

Open Research Online

The Open University's repository of research publications and other research outputs

Applications of numerical analysis in navigation

Thesis

How to cite:

Williams, Roy (1994). Applications of numerical analysis in navigation. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1994 The Author

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's [data policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

APPLICATIONS
OF
NUMERICAL ANALYSIS
IN
NAVIGATION

by

ROY WILLIAMS

Bachelor of Science (London University)

Bachelor of Philosophy (Open University)

THESIS PRESENTED FOR
DOCTOR OF PHILOSOPHY
IN
APPLIED MATHEMATICS

24th MAY 1994

- CONTENTS -

0 - INTRODUCTION 1

- PART 1 -

APPLICATION OF NUMERICAL METHODS IN
NAVIGATION

1 - THE GEOMETRY OF THE EARTH 10

2 - NAVIGATING ALONG A LOXODROMIC CURVE ON THE
SURFACE OF A SPHERE 17

3 - NAVIGATING ALONG A LOXODROMIC CURVE ON THE
SURFACE OF A SPHEROID 37

4 - NAVIGATING ALONG THE GEODESIC PATH BETWEEN
TWO POINTS ON THE SURFACE OF A SPHERE 54

5 - NAVIGATING ALONG THE GEODESIC PATH BETWEEN
TWO POINTS ON THE SURFACE OF A SPHEROID 72

6 - NAVIGATING ALONG THE GEODESIC PATH BETWEEN
TWO NEARLY ANTIPODEAN POINTS ON THE EQUATOR
OF A SPHEROID 96

7 - NAVIGATING ALONG THE GEODESIC PATH BETWEEN
TWO NEARLY ANTIPODEAN POINTS AWAY FROM THE
EQUATOR OF A SPHEROID 106

8 - COMPUTATION OF ASTRONOMICAL FIXES 116

9 - COMPUTING THE OBSERVED POSITION FROM THE
OBSERVATION OF A SINGLE BODY 128

- PART 2 -

DIRECT SPLINE APPROXIMATIONS TO INTEGRALS

10 - DIRECT CUBIC SPLINE APPROXIMATION TO AN INTEGRAL149
11 - HIGHER ORDER DIRECT SPLINE APPROXIMATIONS TO AN INTEGRAL164
12 - CONCLUSION177
- - REFERENCES183

- APPENDICES -

A1 - TRANSFORMATION OF EQUATION (5.4) TO (5.5)187
A2 - A TABLE OF LATITUDE PARTS188
A3 - SOME FORMULAE IN SPHERICAL TRIGONOMETRY189

ABSTRACT OF THESIS
"APPLICATIONS OF NUMERICAL ANALYSIS IN NAVIGATION"

by

ROY WILLIAMS

PRESENTED FOR THE DOCTOR OF PHILOSOPHY DEGREE IN APPLIED MATHEMATICS

July 1994

Part one of the thesis contains an analysis of the methods of computation in navigation. We start with loxodromic navigation and, although this subject is a well documented, we make a positive attempt to analyse the subject matter using the methods of differential geometry. We then turn to the problem of shortest path curves and set out an alternative method of solving the problem of navigating along the arc of a great circle on the surface of a sphere which can be generalised to other surfaces. In particular, a contribution made by the thesis is an analysis of the problem of navigating along shortest path geodesic arcs on the surface of a spheroid which introduces an algebraic representation of the geodesic curve by solving Clairaut's equation using a cylindrical transformation. We are therefore able to compute the the coordinates of the positions of points along the path of the geodesic and the length of arc along the path of the geodesic curve can then be computed step by step between these points by a numerical method - the Direct Cubic Spline method which was first introduced by this author in the Bachelor of Philosophy thesis in 1982 and is developed further in part 2 of this thesis. We apply this method also to the special problem of computing the distance along the shortest path between nearly antipodean points on the surface of a spheroid.

We analyse the problem of computing an observer's position on the surface of the Earth using astronomical observations and show how a position locus is distorted when it is transferred over the surface. We offer a method of computing the observer's position by a series of observations of a single astronomical body taken over a comparatively short period of time and which does not necessarily include an observation at the time of meridian passage of the body.

In part two of the thesis we discuss the Direct Spline approximation to integrals and give some error bounds. The Direct Cubic Spline is a step by step method of fitting a cubic spline to the integral of a function *directly* which computes the value of the integral of the function step by step between the data points which need not be evenly spaced. We extend the idea to splines of higher order and give the formulae from which they may be obtained but we show that, except for a particular special form of the direct quartic spline, the higher order direct splines do not yield algorithms for computing integrals which are as efficient as the Direct Cubic Spline.

- 0 -

INTRODUCTION

The Science of Navigation has entered the electronic age with a great deal of enthusiasm and energy. Orbiting man made satellites have, to a large extent, replaced the stars in the attentions of a modern navigator and it is the orbits of these satellites, rather than the rotation of the heavens, which now tend to control the pattern by which the observer's position is fixed during the day. The science is moving forward at a fast rate. Radio position fixing systems, which were created rapidly in the years immediately following the second world war (many as a result of methods developed during that period out of necessity) and which, for the science of navigation, seemed to have made such great strides forward at the time, have been and gone and are now obsolete, historical even. The new age of space travel (again largely through the necessity) has helped to produce compact computer controlled position fixing systems with built in computational procedures. These are systems which not only keep the observer informed of his present position but which also hold his history file and which update his travel plans.

During this time we have not perhaps paused to think how the classical methods of navigation, particularly the mathematics, could be reviewed to give a modern outlook to the subject. Even today, some educational programs, particularly in seagoing navigation, seem to be designed purposely to avoid the use of the calculus. For this reason methods are still in use which are only approximations. Due to the advent of modern computing devices, which relieve the tedium of long computations and the possibility of errors, this is no longer necessary. Many problems, which seemed algebraically and numerically intractable previously, can now be treated as a routine and, consequently, many aspects of the mathematical analysis can also be reviewed and perhaps improved.

In this thesis, therefore, this is our theme. We have taken the opportunity to introduce some new numerical methods of our own which we find to have particularly useful applications in navigation and

we have used the methods of differential geometry to make what we feel is a reasonably full analysis of the methods of navigation when applied to the sphere and the spheroid. This has, in turn, lead us to make what we feel are some contributions to the mathematical theory.

Part 1 of the thesis is concerned with the application of numerical methods in navigation, paying due attention to the underlying mathematical models used. There does not seem to be any written exposition in navigation which adopts a rigorous mathematical approach to the subject. It is the purpose in the early chapters of this thesis, therefore, to correct this. We adopt two models for the shape of the Earth - the spherical model and the spheroidal model - and, using the methods of differential geometry, show how the familiar navigational formulae are derived. The spherical model is not such a bad approximation when applied to the problems in navigation, *if it used consistently*, but, in practice, this has not been so and the habit developed where elements computed from different models have been used *in the same formula*. For example, this mistake was actually made in two different publications of nautical tables. A correction to apply to the *Mean Latitude* to give *Middle Latitude* for an observer travelling along the arc of a loxodrome (rhumb line) was computed when middle latitude was determined using a formula which was given as

$$\text{Cos(Middle Latitude)} = \frac{\text{Difference of latitude}}{\text{Difference of Meridional Parts}}$$

As is well known, difference of latitude is the number of minutes of arc on the meridian of a sphere but the difference of meridional parts used to compile the above mentioned correction table was a table computed from spheroid data. The middle latitude correction table was published through a number of editions and this went apparently unnoticed but it seems that the theory behind the method of computation known as "Middle Latitude Sailing" was never really widely understood. If difference of latitude is replaced by the

meridian distance then the formula for middle latitude would be correct.

In chapter 1 we define the ways in which a position on the Earth's surface may be defined when both spherical and spheroidal models are used. The spheroidal model is a *regular* spheroid by which we mean that the spheroid is generated by revolving an ellipse about its minor axis and that this spheroidal model is used as a global model for the surface of the Earth. We set the Earth in a spherical coordinate system denoted in the standard form for these coordinates; r, θ, ϕ where θ is the longitude and ϕ is the geocentric latitude. Since arc length and geodetic latitude are intrinsic properties of the surface we use the standard notation for intrinsic coordinates and denote arc length by s and geodetic latitude by ψ .

In chapter 2, using the methods of differential geometry, we analyse the problem of navigating along a loxodrome on the surface of a sphere. We are concerned here to derive with mathematical rigour the familiar navigational formulae for the spherical model and, although the method is now only historical, we try to give a rigorous derivation of the theory behind the method known as "Middle Latitude Sailing". W.N.Smart' presented a complete analysis of Middle Latitude Sailing for both the sphere and the spheroid. In both cases he relies upon the computation of meridional parts but, in fact, on the surface of the sphere the problem can be solved approximately without meridional parts. In the case of the spheroid, however, the problem cannot really be solved without recourse to formulae provided by the method of Mercator Sailing so that the method does not exist independently.

Chapter 3 deals with the same problem for navigating on the surface of a spheroid. For comparatively small gains in accuracy this problem on the surface of the spheroid needs some fairly high powered numerical methods to aid its solution but it does stimulate a lot of theoretical interest and the best way find a numerical

solution still promotes discussion in the pages of the Journal of Navigation. This author had a paper published in 1981² concerning the computation of meridian distance and since then there have been others written on the same theme. The latest contribution was by Kitt C. Carlton Wipperfurth in 1992³.

In chapter 4 we discuss the problem of navigating along the shortest paths on the surface of a sphere. These shortest paths are, of course, great circles and, although the methods of spherical trigonometry are well suited to many aspects of the solution of this problem, we show also that the problem can be also be solved very conveniently by defining a great circle in the form of a differential equation (Clairaut's equation) and solving this equation using numerical integration by the Direct Cubic Spline method. The direct cubic spline method was introduced in the Bachelor of Philosophy thesis by this author and is developed further in this thesis in chapter 10. This method of numerical integration has the convenience of being a step by step method which computes intermediate points along the path sequentially, the distance between them, the overall distance from the starting point and the course at the intermediate points. The intermediate points can be chosen by the navigator to coincide with the "way" points that would normally be chosen. For great circles the method gives results whose accuracy leaves nothing to be desired by the sea-going navigator even though it is a numerical "approximation". Although the methods of spherical trigonometry give exact formulae from which to compute the elements of the great circle, the numerical results from the direct cubic spline are indistinguishable and it can be argued that there is more computation to be done in the methods using spherical trigonometry. It is not the purpose, however, to *compete* with the methods of spherical trigonometry but to provide an alternative method of solution which can be generalised to other surfaces.

This, then, leads into chapter 5 where we solve the same shortest path problem on the surface of a spheroid by defining the geodesic

arc by means of Clairaut's equation as in chapter 4. There is, of course, a great deal more numerical computation needed. For navigators the most widely promulgated methods of solving this problem seem to be the *correction* methods where a correction is applied to the great circle distance on the sphere in order to determine the shortest distance between two points with corresponding positions on the surface of the spheroid. See, for instance, the publication by Paul D. Thomas⁴ published by the U.S. Naval Oceanographic Office. In the publications concerned with Geodetics the problem of computing the shortest distance on the surface of a spheroid is generally posed in two ways - the "direct" problem and the "inverse" problem. The "direct" problem is posed so that, given a starting position on the surface of the spheroid, an initial course (azimuth) and a distance travelled, we compute the final position when the path taken is the shortest path. The "inverse" problem is posed so that, given two points on the surface of the spheroid we compute the shortest distance between them. The direct problem is not one that is often posed by the navigator and the inverse problem is only a partial answer to the navigator's quest. What the navigator needs is a system of computation which will plot the path of the geodesic arc so that this path can be considered for suitability and so that the intermediate points can be plotted on a chart. In other words, given any longitude, the navigator would like to know the corresponding latitude where the geodesic arc crosses the meridian. One of the major expositions written on the problem of geodesic arcs on the surface of the spheroid was produced by Fichot⁵. In that work Fichot says that, in general, the geodesic arc is not expressible in algebraic form :

" ..Exception faite pour l'equator et les ellipses méridiennes, aucune de ces courbes n'est algebrique.. ". We have found no other author who contradicts this statement but we have opted to solve the problem by a method different to that which appears in the literature and to solve Clairaut's equation using transformations which, in effect, project the geodesic arc onto the surface of a cylinder coaxial with the spheroid. There is a lot of manipulative algebra (details of which are shown in appendix 1) but, for the

effort, we find that, for a geodesic curve which reaches its vertex where the geocentric latitude is ϕ_v and the longitude is θ_v , when the geodesic curve is *not* the equator or a meridian then ϕ and θ are connected by the relationship

$$\tan \phi = \tan \phi_v \cos[\mu(\phi)(\theta_v - \theta)]$$

where

$$\mu(\phi) = \frac{1}{u'} \int_0^{u'} \sqrt{\frac{a^2(1-e^2)^2 + \tan^2 \phi_v \sin^2 u}{a^2(1-e^2) + \tan^2 \phi_v \sin^2 u}} du$$

$$u' = \sin^{-1}\left(\frac{\tan \phi}{\tan \phi_v}\right)$$

and a is the equatorial radius of the spheroid. When the surface is a sphere then the same equation applies and $\mu(\phi) \equiv 1$.

Lambert⁶, paying tribute to Fichot, stated that, at that time (1942) the work by Fichot was the fullest exposition so far written on the subject of geodesic arcs on the surface of the spheroid and that there was no comparable work written in English. He went on to say that it was his intention to publish further work on the subject himself and, in particular, to publish work on the special problem of the shortest distance between nearly antipodean points. As far I can see, such a full exposition in English has not appeared. Papers appear regularly which continue to recommend improvements to the solution of the direct and inverse problems following an analysis of the problem similar to that of Fichot. These methods of computation are really now very well developed. One of the more recent of these was written by Bowring⁷ who also deals with the special problem of the nearly antipodean points.

In chapter 6 we define what we mean by *nearly antipodean points* and show how, from our solution of Clairaut's equation, we can compute the shortest distance between two nearly antipodean points which both lie on the Equator. We find a particularly useful expression in simple form which gives the half period, θ_p (the difference of longitude between two successive transits of the equator by the

geodesic), of a geodesic arc given the geocentric latitude, ϕ_v , of its vertex. This expression is

$$\theta_p = 2 \int_0^{\pi/2} \sqrt{\frac{a^2(1-e^2)^2 + \tan^2 \phi_v \sin^2 u}{a^2(1-e^2) + \tan^2 \phi_v \sin^2 u}} du$$

We also extend the application of the method to the ellipsoids with varying eccentricities and compute the periods of the geodesic curves on such surfaces. In chapter 7 we then formulate the general problem of computing the shortest distance between nearly antipodean points using the method of chapter 5 and the results of chapter 6 with respect to the values of the periods of geodesics with given vertices and show how this can be utilised in voyage planning at sea.

In chapter 8 we discuss the ways in which numerical methods can be applied to the computation of position from astronomical observations. There has been much interest shown in this aspect of navigation recently due, it seems, to the development of the leisure industry and the interest shown by people of strong technical background in scientific fields who are part of this new development. Although much of astronomical navigation may seem obsolete in light of modern improvements to navigation brought on by the introduction of global position fixing systems from satellites it is still an interesting theoretical pursuit. One of this author's contributions in this respect was contained in a paper published by the Journal of Navigation⁶ in which, in response to two other authors who made an over simplification of the problem of transferring a position circle, it was shown mathematically that the original position circle, when transferred, is no longer a circle but suffers a distortion. This may well account for problems which are encountered by an observer close to a pole of the Earth when taking astronomical observations - a large distortion of a transferred position circle takes place close to the pole when the

observer travels at an oblique angle to the meridian and this can be seen clearly from the plots of the transferred circles.

We extend this work on astronomical navigation into chapter 9 where we are particularly concerned with observations taken of a single body over a comparatively short period of time and from which, at a time chosen, we can compute the altitude and its rate of change. Given these two pieces of information we can then fix the observer's position. The classical situation in which this technique is applied is at the time of culmination when the maximum altitude of the observed body and the time that it occurs is computed from a sequence of observations taken over a period surrounding the time of the maximum. The technique is usually described graphically but was set in mathematical terms by Matti Ranta⁹. This idea was extended by this author¹⁰ to find the altitude and its rate of change for any fixed time that the body was visible to the observer by observing the body over a period of about forty minutes surrounding the fixed time for which the observed position is required. The method described requires the smoothing of the observed altitude data and the fitting of a least squares function approximation which we then differentiate to find the rate of change. We try out two smoothing techniques; we experiment with a least squares trigonometric approximation and with the least squares orthogonal polynomial approximations of Forsythe¹¹. The least squares trigonometric approximations give the better results on the data used.

In part 2 of this thesis we describe the *Direct Cubic Spline*. This is a variant of the familiar spline, used in a way that makes it particularly convenient for numerical integration. Its derivation will be found in chapter 10 with a more comprehensive error analysis than has appeared hitherto. We show that the direct cubic spline is a generalisation of Simpson's Rules which can be applied to compute the integral between any two ordinates and that the ordinates need not be evenly spaced. The method has been applied on many occasions in part one of the thesis, in particular, in chapter 3, to the computation of meridian distance on the surface of the spheroid, in

chapter 4 to the step by step solution of the great circle problem and, in chapter 5, to the step by step solution of the problem of navigating along the path of a geodesic curve on the surface of a spheroid. In part two there is also a discussion of some possible extensions of this idea to the higher order splines. The approach is described in detail in chapter 11. None of the algorithms for computing integrals using these direct higher order spline approximations produce very worthwhile results, however, with the exception of the quartic spline, used with a "borrowed" form for computing the "moments", which is efficient and which has been used by this author when applied to certain problems in ship stability.

For the time being, then, this is as far as we go. In particular (perhaps because of our own familiarity) we have found the Direct Cubic Spline Approximation to be a useful tool in our researches, particularly when applied to the solution of numerical problems in navigation and we can only hope that what is written here will also be of use or of interest to others.

This work began in collaboration with Professor J.E.Phythian before it was thought that I should enter a formal course for the Doctor of Philosophy degree. Professor Phythian then became my supervisor but, when he took up an academic post abroad, the direct supervision was taken over by Professor C.W.Clenshaw. I am greatly indebted to these two gentlemen for their meticulous reading, creative criticism and supportive suggestions. Thus encouraged, the work has, I believe, been undertaken and completed in a spirit of true philosophy.

PART ONE
APPLICATIONS
OF
NUMERICAL METHODS
IN
NAVIGATION

- 1 -

THE GEOMETRY
OF
THE EARTH

1.1 THE SHAPE OF THE EARTH.

In the study of Navigation the simplest approximation adopted for the shape of the Earth is a sphere. This is a reasonable approximation, if it is used consistently, and introduces no serious errors in most cases. There are, however, some surprises when shortest path curves are required, for example when we need to find the shortest path between two *nearly* antipodean points. (Antipodean points are points which lie at the opposite extremities of a diameter). For reasons such as this we require a better approximation to the shape of the Earth and the next simplest approximation is a regular oblate spheroid. Generally, a spheroid is simply defined as a "sphere-like" surface but the *regular* spheroid is generated by revolving an ellipse, whose eccentricity is small, around the minor axis. This axis coincides with the axis of revolution of the Earth. The poles of the Earth, whether it is considered to be a sphere or a spheroid, are found at the points where the axis of revolution cuts the surface. The poles are designated North (N) and South (S). An observer sited above the North pole and looking down upon the surface would view the Earth to be rotating anticlockwise. See Figure 1.1.

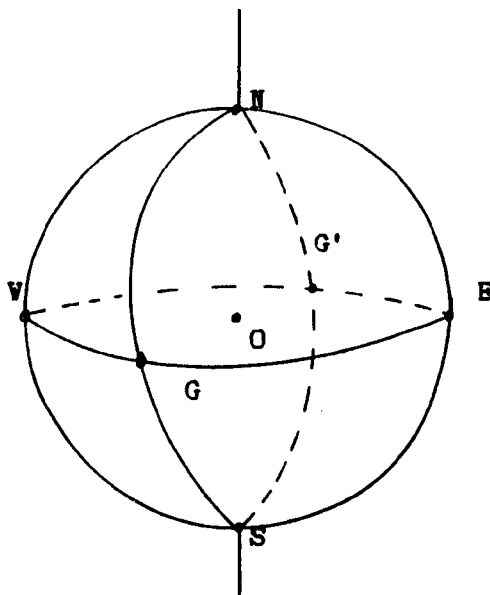


FIGURE 1.1 - THE EARTH AS A SPHERE.

In Figure 1.1 O is the centre of the Earth and the equator (WEGE') is the circle on the surface of the Earth which is the locus of all points which are equidistant from both poles. The semicircle NGS is a meridian. A meridian is the curve of intersection with the Earth's surface of a plane through the axis of revolution. The equator bisects the meridians.

The second approximation, the spheroid, is a better approximation to the shape of the Earth but it does make the computations in Navigation much more complicated. Indeed, for manual computations, it is often too difficult to assume anything else but that the Earth is a sphere.

There is some argument as to which single spheroid best suits the Earth's shape. Slightly different values are assigned to flattening, f , of the different spheroids and the agencies responsible to different governments for surveying sometimes each adopt a distinct spheroid which, they feel, best suits the shape of the Earth as they recognise it. The flattening, f is given by

$$f = \frac{a - b}{a}$$

where a is the length of the semi major axis of the selected generating ellipse and b is the length of its semi minor axis .

It is a fact that, in different geographical locations, the surface of the Earth is better approximated locally by one spheroid than another and so, in effect, the Earth is best approximated by a smooth union of different spheroids. The ocean basins, for instance, may each be fitted better by different spheroids although most books of nautical tables do stick to one spheroid which approximates the whole surface of the Earth. Globally, therefore, the Earth is almost invariably approximated by a *regular* spheroid, it is in the science of surveying that more local refinements are needed and practised. Some examples of the values assigned to the eccentricity of the meridional ellipse of the Terrestrial Spheroid as determined at

different times and places can be found in the book by G. Bomford¹⁴ or in the publication known familiarly as "Bowditch"¹² the full title of which is THE AMERICAN PRACTICAL NAVIGATOR.

On the surface of a spheroid the Equator and a meridian are defined in the same way as on the surface of a sphere. On the surface of a spheroid, however, whilst the Equator is still a circle, the meridians are ellipses. Instead of defining a spheroid by its flattening we more often (as we will here) define the spheroid by referring to the eccentricity of its meridional ellipses which are all identical in shape. The flattening, f , of the spheroid is related to the eccentricity, e , of its meridional ellipse by

$$f = 1 - \sqrt{1-e^2}$$

We use the term "spheroid" because, in the case of the Earth, the eccentricity of the meridional ellipse is small (≈ 0.08) and, hence, the surface is still "Sphere-like". In the case of other planets, such as Jupiter, where the eccentricity of the meridional ellipse is larger (≈ 0.3) then the surface is more aptly referred to as an Ellipsoid.

1.2 DEFINING POSITION ON THE SURFACE OF THE SPHERICAL EARTH MODEL.

On the surface of a sphere a circle defined by the intersection of the sphere with a plane through its centre is known as a GREAT CIRCLE and the unique great circle whose plane is perpendicular to the axis of revolution of the Earth is known as the equator. Any other circle on the surface of a sphere which is not a great circle is called a SMALL CIRCLE.

In the science of Navigation, where, in certain circumstances, the sphere is still often used as an approximation to the shape of the Earth, the position of a point is expressed in terms of its LATITUDE and LONGITUDE. The latitude of all points on the equator is zero and the latitude of any other point P , on the surface of the sphere is defined by the angle at the centre of the sphere subtended by the

arc of the meridian through P from the point P to the point where the meridian through P cuts the Equator. See Figure 1.2. In the Figure, N is the North pole, NO is the axis of revolution and E is the point where the meridian through P cuts the Equator. The angle ϕ is the LATITUDE of the point P.

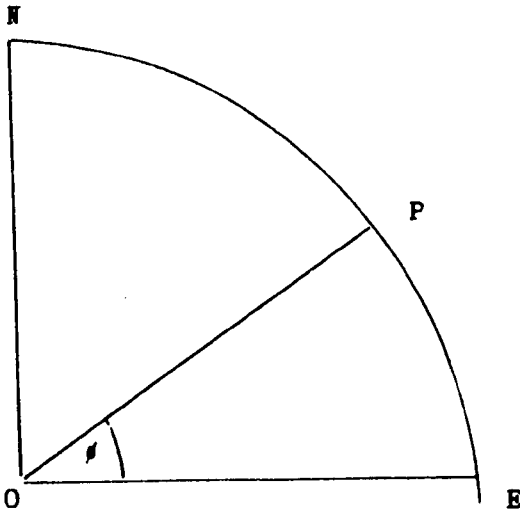


FIGURE 1.2

A particular meridian is selected for which the longitude is zero. On the surface of the Earth this meridian is the Greenwich Meridian - the meridian which passes through the Greenwich Observatory. The longitude, θ , then of any other point P on the surface of the sphere representing the Earth is the angle between the planes through the axis of revolution of the Earth one of which contains the Greenwich meridian and the other the meridian through the point P. See Figure 1.3.

A circle on the surface of the sphere whose plane is parallel to the equatorial plane is known as a PARALLEL OF LATITUDE so called because all points on this circle are in the same latitude. Except for the equator, parallels of latitude are SMALL circles. In Figure 1.3 the arc PG'P' is a small circle.

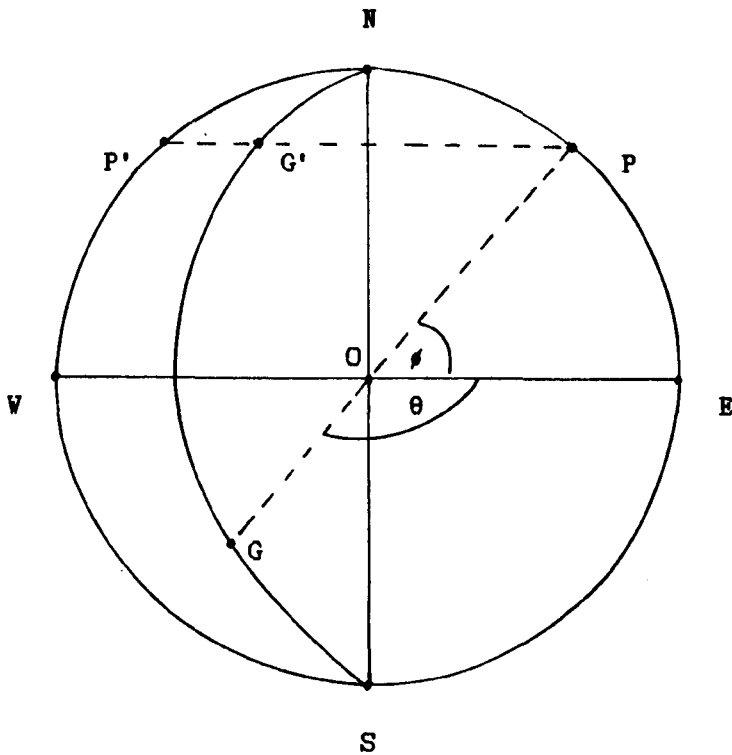


FIGURE 1.3

In navigation and in geography latitude and longitude are expressed in degrees and minutes - latitude as north or south of the equator and longitude as east or west of the Greenwich meridian. The ranges are

$$90^{\circ}\text{S} < \phi < 90^{\circ}\text{N}$$

$$180^{\circ}\text{W} < \theta < 180^{\circ}\text{E}$$

To give a mathematical treatment to the methods in Navigation we should express the angles in radians and, preferably, in the ranges

$$-\frac{1}{2}\pi < \phi < \frac{1}{2}\pi \quad (\text{North Positive})$$

and

$$0 < \theta < 2\pi \quad (\text{East Positive})$$

Distances on the surface of a sphere or a spheroid are expressed in a natural way in units of the length of one minute of arc of the Equator. The accepted value for this unit is 6087.2 feet or 1852 metres. This unit is known as the Geographical Mile and we will use it throughout in our treatment of navigational methods. There is another "mile" which is sometimes referred to as the "Nautical

Mile". This mile is a standard mile of 6080 feet and is the approximate length of one minute of arc of the meridian in the vicinity of Southern England. Some confusion has arisen in practice at sea because distances computed from engine revolutions or rotator logs have not been consistent. The errors have, however, been small and negligible in most cases.

1.3 DEFINING POSITION ON THE SURFACE OF THE SPHEROIDAL EARTH MODEL.

Let the surface of the Earth be modelled by a regular spheroid whose meridians are a family of ellipses of the same eccentricity and which share a common minor axis. At a point P on this surface the angle subtended at the centre of the Earth by the arc of the meridian through P from the equator to P is the GEOCENTRIC LATITUDE. At the same point P the angle at which the normal to the meridian at P cuts the equatorial plane is the GEODETIC LATITUDE. See Figure 1.4.

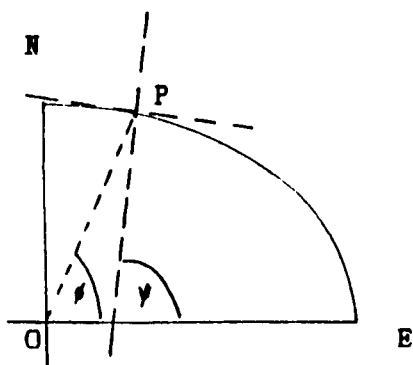


FIGURE 1.4

In the Figure 1.4 :

ϕ is the GEOCENTRIC latitude
 ψ is the GEODETIC latitude and
 ϕ and ψ are connected by

$$\tan \phi = (1 - e^2) \tan \psi .$$

Where e is the eccentricity of the meridional ellipse.

The longitude of a point P is expressed in the same way as on the surface of a sphere.

On the surface of a *regular* spheroid the Geodetic Latitude is also the Astronomical (Geographical) Latitude. The Geodetic Latitude is the term used by the surveyor. In local conditions such as mountainous regions, where abnormal gravitational effects upon a plumb line may be observed the normal to the surface is considered to lie along this plumb line. On the side of a mountain ,therefore,

a geographer, who is using a *global* approximation to the surface of the Earth, may find a slightly different value for the latitude of a particular point than the surveyor who is considering a *local* approximation to the same surface. For a complete discussion of this one should read the books by Cotter'³ and Bonford'⁴. However, in Navigation we confine ourselves to computations on the surface of a *regular* spheroid which is a *global* approximation to the surface of the Earth.

-2-

**NAVIGATING ALONG
A LOXODROMIC CURVE ON THE
SURFACE OF A SPHERE**

2.1 INTRODUCTION.

The science of navigation, although it is a numerical science, is presented in the reference texts in a manner which tends to be empirical rather than analytical. There is little evidence of mathematical rigour and little use is made of the techniques of the calculus. It is the purpose in this chapter, therefore, to analyse the methods of computation used in navigation by relating them to the results obtained by considering the differential geometry of the sphere and thereby to prove the formulae that are in common use. Some of these results and proofs can be found scattered through the pages of books on differential geometry but not always in a concise or complete form in books on navigation.

2.2 THE LOXODROMIC CURVE.

A LOXODROME, known more familiarly to seafaring Navigators as a RHUMB LINE, is a curve on a surface of revolution which cuts all the meridians at the same angle.

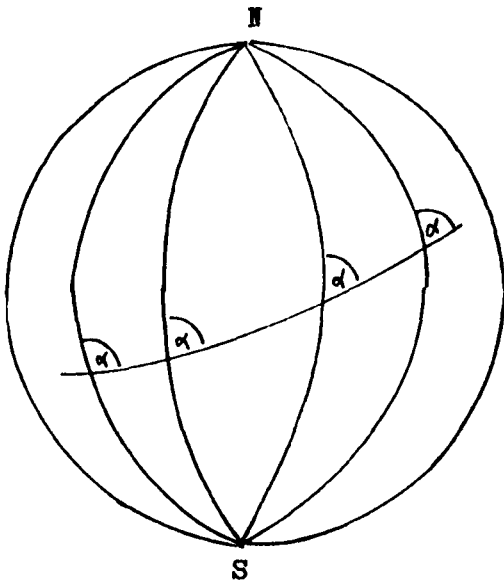


FIGURE 2.1
A LOXODROME ON THE
SURFACE OF A SPHERE.

Figure 2.1 shows a loxodrome on the surface of a sphere which cuts the meridians at an angle α . To the seafaring Navigator it is a line of constant course and there are obvious reasons why it is desirable to cross the Oceans on such a line rather than to take the shortest

route which would be the first choice in most circumstances. Weather conditions are one such reason why a navigator might choose not to follow the shortest route to his destination and the flow of the Ocean currents are another. If we consider the particular example of the North Atlantic Ocean, for instance; from the Gulf of Mexico to the Western Approaches of the British Isles we have the "Gulf Stream" which is a flow of warm water between the tropics and cooler northern climes. This current of water closely follows the great circle path so that any traffic from the UK or Northern Europe headed out from the English Channel to North American or the Caribbean Islands will oppose this current for a large part of the journey. It turns out that, on the direct route to the Caribbean Islands, in order to avoid the Gulf Stream, most ships will follow one constant course once they have cleared the English Channel even though this path will be longer in distance. In winter this has the added advantage that the ship will make quicker progress to the south and thus clear the stormy regions earlier. Ocean Routing is a sub branch of the science of Navigation which studies these matters in detail.

The loxodrome is an endless curve of finite length which, for every value of α : $0 < \alpha < \pi/2$, spirals to end limit point at the pole. When $\alpha = \pi/4$, for instance, the length of the loxodrome from any initial point on the Equator to the pole is, for the spherical model,

$$\frac{a\pi\sqrt{2}}{2}$$

where a is the radius of the sphere. (See Lipschutz's).

2.3 STEREOGRAPHIC PROJECTION OF A LOXODROMIC CURVE.

In general, under the Stereographic Projection, a point P on the surface of a sphere is projected onto a tangent plane. The source of the projection is the point which is the antipode of the point of tangency. (Two points on the surface of a sphere are antipodes to each other when they lie at the opposite ends of the same diameter).

The stereographic projection is a conformal mapping and, in the stereographic projection of the Terrestrial Sphere onto a polar plane, (a plane tangent to the pole) where the source of the projection is the opposite pole, the image of a loxodrome is an equiangular spiral. This is because, under the projection, angles are preserved and this is the one single and common property by which both curves can be defined. Figure 2.1 shows the projection from the South Pole, under the stereographic projection, of a point P on the surface of the Terrestrial Sphere to the point P' in the plane tangent to the North Pole.

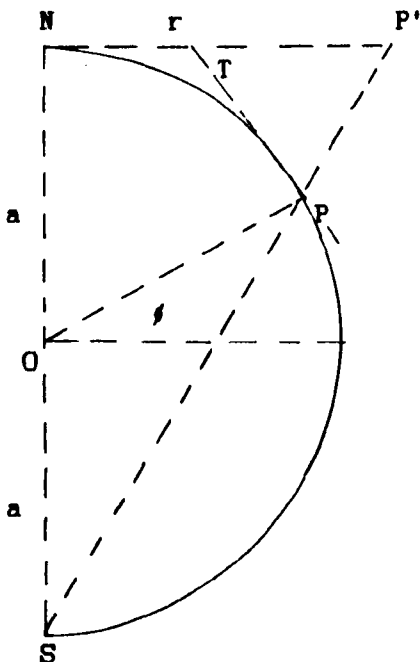


FIGURE 2.2
STEREOGRAPHIC PROJECTION.

The stereographic projection is often used to map the Earth's polar regions. The meridians on the surface of the sphere are projected into the radial lines (in the polar plane) which are the intersections of the polar plane with the planes containing the meridians. If the initial line ($\theta=0$) of polar coordinates in the polar plane corresponds with the Greenwich Meridian then the longitude on the surface of the Terrestrial Sphere corresponds to the angular coordinate in the polar plane. Simple considerations of

geometry in Figure 2.2 show that the radius, r , in the polar plane of the point P' from the pole, N , is given by

$$r = 2a \tan(\pi/4 - \phi/2) \quad \dots\dots (2.1)$$

where a is the radius of the Terrestrial Sphere and ϕ is the latitude of the point P .

The stereographic projection of the surface of a sphere onto a tangent plane is a conformal mapping which can be defined analytically (Struick¹⁶). We now show geometrically how angles are preserved under the stereographic projection of the surface of a sphere onto the polar plane. We can use the essential geometrical features to show that, in a similar manner, we can project the surface of a spheroid onto the polar plane; this is also a conformal mapping.

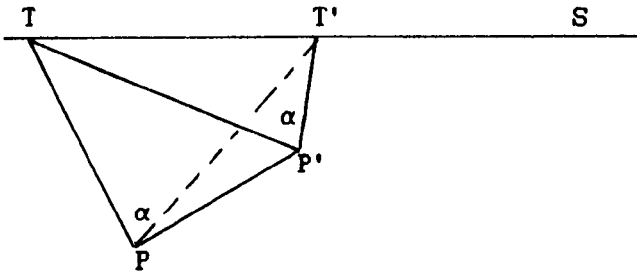


FIGURE 2.3

The tangent plane to the sphere at point P cuts the polar plane in a line through the point T in the polar plane. This line (TS in Figure 2.3) is at right angles to the plane of the meridian through P on the sphere and also at right angles to the projection of this meridian in the polar plane. From Figure 2.2 above we can deduce that the triangle PTP' , which lies in the plane of the meridian through P , is isosceles so that $PT = P'T$. This is a consequence of the fact that the angles TPP' and $TP'P$ are both equal to $\pi/4 + \phi/2$.

The tangent at P to the loxodrome lies in the tangent plane through P . Let this tangent cut the polar plane in T' . Clearly, the point T' lies on the line TS . See Figure 2.3.

Let the angle at P in the triangle TPT' (Figure 2.3) be α . If we consider the triangles PTT' and P'TT' then they are congruent since both are right angled at T, the side TT' is common and $PT = P'T'$ as we have seen. Hence the angle at P' in triangle P'TT' is equal to α also. This is a general result because the tangent PT' could be the tangent to any curve on the surface of the sphere through the point P.

This geometric proof can, of course, be applied just as well for a point P in the Southern Hemisphere.

Conversely, it can be deduced from the foregoing that the above projection of the surface of the sphere is a conformal mapping if, and only if, the triangle PTP' in Figure 2.2(1) is isosceles. In the projection of the surface of the spheroid onto its polar plane we can therefore make the mapping conformal if the equivalent of the triangle PTP' is also isosceles. This mapping is described in Chapter 3.

Under the stereographic projection of the surface of the sphere onto the polar plane, the equation of the image of the loxodrome in polar coordinates is

$$\boxed{r = r_0 \exp[-(\theta - \theta_0) \cot \alpha]} \quad \dots\dots (2.2)$$

where (r_0, θ_0) are the coordinates of the projection of some initial point P_0 on the loxodrome and α is the angle at which the loxodrome cuts the meridians on the surface of the sphere.

To prove this let ΔE be an element of the image of the loxodrome in the polar plane. Since angles are preserved under the stereographic projection then the angle, α , at which the loxodrome cuts the meridians must be the same angle at which the image of the loxodrome cuts the radial lines in the polar plane.

Figure 2.5 shows an element of the image of the loxodrome in the polar plane. Since the loxodrome is directed towards the pole then so is its image and, hence, corresponding to the element Δs of the

loxodrome, we have $\Delta\epsilon$ (=PR) (triangle PQR - Figure 2.5). Also in Figure 2.5 we have $r \cdot \Delta\theta$ (=PQ) along the arc of the circle

$$r = \text{constant}$$

and $-\Delta r$ (=QR) along the radial line $\theta = \text{constant}$.

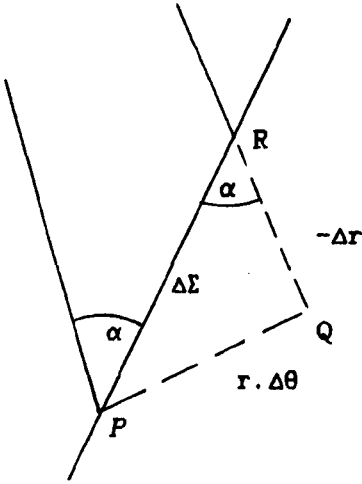


FIGURE 2.5

AN ELEMENT OF THE IMAGE OF A LOXODROME.

From the triangle PQR (Figure 2.5) we see that

$$\tan \alpha = - \frac{r \cdot \Delta\theta}{\Delta r}$$

And, in the limit as $\Delta\epsilon \rightarrow 0$ we have

$$\frac{dr}{r} = - \cot \alpha \, d\theta$$

Integrating from an initial point P_0 where $r=r_0$ and $\theta=\theta_0$ we find

$$\ln r - \ln r_0 = - \cot \alpha (\theta - \theta_0)$$

which gives $r = r_0 \exp[-(\theta - \theta_0) \cot \alpha]$ as required.

2.4 NAVIGATING ALONG THE PATH OF A LOXODROME ON THE SURFACE OF A SPHERE.

Let us consider that the Earth is a sphere on which the position of a point P is determined by its latitude, ϕ , and its longitude, θ , where ϕ and θ are measured in radians. If a is the radius of the Earth then the distance along the meridian from the Equator to P is $a\phi$ and the distance along the parallel of latitude from the Greenwich Meridian to P is $a\theta \cos \phi$. See Figure 2.6 where N is the North Pole, NG is an arc of the Greenwich Meridian, GE is an arc of the Equator and PG' is an arc of the parallel of latitude through P. PM, the perpendicular from P to the axis of revolution is the radius of the parallel of latitude and $PM = a \cos \phi$.

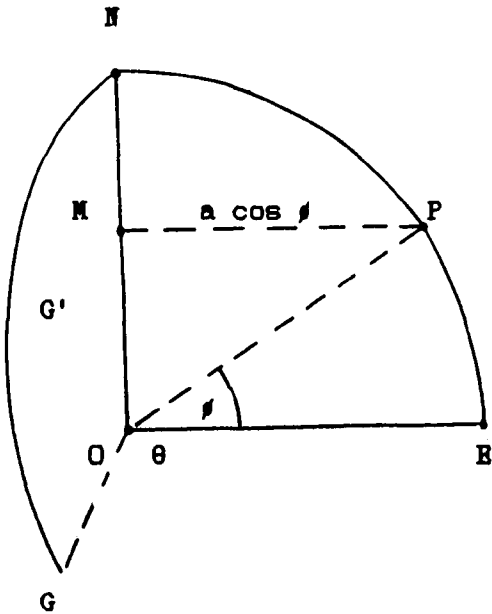


FIGURE 2.6

Let a be the radius of the Earth. The distance between two points P_1 and P_2 on the same meridian whose latitudes are ϕ_1 and ϕ_2 , respectively, and which is equal to $a(\phi_2 - \phi_1)$ ($\phi_2 > \phi_1$) is known as the DIFFERENCE OF LATITUDE. The difference of latitude is, in fact, the angle $\phi_2 - \phi_1$ expressed in minutes.

Similarly, the distance along the equator between the feet of the meridians through two points P_1 and P_2 whose longitudes are θ_1 and θ_2

θ_2 , respectively, and which is equal to $a(\theta_2 - \theta_1)$ ($\theta_2 > \theta_1$) is known as the DIFFERENCE OF LONGITUDE. The difference of longitude is the angle $\theta_2 - \theta_1$ expressed in minutes.

If we now consider a small element of length Δs of a loxodrome through P which cuts the meridians at an angle α then $\Delta \phi$ is the small increment in the latitude and $\Delta \theta$ is the small increment in the longitude which results from moving the small distance Δs . Corresponding to this element Δs we have a triangle PQR (Figure 2.7) in which PQ ($= \Delta s$) is along the arc of the loxodrome, QR ($= a \Delta \phi$) is along the meridian through Q and R and PR ($= a \cos \phi \Delta \theta$) is along the parallel of latitude through P. The angle at P is equal to α .

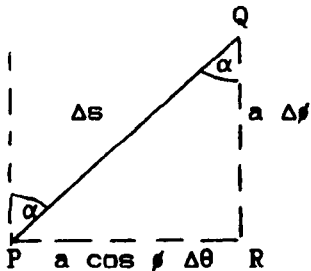


FIGURE 2.7
A SMALL ELEMENT Δs OF THE LOXODROME.

In the limit as $\Delta s \rightarrow 0$ the triangle PQR becomes a right angled triangle in which

$$PQ = ds \quad PR = a \cos \phi d\theta \quad \text{and} \quad QR = a d\phi .$$

It is from this triangle that we deduce the formulae that are relevant to an observer travelling along the path of a loxodrome.

2.5 THE MERCATOR PROJECTION.

The Mercator Projection is so called because it was introduced by Gerhard Mercator early in the 17th century. According to Bowditch¹⁴, Mercator constructed his chart purely to satisfy the needs of the sea-going navigator who needed a chart on which his line of constant course (his "Rhumb Line") would be a straight line.

On the surface of a sphere, along a parallel where the latitude is ϕ , the ratio of the length of one minute of arc of the meridian to the length of one minute of longitude along the parallel is

$$\sec \phi : 1 .$$

If, then, the surface of the sphere is mapped onto another surface so that this ratio is preserved and, at the same time, the images of the meridians and parallels of latitude are also orthogonal then the angle of the rhumb line (loxodrome) will also be preserved.

Mercator achieved this by mapping the surface of the sphere onto the surface of an infinite cylinder, coaxial with the sphere and of the same radius. Under this mapping, the meridians on the surface of the sphere are mapped into the meridians on the surface of the cylinder (which are straight parallel lines) and the parallels of latitude on the sphere are mapped into circles of equal radius on the surface of the cylinder. These circles are contained in planes perpendicular to the meridians. The length of arc of one minute of longitude is then constant on the surface of the cylinder. When the latitude is ϕ , therefore, the length of image on the cylinder of one minute of arc of the meridian on the sphere in units of the length of one minute of longitude on the cylinder must be equal to $\sec \phi$.

These image elements on the surface of the cylinder of the minutes of arc of the meridian on the surface of the sphere became known as "MERIDIONAL PARTS". The first table of meridional parts was compiled by Edward Certaine in 1599 and it was from this table of meridional parts that Mercator constructed his chart.

Considering that Certaine compiled his table in the days before Leibniz or Newton we must assume that the entries in his table were computed in a manner similar to the following:

When the latitude is ϕ_n and n is the number of minutes of arc, then, if $M(\phi_n)$ is the sum of the meridional parts, we find

$$M(\phi_n) = \sum_{i=1}^n \sec \phi_i$$

The mapping that Mercator would then have used to construct his chart would have been defined by

$$\phi_n \rightarrow \sum_{i=1}^n \sec \phi_i \qquad \theta \rightarrow \theta$$

In modern terms, if we subdivide the meridian of the sphere by a finer mesh so that $\phi_i - \phi_{i-1} = h$ and let $h \rightarrow 0$ then we find:

$$M(\phi) = \int_0^{\phi} \sec u \, du \qquad \dots\dots (2.3)$$

and the Mercator projection of the surface of the sphere onto the surface of the cylinder is given by

$$\phi \rightarrow \int_0^{\phi} \sec u \, du \qquad \theta \rightarrow \theta$$

so that, on the chart, (which is the cylinder "unrolled") the meridians are mapped into parallel lines and a loxodrome is mapped into a straight line which cuts the mappings of the meridians at the same constant angle as the loxodrome itself cuts the meridians on the surface of the sphere. Figure 2.8 is an illustration of a Mercator chart.

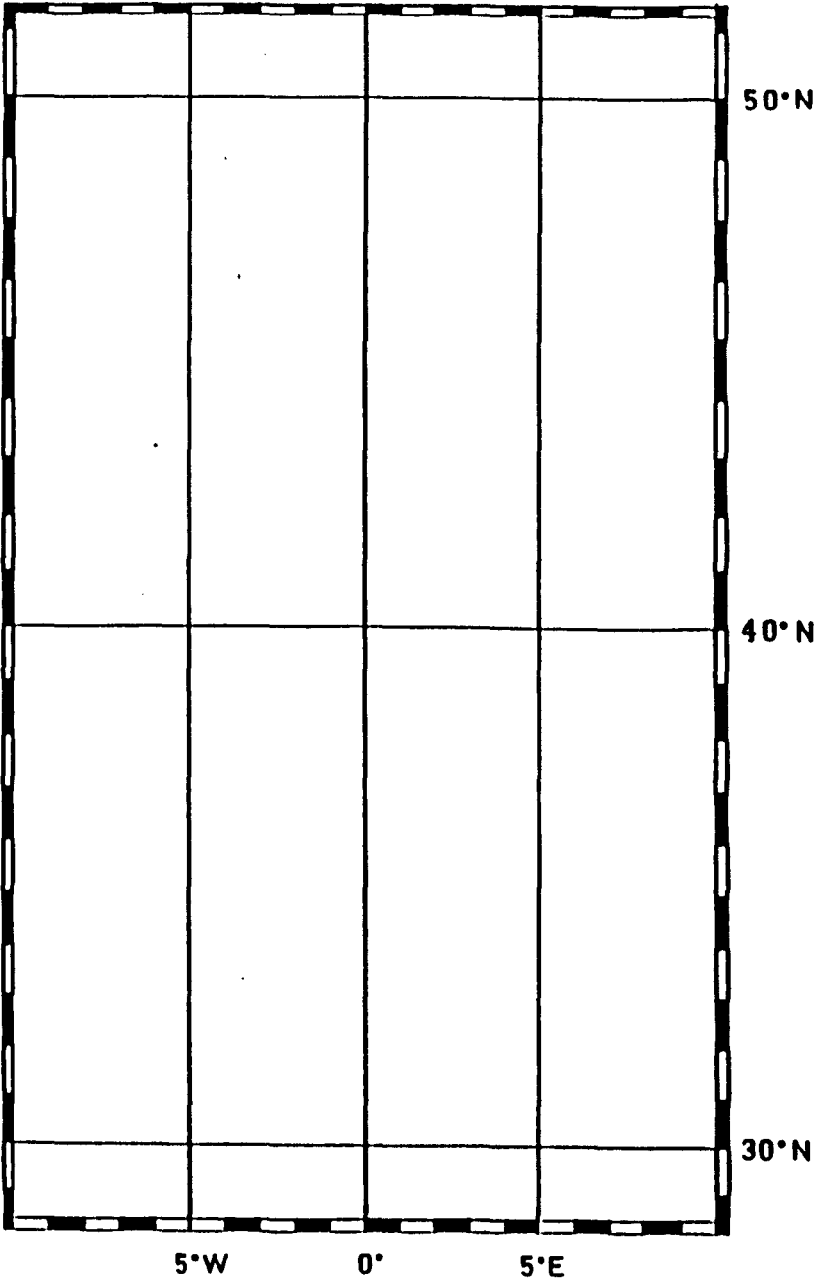


FIGURE 2.8 - A MERCATOR CHART

The Mercator projection is sometimes referred to as a "cylindrical projection" which can be misleading, for, although the Mercator projection is the mapping of the surface of a sphere onto the surface of the infinite coaxial cylinder of the same radius, we cannot present this mapping visually. The Cylindrical Projection is the particular projection which is the mapping of the surface of a sphere onto the surface of the same infinite coaxial cylinder as the Mercator projection but defined by the mapping

$$\phi \rightarrow a \tan \phi \qquad \theta \rightarrow \theta$$

2.6 METHODS OF COMPUTATION IN MERCATOR SAILING.

The navigational formula which are derived for calculating the components of the right angled triangles defined by the intersection of a loxodrome with the meridians and the parallels of latitude are collectively known as the formulae for MERCATOR SAILING.

From the limiting form of triangle PQR in Figure 2.7, we find

$$\frac{a \cos \phi \, d\theta}{a \, d\phi} = \tan \alpha$$

Separating the variables gives

$$a \, d\theta = (\tan \alpha) a \sec \phi \, d\phi$$

where it will be noticed that we have not cancelled a from both sides of the equation.

If the observer has travelled along the path of the loxodrome from the point P_0 (longitude θ_0 , latitude ϕ_0) to the point P_1 (longitude θ_1 , latitude ϕ_1) we integrate to find

$a(\theta_1 - \theta_0) = \left(\int_{\phi_0}^{\phi_1} a \sec \phi \, d\phi \right) \tan \alpha$ (2.4)
---	-------------

This equation - (2.4) - is fundamental for a ship sailing along the arc of a loxodrome.

The expression $a(\theta_1 - \theta_0)$ is the DIFFERENCE OF LONGITUDE between P_0 and P_1 and is conventionally measured in minutes. The abbreviation for the difference of longitude is D'LONG.

For a point P, where the latitude is ϕ , we also define the function $M(\phi)$:

$$M(\phi) = a \int_0^{\phi} \sec u \, du = a \ln[\tan(\phi/2 + \pi/4)] \dots\dots (2.5)$$

The function, $M(\phi)$, is known as the MERIDIONAL PARTS of the point P being a function of the latitude, ϕ . This is the distance from the Equator to the image of the point P in Mercator's Projection. Meridional Parts are expressed in units of the length of one minute of arc of the Equator. In Nautical Tables, Meridional Parts are tabulated for each one minute of arc of the meridian and, even until the middle of the twentieth century, this tabulation was for the sphere. On a Mercator chart at the parallel where the latitude is ϕ the ratio of the scale of longitude to the scale of latitude is locally 1 : sec ϕ . The integral

$$a \int_{\phi_0}^{\phi_1} \sec \phi \, d\phi$$

is the DIFFERENCE OF MERIDIONAL PARTS and, in navigational notation, is abbreviated to D'MP. In navigational notation, therefore, equation (2.4) is written

$$D'LONG = D'MP \times \tan(\text{Course}) \dots\dots (2.6)$$

To compute the distance, s , from P_0 to P_1 along the arc of the loxodrome we use triangle PQR to find

$$ds = a \sec \alpha \, d\phi$$

then we have, on integration

$$s = a(\phi_1 - \phi_0) \sec \alpha \dots\dots (2.7)$$

As we saw, the quantity $a(\phi_1 - \phi_0)$ is the difference of latitude measured in minutes and, in navigational notation is abbreviated to D'LAT so that equation (2.7) is written in the form

$$\boxed{\text{DISTANCE} = \text{D'LAT} \times \text{SEC}(\text{Course})}$$

Distance cannot, in general, be measured directly between two points on a Mercator chart but, *locally*, a good approximation can be obtained. This can be seen by noting that, using the mean value theorem, D'MP may be written as

$$a \sec \phi_m (\phi_1 - \phi_0)$$

where ϕ_m is some value of ϕ such that $\phi_0 < \phi_m < \phi_1$.

When $a(\phi_1 - \phi_0) = 1$ the length of the image of one minute of latitude is seen to be $\sec \phi_m$. We find, however, that, in the limit, as $\phi_1 \rightarrow \phi_0$ the image length of one geographical mile along the meridian is equal to $\sec \phi_0$. At the same time the image length of one mile along the parallel of latitude $\phi = \phi_0$ is also equal to $\sec \phi_0$.

2.6 EXAMPLES OF CALCULATIONS USING MERCATOR SAILING.

The navigator uses the method of Mercator Sailing in two ways - to calculate the final position, P_1 , after sailing a given distance along the arc of a loxodrome from a point P_0 or to find the course and distance made good between two observed positions, P_0 and P_1 .

The method is reduced to the simple application of plane trigonometry to the right angled triangles formed by the intersection of the loxodrome with the meridian through P_0 and parallel of latitude through P_1 , shown in Figures 2.9(i) and 2.9(ii). Figure 2.9 (ii) is a triangle on the surface of the sphere and Figure 2.9 (i) is a triangle on the surface of the coaxial cylinder. An example of each of these calculations follows.

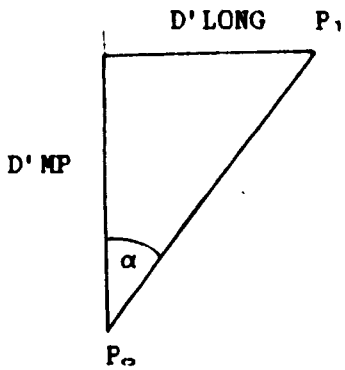


FIGURE 2.9(1)

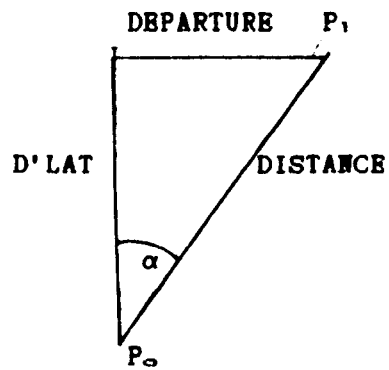


FIGURE 2.9(11)

EXAMPLE 1. At noon on one day the observer's position was $31^{\circ}45'N$ $32^{\circ}35'E$ and on the next day at noon the position was $36^{\circ}30'N$ $40^{\circ}20'E$. Find the course and distance made good in the twenty four hour period and the average speed for the day.

From nautical tables we find the values of the meridional parts function $M(\phi)$:

$M(36^{\circ}30'N)$	2355.19
$M(31^{\circ}45'N)$	2010.72
D' MP	344.47

Final Longitude	$40^{\circ}20'$
Initial Longitude	$32^{\circ}35'$
D' Long	$7^{\circ}45'$ (= 465 minutes of arc)

From equation (2.4) or (2.6) and Figure 2.7(1) :

$$\tan \alpha = \frac{D' LONG}{D' MP} = \frac{465}{344.47} = 1.3498998$$

$$\alpha = 53^{\circ}28.1'$$

Final Latitude	$36^{\circ}30.00'$
Initial Latitude	$31^{\circ}45.00'$
D' LAT	$4^{\circ}45.00'$ (= 285 minutes of arc)

From equation (2.7)

$$s = a(\phi_1 - \phi_0) \sec \alpha = 478.79$$

COURSE = 053°28.1' DISTANCE = 478.79 SPEED = 19.95 knots.

EXAMPLE 2. An observer in position 30°00'N 30°00'E travels a distance of 500 nautical miles along a loxodrome on a course of 045°. Find the final position.

From equation (2.7) we find the D'LAT [= $a(\phi_1 - \phi_0)$]

$$D'LAT = s \cos \alpha = 500 \cdot \cos 45^\circ = 353.55 \text{ minutes of arc}$$

On a course of 045° the observer increases latitude and so the Difference of Latitude is Northward

Initial Latitude	30°00'
D'Lat	+ 5°53.55'
Final Latitude	35°53.55'

From Nautical Tables we find

	M(35°53.55')	2309.47
	M(30°00.00')	1888.38
so that	D'MP	421.09

From equation (2.4) or (2.6) we now find

$$D'LONG = D'MP \tan 45^\circ$$

hence $D'LONG = 421.09$ minutes of arc.

On a course of 045° the observer increases the longitude to the East so that we add the difference of longitude

Initial Longitude	$30^\circ 00.00'$
D'Long	$7^\circ 01.09'$
Final Longitude	$37^\circ 01.09'$

FINAL POSITION : $35^\circ 53.55' N$ $37^\circ 01.09' E$

There is, of course, no need to perform these computations this way in this day and age since we have computing devices which make redundant the need to keep tables. The integral in equation (2.3) is a standard integral and does integrate exactly to a form which can be computed easily. However, as another consequence of the availability of the computer we no longer use the spherical approximations and concentrate more on the performing the same computations on the surface of the Spheroidal Earth.

It is interesting to note that the set of *Horie's Nautical Tables*¹⁷ that were used to complete the above computation are a set published in 1948 and give tables of Meridional parts for both the sphere and the spheroid and that the table for Meridional Parts for the sphere *is given preference* where one would have thought the more accurate version for the spheroid would have had pride of place.

In the first set of *Horie's Nautical Tables*¹⁸ that the author acquired as an officer cadet in 1954 (and which were published in 1954) there is only one table of Meridional Parts - that for the Terrestrial Spheroid.

*Inman's Nautical Tables*¹⁹ of 1952 give a table of Meridional Parts for the sphere with a footnote explaining that the Meridional Parts for the spheroid can be obtained by entering the table with the "reduced" latitude.

2.6 MIDDLE LATITUDE SAILING.

On the surface of a sphere Middle Latitude Sailing is a method of computing course and distance along the arc of a loxodrome without involving Mercator's Projection. On a surface of a spheroid, however, Middle Latitude sailing cannot be used independently although the method is still documented.

Consider, once again, an observer travelling along a loxodrome on course α on the surface of a sphere. We have, at a point P on the loxodrome where the latitude is ϕ and the longitude is θ , the differential triangle PQR (Figure 2.7) where a is the equatorial radius of the Earth and s is the arc length along the loxodrome. The angles are measured in radians and the distances in the unit of one minute of arc of the Equator.

In the triangle PQR the side PQ is along the tangent to the loxodrome at P, the side QR is along the tangent to the meridian and the departure, PR (= $d\lambda$ say) which is along the tangent to the parallel of latitude, is given by

$$d\lambda = a \cos \phi \, d\theta .$$

Thus when the observer travels along the loxodrome from a point P_0 (latitude ϕ_0 , longitude θ_0) to the point P_n (latitude ϕ_n , longitude θ_n) then the departure made good, λ , is given by

$$\lambda = a \int_{\theta_0}^{\theta_n} \cos \phi \, d\theta \quad \dots\dots (2.7)$$

Departure is the length of the side opposite to the course angle, α , in Figure 2.9(ii).

Now ϕ is a function of θ and, indeed, they are related by the equation

$$\boxed{\tan(\pi/4 - \phi/2) = \tan(\pi/4 - \phi_0/2) e^{-\sec \alpha (\theta - \theta_0)}} \quad \dots (2.8)$$

This equation, (2.8) can be deduced from equations (2.1) and (2.2). It is difficult from this relationship to express ϕ explicitly in terms of θ and perform the integration in equation (2.7) so we proceed without doing so. Since $\cos \phi$ is positive and also decreasing in the interval $[0, \pi/2]$, we can apply the Second Mean Value Theorem for Integrals through which we find that for some value χ of ϕ such that $\phi_0 \leq \chi \leq \phi_n$ we have

$$\lambda = a \cos \chi \int_{\theta_0}^{\theta_n} d\theta$$

or $\lambda = a(\theta_n - \theta_0) \cos \chi \quad \dots \dots (2.9)$

This is saying that

$$\boxed{\text{DEPARTURE} = D' \text{ LONG} \times \text{MEAN OF COS LAT}}$$

whereas, in common parlance, this formula is often quoted as

$$\text{DEPARTURE} = D' \text{ LONG} \times \text{COS OF MEAN LAT}$$

and even applied that way.

Strictly speaking of course, $\cos \chi$ is not the mean of $\cos \phi$ when $\phi_0 \leq \chi \leq \phi_n$ but, rather, some value of $\cos \phi$ which satisfies equation (2.9). In the case of a sphere, however, if we use the simple form of the mean value we find that

$$\cos \chi = \frac{\int_{\phi_0}^{\phi_n} \cos \phi \, d\phi}{\phi_n - \phi_0}$$

and that, over short distances, the error in so doing is not serious. Hence, from equation (2.9) we find

$$\lambda = \frac{a(\theta_n - \theta_0)}{(\phi_n - \phi_0)} \int_{\phi_0}^{\phi_n} \cos \phi \, d\phi$$

or

$$\lambda = \frac{a(\theta_n - \theta_o)}{(\phi_n - \phi_o)} (\sin \phi_n - \sin \phi_o) \dots\dots (2.10)$$

and this is a fairly simple formula from which to compute departure on the surface of a sphere.

As an example let us compute the course and distance made good between the points P_o (31°45'N 32°35'E) and P_n (36°30'N 40°20'E) making use of the above formula (4) and also by the method of Mercator Sailing to compare the results.

From the Middle Latitude Sailing method we find

$$\text{Course Made Good} = 053^{\circ}28.6' \quad \text{Distance} = 478.86$$

and from Mercator Sailing we find

$$\text{Course Made Good} = 053^{\circ}28.1' \quad \text{Distance} = 478.79$$

The value used for the radius of the Earth in both cases was a = 3437.7468 geographical miles.

Over longer distances such as the passage across an ocean, equation (2.10) is not accurate enough and we must then use some numerical method to evaluate the integral in equation (2.7) expressing ϕ in terms of θ by equation (2.8).

-3-

NAVIGATING ALONG
A LOXODROMIC CURVE ON THE
SURFACE OF A SPHEROID

3.1 INTRODUCTION.

The shape of the Earth is better approximated by a regular spheroid rather than a sphere by which we mean, in this case, a spheroid which is a surface of revolution generated by revolving an ellipse about its minor axis. On such a surface the meridians are ellipses. The eccentricity of the meridians is small (≈ 0.08) and there is no general agreement as to which exact value is the "best" value to effect a global fit of this surface to the surface of the Earth. For a mathematical treatment of the methods of navigating along a loxodromic curve on the surface of a spheroid it does not really matter which precise value we use. We will use the letter "e" to denote the eccentricity and, where numerical results are required, we will use the value $e = 0.08227$ which is the value (attributed to Clarke) computed in 1866 from survey data gathered in North and Central America and Greenland. This particular value was chosen because this author became familiar with it through using the "American Practical Navigator"¹².

Despite the fact that the knowledge that the Earth is ellipsoidal in shape and not spherical has long been well known it has not even yet been fully embraced in the science of Navigation. In this chapter we will review the work done in recent years.

3.2 COMPUTING THE LENGTH OF A MERIDIAN ON THE SURFACE OF A SPHEROID.

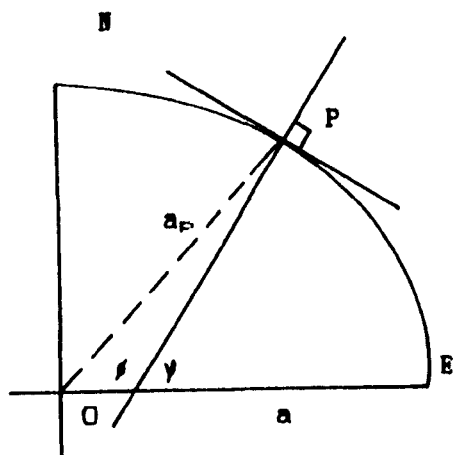


FIGURE 3.1
SECTION OF A SPHEROID IN THE PLANE
OF A MERIDIAN.

Let us suppose that the Earth is a regular spheroid and that the eccentricity of the meridional ellipse is e . Figure 3.1 shows part of the section of the spheroid in the plane of a meridian. In Figure 3.1 the centre of the spheroid is at O and N is the North Pole. ON ($= a$) is the Equatorial radius. At the point P the geocentric latitude is ϕ and the geodetic latitude is ψ . ψ and ϕ are connected by

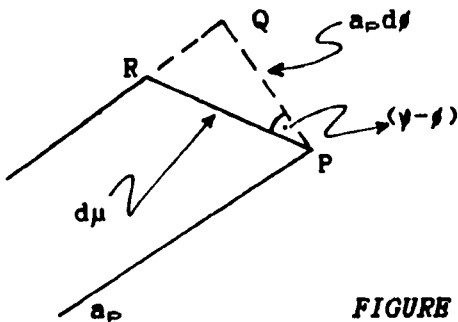
$$(1 - e^2)\tan \psi = \tan \phi \quad \dots\dots (3.1)$$

The radius, a_p , ($= OP$ in Figure 3.1) of the spheroid at point P is given by

$$a_p = a \sqrt{\frac{1 - e^2}{1 - e^2 \cos^2 \phi}} \quad \dots\dots (3.2)$$

where a is the equatorial radius of the spheroid. This can be seen from elementary consideration of the geometry of the ellipse.

Let $\Delta\mu$ ($= PR$ in Figure 3.2) be an small element of the meridian at P . In the triangle PQR of Figure 3.2 PQ is the arc of a circle of radius a_p centred at the centre of the spheroid and we have $PQ = a_p \Delta\phi$. The angle QPR is $(\psi - \phi)$ which is the angle between the normal at P and the radius of the spheroid at P . In the limit as $\Delta\mu \rightarrow 0$ the triangle PQR is a right angled triangle with the right angle at Q and we find



$$\begin{aligned} PR &= PQ \sec (\psi - \phi) \\ &= a_p \sec (\psi - \phi) d\phi \end{aligned}$$

FIGURE 3.2

The length of the meridian, therefore, from the Equator to the point P (where the latitude is ϕ_p) is given by $L(\phi_p)$ and the functional notation $L(\phi)$ is used to denote the Length of the meridian from the

equator to the point on the meridian where the latitude is ϕ . We find then

$$L(\phi_P) = \int_0^{\phi_P} a_P \sec(\psi - \phi) d\phi \quad \dots\dots (3.3)$$

Since the latitude of the point P on the terrestrial spheroid is expressed formally as a geodetic latitude then the integral (3.3) must also be expressed in terms of the geodetic latitude.

If we use equation (3.1) to find ψ in terms of ϕ and differentiate to find $d\psi$ in terms of $d\phi$ then, after some manipulation, the integral (3.3) becomes

$L(\psi_P) = \int_0^{\psi_P} \frac{a(1 - e^2)}{\sqrt{(1 - e^2 \sin^2 \psi)^3}} d\psi$	$\dots\dots (3.4)$
---	--------------------

so that when the geodetic latitude of the point P is ψ_P then the distance along the meridian from the Equator to P is given by $L(\psi_P)$ in equation (3.4).

As an analogy with Meridional Parts we have called this function value $L(\psi_P)$ the LATITUDE PARTS at the point P. In the past it has been the practice to compute the length of the meridian on the surface of the Terrestrial Spheroid by computing the number of minutes of arc in the Difference of Geodetic Latitude and declaring this to be the length of arc in Geographical Miles. Indeed, even now, none of the standard text books recommend anything different but simple numerical methods such as the Direct Cubic Spline (which we will describe in Part 2 of this thesis) can give the value of the integral (3.4) in geographical miles correct to two decimal places when the step length of the procedure is as much as 5° . A table of LATITUDE PARTS was computed by the author²⁰, and is shown in Appendix 2. The computation is based on the Clarke (1866) spheroid where the eccentricity, e , of the meridional ellipse is taken as 0.08227. The paper, detailing the method of computation of Latitude

Parts, was published by the Journal of Navigation. At that time it was still the practice at sea to compute manually but now it is clear that with a modern small PC the values of the integral (3.4) can be computed directly as a subroutine in a computer programme while solving the general problem of navigating along a loxodromic curve. Other papers have been published along the same theme. In 1948 D.H.Sadler²¹ and then in 1950 J.E.D.Williams²² both published works which recommended a more accurate solution to the problem of computing the course and distance along the arc of a loxodromic curve between two points on the surface of a spheroid. Both of these authors each recommended and produced a table of corrections to be applied to the Difference of Latitude between two points on a meridian, when the Difference of Latitude is expressed in minutes of arc, in order to find the actual distance along the meridian between the two points in geographical miles. Implicit in their approach was the fact that the computations would be performed manually. After the publication of the Table of Latitude Parts by the author²⁰, Hairawa²³ acknowledged the introduction of the term "Latitude Parts" and published his own table for the Bessel spheroid for which the eccentricity of the meridional ellipse is 0.081697 . This is the spheroid computed by Bessel based on survey data collected in China, Korea and Japan.

Since the total length of arc of the meridian from the equator to the pole is , in the Bessel spheroid, 5390.96 geographical miles while the distance from the equator to the pole in the sphere with the same radius (3437.7468 geographical miles) is 5400 geographical miles, one might be tempted to assume that the error per degree of latitude is only about 0.1 geographical miles but the form of the meridian is, in fact, such that, from the Equator up to latitude approximately 55°, the length of one minute of arc of the meridian is less than one geographical mile and, above latitude approximately 55° to the Pole, the length of one minute of arc of the meridian is greater than one geographical mile. To the sea-going navigator this is important since most commercial activity lies within latitudes 55°N and 45°S. The error in using the Difference of Latitude instead

of Latitude Parts is therefore nearly always in the same sense. On the surface of the Clarke spheroid, for instance, where the eccentricity $e=0.08227$ approximately, the distance from the Equator to a point P where the geodetic latitude is 1° is equal to 59.56 geographical miles correct to two decimal places. The difference of latitude in minutes of arc would approximate this distance as 60 geographical miles - an error of 0.44 geographical miles and much greater than the average.

3.3 THE MERCATOR PROJECTION OF THE SURFACE OF A SPHEROID.

The projection favoured by Navigators for drawing charts is the Mercator Projection. This projection was discussed in Chapter 2 with reference to the sphere. In a similar manner the surface of the spheroid is projected onto a cylinder whose radius is equal to the equatorial radius of the spheroid and whose axis coincides with the axis of revolution of the spheroid. For the spheroid the mapping is defined by the equations

$$\phi \rightarrow \int_0^\phi a_p \sec(\psi - \phi) \sec \phi \, d\phi \qquad \theta \rightarrow \theta$$

The Mercator Projection of the surface of the spheroid onto the coaxial cylinder of the same radius as the spheroid is a conformal mapping in which the length of one minute of arc of longitude is constant and angles are preserved. The parallels of latitude on the surface of the spheroid are transformed into circles on the surface of the cylinder but they are all of radius equal to the Equator and contained in parallel planes. The Meridians of the Spheroid transform into the meridians of the cylinder and, of course, they cut the images of the parallels of latitude at right angles.

Let ds be the differential element along the arc of the loxodrome on the surface of the spheroid between the points P and Q which cuts the meridians at a constant angle α . See Figure 3.3.

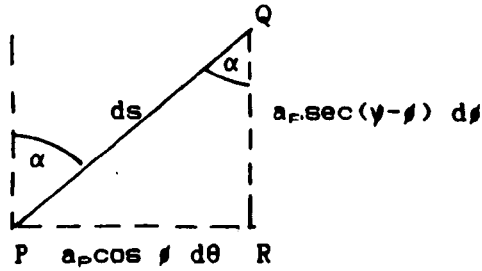


FIGURE 3.3

Corresponding to ds , the distance (PR) along the meridian through P to the parallel of geocentric latitude (QR) is $a_p \sec(\gamma - \phi) d\phi$ (as determined in section 3.2) and the distance (QR) along the parallel where the geocentric latitude is ϕ is $a_p \cos \phi d\theta$. Thus, by elementary considerations of trigonometry,

$$\tan \alpha = \frac{a_p \cos \phi d\theta}{a_p \sec(\gamma - \phi) d\phi} \quad \dots\dots (3.5)$$

We find that the differential (a $d\theta$) of the Difference of Longitude (D'LONG) to be

$$a d\theta = (\tan \alpha) a \sec(\gamma - \phi) \sec \phi d\phi \quad \dots\dots (3.6)$$

If we now consider that an observer has travelled along the loxodrome from the Equator where $\phi=0$ and $\theta=\theta_0$ to a point P where $\phi=\phi_p$ and $\theta=\theta_p$ then we find by integrating equation (3.6)

$$a(\theta_p - \theta_0) = (\tan \alpha) \int_0^{\phi_p} a \sec(\gamma - \phi) \sec \phi d\phi \quad \dots\dots (3.7)$$

If, therefore, we map the point P whose geocentric latitude is ϕ_p onto the surface of the cylinder so that the distance, $N(\phi_p)$, along the meridian from the Equator to the image of P is given by

$$N(\phi_p) = \int_0^{\phi_p} a \sec(\gamma - \phi) \sec \phi d\phi$$

hence

$a(\theta_p - \theta_0) = N(\phi_p) \tan \alpha$	$\dots\dots (3.8)$
--	--------------------

We see then that, under the transformation

$$\phi \rightarrow N(\phi) \quad \theta \rightarrow \theta$$

the angle α is preserved and that the mapping is conformal.

3.4 COMPUTATION OF MERIDIONAL PARTS.

The length of one minute of arc of latitude on the surface of the cylinder varies as a function of latitude and, for a given value of the geodetic latitude, γ , on the surface of the spheroid, the distance, along the meridian on the surface of the cylinder is given by $N(\gamma)$, the MERIDIONAL PARTS for the geodetic latitude γ determined below. In nautical tables the meridional parts are expressed in units of the length of one minute of arc of the equator.

The image of the loxodrome under the Mercator Projection is a circular helix which cuts the meridians of the cylinder at the same constant angle as the loxodrome cuts the meridians on the surface of the spheroid.

Equation (3.8) is the Navigator's formula

$$D'LONG = D'MP \times TAN(Course) \quad \text{for the spheroid.}$$

In the mapping, therefore in which equation (3.8) holds, the angle, α , is preserved, the meridians on the spheroid are mapped into the meridians of the cylinder and the perpendicular distance between the meridians on the surface of the cylinder is constant.

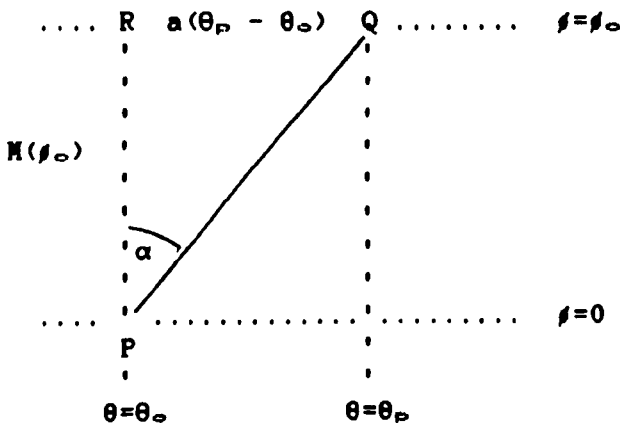


FIGURE 3.4

If the cylinder is then "unfolded" onto a plane then the loxodrome is a straight line which cuts the meridians at the angle α and we have a right angled triangle. The right angle is at the point of intersection of the meridian $\theta = \theta_0$ with the parallel of geocentric latitude $\phi = \phi_0$. $a(\theta_P - \theta_0)$ (which is the perpendicular distance between the meridians on the flat plane) is the side opposite the angle α and $M(\phi_P)$ is the side adjacent to the angle α (Triangle PQR in Figure 3.4).

If we differentiate equation (3.1), solve for $d\phi$ and substitute into equation (3.7), then, along an arc of the loxodrome between two positions P_0 and P_n where the geodetic latitudes are γ_0 and γ_n and the longitudes are θ_0 and θ_n , respectively, and the course is α , we find

$$a(\theta_n - \theta_0) = (\tan \alpha) \int_{\theta_0}^{\theta_n} a(1-e^2) \sec(\gamma-\phi) \cos \phi \sec^2 \gamma \, d\gamma$$

In the integral on the right hand side we express ϕ in terms of γ by means of equation (3.1) and rearrange to find

$$a \int_{\gamma_0}^{\gamma_n} \sec \gamma \, d\gamma = a \int_{\gamma_0}^{\gamma_n} \frac{e^2 \cos \gamma}{1 - e^2 \sin^2 \gamma} \, d\gamma \quad \dots \dots (3.9)$$

which integrates exactly to

$\int_{\gamma_0}^{\gamma_n} \left a \ln \left[\tan \left(\frac{\pi}{4} + \frac{\gamma}{2} \right) \right] - (Mae) \ln \left(\frac{1 + e \sin \gamma}{1 - e \sin \gamma} \right) \right $
--

$$0 < |\gamma| < \pi/2 \quad \dots \dots (3.10)$$

The integral (3.9) and its solution given by (3.10) give the length of the meridian *IN MERCATOR'S PROJECTION OF THE SPHEROID* between two points where the geodetic latitudes are γ_0 and γ_n . This distance is known as the *DIFFERENCE OF MERIDIONAL PARTS* and is measured in the units of the length of one minute of arc of the Equator. The integral over the interval $[0, \gamma]$ for $0 < \gamma < \pi/2$ is tabulated into a *Table of Meridional Parts for the Terrestrial Spheroid* and is now

published in books of Nautical tables to aid manual computations. The evaluation of the expression (3.10) to give the Difference of Meridional Parts (D'MP) presents no problems to the modern computer.

3.5 PROJECTION OF THE SURFACE OF THE SPHEROID ONTO THE POLAR PLANE.

In Chapter 2 we have shown how the surface of the sphere is projected from the South Pole onto the plane tangent to the North Pole under the stereographic projection. In a similar manner we can show how the surface of the spheroid can also be mapped onto the plane tangent to the North Pole so that the mapping is conformal.

We do not project the surface from a single point but the point of projection moves along the axis of the spheroid. See Figure 3.51.

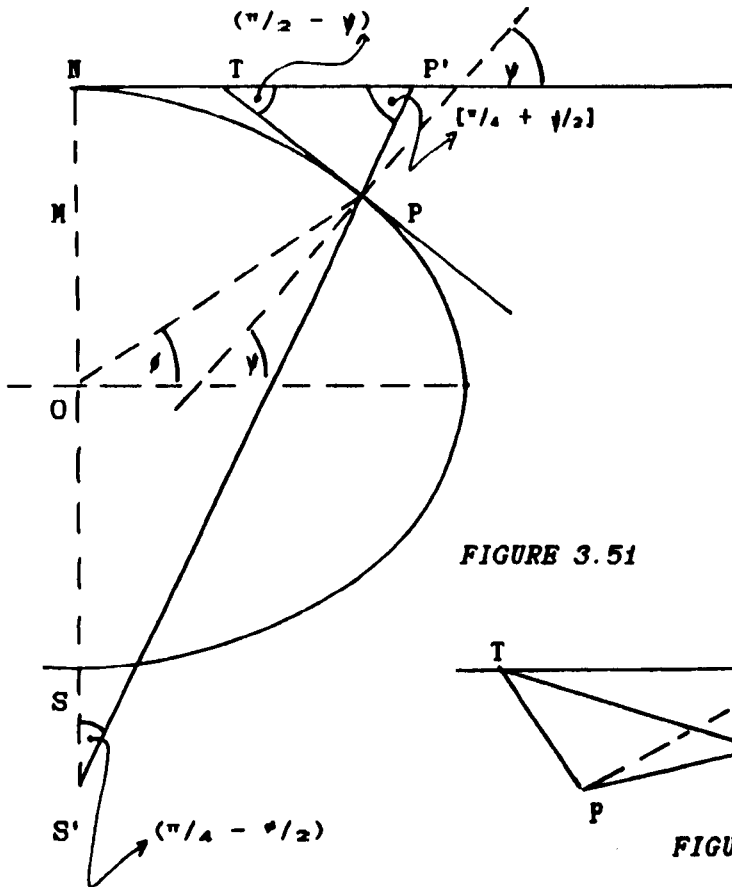


FIGURE 3.51

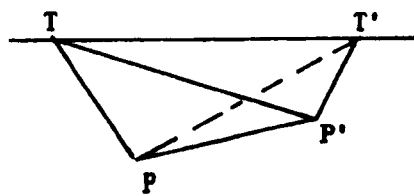


FIGURE 3.511

The geometrical basis of the projection is that the triangle PTP' must always be isocoles. When this is so then the faces PTT' and P'TT' of the wedge PTT'P' (Figure 3.5ii) are congruent triangles in which the angle TPT' (the angle at which the loxodrome cuts the meridian) and the angle TP'T' (the angle at which the image of the loxodrome cuts the radial line in the polar plane) are equal and therefore preserved by the projection. The line TT' is the intersection of the plane through P which is tangent to the spheroid and the polar plane (the plane which is tangent to the pole).

If ψ is the geodetic latitude then the angle at T in the triangle PTP' is equal to $\frac{1}{2}\pi - \psi$ and, hence, when triangle PTP' is isocoles, it can be deduced that the angle at S' (the point where the line P'P produced cuts the axis of the spheroid) is equal to $\frac{1}{2}\pi - \frac{1}{2}\psi$. The length, therefore, of the radius vector r (=NP') in the polar plane is given by

$$r = NP' = K(\psi) \tan(\frac{1}{2}\pi - \frac{1}{2}\psi)$$

where

$$k(\psi) = b - a_p \sin \phi + a_p \cos \phi \cot(\frac{1}{2}\pi - \frac{1}{2}\psi)$$

The stereographic projection is useful in mapping the Polar Regions.

3.6 COMPUTATION OF COURSE AND DISTANCE ALONG A LOXODROMIC CURVE ON THE SURFACE OF A SPHEROID.

If we refer again to Figure 3.3 and to the limiting quantities that result from the triangle PQR we see that

$$ds = (\sec \alpha) a_p \sec(\psi - \phi) d\phi$$

so that the distance, s , between two observed position where the geocentric latitudes are ϕ_0 and ϕ_n is given by

$$s = (\sec \alpha) \int_{\phi_0}^{\phi_n} a_p \sec(\psi - \phi) d\phi$$

If the latitudes are expressed in geodetic form (which is usual) then

$$s = (\sec \alpha) \int_{\psi_0}^{\psi_n} \frac{a(1 - e^2)}{\sqrt{(1 - e^2 \sin^2 \psi)^3}} d\psi \quad \dots\dots (3.11)$$

The integral on the right hand side is the Difference of Latitude Parts and is abbreviated to D'LP. Equation 3.11 is written

$$\text{DISTANCE} = \text{D'LP} \times \text{SEC}(\text{Course})$$

We will now rework the examples used in chapter 2 using the same positions but noting that, on the surface of the spheroid, the latitudes are now geodetic latitudes.

The "Sailing" Triangles are now shown in Figures 3.6(1) and 3.6(11).

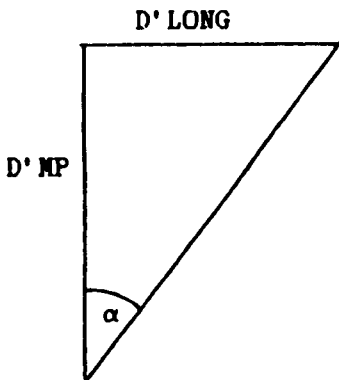


FIGURE 3.6(1)

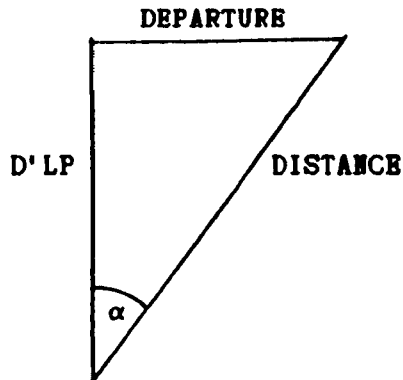


FIGURE 3.6(11)

Example 1. At noon on one day the observer's position was 31°45'N 32°35'E and on the next day at noon the position was 36°30'N 40°20'E. Find the course made good in the twenty four hour period and the average speed for the day.

From Nautical Tables:	M(36°30')	2341.27
	M(31°45')	1998.40
	D'MP	342.97

Final Longitude	40°20'
Initial Longitude	32°35'
D'LONG	7°45' (=465 minutes)

$$\tan \alpha = \frac{D'LONG}{D'MP} = \frac{465}{342.97} = 1.3558037$$

$$\alpha = 53^{\circ}35.3'$$

From Table of Latitude Parts (Appendix 2)

L(36°30')	2177.94
L(31°45')	1893.96
D'LP	283.98

$$s = D'LP \sec \alpha = 478.42$$

COURSE = 053°35.3' DISTANCE = 478.42 SPEED = 19.93 knots

The results are not very different from those obtained for the spherical Earth model (053°28.1', 478.79, 19.95) but, if we had used *Difference of Latitude* instead of *Difference of Latitude Parts* we would have found

$$s = D'LAT \sec \alpha = a(\psi_n - \psi_o) \sec \alpha = 480.14$$

This justifies the opening remark of Chapter 1 - that if the spherical model is used *consistently* then the model is an acceptable approximation but, in fact, the model has not been used consistently and it *has* been the practice to use D'LAT (for a sphere) in the same formula as the course angle, α , determined from spheroid data.

We now rework the second example again where, given the initial position and the course and distance steamed, we need to find the final position. This not such a straightforward problem on the surface of a spheroid. We will perform the computation using a PC with the following procedure. We can compute the D'LP from the formula

$$D'LP = s \cos \alpha$$

and let us suppose that we have found $D'LP = \Delta$ (say) . Then

$$\int_{\psi_0}^{\psi_n} \frac{a(1 - e^2)}{\sqrt{(1 - e^2 \sin^2 \psi)^3}} d\psi = \Delta \quad \dots\dots (3.12)$$

We know ψ_0 but we need to determine ψ_n from (3.12). This can be done by defining $F(\psi_n)$ by

$$F(\psi_n) = \int_{\psi_0}^{\psi_n} \frac{a(1 - e^2)}{\sqrt{(1 - e^2 \sin^2 \psi)^3}} d\psi - \Delta$$

and solving the non-linear equation $F(\psi_n) = 0$ using Newton's method.

We have
$$F'(\psi_n) = f(\psi_n) = \frac{a(1 - e^2)}{\sqrt{(1 - e^2 \sin^2 \psi_n)^3}}$$

and we can determine ψ_n from the iterative scheme

$$\psi_{r+1} = \psi_r - \frac{F(\psi_r)}{f(\psi_r)}$$

The initial approximation ψ_1 for ψ_n can be obtained by assuming that Δ is equal to the Difference of Latitude in minutes of arc and then

$$\psi_1 = \psi_0 + \frac{\Delta}{a}$$

Given, then, that we have determined the final latitude, ψ_n , we can determine the Meridional Parts, $M(\psi_n)$, and the Difference of Meridional Parts $D'MP = M(\psi_n) - M(\psi_o)$. We then find the Difference of Longitude, $D'LONG, [= a(\theta_n - \theta_o)]$ from

$$a(\theta_n - \theta_o) = D'MP \cot \alpha$$

and the final longitude, θ_n , is

$$\theta_n = \frac{1}{a} (D'MP \cot \alpha) + \theta_o$$

The details of the computed solution for Example 2 with the computer program now follow overleaf.

3.7 THE COMPUTER PROGRAM TO SOLVE EXAMPLE 2.

EXAMPLE 2. An observer in position 30°00'N 30°00'E travels a distance of 500 nautical miles along a loxodrome on a course of 045°. Find the final position.

```

10 REM:      PROGRAM TO COMPUTE FINAL POSITION
20 REM:      GIVEN COURSE AND DISTANCE MOVED
30 INPUT "lat1?",x,y:INPUT "Distance?",d:INPUT "course?",c
40 INPUT "Long1?",u,v
50 LPRINT "Initial Lat ";x;y:LPRINT "Initial long";u;v
60 LPRINT "Distance ";d:LPRINT "Course ";c
70 REM: Latitude in degrees (x) and Minutes (y)
80 a=3437.74682:e=0.0824834:pi=3.1415926542
90 REM: This value of e is used in Nories Tables
100 x=x+(y/60):x=x*pi/180:c=c*pi/180:DLP=d*COS(c)
110 REM: Latitude and course angle expressed in radians
120 g=a*(1-e^2)/((1-(e*SIN(x))^2)^(3/2))
130 REM: g is the integrand of L(x)
140 n=3*a*(1-e^2)*(e^2)*SIN(2*x)
150 n=n/(2*((1-(e*SIN(x))^2)^(5/2)))
160 REM: n is the derivative of g
170 y=x+(d/p/a):FOR i= 1 TO 10:h=y-x
180 REM:y is the 1st approximation to final latitude
190 f=a*(1-e^2)/((1-(e*SIN(y))^2)^(3/2))
200 REM: f is the integrand of L(y)
210 m=2*(f-g)/h-n:s=h*(f+g)/2-(h^2)*(m-n)/12
220 REM: m is the approximate derivative of f
230 REM: s is the direct cubic spline
240 REM: approximation to L(y)
250 s=s-DLP:y=y-(s/f):NEXT i
260 REM: we iterate to find y using
270 REM: Newton's Method
280 lat2=y:lat1=x:y=y*180/pi:x=INT(y):y=(y-x)*60
290 LPRINT
300 y=ROUND(y,3):LPRINT "FINAL LATITUDE = ";x;y:
310 REM: We now calculate the final longitude
320 y=lat2:x=lat1:MP1=LOG(TAN((x/2)+(pi/4)))
330 MP1=MP1-(e/2)*LOG((1+e*SIN(x))/(1-e*SIN(x)))
340 MP1=a*MP1:MP2=LOG(TAN((y/2)+(pi/4)))
350 MP2=MP2-(e/2)*LOG((1+e*SIN(y))/(1-e*SIN(y)))
360 MP2=a*MP2:DLONG=(MP2-MP1)*TAN(c)
370 x=u:y=v:x=(x*60)+y
380 x=(x+DLONG)/60:y=(x-INT(x))*60:x=INT(x)
390 y=ROUND(y,2)
400 LPRINT:LPRINT "FINAL LONGITUDE =";x;y:END

```

Initial Lat 30 0 Initial long 30 0
Distance 500 Course 45

FINAL LATITUDE = 35 54.9
FINAL LONGITUDE = 37 1.28

In chapter 2 using the spherical approximation to the shape of the Earth we found

FINAL POSITION : 35°53.55'N 37°01.09'W.

The difference in the latitudes is quite significant.

3.8 THE METHOD OF CARLTON WIPPERN USING ELLIPTICAL INTEGRALS.

The function $L(\gamma)$ gives us the length of arc of the meridian on the surface of a spheroid from the equator to the parallel where the geodetic latitude is γ . This the length of the arc of an ellipse and can also be evaluated using an elliptical integral. The use of the elliptical integral to compute the length of arc of the meridian was featured in the paper by Carlton-Wippern²⁴ published in the Journal of Navigation in May 1992. It does mean that the latitude must be transformed to suit the form in which the elliptical integral is expressed .

Let a meridian on the surface of a spheroid whose equatorial radius is a and polar radius b be expressed in cartesian coordinates in the usual manner so that the origin is at the centre of the spheroid. Its equation is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

and the length of arc from the point P on the ellipse where the x coordinate is x_p to the extremity of the major axis is given by the integral

$$\int_{x_P}^a \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx$$

If we use the substitution $\sin \chi = \frac{x}{a}$ then this integral becomes

$$a \int_{\chi_P}^{\pi/2} \sqrt{1 - e^2 \sin^2 \chi} d\chi$$

which is an elliptical integral.

The angle χ is the angle subtended at the point P' on the auxiliary circle by the abscissa of the point P on the ellipse. χ is connected to the geocentric latitude, ϕ , and then the geodetic latitude, ψ , by the equations

$$\tan \phi = (1 - e^2) \tan \psi$$

$$\sin \chi = \cos \phi \sqrt{\frac{1 - e^2}{1 - e^2 \cos^2 \phi}}$$

See Figure 3.3.

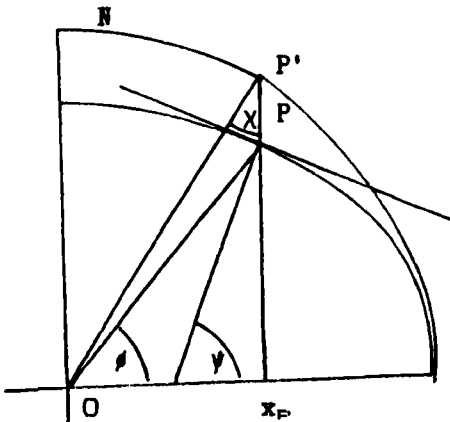


FIGURE 3.3

Carlton-Wipperfurth did not give any numerical results to his method probably because the basic formulae that he uses are well founded. His method follows the pattern of the method used in section 3.6 and in the computer program of section 3.7 here.

- 4 -

NAVIGATING ALONG THE GEODESIC
PATH BETWEEN TWO POINTS
ON THE SURFACE OF A SPHERE

4.1. GEODESIC ARCS.

While it is often convenient, in order to avoid the flow of current or adverse prevailing weather conditions, to navigate across the oceans between two points along a line of constant course it would seem more logical to seek the shortest path. Over long distances often the result is a compromise, part of the journey along a loxodromic curve and part along the shortest path. We have discussed the problem of navigating along the arc of a loxodromic curve in Chapters 2 and 3 - now we will consider the problem of navigating along the shortest path.

On any surface of suitable continuity class the shortest path between two points on the surface is along the arc of a GEODESIC CURVE. The definition of a geodesic curve on a surface is a curve along whose length, at every point, the normal to the curve is also the normal to the surface at that point. There may be arcs of more than one geodesic curve through the two points and the corresponding arcs of these geodesic curves may be of different length. There may also be more than one geodesic arc between the two points which are of the same length.

We are particularly concerned with the problems of determining the shortest path between two points on the surface of a SPHERE and on the surface of an OBLATE SPHEROID both of which are used as approximations to the shape of the Earth. The sphere and the oblate spheroid are surfaces of revolution. An oblate spheroid is generated by revolving an ellipse about its minor axis and a sphere is the special case in which that ellipse is a circle. In this chapter we consider the shortest path on the surface of a SPHERE.

4.2 GEODESIC ARCS ON THE SURFACE OF A SPHERE.

On the surface of a sphere all geodesic curves are great circles. A great circle on the surface of a sphere is a circle whose plane passes through the centre of the sphere and, except for points which are antipodean, the great circle arc which passes through two points

is unique . The plane of this great circle is generally defined by the two points together with the point at the centre of the sphere so that, when the two points are antipodean, the three points are then colinear and there are an infinite number of planes which pass through these points. Consider, for instance, the family of circles whose common diameter lies along the axis of revolution of the sphere.

Figure 4.1 shows a plane intersecting a sphere through its centre O. The closed curve WQER is a great circle.

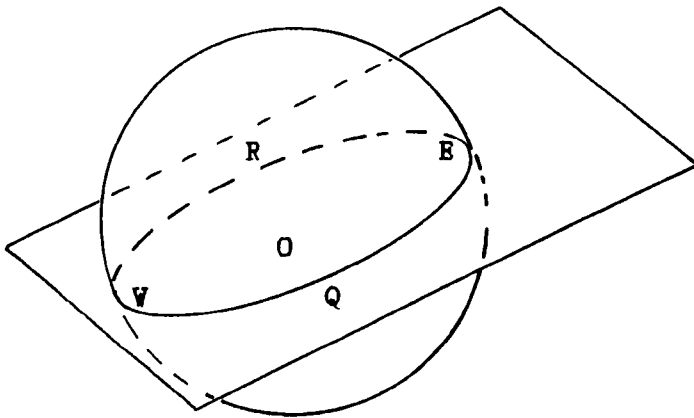


FIGURE 4.1

4.3 SURFACES OF REVOLUTION - CLAIRAUT'S EQUATION.

On a surface of revolution which satisfies the required continuity conditions a special set of plane geodesic curves are defined by the intersection of the surface with a plane through the axis of revolution. These geodesic curves are known as meridians. Figure 4.2 shows meridians on the surface of a sphere. In the case of a sphere all the meridians begin and end on the axis of revolution at the extremities of a diameter. In the figure these extremities are N and S, known as the Poles, and, on the surface of the Earth these are designated as the North Pole and the South Pole.

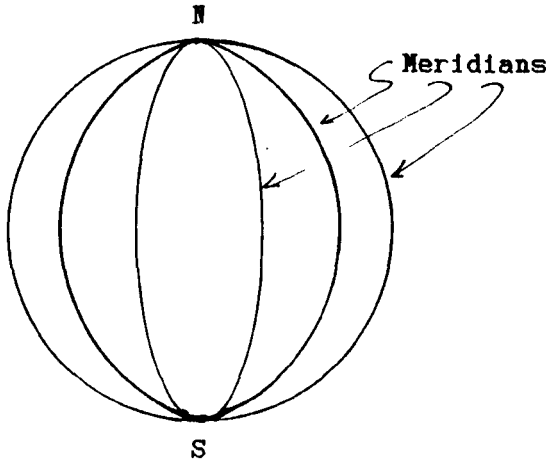


FIGURE 4.2
MERIDIANS ON THE
SURFACE OF A SPHERE

If any other geodesic on the surface of revolution cuts the meridians m_0, m_1, \dots, m_n in points P_0, P_1, \dots, P_n at angles $\gamma_0, \gamma_1, \dots, \gamma_n$, respectively, then we find

$$\boxed{r_1 \sin \gamma_1 = \text{constant}} \quad \dots \quad (4.1)$$

where r_1 is the perpendicular distance from P_1 to the axis of revolution. See Figure 4.3

Equation (4.1) is known as CLAIRAUT'S EQUATION. (Lyusternik²⁵.)

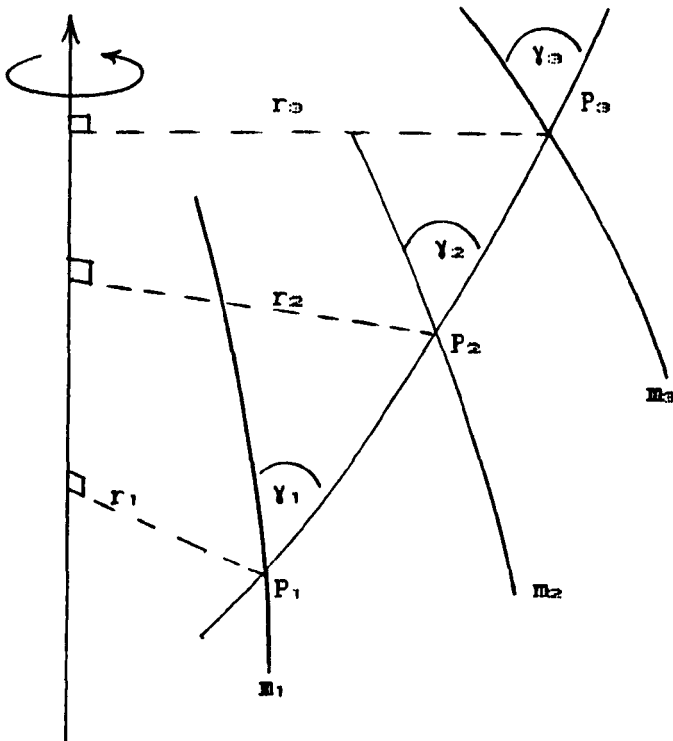


FIGURE 4.3

The proof of Clairaut's Equation which follows is taken from the book by Bell²⁶. It applies to any surface of revolution which satisfies the required continuity properties and it follows from the basic definition of a geodesic curve on a surface as stated in section 4.1 i.e. a geodesic curve is a curve on a surface along which, at every point, the normal to the curve is also the normal to the surface at that point.

In the xyz coordinate frame let the surface be represented by the equation

$$F(x, y, z) = 0$$

and let the curve on this surface be defined in terms of the arc length parameter s so that, at a point P on the curve the vector $OP (= r)$ is given by

$$\vec{r} = x(s) \hat{i} + y(s) \hat{j} + z(s) \hat{k} .$$

The normal to the surface is along the direction of the vector

$$(F_x, F_y, F_z) .$$

where F_x , F_y , and F_z are the first partial derivatives of the function $F(x, y, z)$ with respect to x, y, z , respectively. The normal to the curve is along the direction of the vector

$$\frac{d^2\vec{r}}{ds^2} = \frac{d^2x}{ds^2} \hat{i} + \frac{d^2y}{ds^2} \hat{j} + \frac{d^2z}{ds^2} \hat{k}$$

From this we see that, if the curve is a geodesic curve,

$$\frac{d^2x}{ds^2} = \lambda F_x \quad \frac{d^2y}{ds^2} = \lambda F_y \quad \frac{d^2z}{ds^2} = \lambda F_z \quad \dots (4.2)$$

for some constant λ .

Now, on a surface of revolution, the z coordinate can be expressed in terms of x and y and the implicit function $F(x, y, z)$ may thus be written in the form

$$F(x, y, z) = f(\sqrt{x^2 + y^2}) - z$$

so that, substituting $u = \sqrt{x^2 + y^2}$ we find

$$F_x = \frac{x}{u} f'(u) \quad F_y = \frac{y}{u} f'(u)$$

hence from equation (4.2)

$$x \frac{d^2y}{ds^2} - y \frac{d^2x}{ds^2} = 0$$

or
$$\frac{d}{ds} \left(x \frac{dy}{ds} - y \frac{dx}{ds} \right) = 0$$

so that
$$x \frac{dy}{ds} - y \frac{dx}{ds} = \text{constant} \quad \dots\dots (4.3)$$

In polar coordinates $x = u \cos \theta$ and $y = u \sin \theta$

$$\frac{dx}{ds} = \frac{du}{ds} \cos \theta - u \frac{d\theta}{ds} \sin \theta$$

$$\frac{dy}{ds} = \frac{du}{ds} \sin \theta + u \frac{d\theta}{ds} \cos \theta$$

Substituting these in equation (4.3) and rearranging gives

$u^2 \frac{d\theta}{ds} = \text{constant}$ (4.4)
--	-------------

and this is the general result for a surface of revolution as given by Bell²⁶.

In the particular case of the sphere consider the differential element ds of the geodesic arc between points P and R (Figure 4.4). At P let the latitude be ϕ and the longitude θ . Corresponding to ds we have $u \sec \phi d\phi$ (= PQ) along the arc of the meridian through P and $u d\theta$ (= QR) along the parallel of latitude through Q and R.

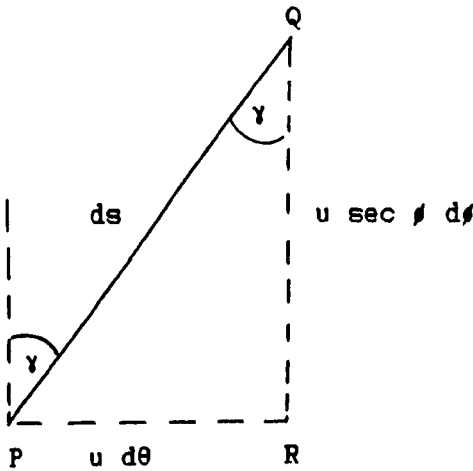


FIGURE 4.4

The angle \hat{PQR} ($= \gamma$) is given by

$$\sin \gamma = \frac{u \, d\theta}{ds}$$

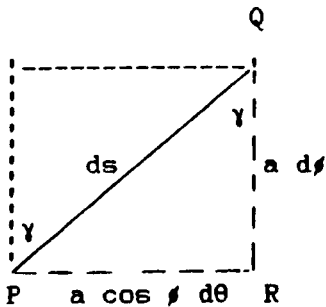
so that equation (4.4) becomes

$$u \sin \gamma = \text{constant}$$

which is Clairaut's Equation as given by equation (4.1).

7.5 NAVIGATING ALONG THE ARC OF A GREAT CIRCLE ON THE SURFACE OF THE SPHERICAL EARTH.

Let us consider a point P on the arc of a great circle which cuts the meridian through P at an angle γ . Let ds be the differential element of the great circle at P. See Figure 4.5. In the triangle PQR corresponding to ds ($=PQ$) we have $PQ = a \, d\phi$ and $QR = a \cos \phi \, d\theta$ where a is the radius of the sphere.



From triangle PQR we see that

$$\sin \gamma = \frac{a \cos \phi \, d\theta}{ds}$$

FIGURE 4.5

and, substituting this into Equation (4.1), we find the differential equation

$$a^2 \cos^2 \phi \left(\frac{d\theta}{ds} \right) = \text{constant}$$

since the length of the perpendicular from P to the axis of revolution is $a \cos \phi$.

At the vertex, V , of the great circle, (the point on the great circle at which the latitude is a maximum - see figure 4.6), let the latitude be ϕ_v . At this point the great circle cuts the meridian at right angles and for this curve we then have $\gamma = \pi/2$ and $\sin \gamma = 1$ so that the constant in equation (4.1) is equal to $a \cos \phi_v$.

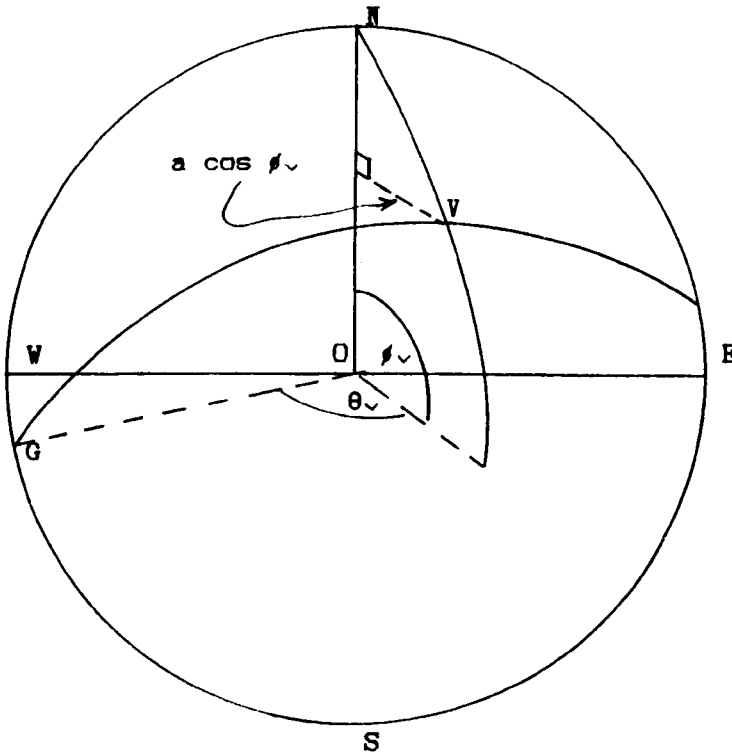


FIGURE 4.6

The great circle that passes through the point P and reaches its vertex, V , in latitude ϕ_v is therefore defined by the differential equation

$$\boxed{a \cos^2 \phi \left(\frac{d\theta}{ds} \right) = \cos \phi_v} \quad \dots\dots (4.5)$$

and equation (4.1) can also be written in the form

$$\boxed{\cos \phi \sin \gamma = \cos \phi_v} \quad \dots\dots (4.6)$$

We can solve equation (4.5) using the substitution $y = a \tan \phi$ (See Figure 4.7) and by expressing ds in terms of dy and $d\theta$.

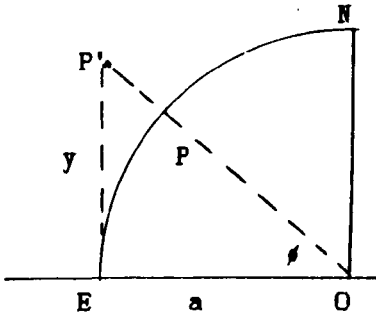


FIGURE 4.7

$$EP' = y = a \tan \phi$$

We find

$$\cos^2 \phi = \frac{a^2}{(y^2 + a^2)}$$

$$d\phi = \frac{a \, dy}{(y^2 + a^2)}$$

and, from Figure 4.5, we see that

$$ds^2 = a^2 d\phi^2 + a^2 \cos^2 \phi \, d\theta^2$$

Using these substitutions Equation (4.5) becomes

$$d\theta = \frac{dy}{\sqrt{(y_{\checkmark}^2 - y^2)}}$$

and its solution is

$$y = y_{\checkmark} \sin(\theta - \theta_E) \quad \dots\dots (4.7)$$

where $y_{\checkmark} = a \tan \phi_{\checkmark}$ and θ_E is the longitude in which the great circle crosses the equator. Equation (4.7) represents the cylindrical projection (not the Mercator Projection) of the great circle onto an infinite cylinder which is coaxial with the sphere.

If the great circle passes through the points P_A , where $y = y_A$ and $\theta = \theta_A$, and P_B , where $y = y_B$ and $\theta = \theta_B$, then we can find y_v and θ_E from equation (4.7)

$$\tan \theta_E = \frac{y_A \sin \theta_B - y_B \sin \theta_A}{y_A \cos \theta_B - y_B \cos \theta_A} \dots (4.8)$$

and $y_v = y_A \operatorname{cosec}(\theta_A - \theta_E) \dots (4.9)a$

or $y_v = y_B \operatorname{cosec}(\theta_B - \theta_E) \dots (4.9)b$

Replacing $y = \tan \phi$ in (4.7) we find the general equation which gives the latitude, ϕ , in terms of the longitude, θ , along the arc of a great circle :

$$\boxed{\tan \phi = \tan \phi_v \sin(\theta - \theta_v)} \dots (4.10)$$

To determine the distance, s , along the arc of the great circle between P_A and P_B we separate the variables in equation (4.5) and integrate to give

$$\boxed{s = a \int_{\theta_A}^{\theta_B} \frac{\cos^2 \phi}{\cos \phi_v} d\theta} \dots (4.11)$$

Although this integral does have an analytical solution it can be just as easy to compute numerically since, for navigational purposes, it is desirable to compute a number of intermediate points along the path of the Great Circle. The step-by-step method of the Direct Cubic Spline (which is described in Part 2 of this thesis) suits this purpose well and we also find that, when $\phi = \phi_v$ and $\phi = 0$, the derivatives of the integrand in equation (4.11) are both equal to zero and this provides us with the boundary conditions which are a desirable (but not essential) part of the Direct Cubic Spline computational scheme.

Subdividing the interval $[\theta_A, \theta_B]$ at points where the longitudes are

$$\theta_1, \theta_2, \dots, \theta_{n-1}$$

and such that $\theta_A = \theta_0 < \theta_1 < \dots < \theta_{n-1} < \theta_n = \theta_B$

we can determine the corresponding latitudes $\phi_1, \phi_2, \dots, \phi_{n-1}$ of intermediate points between P_A and P_B using whichever of the equations (4.9)a or (4.9)b is relevant.

If the vertex of the great circle lies on the arc between P_A and P_B then we can express the integral in equation (4.11) as the sum of two parts :

$$\int_{\theta_A}^{\theta_B} f(\theta) d\theta = \int_{\theta_A}^{\theta_V} f(\theta) d\theta + \int_{\theta_V}^{\theta_B} f(\theta) d\theta$$

or
$$\int_{\theta_A}^{\theta_B} f(\theta) d\theta = \int_{\theta_V}^{\theta_B} f(\theta) d\theta - \int_{\theta_V}^{\theta_A} f(\theta) d\theta$$

and the boundary condition $f'(\theta_V) = 0$ can be applied.

Similarly, if the great circle arc between P_A and P_B crosses the Equator then

$$\int_{\theta_A}^{\theta_B} f(\theta) d\theta = \int_{\theta_E}^{\theta_B} f(\theta) d\theta - \int_{\theta_E}^{\theta_A} f(\theta) d\theta$$

and the boundary condition $f'(\theta_E) = 0$ can be used.

From equation (4.6) we can determine the angle between the great circle and the meridian at any point along the path and, hence, for the navigator, the course to steer.

The spherical model is the one most often used by seagoing navigators as the model for the shape of the Earth particularly in the manual method of computation of shortest distance. This computation is usually called "Great Circle Sailing" and its solution is usually effected using the methods of Spherical

Trigonometry - the Spherical Cosine Formula and Napier's Rules. The methods described here, however, can be just as simple to apply and do give us a lead into to solving the more complicated problem of determining the shortest path on the surface of a spheroid.

4.5 COMPUTATION OF POSITIONS ALONG THE GREAT CIRCLE PATH AND THE COURSE AND DISTANCE BETWEEN THEM.

As an example of the application of the Direct Cubic Spline Approximation to the computation of intermediate points along the path of a great circle arc on the surface of a sphere and the courses and distances between them, let us consider the arc of the great circle which starts at the Equator in longitude 0° and reaches its vertex in latitude 45° at longitude 90° .

At a point P on this particular great circle arc the latitude, ϕ , and the longitude, θ , are related by

$$\tan \phi = \sin \theta.$$

This is a consequence of equation (4.10).

Let us choose a set of intermediate points $\{P_i\}$ along the path of the great circle where the longitudes $\{\theta_i\}$ of these points are evenly spaced at 5° intervals. The latitudes $\{\phi_i\}$ are given by

$$\phi_i = \tan^{-1}(\sin \theta_i)$$

At the point P_i the integrand, f_i (the value of $f(\theta)$ when $\theta=\theta_i$), in equation 4.11 is, in this case, given by

$$f_i = a \sqrt{2} \cos^2 \phi_i$$

and the course angle, γ_i , is, from equation (4.6), given by

$$\gamma_i = \sin^{-1}(\frac{1}{2} \sqrt{2} \sec \phi_i).$$

If S_i is the approximation to the distance P_0P_i along the arc of the great circle and the $\{M_i\}$ are the moments of the cubic spline, then the computational scheme (Part 2 - chapter 1 : section 1.4) for the latitudes of the points along the path, the course angles and the distances is given by

$$S_0 = 0 : f_0 = a \sqrt{2} : M_0 = 0 : h = \frac{\pi}{36} : \theta_0 = 0 : \phi_0 = 0$$

For $i = 1$ to 18:

$$\theta_i = ih$$

$$\phi_i = \tan^{-1}(\sin \theta_i)$$

$$\gamma_i = \sin^{-1}(\frac{1}{2} \sqrt{2} \sec \phi_i)$$

$$f_i = a \sqrt{2} \cos^2 \phi_i$$

$$M_i = \frac{2(f_i - f_{i-1})}{h} - M_{i-1}$$

$$S_i = S_{i-1} + \frac{h}{2}(f_i + f_{i-1}) - \frac{h^2}{12}(M_i - M_{i-1})$$

The results from the computation are shown in Table 4.1 below.

1	θ_i	ϕ_i	COURSE	DISTANCE
1	5	4°58.9'	045.22°	423.20
2	10	9°51.1'	045.86°	840.12
3	15	14°30.6'	046.92°	1245.22
4	20	18°52.9'	048.36°	1634.19
5	25	22°54.6'	050.14°	2004.20
6	30	26°33.9'	052.24°	2353.90
7	35	29°50.3'	054.60°	2683.15
8	40	32°43.9'	057.20°	2992.76
9	45	35°15.9'	060.00°	3284.14
10	50	37°27.22	062.97°	3559.08
11	55	39°19.4'	066.07°	3819.54
12	60	40°53.6'	069.30°	4067.54
13	65	42°11.2'	072.61°	4305.07
14	70	43°13.2'	076.00°	4534.03
15	75	44°00.4'	079.45°	4756.28
16	80	44°33.7'	082.95°	4973.56
17	85	44°53.4'	086.47	5187.60
18	90	45°00.0'	090.00	5400.00

TABLE 4.1

It is interesting to note that when $i=18$ the actual results for the latitude and the course were correct to 7 decimal places and the distance was correct to 4 decimal places. For the purpose of the table the results have been rounded to two decimal places.

The work in this section appeared as a paper published in the Journal of Navigation²⁷.

4.6 SOLUTION OF GREAT CIRCLE SAILING PROBLEM USING SPHERICAL TRIGONOMETRY.

A spherical triangle is the area enclosed by the intersection of three great circles. Figure 4.8 shows a spherical triangle PAB on the surface of a sphere. In the triangle the angles are denoted by P, X and Y and the sides as p, x and y. The sides are expressed as angles where p, x and y are the angles subtended at the centre of the sphere by the arcs XY, PY and PX respectively .

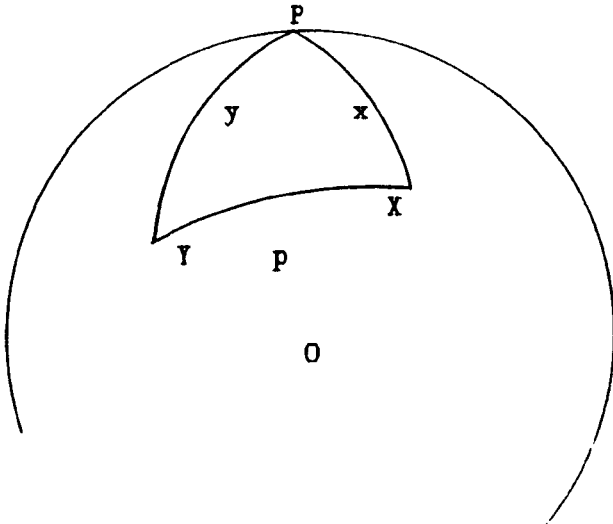


FIGURE 4.8

To solve the triangle for navigational purposes we use the SPHERICAL COSINE FORMULA.

Given a spherical triangle PXY with angles at P,X and Y and sides p,x,y where side p is opposite angle P, etc., we would use the Cosine Formula in the form

$$\cos P = \frac{\cos p - \cos x \cos y}{\sin x \sin y}$$

Or, to find a side (p, say) we would use the Cosine Formula in the form

$$\cos p = \cos x \cos y + \sin x \sin y \cos P$$

In Navigation the point P would normally be used to denote the position of the Pole. The initial position on the great circle would be X and the final position at Y. We will know the angle at P which is the difference of longitude between X and Y and we will know the angles subtended by the arcs PX (=y) and PY (=x)

$$y = \frac{1}{2}\pi - \phi_x \qquad x = \frac{1}{2}\pi - \phi_y$$

where ϕ_x and ϕ_y are the latitudes of the points X and Y respectively.

Using the cosine formula to find the angle X, which will be the initial course along the path of the great circle, we would find

$$\cos X = \frac{\cos x - \cos p \cos y}{\sin p \sin y} \qquad \dots\dots (4.12)$$

and to determine the angle subtended by the arc which is the side p of the triangle, we would use

$$\cos p = \cos x \cos y + \sin x \sin y \cos P \quad \dots (4.13)$$

If the point P, then, is a pole of the Spherical Earth, and the path of the great circle is from X to Y we would first determine the angle at X which is the initial course of the great circle using the Equation (4.12). We would then determine the angle, p, subtended by the arc XY so that the distance, s, in geographical miles along the arc of the great circle between X and Y using Equation (4.13) is

$$s = ap$$

where a is the radius of the spherical Earth and p is expressed in radians.

Using the information we have now found we would then compute the position of the vertex, V, (the point at which the latitude along

the arc of the great circle is a maximum) which may or may not lie between X and Y. See Figure 4.9.

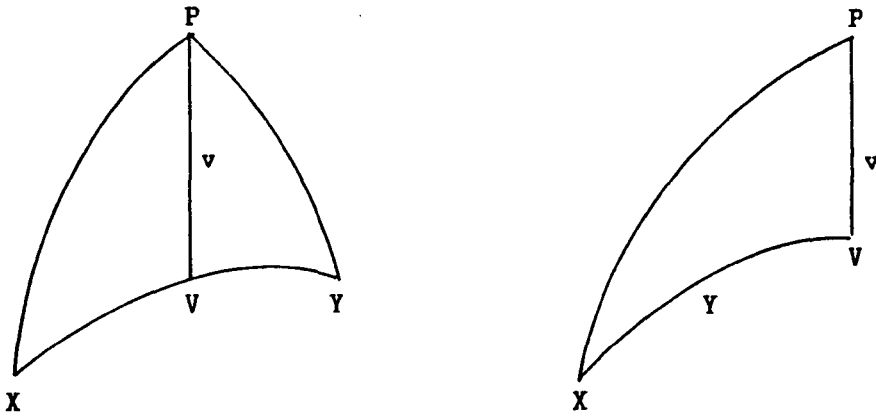


FIGURE 4.8

The angle at V in spherical triangles PVX and PYV are right angles and this simplifies the application of the Cosine Formula. In triangle PVX we know the angle subtended by the arc PX and we have found the angle at X above. We therefore find the angle, v , subtended by the arc PV from

$$\sin v = \sin y \sin X$$

and the angle at P (P_{vx}) in the triangle PVX from

$$\cot P_{vx} = \cos y \tan X$$

We then find the latitude, ϕ_v , and longitude, θ_v , of the vertex

$$\phi_v = \frac{1}{2}\pi - v \quad \theta_v = \theta_x + P_{vx}$$

Given then that we wish to find the latitudes (ϕ_1) of intermediate points lying in longitudes (θ_1) we have

$$\tan \phi_1 = \tan \phi_v \cos (\theta_v - \theta_1)$$

We would compute any information required in triangle PYV in a similar manner.

The methods employing spherical trigonometry are well tried and tested. It is not therefore necessary to show the results that would be obtained by the methods of spherical trigonometry as a comparison with the Direct Cubic Spline method although this is done indirectly in Chapter 5. It is hard to imagine, in any case, that any results better than those shown in table 4.1 could be achieved by any method.

There is not much to choose in the amount of computation required in either of the above methods of computing the distance along the great circle path but the first method, utilising the Direct Cubic Spline Approximation, we do find a step by step method of computing the intermediate positions and the course and distance between them along the great circle arc during the process of computing the overall distance. This saves a lot of time. The same method also serves to solve the similar problem of computing the shortest distance along a geodesic arc on the surface of a spheroid which is a better approximation to the shape of the Earth and where the methods of spherical trigonometry do not apply.

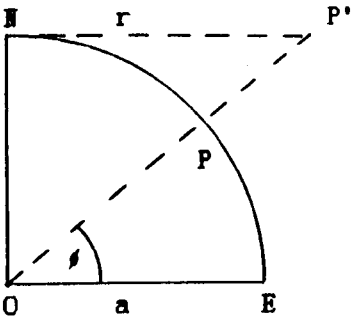
4.6 THE GNOMIC PROJECTION OF THE GREAT CIRCLE.

If, from the centre of the sphere, we project the arc of a great circle onto a tangent plane then the resulting image is always a straight line. This projection, known as the Gnomonic projection, has many uses in navigation and many large scale navigational charts of ports and harbours are constructed using this projection. There are, however, gnomonic projections of the Ocean Basins from which a navigator can lay off his great circle track as a straight line and then pick off points to transfer to the Mercator Chart. This is sometimes done instead of the computations above but there is a consequent loss of accuracy.

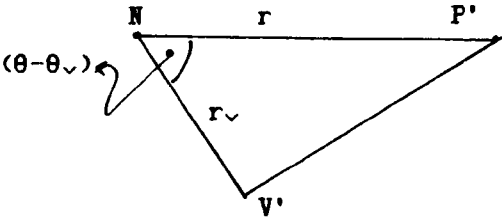
In the case where the tangent plane is the polar plane then, with the origin at the pole, using polar coordinates (r, θ) where $r = a \cot \phi$ and θ is the longitude, the equation of the gnomonic projection of a great circle is

$$r = r_v \sec(\theta - \theta_v) \quad \dots\dots (4.14)$$

where $r_v = a \cot \phi_v$. See Figure 4.9 .



(i)



(ii)

FIGURE 4.9

Figure 4.9 (i) shows a section through the axis of the sphere and the point P. The point P' is the projection of P in the polar plane. The straight line V'P' in the tangent plane (Figure 4.9 (ii)) is the projection onto the polar plane of the arc VP of the great circle.

- 5 -

NAVIGATING ALONG THE GEODESIC
PATH BETWEEN TWO POINTS
ON THE SURFACE OF A SPHEROID

5.1 GEODESIC PATHS ON THE SURFACE OF A SPHEROID.

On the surface of a sphere the geodesic paths are the great circles but on the surface of an oblate spheroid the geodesic paths are not so easily defined except that the Equator of the spheroid is a circle and its meridians are ellipses. An oblate spheroid is generated by revolving an ellipse about its minor axis. On the surface of a spheroid the shortest path between two points, P_0 and P_n , is along the arc of a geodesic curve but this curve, unlike a great circle on the surface of a sphere, is not always a plane curve nor is it necessarily part of a closed curve. This means that if we project the arc of the geodesic curve from the centre of a spheroid onto a plane tangent to the spheroid then the resulting locus is *not*, in general, a straight line as it is in the case of the gnomonic projection of the great circle on a sphere.

A spheroid whose meridians are ellipses of fixed eccentricity is a better approximation to the shape of the Earth than the sphere and it is the approximation that we shall use here. In fact, the Earth may be better approximated by the smooth union of a number of spheroids and, in the science of navigation, the distances across the different ocean basins may be calculated using a different value for e , the eccentricity of the meridional ellipse. Corresponding to the spheroid which we will adopt as the approximation to the shape of the Earth, there is a sphere whose equator coincides with the equator of the spheroid. We will refer to this sphere as the CORRESPONDING SPHERE. We have chose to call this sphere the "corresponding sphere" because the latitude ϕ on the surface of the sphere corresponds directly to the geodetic latitude ψ on the surface of the spheroid.

There is another sphere which is known as the JACOBI (AUXILLIARY) SPHERE. This sphere is, physically, the same sphere as the corresponding sphere but the relationship between the latitude ϕ on the Jacobi sphere and the geodetic latitude ψ on the surface of the spheroid is given by

$$\tan \phi = \sqrt{(1-e^2)} \tan \psi .$$

The latitude ϕ is known as the "Reduced" latitude of the spheroid. There is a special relationship between the geodesic arc on the surface of the spheroid whose vertex lies in latitude ψ and the great circle on the surface of the Jacobi sphere whose vertex lies in corresponding latitude ϕ - at points with corresponding values of the latitude the azimuth angle of the spheroidal geodesic is the same as the azimuth angle of the great circle. This result is due to Jacobi and hence the auxiliary sphere is sometimes known by his name. At points where latitudes correspond the longitudes, however, do not, in general, correspond. We do not make use of the Jacobi sphere in this analysis; we make use exclusively of the relationship between the spheroid and the corresponding sphere .

On the surface of a spheroid the number of geodesic arcs joining two points P_1 and P_2 differs according to the relative positions of the two points on the surface. As in the case of the sphere there are, for instance, an infinite number of meridians which join the poles but this is not so for any other antipodean points. We have, however, the new problem of *NEARLY* antipodean points, i.e. those points for which the difference of longitude exceeds a certain fixed value ($\approx 179^\circ 24'$ on the surface of the terrestrial spheroid). These points form a special case and with which we will deal in the next chapter. In the case of antipodean points which are not the poles the shortest paths are easily defined on the surface of the spheroid - there are two equal shortest paths, both of which coincide with the path of the meridian one of which passes through the north pole and the other through the the south pole. These paths are impractical for the use of seafaring navigators but are very useful to aviators.

In this chapter we shall deal with the computation of the shortest path along a geodesic arc between two points which are not antipodean or "nearly antipodean" and for which it turns out that the geodesic arc is unique. We shall use the standard notation for

spherical coordinates with the origin of the coordinate system at O , the centre of the spheroid. The GEOCENTRIC LATITUDE is then denoted by ϕ and the LONGITUDE by θ . The GEODETIC (ASTRONOMICAL) LATITUDE at a point P on the surface is the angle between the normal to the surface at P and the plane of the Equator. This is an intrinsic property of the surface and we will denote it by ψ . The ranges of values of ϕ , ψ and θ are

$$\begin{aligned}
 -\frac{1}{2}\pi &\leq \phi, \psi \leq \frac{1}{2}\pi && \text{(North positive)} \\
 0 &\leq \theta \leq 2\pi && \text{(East positive)} .
 \end{aligned}$$

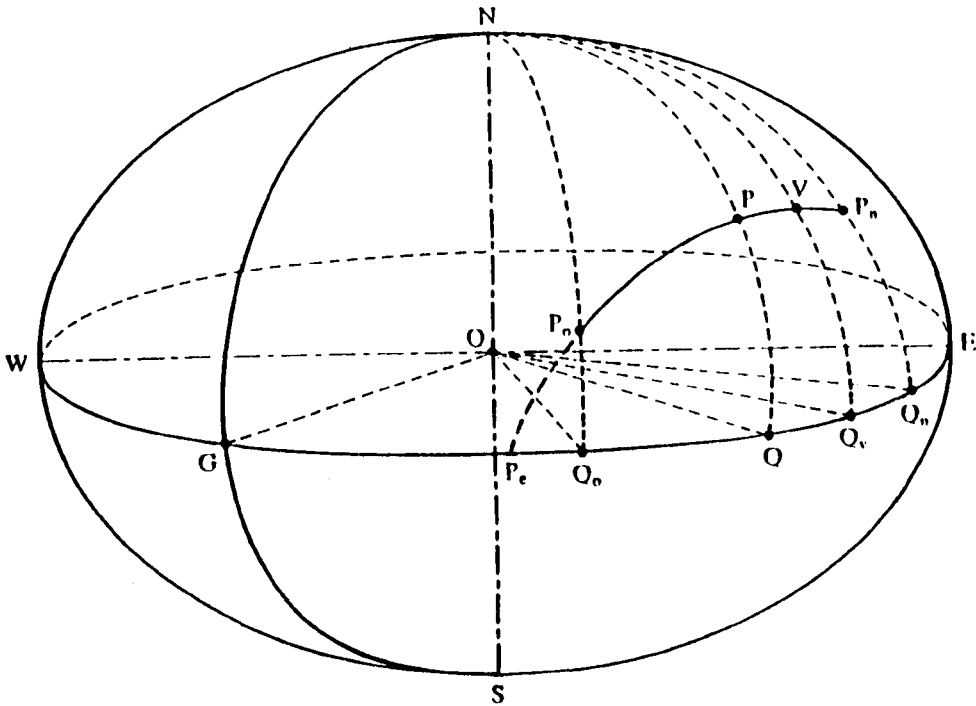


FIGURE 5.1 - THE GEODESIC PATH P_0PVP_n ON THE SPHEROIDAL EARTH.

Figure 5.1 is a representation of the Earth as a regular oblate spheroid. O is the centre of the spheroid, N is the North Pole, S is the South Pole and the line NOS is the axis of the spheroid. The meridian NGS is the Greenwich Meridian, WGE is the Equator. The arc P_0PVP_n is the geodesic arc joining the points P_0 and P_n and V is the vertex of the geodesic. V may not necessarily lie between P_0 and P_n . P is a general point along the path of the geodesic between P_0 and P_n and P_E is the point where the geodesic, when extended, crosses the Equator, or, if P_0 and P_n lie in opposite hemispheres, P_E is the point between P_0 and P_n .

Q_0, Q, Q_V, Q_n are the points on the Equator where the meridians through P_0, P, P_V, P_n , respectively, cut the Equator. The angles GOQ_0, GOQ, GOQ_V, GOQ_n are therefore the longitudes of the points P_0, P, P_V, P_n , respectively and the angles $Q_0OP_0, QOP, Q_VOP_V, Q_nOP_n$ are the geocentric latitudes.

5.2 THE EQUATION OF A GEODESIC CURVE ON THE SURFACE OF THE SPHEROIDAL EARTH.

Let us consider that the Earth is a regular spheroid with equatorial radius a , polar radius b and that the eccentricity of the meridional ellipse is e . See Figure 5.2.

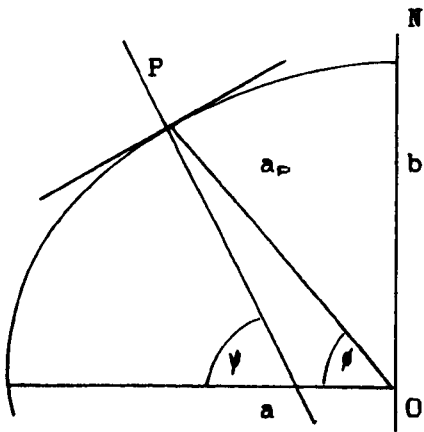


FIGURE 5.2

Let a position on the surface of the spheroid be determined by its geocentric latitude, ϕ , and its longitude, θ . The distances are measured in units of one minute arc of the equator and the angles

are in radians. Let the geodetic latitude be ψ . ϕ and ψ are connected by the relationship

$$\tan \phi = (1-e^2)\tan \psi \quad \dots\dots (5.1)$$

Since a spheroid, like the sphere, is a surface of revolution fulfilling the required continuity conditions we can, once again, apply Clairaut's Equation.

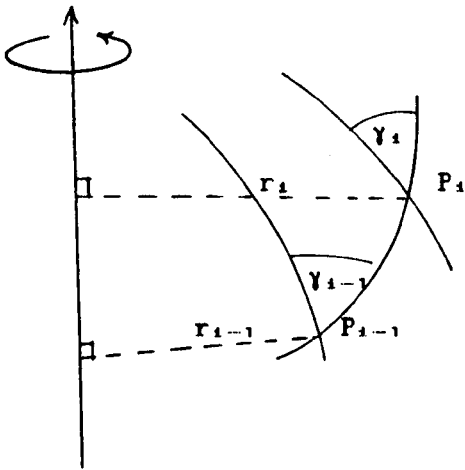


FIGURE 5.3

Let there be a geodesic arc joining two points P_0 and P_n which passes through intermediate points P_1, P_2, \dots, P_{n-1} . See Figure 5.3. Let the geodesic arc cut the meridians through these points at angles $\gamma_1, \gamma_2, \dots, \gamma_{n-1}$ respectively, then Clairaut's equation gives

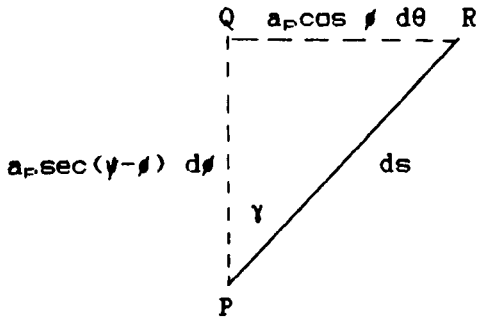
$$r_i \sin \gamma_i = \text{constant} \quad \dots\dots (5.2)$$

where r_i is the perpendicular distance from P_i to the axis of revolution.

Let P be a point on the geodesic arc joining points P_0 and P_n on the surface of a spheroid. The radius of the spheroid at point P is given by

$$a_P = a \sqrt{\frac{(1 - e^2)}{(1 - e^2 \cos^2 \phi)}} \quad \dots\dots (5.3)$$

Let the geodesic arc through P cut the meridian at an angle γ . Let ds be the differential element of the geodesic at P. If we consider the differential triangle PQR (Figure 5.4) - PQ is along the tangent to the meridian and PR is along the tangent to the geodesic - we have



$$PQ = a_p \sec(\gamma - \phi) d\phi$$

$$PR = ds$$

$$QR = a_p \cos \phi d\theta$$

FIGURE 5.4

Since, on the surface of a spheroid, the perpendicular distance, r , of the point P from the axis of revolution is given by $r = a_p \cos \phi$ then, from triangle PQR, we see that

$$\sin \gamma = \frac{a_p \cos \phi d\theta}{ds}$$

and so

$$\boxed{r \sin \gamma = a_p^2 \cos^2 \phi \left(\frac{d\theta}{ds} \right) = \text{constant} = c, \text{ say,}} \quad \dots (5.4)$$

This equation (5.4) is the fundamental equation which we will use for the geodesic on the surface of the spheroid and, in the following sections, we will transform this equation to find a solution which will give us the relationship between the geocentric latitude, ϕ , and the longitude, θ , at any point along the path of the geodesic.

5.3 THE EQUATION OF THE CYLINDRICAL PROJECTION OF A GEODESIC ON THE SURFACE OF A SPHEROID.

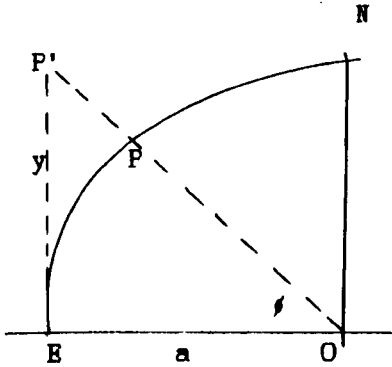


FIGURE 5.5

From the centre, O, of the spheroid, let us project the surface of the spheroid onto the surface of a coaxial cylinder of the same radius using the simple cylindrical projection (not the Mercator projection) so that the point P on the surface of the spheroid corresponds to the point P' on the cylinder. See Figure 5.5.

The points O, P and P' lie on a straight line and y is the perpendicular distance EP' from the Equator to the point P'.

Hence $y = a \tan \phi$ and $d\phi = \frac{a}{a^2 + y^2} dy$

while θ remains unchanged.

Using these substitutions in Equation (5.4) we show, in Appendix 1, that this leads to the form

$$\boxed{dy^2 = \frac{[a^2(1-e^2) + y^2]}{[a^2(1-e^2)^2 + y^2]} (y_{\checkmark}^2 - y^2) d\theta^2} \dots\dots (5.5)$$

where $y_{\checkmark}^2 = \frac{a^2(1-e^2)(a^2 - c^2)}{c^2} \dots\dots (5.6)$

and c is the constant in Equation (5.4). The details of the algebra involved in the transformation of Equation (5.4) into (5.5) are quite lengthy.

Clearly $a_p \leq a$ and $\left| \frac{d\theta}{ds} \right| < 1$ so that $a^2 > c^2$ and hence $y_v^2 > 0$ and y_v is real.

Note also that when $y = y_v$ then $\frac{dy}{d\theta} = 0$ and that the point (θ_v, y_v) on the cylindrical projection of the geodesic arc is a turning point and so this corresponds to a vertex of the geodesic itself.

If we let V denote the point on the geodesic where $y = y_v$ at $\theta = \theta_v$ and $\phi = \phi_v$ then V is the vertex of the geodesic - the point at which the geodesic approaches nearest to the pole. From Equation (5.4) we see that, since $\gamma = 90^\circ$ at the vertex, then

$$c = r_v = a_v \cos \phi_v \quad \dots \dots (5.7)$$

where a_v is the radius of the spheroid at the vertex.

Taking positive square roots and separating the variables in equation (5.5) and then integrating, we find

$$\theta' - \theta_E = \int_0^{y'} \frac{\sqrt{[a^2(1-e^2)^2 + y^2]}}{\sqrt{[a^2(1-e^2) + y^2]}} \frac{dy}{\sqrt{(y_v^2 - y^2)}} \quad \dots (5.8)a$$

where y' lies in the interval $[0, y_v]$, θ' is the corresponding value of the longitude and θ_E is the value of the longitude when $y = 0$ (at the point P_E on the geodesic - the point where the geodesic crosses the Equator).

If
$$f(y) = \frac{\sqrt{[a^2(1-e^2)^2 + y^2]}}{\sqrt{[a^2(1-e^2) + y^2]}}$$

then, in the interval $[0, y_v]$, $f(y)$ is continuous and non-negative so that we can apply the Second Mean Value Theorem for Integrals through which, for any interval $[0, y']$ contained in $[0, y_v]$, there exists $\chi_y : 0 \leq \chi_y \leq y'$ such that

$$\theta' - \theta_E = f(\chi_v) \int_0^{y'} \frac{dy}{\sqrt{(y_v^2 - y^2)}} \quad \dots\dots (5.9)$$

From equations (5.8) & (5.9) we find

$$\int_0^y \frac{\sqrt{[a^2(1-e^2)^2 + t^2]}}{\sqrt{[a^2(1-e^2) + t^2]}} \frac{dt}{\sqrt{(y_v^2 - t^2)}} = f(\chi_v) \int_0^y \frac{dt}{\sqrt{(y_v^2 - t^2)}}$$

Substitute $t = y_v \sin u$ on the left hand side and simplifying the right hand side gives

$$\int_0^u f(y_v \sin u) du = f(\chi_v) \sin^{-1} \left(\frac{y}{y_v} \right)$$

from which we find

$$f(\chi_v) = \frac{1}{u} \int_0^u f(y_v \sin u) du$$

If we write $f(\chi_v) = \frac{1}{\lambda(\chi_v)}$ and $y = a \tan \phi$

then the solution of the differential equation (5.4) may be written

$\tan \phi = \tan \phi_v \sin [\lambda(\chi_v) \cdot (\theta - \theta_E)]$	$\dots (5.10)a$
--	-----------------

where $\frac{1}{\lambda(\chi_v)} = \frac{1}{u} \int_0^u f(y_v \sin u) du$

and $u = \sin^{-1} \left(\frac{y}{y_v} \right)$

Alternatively, as a solution of equation (5.5) we find

$$\theta_v - \theta' = \int_0^{y_v} \frac{\sqrt{[a^2(1-e^2)^2 + y^2]}}{y \sqrt{[a^2(1-e^2) + y^2]}} \frac{dy}{\sqrt{(y_v^2 - y^2)}} \quad \dots (5.8)b$$

where we now find that θ_v is the value of θ at the vertex of the geodesic.

Applying the second mean value theorem to equation (5.8)b there exists a value χ'_v in the interval $[y', y_v]$ such that

$$\theta_v - \theta' = f(\chi'_v) \int_{y'}^{y_v} \frac{dy}{\sqrt{(y_v^2 - y^2)}}$$

making the substitution $y = y_v \cos u$ and writing $y = a \tan \phi$ then we find that the solution of equation (5.4) may be written in the alternative form

$\tan \phi = \tan \phi_v \cos[\mu(\chi'_v) \cdot (\theta_v - \theta)]$ (5.10)b
--	--------------

where
$$\mu(\chi'_v) = \frac{1}{u} \int_0^u f(y_v \cos u) du$$

and
$$u = \sin^{-1} \left(\frac{y}{y_v} \right)$$

5.4 THE PERIOD OF THE GEODESIC ARC.

Through either of equations (5.10) we have thus defined the position of a point on the path of the geodesic curve and the general nature of this equation shows that the projected path of the geodesic follows a sinusoidal curve whose period is less than 360° (since λ is always just a little greater than unity) and whose amplitude is y_v . We can determine the period of the geodesic from equation (5.8). Let the value of the longitude at the vertex of the geodesic be θ_v . We know then that $\theta_v - \theta_e$ is one quarter of the sinusoidal period. If we use the substitution $y = y_v \sin u$ on the right hand side of equation (5.8) we then find that the period of the geodesic which reaches its vertex in latitude ϕ_v is

$$4 \int_0^{\pi/2} \sqrt{\frac{a^2(1-e^2)^2 + y_v^2 \sin^2 u}{a^2(1-e^2) + y_v^2 \sin^2 u}} du$$

In Figure 5.6 we show the path of the cylindrical projection of the geodesic arc over two cycles. This shows that the geodesic arc on the surface of a spheroid which is not a meridian or the Equator winds around the spheroid and does not meet itself again unless the value of $1/(1-\lambda_v)$ is an integer.

We will find it convenient to use the value of the half period of the geodesic which we will denote by θ_p . θ_p is therefore given by

$$\theta_p = \frac{\pi}{\lambda_v} = 2 \int_0^{\pi/2} \sqrt{\frac{a^2(1-e^2)^2 + y_v^2 \sin^2 u}{a^2(1-e^2) + y_v^2 \sin^2 u}} du \quad \dots (5.11)$$

where λ_v is the value of $\lambda(\chi_v)$ when $y=y_v$.

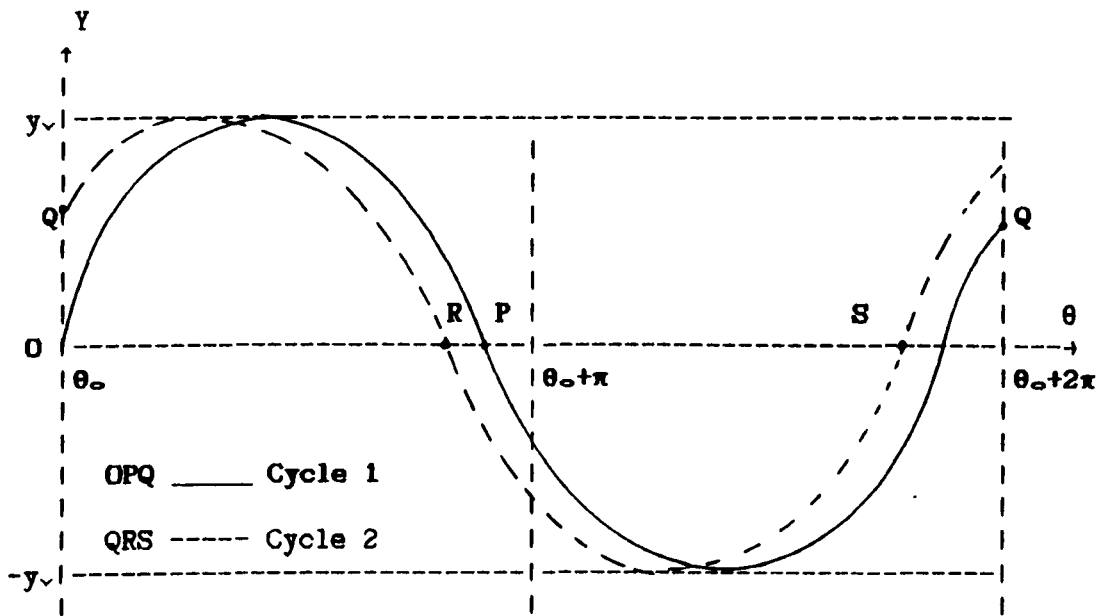


FIGURE 5.6 - PROJECTION OF A GEODESIC CURVE ON THE SURFACE OF A SPHEROID ONTO A COAXIAL CYLINDER OF EQUAL DIAMETER.

5.5 THE DISTANCE ALONG THE GEODESIC ARC.

Let s be the distance along the geodesic arc from the point P_o (θ_o, ϕ_o) to the point P_n (θ_n, ϕ_n) where the difference of longitude between the two points is less than 180° and such that the two points cannot be considered to be "Nearly Antipodean" . Since with equation (5.7) we have determined the value of the constant in equation (5.4) we can rewrite equation (5.4) as

$$a_p^2 \cos^2 \phi \left(\frac{d\theta}{ds} \right) = a_v \cos \phi_v$$

Separating the variables and integrating then gives

$s = \int_{\theta_o}^{\theta_n} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta$ (5.12)
--	--------------

Since we have $y = a \tan \phi$ we can use equation (5.10) to express ϕ as a function of θ by

$$\tan \phi = \tan \phi_v \sin [\lambda(\theta - \theta_o)] \quad \text{..... (5.13)}$$

where $\frac{1}{\lambda} = \frac{1}{u} \int_0^u f(y_v \sin u) du$

and $u = \sin^{-1} \frac{y}{y_v}$.

In navigation the popular existing methods of computing the shortest distance on the surface of a spheroid which seem to be most widely referred to are due to Andoyer²⁸ and Lambert²⁹. Both of these authors use a method which involves applying a correction to the great circle distance on the corresponding sphere and both methods can be handled by manual computation. T. Hairawa²³ also applied the same method in a recent publication. Here, however, we use a Direct Method applying the equations that we have derived above and use iterative procedures where, when we are given the longitude, θ , we compute ϕ , y and then s . This is described in detail in section 5.7.

5.6 THE COURSE ALONG THE PATH OF THE GEODESIC.

It has already been shown as a direct consequence of Clairaut's equation that, if the geodesic cuts the meridian where the latitude is ϕ at an angle γ , then we have

$$a_p \cos \phi \sin \gamma = a_v \cos \phi_v$$

The angle γ is the AZIMUTH and the course in the 360° notation is derived from it.

5.7 COMPUTATIONAL PROCEDURES.

To begin the computation of the distance, s , along the path of the geodesic arc between the two points $P_o (\phi_o, \theta_o)$ and $P_n (\phi_n, \theta_n)$ we first need to know the position $P_E (0, \theta_E)$, where the geodesic arc between the points (or its extension) crosses the Equator, and $P_v (\phi_v, \theta_v)$, the position of the vertex. We will also need approximations for the values of λ_o and λ_n . These are the values of λ when $\phi = \phi_o$ and $\phi = \phi_n$, respectively. An approximation for λ_1 when $\phi = \phi_1$ can be determined from the application of numerical methods to the scheme :

$$y_1 = a \tan \phi_1 \quad (y_v = a \tan \phi_v)$$

$$u_1 = \sin^{-1} \frac{y_1}{y_v}$$

$$\frac{1}{\lambda_1} = \frac{1}{u_1} \int_0^{u_1} f(y_v \sin u) du$$

From equation (5.9) we find, at P_o and P_n , respectively,

$$y_v \sin[\lambda_o(\theta_o - \theta_E)] - y_o = 0 \quad \dots\dots (5.15)$$

and $y_v \sin[\lambda_n(\theta_n - \theta_E)] - y_n = 0 \quad \dots\dots (5.16)$

We can solve these equations (5.15) and (5.16) simultaneously for y_v and θ_E using the two dimensional form of Newton's method.

If we compute the values of θ_E and y_v which would be found from equations (4.7) and (4.8) of Chapter 4 by considering, as a first approximation, that the Earth is a sphere, then we can use the values of y_v and θ_E so found as the first approximation to the solution of equations (5.15) and (5.16) by Newton's Method. During the iterative procedure we must also compute values of λ_o and λ_n corresponding to y_v .

Having found satisfactory values of θ_E and y_v we can then find ϕ_v :

$$\phi_v = \tan^{-1} \left(\frac{y_v}{a} \right)$$

It may be more convenient to find the value of θ_v rather than θ_E in which case we find the simultaneous solutions of the equations

$$y_v \cos[\mu_o(\theta_v - \theta_o)] - y_o = 0 \quad \dots\dots (5.17)$$

$$y_v \cos[\mu_n(\theta_v - \theta_n)] - y_n = 0 \quad \dots\dots (5.18)$$

To start the actual computation of the distance, s , along the geodesic arc we must first subdivide the longitude interval $[\theta_o, \theta_n]$ in some manner; we introduce convenient intermediate points $\theta_1, \theta_2, \dots, \theta_{n-1}$ and find the corresponding latitudes $\phi_1, \phi_2, \dots, \phi_{n-1}$.

To do this we set up an iterative scheme with first approximations found from the great circle on the corresponding sphere.

Given θ_1, y_v and θ_E we can find the $(y_i)_o$ values (the initial approximation to y_i) from equation 4.6 :

$$(y_i)_o = y_v \sin (\theta_i - \theta_E) \quad \text{for } i=1,2,\dots,n-1 .$$

The iteration steps then involve repeatedly computing $\{\lambda_i\}_j$ and $\{y_i\}_j$ using

$$\{u_i\}_{j-1} = \sin^{-1} \frac{\{y_i\}_{j-1}}{y_v}$$

$$\frac{1}{\{\lambda_i\}_j} = \frac{1}{\{u_i\}_{j-1}} \int_0^{\{u_i\}_{j-1}} f(y_v \sin u) du$$

and $\{y_i\}_j = y_v \sin \{\lambda_i\}_j (\theta_i - \theta_E)$

for $j=1,2,\dots$ until convergence is deemed to have occurred when

$$|\{y_i\}_j - \{y_i\}_{j-1}| < \epsilon$$

for each i and for some preassigned tolerance $\epsilon > 0$.

Once the iteration has converged to give values for λ_i and y_i then ϕ_i can be calculated from

$$\phi_i = \tan^{-1} \left(\frac{y_i}{a} \right)$$

Given now the values of the geocentric latitudes $\{\phi_i\}$ ($i=0,1,\dots,n$) corresponding to the longitudes $\{\theta_i\}$, we can compute the ordinate values for the integrand in equation (5.12). We can, for convenience, rewrite the equation in the form

$$s = \int_{\theta_E}^{\theta_n} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta - \int_{\theta_E}^{\theta_o} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta$$

if $\theta_E \leq \theta_o \leq \theta_n$

or in the form

$$s = \int_{\theta_o}^{\theta_E} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta + \int_{\theta_E}^{\theta_n} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta$$

if $\theta_o \leq \theta_E \leq \theta_n$.

This enables us to take advantage of the fact that the derivative of the integrand is zero when $\theta = \theta_E$ which is helpful in the application of the step by step method of the Direct Cubic Spline.

In most problems in *sea-going* navigation where ships are usually confined to what are termed the "Navigable Latitudes" and to the main ocean basins the method of determining the distance as described above is generally applicable. There are, however, circumstances, in theory, when the method will become ill-conditioned. Consider, for instance, the case when the the vertex of the geodesic is in high latitude and when, as a consequence, in low latitude, the initial azimuth of the geodesic path is very small i.e. when the angle, γ , between the meridian and the geodesic arc is small. There is then a large change in latitude for a small change in longitude and it becomes necessary to transform equation (5.12) to allow for this.

If we differentiate with respect to $d\theta$ in equation (5.12) :

$$\frac{ds}{d\theta} = \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v}$$

and, from equation (5.8) we find

$$\theta - \theta_e = \int_0^{\phi} \frac{f(a \tan u) \sec^2 u}{\sqrt{(\tan^2 \phi_v - \tan^2 u)}} du$$

whence
$$\frac{d\theta}{d\phi} = \frac{f(a \tan \phi) \sec^2 \phi}{\sqrt{(\tan^2 \phi_v - \tan^2 \phi)}}$$

Taking these together we find

$$\frac{ds}{d\phi} = \frac{ds}{d\theta} \cdot \frac{d\theta}{d\phi} = \frac{f(a \tan \phi) a_p^2}{a_v \cos \phi_v \sqrt{(\tan^2 \phi_v - \tan^2 \phi)}}$$

The distance, s' , along the arc of the geodesic in an interval $[\phi_0, \phi']$ contained in the interval $[\phi_0, \phi_v]$ is then

$s' = \int_{\phi_0}^{\phi'} \frac{f(a \tan \phi) a_p^2}{a_v \cos \phi_v \sqrt{(\tan^2 \phi_v - \tan^2 \phi)}} d\phi$... (5.18)
--	------------

The reason for the restriction is that there is a singularity in the integrand in equation (5.18) when $\phi = \phi_v$ at the point where the geodesic arc reaches its vertex.

It is only necessary to use equation (5.18) when the azimuth angle, γ , is small and there is a large change in the latitude over a comparatively small interval in the longitude. As the azimuth angle increases along the path of the geodesic then, clearly, we can revert to using equation (5.12) for computing the distance over the part of the geodesic curve which includes the vertex before the geodesic arc reaches its vertex where $\gamma = 90^\circ$. In practice the point at which we revert to using equation (5.12) can be decided on the value of γ . It might be logical, for example, to choose $\gamma = 45^\circ$ but, in most circumstances, this procedure is not necessary and will only be required in those cases where the vertex of the geodesic is in high latitude. In the next chapters where we discuss the computation of the shortest distance between Nearly Antipodean points, where the geodesic paths do reach high latitude, then we will need to use equation (5.18) to compute the distance along the relevant arc.

5.6 THE CORRECTION METHOD OF LAMBERT.

Let us consider two points X and Y on the surface of the spheroidal Earth Model whose geodetic (astronomical) latitudes are ψ_0 and ψ_n , respectively. Let PXY be the spherical triangle on the corresponding sphere (Figure 5.7) .

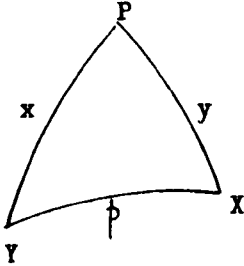


FIGURE 5.7

In Figure 5.7 P is the pole and p, x, y are the sides of the triangle PXY expressed as the angles subtended at the centre of the sphere by the arcs XY, PY and PX, respectively. Having computed s, the great circle distance in geographical miles from X to Y then the correction, Δs , to apply to this great circle distance to give us the distance along the geodesic arc joining the points X and Y on the surface of the spheroid is, according to Lambert²⁹,

$$\Delta s = \frac{1}{2}af(3 \sin p - p) \frac{\sin^2 \psi \cos^2 \Delta \psi}{\cos^2 \frac{1}{2}p} - \frac{1}{2}af(3 \sin p + p) \frac{\cos^2 \psi \sin^2 \Delta \psi}{\sin^2 \frac{1}{2}p}$$

where $f [= 1 - \sqrt{(1-e^2)}]$ is the flattening of the spheroid and

where $\psi = \frac{1}{2}(\psi_0 + \psi_n)$ and $\Delta \psi = \frac{1}{2}(\psi_0 - \psi_n)$.

5.7 EXPERIMENTAL RESULTS.

As an example we will compute the path of the geodesic arc from a position 51°46'N 55°22'W , off Belle Island in Newfoundland, to a position 55°32'N 7°14'W , off Inistrahull in Ireland. Treating the Earth as a sphere and using the great circle method the distance is found to be 1691.61 geographical miles but, along the shortest path, treating the Earth as regular spheroid, we find that the distance is 1695.24 geographical miles . Table 5.1 below gives the details of the spheroidal computation. The way points are chosen to include the vertex of the geodesic (which is outlined in the table by the dotted lines) and then points along the path so that the longitude interval from the initial position to the vertex is subdivided by evenly spaced points and then the longitude interval from the vertex to the final position subdivided by evenly spaced points also.

WAY POINT	LONGITUDE	LATITUDE		DISTANCE	AZIMUTH (COURSE)
		GEOCENTRIC ϕ	GEODETTIC ψ		
0	55°22.00'	51°34.80'	51°46.00'	-	063.14
1	49°53.96'	53° 6.98'	53°18.02'	220.052	067.49
2	44°25.93'	54°18.87'	54°29.76'	426.426	071.91
3	38°57.89'	55°12.67'	55°23.44'	622.813	076.38
4	33°29.86'	55°50.01'	56° 0.69'	811.719	080.90
5	28° 1.82'	56°11.98'	56°22.61'	995.336	085.44
6	22°33.77'	56°19.23'	56°29.85'	1177.787	090.00
7	17°27.18'	56°12.90'	56°23.53'	1347.730	094.26
8	12°20.41'	55°53.73'	56° 4.41'	1519.554	098.51
9	7°14.00'	55°21.25'	55°32.00'	1695.240	102.73

TABLE 1 - PATH OF GEODESIC FROM BELLE ISLAND TO INISTRHULL.

It was considered of interest to perform the same experiments as were done by T.Hairawa²³ since some of this work written here was presented in a paper to the Journal of Navigation by the author with J.E.Phythian²⁰ as a response to the paper by Hairawa in the same Journal . In his paper Hairawa drew comparisons between the methods of computing distance on the surfaces of the sphere and the spheroid along the arcs of loxodromic curves and geodesics. In the case of the geodesics he used the correction method of Lambert. Our

computations are based on the same spheroid data used also by Hairawa (Bessel's Spheroid : $e = 0.081697$, $a = 3437.7468$). The method presented by us is referred to as the "Direct Method" and the method used by Hairawa as the "Correction Method" . The experiment consisted of computing the shortest distance between pairs of points which are 100° of longitude apart and lying on the same parallel of latitude and letting the latitude vary from 10° to 80° . The results are shown in Table 5.2 below.

TABLE 5.2

GEOGRAPHICAL LATITUDE	D_1	D_2	D_3	D_4	D_5
10°	5876.82	5876.83	5877.3	5877.33	5877.33
20°	5525.02	5525.02	5526.9	5526.96	5526.96
30°	4987.29	4987.29	4991.2	4991.21	4990.21
40°	4311.84	4311.84	4317.6	4317.61	4317.61
50°	3539.84	3539.86	3546.7	3546.71	3546.70
60°	2702.52	2702.55	2709.3	2709.29	2709.26
70°	1822.67	1822.71	1828.1	1828.08	1828.04
80°	917.31	917.34	920.3	920.32	920.28

In Table 5.2 :

D_1 is the Great Circle Distance on the surface of the corresponding sphere as computed using the Cosine Formula.

D_2 is the Great Circle Distance on the surface of the corresponding sphere as computed by the Direct Method with $e=0$ and intermediate points taken along the Great Circle at 5° intervals in the longitude.

D_3 is the shortest distance on the surface of the spheroid as found by Hairawa³ with the Correction Method.

D_4 is the shortest distance on the surface of the spheroid as found by the Direct Method with points intermediate along the geodesic arc taken at intervals of 5° in the longitude.

D_s is the shortest distance on the surface of the spheroid as found by the Direct method with intermediate points along the geodesic arc taken at intervals of 1° in the longitude.

The distances computed by Hairawa were probably taken between points on the same parallel of latitude because, in the Correction Method, the formula takes a particularly simple form in this instance. The latitudes of the two end points of the geodesic arc appear in the formula symmetrically and we also find that the best results are obtained from the correction method when these latitudes have the same absolute value.

In the example of the distance from the position off Belle Island in Newfoundland, to a position off Inistrahull using the Direct Method the shortest distance, found taking intermediate points along the path at intervals of approximately equal to 5° in the longitude, is found to be 1695.24 geographical miles - a difference of 3.63 geographical miles from the great circle method. Lambert's method in this case gives a distance of 1695.25 geographical miles.

In response to the paper presented to the Journal of Navigation by Williams & Phythian³⁰ a paper was published by Roger Bourbon³¹ describing an improvement to the Correction Method of Lambert and involving further iterations. He used the example of the shortest distance between the points of Belle Island and Inistrahull and gave this distance as 1695.27 geographical miles.

5.8 REMARKS ON THE EXPERIMENTAL RESULTS.

The Mathematical problem of describing the geodesic paths on the surface of a spheroid by analytical equations is a classical one, first solved by Clairaut and given in the book by Todhunter³². However, it has never been easy to give a numerical solution to Clairaut's Equation (5.4) for all cases. Most solutions use iterative methods and this has required the use of computers. In the past, therefore, alternative solutions, such as the Correction

Method, have had to suffice. By modern standards the Great Circle solution is not really good enough and, since electronic computing devices are now widely used, it is possible now to obtain accurate numerical solutions. There has also been the danger, as pointed out by Hairawa²³ that, in seagoing navigation, it is a practice to use formulae which are not correct theoretically. There is really no need to continue this practice now that algorithms are available which can compute distances to very acceptable accuracy. Indeed, as was mentioned in the introduction to the thesis, much has been published on the single problem of computing the distance along the geodesic arcs of a spheroid and claims that it can be done correct to a millimetre for a regular spheroid are justified. For navigational purposes, however, this level of accuracy is not strictly necessary and we have concentrated more on the computation of the positions along the arc of the geodesic which requires a method such as the direct method as described in this chapter.

For the positions chosen by Hairawa the results obtained from the Direct Method and the Correction Method of Lambert compare favourably. They both differ from the Great Circle Distance and, where they differ from each other, that difference is small compared to the difference each has with the Great Circle Distance.

No error analysis has been done here on the Correction Method. As to the Direct Method, we would expect that the main source of error would be due to round off, particularly in the approximation to the value of λ . The error due to round off which affects D_4 in Table 5.1 can be estimated by the difference between D_1 and D_2 which at its maximum is 0.04 geographical miles. We are confident that, by iteration, the error in the final result due to the value of λ is reduced to a negligible amount. It is our belief, then, that the distance D_4 in Table 5.2 is as good an estimate to the shortest distance between the specified points on the surface of the spheroid used as the approximation to the shape of the Earth that we need to make.

When second approximations are taken the correction methods of computing the length of arc along a geodesic curve on the surface of the *terrestrial* spheroid are quite adequate for the purpose but, from this method, we do not gain any information about the *path* that the geodesic takes. This is important to the navigator and this is the reason that we have chosen to use this direct method - so that by applying a step by step method of solution of Clairaut's equation we can determine intermediate points along the path, the distance between them and the angle at which the geodesic curve cuts the meridians at the intermediate points. The path of the geodesic curve can then be plotted on a Mercator chart in the traditional manner. Automatic systems for voyage planning also use intermediate points along the projected path which are called "Way Points". The positions of these "way points" have to be determined and it is always better for the navigator to achieve the "way points" along the path as near as possible.

-6-

THE SHORTEST DISTANCE BETWEEN
NEARLY ANTIPODEAN POINTS
ON THE EQUATOR

6.1. INTRODUCTION.

There is a special problem in finding the shortest path geodesic arc that joins two Nearly Antipodean points on the surface of the spheroidal Earth. Nearly Antipodean Points are points which are *almost* 180° apart in longitude but which have the same latitude in opposite hemispheres. The specific longitude at which this problem occurs is defined below. It turns out that, in this case, the shortest path can deviate quite considerably from the path predicted by the method described in Chapter 5 and from the great circle path on the corresponding sphere. When points are exactly 180° apart in longitude then the shortest path between them is along the meridian. Lambert²⁹, in his paper entitled "THE DISTANCE BETWEEN TWO WIDELY SEPARATED POINTS ON THE SURFACE OF THE EARTH", published in 1942, having given a solution which was applicable in *most* cases, stated that there was a particular difficulty in using the Correction Method to determine the shortest distance between two *nearly* antipodean points on the surface of the spheroidal Earth but this was not fully analysed in that paper. The problem has been solved by several authors since and one of the more recent is Bowring⁷. The method described in Chapter 5 in this thesis computes the shortest distance between two points on the spheroidal Earth model but this distance is along a geodesic arc between the points which closely follows the great circle arc on the corresponding sphere. Indeed the method here in Chapter 5 uses the great circle values as initial approximations in subsequent iterative procedures and, for similar reasons, whilst the method gives the shortest distance for most cases, it does not give the shortest distance for the case when two points are nearly antipodean. We will now adapt our procedure to correct this.

The theory of this subject has been discussed before by Helmert³³ and Fichot³⁴ (among others) but it seems that there has been no such full exposition in English. Lambert, in his paper, stated that it was his intention to publish further work on this problem, in English, but, to our knowledge, this has not appeared and, as far

as we can see, no work similar to that of Fichot by another author has appeared either. We will recapitulate the statement of the problem posed by Lambert and present our solution .

6.2 NEARLY ANTIPODEAN POINTS ON THE EQUATOR.

Consider first, as Lambert did, two points which both lie on the Equator. Although the Equator is a geodesic curve the shortest distance between two points on the Equator is not, as one might imagine, always along the Equator. When the difference of longitude between them is greater than a certain fixed value (which according to Lambert, is approximately 179°24') then the shortest path between them is along another geodesic arc which takes a northerly (or southerly) route in order to take advantage of the flattening of the spheroid.

At each point on the surface of a spheroid a family of geodesic curves is defined and these geodesic curves emanate in all directions from the point. From the poles these families are the meridians but from all other points the family is defined by Clairaut's Equation (Lyusternik²⁶) written in the form

$$\boxed{a_P \cos \phi \sin \gamma = a_V \cos \phi_V} \quad \dots\dots (6.1)$$

where, at a point P along a geodesic arc, a_P is the radius of the spheroid, ϕ is the geocentric latitude and γ is the angle between the geodesic and the meridian through P. This equation (6.1) was derived in Chapter 5. A particular member of the family is specified by ϕ_V (the geocentric latitude of the vertex, V, of the geodesic curve) and each member of the family has a slightly different sinusoidal period. The sinusoidal period of a geodesic is related to the geocentric latitude of its vertex so that the higher the latitude of the vertex the greater is the sinusoidal period. The half period of a geodesic is the difference of longitude between two successive passes of the geodesic through the Equator. The half

period, θ_p , of a geodesic which reaches its vertex in latitude ϕ_v is given by equation (5.11).

Let us consider the family of geodesic arcs which start at the point with latitude 0° and longitude 0° , which have their end points on the Equator also and which have their vertices in the Northern Hemisphere. See Figure 6.1. The family of geodesics in the Northern Hemisphere are shown by the continuous arcs in the figure. To each member of the family, starting at the same point, there corresponds a symmetrical geodesic of equal length in the Southern Hemisphere and terminating at the same point also. The members of this family of geodesics in the Southern Hemisphere are marked by the broken lines in Figure 6.1.

The longitudes of the end points of the geodesics all fall in the interval $\theta_L \leq \theta \leq 180^\circ$, where θ_L is Lambert's value ($\approx 179^\circ 24'$ which is given as $2\pi \sqrt{1-e^2}$ in chapter 5) but the longitudes of these end points are different for each member of the family. The extreme members of the family are that part of the Equator itself between longitudes 0° and θ_L (vertex 0° , half period θ_L) and the meridian with end points in longitudes 0° and 180° (vertex 90° , half period 180°). Note then, that when the difference of longitude, θ_d , falls in the interval $\theta_L < \theta_d \leq 180^\circ$, there are three geodesic arcs joining the two points - the Northerly geodesic, the Southerly geodesic and the Equator - but *only two* of them - the Northerly geodesic and the Southerly geodesic - give us the equal shortest distance. The Equator will give us the shortest path when $\theta_d \leq \theta_L$, for then all three paths will coincide. This means that if the difference of longitude, θ_d , between two points on the Equator is less than or equal to this figure then the shortest path between them is along the Equator. If, however,

$$179^\circ 23.898' \leq \theta_d \leq 180^\circ$$

then the shortest path is along a northerly or southerly route - either are possible and both give the same shortest distance.

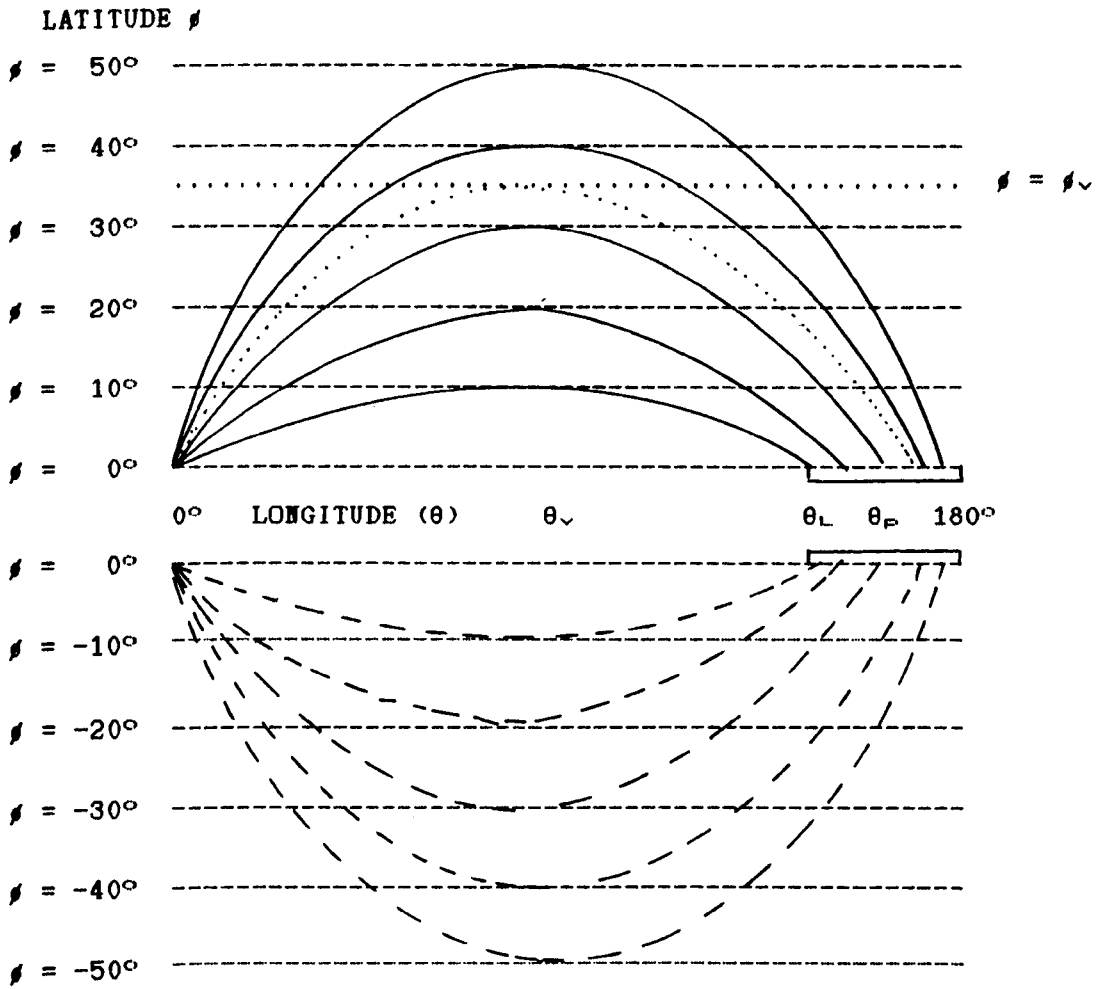


FIGURE 6.1. CYLINDRICAL PROJECTION OF GEODESIC PATHS BETWEEN TWO NEARLY ANTIPODEAN POINTS ON THE EQUATOR.

It was shown Chapter 5 that, along the arc of a geodesic which crosses the Equator in longitude θ_E and reaches its vertex in latitude ϕ_V , longitude θ_V , at any point along the geodesic arc where the latitude is ϕ and the longitude is θ , then ϕ and θ are related by the equations

$$y = a \tan \phi \quad (y_V = a \tan \phi_V)$$

$$\theta - \theta_E = \int_0^y \frac{\sqrt{a^2(1-e^2)^2 + t^2}}{a^2(1-e^2) + t^2} \frac{dt}{\sqrt{y_V^2 - t^2}}$$

whose solution may be expressed in the form

$$\boxed{\tan \phi = \tan \phi_V \sin[\lambda(\theta - \theta_E)]} \quad \dots\dots (6.2)$$

where

$$\frac{1}{\lambda} = \frac{1}{u} \int_0^u f(y_V \sin u) du$$

and

$$u = \sin^{-1} \frac{y}{y_V}$$

This is found from the differential equation

$$a_p^2 \cos^2 \phi \left(\frac{d\theta}{ds} \right) = a_V \cos \phi_V \quad \dots\dots (6.3)$$

which was also derived in Chapter 5 and defines the path of a geodesic on the surface of a spheroid by means of Clairaut's equation expressed as a differential equation.

It was shown in chapter 5 that the half period, θ_P , of the geodesic is given exactly by equation (5.11) which is

$$\boxed{\theta_P = \int_0^{\pi/2} \frac{\sqrt{a^2(1-e^2)^2 + y_V^2 \sin^2 u}}{a^2(1-e^2) + y_V^2 \sin^2 u} du} \quad \dots\dots (6.4)$$

where

$$\pi \sqrt{1-e^2} < \theta_P \leq \pi$$

A typical geodesic whose vertex is at latitude ϕ_v and whose half period is θ_p is marked by the dotted line in Figure 6.1.

Table 6.1, in column 2, shows the value of the half period, θ_p , of the geodesic with its vertex in the latitude given in column 1. This value is the value of the longitude where the geodesic, which started out in latitude 0° , longitude 0° , crosses the Equator again. θ_p is computed from equation (6.4) with $y_v = a \tan \phi_v$, $a = 3437.7468$ geographical miles and $e = 0.081697$.

LATITUDE OF VERTEX (ϕ_v)	HALF PERIOD (θ_p)	DISTANCE P_oP_n ALONG EQUATOR (g.m.)	DISTANCE P_oP_n ALONG GEODESIC (g.m.)	INITIAL AZIMUTH (γ)
10°	179°24.451'	10764.45	10764.44	079.97°
20°	179°26.091'	10766.09	10765.99	069.94°
30°	179°28.769'	10768.77	10768.38	059.92°
40°	179°32.391'	10772.39	10771.32	049.91°
50°	179°36.850'	10776.85	10774.42	039.91°
60°	179°42.007'	10782.01	10777.43	029.92°
70°	179°47.704'	10787.70	10779.77	019.94°
80°	179°53.871'	10793.87	10781.40	009.97°

TABLE 6.1. SHORTEST DISTANCES FOR NEARLY ANTIPODEAN POINTS ON THE EQUATOR

Table 6.1 also shows, in column 3, the distance along the Equator between two points P_o and P_n when the difference of longitude between them is equal to θ_p . This value is simply the angle θ_p expressed in minutes of arc and gives the result in geographical miles. Column 4 of Table 6.1 gives the shortest distance between P_o and P_n . It was shown in Chapter 5 that the distance, s , along the geodesic arc between two points P_o and P_n whose longitudes are θ_o and θ_n , respectively, is

$$s = \int_{\theta_o}^{\theta_n} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta \quad \dots\dots (6.6)$$

where ϕ is expressed in terms of θ by equation (6.2).

Since, in this case, both points lie on the Equator and we have

$$\theta_p = \theta_n - \theta_s$$

or
$$\theta_p = 2(\theta_v - \theta_s)$$

then the distance, s , in Equation (6.6) is, in practice, more conveniently computed from

$$s = 2 \int_{\theta_s}^{\theta_v} \frac{a_p^2 \cos^2 \theta}{a_v \cos \theta_v} d\theta \quad \dots\dots (6.7)$$

Given θ_v from column 1 of Table 6.1, if this $\theta_v \geq 60^\circ$ then we have used a combination of Equations (5.18) and (6.7) to compute the distance, s . We find a value θ' , and a corresponding value θ somewhere near to the point on the geodesic where $\gamma = 45^\circ$ so that

$$s = \int_0^{\theta'} \frac{f(\gamma) a_p^2}{a_v \cos \theta_v \sqrt{(\tan^2 \theta_v - \tan^2 \theta)}} d\theta + \int_{\theta'}^{\theta_v} \frac{a_p^2 \cos^2 \theta}{a_v \cos \theta_v} d\theta \quad \dots\dots (6.8)$$

The procedure for evaluating the integrals in equation (6.8) were described in chapter 5.

The entry in column 5 gives the initial azimuth, γ , of the path of the geodesic from the point on the Equator. This depends also upon the geocentric latitude of the vertex, θ_v , and is given by :

$$\sin \gamma = \frac{a_v}{a} \cos \theta_v \quad \dots\dots (6.5)$$

where a_v is the radius of the spheroid at the vertex.

6.3 GEODESIC CURVES ON THE SURFACES OF OTHER PLANETS.

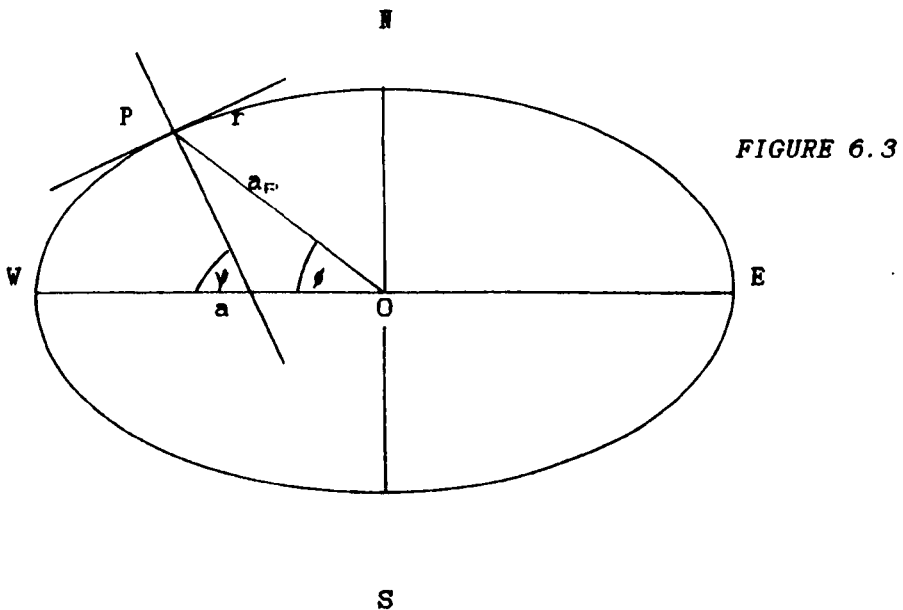
The Earth is usually considered to be "spheroidal" where the term *spheroidal* is used in the sense that the Earth is "sphere-like" because the flattening is small. The inner planets (Mercury, Mars and Venus), like the Earth, are also spheroids since the flattening of these planets is small too but, in the case of some of the outer planets, such as Jupiter, the information so far gathered would suggest that the flattening is much more distinct. Measurements would indicate that the eccentricity of the meridional ellipse of Jupiter is approximately equal to 0.3. The planet Jupiter is therefore more aptly referred to as an *ellipsoid of revolution*. There is no *geometrical* difference between a spheroid and an ellipsoid of revolution and here we will use the terms freely to mean the same thing.

In the case of the Earth and the planets the axis of rotation of the planet is also the axis of revolution about which the ellipsoidal shape of the planet is generated. Indeed, there are good reasons why this should be so - the spheroidal or ellipsoidal shape is the condition of hydrostatic equilibrium of the planet due to its circular motion about its axis. The conditions of hydrostatic equilibrium depend upon its physical constitution and its angular velocity. See, for instance, the book by Samuel Glasstone²⁵.

Although the mathematical analysis of geodesic arcs on the surface of a spheroid as described in Chapter 5 and in the foregoing sections of this chapter have been written with the Earth in mind, where the flattening of the surface is small and the eccentricity of the meridional ellipse is approximately equal to 0.08, it is not difficult to see that it will work equally well for a surface such as Jupiter where the eccentricity of the meridional ellipse is approximately equal to 0.3.

We can set up a coordinate system on the surface of a planet with latitudes and longitude defined in a similar manner to that on the Earth. The Poles of a planet will be those points on the surface

where the axis of rotation cuts the surface and the *North Pole* will be the pole above which an observer, looking down upon the pole, will observe the planet to be rotating anticlockwise. The equator of a planet will be a circle which is the locus of points which are equidistant from both poles. At a point P on the surface of the planet the *geocentric* latitude, ϕ , will be the angle subtended at the centre of the ellipsoidal planet from the Equator to P and the *geodetic* latitude will be the angle between the normal to the surface at P and the Equatorial plane. See Figure 6.3.



The longitude, θ , will be the angle between the plane through a selected meridian where $\theta=0$ and the plane through the meridian at P measured Eastwards. The relationship between the coordinates of a point on a geodesic curve on the surface of an ellipsoid is therefore

$$\tan \phi = \tan \phi_v \sin [\lambda(\theta - \theta_E)]$$

where $\lambda(\chi_v) = \frac{1}{f(\chi_v)}$

as defined in chapter 5, ϕ_v is the latitude of the vertex and θ_E is the longitude where the geodesic crosses the Equator. The half

period of a geodesic curve on the surface of an ellipsoid is given exactly by equation (6.4) .

Table 6.2 shows the values of the half period of a geodesic for given eccentricity of the meridional ellipse of an ellipsoid and at intervals of 20° in the latitude of the vertex. On the surface of a planet in which the eccentricity of the meridional ellipse is 0.25, for instance, the geodesic curve which reaches its vertex at a point where the geocentric latitude is 40° will have a half period of 175°41.7' .

TABLE 6.2 - HALF PERIODS OF GEODESIC CURVES ON THE SURFACES OF ELLIPSOIDS OF REVOLUTION.

e	LATITUDE OF VERTEX OF GEODESIC				
	0°	20°	40°	60°	80°
0.05	179°46.5'	179°47.3'	179°49.7'	179°53.3'	179°57.7'
0.10	179° 5.9'	179° 9.2'	179°18.6'	179°33.1'	179°50.8'
0.15	177°57.8'	178° 5.4'	178°27.0'	178°59.6'	179°39.5'
0.20	176°21.8'	176°35.6'	177°14.6'	178°13.0'	179°23.8'
0.25	174°17.1'	174°39.3'	175°41.7'	177°13.7'	179° 3.9'
0.30	171°42.2'	172°15.9'	173°48.4'	176° 2.2'	178°40.2'
0.35	168°36.9'	169°24.7'	171°34.9'	174°39.1'	178°13.0'
0.40	164°58.4'	166° 4.6'	169° 1.7'	173° 5.4'	177°42.7'

Although they are very close to it, none of the planets are really regular ellipsoids, and, indeed, information about the details of the shapes of the planets (and of the Earth even) is constantly being updated. For Navigational purposes, however, it is probably accurate enough to assume that the Earth is an oblate spheroid whose meridians are ellipses whose eccentricity is approximately equal to 0.082. The eccentricities of the meridians of Mercury, Mars and Venus are considered to be less than this value and the eccentricities of the meridians of Jupiter and Saturn somewhat greater.

-7-

THE SHORTEST DISTANCE BETWEEN
NEARLY ANTOPODEAN POINTS
OFF THE EQUATOR

7.1. NEARLY ANTIPODEAN POINTS OFF THE EQUATOR.

We have analysed the case when two nearly antipodean points both lie on the Equator and a similar situation occurs when two nearly antipodean points have the same latitude in opposite hemispheres. Let the two points be P_o and P_n . If the latitude of one of them is ϕ_o , say, then the latitude of the other point will therefore be $-\phi_o$. As was stated in section 6.2 of Chapter 6, from P_o a family of geodesic curves emanates in all directions and are defined by the equation

$$a_p \cos \phi \sin \gamma = a_v \cos \phi_v$$

but, in this case, $\phi_v > \phi_o$.

Let us consider the geodesic whose vertex latitude, ϕ_v , is equal to ϕ_o . The half period of the geodesic whose vertex is in latitude ϕ_v ($=\phi_o$) will be θ_p as determined from column 2 of Table 6.1 corresponding to the vertex latitude ϕ_v ($=\phi_o$) in column 1. If, then, the difference of longitude between P_o and P_n is less than or equal to θ_p (see Figure 6.3) the shortest distance between them will be along the geodesic arc as computed by the method described in Chapter 5.

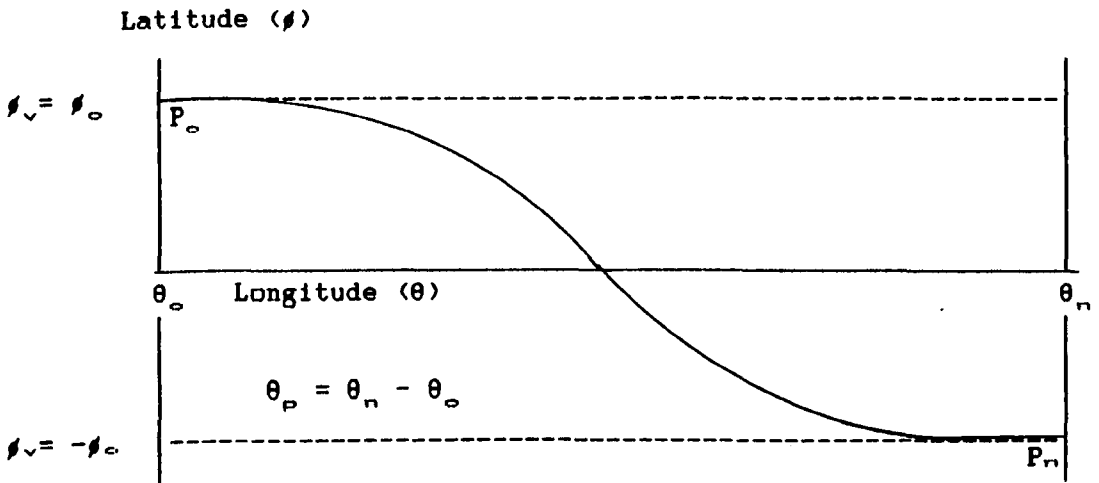


FIGURE 7.1. CYLINDRICAL PROJECTION OF GEODESIC ARC (HALF PERIOD = θ_p)

If, however, $\theta_P < \theta_A \leq \pi$ then the shortest path will be along a different route which takes advantage of the flattening of the spheroid and this path will be along a geodesic arc whose half period will be equal to the difference of longitude between the two points P_o and P_n . See Figure 7.2. Once again it turns out that, when $\theta_P < \theta_A \leq \pi$, there are three geodesic arcs joining two nearly antipodean points *off* the Equator just as there are three geodesic arcs joining two nearly antipodean points *on* the Equator but *only two* of them in each case give the equal shortest paths. Between two nearly antipodean points *off* the Equator there are, therefore, two equal shortest paths along the arcs of geodesics whose half period is equal to θ_A , one of which takes a northerly route through its vertex in the northern hemisphere and the other takes a southerly route through its vertex in the southern hemisphere. There is also a third geodesic, which we will call the *INTERMEDIATE GEODESIC*, whose half period is equal to θ_P , which closely follows what would be the great circle path on the corresponding sphere and which plays a role which would be the equivalent to the Equator in the case of two nearly antipodean points *on* the Equator.

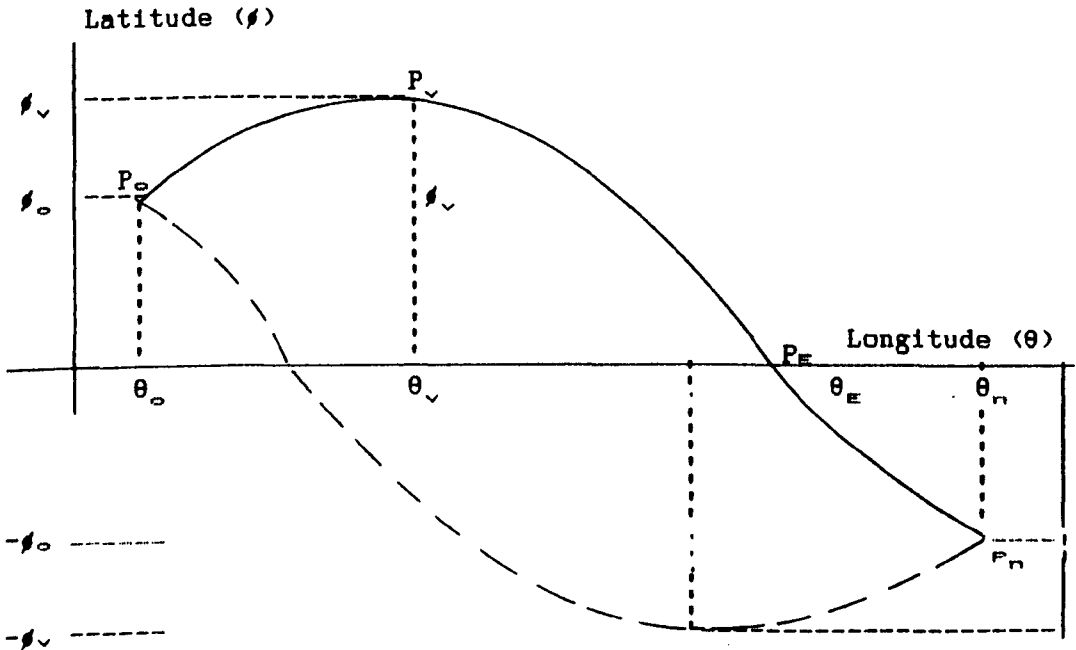


FIGURE 7.2. CYLINDRICAL PROJECTION OF THE GEODESIC PATH BETWEEN TWO NEARLY ANTIPODEAN POINTS OFF THE EQUATOR.

Figure 7.2 shows the cylindrical projection of the three geodesic arcs between two typical nearly antipodean points off the Equator. The continuous line shows the shortest path geodesic which passes through its vertex in the northern hemisphere and the broken line shows the shortest path geodesic which passes through its vertex in the southern hemisphere. The dotted line shows the path of the intermediate geodesic whose half period is equal to θ_p and which has its vertices at geocentric latitudes $\pm\phi_0$. The three geodesics in Figure 7.2 all start at the point P_0 , latitude ϕ_0 , longitude θ_0 , and end at P_n , latitude $-\phi_0$, longitude θ_n . The northerly geodesic reaches its vertex at P_v , latitude ϕ_v , longitude θ_v , and crosses the Equator at P_E where the longitude is θ_E . From the point P_0 there is, in fact, a family of geodesics to points near P_n similar to the set described in Section 6.2 and illustrated in Figure 6.1. This family has one of its end points at P_0 and the other at a point in latitude $-\phi_0$. Once again, the latitude of the vertex of the geodesic is related to its half period and, although all the geodesics leave P_0 in latitude ϕ_0 , longitude θ_0 , when they arrive in latitude $-\phi_0$ they are not all in the same longitude. This longitude will depend upon the half period of the geodesic. See Figure 7.3. The continuous arcs are the shortest path geodesics which pass through their vertices in the northern hemisphere and the broken arcs are the shortest path geodesics which pass through their vertices in the southern hemisphere. The dotted line shows the Intermediate Geodesic for this family. The limiting members of the family are the Intermediate Geodesic (vertex in latitude ϕ_0 , half period θ_p) and the meridian through P_0 (vertex at latitude 90° , half period 180°).

When $\theta_p < \theta_d \leq \pi$, then the absolute value of the geocentric latitude, ϕ_v , of the vertices of the two equal shortest path geodesics between two nearly antipodean points off the Equator can be found by inverse interpolation of Table 6.1, taking θ_d as equal to the half period in column 2. Table 6.1 is, in fact, held on file on our computer at intervals of 1° of the geocentric latitude of the vertex, ϕ_v .

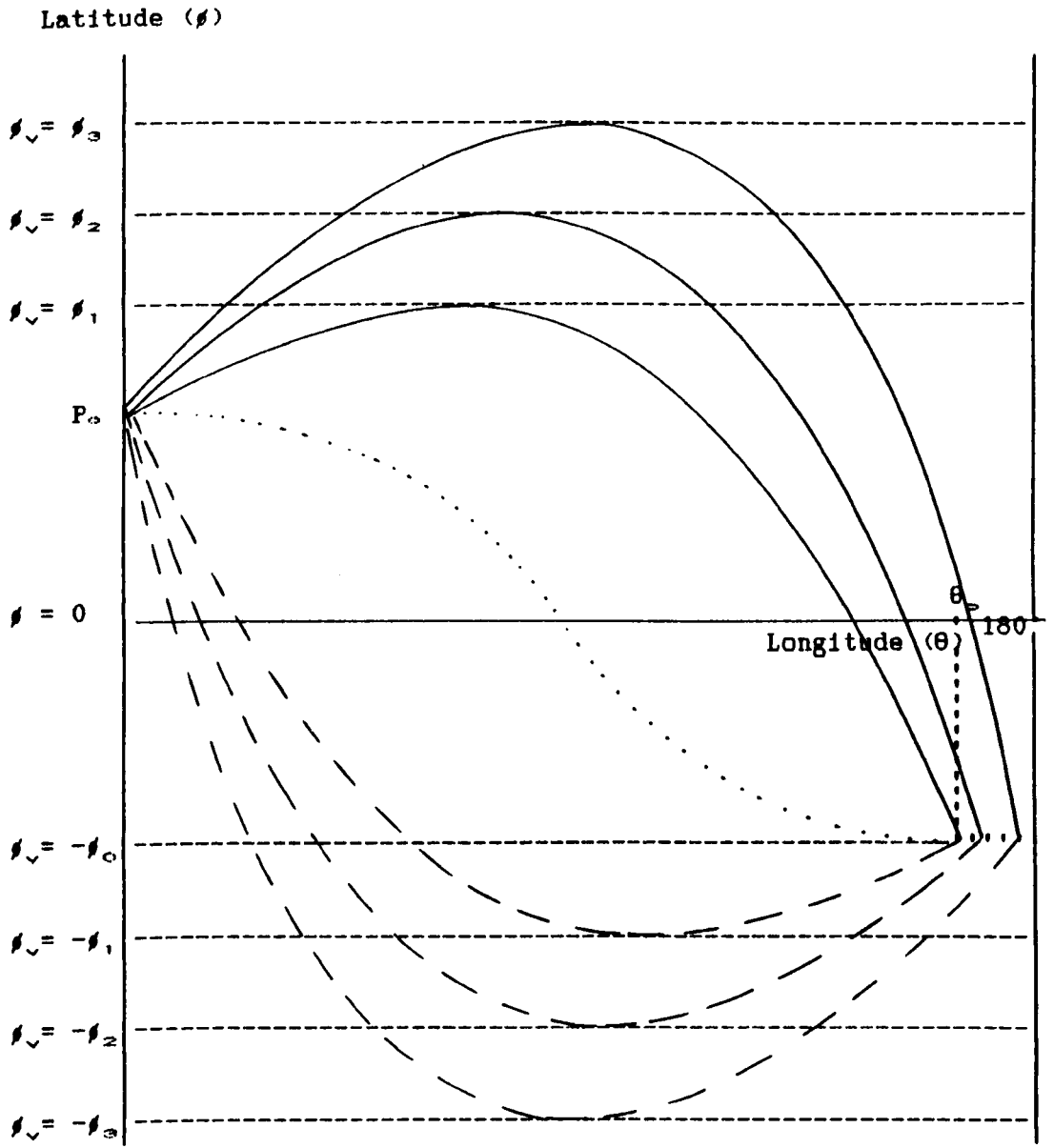


FIGURE 7.3. CYLINDRICAL PROJECTION OF GEODESIC ARCS THROUGH P_0 .

It was shown in Chapter 5 , and in Section 6.2 that, along a geodesic whose vertex lies in latitude ϕ_v which crosses the Equator in longitude θ_E ; the latitude, ϕ , and the longitude, θ , of a point on the geodesic are related by

$$\tan \phi = \tan \phi_v \sin[\lambda(\theta - \theta_E)]$$

where
$$\frac{1}{\lambda} = \frac{1}{u} \int_0^u f(y_v \sin u) du \quad \dots\dots (7.1)$$

$$u = \sin^{-1} \frac{y}{y_v}$$

and
$$y = a \tan \phi$$

Given then, that we have two points : P_o , in latitude ϕ_o , longitude θ_o ; and P_n , in latitude $-\phi_o$, longitude θ_n , and that, from Table 6.1, we determined ϕ_v , the latitude of the vertex of the geodesic whose half period is equal to the difference of longitude, $\theta_d (= \theta_n - \theta_o)$ then we find

$\theta_E = \theta_o - \frac{1}{\lambda_o} \sin^{-1} \frac{y_o}{y_v}$	\dots\dots (7.2)
---	------------------

The shortest path, therefore, between P_o and P_n is along a geodesic arc which reaches its vertex in latitude ϕ_v (which corresponds to a geodesic whose half period is equal to θ_d) and which crosses the Equator in longitude θ_E .

We have determined four points along the path of the geodesic which are

- $P_o \quad (\phi_o, \theta_o)$
- $P_v \quad (\phi_v, \theta_v)$
- $P_E \quad (0, \theta_E)$
- $P_n \quad (-\phi_o, \theta_n)$

See Figure 7.2.

We can compute the distance P_0P_n piecewise along the three separate arcs P_0P_v , P_vP_E , P_EP_n . The distance along the arc P_0P_n is then

$$s = \int_{\theta_0}^{\theta_v} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta + \int_{\theta_v}^{\theta_E} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta + \int_{\theta_E}^{\theta_n} \frac{a_p^2 \cos^2 \phi}{a_v \cos \phi_v} d\theta \quad \dots (7.3)$$

Each integral in (7.3) will be evaluated numerically by the same method that was used in Section 5.6 for the integral in equation (5.12), subdividing the intervals of integration by points where the longitudes are $\theta_i (i=1,2,\dots,n-1)$ and determining the corresponding latitudes $\{\phi_i\}$ using the same iterative procedure.

7.2 AN EXAMPLE OF SHORTEST DISTANCE BETWEEN NEARLY ANTIPODEAN POINTS.

As an example, let us consider the shortest distance from positions off Fremantle in Western Australia (latitude 32°S , longitude $115^\circ 30'\text{E}$ approximately) to a fixed position (latitude 32°N , longitude 64°W) off the island of Bermuda in the Atlantic. It is customary practice for seagoing navigators to choose departure and arrival positions for sea passages and we would not be stretching the credulity of this practice in this case if we decide that both positions should be on parallels where the absolute value of the geocentric latitude is 32° . By the same principle, let us fix the longitude of the arrival position off Bermuda at $64^\circ 00'$ and adjust the departure position off Fremantle to suit our purposes.

Table 7.1 shows, for two points in opposite hemispheres with geocentric latitude $\pm 32^\circ$, the equivalent data as shown in Table 6.1 for two points on the Equator. Given ϕ_v , the latitude of the vertex, we find θ_p from equation (6.6). The distance along the intermediate geodesic is computed by the method of Williams & Phythian³⁰ and is the distance along the path which closely follows what would be the great circle path on the corresponding sphere. The distance along

the shortest path is given by Table 7.1 since the shortest distance along a geodesic path with given vertex is the same for any two points between which the difference of longitude is equal to the half period of the geodesic for any geodesic on the spheroidal Earth.

VERTEX OF GEODESIC (ϕ_v)	HALF PERIOD OF GEODESIC (θ_p)	DISTANCE ALONG INTERMEDIATE GEODESIC	DISTANCE ALONG SHORTEST PATH (s)	INITIAL AZIMUTH (γ_o)
32°	179°29.420'	10768.9	10768.9	090.00°
35°	179°30.468'	10769.7	10769.7	074.97°
40°	179°32.391'	10771.5	10771.3	064.54°
45°	179°34.526'	10773.3	10773.0	056.43°
50°	179°36.850'	10775.1	10774.4	049.22°
60°	179°42.007'	10779.6	10777.4	036.06°

TABLE 7.1 SHORTEST DISTANCES FOR NEARLY ANTIPODEAN POINTS IN LATITUDE $\pm 32^\circ$.

Let us focus our attention on the geodesic path whose vertex lies in latitude 45° . Since the difference of longitude between the departure and arrival positions must be chosen to be equal to the half period of this geodesic and we have fixed the longitude of the arrival position at 64°W , then the longitude of the departure position must be chosen as $115^\circ 34.526'$. Thus the coordinates of P_o are

$$\phi_o = -32^\circ 00' \quad \theta_o = 115^\circ 34.526' .$$

Corresponding to $\phi_o = -32^\circ$ we find $\tan \phi_o = -0.624869352$

and, since $\phi_v = 45^\circ$ then $\tan \phi_v = 1$

so that
$$\frac{y_o}{y_v} = \frac{a \tan \phi_o}{a \tan \phi_v} = \tan \phi_o$$

From equation (7.1) we also find $\lambda_o = 1.003186 .$

Hence, using these values of $\tan \phi_0$ and λ_0 in equation (7.2), we find a value for θ_E , the longitude of the point where the geodesic crosses the Equator :

$$\theta_E = \theta_0 - \frac{1}{\lambda_0} \sin^{-1} \left(\frac{y_0}{y_v} \right)$$

$$\theta_E = 115^\circ 34.526' - \frac{\sin^{-1}(-0.62469352)}{1.003186}$$

or $\theta_E = 115^\circ 34.526' + 38^\circ 32.215'$

i.e. $\theta_E = 154^\circ 6.741' \text{ (E)}$.

This geodesic also crosses the Equator at longitude θ'_E where

$$\theta'_E = \theta_E + \frac{\pi}{\lambda_v}$$

where $\frac{\pi}{\lambda_v}$ is given by equation (5.11) .

Hence $\theta'_E = 154^\circ 6.741' + 179^\circ 34.526'$

$$\theta'_E = 333^\circ 41.267'$$

which is $26^\circ 18.733' \text{ (W)}$.

The longitude of the vertex, θ_v , is given by

$$\theta_v = \theta_E - \frac{\pi}{2\lambda_v}$$

$$\theta_v = 154^\circ 6.741' - 89^\circ 47.263'$$

$$\theta_v = 64^\circ 19.478' \text{ (E)}$$
 .

Computing the distance along the intermediate geodesic path by the method of Williams & Phythian³⁰ gives 10773.3 geographical miles. The intermediate path is the dotted line in Figure 7.3 - it crosses the Equator, halfway between the departure and arrival points, somewhere on the African Continent and has its vertices at latitude $\pm 32^\circ$ very nearly. It is symmetric about the Equator and closely follows what would be the great circle path on the corresponding sphere. We find, however, that the distance along the southerly geodesic, which is one of the two shortest paths, is 10773.0 geographical miles. This distance is only marginally smaller than the distance along the intermediate geodesic but what is surprising

and important is that the path taken is along an entirely different route and navigable entirely by sea. The route heads southwards initially to its vertex in the South Indian Ocean at 45°S and $64^{\circ}19'\text{E}$ (approximately). It passes close to the west of Capetown, crossing the parallel of 35°S in longitude $17^{\circ}56'\text{E}$ (approximately), proceeds northward through the South Atlantic Ocean and crosses the Equator in longitude $26^{\circ}23'\text{W}$ (approximately). In Figure 7.3 the continuous line shows this route. There is, of course, another route (along the northerly geodesic) of equal length which passes through its vertex in latitude 45°N . This route is marked by the broken line in Figure 7.3.

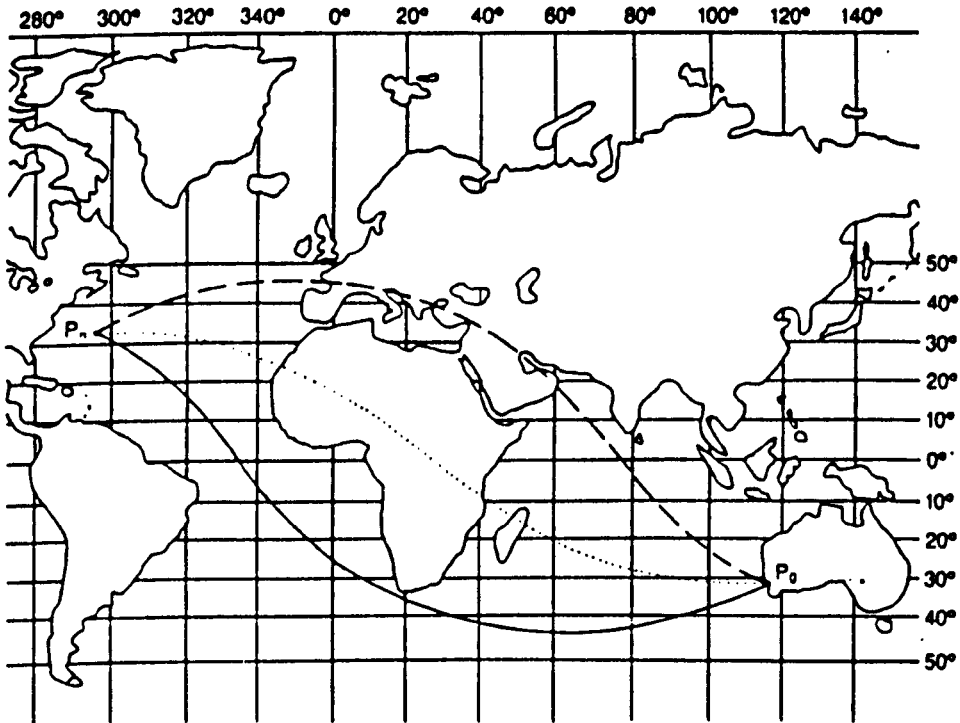


FIGURE 7.3. GEODESIC PATHS FROM FREMANTLE TO BERMUDA.

Surprising again is the fact that a small change in the difference of longitude between the departure and arrival positions will result in another path which will differ considerably from the above path. This does mean that, in fact, the computation of position along the geodesic path is relatively unstable although the distance calculation itself is not.

7.3. CONCLUSION.

Theoretically speaking, nearly antipodean points, in the sense used here, are those points which have exactly the same latitude in opposite hemispheres and whose longitudes fall within the specified range which is applicable to that latitude. Naturally, points whose latitudes differ in absolute value by only small amounts which might fall inside the error bounds for the computation could also, in practice, be considered nearly antipodean for all practical purposes.

The work in chapters 6 and 7 is an expansion of the paper by Williams & Pythian³⁶ published in the Journal of Navigation.

As has been shown, the saving in distance by using a shortest path geodesic between two nearly antipodean points instead of the path along the intermediate geodesic is only small, and, for most practical purposes, negligible. The routes taken by the shortest path geodesics are, however, geographically significant in many cases and could provide advantageous alternatives to the navigator.

- 8 -

COMPUTATION
OF
ASTRONOMICAL FIXES

8.1 THE GEOGRAPHICAL POSITION OF A HEAVENLY BODY.

At any instant in time each heavenly body (the Sun, Moon, Planets and Stars) is directly above some point on the surface of the Earth, that is, it lies on the line which is normal to the surface of the Earth at that point. This point on the surface of the Earth is known as the GEOGRAPHICAL POSITION of the heavenly body and the heavenly body is also at the ZENITH of an observer placed in that position. The point X in Figure 8.1 is the Geographical Position of the astronomical body *. The Zenith for an observer at a point Z on the surface of the Earth is the point in space at the extremity of the normal to the surface at Z. The body *, for instance lies at the zenith of the point X.

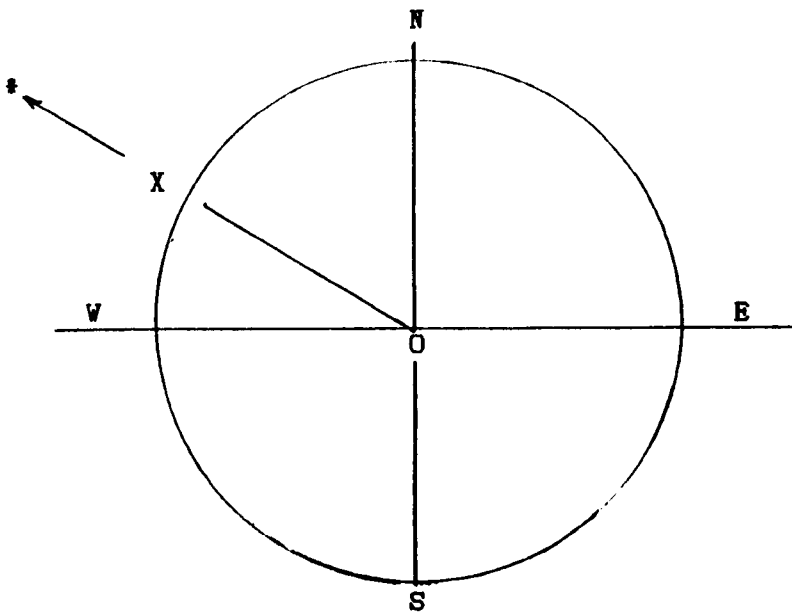


FIGURE 8.1

The CELESTIAL SPHERE is an imaginary sphere of no fixed radius but centred at the centre of the Earth and onto whose surface the positions of all the heavenly bodies are mapped. The Celestial Equator is the intersection of the Celestial Sphere by the plane through the Earth's Equator. Positions on the surface of the Celestial Sphere are described by angular coordinates. Declination (which is the exact equivalent of latitude) is the angular coordinate of the heavenly body North or South of the Celestial

Equator. The declination of a heavenly body immediately gives the latitude of the geographical position of the heavenly body. The angular coordinate on the surface of the Celestial Sphere from which we obtain the longitude of the Geographical Position of the heavenly body on the surface of the Earth is the HOUR ANGLE. Hour Angle is expressed in degrees from 0° to 360° and is measured westwards because the bodies in the heavens themselves move westwards relative to an observer on the surface of the Earth. The GREENWICH HOUR ANGLE (GHA) of a heavenly body is the angle between the image of the Greenwich meridian on the Celestial Sphere and the meridian through the body. The longitude of the Geographical Position of the body is obtained from the GHA :

$$\text{LONGITUDE} = 2\pi - \text{GHA}$$

Nautical Almanacs provide us with the information from which to compute the Geographical Position. The Nautical Almanacs are published yearly and the Declination and GHA of the Sun, Moon, and major Planets are given for hourly intervals throughout the year. Interpolation Tables to interpolate the GHA and Declination of these bodies for minutes and seconds are also given where it assumed that the hourly changes are more or less constant.

The GHA is also tabulated hourly for a reference point on the Celestial Sphere which is known as the FIRST POINT OF ARIES. This point is a reference point established by astronomers. It is the point on the celestial equator at which the sun crosses at Vernal (Spring) Equinox. The position of the First Point of Aries varies very slightly and is so named because, at the time that this point was first established, it lay in the constellation of Aries. In the case of stars, Hour Angle is then measured from the FIRST POINT OF ARIES. The angle between the meridian on the Celestial Sphere which passes through the First Point of Aries and the meridian through a particular star is known as the SIDEREAL HOUR ANGLE (SHA). For a star the values of the Declination and the SHA vary little and are tabulated once every few days. We obtain the GHA of star (*) from

$$\text{GHA}(\ast) = [\text{GHA}(\text{Aries}) + \text{SHA}(\ast)] \bmod(2\pi)$$

8.2 ASTRONOMICAL OBSERVATIONS

In Navigational terms an astronomical observation means an observation of any heavenly body - stars, sun, moon or planets. These observations are the altitudes of the heavenly bodies above the horizon taken using a sextant. For a seafaring Navigator this horizon is usually the sea horizon (VISIBLE HORIZON) and for an aviator the horizon is usually an artificial horizon (SENSIBLE HORIZON) in a bubble sextant.

Figure 8.2 shows the Visible Horizon, VV' , and Sensible Horizon SS' , for an observer at the point Z whose height above the surface of the Earth is h . The angle SZV is known as the DIP of the Sea Horizon.

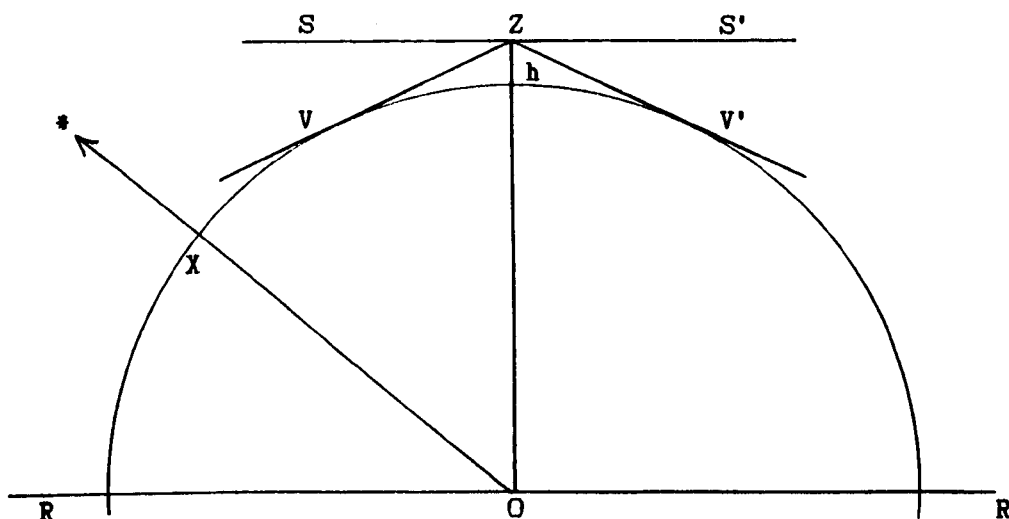


FIGURE 8.2 - OBSERVER'S VISIBLE AND SENSIBLE HORIZONS.

The plane through the centre of the Earth which is parallel to the Sensible Horizon is the RATIONAL HORIZON. (RR' in Figure 8.2).

After the altitude has been read from the sextant it is corrected to give the altitude of the centre of the body from the centre of the Earth above the RATIONAL HORIZON.

The altitude of the body (*) is RO^* in Figure 8.2. The point X is the Geographical Position of *.

Having found the altitude of a heavenly body, *, we subtract this from 90° to give us the ZENITH DISTANCE of * which is the angle ZO^* in Figure 8.2. When the Zenith Distance is expressed in minutes of

arc it gives us the distance in geographical miles along the Great Circle arc between the observer and the Geographical Position of the heavenly body.

This zenith distance then gives a POSITION CIRCLE on which the observer must lie. The circle has its centre at the geographical position of the heavenly body and its radius is zenith distance.

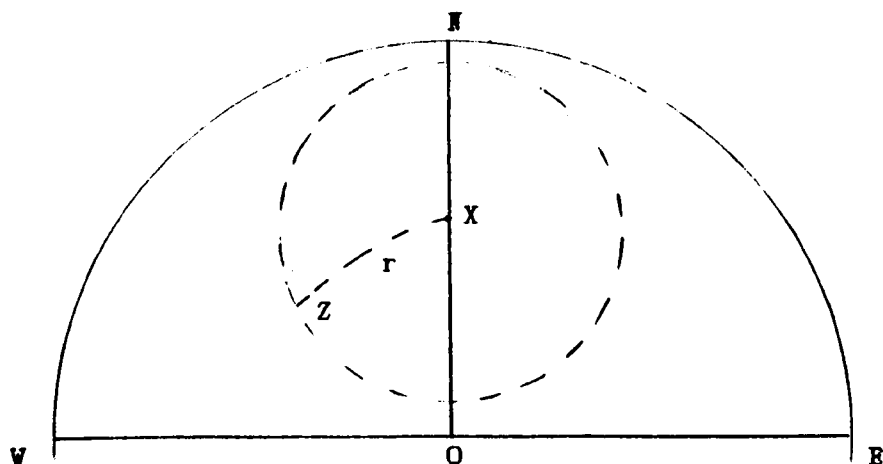


FIGURE 8.3 - POSITION CIRCLE ON THE SURFACE OF THE SPHERICAL EARTH.

If we take two simultaneous observations of two different heavenly bodies then this defines two position circles, and, at one of the points of intersection of these position circles, we will find the observer's position. The observer will usually have a good estimate of his position beforehand and this should determine, with very little possibility of ambiguity, which of the two points of intersection of the position circles is the observer's position. If it happens that there might be ambiguity, then this can be resolved by taking three observations of three distinct heavenly bodies.

8.3 DEFINING THE EQUATION OF A POSITION CIRCLE USING THE METHODS OF SPHERICAL TRIGONOMETRY.

A spherical triangle on the surface of a sphere is the area enclosed by the intersections of three great circles. Figure 8.4 shows the spherical triangle PZX. In the triangle the angles are at P, Z and X and the sides are p, z, and x. The sides are expressed as angles where p, z and x are the angles subtended at the centre of the sphere by the arcs ZX, PZ and PX on the surface of the sphere, respectively.

To determine a side of the triangle, p say, we would use the SPHERICAL COSINE FORMULA from which we would find

$$\boxed{\cos p = \cos z \cos x + \sin z \sin x \cos P} \quad \dots\dots (8.1)$$

The details of how this formula is derived may be found in the book by Gow³⁷.

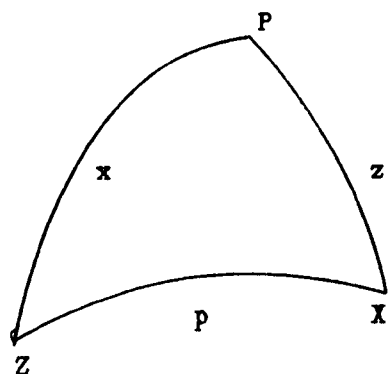


FIGURE 8.4 - A SPHERICAL TRIANGLE.

We can give a particular physical significance to triangle PZX and show that equation (8.1) is the equation of the position circle on the surface of the spherical Earth.

See Figure 8.5. In the figure the point P is at the Pole and the arc PG is the Greenwich meridian. Let the point X be the geographical position of the heavenly body and let the point Z be the variable position of the observer who lies on a circle around the point X the radius of which is the arc ZX (= ap where a is the equatorial radius of the Earth in nautical miles).

If the geographical position of the point X is (χ, λ) where χ is the latitude of the geographical position of the observed body (its declination) and λ is the longitude ($2\pi - \text{GHA}$) and (ϕ, θ) are the coordinates of the variable point X then the angles subtended by the arcs PX ($=z$) and PZ ($=x$) are

$$z = \frac{1}{2}\pi - \chi \qquad x = \frac{1}{2}\pi - \phi$$

The angle at P is $\lambda - \theta$ and equation (8.1) becomes

$$\boxed{\cos p = \sin \phi \sin \chi + \cos \phi \cos \chi \cos (\lambda - \theta)} \quad \dots (8.2)$$

p is the angle obtained from the observation and equation (8.2) will give us the locus of the variable point (ϕ, θ) . The angles are all expressed in radians with latitude, ϕ , and longitude, θ bounded by

$$-\frac{1}{2}\pi \leq \phi \leq \frac{1}{2}\pi \quad (\text{North positive})$$

$$0 \leq \theta \leq 2\pi \quad (\text{East positive}).$$

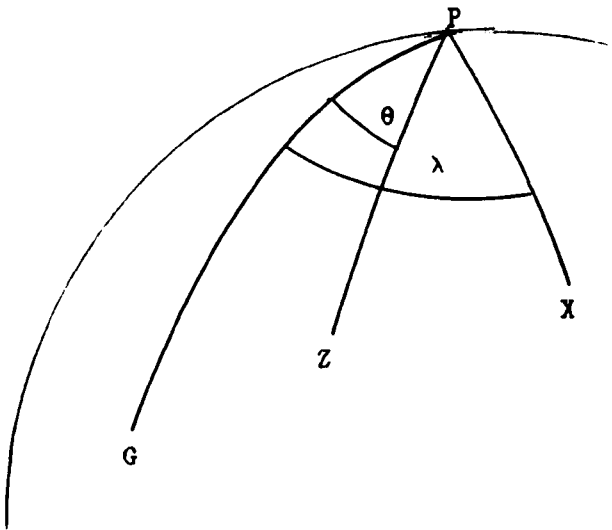


FIGURE 8.5 - POSITION CIRCLE

If we now obtain a second observation of a heavenly body whose geographical position is (χ', λ') and the resulting zenith distance angle is p' then we find

$$\boxed{\cos p' = \sin \phi \sin \chi' + \cos \phi \cos \chi' \cos (\lambda' - \theta)} \quad \dots (8.3)$$

The observer's position is therefore the selected solution for ϕ and θ of the pair of simultaneous non-linear equations (8.2) and (8.3). To solve these equations analytically by elimination is a lengthy procedure and its final solution requires a lot of function evaluations. One of the ways of computing the coordinates of the points of intersection of the position circles directly was presented in a paper by Arturo and Raphaele Chiesa³⁸ in the Journal of Navigation. The derivation of the equations is lengthy but, in the end, the longitude, θ , and then the latitude, ϕ , of the points of intersection of the position circles are given explicitly. The method involves determining the chord which is the arc of a great circle and which is common to both position circles. This was also done by Bernard Spencer³⁹. The solution to the problem can, however, be effected more economically using Newton's Method for two dimensions with the Dead Reckoning Position of the observer used as a first approximation.

8.4. COMPUTATION OF AN ASTRONOMICAL RUNNING FIX.

In a *RUNNING FIX* the second observation, probably of the same heavenly body as the first, is taken after the observer has travelled an appreciable distance along the arc of a loxodrome. This is a common practice and, indeed, it was customary, on almost every British Ship on Ocean passage, to compute the "noon position" (a serious ritual) from the intersection of a position line obtained from the sun approximately three hours before noon with a latitude obtained directly from the meridian passage of the sun at noon. In their paper Chiesa & Chiesa³⁸ state that it is simply a case of transferring the geographical position of the heavenly body at the first observation along the arc of the loxodrome that the observer has experienced, drawing the position circle around it and then

computing the intersection of this transferred position circle with the position circle obtained from the second observation. The mathematics does, however, show that the problem of transferring a position circle is more complicated than that. This was initially pointed out to this author by a colleague in a private communication. There are certain theoretical implications in the problem of transferring a position circle which were not mentioned by Chiesa & Chiesa³⁹. In fact, the transferred locus is no longer a circle but suffers a distortion and, in response to the the paper by Chiesa and Chiesa, this author⁴⁰ wrote a paper which was also published by the Journal of Navigation.

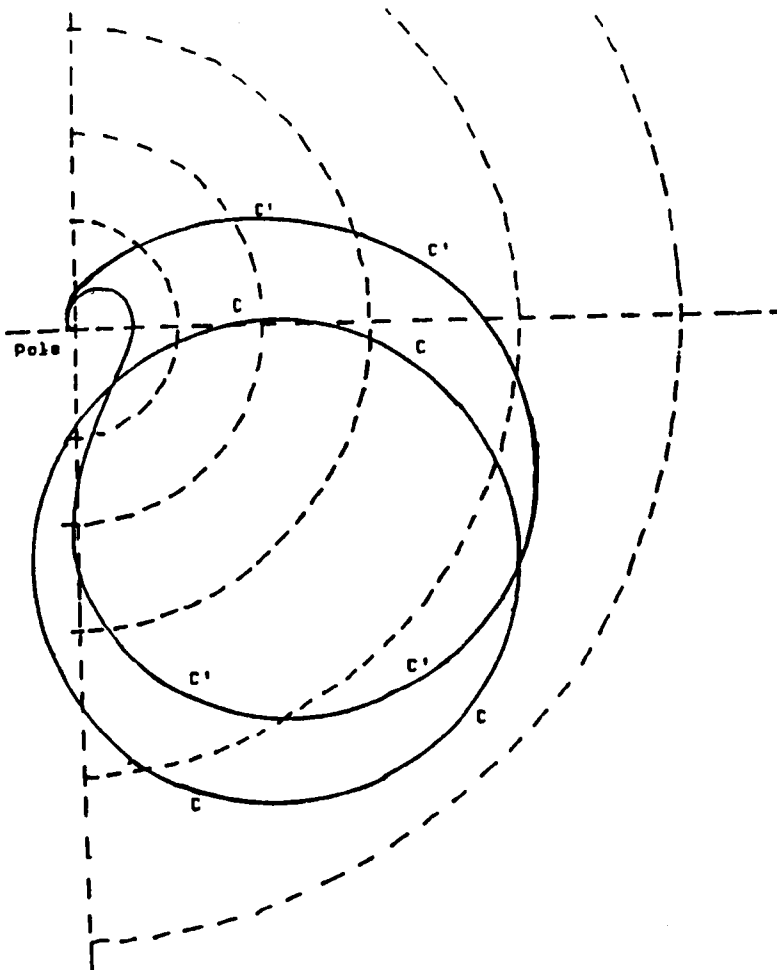


FIGURE 8.6 DISTORTION OF A TRANSFERRED POSITION CIRCLE.

It is, after all, the position of the observer which has changed and so, if the observer lies in a certain position on a position locus then it is *that point* on the locus which must be transferred. Since we are uncertain as to which exact point on the locus the observer lay at the time of the first observation then we must transfer *each* point of the locus the same distance on the same course along a loxodrome *from that point*. Now each point will transfer in a slightly different manner as can be seen by actual plotting so that the original position circle does not retain its shape. See Figure 8.6. which illustrates the extreme distortion of a transferred position circle in the vicinity of the pole.

If we move every point of the position circle (8.2) through a distance s (in nautical miles) along a loxodrome on course α , then the general point (ϕ, θ) on the position circle (8.2) transfers to the point (ϕ', θ') . These coordinates are related by

$$\boxed{\phi = \phi' - \frac{s}{a} \cos \alpha} \quad \dots\dots (8.4)$$

$$\boxed{\theta = \theta' - \tan \alpha \int_{\phi}^{\phi'} \sec u \, du} \quad \dots\dots (8.5)$$

where ϕ' is the latitude and θ' is the longitude of a point on the transferred locus .

Using the substitutions (8.4) and (8.5) in equation (8.2) we find the equation of the transferred position locus:

$$\boxed{\cos(\lambda - k - \theta') = \frac{\cos p - \sin(\phi' - \frac{s}{a} \cos \alpha) \sin \chi}{\cos(\phi' - \frac{s}{a} \cos \alpha) \cos \chi}} \quad \dots\dots (8.6)$$

where

$$k = \tan \alpha \int \sec u \, du$$

k is the difference of longitude experienced by the observer travelling along the loxodrome and is expressed in radians.

Figure 8.7 shows the actual distortion that takes place when each point of the arc AA' of a position circle is transferred 300 nautical miles on a course of 045° along the arc of a loxodrome. The arc BB' is the image. The diagram is drawn to scale in the stereographic projection where a circle on the surface of the sphere is projected into a circle in the stereographic plane.

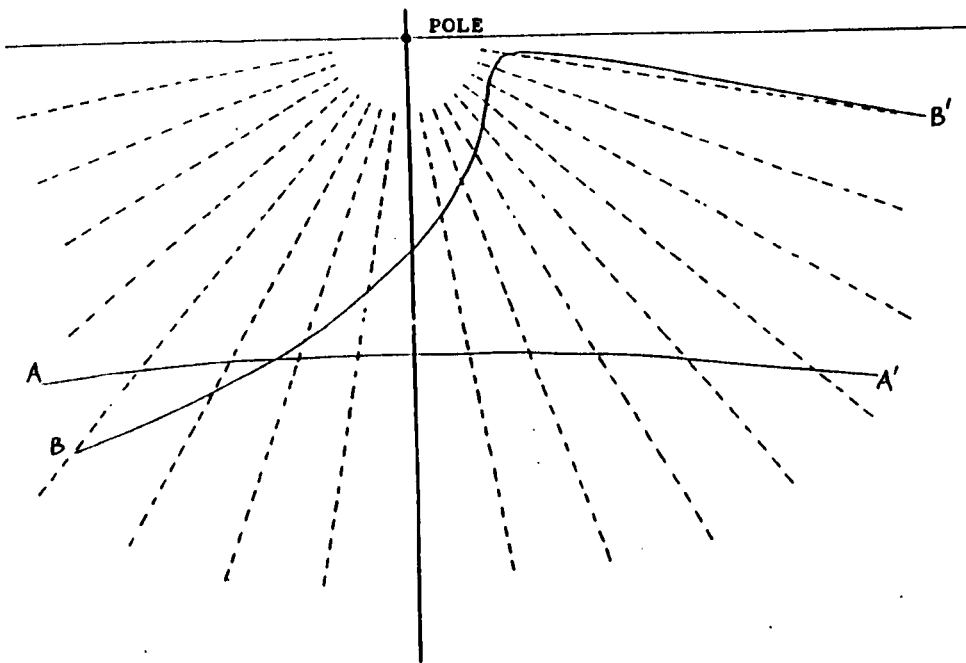


FIGURE 8.7 ARC OF POSITION CIRCLE AND TRANSFERRED POSITION CIRCLE.

Now we have shown in Chapter 2 that a loxodromic curve on the surface of a sphere is a curve of finite length which spirals towards an end limit point at each pole, therefore the loxodrome, which passes through a point P on a position circle defined by equation 8.1, where the coordinates of P are (ϕ_P, θ_P) , on course α , is of finite length to the pole so that, if the distance, d, steamed by the observer, is greater than the distance along the loxodrome on course α from P to the pole, then the point P has no image point on the transferred locus defined by equation 8.6. A point (ϕ, θ) on the position circle will therefore have an image on the transferred locus only if

$$|\phi + \frac{d}{a} \cos \alpha| < \frac{1}{2}\pi$$

8.5 COMPUTING OBSERVED POSITION FROM A RUNNING FIX.

The observed position at the time of the second observation is the selected solution of the pair of simultaneous non-linear equations 8.3 and 8.6. Since a direct method of doing this in this case would seem to be out of the question, we use a numerical method such as Newton's method. We know that there are two solutions satisfying the pair of equations; Newton's method will find one of them. The observer will have knowledge of his Dead Reckoning position and if this is used as the first approximation then the iterative procedure will give the observer's true position.

In the standard method of determining the observer's position in the days before the advent of the computer the position was found after an elaborate method of plotting position lines which were, in effect, those small arcs of the position circle which lay close to the dead reckoning position. Early computer methods systemised this plotting procedure. Our chosen method of finding the observer's position will require a computer. Indeed, in a system designed to find the observer's position from astronomical observations using a computer the method of finding the position directly from the intersection of the position loci is probably the most efficient.

Simultaneous and running fixes can be solved equally well in one computer programme designed for both cases.

In most cases, and in sea-going Navigation particularly, where activity is restricted to what are termed the "Navigable Latitudes", a representation of the whole of the transferred position locus is not necessary to find its intersection with a position circle. Moreover, that portion of the position circle expressed by equation 8.1, which is relevant to the observer and on which he might be expected to lie, will almost certainly have a complete image on the locus represented by equation 8.6.

The latitude obtained by astronomical observation is known as the ASTRONOMICAL LATITUDE, and is, in fact, the geodetic latitude and not the geocentric latitude even though it is measured from the centre of the Earth. This is because the Earth itself is only considered to be a point at the centre of the Celestial Sphere.

It must be admitted that there is often little practical value in allowing for the distortion of a position circle when the observer travels along a loxodrome, since the error in estimating the course and distance made good between the observations may well exceed any correction made by allowing for the distortion. Nevertheless we believe that it is advisable to use accurate formulae whenever possible.

- 9 -

COMPUTING POSITION
FROM OBSERVATION
OF A SINGLE BODY

9.1 COMPUTING THE POSITION BY OBSERVATION AT MERIDIAN PASSAGE.

We have considered the computation of the observer's position by taking a "Running Fix", by which method the position is fixed by taking two observations of the same heavenly body some hours apart, but now we will consider the possibility of fixing the observer's position by taking observations of a single heavenly body over a much shorter interval of time.

When the geographical positions of the observer and a heavenly body have the same longitude then, provided that the observer is stationary and that the rate of change in the declination of the body is negligible, the body will be at its maximum or minimum altitude. We refer to this as the MERIDIAN PASSAGE of the heavenly body through the observer's meridian. In this circumstance the latitude of the observer can be computed directly from the altitude. See Figure 9.1 which shows a section through the Celestial Sphere in the plane of the observer's meridian at the time of meridian passage.

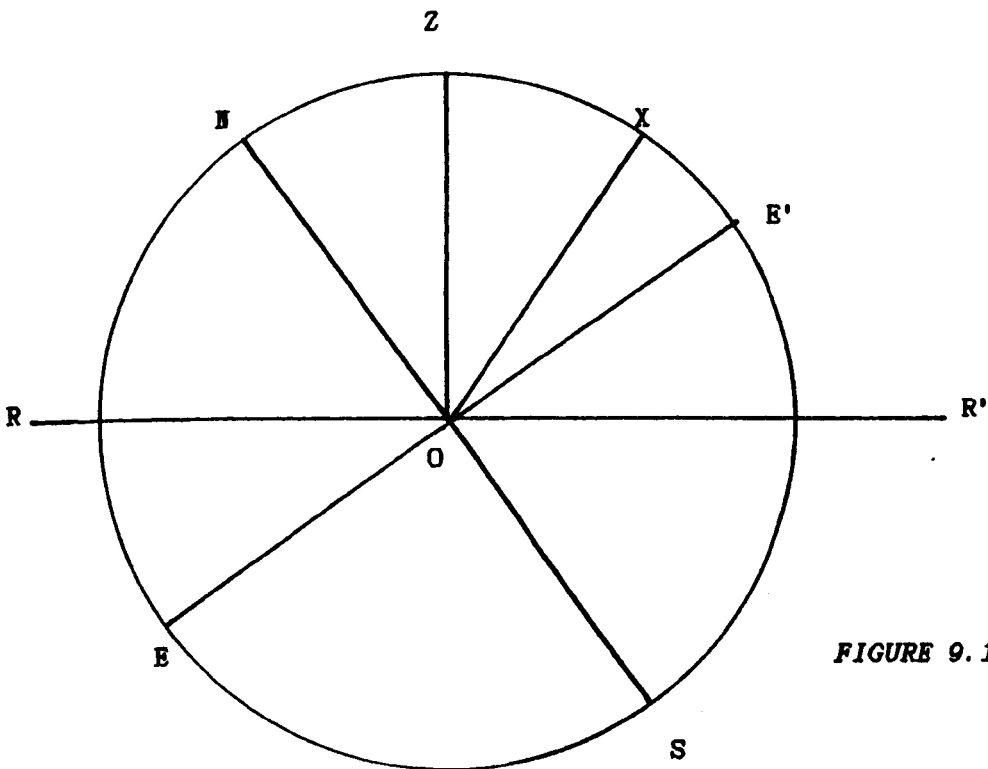


FIGURE 9.1

The Earth, and hence the observer, is just a point and lies at O which is at the centre of the sphere. The heavenly body is at X and Z is the zenith. The line EE' is the plane of the Equator and the angle E'OX is the declination which, at the same time, is the latitude, χ , of the geographical position of the body. The line NS is the axis of rotation of the Earth. The line ROR' is the plane of the Rational Horizon (defined in chapter 8) and the angle XOR' is the altitude of the heavenly body.

In equation (8.2) $\lambda - \theta = 0$ and the equation reduces to the simple form

$$\cos p = \cos (\phi - \chi) \quad \dots\dots (9.1)$$

whence we find that either

$$p = \phi - \chi$$

or $p = - (\phi - \chi)$

If α is the true altitude of the heavenly body then $p = \frac{1}{2}\pi - \alpha$ so that

$$\phi = (\frac{1}{2}\pi - \alpha) + \chi \quad \dots\dots (9.2)a$$

or $\phi = \chi - (\frac{1}{2}\pi - \alpha) \quad \dots\dots (9.2)b$

If, at the time of meridian passage, the exact Greenwich Mean Time can be determined then we also have a method of finding the longitude and hence the observer's position. Each one minute of time difference between the meridian passage of the heavenly body at Greenwich and the passage of the same body through the observer's meridian will be equal to fifteen minutes of arc in the difference in the longitude. This is due to the angular velocity of rotation of the Earth. The difficulty arises in the estimation of the exact moment at which the body reaches its maximum altitude. From experience, even when a heavenly body is at high altitude and where it is changing its altitude fairly rapidly, it is difficult to estimate the time of the maximum to the nearest minute. For a yachtsman, on an ocean passage and far from land where sea

conditions make astronomical observation difficult anyway, this might be an acceptable approximation - one minute would give the longitude to within fifteen minutes of arc - but for the navigator engaged in commercial business the process would have to be refined and this can be done.

If the observer is stationary, one way is to take an altitude of the heavenly body some minutes before meridian passage and while it is still changing altitude rapidly enough that this change is easily observable. After noting the exact time of this observation, we then wait until the body has crossed the meridian and then note the exact time that the body is again at the same altitude after meridian passage. The mid point of this time interval will then be the time of meridian passage and the longitude can then be determined. See Figure 9.2 in which * represents the body. The figure is a graph of altitude, α , against time, t .

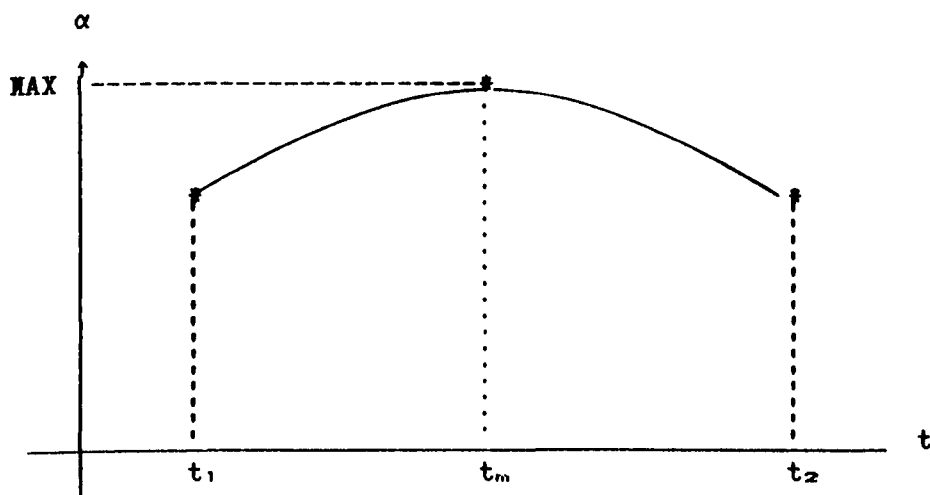


FIGURE 9.2

9.2 COMPUTING THE OBSERVED POSITION AT THE TIME OF CULMINATION.

The accuracy of the above method of determining the maximum altitude of the astronomical body will depend also upon the fact that the body is not changing the latitude of its geographical

position at too great a rate. In the case of the sun, which is the body that is most often used in the observation of meridian passage, the maximum hourly rate of change in the declination of the sun that takes place is approximately 1 minute of arc and this occurs at the time of the equinox. At the time of the solstices this rate of change in declination of the sun is zero. At certain times, therefore, allowance should be made for this rate of change. The moon moves much more rapidly and the hourly rate of change in its declination can be as much as nearly 17 minutes of arc. Even when the moon is close to the maximum absolute value of its declination there is still an appreciable hourly rate of change. When there is an appreciable hourly rate of change in the declination of the observed body then the maximum altitude will not occur exactly at the time of meridian passage, and, when this is the case, equations (9.2)a and (9.2)b no longer hold so that we must go back to equation (8.2). We refer to the occurrence of the maximum or minimum altitude as CULMINATION.

In most cases the problem is a dual one - not only do we have to determine the difference between the time that the heavenly body crosses the observer's meridian and the time of occurrence of maximum altitude due to the change in geographical position but we have the additional problem that the observer may be moving at an appreciable speed and the time of the maximum altitude might no longer coincide with the time of meridian passage due to this effect also.

Solutions to this problem have been published recently by J.N.Wilson⁴¹ and Matti Ranta⁴². Both authors use more or less the same mathematical analysis but, whereas Wilson offers a mainly graphical solution, Ranta uses numerical analysis. Let us first analyse the method of solution of the problem as presented by Ranta⁴².

We need to know the value of α at culmination. This is estimated by Ranta⁴² by taking a series of observations of the altitude, α , of

the body during a time interval which includes the time of culmination, fitting a polynomial

$$\alpha(t) = \sum_{r=0}^n c_r t^r$$

and computing the value of α when $\alpha'(t) = 0$.

If a least squares quadratic is fitted then clearly the time of culmination, t_c , is given at

$$t_c = -\frac{c_1}{2c_2}$$

and the altitude, α_c , at this time is

$$\alpha_c = c_0 - \frac{c_1^2}{4c_2}$$

Let us consider an observer in a position (Z) on the surface of the Earth where the latitude is ϕ and the longitude is θ . The observer finds the altitude, α , of a heavenly body (X) whose declination is χ and whose Greenwich Hour Angle is $2\pi - \lambda$. The spherical triangle, PZX, which results from this observation is shown in Figure 9.3.

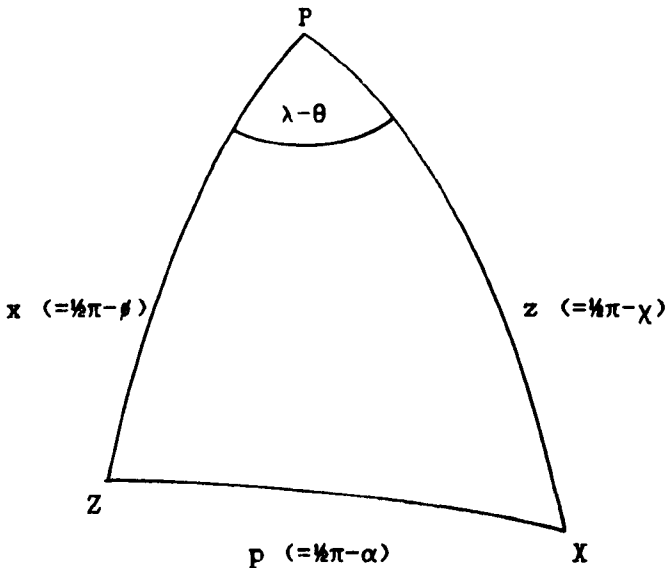


FIGURE 9.3

In Figure 9.3 P is at the North Pole and, as a result, equation (8.2) gives us

$$\cos p = \sin \phi \sin \chi + \cos \phi \cos \chi \cos (\lambda - \theta)$$

Since $p = \frac{1}{2}\pi - \alpha$ this can be written

$$\sin \alpha = \sin \phi \sin \chi + \cos \phi \cos \chi \cos (\lambda - \theta) \quad \dots\dots (9.3)$$

Now $\alpha, \phi, \chi, \theta$ and λ are all functions of the time, t , so that, differentiating equation (9.3) with respect to t gives

$$\begin{aligned} \cos \alpha \frac{d\alpha}{dt} &= \cos \phi \sin \chi \frac{d\phi}{dt} + \sin \phi \cos \chi \frac{d\chi}{dt} \\ &\quad - \sin \phi \cos \chi \cos (\lambda - \theta) \frac{d\phi}{dt} \\ &\quad - \cos \phi \sin \chi \cos (\lambda - \theta) \frac{d\chi}{dt} \\ &\quad - \cos \phi \cos \chi \sin (\lambda - \theta) \left(\frac{d\lambda}{dt} - \frac{d\theta}{dt} \right) \end{aligned}$$

Re-arranging gives

$$\begin{aligned} \cos \alpha \frac{d\alpha}{dt} &= [\cos \phi \sin \chi - \sin \phi \cos \chi \cos (\lambda - \theta)] \frac{d\phi}{dt} \\ &\quad + [\sin \phi \cos \chi - \cos \phi \sin \chi \cos (\lambda - \theta)] \frac{d\chi}{dt} \\ &\quad - \cos \phi \cos \chi \sin (\lambda - \theta) \left(\frac{d\lambda}{dt} - \frac{d\theta}{dt} \right) \quad \dots\dots (9.4) \end{aligned}$$

From the spherical cosine formula we see that

$$\cos (\lambda - \theta) = \frac{\sin \alpha - \sin \phi \sin \chi}{\cos \phi \cos \chi} \quad \dots\dots (9.5)a$$

and from the spherical sine formula

$$\sin (\lambda-\theta) = \frac{\cos \alpha \sin Z}{\cos \chi} \quad \dots \dots (9.5)b$$

where Z is the azimuth (the angle at the position of the observer).

If we substitute equation (9.5)a in the coefficients of $\frac{d\phi}{dt}$ and $\frac{d\chi}{dt}$

and substitute equation (9.5)b in the coefficient of $(\frac{d\lambda}{dt} - \frac{d\theta}{dt})$ in

equation (9.4) and divide the equation through by $\cos \alpha$ the result is

$\frac{d\alpha}{dt} = \cos Z \frac{d\phi}{dt} + \cos \chi \frac{d\chi}{dt} - \sin Z \cos \phi \left(\frac{d\lambda}{dt} - \frac{d\theta}{dt} \right)$... (9.6)
--	-----------

where we have substituted

$$\cos \chi = \frac{\sin \phi - \sin \alpha \sin \lambda}{\cos \alpha \cos \lambda}$$

and

$$\cos Z = \frac{\sin \lambda - \sin \alpha \sin \phi}{\cos \alpha \cos \phi}$$

Now $\cos \chi$ can also be expressed in the form

$$\cos \chi = - \cos (\lambda-\theta) \cos Z + \sin (\lambda-\theta) \sin Z \sin \phi$$

and, if the observer is moving at a speed V with northerly component V_n and easterly component V_e , then

$$\frac{d\phi}{dt} = V_n \quad , \quad \frac{d\theta}{dt} = V_e \sec \phi$$

so that, if we substitute all this into equation (9.6) and re-arrange we find

$$\frac{d\alpha}{dt} = [V_{\phi} - \cos(\lambda-\theta) \frac{d\lambda}{dt}] \cos Z + [V_{\theta} + \sin(\lambda-\theta) \sin \phi \frac{d\lambda}{dt} - \cos \phi \frac{d\lambda}{dt}] \sin Z$$

..... (9.7)

At the time of culmination when $\frac{d\alpha}{dt} = 0$ we can solve equation to give us the value of the angle Z :

$$Z = \tan^{-1} \frac{V_{\phi} - \cos(\lambda-\theta) \frac{d\lambda}{dt}}{V_{\theta} + \sin(\lambda-\theta) \sin \phi \frac{d\lambda}{dt} - \cos \phi \frac{d\lambda}{dt}}$$

..... (9.8)

where $-\frac{1}{2}\pi < Z < \frac{1}{2}\pi$ if the geographical position of the body is to the north of the observer and $\frac{1}{2}\pi < Z < \frac{3}{2}\pi$ if the geographical position of the body is to the south of the observer. We can find then that the angle at P [= $(\lambda-\theta)$] in the spherical triangle PZX is found from the sine formula and

$$\lambda - \theta = \sin^{-1} \frac{\sin Z \cos \alpha}{\cos \chi}$$

..... (9.9)

The Greenwich Hour Angle, GHA, of the body can be found from the Nautical Almanac for the time of culmination whence $\lambda = 2\pi - \text{GHA}$ and, hence, from equation (9.9), we can determine the longitude, θ , at this time.

We can now solve equation (9.3) to give us the latitude, ϕ , at the time of culmination. Ranta⁴² does this by writing

$$\sin \alpha = R \cos (\phi - k)$$

which gives

$$R = \sqrt{1 - \sin^2 Z \cos^2 \alpha}$$

and

$$k = \tan^{-1} \frac{\tan \chi}{\cos(\lambda - \theta)}$$

Hence

$$\phi = \cos^{-1} \frac{\sin \alpha}{\sqrt{1 - \sin^2 Z \cos^2 \alpha}} + \tan^{-1} \frac{\tan \chi}{\cos(\lambda - \theta)}$$

..... (9.10)

To summarise, then, the procedure for finding the observer's position at the time of culmination of a heavenly body is to take a series of altitudes of the body surrounding the time of culmination, to fit an interpolating polynomial to the data thus found giving the altitude as a function of time and, from this interpolating polynomial, to find the maximum altitude and the time at which it occurs. We then find, by iteration, the latitude, ϕ , and the longitude, θ , of the observer at the time of culmination by using equations (9.8), (9.9) and (9.10). The first approximations for ϕ and θ will be the Dead Reckoning position of the observer at the estimated time of culmination.

9.3 COMPUTATIONAL EXPERIMENT.

James N. Wilson⁴ provided useful data from which to experiment with the computational formula of the previous section. His data was obtained on voyages in a yacht off the coast of California near Catalina Island. He observed the altitude of the sun for approximately one hour during which time it reached its highest altitude. He himself only used those observations taken in a ten minute period which included culmination from which he estimated, graphically, the maximum altitude. We will use the same observations over the same period but we will fit a least squares quadratic to this data and, from this quadratic approximation, we will compute both the time of culmination and the corresponding altitude. The data is shown in Table 9.1 below.

	TIME	ALTITUDE	
→	11 50 39	32°55.1'	†
	11 51 41	32°56.1'	†
→	11 52 21	32°56.2'	
	11 53 31	32°58.5'	†
→	11 54 30	32°56.9'	
	11 55 04	32°57.5'	†
→	11 56 16	32°57.5'	
	11 56 52	32°57.6'	†
→	11 57 42	32°57.6'	
	11 58 42	32°59.9'	†
→	11 59 20	32°57.5'	
→	12 00 58	32°55.1'	†

TABLE 9.1

In the first experiment the least squares quadratic polynomial was fitted to *all* the data points of Table 9.1 and, from this we found that

$$(1) \quad t_c = 11\ 56\ 23 \qquad \alpha_c = 32^\circ 58.1'$$

In the second experiment the same type of curve was fitted to the seven points marked in Table 9.1 by arrows on the left. In this case we found

$$(2) \quad t_c = 11\ 56\ 14 \qquad \alpha_c = 32^\circ 57.6'$$

In the third experiment we fitted the least squares quadratic to the data points in Table 9.1 marked by an arrow on the right. we found

$$(3) \quad t_c = 11 \ 56 \ 11 \qquad \alpha_c = 32^\circ 58.6'$$

Using the first of these results - (1) - we then find the latitude, ϕ , and the longitude, θ , of the observer using the iterative scheme

$$\tan Z_n = \frac{-V \cos \gamma}{\Omega \cos \phi_{n-1} + V \sin \gamma} \qquad \dots\dots (9.11)a$$

$$\sin(\theta-\lambda)_n = \frac{\sin Z_n \cos \alpha}{\cos \chi} \qquad \dots\dots (9.11)b$$

$$\cos \phi_n = \frac{\sin \alpha - \sin \phi_{n-1} \sin \chi}{\cos \chi \cos(\theta-\lambda)_n} \qquad \dots\dots (9.11)c$$

γ is the course angle on which the observer is moving at speed V knots. Since the observations of the sun were taken near to the time of the winter solstice then the rate of change in declination is zero. Wilson⁴¹ states that he was steering a course of 210° at a speed of 6 knots. The altitude, α , as given in Table 9.1 is the altitude as read from the sextant and is the altitude of the sun above the visible horizon. To this altitude we must add a correction of $13.9'$ to give the altitude of the sun above the rational horizon. This gives $\alpha_c = 33^\circ 12'$.

In the iterative scheme (9.11) we use the initial values

$$\phi_0 = 33^\circ 40.0'$$

$$\chi = 23^\circ 8.9'$$

The results that we obtained for ϕ and $(\lambda-\theta)$ were

$$\phi = 33^\circ 39.1' \qquad \lambda-\theta = 0^\circ 21.8'$$

At the time of culmination, from the Nautical Almanac, we find that the Greenwich Hour Angle of the sun was $118^\circ 26.8'$ so that

$$\lambda = 360^\circ - \text{GHA} = 241^\circ 33.2'$$

and $\theta = 241^\circ 33.2' + 0^\circ 21.8' = 241^\circ 55.0'$

West of Greenwich this gives

$$\text{Longitude} = 118^\circ 5.0' \text{ W .}$$

This longitude does not agree very well with the position found by Wilson³⁰. He finds that the longitude of the observed position is $118^\circ 15.3' \text{ W}$. The difference occurs because we disagree about the time of culmination. His time of culmination, found graphically using in a way which will be described below, is 11 57 09. The time of culmination that we determined here in the first experiment using all the observations was 11 56 23 .

Wilson⁴¹ did not estimate the time of culmination from the data in Table 9.1 but took two further sets of altitudes of the sun - one set before culmination and one set after culmination. These two sets of altitudes are shown in Tables 9.2 and 9.3 :

TIME	ALTITUDE
11 28 21	32°34.0'
11 28 59	32°35.5'
11 30 00	32°34.0'
11 30 44	32°36.9'
11 31 24	32°38.5'
11 32 06	32°37.5'
11 32 42	32°40.3'
11 33 12	32°40.3'

TABLE 9.2

TIME	ALTITUDE
12 19 54	32°42.6'
12 21 10	32°41.2'
12 22 51	32°37.6'
12 23 47	32°36.6'
12 24 50	32°35.0'
12 26 01	32°33.9'

TABLE 9.3

From these tables, 9.2 and 9.3, Wilson⁴¹ faired straight lines, graphically, and from their point of intersection, estimated the time of highest altitude - the time of culmination. He estimated the time of culmination at 11 57 09 . However, fitting least squares straight lines to the data of tables 9.2 and 9.3 we find, algebraically, that the point of intersection of the straight lines estimated the time of culmination at 11 58 51 .

In conclusion, it must be said that the data is not really of the best kind to demonstrate the method of computing the ship's position at the time of culmination although Wilson⁴¹ obtains good results from it in that he puts his observed position very close to his dead reckoning position. The data was obtained by taking observations of the sun aboard a small ketch in the open sea off Catalina Island, California. The observations were also taken in December when the sun was almost at its maximum southerly declination. Wilson⁴¹ does not say what criteria he has used to fair his straight lines through the points on his graph which he reproduces in his paper but it would appear that another individual might just as well find a different result since the angle between the straight lines is small. This author, having faith in the general approach adopted by Ranta⁴², will endeavour to obtain data for himself at sea.

9.4 GENERALISATION OF THE METHOD OF RANTA⁴².

It would seem that at any time other than the time of culmination we can just as well take a series of observations of the altitude, α , of a heavenly body and, after fitting a least squares function approximation, $\alpha(t)$, through the data points, we would differentiate to find $\alpha'(t_0)$ at a time, t_0 , which we would choose for making an observed position. Substituting this value of $\alpha'(t_0)$ in equation (9.7) we could use this equation to find our first and subsequent approximations to Z_n combined then with equations (9.11)b and (9.11)c to form an iterative scheme for computing $(\lambda-\theta)$ and ϕ at time t_0 .

Let us consider first that the observer is stationary - that $V=0$ in equation (9.7). The equation then becomes

$$\frac{d\alpha}{dt} = [\sin(\lambda-\theta) \sin \phi \frac{d\chi}{dt} - \cos \phi \frac{d\lambda}{dt}] \sin Z - [\cos(\lambda-\theta) \frac{d\chi}{dt}] \cos Z$$

..... (9.12)

If we express equation (9.12) in the form

$$R \sin(Z-k) = \frac{d\alpha}{dt} \quad \text{then} \quad Z = k + \sin^{-1} \left[\frac{1}{R} \frac{d\alpha}{dt} \right]$$

where

$$\tan k = \frac{\cos(\lambda-\theta) \frac{d\chi}{dt}}{\sin(\lambda-\theta) \sin \phi \frac{d\chi}{dt} - \cos \phi \frac{d\lambda}{dt}}$$

$$R^2 = [\sin(\lambda-\theta) \sin \phi \frac{d\chi}{dt} - \cos \phi \frac{d\lambda}{dt}]^2 + [\cos(\lambda-\theta) \frac{d\chi}{dt}]^2$$

the procedure would then be to take a series of observations $\alpha_1, \alpha_2, \dots, \alpha_n$ at times t_1, t_2, \dots, t_n , respectively, such that $t_1 < t_0 < t_n$, to fit a function approximation $\alpha(t)$, and differentiate this to find $\alpha'(t)$.

Using the values of $\alpha(t_0)$ and $\alpha'(t_0)$ so found we can compute the angle Z as above and then the LHA ($= \lambda - \theta$) from

$$\sin(\lambda - \theta) = \frac{\sin Z \cos \alpha(t_0)}{\cos \chi} \dots\dots (9.13)$$

To a stationary observer on the surface of the Earth the altitude, α , of a heavenly body is given by equation (9.3) :

$$\sin \alpha = \sin \phi \sin \chi + \cos \phi \cos \chi \cos(\lambda - \theta)$$

If we write $y = \sin \alpha$ we may, therefore, be able to approximate efficiently with a sinusoidal function of the form

$$y = A \sin x + B \cos x + C$$

where $x = \lambda - \theta$. The period of the trigonometric functions is the length of the apparent day (the period elapsing between two successive transits of the observer's meridian by the heavenly body). If the body is observed at times (t_1) with corresponding altitudes (α_1) then we fit y to this data in the least squares sense so that

$$S = \sum (y_1 - (A \cos x_1 + B \sin x_1 + C))^2 \dots\dots (9.14)$$

is a minimum.

If we find the partial derivatives of S with respect to A, B and C then these are the "normal" equations and furnish a set of linear equations (9.15) from which to determine the A, B, C :

$$\begin{bmatrix} \Sigma \cos^2 x_i & \Sigma \cos x_i \sin x_i & \Sigma \cos x_i \\ \Sigma \cos x_i \sin x_i & \Sigma \sin^2 x_i & \Sigma \sin x_i \\ \Sigma \cos x_i & \Sigma \sin x_i & \Sigma 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} \Sigma y_i \cos x_i \\ \Sigma y_i \sin x_i \\ \Sigma y_i \end{bmatrix} \dots\dots (9.15)$$

There are certain circumstances where the matrix defining the system (9.15) is singular or nearly so. This occurs, for instance, when the hour angle $\lambda - \theta$ ($= x$) and the altitude α ($= \sin^{-1} y$) are both close to 45° . In such a case we will perhaps overcome this by using the least squares orthogonal polynomial approximations. (See Forsythe⁴³). This approach is very similar to that of Ranta⁴², but it does not restrict the polynomial approximation to quadratics.

9.5 EXPERIMENTAL RESULTS.

In the absence of observed data we can test the computational procedure by using the values of the altitude of the sun given in Davis's Tables⁴⁴. These tables are intended to give the values of the true altitude of the sun (correct to the nearest one minute of arc) at intervals of eight minutes but the tables are very old and some of the entries are not, in fact, correct to the nearest minute. We have, in fact, corrected any entry from Davis that we have found to be wrong. We have then applied the trigonometric least squares approximation to this data.

On the bridge of an ocean-going cargo ship in good observing conditions one would normally expect the readings from the sextant to give altitudes at better accuracy than one minute. The common form of the sextant used by seagoing navigators is usually read to 0.1 of a minute and, although one might not expect the altitudes always to accurate to that level, it might not be unreasonable to expect accuracy to 0.2 of a minute. Using the spherical cosine

formula we have therefore compiled a second table of altitudes correct to 0.25 of a minute in order to test the application of the Forsythe polynomials to our least squares approximation. The Forsythe polynomials did not give such good results when the data was expressed to the level of accuracy of one minute but gave acceptable results at the level of accuracy of 0.25 minutes.

For the purposes of testing the procedure of computing the position of an observer we have sited the observer on the Greenwich meridian (longitude 0°) and at latitudes ranging from 30° to 55° North. The three examples have been chosen to demonstrate the method in the extreme cases that are likely to occur. The first is chosen when the sun is at high altitude, the second is chosen when the sun is at low altitude and the third is chosen when the sun is changing its declination at the maximum rate.

EXAMPLE 1. June 10 1990 - 1100 GMT . Declination 23.0117°N. Rate of change of declination 0.001597°/hour. Hour Angle 345°9.9' Rate of change of Hour Angle -14.998 °/hour. Longitude 0°.

LAT	HOUR ANGLE (λ)					
	349°	347°	345°	343°	341°	339°
30°	77°56'	76°27'	74°53'	73°17'	71°39'	69°59'
35°	74°39'	73°31'	72°16'	70°57'	69°34'	68°09'
40°	70°38'	69°46'	68°48'	67°45'	66°38'	65°26'
45°	66°15'	66°35'	64°50'	64°00'	63°06'	62°08'
50°	61°40'	61°09'	60°34'	59°55'	59°11'	58°24'
55°	56°59'	56°35'	56°07'	55°36'	55°01'	54°24'

TABLE 9.5

The picture presented by the data of Table 9.5 is the "static" picture - the declination of the sun is kept fixed. Allowance for the effect upon the rate of change of altitude by the rate of change of declination is made by the added term

$$\left(\frac{\sin \phi \cos \chi - \cos \phi \sin \chi \cos(\lambda - \theta)}{\cos \alpha} \right) \left(\frac{d\chi}{dt} \right)$$

The results from the computation using the above data from Table 9.5 are shown below in Table 9.6.

LAT	TRUE ALTITUDE	COMPUTED ALTITUDE	TRUE RATE	CALC RATE	HOUR ANGLE	LONGITUDE
30°	75.0167	75.0180	11.8401	11.8377	14°49.8'	0°0.3'E
35°	72.3749	72.3724	9.5632	9.5662	14°50.5'	0°0.4'W
40°	68.8854	68.8858	7.5174	7.5194	14°50.3'	0°0.2'W
45°	64.8996	64.8971	5.8930	5.8964	14°50.7'	0°0.6'W
50°	60.6192	60.6192	4.6322	4.6256	14°48.7'	0°1.4'E
55°	56.1562	56.1577	3.6416	3.6575	14°54.0'	0°3.9'W

TABLE 9.6

Table 9.7 below corrects the altitudes of the sun given in Table 9.5 to 0.25 of a minute

LAT	HOUR ANGLE (λ)					
	349°	347°	345°	343°	341°	339°
30°	77°56¼'	76°26¼'	74°53¼'	73°17'	71°39'	69°59¼'
35°	74°39¼'	73°30¼'	72°16¼'	70°57'	69°34¼'	68°08¼'
40°	70°38'	69°46'	68°48¼'	67°45¼'	66°37¼'	65°26¼'
45°	66°15'	65°35¼'	64°50'	64°00¼'	63°06'	62°08'
50°	61°40¼'	61°09¼'	60°34'	59°54¼'	59°11¼'	58°24¼'
55°	56°58¼'	56°34¼'	56°07'	55°35¼'	55°01¼'	54°24'

TABLE 9.7

Table 9.8 shows the results from using the Forsythe polynomials on the data in Table 9.7

LAT	TRUE ALTITUDE	COMPUTED ALTITUDE	TRUE RATE	COMP RATE	HOOR ANGLE	LONGITUDE
30°	75.0167	75.0165	11.8401	11.8280	14°49.2'	0°0.9'E
35°	72.3749	72.3750	9.5632	9.5605	14°49.8'	0°0.3'E
40°	68.8854	68.8869	7.5174	7.5042	14°48.4'	0°1.7'E
45°	64.8996	64.8997	5.8930	5.9012	14°51.4'	0°1.3'W
50°	60.6192	60.6188	4.6322	4.6147	14°46.6'	0°3.5'E
55°	56.1562	56.1556	3.6416	3.6370	14°49.0'	0°1.1'E

TABLE 9.8

EXAMPLE 2. December 11 1990 - 1100 GMT - Declination 23°S - Rate of change of declination -0.003611 °/hour. Hour Angle 13°17.3' - Rate of change of HA -14.9605 °/hr.

LAT	HOOR ANGLE (λ)					
	9°	11°	13°	15°	17°	19°
30°	36°18'	35°57'	35°33'	35°05'	34°32'	33°57'
35°	31°22'	31°04'	30°42'	30°17'	29°48'	29°16'
40°	26°27'	26°10'	25°51'	25°28'	25°02'	24°33'
45°	21°30'	21°16'	20°58'	20°38'	20°15'	19°49'
50°	16°34'	16°21'	16°06'	15°48'	15°27'	15°05'
55°	11°37'	11°26'	11°13'	10°57'	10°39'	10°19'

TABLE 9.9

The results found using the data from Table 9.9 are shown in Table 9.10 below :

LAT	TRUE ALTITUDE	COMPUTED ALTITUDE	TRUE RATE	COMPUTED RATE	COMPUTED HR ANGLE	LONGITUDE
30°	35.4837	35.4855	3.3630	3.3690	13°18.7'	0°1.4'W
35°	30.6459	30.6450	3.0103	3.0173	13°17.2'	0°0.1'E
40°	25.7923	29.7906	2.6895	2.6883	13°16.9'	0°0.4'E
45°	20.9271	20.9273	2.3927	2.4042	13°21.7'	0°3.4'W
50°	16.0532	16.0546	2.1136	2.1202	13°19.7'	0°2.4'W
55°	11.1732	11.1742	1.8471	1.8478	13°17.6'	0°0.9'W

TABLE 9.10

EXAMPLE 3. Sept 21st 1990 - 1000 GMT. Declination $\chi=0.728333^{\circ}\text{N}$.
 Longitude $\theta=0^{\circ}$
 Hour Angle $\lambda=331^{\circ}42.3'$. $d\lambda/dt = -15.003$ $^{\circ}/\text{hour}$
 $d\chi/dt = -0.01618$ $^{\circ}/\text{hour}$

LAT	HOUR ANGLE (λ)					
	327°	329°	331°	333°	335°	337°
30°	47°06'	48°28'	49°48'	51°04'	52°18'	53°28'
35°	43°58'	45°11'	46°22'	47°29'	48°36'	49°35'
40°	40°35'	41°40'	42°42'	43°41'	44°37'	45°30'
45°	37°01'	37°57'	38°52'	39°43'	40°32'	41°17'
50°	33°17'	34°06'	34°53'	35°37'	36°19'	36°58'
55°	29°26'	30°08'	30°48'	31°26'	32°01'	32°34'

TABLE 9.11

The results obtained using the data from Table 9.11 are shown in Table 9.12 below :

LAT	TRUE ALTITUDE	COMPUTED ALTITUDE	TRUE RATE	CALC RATE	LHA	LONG
30°	50.2503	50.2507	9.6189	9.6333	28°20.4'	0°2.7'W
35°	46.7617	46.7645	8.4903	8.5017	20.1'	2.4'W
40°	43.0497	43.0521	7.4407	7.4389	17.3'	0.4'E
45°	39.1652	39.1631	6.4710	6.4698	17.4'	0.3'E
50°	35.1472	35.1459	5.5753	5.5731	17.0'	0.7'E
55°	31.0253	31.0273	4.7445	4.7414	16.4'	1.3'E

TABLE 9.12

9.6 CONCLUSION.

In good observing conditions at sea, the kind of conditions which would be necessary in any case for taking astronomical observations, the results above would be achievable. Indeed, a good observer would expect altitudes to be of greater accuracy than one minute of arc. Most micrometer sextants are read to 0.1' of arc and all computations assume this accuracy. As they stand, most of the results for the longitude found above would be very acceptable to an observer on an ocean passage.

PART TWO
DIRECT SPLINE APPROXIMATIONS
TO
INTEGRALS

- 10 -

DIRECT CUBIC SPLINE
APPROXIMATIONS
TO
INTEGRALS

10.1 INTRODUCTION.

The theme of the the Bachelor of Philosophy thesis by this author⁴⁵ was the development of algorithms from which to compute the integral of a function by fitting cubic splines .The thesis was conceived originally to describe the methods by which the cubic spline and the bicubic spline approximations could be applied to the problem of numerical integration in one and two dimensions. During the experiments it became apparent that it was possible to approximate to the integral of a function *directly*, that is, to fit a cubic spline to the indefinite integral of a function with given initial conditions. The result is an algorithm which is suitable for both manual or automatic computation. Further, the method can be applied to problems in both one or two dimensions and it can be shown that its error bounds are of the same order as those for the integrated cubic splines. We use the term Direct Cubic Spline Approximation for this algorithm and the analysis used in the thesis to prove the results which lead to the steps in the computational procedure was lengthy. It is clear now, however, that the Direct Cubic Spline Approximation could be developed by simply fitting a quadratic polynomial to the integrand and then integrating. In either approach, in addition to the function values of the integrand, an extra initial condition is required to start the computation: we will, therefore assume that the derivative of the integrand at the starting point is known. It will emerge, however, that such knowledge is not always critical in practice. In section 10.6, for instance, we show that if the subintervals are equal in length and their number is even, then this is, in theory, irrelevant. The method of the Direct Cubic Spline Approximation results in a 'step by step' method of numerical integration through which the integral between any two ordinates can be found and applicable where the ordinates are randomly spaced.

10.2 NUMERICAL INTEGRATION USING CUBIC SPLINES.

The usual method of computing the integral of a function $f(x)$ over an interval $[a,b]$ using spline functions is achieved by fitting a

spline function $s(x)$ to the integrand $f(x)$ over the interval $[a, b]$ at the points of subdivision when $[a, b]$ has been subdivided by points x_i ($i=1, \dots, n-1$) such that

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$$

and integrating $s(x)$ in place of $f(x)$.

The cubic spline is the popular spline function used for this purpose and, by applying the above procedure using the cubic spline, Davis & Rabinowitz⁴⁶ derive the formula

$$\int_a^b f(x) dx \approx \int_{x_0}^{x_n} s(x) dx = \sum_{i=1}^n \frac{h_i}{2} (f_{i-1} + f_i) - \sum_{i=1}^n \frac{h_i^3}{24} (m_{i-1} + m_i) \quad \dots (10.1)$$

where $f_i = f(x_i)$, $h_i = x_i - x_{i-1}$ and the m_i are the "Moments" of the spline - $m_i = s''(x_i)$.

The details of the computation of the moments of the cubic spline can be found in the books by Ahlberg, Nilson & Walsh⁴⁷ and P.M. Prenter⁴⁸. In this dissertation, however, we are concerned with *DIRECT* Spline Approximation to the integral function $F(x)$ in the interval $[a, b]$ where

$$F(x) = \int_a^x f(x) dx$$

and $f(x)$ belongs to a suitable continuity class.

We begin with the Direct Cubic Spline Approximation and develop a computational scheme which is both efficient and easy to apply.

10.3 DEFINITION OF THE DIRECT CUBIC SPLINE APPROXIMATION.

In the B.Phil thesis by Williams⁴⁵ a method of computing the integral

$$\int_a^b f(x) dx \quad \dots\dots (10.2)$$

where $f(x)$ is a continuous function which is assumed to have a continuous derivative, is found by fitting a cubic spline approximation $S(x)$ directly to the function $F(x)$ over the interval $[a, b]$ subdivided by points (x_i) ($i=1, \dots, n-1$) where

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$$

and where

$$F(x) = \int_a^x f(x) dx \quad (x \in [a, b]) \quad \dots (10.3)$$

$S(x)$ is referred to as the DIRECT CUBIC SPLINE Approximation to the function $F(x)$.

In the paper by Pythian & Williams⁴⁹, to determine $S(x)$, we defined, at each mesh point, x_i , a "Moment", M_i , where $M_i = S''(x_i)$ and then expressed $S''(x)$ as a piecewise continuous linear function so that, in each sub-interval $[x_{i-1}, x_i]$, we have

$$S''(x) = \frac{M_{i-1}}{h_i} (x_i - x) + \frac{M_i}{h_i} (x - x_{i-1}) \quad \dots\dots (10.4)$$

where $h_i = x_i - x_{i-1}$.

We then integrated twice and determined the constants of integration but this is a lengthy procedure and it is apparent that, since $S(x)$ is cubic, we can approximate to $f(x)$ using a suitable quadratic polynomial.

In the interval $[x_0, x_1]$ let $f(x)$ be approximated by the quadratic polynomial $q(x)$ where

$$q(x) = f(x_0) + (x-x_0)M_0 + A(x-x_0)^2 \quad \dots\dots (10.5)$$

This ensures that $q(x_0) = f(x_0)$ and $q'(x_0) = M_0 = f'(x_0)$.

We also have

$$q(x_1) = f(x_1) = f(x_0) + h_1 M_0 + Ah_1^2 \quad \dots\dots (a)$$

$$q'(x_1) = M_1 = M_0 + 2Ah_1 \quad \dots\dots (b)$$

so that, eliminating A from the equations (a) and (b) gives the relationship

$$M_1 = 2 \left(\frac{f_1 - f_0}{h_1} \right) - M_0$$

In any interval $[x_{i-1}, x_i]$ we can similarly define a quadratic polynomial $q_i(x)$ using the values f_{i-1} , f_i and M_{i-1}

(where $f_i = f(x_i)$) so that

$$q_i(x) = f_{i-1} + (x-x_{i-1})M_{i-1} + A_i(x-x_{i-1})^2 \quad \dots\dots (10.6)$$

and thereby find the recursive scheme from which to determine the M_i

$$M_i = 2 \left(\frac{f_i - f_{i-1}}{h_i} \right) - M_{i-1} \quad \dots\dots (10.7)$$

If we now integrate $q_i(x)$ over the interval $[x_{i-1}, x_i]$ we find

$$S_i - S_{i-1} = h_i f_{i-1} + \frac{1}{2} M_{i-1} h_i^2 + \frac{1}{6} A_i h_i^3$$

where $S_i = S(x_i)$.

We find that

$$A_i = \left(\frac{M_i - M_{i-1}}{2h_i} \right)$$

so that

$$S_i - S_{i-1} = \frac{h_i^2}{6} (M_i + 2M_{i-1}) + h_i f_{i-1} \quad \dots\dots (10.8)$$

This is the form of the recursive scheme for S_1 that was first arrived at. If, however, we find $q_1(x_1)$ from equation (10.6), multiply this by $\frac{1}{2}h_1$, substitute for A_1 and add this to equation (10.8) we find the "symmetric" form of the recursive scheme for S_1 :

$$S_1 - S_{1-1} = \frac{h_1}{2}(f_1 + f_{1-1}) - \frac{h_1^2}{12}(M_1 - M_{1-1})$$

and this is the form that we now use in the computational scheme. Not only does this provide a more efficient algorithm for computing the S_1 but it also facilitates the estimation of its error bound.

10.4 THE COMPUTATIONAL SCHEME.

The method of computation used to evaluate integral in Equation (10.1) using the Direct Cubic Spline Approximation on the mesh

$$a = x_0 < x_1 < \dots < x_n = b$$

is given by

(i) $S_0 = 0 : M_0 = f'(a)$

(ii) $h_1 = x_1 - x_{1-1}$

(iii) $M_1 = \frac{2}{h_1}(f_1 - f_{1-1}) - M_{1-1}$ for $i=1, \dots, n$

(iv) $S_1 = S_{1-1} + \frac{h_1}{2}(f_1 + f_{1-1}) - \frac{h_1^2}{12}(M_1 - M_{1-1})$

..... (10.9)

Ultimately

$$S_n = \int_a^b f(x) dx + E_n$$

where E_n is the error term.

It can be shown by direct summation that the scheme (10.8) is equivalent to

$$S_n = \frac{1}{2} \sum_{i=1}^n h_i (f_i + f_{i-1}) - \frac{1}{12} \sum_{i=1}^n h_i^2 (M_i - M_{i-1}) \dots\dots (10.10)$$

Equation (10.10) is of the same form as the truncated Euler MacLaurin integration formula and is in "symmetric" form. If $f \in C^4[a,b]$ and we expand the integral

$$\int_{x_{i-1}}^{x_i} f(x) dx$$

using the Euler-MacLaurin expansion formula (which can be found in a book such as that by Scheid⁵⁰), summing over all the intervals gives the form

$$\int_a^b f(x) dx = \sum_{i=1}^n \left[\frac{h_i}{2} (f_i + f_{i-1}) - \frac{h_i^2}{12} (f'_i - f'_{i-1}) \right] + \sum_{i=1}^n \frac{h_i^4}{720} (f^{(3)}_i - f^{(3)}_{i-1}) + O(h^7) \dots\dots (10.11)$$

where h is the maximum length of any subinterval, h_i , in the interval $[a,b]$.

If we compare Equations (10.10) and (10.11) we see that the first two terms of the Euler-MacLaurin Integral Expansion have the same coefficients as the Direct Cubic Spline Approximation.

10.5 ERROR ESTIMATE FOR THE DIRECT CUBIC SPLINE

If, then, we subtract Equation (10.11) from (10.10), we find

$$\begin{aligned}
 S_n - \int_a^b f(x) dx &= \frac{1}{12} \sum_{i=1}^n h_i^2 [(f'_i - M_i) - (f'_{i-1} - M_{i-1})] \\
 &= \frac{1}{720} \sum_{i=1}^n h_i^4 (f_i^{(3)} - f_{i-1}^{(3)}) + O(h^7) \\
 &\dots\dots (10.12)
 \end{aligned}$$

Let us define

$$\begin{aligned}
 F_1 &= \int_a^{x_1} f(x) dx \\
 E_1 &= S_1 - F_1 \\
 e_1 &= f'_1 - M_1
 \end{aligned}$$

From the computational scheme (10.9) we have

$$S_1 - S_{1-1} = \frac{h_1}{2} (f_1 + f_{1-1}) - \frac{h_1^2}{12} (M_1 - M_{1-1})$$

and the Euler MacLaurin integral expansion gives

$$\begin{aligned}
 F_1 - F_{1-1} &= \frac{h_1}{2} (f_1 + f_{1-1}) - \frac{h_1^2}{12} (f'_1 - f'_{1-1}) + \frac{h_1^4}{720} (f_i^{(3)} - f_{i-1}^{(3)}) \\
 &\quad + O(h^5)
 \end{aligned}$$

So that if we write

$$\begin{aligned}
 E_1 &= S_1 - F_1 && \text{(the error in } S_1) \\
 e_1 &= f'_1 - M_1 && \text{(the error in } M_1)
 \end{aligned}$$

we have

$$\begin{aligned}
 E_1 - E_{1-1} &= \frac{h_1^2}{12} (e_1 - e_{1-1}) - \frac{h_1^4}{720} (f_i^{(3)} - f_{i-1}^{(3)}) + O(h^7) \\
 &\dots\dots (10.13)
 \end{aligned}$$

Let us suppose that

$$e_1 = f'_1 - M_1 = A_1 f_i^{(3)} + B_1 f_i^{(4)} + C_1 f_i^{(5)} + \dots\dots (10.14)$$

From $M_1 + M_{1-1} = \frac{2}{h_1} (f_1 - f_{1-1})$

we have $e_1 + e_{1-1} = (f'_1 + f'_{1-1}) - \frac{2}{h_1} (f_1 + f_{1-1})$

Expanding this about $x=x_{i-n}$ writing $t=nh_1$ and $f=f_{i-n}$ etc.,

$$\begin{aligned}
 e_i + e_{i-1} &= [(f' + tf'' + \frac{1}{2}t^2f''' + \dots) + (f' - tf'' + \frac{1}{6}t^3f^{(3)} + \dots)] \\
 &\quad + \frac{1}{t} [(f + tf' + \frac{1}{2}t^2f'' + \frac{1}{6}t^3f^{(3)} + \dots) \\
 &\quad \quad - (f - tf' + \frac{1}{2}t^2f'' - \frac{1}{6}t^3f^{(3)} + \dots)] \\
 &= 2(f' + \frac{1}{3}t^2f^{(3)} + \frac{1}{24}t^4f^{(5)} + \dots - f' - \frac{1}{6}t^3f^{(3)} - \frac{1}{120}t^5f^{(5)} - \dots) \\
 &= \frac{2}{3}t^2f^{(3)} + \frac{1}{15}t^4f^{(5)} + \dots
 \end{aligned}$$

giving
$$e_i + e_{i-1} = \frac{h_1^2}{6} f_{i-n}^{(3)} + \frac{h_1^4}{240} f_{i-n}^{(5)} + O(h^6) \dots \dots \dots (10.15)$$

If we now expand the expression on the right hand side of Equation (10.14) about $x=x_{i-n}$ using the same notation for t and f we find

$$\begin{aligned}
 e_i + e_{i-1} &= \\
 &A_i[f^{(3)} + tf^{(4)} + \frac{1}{2}t^2f^{(5)} \dots] + A_{i-1}[f^{(3)} - tf^{(4)} + \frac{1}{2}t^2f^{(5)} \dots] \\
 &+ B_i[f^{(4)}t + tf^{(5)} + \frac{1}{2}t^2f^{(6)} \dots] + B_{i-1}[f^{(4)} - tf^{(5)} + \frac{1}{2}t^2f^{(6)} \dots] \\
 &+ C_i[f^{(5)} + tf^{(6)} + \frac{1}{2}t^2f^{(7)} \dots] + C_{i-1}[f^{(5)} - tf^{(6)} + \frac{1}{2}t^2f^{(7)} \dots] \\
 &+ \dots \dots \dots (10.16)
 \end{aligned}$$

Comparing this with Equation (10.15) we find the relationships

$$A_i + A_{i-1} = \frac{h_1^2}{6}$$

$$\frac{h_1}{2} (A_i - A_{i-1}) + (B_i - B_{i-1}) = 0$$

$$\frac{h_1^2}{8} (A_i + A_{i-1}) + \frac{h_1}{2} (B_i - B_{i-1}) + (C_i + C_{i-1}) = \frac{h_1^4}{240}$$

and from these relationships we can generate the A_i , B_i and C_i once we have suitable starting values.

We can see that $e_i = f'_i - M_i$

$$\begin{aligned}
 &= f'_i - \frac{2}{h_1}(f_i - f_0) + f'_0 \\
 &= (f'_i + f'_i - h_1f''_i + \frac{h_1^2}{2}f'''_i - \frac{h_1^3}{6}f^{(4)}_i + \dots) \\
 &\quad - \frac{2}{h_1}[f_i - (f_i - h_1f'_i + \frac{h_1^2}{2}f''_i - \frac{h_1^3}{6}f^{(3)}_i + \frac{h_1^4}{24}f^{(4)}_i + \dots)]
 \end{aligned}$$

$$= \frac{h_1^2}{6} f^{(3)} - \frac{h_1^3}{12} f^{(4)} + \dots$$

This gives us (since $A_0 = 0$) $A_1 = \frac{h_1^2}{6}$

$$\text{and } A_1 = \frac{1}{6}(h_1^2 - h_{1-1}^2 + \dots \pm h_2^2 \mp h_1^2)$$

Now equation (1.13) gives us

$$E_1 - E_{1-1} = \frac{h_1^2}{12} (e_1 - e_{1-1}) - \frac{h_1^4}{720} (f^{(3)} - f^{(3)}) + O(h^7)$$

which, since $E_0 = 0$ and $e_0 = 0$ yields directly

$$E_1 = \frac{h_1^4}{72} f_{1-1}^{(3)} - \frac{h_1^5}{120} f_{1-1}^{(4)} + O(h^6)$$

When $i > 1$ and we expand about $x = x_{i-1}$ and using equation (10.14) we find

$$E_i - E_{i-1} = \frac{h_i^2}{12} (A_i - A_{i-1}) f_{i-1}^{(3)} - \frac{h_i^5}{720} f_{i-1}^{(4)} + O(h^7) \quad \dots \dots (10.17)$$

so that

$$E_i - E_{i-1} = \frac{h_i^2}{72} (h_i^2 - 2h_{i-1}^2 + \dots + 2h_2^2 \pm 2h_1^2) f_{i-1}^{(3)} + O(h^6)$$

$$\text{Now } E_1 = \sum_{r=1}^1 (E_r - E_{r-1}) \quad (E_0 = 0)$$

hence

$$E_1 = \sum_{r=2}^1 \left[\frac{h_r^2}{72} f_{i-1}^{(3)} (h_r^2 - 2h_{r-1}^2 + \dots \mp 2h_2^2 \pm 2h_1^2) \right] + E_1 + O(h^6)$$

If h is the maximum absolute value of subinterval and $f^{(4)}$ is the maximum absolute value of $f^{(4)}(x)$ in $[a, b]$ then

$$|E_1| \leq \frac{h^2}{72} f^{(3)} \left| \sum_{r=2}^1 (h_r^2 - 2h_{r-1}^2 + \dots \mp 2h_2^2 \pm 2h_1^2) \right| + |E_1| + O(h^6) \quad \dots \dots (10.18)$$

$$\text{Now } \sum_{r=2}^1 (h_r^2 - 2h_{r-1}^2 + \dots \mp 2h_2^2 \pm 2h_1^2)$$

$$= \sum_{r=2}^i h_r^2 + 2 \sum_{r=2}^{i-1} \sum_{j=1}^r (-1)^{(r-j+1)} h_j^2$$

and the second term of this is a triangular sum. Hence

$$\begin{aligned} & \sum_{r=2}^{i-1} \sum_{j=1}^r (-1)^{(r-j+1)} h_j^2 \\ = & -h_{i-1}^2 + h_{i-2}^2 - h_{i-3}^2 + \dots \pm h_3^2 \mp h_2^2 \pm h_1^2 \\ & -h_{i-2}^2 + h_{i-3}^2 - \dots \mp h_3^2 \pm h_2^2 \mp h_1^2 \\ & -h_{i-3}^2 + \dots \pm h_3^2 \mp h_2^2 \pm h_1^2 \\ & \dots \\ & -h_3^2 + h_2^2 - h_1^2 \\ & -h_2^2 + h_1^2 \\ & -h_1^2 \\ = & -h_{i-1}^2 - h_{i-3}^2 - \dots - h_2^2 \quad \text{if } i \text{ is even} \\ = & -h_{i-1}^2 - h_{i-3}^2 - \dots - h_1^2 \quad \text{if } i \text{ is odd.} \end{aligned}$$

Now we can write the inequality (10.18) in the form

$$|E_1| \leq \frac{h^2}{72} f^{(3)} \left| \sum_{r=1}^i h_r^2 + 2 \sum_{r=2}^{i-1} \sum_{j=1}^r (-1)^{(r-j+1)} h_j^2 \right| + O(h^5)$$

and, if i is even, we find

$$\begin{aligned} & \left| h_i^2 + h_{i-1}^2 + h_{i-2}^2 + h_{i-3}^2 + \dots + h_3^2 + h_2^2 + h_1^2 \right. \\ & \quad \left. - 2h_{i-1}^2 \quad - 2h_{i-3}^2 \quad - 2h_2^2 \right| \\ = & \left| (h_i^2 - h_{i-1}^2) + (h_{i-2}^2 - h_{i-3}^2) + \dots + (h_3^2 - h_2^2) + h_1^2 \right| \leq \frac{1}{2} h^2 \end{aligned}$$

or, if i is odd;

$$\begin{aligned} & \left| h_i^2 + h_{i-1}^2 + h_{i-2}^2 + h_{i-3}^2 + \dots + h_3^2 + h_2^2 + h_1^2 \right. \\ & \quad \left. - 2h_{i-1}^2 \quad - 2h_{i-3}^2 \quad - 2h_3^2 \quad - 2h_1^2 \right| \\ = & \left| (h_i^2 - h_{i-1}^2) + (h_{i-2}^2 - h_{i-3}^2) + \dots + (h_2^2 - h_1^2) \right| \leq \frac{1}{2} h^2 \end{aligned}$$

so that this gives us

$$|E_1| \leq \frac{1h^4}{144} f^{(3)} + \frac{1h^5}{720} f^{(4)} + O(h^7)$$

When h is constant simplifications of this error bound take place in particular cases (as will be shown in the following sections) but it has been pointed out that there is also one interesting case which occurs when we make the arbitrary choice $A_n = h^2/12$ for then we find that $A_i = h^2/12$ for all i and equation (10.17) yields :

$$E_1 = \sum_{r=1}^1 (E_r - E_{r-1}) = \frac{h^2}{12} \sum_{r=1}^1 (A_r - A_{r-1}) f_{r-h}^{(3)} - \frac{h^5}{720} \sum_{r=1}^1 f_{r-h}^{(4)} + O(h^7)$$

$$E_1 = - \frac{h^5}{720} \sum_{r=1}^1 f_{r-h}^{(4)} + O(h^7)$$

It is not clear whether or not this can be put to practical advantage in any way but it does pose an interesting question.

10.6 RELATIONSHIPS BETWEEN THE DIRECT CUBIC SPLINE APPROXIMATION AND OTHER INTEGRATION RULES.

In the special case where h is constant and n is fixed, the formula that results from (10.9) is

$$S_n = \frac{h}{2} \sum_{i=1}^n (f_i + f_{i-1}) - \frac{h^2}{12} (M_n - f'_0) \quad \dots \dots (10.19)$$

Under the same circumstances the Euler-Maclaurin formula for the function $f(x) \in C^4[a, b]$ is

$$\int_a^b f(x) dx = \frac{h}{2} \sum_{i=1}^n (f_i + f_{i-1}) - \frac{h^2}{12} (f'_n - f'_0) + \frac{nh^5}{720} f^{(4)}(\chi)$$

for some $\chi : a \leq \chi \leq b$.

If f'_n is known, and we replace M_n in (10.19) by this known value, then this will give us the truncated form of the Euler-Maclaurin formula and then the error in S_n , if $f(x) \in C^4[a, b]$, as an approximation to the integral of $f(x)$ in $[a, b]$ is

$$\frac{(b-a)h^4}{720} f^{(4)}(\chi)$$

It is interesting to note too that if we consider the usual cubic spline approximation, $s(x)$, to $f(x)$ in $[a, b]$, it was shown in the B.Phil thesis by Williams⁴⁵ (p.10) that the formula for the integral given by Davis & Rabinowitz⁴⁶ can be written as

$$\int_a^b f(x) dx \approx \int_a^b s(x) dx = \frac{h}{2} \sum_{i=1}^n (f_i + f_{i-1}) - \frac{h^2}{12} (f'_n - f'_0)$$

so that the formula for the Direct Cubic Spline Approximation is, when f'_n is known, the same as that for the integrated Cubic Spline.

Moreover, if we consider the special case of the scheme using equation (10.8) or equation (10.9) for S_i computed for $n=2$ then we find that

$$S(a) = 0 : M_0 = f'(a) : h = \frac{1}{2}(b-a)$$

$$M_i = 2(f_i - f_{i-1})/h - M_{i-1}$$

$$S_i = S_{i-1} + h^2(M_i + 2M_{i-1})/6 + hf_{i-1}$$

$$\text{or } S_i = S_{i-1} + \frac{h}{2} (f_i + f_{i-1}) - \frac{h^2}{12} (M_i - M_{i-1}) \quad i=1,2$$

is exactly Simpson's First Rule so that

$$S_2 = \frac{h}{3} (f_0 + 4f_1 + f_2)$$

See Phythian and Williams⁵¹.

The above case illustrates the more general rule that, in the equal interval case, if the number of intervals, n , is even, then S_n is independent of $f'(a)$ so that, in theory, provided n is always even, any value of $f'(a)$ would serve as the initial value of M_0 . To prove this let

$$M_{i+1} + M_i = \frac{2(f_{i+1} - f_i)}{h}$$

$$\text{and } M_i + M_{i-1} = \frac{2(f_i - f_{i-1})}{h}$$

Adding these gives

$$\begin{aligned} M_{i+1} + 2M_i &= \frac{2(f_{i+1} - f_i)}{h} + \frac{2(f_i - f_{i-1})}{h} - M_{i-1} \\ &= \frac{2(f_{i+1} - f_i)}{h} + \frac{2(f_i - f_{i-1})}{h} - \frac{2(f_{i-1} - f_{i-2})}{h} + M_{i-2} \\ &= \frac{2(f_{i+1} - f_i)}{h} + \sum_{r=1}^i \frac{(-1)^{i-r} 2(f_r - f_{r-1})}{h} + (-1)^{i-r+1} f'_0 \end{aligned}$$

for $i \geq 1$.

When $i=1$ we have

$$M_1 + 2f'_0 = \frac{2(f_1 - f_0)}{h} + f'_0$$

The integral S_n is found from

$$S_n = \sum_{i=0}^{n-1} \left(\frac{h^2}{6} (M_{i+1} + 2M_i) + hf'_i \right) \dots\dots (10.20)$$

and, when the sum over i is taken, the terms $(-1)^{i-n+1}f'_i$ cancel in pairs so that, if n is even, the aggregate of these terms is zero. It can be shown from (10.20) that, when $n=4$,

$$S_4 = \frac{h}{3} (f_4 + 4f_3 + 2f_2 + 4f_1 + f_0)$$

which is the compound form of Simpson's Rule and, for all n while $2nh$ is less than the length of the interval of integration, we also find

$$S_{2n} = \frac{h}{3} (f_{2n} + 4f_{2n-1} + 2f_{2n-2} + \dots + 2f_2 + 4f_1 + f_0)$$

Conversely, if a function $f(x)$, which is assumed to have a continuous first derivative and which is defined over an interval which is subdivided by a regular mesh, is integrated numerically by the compound form of Simpson's Rule, then the values (S_{2i}) so determined are, in fact, node values at the even numbered node points of the Direct Cubic Spline Approximation to the Integral of $f(x)$ over the same mesh. The Direct Cubic Spline, in a sense therefore, "interpolates" Simpson's Rule and the interpolating function, $S(x)$, is given by

$$S(x) = S_{i-1} + \frac{M_i}{6h} (x-x_{i-1})^3 + \frac{M_{i-1}}{2h} [h^2(x-x_{i-1})^2 - \frac{1}{3}h^3 + \frac{1}{3}(x-x_{i-1})^3] + (x-x_{i-1})f_{i-1}$$

10.7 THE PERFORMANCE OF THE DIRECT CUBIC SPLINE APPROXIMATION.

If we compare the error bound (10.14) of the Direct Cubic Spline Approximation with the error term for Simpson's First Rule over two intervals then we find that they are of the same form. Also, in light of the above relationship between the Direct Cubic Spline computational scheme, with h constant and $n=2$, and Simpson's First Rule, we can infer that the Direct Cubic Spline Approximation to an Integral Function is a generalisation of Simpson's First Rule.

The advantages that the Direct Cubic Spline Approximation has over other similar integration rules (Simpson and low order Gaussian rules) is that, for the Direct Cubic Spline, the ordinates do not have to be defined at specific points, the integral can be computed step by step between any two ordinates and the number of ordinates need not be specified in advance.

On the debit side, however, we find that, employing computational scheme (10.8) for the Direct Cubic Spline, since $S(x)$ is cubic, the integrand, $f(x)$ is approximated by a piecewise quadratic polynomial and therefore its derivative, $f'(x)$, by a piecewise straight line. Experience shows that the method gives its best results over intervals where the derivative, $f'(x)$, is monotonic. We can, however, generalise by going one step further - fitting a direct quartic spline to the integral and using the derivative computed from the X-Spline of Clenshaw and Negus^{5,2}. This is dealt with in the next chapter.

10.8 SOME APPLICATIONS OF THE DIRECT CUBIC SPLINE APPROXIMATION.

A number of applications of the use of the Direct Cubic Spline have appeared in Part One of this thesis. The method is particularly useful in the procedure used to compute the points along the arcs of geodesic curves on the surfaces of the Sphere and the Spheroid. For Navigational purposes it is quite usual, in practice, to compute points step by step along these arcs and the distances between them and, since we also know the boundary derivatives in most cases, the method suits the situation well. There are also some applications in ship technology to which the method is very well suited.

- 11 -

**HIGHER ORDER DIRECT SPLINE
APPROXIMATIONS
TO
INTEGRALS**

11.1 INTRODUCTION.

Having achieved some measure of success with the Direct Cubic Spline approximation to an integral it would seem logical to investigate whether or not any higher order splines can be used in the same way to compute the values of integrals by a recursive scheme similar to that developed for the Cubic Spline and whether or not any improvement in results can be achieved thereby. Experience shows, for instance, that the cubic spline gives best results when the first derivative of the integrand is monotonic - the computational scheme (1.9) does not respond well when a point of inflexion in the integrand occurs in the interval of integration - a higher order spline may well give better results in this respect.

11.2 DERIVATION OF THE DIRECT QUARTIC SPLINE APPROXIMATION .

Let $S(x)$ be the Quartic Spline Approximation to the function $F(x)$ in the interval $[a, b]$ where

$$F(x) = \int_a^x f(t) dt$$

and $f(x)$ is assumed to be continuous with continuous first and second derivatives.

Let the interval $[a, b]$ be subdivided by points x_1, x_2, \dots, x_{n-1} such that

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b .$$

$S(x)$ is the piecewise quartic polynomial in $[a, b]$ such that, in the subinterval $[x_{i-1}, x_i]$

$$S(x) = s_i(x) = a_i x^4 + b_i x^3 + c_i x^2 + d_i x + e_i$$

and $S(x)$ has continuous derivatives $S'(x)$, $S''(x)$ and $S^{(3)}(x)$ throughout $[a, b]$. Since the indefinite integral $S(x)$ is a quartic polynomial then the integrand will be a cubic. In the interval $[x_0, x_1]$ this cubic can be suitably defined by $c_0(x)$ where

$$c_0(x) = f_0 + M_0(x-x_0) + \frac{1}{2}N_{1-1}(x-x_0)^2 + A(x-x_0)^3 \dots (11.1)$$

this ensures that

$$c_0(x_0) = f_0$$

$$c_0'(x_0) = M_0 = f_0'$$

$$c_0''(x_0) = N_0 = f_0''$$

Where, in addition to the function values at the mesh points we assume that at $x=x_0$ the values of $f'(x_0)$ and $f''(x_0)$ are known. If we now differentiate $c_0(x)$ twice and determine $c_0''(x_1)$ we find

$$A = \frac{(N_1 - N_0)}{6h_1}$$

where $h_1 = x_1 - x_0$.

From $c_0(x_1)$ and $c_0'(x_1)$ we then find recursive equations through which we can determine N_1 and M_1 respectively. We then integrate $c_0(x)$ so that

$$S_1 = \int_{x_0}^{x_1} c_0(x) dx$$

We find: $N_1 = \frac{6}{h_1^2}(f_1 - f_0) - \frac{6}{h_1} M_0 - 2N_0$

$$M_1 = M_0 + \frac{h_1}{2} (N_1 - N_0)$$

$$S_1 = S_0 + \frac{h_1^3}{24} (N_1 + 3N_0) + \frac{h_1^2}{2} M_0 + h_1 f_0 \quad \dots \dots (11.2)$$

11.3 COMPUTATIONAL SCHEME FOR THE DIRECT QUARTIC SPLINE APPROXIMATION.

Given the initial values

$$S_0 = 0 ; f_0 = f(x_0) ; M_0 = f'(x_0) ; N_0 = f''(x_0)$$

then we can compute N_1 , M_1 and S_1 from the computational scheme (11.2) and then, if we define a suitable cubic polynomial $c_1(x)$ in the interval $[x_{1-1}, x_1]$:

$$c_1(x) = f_{1-1} + N_{1-1}(x-x_{1-1}) + M_{1-1}(x-x_{1-1})^2 + A_1(x-x_{1-1})^3$$

then we find a possible scheme for computing the S_i .

The scheme here is :

$$(i) \quad S_0 = 0 ; f_0 = f(x_0) ; M_0 = f'(x_0) ; N_0 = f''(x_0)$$

$$(ii) \quad h_i = x_i - x_{i-1}$$

$$(iii) \quad N_i = \frac{6}{h_i^2}(f_i - f_{i-1}) - \frac{6}{h_i}M_{i-1} - 2N_{i-1}$$

$$(iv) \quad M_i = M_{i-1} + \frac{h_i}{2}(N_i + N_{i-1}) \quad i=1, \dots, n$$

$$(v) \quad S_i = S_{i-1} + \frac{h_i^3}{24}(N_i + 3N_{i-1}) + \frac{h_i^2}{2}M_{i-1} + h_i f_{i-1} \quad \dots \quad (11.3)$$

where, ultimately,
$$S_n = \int_a^b f(x) dx + E_n$$

and E_n is the error term.

11.4 EXPERIMENTAL RESULTS FOR DIRECT QUARTIC SPLINE APPROXIMATION.

The results found from applying the computational scheme (11.3) are rather disappointing. It would seem that this recursive scheme is unstable for the computation of the N_i as i increases although it is quite good in the early stages.

When $f(x) = \sin x$, for instance, we compute the approximation

$$S(x) \approx F(x) = \int_0^{4\pi} \sin x dx = 0.2928932188$$

using 5 intervals each of length $\pi/20$ and find that $S_5 = 0.292895509$ which gives an error of 0.00000229. Yet, when we compute the same integral, using 10 intervals each of length $\pi/40$, we find that $S_{10} = 0.2928647048$ which gives an error of 0.000028514 - a ten fold increase in the error for half the interval spacing! However, if we modify the method, taking advantage of the self starting capability and then, when $i \geq 3$, computing $L'(x_i)$ and $L''(x_i)$ - the first and second derivatives of the cubic Lagrange Polynomial approximation to $f(x)$ over the

interval $[x_{i-1}, x_i]$ and setting $M_i = L'(x_i)$, $N_i = L''(x_i)$ then we find

$$S_{\infty} = 0.292878578$$

and $S_{1,0} = 0.2928912169$

which is more in line with what we might have expected. It should be noticed though that the S_{∞} computed this way is not so good as the S_{∞} found by applying the scheme (11.3) without modification. The problem, without doubt, seems to be that the equation (iii) of scheme (11.3) is unstable as i increases. If, however, we modify the scheme in such a way that

$$(i) \quad M_i = \frac{2}{h_i} (f_i - f_{i-1}) - M_{i-1},$$

$$(ii) \quad N_i = \frac{2}{h_i} (M_i - M_{i-1}) - N_{i-1},$$

..... (11.4)

and so that equation (v) of scheme (11.3) remains unchanged then we will find

$$S_{\infty} = 0.2929580156$$

and $S_{1,0} = 0.2929128574$.

Although the result for $S_{1,0}$ is an improvement, and, in further experiments, as h is made smaller, the results show a definite tendency to converge, these results are not as good, even so, as the results obtained by applying scheme (10.9) for the Direct Cubic Spline defined on the same interval with the same mesh.

If, then, following along the lines taken in chapter 10, we look for the "symmetric" form for the Direct Quartic Spline approximation, we find

$$S_i = S_{i-1} + \frac{h_i}{2} (f_i + f_{i-1}) - \frac{h_i^2}{4} (M_i - M_{i-1}) + \frac{h_i^3}{12} (N_i + N_{i-1})$$

..... (11.5)

but then, substituting

$$N_i + N_{i-1} = \frac{2}{h_i} (M_i - M_{i-1}) \quad \text{in (11.5) results in}$$

$$S_i = S_{i-1} + \frac{h_i}{2} (f_i + f_{i-1}) - \frac{h_i^2}{12} (M_i - M_{i-1}) \quad \text{..... (11.6)}$$

which is exactly the same as the "symmetric" form of the Direct CUBIC Spline approximation since $M_i = S''(x_i)$.

This would imply that the Direct Quartic Spline approximation *does not exist independently*. Indeed, as it will be shown below, every even order Direct Spline Approximation can, in the "symmetric" form, be shown to be the same as the lower order Direct Spline Approximation, and this would imply, generally, that *none* of the even order Direct Spline Approximations exist independently.

This, however, prompts the question as to whether there might be another way in which we can improve our estimate of the derivative of the spline given at the mesh points by the M_i . One possibility is given by Clenshaw & Negus⁴¹

In an interval $[x_0, x_n]$ subdivided by points x_1, x_2, \dots, x_{n-1} such that

$$x_0 < x_1 < \dots < x_{n-1} < x_n$$

We compute

$$h_i = x_i - x_{i-1}$$

$$\beta_i = \frac{h_{i+1}}{h_i + h_{i+1}}$$

$$Y_i = \frac{f_i - f_{i-1}}{h_i}$$

then

$$M_i = (1-\beta_i)^2 Y_{i+1} + \beta_i(3-\beta_i) Y_i - \beta_i M_{i-1}, \dots \quad (11.7)$$

We find this by fitting a suitable cubic polynomial approximation to $f(x)$ in the interval $[x_{i-1}, x_{i+1}]$ where we have the given values f_{i-1}, f_i, f_{i+1} and a computed value of M_{i-1} . In the interval $[x_0, x_2]$ these values will be f_0, f_1, f_2 and f'_0 .

Let the cubic polynomial be $c_i(x)$ where

$$c_i(x) = f_{i-1} + M_{i-1}(x-x_{i-1}) + A(x-x_{i-1})^2 + B(x-x_{i-1})^3$$

This ensures that $c_i(x_{i-1}) = f_{i-1}$ and $c'_i(x_{i-1}) = M_{i-1}$.

When we equate $c_i(x)$ at $x=x_i$ and $x=x_{i+1}$ to f_i and f_{i+1} , respectively, then we find simultaneous linear equations from which to determine the constants A and B. We omit the details here.

The computational scheme which results when we use this form of the "moment", M_i , as found from the limiting form of the Clenshaw-Negus X-spline is given by the scheme which leads to the equation (11.7) with the symmetric form of the quartic/cubic spline approximation to S_i :

We compute

$$h_i = x_i - x_{i-1}$$

$$\beta_i = \frac{h_{i-1}}{h_i + h_{i+1}}$$

$$Y_i = \frac{f_i - f_{i-1}}{h_i}$$

$$M_i = (1-\beta_i)^2 Y_{i+1} + \beta_i(3-\beta_i) Y_i - \beta_i M_{i-1}$$

$$S_i = S_{i-1} + \frac{h_i}{2} (f_i + f_{i-1}) - \frac{h_i^2}{12} (M_i - M_{i-1})$$

..... (11.8)

The results found from experiments with constant h , conducted using the two computational schemes; (10.9) and (11.8), are shown in Table 11.1.

	n	SIMPLE FORM		CLENSHAW-NEGUS	
		M_n	S_n	M_n	S_n
$\int_0^{\pi/2} \cos x \, dx$ $= 1.0$	18	-1.0006351	1.00000032	-1.00000022	0.99999992
	19	-1.0005700	1.00000026	-1.00000017	0.99999994
	20	-1.0005144	1.00000021	-1.00000014	0.99999995
	21	-1.0004665	1.00000017	-1.00000012	0.99999996
	22	-1.0004250	1.00000014	-1.00000010	0.99999996
$\int_0^{\pi/2} \sin x \, dx$ $= 1.0$	18	-0.0006351	1.00000032	0.00001847	0.99999991
	19	+0.0005700	0.99999961	0.00001570	0.99999993
	20	-0.0005144	1.00000021	0.00001346	0.99999994
	21	+0.0004665	0.99999974	0.00001163	0.99999995
	22	-0.0004250	1.00000014	0.00001011	0.99999996
	23	+0.0003889	0.99999982	0.00000885	0.99999997
$\int_0^1 e^x \, dx$ $= 1.718282$	8	2.7160480	1.71828416	2.71843142	1.71828105
	9	2.7144612	1.71828540	2.71838695	1.71828136
	10	2.7168514	1.71828278	2.71835827	1.71828153
	11	2.7157231	1.71828343	2.71833923	1.71828163
	12	2.7172881	1.71828229	2.71832599	1.71828169
$\text{erf}(1)$ $= 0.8427008$	8	-0.8319192	0.8427023	-0.8305305	0.8427008
	9	-0.831596	0.8427018	-0.8304458	0.8427008
	10	-0.8313535	0.8427015	-0.8303885	0.8427008
	11	-0.8311726	0.8427013	-0.8303479	0.8427008
	12	-0.8310303	0.8427011	-0.8303202	0.8427008

TABLE 11.1

In the last example in Table 11.1 the correct value of $F'(1)$ to 7D is -0.8302150 .

When the Clenshaw-Negus form of the moment is used it can be seen that the results converge consistently for the values of both M_1 and S_1 but, when the simple form of the moment is used, two sequences of values of M_1 and S_1 seem to appear, one corresponding to the odd number of intervals and one corresponding to the even number of intervals, and that each sequence converges separately with the sequence corresponding to the even intervals giving the better results.

To show the advantage that can be found from using unequal intervals consider the case where we compute the integral

$$\int_0^{2.0} e^{-x} dx$$

using 16 intervals - first when they are all equal to 0.175 and second when they are defined according to the sequence

$$h_i = 0.1 + (i-1)0.01 \quad i = 1, \dots, 16$$

The results are shown in Table 11.2 :

	SIMPLE FORM		CLENSHAW NEGUS	
	M_n	S_n	M_n	S_n
EQUAL	-0.06319964	0.93919481	-0.06080124	0.93918869
UNEQUAL	-0.06123427	0.93919447	-0.06078267	0.93918890
TRUE	-0.06081006	0.93918994	-0.06081006	0.93918994

TABLE 11.2

11.5 DIRECT QUINTIC SPLINE APPROXIMATION TO AN INTEGRAL FUNCTION .

The obvious next step is to carry the idea of the Direct Spline approximations even further and to investigate the possibility of deriving the recursive scheme which will give the Direct *Quintic* Spline approximation to the integral of a function $f(x)$.

Proceeding as before, let us define $S^{(4)}(x)$ as the linear function

$$S^{(4)}(x) = \frac{P_1}{h_1}(x - x_{i-1}) + \frac{P_{i-1}}{h_1}(x_i - x) \quad \dots\dots (11.9)$$

where $P_1 = S^{(4)}(x_1)$.

If we are given the initial values

$$S^{(4)}(a) = P_0 = f^{(4)}(a)$$

$$S^{(3)}(a) = N_0 = f'''(a)$$

$$S''(a) = M_0 = f''(a)$$

$$S'(a) = f_0 = f(a)$$

$$S(a) = 0$$

where $N_1 = S^{(3)}(x_1)$ and $M_1 = S''(x_1)$, then, after integrating successively and determining the constants of integration, we will find the recursive scheme

$$(i) \quad P_1 = \frac{24}{h_1^3}(f_1 - f_{i-1}) - \frac{24}{h_1^2} M_{i-1} - \frac{12}{h_1} N_{i-1} - 3P_{i-1},$$

$$(ii) \quad N_1 = N_{i-1} + \frac{h_1}{2}(P_1 + P_{i-1})$$

$$(iii) \quad M_1 = M_{i-1} + \frac{h_1^2}{6}(P_1 + P_{i-1}) + h_1 N_{i-1}$$

$$(iv) \quad S_1 = S_{i-1} + \frac{h_1}{2}(f_1 + f_{i-1}) - \frac{h_1^2}{4}(M_1 - M_{i-1}) + \frac{h_1^3}{12}(N_1 + N_{i-1}) - \frac{h_1^4}{80}(P_1 - P_{i-1})$$

$$\text{for } i = 1, \dots, n \quad \dots\dots (11.10)$$

This recursive scheme is even more unstable than the scheme (11.3) and gives virtually no worthwhile results at all. The obvious cause is the calculation of the highest derivative - P_1 - and, again, we can try to alleviate this by modifying the scheme and find an alternative method of computing the derivatives. The first attempt is to replace (i), (ii), and (iii) in scheme (11.10) by

$$(i) \quad M_1 = \frac{2}{h_1}(f_1 - f_{i-1}) - M_{i-1},$$

$$(ii) \quad N_1 = \frac{2}{h_1}(M_1 - M_{i-1}) - N_{i-1},$$

$$(iii) \quad P_i = \frac{2}{h_i}(N_i - N_{i-1}) - P_{i-1},$$

and, although this is effective in reducing the instability and results in a convergent computation as h is made smaller, it is still rather crude. We can improve this, however, by using these alternative approximations to the derivatives only as first approximations then making successive approximations finding

$$(i)' \quad M_i' = N_{i-1} + \frac{h_i^2}{6}(P_i + 2P_{i-1}) + h_i N_{i-1},$$

$$(ii)' \quad N_i' = \frac{2}{h_i}(M_i' - M_{i-1}') - N_{i-1},$$

$$(iii)' \quad P_i' = \frac{2}{h_i}(N_i' - N_{i-1}') - P_{i-1},$$

Applying the above to compute the integral

$$\int_0^{4\pi} \sin x \, dx$$

we find that, with interval spacing $h = \pi/20$, $S_5 = 0.2928165725$ which gives an error of 0.0000766463 and when we reduce h to $h = \pi/40$ we find $S_{10} = 0.2928757802$ which gives an error of 0.0000174386 - results which are still not as good as those which can be achieved using the Direct Cubic Spline Approximation !

11.6 GENERALISATION TO HIGHER ORDER DIRECT SPLINE APPROXIMATIONS.

The pattern of the recursive formulae which give us the algorithms for applying the Direct Cubic, Quartic and Quintic Spline approximations becomes familiar. The formula which gives S_1 in the Direct Cubic Spline scheme, for instance, shows itself as the formula for $S''(x_1)$ in the Direct Quintic Approximation scheme, and it can therefore be predicted that the the formula which will give $S''(x_1)$ in the Direct Sextic Approximation scheme will be of the same form as that which gives $S(x_1)$ in the Direct Quartic Approximation scheme. Thus we can see then that a recursive scheme can be developed quite easily for higher order Direct Spline Approximations

although their efficiency as algorithms for computing the integral is very much in doubt. The problem in the cases of the Direct Quartic and Direct Quintic Approximation schemes, (11.3) and (11.10), respectively, is the computation of the highest order derivatives - M_1 and P_1 . . The formula which gives the highest order derivative is, in fact, a transposition of the formula which gives the known first derivative of the spline, $S'(x_1)$ [= $f(x_1)$], in terms of the higher derivatives. These formulae are obviously very unstable, particularly for the quintic, while the other formulae in the schemes are, really, very stable.

Let ${}_nS(x)$ be the n th degree Direct Spline Approximation to the function $F(x)$ over the interval $[a, b]$, where $F(x)$ is the integral of the function $f(x) \in C^{n-1}[a, b]$ over the subinterval $[a, x]$ contained in $[a, b]$.

The formula giving ${}_nS(x)$ in the subinterval $[x_{i-1}, x_i]$ can be predicted to be

$$\begin{aligned}
 {}_nS(x) = & {}_nS(x_{i-1}) + {}_nS_1^{(n-1)} \frac{(x-x_{i-1})^n}{n!h_1} + (-1)^{n-1} {}_nS_1^{(n-1)} \frac{(x_i-x)^n}{n!h_1} \\
 & + {}_nS_1^{(n-1)} \sum_{r=2}^n (-1)^r h_1^{r-1} \frac{(x-x_{i-1})^{n-r}}{r!(n-r)!} \\
 & + \sum_{r=1}^{n-2} \frac{(x-x_{i-1})^r}{r!} {}_nS_1^{(r)} \dots\dots (11.11)
 \end{aligned}$$

where ${}_nS_1^{(r)}$ is the r th derivative of ${}_nS(x)$ at $x=x_1$.

We can prove this result (11.11) by Induction. Let us assume that the result is true.

The first derivative of the $(n+1)$ th Direct Spline Approximation - ${}_{n+1}S'(x)$ - is of the same form as (11.11) with ${}_{n+1}S'(x_{i-1}) = f_{i-1}$ and we therefore find :

$$\begin{aligned}
{}_{n+1}S'(x) &= f_{i-1} + {}_{n+1}S_i^{(n)} \frac{(x-x_{i-1})^n}{n! h_i} + (-1)^n {}_{n+1}S_{i-1}^{(n)} \frac{(x_i-x)^n}{n! h_i} \\
&+ {}_{n+1}S_i^{(n)} \sum_{r=2}^n (-1)^r h_i^{r-1} \frac{(x-x_{i-1})^{n-r}}{r! (n-r)!} \\
&+ \sum_{r=1}^{n-2} \frac{(x-x_{i-1})^r}{r!} {}_{n+1}S_{i-1}^{(r+1)}
\end{aligned}$$

If we then integrate ${}_{n+1}S'(x)$ with respect to x we will find

$$\begin{aligned}
{}_{n+1}S(x) &= x f_{i-1} + {}_{n+1}S_i^{(n+1)} \frac{(x-x_{i-1})^{n+1}}{(n+1)n! h_i} + (-1)^n {}_{n+1}S_{i-1}^{(n+1)} \frac{(x_i-x)^{n+1}}{(n+1)n! h_i} \\
&+ {}_{n+1}S_i^{(n)} \sum_{r=2}^{n-1} (-1)^r h_i^{r-1} \frac{(x-x_{i-1})^{n-r+1}}{(n-r+1)r! (n-r)!} + (-1)^n {}_{n+1}S_{i-1}^{(n)} \frac{h_i^{n-1} x}{n!} \\
&+ \sum_{r=2}^{n-1} \frac{(x-x_{i-1})^{r+1}}{(r+1)r!} {}_{n+1}S_{i-1}^{(r+1)} + C_{n1}
\end{aligned}$$

where C_{n1} is the constant of integration .

At $x=x_1$

$$\begin{aligned}
{}_{n+1}S(x_{i-1}) &= {}_{n+1}S_{i-1} = x_{i-1} f_{i-1} + {}_{n+1}S_{i-1}^{(n+1)} \frac{(-1)^n h_i^{n+1}}{(n+1)! h_i} \\
&+ {}_{n+1}S_{i-1}^{(n)} \frac{(-1)^n h_i^{n-1}}{n!} x_{i-1} + C_{n1}
\end{aligned}$$

and, transposing this equation to give C_{n1} we find that

$$\begin{aligned}
C_{n1} &= {}_{n+1}S_{i-1} - x_{i-1} f_{i-1} + {}_{n+1}S_{i-1}^{(n+1)} \frac{(-1)^{n+1} h_i^n}{(n+1)!} \\
&- {}_{n+1}S_{i-1}^{(n)} \frac{(-1)^n h_i^{n-1}}{n!} x_{i-1}
\end{aligned}$$

and hence

$$\begin{aligned}
{}_{n+1}S(x) &= (x-x_{i-1})f_{i-1} + {}_{n+1}S_{i-1}^{(n)} \frac{(x-x_{i-1})^{n+1}}{(n+1)!h_i} \\
&+ {}_{n+1}S_{i-1}^{(n)} (-1)^n \frac{(x-x_{i-1})^{n+1}}{(n+1)!h_i} \\
&+ {}_{n+1}S_{i-1}^{(n)} \sum_{r=2}^{n-1} (-1)^r \frac{h_i^{r-1} (x-x_{i-1})^{n-r+1}}{r!(n-r+1)!} \\
&+ {}_{n+1}S_{i-1}^{(n)} (-1)^n \frac{h_i^{n-1} (x-x_{i-1})}{n!} + {}_{n+1}S_{i-1}^{(n)} (-1)^n \frac{h_i^n}{(n+1)!} \\
&+ \sum_{r=1}^{n-2} {}_{n+1}S_{i-1}^{(r+1)} \frac{(x-x_{i-1})^{r+1}}{(r+1)!} + S_{i-1}
\end{aligned}$$

The first and seventh terms can be collected to give

$$\sum_{r=1}^{n-1} \frac{(x-x_{i-1})^r}{r!} {}_{n+1}S_{i-1}^{(r)}$$

since $f_{i-1} = {}_{n+1}S_{i-1}'$, and then the fourth, fifth and sixth terms can also be collected to give

$${}_{n+1}S_{i-1}^{(n)} \sum_{r=2}^{n+1} (-1)^r h_i^{r-1} \frac{(x-x_{i-1})^{n-r}}{r!(n-r)!}$$

so that

$$\begin{aligned}
{}_{n+1}S(x) &= {}_{n+1}S(x_{i-1}) + {}_{n+1}S_{i-1}^{(n)} \frac{(x-x_{i-1})^{n+1}}{(n+1)!h_i} \\
&+ {}_{n+1}S_{i-1}^{(n)} (-1)^n \frac{(x-x_{i-1})^{n+1}}{(n+1)!h_i} \\
&+ {}_{n+1}S_{i-1}^{(n)} \sum_{r=2}^{n+1} (-1)^r h_i^{r-1} \frac{(x-x_{i-1})^{n-r}}{r!(n-r)!} \\
&+ \sum_{r=1}^{n-1} \frac{(x-x_{i-1})^r}{r!} {}_{n+1}S_{i-1}^{(r)} \dots\dots (11.12)
\end{aligned}$$

Which is the formula (11.10) in which $n+1$ replaces n . The formula is true for $n=3$ and $n=4$ hence, by induction, it is true for all n .

11.7 COMPUTATIONAL SCHEME FOR NTH ORDER DIRECT SPLINE APPROXIMATION

Following along with our theoretical consideration of the nth order spline approximation to an integral function we can develop the computational scheme which is equivalent to the scheme (1.8) for the Direct Cubic Spline even though, when $n > 4$, their efficiency as an algorithm for computing the integral is in doubt. The scheme is :

$$\text{Given } {}_nS_i^{(0)} = 0 : {}_nS_i^{(k)} = f^{(k-1)} \quad k=1,2,\dots,n-2$$

then

$$\begin{aligned}
 {}_nS_i^{(k)} = & {}_nS_{i-1}^{(k)} + \frac{h_i^{(n-k-1)}}{(n-k)!} [{}_nS_i^{(n-1)} + S_{i-1}^{(n-1)}] \\
 & + \sum_{r=1}^{n-k-2} \frac{h_i^r}{r!} {}_nS_{i-1}^{(k+r)} \\
 & + \sum_{r=2}^{n-k} (-1)^r \frac{(n-k)!}{r!(n-k-r)!}
 \end{aligned}$$

$$\text{for } k = n-2, \dots, 1, 0 \quad \dots \dots (11.13)$$

11.8 CONCLUSIONS ON HIGHER ORDER DIRECT SPLINE APPROXIMATIONS TO INTEGRALS.

None of the schemes of computation for the higher order Direct Spline Approximations investigated in this chapter, and that we have tried, have proved to of any great value when applied to the particular problems that we have chosen. The problem seems to be focussed on the computation of the derivatives, which is not unexpected since the computation of derivatives often does result in instability.

- 12 -

CONCLUSION

The aim of part one of this thesis has been to present a mathematical analysis of the methods of computation in navigation and to effect some new numerical solutions. In the past the methods of computation in navigation have generally been simplified to suit manual computations. However, with the advent of computer software and the introduction of the electronic chart these simplifications can no longer be considered adequate or necessary. As we stated in the introduction, the spherical model for the shape of the Earth is, for most practical purposes in navigation, quite adequate, if used consistently, but, with the computer to do the work for us, the spheroidal model is obviously more fitting and this model has become the focus of our attention. It was considered useful, however, to analyse the properties of the curves on the surface of the spherical Earth model since this simplification serves to illustrate the same pattern of mathematical analysis which we can use to develop the equations for navigating along curves on the surface of the spheroid. All the chapters from chapter 3 onwards contain work which has been published .

The methods of navigation along the arc of a loxodrome are now fully developed for the spheroidal model and most applications embrace this. The single most important aspect of navigating along the arc of the loxodrome is the computation of meridian distance. Carlton Wippen²⁴ defines this distance in terms of elliptical integrals since the meridian is but the arc of an ellipse but does not do any actual computation. Bowring⁵³ gives a new method which will compute this distance to a very high degree of accuracy but we use the direct cubic spline which, for our purposes gives us the accuracy we require (two decimal places of a geographical mile) at a step length of 5° in latitude along the meridian and because, in this case, the algorithm is easy to apply.

The great circle method on the spherical model still seems to be the popular method, in practice, of computing the shortest path geodesic arc. This is because the navigator requires more than just the

straight forward computation of distance (the methods for which are well developed for the terrestrial spheroid) so that the great circle method has been the only method available which will give the intermediate points along the path of the geodesic. We have therefore presented a solution to the problem of computing the geodesic arc on the surface of the spheroid in a different way by starting from the definition of the geodesic by means of Clairaut's equation. We have solved the equation to give us a relationship between the geocentric latitude and the longitude of a point along the path of the geodesic then, using the step by step method of the Direct Cubic Spline, given the step values in the longitude, we have found the corresponding values of the latitude and the distance along the geodesic arc between them. We have applied this same method to the shortest path problems on the surfaces of both the sphere and the spheroid and, in the case of the sphere, the algorithm is quick, efficient and in every way comparable with the methods of spherical trigonometry. In its application to the surface of the spheroid the algorithm for doing this has also proved to be simple and efficient even though it involves a fair number of iterations. The results can be condensed neatly as shown in Table 4.1, chapter 4, for the sphere and Table 5.1, Chapter 5, for the spheroid. The positions from these tables can be plotted on a chart to give visual representations of the paths. If the chart is a standard paper chart then the way points can be plotted and the usual convention would then be to join the way points with short rhumb lines. We can expect, however, that the electronic chart will soon be in general use and the application can be refined considerably. On the electronic chart we will be able to represent the portions of the geodesic between the way points by curved arcs. Indeed, the relevant way to do this will be to fit a conventional cubic spline approximation between the end points of the geodesic arc with the way points as the "knots" of the spline.

In chapters 5,6 and 7, we believe that we have presented a full analysis of the solution by a direct method of the problem of computing the path of geodesic arcs on the surface of a spheroid. We

have defined our distances along the arcs of the geodesics by line integrals computed using the method of the direct cubic spline approximation. All the results have, we believe, been computed correct to the first decimal place of a geographical mile. For the purposes of navigation, either in the air or at sea, this is adequate. In the future this computed solution can be linked with an electronic position fixing system so that, from the input of the observed position, the path to the destination, initial course and distance, with its intermediate way points can be quickly updated at any time. In coastal navigation, where course lines are all short arcs of a loxodrome, this is done already with the Decca Navigator system. In a system designed for navigating world wide a choice of routes - loxodromic, geodesic or composite - can be offered from any observed position to the destination.

Although the methods of analysis and computation of the properties of geodesic and loxodromic arcs have been developed with the problem of navigating on the surface of the Earth uppermost in the mind the methods are not strictly "Earthbound" at all. The analysis is general to any ellipsoidal surface. Should it soon be possible in a space vehicle to skim the surface of an outer planet where the flattening of the surface is more pronounced, then, by our methods, for navigational purposes, we will be able to compute the path of the geodesic or loxodromic arcs on such a surface also. For the shortest path geodesic arc between two points on a surface such as Jupiter, for instance, where the eccentricity of the meridian ellipse is approximately equal to 0.3, then, the initial approximation to the path which we can use for the iterative procedures in the computation will be the "great ellipse" - the ellipse defined by the intersection of the surface with a plane through the two points and the centre of the ellipsoid.

At first sight it might seem that our achievement in chapter 8 is simply a tidy representation of the formulae that are used in astronomical navigation and, that, since astronomical navigation is, to some extent, only a sideshow these days, little has been achieved

by it. There is, however, a little more to it than that. A work of philosophy must contain some elements which are there because, if little else, they present an interesting theoretical problem but, even so, this is not entirely the case in chapter 8. The consequences of the distortion in the transferred position circle are worth noting and, in the past, might well have caused some problems for polar navigators because it is close to the pole that this distortion is most pronounced. At some stage on the polar journey it must be necessary to travel at an oblique angle to the meridians and, since the sun is the only astronomical body visible, transferred position circles must have been used to fix position. It would be interesting to analyse the results from such observations that were taken by polar explorers and to see whether or not, in general, such allowances were made.

We have not made any mention of the way in which astronomical observations are affected by the spheroidal shape of the Earth. In truth the position "circle" is the locus of intersection of the surface of a cone and the surface of the spheroid. It would appear that the cone has its apex at the centre of curvature of the spheroid at the point which is the geographical position of the observed body and its axis is along the corresponding radius of curvature. The position "circle" is not, therefore, in general, a circle. The observer's position when two simultaneous observations are taken is then at a point of intersection of three surfaces - two cones which are defined by the astronomical observations and the surface of the spheroid. This problem is now being studied with the purpose of finding out whether there is a solution which will give worthwhile results to the navigator.

In chapter 9 we concern ourselves with computing position from the observation of a single astronomical body over a short period of time. This method would have an application at sea in good observing conditions and would give positions which should be better than the running fix but it is implicit that the computations are only possible using a powerful computing device. It is, however, in the

possibility of mechanisation that this method might find its place. Judging by the demonstrations of the accuracy of gyroscopic stabilisation that have been made recently in weapons technology it might be more than just feasible to design an instrument which can lock on to a source such as the sun or a star and measure the instantaneous altitude and its rate of change. From this the observer's position can be computed, automatically, by the method given in chapter 9 and displayed for the observer. This is only conjecture, of course, but one is lead to believe that this should be possible. Such a system, if it is feasible, would reduce the reliance on the orbits of the man made sattelites.

Part two of the thesis has been devoted to the Direct Spline Approximations to integrals. Chapter 10 is concerned exclusively with the Direct Cubic Spline Approximation and it is this approximation to integrals that is applied to such good effect throughout part one. The method is a step by step method and is particularly effective in application to line integrals because of the in built property of generating points along the path. One disadvantage of the method in this respect, however, is that it does not converge very rapidly as the interval length is decreased.

We have shown that the direct cubic spline is a generalisation of Simpson's Rule and a truncated approximation of Euler's Integral Expansion method. It has the distinct advantage too of being within the realm of manual computation. The requirement that the derivative at one boundary should be known is not, in fact, too much of a disadvantage to the method. In our applications we have found many cases where the derivative, or a simple approximation to it, have been easily found. While, in theory, an approximation to the derivative will reduce the order of the error bound it is not a serious practical consideration. If we are to strictly maintain the step by step property of the method it does sometimes mean that in the absence of true value of the derivative we might have to make a linear approximation, but, in general, we are not so restricted in practice. In the case of the geodesic arcs, for instance, we know

our destination and can therefore choose the intermediate points in advance. In the computer program to compute the length of the geodesic arc passing through these intermediate points we begin by fitting a Lagrange cubic polynomial to the first four intermediate points and we have differentiated this Lagrange polynomial to find a quadratic approximation to the derivative.

In Chapter 11 we hope to carry the idea of the Direct Cubic Spline into the development of higher order direct spline approximations to integrals. We have generated the formulae which will, by the same analysis, give us expressions for these higher order direct splines but, except in the case of the Direct Quartic Spline, we have not demonstrated that there are any useful algorithms. For the most part, then, at the moment, this chapter is just a theoretical demonstration which is left as an open question. For the direct quartic spline the "moments" of the spline are provided for us by the limiting form of the X-spline developed by Clenshaw and Negus⁵². The integral formula in the direct quartic spline is, however, exactly the same as that in the direct cubic spline and this is general for all the higher even order direct splines - the integral formula is the same as that of the next lower odd order direct spline.

There are other particular areas of study where the direct spline approximations have a relevant application. We are also using these methods in the theory of ship stability where the evaluation of integrals with known boundary derivatives are commonly required. As a professional navigator it has been particularly rewarding, however, to find that the direct cubic spline has been so particularly useful in the science of navigation.

- REFERENCES -

REFERENCES.

1. W.M.SMART: "On a Problem in Navigation" - *Monthly Notices of the Royal Astronomical Society* 106,2,pp124-127. (1946)
2. R.WILLIAMS: "A Table of Latitude Parts" - *Journal of Navigation* 34,247 (1981)
3. KITT C. CARLTON WIPPERN: "On Loxodromic Navigation" - *Journal of Navigation* 45,292 (1992)
4. P.D.THOMAS: "Mathematical Models For Navigation Sytems" - U.S. Naval Oceanographic Office, Washington D.C. (October 1965).
5. E.FICHOT: "Sur Les Systemes Geodesiques Equaliteres a la Surface du Spheroides Terrestre" - *Annales Hydrographique* pp 99-192 (1926)
6. W.D.LAMBERT: "The Distance Between Two Widely Separated Points on the Surface of the Earth" - *Journal of the Washington Academy of Science* 32,125 (1942)
7. B.R.BOWRING: "The Geodesic Inverse Problem" - *Bulletin Geodesique* 57,109 (1983)
8. R.WILLIAMS: "Computation of an Astronomical Running Fix" - *Journal of Navigation* 43,3 (1990)
9. MATTI RANTA: "Position Fixing in a Fast Moving Ship by Culmination of a Celestial Body" - *Journal of Navigation* - 44,2 (1990)
10. R.WILLIAMS: "Position by Observation of a Single Body" - *Journal of Navigation* - 47,1 (1994)
11. G.E.FORSYTHE: " Generation and Use of Orthogonal Polynomials for Data Fitting with a Digital Computer" - *Journal of Industrial Applied Mathematics* 5,2 (1957)
12. NATHANIEL BOWDITCH - "The American Practical Navigator." - Defense Mapping Agency Hydrographic Center (US) - Pub No.9 (1977)
13. CHARLES H.COTTER - "The Astronomical and Mathematical Foundations of Geography" - Hollis and Carter.
14. G.BOMFORD - "Geodesy" - OUP
15. MARTIN.M.LIFSCHUTZ - "Differential Yeometry" - Schaum MacGraw Hill (1969)

16. DIRK J. STRUICK - "Lectures in Classical Differential Geometry" -
Dover (1988)
17. NORIES NAUTICAL TABLES - Imray, Norie & Wilson (1948)
18. NORIES NAUTICAL TABLES - Imray, Norie & Wilson (1954)
19. INMANS NAUTICAL TABLES - J.D.Potter (London 1952)
20. R.WILLIAMS: "A Table of Latitude Parts" - *Journal of Navigation*
34,247 (1981)
21. D.H.SADLER: "Rhumb Line Navigation" - *Journal of Navigation*
9,371 (1956)
22. J.E.D.WILLIAMS: "Loxodromic Distances On The Terestial Spheroid"
- *Journal of Navigation* 3,133 (1950)
23. T.HAIRAWA: "On Modification Of Sailing Calculations" - *Journal*
of Navigation 40,138 (1987)
24. KITT C.CARLTON WIPPERN: "On Loxodromic Navigation" - *Journal of*
Navigation 45,292 (1992)
25. LYUSTERNIK L.A. "Shortest Paths" - Pergammon Press (1964)
26. BELL R.J.T. "Coordinate Geometry of Three Dimensions"
MacMillan ()
27. WILLIAMS.R. "A Step by Step Method of Computing Points Along
the Arc of a Great Circle" - *Journal of Navigation* (Vol 46, No.2
May 1993)
28. ANDOYER H. "Formule Donnant la Longueur de la Geodesique
Joignant Deux Points de l'Ellipsoide donnees Par leur Coordonnes
Geographique." *Bulletin Geodesique* , 34,77 (1932)
29. LAMBERT W.D. "The Distance Between two Widely Separated
Points on the Surface of the Earth" *Journal of the Washington*
Academy of Science - 32,125 (1942)
30. WILLIAMS R & PHYTHIAN J.E. "Navigating Along Geodesic Paths on
the Surface of a Spheroid" - *Journal of Navigation* - 42, 129 (1989)
31. BOURBON R. "Geodesic Line On The Surface Of A Spheroid" -
Journal of Navigation 43.1 (1990)
32. TODHUNTER I "A History of The Mathematical Theories of
Attraction and the Figure of the Earth" - Dover (1962)
33. HELMBERT " Die Mathematischen und Physikalischen
Theorieen der hoheren Geodesie" Chapters 5 & 7 - Leipzig (1880)

34. FICHOT E. "Sur les systemes geodesiques equilateres a la surface du spherode terrestre." - *Annales Hydrographiques* (sr 3) 4: 99,1921
35. GLASSTONE S. "Sourcebook On The Space Sciences" - Van Noorstrand (in cooperation with NASA) (1965)
36. WILLIAMS R & PHYTHIAN J.E. "The Shortest Distance Between Two Nearly Antipodean Points on the Surface of a Spheroid" - *Journal of Navigation* 45.1, 114 (1992).
37. MARGARET GOW - "A Coures In Pure Mathematics" - English Universities Press (1960)
38. ARTURO & RAFAELE CHIESA - "A Mathematical Method of Obtaining an Astronomical Vessel Position" - *Journal of Navigation* 43,1 (1990)
39. BERNARD SPENCER - "Astronomical Fixes Without an Assumed Position" - *Journal of Navigation* 43,3 (1990)
40. ROY WILLIAMS - "Computation of an Astronomical Running Fix" - *Journal of Navigation* 43,3 (1990)
41. JAMES .N. WILSON - "Position From Observation Of A Single Body" - "Navigation" - *Journal of the American Institute of Navigation* - Vol 32 1985.
42. MATTI RANTA - "Position Fixing In A Fast Moving Ship By Culmination Of A Celestial Body" - *Journal of Navigation* - The Royal Institute of Navigation - Vol 44,2 1991.
43. GEORGE E. FORSYTHE - "Generation and use of Orthogonal Polynomials for Data-fitting with a Digital computer." - *Journal of Industrial Applied Mathematics* Vol 5 No.2, June 1957
44. PERCY. L. H. DAVIS - "Altitude - Azimuth Tables" - HNSO (1918)
45. WILLIAMS R. "Double Integration Using Cubic Splines" - Bachelor of Philosophy Thesis - 1982 - Open University Library - Chapter 3.
46. DAVIS P. J. & RABINOWITZ P. "Methods of Numerical Integration" - Academic Press (1975).
47. AHLBERG J. H., NILSON E. N. & WALSH J. L. " The Theory of Splines and their Applications" - Academic Press (1967).
48. PRENTER P. M. "Splines and Variational Methods" - John Wiley & Sons (1975).

49. PHYTHIAN J.E. & WILLIAMS R. "Cubic Spline Aproximation to an Integral Function" - *Bulletin of the Institute of Mathematics & its Applications* - Volume 21 July/August 1985 - pp 130 & 131.
50. SCHEID F. "Numerical Analysis" - Schaum MacGraw Hill - 1968.
51. PHYTHIAN J.E. & WILLIAMS R. "Some Relationships between the Direct Cubic Spline Approximation and other Integration Formulae" - *Bulletin of the Institute of Mathematics and its Applications Vol 24* 9&10 (1988)
52. CLENSHAW C.W. & NEGUS B. "The Cubic X-Spline and its Application to Interpolation" *Journal of the Institute of Mathematics and its Applications* 22 (1978) pp109-119.
53. BOWRING B.R. "New Equations For Meridian Distance" - *Bulletin Geodesique* (1984)

APPENDIX 1
TRANSFORMATION OF EQUATION
(5.4) TO EQUATION (5.5)

Using the substitutions $x = a\theta$ $y = a \tan \phi$ we find $dx = a d\theta$ and $dy = a \sec^2 \phi d\phi$

We also have $\tan \phi = \frac{y}{a}$ $\sin \phi = \frac{y}{\sqrt{a^2 + y^2}}$ $\cos \phi = \frac{a}{\sqrt{a^2 + y^2}}$

and $\sec^2 \phi = \frac{a^2 + y^2}{a^2}$ so that $dy = \frac{a^2 + y^2}{a} d\phi$ and $d\phi = \frac{\cos^2 \phi}{a} dy$

The equation $a_p^2 \cos^2 \phi \left(\frac{d\theta}{ds}\right) = a_p \cos \phi$ may be written $a_p^2 \cos^2 \phi d\theta = \lambda ds$ where we have written $\lambda = a_p \cos \phi$.

Squaring this equation and substituting $ds^2 = a_p^2 \cos^2 \phi d\theta^2 + a_p^2 \sec^2(\psi - \phi) d\phi^2$ we find

$$a_p^4 \cos^4 \phi d\theta^2 = \lambda^2 [a_p^2 \cos^2 \phi d\theta^2 + a_p^2 \sec^2(\psi - \phi) d\phi^2] = \lambda^2 \cos^2 \phi [a_p^2 d\theta^2 + a_p^2 \frac{\cos^2 \phi}{a^2} dy^2]$$

Dividing through by $a_p^2 \cos^2 \phi$ and substituting for $\cos^2 \phi$ from above we find

$$\frac{a_p^2 a^2}{(a^2 + y^2)} d\theta^2 = \lambda^2 \left[d\theta^2 + \sec^2(\psi - \phi) \frac{dy^2}{(a^2 + y^2)} \right] \text{ which, after rearranging, gives}$$

$$\cos^2(\psi - \phi) [a_p^2 a^2 - \lambda^2 (a^2 + y^2)] d\theta^2 = \lambda^2 dy^2 \dots\dots\dots (A1.1)$$

From the equation $\tan \phi = (1 - e^2) \tan \psi$ we find $\tan \psi = \frac{y}{a(1 - e^2)}$ $\cos \psi = \frac{a(1 - e^2)}{\sqrt{[a^2(1 - e^2)^2 + y^2]}}$

and $\sin \psi = \frac{y}{\sqrt{[a^2(1 - e^2)^2 + y^2]}}$. Using these results we find

$$\cos^2(\psi - \phi) = \frac{[a^2(1 - e^2) + y^2]}{[a^2(1 - e^2)^2 + y^2](a^2 + y^2)} \dots\dots\dots (A1.2)$$

Now $\lambda^2 = a_p^2 \cos^2 \psi = \frac{a^2(1 - e^2)}{(1 - e^2 \cos^2 \psi)} \cos^2 \psi = \frac{a^4(1 - e^2)}{[a^2(1 - e^2) + y^2]} \dots\dots\dots (A1.3)$

Also $a_p^2 = \frac{a^2(1 - e^2)}{(1 - e^2 \cos^2 \psi)} = \frac{a^2(1 - e^2)(a^2 + y^2)}{[a^2(1 - e^2) + y^2]} \dots\dots\dots (A1.4)$

Substituting (A1.2), (A1.3) and (A1.4) into (A1.1) and rearranging results in

$$dy^2 = \frac{[a^2(1 - e^2) + y^2]}{[a^2(1 - e^2)^2 + y^2]} (y_c^2 - y^2) d\theta^2$$

as required,

APPENDIX 2
A TABLE OF LATITUDE PARTS

	0'	10'	20'	30'	40'	50'
0	0.00	9.93	19.86	29.80	39.87	49.66
1	59.59	69.53	79.46	89.39	99.32	109.26
2	119.19	129.12	139.05	148.99	158.92	168.85
3	178.78	188.72	198.65	208.58	218.58	228.45
4	238.38	248.31	258.25	268.18	278.11	288.04
5	297.98	307.91	317.84	327.78	337.71	347.64
6	357.58	367.51	377.44	387.38	397.31	407.24
7	417.18	427.11	437.05	446.98	456.91	466.85
8	476.78	486.72	496.65	506.59	516.52	526.45
9	536.39	546.32	556.26	566.19	576.13	586.06
10	596.00	605.94	615.87	625.81	635.74	645.68
11	655.61	665.55	675.49	685.42	695.36	705.29
12	715.23	725.17	735.11	745.04	754.98	764.92
13	774.85	784.79	794.73	804.67	814.60	824.54
14	834.48	844.42	854.36	864.30	874.23	884.17
15	894.11	904.05	913.99	923.93	933.87	943.81
16	953.75	963.69	973.63	983.57	993.51	1003.45
17	1013.39	1023.33	1033.27	1043.22	1053.16	1063.10
18	1073.04	1082.98	1092.93	1102.87	1112.81	1122.75
19	1132.70	1142.64	1152.58	1162.53	1172.47	1182.41
20	1192.36	1202.30	1212.25	1222.19	1232.14	1242.08
21	1252.03	1261.97	1271.92	1281.86	1291.81	1301.75
22	1311.70	1321.65	1331.59	1341.54	1351.49	1361.44
23	1371.38	1381.33	1391.28	1401.23	1411.18	1421.12
24	1431.07	1441.02	1450.97	1460.92	1470.87	1480.82
25	1490.77	1500.72	1510.67	1520.62	1530.57	1540.53
26	1550.48	1560.43	1570.38	1580.33	1590.29	1600.24
27	1610.19	1620.15	1630.10	1640.05	1650.01	1659.96
28	1669.92	1679.87	1689.82	1699.78	1709.74	1719.69
29	1729.65	1739.60	1749.56	1759.52	1769.47	1779.43
30	1789.39	1799.35	1809.30	1819.26	1829.22	1839.18
31	1849.14	1859.10	1869.06	1879.02	1888.98	1898.94
32	1908.90	1918.86	1928.82	1938.78	1948.74	1958.70
33	1968.67	1978.63	1988.59	1998.55	2008.52	2018.48
34	2028.45	2038.41	2048.37	2058.34	2068.30	2078.27
35	2088.23	2098.20	2108.17	2118.13	2128.10	2138.07
36	2148.03	2158.00	2167.97	2177.94	2187.90	2197.87
37	2207.84	2217.81	2227.78	2237.75	2247.72	2257.69
38	2267.66	2277.63	2287.60	2297.57	2307.54	2317.52
39	2327.49	2337.46	2347.43	2357.41	2367.38	2377.35
40	2387.33	2397.30	2407.28	2417.25	2427.23	2437.20
41	2447.18	2457.15	2467.13	2477.11	2487.08	2497.06
42	2507.04	2517.02	2526.99	2536.97	2546.95	2556.93
43	2566.91	2576.89	2586.87	2596.85	2606.83	2616.81
44	2626.79	2636.77	2646.75	2656.74	2666.72	2676.70
45	2686.68	2696.67	2706.65	2716.63	2726.62	2736.60
46	2746.59	2756.57	2766.56	2776.54	2786.53	2796.51
47	2806.50	2816.49	2826.47	2836.46	2846.45	2856.44
48	2866.42	2876.41	2886.40	2896.39	2906.38	2916.37
49	2926.36	2936.35	2946.34	2956.33	2966.32	2976.31
50	2986.31	2996.30	3006.29	3016.28	3026.27	3036.27
51	3046.26	3056.25	3066.25	3076.24	3086.24	3096.23
52	3106.23	3116.22	3126.22	3136.21	3146.21	3156.21
53	3166.20	3176.20	3186.20	3196.20	3206.19	3216.19
54	3226.19	3236.19	3246.19	3256.19	3266.19	3276.19
55	3286.19	3296.19	3306.19	3316.19	3326.19	3336.19
56	3346.20	3356.20	3366.20	3376.20	3386.21	3396.21
57	3406.21	3416.22	3426.22	3436.22	3446.23	3456.23
58	3466.24	3476.24	3486.25	3496.26	3506.26	3516.27
59	3526.28	3536.28	3546.29	3556.30	3566.31	3576.31
60	3586.32	3596.33	3606.34	3616.35	3626.36	3636.37
61	3646.38	3656.39	3666.40	3676.41	3686.42	3696.43
62	3706.44	3716.45	3726.46	3736.48	3746.49	3756.50
63	3766.51	3776.53	3786.54	3796.55	3806.57	3816.58

APPENDIX 3
SOME RESULTS
IN
SPHERICAL TRIGONOMETRY

A spherical triangle on the surface of a sphere is defined by the area enclosed by the intersection of three great circles. Figure A3.1 shows the spherical triangle with vertices A, B and C which are the points of intersection of the three great circles the arcs of which are $AB(=c)$, $BC(=a)$ and $CA(=b)$. a , b and c are expressed as angles where a , for instance, is the angle subtended at the centre of the sphere by the arc BC.

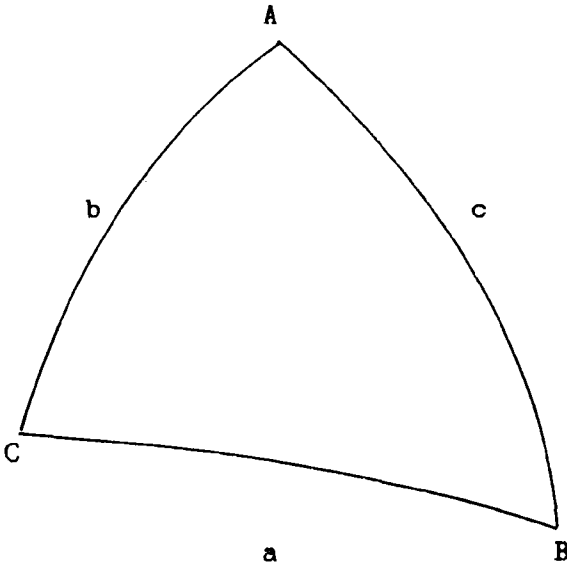


FIGURE A3.1

The *SPHERICAL COSINE FORMULAE* read

$$\cos a = \cos b \cos c + \sin b \sin c \cos A$$

or
$$\cos b = \cos a \cos c + \sin a \sin c \cos B$$

or
$$\cos c = \cos a \cos b + \sin a \sin b \cos C$$

and, interchanging angles for sides we find also

$$\cos A = \sin B \sin C \cos a - \cos B \cos C$$

or
$$\cos B = \sin A \sin C \cos b - \cos A \cos C$$

or
$$\cos C = \sin A \sin B \cos c - \cos A \cos B$$

The SPHERICAL SINE FORMULAE read

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

These results are proved in the book by Margaret Gow³⁷.