NOTES

An Application of Geography to Mathematics: History of the Integral of the Secant

V. FREDERICK RICKEY
Bowling Green State University
Bowling Green, OH 43403

PHILIP M. TUCHINSKY
Ford Motor Company
Engineering Computing Center
Dearborn, MI 48121

Every student of the integral calculus has done battle with the formula

$$\int \sec\theta\, d\theta = \ln|\sec\theta + \tan\theta| + c. \tag{1}$$

This formula can be checked by differentiation or "derived" by using the substitution $u = \sec\theta + \tan\theta$, but these ad hoc methods do not make the formula any more understandable. Experience has taught us that this troublesome integral can be motivated by presenting its history. Perhaps our title seems twisted, but the tale to follow will show that this integral should be presented not as an application of mathematics to geography, but rather as an application of geography to mathematics.

The secant integral arose from cartography and navigation, and its evaluation was a central question of mid-seventeenth century mathematics. The first formula, discovered in 1645 before the work of Newton and Leibniz, was

$$\int \sec\theta\, d\theta = \ln\left|\tan\left(\frac{\theta}{2} + \frac{\pi}{4}\right)\right| + c, \tag{2}$$

which is a trigonometric variant of (1). This was discovered, not through any mathematician's cleverness, but by a serendipitous historical accident when mathematicians and cartographers sought to understand the Mercator map projection. To see how this happened, we must first discuss sailing and early maps so that we can explain why Mercator invented his famous map projection.

From the time of Ptolemy (c. 150 A.D.) maps were drawn on rectangular grids with one degree of latitude equal in length to one degree of longitude. When restricted to a small area, like the Mediterranean, they were accurate enough for sailors. But in the age of exploration, the Atlantic presented vast distances and higher latitudes, and so the navigational errors due to using the "plain charts" became apparent.

The magnetic compass was in widespread use after the thirteenth century, so directions were conveniently given by distance and compass bearing. Lines of fixed compass direction were called **rhumb** lines by sailors, and in 1624 Willebrord Snell dubbed them **loxodromes**. To plan a journey one laid a straightedge on a map between origin and destination, then read off the compass bearing to follow. But rhumb lines are spirals on the globe and curves on a plain chart —facts sailors had difficulty understanding. They needed a chart where the loxodromes were represented as straight lines.

It was Gerardus Mercator (1512–1594) who solved this problem by designing a map where the lines of latitude were more widely spaced when located further from the equator. On his famous world map of 1569 ([1], p. 46), Mercator wrote:

> In making this representation of the world we had...to spread on a plane the surface of the sphere in such a way that the positions of places shall correspond on all sides with each other both in so far as true direction and distance are concerned and as concerns correct longitudes and latitudes... . With this intention we have had to employ a new proportion and a new arrangement of the meridians with reference to the parallels. ... It is for these reasons that we have progressively increased the degrees of latitude towards each pole in proportion to the lengthening of the parallels with reference to the equator.

Mercator wished to map the sphere onto the plane so that both angles and distances are preserved, but he realized this was impossible. He opted for a conformal map (one which preserves angles) because, as we shall see, it guaranteed that loxodromes would appear on the map as straight lines.

Unfortunately, Mercator did not explain how he "progressively increased" the distances between parallels of latitude. Thomas Harriot (c. 1560–1621) gave a mathematical explanation in the late 1580's, but neither published his results nor influenced later work (see [6], [11]-[15]). In his *Certaine Errors in Navigation...* [22] of 1599, Edward Wright (1561–1615) finally gave a mathematical method for constructing an accurate Mercator map. The Mercator map has its meridians of longitude placed vertically and spaced equally. The parallels of latitude are horizontal and unequally spaced. Wright's great achievement was to show that the parallel at latitude θ should be stretched by a factor of $\sec\theta$ when drawn on the map. Let us see why.

FIGURE 1 represents a wedge of the earth, where AB is on the equator, C is the center of the earth, and T is the north pole. The parallel at latitude θ is a circle, with center P, that includes arc MN between the meridians AT and BT. Thus BC and NP are parallel and so angle $PNC=\theta$. The "triangles" ABC and MNP are similar figures, so

$$\frac{AB}{MN}=\frac{BC}{NP}=\frac{NC}{NP}=\sec\theta,$$

or $AB=MN\sec\theta$. Thus when MN is placed on the map it must be stretched horizontally by a factor $\sec\theta$. (This argument is not the one used by Wright [22]. His argument is two dimensional and shows that $BC=NP\sec\theta$.)

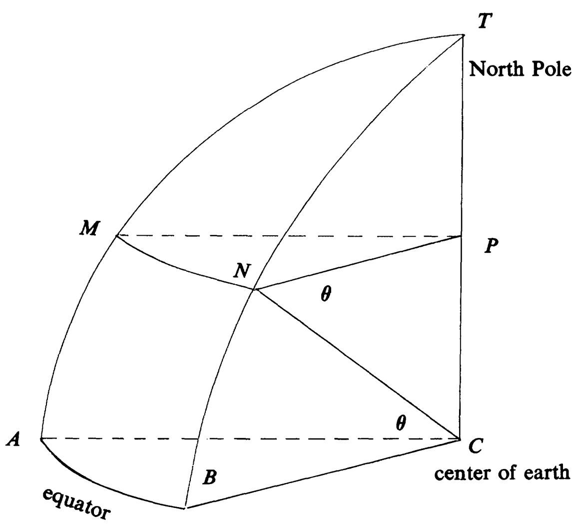

FIGURE 1.

Suppose we can construct a map where angles are preserved, i.e., where the globe-to-map function is conformal. Then a loxodrome, which makes the same angle with each meridian, will appear on this map as a curve which cuts all the map's meridians (a family of parallel straight lines) at the same angle. Since a curve that cuts a family of parallel straight lines at a fixed angle is a straight line, loxodromes on the globe will appear straight on the map. Conversely, if loxodromes are mapped to straight lines, the globe-to-map function must be conformal.

In order for angles to be preserved, the map must be stretched not only horizontally, but also vertically, by $\sec\theta$; this, however, requires an argument by infinitesimals. Let $D(\theta)$ be the distance *on the map* from the equator to the parallel of latitude θ, and let dD be the infinitesimal change in D resulting from an infinitesimal change $d\theta$ in θ. If we stretch vertically by $\sec\theta$, i.e., if

$$dD = \sec\theta \, d\theta$$

then an infinitesimal region on the globe becomes a similar region on the map, and so angles are preserved. Conversely, if the map is to be conformal the vertical multiplier must be $\sec\theta$.

Finally, "by perpetuall addition of the Secantes," to quote Wright, we see that the distance on the map from the equator to the parallel at latitude θ is

$$D(\theta) = \int_0^\theta \sec\theta \, d\theta.$$

Of course Wright did not express himself as we have here. He said ([2], pp. 312–313):

> the parts of the meridian at euery poynt of latitude must needs increase with the same proportion wherewith the Secantes or hypotenusae of the arke, intercepted betweene those pointes of latitude and the aequinoctiall [equator] do increase. ... For...by perpetuall addition of the Secantes answerable to the latitudes of each point or parallel vnto the summe compounded of all former secantes,...we may make a table which shall shew the sections and points of latitude in the meridians of the nautical planisphaere: by which sections, the parallels are to be drawne.

Wright published a table of "meridional parts" which was obtained by taking $d\theta = 1'$ and then computing the Riemann sums for latitudes below 75°. Thus the methods of constructing Mercator's "true chart" became available to cartographers.

Wright also offered an interesting physical model. Consider a cylinder tangent to the earth's equator and imagine the earth to "swal [swell] like a bladder." Then identify points on the earth with the points on the cylinder that they come into contact with. Finally unroll the cylinder; it will be a Mercator map. This model has often been misinterpreted as the cylindrical projection (where a light source at the earth's center projects the unswollen sphere onto its tangent cylinder), but this projection is not conformal.

We have established half of our result, namely that the distance on the map from the equator to the parallel at latitude θ is given by the integral of the secant. It remains to show that it is also given by $\ln|\tan(\frac{\theta}{2} + \frac{\pi}{4})|$.

In 1614 John Napier (1550–1617) published his work on logarithms. Wright's authorized English translation, *A Description of the Admirable Table of Logarithms*, was published in 1616. This contained a table of logarithms of sines, something much needed by astronomers. In 1620 Edmund Gunter (1581–1626) published a table of common logarithms of tangents in his *Canon triangulorum*. In the next twenty years numerous tables of logarithmic tangents were published and so were widely available. (Not even a table of secants was available in Mercator's day.)

In the 1640's Henry Bond (c. 1600–1678), who advertised himself as a "teacher of navigation, survey and other parts of the mathematics," compared Wright's table of meridional parts with a log-tan table and discovered a close agreement. This serendipitous accident led him to conjecture that $D(\theta) = \ln|\tan(\frac{\theta}{2} + \frac{\pi}{4})|$. He published this conjecture in 1645 in Norwood's *Epitome of Navigation*. Mainly through the correspondence of John Collins this conjecture became widely known. In fact, it became one of the outstanding open problems of the mid-seventeenth century, and was attempted by such eminent mathematicians as Collins, N. Mercator (no relation), Wilson, Oughtred and John Wallis. It is interesting to note that young Newton was aware of it in 1665 **[18]**, **[21]**.

The "Learned and Industrious *Nicolaus Mercator*" in the very first volume of the *Philosophical Transactions* of the Royal Society of London was "willing to lay a *Wager* against any one or more persons that have a mind to engage... *Whether the Artificial* [logarithmic] *Tangent-line be the true Meridian-line*, yea or no?" (**[9]**, pp. 217–218). Nicolaus Mercator is not, as the story is often told, wagering that he knows more about logarithms than his contemporaries; rather, he is offering a prize for the solution of an open problem.

The first to prove the conjecture was, to quote Edmund Halley, "the excellent Mr. *James Gregory* in his *Exercitationes Geometricae*, published *Anno* 1668, which he did, not without a long train of Consequences and Complication of Proportions, whereby the evidence of the Demonstration is in a great measure lost, and the Reader wearied before he attain it" (**[7]**, p. 203). Judging by Turnbull's modern elucidation **[19]** of Gregory's proof, one would have to agree with Halley. At any rate, Gregory's proof could not be presented to today's calculus students, and so we omit it here.

Isaac Barrow (1630–1677) in his *Geometrical Lectures* (Lect. XII, App. I) gave the first "intelligible" proof of the result, but it was couched in the geometric idiom of the day. It is especially noteworthy in that it is the earliest use of partial fractions in integration. Thus we reproduce it here in modern garb:

$$
\begin{aligned}
\int \sec\theta\, d\theta &= \int \frac{1}{\cos\theta}\, d\theta \\
&= \int \frac{\cos\theta}{\cos^2\theta}\, d\theta \\
&= \int \frac{\cos\theta}{1-\sin^2\theta}\, d\theta \\
&= \int \frac{\cos\theta}{(1-\sin\theta)(1+\sin\theta)}\, d\theta \\
&= \frac{1}{2}\int \frac{\cos\theta}{1-\sin\theta} + \frac{\cos\theta}{1+\sin\theta}\, d\theta \\
&= \frac{1}{2}[-\ln|1-\sin\theta| + \ln|1+\sin\theta|] + c \\
&= \frac{1}{2}\ln\left|\frac{1+\sin\theta}{1-\sin\theta}\right| + c \\
&= \frac{1}{2}\ln\left|\frac{1+\sin\theta}{1-\sin\theta}\cdot\frac{1+\sin\theta}{1+\sin\theta}\right| + c \\
&= \frac{1}{2}\ln\left|\frac{(1+\sin\theta)^2}{1-\sin^2\theta}\right| + c \\
&= \frac{1}{2}\ln\left|\frac{(1+\sin\theta)^2}{(\cos\theta)^2}\right| + c \\
&= \ln\left|\frac{1+\sin\theta}{\cos\theta}\right| + c \\
&= \ln|\sec\theta + \tan\theta| + c.
\end{aligned}
$$

We became interested in this topic after noting one line of historical comment in Spivak's excellent *Calculus* (p. 326). As we ferreted out the details and shared them with our students, we found an ideal soapbox for discussing the nature of mathematics, the process of mathematical discovery, and the role that mathematics plays in the world. We found this so useful in the classroom that we have prepared a more detailed version for our students **[17]**.

References

The following works contain interesting information pertaining to this paper. The best concise source of information about the individuals mentioned in this paper is the excellent *Dictionary of Scientific Biography*, edited by C. C. Gillespie.

[1] Anonymous, Gerard Mercator's Map of the World (1569), Supplement no. 2 to Imago Mundi, 1961.

[2] Florian Cajori, On an integration ante-dating the integral calculus, Bibliotheca Mathematica, 3rd series, 14(1915) 312–319.

[3] H. S. Carslaw, The story of Mercator's map. A chapter in the history of mathematics, Math. Gaz., 12(1924) 1–7.

[4] Georgina Dawson, Edward Wright, mathematician and hydrographer, Amer. Neptune, 37(1977) 174–178.

[5] Jacques Delevsky, L'invention de la projection de Mercator et les enseignements de son histoire, Isis, 34(1942) 110–117.

[6] Frank George, Hariot's meridional parts, J. Inst. Navigation, London, 21(1968) 82–83.

[7] E. Halley, An easie demonstration of the analogy of the logarithmick tangents to the meridian line or sum of secants: with various methods for computing the same to the utmost exactness, Philos. Trans., Roy. Soc. London, 19(1695–97) 202–214.

[8] Johannes Keuning, The history of geographical map projections until 1600, Imago Mundi, 12(1955) 1–24.

[9] Nicolaus Mercator, Certain problems touching some points of navigation, Philos. Trans., Roy. Soc. London, 1(1666) 215–218.

[10] E. J. S. Parsons and W. F. Morris, Edward Wright and his work, Imago Mundi, 3(1939) 61–71.

[11] Jon V. Pepper, Hariot's calculation of the meridional parts as logarithmic tangents, Archive for History of Exact Science, 4(1967) 359–413.

[12] ______, A note on Hariot's method of obtaining meridional parts, J. Inst. Navigation, London, 20(1967) 347–349.

[13] ______, The study of Thomas Hariot's manuscripts, II: Hariot's unpublished papers, History of Science, 6 (1967) 17–40.

[14] ______, Hariot's earlier work on mathematical navigation: theory and practice. With an appendix, 'The early development of the Mercator chart,' in Thomas Hariot: Renaissance Scientist, Clarendon Press, Oxford, 1974, John W. Shirley, editor, pp. 54–90.

[15] D. H. Sadler and Eva G. R. Taylor, The doctrine of nauticall triangles compendious. Part I–Thomas Hariot's manuscript (by Taylor). Part II–Calculating the meridional parts (by Sadler), J. Inst. Navigation, London, 6(1953) 131–147.

[16] Eva G. R. Taylor, The Haven-Finding Art, Hollis and Carter, London, 1971.

[17] P. M. Tuchinsky, Mercator's World Map and the Calculus, Modules and Monographs in Undergraduate Mathematics and its Applications (UMAP) Project, Education Development Center, Newton, Mass., 1978.

[18] H. W. Turnbull, editor, The Correspondence of Isaac Newton, Cambridge Univ. Press, 1959–1960, vol. 1, pp. 13–16, and vol. 2, pp. 99–100.

[19] H. W. Turnbull, James Gregory Tercentenary Memorial Volume, G. Bell & Sons, London, 1939, pp. 463–464.

[20] D. W. Waters, The Art of Navigation in England in Elizabethan and Early Stuart Times, Yale Univ. Press, New Haven, 1958.

[21] D. T. Whiteside, editor, The Mathematical Papers of Isaac Newton, vol. 1, Cambridge Univ. Press, 1967, pp. 466–467, 473–475.

[22] Edward Wright, Certaine Errors in Navigation, Arising either of the ordinaire erroneous making or vsing of the sea Chart, Compasse, Crosse staffe, and Tables of declination of the Sunne, and fixed Starres detected and corrected, Valentine Sims, London, 1599. Available on microfilm as part of Early English Books 1475–1640, reels 539 and 1018 (these two copies from 1599 have slightly different title pages). The preface and table of meridional parts have been reproduced as "Origin of meridional parts," International Hydrographic Review, 8(1931) 84–97.