Notes on Differentiation

1 The Chain Rule

This is the following famous result:

1.1 Theorem. *Suppose U and V are open sets with f and g complex-valued functions defined on U and V respectively, where $f(U) \subset V$. Suppose that $z_0 \in U$ (so that $f(z_0) \in V$). If f is (complex) differentiable at z_0 and g is differentiable at $f(z_0)$, then $g \circ f : U \to \mathbb{C}$ is differentiable at z_0, and $(g \circ f)'(z_0) = g'(f(z_0))f(z_0)$.*

Proof. Let $w_0 = f(z_0)$. Our hypotheses are that

$$f(z) - f(z_0) = f'(z_0)(z - z_0) + R(z) \quad \text{where} \quad \frac{|R(z)|}{|z - z_0|} \to 0 \text{ as } z \to z_0, \tag{1}$$

and

$$g(w) - g(w_0) = g'(w_0)(w - w_0) + S(w) \quad \text{where} \quad \frac{|R(w)|}{|w - w_0|} \to 0 \text{ as } w \to w_0. \tag{2}$$

Our goal is to show that if $h = g \circ f$ then

$$h(z) - h(z_0) = g'(f(z_0))(z - z_0) + T(z) \quad \text{where} \quad \frac{|T(z)|}{|z - z_0|} \to 0 \text{ as } z \to z_0. \tag{3}$$

To this end, substitute $w = f(z)$ into (2) (legal because $f(U) \subset V$) to get

$$h(z) - h(z_0) = g'(f(z_0))[f(z) - f(z_0)] + S(f(z)),$$

and then substitute the result of (1) into this equation to get:

$$h(z) - h(z_0) = g'(f(z_0))f'(z_0)(z - z_0) + g'(f(z_0))R(z) + S(f(z)) \tag{4}$$

The second term on the right side of (4) is $o(|z - z_0|)$ as $z \to z_0$, so all we have to do is show that the same is true for the third term on the right.

For this, let $\varepsilon(w) = S(w)/|w - w_0|$, so that $S(w) = \varepsilon(w)|w - w_0|$ where $\varepsilon(w) \to 0$ as $w \to w_0$. Then substituting $w = f(z)$ and using (1) one last time:

$$S(f(z)) = \varepsilon(f(z))|f(z) - f(z_0)| = \varepsilon(f(z))|f'(z_0)|\,|z - z_0| + \varepsilon(f(z))R(z),$$

where, on the right-hand side of the equation, the first term on the is $o(|z - z_0|)$ because $f(z) \to f(z_0)$ as $z \to z_0$ (remember: differentiability at a point implies continuity there), and the second term is $o(|z - z_0|)$ because $R(z)$ has this property, and ε is bounded. Thus $S(f(z))$ is $o(|z - z_0|)$ and $z \to z_0$, and the proof of the chain rule is complete. $\square$

1.2 Exercise. Suppose f obeys the hypotheses above, $[a, b]$ is a finite, closed real interval contained in $f(U)$, and $\gamma : [a, b] \to \mathbb{C}$ is (real) differentiable at a point $t_0 \in (a, b)$. show that $f \circ \gamma$ is (real) differentiable at t_0, and that $(f \circ \gamma)'(t_0) = f'(\gamma(t_0))\gamma'(t_0)$.

2 Sufficient condition for differentiability.

We begin with a concrete situation. Suppose G is an open subset of R^2 and $u : G \to \mathbb{R}$ is a real-valued function defined on G. Let $z_0 = (x_0, y_0)$ be a point of G.

2.1 Theorem. *If the first partial derivatives of u exist at every point of G and are continuous at z_0, the u is differentiable at z_0.*

PROOF. From our discussion in class, it is enough to show that if

$$R(h) \stackrel{\text{def}}{=} u(z_0 + h) - u(z_0) - [u_x(z_0)h_1 + u_y(z_0)h_2]$$

(for $h = (h_1, h_2) \in \mathbb{R}^2$ with $|h|$ sufficiently small, then $|R(h)| = \mathrm{o}(|h|)$ as $|h| \to 0$.

The first step is to write

$$u(z_0 + h) - u(z_0) = [u(x_0 + h_1, y_0 + h_2) - u(x_0, y_0 + h_2)] + [u(x_0, y_0 + h_2) - u(x_0, y_0)],$$

and then apply the (one-variable) Mean Value Theorem of differential calculus to each of the square-bracketed terms on the right. With this you see that the right-hand side of the equation above is:

$$u_x(x_0 + h_1, y_0 + t_1)h_1 + u_y(x_0, y_0 + t_2),$$

where t_1 lies between 0 and h_1 and t_2 lies between 0 and h_2. Thus $z_1 = (x_0+h_1, y_0+t_1)$ and $z_2 = (x_0, y_0 + t_2)$ both $\to z_0$ as $h \to 0$, and

$$R(h) = [u_x(z_1) - u_x(z_0)]h_1 + [u_y(z_2) - y_y(z_0)]h_2, \qquad (5)$$

for all h sufficiently small.

Let $\varepsilon_1(h)$ and $\varepsilon_2(h)$ denote, respectively, the two terms in square brackets on the right-hand side of (5) (these terms are functions of h because the points z_1 and z_2 depend only on h). The continuity of u_x and u_y at z_0 (used here for the first and only time) guarantees that that both $\varepsilon_1(h)$ and $\varepsilon_2(h) \to 0$ as $h \to 0$. Thus the same is true of

$$\varepsilon(h) \stackrel{\text{def}}{=} \max\{|\varepsilon_1(h)|, |\varepsilon_2(h)|\}.$$

Now an easy estimate starting with (5) yields

$$|R(h)| \le \varepsilon(h)|h|,$$

which shows that u is differentiable at z_0 because (as we just noted) $\varepsilon(h) \to 0$ as $h \to 0$. □

2.2 Exercise. Define "differentiable" for real-valued functions defined on open subsets of $\mathbb{R}^n$. State and prove a sufficient condition for differentiability of such functions that generalizes Theorem 2.1.

3 The "cosmic truth" about differentiation

Suppose U is an open subset of $\mathbb{R}^n$, and that $f : U \to \mathbb{R}^m$.

3.1 Definition. We say that f is *differentiable* at a point p_0 of U if there is a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(p_0 + h) = f(p_0) + Th + R(h) \quad \forall h \in \mathbb{R}^n \text{ sufficiently close to } p_0, \tag{6}$$

where $|R(h)| = \mathrm{o}(|h|)$ as $|h| \to 0$ in $\mathbb{R}^n$.

3.2 Notation. If f is differentiable at p_0 then the linear transformation T in the above definition is called the *derivative* of f at p_0. We will write $T = Df(p_0)$, preferring (almost always) to reserve the "prime" notation for the complex derivative.

3.3 Exercises. You should go back to the definition of differentiability for functions $R^2 \to \mathbb{R}$ and identify this linear transformation. Do the same for functions from an open subset of $\mathbb{R}^n$ to $\mathbb{R}$. Do the same for differentiable functions from intervals of the real line to $\mathbb{R}^2$, or more generally R^n (the so-called "vector-valued functions). Finally, how do you fit the definition of "complex differentiability" into this "linear transformation" context?

3.4 Exercise. *If f is differentiable at p_0 then f is continuous at p_0.*

3.5 The matrix of $\mathbf{Df(p_0)}$. Let $\{e_1, e_2, \ldots e_n\}$ be the standard unit vector basis for $\mathbb{R}^n$ (i.e., e_j is the vector with 1 in the j-th position and zeros elsewhere). Then upon fixing j and substituting $h = te_j$ into (6) and letting $t \to 0$, an argument entirely similar to the one we used in class for the case $n = 2, m = 1$ shows that

$$Te_j = \frac{\partial f}{\partial x_j}(p_0) = \left(\frac{\partial f_1}{\partial x_j}(p_0), \ldots , \frac{\partial f_m}{\partial x_j}(p_0)\right),$$

where $f_j : U \to \mathbb{R}$ is the j-th coordinate function of f:

$$f = (f_1, f_1, \ldots , f_n).$$

Thus the *matrix* of $Df(p_0)$ with respect to the standard bases in R_n and R_m respectively is the one whose j-th column is $\frac{\partial f}{\partial x_j}(p_0)$ (written as a column vector, rather than as the usual row vector). Let's call this matrix $[Df(p_0)]$. Thus $[Df(p_0)] = [\frac{\partial f_i}{\partial x_j}(p_0)]_{i=1,j=1}^{m,n}$. The image $Df(p_0)h$ of a vector $h \in \mathbb{R}^n$ is the m-dimensional vector found from the equation

$$[Df(p_0)h] = [Df(p_0)][h],$$

where square brackets around a *vector* denote the corresponding column matrix, i.e. the transpose of the original (row) vector.

3.6 The Chain Rule revisited. Now suppose that, in addition to the setup above, V is an open subset of $\mathbb{R}^m$ contained in $f(U)$, and $g : V \to \mathbb{R}^p$ is differentiable at $f(p_0)$. Then $g \circ f : U \to \mathbb{R}^p$ is differentiable at p_0, and

$$D(g \circ f)(p_0) = (Dg)(f(p_0))Df(p_0)$$

where the product between the derivatives on the right is the product of linear linear transformations, i.e. their *composition*.

3.7 Exercise. Adopt the argument of §1 to prove this version of the Chain Rule.

Suggestion: If you make a suitable definition of the *norm* $|T|$ of a linear transformation, say

$$|T| = \max\{|Tx| : x \in \mathbb{R}^n, |x| = 1\},$$

then the maximum in question exists (and is finite) because the map $x \to |Tx|$ is a continuous real-valued function on the (compact) unit sphere of $\mathbb{R}^n$, and you can easily prove that

$$|Tx| \leq |T||x| \qquad \forall x \in \mathbb{R}^n, \tag{7}$$

From this you should be able to write down a proof of the chain rule that is almost word-for-word the same as the one in §1. If you want a more concretely defined norm for linear transformations you can take $|T|$ to be the square root of sum of the squares of the entries of $[T]$, a quantity that is, in general, larger than the previously defined norm. For this one the Cauchy-Schwarz inequality gives (7).

3.8 Corollary *Suppose that:*

- $f : U \to \mathbb{R}^m$ *is differentiable at at each point of an open set* $U \subset \mathbb{R}^n$,
- I *is an open interval of the real line, and*
- $\gamma : I \to \mathbb{R}^n$ *is a differentiable function with* $\gamma(I) \subset U$.

Suppose $t_0 \in I$, v *is a nonzero vector in* R^n *that is tangent to the curve* γ *at* $\gamma(t_0)$. *Then* $Df(\gamma(t_0))\gamma(t_0)$ *is a vector tangent to the image-curve* $f \circ \gamma$ *at* $f(\gamma(t_0))$ (*as long as this vector is nonzero*).

Proof. A tangent vector to a curve γ at one of its points $\gamma(t_0)$ is just $\gamma'(t_0)$, which you can think of as a vector (with coordinates equal to the derivatives of the coordinate functions). (As a linear transformation $R \to \mathbb{R}^n$, this derivative would just be the map that takes $h \in \mathbb{R}$ to h times the tangent vector.) With this in hand, the Corollary becomes a restatement of the Chain Rule—try it! □

Corollary 3.8 says, roughly, that the linear map $Df(p_0)$ takes each vector tangent to a given curve through p_0 into a vector tangent to the image curve at $f(p_0)$. In this context, when we assert that $Df(p_0)$ is a linear transformation from $\mathbb{R}^n$ to R^n we should actually think of $\mathbb{R}^n$ as the space of vectors tangent to all possible differentiable curves through p_0, and $\mathbb{R}^m$ as the corresponding "tangent space" at $f(p_0)$. This point of view will show up again in the next section.

3.9 Exercise. Suppose G is an open subset of $\mathbb{R}^n$, $p_0 \in G$, and $u_j : G \to R$ is a real-valued function defined on G $(j = 1, 2, \dots, m)$. Then $f = (u_1, \dots, u_m) : G \to R^m$, and every R^m-valued function on G has this form. State and prove a sufficient condition for differentiability of f at p_0 that generalizes Theorem 2.1 to this situation. Suggestion: The problem quickly reduces to Exercise 2.2.

4 Conformality of the Stereographic Projection

We apply the ideas of the previous sections to prove that the stereographic projection is conformal in that differentiable curves in the plane that meet at a point z, get projected to curves on S^2 that meet at z^* and make the same angle (with the same sense) there.

4.1 Notation. For this discussion we denote:

- Points of $\mathbb{R}^3$ by (ξ, η, ζ), and those of $\mathbb{R}^2$ by (x, y),
- The North Pole $(0, 0, 1)$ of S^2 by N,
- The inner product (i.e. dot product) of two vectors v and w in the same Euclidean space by $\langle v, w\rangle$.

4.2 The "stereographic extension." Consider the natural extension σ of the stereographic projection to $R^3\backslash\{\zeta = 1\}$, defined by:

$$\sigma(\xi, \eta, \zeta) = (\frac{\xi}{1-\zeta}, \frac{\eta}{1-\zeta}). \tag{8}$$

By the sufficient condition of Exercise 3.9, σ is differentiable on $M \stackrel{\text{def}}{=} R^3\backslash\{\zeta = 1\}$, and by our work in class the matrix of the derivative of $\sigma = (\sigma_1, \sigma_2)$ at a point $p = (\xi, \eta, \zeta) \in M$ is obtained by placing the (vector) partial derivative of σ with respect to each coordinate down the respective columns of a two by three matrix:

$$[D\sigma(p)] = \begin{bmatrix} \frac{\partial\sigma_1}{\partial\xi}(p) & \frac{\partial\sigma_1}{\partial\eta}(p) & \frac{\partial\sigma_1}{\partial\zeta}(p) \\ \frac{\partial\sigma_2}{\partial\xi}(p) & \frac{\partial\sigma_2}{\partial\eta}(p) & \frac{\partial\sigma_2}{\partial\zeta}(p) \end{bmatrix} = \frac{1}{1-\zeta}\begin{bmatrix} 1 & 0 & \frac{\xi}{1-\zeta} \\ 0 & 1 & \frac{\eta}{1-\zeta} \end{bmatrix}$$

From our discussion of the chain rule, the problem is to show that, for every point $p = (\xi, \eta, \zeta) \in S^2\backslash\{N\}$, the linear transformation $D\sigma(p)$ preserves angles between vectors *tangent* to S^2 at p! In other words, if v and w are three-dimensional vectors tangent to S^2 at p (analytically: their inner products with p are both zero), then the angle from v to w is the same as the angle from $D\sigma(p)v$ to $D\sigma(p)w$.

4.3 Analytic statement of the problem. It's pretty clear that the stereographic projection preserves the *sense* of angles, so we will concentrate on preservation of *magnitudes* of angles. Here is the analytic expression of what needs to be done.

Suppose $p \in S^2\backslash\{N\}$ *and* $v, w \in \mathbb{R}^3$ *with*

$$\langle p,\, v\rangle = 0 \quad and \quad \langle p,\, w\rangle = 0, \tag{9}$$

(i.e. v and w are orthogonal to the line from the origin to p, and hence tangent to S^2 at p). Let $A = [D\sigma(p)]$, the matrix of $Df(p)$ with respect to the standard basis. We desire to show that

$$\frac{\langle Av,\, Aw\rangle}{|Av|\,|Aw|} = \frac{\langle v,\, w\rangle}{|v|\,|w|}. \tag{10}$$

The quantities that show up on the left and right hand sides of this last equation are, you will recall, the *cosines* of the angles between the respective pairs of vectors[1].

4.4 Exercise—reduction of problem. Show that in order to prove (10) you need only show that there is a positive constant c such that

$$\langle Av,\, Aw\rangle = c\langle v,\, w\rangle \tag{11}$$

for all $v, w \in \mathbb{R}^3$ satisfying (9).

In order to prove (11) we observe that

$$\langle Av,\, Aw\rangle = \langle A^*Av,\, w\rangle$$

where A^* is the linear transformation $R^2 \to R^3$ whose matrix is the transpose of the matrix of A. This is just a matrix calculation based on the fact that $\langle x,\, y\rangle = [x]^T[y]$ (matrix product), where x and y are any vectors in the same Euclidean space, $[x]$ and $[y]$ are their respective column vectors, and the superscript "T" denotes matrix transpose. Thus:

$$\begin{aligned}\langle Aw,\, Av\rangle &= [Aw]^T[Av] = ([A][w])^T[A][v] \\ &= [w]^T[A]^T[A][v] = [w]^T[A^*A][v] \\ &= \langle w,\, A^*Av\rangle\end{aligned}$$

from which the desired result follows by the symmetry of the inner product (it is unchanged if the order of its entries is reversed). Note that this calculation works as well for any $m \times n$ matrix and any pair of vectors in $\mathbb{R}^n$.

[1]Remember that the inner product on the left-hand side of the equation is the one for $\mathbb{R}^3$, and the one on the right is that of $\mathbb{R}^2$.

4.5 Computations with A*A. For $p = (\xi, \eta, \zeta) \in S^2 \backslash \{N\}$ let's write

$$x = \frac{\xi}{1-\zeta} \quad \text{and} \quad y = \frac{\eta}{1-\zeta}. \tag{12}$$

We use x and y simply as notational conveniences here, but nevertheless, note that $x + iy$ is the stereographic image, in the complex plane, of the point $p = (\xi, \eta, \zeta)$ of S^2. From the work of §4.2 the matrix of $A \overset{\text{def}}{=} D\sigma(p)$ can be written

$$[A] = \frac{1}{1-\zeta} \begin{bmatrix} 1 & 0 & x \\ 0 & 1 & y \end{bmatrix},$$

whereupon

$$[A^*A] = [A]^T[A] = c \begin{bmatrix} 1 & 0 & x \\ 0 & 1 & y \\ x & y & x^2+y^2 \end{bmatrix},$$

where $c = \frac{1}{(1-\zeta)^2}$.

Suppose for the moment that $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$ are any vectors in R^3. Then using the fact that $\langle Av,\, Aw\rangle = \langle A^*Av,\, w\rangle = [w]^T[A^*A][v]$ we obtain after a little calculation:

$$\begin{aligned} c^{-1}\langle Av,\, Aw\rangle &= \langle A^*Av,\, w\rangle = [w]^T[A^*A][v] \\ &= (v_1 + xv_3)w_1 + (v_2 + yv_3)w_2 + [xv_1 + yv_2 + (x^2+y^2)v_3]w_3 \\ &= v_1w_1 + v_2w_2 + v_3w_3 + (xw_1 + yw_2)v_3 + [xv_1 + yv_2 + (x^2+y^2-1)v_3]w_3 \\ &= \langle v,\, w\rangle + \Delta, \end{aligned}$$

where

$$\Delta = (xw_1 + yw_2)v_3 + [xv_1 + yv_2 + (x^2+y^2-1)v_3]w_3.$$

We claim that if v and w are tangent to S^2 at p then $\Delta = 0$, which will finish our proof. For this, go back to equation (12) describing x and y in terms of the coordinates of p, and note that

$$x^2 + y^2 - 1 = \frac{2\zeta}{1-\zeta},$$

hence

$$\begin{aligned} (1-\zeta)\Delta &= (\xi w_1 + \eta w_2)v_2 + (\xi v_1 + \eta v_2 + 2\zeta v_3)w_3 \\ &= (\xi w_1 + \eta w_2 + \zeta w_3)v_3 + (\xi v_1 + \eta v_2 + \zeta v_3)w_3 \\ &= \langle p,\, w\rangle v_3 + \langle p,\, v\rangle w_3 \end{aligned}$$

If v and w are tangent to S^2 at p, then the inner products in the last line are both zero, hence $\Delta = 0$ (since $p \neq N \Rightarrow \zeta \neq 1$). This completes the proof that the stereographic projection is conformal. □

5 The Mercator Projection.

This section borrows heavily from the beautiful book by Eli Maor: *Trigonometric Delights,* Princeton University Press, 1998, especially Chapters 13 and 14.

5.1 Mercator, conformality, and calculus. In order to make a map of the earth one has to address the problem of representing the sphere S^2 with some accuracy on a flat plane. The Flemish map maker Gerardus Mercator attacked the problem of making a map that would represent $S^2\backslash\{|\zeta| < 1-\varepsilon\}$, the unit sphere of $\mathbb{R}^3$ with equal spherical caps removed around the north and south poles, on a rectangle $[-\pi,\pi]\times[-h,h]$, where ε is some small positive number, and h is positive. Mercator wanted his map to have the following properties:

(a) The circles of latitude should be represented by horizontal lines of length $2h$, with the equator (latitude zero degrees) the horizontal axis of symmetry.

(b) Equally spaced circles of longitude (great circles on S^2 through the north and south poles) should be represented by equally spaced vertical lines.

(c) The correspondence between points of S^2 and points of Mercator's map should be *conformal.*

Conformality is the most important property—it guarantees that, at least in principal, a traveler wishing to go from point A on the globe to point B need only draw a line between the corresponding points on Mercator's map, measure the angle between this line and the vertical axis (true north), and using a compass, travel along a path that always makes the desired constant compass heading.

Such a map cannot, of course, accurately represent distances. Indeed, the circles of latitude on S^2 get smaller as their centers approach the poles, and the Mercator representation must stretch distances along these circles to make the corresponding horizontal lines the same length. This is why Greenland, for example, looks immense on a Mercator projection, when in reality it is not. In order to preserve conformality in the face of such horizontal stretching, Mercator had to correspondingly distort the distances between the lines of latitude: for example the vertical distance on Mercator's map between the lines representing latitude $15°$ and $30°$ will be larger than that between the equator and the line representing latitude $15°$. Mercator's great triumph was to figure out how to accomplish the vertical stretching that insures conformality. He published his map of the world in 1569, but unfortunately never explained his method, an omission that contributed to a certain skepticism about the value of his accomplishment.

In 1599 Edward Wright lifted the veil of mystery from Mercator's method, publishing an accurate account of its underlying mathematics. Nowadays the mathematical foundation of Mercator's feat reduces to an exercise in freshman calculus, as is described succinctly in Chapter 13 of the above-mentioned book by Maor. There it's shown how, in order for spherical rectangles to get mapped to "similar" plane rectangles, the height y of the line representing the circle of latitude λ on the sphere must obey the differential equation

$$dy = \sec\lambda\, d\lambda, \tag{13}$$

in other words,

$$y = \int_0^\lambda \sec\, t\, dt \tag{14}$$

Now Wright did not phrase his solution in terms of calculus—indeed, his work appeared long before Newton and Leibnitz fully developed the subject! What Wright did was solve the discrete analogue of (13) in increments of one degree. Nowadays we prove in freshman calculus (at least we used to, before Calculus "Reform") that

$$\int_0^\lambda \sec\, t\, dt = \ln\left|\tan\left(\frac{\lambda}{2}+\frac{\pi}{4}\right)\right|, \tag{15}$$

but note that it wasn't until 1614, fifteen years after Wright's treatise on Mercator's projection, that Napier published his invention of logarithms!

In 1645 Henry Bond, based on both the work of Wright and recently published tables of logarithms, conjectured (15), and this became one of the outstanding mathematical problems of the latter half of the seventeenth century. James Gregory proved (15) in 1668, but his proof was so complicated that it was viewed as suspicious at best. Two years later Isaac Barrow, Newton's predecessor at Cambridge, gave a comprehensible proof—essentially the same one you find in the calculus books of today. For this Barrow invented the technique of partial fraction decompositions, which he applied to the evaluation of many other integrals.

5.2 The Mercator Projection and the complex logarithm. The work we have done in these notes on conformality of the stereographic projection, the corresponding conformality of holomorphic functions done in class, and the holomorphicness of the Principal Branch of the logarithm function result in a quick solution of Mercator's problem:

(a) First note that the stereographic projection maps the spherical region $S_\varepsilon \stackrel{\text{def}}{=} S^2\backslash\{|\zeta| < 1-\varepsilon\}$ onto the annular open subset of $\mathbb{C}$ described by

$$A_R = \{z \in \mathbb{C} : 1/R < |z| < R\},$$

where $R = \sqrt{(2-\varepsilon)/\varepsilon} > 1$. This is a little exercise in right triangles, along with the formulas describing the stereographic projection, and one of the reflection results you obtained in the problem set about that projection. I leave it to you. The actual relationship between $\mathbb{R}$ and ε is not important: what *is* important is that the inner radius of the annulus is the reciprocal of the outer radius.

(b) Recall that, under the stereographic projection, the circles of longitude on the sphere go to rays in the plane, and the circles of latitude go to circles in the plane centered at the origin.

(c) The conformal map $w = \phi(z) = i\text{Log}\, z$ takes the slit annulus $A_R \backslash (-R, -1/R)$ one-to-one onto the rectangle $\mathcal{R}_h = [-\pi, \pi] \times [-\ln h, \ln h]$, taking the concentric circles that are images of the latitudes on the sphere onto horizontal lines in the rectangle, and the rays that are the images of the circles of longitude to vertical lines in the rectangle.

(d) Since both the stereographic projection σ and the logarithm map ϕ defined above are conformal, so is their composition $\phi \circ \sigma : S_\varepsilon \to \mathcal{R}_h$. *This* is the Mercator projection.

5.3 Position of the lines of latitude. A careful analysis of our representation $\phi \circ \sigma$ of the Mercator projection gives another way of seeing how the integral of the secant enters the picture. First note that the stereographic projection sends a point P^* of latitude λ on the sphere to a point P of the plane that lies $\tan(\frac{\lambda}{2} + \frac{\pi}{4})$ units from the origin. To see this, just look at the picture below, which takes place in the plane of P^* and the vertical axis of $\mathbb{R}^3$.

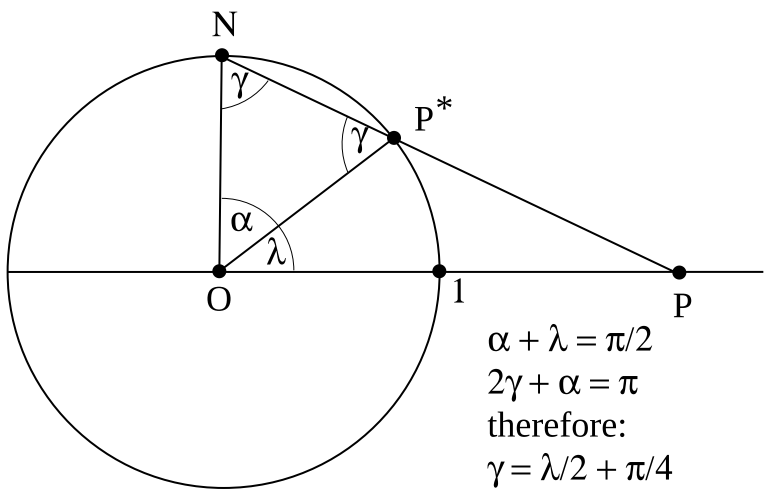

Figure 1: Latitude λ of P^* vs. magnitude of P

The distance in question, OP, is the tangent of the angle $\gamma = \angle ONP^*$ (vertex at N), but because the sides ON and OP^* of triangle ONP^* have the same length (namely, 1), you can easily see that $\gamma = \frac{\lambda}{2} + \frac{\pi}{4}$. We are interested in the imaginary part of $\phi(P)$, and this is just

$$\ln \left| \tan \left(\frac{\lambda}{2} + \frac{\pi}{4} \right) \right|,$$

which, as we noted earlier, equals $\int_0^\lambda \sec t\, dt$.