Mercator's Projection: A Comparative Analysis of Rhumb Lines and Great Circles

Karen Vezie

Whitman College

May 13, 2016

Abstract

This paper provides an overview of the Mercator map projection. We examine how to map spherical and ellipsoidal Earth onto 2-dimensional space, and compare two paths one can take between two points on the earth: the great circle path and the rhumb line path. In looking at these two paths, we will investigate how latitude and longitude play an important role in the proportional difference between the two paths.

Acknowledgements

Thank you to Professor Albert Schueller for all his help and support in this process, Nina Henelsmith for her thoughtful revisions and notes, and Professor Barry Balof for his mentorship and instruction on mathematical writing. This paper could not have been written without their knowledge, expertise and encouragement.

Contents

1 Introduction

The figure of the earth has been an interesting topic in applied geometry for many years. What people know about the earth's shape has evolved tremendously; once considered to be a flat disk floating in the ocean, people now understand the earth to be an oblate spheroid. The next important milestone for cartography after discovering the earth's curved shape became understanding how to map the earth onto two dimensional space. The difficulty was deciding not only how to map from 3-dimensions to 2-dimensions but also deciding which way was most effective and useful. In Section 2 we examine one type of map construction: the Mercator Map.

The importance of this map is that it is conformal. It preserves all angles between pairs of intersecting paths [6]. Thus, the angles between paths on the map are the same as angles between paths on the earth. Therefore, shapes are also preserved. This type of mapping was very useful for navigators in 16^{th} century. They could measure the bearing they wanted to travel on by measuring the bearing on the map. We call the lines on Mercator's Map that allow this to be possible *rhumb lines*. Rhumb lines are lines of constant bearing. They are curved on the earth but straight lines on Mercator's Map. This paper investigates how to construct Mercator's Map by projecting these curved rhumb lines on 3-dimensional Earth as straight lines in 2-dimensions. In Section 3 we calculate the distance between two points on a rhumb line and compare this distance with the shortest distance possible between two points on the earth. This shortest path is called a *great circle*. We then examine the proportional difference between rhumb lines and great circles to discover some interesting properties of these different paths.

1.1 History

Water transport was a very important form of transportation in the 16^{th} and 17^{th} centuries and we still use it nowadays for exports such as oil and food. Countries in early Renaissance Europe relied on water routes for trade primarily in spices [1]. They relied on navigation for the economic livelihood of their country. For example, the Portuguese sailed the coast of Africa many times under sponsorship of Prince Henry [9] and Christopher Columbus discovered America in 1492 in his voyage across the Atlantic [1].

However, sailing the Atlantic was much more challenging than sailing the Mediterranean.

There is no land to guide one's path. Relying on a compass and constellations was the primary way to manage navigating. A common practice for navigators in the Northern Hemisphere was to use the North Star as a guide to make sure they stayed on the correct bearing. That is, mariners would measure the height of the North Star Polaris normally with a sextant and sight reduction tables but some merely used their own arm to measure the height [9]. They would then make sure they were always at the same height because this corresponded to staying on a constant latitude. However, this meant the bearing from North continuously changed. This technique was not very reliable due to cloudy nights and there was no guidance for navigation during the day.

Therefore, creating succinct and understandable ways to map the earth and understand its features was of great importance to navigation in the 16^{th} and 17^{th} centuries [1]. However, at the time people believed the earth to be a perfect sphere. In fact, the concept of spherical Earth dates back to 6^{th} century B.C. when Anaximander and Thales of Miletus, Greek geometers, postulated the earth was a sphere positioned at the centre of the universe, which itself was also a sphere [6]. Pythagoras, Aristotle and Plato all postulated this idea and philosophised that the earth was perfectly created and thus must be the perfect geometric shape: a sphere. It was not until the 17th century that Isaac Newton argued the earth was in fact an oblate spheroid or ellipsoid: a sphere that is flattened at the poles. Newton posited this on account of his research on gravitation and planetary motion. He further classified the earth as an ellipsoid of revolution, which means it rotates about its shorter axis (the axis between the North and South pole).

Newton built on the law of pendulum motion to prove there is an increase in gravity from the equator to the poles and developed a model of the earth to incorporate this flattening,

$$f = \frac{a - b}{a}, \tag{1}$$

which is simply the proportional difference of one radius versus the other. Here, a is the semi-major axis and b is the semi-minor axis [14]. It follows that this flattening equation measures the compression of a sphere along the diameter to form an ellipsoid. Newton hypothesised f to be $\frac{1}{230}$. This low number would mean the earth is very similar to a sphere. For example, when there is no difference between a and b and so the earth is perfectly round, $f = 0$. The bigger the difference between a and b, the bigger the value of f. As such, the biggest value of f is when $b = 0$ and thus $f = 1$. This would have been the value of f if the earth were flat. In order to verify this flattening and ellipsoidal calculations, one had to compute the arc-measurements at different latitudes and longitudes. That is, measure the distance between two points with the same longitude, or the same latitude.

Indeed, a group of mathematicians and physicists, including A.C. Clairaut, went on an exhibition to Lapland in 1736 to do just this. They showed the flattening of the earth is $\frac{1}{210}$, so Newton's hypothesis was very close. After this exhibition, Clairaut in fact wrote a synthesis of the physical and geodetic evidence of the earth's ellipsoidal shape [14]. He related gravity at any point on the earth's surface to the location of that point and so the earth's shape was justified by examining the measurements of gravity at different latitudes. A stronger measure of gravity at the North Pole compared to the equator indicates the length from centre of the earth at the North Pole is smaller than the length from centre of the earth at the equator. In doing this Clairaut developed a formula for the acceleration due to gravity on the surface of an ellipsoid at latitude θ:

$$g(\theta) = G_e \left[1 + \left(\frac{5m}{2} - f \right) \sin^2 \theta \right],$$

where G_e is value of acceleration of gravity at the equator, m the ratio of centrifugal force to gravity at the equator, and f is defined above in Equation (1). The acceleration of gravity at the equator is slightly less than the acceleration of gravity at the poles (G_p) where $G_p \approx 9.8322\text{ms}^{-2}$. The value of gravity that most people are familiar with, $g = 9.80665\text{ms}^{-2}$, is the acceleration due to gravity at latitude 45.32°.

Laplace [14] used this information to prove $f = \frac{1}{330}$. With modern satellite technology this number is now $\frac{1}{298.2}$ [6]. A very small number such as this one demonstrates just how similar the earth is to a sphere. We are able to make fairly accurate calculations of the earth, such as distances between two points by assuming it is a sphere. However, it is important to note these calculations do not account for the various bumps, bulges and depressions on Earth's surface. Therefore, geoid is a term used to define the shape of the earth that best approximates the earth's actual shape, smoothing the mountains, valleys, canyons and oceans [4] so that we are left with a smooth but highly irregular shape. Today's GPS systems use the reference ellipsoid WGS84 (World Geodetic System of 1984) to calculate its measurements [6].

1.2 Eccentricity and Latitude

From Equation (1) we get a very important variable, e, defined as

$$e^2 = 2f - f^2$$

where e is the eccentricity of the earth and $0 \leq e \leq 1$. More formally, we can define eccentricity

$$e = \sqrt{\frac{a^2 - b^2}{a^2}},$$

which specifies how deviant the earth is from a sphere. Note that when $e = 0$, the earth is a sphere.

Definition 1.1. [***Latitude***] Latitude is the angle between the normal to the earth's surface and the equatorial plane.

On a sphere, the normal passes through the centre and thus latitude is the angle between the normal the radius and the equatorial plane; we call this geocentric latitude. However, on an ellipsoid, the normal does not necessarily pass through the centre, and thus we must use different terminology to distinguish between the different types of latitude. Geodetic latitude is the latitude on an ellipsoid, and it is still the angle between the normal and equatorial plane. However, where the normal crosses the axis between the North and South Poles varies. Therefore, when exploring the properties of spheroidal Earth later in this paper, it is imperative to find a mathematical relationship between geodetic latitude and geocentric latitude. Figure 1 displays the difference between the two latitudes:

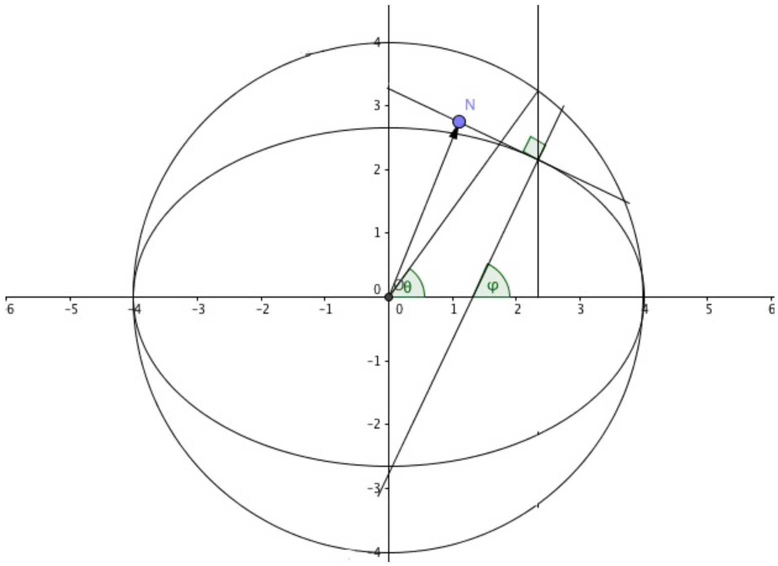


Figure 1: Geodetic (ψ) and geocentric (θ) latitude.

Firstly, recall a parametrisation of an ellipsoid is:

$$\begin{aligned} x &= a\cos\phi\cos\theta \\ y &= a\sin\phi\cos\theta \\ z &= b\sin\theta. \end{aligned} \tag{2}$$

Note here, ϕ is longitude ($0 \leq \phi \leq 2\pi$) and θ is geocentric latitude ($0 \leq \theta \leq \pi$).

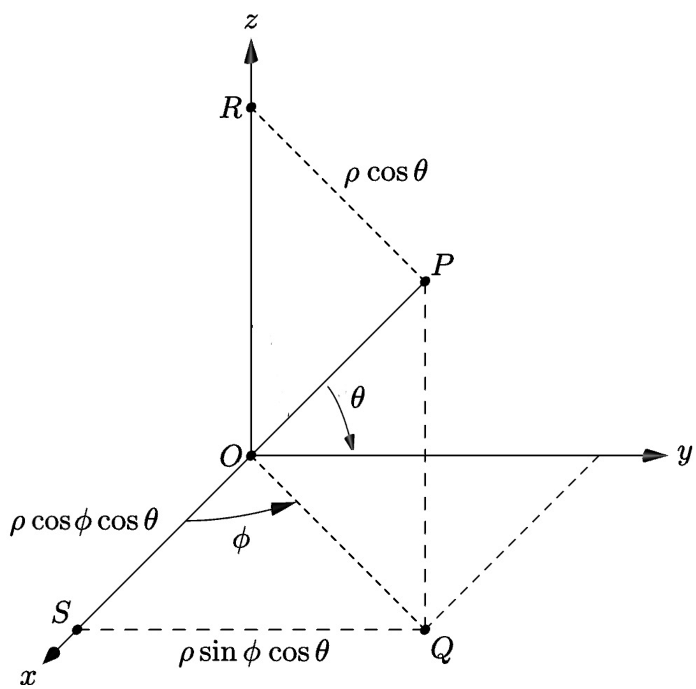


Figure 2: Spherical coordinates with radius ρ.

An outward-pointing normal to the ellipsoid at (ϕ, θ) is $(b\cos\phi\cos\theta, b\sin\phi\cos\theta, a\sin\theta)$. Taking $\phi = 0$ we find $N = \langle b\cos\theta, 0, a\sin\theta\rangle$ and $\vec{i} = \langle 1, 0, 0\rangle$. Since we want to find a relationship between θ and ψ we calculate

$$\cos\psi = \frac{N\cdot\vec{i}}{|N|} = \frac{b\cos\theta}{\sqrt{b^2\cos^2\theta + a^2\sin^2\theta}}.$$

It follows,

$$\begin{aligned}
\cos^2\psi &= \frac{b^2\cos^2\theta}{b^2\cos^2\theta + a^2\sin^2\theta} \\
\frac{1}{\cos^2\psi} &= \frac{b^2\cos^2\theta + a^2\sin^2\theta}{b^2\cos^2\theta} \\
\frac{b^2}{\cos^2\psi} &= b^2 + a^2\tan^2\theta \\
b^2(\sec^2\psi - 1) &= a^2\tan^2\theta \\
b\tan\psi &= a\tan\theta. \qquad (3)
\end{aligned}$$

Since

$$e^2 = 1 - \frac{b^2}{a^2},$$

we can easily show Equation (3) is equivalent to,

$$\tan\psi\sqrt{1-e^2} = \tan\theta. \tag{4}$$

This provides a simple formula between geodetic latitude and geocentric latitude. For much of this paper we will look at spherical Earth, so we assume geocentric latitude, but it is useful to understand the relationship when we discuss ellipsoidal Earth. Firstly however, we will examine spherical Earth and how to map it onto 2-dimensions. There were two major discoveries that occurred in the early 16^{th} century that enabled this mapping: Rhumb lines and Mercator's Map.

2 Mercator's Map

2.1 Rhumb Lines

Portuguese mathematician and cosmographer, Pedro Nunes, first introduced the rhumb line in 1537. However, before defining the rhumb line, it is important to understand what meridians and great circles are.

Definition 2.1. [***Great Circle***] A great circle is the shortest distance on Earth from one place to the next.

Definition 2.2. [***Meridian***] A meridian, or line of longitude, is any line that is half of an imaginary great circle that goes from North Pole to South Pole.

Definition 2.3. [***Rhumb Line***] The geometric representation of the curve on the spherical surface of the earth that intersects all meridians at the same angle [9].

Therefore, a rhumb line is a line of constant bearing [1]. Some refer to rhumb lines as loxodromes but for the purpose of this paper we will continue using the rhumb lines terminology. Figure 3 demonstrates the rhumb line path on spherical Earth.

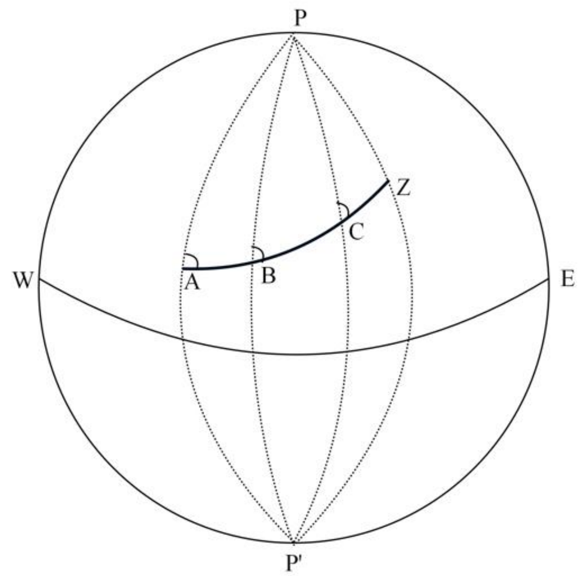


Figure 3: Rhumb line path intersecting meridians at same angle [3].

Nunes introduced the difference between following a great circle and following a rhumb line. This can be seen in Figure 4:

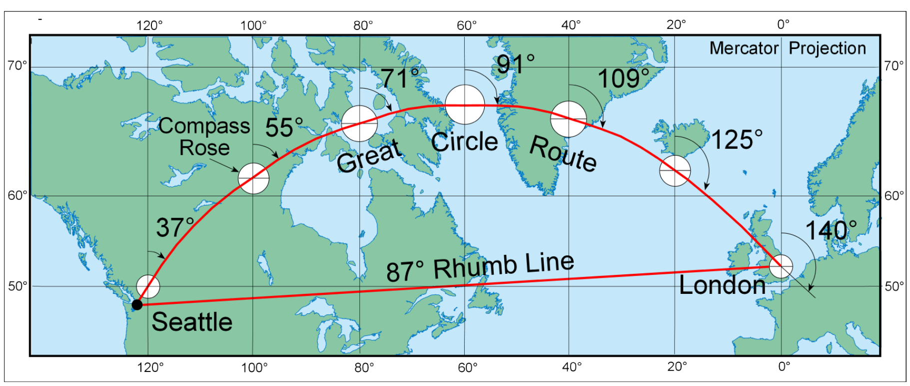


Figure 4: The difference between a great circle (4791.27 miles) and a rhumb line path (5485.6 miles) on a 2D map [8].

It is useful to note that the East-West passage along the equator (90°) is the same distance for both a great circle route and a rhumb line route. Similarly, when the angle is 0° longitude, the line will be along The Greenwich meridian, the prime meridian. Another interesting point is that if one continues to circle the earth either continuing North or continuing South, the non-great circle rhumb lines spiral to the poles but never reach the poles. We will discuss this in greater detail later in the paper but Figure 5 demonstrates this.

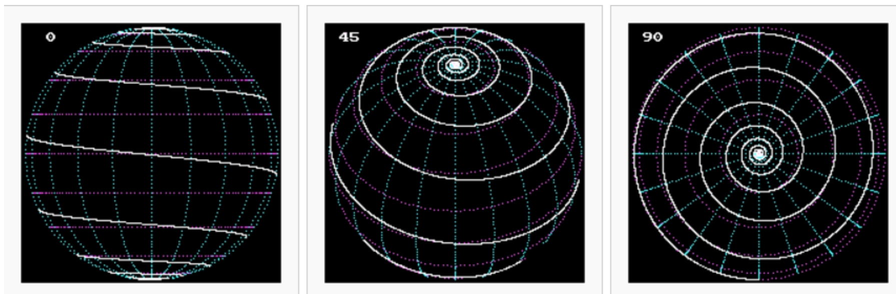


Figure 5: Rhumb lines spirally infinitely around North Pole [16].

2.2 The Mercator Map

The next important accomplishment was the invention of Mercator's map [9]. The Flemish geographer and cartographer Gerardus Mercator presented it in 1569. As seen in Figure 4, this map demonstrated rhumb lines on a flat, straight space. The map enabled navigators to draw a straight line, measure the bearing they wanted to travel in and then sail away. The invention of a magnetic compass made this navigation relatively easy.

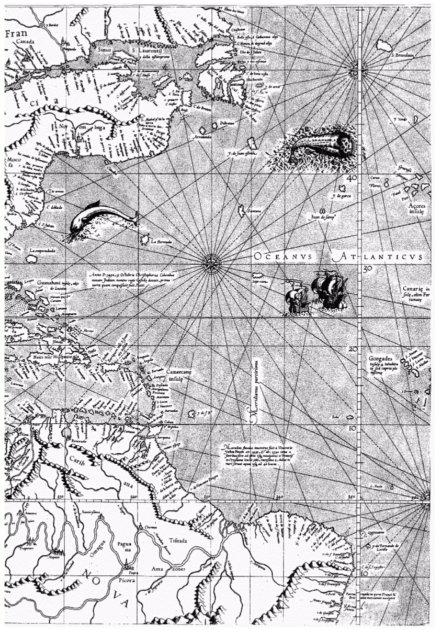


Figure 6: Original Mercator Map showing part of the Atlantic Ocean.

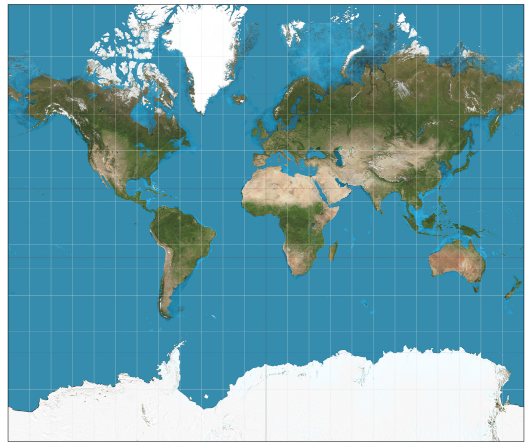


Figure 7: Modern day Mercator Map.

In these figures, especially Figure 7, we see that the northern and southern lines of latitude are extremely distorted. This accounts for why Greenland appears much bigger on the Mercator Projection than it is in reality. The next section derives the Mercator mapping and shows how the distance between lines of latitude on the map are distorted.

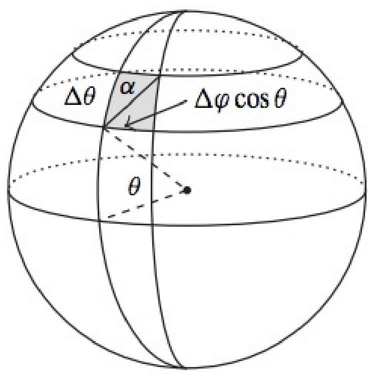
2.3 Constructing the Map

When Mercator first constructed his map, he assumed the earth was a sphere [5] since he did not know anything to the contrary. Thus, when constructing the map, we will assume the same. Mercator wanted the map to satisfy three conditions [5]:

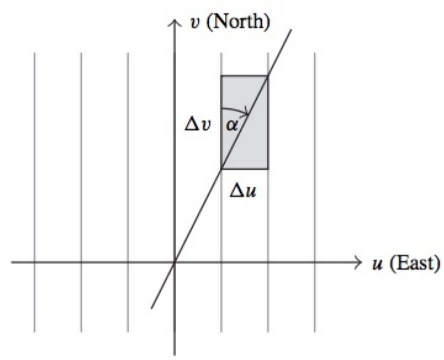
1. The north-south direction is the vertical direction;
2. The east-west direction is the horizontal direction, where the length of the equator is preserved;
3. All straight lines on the map are lines of constant bearing.

Aligning with the first two conditions, the meridians (lines of longitude) on the earth are straight vertical lines on Mercator's map and the lines of latitude are straight horizontal lines. For simplicity sake, let's examine the spherical Earth with radius one. Therefore, the equator has length 2π and thus the horizontal length of the Mercator map is 2π. Now, we must determine the spacing of the lines of latitude (parallels).

First, recall spherical coordinates where ϕ is the angle of longitude, θ is the angle of latitude from Equation (2). Here, we introduce a rectangular coordinate system where $u(\phi, \theta)$ is the horizontal direction and $v(\phi, \theta)$ is the vertical direction [5]. We want to map the rectangle on the sphere in Figure 8a to rectangular coordinates in Figure 8b.



(a) Rectangle on a sphere [5].



(b) Rectangle on Mercator's projection [5].

Figure 8: Transformation from spherical to rectangular coordinates.

To do this, we need to satisfy the third condition. Assume α in Figure 8a is the bearing from due north. Condition 2 gives us that $u = \phi$; if θ is zero, we just travel along the horizontal axis and $\Delta\phi = \Delta u$. Further, the parallel at latitude θ has radius $\cos\theta$ due to trigonometry, given we are considering the sphere. If we examine the shaded rectangle in Figure 8a we see that the parallel is $\Delta\phi\cos\theta$ and the vertical length is $\Delta\theta$. Therefore,

$$\tan\alpha = \frac{\Delta\phi\cos\theta}{\Delta\theta}. \tag{5}$$

However, since we are mapping the sphere to a $2D$ plane, this rectangle becomes Δu by Δv. So,

$$\tan\alpha = \frac{\Delta u}{\Delta v} = \frac{\Delta\phi\cos\theta}{\Delta\theta}. \tag{6}$$

Thus, substituting $\Delta\phi$ for Δu we get

$$\begin{aligned}
\frac{\Delta\phi}{\Delta v} &= \frac{\Delta\phi\cos\theta}{\Delta\theta} \\
\frac{1}{\Delta v} &= \frac{\cos\theta}{\Delta\theta} \\
\Delta v &= \Delta\theta\sec\theta.
\end{aligned}$$

Letting $\Delta\theta$ tend to zero we have:

$$\frac{dv}{d\theta} = \sec\theta. \tag{7}$$

Thus, integrating we determine a function for v where

$$v(\theta) = \ln(\tan\theta + \sec\theta) + c.$$

Since $v(0) = 0$, $c = 0$ and the map

$$\begin{cases} u(\phi,\theta) = \phi, \\ v(\phi,\theta) = \ln(\tan\theta + \sec\theta) \end{cases} \tag{8}$$

satisfies all three conditions.

This means that as we travel further north on Mercator's map, the distance between the lines of latitude increases also. For example, Greenland on the map appears to be the same size as Africa when in fact in reality it is much smaller than Africa.

2.4 Stretching Factor

Equation (8) accounts for the fact that latitudes are parallel circles [1] shrinking in radius away from the equator. This means that the circumference of each circular line of latitude is smaller the closer you get to the north pole. This can be seen in Figure 9.

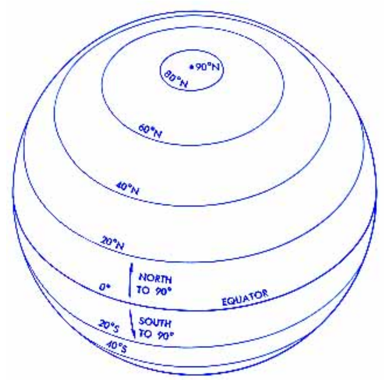


Figure 9: Lines of latitude circles on the earth [12].

Alexander [1] defines $\sigma(L)$ as the stretching factor at latitude L, and constructs the equation:

$$\sigma(L) = \frac{(1-e^2)\sec L}{1-e^2\sin^2 L}, \tag{9}$$

where eccentricity $e \approx 0.081$ and accounts for the amount the earth is deviant from a perfect sphere [1]. We first make sense of this by examining the representations of the lines of latitudes on Earth as Mercator first constructed it, a perfect sphere. Figure 10 displays the difference between the radius at the equator and the radius of a line of latitude nearer the North Pole when the earth is considered a perfect sphere [8].

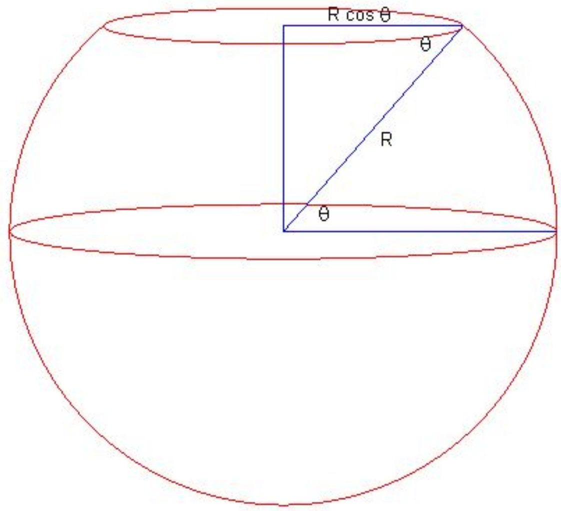


Figure 10: Radius of parallels shrink by $\cos\theta$ further north.

Assume the earth has radius R. On Mercator's map all the parallels are the same length so we need to adjust for this change in length by using the stretching factor. When the earth is a perfect sphere, $e = 0$ and

$$\sigma(L) = \frac{1(\sec L)}{1} = \sec L.$$

This is consistent with how we derived Equation (8). Suppose we want to find the legend length of 1 inch on the equator versus 1 inch in Walla Walla on spherical Earth. Walla Walla is at $46.0650°N$ [15]. Thus, in radians, $L = 0.804$ radians. So, if 1 inch = 1 mile in Walla Walla then 1 inch on the equator is

$$\sigma(L) = \sec^{-1}(0.804) = 1.441\,\text{miles}.$$

This accounts for why the north and south parts on Mercator's map are magnified.

Furthermore, we can use Mercator's stretching factor to describe rhumb lines on Mercator's Map. As we have shown already, a straight line on the map is a rhumb line and this the rhumb line is:

$$\Sigma_2 = m(\phi_2 - \phi_1) + c, \tag{10}$$

where $\Sigma(L) = \int_0^L \sigma(L)\,dl$ and

$$\sigma(L) = \frac{(1-e^2)\sec L}{1 - e^2\sin^2 L}.$$

To find the rhumb line bearing from one location to another, we first derive m in terms of latitude and longitude,

$$m = \frac{\Sigma_2 - \Sigma_1}{\phi_2 - \phi_1},$$

where $c = \Sigma_1$. Thus, the bearing from location 1 to location 2 is

$$\alpha = \cot^{-1}\left(\frac{\Sigma_2 - \Sigma_1}{\phi_2 - \phi_1}\right). \tag{11}$$

Note as $L \to \pi/2$, $\Sigma \to \infty$ and the parallels of latitude spread further and further apart.

2.5 Mapping Ellipsoidal Earth

Now that we have derived Mercator's map assuming the earth is a sphere, we can construct Mercator's projection assuming the earth is its actual shape: an oblate spheroid. We know that $u(\theta, \phi)$ stays the same but need to find $v(\theta, \phi)$. To derive the equation, we implement a different method. We describe the path on an ellipsoid by making ϕ a function of θ, so $\phi = \phi(\theta)$ is the path [13]. Remember that we parametrised the surface of the ellipsoid as $(a \cos\phi(\theta)\cos\theta, a \sin\phi(\theta)\cos\theta, b \sin\theta)$. Thus, we have a tangent vector $\vec{t}$ such that:

$$\vec{t} = \left(\frac{dx}{d\theta}, \frac{dy}{d\theta}, \frac{dz}{d\theta}\right) = \vec{n} + \vec{v}\frac{d\phi}{d\theta}, \tag{12}$$

where using the multivariable chain rule shows us that

$$\vec{n} = \left(\frac{\partial x}{\partial\theta}, \frac{\partial y}{\partial\theta}, \frac{\partial z}{\partial\theta}\right), \tag{13}$$

and

$$\vec{v} = \left(\frac{\partial x}{\partial\phi}, \frac{\partial y}{\partial\phi}, \frac{\partial z}{\partial\phi}\right). \tag{14}$$

Using the product rule we find that

$$\vec{n} = (-a\cos\phi(\theta)\sin\theta, -a\sin\phi(\theta)\sin\theta, b\cos\theta)$$

and

$$\vec{v} = (-a\sin\phi(\theta)\cos\theta, a\cos\phi(\theta)\cos\theta, 0).$$

We used $\phi'(\theta)$ from differentiation with respect to θ to get $\vec{v}$. Further, since a meridian is a curve where θ is constant, $\vec{n}$ points due north. Similarly, $\vec{v}$ points due east and so $\vec{n}$ and $\vec{v}$ are perpendicular, their dot product is 0 and they are orthogonal. When $\phi = 0$, $\vec{n} = (-a\sin\theta, 0, b\cos\theta)$ and

$$||\vec{n}||^2 = a^2\sin^2\theta + b^2\cos^2\theta. \tag{15}$$

Remembering that

$$e^2 = \frac{a^2 - b^2}{a^2},$$

we can write

$$\begin{aligned} ||\vec{n}||^2 &= a^2(1 - \cos^2\theta) + b^2\cos^2\theta \\ &= a^2 - a^2\cos^2\theta + b^2\cos^2\theta \\ &= a^2 - \frac{a^2(a^2 - b^2)}{a^2}\cos^2\theta \\ &= a^2 - a^2(e^2)\cos^2\theta \\ &= a^2(1 - e^2\cos^2\theta). \end{aligned} \tag{16}$$

Similarly when $\phi = 0$,

$$||\vec{v}||^2 = a^2\cos^2\theta. \tag{17}$$

We also know

$$||\vec{t}||^2 = ||\vec{n}||^2 + ||\vec{v}||^2(\phi)'^2. \tag{18}$$

In mapping the earth onto 2D plane using Mercator's map conditions, we want to find a path of constant bearing (rhumb line), so the path for which the angle between $\vec{t}$ and $\vec{n}$ is constant [13]. Let us call this angle α. Such an angle is determined by

$$\vec{n}\cdot\vec{t} = ||\vec{n}||||\vec{t}||\cos\alpha.$$

To simplify, observe from the orthogonality of $\vec{v}$ and $\vec{n}$ that

$$\begin{aligned} \vec{n}\cdot\vec{t} &= (\vec{n} + \vec{v}\phi')\cdot\vec{n} \\ &= \vec{n}\cdot\vec{n} + \vec{v}\cdot\vec{n}\phi' \\ &= ||\vec{n}||^2. \end{aligned}$$

So,

$$||\vec{n}|| = ||\vec{t}||\cos\alpha. \tag{19}$$

We can use substitution and rearrange the equations 16 and 17 to get a formula for ϕ' as follows:

$$\begin{aligned}
||\vec{v}||^2(\phi')^2 &= ||\vec{t}||^2 - ||\vec{n}||^2 \\
&= \frac{||\vec{n}||^2}{\cos^2\alpha} - ||\vec{n}||^2 \\
&= ||\vec{n}||^2\Big[\frac{1}{\cos^2\alpha} - 1\Big] \\
(\phi')^2 &= \frac{||\vec{n}||^2}{||\vec{v}||^2}(\sec^2\alpha - 1) \\
&= \frac{||\vec{n}||^2}{||\vec{v}||^2}\tan^2\alpha \\
\phi' &= \frac{||\vec{n}||}{||\vec{v}||}\tan\alpha. \qquad (20)
\end{aligned}$$

We can take the square root because ϕ and θ both increase or both decrease, implying α is either in the range $(0, \pi/2)$ or $(-\pi/2, -\pi)$. Therefore, using the equations for $||\vec{n}||$ and $||\vec{t}||$, and since $\cos\theta \geq 0$ for $-\pi/2 \leq \theta \leq \pi/2$, we get

$$\phi' = (\tan\alpha)\frac{a\sqrt{1 - e^2\cos^2\theta}}{a\cos\theta} = (\tan\alpha)\sqrt{\sec^2\theta - e^2}. \qquad (21)$$

Integrating with respect to θ we get

$$\phi = \tan\alpha \int \sqrt{\sec^2\theta - e^2}\, d\theta + c.$$

When $e = 0$ and the earth is spherical it simplifies to

$$\phi = \tan\alpha \ln(\sec\theta + \tan\theta) + c.$$

We can use this result to determine the bearing of one point to another, given the latitude and longitude of those two points. To see this consider Figure 4 where the path from London to Seattle is mapped. London lies on the prime meridian and so its coordinates are $L(0, 0.899)$. Seattle is slightly southwest of London, which you can see in its coordinates $S(2.1329, 0.831)$. Now we determine the angle α, which is the bearing. It follows,

$$\begin{aligned}
\phi_2 - \phi_1 &= \tan\alpha \int_{\theta_1}^{\theta_2} \sqrt{\sec^2\theta - e^2}d\theta \\
2.1329 &= \tan\alpha \int_{0.899}^{0.831} \sqrt{\sec^2\theta - e^2}d\theta \\
2.1329 &= \tan\alpha[-0.10491] \\
\alpha &= -1.52165.
\end{aligned}$$

This verifies we are travelling in a southwesterly direction.

To compute the length D of the line, we assume $\theta_2 > \theta_1$ and evaluate the integral $\int_{\theta_1}^{\theta_2} \sqrt{||\vec{t}||}\, d\theta$. Therefore when $e = 0$:

$$
\begin{aligned}
D &= a|\sec\alpha| \int_{\theta_1}^{\theta_2} \sqrt{1 - e^2\cos^2\theta}\, d\theta \\
&= a|\sec(-1.52165)| \int_{0.831}^{0.899} \sqrt{1}\, d\theta \\
&= a(20.356)[0.068] \\
&= 1.3842.
\end{aligned}
$$

Note, we are considering the unit sphere here, so $a = 1$. If we use the radius of the earth, 3963 miles, we find that this distance is 5485.6 miles.

When $e > 0$, the integration becomes much harder to calculate. To rectify this problem we can estimate the integral by using Maple to calculate a numeric table of varying e.

e	$\lvert\alpha\rvert$	D	Distance (miles)
0	1.522	1.3842	5485.6
0.081	**1.5217**	**1.3841**	**5485.4**
0.1	1.5218	1.3753	5450.5
0.2	1.52206	1.3692	5426.2
0.3	1.52259	1.3627	5400.4
0.4	1.52333	1.3556	5372.2
0.5	1.5243	1.3476	5340.6
0.6	1.5255	1.3384	5304.1
0.7	1.527	1.3274	5260.67
0.8	1.5288	1.3139	5207

Table 1: Table displaying rhumb line distance for varying eccentricity; eccentricity of Earth bolded.

3 Distances Between Points On Earth

3.1 Rhumb Line distance

We can use Equation (8) to calculate rhumb line distance between two points on the earth assuming the earth is a sphere. Suppose we are travelling from London to Seattle, as shown in Figure 4. As before, let

$$v = \ln(\tan\theta + \sec\theta), \tag{22}$$

where θ represents the latitude. Therefore, plugging θ_1 and θ_2 into Equation (22), we get $v_1 = 1.052$ and $v_2 = 0.9473$ respectively measuring the height on the map corresponding to latitude. Therefore, now we have our points on the map $L(v = 1.052, \phi = 0)$ and $S(v = 0.9473, \phi = 2.1329)$.

Note that $\sec\theta + \tan\theta = \tan\left(\frac{\pi}{4} + \frac{x}{2}\right)$ [5] so,

$$\theta = 2\arctan e^v - \frac{\pi}{2} = G(v).$$

On Mercator's map we can parametrise the straight line between these two points as a function of t [2] to get the vector $< v_1 - (v_2 - v_1)t, \phi t >$ where $0 \leq t \leq 1$. Hence, in our case, we get the vector t such that $\left\langle 1.052 - 0.1047t, 2.1329t \right\rangle$, $0 \leq t \leq 1$. Thus, the rhumb line on a sphere is

$$\theta = G(1.052 - 0.1047t),$$

$$\phi = 2.1329t\,[2].$$

To calculate the distance we recall the two-dimensional arc formula is:

$$\int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dt\,[7].$$

We derive the spherical coordinate arc length formula from this equation, with radius one and get the length of the rhumb line is:

$$D = \int_0^1 \sqrt{\left(\frac{d\theta}{dt}\right)^2 + \cos^2(\theta) d\phi^2}\, dt\,[2]. \tag{23}$$

Since $G(v)$ is the inverse of $\ln(\sec\theta + \tan\theta)$ with derivative $\sec\theta$,

$$G'(v) = \frac{1}{\sec(G(v))} = \cos(G(v)).$$

Thus we have,

$$\begin{aligned}
D &= \int_0^1 \sqrt{(-0.1047)^2\cos^2(G(1.052-0.1047t)) + 2.1329^2\cos^2(G(1.052-0.1047t))}\,dt \\
&= \int_0^1 2.1355\cos(G(1.052-0.1047t))\,dt \\
&= \int_0^1 2.1355\sin(2\arctan e^{(1.052-0.105t)})\,dt \\
&= 4.2709\int_0^1 \frac{e^{(1.052-0.105t)}}{1+e^{2(1.052-0.105t)}}\,dt.
\end{aligned}$$

Using u-substitution to evaluate this integral, we find $D = 1.384$, the same value we found in Section 2.5. We can convert this length to miles and find that the rhumb line distance is 5485.58 miles between London and Seattle on spherical Earth. Next we compare this to the great circle distance.

3.2 Great Circle Arcs

Recall a circle on a sphere is called a great circle if it is the shortest path between two points. Comparatively, the shortest path between two points on a spheroid is called a *geodesic*. However, this path is much harder to calculate so in this paper we will instead solve only great circle distances. We find the great circle path by examining the intersection of the sphere and a plane containing the centre of the sphere [6]. Therefore in order to obtain the great circle we first need to find the plane determined by these two points and the centre of the sphere. The shorter of the two great circle arcs is the shortest path between the two points. Reflecting back to the figure of the earth, meridians are great circles and the only line of latitude that is a great circle is the equator. This makes sense since the equator is itself a plane that passes through the centre of a sphere.

In 2D, the shortest path between two points is the straight line between them. Similarly on the sphere, the shortest path is the "straighter" path. The larger a circle's radius, the less curvature the circle has and the straighter it is. Therefore, the largest circles on the earth are the great circles. This can be seen in Figure 11 where the great circle arc is the larger of the arcs.

However, before calculating the great circle distance we will prove that the great circle arc is indeed the shortest distance between two points. We can use vector calculus to do this. First, assume we are examining the unit sphere where the centre of the sphere is the origin and one of our points is on the North Pole. Define the other point as θ_1 latitude (measured from the equator) and ϕ_1 longitude (measured from the prime meridian). The great circle arc will travel from the north pole along the meridian to where the other point lies, specifically longitude ϕ_1. Therefore the angle the path will travel along is $\pi/2-\theta_1$ since the latitude of the north pole is $\pi/2$. Thus the length of the arc will also be $\pi/2-\theta_1$. We must now show an arbitrary path between any two points is greater than or equal $\pi/2-\theta_1$.

Let us parametrise our path as functions of t. So, define $\theta(t)$ as latitude at time t and define $\phi(t)$ as longitude at time t. Hence, our position vector is

$$\mathbf{r}(t) = \big\langle \cos\theta(t)\cos\phi(t), \cos\theta(t)\sin\phi(t), \sin\theta(t)\big\rangle.$$

We differentiate this to find the velocity at time t:

$$\begin{aligned}\mathbf{r}'(t) = \Bigg\langle & -\sin\theta(t)\frac{d\theta}{dt}\cos\phi(t) - \cos\theta(t)\sin\phi(t)\frac{d\phi}{dt}, \\ & -\sin\theta(t)\frac{d\theta}{dt}\sin\phi(t) + \cos\theta(t)\cos\phi(t)\frac{d\phi}{dt}, \cos\theta(t)\frac{d\theta}{dt}\Bigg\rangle. \end{aligned} \tag{24}$$

Therefore, our speed at time t is the magnitude of the velocity vector:

$$\begin{aligned}||\mathbf{r}'(t)|| = \Bigg(& \cos^2(t)\left[\sin^2\phi(t)\left(\frac{d\phi}{dt}\right)^2 + \cos^2\phi(t)\left(\frac{d\phi}{dt}\right)^2 + \left(\frac{d\theta}{dt}\right)^2\right] \\ & + \left(\frac{d\theta}{dt}\right)^2\left[\sin^\theta(t)\cos^2\phi(t) + \sin^2\theta(t)\sin^2\phi(t)\right]\Bigg)^{\frac{1}{2}},\end{aligned}$$

$$\begin{aligned}||\mathbf{r}'(t)|| &= \left(\cos^2\theta(t)\left[\left(\frac{d\phi}{dt}\right)^2 + \left(\frac{d\theta}{dt}\right)^2\right] + \left(\frac{d\theta}{dt}\right)^2\left[\sin^2\theta(t)\right]\right)^{\frac{1}{2}} \\ &= \sqrt{\left(\frac{d\theta}{dt}\right)^2 + \cos^2\theta(t)\left[\frac{d\phi}{dt}\right]^2}.\end{aligned}$$

To find the distance we need to integrate this speed function. So assume we are travelling from time $t=0$ and end at time t_1. Thus, $\theta(0)=\pi/2$, $\theta(t_1)=\theta_1$ and $\phi(t_1)=\phi_1$. Therefore the distance is

$$D = \int_{t=0}^{t_1} ||\mathbf{r}'(t)||\, dt. \tag{25}$$

Observe,

$$||\mathbf{r}'(t)|| \geq \sqrt{\left(\frac{d\theta}{dt}\right)^2} = \left|\frac{d\theta}{dt}\right| \geq -\frac{d\theta}{dt}.$$

Therefore,

$$D \geq \int_{t=0}^{t_1} -\frac{d\theta}{dt}dt = \theta(0) - \theta(t_1) = \pi/2 - \theta_1.$$

We can conclude that the great circle path is the shortest possible path.

3.3 Calculating Great Circle Distance

Consider the points L and S:

$$L = (\cos\theta_1 \cos\phi_1, \cos\theta_1 \sin\phi_1, \sin\theta_1),$$

$$S = (\cos\theta_2 \cos\phi_2, \cos\theta_2 \sin\phi_2, \sin\theta_2).$$

These can be converted to vectors $\vec{OL}$ and $\vec{OS}$ by using the origin of the sphere, which is also the centre of the great circle connecting the two points. We want to find the angle α between them in radians and then calculate the distance. We implement the standard formula:

$$\cos(\alpha) = \frac{\vec{OL} \cdot \vec{OS}}{|\vec{OL}||\vec{OS}|}. \tag{26}$$

Since we are considering the unit sphere, it follows,

$$\begin{aligned}
\cos(\alpha) &= \cos\theta_1 \cos\phi_1 \cos\theta_2 \cos\phi_2 + \cos\theta_1 \sin\phi_1 \cos\theta_2 \sin\phi_2 + \sin\theta_1 \sin\theta_2 \\
&= \cos\theta_1 \cos\theta_2 (\cos\phi_1 \cos\phi_2 + \sin\phi_1 \sin\phi_2) + \sin\theta_1 \sin\theta_2 \\
&= \cos\theta_1 \cos\theta_2 (\cos\phi_1 - \phi_2) + \sin\theta_1 \sin\theta_2.
\end{aligned}$$

Therefore,

$$\alpha = \arccos(\cos\theta_1 \cos\theta_2 (\cos\phi_1 - \phi_2) + \sin\theta_1 \sin\theta_2). \tag{27}$$

On a sphere with radius one the distance between two points is the same as the angle between the vectors. On Earth with radius R [6], the distance will be given by

$$D = R\alpha.$$

Let us now examine our path between London and Seattle. Recall London is on the prime meridian with coordinates $(0.8988, 0)$ and Seattle is $(0.83, 2.1329)$. It follows,

$$\alpha = \arccos[\cos(0.8988)\cos(0.83)\cos(-2.1329) + \sin(0.8988)\sin(0.83)] = 1.209\,\text{radians}.$$

With radius of Earth $R = 3963$ miles, the distance is

$$\begin{aligned}
D &= (1.209)(3963) \\
&= 4791.27\,\text{miles}.
\end{aligned}$$

This is indeed shorter than the rhumb line distance between the two points. The question that follows is how much do rhumb line paths and great circle paths differ and how do they vary as latitude and longitude change?

3.4 Comparing Rhumb Line and Great Circle

As we have seen, London and Seattle are very similar in latitude, but let us explore what happens when two points lie on exactly the same parallel. We compare rhumb lines to great circles for fixed longitude but varying latitude.

3.5 On The Equator

Let's examine two points arbitrarily $\pi/2$ longitudinal distance away from each other on the equator. Calculating the rhumb line distance and great circle distance in this case is very simple since the equator is a great circle, and the equator maps identically to Mercator's map. All we have to do is multiply the angle of longitudinal difference by R where R is equatorial radius, 3963 miles. Thus, the distance is

$$D_{gc} = 3963(\pi/2) = 6103 \text{ miles.}$$

What about two points on the same parallel that is not the equator?

3.6 Increasing Latitude Pairs, Constant Longitude

Let us examine two points on the same parallel but both increasing in latitude away from the equator. To do this we use the equations derived already for rhumb line and great circle distances.

Great Circle: Our equations are:

$$\alpha = \arccos(\cos\theta_1 \cos\theta_2(\cos\phi_1 - \phi_2) + \sin\theta_1 \sin\theta_2)$$

and

$$D_{gc} = R\alpha,$$

with θ latitude and ϕ longitude. Therefore, if we examine two points on latitude $\pi/12$, with $\Delta\phi = \pi/2$, plugging in our values into the equations we find

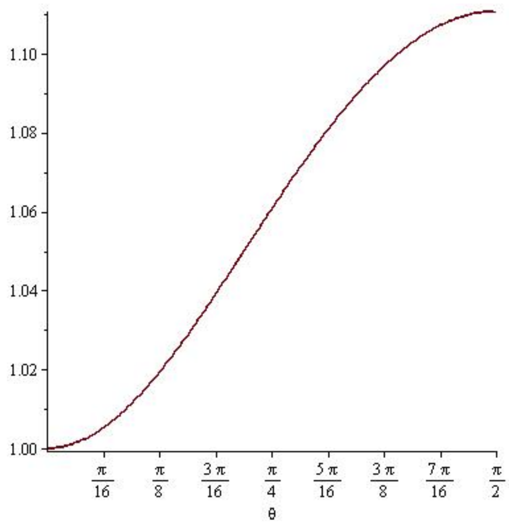
$$\begin{aligned} D_{gc} &= 3963(1.475) \\ &= 5845.2 \text{ miles.} \end{aligned} \tag{28}$$

Rhumb Line: The rhumb line distance is simply the distance along the parallel. To calculate, we multiply the longitudinal difference by the equatorial radius and then multiply by a scale factor of $\cos(\theta)$. This scale factor we found when constructing Mercator's map and from Equation (9). Thus,

$$\begin{aligned} D_{rl} &= (\Delta\phi)\cos(\theta)R \\ &= (\pi/2)(\cos(\pi/12))(3963) = 5895.1 \text{ miles.} \end{aligned} \tag{29}$$

The great circle distance is shorter than the rhumb line distance, but only by a small amount.

We can continue to do this calculation for increasing latitude. To make sense of our data, we calculate the proportional difference between rhumb line and great circle path lengths by dividing rhumb line distance by the great circle distance. For example, when we measure D_{rl} and D_{gc} between $(\pi/12, 0)$ and $(\pi/12, \pi/2)$, the rhumb line difference is 1.0085 times larger than the great circle difference. Figure 11 plots latitude on the x-axis and D_{rl}/D_{gc} on the y-axis.



(a) Proportional difference versus latitude.

Point (θ)	$\frac{D_{rl}}{D_{gc}}$
0	1
$\frac{\pi}{12}$	1.0085
$\frac{\pi}{6}$	1.0305
$\frac{\pi}{4}$	1.0579
$\frac{\pi}{3}$	1.08295
$\frac{5\pi}{12}$	1.10001

(b) Table of values from graph.

Figure 11: Proportional difference between rhumb lines and great circles versus increasing latitude.

Note that when $\theta = \pi/2$, we cannot compute both great circle and rhumb line distances as we are left with 0/0. Therefore, we can compute the limit of the ratio as θ approaches $\pi/2$ or more specifically:

$$\lim_{\theta \to \frac{\pi}{2}} \frac{\Delta\phi(\cos\theta)}{\cos^{-1}(\cos\Delta\phi\cos^2\theta + \sin^2\theta)}. \tag{30}$$

L'Hopital's rule does not work with this limit because no matter how many times you differentiate, the answer is still 0/0. Therefore, we can use Maple to approximate the limit. Firstly, when $\Delta\phi = \frac{\pi}{2}$, the limit changes to:

$$\lim_{\theta \to \frac{\pi}{2}} \frac{1}{2} \frac{\pi(\cos\theta)}{\cos^{-1}(\sin^2\theta)} = 1.1107.$$

The next question to ask is what does this number mean? Is there a mathematical significance to this number? One can plot the evaluated limit at different longitudinal differences to see the relationship. The following graph demonstrates the limit as θ approaches $\pi/2$ with longitudinal difference ranging from 0 to π.

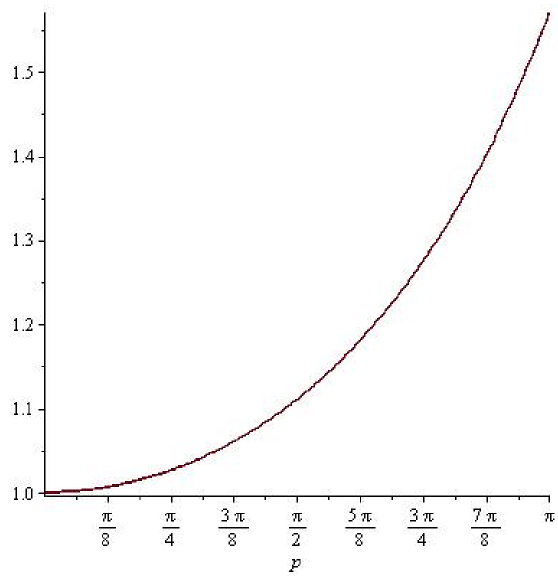


Figure 12: Limit as $\theta \to \pi/2$ in Equation 30 versus $\Delta\phi$.

This relationship makes sense: the further apart two points are from each other longitudinally, the larger the proportional difference between rhumb line and great circle paths are. Furthermore, this finding is similar to the finding in Figure 11 where the increase in latitude caused an increase in proportional difference too. If we change the limit of θ to a smaller value of $\pi/6$, we get the graph in Figure 13.

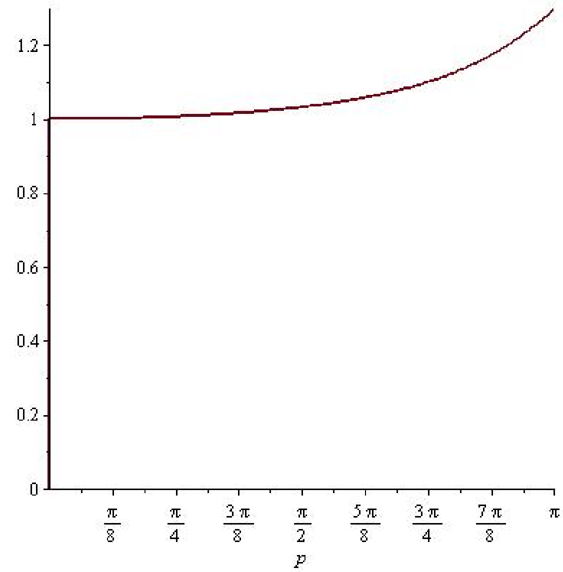


Figure 13: Limit as $\theta \to \pi/6$ in Equation (30) versus $\Delta\phi$.

As we would expect, the shallower increase in the proportion relates to the smaller change in latitude.

3.7 Increasing Latitude, Constant Longitude, Single Point

Now, let's investigate what happens when we keep one point exactly the same ($P_1(\theta = \pi/4, \Delta\phi = \pi/2)$) and calculate rhumb line and great circle distances to points on different latitudes, keeping longitudinal difference the same.

Using the methods and formulas previously derived we find the distance between P_1 and the following points in Table 2.

Point (θ)	D_{rl}(miles)	D_{gc}(miles)	$\frac{D_{rl}}{D_{gc}}$
0	6257	6132	1.0205
$\frac{\pi}{12}$	5577	5405	1.0319
$\frac{\pi}{6}$	4916	4708	1.0443
$\frac{\pi}{4}$	4310	4074	1.058
$\frac{\pi}{3}$	3808	3554	1.0713
$\frac{5\pi}{12}$	3472	3211	1.0811

Table 2: Table displaying distance between P_1 and points of varying latitude.

Figures 14 to 19 show the rhumb line and great circle paths between some of these values on the Mercator Projection [10].

Figure 14: Great circle distance between P_1 and point on equator.

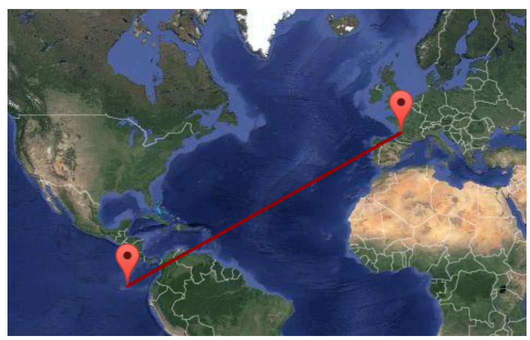


Figure 15: Rhumb line distance.

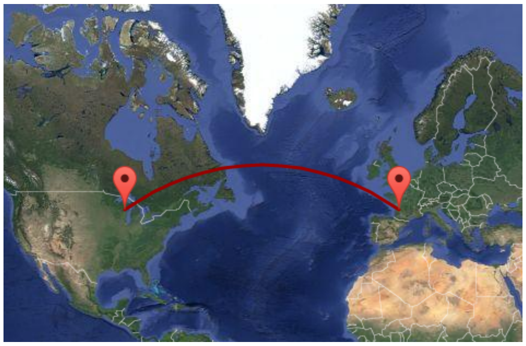


Figure 16: Great circle distance between P_1 and point on same parallel.

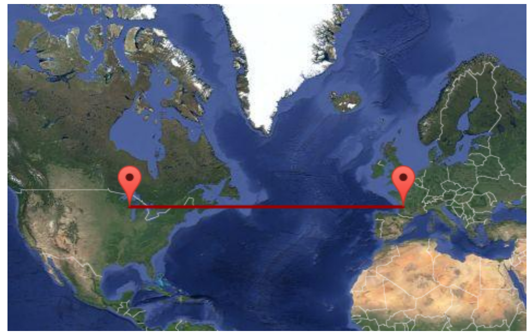


Figure 17: Rhumb line distance.

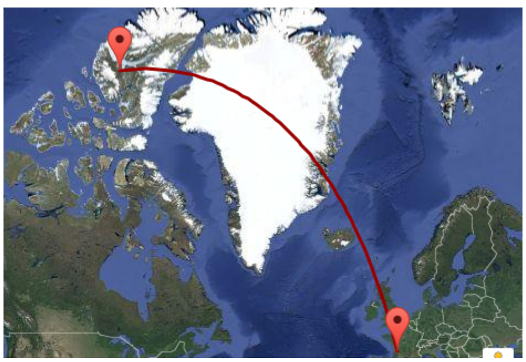


Figure 18: Great circle distance between P_1 and point 80° N.

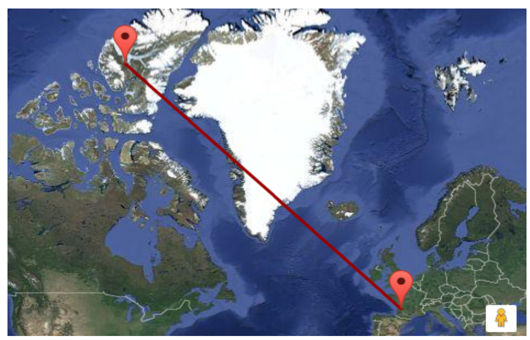


Figure 19: Rhumb line distance.

The great circle appears curved and seems to vary greatly from the rhumb line when the distances are not that deviant from each other. We also see that the proportional difference is small but increases as latitude increases. This goes against the conjecture in Alexander's paper [1], which stated the proportional difference is greatest when the two points are on the same latitude. This implies longitudinal difference may have an effect since we are using $\Delta\phi \approx \pi/2$ longitudinal difference and Alexander used π longitudinal difference. Another issue may be that meridians get closer together as they approach the North Pole, and so computing great circle distance keeping $\Delta\phi$ constant may not actually be accurate. However, with the problem we are examining, the next question we must ask is: is there a maximum proportional difference as latitude between two points increases? To answer this question we can look at smaller increasing increments of latitude, as shown in Table 3.

Point (θ)	D_{rl}(miles)	D_{gc}(miles)	$\frac{D_{rl}}{D_{gc}}$
$\frac{5\pi}{12}$	3472	3211	1.08107
1.396	3403	3148	1.08113
1.48	3349	3114	1.0756

Table 3: Table displaying distance between P_1 and points of varying latitude.

We find that increasing the latitude by 5°, after 80° ≈ 1.396 latitude, the proportional difference decreases again. The maximum proportional difference is therefore between 80° and 85°. It is interesting to see that there is indeed a maximum point and we could look further to understand why this maximum occurs at this point.

3.8 Increasing Longitude, Constant Latitude

Now we will examine points that lie on the same latitude but the longitude changes. For example, let's examine the parallel $\theta = 0.3329$. Using the same process as above, the great circle distance from a point $\phi = \pi/6$ is

$$\begin{aligned} D_{gc} &= 3963(0.4942) \\ &= 1958.6\,\text{miles}. \end{aligned}$$

The rhumb line distance is

$$D_{rl} = (\pi/6)(3963)(\cos(0.3329)) = 1961.1 \text{ miles}.$$

Table 4 shows different longitudinal differences.

Point (ϕ)	D_{rl}(miles)	D_{gc}(miles)	$\frac{D_{rl}}{D_{gc}}$
$\frac{\pi}{6}$	1959	1961	1.0013
$\frac{\pi}{3}$	3901	3922	1.0053
$\frac{2\pi}{3}$	7599	7844	1.032

Table 4: Constant latitude but changing longitude.

The difference in length between the two paths increases more rapidly as longitudinal separation increases. This is not surprising given what we have discovered about great circle paths.

3.9 Same Longitude

Recall Definition 2.2. that states a meridian is a great circle. Therefore, calculating distance between two points on the same longitude only requires us to calculate the great circle distance between them. Let us consider two points on the same meridian that are $\theta = \frac{\pi}{12}$ radians apart. This time we do not have to calculate α because we already have the change in direction, namely $\frac{\pi}{12}$. Therefore, we need to calculate $(\frac{\pi}{12})(R)$ but instead use the polar radius of the earth (the semi-minor axis) and so we get

$$D_{gc} = 3950\left(\frac{\pi}{12}\right) = 1034 \text{ miles}.$$

Since the form of this calculation is $y = mx + c$ where $y = D$, $m = 3950$, $\theta = x$ and $c = 0$, the distance between two points on the same longitude but with varying latitude will change linearly between latitudes. Figure 20 demonstrates this.

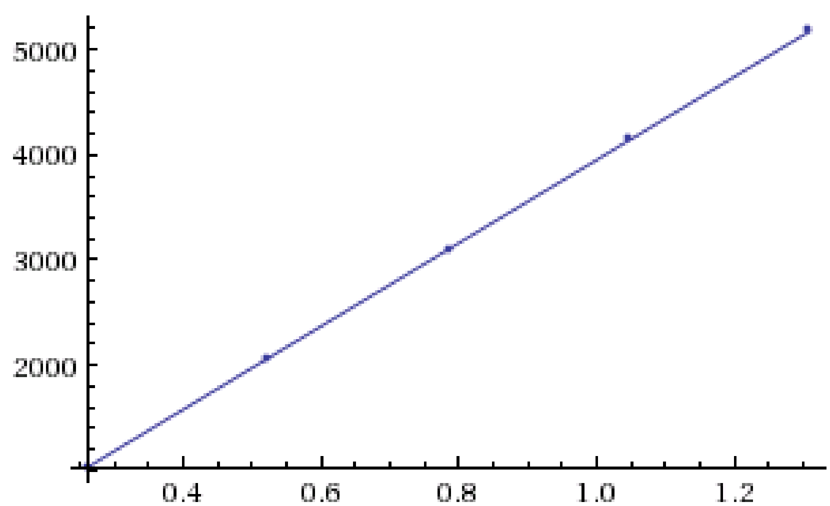


Figure 20: The linear change in distance (miles) on the y-axis as θ increases on the x-axis.

4 Conclusion

This project provides a brief comparison of rhumb lines and great circles. It overviews the mathematics behind the construction of Mercator's map both for the spherical case and the ellipsoidal case. It begins to analyse the properties behind computing different path lengths on spherical Earth and how those paths are related mathematically. Unfortunately, the greatest limitation of this paper and of many papers examining the mapping of the earth is that we assume the earth is spherical when in fact it is not. Future work should apply methods derived here to spheroidal Earth, to improve the accuracy of calculations as well as investigate further the properties of the stretching factor displayed in Section 2.4. Furthermore, future senior projects could investigate other map projections than the Mercator projection considered in this paper.

References

[1] Alexander, James. *"Loxodromes: A rhumb way to go."* Mathematics magazine 77, no. 5 (2004): 349-356.

[2] http://www.ams.org/samplings/feature-column/fcarc-navigation5

[3] Astro Navigation Demystified. https://astronavigationdemystified.com/the-rhumb-line/

[4] Chatfield, Averil Burton Chatfield. *"Fundamentals of high accuracy inertial navigation. Vol. 174."* Aiaa, 1997.

[5] Daners, Daniel. *"The Mercator and stereographic projections, and many in between."* The American Mathematical Monthly 119, no. 3 (2012): 199-210.

[6] Feeman, Timothy G. *Portraits of the earth: A mathematician looks at maps.* Vol. 18. American Mathematical Soc., 2002.

[7] Guichard, David *"Community Calculus"* (2013).

[8] Israel, Robert, Math UBC, http://www.math.ubc.ca/ israel/m103/mercator/mercator.html

[9] Leitão, Henrique, and Joaquim Alves Gaspar. *"Globes, Rhumb Tables, and the Pre-History of the Mercator Projection."* Imago Mundi 66, no. 2 (2014): 180-195.

[10] Movable Type Scripts, http://www.movable-type.co.uk/scripts/latlong.html

[11] *"Oxford Dictionary":* https://www.oxforddictionaries.com/definition/english/geodesic

[12] http://web.csulb.edu/ rodrigue/geog140/parallel.jpg

[13] Schechter, Murray. *"Which Way Is Jerusalem? Navigating on a Spheroid."* The College Mathematics Journal 38, no. 2 (2007): 96-105.

[14] Torge, Wolfgang, and Jürgen Müller. *"Geodesy."* Walter de Gruyter, 2012.

[15] Travel Math, http://www.travelmath.com/cities/Walla+Walla,+WA

[16] Wikipedia. https://en.wikipedia.org/wiki/Rhumbline

[17] Wolfram Math World, http://mathworld.wolfram.com/OblateSpheroid.html

[18] Zhang, Keke, Kit H. Chan, and Xinhao Liao. *"On precessing flow in an oblate spheroid of arbitrary eccentricity."* Journal of Fluid Mechanics 743 (2014): 358-384.

Alphabetical Index