Paths Between Points on Earth: Great Circles, Geodesics, and Useful Projections

James R. Clynch
February 2006

I. Historical and Common Navigation Methods

There are an infinite number of paths between two points on the earth. For navigation purposes there are only a few historically common choices.

1. **Sailing a latitude**.

Since ancient times, latitude has been measured using observations of stars. Longitude could only be estimated by dead reckoning when far at sea until the 18^{th} century. Prior to a good method of finding longitude at sea, you sailed north or south until you came to the latitude of your destination then you sailed due east or west until you reached your goal. Of course if you choose the wrong direction, you can get in big trouble. (This happened in 1707 to a British fleet that sunk when it hit rocks. This started the major British effort for a method to find longitude at sea.)

In this method the route follows the legs of a right triangle on the sphere. Just as in right triangles on the plane, the hypotenuse is always shorter than the sum of the other two sides. This method is very inefficient, but it was all that worked for open ocean sailing before 1770.

2. Rhumb Line

A **rhumb line** is a line of constant bearing or azimuth. This is not the shortest distance on a long trip, but it is very easy to follow. You just have to know the correct azimuth to use to get between two sites, such as New York and London. You can get this from a straight line on a Mercator projection. This is the reason the Mercator projection was invented. On other projections straight lines are not rhumb lines.

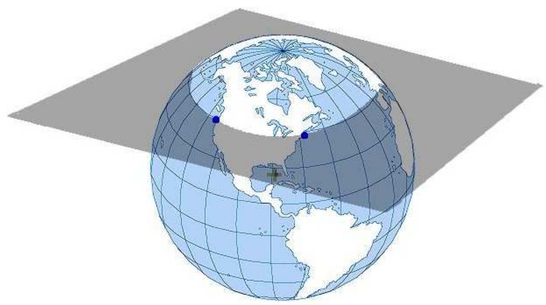
3. Great Circle

On a spherical earth, a **great circle** is the shortest distance between two locations. A great circle is a straight line on a Gnomonic projection. Gnomonic's don't look nice, but this is their useful property. In the past the latitude and longitude of evenly spaced points on a Gnomonic chart straight line were transferred to a Mercator. This formed a path of a set of straight segments of constant bearing with small turns between the segments.

II. Great Circles

There are several ways to define a **great circle**. One of the most useful in understanding its properties is to look at the intersection of a plane and a sphere. This will always be a circle, but usually not a great circle.

As an example, consider the paths between Portland Oregon and Portland Maine. Both are at about 45 degrees North. Look at the plane that is perpendicular to the z-axis and intersects the earth at 45 N.

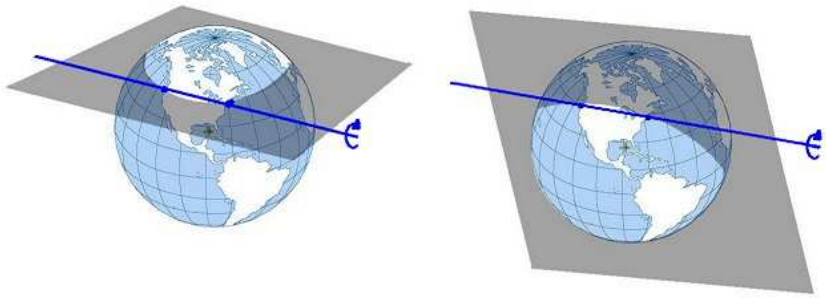


Plane and Earth

Constant Latitude Plane of a Parallel

The approximate positions of the two Portlands are shown as blue dots. The circle that the plane makes is called a **small circle**. It passes through both cities, but is not the shortest path. It is at the same latitude everywhere. (Because in this example we assume both cities are on the same latitude, this is also a rhumb line at azimuth of 90 degrees.)

One definition of <u>a great circle is the intersection of a plane and a sphere where the origin of the sphere is on the plane.</u> Thus you can tilt the plane about the line that goes through the two cities until the plane passes through the center of the earth. This procedure works to connect any two points on a spherical earth with a great circle. The correct plane is unique except for two points directly opposite each other as seen from the earth center. These are called antipodal points. The North and South Poles are one example of an antipodal pair of points.

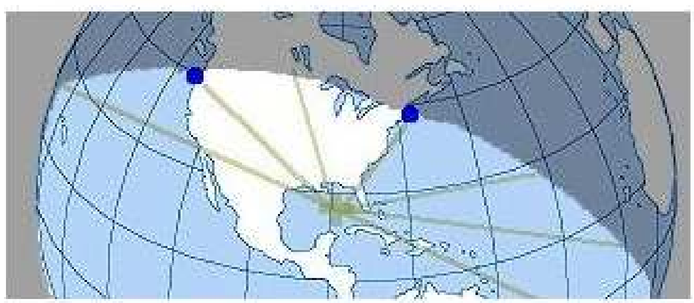


Making a Great Circle

Pass Plane Through End Points
Rotate Plane About Line Through Ends
Stop When Plane Include Center of Earth

The resulting intersection from this procedure is now a great circle. (Mathematically you can take the three points of the end points and the earth center and define this plane.)

The great circle is the shortest distance between the two points along the surface of the spherical earth. On the real earth this method produces a path adequate for navigation. It is not the absolute shortest path though.



Great Circle Path

Notice that the intersection between the plane and the earth goes considerably north (pole ward) of the constant latitude path. At the maximum the difference in the two paths is about 350 km. The path is shorter by only 100 km though. For longer paths, the distance reduction is larger.

The equations for a great circle distance are fairly straightforward if you know the end points.

If θ is the earth central angle of the arc between the two points, then

$$\cos\theta \quad = \quad \sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos(\lambda_1 - \lambda_2) .$$

This result can be obtained from spherical trigonometry. You can also obtain by considering the "up" vector at each point. The up direction at any point is given by the unit vector:

$$\hat{e}_u \quad = \quad (\cos\phi\cos\lambda, \cos\phi\sin\lambda, \sin\phi) .$$

The cosine of the arc between them, on a sphere, is the cosine of the angle between the up vectors. And this cosine is just the inner product of the two vectors!

$$\cos\theta = \hat{e}_{u1} \bullet \hat{e}_{u2} .$$

Using this method and a the formula for the cosine of the difference of two angles give the same result.

After finding θ (in radians) you just multiply by the radius of the earth to obtain the distance. The correct radius to use is discussed in detail below. In general R_N at the midpoint is best.

The azimuth, α_1, at point 1 is given from the equation:

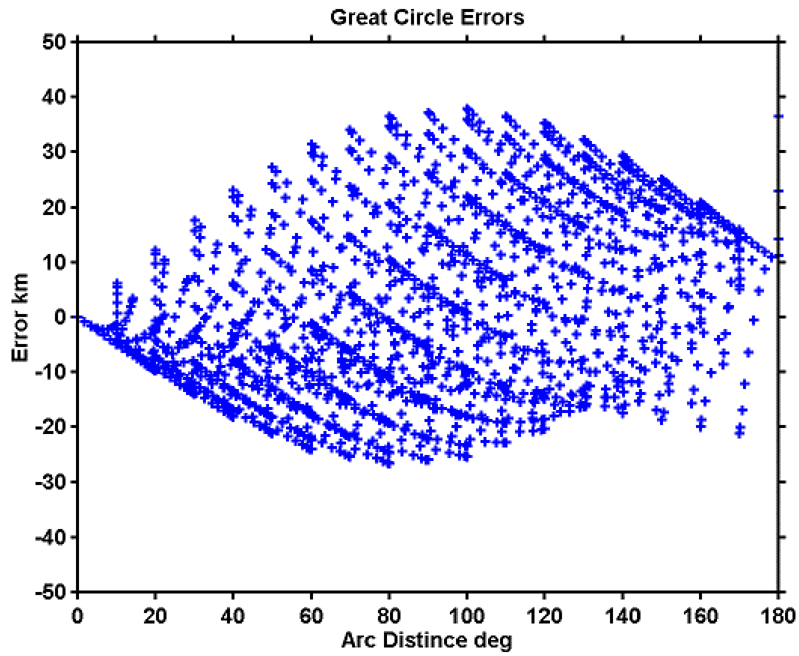
$$\begin{aligned} \cos\phi_2 \sin(\lambda_1 - \lambda_2) \quad &= \quad \sin\theta \sin\alpha_0 \\ \alpha_1 \quad &= \quad \begin{Bmatrix} \alpha_0, & \phi_1 \le \phi_2 \\ \pi - \alpha_0, & \phi_1 > \phi_2 \end{Bmatrix} \end{aligned} .$$

You have to take care in establishing the quadrant of the azimuth from the difference in latitude.

III. Geodesics

For an ellipsoidal earth, the situation is much more complex. In this case the shortest path has the name of a geodesic. The mathematics are complex and the result is almost the same as the great circle. Using some fixed radius results in an error up to 50 km in distance traveled for very long paths.

The figure below shows the error in using a good choice for a fixed radius to determine the true shortest distance from a from a great circle. The computations were made at steps of 10 deg in latitude and longitude between all point pairs. The radius used was $\sqrt[3]{a^2 b}$, which is the radius of a sphere with the same volume as the ellipsoid. The arc distance is the great circle earth central angle. The area covered by the points would be filled in if a smaller step size had been used.



Notice that there are some points at the right that seem to deviate from the pattern. These represent very long arcs where the geodesic takes a far different path from the great circle. For antipodal points there are an infinite number of great circles between the two points and the length is half the circumference of the earth. But for the ellipsoidal earth there are just two paths for antipodal points that have opposite directions. For near antipodal points there is only one geodesic and it often takes off at an angle far from the great circle direction.

IV Rhumb Lines - Constant Azimuth Paths

Officially the line of constant azimuth is called a loxodrome, but here we will just call it a rhumb line. The rhumb line is a straight line on a Mercator projection. Therefore the equations of that projection can be used to solve for the azimuth needed. The equations for the sphere follow. For an ellipsoidal earth they are much more complex. However the azimuth can be for a spherical earth can be found from the formula for the points on the Mercator projection. These will be close to the values for the ellipsoid. (The easy generation of rhumb lines as the reason the Mercator projection was invented. Mercator had a good business in producing maps for navigators in his new projection.)

The paper coordinates (x, y) of a point (ϕ, λ) in a Mercator projection of the sphere are given by:

$$
\begin{aligned}
x &= R\,(\lambda - \lambda_0) \\
y &= R \ln\left[\tan\left(\frac{\phi}{2} - \frac{\pi}{4}\right)\right] \\
&= \frac{R}{2} \ln\left[\frac{1+\sin\phi}{1-\sin\phi}\right]
\end{aligned}
$$

where the center of the map is at the longitude λ_0 and R is the scaled radius of the earth. The slope of a line will be the tangent of the rhumb line azimuth α. Therefore the rhumb line azimuth from point 1 to point 2 is given by:

$$
\begin{aligned}
\tan\alpha &= \frac{y_2 - y_1}{x_2 - x_1} \\
&= \frac{\left\{\ln\left[\tan\left(\frac{\phi_2}{2} - \frac{\pi}{4}\right)\right] - \ln\left[\tan\left(\frac{\phi_1}{2} - \frac{\pi}{4}\right)\right]\right\}}{(\lambda_2 - \lambda_1)}, \\
&= \frac{1}{2} \frac{\ln\left[\frac{(1+\sin\phi_2)(1-\sin\phi_1)}{(1-\sin\phi_2)(1+\sin\phi_2)}\right]}{(\lambda_2 - \lambda_1)}
\end{aligned}
$$

or the second form for y can also be used. Given the azimuth, the distance, s, is given by:

$$
s = R_e \left|\frac{\phi_2 - \phi_1}{\cos\alpha}\right| .
$$

Clearly the longitudes must be in radians here.