Computing Great Circles

Leon Tabak
Department of Computer Science
Cornell College
600 First Street West
Mount Vernon, IA 52314–1098
E-mail: l.tabak@ieee.org
URL: http://people.cornellcollege.edu/ltabak

Abstract

The shortest path between two points on the surface of the globe is a circular arc. It can take a traveler through points that are closer to a pole than either endpoint. Hence, a journey that begins and ends in temperate latitudes can take a traveler through the Arctic.

A curious tourist might only want to know how closely a flight will approach the North Pole. The pilot will want to know distance and the succession of compass headings to follow. On the ground, controllers who track the flight will want to know where the aircraft is at each moment. The problem of computing great circles appeals because it is simply stated and also offers enough variations to elicit creative responses.

This paper describes a case study of design and testing. Exercises invite students to critique a given program, adapt it, and enhance it. Instructions direct students to extend suites of automated tests. Examples show students how to elaborate text and diagrams that describe the given design as they add and edit features.

1 Great Circles

A sailor leaves Cape May, New Jersey bound for Lisbon, Portugal. The two cities lie at nearly the same latitude. The sailor steers always east and follows that line of latitude from home port to destination. Yet this simplest course is not the shortest course.

A person in Eau Claire, Wisconsin wants to face Mecca when praying. In which direction should this person face? Mecca is south and east of Eau Claire, but the shortest route from Eau Claire to Mecca begins with a northeasterly trajectory.

The shortest path between two points on the surface of the earth can take a traveler through points that are closer to a pole than either endpoint. Hence, a journey that begins and ends in temperate latitudes might take a traveler through the Arctic. In the general case, the compass heading will vary continuously along the shortest course.

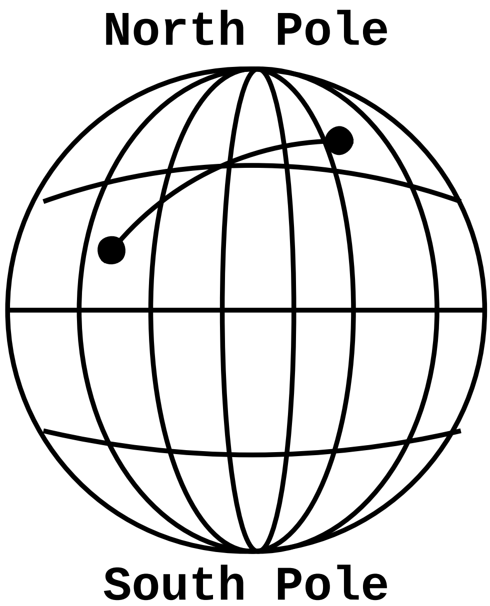

Figure 1: A great circle route.

Let us suppose that the starting point of a journey and the destination are two points on the surface of the earth. Let us suppose that a traveler is bound to remain on the surface of the earth. Finally, let us approximate the shape of the earth with a sphere.

Then the journey's starting point, its destination, and the center of the earth define a plane if the three points are not colinear. The intersection of that plane with the surface of the earth. defines a circle. The starting point and destination lie on the circle. They divide the circle into two arcs. The smaller arc is the shortest path over the earth's surface between the starting point and destination.

By choosing to fly, a traveler increases the length of a journey. To account for the ascent at the start of the journey and the descent at the end, the distance traveled then includes twice the altitude at which the airplane cruises. By traveling above the earth, and therefore farther from the earth's center, the air traveler also increases the radius of the great circle.

We will assume that travelers fly at such low altitudes that these adjustments do not add enough to the total distance traveled to be worth counting. Neither will we consider in the first draft of our program how the rotation of the earth beneath an aircraft in flight changes the problem.

1.1 Easy Cases

The equator is the only line of latitude that is a great circle. Other lines of latitude define circles whose center lies on the earth's axis but away from the earth's center. If a journey begins and ends on the equator, then the shortest course between the two points keeps the traveler on the equator. The navigator will either steer always east or always west.

Every line of longitude is one half of a great circle. Each is a half circle that connects the North Pole to the South Pole. If the starting point and destination of a journey lie between the poles on the same line of longitude, then the optimal route follows that line north or south. If the destination is the North Pole and the starting point is anywhere other than the South Pole, then a unique shortest path will keep the traveler headed north at every point along the way.

A pair of complementary lines of longitude together make a complete circle. Separation by 180 degrees makes two longitudes complementary. For example, `90 W` and `90 E` are complementary longitudes. The two lines of longitude meet at the North Pole and again at the South Pole. Longitudes `75 W` and `105 E` are also complementary because $75 + 105 = 180$.

If the starting point and destination of a journey lie on complementary longitudes, then the shortest path between the two points follows first one, then the other line of longitude. For example, if the journey's endpoints both lie in the northern hemisphere, the shortest route leads first north to the North Pole and then south to the destination.

1.2 Ambiguous Cases

Many equally short routes connect the North Pole to the South Pole. The compass will point south at every step along every route but different routes will cross the equator at different locations. Every line of longitude connects the two poles. Every line of longitude describes an equally short route between the poles.

Ambiguity also arises if the starting point and destination lie at complementary latitudes and longitudes. For example, many equally short routes connect the point at `45 N, 90 W` with the point at `45 S, 90 E`. The first point is in central Wisconsin. The second point is in the Indian Ocean between Durban, South Africa and Perth, Australia. In this case, a navigator can make an arbitrary choice. One possible route crosses the North Pole. Another will take the traveler through Antarctica. Other choices would allow a traveler to visit Hawaii, Rome, or any other favorite port of call en route between Wisconsin and the Indian Ocean.

The ambiguity arises because these two points and the earth's center do not define a unique plane. In the ambiguous case, the earth's center lies on the straight line that connects the journey's endpoints. An infinite number of planes pass through that straight line. The intersection of each of these planes with the earth's surface defines a different great circle.

1.3 The General Case

In the general case...

- neither of a journey's endpoints lies at one of the earth's poles
- at least one endpoint lies off of the equator,
- the journey's endpoints have distinct and non-complementary longitudes

In the general case, the circle that contains the shortest route between two points neither coincides with the equator nor passes through the poles. The great circle intersects the equatorial plane at an angle greater than 0 degrees but less than 90 degrees.

One solution to the problem of computing great circles proceeds by transforming the general case into one of the easy cases. The program rotates the great circle into a simpler frame, computes properties of the great circle in that simpler frame, and then rotates the circle back to its original tilt relative to the equator.

2 Ways to Describe the Problem Mathematically

A navigator will want to know the length of the shortest route between a journey's start and end. A navigator will want the coordinates of points through which the route passes. Calculations of distance and coordinates are easy if the route is a circular arc centered at the origin and situated in the x–y plane.

- Let s a point in the x–y plane and the starting point of a journey.
- Let d, the destination of the journey, be another point in the x–y plane at the same distance from the origin as s.
- Let r be the radius of the circle that is centered at the origin and that joins s and d. It is the radial coordinate of the two points: $s_r = d_r = r$.
- Let s_ϕ be the radial coordinate (measured in radians) of the starting point of a journey.
- Let d_ϕ be the radial coordinate of the destination of a journey.
- Let f be a fraction of the distance between s and d: $0 \leq f \leq 1$.

Then...

- the distance between s and d is $|r \cdot (d_\phi - s_\phi)|$
- the polar coordinates of a point that lies on the circular arc that connects s and d, at a fraction f of the distance between the two points are $(r, s_\phi + f \cdot (d_\phi - s_\phi))$

2.1 Coordinate Systems

A program aids navigators best if it allows navigators to enter coordinates in the form they prefer. It must also print the results of its computations in the navigators' preferred form. Angular coordinates will make the measure and subdivision of arcs easy. Cartesian coordinates will allow the use of affine transformations to move a great circle into a convenient frame.

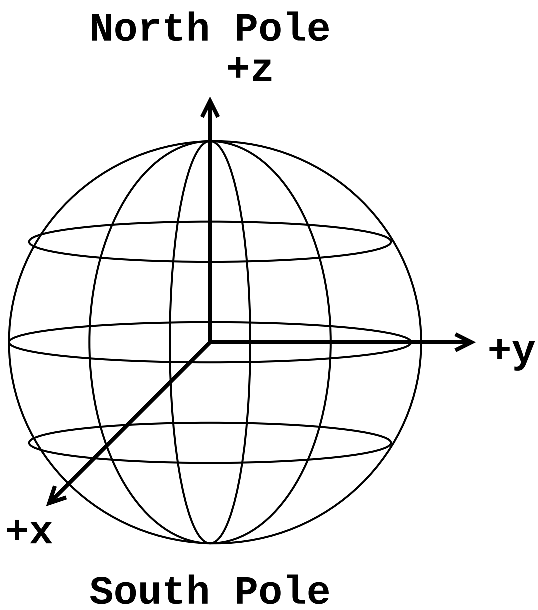

Figure 2: Latitudes, longitudes, and the Cartesian coordinate axes.

Aviators and mariners specify the location of a point on the earth's surface with a latitude and longitude. A latitude is a number and a letter. The letter indicates whether the number measures degrees north ('N') or south ('S') of the equator. A longitude is also a number and a letter. With a longitude, the letter indicates whether the number measures degrees east ('E') or west ('W') of the prime meridian. The prime meridian is the line of longitude that runs from the North Pole through Greenwich, England to the South Pole.

The geographers' system of coordinates resembles the spherical system of coordinates in which all points have the same radial coordinate. In the spherical system, numbers without letters suffice to specify angular coordinates.

Let $\vec{p}$ be the vector that extends from the origin to a point on the earth's surface,

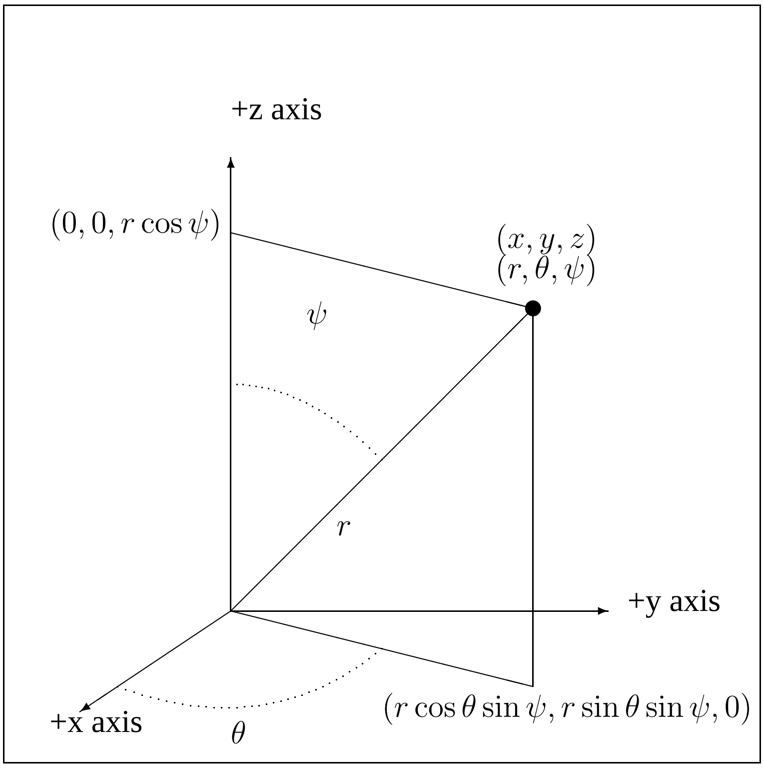

Figure 3: Cartesian and spherical coordinate systems.

Then the first of the spherical coordinates r specifies the length of $\vec{p}$. This is the distance of the point from the origin, always a non-negative number, and, in this application, the earth's radius. By assuming that the earth is a perfect sphere and by choosing to place the origin at the earth's center, we make this radial coordinate the same for all points on the earth.

The second spherical coordinate θ specifies the angle between a projection onto the x–y plane of $\vec{p}$ and the x–z plane. Values range from 0 to 2π. It corresponds to the point's longitude.

The third spherical coordinate ψ specifies the angle between $\vec{p}$ and the z axis. Values range from 0 to π. It corresponds to the point's latitude. The value of θ is arbitrary when $\psi = 0$ or $\psi = \pi$.

Latitude and longitude to spherical coordinates

A point on the equator lies neither in the northern hemisphere nor in the southern hemisphere. The number 0 by itself, without a letter, suffices to name the latitude of points at the equator.

A point on the prime meridian lies neither in the eastern nor the western hemisphere. The number 0 by itself, without a letter, suffices to name the longitude of points on the prime meridian. Similarly, points at longitude $180°$ also lie on a line that separates the eastern and

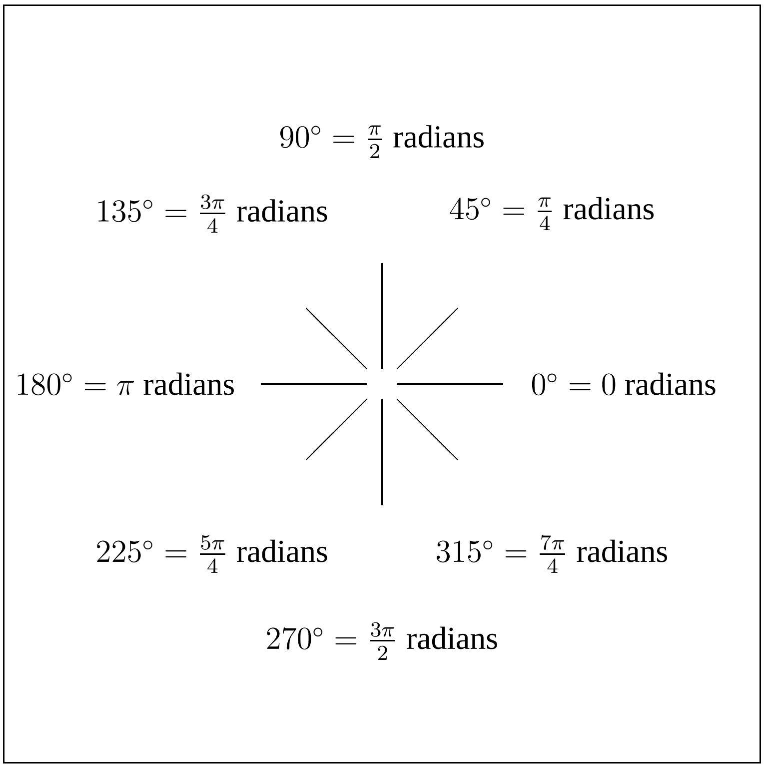

Figure 4: Two measures of angle: radians and degrees.

western hemispheres.

The best established convention measures fractions of a degree in minutes and seconds of arc: 60 minutes in one degree and 60 seconds in one minute. Not all modern cartographers comply with this convention.

To produce the third spherical coordinate ψ, proceed as follows: Subtract the number of degrees of latitude from 90° if the latitude corresponds to a point in the northern hemisphere. If the latitude corresponds to a point in the southern hemisphere, add the number of degrees of latitude to 90°. For points at the equator, subtraction of 0° from 90° yields that same result as addition of 0° to 90°. Then convert this measure in degrees to an equivalent measure in radians.

$$\begin{aligned} degrees &= 180 \cdot \frac{radians}{\pi} \\ radians &= \pi \cdot \frac{degrees}{180} \end{aligned}$$

To produce the second spherical coordinate θ, proceed as follows: Subtract the number of degrees of longitude from 360° if the longitude corresponds to a point in the western hemisphere. For points in the eastern hemisphere or the boundary between the two hemispheres, no addition or subtraction is necessary. Then convert this measure in degrees to an equivalent measure in radians.

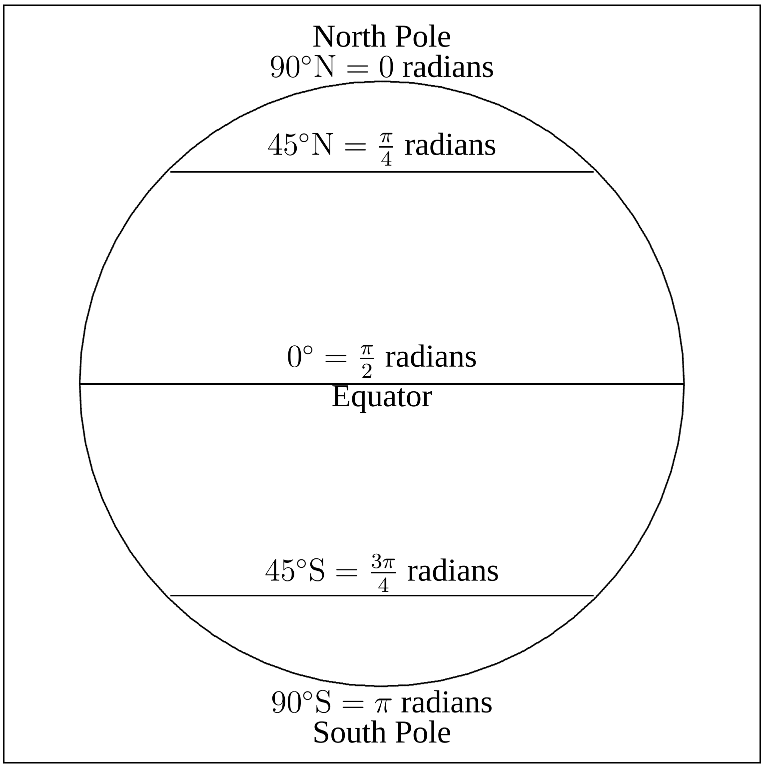

Figure 5: Values of third spherical coordinate ψ at various latitudes.

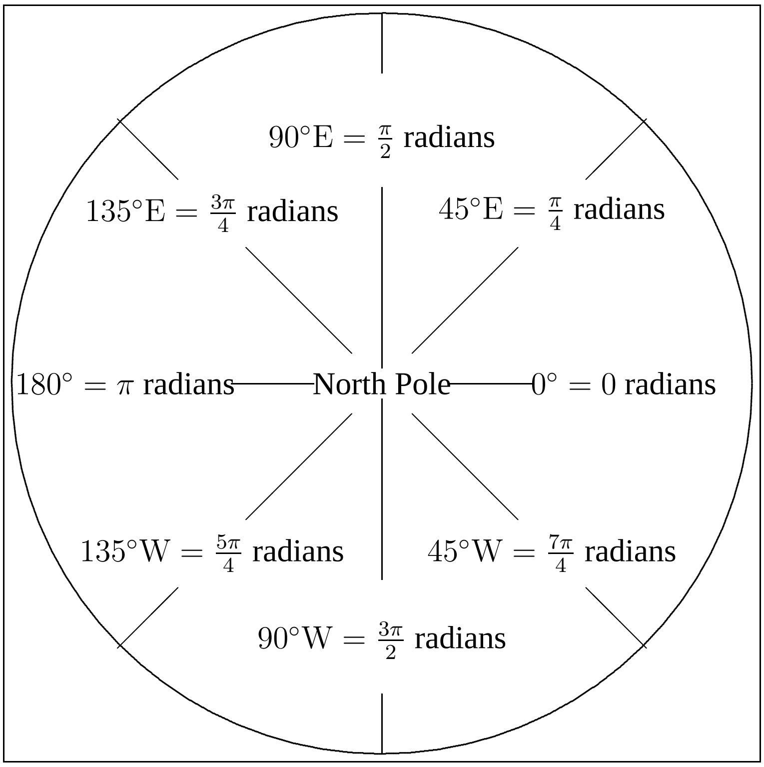

Figure 6: Values of second spherical coordinate θ at various longitudes.

Spherical to Cartesian coordinates

$$
\begin{aligned}
x &= r\cos\theta\sin\psi \\
y &= r\sin\theta\sin\psi \\
z &= r\cos\psi
\end{aligned}
$$

Cartesian to spherical coordinates

$$
\begin{aligned}
r &= \sqrt{x^2+y^2+z^2} \\
\cos\theta &= \frac{x}{\sqrt{x^2+y^2}} \\
\sin\theta &= \frac{y}{\sqrt{x^2+y^2}} \\
\theta &= \arctan\frac{y}{x} \\
\cos\psi &= \frac{z}{r} \\
\sin\psi &= \frac{\sqrt{x^2+y^2}}{r} \\
\psi &= \arctan\frac{\sqrt{x^2+y^2}}{z}
\end{aligned}
$$

Spherical coordinates to latitude and longitude

$$
\begin{aligned}
latitude &= \begin{cases} (90-(180\,\psi)/\pi)^\circ \text{ N} & \text{if } 0 \le \psi < \frac{\pi}{2} \\ 0^\circ & \text{if } \psi = \frac{\pi}{2} \\ ((180\,\psi)/\pi - 90)^\circ \text{ S} & \text{if } \frac{\pi}{2} < \psi \le \pi \end{cases} \\
longitude &= \begin{cases} 0^\circ & \text{if } \theta = 0 \\ ((180\,\theta)/\pi)^\circ \text{ E} & \text{if } 0 < \theta < \pi \\ 180^\circ & \text{if } \theta = \pi \\ (360-(180\,\theta)/\pi)^\circ \text{ W} & \text{if } \pi < \theta < 2\pi \end{cases}
\end{aligned}
$$

3 Algorithms

3.1 Transformation

Two rotations in succession map a great circle onto the equator. The first rotation moves the normal vector of the great circle into the x–z plane. The second rotation moves the normal vector onto the z axis.

From the Cartesian coordinates of two points on the surface of the earth, we can construct vectors that extend from the origin to the points. Products of vectors yield the elements of matrices that describe the rotations needed to move the great circle that connects the points onto the equator.

The cross product of two normalized vector yields a new vector that is perpendicular to both multiplier and multiplicand. In particular, if $\vec{p_0}$ is a vector that extends from the origin to one point on the earth's surface and $\vec{p_1}$ is a vector that extends from the origin to another point on the surface (and the origin is not colinear with $\vec{p_0}$ and $\vec{p_1}$), then $\vec{p_0} \times \vec{p_1}$ is a vector perpendicular to the plane defined by the origin and the two points on the surface.

$$\begin{aligned}\vec{n} &= \vec{p_0} \times \vec{p_1} \\ &= (p_0.y\ p_1.z - p_0.z\ p_1.y, p_0.x\ p_1.z - p_0.z\ p_1.x, p_0.x\ p_1.y - p_0.y\ p_1.x)\end{aligned}$$

Normalization of a vector produces a new vector with the same direction as the old but a length of one.

$$\begin{aligned}|\vec{n}| &= \sqrt{n.x^2 + n.y^2 + n.z^2} \\ \hat{n} &= \frac{1}{|\vec{n}|}\vec{n} \\ &= (\frac{n.x}{|\vec{n}|}, \frac{n.y}{|\vec{n}|}, \frac{n.z}{|\vec{n}|})\end{aligned}$$

The dot product of two normalized vectors produces the cosine of the angle between them. In particular, let $\hat{k} = (0, 0, 1)$ be the vector of unit length that points along the positive z axis. Then $\hat{n} \cdot \hat{k}$ is the cosine of the angle between $\vec{n}$ and the positive z axis. Because $\cos^2 \phi + \sin^2 \phi = 1$ for any angle ϕ, the sine of the angle between $\vec{n}$ and the positive z axis is $\sqrt{1 - (\hat{n} \cdot \hat{k})^2}$.

The projection of $\vec{n}$ onto the x–y plane is $\vec{n}_{xy} = (n.x, n.y, 0)$. Let $\hat{i} = (1, 0, 0)$ be the vector of unit length that points along the positive x axis. Then $\hat{n}_{xy} \cdot \hat{i}$ is the cosine of the angle between the two vectors.

Construction of a matrix that describes a rotation about one of the principal axes does not require a knowledge of the angle of rotation, but only a knowledge of the angle's cosine and sine.

$$\begin{aligned}\mathbf{R_z}(\phi) &= \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \mathbf{R_y}(\phi) &= \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix}\end{aligned}$$

The product $\mathbf{R_z}(\phi)\vec{v}$ is the vector $\vec{v}$ rotated ϕ radians—it is a new vector whose length is that same as that of $\vec{v}$ but whose orientation differs by ϕ radians.

The cross product of two normalized vectors yields the sine of the angle between the vectors. Their dot product yields the cosine of the angle between the vectors. The division of the sine by the cosine yields the tangent of the angle. The arctangent function, applied to the tangent, yields the angle. The product of an angle measured in radians and the radius of an arc yields the length of the arc.

3.2 Rotation within the Plane of the Great Circle

A different algorithm begins by computing the normal to the plane of a great circle and the angle spanned by the path between the endpoints. The normal vector defines an axis. The algorithm constructs a matrix that describes rotations about that axis. Then it rotates the vector that points from the origin to the starting point of the journey in small steps about that axis. Each rotated vector points to a new point on the great circle.

This algorithm constructs its operator by forming the product of the same matrices that helped map a great circle onto the equator in the previous algorithm.

3.3 Recursive Calculation of Midpoints

If $\vec{p_0}$ and $\vec{p_1}$ are vectors that point to the endpoints of a circular arc whose radius is r, then the vector $\vec{mp}$ that points to the midpoint of the arc is:

$$
\begin{aligned}
\vec{v} &= \frac{1}{2}(\vec{p}_0 + \vec{p}_1) \\
\vec{v}_{midpoint} &= \frac{r}{|\vec{v}|}\vec{v}
\end{aligned}
$$

This suggests yet another algorithm for finding points on a great circle. Find the midpoint of the great circle path that connects two points. That divides the path into two halves. Then find the midpoints of the halves. Recursively find midpoints of quarters of the arc, eighths of the arc, and so on until the process has computed points whose number is large enough and whose spacing is close enough to give a picture of a smooth curve.

3.4 Gnomic Projection

Suppose that both the starting point and destination of a journey lie in the northern hemisphere on the earth's surface. The origin **O** of a Cartesian coordinate system is at the earth's center. The z axis coincides with the earth's rotational axis. The positive z axis pierces the earth's surface at the North Pole. Finally, imagine a plane P that is parallel to the x–y plane and tangent to the earth at the North Pole.

Let $\vec{m}_0$ be a vector that points to the intersection of a line from the origin through the starting point of the journey with the plane P. Let $\vec{m}_1$ be a vector that points to the intersection of a line from the origin through the destination with the plane P. The line segment that connects the two points of intersection is a projection of the great circle from the surface of the earth onto the plane P.

Points p on that line segment are:

$$p = \mathbf{O} + ((1-w)\vec{m}_0 + w\vec{m}_1) \quad \text{where} \quad 0 \leq w \leq 1$$

The inverse projection yields points on the great circle. The intersection of a line that connects the origin to any point p with the earth's surface is a point on the great circle.

The inverse projection proceeds as follows. Normalize the vector that points to p and then multiply the normalized vector by the radius of the earth. The result points to a point on the great circle. The elements of the vector are the Cartesian coordinates of a point on the shortest path. The last step produces a latitude and longitude from the Cartesian coordinates.

4 Design, Test, Build, and Extend

Design should precede and guide construction, rather than follow and attempt to justify. Tests written before building a program describe intended outputs and define correct results. A knowledge of actual results biases tests written after writing the program. Actual results differ from intended results when programmers make false assumptions during the long and difficult labor of writing a program. Tests written after writing the program often repeat and so appear to confirm those false assumptions.

The design of a solution to the problem of computing great circles must describe rerpresentations of coordinates and methods for translating between representations.

Tests of a design should apply functions to values inside and outside of allowed ranges. The symmetry of ranges suggests a way of choosing pairs of values. For example, a test might examine two points that are equidistant from the equator but in opposite hemispheres. Tests at the extremes of a range often reveal more than tests of values in the middle of a range. Points on the equator are neither in the northern nor the southern hemisphere. Points on the prime meridian are neither in the eastern nor the western hemisphere.

Singular points deserve special attention—the North and South Poles have no longitude and their latitude has no fractional part.

Fixed points (at which $x = f(x)$) suggest particularly simple tests. For example, the longitude of points on the prime meridian is zero and so is the corresponding spherical coordinate of such points. The function that translates between the two kinds of coordinates maps zero to zero in the case of points on the prime meridian.

Inverse functions aid testing: $x = f^{-1}(f(x))$. They allow round-trip conversions. A test might convert a latitude and longitude to Cartesian coordinates and then the convert those Cartesian coordinates to a latitude and longitude. The latitude and longitude that go into this sequence of transformations should match the latitude and longitude that come out.

The shortest distance between two points on the surface of the earth will never exceed half of the earth's circumference. The distance function is symmetric: the distance from a point a to a point b should always equal the distance from b to a. The triangular inequality applies: the sum of the distances from a to b and from b to c will be at least as great as the length of the path that connects a directly to c. Of course, the distance from any point a to a (the length of a journey that goes nowhere) is zero.

Testing exposes defects in design and directs modifications to design. Extensions of a design require the composition of new tests. Repeated and automated testing checks that progress on one part of a design has not undone progress previously made on another part. Designs grow incrementally and converge upon a solution.

Finally, the solution of one problem often uncovers other problems.

Before the invention of computers, navigators knew how to plot great circles but lacked practical means of following routes that required a new heading at every point along the way. Instead, they approximated great circles by a series of paths along which the compass heading remained constant. Mercator made the identification of the constant compass heading that would carry a traveler from one point to another easy through his invention of a map on which navigators can draw such paths—called loxodromes—with a straight edge. On Mercator's projection, longitudes are parallel and horizontal scale is a function of latitude.

Modern explorers might also consider problems of tracking satellites. At what times will the satellite pass over a given point?

How might the methods described here be adapted for the solution of these or other related problems?

5 Conclusion

This paper has provided the basis for an exercise in design and testing. It challenges readers to select a specific aspect of a problem it has described in general terms. For example, a student who undertakes this exercise might choose to compute the length of a great circle route, or the locations of points on a great circle, or compass headings at points on the route. This paper outlines several algorithms and challenges readers to choose one. By describing problem and solution without the use of a programming language, it demonstrates that design and test can—and should—precede coding.

For students and teachers who want to follow this exercise with an exercise in coding, the author has written a program in Java that computes points on a great circle path. The author invites communication from readers.

Reading List

[1] Ken Alder. *The Measure of All Things: the seven-year odyssey and hidden error that transformed the world*. The Free Press, New York, 2002.

[2] Lloyd A. Brown. *The Story of Maps*. Little, Brown and Company, Boston, 1949.

[3] General Cartography. *General Cartography*. McGraw-Hill Book Company, New York, 1948.

[4] G.R. Crone. *Maps and their Makers: an introduction to the history of cartography*. Hutchinson & Co., London, 1962.

[5] Timothy G. Feeman. *Portraits of the Earth: a mathematician looks at maps*, volume 18 of *Mathematical World*. American Mathematical Society, Providence, Rhode Island, 2002.

[6] Cornelius Lanczos. *Space Through the Ages: the evolution of geometrical ideas from Pythagoras to Hilbert and Einstein*. Academic Press, New York, 1970.

[7] Ivan I. Mueller and Heinrich Eichhorn. *Spherical and Practical Astronomy as Applied to Geodesy*. Frederick Ungar Publishing Co., New York, 1969.

[8] Archie E. Roy. *The Foundations of Astrodynamics*. The Macmillan Company, New York, 1965.

[9] Dava Sobel. *Longitude: the true story of a lone genius who solved the greatest scientific problem of his time*. Walker and Company, New York, 1995.

[10] E.G.R. Taylor. *The Haven-Finding Art: a history of navigation from Odysseus to Captain Cook*. Abelard-Schuman Limited, New York, 1957.