8 January 2019

Can Maths Tell Us Where We Are?

Professor Chris Budd OBE

Ballad of Gresham College v. 26

The College will the whole world measure
Which most impossible conclude,
And Navigation make a pleasure
By finding out the Longitude.
Every Tarpaulin shall then with ease,
Sayle any ship to the Antipodes

Anon (circa 1660)

1. Introduction

Where am I? In my previous life as a mountaineering leader for West of Ireland Camps, this was a question I was frequently asked by the young people I was trying to guide through the mist. Teaching maths to undergraduates is a not dissimilar experience, as we will find out in lecture five of this series. But more seriously, accurately locating ourselves, and then working out where we should go next, has been a question of vital importance to civilisation since its very beginnings and before. Human beings have navigated by looking at the features around them, by using maps and compass, by looking at the stars, by dead reckoning, and most recently by electronic means such as Gee and GPS. Mathematics plays a role in all of these, and in the case of GPS some of the advanced mathematics behind Einstein's General Theory of Relativity. By doing so, maths has changed the course of the world. It is also fair to say that the search for finding out where are has led to developments not only in mathematics itself (such as the theory of spherical trigonometry), but also in the interplay behind mathematics and technology (a favourite subject of mine), and the development of the modern computer.

In this lecture I will take you on a (quite literal in a sense) journey through the history of navigation, from the ancients to the present day. It is excellent to see that Gresham College in general, and the Gresham Professors of Geometry in particular, have played an important role in this story. I will then ask the question, given our modern reliance on GPS navigation of *what happens if it all goes wrong?*, and can we take steps to avoid this calamity?

2. How to use a map and a compass

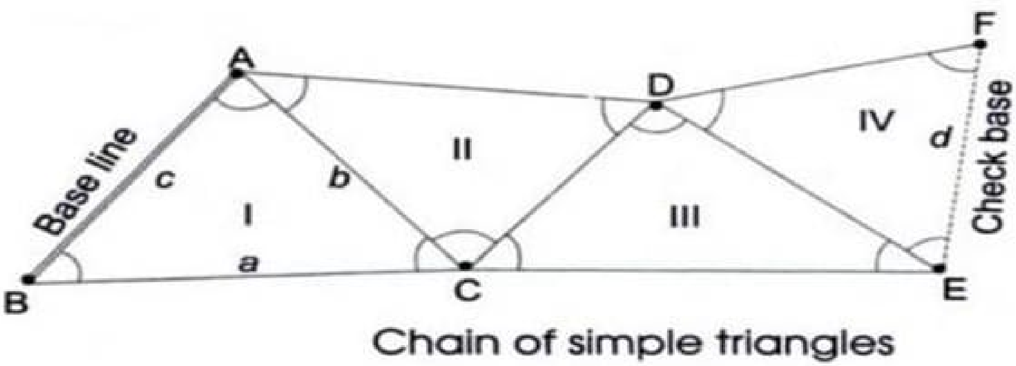
2.1 Early maps

In order to find out where we are we need to know where everything else is. To do this we need an accurate map. Maps have been existence for a long with examples of them being found on Babylonian cuneiform tablets dated to 600 BC. Two early maps are shown below. The one on the left is due to the Greek scholar Claudius Ptolemaeus, better known as Ptolemy. In 150 A.D., he created an eight-volume textbook *Geography* that included some of the first maps constructed using mathematical ideas. This map is remarkably accurate given the limitations of the time. The map on the right is the celebrated Mappa Mundi in Hereford Cathedral. This map places countries according to religious rather than mathematical principles. The British Isles can be seen on the top left, with Jerusalem in the dead centre.

2.2 Maps of the UK and beyond

In contrast to the Mappa Mundi, all modern maps are constructed according to the mathematical principles pioneered by Ptolemy. Perhaps the most famous of these (at least in the UK) are those produced by the Ordnance Survey (OS). As a mountaineering leader I would like to include in this the women and men of this wonderful organisation. Their truly remarkable work opened up the world to me and allowed me to explore the countryside in way that simply would not have been possible without the maps that they produce. The OS is also a major employer of fine mathematicians. It was founded in 1747 with the express purpose of providing high quality maps of the UK for the army following the Battle of Culloden in 1746 In 1747, Lieutenant-Colonel David Watson proposed the compilation of a map of the Highlands to help to subjugate the clans. In response, King George II charged Watson with making a military survey of the Highlands under the command of the Duke of Cumberland. The survey was produced at a scale of 1 inch to 1000 yards (1:36,000) and included "the Duke of Cumberland's Map", primarily by Watson and his assistant Roy. Roy later had an illustrious career, rising to the rank of General in the Royal Engineers, and he was largely responsible for the British share of the work in determining the relative positions of the French and British royal observatories. This work was the starting point of the Principal Triangulation of Great Britain (1783–1853) and led to the creation of the Ordnance Survey itself. In 1801 the first one-inch-to-the-mile (1:63,360 scale) map was published, detailing the county of Kent with Essex following shortly afterwards. For many years the (red) one-inch map was the standard map produced by OS. However, in 1974 It was replaced by the (purple) 50,000 map or 2cm to one km in which the grid squares are one lm wide. More detailed 25,000 (or 2 ½ inch to the mile) maps were introduced shortly afterwards, and have become the standard maps for mountaineering. I have a very complete collection!

Surveying and Triangulation



Traditional map making made extensive use of triangulation, which in turn relies heavily on the mathematics of trigonometry. Triangulation surveying is the tracing and measurement of a series or network of triangles to determine distances and relative positions of points spread over an area. It does this by starting from a very accurately measured baseline, which could be several kilometres in length. This baseline had to be very straight, and its length determined to high precision in all possible weathers. Doing this alone was a far from easy task. The triangulation was then continues by measuring the angles to carefully sited points, often on the tops of hills or on high towers. The measurements themselves were made using a theodolite, which found angles to high precision. The UK itself is covered with a network of triangulation, or trig, points, which will be familiar to any walkers. In the UK, the Ordnance Survey's first trig point was erected on 18 April 1936 near Cold Ashby in Northamptonshire.

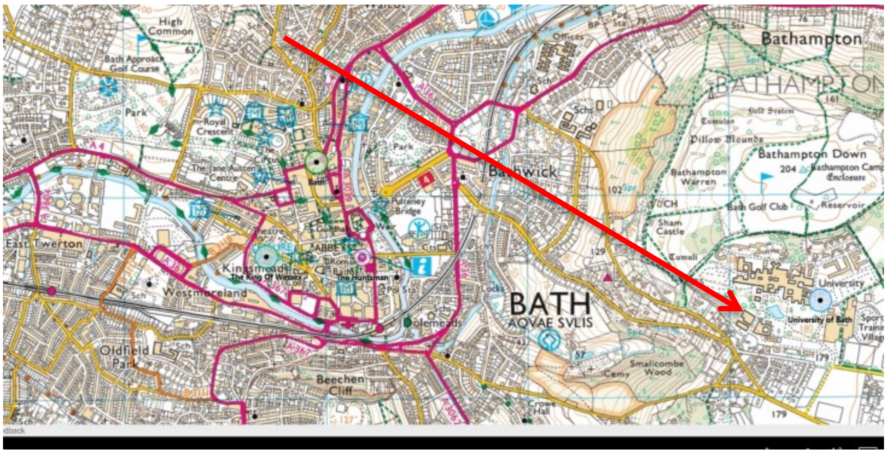
By measuring the angles between the trig points a network of triangles can be constructed linking these points together and hence produce the basis of a map. The triangulation must then be checked periodically by measuring further base lines to high accuracy. From this network of large triangles, smaller triangles can be then constructed and a map built up

A large scale triumph of the use of the triangulation technique was the Great Trigonometrical Survey, which was a project which aimed to measure the entire Indian subcontinent, with scientific precision. This project was begun in 1802, by William Lambton, under the auspices of the East India Company. It was continued by his successor, George Everest (after whom the mountain is named) and completed in 1871. See [1] for a good account of this great achievement.

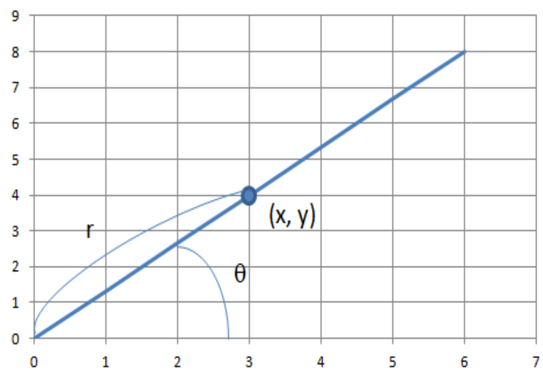
Modern Ordnance Survey maps are largely based on aerial photographs (carefully corrected to avoid the distortions introduced by the effects of perspective) and also a team of surveyors who visit in person and survey areas that cannot be surveyed using photogrammetric methods such as land obscured by vegetation. Recently I have been working as a mathematician with the Canadian equivalent of the OS who are starting to use aerial drones to survey areas of British Columbia, which cannot be surveyed directly due to the activity of local bears.

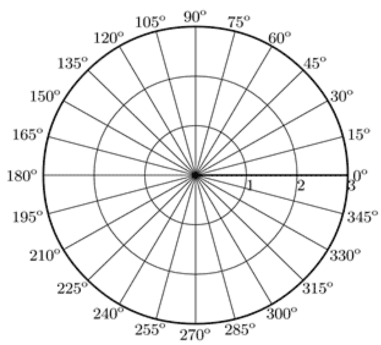
Cartesian and Polar Coordinates

Using an OS map, navigation on a mountain (as I used to tell the young people in my group, although I'm not sure if they appreciated this too much) becomes an exercise in the use of both *Cartesian* and *Polar coordinates.* In the Cartesian system we specify the position of a point by using (x,y) coordinates. In the OS system this becomes a *map reference,* usually expressed as two letter (to indicate which map we are on) and then two sets of three digits to give the (x,y) position on the map in terms of the grid on the map. For example, my office at the *University of Bath* is at grid reference ST 772 645. An OS map (1:25 000 scale) of Bath is given below.



The Cartesian system is used by GPS navigation devices to say where you are on the OS map. However, it does not really help you to find your way if, as is usually the case in mountains in the British Isles, a complete white out with nil visibility. To do this you need to use polar coordinates. In this system instead of specifying a point using (x,y) coordinates you use a distance r and an angle theta to express your position. In terms of navigation, if you want to get to a point on the map, you simply set your compass to the angle on the map (usually expressed in degrees), measure the distance on the map and then simply walk that distance. (You can measure the distance walked by counting the number of double paces that you take. In my case it is 660 double paces per kilometre). If you do this correctly than it is hard to get lost, although as errors always arise, and accumulate, it must be combined with a set of spot checks with way markers as you progress from one point to another.



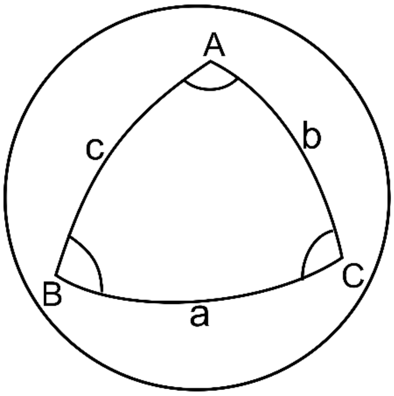


I have found this practical approach a great way not only to navigate, but in my teaching I also it to show the use of different coordinate systems, and why it is important to consider the two different approaches to representing points on the plane.

It could be argued that we no longer need to use map and compass methods now that we have GPS systems, and I will return to the use of GPS in the last part of this lecture. However, these methods have three big advantages over GPS. Firstly, they do not use batteries which can wear out, secondly, they will work even if the GPS system goes down or a satellite is not visible, and finally in a mountainous region, the effect of reflections of radio waves from high cliffs can distort the GPS system to the point where it ceases to have any accuracy at all.

Spherical Trigonometry

The triangles in the Great Trigonometrical Survey were so large that the effect of the Earth's curvature became important. Triangles drawn on the surface of a sphere such as the Earth (spherical triangles) differ from those drawn on a plane sheet of paper in one vital respect. The angles in them do not add up to 180 degrees. The study of such triangles is called spherical trigonometry, and this plays a central role both in cartography and also in celestial navigation. It was perfected in the 19th Century and described in the book by Todhunter [2].



The figure above shows an example of a spherical triangle drawn on the unit sphere. This triangle has sides a,b,c (which correspond directly to angles relative to the centre of the sphere), which meet at corresponding angles A,B,C. These are related through the *law of cosines,* and the corresponding *law of sines which* are respectively given by the expressions

$$\cos(a) = \cos(b)\cos(c) + \sin(b)\sin(c)\cos(A),$$
$$\cos(b) = \cos(c)\cos(a) + \sin(c)\sin(a)\cos(B),$$
$$\cos(c) = \cos(a)\cos(b) + \sin(a)\sin(b)\cos(C).$$

and

$$\frac{\sin(A)}{\sin(a)} = \frac{\sin(B)}{\sin(b)} = \frac{\sin(C)}{\sin(c)}.$$

From these results we can find an expression for the sum of the angles (in Radians) of the triangle, which is given by the expression below. Note that the sum is always greater than pi.

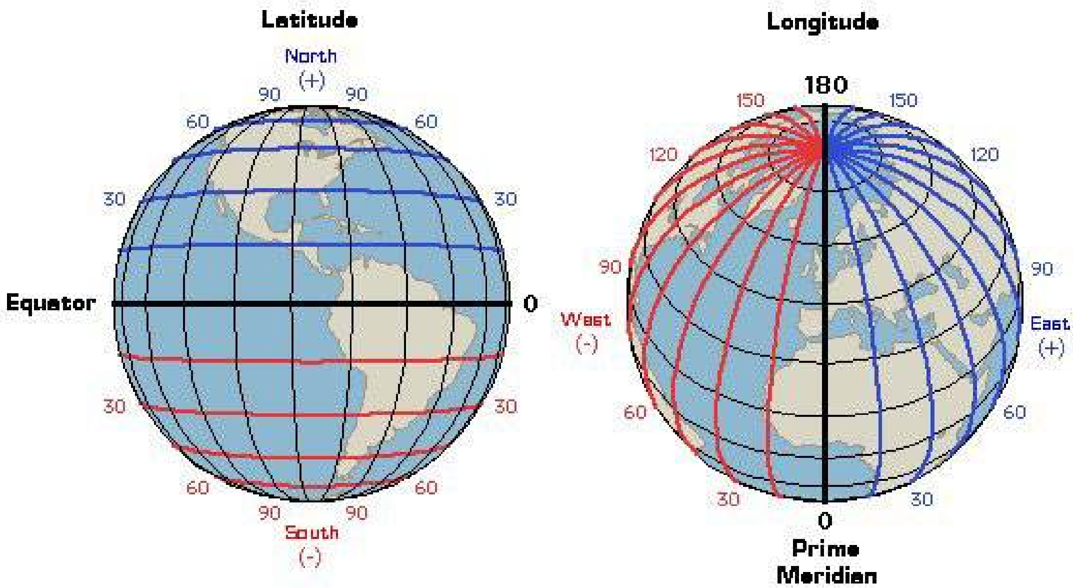
$$A + B + C = \pi + 3 \times \text{Area of the triangle}$$

These results allow us to 'solve the triangle' so that we can find the length of the sides knowing the values of the angles. As we shall see, this is a crucially important result in celestial navigation.

2.3 Maps on the surface of the Earth.

One of the problems we have with all large scale navigation is that the Earth is not flat. As we have already seen with the survey of India, every large scale map has to take into account the fact that the Earth is a sphere. However, in order to print a map onto paper we need to make a projection of the sphere onto a flat surface. To do this we have to make certain compromises, and these lead to interesting mathematics, as well as to interesting politics.

To represent a map on the Earth we immediately need a suitable coordinate system for describing points (x,y,z) on the Earth's surface. A useful such system is given by Latitude and Longitude coordinates. In these the Latitude measures the angle of the point between the centre of the Earth and the Equator, and the Longitude the angle between the points and the *Prime Meridian* which is the great circle running through the Poles and Greenwich. (The reason for the choice of Greenwich will become clear shortly.) See below.

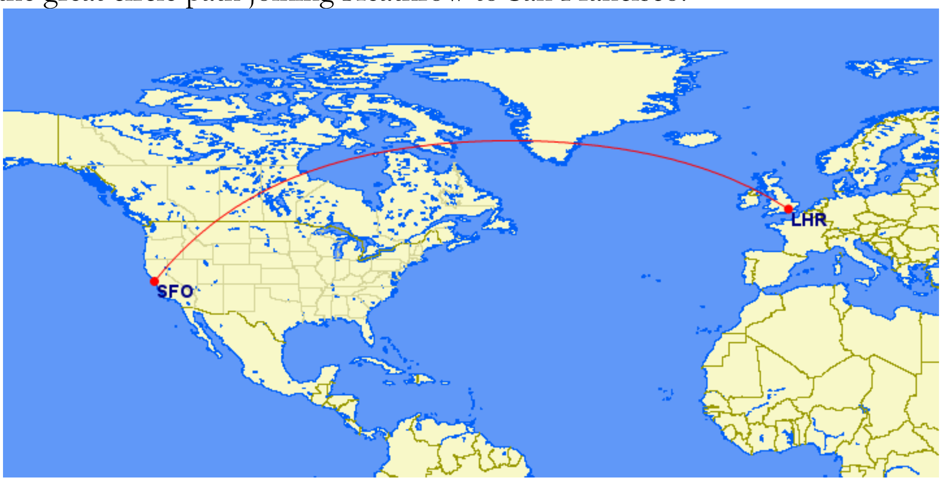


In this system Gresham College is at 51.5175 Degrees North and 0.1098 Degrees West. If theta is the angle of Latitude, phi the angle of Longitude, and R the radius of the Earth then we have

$$\begin{aligned} x &= R\cos(\theta)\cos(\phi), \\ y &= R\cos(\theta)\sin(\phi), \\ z &= R\sin(\theta). \end{aligned}$$

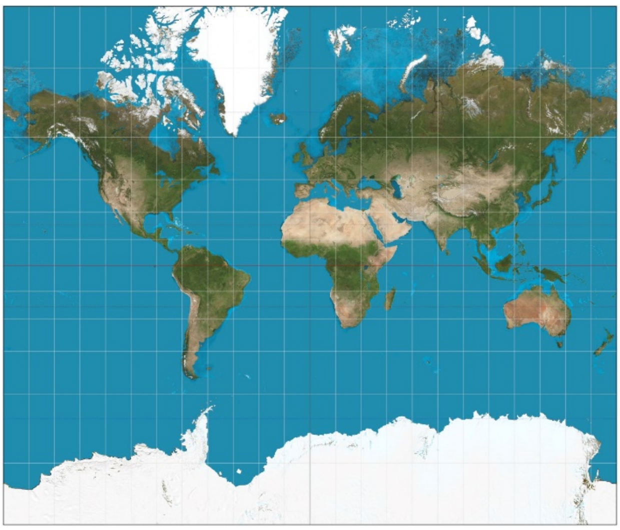
There are many attractive features of the Latitude-Longitude system. One of these is that it is an *orthogonal coordinate* system, so that the lines of constant Latitude and constant Longitude meet at right angles. A (significant) disadvantage is that the coordinate system has singularities at the two poles where all of the lines of Longitude meet. This means that the North Pole has not got a well-defined value Longitude. As we shall see, this causes computational problems. Latitude is always expressed in degrees (from -90 to 90) but Longitude can either be expressed in degrees from 0 to 360, or in hour angles. As the Earth rotates a full amount in 24 Hours, it must rotate 15 degrees per hour. Thus every 15 degrees of Longitude is one Hour Angle. Put another way every minute of arc (which is $1/60^{th}$ of a degree) corresponds to 4s of the Earth's rotation. The Nautical Mile (nm) is defined to be that part of the Earth's surface corresponding to such a minute of arc. Thus the Earth rotate 1 nm every 4s. The implication of this is that if a clock is to be used to measure Longitude, it has to be correct to 4s if the corresponding position is to be within one mile of being correct.

Another important concept when looking at the Earth's surface is that of the *great circle*. This is a curve given by drawing a circle on the surface of the Earth with its centre coincident with the centre of the Earth. If A and B are any two points on the Earth's surface, then (unless they are exactly opposite each other) there is a unique great circle joining them. This is then the shortest distance path (called a geodesic) joining them together. This fact is of vital importance to airlines, which fly great circle paths to minimise distance, flight time and cost. In the figure below we can see the great circle path joining Heathrow to San Francisco.

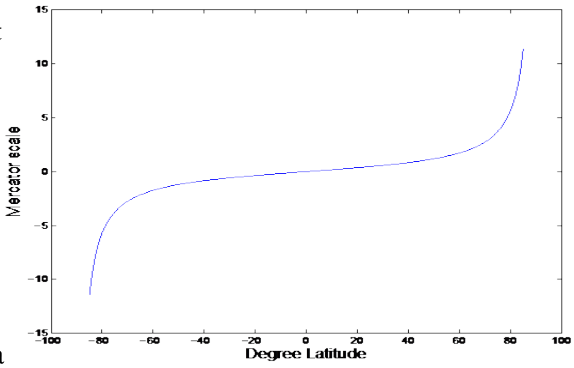


Every meridian, which is a line of constant Longitude, is a great circle. However, lines of constant Latitude are not. It is noticeable that the great circle path is very different from the straight line from San Francisco to Heathrow as it appears on the map. This is due to the distortion in the map scale as we approach the Poles

The most widely used map projection is the familiar *Mercator Projection* see below.

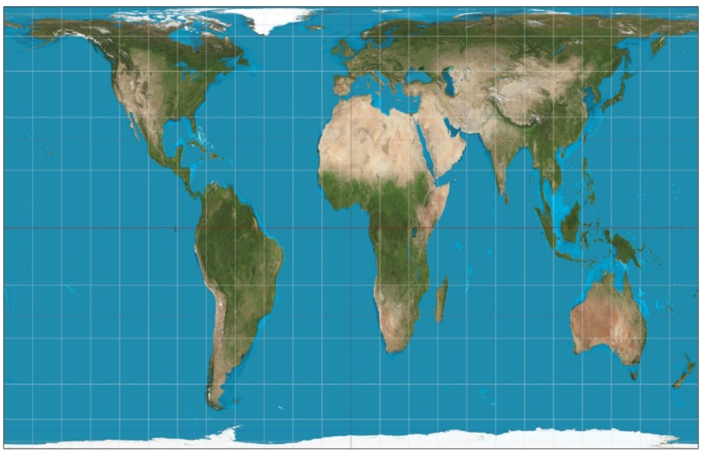


In this projection the lines of constant Latitude and Longitude are straight. This is convenient as it means that if a ship sails at a constant bearing then this is a straight line. Thus the path of a ship could be transferred to the Mercator map with a protractor and a parallel ruler. This type of projection introduced in 1569 by Gerardus Mercator. The meridians are equally spaced parallel vertical lines, and the parallels of latitude are parallel horizontal straight lines that are spaced farther and farther apart as their distance from the Equator increases. In particular a line of latitude at angle theta is plotted at the position tan(theta). It is less practical for world maps, however, because the scale is greatly distorted, and is far from linear, as theta approaches +/- 90 Degrees as we can see below.

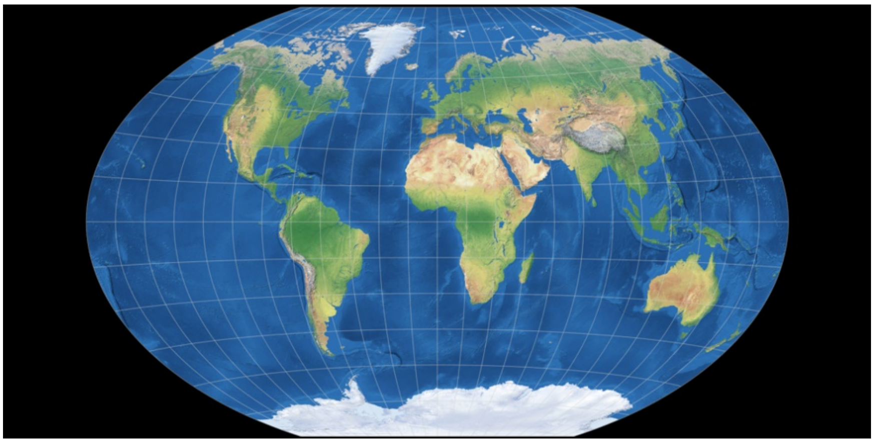


Because of this distorted scale, areas farther away from the Equator appear disproportionately large. On a Mercator projection, for example, the landmass of Greenland appears to be greater than that of the continent of South America in actual area, Greenland is smaller than the Arabian Peninsula. This leads to political issues, with countries near to the pole (such as the UK and North America) appearing larger, and thus more important, than countries (such as those in Africa) nearer to the Equator.

Other projections have been developed to overcome the distortions of the Mercator projection. None are entirely satisfactory, as all of them introduce some form of distortion somewhere or other. This is in fact a consequence of the celebrated mathematical result often called the *Hairy Ball Theorem*. This says (and I paraphrase) that you cannot smoothly comb a head of hair (approximating a head as a sphere) without at least one hair standing upright. In the context of a map this means that if you try to smoothly project the Earth's surface onto a two-dimensional map then there must be a singularity at some point, which will introduce significant distortion at its location. It's a mathematical fact that somewhere, some politician, will be offended by the consequences of this mathematical theorem. Thus all maps of the Earth's surface need to introduce some approximations between accuracy and distortion, at some point. One way to do the projection is to consider a cylinder passing through the Earth and then to project the surface onto the cylinder. In the Mercator projection the cylinder touches the Earth (a tangent cylinder). In others the cylinder passes through the Earth (a secant cylinder). An example of this is given by the Gall-Peters projection (see below). This projection is an example of the cylindrical equal-area projection with latitudes 45° north and south as the regions on the map that have no distortion. The projection is named after Gall and Peters. Gall is credited with describing the projection in 1855 at a science convention. The film-maker and journalist Arno Peters brought the projection to a wider audience beginning in the early 1970's and they are widely used by British schools. True to form, the Gall–Peters projection achieved notoriety in the late 20th century as the centre of a controversy about the political implications of map design. But as we have seen, this is just a consequence of the Hairy Ball Theorem.

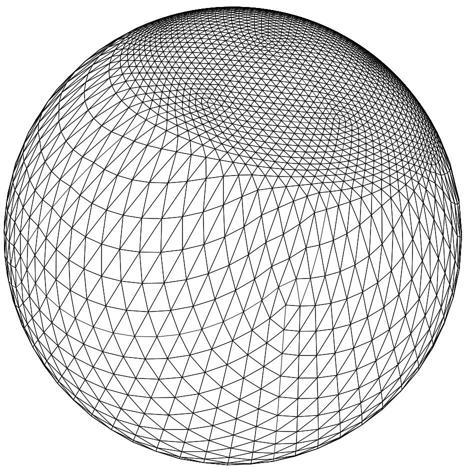


Other projections include the Bartholomew's (used by the Times Atlas of the World [3]), and the Winkel-Tripel projections in which the surface is projected onto a (pseudo-) cone. This is less distorting than the other projections above, but has the disadvantage that the lines of Longitude and Latitude are no longer straight.



Forecasting the weather

As an aside, this discussion has an interesting impact on my own work on weather forecasting. To forecast the weather accurately over the whole of the Earth we need to solve the equations for the weather (which I discussed in the last lecture on the Mathematics of Climate Change) on the sphere. To do this we need to put a mesh on the sphere and to represent the equations on this mesh. An obvious choice of mesh is to use the Latitude-Longitude coordinates, and indeed this is what is currently used for most weather forecasting. However, this choice has the same problems as those, which arose in the Mercator projection. Namely the areas around the Poles get distorted. These meshes also have the problem that they treat the Poles in a special manner and it is hard to vary the mesh to concentrate the points elsewhere. As a piece of ongoing research we are now looking at meshes, which move away from Latitude-Longitude coordinates. One choice is to use mesh points based around an Icosahedral lattice. These offer the prospect of much better resolution, and get away from problems at the Poles. An example of just such a mesh, generated using an algorithm developed in [4] is shown below.



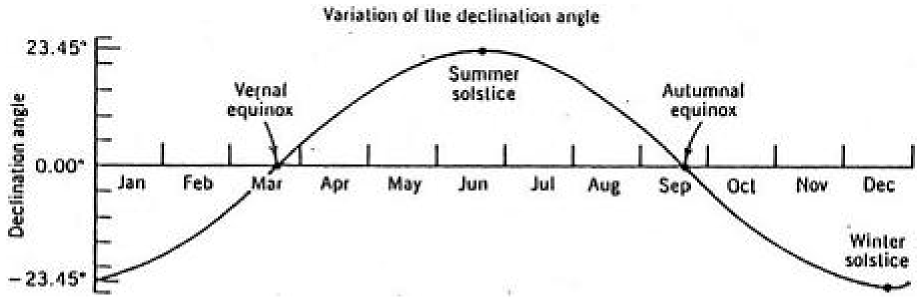
3. Celestial Navigation and a question of Longitude.

As we have seen, a point on the surface of the Earth can be described exactly by its Latitude and Longitude coordinates, phi and theta.

Finding where you are on Earth's surface is this equivalent to determining your exact Latitude and Longitude. In order to find your position to an accuracy of one nautical mile this requires finding position to an accuracy of one minute of a degree.

Calculating Latitude

Determining Latitude is relatively easy provided that the angle of the Sun at mid-day can be found with precision. The regularity of the motion of the Earth around the Sun (or equivalently of the perceived motion of the Sun around the Earth) means that on any day of the year the declination of the Sun at mid-day, defined as its angle d measured above or below the Equator, can be calculated with precision. The declination varies during the year due to the tilt of the Earths pole of 23.45 degrees relative to its orbit, and also because the orbit is elliptical rather than circular. The graph of the declination is shown below.

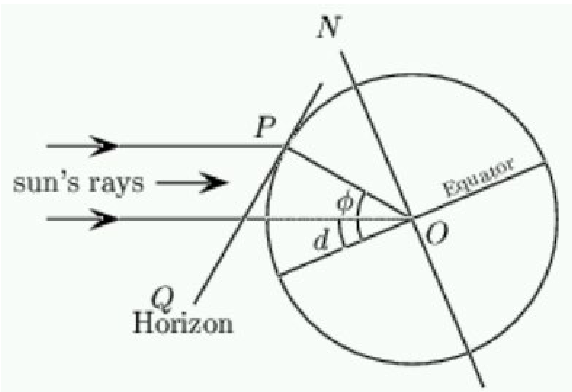


If N is the number of days which have elapsed since 00:01am on 1st January, then (assuming that all angles and calculations are made in degrees) then d is given by the formula

$$d = \arcsin\left(\sin(-23.44)\cos\left(\frac{360}{365.24}(N+10) + 1.91\sin\left(\frac{360}{365.24}\right)\right)\right).$$

This formula is very important for modern day solar energy generation, as the angle of the sun above the Horizon at mid-day determines the amount of energy that can be extracted from photo-voltaic devices.

The value of d is given to high precision in the Ephemerides tables which, in the 17th and 18th Centuries were calculated by human computers.



The declination and the Latitude are shown above. A little geometry shows that the angle Z that the Sun makes to the Horizon at mid-day is given (in degrees) by

$$Z = 90 + d - \phi$$

Hence, if Z is measured accurately, using a sextant, and d is given in the Ephemerides or by the above formula, then the Latitude can be calculated with precision. The same calculation can be used for bright stars, such as Rigel or Antares, provided that the declination of these is tabulated in the Ephemerides

Calculating Longitude

Calculating Longitude is much harder than calculating Latitude, and it became one of the most celebrated questions of its age. The history of the question of the search for Longitude is well known, see [5] for an excellent account. The importance of the search for Longitude can be seen from the opening verses to this lecture in which we see that determining Longitude was considered a suitable occupation for a Gresham professor. Indeed Robert Hooke, my predecessor as Gresham Professor of Geometry, himself worked on this problem. It is possible that these verses were written in jest, as at the time that they were published the search of Longitude had acquired the same notoriety as the quest for the Holy Grail and the turning of Lead into Gold. However, unlike these latter questions, the search for Longitude had huge economic importance and was scientifically possible. So important was the question of finding Longitude that the British Government offered a prize of up to £20,000 (equivalent to £2.89 million in 2018) under the 1714 Longitude Act. This was a huge amount at the time, and not unreasonably, the question attracted enormous interest from the great minds of the time. These included Sir Isaac Newton, and the Astronomers Royal, Flamsteed, Halley, Bradley and Maskelyne.

The key to determining Longitude is that of finding time. We have already seen that each hour corresponds to a change in Longitude of 15 degrees. One way to find the Longitude is to find the time at Greenwich when the Sun is at its highest point in the sky at your own location. The difference, between the Greenwich time and the local time thus gave the Longitude, with 1nm corresponding to a 4 second difference. Thus, if a time-keeper could keep time to an accuracy of 4s or better then navigation to a precision of 1nm was possible. The question was whether such a time-keeper should be celestial or mechanical, and this led to a fierce battle between the astronomers and the clock makers.

Moons of Jupiter and Galileo

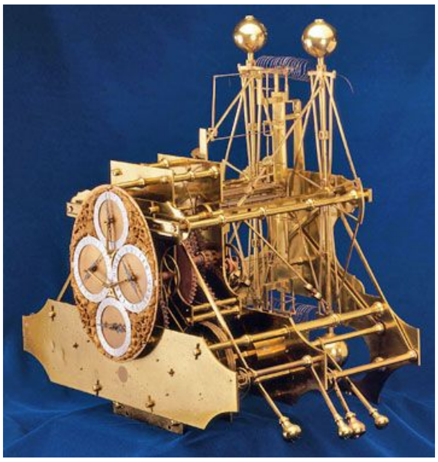
We met Galileo in the last lecture on Chaos Theory. It is well known that he was the first scientist to point a telescope at the Heavens. It is perhaps less well known, that by doing so he devised a useful method of determining Longitude on land. One of the main discoveries that Galileo made was that of the moons of Jupiter. These orbited Jupiter very regularly and Galileo recognised that they provided an accurate celestial time piece. By observing the times of the various transits of the moons across the face of Jupiter, and comparing these with the pre calculated times it was possible to determine the difference between the local time and a time reference. By doing this it was then possible to find the Longitude. This method was adopted after 1650 and was reliable on land. Using it, map makers could make more accurate maps, sometimes leading to significant revisions of existing ideas of the size of nations. Indeed, Louis XIV is quoted as saying that he had lost more of his realm to the astronomers than to any hostile power. Unfortunately, the difficulties in observing the moons of Jupiter ruled this method out for navigation on ships.

The Lunar Distance method

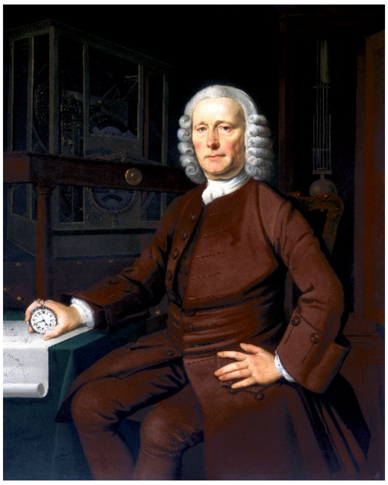
The lunar method was for some time the preferred method for finding Longitude. Whilst now almost completely eclipsed (bad pun) by the use of the Chronometer it was used (and still is today) if a time piece is not available. It also lead to the very important mathematical advance of using numerical methods to solve ordinary differential equations (and thus indirectly to my own position at Bath). The lunar method uses the Moon as a celestial time keeper. Essentially the method works by the navigator observing precisely the location of the Moon in respect to the Sun and to various fixed stars. If the motion of the Moon can be predicted in advance with sufficient accuracy, then by comparing the observations with the predictions it is possible to deduce the time in Greenwich when the Moon is at this location. By comparing this with the local time then the Longitude can be calculated. The key factor in making this system work (apart from the need to make accurate measurements) is the advanced calculation of the position of the Moon. It was the difficultly in doing this, which delayed the construction of effective lunar tables, and hence the introduction of the lunar method. A break through in the 1750s came via the great mathematician Leonhard Euler. Euler is widely regarded as one of the greatest mathematicians of all time. His contributions to mathematics were huge and in many areas. In particular he worked on the theory of differential equations, and applied this theory to the motions of the Sun, Earth and Moon. Using this theory it became possible to predict the motion of the moon with precision and the mapmaker Tobias Mayer used these predictions to produce tables of the Moon's motion. These could then be used to make the lunar distance method practical, and it could find Longitude to within half a degree. The astronomer Royal at the tome was Bradley and he became a supporter of this method. Whilst theoretically sound, and usable by an expert navigator, the lunar distance method had many problems. Firstly, it involved many time consuming and error prone calculations. Secondly, observations of the Moon were themselves subject to errors due to parallax (as it is much closer to the Earth than the Sun) and the refraction of the Earth's atmosphere.

The invention of the Chronometer

The true answer to the problem of Longitude came not from clever mathematics alone but from a combination of mathematics and precision engineering. Instead of telling time from the stars it was easier, and much more effective, to tell the time from a watch. This was, of course, not a new suggestion at the time of the announcement of the prize, for example it had been suggested in 1530 by Frisius, and the objection was obvious. Clocks at the time relied on lubricated mechanical movements, mainly driven by pendulums. These could not be used on a ship, due to the motion of the ship on the sea, and also suffered from changes in temperature. The latter both altered the length of the pendulum and also changed the viscosity of the lubricants. The mathematicians and astronomers of the day thought that it would be impossible to overcome these problems. Fortunately, a clock maker, John Harrison, disagreed with them. Harrison developed a series of clocks with extremely low friction bearings, which were also temperature compensated so that they would keep the same time in all climates. He also developed opposing pendulum actions, which were not affected by the motion of the ship in the sea. Despite opposition from the Board of Longitude, but supported by George Graham the clockmaker to the King, Harrison produced his first clock H1, illustrated below. When this went by sea to Portugal in 1736 it proved to be highly accurate.



Harrison, however, was not satisfied with it. He went on to develop two further large clocks, H2 and H3, before changing the design completely and producing the watch H4 in 1761. H4 was tried on two sea voyages to the West Indies in 1761 and (to make sure) in 1764. It proved sufficiently accurate to replace the Lunar Method, much to the chagrin (and opposition) of Meskelyne, the current Astronomer Royal. As a result of the success of this method not only did Harrison win the Longitude prize, but Greenwich had the honour of being the marker point of the Prime Meridian of zero degrees Longitude.

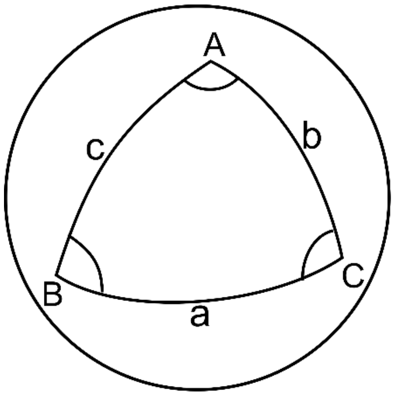


John Harrison is illustrated above. This portrait was painted by James King in 1766 and now hangs in the Science Museum. In the background you can see two of his clocks, including H3. Interestingly enough the watch he is holding is not H4, but an earlier watch made for him by Jefferys, which is said to have inspired H4. H4 itself is illustrated on the right. It was not available at the time of the portrait as it was being investigated by the Board of Longitude at the time.

It is worth saying that by inventing the Chronometer, not only did Harrison solve the problem of Longitude, but he also ushered in the Modern Age with its reliance on sophisticated technology. The Chronometer H4 was truly the super computer of its age.

Modern Celestial Navigation using the intercept method.

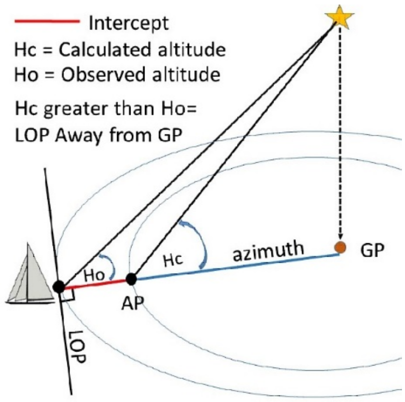
Given an accurate time-piece, modern Celestial Navigation is now routine and is done using the Marc St. Hilaire or Intercept Method. This is an effective method, although great care still has to be taken as errors and corrections which have to made, can still change a ship's location by several miles. This method relies on having a reasonable initial estimate (assumed position AP) of the ship's location (obtained by using dead reckoning) and then finding a correction to this. This greatly simplifies the calculation by linearising the complex geometrical equations about AP. These linear equations are much easier to solve than the full equations. To determine the position of a ship, the altitude angle above the horizon of a convenient celestial body is observed using a sextant and the time of this observation is recorded. Various corrections to this (due to the effects of instrument error, the height of the observer and the refraction of the atmosphere) are then made to give a true altitude Ho. Using the Ephmerides (with appropriate use of interpolation where necessary) it is then possible to determine the geographic position (GP) of the same body at that time. This is the position on the Earth's surface intersected by a line joining the centre of that body to the centre of the Earth. The altitude Hc and azimuth Z of the body as seen from the assumed position AP can then be calculated using spherical trigonometry. In particular returning to the figure we saw earlier, shown below, we can take A to be the North Pole, B to be the GP and C to be the AP. The angle A is the difference between the Longitude of the GP and the AP, c is 90 – the declination of the GP, and b is 90- the Latitude of the AP.



In this figure the angle C is now the azimuth Z of the GP as seen from the AP, which is the angle on a chart relative to North of a line pointing from the AP to the GP. Similarly, the angle a is the declination of the GP as seen from the AP, which is the angle of the celestial body above the Horizon measured by the sextant. Using the spherical trigonometric formulae given earlier we can compute these easily via the formulae

$$\cos(a) = \cos(b)\cos(c) + \sin(b)\sin(c)\cos(A),$$
$$\sin(C) = \sin(c)\sin(A)/\sin(a).$$

The calculated altitude Hc is then given by Hc = 90 – a. Of course in the past these formulae would have been given in tables of Haversines and similar quantities. Now they can be calculated easily using a scientific calculator. These computed values are then compared with the observed value Ho. The navigator marks the AP and draws a line in the direction of the azimuth given by the angle C computed above. They then measure the intercept distance along this azimuth line, towards the body if Ho > Hc, measuring one nautical mile for every minute of arc difference, and away from it if Ho < Hc. At this new point they draws a perpendicular line of position, LOP, to the azimuth line. The position of the ship lies on this line as illustrated below.



To fix the position of the ship this procedure is repeated for one or more celestial bodies, and the ship then lies on the intersection of the various calculated position lines. See [??] for more details of this method.

The Invention of The Computer

It is often said that it was the stimulus of World War 2, in particular the needs to crack the Enigma and other codes, and also the design of the atomic bomb, which led to the invention of the computer. It is certainly true that the modern programmable electronic computer can trace its origins back to the Colossus and ENIAC machines, which were constructed to perform these tasks. However, the basic ideas behind the modern computer predate this by one hundred years, and were mainly driven by the problems of navigation.

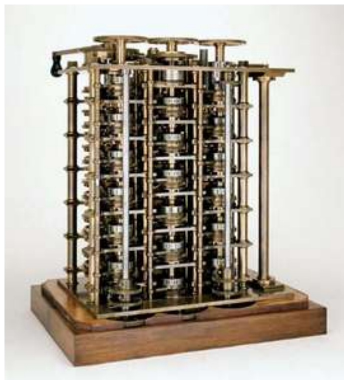
The Ephemerides described above required a huge amount of calculation to produce. These calculations were done by ‘computers’ who were human beings equipped, if they were lucky, with mechanical calculating devices. Many of these calculations were routine, and involved the use of *difference tables* and *Newton’s method of divided differences*. Essentially this method starts with a series of numbers which follow from the evaluation of a polynomial. If differences of these numbers are taken, and then differences of these differences, and so on, then eventually the result is a series of constants. If this process is reversed then it gives a means of evaluating the original polynomial. This process reduced quite complex calculations, such as the evaluation of polynomials, to a large number of systematic additions. Below we can see the finite difference table used to evaluate the polynomial

$$y = 1 + \frac{3}{2}x^2 + \frac{1}{2}x^3$$

For x=0,1,2,3,4. The table starts with a series of 3s in the bottom row. These are added to give the next row up. This process continues to build up the values of the final polynomial.

1		3		11		28		57
	2		8		17		29	
		6		9		12		
			3		3			

A very similar process can be used to construct the tables of numbers in the Ephemerides using extended difference tables. Although these are tables of spherical trigonometric functions, they can be approximated by polynomials, which can in turn be evaluated by difference tables. This procedure is extremely straightforward, but it is also time consuming and prone to error. Up to the 20th century the calculations in the difference tables were made by hand (or at best by simple mechanical calculators) by human *computers*. (In the film *Hidden Figures* you can see a team of computers in action.) However, in 1823 the mathematician Charles Babbage came up with the revolutionary idea that the same routine difference calculations could be done mechanically. He designed a mechanical computer to do this calculation which was called the Difference Engine. A substantial government grant of £17,000 led Babbage and Joseph Clement (also aided by Ada Lovelace) to produce in 1832 a small working model of the calculating section of Difference Engine No. 1, which operated on 6-digit numbers and second-order differences. Unfortunately, the full project was not accomplished successfully, and it was abandoned in 1842.



Babbage then went on from the design of the difference engine to the much more sophisticated analytical engine which could be programmed. Unfortunately, this machine was not constructed either, although the ideas behind it led to the invention of the modern computer. In the science-fiction novel [7] it is speculated what might have happened if Babbage had been successful, and analytical engines had been mass produced in the Victorian age. Fortunately, the story of the Difference Engine has a happy ending. From 1985-1991 the (UK) Science Museum constructed a working calculating section of difference engine No. 2. In 2002, the printer, which Babbage originally designed for the difference engine, was also completed. Both the difference engine and the printer have worked perfectly ever since. I shall return to the operation of the Difference Engine in my Gresham lecture next year on the ‘Future of computing’.

4. Navigation in WW2

During World War 2 the major offensive of the British against the Germans, in the years before D-Day in 1944, was the offensive by RAF Bomber Command. Due to the effectiveness of the German defences, these attacks had to be made at night. The aircraft therefore had to find their targets in the dark over a blacked out country. The original plan for doing this was to use celestial navigation and the stars and heavy bombers such as the Avro Lancaster and Handley Page Halifax had an astrodome (illustrated below), which allowed the navigator to do just this.

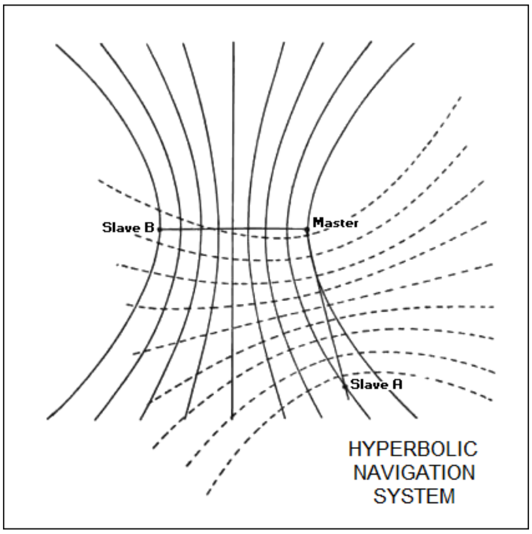
Unfortunately, this navigational method proved inaccurate for bombing due to the speed of the aircraft and the way that stars were usually obscured by cloud. As a result, early bombing was extremely inaccurate with accuracies no better than tens of miles.

As a replacement to celestial navigation it was essential to introduce electronic navigational tools. The German Luftwaffe led the way with this in the invention of the Knickebein and X-Gerat systems, both of which relied on the use of intersecting radio beams. These could direct a single group of aircraft to a single target, but could not be used for general navigation. In contrast the RAF developed the Gee hyperbolic navigational system. This was invented by Robert Dippy and was in introduced in 1942. However, it can trace its roots back to the methods of detection of artillery pioneered in the First World War. In this earlier system sound recorders (now called transducers) were placed at different points in the trenches would record the time at which they hear a gun fired. If the time difference T between two such recordings at locations L1 and L2 (assumed after that at L1) was measured then the difference in distance D of the gun from those two recordings was D = c T, where c was the speed of sound. The possible location of the gun then had to lie on a curve of points z=(x,y) so that |z - L2| - |z – L1| = D (where |z| is the length of the vector z). Such a curve is a *hyperbola* which is one of the classical conic section curves discovered by the Greek geometer Appolonius. This curve has the canonical formula

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

but it can also be rotated into any orientation. If another pair of signals was recorded from two other transducers, then the gun had to lie on a second hyperbola. Its location was then on the intersection of the two hyperbolae.

Essentially the same idea was used for the Gee system. In this case a master station and two slave stations (marked A and B below) based in the UK sent radio pulses to the bomber aircraft over Germany. These were then received in the bomber and the time difference between the slaves and the master signals was recorded on the Gee receiver. This was then plotted on a chart of Hyperbolae (carefully corrected to allow for the curvature of the Earth). The intersection of the hyperbolae corresponding to the two slave stations then gave the position of the aircraft. This is shown below (left) with a Gee receiver shown on the right.



Experimental systems, which were set up in June 1940, showed that the Gee was usable to a range of at least 300 miles at altitudes of 10,000 feet. The system then went into use the following year, and was used to the end of the war, although it became much less effective due to the use of jamming by the Germans. The accuracy of the Gee system depended on two factors. One was the accuracy in measuring the time difference between the two received signals. It was estimated in practice that this could be done to about 0.666 micro seconds ($1/10^{th}$ of a Gee unit) corresponding to a distance of 0.125 miles. The accuracy of the system also changes with distance as the hyperbolae become closer to straight and the crossing point harder to find accurately. At short ranges accuracies of 165 yards (151 m) were possible, while at long range over Germany the accuracy was about 1 mile (1.6 km). This meant that Gee could be used to get a bomber into the right location, which was a vast improvement on Celestial Navigation, but it could not be used as a blind bombing aid. A later development, the Oboe system, which also relied on radio navigation, was much more accurate and could be used in this way. Oboe could only be used for one aircraft (typically a Mosquito bomber). Guided by Oboe, this would then release markers over the target, which the rest of the bomber force could then be used as target indicators. The Gee system is still in use, in a more sophisticated form, as the Loran navigational system.

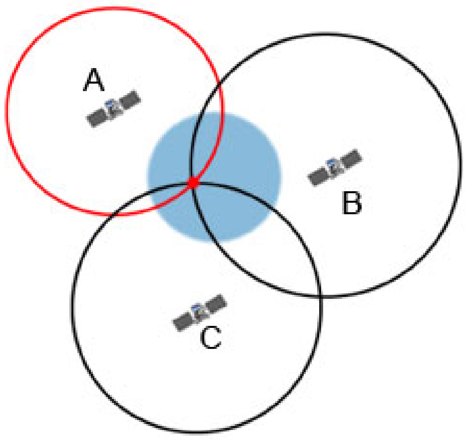
5. The GPS revolution

If we want to find out where we are in the 21^{st} Century, the chances are that will use the Global Positioning System (GPS). This has an accuracy of around 10-30 m errors, allowing navigation in real time. It is widely used in mobile phones, SatNav and many other navigational devices. Like celestial navigation, GPS relies on careful time measurements. It also needs to take many factors (including special and general relativity) into account to obtain accuracy. GPS uses more than 24 satellites in 12 hour orbits at 20183 km, covering 6 orbital planes at 55 degrees. From any point on the Earth's surface there are always five such satellites in view. Each satellite transmits two signals at 1227 MHz and at 1575 MHz, this allows for interference from one signal to be corrected by the other. The signal contains precise the precise time and location of the satellite at transmission.

At a position X = (x,y,z) on the Earth's surface the time of reception of the signals from the visible satellites is recorded measured against the clock in the receiving device. The satellite times for the transmission of each of these signals are carefully synchronised with each other and are at (say) times T1,T2,T3,T4,T5. At each of these times the satellite position is known (by careful calculation) to be at locations X1,X2,X3,X4,X5. We suppose that the time of reception of these signals as measure at the receiver is TR. However, the receiver is not quite correct and the true time T = TR + TO where TO is the (unknown) receiver time offset. The distance Di from the receiver to the satellite Si is then given by

$$Di = c\,(T - Ti).$$

Where c is the speed of light. To find the position of the receiver we have four unknowns which are the position (x,y,z) and the time offset TO. Hence if we measure Di for four satellites then we can locate the receiver. The fifth satellite allows for the assessment and correction of any positioning errors.



The GPS positioning accuracy is dependent on a number of interacting factors. Some of these are very similar to those involved in Celestial Navigation and include errors in calculating the position of the satellite and in the timing of the clocks on board both the satellite and the receiver. The difference is one of scale. As all errors get multiplied by the (very large speed of light), a high precision is required. For example, whereas Harrison's chronometer had to be accurate to seconds per day, the clock in a GPS satellite has to be accurate to micro seconds per day. Similarly, the position of the satellite has to be known to very high precision.

Some of the errors can be determined in advance, or as part of the calibration process. These are typically systematic errors in the satellite transmitter. Other errors can be calculated in advance. For example, A very interesting feature of GPS is that it is one of the few technologies which makes direct use of Einstein's Special and General theories of Relativity. In 1905 Einstein published his Special Theory of relativity. This made the prediction that a fast-moving clock runs slower than a stationary one. As a consequence, the clock on the fast-moving satellite in orbit around the Earth runs more slowly than one on the surface. The difference is about 7 micro seconds per day. Einstein's General Theory of Relativity was published in 1915 and made the prediction that the time measured on a clock is altered by the gravitational field it is in. This field changes as the satellite moves further away from the Earth. The equations of General relativity accordingly predict that a clock at 20,000 km altitude will run more quickly than those on the surface of the Earth. The difference due to this is about 45 micro seconds per day. Taken together this means there is a 38 micro second per day error due to relativistic effects. This may not seem much, but when multiplied by the speed of light, this amounts to an error of about 11 km per day in distance, which is very significant. Fortunately, this error is predictable in advance and can be easily corrected for.

Other errors which are much harder to assess come from estimating the propagation in the signal. The calculation above assumed firstly that the signals from the satellites travel in straight lines, and secondly that the speed of light c is constant along the path from the satellite to the receiver. As we have seen earlier in this talk, the GPS signal path in the mountains may be far from a straight line, as it may bounce several times off the faces of cliffs. This leads to a much larger estimate of the distance of the receiver from the satellite which can, in turn, lead to dramatically large errors in the position estimate. Fortunately, these tend to self correct as the receiver moves, although they may see their location jump abruptly. Uses of SatNav devices may be familiar with this happening to them as they pass through the urban caverns encountered in cities, with many reflections possible off buildings. The change of the speed of light is harder to correct for. Typically this is due to changes in the electron density of the ionosphere in the GPS path. This is the reason for sending two signals at two different frequencies. These are affected differently from each other by the ionosphere, and by looking at the difference in the arrival time, not only can the propagation error be assessed but also the electron density along the path. The latter information can then be used to asses the overall state of the ionosphere and to help predict ionospheric storms [8].

6. Solar flares. An end to civilisation as we now know it?

In November 2018, the Times ran a story about space weather [9]. This is a very important topic for modern navigation. In a typical solar storm charged particles are emitted from the Sun in Coronal Mass Ejections or CMEs. These particles then travel towards the Earth. Most of the time these are harmless and lead to phenomena such as the Aurora Borealis. However, a large solar ejection would not only disrupt the ionosphere, leading to the navigational errors described above but could also disable the GPS system entirely, and even the power grid network. This possibility is so important that significant resource is now in place to both predict the CMEs and to track the emitted particles as they make their way to the Earth. For example, the Met Office has a team of space weather experts, who monitor the Sun 24 hours a day and issue warnings of heightened solar activity. If we did have a major solar storm (the last one being on 13th March 1989) it is quite possible that our entire GPS system would cease to function for some time. In this case we may have to go back to Celestial Navigation to find out where we are.

© Professor Chris Budd, 2018

References

[1] J. Keay, (2000), *The Great Arc,* Harper Collins.

[2] I. Todhunter (1886), *Spherical Trigonometry*, MacMillan and Co.

[3] The Times Atlas of the World, Times Books

[4] C. Budd, A. McRae, and C. Cotter, (2018), The geometry of optimally transported meshes on the sphere, SIAM J. Sci. Comp.

[5] D. Sobel, (1995), *Longitude,* Fourth Estate.

[6] M. Blewitt, *Celestial navigation for yachtsman*, (1971), Edward Stanford Ltd.

[7] W. Gibson and B. Sterling, (1990), *The Difference Engine*, Victor Gollancz

[8] N. D. Smith, D. Pokhotelov, C. N. Mitchell, and C. J. Budd, (2010), *Image-model coupling: application to an ionospheric storm,* Nonlin. Processes Geophys., 17, 361-369

[9] Times article (November, 18 2018), *Solar blast could knock out the power grid — and your computer.* https://www.thetimes.co.uk/article/solar-blast-could-knock-out-the-power-grid-and-your-computer-jxl02j755