appendix G we show that if the Cauchy-Riemann conditions are satisfied then the transformation preserves angles. Two lines intersecting at an angle α at a point P in the ζ-plane intersect at the same angle when transformed to the z-plane: the definition of conformality.

4.2 Closed formulae for the TMS transformation

The transformation $z(\zeta)$ from the ζ-plane to the z-plane must satisfy boundary conditions

$$x(0,\psi) = 0, \qquad y(0,\psi(\phi)) = m(\phi) = R\phi, \tag{4.16}$$

$$\text{or} \qquad z(0,\psi) = iR\phi = iR\,\mathrm{gd}\,\psi = R\,\mathrm{gd}^{-1}(i\psi) = R\,\mathrm{gd}^{-1}(\mathrm{Im}\,\zeta), \tag{4.17}$$

where we have set $\phi = \mathrm{gd}\,\psi$ (equation 2.56) and $i\,\mathrm{gd}(\psi) = \mathrm{gd}^{-1}(i\psi)$ (equation C.63). Therefore it is "obvious" that we should consider the transformation

$$\boxed{z(\zeta) = R\,\mathrm{gd}^{-1}(\zeta)} \tag{4.18}$$

or, in full,

$$x(\lambda,\psi) + iy(\lambda,\psi) = R\,\mathrm{gd}^{-1}(\lambda + i\psi). \tag{4.19}$$

When λ=0, this becomes

$$x(0,\psi) + iy(0,\psi) = R\,\mathrm{gd}^{-1}(i\psi) = iR\,\mathrm{gd}(\psi) = iR\phi,$$

reproducing the boundary condition as above.

Comment on notation. In more advanced treatments, such as Lee (1976) and Karney (2011), the complex planes are chosen with the imaginary axes along the equator and the real axis along the central meridian. As a result the transformation is simply $z = R\,\mathrm{gd}(\zeta)$ with $\zeta = \psi + i\lambda$ and $z = y + ix$. The final results are the same.

Evaluating the real and imaginary parts of Equation 4.18 and its inverse, $\zeta = \mathrm{gd}(z/R)$, is straightforward but non-trivial. The details are given in Appendix G, Equations G.24–G.40. The results are

$$\begin{aligned} x &= R\tanh^{-1}[\sin\lambda\ \mathrm{sech}\,\psi], & \lambda &= \tan^{-1}[\sinh(x/R)\sec(y/R)], \\ y &= R\tan^{-1}[\sec\lambda\ \sinh\psi], & \psi &= \tanh^{-1}[\mathrm{sech}(x/R)\sin(y/R)]. \end{aligned} \tag{4.20}$$

From Equation 2.55 we substitute $\mathrm{sech}\,\psi = \cos\phi$, $\sinh\psi = \tan\phi$ and $\tanh\psi = \sin\phi$ to give the direct and inverse equations for TMS, 3.13 and 3.22.

The Cauchy–Riemann conditions

To check the Cauchy–Riemann equations (4.15) we evaluate the partial derivatives of x and y from Equations 4.20 (setting R=1 for clarity),

$$\tanh x = \sin\lambda\,\operatorname{sech}\psi, \qquad \tan y = \sec\lambda\,\sinh\psi, \tag{4.21}$$

$$(\operatorname{sech}^2 x)\,x_\lambda = \cos\lambda\,\operatorname{sech}\psi, \qquad (\sec^2 y)\,y_\lambda = \sin\lambda\,\sec^2\lambda\,\sinh\psi, \tag{4.22}$$

$$(\operatorname{sech}^2 x)\,x_\psi = -\sin\lambda\,\sinh\psi\,\operatorname{sech}^2\psi, \qquad (\sec^2 y)\,y_\psi = \sec\lambda\,\cosh\psi \tag{4.23}$$

and simplify using

$$\operatorname{sech}^2 x = 1 - \sin^2\lambda\,\operatorname{sech}^2\psi = \operatorname{sech}^2\psi\left[\cosh^2\psi - \sin^2\lambda\right], \tag{4.24}$$

$$\sec^2 y = 1 + \sec^2\lambda\,\sinh^2\psi = \sec^2\lambda\left[\cosh^2\psi - \sin^2\lambda\right]. \tag{4.25}$$

4.3 Transformation to the TMS series

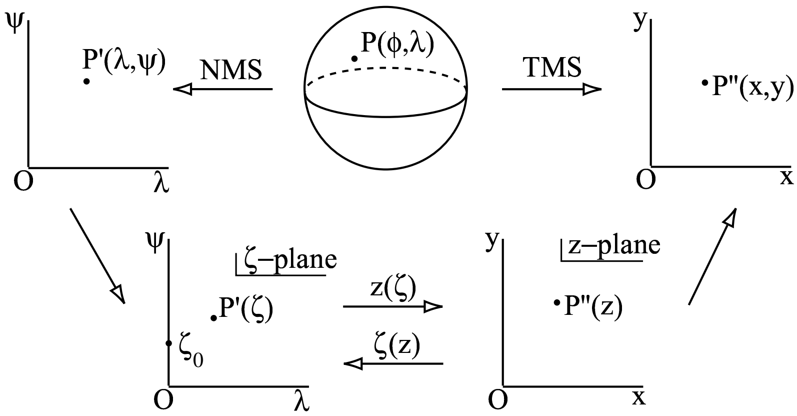


Figure 4.3

The steps of the transformation are summarized in the figure: anti-clockwise from the sphere to the normal Mercator projection, to the (λ,ψ) complex plane and then to the (x,y) complex plane by means of the function $z(\zeta)$. This transformation is expressed as a complex Taylor series expansion about a point $\zeta_0 = i\psi_0$ on the imaginary axis of the ζ-plane. The real and imaginary parts of the complex series generate the direct series Equations 4.4, 4.4. We expect that the series to be valid for small values of $|\zeta-\zeta_0|$.

The coefficients in the Taylor series are related to the values of $z(\zeta)$ and its derivatives at ζ_0. The clever trick is to use the fact that the derivatives of an analytic function may be evaluated in any direction in the complex plane. By taking them along the imaginary axis we shall find that they reduce to derivatives of the meridian distance function $M(\psi)$ with respect to ψ. These are the derivatives evaluated in Section 4.1.

The inverse transformations $\lambda(x,y)$ and $\psi(x,y)$ are evaluated by inverting the complex Taylor series and then taking the real and imaginary parts. Finally $\phi(x,y)$ is calculated from $\psi(x,y)$ by a further Taylor series.

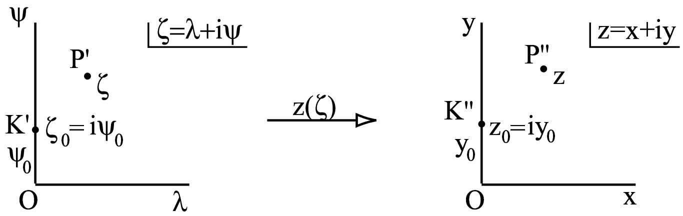


Figure 4.4

The direct complex Taylor series

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

$$z(\zeta) = z_0 + (\zeta - \zeta_0)z'(\zeta_0) + \frac{1}{2!}(\zeta - \zeta_0)^2 z''(\zeta_0) + \frac{1}{3!}(\zeta - \zeta_0)^3 z'''(\zeta_0) + \cdots \tag{4.26}$$

and it must satisfy boundary conditions

$$x(0, \psi) = 0, \tag{4.27}$$

$$y(0, \psi) = M(\psi). \tag{4.28}$$

so that at $\zeta = \zeta_0 = i\psi_0$

$$z_0 = z(\zeta_0) = iy_0 = iM(\psi_0) = iM_0 \tag{4.29}$$

Construct the derivatives of $z(\zeta)$ from first principles, (as in G.45):

$$z'(\zeta) = \lim_{\delta\zeta \to 0} \frac{z(\zeta + \delta\zeta) - z(\zeta)}{\delta\zeta}. \tag{4.30}$$

Conformality demands that $z(\zeta)$ be analytic and this limit must be independent of direction: we choose to take it in the ψ direction so that $\lambda = 0$ and $\delta\zeta = i\delta\psi$. Therefore we have

$$z'(\zeta) = \left(\frac{1}{i}\frac{d}{d\psi}\right) z(\zeta)\bigg|_{\lambda=0} = \left(\frac{1}{i}\frac{d}{d\psi}\right) iM(\psi) \tag{4.31}$$

Therefore the derivatives at ζ_0 are

$$\begin{aligned}
z'(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\left(iM(\psi)\right)\bigg|_{\psi_0} &&= M'(\psi_0),\\
z''(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\left(M'(\psi)\right)\bigg|_{\psi_0} &&= -iM''(\psi_0),\\
z'''(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\left(-iM''(\psi)\right)\bigg|_{\psi_0} &&= -M'''(\psi_0),\\
z''''(\zeta_0) &= \left(-i\frac{d}{d\psi}\right)\left(-M'''(\psi)\right)\bigg|_{\psi_0} &&= iM''''(\psi_0).
\end{aligned} \tag{4.32}$$

Finally, if we abbreviate $M'(\psi_0) = M'_0$, $M''(\psi_0) = M''_0$ *etc.*, the Taylor series (4.26) may be written as

$$\boxed{z = z_0 + (\zeta - \zeta_0)M'_0 - \frac{i}{2!}(\zeta - \zeta_0)^2 M''_0 - \frac{1}{3!}(\zeta - \zeta_0)^3 M'''_0 + \frac{i}{4!}(\zeta - \zeta_0)^4 M''''_0 + \cdots} \tag{4.33}$$

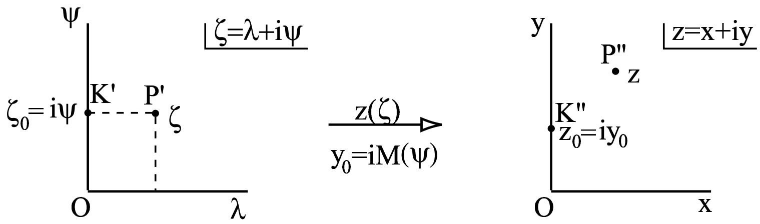


Figure 4.5

The direct series for x and y

When we derived the direct series for TMS in Section 3.8 we expanded x and y as power series in λ keeping ϕ constant, corresponding to ψ constant in the ζ-plane. Therefore, if we start from a given point at $\zeta = \lambda + i\psi$, we must choose ζ_0 at K' with the *same* ordinate, that is $\zeta_0 = i\psi$: see Figure 4.5. In the Taylor series (4.33) we therefore set $\zeta - \zeta_0 = (\lambda + i\psi) - i\psi = \lambda$ so that the complex series reduces to a power series in λ with complex coefficients. Since M and its derivatives are evaluated at $\psi_0 = \psi$ we have $z_0 = iM_0 \to iM$ and $M'_0 \to M'$ *etc.* The series is now

$$z = x + iy = iM + \lambda M' - \frac{i}{2!}\lambda^2 M'' - \frac{1}{3!}\lambda^3 M''' + \frac{i}{4!}\lambda^4 M'''' + \cdots \tag{4.34}$$

The real and imaginary parts of equation (4.34) give x and y as functions of λ and ψ. The derivatives of M are real so that the transformations from NMS→TMS are

$$x(\lambda,\psi) = \lambda M' - \frac{1}{3!}\lambda^3 M''' + \cdots \tag{4.35}$$

$$y(\lambda,\psi) = M - \frac{1}{2!}\lambda^2 M'' + \frac{1}{4!}\lambda^4 M'''' + \cdots . \tag{4.36}$$

On substituting for M and its derivatives using equations (4.13), we obtain the corresponding expressions in terms of λ and ϕ (with $s = \sin\phi$ *etc.*)

$$x(\lambda,\phi) = Rc\lambda + \frac{1}{3!}Rc^3(1-t^2)\lambda^3 + \cdots , \tag{4.37}$$

$$y(\lambda,\phi) = R\phi + \frac{1}{2!}Rsc\,\lambda^2 + \frac{1}{4!}Rsc^3(5-t^2)\,\lambda^4 + \cdots . \tag{4.38}$$

These results agree with the expansions obtained in equations (3.59, 3.60).

The Cauchy–Riemann equations

It is instructive to verify that the (plane to plane) Cauchy–Riemann equations (4.15) are indeed satisfied by equations (4.35) and (4.36) (at least if the series are continued to infinity). Evaluating the four partial derivatives we have

$$x_\lambda = \quad y_\psi = M' - \frac{1}{2!}\lambda^2 M''' + \cdots \tag{4.39}$$

$$y_\lambda = -x_\psi = -\lambda M'' + \frac{1}{3!}\lambda^3 M'''' + \cdots . \tag{4.40}$$

Note also that the equations (4.37, 4.38) satisfy the Cauchy–Riemann equations (3.53) which apply to the transformation from the *sphere* to the TMS plane. Explicitly

$$x_\lambda = \quad \cos\phi\, y_\phi = Rc + \frac{1}{2!}Rc^3(1-t^2)\lambda^2 \cdots, \tag{4.41}$$

$$y_\lambda = -\cos\phi\, x_\phi = Rsc\,\lambda + \frac{1}{3!}Rsc^3(5-t^2)\,\lambda^3 + \cdots. \tag{4.42}$$

The inverse complex series: method of Lagrange series reversion

The simplest method of obtaining the inverse series is to use the Lagrange method described in Appendix B; in particular we use the reversion 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 (4.33) as

$$\frac{z-z_0}{M_0'} = (\zeta-\zeta_0) + \frac{b_2}{2!}(\zeta-\zeta_0)^2 + \frac{b_3}{3!}(\zeta-\zeta_0)^3 + \frac{b_4}{4!}(\zeta-\zeta_0)^4 + \cdots \tag{4.43}$$

where

$$b_2 = -i\frac{M_0''}{M_0'}, \qquad b_3 = -\frac{M_0'''}{M_0'}, \qquad b_4 = i\frac{M_0''''}{M_0'}. \tag{4.44}$$

The series (4.43) 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 (4.43):

$$\boxed{\zeta-\zeta_0 = \left(\frac{z-z_0}{M_0'}\right) - \frac{p_2}{2!}\left(\frac{z-z_0}{M_0'}\right)^2 - \frac{p_3}{3!}\left(\frac{z-z_0}{M_0'}\right)^3 - \frac{p_4}{4!}\left(\frac{z-z_0}{M_0'}\right)^4 + \cdots} \tag{4.45}$$

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

$$\begin{aligned} p_2 &= b_2 &&= -\frac{iM_0''}{M_0'} \\ p_3 &= b_3 - 3b_2^2 &&= -\frac{M_0'''}{M_0'} + 3\frac{(M_0'')^2}{(M_0')^2} \\ p_4 &= b_4 - 10b_2b_3 + 15b_2^3 &&= \frac{iM_0''''}{M_0'} - 10i\frac{M_0''M_0'''}{(M_0')^2} + 15i\frac{(M_0'')^3}{(M_0')^3}. \end{aligned} \tag{4.46}$$

Using equations (4.13), these become

$$\begin{aligned} p_2 &= \ is_0 \\ p_3 &= \ c_0^2(1+2t_0^2) \\ p_4 &= -is_0c_0^2(5+6t_0^2), \end{aligned} \tag{4.47}$$

with $c_0 = \cos\phi_0$ *etc.* where $\phi_0 = \operatorname{gd}\psi_0$.

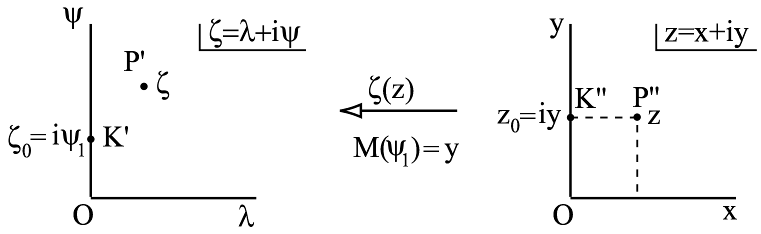


Figure 4.6

The inverse series for the Mercator parameter

For the direct series we kept ψ constant in the ζ plane, Figure 4.5. For the inverse we keep y constant in the z-plane, Figure 4.6: K'' is the footpoint for P''. Therefore $z_0 = iy$ and $z - z_0 = (x+iy) - iy = x$ and therefore the complex series (4.45) reduces to a power series in x with the complex coefficients given in Equations 4.46. The coefficients are expressed in terms of the derivatives of the meridian function $M(\psi)$ evaluated at ψ_1 but if we use equations (4.13) the derivatives must be evaluated at the footpoint latitude ϕ_1. Setting $M_1 = M(\psi_1) = m(\phi_1) = y$ *etc.* equation (4.45) becomes

$$(\lambda + i\psi) - i\psi_1 = \frac{x}{M_1'} - \frac{p_2}{2!}\left(\frac{x}{M_1'}\right)^2 - \frac{p_3}{3!}\left(\frac{x}{M_1'}\right)^3 - \frac{p_4}{4!}\left(\frac{x}{M_1'}\right)^4 + \cdots \tag{4.48}$$

and the real and imaginary parts are

$$\lambda(x,y) = \frac{x}{M_1'} - \frac{1}{3!}p_3\left(\frac{x}{M_1'}\right)^3 + \cdots \tag{4.49}$$

$$\psi(x,y) = \psi_1 - \frac{1}{2!}\operatorname{Im}(p_2)\left(\frac{x}{M_1'}\right)^2 - \frac{1}{4!}\operatorname{Im}(p_4)\left(\frac{x}{M_1'}\right)^4 + \cdots. \tag{4.50}$$

Substitute for the p-coefficients (4.47) and M_1' (from 4.13), all evaluated at the footpoint latitude $\phi_1 = \phi(\psi_1) = \operatorname{gd}\psi_1$.

$$\lambda(x,y) = \frac{1}{c_1}\frac{x}{R} - \frac{1}{3!}\frac{1}{c_1}\left[1+2t_1^2\right]\left(\frac{x}{R}\right)^3 + \cdots, \tag{4.51}$$

$$\psi(x,y) = \psi_1 - \frac{1}{2!}\frac{t_1}{c_1}\left(\frac{x}{R}\right)^2 + \frac{1}{4!}\frac{t_1}{c_1}\left[5+6t_1^2\right]\left(\frac{x}{R}\right)^4 + \cdots, \tag{4.52}$$

where $c_1 = \cos\phi_1$, $t_1 = \tan\phi_1$, $m(\phi_1) = R\phi_1 = y$ and $\psi_1 = \operatorname{gd}^{-1}\phi_1 = \tanh^{-1}\sin\phi_1$. Term by term the derivatives satisfy the Cauchy-Riemann conditions for the inverse function $\zeta(z)$:

$$\lambda_x = \psi_y = \frac{1}{Rc_1} - \frac{1+2t_1^2}{2Rc_1}\left(\frac{x}{R}\right)^2 + \cdots, \tag{4.53}$$

$$\lambda_y = -\psi_x = \frac{t_1}{Rc_1}\frac{x}{R} - \frac{t_1}{6Rc_1}\left[5+6t_1^2\right]\left(\frac{x}{R}\right)^3 + \cdots. \tag{4.54}$$

Relating latitude and the Mercator parameter

The series (4.53) for λ is in agreement with equation (3.61) but we must now derive the series for ϕ from that for ψ.

Equation 4.52 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:

$$\phi(\psi) = \phi(\psi_1) + (\psi - \psi_1)\left.\frac{d\phi}{d\psi}\right|_1 + \frac{1}{2!}(\psi - \psi_1)^2 \left.\frac{d^2\phi}{d\psi^2}\right|_1 + \cdots. \tag{4.55}$$

The derivatives of $\phi(\psi)$ are easily obtained from the defining equation for ψ, (2.25):

$$\frac{d\psi}{d\phi} = \sec\phi, \qquad \frac{d\phi}{d\psi} = \left(\frac{d\psi}{d\phi}\right)^{-1} = \cos\phi, \tag{4.56}$$

$$\frac{d^2\phi}{d\psi^2} = \frac{d}{d\psi}(\cos\phi) = -\sin\phi\,\frac{d\phi}{d\psi} = -\sin\phi\,\cos\phi. \tag{4.57}$$

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

$$\phi = \phi_1 + (\psi - \psi_1)\cos\phi_1 - \frac{1}{2!}(\psi - \psi_1)^2 \sin\phi_1 \cos\phi_1 + \cdots. \tag{4.58}$$

The inverse series for latitude

In Equation 4.58 we substitute for $\psi - \psi_1$ from equation (4.52). To order $(x/a)^4$,

$$\phi(x,y) = \phi_1 + \left[-\frac{1}{2!}\left(\frac{x}{R}\right)^2 + \frac{1}{4!}\left(\frac{x}{R}\right)^4 \left[5+6t_1^2\right]\right]\frac{t_1}{c_1}c_1$$
$$-\frac{1}{2!}\left[-\frac{1}{2!}\left(\frac{x}{R}\right)^2 + \cdots\right]^2 \left(\frac{t_1}{c_1}\right)^2 s_1 c_1$$

which simplifies to

$$\phi(x,y) = \phi_1 - \frac{t_1}{2}\left(\frac{x}{R}\right)^2 + \frac{t_1}{24}\left[5+3t_1^2\right]\left(\frac{x}{R}\right)^4 + \cdots, \tag{4.59}$$

where $m(\phi_1) = R\phi_1 = y$, in agreement with equation (3.62).

4.4 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 4.3 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) $\psi(0,y)$ is prescribed. Therefore

$$\zeta(z) = \lambda(x,y) + i\psi(x,y), \tag{4.60}$$

$$\lambda(0,y) = 0, \tag{4.61}$$

$$\psi(0,y) = \overline{M}(y), \tag{4.62}$$

where $\overline{M}(y)$ is an inverse to $M(\psi)$ in the sense that $M\big(\overline{M}(y)\big) = y$ and $\overline{M}\left(M(\psi)\right) = \psi$. The Taylor series analogous to (4.33) is then an expansion of $\zeta(z)$ about a point on the $z_0 = iy_0$ on the y-axis of the z-plane:

$$\zeta(z) = \zeta_0 + (z-z_0)\overline{M}_0' - \frac{i}{2!}(z-z_0)^2\overline{M}_0'' - \frac{1}{3!}(z-z_0)^3\overline{M}_0''' + \frac{i}{4!}(z-z_0)^4\overline{M}_0'''' + \cdots, \tag{4.63}$$

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

$$y = M(\overline{M}(y)) = M(\psi), \tag{4.64}$$

$$\psi = \overline{M}(M(\psi)) = \overline{M}(y), \tag{4.65}$$

to give

$$\frac{dy}{d\psi} = M'(\psi), \qquad \frac{d\psi}{dy} = \overline{M}'(y). \tag{4.66}$$

Therefore, since $M'(\psi) \neq 0$ (4.10) we have

$$\overline{M}'(y) = \frac{d\psi}{dy} = \frac{1}{M'(\psi)}, \tag{4.67}$$

and in general

$$\frac{d(\,)}{dy} = \frac{1}{M'(\psi)}\frac{d(\,)}{d\psi}. \tag{4.68}$$

We now calculate all the derivatives in equation (4.63). 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 (4.46) we find

$$\begin{aligned}
\overline{M}'(y) &= \frac{1}{M'} && = \frac{1}{M'},\\
\overline{M}''(y) &= \frac{1}{M'}\frac{d}{d\psi}\left[\overline{M}'(y)\right] = \frac{1}{M'}\frac{d}{d\psi}\left[\frac{1}{M'}\right] = -\frac{M''}{(M')^3} && = \frac{-ip_2}{M'^2},\\
\overline{M}'''(y) &= \frac{1}{M'}\frac{d}{d\psi}\left[\overline{M}''(y)\right] = \frac{1}{M'}\frac{d}{d\psi}\left[\frac{-M''}{(M')^3}\right] = -\frac{M'''}{(M')^4} + 3\frac{(M'')^2}{(M')^5} && = \frac{p_3}{M'^3},\\
\overline{M}''''(y) &= \frac{1}{M'}\frac{d}{d\psi}\left[\overline{M}'''(y)\right] = -\frac{M''''}{(M')^5} + 10\frac{M''M'''}{(M')^6} - 15\frac{(M'')^3}{(M')^7} && = \frac{ip_4}{M'^4}.
\end{aligned} \tag{4.69}$$

Substituting these derivatives (evaluated at y_0 for $\overline{M}$ and at ϕ_0 for M) into equation (4.63) gives a complex series identical to that of (4.45) 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.

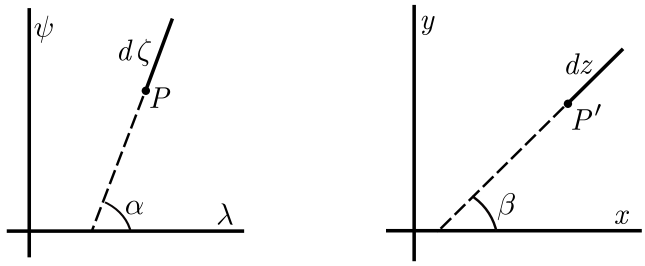


Figure 4.7

4.5 Scale and convergence in TMS

Units: We set R=1 throughout this section.

Consider an arbitrary conformal transformation described by the analytic function $z(\zeta)$. The scale factor m is the ratio of the lengths of small elements dz and $d\zeta$:

$$m = \frac{|dz|}{|d\zeta|} = |z'(\zeta)|, \qquad \frac{1}{m} = \frac{|d\zeta|}{|dz|} = |\zeta'(z)|. \tag{4.70}$$

The complex displacements may be written as

$$d\zeta = |d\zeta| \exp i\alpha, \qquad dz = |dz| \exp i\beta, \tag{4.71}$$

so that the angle of rotation between the elements is $\beta - \alpha = \arg z'(\zeta)$. This result apples to any small element but in particular it applies to an element along the ψ-axis in the ζ-plane. If that element coincides with a meridian, as it does for NMS, then the corresponding angle in the projection z-plane defines the angle between the projected meridian and the y-axis. This is just the convergence as defined in Section 3.6. Therefore

$$\gamma = \arg z'(\zeta) = -\arg \zeta'(z). \tag{4.72}$$

The derivative $z'(\zeta)$ determines both the scale factor and the convergence.

The freedom to choose the derivatives of an analytic function in any direction allows us to take them along the real axis. Therefore

$$z'(\zeta) = x_\lambda + i y_\lambda, \qquad \zeta'(z) = \lambda_x + i\psi_x. \tag{4.73}$$

Since x and y are functions of λ and ψ, and vice-versa, we have

$$m(\lambda,\psi) = \sqrt{x_\lambda^2 + y_\lambda^2}, \qquad \frac{1}{m(x,y)} = \sqrt{\lambda_x^2 + \psi_x^2}, \tag{4.74}$$

and

$$\tan\gamma(\lambda,\psi) = \frac{y_\lambda}{x_\lambda}, \qquad \tan\gamma(x,y) = -\frac{\psi_x}{\lambda_x}. \tag{4.75}$$

Note that

$$m(\lambda,\psi) = x_\lambda \sec\gamma(\lambda,\psi), \qquad \frac{1}{m(x,y)} = \lambda_x \sec\gamma(x,y). \tag{4.76}$$

Exact formulae for NMS to TMS scale factor and convergence

The real and imaginary parts of $z(\zeta) = \mathrm{gd}^{-1}\zeta$ are given in (4.20). Setting R=1 for clarity,

$$\tanh x = \sin\lambda\,\mathrm{sech}\,\psi, \qquad \tan y = \sec\lambda\,\sinh\psi, \tag{4.77}$$
$$(\mathrm{sech}^2 x)\,x_\lambda = \cos\lambda\,\mathrm{sech}\,\psi, \qquad (\sec^2 y)\,y_\lambda = \sin\lambda\,\sec^2\lambda\,\sinh\psi, \tag{4.78}$$
$$\tan\lambda = \sinh x\sec y \qquad \tanh\psi = \mathrm{sech}\,x\sin y \tag{4.79}$$
$$(\sec^2\lambda)\lambda_x = \cosh x\sec y \qquad (\mathrm{sech}^2\,\psi)\,\psi_x = -\sinh x\,\mathrm{sech}^2 x\sin y. \tag{4.80}$$

Simplify using

$$\mathrm{sech}^2 x = 1 - \sin^2\lambda\,\mathrm{sech}^2\psi = \mathrm{sech}^2\psi\left[\cosh^2\psi - \sin^2\lambda\right], \tag{4.81}$$
$$\sec^2 y = 1 + \sec^2\lambda\,\sinh^2\psi = \sec^2\lambda\left[\cosh^2\psi - \sin^2\lambda\right], \tag{4.82}$$
$$\sec^2\lambda = 1 + \sinh^2 x\sec^2 y = \sec^2 y\,[\cosh^2 x - \sin^2 y], \tag{4.83}$$
$$\mathrm{sech}^2\,\psi = 1 - \mathrm{sech}^2 x\sin^2 y = \mathrm{sech}^2 x\,[\cosh^2 x - \sin^2 y]. \tag{4.84}$$

The scale factors and convergence for the transformation $z(\zeta)$ are therefore.

$$m(\lambda,\psi) = \sqrt{x_\lambda^2 + y_\lambda^2} = \frac{1}{\sqrt{\cosh^2\psi - \sin^2\lambda}} \tag{4.85}$$
$$m(x,y) = \frac{1}{\sqrt{\lambda_x^2 + \psi_x^2}} = \sqrt{\cosh^2 x - \sin^2 y}, \tag{4.86}$$
$$\tan\gamma(\lambda,\psi) = \frac{y_\lambda}{x_\lambda} = \tan\lambda\tanh\psi, \tag{4.87}$$
$$\tan\gamma(x,y) = -\frac{\psi_x}{\lambda_x} = \tanh x\tan y. \tag{4.88}$$

Exact formulae for sphere to TMS scale factor and convergence

The scale factor from the sphere to TMS is the product of the scale factor from the sphere to NMS and that for transformation $z(\zeta)$ from NMS to TMS. Using the NMS scale factors (2.61), (2.62) and equations (2.55), (4.79)b to express $\cosh\psi$ in terms of ϕ and (x,y) respectively

$$k(\lambda,\phi) = \sec\phi\, m(\lambda,\psi) = \frac{\sec\phi}{\sqrt{\cosh^2\psi - \sin^2\lambda}} = \frac{1}{\sqrt{1 - \sin^2\lambda\cos^2\phi}} \tag{4.89}$$
$$k(x,y) = \cosh\psi(x,y)\,m(x,y) = \frac{\sqrt{\cosh^2 x - \sin^2 y}}{\sqrt{1 - \mathrm{sech}^2 x\sin^2 y}} = \cosh x, \tag{4.90}$$

Convergence is additive but from sphere to NMS it is zero. Therefore equations 4.87, 4.88 give the sphere to TMS convergence. Setting $\tanh\psi = \sin\phi$ in 4.87

$$\gamma(\lambda,\phi) = \tan^{-1}[\tan\lambda\sin\phi], \tag{4.91}$$
$$\gamma(x,y) = \tan^{-1}[\tanh x\tan y]. \tag{4.92}$$

These results are in agreement with Equations 3.72–3.75 (after restoring R).

Direct series for NMS to TMS scale factor and convergence

We work from the series solutions (4.37) and (4.38) in which we neglected terms of $O(\lambda^5)$. Therefore we must neglect terms of $O(\lambda^4)$ in the expressions for the derivatives, (4.41) and (4.42), and any expressions derived from them. Inserting these derivatives into equation (4.87) gives (R=1, s=$\sin\phi$ *etc.*)

$$\tan\gamma = \frac{y_\lambda}{x_\lambda} = \frac{sc\lambda + \frac{1}{6}sc^3[5-t^2]\lambda^3 + \cdots}{c + \frac{1}{2}c^3[1-t^2]\lambda^2\cdots} \tag{4.93}$$

$$= s\lambda\left(1 + \frac{1}{3}[1+t^2]c^2\lambda^2 + \cdots\right). \tag{4.94}$$

Using the series for arctan (equation E.20),

$$\gamma = \tan^{-1}(\tan\gamma) = \tan\gamma - \frac{1}{3}(\tan\gamma)^3 + \cdots \tag{4.95}$$

$$= s\lambda\left(1 + \frac{1}{3}c^2\lambda^2 + \cdots\right). \tag{4.96}$$

To the same order

$$\sec\gamma = \sqrt{1+\tan^2\gamma} = 1 + \frac{1}{2}\tan^2\gamma + \cdots = 1 + \frac{1}{2}s^2\lambda^2 + \cdots. \tag{4.97}$$

We calculate the NMS to TMS scale factor by using (4.76) rather than (4.74). (The method is easier for the more involved series for TME.)

$$m(\lambda,\phi) = x_\lambda\sec\gamma = c\left(1 + \frac{1}{2}c^2\lambda^2 + \cdots\right) \tag{4.98}$$

Direct series for sphere to TMS scale factor and convergence

The sphere to TMS scale factor is given by multiplying this result by the sphere to NMS scale factor, $\sec\phi = 1/c$.

$$k(\lambda,\phi) = \frac{1}{c}m(\lambda,\phi) = \left(1 + \frac{1}{2}c^2\lambda^2 + \cdots\right) \tag{4.99}$$

in agreement with the leading terms of the expansion in equation 3.74. The series (4.96) for the convergence is unchanged:

$$\gamma(\lambda,\phi) = s\lambda\left(1 + \frac{1}{3}c^2\lambda^2 + \cdots\right). \tag{4.100}$$

where s=$\sin\phi$ *etc.* This result in agreement with equation 3.72, neglecting $O(\lambda^4)$.

Inverse series for NMS to TMS scale factor and convergence

We neglected terms of $O(x^5)$ in the series for the inverse transformation, (4.51) and (4.52). Therefore we must neglect terms of $O(x^4)$ in the expressions for the derivatives, (4.53) and (4.54) and any expressions derived from them. Inserting the derivatives into equation (4.87) gives (R=1, $s_1 = \sin\phi_1$ *etc.*)

$$\tan\gamma = -\frac{\psi_x}{\lambda_x} = \frac{\dfrac{t_1}{c_1}x - \dfrac{t_1}{6c_1}\left[5+6t_1^2\right]x^3 + \cdots}{\dfrac{1}{c_1} - \dfrac{1}{2c_1}[1+2t_1^2]x^2 + \cdots} \tag{4.101}$$

$$= t_1 x\left(1 - \frac{1}{3}x^2 + \cdots\right). \tag{4.102}$$

Using the series for arctan (equation E.20),

$$\gamma = \tan^{-1}(\tan\gamma) = \tan\gamma - \frac{1}{3}(\tan\gamma)^3 + \cdots \tag{4.103}$$

$$= t_1 x\left(1 - \frac{1}{3}[1+t_1^2]x^2 + \cdots\right). \tag{4.104}$$

To the same order

$$\sec\gamma = \sqrt{1+\tan^2\gamma} = 1 + \frac{1}{2}\tan^2\gamma + \cdots = 1 + \frac{1}{2}t_1^2 x^2 + \cdots. \tag{4.105}$$

Therefore the NMS to TMS scale factor expressed in projection coordinates is given by

$$\frac{1}{m(x,y)} = \lambda_x \sec\gamma = \frac{1}{c_1}\left(1 - \frac{1}{2}[1+t_1^2]x^2 + \cdots\right), \tag{4.106}$$

$$m(x,y) = \frac{1}{\lambda_x \sec\gamma} = c_1\left(1 + \frac{1}{2}[1+t_1^2]x^2 + \cdots\right). \tag{4.107}$$

Inverse series for sphere to TMS scale factor and convergence

For the TMS scale factor we must multiply the above result by the sphere to NMS scale factor, $\cosh\psi$ in (2.62). We can no longer simply use (4.79)b to express this factor in terms of x and y: we have only the series for ψ at (4.52) with R=1:

$$\psi(x,y) = \psi_1 - \frac{1}{2!}\frac{t_1}{c_1}x^2 + \frac{1}{4!}\frac{t_1}{c_1}\left[5+6t_1^2\right]x^4 + \cdots, \tag{4.108}$$

We then substitute $(\psi - \psi_1)$ in a Taylor series expansion of $\cosh\psi$ about ψ_1. Since the derivatives of cosh are simply sinh, cosh, ... alternating, we have

$$\cosh\psi = \cosh\psi_1 + \frac{1}{1!}\sinh\psi_1(\psi-\psi_1) + \frac{1}{2!}\cosh\psi_1(\psi-\psi_1)^2 + \cdots. \tag{4.109}$$

Substitute $\psi-\psi_1$ from 4.108 and neglect terms of order $O(x^4)$ and, at the same time, set $\cosh\psi_1 = \sec\phi_1$ and $\sinh\psi_1 = \tan\phi_1$ from (2.55):

$$\cosh\psi = \frac{1}{c_1}\left(1-\frac{1}{2}t_1^2\left(\frac{x}{R}\right)^2+\cdots\right) \tag{4.110}$$

Therefore the sphere to TMS scale factor is

$$k(x,y) = \cosh\psi\, m(x,y) = \left(1+\frac{1}{2}x^2+\cdots\right), \tag{4.111}$$

in agreement with the two leading terms of equation 3.75 (after restoring R) which are just the leading terms of the expansion of $\cosh x$. This result is of course trivial, but we it demonstrates the method we shall follow for the TME series.

The series (4.96) for the convergence is unchanged:

$$\gamma(x,y) = t_1 x\left(1-\frac{1}{3}(1+t_1^2)x^2+\cdots\right). \tag{4.112}$$

where $t_1{=}\tan\phi_1$ and $m(\phi_1) = y$. This result is in agreement with equation 3.73.

Chapter 5

The geometry of the ellipsoid

Abstract

Geodetic and geocentric latitude. Parameters of the ellipsoid. Relation of Cartesian and geographical coordinates. Reduced or parametric latitude. Curvature. Distances on the ellipsoid. Meridian distance and its inverse. Auxiliary latitudes: conformal, rectifying and authalic.

5.1 Coordinates on the ellipsoid

The Earth is more accurately modelled as an oblate ellipsoid of revolution. If the symmetry axis is taken as OZ the Cartesian equation with respect to its centre is

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

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

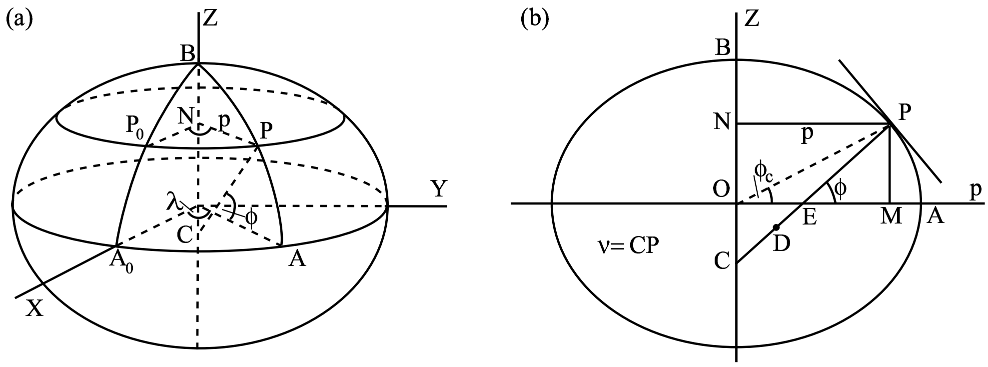


Figure 5.1

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

5.2 The parameters of the ellipsoid

The parameter a is the **equatorial radius** and the parameter b is the distance from centre to pole. The latter is often called the **polar radius**: this is misleading for there is no circle involved. These parameters are the **major and minor semi-axes** of a meridional ellipse defined by any meridian and its continuation over the poles. Instead of using (a,b) as the basic parameters of the ellipse we can use the combination (a,e) where e is the **eccentricity**, or (a,f) where f is the (first) **flattening**. These parameters are defined and related by

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

For numerical examples we use the values for the WGS84 ellipsoid:

$$\begin{aligned} a &= 6378137.0\text{m}, & e &= 0.0818191908, & f &= 0.003352810, \\ b &= 6356752.314\text{m}, & e^2 &= 0.0066943799, & \frac{1}{f} &= 299.3249753. \end{aligned} \tag{5.4}$$

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 an ellipsoid with equatorial radius a would be about the width of one of the lines in the figure. The ellipses shown here, and elsewhere, are greatly exaggerated.

Other parameters used to describe an ellipse

Several other small parameters arise naturally in the study of the ellipse. We shall need the second eccentricity, e', and the **third flattening** n (rarely e_1). They are defined by

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

(The second flattening, defined as $(a-b)/b$ is rarely used). There are many 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 any meridian 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

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

Eliminating Z from equations 5.10 and 5.12 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 $\nu = CP = PN\sec\phi = p\sec\phi$ 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

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

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

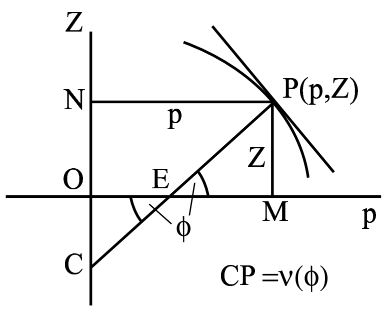


Figure 5.2

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 geodetic and geocentric latitudes

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

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

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 WGS ellipsoid (equation 5.4) shows that the maximum difference occurs at $\phi \approx 45^\circ.095$, for which $\phi_c \approx 44^\circ.904$ 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 (the Mercator parameter in the notation of Sections 2.4 and 6.1) and we shall meet four further latitude definitions. These are the reduced (or parametric) latitude U (Section 5.5), and, in Section 5.10, three auxiliary latitudes: the rectifying latitude μ, the conformal latitude χ and the authalic latitude ξ. 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 (where it is zero) 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

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

The two dimensional coordinate system describing points on the surface may be extended to a three dimensional coordinate system. Let H be a point 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 (5.19), we have $EP = CP - CE = \nu(1-e^2)$. The coordinates of H are

$$X(\phi) = \big(\nu(\phi) + h\big)\cos\phi\cos\lambda, \tag{5.27}$$
$$Y(\phi) = \big(\nu(\phi) + h\big)\cos\phi\sin\lambda, \tag{5.28}$$
$$Z(\phi) = \big((1-e^2)\,\nu(\phi) + h\big)\sin\phi. \tag{5.29}$$

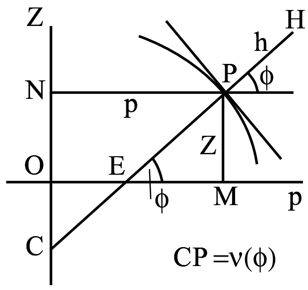


Figure 5.3

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):

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

The formulae developed in this section are used in the OSGB (1999) publications.

5.5 The reduced or parametric 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. 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

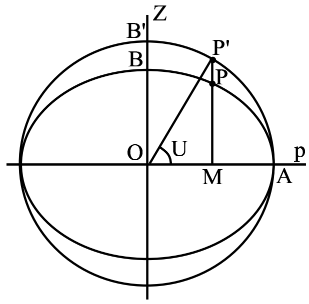


Figure 5.4

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

The pair of equations

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

constitutes the required parametric representation of the ellipse. At a general point the ordinates $OP = Z = b\sin U$ and $OP' = a\sin U$ are in the ratio of b/a: the ellipse is a uniformly squashed circle.

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 $b = a\sqrt{1-e^2}$)

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

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

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

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

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 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 the meridian

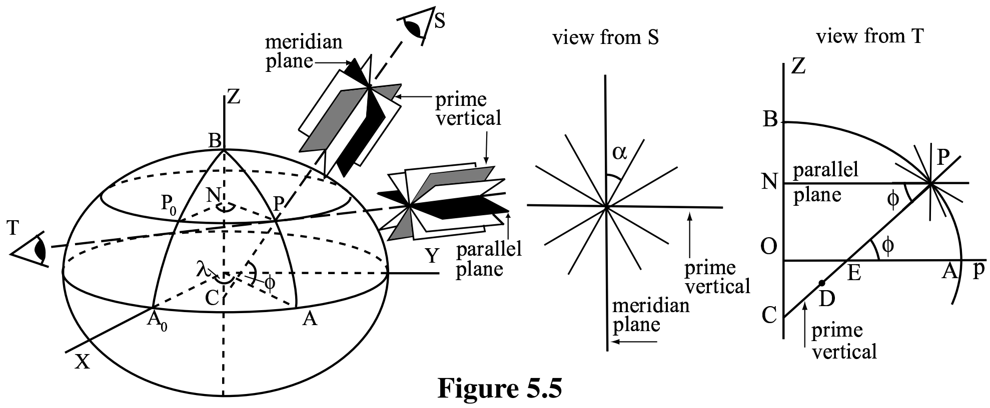


Figure 5.5

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 $\rho \le \nu$ and therefore $\rho^{-1} \ge \nu^{-1}$ 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.

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

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

Themetric may be written as

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

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

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

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

Finite distances on parallel and meridian

Equations (5.63) and (5.64) may be integrated to give

$$s_{\text{parallel}} = \int_{\lambda_1}^{\lambda_2} ds_{\text{parallel}} = \int_{\lambda_1}^{\lambda_2} \nu(\phi)\cos\phi\, d\lambda = \nu(\phi)\cos\phi\,\big(\lambda_2-\lambda_1\big), \tag{5.65}$$

$$s_{\text{meridian}} = \int_{\phi_1}^{\phi_2} ds_{\text{meridian}} = \int_{\phi_1}^{\phi_2} \rho(\phi) d\phi = a(1-e^2)\int_{\phi_1}^{\phi_2} \frac{d\phi}{\big(1-e^2\sin^2\phi\big)^{3/2}}. \tag{5.66}$$

The meridian integral is related to a special case of an incomplete elliptic integral of the third kind. In the notation of the NIST handbook (Olver *et al.*, 2010) at Section 19.2(ii)) it is equal to $\Pi(\phi_2,e^2,e)-\Pi(\phi_1,e^2,e)$. Since the evaluation of such elliptic integrals requires series computation we prefer to evaluate the integral from first principles: see Section 5.8.

Geodesics

A geodesic on the ellipsoid is the shortest distance between two points.

The direct problem. Given a geodesic starting at $P_1(\phi_1,\lambda_1)$ with an azimuth α_1 find the coordinates of the point $P_2(\phi_2,\lambda_2)$ at a distance s measured along the geodesic; find also the azimuth α_2 of the geodesic at P_2.

The inverse problem. Given the points $P_1(\phi_1,\lambda_1)$ and $P_2(\phi_2,\lambda_2)$ find s, the geodesic distance between them, and the azimuths α_1, α_2 at the end points.

These problems were first solved by Bessel (1825)) and implementations of his solution were given by Vincenty (1976) but without any background theory. More recently Karney (2012) has published a full outline of the Bessel theory together with more accurate algorithms. There is no space to describe those methods here.

The infinitesimal element on the ellipsoid

The infinitesimal element on the sphere was discussed in Section 2.1, Figure 2.5. From equations (5.63) and (5.64) we see that the infinitesimal element on the ellipsoid is approximated by a planar rectangular quadrilateral with sides of length $\rho\,\delta\phi$ on the meridians and $\nu\cos\phi\,\delta\lambda$ on a parallel, (Figure 5.6).

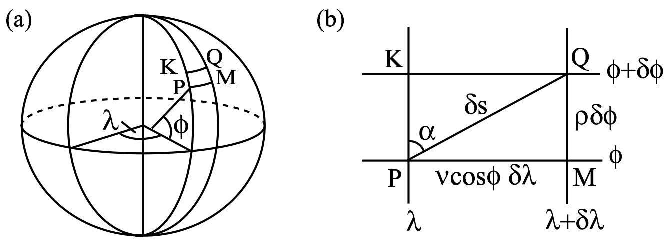


Figure 5.6

5.8 The meridian distance on the ellipsoid

On the sphere the meridian distance was simply $m(\phi) = a\phi$. On the ellipsoid we use the same notation but the definition follows from equation (5.66):

$$m(\phi) = \int_0^\phi ds_{\text{meridian}} = \int_0^\phi \rho(\phi)d\phi = a(1-e^2)\int_0^\phi \frac{d\phi}{\left(1-e^2\sin^2\phi\right)^{3/2}}. \tag{5.67}$$

Following Delambre (1799), expand the denominator by the binomial theorem, giving

$$m(\phi) = a(1-e^2)\int_0^\phi \left(1 + b_2e^2s^2 + b_4e^4s^4 + b_6e^6s^6 + b_8e^8s^8 + \cdots\right) d\phi, \tag{5.68}$$

where we have set $s = \sin\phi$. Using (E.30) we find

$$b_2 = \frac{3}{2}, \qquad b_4 = \frac{15}{8}, \qquad b_6 = \frac{35}{16}, \qquad b_8 = \frac{315}{128}. \tag{5.69}$$

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

$$\boxed{m(\phi) = A_0\phi + A_2\sin 2\phi + A_4\sin 4\phi + A_6\sin 6\phi + A_8\sin 8\phi + \cdots,} \tag{5.70}$$

where the coefficients are given by

$$\begin{aligned}
A_0 &= a(1-e^2)\left(1+\frac{b_2e^2}{2}+\frac{3b_4e^4}{8}+\frac{5b_6e^6}{16}+\frac{35b_8e^8}{128}\right) = a\left(1-\frac{e^2}{4}-\frac{3e^4}{64}-\frac{5e^6}{256}-\frac{175e^8}{16*1024}\right)\\
A_2 &= \frac{a(1-e^2)}{2}\left(-\frac{b_2e^2}{2}-\frac{b_4e^4}{2}-\frac{15b_6e^6}{32}-\frac{7b_8e^8}{16}\right) = a\left(-\frac{3e^2}{8}-\frac{3e^4}{32}-\frac{45e^6}{1024}-\frac{420e^8}{16*1024}\right)\\
A_4 &= \frac{a(1-e^2)}{4}\left(\frac{b_4e^4}{8}+\frac{3b_6e^6}{16}+\frac{7b_8e^8}{32}\right) = a\left(\frac{15e^4}{256}+\frac{45e^6}{1024}+\frac{525e^8}{16*1024}\right),\\
A_6 &= \frac{a(1-e^2)}{6}\left(-\frac{b_6e^6}{32}-\frac{b_8e^8}{16}\right) = a\left(-\frac{35e^6}{3072}-\frac{175e^8}{12*1024}\right),\\
A_8 &= \frac{a(1-e^2)}{8}\left(\frac{b_8e^8}{128}\right) = a\left(\frac{315e^8}{128*1024}\right).
\end{aligned} \tag{5.71}$$

If we use the numerical values for the WGS84 ellipsoid (5.4), then, in metres,

$$m(\phi) = 6367449{\cdot}146\phi - 16038{\cdot}509\sin 2\phi + 16{\cdot}833\sin 4\phi - 0{\cdot}022\sin 6\phi + 0{\cdot}00003\sin 8\phi \tag{5.72}$$

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 can usually drop $O(e^8)$ terms in expressions for the meridian distance . Note that for the first term ϕ must be expressed in radians.

Meridian distance II: expansion in flattening

There are other ways of obtaining a series expansion. For example, the OSGB (1999) transverse Mercator formulae use an expansion in terms of the flattening parameter, n. Following Bessel (1825) we use the relations between a, b, e, n given in Section 5.2 to write the meridian distance as

$$m(\phi) = b(1-n)(1+n)^2 \int_0^\phi \frac{d\phi}{\left(1+2n\cos 2\phi + n^2\right)^{3/2}}. \tag{5.73}$$

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 $O(n^3)$,

$$\begin{aligned}
&\left(1+2n\cos 2\phi + n^2\right)^{-3/2} \\
&\quad = (1+nz)^{-3/2}\left(1+nz^{-1}\right)^{-3/2} \\
&\quad = \left(1+a_1nz+a_2n^2z^2+a_3n^3z^3+\right)\left(1+a_1nz^{-1}+a_2n^2z^{-2}+a_3n^3z^{-3}\right) \\
&\quad = 1+a_1^2n^2+\left(a_1n+a_1a_2n^3\right)\left[z+\frac{1}{z}\right]+a_2n^2\left[z^2+\frac{1}{z^2}\right]+a_3n^3\left[z^3+\frac{1}{z^3}\right]+O(n^4)
\end{aligned}$$

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

$$a_1 = -\frac{3}{2}, \qquad a_2 = \frac{15}{8}, \qquad a_3 = -\frac{35}{16}. \tag{5.74}$$

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

$$\begin{aligned}
&\int_1^z \frac{dz}{2iz}\left(1+a_1^2n^2+\left(a_1n+a_1a_2n^3\right)\left[z+\frac{1}{z}\right]+a_2n^2\left[z^2+\frac{1}{z^2}\right]+a_3n^3\left[z^3+\frac{1}{z^3}\right]\right) \\
&= \frac{1}{2i}\left(\left(1+a_1^2n^2\right)\ln z+\left(a_1n+a_1a_2n^3\right)\left[z-\frac{1}{z}\right]+\frac{a_2n^2}{2}\left[z^2-\frac{1}{z^2}\right]+\frac{a_3n^3}{3}\left[z^3-\frac{1}{z^3}\right]\right)\Bigg|_1^z
\end{aligned}$$

All terms vanish at the lower limit. 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

$$m(\phi) = B_0\phi + B_2\sin 2\phi + B_4\sin 4\phi + B_6\sin 6\phi + \cdots, \tag{5.75}$$

where the coefficients are given to order n^3 by

$$\begin{aligned}
B_0 &= b(1-n)(1+n)^2\left(1+\frac{9}{4}n^2\right) &&= b\left(1+n+\frac{5}{4}n^2+\frac{5}{4}n^3\right), \\
B_2 &= b(1-n)(1+n)^2\left(-\frac{3n}{2}-\frac{45n^3}{16}\right) &&= -b\left(\frac{3}{2}n+\frac{3}{2}n^2+\frac{21}{16}n^3\right), \\
B_4 &= b(1-n)(1+n)^2\left(\frac{15n^2}{16}\right) &&= b\left(\frac{15}{16}n^2+\frac{15}{16}n^3\right), \\
B_6 &= b(1-n)(1+n)^2\left(-\frac{35n^3}{48}\right) &&= -b\left(\frac{35}{48}n^3\right).
\end{aligned} \tag{5.76}$$

Meridian distance from a reference latitude

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

$$\begin{aligned}\Delta m &= m(\phi) - m(\phi_0)\\ &= B_0(\phi-\phi_0) + B_2(\sin 2\phi - \sin 2\phi_0) + B_4(\sin 4\phi - \sin 4\phi_0) + B_6(\sin 6\phi - \sin 6\phi_0)\\ &= B_0(\phi-\phi_0) + 2B_2 \sin(\phi-\phi_0)\cos(\phi+\phi_0) + 2B_4 \sin 2(\phi-\phi_0)\cos 2(\phi+\phi_0)\\ &\qquad + 2B_6 \sin 3(\phi-\phi_0)\cos 3(\phi+\phi_0) + \cdots .\end{aligned} \tag{5.77}$$

with the coefficients given by (5.76). This is the expression used by the OSGB (1999).

Meridian distance III: Helmert's formula

Helmert (1880) writes the meridian distance as:

$$m(\phi) = \frac{a}{1+n}\left(h_0\,\phi + h_2 \sin 2\phi + h_4 \sin 4\phi + h_6 \sin 6\phi + h_8 \sin 8\phi + \cdots\right). \tag{5.78}$$

Comparing this with the series 5.75 shows that $h_n = (1+n)B_n/a$. If the series is extended to include $O(n^4)$ terms (exercise for the reader, or use the maxima code in Appendix H.3) the coefficients may be evaluated directly from (5.76). Use equation 5.6: $(1+n)b/a = (1-n)$.

$$\begin{aligned} h_0 &= 1 + \frac{n^2}{4} + \frac{n^4}{64},\\ h_2 &= -\frac{3n}{2} + \frac{3n^3}{16},\\ h_4 &= \frac{15n^2}{16} - \frac{15n^4}{64},\\ h_6 &= -\frac{35n^3}{48},\\ h_8 &= \frac{315n^4}{512}. \end{aligned} \tag{5.79}$$

Note that these coefficients for $m(\phi)$ achieve the same accuracy with far less terms than the previous expressions, (5.71) and (5.76).

The polar distance

The distance from equator to pole may be obtained from any of the above series. It is defined by

$$m_p = m(\pi/2) = \frac{1}{2}\pi A_0 = \frac{1}{2}\pi B_0 = \frac{1}{2}\pi\frac{ah_0}{1+n} = 10{,}001{,}965{\cdot}730 \text{ metres}. \tag{5.80}$$

where the numerical value is calculated for WGS84. The corresponding mean radius is defined to be $R = 2m_p/\pi = 6367449.146$m

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 $m_{\mathrm{sph}}(\phi_1) = a\phi_1 = y$ for a given point (x, y) on the projection. Trivially, $\phi_1 = y/a$. For the ellipsoid we must invert one of the series (5.70), (5.75) or (5.78) . We outline three methods.

Inverse meridian distance I: fixed point iteration

To solve $m(\phi) = y$ when $m(\phi)$ is given by one of the above series 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 $\phi_0 = y/a$. 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 $g'(\phi) \approx 1 - B_0/a = O(e^2) \ll 1$ the iteration will converge quickly.

Inverse meridian distance II: Newton Raphson

Solving $m(\phi) = y$ for a given y is equivalent to finding the zero of the function

$$q(\phi) \equiv -y + m(\phi). \tag{5.83}$$

Taking the series 5.78 as an example we have

$$q(\phi) = -y + \frac{a}{1+n}\left(h_0\,\phi + h_2 \sin 2\phi + h_4 \sin 4\phi + h_6 \sin 6\phi + \cdots\right) \tag{5.84}$$

$$q'(\phi) = \frac{a}{1+n}\left(h_0 + 2h_2 \cos 2\phi + 4h_4 \cos 4\phi + 6h_6 \cos 6\phi + \cdots\right). \tag{5.85}$$

The initial value may be taken as that for the spherical approximation, namely $\phi_0 = y/a$. The Newton-Raphson method then gives the required value of ϕ as the limit of the iteration:

$$\phi_{n+1} = \phi_n - \frac{q(\phi_n)}{q'(\phi_n)}, \qquad n = 0, 1, 2 \ldots. \tag{5.86}$$

Inverse meridian distance III: via the rectifying latitude

The **rectifying latitude**, μ, is a scaled version of the meridian distance (5.71, 5.76, 5.79):

$$\mu(\phi) = \frac{\pi}{2}\frac{m(\phi)}{m_p} = \frac{m(\phi)}{A_0} = \frac{m(\phi)}{B_0} = \frac{(1+n)m(\phi)}{ah_0}. \tag{5.87}$$

A given m determines μ and hence ϕ follows from the series given in Section 5.12.

5.10 Auxiliary latitudes double projections

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

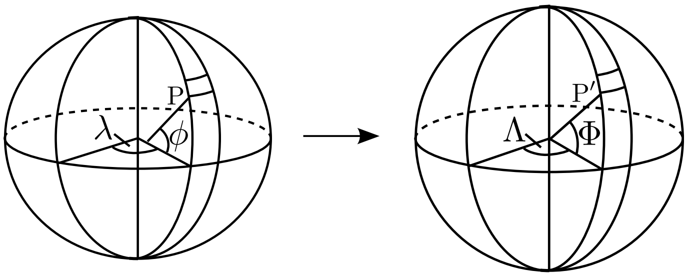


Figure 5.7

If we denote the latitude and longitude coordinates 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. For a double projection we would define coordinates (x, y) on the plane by specifying two further functions $x(\Phi, \Lambda)$ and $y(\Phi, \Lambda)$. Note that we have not yet specified the radius of the sphere.

We consider only **restricted projections** in which $\Lambda = \lambda$ and Φ is a function of ϕ only. This restricted set includes the main practical applications.

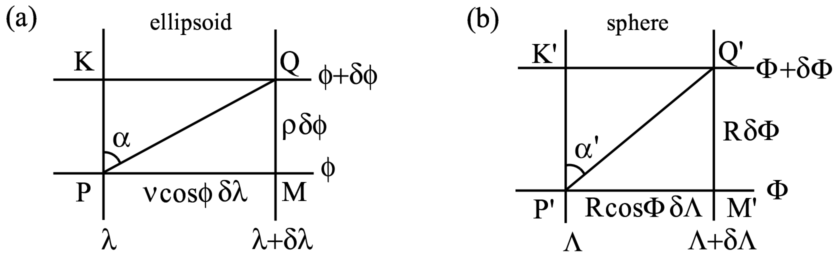


Figure 5.8

The basic properties of the projection from the ellipsoid to the sphere can be investigated by comparing the infinitesimal elements shown in Figure 5.8. In particular the geometry of the infinitesimal elements gives

$$\text{(a)} \quad \tan\alpha = \frac{\nu\cos\phi\,\delta\lambda}{\rho\,\delta\phi} \qquad \text{and} \qquad \text{(b)} \quad \tan\alpha' = \frac{R\cos\Phi\,\delta\Lambda}{R\delta\Phi}. \tag{5.88}$$

so that

$$\tan\alpha' = \frac{\cos\Phi}{\Phi'(\phi)}\,\frac{\rho}{\nu\cos\phi}\tan\alpha. \tag{5.89}$$

We shall also require the meridian scale factor:

$$h(\phi) = \frac{R\delta\Phi}{\rho\,\delta\phi} = \frac{R\Phi'(\phi)}{\rho}. \tag{5.90}$$

Conserving angles: 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 $\chi(\phi)$, (Snyder, 1987). There are many other nomenclatures in the literature. Beware particularly that Adams (1921) uses the same notation χ for the conformal latitude as defined here but he gives it the name 'isometric latitude'. In modern works isometric latitude refers to the term which we have called the Mercator parameter in Chapters 2 and 6.

Equation (5.29) shows that the projection is conformal, that is $\alpha{=}\alpha'$, if $\Phi = \chi$ satisfies the condition

$$\sec\chi\,\chi'(\phi) = \frac{\rho(\phi)\sec\phi}{\nu(\phi)}. \tag{5.91}$$

We defer the integration of this equation until the Section 5.11. Note that the conformality condition imposes no constraint on the R, the radius of the sphere. All of the following choices have been used: (1) the semi-major axis of the ellipsoid; (2) an arithmetic or geometric mean of the semi-axes; (3) the meridian radius of curvature, ρ, at a latitude where we seek the best fit; (4) the Gaussian radius of curvature, $\sqrt{\rho\nu}$, at a latitude where we seek the best fit; (5) radius of equal volume sphere; (6) radius of equal polar distance sphere (Section 5.12); (7) radius of equal area sphere (Section 5.13). See also Section 2.1.

Conserving meridian length: the rectifying latitude

The function of $\Phi(\phi)$ which preserves meridian length is called the **rectifying latitude**, for which we use the notation $\mu(\phi)$. In Figure 5.8 $PK = PK'$ so that

$$\rho\, d\phi = R\, d\mu. \tag{5.92}$$

This integrates immediately (see 5.67) to give

$$\mu(\phi) = \frac{1}{R}\int_0^{\phi} \rho(\phi)\,d\phi = \frac{m(\phi)}{R} = \frac{\pi}{2}\frac{m(\phi)}{m_p} \tag{5.93}$$

where we have imposed $\mu(\pi/2) = \pi/2$ and set $m(\pi/2) = m_p = \pi R/2$. The projection must be made to a sphere with this **rectifying radius**. Further details are in Section 5.12

Conserving area: the authalic latitude

The function of $\Phi(\phi)$ which preserves area is called the **authalic latitude** (Greek for 'same area'), for which we use the notation $\xi(\phi)$. In Figure 5.8 gives

$$\delta A_{\text{ellipsoid}} = \nu\cos\phi\,\delta\lambda\cdot\rho\,\delta\phi = R\cos\xi\,\delta\Lambda\cdot R\delta\xi = \delta A_{\text{sphere}}, \tag{5.94}$$

so that

$$\cos\xi\,\frac{d\xi}{d\phi} = \frac{\rho\nu\cos\phi}{R^2}, \tag{5.95}$$

We defer the evaluation of this integral and its properties until Section 5.13 but we note here that R must be chosen appropriately if the total areas of ellipsoid and sphere are the same.

5.11 The conformal latitude

The defining equation, (5.91), integrates to

$$\int_0^{\chi(\phi)} \sec\chi \, d\chi = \int_0^{\phi} \frac{\rho(\phi)\sec\phi}{\nu(\phi)} d\phi. \tag{5.96}$$

The integral on the left is the same as that for the Mercator parameter on the sphere, equation (2.26), so that

$$\text{LHS} = \ln\left[\tan\left(\frac{\chi(\phi)}{2}+\frac{\pi}{4}\right)\right] = \tanh^{-1}\sin\chi = \text{gd}^{-1}\chi. \tag{5.97}$$

For the integral on the right we substitute for the functions ν and ρ from equations (5.16):

$$\text{RHS} = \int_0^{\phi} \frac{(1-e^2)}{\cos\phi}\frac{1}{1-e^2\sin^2\phi} d\phi. \tag{5.98}$$

Splitting the integrand into partial fractions and noting that the first term gives the same integral as (2.26), we

$$\begin{aligned}
\text{RHS} &= \int_0^{\phi}\left[\frac{1}{\cos\phi} - \frac{e^2\cos\phi}{2}\left(\frac{1}{1+e\sin\phi}+\frac{1}{1-e\sin\phi}\right)\right]d\phi \\
&= \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\right] - \frac{e}{2}\left[\ln\left(\frac{1+e\sin\phi}{1-e\sin\phi}\right)\right] && (5.99)\\
&= \tanh^{-1}\sin\phi - e\tanh^{-1}[e\sin\phi] && (5.100)\\
&= \text{gd}^{-1}\phi - e\tanh^{-1}[e\sin\phi] && (5.101)
\end{aligned}$$

Therefore

$$\begin{aligned}
\chi(\phi) &= 2\arctan\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right] - \frac{\pi}{2} && (5.102)\\
&= \sin^{-1}\left[\tanh\left(\tanh^{-1}\sin\phi - e\tanh^{-1}[e\sin\phi]\right)\right] && (5.103)\\
&= \text{gd}\left(\text{gd}^{-1}\phi - e\tanh^{-1}[e\sin\phi]\right). && (5.104)
\end{aligned}$$

This rather complicated transformation (along with $\Lambda = \lambda$) has been constructed to guarantee a conformal projection from ellipsoid to sphere. Note that equations (5.90), (5.91) and (5.102) show that the scale cannot be uniform on any meridian of the conformal 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. Such a double projection is not TME where we demand that the scale be constant on the central meridian.

Series for the conformal latitude

Equation (5.102) shows that χ and ϕ are equal at the equator and at the pole and elsewhere they differ by terms of order $O(e^2)$. They are related by Fourier series in $\sin 2k\phi$:

$$\chi(\phi) = \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots, \tag{5.105}$$

$$\phi(\chi) = \chi + d_2 \sin 2\chi + d_4 \sin 4\chi + d_6 \sin 6\chi + d_8 \sin 8\chi + \cdots. \tag{5.106}$$

Low order analytic derivations may be found on line in Adams (1921). Note that he uses 'isometric latitude' instead of 'conformal latitude' (although he does use the same notation). He gives six methods for the first series and seven for the second (pages 16–59). A more modern notation is used by Deakin (2010) (paper 6). Maxima code can also be found in Section H.4: this can generate as many terms as required. The results are as follows:

Conformal latitude from geodetic latitude

$$\chi(\phi) = \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots \tag{5.107}$$

$$\begin{aligned}
b_2 &= -\frac{e^2}{2} - \frac{5e^4}{24} - \frac{3e^6}{32} - \frac{281e^8}{5760} &&= -2n + \frac{2n^2}{3} + \frac{4n^3}{3} - \frac{82n^4}{45},\\
b_4 &= \frac{5e^4}{48} + \frac{7e^6}{80} + \frac{697e^8}{11520} &&= \frac{5n^2}{3} - \frac{16n^3}{15} - \frac{13n^4}{9},\\
b_6 &= -\frac{13e^6}{480} - \frac{461e^8}{13440} &&= -\frac{26n^3}{15} + \frac{34n^4}{21},\\
b_8 &= \frac{1237e^8}{161280} &&= \frac{1237n^4}{630}.
\end{aligned} \tag{5.108}$$

Geodetic latitude from conformal latitude

$$\phi(\chi) = \chi + d_2 \sin 2\chi + d_4 \sin 4\chi + d_6 \sin 6\chi + d_8 \sin 8\chi + \cdots \tag{5.109}$$

$$\begin{aligned}
d_2 &= \frac{e^2}{2} + \frac{5e^4}{24} + \frac{e^6}{12} + \frac{13e^8}{360} &&= 2n - \frac{2n^2}{3} - 2n^3 + \frac{116n^4}{45},\\
d_4 &= \frac{7e^4}{48} + \frac{29e^6}{240} + \frac{811e^8}{11520} &&= \frac{7n^2}{3} - \frac{8n^3}{5} - \frac{227n^4}{45},\\
d_6 &= \frac{7e^6}{120} + \frac{81e^8}{1120} &&= \frac{56n^3}{15} - \frac{136n^4}{35},\\
d_8 &= \frac{4279e^8}{161280} &&= \frac{4279n^4}{630}.
\end{aligned} \tag{5.110}$$

5.12 The rectifying latitude

The **rectifying latitude** was defined in equation 5.87 as a projection from the ellipsoid to a sphere of specific radius, namely $R = 2m_p/\pi$.

$$\mu(\phi) = \frac{\pi}{2}\frac{m(\phi)}{m_p} = \frac{m(\phi)}{A_0} = \frac{m(\phi)}{B_0} = \frac{(1+n)m(\phi)}{ah_0} \tag{5.111}$$

using the series (5.70), (5.75) or (5.78).

$$\mu(\phi) = \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots \tag{5.112}$$

where the coefficients $b_n = A_n/A_0 = B_n/B_0 = h_n/h_0$ are given in (5.71), (5.76) or (5.79). (See also Adams (1921), pages 123-128: he uses ω for the rectifying latitude.The series were confirmed by the Maxima code given in Section H.3. The results for the direct and inverse series are

Rectifying latitude from geodetic latitude

$$\mu(\phi) = \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots \tag{5.113}$$

$$\begin{aligned}
b_2 &= -\frac{3\,e^2}{8} - \frac{3\,e^4}{16} - \frac{111\,e^6}{1024} - \frac{141\,e^8}{2048} &&= -\frac{3\,n}{2} + \frac{9\,n^3}{16}, \\
b_4 &= \frac{15\,e^4}{256} + \frac{15\,e^6}{256} + \frac{405\,e^8}{8192} &&= \frac{15\,n^2}{16} - \frac{15\,n^4}{32}, \\
b_6 &= -\frac{35\,e^6}{3072} - \frac{35\,e^8}{2048} &&= -\frac{35\,n^3}{48}, \\
b_8 &= \frac{315\,e^8}{131072} &&= \frac{315\,n^4}{512}.
\end{aligned} \tag{5.114}$$

Geodetic latitude from rectifying latitude

$$\phi(\mu) = \mu + d_2 \sin 2\mu + d_4 \sin 4\mu + d_6 \sin 6\mu + d_8 \sin 8\mu + \cdots \tag{5.115}$$

$$\begin{aligned}
d_2 &= \frac{3\,e^2}{8} + \frac{3\,e^4}{16} + \frac{213\,e^6}{2048} + \frac{255\,e^8}{4096} &&= \frac{3\,n}{2} - \frac{27\,n^3}{32}, \\
d_4 &= \frac{21\,e^4}{256} + \frac{21\,e^6}{256} + \frac{533\,e^8}{8192} &&= \frac{21\,n^2}{16} - \frac{55\,n^4}{32}, \\
d_6 &= \frac{151\,e^6}{6144} + \frac{151\,e^8}{4096} &&= \frac{151\,n^3}{96}, \\
d_8 &= \frac{1097\,e^8}{131072} &&= \frac{1097\,n^4}{512}.
\end{aligned} \tag{5.116}$$

5.13 The authalic latitude

The definition of the authalic latitude was given in equation 5.95:

$$R^2 \cos\xi \frac{d\xi}{d\phi} = \rho\nu\cos\phi, \tag{5.117}$$

Substituting for the functions ν and ρ from equations (5.16), (5.45) and integrating

$$\int_0^\xi \cos\xi \, d\xi = \int_0^\phi \frac{\rho\nu\cos\phi}{R^2} d\phi, \tag{5.118}$$

$$\sin\xi = \frac{a^2(1-e^2)}{R^2} \int_0^\phi \frac{\cos\phi \, d\phi}{(1-e^2\sin^2\phi)^2} \tag{5.119}$$

$$\equiv \frac{a^2}{2R^2} q(\phi). \tag{5.120}$$

If we demand that $\xi = \pi/2$ when $\phi = \pi/2$ we must fix $R = a\sqrt{q_p/2}$ where $q_p = q(\pi/2)$: this defines the **authalic radius**.

The integral for $q(\phi)$ may be evaluated by the substitution $e\sin\phi = \tanh\alpha$:

$$\begin{aligned}
q(\phi) &= 2(1-e^2)\int_0^\phi \frac{\cos\phi \, d\phi}{(1-e^2\sin^2\phi)^2} \\
&= 2e^{-1}(1-e^2)\int_0^\alpha \cosh^2\alpha \, d\alpha \\
&= e^{-1}(1-e^2)\int_0^\alpha (\cosh 2\alpha + 1) d\alpha \\
&= e^{-1}(1-e^2)(\sinh\alpha\cos\alpha + \alpha) \\
&= e^{-1}(1-e^2)\left(\frac{e\sin\phi}{1-e^2\sin\phi} + \tanh^{-1}[e\sin\phi]\right)
\end{aligned} \tag{5.121}$$

The final result for the authalic latitude is therefore

$$\xi(\phi) = \sin^{-1}\left(\frac{q(\phi)}{q_p}\right), \tag{5.122}$$

$$q(\phi) = \frac{(1-e^2)\sin\phi}{1-e^2\sin^2\phi} + \frac{1-e^2}{e}\tanh^{-1}(e\sin\phi), \tag{5.123}$$

$$q_p = q(\pi/2) = 1 + \frac{1-e^2}{e}\tanh^{-1} e. \tag{5.124}$$

Area of an oblate ellipse

By construction, the oblate ellipsoid and the authalic sphere have the same area. Therefore

$$A_{\text{ellipsoid}} = 4\pi R^2 = 2\pi a^2 q_p = 2\pi a^2\left(1 + \frac{1-e^2}{e}\tanh^{-1} e\right). \tag{5.125}$$

NB: the result for the prolate ellipsoid is quite different. See mathworld.

Series for the authalic latitude

Once again, low order analytic derivations of the series may be found on line in Adams (1921), pages 60–83: he uses β for the authalic latitude. Maxima code is given in Section H.5: this can generate as many terms as required. The results are as follows:

Authalic latitude from geodetic latitude

$$\xi = \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots \tag{5.126}$$

$$\begin{aligned}
b_2 &= -\frac{e^2}{3} - \frac{31\,e^4}{180} - \frac{59\,e^6}{560} - \frac{42811\,e^8}{604800} &&= -\frac{4\,n}{3} - \frac{4\,n^2}{45} + \frac{88\,n^3}{315} + \frac{538\,n^4}{4725},\\
b_4 &= \frac{17\,e^4}{360} + \frac{61\,e^6}{1260} + \frac{76969\,e^8}{1814400} &&= \frac{34\,n^2}{45} + \frac{8\,n^3}{105} - \frac{2482\,n^4}{14175},\\
b_6 &= -\frac{383\,e^6}{45360} - \frac{3347\,e^8}{259200} &&= -\frac{1532\,n^3}{2835} - \frac{898\,n^4}{14175},\\
b_8 &= \frac{6007\,e^8}{3628800} &&= \frac{6007\,n^4}{14175}.
\end{aligned} \tag{5.127}$$

Geodetic latitude from authalic latitude

$$\phi = \xi + d_2 \sin 2\xi + d_4 \sin 4\xi + d_6 \sin 6\xi + d_8 \sin 8\xi + \cdots \tag{5.128}$$

$$\begin{aligned}
d_2 &= \frac{e^2}{3} + \frac{31\,e^4}{180} + \frac{517\,e^6}{5040} + \frac{120389\,e^8}{1814400} &&= \frac{4\,n}{3} + \frac{4\,n^2}{45} - \frac{16\,n^3}{35} - \frac{2582\,n^4}{14175},\\
d_4 &= \frac{23\,e^4}{360} + \frac{251\,e^6}{3780} + \frac{102287\,e^8}{1814400} &&= \frac{46\,n^2}{45} + \frac{152\,n^3}{945} - \frac{11966\,n^4}{14175},\\
d_6 &= \frac{761\,e^6}{45360} + \frac{47561\,e^8}{1814400} &&= \frac{3044\,n^3}{2835} + \frac{3802\,n^4}{14175},\\
d_8 &= \frac{6059\,e^8}{1209600} &&= \frac{6059\,n^4}{4725}.
\end{aligned} \tag{5.129}$$

Authalic radius

$$R_q = a\left(1 - \frac{e^2}{6} - \frac{17\,e^4}{360} - \frac{67\,e^6}{3024} - \frac{23123\,e^8}{1814400}\right) \tag{5.130}$$

$$= a\left(1 - \frac{2n}{3} + \frac{26n^2}{45} - \frac{374n^3}{945} + \frac{722n^4}{2025}\right). \tag{5.131}$$

5.14 Ellipsoid: summary

Equation: ellipsoid and cross-section

$$\frac{X^2}{a^2}+\frac{Y^2}{a^2}+\frac{Z^2}{b^2}=1, \qquad \frac{p^2}{a^2}+\frac{Z^2}{b^2}=1. \tag{5.132}$$

Parameters

$$b^2=a^2(1-e^2), \qquad f=\frac{a-b}{a}, \qquad e^2=2f-f^2, \tag{5.133}$$

$$e'^2=\frac{a^2-b^2}{b^2}=\frac{e^2}{1-e^2}, \qquad e_1=n=\frac{a-b}{a+b}. \tag{5.134}$$

WGS84 ellipsoid

$$\begin{aligned}
&a=6378137.0\text{m}, && e=0.0818191908, && f=0.003352810,\\
&b=6356752.314\text{m}, && e^2=0.0066943799, && \frac{1}{f}=299.3249753.
\end{aligned} \tag{5.135}$$

Cartesian coordinates

$$\begin{aligned}
X(\phi)&=p(\phi)\cos\lambda=\nu(\phi)\,\cos\phi\cos\lambda,\\
Y(\phi)&=p(\phi)\sin\lambda=\nu(\phi)\,\cos\phi\sin\lambda,\\
Z(\phi)&=(1-e^2)\,\nu(\phi)\,\sin\phi.
\end{aligned} \tag{5.136}$$

Coordinate derivatives

$$\frac{dp}{d\phi}=-\rho\sin\phi, \qquad \frac{dZ}{d\phi}=\rho\cos\phi. \tag{5.137}$$

Radii of curvature and their ratio

$$\nu(\phi)=\frac{a}{[1-e^2\sin^2\phi]^{1/2}}, \qquad \rho(\phi)=\frac{\nu^3}{a^2}\left(1-e^2\right), \qquad \beta(\phi)=\frac{\nu}{\rho}=\frac{1-e^2\sin^2\phi}{1-e^2}. \tag{5.138}$$

Curvature derivatives

$$\frac{d\nu}{d\phi}=(\beta-1)\rho\tan\phi, \qquad \frac{d\rho}{d\phi}=3\frac{(\beta-1)}{\beta}\rho\tan\phi, \qquad \frac{d\beta}{d\phi}=-2(\beta-1)\tan\phi. \tag{5.139}$$

Metric

$$ds^2=\rho^2\,d\phi^2+\nu^2\cos^2\phi\,d\lambda^2. \tag{5.140}$$

/continued

Meridian distance

$$
\begin{aligned}
m(\phi) &= A_0\phi + A_2 \sin 2\phi + A_4 \sin 4\phi + A_6 \sin 6\phi + \cdots, && (5.141)\\
&= B_0\phi + B_2 \sin 2\phi + B_4 \sin 4\phi + B_6 \sin 6\phi + \cdots && (5.142)\\
&= \frac{a}{1+n}\left(h_0\,\phi + h_2 \sin 2\phi + h_4 \sin 4\phi + h_6 \sin 6\phi + \cdots\right). && (5.143)
\end{aligned}
$$

Rectifying latitude

$$
\begin{aligned}
\mu(\phi) &= \frac{\pi}{2}\frac{m(\phi)}{m_p}, && (5.144)\\
&= \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + \cdots, && (5.145)\\
\phi(\mu) &= \mu + d_2 \sin 2\mu + d_4 \sin 4\mu + d_6 \sin 6\mu + \cdots. && (5.146)
\end{aligned}
$$

where A_n, B_n, h_n, b_n and d_n are given by (5.71), (5.76), (5.79), (5.114) and (5.116).

Conformal latitude

$$
\begin{aligned}
\chi(\phi) &= \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots, && (5.147)\\
\phi(\chi) &= \chi + d_2 \sin 2\chi + d_4 \sin 4\chi + d_6 \sin 6\chi + d_8 \sin 8\chi + \cdots, && (5.148)
\end{aligned}
$$

where b_n and d_n are given by (5.108) and (5.110).

Authalic latitude

$$
\begin{aligned}
\xi &= \phi + b_2 \sin 2\phi + b_4 \sin 4\phi + b_6 \sin 6\phi + b_8 \sin 8\phi + \cdots && (5.149)\\
\phi &= \xi + d_2 \sin 2\xi + d_4 \sin 4\xi + d_6 \sin 6\xi + d_8 \sin 8\xi + \cdots && (5.150)
\end{aligned}
$$

where b_n and d_n are given by (5.127) and (5.129).

Chapter 6

Normal Mercator on the ellipsoid (NME)

Abstract

Derivation by analogy with NMS. Alternative forms of the projection. The inverse projection using (a) numerical methods, (b) Taylor series expansions, (c) the inverse series for the conformal latitude Scale factor in geographical coordinates and projection coordinates.

6.1 Introduction

The normal Mercator projection on the ellipsoid (NME) is a straightforward, but non-trivial, generalisation of the normal projection on the sphere (NMS in Chapter 2) and shares the same advantages and disadvantages. 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. The conformality guarantees that the scale at any point is isotropic (independent of direction) so that the projection is locally orthomorphic, preserving small shapes approximately. 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 using a secant form of the projection. The projection stretches to infinity and it is greatly distorted at high latitudes. The projection does not preserve area.

We shall find that the quantitative differences between NMS and NME are of order e^2, about 0.007, and thus less than 1%.

The projection equations are written in terms of a modified Mercator parameter ψ (usually called the isometric latitude in the literature):

$$x(\lambda,\phi) = a\lambda, \qquad y(\lambda,\phi) = a\psi(\phi). \tag{6.1}$$

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 overleaf.

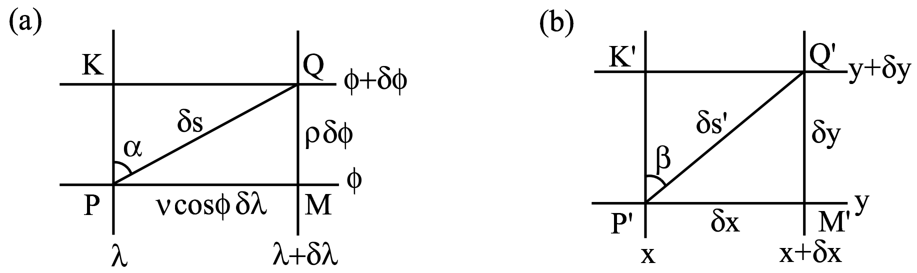


Figure 6.1

6.2 The direct transformation for NME

The modified Mercator parameter is derived by comparing the infinitesimal element on the ellipsoid (Figure 5.6 and above) and the projection plane and imposing the conformality condition. The geometry of the infinitesimal elements gives

$$\text{(a)}\quad \tan\alpha = \frac{\nu\cos\phi\,\delta\lambda}{\rho\,\delta\phi}, \qquad \text{and} \qquad \text{(b)}\quad \tan\beta = \frac{\delta x}{\delta y} = \frac{a\,\delta\lambda}{a\,\psi'(\phi)\delta\phi}, \tag{6.2}$$

so that

$$\tan\beta = \frac{\rho\sec\phi}{\nu\psi'(\phi)}\tan\alpha. \tag{6.3}$$

The projection is conformal if $\alpha = \beta$. Therefore

$$\frac{d\psi}{d\phi} = \frac{\rho(\phi)\sec\phi}{\nu(\phi)}. \tag{6.4}$$

The functions ν and ρ are given in equations (5.16) and (5.45) as

$$\nu(\phi) = \frac{a}{[1-e^2\sin^2\phi]^{1/2}}, \qquad \rho(\phi) = \frac{a(1-e^2)}{\left(1-e^2\sin^2\phi\right)^{3/2}} \tag{6.5}$$

and therefore

$$\psi(\phi) = \int_0^\phi \frac{(1-e^2)}{\cos\phi}\,\frac{1}{1-e^2\sin^2\phi}\,d\phi. \tag{6.6}$$

This integral differs from that for the spherical case by terms of order $O(e^2)$. It was also evaluated in the course of deriving the expression for the conformal latitude and it is directly related to that latitude. From equations 5.98–5.104 we have

$$\psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right] = \ln\left[\tan\left(\frac{\chi(\phi)}{2}+\frac{\pi}{4}\right)\right] \tag{6.7}$$

$$= \tanh^{-1}\sin\phi - e\tanh^{-1}(e\sin\phi) = \tanh^{-1}\sin\chi(\phi) \tag{6.8}$$

$$= \operatorname{gd}^{-1}\phi - e\tanh^{-1}(e\sin\phi) = \operatorname{gd}^{-1}\chi(\phi). \tag{6.9}$$

Equation 6.7 clearly shows that this direct projection for NME may be considered as a double projection (Section 5.10) from the ellipsoid to the conformal sphere, $\phi \to \chi(\phi)$, and then applying the spherical Mercator projection, equation 2.28, to the conformal sphere.

6.3 The inverse transformation for NME

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 invert equations 6.7–6.9 for $\psi(\phi)$ to give $\phi(\psi)$ in a closed form. Attempting to find a series for $\psi(\phi)$ and inverting by the Lagrange reversion method, as we did for the auxiliary latitudes in Sections 5.11–5.13, is doomed to failure because ψ becomes infinite and is close to ϕ only near the equator. We outline three methods of proceeding.

Via the conformal latitude

The simplest method of inverting $\psi(\phi)$ is to first calculate the conformal latitude χ which corresponds to a given value of ψ by using any of the three equations 6.7–6.9 and then calculating $\phi(\chi)$ by using the series 5.109.

Fixed point iteration

Any of the equations 6.7–6.9 may be rewritten in a way which permits a solution by fixed point iteration. For example, corresponding to equation 6.9, we construct the following iteration scheme for a given value of $\psi = y/a$.

$$\phi_{n+1} = \mathrm{gd}\left[\psi + e\tanh^{-1}(e\sin\phi_n)\right], \qquad n = 0, 1, 2, \ldots, \tag{6.10}$$

with initial value taken as the spherical approximation $\phi_0 = \mathrm{gd}(y/a)$, equation 2.35.

Taylor series expansions

Although there are no series analogous to those derived for the auxiliary latitudes in Sections 5.11–5.13 we must develop a series which will be used in the next chapter on TME (transverse Mercator on the ellipsoid). Suppose we a given a specific latitude value ϕ_1 (which will be the footpoint latitude) and the corresponding Mercator parameter ψ_1 calculated from any of equations 6.7–6.9. The values of ϕ corresponding to nearby points can then be deduced from the Taylor series of the function $\phi(\psi)$ about ψ_1. To fourth order the series is:

$$\phi(\psi) = \phi_1 + (\psi-\psi_1)\left.\frac{d\phi}{d\psi}\right|_1 + \frac{(\psi-\psi_1)^2}{2!}\left.\frac{d^2\phi}{d\psi^2}\right|_1 + \frac{(\psi-\psi_1)^3}{3!}\left.\frac{d^3\phi}{d\psi^3}\right|_1 + \frac{(\psi-\psi_1)^4}{4!}\left.\frac{d^4\phi}{d\psi^4}\right|_1, \tag{6.11}$$

We do not know the function $\phi(\psi)$ explicitly but we do know its first derivative *as a function of* ϕ. Equation (6.4) gives

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

We now construct expressions for all the derivatives in the Taylor series as functions of ϕ. Using $\beta' = (2-2\beta)t$ from equation (5.54) and setting $s = \sin\phi$ *etc.* ,

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

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

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

$$\begin{aligned}
\frac{d^4\phi}{d\psi^4} &= \frac{d}{d\phi}\left\{c^3\left[\beta^3(-3+15t^2)+\beta^2(2-18t^2)+\beta(4t^2)\right]\right\}\frac{d\phi}{d\psi} \\
&= c^4 t\left[\beta^4(57-105t^2) + \beta^3(-68+180t^2) + \beta^2(16-84t^2) + \beta(8t^2)\right] \qquad (6.13)
\end{aligned}$$

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

$$\phi - \phi_1 = (\psi-\psi_1)\,\beta_1 c_1 + \frac{(\psi-\psi_1)^2}{2!}\beta_1 c_1^2 t_1 D_2 + \frac{(\psi-\psi_1)^3}{3!}\beta_1 c_1^3 D_3 + \frac{(\psi-\psi_1)^4}{4!}\beta_1 c_1^4 t_1 D_4, \qquad (6.14)$$

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

$$\begin{aligned}
D_2 &= -3\beta_1 + 2 \\
D_3 &= \beta_1^2(-3+15t_1^2) + \beta_1(2-18t_1^2) + 4t_1^2 \\
D_4 &= \beta_1^3(57-105t_1^2) + \beta_1^2(-68+180t_1^2) + \beta_1(16-84t_1^2) + 8t_1^2. \qquad (6.15)
\end{aligned}$$

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

$$\begin{aligned}
\overline{D}_2 &= -1 \\
\overline{D}_3 &= -1 + t_1^2 \\
\overline{D}_4 &= 5 - t_1^2. \qquad (6.16)
\end{aligned}$$

6.4 The scale factor

The scale factor for NME follows immediately from Figure 6.1. Having imposed conformality the angles α and β are equal and the triangles PQM and $P'Q'M'$ are similar so that the scale factor, defined as $P'Q'/PQ$, is equal to the ratio $P'M'/PM$. Since $\delta x = a\delta\lambda$ we see that the scale factor is

$$k(\phi) = \frac{a\sec\phi}{\nu(\phi)}. \tag{6.17}$$

This scale factor is isotropic: it is independent of the angle α and depends only on latitude. It differs from the isotropic scale factor for TMS, $\sec\phi$, by a factor of a/ν which gives rise to differences of order e^2 which are less than 1%.

There is no simple expression for the scale factor expressed in terms of the projection coordinate y, or equivalently $\psi = y/a$. The best we can do is to find the value of ϕ corresponding to given a $\psi_1 = y/a$ by the methods of the previous section and then use equation 6.17. We can then use a Taylor series to evaluate further values of k corresponding to values of ψ close to ψ_1.

Consider the Taylor expansion

$$k(\psi) = k(\psi_1) + (\psi-\psi_1)\left.\frac{dk(\psi)}{d\psi}\right|_1 + \frac{(\psi-\psi_1)^2}{2!}\left.\frac{d^2k(\psi)}{d\psi^2}\right|_1 + \frac{(\psi-\psi_1)^3}{3!}\left.\frac{d^3k(\psi)}{d\psi^3}\right|_1 + \cdots. \tag{6.18}$$

In the next Chapter we shall need this series where ψ_1 will be identified as the footpoint parameter defined as in Figure 4.2 and equation 4.14 generalised to the ellipsoid. See next chapter. The third order terms will prove adequate.

Once again, we shall construct the derivatives in the coefficients as functions of ϕ evaluated at the particular value ϕ_1 corresponding to ψ_1. Using the value of $d\phi/d\psi$ given equation 6.4 we find (with the usual abbreviations for $\sin\phi$ *etc.* and also setting $\beta = \nu/\rho$ with derivatives 5.54)

$$\begin{aligned}
\frac{dk(\psi)}{d\psi} &= \frac{dk(\phi)}{d\phi}\frac{d\phi}{d\psi} = \frac{d}{d\phi}\left(\frac{a\sec\phi}{\nu(\phi)}\right)\frac{\nu(\phi)}{\rho\sec\phi} = (-1)\frac{\nu' c - \nu s}{\nu^2 c^2}\beta c = \frac{t}{\nu},\\
\frac{d^2k(\psi)}{d\psi^2} &= \frac{d}{d\phi}\left(\frac{t}{\nu}\right)\frac{d\phi}{d\psi} = \frac{c}{\nu}(\beta + t^2),\\
\frac{d^3k(\psi)}{d\psi^3} &= \frac{c^2 t}{\nu}\left(5\beta - 4\beta^2 + t^2\right).
\end{aligned} \tag{6.19}$$

$$k(\psi) = \frac{a}{\nu_1 c_1}\left[1 + E_1(\psi-\psi_1) + E_2(\psi-\psi_1)^2 + E_3(\psi-\psi_1)^3 + \cdots\right], \tag{6.20}$$

$$\begin{aligned}
E_1 &= s_1 & \bar{E}_1 &= s_1,\\
E_2 &= c_1^2(\beta_1 + t_1^2) & \bar{E}_2 &= c_1^2(1 + t_1^2),\\
E_3 &= c_1^3 t_1(5\beta_1 - 4\beta_1^2 + t_1^2) & \bar{E}_3 &= c_1^3 t_1(1 + t_1^2).
\end{aligned} \tag{6.21}$$

6.5 Rhumb lines

Since the NME projection is a conformal cylindrical projection, rhumb lines making a constant angle with the ellipsoid meridians are transformed into straight lines. The treatment of rhumb lines on the sphere given in Section 2.5 carries through to the ellipsoid and NME with very little modification. Basically Figure 2.15 is replaced by Figure 6.1 and the meridian distance, $R\phi$ on the sphere, becomes $m(\phi)$ on the ellipsoid. The distances along rhumb lines become

$$r_{12} = \nu(\phi)\cos\phi\,(\lambda_2-\lambda_1), \qquad \text{parallel}, \tag{6.22}$$
$$r_{12} = m(\phi_2)-m(\phi_1), \qquad \text{meridian}, \tag{6.23}$$
$$r_{12} = \sec\alpha\,\big(m(\phi_2)-m(\phi_1)\big), \qquad \text{loxodrome}. \tag{6.24}$$

The equation of the loxodrome through the point (ϕ_1,λ_1) at an azimuth α is given by taking the the straight line

$$y-y_1 = (x-x_1)\cot\alpha\,. \tag{6.25}$$

and setting $y = a\psi(\phi)$ and $x = a\lambda$ with ψ defined by any of equations 6.7–6.9. Therefore

$$\psi(\phi) = \psi(\phi_1) + (\lambda-\lambda_1)\cot\alpha\,, \tag{6.26}$$

$$\lambda(\phi) = \lambda_1 + \tan\alpha\left[\tanh^{-1}\sin\phi - e\tanh^{-1}(e\sin\phi)\right]_{\phi_1}^{\phi}. \tag{6.27}$$

There is no equation corresponding to equation 2.43 because $\psi(\phi)$ cannot be inverted in closed form.

6.6 Modified NME

NME can be modified exactly as NMS (in Section 2.7) 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 by terms of order e^2, a cange of less than 1%. We simply introduce a factor of k_0 into the transformations.

$$x = k_0 a\lambda, \qquad y = k_0 a\psi(\phi), \tag{6.28}$$
$$\lambda = \frac{x}{k_0 a}, \qquad \phi = \phi(\psi) \quad \text{with} \quad \psi = y/k_0 a. \tag{6.29}$$

where

$$\psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2}+\frac{\pi}{4}\right)\left(\frac{1-e\sin\phi}{1+e\sin\phi}\right)^{e/2}\right] = \ln\left[\tan\left(\frac{\chi(\phi)}{2}+\frac{\pi}{4}\right)\right] \tag{6.30}$$
$$= \tanh^{-1}\sin\phi - e\tanh^{-1}(e\sin\phi) \qquad = \tanh^{-1}\sin\chi(\phi) \tag{6.31}$$
$$= \mathrm{gd}^{-1}\phi - e\tanh^{-1}(e\sin\phi) \qquad = \mathrm{gd}^{-1}\chi(\phi). \tag{6.32}$$

and $\phi(\psi)$ is calculated by the methods of Section 6.3.

Chapter 7

Transverse Mercator on the ellipsoid (TME)

The longitude (λ) series

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.

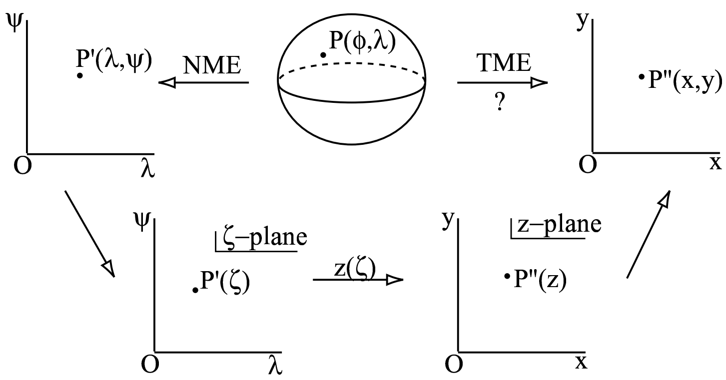


Figure 7.1

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 (6.7–6.9).

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. This chapter presents truncated power series in λ for 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 $x = 0$. The method parallels that of Chapter 4 with the amended definitions of the Mercator parameter and meridian distance for the ellipsoid. These series, originally investigated by Gauss (c1822), were published by Krüger (1912); they were rebublished in English by Lee (1945) and extended by Redfearn (1948) and Thomas (1952)

The meridian distance

The meridian distance on the ellipsoid was obtained as a (various) series in Section 5.8: the most useful for practical work is the Helmert form of equation (5.78):

$$m(\phi) = \frac{a}{1+n}(h_0\phi + h_2 \sin 2\phi + h_4 \sin 4\phi + h_6 \sin 6\phi + h_8 \sin 8\phi + \cdots, \tag{7.2}$$

where the h-coefficients are given in equations (5.79). 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, use the series (5.115) with $\mu = (1+n)m(\phi)/h_0 = (1+n)y/ah_0$, (5.111) ,

$$\phi = \mu + d_2 \sin 2\mu + d_4 \sin 4\mu + d_6 \sin 6\mu + \cdots, \qquad \mu = \frac{y}{B_0}, \tag{7.6}$$

where the d-coefficients are given in (5.116).

The Mercator parameter: derivative and inverse

The Mercator parameter on the ellipsoid is given in equations (6.7)–(6.9) as, for example,

$$\psi(\phi) = \ln\left[\tan\left(\frac{\phi}{2} + \frac{\pi}{4}\right)\left(\frac{1 - e\sin\phi}{1 + e\sin\phi}\right)^{e/2}\right]. \tag{7.7}$$

We do not need this explicit form, only its derivatives. From (6.4)

$$\frac{d\psi}{d\phi} = \frac{\rho(\phi)}{\nu(\phi)\cos\phi}, \qquad \frac{d\phi}{d\psi} = \frac{\nu(\phi)\cos\phi}{\rho(\phi)}. \tag{7.8}$$

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 (6.14).

$$\phi-\phi_1 = (\psi-\psi_1)\,\beta_1 c_1 + \frac{(\psi-\psi_1)^2}{2!}\beta_1 c_1^2 t_1 D_2 + \frac{(\psi-\psi_1)^3}{3!}\beta_1 c_1^3 D_3 + \frac{(\psi-\psi_1)^4}{4!}\beta_1 c_1^4 t_1 D_4 \tag{7.9}$$

where the D-coefficients are given in (6.15) and (6.16). The '1' suffix 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 (5.67) 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):

$$\frac{d\nu}{d\phi} = (\beta-1)\rho\tan\phi, \qquad \frac{d\beta}{d\phi} = -2(\beta-1)\tan\phi. \tag{7.12}$$

$$M^{(1)} = \frac{dM}{d\psi} = \nu c. \tag{7.13}$$

$$M^{(2)} = \frac{d^2M}{d\psi^2} = \frac{d}{d\phi}\left(M^{(1)}\right)\frac{d\phi}{d\psi} = \frac{d}{d\phi}(\nu c)\frac{d\phi}{d\psi} = [(\beta-1)\rho t c - \nu s]\,\frac{\nu c}{\rho}$$
$$= -\nu s c. \tag{7.14}$$

$$M^{(3)} = \frac{d^3M}{d\psi^3} = -\left[(\beta-1)\rho t s c + \nu(c^2-s^2)\right]\frac{\nu c}{\rho} = -\nu c^3\left(\beta - t^2\right) \tag{7.15}$$
$$\equiv -\nu c^3 W_3.$$

/cont.

$$
\begin{aligned}
M^{(4)} = \frac{d^4M}{d\psi^4} &= -\Big[\big\{(\beta-1)\rho t c^3 - 3\nu c^2 s\big\}\,(\beta-t^2) + \nu c^3\,\big\{-2(\beta-1)t - 2t(1+t^2)\big\}\Big]\,\frac{\nu c}{\rho}\\
&= \nu s c^3\left[4\beta^2+\beta-t^2\right] \qquad (7.16)\\
&\equiv \nu s c^3 W_4.
\end{aligned}
$$

$$
\begin{aligned}
M^{(5)} = \frac{d^5M}{d\psi^5} &= \Big[\big\{(\beta-1)\rho t s c^3 + \nu(c^4-3s^2c^2)\big\}\,(4\beta^2+\beta-t^2)\\
&\qquad + \nu s c^3\,\big\{(8\beta+1)(-2\beta+2)t - 2t(1+t^2)\big\}\Big]\,\frac{\nu c}{\rho}\\
&= \nu c^5\left[4\beta^3(1-6t^2)+\beta^2(1+8t^2)-2\beta t^2+t^4\right] \qquad (7.17)\\
&\equiv \nu c^5 W_5.
\end{aligned}
$$

$$
\begin{aligned}
M^{(6)} = \frac{d^6M}{d\psi^6} &= \Big[\big\{(\beta-1)\rho t c^5 + \nu(-5sc^4)\big\}\,W_5 + \nu c^5 W_5'\Big]\,\frac{\nu c}{\rho}\\
&= \nu s c^5\Big[(-4\beta-1)\,\big\{4\beta^3(1-6t^2)+\beta^2(1+8t^2)-2\beta t^2+t^4\big\}\\
&\qquad + \beta t^{-1}\,\big\{\big(12\beta^2(1-6t^2)+2\beta(1+8t^2)-2t^2\big)\,(-2\beta+2)\,t\\
&\qquad\qquad - \big(24\beta^3-8\beta^2+2\beta\big)\,\big(2t(1+t^2)\big)+4t^3(1+t^2)\big\}\Big]\\
&= -\nu s c^5\left[8\beta^4(11-24t^2)-28\beta^3(1-6t^2)+\beta^2(1-32t^2)-2\beta t^2+t^4\right]\\
&\equiv -\nu s c^5 W_6 \qquad (7.18)
\end{aligned}
$$

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 $e^2\lambda^8$. (The full terms of this order are in Thomas (1952), pages 95,96.) 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

$$
\begin{aligned}
M^{(7)} = \frac{d^7M}{d\psi^7} &= -\Big[\nu\left(c^6-5s^2c^4\right)\,W_6\big|_{\beta=1} + \nu s c^5\,W_6'\big|_{\beta=1}\Big]\,\frac{\nu c}{\rho}\\
&= -\nu c^7\left[\left(1-5t^2\right)\left(61-58t^2+t^4\right)+t\left(-116t+4t^3\right)(1+t^2)\right]\\
&= -\nu c^7\left(61-479t^2+179t^4-t^6\right).\\
&\equiv -\nu c^7\overline{W}_7. \qquad (7.19)
\end{aligned}
$$

$$
\begin{aligned}
M^{(8)} = \frac{d^8M}{d\psi^8} &= \Big[7\nu c^6 s\overline{W}_7 - \nu c^7\overline{W}_7{}'\Big]\,\frac{\nu c}{\rho}\\
&= \nu s c^7\left[7\left(61-479t^2+179t^4-t^6\right)-\frac{1}{t}\left(-958t+716t^3-6t^5\right)(1+t^2)\right]\\
&= \nu s c^7\left(1385-3111t^2+543t^4-t^6\right).\\
&\equiv \nu s c^7\overline{W}_8. \qquad (7.20)
\end{aligned}
$$

Note the minus signs introduced in the definitions of W_3, W_6 and $\overline{W}_7$.

Summary of derivatives

$$
\begin{aligned}
M^{(1)} &= \nu c & W_1 &= 1\\
M^{(2)} &= -\nu sc & W_2 &= 1\\
M^{(3)} &= -\nu c^3\, W_3 & W_3(\phi) &= \beta - t^2\\
M^{(4)} &= \nu sc^3 W_4 & W_4(\phi) &= 4\beta^2 + \beta - t^2\\
M^{(5)} &= \nu c^5\, W_5 & W_5(\phi) &= 4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4\\
M^{(6)} &= -\nu sc^5 W_6 & W_6(\phi) &= 8\beta^4(11-24t^2) - 28\beta^3(1-6t^2) + \beta^2(1-32t^2) - 2\beta t^2 + t^4\\
M^{(7)} &= -\nu c^7\, \overline{W}_7 & \overline{W}_7(\phi) &= 61 - 479t^2 + 179t^4 - t^6 + O(e^2)\\
M^{(8)} &= \nu sc^7 \overline{W}_8 & \overline{W}_8(\phi) &= 1385 - 3111t^2 + 543t^4 - t^6 + O(e^2).
\end{aligned}
\tag{7.21}
$$

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

$$
\begin{aligned}
W_3(\phi) &\to \overline{W}_3(\phi) = 1 - t^2,\\
W_4(\phi) &\to \overline{W}_4(\phi) = 5 - t^2,\\
W_5(\phi) &\to \overline{W}_5(\phi) = 5 - 18t^2 + t^4,\\
W_6(\phi) &\to \overline{W}_6(\phi) = 61 - 58t^2 + t^4.
\end{aligned}
\tag{7.22}
$$

These derivatives are more easily calculated by using Maxima (2009)—see Appendix H.

7.2 Derivation of the Redfearn series

The direct complex series

Following Section 4.3, 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

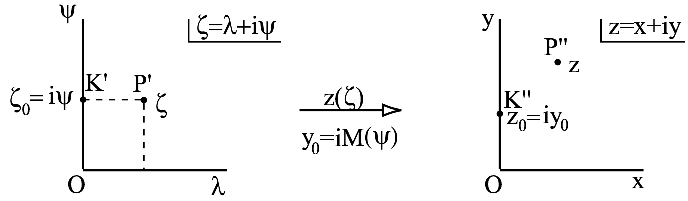


Figure 7.2

set $\zeta - \zeta_0 = \lambda$ 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 (4.34) 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):

$$x(\lambda,\phi) = \lambda\nu c + \frac{\lambda^3 \nu c^3}{3!}W_3 + \frac{\lambda^5 \nu c^5}{5!}W_5 + \frac{\lambda^7 \nu c^7}{7!}\overline{W}_7, \tag{7.28}$$

$$y(\lambda,\phi) = m(\phi) + \frac{\lambda^2 \nu s c}{2} + \frac{\lambda^4 \nu s c^3}{4!}W_4 + \frac{\lambda^6 \nu s c^5}{6!}W_6 + \frac{\lambda^8 \nu s c^7}{8!}\overline{W}_8. \tag{7.29}$$

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 (4.15)

$$x_\lambda = y_\psi = M^{(1)} - \frac{1}{2!}\lambda^2 M^{(3)} + \frac{1}{4!}\lambda^4 M^{(5)} - \frac{1}{6!}\lambda^6 M^{(7)} + \cdots, \tag{7.30}$$

$$x_\psi = -y_\lambda = \lambda M^{(2)} - \frac{1}{3!}\lambda^3 M^{(4)} + \frac{1}{5!}\lambda^5 M^{(6)} - \frac{1}{7!}\lambda^7 M^{(8)} + \cdots. \tag{7.31}$$

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

$$\frac{z-z_0}{\nu_0 c_0} = (\zeta-\zeta_0) + \frac{b_2}{2!}(\zeta-\zeta_0)^2 + \frac{b_3}{3!}(\zeta-\zeta_0)^3 + \cdots + \frac{b_8}{8!}(\zeta-\zeta_0)^8 + \cdots \tag{7.32}$$

where we have set $z_0 = iy_0 = iM_0$. The b-coefficients are

$$\begin{aligned}
b_2 &= \frac{-iM_0^{(2)}}{\nu_0 c_0} &&= is_0\\
b_3 &= \frac{-M_0^{(3)}}{\nu_0 c_0} &&= c_0^2 W_3(\phi_0)\\
b_4 &= \frac{iM_0^{(4)}}{\nu_0 c_0} &&= is_0 c_0^2 W_4(\phi_0)\\
b_5 &= \frac{M_0^{(5)}}{\nu_0 c_0} &&= c_0^4 W_5(\phi_0)\\
b_6 &= \frac{-iM_0^{(6)}}{\nu_0 c_0} &&= is_0 c_0^4 W_6(\phi_0)\\
b_7 &= \frac{-M_0^{(7)}}{\nu_0 c_0} &&= c_0^6 W_7(\phi_0)\\
b_8 &= \frac{iM_0^{(8)}}{\nu_0 c_0} &&= is_0 c_0^6 W_8(\phi_0)
\end{aligned} \tag{7.33}$$

where the functions on the right hand sides are evaluated at ϕ_0 such that $\psi_0 = \psi(\phi_0)$.

The Lagrange reversion 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

$$\zeta-\zeta_0 = \left(\frac{z-z_0}{\nu_0 c_0}\right) - \frac{p_2}{2!}\left(\frac{z-z_0}{\nu_0 c_0}\right)^2 - \frac{p_3}{3!}\left(\frac{z-z_0}{\nu_0 c_0}\right)^3 - \cdots - \frac{\bar{p}_8}{8!}\left(\frac{z-z_0}{\nu_0 c_0}\right)^8, \tag{7.34}$$

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

$$\begin{aligned}
p_2 &= ic_1 t_1, \\
p_3 &= c_1^2 V_3 & V_3 &= \beta_1 + 2t_1^2,\\
p_4 &= ic_1^3 t_1 V_4 & V_4 &= 4\beta_1^2 - 9\beta_1 - 6t_1^2,\\
p_5 &= c_1^4 V_5 & V_5 &= 4\beta_1^3(1-6t_1^2) - \beta_1^2(9-68t_1^2) - 72\beta_1 t_1^2 - 24t_1^4,\\
p_6 &= ic_1^5 t_1 V_6 & V_6 &= 8\beta_1^4(11-24t_1^2) - 84\beta_1^3(3-8t_1^2) + 225\beta_1^2(1-4t_1^2) + 600\beta_1 t_1^2 + 120t_1^4,\\
\bar{p}_7 &= c_1^6 \bar{V}_7 & \bar{V}_7 &= 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6,\\
\bar{p}_8 &= ic_1^7 t_1 \bar{V}_8 & \bar{V}_8 &= -1385 - 7266t_1^2 - 10920t_1^4 - 5040t_1^6.
\end{aligned} \tag{7.35}$$

The inverse series for ψ and λ

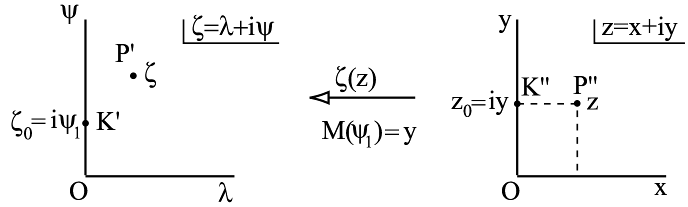


Figure 7.3

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). The the real and imaginary parts are

$$\lambda(x,y)=\frac{x}{\nu_1c_1}-\frac{x^3}{3!\,\nu_1^3c_1}V_3-\frac{x^5}{5!\,\nu_1^5c_1}V_5-\frac{x^7}{7!\,\nu_1^7c_1}\overline{V}_7,\quad\text{where } m(\phi_1)=y, \tag{7.37}$$

$$\psi-\psi_1=-\frac{x^2t_1}{2!\,\nu_1^2c_1}-\frac{x^4t_1}{4!\,\nu_1^4c_1}V_4-\frac{x^6t_1}{6!\,\nu_1^6c_1}V_6-\frac{x^8t_1}{8!\,\nu_1^8c_1}\overline{V}_8. \tag{7.38}$$

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 $\beta_1\to1$ as $e\to1$ 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 (6.14) 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 (6.15). 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

$$\begin{aligned}
\phi-\phi_1 = &-\frac{1}{2!}\widetilde{x}^2\beta_1 t_1\,[1]\\
&-\frac{1}{4!}\widetilde{x}^4\beta_1 t_1\left[V_4-3t_1^2D_2\right]\\
&-\frac{1}{6!}\widetilde{x}^6\beta_1 t_1\left[V_6-15t_1^2D_2V_4+15t_1^2D_3\right]\\
&-\frac{1}{8!}\widetilde{x}^8\beta_1 t_1\left[\overline{V}_8-28t_1^2\overline{D}_2\overline{V}_6-35t_1^2\overline{D}_2\overline{V}_4^{\,2}+210t_1^2\overline{D}_3\overline{V}_4-105t_1^4\overline{D}_4\right],
\end{aligned} \tag{7.42}$$

where we use the spherical approximation in evaluating the eighth order term. Substituting for the D-coefficients from equations (6.15, 6.16) and the V-coefficients from equations (7.35,7.39) our final result for ϕ is

$$\phi(x,y) = \phi_1-\frac{x^2\beta_1 t_1}{2\nu_1^2}-\frac{x^4\beta_1 t_1}{4!\nu_1^4}U_4-\frac{x^6\beta_1 t_1}{6!\nu_1^6}U_6-\frac{x^8\beta_1 t_1}{8!\nu_1^8}\overline{U}_8, \tag{7.43}$$

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:

$$\begin{aligned}
U_4 \to \overline{U}_4 &= -5-3t_1^2,\\
U_6 \to \overline{U}_6 &= \ 61+90t_1^2+45t_1^4,
\end{aligned} \tag{7.45}$$

7.3 Convergence and scale in TME

This section parallels the calculation of convergence and scale factor series given in Section 4.5, starting from equations 4.75 and 4.76:

$$\tan\gamma(\lambda,\psi)=\frac{y_\lambda}{x_\lambda}, \qquad \tan\gamma(x,y)=-\frac{\psi_x}{\lambda_x}, \tag{7.46}$$

$$m(\lambda,\psi)=x_\lambda\sec\gamma(\lambda,\psi), \qquad \frac{1}{m(x,y)}=\lambda_x\sec\gamma(x,y). \tag{7.47}$$

The scale factors m are those for the transformation between the (λ,ψ) and (x,y) complex planes. They must be multiplied by the the scale factor from the ellipsoid to the plane:

$$k(\lambda,\phi)=k_{\text{NME}}(\lambda,\phi)\,m(\lambda,\psi(\phi)), \qquad k(x,y)=k_{\text{NME}}(x,y)\,m(x,y), \tag{7.48}$$

where we have introduced new notation for the scale factors of NME defined in equations 6.17 and 6.18. In the latter ψ is a function of (x,y) through the inverse series 7.38 and the definition of the footpoint latitude such that $m(\phi_1)=y$.

Series for partial derivatives

The above expressions for the convergence and scale factors depend on the partial derivatives x_λ, y_λ, λ_x, ψ_x of the series 7.28, 7.29, 7.37, 7.38. Setting $\widetilde{\lambda}=\lambda c$ and $\widetilde{x}=x/\nu_1$

$$x_\lambda=\frac{\partial x}{\partial\lambda}=\quad \nu c\left[1+\frac{1}{2}\widetilde{\lambda}^2W_3+\frac{1}{24}\widetilde{\lambda}^4W_5+\frac{1}{720}\widetilde{\lambda}^6\overline{W}_7,\right], \qquad \widetilde{\lambda}=\lambda c \tag{7.49}$$

$$y_\lambda=\frac{\partial y}{\partial\lambda}=\quad \nu s\widetilde{\lambda}\left[1+\frac{1}{6}\widetilde{\lambda}^2W_4+\frac{1}{120}\widetilde{\lambda}^4W_6+\frac{1}{5040}\widetilde{\lambda}^6\overline{W}_8\right], \tag{7.50}$$

$$\lambda_x=\frac{\partial\lambda}{\partial x}=\quad \frac{1}{\nu_1c_1}\left[1-\frac{1}{2}\widetilde{x}^2V_3-\frac{1}{24}\widetilde{x}^4V_5-\frac{1}{720}\widetilde{x}^6\overline{V}_7\right], \qquad \widetilde{x}=\frac{x}{\nu_1} \tag{7.51}$$

$$\psi_x=\frac{\partial\psi}{\partial x}=-\frac{t_1\widetilde{x}}{\nu_1c_1}\left[1+\frac{1}{6}\widetilde{x}^2V_4+\frac{1}{120}\widetilde{x}^4V_6+\frac{1}{5040}\widetilde{x}^6\overline{V}_8\right]. \tag{7.52}$$

Note that, apart from the overall multiplicative terms, the series for λ_x and ψ_x are obtained from those for x_λ and y_λ by $\widetilde{\lambda}\to\widetilde{x}$ and $W_n\to -V_n$ (n odd) and $W_n\to V_n$ (n even).

NB. In constructing the transformation series 7.28, 7.29, 7.37, 7.38 we discarded terms of order greater than λ^8 and $(x/a)^8$ so we must discard terms of order greater than λ^7 and $(x/a)^7$ 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 transformation 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 $\widetilde{\lambda}^6$, $\widetilde{\lambda}^7$, $\widetilde{x}^6$, and $\widetilde{x}^7$ wherever they arise in the manipulation of the series for the derivatives.

The quotient of $\mathbf{y}_\lambda$ and $\mathbf{x}_\lambda$

Using (E.31) we find the inverse of x_λ in (7.49) as

$$\frac{\nu c}{x_\lambda} = 1 - \widetilde{\lambda}^2\left(\frac{W_3}{2}\right) - \widetilde{\lambda}^4\left(\frac{W_5}{24} - \frac{W_3^2}{4}\right) - \widetilde{\lambda}^6\left(\frac{\overline{W}_7}{720} - \frac{\overline{W}_3\overline{W}_5}{24} + \frac{\overline{W}_3{}^3}{8}\right). \tag{7.53}$$

Note that the W_3 and W_5 which the inversion casts into the coefficient of λ^6 have been replaced by their spherical limits. The product of the above with (7.50) gives

$$\frac{y_\lambda}{x_\lambda} = t\,\widetilde{\lambda}\left[1 + a_2\widetilde{\lambda}^2 + a_4\widetilde{\lambda}^4 + \bar{a}_6\widetilde{\lambda}^6\right]. \tag{7.54}$$

The a_n coefficients (and their spherical limits) are calculated using (7.21, 7.22)

$$\begin{aligned}
a_2 &= \frac{W_4}{6} - \frac{W_3}{2} \quad = \frac{1}{3}\left[2\beta^2 - \beta + t^2\right], & \bar{a}_2 &= \frac{1}{3}\left(1 + t^2\right)\\
a_4 &= \frac{W_6}{120} - \frac{W_3W_4}{12} + \frac{W_3^2}{4} - \frac{W_5}{24},\\
&= \frac{1}{15}\left[\beta^4(11-24t^2) - \beta^3(11-36t^2) + \beta^2(2-4t^2) - 4\beta t^2 + 2t^4\right], & \bar{a}_4 &= \frac{1}{15}\left(2+4t^2+2t^4\right)\\
\bar{a}_6 &= \frac{\overline{W}_8}{5040} - \frac{\overline{W}_3\overline{W}_6}{240} + \frac{\overline{W}_3{}^2\overline{W}_4}{24} - \frac{\overline{W}_4\overline{W}_5}{144} - \frac{\overline{W}_3{}^3}{8} + \frac{\overline{W}_3\overline{W}_5}{24} - \frac{\overline{W}_7}{720}\\
&= \frac{1}{315}\left[17 + 51t^2 + 51t^4 + 17t^6\right].
\end{aligned} \tag{7.55}$$

The quotient of ψ_x and λ_x

From the comment immediately after equation (7.52) we see that if we define

$$\frac{\psi_x}{\lambda_x} = -t_1\,\widetilde{x}\left[1 + r_2\widetilde{x}^2 + r_4\widetilde{x}^4 + \bar{r}_6\widetilde{x}^6\right] \tag{7.56}$$

then the r_n coefficients follow by analogy with equations (7.55). Using (7.35, 7.39) we find that coefficients and their spherical limits are

$$\begin{aligned}
r_2 &= \frac{V_4}{6} + \frac{V_3}{2} \quad = \frac{1}{3}\left[2\beta_1^2 - 3\beta_1\right], & \bar{r}_2 &= -\frac{1}{3}\\
r_4 &= \frac{V_6}{120} + \frac{V_3V_4}{12} + \frac{V_3^2}{4} + \frac{V_5}{24},\\
&= \frac{1}{15}\left[\beta_1^4(11-24t_1^2) - 3\beta_1^3(8-23t_1^2) + 15\beta_1^2(1-4t_1^2) + 15\beta_1 t_1^2\right], & \bar{r}_4 &= \frac{2}{15}\\
\bar{r}_6 &= \frac{\overline{V}_8}{5040} + \frac{\overline{V}_3\overline{V}_6}{240} + \frac{\overline{V}_3{}^2\overline{V}_4}{24} + \frac{\overline{V}_4\overline{V}_5}{144} + \frac{\overline{V}_3{}^3}{8} + \frac{\overline{V}_3\overline{V}_5}{24} + \frac{\overline{V}_7}{720} = -\frac{17}{315}.
\end{aligned} \tag{7.57}$$

Note the absence of terms in t^2, t^4 or t^6 in the coefficients $\bar{r}_2$, $\bar{r}_4$ or $\bar{r}_6$.

7.4 Convergence in geographical coordinates

From equations (7.46) and (7.54) we have

$$\tan\gamma(\phi,\lambda) = \widetilde{\lambda} t\left[1 + a_2\widetilde{\lambda}^2 + a_4\widetilde{\lambda}^4 + \bar{a}_6\widetilde{\lambda}^6\right] \tag{7.58}$$

and we calculate γ as $\arctan(\tan\gamma)$ by using the series (E.20):

$$\gamma = \tan\gamma - \frac{1}{3}\tan^3\gamma + \frac{1}{5}\tan^5\gamma - \frac{1}{7}\tan^7\gamma + \cdots. \tag{7.59}$$

To order $\widetilde{\lambda}^7$ the higher powers of $\tan\gamma$ are given by

$$\begin{aligned}
\tan^2\gamma &= \widetilde{\lambda}^2 t^2\left[1 + 2a_2\widetilde{\lambda}^2 + (2\bar{a}_4 + \bar{a}_2^2)\widetilde{\lambda}^4\right]\\
\tan^3\gamma &= \widetilde{\lambda}^3 t^3\left[1 + 3a_2\widetilde{\lambda}^2 + 3(\bar{a}_4 + \bar{a}_2^2)\widetilde{\lambda}^4\right]\\
\tan^4\gamma &= \widetilde{\lambda}^4 t^4\left[1 + 4\bar{a}_2\widetilde{\lambda}^2\right]\\
\tan^5\gamma &= \widetilde{\lambda}^5 t^5\left[1 + 5\bar{a}_2\widetilde{\lambda}^2\right]\\
\tan^6\gamma &= \widetilde{\lambda}^6 t^6\\
\tan^7\gamma &= \widetilde{\lambda}^7 t^7
\end{aligned} \tag{7.60}$$

so that

$$\gamma = \widetilde{\lambda} t + \frac{\widetilde{\lambda}^3 t}{3}\left[3a_2 - t^2\right] + \frac{\widetilde{\lambda}^5 t}{5}\left[5a_4 - 5a_2t^2 + t^4\right] + \frac{\widetilde{\lambda}^7 t}{7}\left[7\bar{a}_6 - 7\left(\bar{a}_4 + \bar{a}_2^2\right)t^2 + 7\bar{a}_2t^4 - t^6\right]. \tag{7.61}$$

Substituting the a-coefficients from (7.55) gives the final result

$$\gamma(\lambda,\phi) = \widetilde{\lambda} t + \frac{1}{3}\widetilde{\lambda}^3 t H_3 + \frac{1}{15}\widetilde{\lambda}^5 t H_5 + \frac{1}{315}\widetilde{\lambda}^7 t \bar{H}_7, \qquad \widetilde{\lambda} = \lambda c. \tag{7.62}$$

$$\begin{aligned}
H_3 &= 2\beta^2 - \beta,\\
H_5 &= \beta^4(11 - 24t^2) - \beta^3(11 - 36t^2) + \beta^2(2 - 14t^2) + \beta t^2,\\
\bar{H}_7 &= 17 - 26t^2 + 2t^4.
\end{aligned} \tag{7.63}$$

We also require an expression for $\sec\gamma$ for equation 7.47:

$$\begin{aligned}
\sec\gamma &= \left\{1 + \tan^2\gamma\right\}^{1/2}\\
&= 1 + \frac{1}{2}\tan^2\gamma - \frac{1}{8}\tan^4\gamma + \frac{1}{16}\tan^6\gamma + \cdots\\
&= 1 + \frac{1}{2}\widetilde{\lambda}^2 t^2 + \frac{1}{8}\widetilde{\lambda}^4\left[8a_2t^2 - t^4\right] + \frac{1}{16}\widetilde{\lambda}^6\left[16\bar{a}_4t^2 + 8\bar{a}_2^2t^2 - 8\bar{a}_2t^4 + t^6\right]
\end{aligned} \tag{7.64}$$

7.5 Convergence in projection coordinates

From (7.46) and (7.56) and

$$\tan\gamma(x,y) = t_1\widetilde{x}\left[1 + r_2\widetilde{x}^2 + r_4\widetilde{x}^4 + \bar{r}_6\widetilde{x}^6\right]. \tag{7.65}$$

Comparing this equation with (7.58) we see from equation (7.61) that

$$\gamma = \widetilde{x}t_1 + \frac{\widetilde{x}^3 t_1}{3}\left[3r_2 - t_1^2\right] + \frac{\widetilde{x}^5 t_1}{5}\left[5r_4 - 5r_2t_1^2 + t_1^4\right] + \frac{\widetilde{x}^7 t_1}{7}\left[7\bar{r}_6 - 7\left(\bar{r}_4 + \bar{r}_2^2\right)t_1^2 + 7\bar{r}_2t_1^4 - t_1^6\right]. \tag{7.66}$$

Substituting the r-coefficients from (7.57) gives

$$\gamma(x,y) = \widetilde{x}t_1 + \frac{1}{3}\widetilde{x}^3 t_1 K_3 + \frac{1}{15}\widetilde{x}^5 t_1 K_5 + \frac{1}{315}\widetilde{x}^7 t_1 K_7, \qquad \widetilde{x} = \frac{x}{\nu_1}, \tag{7.67}$$

where

$$\begin{aligned} K_3 &= 2\beta_1^2 - 3\beta_1 - t_1^2, \\ K_5 &= \beta_1^4(11 - 24t_1^2) - 3\beta_1^3(8 - 23t_1^2) + 5\beta_1^2(3 - 14t_1^2) + 30\beta_1 t_1^2 + 3t_1^4, \\ \bar{K}_7 &= -17 - 77t_1^2 - 105t_1^4 - 45t_1^6. \end{aligned} \tag{7.68}$$

Setting $\beta_1 = \nu_1/\rho_1$ gives the Redfearn series (apart from minor typos where he writes ν rather than ν_1 *etc.*

We also need an expression for $\sec\gamma(x,y)$. Comparison with 7.64 gives

$$\sec\gamma(x,y) = \quad = 1 + \frac{1}{2}\widetilde{x}^2 t_1^2 + \frac{1}{8}\widetilde{x}^4\left[8r_2t_1^2 - t_1^4\right] + \frac{1}{16}\widetilde{x}^6\left[16\bar{r}_4t_1^2 + 8\bar{r}_2^2t_1^2 - 8\bar{r}_2t_1^4 + t_1^6\right] \tag{7.69}$$

7.6 Scale factor in geographical coordinates

In 7.48 substitute for $k_{\text{NME}}(\lambda,\phi)$ and $m(\lambda,\psi(\phi))$ from 6.17 and 7.47 respectively:

$$k(\lambda,\phi) = k_{\text{NME}}(\lambda,\phi)\,m(\lambda,\psi(\phi)) = \frac{x_\lambda \sec\gamma(\lambda,\phi)}{\nu\cos\phi}, \tag{7.70}$$

where x_λ is given in equation (7.49) and $\sec\gamma$ in (7.64):

$$\frac{x_\lambda}{\nu c} = 1 + \frac{1}{2}\widetilde{\lambda}^2 W_3 + \frac{1}{24}\widetilde{\lambda}^4 W_5 + \frac{1}{720}\widetilde{\lambda}^6 \overline{W}_7, \tag{7.71}$$

$$\sec\gamma = 1 + \frac{1}{2}\widetilde{\lambda}^2 t^2 + \frac{1}{8}\widetilde{\lambda}^4\left[8a_2t^2 - t^4\right] + \frac{1}{16}\widetilde{\lambda}^6\left[16\bar{a}_4t^2 + 8\bar{a}_2^2t^2 - 8\bar{a}_2t^4 + t^6\right] \tag{7.72}$$

Therefore

$$k(\lambda,\phi) = 1+\frac{1}{2}\widetilde{\lambda}^2\big(W_3+t^2\big)+\frac{1}{24}\widetilde{\lambda}^4\big(W_5+6t^2W_3+24a_2t^2-3t^4\big)$$
$$+\frac{\widetilde{\lambda}^6}{720}\big(\overline{W}_7+15t^2\overline{W}_5+360\bar{a}_2t^2\overline{W}_3-45t^4\overline{W}_3+720\bar{a}_4t^2+360\bar{a}_2^2t^2-360\bar{a}_2t^4+45t^6\big) \tag{7.73}$$

Using the W-coefficients from (7.21 7.22) and the a-coefficients from (7.55),

$$k(\lambda,\phi) = 1+\frac{1}{2}\widetilde{\lambda}^2H_2+\frac{1}{24}\widetilde{\lambda}^4H_4+\frac{1}{720}\widetilde{\lambda}^6\overline{H}_6, \qquad \widetilde{\lambda}=\lambda c, \tag{7.74}$$

$$H_2 = \beta$$
$$H_4 = 4\beta^3(1-6t^2)+\beta^2(1+24t^2)-4\beta t^2$$
$$\overline{H}_6 = 61-148t^2+16t^4. \tag{7.75}$$

7.7 Scale factor in projection coordinates

In 7.48 substitute for $k_{\mathrm{NME}}(x,y)$ and $m(x,y)$ from 6.20 and 7.47 respectively:

$$k(x,y) = k_{\mathrm{NME}}(x,y)\,m(x,y) = \frac{k(\psi)}{\lambda_x\sec\gamma}. \tag{7.76}$$

Using 6.20 and (7.38) for $\psi-\psi_1$ gives

$$k(\psi) = \frac{1}{\nu_1c_1}\left[1+E_1(\psi-\psi_1)+E_2(\psi-\psi_1)^2+E_3(\psi-\psi_1)^3+\cdots\right] \tag{7.77}$$
$$= \frac{1}{\nu_1c_1}\left[1-E_1\frac{x^2t_1}{2\nu_1^2c_1}\left(1+\frac{x^2}{12\nu_1^2}V_4+\frac{x^4}{360\,\nu_1^4}\overline{V}_6\right)\right.$$
$$\left.+\frac{1}{2}E_2\frac{x^4t_1^2}{4\nu_1^4c_1^2}\left(1+\frac{x^2}{6\nu_1^2}\overline{V}_4\right)-\frac{1}{6}\overline{E}_3\frac{x^6t_1^3}{8\nu_1^6c_1^3}+\cdots\right],$$
$$= \frac{1}{\nu_1c_1}\left[1+\frac{1}{2}\widetilde{x}^2q_2+\frac{1}{24}\widetilde{x}^4q_4+\frac{1}{720}\widetilde{x}^6\overline{q}_6\right], \qquad \widetilde{x}=\frac{x}{\nu_1}. \tag{7.78}$$

With the E coefficients from 6.21 and the V coefficients from 7.35 and 7.39,

$$q_2 = -E_1\frac{t_1}{c_1} = -t_1^2,$$
$$q_4 = -E_1\frac{t_1}{c_1}V_4+3E_2\frac{t_1^2}{c_1^2} = 12\beta_1t_1^2-4\beta_1^2t_1^2+9t_1^4,$$
$$\overline{q}_6 = -E_1\frac{t_1}{c_1}\overline{V}_6+15\overline{E}_2\frac{t_1^2}{c_1^2}\overline{V}_4-15\overline{E}_3\frac{t_1^3}{c_1^3} = -136t_1^2-360t_1^4-225t_1^6, \tag{7.79}$$

where the coefficient of x^6 is evaluated in the spherical limit ($\beta_1=1$).

Combining the results for λ_x, $\sec\gamma(x,y)$ and the r_n from equations 7.51, 7.69 and 7.57 respectively gives

$$\lambda_x \sec\gamma = \frac{1}{\nu_1 c_1}\left[1+\frac{1}{2}\tilde{x}^2 p_2+\frac{1}{24}\tilde{x}^4 p_4+\frac{1}{720}\tilde{x}^6 \bar{p}_6\right], \tag{7.80}$$

where

$$\begin{aligned}
p_2 &= -\beta_1 - t_1^2 & \bar{p}_2 &= -1 - t_1^2\\
p_4 &= -4\beta_1^3(1-6t_1^2)+\beta_1^2(9-52t_1^2)+42\beta_1 t_1^2+9t_1^4 & \bar{p}_4 &= 5+14t_1^2+9t_1^4\\
\bar{p}_6 &= -61-331t_1^2-495t_1^4-225t_1^6.
\end{aligned} \tag{7.81}$$

Using (E.32) to evaluate the inverse

$$\frac{1}{\lambda_x \sec\gamma} = \nu_1 c_1\left[1+\frac{1}{2}\tilde{x}^2 g_2+\frac{1}{24}\tilde{x}^4 g_4+\frac{1}{720}\tilde{x}^6 \bar{g}_6\right], \tag{7.82}$$

where

$$\begin{aligned}
g_2 &= -p_2 && = \beta_1 + t_1^2,\\
g_4 &= -p_4 + 6p_2^2 && = 4\beta_1^3(1-6t_1^2)+\beta_1^2(-3+52t_1^2)-30\beta_1 t_1^2-3t_1^4,\\
\bar{g}_6 &= -\bar{p}_6+30\bar{p}_2\bar{p}_4-90\bar{p}_2^3 = 1+31t_1^2+75t_1^4+45t_1^6.
\end{aligned} \tag{7.83}$$

Therefore, combining 7.78 and 7.82,

$$k(x,y) = \frac{k(\psi)}{\lambda_x \sec\gamma} = \left[1+\frac{1}{2}\tilde{x}^2 K_2+\frac{1}{24}\tilde{x}^4 K_4+\frac{1}{720}\tilde{x}^6 \bar{K}_6\right], \qquad \hat{x} = \frac{x}{\nu_1}, \tag{7.84}$$

with

$$\begin{aligned}
K_2 &= q_2 + g_2 && = \beta_1,\\
K_4 &= q_4 + 6q_2 g_2 + g_4 && = 4\beta_1^3(1-6t_1^2)-3\beta_1^2(1-16t_1^2)-24\beta_1 t_1^2,\\
\bar{K}_6 &= q_6 + 15q_4 g_2 + 15 q_2 g_4 + g_6 = 1.
\end{aligned} \tag{7.85}$$

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.

Comment: all of the preceeding series may be modified for a secant version of the projection which has a wider region within a prescribed scale tolerance. The series may also be applied to an arbitrary central meridian by replacing λ by $\lambda-\lambda_0$. The following summary includes both of these factors, k_0 and λ_0.

7.8 Redfearn's modified (secant) TME series

Direct series

As for NMS, TMS and NME simply multiply (7.28) and (7.29) by a factor of k_0.

$$x(\lambda,\phi) = k_0\nu\left[\widetilde{\lambda} + \frac{\widetilde{\lambda}^3}{3!}W_3 + \frac{\widetilde{\lambda}^5}{5!}W_5 + \frac{\widetilde{\lambda}^7}{7!}\overline{W}_7\right], \qquad \widetilde{\lambda} = (\lambda-\lambda_0)c \tag{7.86}$$

$$y(\lambda,\phi) = k_0\left[m(\phi) + \frac{\widetilde{\lambda}^2\nu t}{2} + \frac{\widetilde{\lambda}^4\nu t}{4!}W_4 + \frac{\widetilde{\lambda}^6\nu t}{6!}W_6 + \frac{\widetilde{\lambda}^8\nu t}{8!}\overline{W}_8\right]. \tag{7.87}$$

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 (7.5) or (7.6) with $y\to y/k_0$: it is the solution of $m(\phi_1) = y/k_0$. Equations 7.37, 7.38 and 7.43 become

$$\lambda(x,y) = \lambda_0 + \frac{\widehat{x}}{c_1} - \frac{\widehat{x}^3}{3!c_1}V_3 - \frac{\widehat{x}^5}{5!c_1}V_5 - \frac{\widehat{x}^7}{7!c_1}\overline{V}_7, \qquad \widehat{x} = \frac{x}{k_0\nu_1} \tag{7.88}$$

$$\psi(x,y) = \psi_1 - \frac{\widehat{x}^2 t_1}{2c_1} - \frac{\widehat{x}^4 t_1}{4!c_1}V_4 - \frac{\widehat{x}^6 t_1}{6!c_1}V_6 - \frac{\widehat{x}^8 t_1}{8!c_1}\overline{V}_8. \qquad m(\phi_1) = \frac{y}{k_0} \tag{7.89}$$

$$\phi(x,y) = \phi_1 - \frac{\widehat{x}^2\beta_1 t_1}{2} - \frac{\widehat{x}^4\beta_1 t_1}{4!}U_4 - \frac{\widehat{x}^6\beta_1 t_1}{6!}U_6 - \frac{\widehat{x}^8\beta_1 t_1}{8!}\overline{U}_8, \tag{7.90}$$

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 7.74, 7.62, 7.84 and 7.67 become

$$k(\lambda,\phi) = k_0\left[1 + \frac{1}{2}\widetilde{\lambda}^2 H_2 + \frac{1}{24}\widetilde{\lambda}^4 H_4 + \frac{1}{720}\widetilde{\lambda}^6\overline{H}_6\right], \qquad \widetilde{\lambda} = (\lambda-\lambda_0)c, \tag{7.91}$$

$$\gamma(\lambda,\phi) = \widetilde{\lambda}t + \frac{1}{3}\widetilde{\lambda}^3 t H_3 + \frac{1}{15}\widetilde{\lambda}^5 t H_5 + \frac{1}{315}\widetilde{\lambda}^7 t\overline{H}_7, \tag{7.92}$$

$$k(x,y) = k_0\left[1 + \frac{1}{2}\widehat{x}^2 K_2 + \frac{1}{24}\widehat{x}^4 K_4 + \frac{1}{720}\widehat{x}^6\overline{K}_6\right], \qquad m(\phi_1) = \frac{y}{k_0}, \tag{7.93}$$

$$\gamma(x,y) = \widehat{x}t_1 + \frac{1}{3}\widehat{x}^3 t_1 K_3 + \frac{1}{15}\widehat{x}^5 t_1 K_5 + \frac{1}{315}\widehat{x}^7 t_1\overline{K}_7, \qquad \widehat{x} = \frac{x}{k_0\nu_1}. \tag{7.94}$$

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 such that $m(\phi_1) = y/k_0$. For convenience all of the required coefficients are collected on the following page.

All coefficients

$$x(\lambda,\phi)\qquad W_3 = \beta - t^2$$
$$W_5 = 4\beta^3(1-6t^2)+\beta^2(1+8t^2)-2\beta t^2+t^4$$
$$\overline{W}_7 = 61-479t^2+179t^4-t^6+O(e^2)$$

$$y(\lambda,\phi)\qquad W_4 = 4\beta^2+\beta-t^2$$
$$W_6 = 8\beta^4(11-24t^2)-28\beta^3(1-6t^2)+\beta^2(1-32t^2)-2\beta t^2+t^4$$
$$\overline{W}_8 = 1385-3111t^2+543t^4-t^6+O(e^2)$$

$$\lambda(x,y)\qquad V_3 = \beta_1+2t_1^2$$
$$V_5 = 4\beta_1^3(1-6t_1^2)-\beta_1^2(9-68t_1^2)-72\beta_1 t_1^2-24t_1^4$$
$$\overline{V}_7 = 61+662t_1^2+1320t_1^4+720t_1^6$$

$$\psi(x,y)\qquad V_4 = 4\beta_1^2-9\beta_1-6t_1^2$$
$$V_6 = 8\beta_1^4(11-24t_1^2)-84\beta_1^3(3-8t_1^2)+225\beta_1^2(1-4t_1^2)+600\beta_1 t_1^2+120t_1^4$$
$$\overline{V}_8 = -1385-7266t_1^2-10920t_1^4-5040t_1^6$$

$$\phi(x,y)\qquad U_4 = 4\beta_1^2-9\beta_1(1-t_1^2)-12t_1^2$$
$$U_6 = 8\beta_1^4(11-24t_1^2)-12\beta_1^3(21-71t_1^2)+15\beta_1^2(15-98t_1^2+15t_1^4)$$
$$+180\beta_1(5t_1^2-3t_1^4)+360t_1^4$$
$$\overline{U}_8 = -1385-3633t_1^2-4095t_1^4-1575t_1^6$$

$$k(\lambda,\phi)\qquad H_2 = \beta$$
$$H_4 = 4\beta^3(1-6t^2)+\beta^2(1+24t^2)-4\beta t^2$$
$$\overline{H}_6 = 61-148t^2+16t^4$$

$$\gamma(\lambda,\phi)\qquad H_3 = 2\beta^2-\beta$$
$$H_5 = \beta^4(11-24t^2)-\beta^3(11-36t^2)+\beta^2(2-14t^2)+\beta t^2$$
$$\overline{H}_7 = 17-26t^2+2t^4$$

$$k(x,y)\qquad K_2 = \beta_1$$
$$K_4 = 4\beta_1^3(1-6t_1^2)-3\beta_1^2(1-16t_1^2)-24\beta_1 t_1^2$$
$$\overline{K}_6 = 1$$

$$\gamma(x,y)\qquad K_3 = 2\beta_1^2-3\beta_1-t_1^2$$
$$K_5 = \beta_1^4(11-24t_1^2)-3\beta_1^3(8-23t_1^2)+5\beta_1^2(3-14t_1^2)+30\beta_1 t_1^2+3t_1^4$$
$$\overline{K}_7 = -17-77t_1^2-105t_1^4-45t_1^6 \qquad (7.95)$$

Comments

1. The series given on the previous page are in full accordance with those printed in Redfearn (1948). They differ slightly in format since Redfearn writes the series in terms of $\rho(\phi)$ and $\nu(\phi)$ rather than $\beta(\phi)$ and $\nu(\phi)$. In Redfearn's paper a few '1' subscripts are omitted in the inverse series.

2. The series are also in full accordance with those printed in Thomas (1952) but he uses a different notation which will be exhibited in the following chapter. Thomas gives the full seventh and eighth order terms without the spherical approximation. (See page 95 *et. seq.* Beware his use of ρ as a conversion factor to radian measure.)

3. In Chapter 8 we discuss the series in the context of two important applications, UTM (1989) and OSGB (1999).

4. The TME projection is used with $k_0 = 0.9996$ by both UTM and OSGB.

5. For OSGB the central meridian is at $\lambda_0 = -2$. UTM is a set of 60 different TME projections with $\lambda_0 = -177, -171, \ldots -3, 3, \ldots 171, 177$ degrees.

6. For computational purposes the series should be written in a 'nested' form. For example equation (7.86) can be written as

$$x(\lambda,\phi) = \nu\tilde{\lambda}\left[1+\tilde{\lambda}^2\left\{\frac{W_3(\phi)}{3!}+\tilde{\lambda}^2\left(\frac{W_5(\phi)}{5!}+\tilde{\lambda}^2\frac{\overline{W}_7(\phi)}{7!}\right)\right\}\right], \qquad \tilde{\lambda} = (\lambda-\lambda_0)c. \tag{7.96}$$

Implementation

The transformations presented in this chapter are incorporated into several online calculators:

1. Geotrans (2010) is provided by the National geospatial intelligence agency. The code is based on the slightly different formulae of Thomas (1952, page 2) who uses the parameter η^2 instead of the β used here. The relation between them is

$$\eta^2 = \beta - 1 = \frac{\nu}{\rho} - 1 = \frac{e^2\cos^2\phi}{1-e^2} = e'^2\cos^2\phi. \tag{7.97}$$

where e' is the second eccentricity. To complicate matters the actual code of Geotrans uses `eta` in place of the η^2 of Thomas. (See code at lines 223, 472–503). Geotrans converts between geodetic and UTM coordinates.

Chapter 8

Applications of TME

Abstract

Coordinates, grids and origins. UTM. The British grid (NGGB). Scale variation and convergence. Accuracy of TME series. Approximations to the series. The OSGB series.

8.1 Coordinates, grids and origins

The transverse Mercator projection of the ellipsoid (TME) is used officially in the United Kingdom (OSGB, 1999), the United States, Ireland, Sweden, Norway, Finland, Poland, Russia, China, ... (For a more complete list see Stuifbergen (2009), page 21). TME is also the basis for world-wide systematic coverage by the 60 Universal Transverse Mercator projections (UTM, 1989). In each case the projections are applied to geographical coordinates (λ, ϕ) defined with respect to different datums for each of which the ellipsoid parameters (a, e) and the longitude of the central meridian must be specified before applying the Redfearn formulae of the previous chapter.

Countries of narrow extent may be covered by a single zone: for example the National Grid of Great Britain (NGGB) has a width of almost 10°. Most countries require several zones. Zone widths vary but few exceed a value of 10° at which errors in the transformations start to become significant. The scale factor on the central meridian also varies: NGGB and UTM use 0.9996, South Africa uses 1, Canada uses 0.9999 *etc.*

The Redfearn formulae determine coodinates (x, y) which we shall refer to as **projection coordinates**; their origin is always at the intersection of the central meridian and the equator and their axes are usually designated as x along the equator and y normal to the equator. Projection coordinates must be distinguished from **grid coordinates**. Grids are normally aligned to the underlying Cartesian system of the projection coordinates but the **true origin** of the grid may differ from the origin of the projection coordinates. For example the true origin of a UTM grid coincides with that of the projection coordinates but for the British grid the true origin is at a latitude of 49°N on the central meridian.

Since the true origin of a grid is always on the central meridian it follows that grid coordinate for points west or south of the true origin are negative. To avoid such negative coordinates the grid coordinates are usually referred to a **false origin** chosen so that all values are positive. Such grid coordinates, rounded to the nearest metre, are normally called the **easting** and **northing**, denoted by E and N. Examples of true and false origins will be given in the following sections.

8.2 The UTM projection

UTM is a set of 60 TME projections based on the WGS (1984) ellipsoid; each is restricted to 3° in longitude from the central meridians at $\lambda_0 = -177, -171, -165\ldots.$ Zones 16, 34 and 53, centred on $\lambda_0 = -87, 21, 135$, are shown below alongwith an enlarged view of zone 34. In each case the UTM grid is restricted to the red rectangle but the actual zones are bounded by meridians at 3°from the central meridian and parallels at 84°N and 80°S.

$\lambda_0 = -87°$

$\lambda_0 = 21°$

$\lambda_0 = 135°$

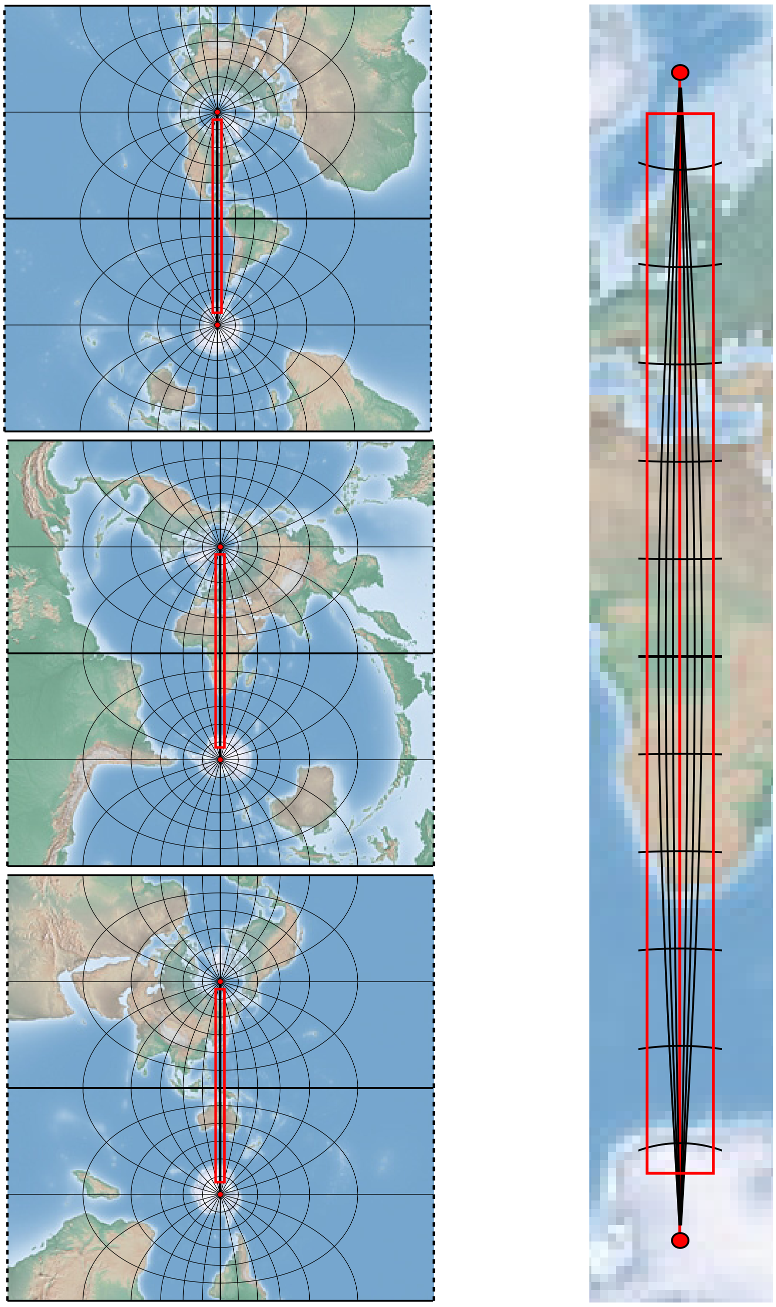


Figure 8.1

.

8.3 UTM coordinate systems

Figure 8.2 shows UTM 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 figure is not to scale: the true shape of any zone is shown on the right of Figure 8.1. The true origin T of the UTM grid coincides with the origin of the projection at O where the central meridian meets the equator at 3°W. We treat the hemispheres differently and introduce **two false origins**; both are 500000m 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 10000000m below the equator on the projection.Therefore the eastings and northings for points in the northern and southern hemispheres are

$$E = E0 + x(\lambda,\phi), \qquad E0 = 500000 \tag{8.1}$$

$$N = N0 + y(\lambda,\phi) \qquad N0 = \begin{cases} 0 & \text{(N)} \\ 10000000 & \text{(S)} \end{cases} \tag{8.2}$$

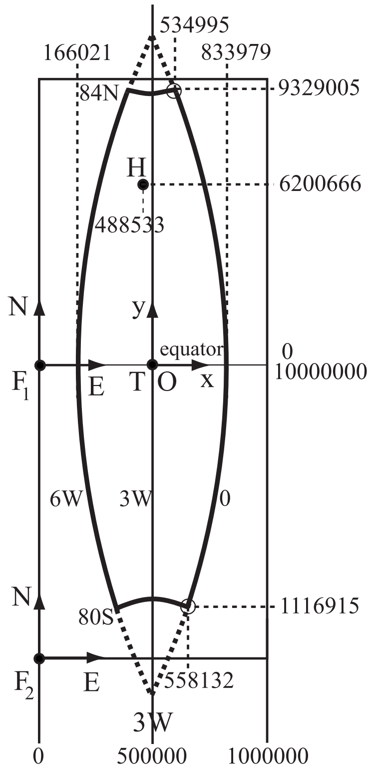


Figure 8.2

where x, y are calculated (to the nearest metre) using the series of Section 7.8 with $\lambda_0 = -3°$, k_0 =0.9996 and the parameters (a,e) of the WGS ellipsoid given in 5.2. Points on the equator are defined to be part of the northern grid so they must have $N = 0$. Note that $(E0, N0)$ are the coordinates of the true origin (at T or O) relative to the false origin on each grid. The inverse relations are calculated from Section 7.8 with the replacement of x by $E - E0$ and coefficients evaluated at a footpoint latitude such that

$$m(\phi_1) = \frac{y}{k_0} = \frac{N - N0}{k_0}. \tag{8.3}$$

The figure shows eastings and northings of several points on the perimeter of the zone. The coordinates may be calculated by using Geotrans (2010) (see also Implementation) or by the web conversion program GeoConvert which is downloadable from Karney (2010) The most easterly (respectively westerly) points of zone 30 have $\phi = 0$ and $\lambda - \lambda_0 = \pm 3° = \pm 0.05236$rad, for which E=166021 (respectively 833979). Therefore the maximum width of a UTM zone is 667.958km in terms of the projection coordinates. This is not exactly the same as the corresponding distance on the WGS ellipsoid, *viz.* $2\pi a/60$=667.917km because the scale factor on the equator is not unity (and not constant). See Section 8.5.

The figure also shows the coordinates at the most northerly and southerly points of the projection where the bounding meridians at $\lambda = \lambda_0 \pm 3$ meet the parallels at 84°N and 80°S. The northings of the two northernmost points are given by $N = 9329005$ on the northern grid. Their eastings are given by $E = 465005$ and $E = 534995$. The coordinate separation of these points is 69.990km whereas the rhumb line distance measured along the parallel at latitude 84°N is $\nu \cos\phi\, \delta\lambda = 70.049$km and the great circle distance is 70.017km.

UTM grid reference system

Eastings and Northings are measured in metres and provide a purely numeric grid reference system within each UTM zone. To specify a point uniquely on the the Earth, at least between latitudes 84°N and 80°S, requires a zone number, a hemisphere specifier (N or S), the easting and the northing. Therefore my armchair location is at

30N 488533 6200666

In this work we do not discuss the coordinate systems used in the polar regions. These are discussed in the DMA publications, UTM (1989).

The military grid reference system (MGRS)

MGRS is alpha-numeric with the following components in addition to the zone number.

- 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′4.4″N, 3°11′1.1″W using WGS coordinates) is just within the northern limit of zone/band 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. There are minor adjustments of the zone/band scheme in northern Europe.
- Within the sub-zones the 100km squares are labelled with row and column letters. For a full description of the labelling of the 100km squares see the DMA manual: (UTM, 1989, Volume 1, Chapter 3). For example Edinburgh is in the 100km square labelled UG inzone 30 and band U. Thus 30UUG fixes my home to within 100km. Boundary points are included in a square if they are on the west or south edges.
- 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 30UVH 88533 00666. The leading zeroes must be specified.
- Such precision is often superfluous and two numbers of 4, 3, 2, 1 digits may be used for accuracy to within 10m, 100m, 1km, 10km. Thus
  - 30UVH 8853 0066 fixes my home to within 10m
  - 30UVH 885 006 . . . . 100m
  - 30UVH 88 00 . . . . 1km
  - 30UVH 8 0 . . . . 10km

8.4 The British national grid: NGGB

The British national grid (NGGB) (OSGB, 1999) is a grid overlain on the TME projection centred on longitude 2°W with coordinates defined on the Airy 1830 ellipsoid, for which the equatorial radius is 6377563.396m and the inverse flattening is 299.3249753. (The derived minor axis is 6356256.910m and the derived eccentricity is 0.0816733724.) It is a secant projection with a value of $k_0 = 0.9996012717$ on the central meridian. The pre-1936 value was 0.9996 but after the 1936 re-survey the value of k_0 was adjusted so that the coordinates of a selected group of locations were as close as possible to their previous values. For most practical purposes, and in the remainder of this chapter, we will take $k_0 = 0.9996$.

The figure shows the central and northern part of the TME projection with $\lambda_0 = -2°$, (2°W). The origin of the projection coordinates is at the point P where the central meridian intersects the equator. The graticule meridians shown are at intervals of 15°measured from the central meridian; the Greenwich meridian is also indicated. The NGGB grid covers only a small part of the projection, just the small box shown in the figure. The true origin for the grid is taken at the point T with latitude $\phi_0 = 49°$ (0.855rad) on the central meridian. The false origin is then chosen west and north of the true origin with E0=400000m and N0=−100000m. The 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.

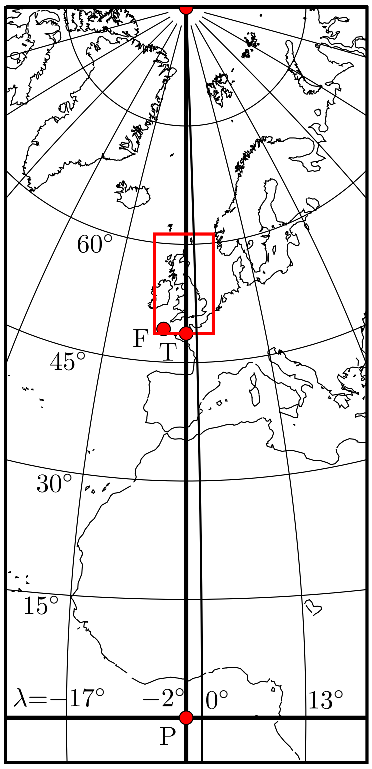


Figure 8.3

Figure 8.4 shows the NGGB grid in greater detail. It covers longitudes from 2°E to almost 8°W, an interval of 10 degrees compared to the 6 degrees of UTM. The longitude range is not symmetric about the central meridian—4 degrees to the east but 6 degrees to the west. This means that NGGB requires use of the Redfearn transformation formulae at up to 6 degrees from the central meridian when we consider locations on the Outer Hebrides.

The grid extends 700km east of the false origin and 1300km north. The actual grid area is divided into 100km squares coded with letter pairs which are used in the alpha-numeric grid reference scheme. That part of the grid which is excluded by virtue of the overlap with the Irish grid is shown by the dashed line in the figure. There are two Irish grid systems in use at the moment. The old system is based on a slightly modified Airy ellipsoid with a true origin on at 53°30′N, 8° W and a false origin such that E0=200000m and N0=250000m. The scale modification is $k_0 = 1.000035$. The new system is based on the WGS84 ellipsoid with the same true origin but E0=600000m and N0=750000m.

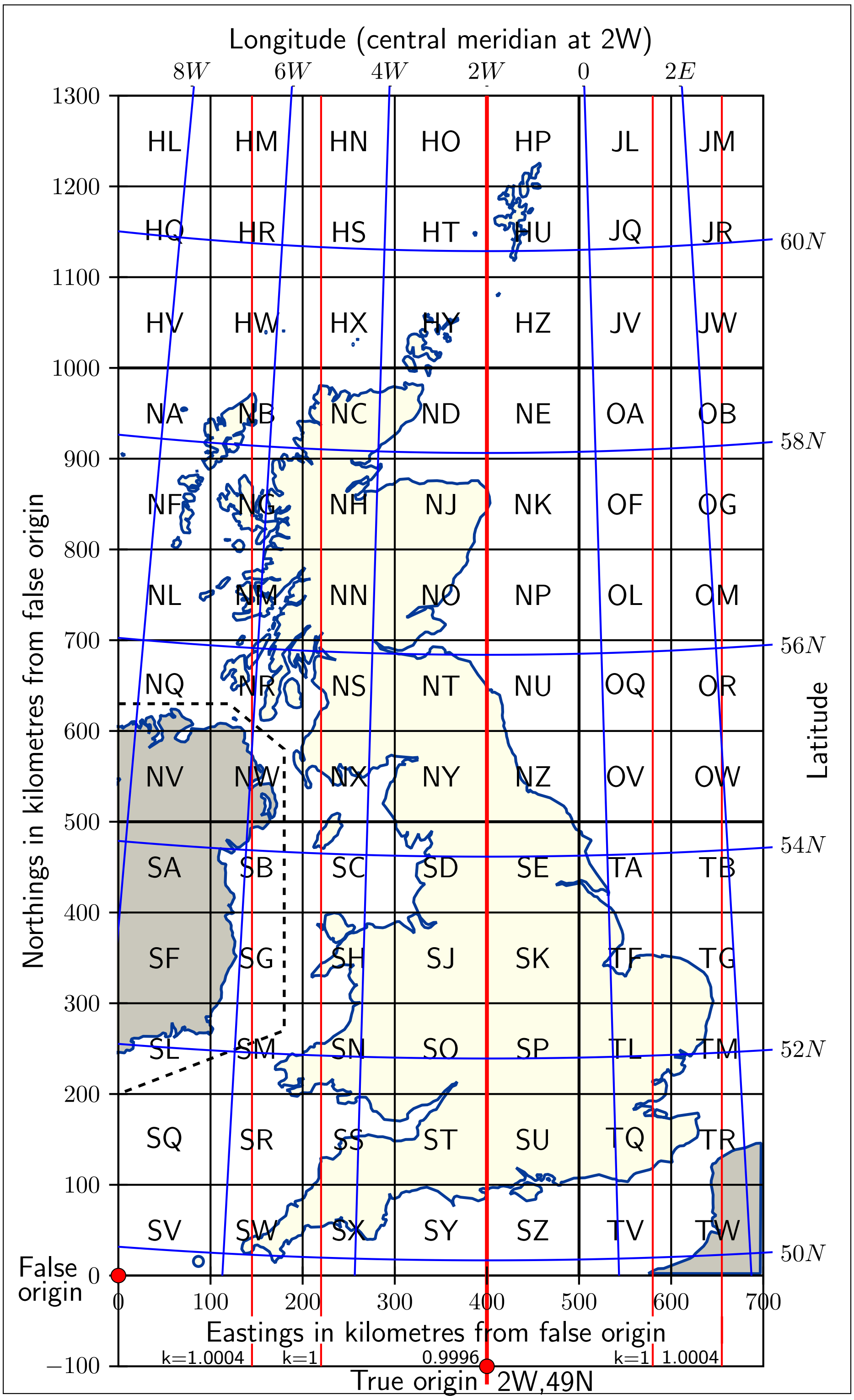


Figure 8.4 The region covered by the British National Grid (NGGB).

The scale factor varies by small amounts over the grid. On the central meridian it is constant at the nominal value 0.9996. The scale increase by 0.0004 to $k = 1$ on two lines which are approximately straight and approximately 180km from the central meridian. The curvature of these isoscale lines is hardly detectable in the figure. The scale increases by a further factor of 0.0004 on two curved lines approximately 255km from the central meridian, a further distance of 75km compared with the first 180km. Outside of these lines the scale factor starts to increase rapidly showing that the TME projections must be restricted to limited longitude ranges. See Section 8.5 for more detail.

With the exception of the central meridian all meridians and parallels are *curved* with values of latitude and longitude indicated on the right and top sides of the outer rectangle. As a consequence true north differs from grid north except on the central meridian, the difference being given by the convergence. All parallels are concave upwards and true east agrees with grid lines only at the point where the parallel crosses the central meridian. The difference between true east and grid east is also determined by the convergence. The curvature of meridians and parallels is small but noticeable and important in high accuracy survey work. Numerical values are discussed in Section 8.6

NGGB grid references

The relation between projection coordinates and eastings and northings must take into account the fact that the true origin is not on the equator. Since $y(0,\phi_0) = k_0 m(\phi_0)$

$$E = E0 + x(\lambda,\phi) \tag{8.4}$$

$$N = N0 + [y(\lambda,\phi) - k_0 m(\phi_0)], \tag{8.5}$$

$\phi_0 = 49° = 0.8552$ radians and the distance of the true origin from the equator (on the Airy ellipsoid) is $m(\phi_0) = m(0.8552) = 5429228.6$m. The true origin is E0=400000m east and N0=-100000m north of the false origin. Therefore

$$E = x + 400000, \tag{8.6}$$

$$N = y - 5527064, \tag{8.7}$$

where x and y are calculated from the series of Section 7.8 The inverse transformations are also calculated from Section 7.8 with x replaced by $E - E0$ and ϕ_1 calculated from

$$m(\phi_1) = \frac{y}{k_0} = \frac{N - N0}{k_0} + m(\phi_0) \tag{8.8}$$

My position on the NGGB datum is 55° 57′ 4.6″N, 3° 10′ 55.9″W (obtained from the WGS value by using Geotrans (2010)). Using the above formulae I calculate that, to the nearest metre, I am sitting at E=326185, N=673765. This is in the 100km square labelled NT so the alpha-numeric grid reference is NT 26185 73765 to within 1m. As for UTM we can use shorter grid references such as NT26187376, NT261737, NT2673, NT27 for accuracy to within 10m, 100m, 1km, 10km. (Grid references truncated).

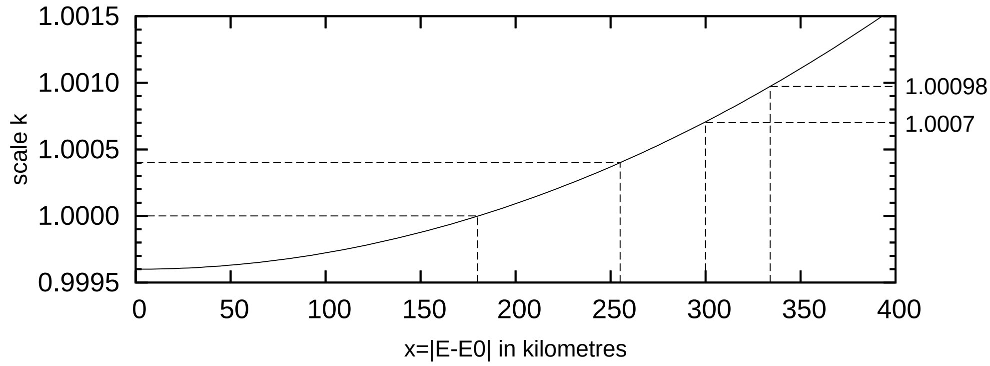


Figure 8.5

8.5 Scale variation in TME projections

Scale variation in TMS

Since the differences between TMS and TME are small it is instructive to consider the TMS scale factor (equation 3.75 and Section 3.4):

$$k(x,y)\Big|_{\text{TMS}} = k_0 \cosh\widehat{x} = k_0\left[1+\frac{1}{2}\widehat{x}^2+\frac{1}{24}\widehat{x}^4+\frac{1}{720}\widehat{x}^6+\cdots\right] \qquad \widehat{x}=\frac{x}{k_0 R} \tag{8.9}$$

This function is plotted above for $k_0 = 0.9996$ and $R = 6372$km (the tri-axial mean for the WGS ellipsoid). The values of k at which the scale factor equals 1 and 1.0004 (**isoscale** lines) are given by inverting equation 8.9 to $x = k_0 R\cosh^{-1}(k/k_0)$. The values are 180km and 255km respectively. The other values indicated at $x=$300km and $x=$333km correspond to the maximum values of x attained in a UTM zone and the NGGB grid respectively.

There is no variation of scale on a grid line with constant y. However, as we move northwards from the equator on a meridian λ remains constant, x decreases and therefore the scale factor decreases. The variation of scale on the meridian is given by the segment of the above graph which lies between the x-values at the ends of the meridian segment. For example, if $\lambda = 3°$ then it has $x = 333$km on the equator and $x = 34.9$km at latitude 84°N: the scale decreases from 1.00098 to 0.999614 on the meridian between these values.

Scale variation in TME

The scale factor in projection coordinates for TME is given by equation (7.93):

$$k(x,y) = k_0\left[1+\frac{1}{2}\widehat{x}^2 K_2+\frac{1}{24}\widehat{x}^4 K_4+\frac{1}{720}\widehat{x}^6\overline{K}_6\right] \qquad \widehat{x}=\frac{x}{k_0\nu_1} \tag{8.10}$$

$$\begin{aligned} &K_2 = \beta_1\,,\\ &K_4 = 4\beta_1^3(1-6t_1^2)-3\beta_1^2(1-16t_1^2)-24\beta_1 t_1^2,\\ &\overline{K}_6 = 1, \end{aligned} \tag{8.11}$$

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 (8.2) or (8.7) for UTM or NGGB respectively. To simplify the comparison of the TMS and TME scale factors it is useful, to write K_2 and K_4 in terms of the η parameter which is used in the TME series as given by Thomas (1952) and OSGB (1999). It is defined as:

$$\eta^2 = \beta - 1 = \frac{\nu}{\rho} - 1 = \frac{e^2\cos^2\phi}{1-e^2} = e'^2\cos^2\phi. \tag{8.12}$$

The non-trivial coefficients become

$$\begin{aligned} K_2 &= 1 + \eta_1^2 \\ K_4 &= 1 + 6\eta_1^2 + \eta_1^4(9 - 24t_1^2) + \eta_1^6(4 - 24t_1^2) \end{aligned} \tag{8.13}$$

As $e^2, \eta^2 \to 0$ the TME scale factor reduces to the TMS scale factor. Their difference is given by terms of order $\eta_1^2\widehat{x}^2$ and $\eta_1^2\widehat{x}^4$ which are never more than 10^{-5} and 10^{-8} respectively over the projection region. Therefore Figure 8.5 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 TME series we must not assume that the t_1 terms are small: at a footpoint latitude of 80° we have $t_1 = \tan\phi_1 \approx 5.7$ and $t_1^2 \approx 32$. On the other hand at large latitudes the values of x are much less than their maximum values at the equator. These two factors compensate to some extent.

Isoscale lines of TME

To find the isoscale lines in TME we must invert invert equation 8.10. To do this we write the equation as

$$w \equiv \frac{2}{K_2}\left(\frac{k}{k_0} - 1\right) = \widehat{x}^2 + \frac{1}{12}\widehat{x}^4\frac{K_4}{K_2} + \frac{1}{360}\widehat{x}^6\frac{\overline{K}_6}{K_2} \qquad \widehat{x} = \frac{x}{k_0\nu_1}. \tag{8.14}$$

The inverse is evaluated by using the Lagrange series reversion given in equations (B.10–B.11) with z replaced by $\widehat{x}^2$:

$$\widehat{x}^2 = w - \left(\frac{K_4}{12K_2}\right)w^2 - \left(\frac{K_2\overline{K}_6 - 5K_4^2}{360K_2^2}\right)w^3 \qquad w = \frac{2}{K_2}\left(\frac{k}{k_0} - 1\right), \tag{8.15}$$

from which we can find $x = k_0\nu_1\widehat{x}$, and hence $(E - E0)$, as a function of k at any given value of N, *i.e.* $E(k,N)$. Results are given below.

Numerical scale factors for UTM

The UTM zone has greatest width on the equator where $x \approx 334$km and the scale factor at the point (E=833979 N=0) is just under 1.001. Comparing this with the scale factor 0.9996 at the origin indicates a scale variation of about 0.15% over the width of the projection. The scale factor at the upper right corner of the projection (E=534995, N=9329005) is 0.999615 and on the equator at (E=534995, N=0) it is 0.9996152: the variation with latitude along a grid line is almost zero. The variation along the meridian is about 0.15%.

To draw the isoscale lines we have taken four values of N and calculated y and m from equation 8.3 and inverted this to obtain the footpoint latitude by the methods of Section 5.9. The coefficients K_2 are then calculated from 8.13 so that x and $E-E_0$ follow from 8.15 with $k=1$ and then $k=1.0004$. The results are shown in the following table.

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

Table 8.1

From the fourth column we see that the eastings on the $k=1$ isoscale increase by just over 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 almost parallel to the central meridian but curve *outwards* a little as they move from the equator..

Numerical scale factors for NGGB

For NGGB the greatest value of scale are encountered on the coast of East Anglia coast and the Outer Hebrides where $|x|$ reaches 300km and $k\approx1.0007$. However, the bulk of the NGGB grid is within approximately 255km of the central meridian where $|k-1|<0.0004$. North-south variation is negligible over the NGGB. Note that a typical map 1:50000 sheet corresponds to a small section of the grid. My 'local' sheet is bounded by E=316km and E=356km and is 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. The isoscale lines of NGGB are found in exactly the same way as for UTM. The curvature is negligible, giving a deviation of only 212m over the full northings extent of 1200km.

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

Table 8.2

8.6 Convergence in the TME projection

When we 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 $\lambda_0+3°$) and on the extreme western meridian for NGGB at 7°W=$\lambda_0-5°$ (in square NF)..

Convergence along a projected meridian			
UTM at $\lambda_0+3°$		NGGB at $\lambda_0-5°$	
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 of convergence approach the limiting value of 3° which is 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 NGGB 1:50000 map sheets. For example exact calculations give the convergence at the corners of the Edinburgh sheet bounded by E=316km, E=356km, N=650km and N=690km as

Convergence on the 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 of the sheet may be approximated by linear interpolation.

It is important to observe that convergence 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 3.5. This correction is important in high accuracy applications.

8.7 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 (7.86, 7.87) and then reverse the transformation with the inverse series (7.88, 7.90) 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 Table (2.1) that $0.0001''$ is equivalent to 3mm along the meridian and less than 2mm along a parallel for locations 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 tests 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 (8.6, 8.7) and the footpoint is to be calculated from (8.8).

	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 below.

	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 $0.0001''$. This applies to a typical point on the mainland but the errors in the Hebrides are more appreciable.

When Redfearn (1948) published his series he was 'simply' extending the series that had been published earlier by Lee (1945) 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 proved to be the case. Redfearn's analysis to higher order makes clear which terms can be safely omitted, as is in OSGB (1999), Thomas (1952) and Snyder (1987) To investigate the size of the terms we again introduce the parameter η^2 defined in equation (8.12). Using (8.6–8.8) the equations of Section 7.8 become

$$E(\lambda,\phi) = E0 + k_0\nu\left[\tilde{\lambda} + \frac{\tilde{\lambda}^3}{3!}W_3 + \frac{\tilde{\lambda}^5}{5!}W_5 + \frac{\tilde{\lambda}^7}{7!}\overline{W}_7\right], \qquad \tilde{\lambda} = (\lambda - \lambda_0)c \tag{8.16}$$

$$N(\lambda,\phi) = N0 + k_0\left[m(\phi) - m(\phi_0) + \frac{\tilde{\lambda}^2\nu t}{2} + \frac{\tilde{\lambda}^4\nu t}{4!}W_4 + \frac{\tilde{\lambda}^6\nu t}{6!}W_6 + \frac{\tilde{\lambda}^8\nu t}{8!}\overline{W}_8\right], \tag{8.17}$$

$$\lambda(E,N) = \frac{\hat{x}}{c_1} - \frac{\hat{x}^3}{3!\,c_1}V_3 - \frac{\hat{x}^5}{5!\,c_1}V_5 - \frac{\hat{x}^7}{7!\,c_1}\overline{V}_7, \qquad m(\phi_1) = \frac{N - N0}{k_0} + m(\phi_0) \tag{8.18}$$

$$\phi(E,N) = \phi_1 - \frac{\hat{x}^2\beta_1 t_1}{2} - \frac{\hat{x}^4\beta_1 t_1}{4!}U_4 - \frac{\hat{x}^6\beta_1 t_1}{6!}U_6 - \frac{\hat{x}^8\beta_1 t_1}{8!}\overline{U}_8, \qquad \hat{x} = \frac{E - E0}{k_0\nu_1} \tag{8.19}$$

with coefficients

$$\begin{aligned}
W_3 &= 1 - t^2 + \eta^2 \;\big|\\
W_4 &= 5 - t^2 + 9\eta^2 \;\big|\; + 4\eta^4\\
W_5 &= 5 - 18t^2 + t^4 + \eta^2(14 - 58t^2) \;\big|\; + \eta^4(13 - 64t^2) + \eta^6(4 - 24t^2)\\
W_6 &= 61 - 58t^2 + t^4 \;\big|\; + \eta^2(270 - 330t^2) + \eta^4(445 - 680t^2)\\
&\qquad\qquad + \eta^6(324 - 600t^2) + \eta^8(88 - 192t^2)\\
\overline{W}_7 &= \;\big|\; 61 - 479t^2 + 179t^4 - t^6\\
\overline{W}_8 &= \;\big|\; 1385 - 3111t^2 + 543t^4 - t^6
\end{aligned} \tag{8.20}$$

$$\begin{aligned}
V_3 &= 1 + 2t_1^2 + \eta_1^2 \;\big|\\
V_5 &= -5 - 28t_1^2 - 24t_1^4 \;\big|\; - \eta_1^2(6 + 8t_1^2) + \eta_1^4(3 - 4t_1^2) + \eta_1^6(4 - 24t_1^2)\\
\overline{V}_7 &= 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6 \;\big|
\end{aligned} \tag{8.21}$$

$$\begin{aligned}
U_4 &= -5 - 3t_1^2 - \eta_1^2(1 - 9t_1^2) \;\big|\; + 4\eta_1^4\\
U_6 &= 61 + 90t_1^2 + 45t_1^4 \;\big|\; + \eta_1^2(46 - 252t_1^2 - 90t_1^4) + \eta_1^4(-3 - 66t_1^2 + 225t_1^4)\\
&\qquad\qquad + \eta_1^6(100 + 84t_1^2) + \eta_1^8(88 - 192t_1^2)\\
\overline{U}_8 &= \;\big|\; - 1385 - 3633t_1^2 - 4095t_1^4 - 1575t_1^6
\end{aligned} \tag{8.22}$$

The significance of the vertical rules in the rewritten coefficients is discussed overleaf.

Consider a worst case example: $\phi = 58°$N and $\lambda = 7°$W, equivalent to projection co-ordinates 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 below 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. Observe that all terms to the right of the rules are smaller than our limits (1mm or 0.0001″) and may be neglected in the series.

In: $\phi = 58°$ $\lambda = -7°$				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: $\phi = 58°$ $\lambda = -7°$				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″				

Table 8.3

8.8 The truncated TME series

Dropping the small terms to the right of the vertical rules in equations (8.16–8.19) gives the truncated series used by OSGB (1999)—see next section. A slightly smaller set of terms is omitted in the series in Snyder (1987) because UTM extends to much higher latitudes than NGGB.

$$E(\phi,\lambda) = E0 + k_0\nu\left[\widetilde{\lambda} + \frac{\widetilde{\lambda}^3}{3!}W_3^T + \frac{\widetilde{\lambda}^5}{5!}W_5^T\right], \tag{8.23}$$

$$N(\phi,\lambda) = N0 + k_0\left[m(\phi) - m(\phi_0)\right] + k_0\left[\frac{\widetilde{\lambda}^2\nu t}{2} + \frac{\widetilde{\lambda}^4\nu t}{4!}W_4^T + \frac{\widetilde{\lambda}^6\nu t}{6!}W_6^T\right], \tag{8.24}$$

$$\lambda(E,N) = \frac{\widehat{x}}{c_1} - \frac{\widehat{x}^3}{3!c_1}V_3^T - \frac{\widehat{x}^5}{5!c_1}V_5^T - \frac{\widehat{x}^7}{7!c_1}\overline{V}_7^T, \tag{8.25}$$

$$\phi(E,N) = \phi_1 - \frac{\widehat{x}^2\beta_1 t_1}{2} - \frac{\widehat{x}^4\beta_1 t_1}{4!}U_4^T - \frac{\widehat{x}^6\beta_1 t_1}{6!}U_6^T, \tag{8.26}$$

where

$$\widetilde{\lambda} = (\lambda - \lambda_0)c, \qquad \widehat{x} = \frac{E - E0}{k_0\nu_1}, \qquad m(\phi_1) = \frac{N - N0}{k_0} + m(\phi_0), \tag{8.27}$$

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 (8.20–8.22)

$$\begin{aligned}
W_3^T &= 1 - t^2 + \eta^2\\
W_4^T &= 5 - t^2 + 9\eta^2\\
W_5^T &= 5 - 18t^2 + t^4 + \eta^2(14 - 58t^2)\\
W_6^T &= 61 - 58t^2 + t^4
\end{aligned} \tag{8.28}$$

$$\begin{aligned}
V_3^T &= 1 + 2t_1^2 + \eta_1^2\\
V_5^T &= -5 - 28t_1^2 - 24t_1^4\\
\overline{V}_7^T &= 61 + 662t_1^2 + 1320t_1^4 + 720t_1^6
\end{aligned} \tag{8.29}$$

$$\begin{aligned}
U_4^T &= -5 - 3t_1^2 - \eta_1^2(1 - 9t_1^2)\\
U_6^T &= 61 + 90t_1^2 + 45t_1^4
\end{aligned} \tag{8.30}$$

where

$$\eta^2 = \beta - 1 = \frac{\nu}{\rho} - 1 = \frac{e^2\cos^2\phi}{1 - e^2} = e'^2\cos^2\phi. \tag{8.31}$$

8.9 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 (5.76, 5.77) to set

$$\begin{aligned} M &= k_0\left[m(\phi)-m(\phi_0)\right] \\ &= k_0 b\bigg[\left(1+n+\frac{5}{4}n^2+\frac{5}{4}n^3\right)(\phi-\phi_0)-\left(3n+3n^2+\frac{21}{8}n^3\right)\sin(\phi-\phi_0)\cos(\phi+\phi_0) \\ &\quad+\left(\frac{15}{8}n^2+\frac{15}{8}n^3\right)\sin 2(\phi-\phi_0)\cos 2(\phi+\phi_0)-\left(\frac{35}{24}n^3\right)\sin 3(\phi-\phi_0)\cos 3(\phi+\phi_0)\bigg] \end{aligned} \tag{8.32}$$

where n is defined in equation (5.5) as $(a-b)/(a+b)$.

Equations (8.24), (8.23), (8.26) and (8.25) may be written as

$$N = \mathrm{I}+\mathrm{II}(\lambda-\lambda_0)^2+\mathrm{III}(\lambda-\lambda_0)^4+\mathrm{IIIA}(\lambda-\lambda_0)^6 \tag{8.33}$$

$$E = E0+\mathrm{IV}(\lambda-\lambda_0)+\mathrm{V}(\lambda-\lambda_0)^3+\mathrm{VI}(\lambda-\lambda_0)^5, \tag{8.34}$$

$$\phi = \phi_1-\mathrm{VII}(E-E0)^2+\mathrm{VIII}(E-E0)^4-\mathrm{IX}(E-E0)^6, \tag{8.35}$$

$$\lambda = \lambda_0+\mathrm{X}(E-E0)-\mathrm{XI}(E-E0)^3+\mathrm{XII}(E-E0)^5-\mathrm{XIIA}(E-E0)^7, \tag{8.36}$$

where ϕ_1 must be calculated from equation

$$m(\phi_1) = \frac{y}{k_0} = \frac{N-N0}{k_0}+m(\phi_0) \tag{8.37}$$

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 (8.28–8.30) as

$$\begin{array}{llll} \mathrm{I} = M+N0 & \mathrm{II} = \dfrac{\widetilde{\nu}sc}{2} & \mathrm{III} = \dfrac{\widetilde{\nu}sc^3W_4^T}{4!} & \mathrm{IIIA} = \dfrac{\widetilde{\nu}sc^5W_6^T}{6!} \\ \mathrm{IV} = \widetilde{\nu}c & \mathrm{V} = \dfrac{\widetilde{\nu}c^3W_3^T}{3!} & \mathrm{VI} = \dfrac{\widetilde{\nu}c^5W_5^T}{5!} & \\ \mathrm{VII} = \dfrac{t_1}{2\widetilde{\rho}_1\widetilde{\nu}_1} & \mathrm{VIII} = \dfrac{-t_1U_4^T}{4!\widetilde{\rho}_1\widetilde{\nu}_1^3} & \mathrm{IX} = \dfrac{t_1U_6^T}{6!\widetilde{\rho}_1\widetilde{\nu}_1^5} & \\ \mathrm{X} = \dfrac{1}{\widetilde{\nu}_1c_1} & \mathrm{XI} = \dfrac{V_3^T}{3!\widetilde{\nu}_1^3c_1} & \mathrm{XII} = \dfrac{-V_5^T}{5!\widetilde{\nu}_1^5c_1} & \mathrm{XIIA} = \dfrac{\overline{V}_7^T}{7!\widetilde{\nu}_1^7c_1} \end{array}$$

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.

8.10 Concluding remarks

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 use the Redfearn series as they stand for they are simple to encode on any computer and they take very little extra computation time. Remember, however, that when these series were first developed it was imperative to simplify the working as much as possible for hand(-machine) calculations. 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 $(x/a)^5e^2$ in the inverse series for λ although we now see that it is negligible.

Nowadays no 'hard graft' is required for the series can be readily implemented on a computer algebra program such as maxima. Such programs are exhibited in Section H: they can easily be extended to higher orders.

The series developed here are simply extensions of the work of Krüger (1912) and Gauss. However, Gauss developed another form of the TM projection which was clarified and actually used by Schreiber in the survey of Prussia. This Gauss-Screiber projection is derived by the transformation of the ellipsoid to the conformal sphere (Sections 5.9,5.10 followed by the TMS from conformal sphere to the plane. This double projection does not have a uniform scale on the central meridian but Kruüger showed that this could be achieved by a further conformal transformation which is based on series linking the conformal and rectifying latitudes. This form of the TME is much more accurate than the Readfearn series. An implementation is described in Karney (2011) and the computer code code is available at Karney (2010). Such implementations can be used for accurate projections to much wider domains.

Both the Redfearn method and this second method based on Gauss-Screiber involve truncated series. They can be extended to any finite order but there are no general series which permit the attainment of arbitrary precision. This is available by using an *exact* version of the TME developed by E. H. Thompson and communicated to Lee (1976). This solution is exact in the sense that it can be computed to arbitrary precision and it provides a yard stick by which other methods can be assessed. Such comparisons permit one to say that the Gauss-Schreiber method is more accurate than the Redfearn series. One remarkable property of the solution is that it is a finite projection which does not tend to infinity as the longitude separation from the central meridian increases. This method is implemented by Dozier (1980), Stuifbergen (2009) and Karney (2011).

Blank page. A contradiction.

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

$$\sec^2\theta \frac{d\theta}{ds} = \frac{d[y'(x)]}{ds} = y''(x)\frac{dx}{ds}$$
$$= y''(x)\cos\theta. \tag{A.3}$$

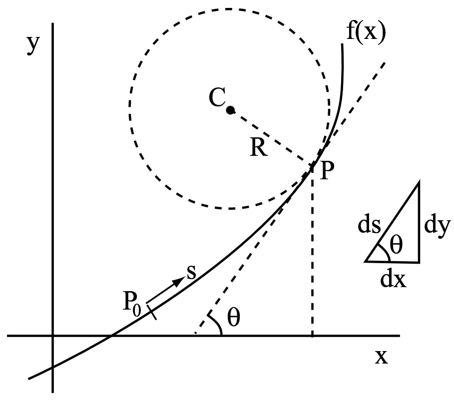


Figure A.1

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 unsigned curvature is given by taking the modulus.

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.)

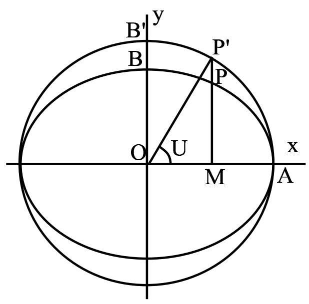


Figure A.2

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.

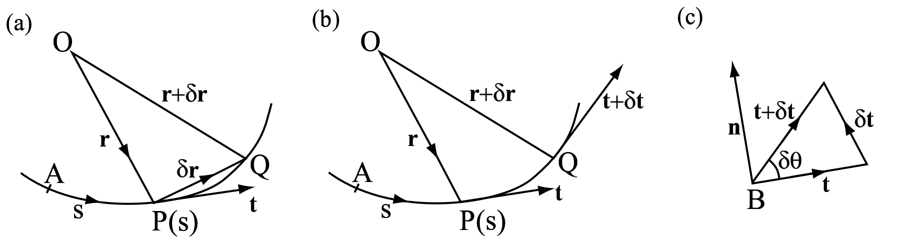


Figure A.3

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

$$\mathbf{b} = \mathbf{t}\times\mathbf{n}, \tag{A.16}$$

$$\begin{aligned} &\mathbf{t}\cdot\mathbf{n} = 0 && \mathbf{n}\cdot\mathbf{b} = 0 && \mathbf{b}\cdot\mathbf{t} = 0, \\ &\mathbf{t}\times\mathbf{n} = \mathbf{b}, && \mathbf{n}\times\mathbf{b} = \mathbf{t}, && \mathbf{b}\times\mathbf{t} = \mathbf{n}. \end{aligned} \tag{A.17}$$

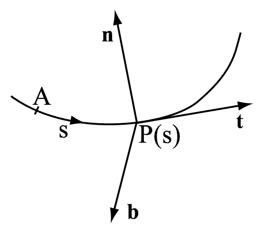


Figure A.4

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 $\mathbf{N}$ 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 $O(x^2)$).

$$w(x) = \sec\phi\left(\frac{1}{2}Ax^2 + O(x^3)\right). \tag{A.26}$$

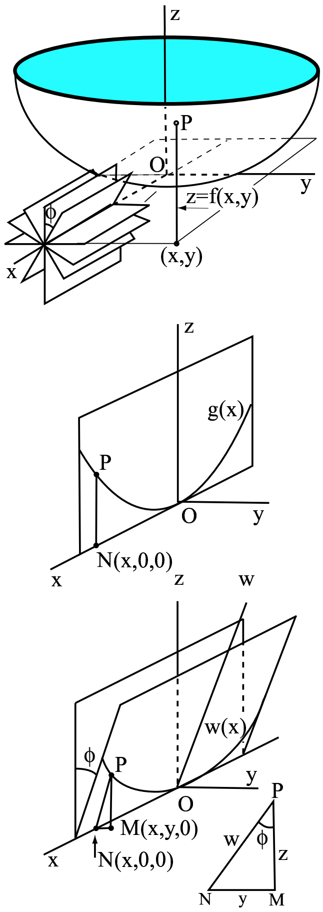


Figure A.5

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:

$$z = f(x,y) = \frac{1}{2}Ax^2 + Bxy + \frac{1}{2}Cy^2 + \cdots. \qquad \text{(A.28)}$$

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

$$g(x) = f(x,0) = \frac{1}{2}Ax^2 + O(x^3), \qquad \text{(A.29)}$$

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.

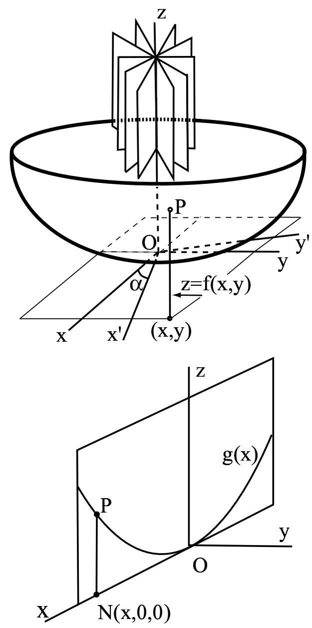


Figure A.6

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

$$x = \cos\alpha\, x' - \sin\alpha\, y', \qquad \text{(A.30)}$$

$$y = \sin\alpha\, x' + \cos\alpha\, y'. \qquad \text{(A.31)}$$

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

$$z = h(x',y') = \frac{1}{2}A(cx'-sy')^2 + B(cx'-sy')(sx'+cy') + \frac{1}{2}C(sx'+cy')^2 + \cdots. \qquad \text{(A.32)}$$

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.

Blank page. A contradiction.

Appendix B

Lagrange expansions

B.1 Introduction

We wish to investigate the inversion of a finite series such as

$$w = z + a_2 z^2 + a_3 z^3 + a_4 z^4 + a_5 z^5 \cdots \tag{B.1}$$

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

$$z = b_1 w + b_2 w^2 + b_3 w^3 + b_4 w^4 + b_5 w^5 + \cdots . \tag{B.2}$$

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

$$\begin{aligned} w = {} & (b_1 w + b_2 w^2 + b_3 w^3 + b_4 w^4) + a_2 w^2 (b_1 + b_2 w + b_3 w^2 + \cdots)^2 \\ & + a_3 w^3 (b_1 + b_2 w + \cdots)^3 + a_4 w^4 (b_1 + \cdots)^4, \\ = {} & (b_1 w + b_2 w^2 + b_3 w^3 + b_4 w^4) + a_2 w^2 (b_1^2 + 2b_1 b_2 w + b_2^2 w^2 + 2b_1 b_3 w^2 + \cdots) \\ & + a_3 w^3 (b_1^3 + 3b_1^2 b_2 w + \cdots) + a_4 w^4 b_1^4 + O(w^5). \end{aligned}$$

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

$$z = w - w^2 [a_2] - w^3 \left[a_3 - 2a_2^2\right] - w^4 \left[a_4 - 5a_2 a_3 + 5a_2^3\right] - \cdots. \tag{B.9}$$

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

$$w = z + a_2z^2 + a_3z^3 + a_4z^4 + a_5z^5 + a_6z^6 + a_7z^7 + a_8z^8. \tag{B.19}$$

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 $f(w) = a_2w^2 + a_3w^3 \cdots$ 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.

$$\begin{aligned}
f(w) &= a_2w^2 + a_3w^3 + a_4w^4 + a_5w^5 + a_6w^6 + a_7w^7 + a_8w^8, \\
[f(w)]^2 &= w^4(a_2^2) \\
&\quad + w^5(2a_2a_3) \\
&\quad + w^6(2a_2a_4 + a_3^2) \\
&\quad + w^7(2a_2a_5 + 2a_3a_4) \\
&\quad + w^8(2a_2a_6 + 2a_3a_5 + a_4^2) \\
&\quad + w^9(2a_2a_7 + 2a_3a_6 + 2a_4a_5) + O(w^{10}), \\
[f(w)]^3 &= w^6(a_2^3) \\
&\quad + w^7(3a_2^2a_3) \\
&\quad + w^8(3a_2^2a_4 + 3a_2a_3^2) \\
&\quad + w^9(3a_2^2a_5 + 6a_2a_3a_4 + a_3^3) \\
&\quad + w^{10}(3a_2^2a_6 + 6a_2a_3a_5 + 3a_2a_4^2 + 3a_3^2a_4) + O(w^{11}), \\
[f(w)]^4 &= w^8(a_2^4) \\
&\quad + w^9(4a_2^3a_3) \\
&\quad + w^{10}(4a_2^3a_4 + 6a_2^2a_3^2) \\
&\quad + w^{11}(4a_2^3a_5 + 12a_2^2a_3a_4 + 4a_2a_3^3) + O(w^{12}), \\
[f(w)]^5 &= w^{10}(a_2^5) \\
&\quad + w^{11}(5a_2^4a_3) \\
&\quad + w^{12}(5a_2^4a_4 + 10a_2^3a_3^2) + O(w^{13}), \\
[f(w)]^6 &= w^{12}(a_2^6) \\
&\quad + w^{13}(6a_2^5a_3) + O(w^{14}), \\
[f(w)]^7 &= w^{14}(a_2^7) + O(w^{15}), \\
[f(w)]^8 &= O(w^{16}).
\end{aligned}$$

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

$$\begin{aligned}
z = w &- w^2 [a_2] \\
&- w^3 \left[a_3 - 2a_2^2\right] \\
&- w^4 \left[a_4 - 5a_2 a_3 + 5a_2^3\right] \\
&- w^5 \left[a_5 - 3(2a_2 a_4 + a_3^2) + 21 a_2^2 a_3 - 14 a_2^4\right] \\
&- w^6 \left[a_6 - 7(a_2 a_5 + a_3 a_4) + 28(a_2^2 a_4 + a_2 a_3^2) - 84 a_2^3 a_3 + 42 a_2^5\right] \\
&- w^7 \left[a_7 - 4(2a_2 a_6 + 2a_3 a_5 + a_4^2) + 12(3a_2^2 a_5 + 6a_2 a_3 a_4 + a_3^3)\right. \\
&\qquad \left. - 30(4a_2^3 a_4 + 6a_2^2 a_3^2) + 330 a_2^4 a_3 - 132 a_2^6\right] \\
&- w^8 \left[a_8 - 9(a_2 a_7 + a_3 a_6 + a_4 a_5) + 15(3a_2^2 a_6 + 6a_2 a_3 a_5 + 3a_2 a_4^2 + 3a_3^2 a_4)\right. \\
&\qquad - 165(a_2^3 a_5 + 3a_2^2 a_3 a_4 + a_2 a_3^3) + 99(5a_2^4 a_4 + 10 a_2^3 a_3^2) \\
&\qquad \left. - 1287 a_2^5 a_3 + 429 a_2^7\right]
\end{aligned} \tag{B.22}$$

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

$$\begin{aligned}
p_2 &= [b_2] \\
p_3 &= \left[b_3 - 3b_2^2\right] \\
p_4 &= \left[b_4 - 10 b_2 b_3 + 15 b_2^3\right] \\
p_5 &= \left[b_5 - (15 b_2 b_4 + 10 b_3^2) + 105 b_2^2 b_3 - 105 b_2^4\right] \\
p_6 &= \left[b_6 - (21 b_2 b_5 + 35 b_3 b_4) + (210 b_2^2 b_4 + 280 b_2 b_3^2) - 1260 b_2^3 b_3 + 945 b_2^5\right] \\
p_7 &= \left[b_7 - (28 b_2 b_6 + 56 b_3 b_5 + 35 b_4^2) + (378 b_2^2 b_5 + 1260 b_2 b_3 b_4 + 280 b_3^3)\right. \\
&\qquad \left. - (3150 b_2^3 b_4 + 6300 b_2^2 b_3^2) + 17325 b_2^4 b_3 - 10395 b_2^6\right] \\
p_8 &= [b_8 - (36 b_2 b_7 + 84 b_3 b_6 + 126 b_4 b_5) \\
&\qquad + (630 b_2^2 b_6 + 2520 b_2 b_3 b_5 + 1575 b_2 b_4^2 + 2100 b_3^2 b_4) \\
&\qquad - (6930 b_2^3 b_5 + 34650 b_2^2 b_3 b_4 + 15400 b_2 b_3^3) \\
&\qquad + (51975 b_2^4 b_4 + 138600 b_2^3 b_3^2) - 270270 b_2^5 b_3 + 135135 b_2^7\left.\right]
\end{aligned} \tag{B.25}$$

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

$$\begin{array}{llll} b_2 = is & b_3 = \ c^2W_3 & b_4 = isc^2W_4 & b_5 = \ c^4W_5 \\ b_6 = isc^4W_6 & b_7 = \ c^6\overline{W}_7 & b_8 = isc^6\overline{W}_8 & \end{array} \tag{B.26}$$

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

$$\begin{aligned} p_2 &= ict\ \ [1] \\ p_3 &= \ \ c^2\left[W_3+3t^2\right] \\ p_4 &= ic^3t\left[W_4-10W_3-15t^2\right] \\ p_5 &= \ \ c^4\left[W_5+(15t^2W_4-10W_3^2)-105t^2W_3-105t^4\right] \\ p_6 &= ic^5t\left[W_6-(21W_5+35W_3W_4)-(210t^2W_4-280W_3^2)\ +1260t^2W_3+945t^4\right] \\ p_7 &= \ \ c^6\Big[\overline{W}_7+(28t^2W_6-56W_3W_5+35t^2W_4^2)\ -(378t^2W_5+1260t^2W_3W_4-280W_3^3) \\ &\qquad -(3150t^4W_4-6300t^2W_3^2)+17325t^4W_3+10395t^6\Big] \\ p_8 &= ic^7t\Big[\overline{W}_8-(36\overline{W}_7+84W_3W_6+126W_4W_5) \\ &\qquad -(630t^2W_6-2520W_3W_5+1575t^2W_4^2-2100W_3^2W_4) \\ &\qquad +(6930t^2W_5+34650t^2W_3W_4-15400W_3^3) \\ &\qquad +(51975t^4W_4-138600t^2W_3^2)\ -270270t^4W_3-135135t^6\Big] \end{aligned} \tag{B.28}$$

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

$$
\begin{aligned}
p_2 &= ict\ \left[1\right]\\
p_3 &= \ \ c^2\left[\beta - t^2 + 3t^2\right]\\
p_4 &= ic^3t\left[4\beta^2 + \beta - t^2 - 10(\beta - t^2) - 15t^2\right]\\
p_5 &= \ \ c^4\big[4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4\\
&\qquad + 15t^2(4\beta^2 + \beta - t^2)\\
&\qquad - 10(\beta - t^2)^2 - 105t^2(\beta - t^2) - 105t^4\big]\\
p_6 &= ic^5t\big[8\beta^4(11-24t^2) - 28\beta^3(1-6t^2) + \beta^2(1-32t^2) - 2\beta t^2 + t^4\\
&\qquad - 21\left\{4\beta^3(1-6t^2) + \beta^2(1+8t^2) - 2\beta t^2 + t^4\right\}\\
&\qquad - 35(\beta - t^2)(4\beta^2 + \beta - t^2) - 210t^2(4\beta^2 + \beta - t^2)\\
&\qquad + 280(\beta - t^2)^2 + 1260t^2(\beta - t^2) + 945t^4\big]\\
\bar{p}_7 &= \ \ c^6\big[61 - 479t^2 + 179t^4 - t^6 + 28t^2(61 - 58t^2 + t^4) - 56(1-t^2)(5 - 18t^2 + t^4)\\
&\qquad + 35t^2(5-t^2)^2 - 378t^2(5 - 18t^2 + t^4) - 1260t^2(1-t^2)(5-t^2)\\
&\qquad + 280(1-t^2)^3 - 3150t^4(5-t^2) + 6300t^2(1-t^2)^2 + 17325t^4(1-t^2)\\
&\qquad + 10395t^6\big]\\
\bar{p}_8 &= ic^7t\big[1385 - 3111t^2 + 543t^4 - t^6 - 36(61 - 479t^2 + 179t^4 - t^6)\\
&\qquad - 84(1-t^2)(61 - 58t^2 + t^4) - 126(5-t^2)(5 - 18t^2 + t^4)\\
&\qquad - 630t^2(61 - 58t^2 + t^4) + 2520(1-t^2)(5 - 18t^2 + t^4)\\
&\qquad - 1575t^2(5-t^2)^2 + 2100(1-t^2)^2(5-t^2) + 6930t^2(5 - 18t^2 + t^4)\\
&\qquad + 34650t^2(1-t^2)(5-t^2) - 15400(1-t^2)^3 + 51975t^4(5-t^2)\\
&\qquad - 138600t^2(1-t^2)^2 - 270270t^4(1-t^2) - 135135t^6\big]
\end{aligned}
$$

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

$$
\begin{aligned}
p_2 &= ict\ \left[1\right]\\
p_3 &= \ \ c^2\left[\beta + 2t^2\right]\\
p_4 &= ic^3t\left[4\beta^2 - 9\beta - 6t^2\right]\\
p_5 &= \ \ c^4\left[4\beta^3(1-6t^2) - \beta^2(9-68t^2) - 72\beta t^2 - 24t^4\right]\\
p_6 &= ic^5t\left[8\beta^4(11-24t^2) - 84\beta^3(3-8t^2) + 225\beta^2(1-4t^2) + 600\beta t^2 + 120t^4\right]\\
\bar{p}_7 &= \ \ c^6\left[61 + 662t^2 + 1320t^4 + 720t^6\right]\\
\bar{p}_8 &= ic^7t\left[-1385 - 7266t^2 - 10920t^4 - 5040t^6\right]
\end{aligned}
\tag{B.30}
$$

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

$$z = a = \frac{1}{2\pi i} \oint_C \frac{z p'(z)}{p(z) - t} dz. \tag{B.35}$$

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.

Web resources:

Mathworld series reversion

Mathworld Lagrange inversion

Wikipedia Lagrange inversiontheorem

Blank page. A contradiction.

Appendix C

Plane Trigonometry

C.1 Trigonometric functions

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)	$\sin x \cos y = \frac{1}{2}[\sin(x+y) + \sin(x-y)]$	(C.9)
$x \leftrightarrow y$	$\cos x \sin y = \frac{1}{2}[\sin(x+y) - \sin(x-y)]$	(C.10)
(C.3) + (C.4)	$\cos x \cos y = \frac{1}{2}[\cos(x+y) + \cos(x-y)]$	(C.11)
(C.4) – (C.3)	$\sin x \sin y = \frac{1}{2}[\cos(x-y) - \cos(x+y)]$	(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)
$$\sin^2 x = \frac{1}{2}[1 - \cos 2x] \tag{C.24}$$
from (C.22)
$$\cos^2 x = \frac{1}{2}[1 + \cos 2x] \tag{C.25}$$

use (C.21)
$$\sin^3 x = \frac{1}{2}\sin x[1 - \cos 2x]$$
use (C.9)
$$= \frac{1}{2}\sin x - \frac{1}{4}[\sin 3x - \sin x]$$
$$= \frac{3}{4}\sin x - \frac{1}{4}\sin 3x \tag{C.26}$$

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)
$$= \frac{3}{8}\sin 2x - \frac{1}{8}[\sin 4x + \sin 2x]$$
$$= \frac{1}{8}[2\sin 2x - \sin 4x] \tag{C.27}$$

from (C.24)
$$\sin^4 x = \frac{1}{4}[1 - \cos 2x]^2$$
use (C.25)
$$= \frac{1}{4}\left[1 - 2\cos 2x + \frac{1}{2} + \frac{1}{2}\cos 4x\right]$$
$$= \frac{1}{8}[3 - 4\cos 2x + \cos 4x] \tag{C.28}$$
use (C.25)
$$\cos^3 x = \frac{1}{2}\cos x[1 + \cos 2x]$$
use (C.11)
$$= \frac{1}{2}\cos x + \frac{1}{4}[\cos 3x + \cos x]$$
$$= \frac{3}{4}\cos x + \frac{1}{4}\cos 3x \tag{C.29}$$

NOTATION
$$S_k \equiv \sin kx \tag{C.30}$$
$$C_k \equiv \cos kx \tag{C.31}$$

Hence
$$\sin^2 x = \frac{1}{2}[1 - C_2] \tag{C.32}$$
$$\sin^3 x = \frac{1}{4}[3S - S_3] \tag{C.33}$$
$$\sin^4 x = \frac{1}{8}[3 - 4C_2 + C_4] \tag{C.34}$$
$$\sin^5 x = \frac{1}{16}[10S - 5S_3 + S_5] \tag{C.35}$$
$$\sin^6 x = \frac{1}{32}[10 - 15C_2 + 6C_4 - C_6] \tag{C.36}$$
$$\sin^7 x = \frac{1}{64}[35S - 21S_3 + 7S_5 - S_7] \tag{C.37}$$
$$\sin^8 x = \frac{1}{128}[35 - 56C_2 + 28C_4 - 8C_6 + C_8] \tag{C.38}$$

$$\sin x \cos x = \frac{1}{2}[S_2] \tag{C.39}$$
$$\sin^3 x \cos x = \frac{1}{8}[2S_2 - S_4] \tag{C.40}$$
$$\sin^5 x \cos x = \frac{1}{32}[5S_2 - 4S_4 + S_6] \tag{C.41}$$
$$\sin^7 x \cos x = \frac{1}{128}[14S_2 - 14S_4 + 6S_6 - S_8] \tag{C.42}$$

C.2 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).

$$\cosh(x \pm y) = \cosh x \cosh y \pm \sinh x \sinh y \tag{C.47}$$
$$\sinh(x \pm y) = \sinh x \cosh y \pm \cosh x \sinh y \tag{C.48}$$
$$\tanh(x \pm y) = \frac{\tanh x \pm \tanh y}{1 \pm \tanh x \tanh y} \tag{C.49}$$

$$1 = \cosh^2 x - \sinh^2 x \tag{C.50}$$
$$\operatorname{sech}^2 x = 1 - \tanh^2 x \tag{C.51}$$
$$\operatorname{cosech}^2 x = \coth^2 x - 1 \tag{C.52}$$

$$\begin{aligned} \cosh 2x &= \cosh^2 x + \sinh^2 x \\ &= 1 + 2\sinh^2 x \\ &= 2\cosh^2 x - 1 \end{aligned} \tag{C.53}$$
$$\sinh 2x = 2\sinh x \cosh x \tag{C.54}$$

Since the hyperbolic functions are defined in terms of the exponential function it is not surprising their inverses can be related to the logarithm function. We consider the three main cases in parallel.

$$\begin{aligned} y &= \sinh^{-1} x & y &= \cosh^{-1} x & y &= \tanh^{-1} x \\ x &= \frac{e^y - e^{-y}}{2} & x &= \frac{e^y + e^{-y}}{2} & x &= \frac{e^y - e^{-y}}{e^y + e^{-y}} \\ 0 &= e^{2y} - 2xe^y - 1 & 0 &= e^{2y} - 2xe^y + 1 & x(e^{2y} + 1) &= e^{2y} - 1 \\ e^y &= x \pm \sqrt{x^2 + 1} & e^y &= x \pm \sqrt{x^2 - 1} & e^{2y} &= \frac{1 + x}{1 - x}. \end{aligned}$$

Taking into account of the ranges of these functions we have

$$\sinh^{-1} x = \ln\left(x + \sqrt{x^2 + 1}\right) \qquad -\infty < x < \infty \tag{C.55}$$
$$\cosh^{-1} x = \ln\left(x + \sqrt{x^2 - 1}\right) \qquad x \geq 1 \tag{C.56}$$
$$\tanh^{-1} x = \frac{1}{2} \ln\left(\frac{1 + x}{1 - x}\right) \qquad -1 < x < 1 \tag{C.57}$$

continued

C.3 Gudermannian functions

Gudermannians link trigonometric and hyperbolic functions. They may be used to describe the Mercator projections. For real x we can take the basic definition as

$$\mathrm{gd}\,x = \tan^{-1}\sinh x. \tag{C.58}$$

This definition implies that $\tan(\mathrm{gd}x) = \sinh x$. The following equivalent definitions follow easily from the properties of the trigonometric and hyperbolic functions. For example, for the second of the following use $\sec^2(\mathrm{gd}x) = 1+\tan^2(\mathrm{gd}x) = 1+\sinh^2 x = \cosh^2 x$.

$$\begin{aligned} \mathrm{gd}\,x &= \tan^{-1}\sinh x = \sec^{-1}\cosh x = \sin^{-1}\tanh x \\ &= \cos^{-1}\mathrm{sech}\,x = \cot^{-1}\mathrm{csch}\,x = \csc^{-1}\coth x. \end{aligned} \tag{C.59}$$

Each of these provides an inverse: simply replace x by $\mathrm{gd}^{-1}x$ and set $\mathrm{gd}(\mathrm{gd}^{-1}x) = x$.

$$\begin{aligned} \mathrm{gd}^{-1}x &= \sinh^{-1}\tan x = \cosh^{-1}\sec x = \tanh^{-1}\sin x \\ &= \mathrm{sech}^{-1}\cos x = \mathrm{csch}^{-1}\cot x = \coth^{-1}\csc x. \end{aligned} \tag{C.60}$$

The inverse gudermannians can be expressed in terms of logarithms by using C.57 and elementary operations:

$$\mathrm{gd}^{-1}x = \frac{1}{2}\ln\left[\frac{1+\sin x}{1-\sin x}\right] = \ln(\sec x + \tan x) = \ln\left[\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right]. \tag{C.61}$$

In the last of these replace $x \to \mathrm{gd}\,x$ and invert to give

$$\mathrm{gd}x = 2\tan^{-1}\mathrm{e}^{x} - \frac{\pi}{2} = \frac{\pi}{2} - 2\tan^{-1}\mathrm{e}^{-x}, \tag{C.62}$$

where the second follows by replacing $\tan(x/2+\pi/4)$ by $\cot(\pi/4-x/2)$.

Equations C.45, C.46 may be used for gudemannians with imaginary arguments.

$$\begin{aligned} \tan[\mathrm{gd}(ix)] &= \sinh(ix) = i\sin x, \\ \tanh[i\,\mathrm{gd}(ix)] &= -\sin x, \\ i\,\mathrm{gd}(ix) &= -\tanh^{-1}[\sin x], \\ \mathrm{gd}(ix) &= i\,\mathrm{gd}^{-1}x. \end{aligned} \tag{C.63}$$

Similarly

$$\begin{aligned} \tanh[\mathrm{gd}^{-1}(ix)] &= \sin(ix) = i\sinh x, \\ \tan[i\,\mathrm{gd}^{-1}(ix)] &= -\sinh x, \\ i\,\mathrm{gd}^{-1}(ix) &= -\tan^{-1}[\sinh x], \\ \mathrm{gd}^{-1}(ix) &= i\,\mathrm{gd}\,x. \end{aligned} \tag{C.64}$$

Integral representations:

$$\operatorname{gd} x = \int_0^x \operatorname{sech}\theta \, d\theta \,, \qquad -\infty < x < \infty. \tag{C.65}$$

$$\operatorname{gd}^{-1} x = \int_0^x \sec\theta \, d\theta \,, \qquad -\pi/2 < x < \pi/2 \,. \tag{C.66}$$

For the second integral set

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

The integral becomes

$$\operatorname{gd}^{-1} x = \int_0^x \frac{\sec^2(\theta/2+\pi/4)}{\tan(\theta/2+\pi/4)}\, dx = \ln\left[\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)\right]. \tag{C.67}$$

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 ∞.

The integral for $\operatorname{gd} x$ may be verified by replacing x by ix in equation C.66.

For further properties of the gudermannians see see Appendix G, Equation G.24 onwards.

Web resources

NIST: Sections 4.14–4.45

Mathworld gudermannian

Mathworld: inverse gudermannian

Wikipedia gudermannian

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 good (but old) text books. See Todhunter (1859): it is available on the web.

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.

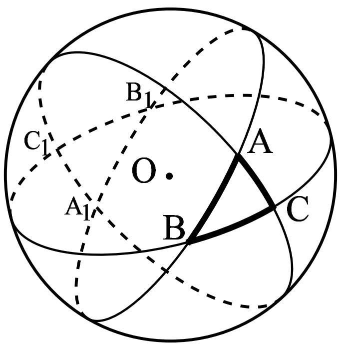


Figure D.1

We now restrict attention to triangles in which the angles are less than or equal to π, with the proviso that not all equal to π: in the latter case the triangle degenerates into 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 (300BC), Book 11, Proposition 21. 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.

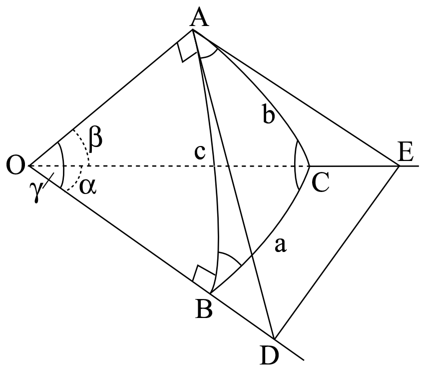


Figure D.2

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

$$\boxed{\begin{aligned}\cos a &= \cos b\,\cos c + \sin b\,\sin c\,\cos A,\\ \cos b &= \cos c\,\cos a + \sin c\,\sin a\,\cos B,\\ \cos c &= \cos a\,\cos b + \sin a\,\sin b\,\cos C.\end{aligned}} \qquad \text{(D.1)}$$

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.

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

$$\begin{aligned}\mathbf{B}&=(\sin c,\ 0,\ \cos c),\\ \mathbf{C}&=(\sin b\cos A, \sin b\sin A,\ \cos b)\end{aligned} \qquad \text{(D.2)}$$

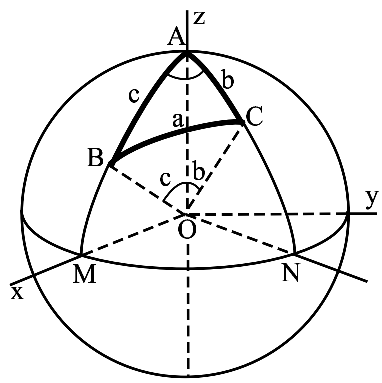


Figure D.3

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

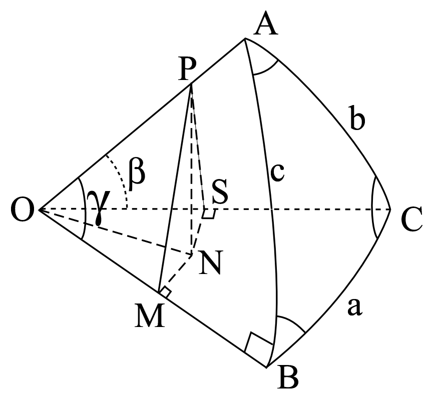


Figure D.4

$$\begin{aligned} PM^2 &= MN^2 + PN^2, \\ OP^2 &= ON^2 + PN^2, \\ ON^2 &= OM^2 + MN^2. \end{aligned}$$

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.

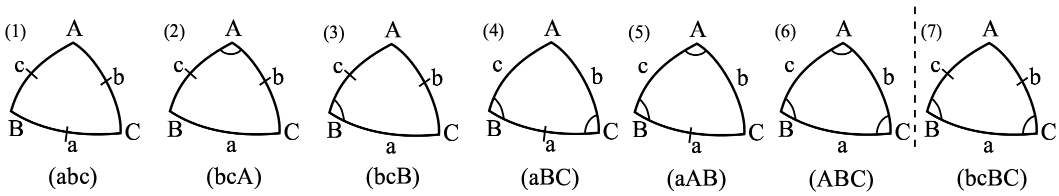


Figure D.5

- **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.

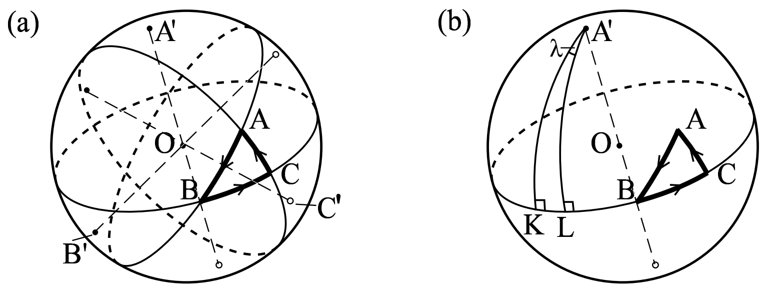


Figure D.6

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.

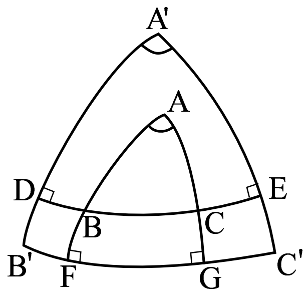


Figure D.7

- 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:

$$\begin{array}{lll} A' = \pi - a & B' = \pi - b & C' = \pi - c, \\ a' = \pi - A & b' = \pi - B & c' = \pi - C. \end{array} \tag{D.8}$$

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 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:

$$
\begin{aligned}
\sin\frac{A}{2} &= \left[\frac{\sin(s-b)\sin(s-c)}{\sin b\sin c}\right]^{1/2} & \sin\frac{a}{2} &= \left[\frac{-\cos S\cos(S-A)}{\sin B\sin C}\right]^{1/2} \\
\cos\frac{A}{2} &= \left[\frac{\sin s\sin(s-a)}{\sin b\sin c}\right]^{1/2} & \cos\frac{a}{2} &= \left[\frac{\cos(S-B)\cos(S-C)}{\sin B\sin C}\right]^{1/2} \\
\tan\frac{A}{2} &= \left[\frac{\sin(s-b)\sin(s-c)}{\sin s\sin(s-a)}\right]^{1/2} & \tan\frac{a}{2} &= \left[\frac{-\cos S\cos(S-A)}{\cos(S-B)\cos(S-C)}\right]^{1/2}
\end{aligned}
\tag{D.13}
$$

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)

(D.1c)	$\cos c = \cos a \cos b,$	(D.11b)	$\tan b = \cos A \tan c,$
(D.7b)	$\sin a = \sin A \sin c,$	(D.11e)	$\tan a = \cos B \tan c,$
(D.7c)	$\sin b = \sin B \sin c,$	(D.10a)	$\cos A = \sin B \cos a,$
(D.11a)	$\tan a = \tan A \sin b,$	(D.10b)	$\cos B = \sin A \cos b,$
(D.11f)	$\tan b = \tan B \sin a,$	(D.10c)	$\cos c = \cot A \cot B.$ (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:

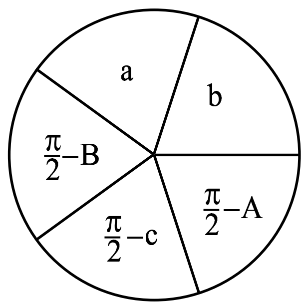


Figure D.8

$$
\boxed{\begin{array}{lcl} \text{sine of middle part} & = & \text{product of tangents of adjacent parts,} \\ \text{sine of middle part} & = & \text{product of cosines of opposite parts.} \end{array}} \qquad \text{(D.17)}
$$

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 $C' = \pi/2$:

$$a' = \pi - A, \quad b' = \pi - B, \quad A' = \pi - a, \quad B' = \pi - b.$$

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:

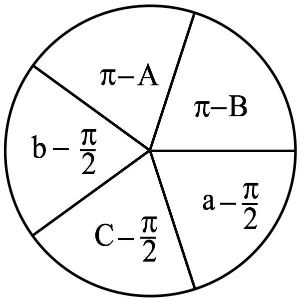


Figure D.9

$$\begin{aligned} \cos C &= -\cos A \cos B, & \tan B &= -\cos a \tan C, \\ \sin A &= \sin a \sin C, & \tan A &= -\cos b \tan C, \\ \sin B &= \sin b \sin C, & \cos a &= \sin b \cos A, \\ \tan A &= \tan a \sin B, & \cos b &= \sin a \cos B, \\ \tan B &= \tan b \sin A, & \cos C &= -\cot a \cot b. \end{aligned} \tag{D.18}$$

Example
As an example of a quadrantal triangle we consider a problem arising in the discussion of geodesics on a sphere. With the following identifications

$$\begin{aligned} a &= s, & b &= \frac{\pi}{2} - \phi, & c &= \frac{\pi}{2}, \\ A &= \lambda, & B &= \alpha_0, & C &= \pi - \alpha. \end{aligned} \tag{D.19}$$

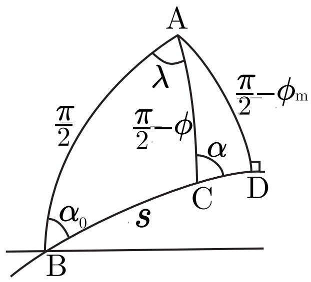


Figure D.10

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.

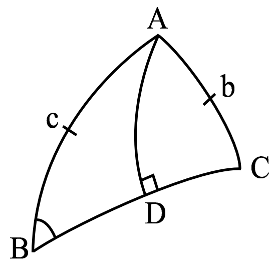


Figure D.11

Appendix E

Power series expansions

E.1 General form of the Taylor and Maclaurin series

Taylor's theorem may be written in the form:

$$f(z) = f(b) + \frac{(z-b)}{1!}f'(b) + \frac{(z-b)^2}{2!}f''(b) + \frac{(z-b)^3}{3!}f'''(b) + \cdots . \qquad \text{(E.1)}$$

or, alternatively,

$$f(b+z) = f(b) + \frac{z}{1!}f'(b) + \frac{z^2}{2!}f''(b) + \frac{z^3}{3!}f'''(b) + \cdots . \qquad \text{(E.2)}$$

When $b = 0$ we obtain Maclaurin's series.

$$f(z) = f(0) + \frac{z}{1!}f'(0) + \frac{z^2}{2!}f''(0) + \frac{z^3}{3!}f'''(0) + \cdots . \qquad \text{(E.3)}$$

E.2 Miscellaneous Taylor series

$$\sin(b+z) = \sin b + z\cos b - \frac{z^2}{2!}\sin b - \frac{z^3}{3!}\cos b + \frac{z^4}{4!}\sin b + \cdots \qquad \text{(E.4)}$$

$$\cos(b+z) = \cos b - z\sin b - \frac{z^2}{2!}\cos b + \frac{z^3}{3!}\sin b + \frac{z^4}{4!}\cos b + \cdots \qquad \text{(E.5)}$$

$$\tan(b+z) = \tan b + z\sec^2 b + z^2\tan b\sec^2 b + \frac{z^3}{3}(1+3\tan^2 b)\sec^2 b + \cdots \qquad \text{(E.6)}$$

$$\tan\left(\frac{\pi}{4}+z\right) = 1 + 2z + 2z^2 + \frac{8}{3}z^3 + \cdots \qquad \text{(E.7)}$$

$$\arcsin(b+z) = \arcsin b + z\frac{1}{(1-b^2)^{1/2}} + \frac{z^2}{2}\frac{b}{(1-b^2)^{3/2}} + \cdots \qquad \text{(E.8)}$$

$$\arctan(b+z) = \arctan b + z\frac{1}{1+b^2} - \frac{z^2}{2!}\frac{2b}{(1+b^2)^2} + \frac{z^3}{3!}\left[\frac{-2}{(1+b^2)^2} + \frac{8b^2}{(1+b^2)^3}\right] - \frac{z^4}{4!}\left[\frac{-24b}{(1+b^2)^3} + \frac{48b^3}{(1+b^2)^4}\right] + \cdots \qquad \text{(E.9)}$$

/over

E.3 Miscellaneous Maclaurin series

- Logarithms

$$\ln(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \frac{1}{4}z^4 + \cdots \qquad -1 < z \le 1 \qquad \text{(E.10)}$$

$$\ln(1-z) = -z - \frac{1}{2}z^2 - \frac{1}{3}z^3 - \frac{1}{4}z^4 + \cdots \qquad -1 \le z < 1 \qquad \text{(E.11)}$$

$$\ln\left(\frac{1+z}{1-z}\right) = 2z + \frac{2}{3}z^3 + \frac{2}{5}z^5 + \frac{2}{7}z^7 + \cdots \qquad |z| < 1 \qquad \text{(E.12)}$$

- Trigonometric

$$\sin z = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \frac{1}{7!}z^7 + \cdots \qquad |z| < \infty \qquad \text{(E.13)}$$

$$\cos z = 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \frac{1}{6!}z^6 + \cdots \qquad |z| < \infty \qquad \text{(E.14)}$$

$$\tan z = z + \frac{1}{3}z^3 + \frac{2}{15}z^5 + \frac{17}{315}z^7 + \cdots \qquad |z| < \frac{\pi}{2} \qquad \text{(E.15)}$$

$$\sec z = 1 + \frac{1}{2}z^2 + \frac{5}{24}z^4 + \frac{61}{720}z^6 + \cdots \qquad |z| < \frac{\pi}{2} \qquad \text{(E.16)}$$

$$\csc z = \frac{1}{z} + \frac{1}{6}z + \frac{7}{360}z^3 + \frac{31}{15120}z^5 + \cdots \qquad 0 < |z| < \pi \qquad \text{(E.17)}$$

$$\cot z = \frac{1}{z} - \frac{1}{3}z - \frac{1}{45}z^3 - \frac{2}{945}z^5 - \cdots \qquad 0 < |z| < \pi \qquad \text{(E.18)}$$

- Inverse trig

$$\arcsin z = z + \frac{1}{6}z^3 + \frac{3}{40}z^5 + \frac{5}{112}z^7 + \cdots \qquad |z| < 1 \qquad \text{(E.19)}$$

$$\arctan z = z - \frac{1}{3}z^3 + \frac{1}{5}z^5 - \frac{1}{7}z^7 + \frac{1}{9}z^9 - \cdots \qquad |z| < 1 \qquad \text{(E.20)}$$

- Hyperbolic

$$\sinh z = z + \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \frac{1}{7!}z^7 + \cdots \qquad |z| < \infty \qquad \text{(E.21)}$$

$$\cosh z = 1 + \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \frac{1}{6!}z^6 + \cdots \qquad |z| < \infty \qquad \text{(E.22)}$$

$$\tanh z = z - \frac{1}{3}z^3 + \frac{2}{15}z^5 - \frac{17}{315}z^7 + \cdots \qquad |z| < \frac{\pi}{2} \qquad \text{(E.23)}$$

$$\operatorname{sech} z = 1 - \frac{1}{2}z^2 + \frac{5}{24}z^4 - \frac{61}{720}z^6 \cdots \qquad |z| < \frac{\pi}{2} \qquad \text{(E.24)}$$

/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

$$\begin{aligned}(1+a_2z^2+a_4z^4+a_6z^6)^{-1} &= 1-(a_2z^2+a_4z^4+a_6z^6)\\ &\qquad +(a_2^2z^4+2a_2a_4z^6+\cdots)^2-(a_2^3z^6+\cdots)^3+O(z^8)\\ &= 1-(a_2)z^2-(a_4-a_2^2)z^4-(a_6-2a_2a_4+a_2^3)z^6+O(z^8).\end{aligned} \tag{E.31}$$

Furthermore

$$\left(1+\frac{a_2z^2}{2}+\frac{a_4z^4}{24}+\frac{a_6z^6}{720}\right)^{-1} = 1-\frac{z^2}{2}(a_2)-\frac{z^4}{24}\left(a_4-6a_2^2\right)-\frac{z^6}{720}\left(a_6-30a_2a_4+90a_2^3\right). \tag{E.32}$$

Blank page. A contradiction.

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.

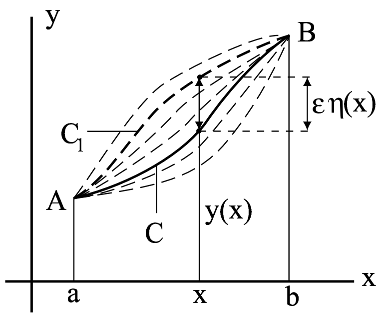


Figure F.1

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 $\bar{y}$ where

$$\bar{y}(x) = y(x) + \varepsilon\eta(x), \tag{F.2}$$

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 $\Phi(\varepsilon)$ such that

$$\Phi(\varepsilon) = J[\bar{y}] = J[y+\varepsilon\eta] = \int_a^b F(x, y+\varepsilon\eta, y'+\varepsilon\eta')dx. \tag{F.3}$$

In this notation the value of the integral on the extremal path is $\Phi(0)$ and the condition that it is an extremum is

$$\left.\frac{d\Phi}{d\varepsilon}\right|_{\varepsilon=0} = 0. \tag{F.4}$$

The Taylor series for the integrand is

$$F(x, y+\varepsilon\eta, y'+\varepsilon\eta') = F + F_y\varepsilon\eta + F_{y'}\varepsilon\eta' + O(\varepsilon^2) \tag{F.5}$$

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

$$\left.\frac{d\Phi}{d\varepsilon}\right|_{\varepsilon=0} = \int_a^b \left[F_y\eta + F_{y'}\eta'\right] dx = 0. \tag{F.6}$$

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:

$$\text{EULER–LAGRANGE} \qquad \boxed{\frac{d}{dx}\left[\frac{\partial F}{\partial y'}\right] - \frac{\partial F}{\partial y} = 0.} \tag{F.10}$$

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

$$\frac{d}{dx}\left[y'F_{y'} - F(x,y,y')\right] = y''F_{y'} + y'\frac{d}{dx}\left[F_{y'}\right] - F_x - F_y y' - F_{y'}y'' = -F_x, \tag{F.12}$$

giving an alternative equation

$$\frac{d}{dx}\left[y'F_{y'} - F\right] + F_x = 0 \tag{F.13}$$

Functions of the form *F(y, y')*

Equation (F.13) shows that if F is independent of x then the equations integrate immediately since $F_x = 0$:

$$y'F_{y'} - F = \text{constant.} \tag{F.14}$$

Functions of the form *F(x, y')*

If F is independent of y then equation (F.10) can be integrated directly since $F_y = 0$:

$$F_{y'} = \text{constant.} \tag{F.15}$$

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

$$\bar{u}(x) = u(x) + \varepsilon\eta^{(u)}(x),$$
$$\bar{v}(x) = v(x) + \varepsilon\eta^{(v)}(x).$$

Equation (F.8) now becomes of the form

$$\int_a^b \left[\eta^{(u)}(x)H^{(u)}(x) + \eta^{(v)}(x)H^{(v)}(x)\right]\, dx = 0. \tag{F.16}$$

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, $x = \mathrm{Re}\,(z)$, and y is the imaginary part, $y = \mathrm{Im}\,(z)$. 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

$$z = x + iy = r(\cos\theta + i\sin\theta). \qquad \text{(G.1)}$$

In this context we say that r is the 'modulus' of z and θ is the 'argument' of z and write

$$r = |z| = \left[x^2 + y^2\right]^{1/2}, \qquad \theta = \arg(z) = \arctan\left(\frac{y}{x}\right). \qquad \text{(G.2)}$$

Note that we can also write $r = |z| = \sqrt{zz^*}$.

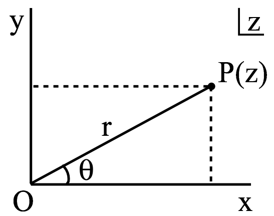


Figure G.1

Complex functions: examples

- The simplest complex function we can consider is a finite polynomial such as:

$$w(z) = 3 + z + z^2. \qquad \text{(G.3)}$$

If we substitute $z = x + iy$ in this expression, using $i^2 = -1$, we obtain

$$w(z) = u(x,y) + iv(x,y) \qquad \text{where} \quad \begin{cases} u(x,y) = 3 + x + x^2 - y^2, \\ v(x,y) = y + 2xy. \end{cases} \qquad \text{(G.4)}$$

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

$$z = r(\cos\theta + i\sin\theta) = r\exp(i\theta) = re^{i\theta}. \tag{G.8}$$

If we raise this result to the n-th power we obtain De Moivre's theorem:

$$z^n = r^n \exp(in\theta) = r^n(\cos n\theta + i\sin n\theta). \tag{G.9}$$

- Trigonometric and hyperbolic functions are defined by

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \cdots, \tag{G.10}$$

$$\cos z = \frac{e^{iz} + e^{-iz}}{2i} = 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \cdots, \tag{G.11}$$

$$\tan z = \frac{e^{iz} - e^{-iz}}{e^{iz} + e^{-iz}} = z + \frac{1}{3}z^3 + \frac{2}{15}z^5 - \cdots, \tag{G.12}$$

$$\sinh z = \frac{e^{z} - e^{-z}}{2} = z + \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \cdots \tag{G.13}$$

$$\cosh z = \frac{e^{z} + e^{-z}}{2} = 1 + \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \cdots, \tag{G.14}$$

$$\tanh z = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} = z - \frac{1}{3}z^3 + \frac{2}{15}z^5 - \cdots. \tag{G.15}$$

Relations between trigonometric and hyperbolic functions:

$$\cos iz = \cosh z, \qquad \sin iz = i\sinh z, \qquad \tan iz = i\tanh z, \tag{G.16}$$

$$\cosh iz = \cos z, \qquad \sinh iz = i\sin z, \qquad \tanh iz = i\tan z. \tag{G.17}$$

- Using the compound angle formulae (C.5,C.3) gives

$$\sin(x+iy) = \sin x\cos iy + \cos x\sin iy \quad = \sin x\cosh y + i\cos x\sinh y, \tag{G.18}$$
$$\cos(x+iy) = \cos x\cos iy - \sin x\sin iy \quad = \cos x\cosh y - i\sin x\sinh y, \tag{G.19}$$
$$\tan(x+iy) = \frac{\sin x\cosh y + i\cos x\sinh y}{\cos x\cosh y - i\sin x\sinh y} \quad = \frac{\sin x\cos x + i\sinh y\cosh y}{\cos^2 x\cosh^2 y + \sin^2 x\sinh^2 y},$$
$$= \frac{\sin 2x + i\sinh 2y}{\cos 2x + \cosh 2y}. \tag{G.20}$$
$$\sinh(x+iy) = \sinh x\cosh iy + \cosh x\sinh iy = \sinh x\cos y + i\cosh x\sin y, \tag{G.21}$$
$$\cosh(x+iy) = \cosh x\cosh iy + \sinh x\sinh iy = \cosh x\cos y + i\sinh x\sin y, \tag{G.22}$$
$$\tanh(x+iy) = \frac{\sinh x\cos y + i\cosh x\sin y}{\cosh x\cos y + i\sinh x\sin y} \quad = \frac{\sinh x\cosh x + i\sin y\cos y}{\cosh^2 x\cos^2 y + \sinh^2 x\sin^2 y},$$
$$= \frac{\sinh 2x + i\sin 2y}{\cosh 2x + \cos 2y}. \tag{G.23}$$

- Let the real and imaginary parts of the gudermannian function, Equation C.59, be λ and ψ. Let $\zeta = \lambda + i\psi$.

$$\zeta = \mathrm{gd}\, z = \sin^{-1}\tanh(x+iy), \tag{G.24}$$
$$\sin(\lambda + i\psi) = \tanh(x+iy). \tag{G.25}$$

From Equations G.18 and G.23 the real and imaginary parts give

$$\sin\lambda\cosh\psi = \frac{\sinh 2x}{\cosh 2x + \cos 2y}, \tag{G.26}$$
$$\cos\lambda\sinh\psi = \frac{\sin 2y}{\cosh 2x + \cos 2y} \tag{G.27}$$

It is simpler to solve these equations for x and y in terms of λ and ϕ. Setting

$$p = \sin\lambda\cosh\psi, \tag{G.28}$$
$$q = \cos\lambda\sinh\psi, \tag{G.29}$$

Equations G.26, G.27 and their quotient may be written as

$$\sinh 2x = p(\cosh 2x + \cos 2y), \tag{G.30}$$
$$\sin 2y = q(\cosh 2x + \cos 2y), \tag{G.31}$$
$$\sinh 2x = \frac{p}{q}\sin 2y. \tag{G.32}$$

Eliminate y from the first and third equations, using $\sin^2 2y + \cos^2 2y = 1$; eliminate x from the second and third equations, using $\cosh^2 2x - \sinh^2 2x = 1$.

$$\tanh 2x = \frac{2p}{1+p^2+q^2} = \frac{2\sin\lambda\,\mathrm{sech}\,\psi}{1+\sin^2\lambda\,\mathrm{sech}^2\psi}, \tag{G.33}$$
$$\tan 2y = \frac{2q}{1-p^2-q^2} = \frac{2\sec\lambda\sinh\psi}{1-\sec^2\lambda\sinh^2\psi}. \tag{G.34}$$

Comparing these equations with formulae, C.49 and C.7 , for $\tanh 2x$ and $\tan 2y$, gives

$$\tanh x = \sin\lambda\,\operatorname{sech}\psi, \tag{G.35}$$
$$\tan y = \sec\lambda\,\sinh\psi. \tag{G.36}$$

These equations may be inverted but note that from Equations G.24 and C.60

$$\begin{aligned} z = \operatorname{gd}^{-1}\zeta &= \sinh^{-1}\tan(\lambda + i\psi), \\ \sinh(x+iy) &= \tan(\lambda + i\psi), \\ \sinh i(y-ix) &= \tan i(\psi - i\lambda), \\ i\sin(y-ix) &= i\tanh(\psi - i\lambda). \end{aligned} \tag{G.37}$$

This is the same equation as G.25 replacing of $\{x, y, \lambda, \psi\}$ by $\{\psi, -\lambda, y, -x\}$ respectively. With the above substitutions (note $\operatorname{sech}\psi = \operatorname{sech}(-\psi)$) Equations G.36 and G.35 become:

$$\tan\lambda = \sec y\,\sinh x, \tag{G.38}$$
$$\tanh\psi = \sin y\,\operatorname{sech} x. \tag{G.39}$$

Equations G.38 and G.39 define the real and imaginary parts of the gudermanian $\operatorname{gd} z$: Equations G.35 and G.36 define the real and imaginary parts of the inverse gudermanian $\operatorname{gd}^{-1}\zeta$.

For the application to TMS we use Equation 2.55 to set $\operatorname{sech}\psi = \cos\phi$, $\sinh\psi = \tan\phi$ and $\tanh\psi = \sin\phi$ to give

$$\begin{aligned} x &= \tanh^{-1}[\sin\lambda\,\cos\phi], & \lambda &= \tan^{-1}[\sinh x\,\sec y], \\ y &= \tan^{-1}[\sec\lambda\,\tan\phi], & \phi &= \sin^{-1}[\operatorname{sech} x\,\sin y]. \end{aligned} \tag{G.40}$$

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

$$f'(x) = \lim_{\delta x \to 0} \frac{f(x+\delta x) - f(x)}{\delta x}. \tag{G.41}$$

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

$$\left(\frac{\partial f}{\partial x}\right)_y = \lim_{\delta x \to 0} \frac{f(x+\delta x, y) - f(x,y)}{\delta x}, \qquad \left(\frac{\partial f}{\partial y}\right)_x = \lim_{\delta y \to 0} \frac{f(x, y+\delta y) - f(x,y)}{\delta y}. \tag{G.42}$$

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,

$$\delta f = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y. \tag{G.43}$$

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

$$\left(\frac{df}{ds}\right)_{\mathbf{n}} = n_x \frac{\partial f}{\partial x} + n_y \frac{\partial f}{\partial y}. \tag{G.44}$$

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

$$w'(z) = \lim_{\delta z \to 0} \frac{w(z+\delta z) - w(z)}{\delta z}. \tag{G.45}$$

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 $\delta z = \delta x + i\delta y$.

$$w'(z) = \lim_{\delta(x+iy)\to 0} \left(\frac{u(x+\delta x, y+\delta y) + iv(x+\delta x, y+\delta y) - u(x,y) - iv(x,y)}{\delta x + i\delta y} \right). \tag{G.46}$$

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:

$$w'(z) = u_x + iv_x. \tag{G.47}$$

Repeating with limit taken along the y-axis, so that $\delta x = 0$, we have

$$w'(z) = \frac{1}{i}(u_y + iv_y) = v_y - iu_y. \tag{G.48}$$

If we now demand that these two derivatives $w'(z)$ are the same we have the Cauchy–Riemann equations:

$$\boxed{u_x = v_y, \qquad u_y = -v_x} \tag{G.49}$$

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

$$\begin{aligned} w = u + iv &= (x+iy)^3 \\ &= x^3 + 3x^2(iy) + 3x(iy)^2 + (iy)^3 \end{aligned} \tag{G.50}$$

  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.49) are indeed satisfied and we can use either (G.47) or (G.48) to identify the derivative as

$$\begin{aligned} w'(z) &= u_x + iv_x = v_y - iu_y \\ &= 3x^2 - 3y^2 + i6xy = 3(x+iy)^2 \\ &= 3z^2. \end{aligned} \tag{G.51}$$

- 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 $r^n e^{in\theta}$.

- Consider the function $w(z) = \sin z$. The real and imaginary parts of the sine function were determined in equation (G.18) 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.19):

$$\begin{aligned} w'(z) &= u_x + iv_x = v_y - iu_y \\ &= \cos x \cosh y - i \sin x \sinh y \\ &= \cos z. \end{aligned} \tag{G.52}$$

- 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.

$$w(z) = w(z_0) + \frac{1}{1!}(z - z_0)w'(z_0) + \frac{1}{2!}(z - z_0)^2 w''(z_0) + \frac{1}{3!}(z - z_0)^3 w'''(z_0) + \cdots. \tag{G.53}$$

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.

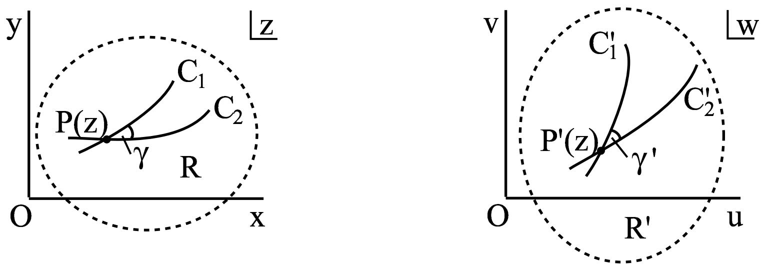


Figure G.2

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 $z-z_0 = re^{i\theta}$. 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

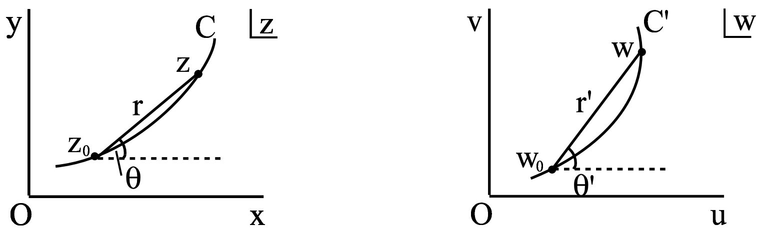


Figure G.3

and w be the corresponding image points and set $w-w_0 = r'e^{i\theta'}$. Taylor's theorem tells us that

$$w(z) = w(z_0) + \frac{1}{1!}(z-z_0)w'(z_0) + \frac{1}{2!}(z-z_0)^2 w''(z_0) + \cdots, \tag{G.54}$$

so that we can write

$$\frac{w-w_0}{z-z_0} = w'(z_0) + \frac{1}{2!}(z-z_0)w''(z_0) + \cdots. \tag{G.55}$$

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.55) becomes

$$\lim_{z\to z_0}\left(\frac{r'}{r}\exp\left[i(\theta'-\theta)\right]\right) = w'(z_0) = A(z_0)\exp(i\alpha(z_0)), \tag{G.56}$$

since the remaining terms on the RHS vanish in the limit. We deduce that

$$\lim_{z\to z_0}\left(\frac{r'}{r}\right) = A(z_0), \tag{G.57}$$

$$\exp i(\theta_0' - \theta_0) = \exp[i\alpha(z_0)], \tag{G.58}$$

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.57,G.58) then imply that the small region is scaled and rigidly rotated, so preserving its shape. This is the property of orthomorphism.

Blank page. A contradiction.

Appendix H

Maxima code

H.1 Common code

Some of the following code is taken from the file 'tmseries.max' at the GeographicLib website (Karney, 2011). The code may be copied from this pdf file and pasted into any version of Maxima but there are slight problems which vary with the chosen pdf viewer. The upquote character (') pastes as a newline-character when copied from Adobe Reader (as of 2013): the caret character (^) is turned into a circumflex (over the next letter) when pasted from the Sumatra viewer. Minimal editing will resolve these problems but they may also be resolved by new versions of the pdf viewers.

The following file is read by all subsequent files. Note that it is a "mac" file whereas the subsequent files are "wxm" files. The path for the load statement in the wxm files should be changed appropriately. It contains trivial initialisations: the parameter maxpow controls the number of terms generated. It also contains code for a maxima function "reverta", a slightly modified of the maxima function revert, which carries out a Lagrange reversion for angle series.

```
/*===== this file is init.mac ======== */
/* INITIALISE*/
maxpow:4$ /* Max power of e*e */
taylordepth:3$
triginverses:'all$
powerdisp:true$
algebraic:true$
debug:false$
/*  LAGRANGE REVERSION
       var2 = expr(var1) = series in eps
    to
       var1 = revertexpr(var2) = series in eps
    Require that expr(var1) = var1 to order eps^0.  This throws in a
    trigreduce to convert to multiple angle trig functions.   */
reverta(expr,var1,var2,eps,pow):=block([tauacc:1,sigacc:0,dsig],
  dsig:ratdisrep(taylor(expr-var1,eps,0,pow)),
  dsig:subst([var1=var2],dsig),
  for n:1 thru pow do (tauacc:trigreduce(ratdisrep(taylor(
     -dsig*tauacc/n,eps,0,pow))),
      sigacc:sigacc+expand(diff(tauacc,var2,n-1))),
      var2+sigacc)$
```

H.2 Lagrange reversion examples

```
/*================ this file is lagrange.wxm =================*/
load ("d:/dropbox/maxima/init.mac")$
load("revert")$
/*===========================================================*/
print("======= POLYNOMIAL REVERSION =======")$
maxpow:8$
/* print out direct, reverted series */
dir:z$  for i: 2 thru maxpow do block(dir: a[i]*z^i+dir)$
print(w=dir)$
kill(w)$
rev:w$  for i: 2 thru maxpow do block(rev: b[i]*(-1)*w^i+rev)$
print(z=rev)$
/* do the reversion */
p:revert(dir,z)$
/* write out coeffts of reversion */
for i: 2 thru maxpow do print(b[i]=expand((-1)*coeff(p,z^i,1)))$

print("")$print("== POLYNOMIAL REVERSION WITH FACTORIALS===")$

/* exhibit factorial factors in coefficients */
dir:z$  for i: 2 thru maxpow do block(dir: a[i]*z^i/i!+dir)$
print(w=dir)$
kill(w)$
rev:w$  for i: 2 thru maxpow do block(rev: b[i]*(-1)*w^i/i!+rev)$
print(z=rev)$
p:revert(dir,z)$
for i: 2 thru maxpow do print(b[i]=expand((-1)*i!*coeff(p,z^i,1)))$

/*=====================================================*/
print("")$ print("======= SINE SERIES REVERSION =======")$

maxpow:4$
/* display series */
s:phi$  for i: 1 thru maxpow do block(s: b[2*i]*sin(2*i*phi)+s)$
print(eta=s)$
v:eta$  for i: 1 thru maxpow do block(v: d[2*i]*e2^i*sin(2*i*eta)+v)$
print(phi=v)$
/* In applications coeffts b_n are order e^(2n). Here assume
   coeffts are O(1) and introduce small eps  to define order
   of coefft before reversion. Then set eps=1 */
/* define series with eps */
s:phi$
for i: 1 thru maxpow do block(s: b[2*i]*eps^i*sin(2*i*phi)+s)$
eta_phi(phi,eps):=s$
p:subst(eps=1,reverta(eta_phi(phi,eps),phi,eta,eps,maxpow))$
/* write coeffts */
for i: 1 thru maxpow do
print(d[2*i]=coeff(p,sin(2*i*eta),1))$
```

H.3 Meridian distance and rectifying latitude

```
/*================ this file is meridian.wxm =================*/
load ("d:/dropbox/maxima/init.mac")$
/*============================================================*/

rho_fn(\phi):=a*(1-e2)*(1-e2*(sin(phi))^2)^(-3/2)$ /* e2=e*e*/
m(phi,e2):=block([mm],
mm:trigreduce(taylor(rho_fn(phi),e2,0,maxpow)),
mm:integrate(mm,phi),
mm:ratdisrep(taylor(mm,e2,0,maxpow)))$

print("====== coeffts of m(phi) in terms of e ======")$
/*write out series for meridian distance m */
m_series:a[0]*%phi$
for i: 1 thru maxpow do m_series: m_series+ a[2*i]*sin(2*i*%phi)$
m=m_series;
/*===== coeffts in terms of e */
m_e(phi,e):=subst(e2=e*e,m(phi,e2))$
a[0]:a*expand((2/%pi)*subst([phi=%pi/2],m_e(phi,e)/a))$
display(a[0])$
for i: 1 thru maxpow do
      a[2*i]:a*coeff(expand(m_e(phi,e)/a),sin(2*i*phi),1)$
for i: 1 thru maxpow do display(a[2*i])$

print("===== coeffts  for Helmert form ======")$
m_h:h[0]*%phi$
for i: 1 thru maxpow do m_h: m_h+ h[2*i]*sin(2*i*%phi)$
m=m_h*(a/(1+n));
m_n(phi,n):=block([t], t:subst(e2=4*n/(1+n)^2,m(phi,e2)),
  expand(ratdisrep(taylor(t*(1+n),n,0,maxpow))))$ /*t/((1-n)^2*(1+n))*/
h[0]:expand((2/%pi)*subst([phi=%pi/2],m_n(phi,n)/a))$
for i: 1 thru maxpow  do
h[2*i]:expand(ratdisrep((1/a)*coeff(m_n(phi,n),sin(2*i*phi),1)))$
display(h[0])$
for i: 1 thru maxpow do display(h[2*i])$

print("====== m_p quadrant (pole-equator) ======")$
m[p]:(%pi/2)*a[0]$
display(m[p])$
m[p]:(a*%pi/2)*(1/(1+n))*h[0]$
display(m[p])$


print("====== rectifying latitude mu(phi) ======")$
/* exhibit series*/
mu_series:%phi$
for i:1 thru maxpow do mu_series: b[2*i]*sin(2*i*%phi)+mu_series$
mu=mu_series;
/* define series and extract coefficients*/
```

```
mu(phi,e2):=expand(ratdisrep(taylor(
  (%pi/2)*m(phi,e2)/subst([phi=%pi/2],m(phi,e2)),  e2,0,maxpow)))$
muphi_e:subst(e2=e*e,mu(phi,e2))$
for i: 1 thru maxpow do print(b[2*i]=coeff(muphi_e,sin(2*i*phi),1))$
muphi_n:subst(e2=4*n/(1+n)^2,mu(phi,e2))$
muphi_n:expand(ratdisrep(taylor(muphi_n,n,0,maxpow)))$
for i: 1 thru maxpow do print(b[2*i]=coeff(muphi_n,sin(2*i*phi),1))$

print("====== revert to phi(mu) ======")$
/* exhibit series*/
phimu:%mu$
for i:1 thru maxpow do phimu: d[2*i]*sin(2*i*%mu)+phimu$
phi=phimu;
/* define series and extract coefficients*/
phimu_e:subst(e2=e*e,reverta(mu(phi,e2),phi,mu,e2,maxpow))$
for i: 1 thru maxpow do print(d[2*i]=coeff(phimu_e,sin(2*i*mu),1))$
phimu_n:subst(e2=4*n/(1+n)^2,reverta(mu(phi,e2),phi,mu,e2,maxpow))$
phimu_n:expand(ratdisrep(taylor(phimu_n,n,0,maxpow)))$
for i: 1 thru maxpow do print(d[2*i]=coeff(phimu_n,sin(2*i*mu),1))$
```

H.4 Conformal latitude

```
/*================ this file is conformal.wxm =================*/
load ("d:/dropbox/maxima/init.mac")$
/*============================================================*/
/* CONFORMAL latitude, chi. Calls LAGRANGE  and ARCTAN

          chi = gd( gd^{-1}phi -e*atanh^{-1}(e*sinh(phi)))

               = asin(tanh( atanh(sin(phi)) - e*atanh(e*sin(phi)) ))

   Notation: tanh(qq) = sin(phi)  OR sinh(qq) = tan(phi)       */

/*-------------------------------------------------------------*/

 print("===================================")$
 print("CONFORMAL (chi) and GEODETIC (phi) ")$
 print("===================================")$

/* define modified atan function */
atanexp(x,eps):=ratdisrep(taylor(atan(x+eps),eps,0,maxpow))$
/* evaluate chi in terms of phi */
psi:qq-e*atanh(e*tanh(qq))$
ratdisrep(taylor(psi,e,0,2*maxpow))$
psi:subst(e=sqrt(e2),psi)$
psi:subst(qq=asinh(tan(phi)),psi)$
tanchi:ratdisrep(taylor(sinh(psi),e2,0,maxpow))$
diff:tanchi-tan(phi)$
```

```
t:subst(x=tan(phi),atanexp(x,eps))$
t:t-subst(eps=0,t)+phi$
t:subst(eps=diff,t)$
t:ratdisrep(taylor(t,e2,0,maxpow))$
t:ratsimp(subst(tan(phi)=sin(phi)/cos(phi),t-phi))+phi$
t:trigreduce(t)$
chi_phi(phi,e2):=t$
print("====== chi(phi) ======")$
s:%phi$ for i: 1 thru maxpow do block(s: b[2*i]*sin(2*i*phi)+s)$
print(chi=s)$
print("====== chi(phi,e) ======")$
chi_e:subst(e2=e*e,t)$
for i: 1 thru maxpow do print(b[2*i]=coeff(chi_e,sin(2*i*phi),1))$
print("====== chi(phi,n) ======")$
chi_n: subst(e2=4*n/(1+n)^2,t)$
chi_n:expand(ratdisrep(taylor(chi_n,n,0,maxpow)))$
for i: 1 thru maxpow do print(b[2*i]=coeff(chi_n,sin(2*i*phi),1))$

/* phi in terms of chi */
 print("")$
 print("====== phi(chi)) ======")$
 s:%chi$
 for i: 1 thru maxpow do block(s: d[2*i]*sin(2*i*chi)+s)$
 print(phi=s)$
 print("====== phi(chi,e) ======")$
 /*phi_chi(chi,e2):=''()$*/
 phi_chi_e: subst(e2=e*e,reverta(chi_phi(phi,e2),phi,chi,e2,maxpow))$
 for i: 1 thru maxpow do print(d[2*i]=coeff(phi_chi_e,sin(2*i*chi),1))$
 print("====== phi(chi,n) ======")$
/* e2_n(n):=4*n/(1+n)^2$*/
 phi_chi_n: subst(e2=4*n/(1+n)^2,
        reverta(chi_phi(phi,e2),phi,chi,e2,maxpow))$
/*phi_chi_n: phi_chi(chi,e2), e2=e2_n(n)$*/
 phi_chi_n:expand(ratdisrep(taylor(phi_chi_n,n,0,maxpow)))$
 for i: 1 thru maxpow do print(d[2*i]=coeff(phi_chi_n,sin(2*i*chi),1))$
```

H.5 Authalic latitude

```
S/*================ this file is authalic.wxm =================*/
load ("d:/dropbox/maxima/init.mac")$
/*===========================================================*/
/* AUTHALIC latitude, xi.
          xi = asin(q(phi)/q_p)
          q(phi) = (1-e2)sin(phi) / 1-e2sin^2(phi)
                    +(1-e2)/e *atanh(e*sin(phi))
             q_p = q(pi/2)
             R_q = a sqrt(q_p/2)
```

```
---------------------------------------------------------------*/
 print("====================================")$
 print("AUTHALIC (XI) and GEODETIC (phi)    ")$
 print("====================================")$

 print("===define series for  xi(phi) ======")$
 s:%phi$
 for i: 1 thru maxpow do block(s: b[2*i]*sin(2*i*phi)+s)$
 %xi=s;
 /* evaluate xi in terms of phi */
 xia:qq+((1-e*e)/e)*atanh(e*sin(phi))$
 xia:ratdisrep(taylor(xia,e,0,2*maxpow))$
 xia:subst(e=sqrt(e2),xia)$
 xia:subst(qq=(1-e2)*sin(phi)/(1-e2*(sin(phi))^2),xia)$
 xia:taylor(xia,e2,0,maxpow)$ if debug then display(xia)$
 q_p:subst(phi=%pi/2,xia)$    if debug then display(q_p)$
 sin_xi:ratdisrep(taylor(xia/q_p,e2,0,maxpow))$
 diff: sin_xi-sin(phi)$
 /* define modified asin function */
 asinexp(x,eps):=ratdisrep(taylor(asin(x+eps),eps,0,maxpow))$
 t:subst(x=sin(phi),asinexp(x,eps))$
 t:t-subst(eps=0,t)+phi$
 t:subst(eps=diff,t)$
 t:ratsimp(t-phi)$
 /* need minus %i in next line to get right answer. Better method?*/
 t:ratsimp(subst(sqrt(-1+(sin(phi))^2)=-%i*cos(phi),t))$
 t:trigreduce(ratdisrep(taylor(t,e2,0,maxpow)))$
 xi_phi(phi,e2):=t+phi$ if debug then display(xi_phi(phi,e2))$
 print("====== xi(phi,e) ======")$
 xi_e: subst(e2=e*e,xi_phi(phi,e2))$ if debug then display(xi_e)$
 xi_e:expand(ratdisrep(taylor(xi_e,e,0,2*maxpow)))$
 for i: 1 thru maxpow do print(b[2*i]=coeff(xi_e,sin(2*i*phi),1))$
 print("====== xi(phi,n) ======")$
 xi_n: subst(e2=4*n/(1+n)^2,xi_phi(phi,e2))$
 xi_n:expand(ratdisrep(taylor(xi_n,n,0,maxpow)))$ xi_n$
 for i: 1 thru maxpow do print(b[2*i]=coeff(xi_n,sin(2*i*phi),1))$

 print("")$
 print("==== series for phi(xi) ======")$
 kill(xi)$ s:%xi$
 for i: 1 thru maxpow do block(s: d[2*i]*sin(2*i*%xi)+s)$
 print(phi=s)$
 kill(xi)$

 /* revert series */
 print("====== phi(xi,e) ======")$
 phi_xi_e: subst(e2=e*e,reverta(xi_phi(phi,e2),phi,xi,e2,maxpow))$
 for i: 1 thru maxpow do print(d[2*i]=coeff(phi_xi_e,sin(2*i*xi),1))$
 print("====== phi(xi,n) ======")$
 phi_xi_n: subst(e2=4*n/(1+n)^2,reverta(xi_phi(phi,e2),phi,xi,e2,maxpow))$
```

```
phi_xi_n:expand(ratdisrep(taylor(phi_xi_n,n,0,maxpow)))$
for i: 1 thru maxpow do print(d[2*i]=coeff(phi_xi_n,sin(2*i*xi),1))$

print("====== authalic radius ======")$
R[q]:a*ratdisrep(subst(e2=e*e,taylor(sqrt(q_p/2),e2,0,maxpow)))$
display(R[q]);
R[q]:ratdisrep(subst(e2=4*n/(1+n)^2,taylor(sqrt(q_p/2),e2,0,maxpow)))$
R[q]:a*expand(ratdisrep(taylor(R[q],n,0,maxpow)))$
display(R[q]);
```

H.6 Redfearn series

In preparation.

Blank page. A contradiction.

Appendix L

Literature and links

This bibliography lists books, journal articles, Wikipedia articles and web sites under a few broad subject headings: map projections, geodesy, geodesics, mathematics and people. Full details of some items have links to the next appendix (References and Bibliography).

The National Geospatial-Intelligence Agency provides access to numerous document.

Wikipedia

The Wikipedia pages are variable in quality but they often include useful references to further material.

Earth radius

Zeno's paradox

Gall-Peters projection

Geoid

Ellipsoid of revolution

Spheroid

Figure of the Earth

meridian arc

French Geodesic Mission to Peru

French Geodesic Mission to Finland

Datum.

North American datum

OSGB36

Ordnance Survey

Latitude

World Geodetic System

Point scale

Theorema Egregium

Retriangulation of Great Britain

Global Positioning System

Triangulation

Global Positioning System

Great-circle distance

George Biddell Airy

Alexander Ross Clarke

George Everest

Carl Friedrich Gauss

Thomas Harriot

William Lambton

Pedro Nunes

Web links

Map production software

Geocart (2008): a commercial map construction program. (Used in the preparation of the TM images in Chapters 3 and 8.)

Online coordinate converters

GeoConvert: an online converter from WGS to UTM. Uses highly accurate transformation equations of Karney (2010).

Geotrans (2010): a free converter which may be downloaded from the National Geospatial-Intelligence Agency. WGS to UTM and many other possibilities. Based directly on the Redfearn transformations of Chapter 7.

Geoscience Australia provides online coordinate conversion and several spreadsheet calculators available through the links on page 2 of the manual. The spreadsheets were designed for use in Australia only but can be converted to global use by minor modifications.

Online calculators for geodesics, distance, rhumbs, etc

GeodSolve: an online calculator for WGS84 only. Solves direct and inverse problems. Uses highly accurate transformation equations of Karney (2010).

Ed Williams' Aviation page. Choice of several ellipsoids as well as mean sphere. Also calculators for way points, rhumb lines etc.

FAI: Choice of WGS ellipsoid or mean sphere.

csgnet Sphere only. Radius such that one arc minute is equal to nautical mile. Way points.

Meridional parts

Map projections

- Maling (1992) *Coordinate Systems and Map Projections*
  Intermediate level text book covering the general features of all projections and details of some. no derivation of the Redfearn formulae. Lots of interesting material.
- Snyder (1987) *Map Projections: a Working Manual*
  Comprehensive summary of most projections in practical use but with no derivations. Free. (See link)
- Snyder (1993) *Flattening the Earth: Two Thousand Years of Map Projections*.
  Comprehensive survey of the history of projections. Inexpensive paperback.
- Bugayevskiy and Snyder (1995) *Map Projections: A Reference Manual*,
  Advanced text book.
- OSGB (1999) *A guide to coordinate systems in Great Britain* here.
  Excellent short survey of geodesy and projections.
- Richardus and Adler (1972) *Map Projections*,
  A highly condensed advanced text.
- Irish National Grid

Geodesy

- Torge (1980) *Geodesy*
  Clear but fairly advanced survey based on modern satellite methods. Available in an inexpensive edition.
- The American Practical Navigator
  Good very elementary discussion. (Chapter 2)
- OSGB (1999) *A guide to coordinate systems in Great Britain* here.
  Excellent short survey of geodesy and projections.
- Clarke (1880) *Geodesy*
  Very clear classic now available as inexpensive reprint (see link). The techniques are outmoded but there are excellent discussions of early determinations of the figure of the Earth and the definition of the metre.
- Laub (1983) *Geodesy and Map Projections*
- WGS (1984) Earth gravitational model.

- WGS (1984) Geodesy for the Layman
- WGS (1984) World Geodetic System 1984 - Its Definition and Relationships with Local Geodetic Systems

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.

Two important papers are included below. The first, by Lee, is the first article in the journal to present a correct derivation of the Transverse Mercator projection formulae using Krüger's lambda expansion. The second paper, by Redfearn, presents a derivation of the series to high enough order to be applicable to all practical problems. These are the papers which are the basis of this article. A number of other papers from the journal are related to the Transverse Mercator projection.

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 technical 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.

- Lee (1945) *The transverse Mercator projection of the spheroid*
- Lee (1946a) *The nomenclature of map projections*,
- Lee (1946b) *The convergence of the meridians*
- Lee (1954) *A transverse Mercator projection of the spheroid alternative to the Gauss-Krüger form*
- Lee (1976) *Conformal projections based on Jacobian elliptic functions*
- Redfearn (1948) *Transverse Mercator formulae*
- Hotine (1946) *The orthomorphic projection of the spheroid, parts I–V*

Geodesics

- Vincenty (1976) *Direct and inverse solutions of geodesics on the ellipsoid with applications of nested tables*
  The formulae encoded into many calculations of ellipsoid geodesics. Now improved by the following:
- Karney (2012) Accurate algorithms which are applicable to near-antipodal points (where Vincenty fails).

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 (1962) *Textbook on spherical astronomy*
  First chapter is a compact survey of Spherical Trigonometry.

- Todhunter (1859) *Spherical Trigonometry*,
  A splendid traditional account of the subject. Many editions. Final edition (1901) revised by J G Leathem is best.

- Euclid (300BC) *The elements*.
  Book 11 contains results required by Todhunter.

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 Reversion

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

More references

Surveying

- RAINSFORD H F, (1957), *Survey adjustments and least squares*, Constable, EUMAIN .52633 RAI

Projections

- JACKSON J E, (1987), *Sphere, Spheroid and Projections for Surveyors*, BSP Professional Books, ISBN: 0632-01867-4. [ROB TA 549 .Jac]

SR

- BOMFORD, A G (1962), Survey Review, Vol **16**, Part 125 pp 318–327.
  Transverse Mercator arc-to-chord and finite distance scale factor formulae.
- COLE, J H (1946), Survey Review, Vol **8**, Part 59 pp 191–194.
  Computation of distances of long arcs for radio purposes.
- HIRVONEN, R A (1953), Bulletin Geodesique, Part **30**, pp 369–392.
  Two papers: (a) Nutshell tables of mathematical functions for interpolation with calculating machines; (b) Tables for the computation of long lines.
- HIRVONEN, R A (1957), Bulletin Geodesique, Part **43**, pp 3–15.
  Computations of triangulations on the ellipsoid by the aid of closed formulas.
- HOTINE, M (1946b), Survey Review, Vol **8**, Part 61 pp 276–277.
  Nomenclature of map projections.
- LAMBERT, W D (1944), Survey Review, Vol **7**, Part 50 pp 172–176.
  The distance between two widely spaced points on the surface of the Earth. (This is actually a review of a paper by Lambert).
- LAUF, G B (1948), Survey Review, Vol **9**, Part 68 pp 259–260.
  The length of the meridian arc.
- LEE, L P (1946b), Survey Review, Vol **8**, Part 60 pp 217–219.
  Nomenclature of map projections.
- LEE , L P (1946c), Survey Review, Vol **8**, Part 61 pp 267–271.
  The convergence of the Meridians. YES
- LEE, L P (1954), Survey Review, Vol **12**, Part 87 pp 12–17.
  A transverse Mercator projection of the spheroid alternative to the Gauss-Krüger form.

- RAINSFORD, H F (1946), Survey Review, Vol **8**, Part 56 pp 53–68 and Part 57 pp 102–114.
  The Clarke formulae for latitude, longitude and reverse azimuth, with corrective terms on very long lines. Part I: Large spheroidal triangles. Part II: The Clarke formulae in detail.
- RAINSFORD, H F (1949), Survey Review, Vol **10**, Parts 71,72 pp 19–29 and 74–81.
  Long lines on the earth: various formulae.
- RAINSFORD, H F (1953), Bulletin Geodesique, Part **37**, pages 12–22.
  Long Geodesics on the ellipsoid.
- ROBBINS, A R (1962), Survey Review, Vol **16**, Part 125 pp 301–309.
  Long lines on the spheroid.

Differential Geometry

- GRAUSTEIN W C, (1966), *Differential Geometry*, Dover reprint of original Macmillan edition. [JCMB QA 641 Gra]
- STRUIK D J, (1950, 1961), *Differential Geometry*, Addison–Wesley (Lib. Cong, 61–12774) [JCMB QA 641 Str]

Geodesy

- CLARKE, A R (1880), *Geodesy*, Clarendon Press, Oxford.

  A classic which is old enough to be very clear! He concludes with interesting chapters showing how (many) survey results are used to define a measured ellipsoid. His ellipsoid is still the basis of many map projections: it was used by the United States until very recently.
- BOMFORD G, (1971), *Geodesy*, Clarendon Press, Oxford.

  A more modern classic: it is heavy going but it is all there.
- LEICK A, (1995), *GPS satellite surveying*, Wiley, ISBN: ??? . [EULIB .526982 Lei.]
- SMITH J R, (1997), *Introduction to geodesy : the history and concepts of modern geodesy*, Wiley, ISBN: ??? . [NATLIB SP2.97.2241]
- VANICEK P, KRAKIWSKY E, (1982), *Geodesy: the concepts*, North-Holland, ISBN: 0444877754, PBK 0444877770 . [EUROB QB281 Tor]
- KEAY, J (2000), *The Great Arc*, Harper-Collins, ISBN: 0 00 257062 9 .

Blank page. A contradiction.

Appendix R

References and Bibliography

ADAMS, Oscar S, 1921, *Latitude Developments Connected With Geodesy and Cartography, (with tables, including a table for Lambert equal area meridional projection)*. Special Publication No. 67 of the US Coast and Geodetic Survey. A facsimile of this publication is available from the US National Oceanic and Atmospheric Administration (http://www.noaa.gov) at http://docs.lib.noaa.gov/rescue/cgs_specpubs/QB275U35no671921.pdf NB: Adams uses the the term isometric latitude for the conformal latitude of this article. [Cited page 36, 103, 105, 106, 108]

BESSEL, F, W, 1825. *Über die Berechnung der geographischen Längen und Breiten aus geodätischen Vermessungen.* Astronomische Nachrichten, volume **4**, Issue 16, p.241–254. Translated into English by Karney C, F, F and Deakin R, E as *The calculation of longitude and latitude from geodesic measurements*. Astronomische Nachrichten, volume 331, Issue 8, p.852861 (2010). A copy of the preprint may be accessed from GeographicLib or directly at Archiv.org. [Cited page 97, 99]

BUGAYEVSKIY, L. M. and SNYDER J, P., 1995, *Map Projections: A Reference Manual*, CRC Press, ISBN: ISBN 0748403043. [Cited page 225]

CLARKE, Alexander Ross, 1880, *Geodesy*. Clarendon Press. Currently (2012) available from Forgotton Books at http://www.forgottenbooks.org/info/Geodesy_1000151390.php. It is also available online at http://books.google.co.uk/books?id=o_IOPzB--OAC&lpg=PA1&dq=alexander%20ross%20clarke&pg=PAI#v=onepage&q&f=false ISBN: http://en.wikipedia.org/wiki/Special:BookSources/9781440088650 [Cited page 10, 10, 225]

CLENSHAW, C. W., 1955, *A note on the summation of Chebyshev series*, Math. Tables Aids Comput., **9**(51), 118–120, http://www.jstor.org/stable/2002068. [Cited page]

DEAKIN, R, E, 2010. A useful selection of articles is available at http://user.gs.rmit.edu.au/rod/publications.htm: (1) Geometric Geodesy A, (2) Bessel's method for Geodesics, (3) The loxodrome on an Ellipsoid, (4) The Geoid, (5) A guide to map projections, (6) The Gauss-Krueger transverse Mercator projection. [Cited page 105]

DELAMBRE, Jean Baptiste Joseph, 1799 *Méthodes Analytiques pour la Dtermination d'un Arc du Méridien; précédées d'un mémoire sur le même sujet par A. M. Legendre* Google books [Cited page 98]

DOZIER, J. 1980, *Improved algorithm for calculation of UTM and geodetic coordinates*, Technical Report NESS 81, NOAA, http://fiesta.bren.ucsb.edu/~dozier/Pubs/DozierUTM1980.pdf. [Cited page 151]

ENGSAGER, K. E. and PODER, K., 2007, *A highly accurate world wide algorithm for the transverse Mercator mapping (almost)*, in *Proc. XXIII Intl. Cartographic Conf. (ICC2007), Moscow*, p. 2.1.2. [Cited page]

EUCLID, 300BC,*The elements*. For references see Wikipedia and David Joyce at Clark university. [Cited page 183, 227]

GAUSS, Karl Friedrich, 1827, *General Investigations of Curved Surfaces*. Published in Latin in the Proceedings of the Royal Society of Gottingen. A translation made in 1902, published by the University of Princeton, is available at the Gutenberg Project: http://www.gutenberg.org/files/36856/36856-pdf.pdf See also http://en.wikipedia.org/wiki/Gauss [Cited page 12]

GEOCART, 2008, *Geocart, version 3.0*, http://www.mapthematics.com. [Cited page 224]

GEOTRANS, 2010, Geographic translator, version 3.0 is available from National Geodesy and Geophysics publications. [Cited page 134, 137, 141, 224]

HALLEY, Edmond, 1696, *An easie demonstration of the analogy of the logarithmick tangents to the meridian line or sum of the tangents, with various methods for computing the same, to the utmost exactness* , Phil. Trans. Royal Soc. London **19**, 1696, pp199–214. Full text [Cited page 17]

HELMERT F, R, 1880, *Die Mathematischen und Physikalischen Theori- een der Höheren Geodäsie*, volume1. Teubner, Leipzig. Google books. Translated into English by Aeronautical chart and information center (St Louis, 1964) as *Mathematical and physical theories of higher geodesy, Part 1*. A copy of this translation is available at GeographicLib. [Cited page 100]

HOLLANDER, Raymond d'., 2005, *Loxodromie et projection de Mercator*. Published in Paris at Institut Océanographique. ISBN: 2903581312 and 9782903581312 [Cited page 16]

HOTINE, M., 1946, *The orthomorphic projection of the spheroid, , parts I–V*, 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. [Cited page 3, 226]

KARNEY, C. F. F., 2010, *GeographicLib, version 1.7*, http://geographiclib.sf.net. This page links to a geographical coordinate conversion program, GEOCONVERT). [Cited page 137, 151, 224, 224]

KARNEY, C. F. F., 2011, http://dx.doi.org/10.1007/s00190-011-0445-3 *Transverse Mercator with an accuracy of a few nanometers*, J. Geodesy 85(8), 475-485 (2011); preprint of paper and C++ implementation of algorithms are available at http://geographiclib.sourceforge.net/tm.html [Cited page 12, 17, 74, 151, 151, 215]

KARNEY, C. F. F., 2011, http://dx.doi.org/10.1007/s00190-012-0578-z *Geodesics on an ellipsoid of revolution*, J. Geodesy (2012); an earlier expanded version is available at http://arxiv.org/abs/1102.1215. For more information see http://geographiclib.sourceforge.net [Cited page 97, 226]

KRÜGER, J. H. L., 1912, *Konforme Abbildung des Erdellipsoids in der Ebene*, New Series 52, Royal Prussian Geodetic Institute, Potsdam, doi:`10.2312/GFZ.b103-krueger28`. [Cited page 17, 117, 151]

KUITTINEN, K., *et al.* , 2006, *ETRS89—järjestelmään liittyvät karttaprojektiot, tasokoordinaatistot ja karttalehtijako*, Technical Report JHS 154, Finnish Geodetic Institute, Appendix 1, Projektiokaavart, http://docs.jhs-suositukset.fi/jhs-suositukset/JHS154/JHS154_liite1.pdf. [Cited page]

LAGRANGE. J. L., 1770, *Nouvelle méthode pour résoudre les équations littérales par le moyen des séries*, in *Oeuvres*, volume 3, pp. 5–73 (Gauthier-Villars, Paris, 1869), reprint of Mém. de l'Acad. Roy. des Sciences de Berlin **24**, 251–326, http://books.google.com/books?id=YywPAAAAIAAJ&pg=PA5. [Cited page]

LAMBERT, J. H., 1772, *Anmerkungen und Zusätze zur Entwerfung der Land- und Himmelscharten*, number 54 in Klassiker ex. Wiss. (Engelmann, Leipzig, 1894). The original text is available at the University of Michigan Historical Math collection: http://quod.lib.umich.edu/u/umhistmath/ABR2581.0001.001. The work was translated into English by W. R. Tobler as *Notes and Comments on the Composition of Terrestrial and Celestial Maps*, and published by the University of Michigan (1972). It has been reprinted (2010) by Esri: http://store.esri.com/esri/showdetl.cfm?SID=2&Product_ID=1284&Category_ID=38. [Cited page 17, 31, 49]

LAUB, G. B., 1983, *Geodesy and Map Projections*, TAEF Publications unit, Collingwood. [Cited page 225]

LEE, L. P., 1945, *The transverse Mercator projection of the spheroid*, Survey Review, **8** (Part 58), pp.142–152. http://www.ingentaconnect.com/content/maney/sre/1945/00000008/00000058/art00004. Errata and comments in Survey Review, **8** (Part 61), pp.277–278. [Cited page 17, 36, 117, 147, 226]

LEE, L. P., 1946a, *The nomenclature of map projections*, Survey Review, **8** (Part 60), pp.217–219. [Cited page 36, 226]

LEE, L. P., 1946b, *The convergence of the meridians*, Survey Review, **8** (Part 61), pp.267–271. [Cited page 47, 226]

LEE, L. P., 1954, *A transverse Mercator projection of the spheroid alternative to the Gauss-Krüger form*, Survey Review, **12** (Part 87), pp.12–17. [Cited page 226]

LEE, L. P., 1976, *Conformal projections based on Jacobian elliptic functions*, Cartographica, **13**(1, Monograph 16), 67–101, doi:10.3138/X687-1574-4325-WM62. [Cited page 17, 74, 151, 226]

LOHNE, J. A., 1965 *Thomas Harriot als Mathematiker*, Centaurus 11, pp19–45. [Cited page 17, 41]

LOHNE, J. A., 1979 *Essays on Thomas Harriot*, Archive for history of exact sciences Vol **20**, Numbers 3-4 (1979), 189-312, doi:10.1007/BF00327737 [Cited page 17]

MALING, D. H., 1992, *Coordinate Systems and Map Projections* (2nd edition). Published by Pergamon Press, ISBN:0-08-037033-3. [Cited page 22, 36, 225]

MAXIMA, 2009, *A computer algebra system, version 5.20.1*, http://maxima.sf.net. [Cited page 121]

MERCATOR, G, 1569, World Map. http://en.wikipedia.org/wiki/Mercator1569worldmap. [Cited page 32, 42]

MONMONIER, Mark, 2004, *Rhumb Lines and Map Wars: A Social History of the Mercator Projection*, University of Chicago Press, ISBN=0-226-53431-6 http://books.google.co.uk/books/about/Rhumb_Lines_and_Map_Wars.html?id=nvwu4Ba_Qp0C [Cited page 32, 40]

NAUTICAL MILE, 1929 The international nautical mile was defined by the First International Extraordinary Hydrographic Conference at Monaco as exactly 1852 metres. See also http://en.wikipedia.org/wiki/Nautical_mile [Cited page 24]

NEWTON, Isaac, 1687, *The Principia*, Book III Proposition XIX Problem III. A modern annotated translation is available at http://17centurymaths.com/contents/newton/book3s1.pdf, page 25 of the pdf, correspondinding to page 746. See also Page 405 in the Andrew Motte translation (1729), available on line at http://www.archive.org/details/100878576#page/404/mode/2up (First American edition by N.W.Chittenden (1846). A pdf version may be downloaded from http://www.archive.org/details/newtonspmathema00newtrich. The bulk of the Chittenden version is also available at http://en.wikisource.org/wiki/The_Mathematical_Principles_of_Natural_Philosophy_(1846). [Cited page]

NIST: Olver F. W. J., Lozier D. W., Boisvert R. F. and Clark C. W., editors, 2010, *NIST Handbook of Mathematical Functions* (Cambridge University Press), http://dlmf.nist.gov. [Cited page 97]

OSGB, 1999, *A guide to coordinate systems in Great Britain* at http://www.ordnancesurvey.co.uk/oswebsite/gps/information/coordinatesystemsinfo/guidecontents. See also OSGB36 and Ordnance Survey [Cited page 11, 13, 14, 46, 49, 91, 99, 100, 134, 135, 139, 143, 147, 149, 225, 225]

PEPPER, J. V., 1967, *Harriot's calculation of the meridional parts as logarithmic tangents*. Archive for the history of exact sciences vol **4** pp359–413. Purchase: doi:10.1007/BF00327697 [Cited page 17]

PRIME MERIDIAN CONFERENCE, 1884, *The proceedings of the International Conference held at Washington (USA) for the purpose of fixing a prime meridian and a universal day.* http://www.gutenberg.org/ebooks/17759. See also the Wikipedia articles at http://en.wikipedia.org/wiki/Prime_meridian. [Cited page 21]

RANDLES, W. G. L., 2000, *Pedro Nunes discovery of the loxodromic curve (1537): how Portuguese sailors in the early sixteenth century, navigating with globes, had failed to solve the difficulties encountered with the plane chart*. Journal of Navigation **50** (1997), pp85–96, reprinted as chap. XIV in W. G. L. Randles, *Geography, Cartography and Nautical Science in the Renaissance: The Impact of the Great Discoveries* , Variorum Collected Studies, Ashgate, 2000. [Cited page 16]

RAPP, Richard H, 1991. Geometric Geodesy, Part I. [Cited page]

REDFEARN, J. C. B., 1948, *Transverse Mercator formulae*, Survey Review, **9** (Part 69), pp. 318–322, http://www.ingentaconnect.com/content/maney/sre/1948/00000009/00000069/art00005 *Transverse Mercator formulae* [Cited page 3, 17, 36, 117, 134, 147, 226]

RICHARDUS P, ADLER R K, 1972, *Map Projections*, North-Holland (Amsterdam), ISBN: 0 7024 50071. [Cited page 225]

SMART W, M., 1962, *Textbook on spherical astronomy* [Cited page 227]

SNYDER, John P, 1987 *Map Projections—A Working Manual*. U.S. Geological Survey Professional Paper 1395 at http://pubs.er.usgs.gov/pubs/pp/pp1395 [Cited page 12, 36, 103, 147, 149, 225]

SNYDER, John P., 1993, *Flattening the Earth: Two Thousand Years of Map Projections*. Published by the University of Chicago Press. ISBN:0-226-76747-7 [Cited page 30, 46, 225]

STEDALL, J. A., 2000 *Robd́ of Glories: The Posthumous Misfortunes of Thomas Harriot and His Algebra*. Archive for History of Exact Sciences, Vol **54** pp455–497. doi:10.1007/s004070050041. [Cited page 17]

STUIFBERGEN, N., 2009, *Wide zone transverse Mercator projection*, Technical Report 262, Canadian Hydrographic Service, http://www.dfo-mpo.gc.ca/Library/337182.pdf. [Cited page 135, 151]

TAYLOR E. G. R., and SADLER, D. H., 1953, *The Doctrine of Nauticall Triangles Compendious*. J. Inst. Navigation **6** pp131–147 [Cited page 17]

THOMAS, Paul D, 1952, *Conformal projections in Geodesy and Cartography*, Special Publication No. 251 of the US Coast and Geodetic Survey. A facsimile of this publication is available from the US National Oceanic and Atmospheric Administration (http://www.noaa.gov) at http://docs.lib.noaa.gov/rescue/cgs_specpubs/QB275U35no2511952.pdf. [Cited page 17, 117, 120, 134, 134, 143, 147]

TODHUNTER I, 1859, *Spherical Trigonometry*, Macmillan (London). 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/ [Cited page 183, 227]

TORGE, W, 1980 *Geodesy*. Published by de Gruyter, ISBN-13: 978-3110170726. (3rd edition, paperback) [Cited page 225]

UTM, Hager J. W., Behensky, J. F. and Drew B. W. , 1989, National Geodesy and Geophysics publications include a two volume manual. Volume 1 : *Datums, Ellipsoids, Grids, and Grid Reference Systems* (Technical Report TM 8358.1). Volume 2: *The universal grids: Universal Transverse Mercator (UTM) and Universal Polar Stereographic (UPS)* (Technical Report TM 8358.2). [Cited page 46, 49, 134, 135, 138, 138]

VINCENTY, T, 1976, *Direct and inverse solutions of geodesics on the ellipsoid with applications of nested tables*. Survey Review, Vol **23**, Part 176 pp 88–93. A copy of this paper is available at http://www.ngs.noaa.gov/PUBS_LIB/inverse.pdf. [Cited page 97, 226]

WALLIS, D. E. 1992, *Transverse Mercator projection via elliptic integrals*, Technical Report NPO-17996, JPL. [Cited page]

WEISSTEIN, Eric, 2012, *Gudermannian* at http://mathworld.wolfram.com/Gudermannian.html. [Cited page 37]

WEISSTEIN, Eric, 2012, *Loxodrome* at http://mathworld.wolfram.com/Loxodrome.html. [Cited page 37]

WHITTAKER, E. T. and WATSON, G. N., 1927, *A Course of Modern Analysis* (Cambridge Univ. Press), 4th edition, reissued in Cambridge Math. Library Series (1996). [Cited page]

WORLD GEODETIC SYSTEM, 1984. The National Geospatial-Intelligence Agency provides access to numerous documents including Geodesy for the Layman, WGS84 definitions and the GEOTRANS: translator between geodetic coordinates and UTM grid refernces. See also Wikipedia: World Geodetic System. [Cited page 22, 136, 225, 226, 226]

WRIGHT, Edward, 1599, *Certaine Errors in Navigation ; The Voyage of ... George Earle of Cumberland to the Azores*. Published (1974) in facsimile in the series *English experience, its record in early printed books* by Theatrum Orbis Terrarum. , Wikipedia: Edward Wright, Google books [Cited page 16, 40]

Blank page. A contradiction.

Index X

Index