Chapter 6 Implicit function theorem

Chapter 5 has introduced us to the concept of manifolds of dimension m contained in $\mathbb{R}^n$. In the present chapter we are going to give the exact definition of such manifolds and also discuss the crucial theorem of the beginnings of this subject. The name of this theorem is the title of this chapter.

We definitely want to maintain the following point of view. An m-dimensional manifold $M \subset \mathbb{R}^n$ is an object which exists and has various geometric and calculus properties which are inherent to the manifold, and which should not depend on the particular mathematical formulation we use in describing the manifold. Since our goal is to do lots of calculus on M, we need to have formulas we can use in order to do this sort of work. In the very discussion of these methods we shall gain a clear and precise understanding of what a manifold actually is.

We have already done this sort of work in the preceding chapter in the case of hypermanifolds. There we discussed the intrinsic gradient and the fact that the tangent space at a point of such a manifold has dimension $n-1$ etc. We also discussed the version of the implicit function theorem that we needed for the discussion of hypermanifolds. We noticed at that time that we were really always working with only the **local** description of M, and that we didn't particularly care whether we were able to describe M with formulas that held for the entire manifold. The same will be true in the present chapter.

A. **Implicit presentation**

We go back to the idea of implicitly defined manifolds introduced briefly on p. 5–26. That is, we are going to have $n-m$ real-valued functions on $\mathbb{R}^n$, and the manifold will be (locally) the intersection of level sets of these functions. Let us denote

$$k = n - m;$$

this integer is frequently called the *codimension* of M. We are assuming $1 \le m \le n-1$, so that $1 \le k \le n-1$. (Hypermanifolds correspond to $k = 1$.) Let us call the "constraint" functions $g_1, \dots, g_k$. Then we are proposing that M be described locally as the set of all points $x \in \mathbb{R}^n$ satisfying the equations

$$\begin{cases} g_1(x) &= 0, \\ \quad\vdots & \\ g_k(x) &= 0. \end{cases}$$

Of course, since we want to do calculus we assume at least that each $g_i \in C^1$. But more is required. We must also have functions which are *independent* from one another. For instance, it would be ridiculous to think that the two equations

$$\begin{cases} x^2 + y^2 + z^2 & = 1, \\ 2x^2 + 2y^2 + 2z^2 & = 2, \end{cases}$$

define a 1-dimensional manifold in $\mathbb{R}^3$!

But more than the qualitative idea of independence is required. Even in the case $n = 2$ and $k = 1$ we have seen that a single C^1 equation $g(x_1, x_2) = 0$ does not necessarily give us a true manifold; see p. 5–2. We have learned in general that significant treatment of hypermanifolds (the case $k = 1$) requires that the gradient of the defining function be nonzero on M. The proper extension of this to our more general situation is that the gradients ∇g_i be *linearly independent* on M. This is the *quantitative* statement of the independence of the defining conditions.

Here are four examples with $n = 3$ and $k = 2$ which are in the spirit of p. 5–2, and are not 1-dimensional manifolds.

EXAMPLE 1. This is a rather artificial example, that comes from the example $x^3 - y^2 = 0$ on p. 5–2. Namely, let

$$\begin{cases} g_1(x, y, z) & = z + x^3 - y^2, \\ g_2(x, y, z) & = z. \end{cases}$$

The set given as $g_1 = g_2 = 0$ is just the curve $x^3 - y^2 = 0$ in the $x - y$ plane. This is not a manifold. Though

$$\nabla g_1 \ = \ (3x^2, -2y, 1) \quad \text{and} \quad \nabla g_2 \ = \ (0, 0, 1)$$

are never zero, they are linearly dependent at the point $(0, 0, 0)$.

EXAMPLE 2. A similar example, but one which is more geometrically significant, is given by

$$\begin{cases} g_1(x, y, z) & = xy - z, \\ g_2(x, y, z) & = z. \end{cases}$$

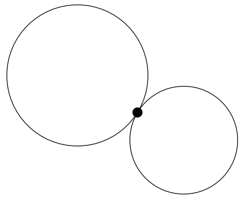
The set $g_1 = g_2 = 0$ is just given as $xy = 0$ in the $x - y$ plane, and this is the intersection of the two coordinate axes. This is not a manifold, the origin being the "bad" point. Again,

$$\nabla g_1 \ = \ (y, x, -1) \quad \text{and} \quad \nabla g_2 \ = \ (0, 0, 1)$$

are never zero, but are linearly dependent at the origin.

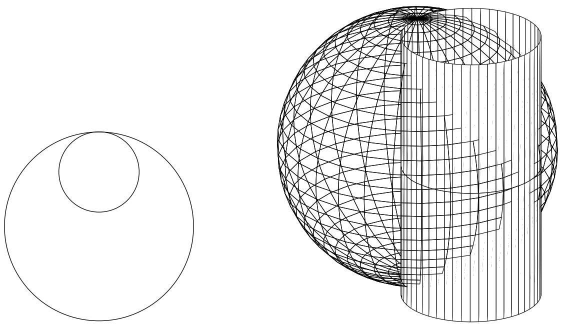
The nice surface $g_1 = 0$ is called a *hyperbolic paraboloid.* The nice surface $g_2 = 0$ is the $x - y$ plane. So this result shows that two nice surfaces may intersect in a bad "curve."

EXAMPLE 3. Here is a rather obvious example, but also it illustrates the point. Two spheres in $\mathbb{R}^3$ may intersect in a single point. Thus the intersection is not a 1-dimensional manifold.



Notice that it is geometrically clear that the two relevant gradients are linearly dependent at the bad point.

EXAMPLE 4. Here is quite an elegant example. We form the intersection of a certain sphere and right circular cylinder. Here are a "top" and 3-D view of the defining equations:



Notice that the gradients are

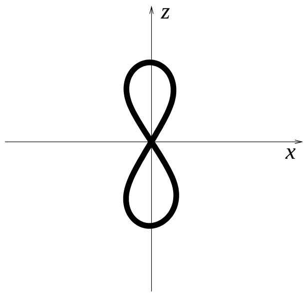
$$(2x, 2y, 2z) \quad \text{and} \quad (2x, 2y-1, 0),$$

and these are linearly dependent at $(0, 1, 0)$. So we expect "trouble" near $(0, 1, 0)$. Indeed, if we look at this intersection from far on the positive y axis and project onto the $x - z$ plane (like the tangent plane to the sphere and the cylinder), we see this equation:

$$x^2 + (1 - x^2 - z^2) - \sqrt{1 - x^2 - z^2} \;=\; 0;$$

solving this for x^2,

$$\begin{aligned} 1 - z^2 &= \sqrt{1 - x^2 - z^2}; \\ (1 - z^2)^2 &= 1 - x^2 - z^2; \\ x^2 &= z^2 - z^4 : \end{aligned}$$



This picture shows clearly the nonmanifold nature of the intersection of these two surfaces.

PROBLEM 6–1. Show that the length of the intersection curve just discussed is equal to

$$4\int_0^1 \sqrt{\frac{2-z^2}{1-z^2}}dz.$$

DEFINITION. Let $1 \le m \le n-1$ and let $M \subset \mathbb{R}^n$ be a set that we want to present as an m-dimensional manifold. Let $k = n - m$. Suppose that in a neighborhood of any point $x_0 \in M$ there are C^1 functions $g_1, \dots, g_k$ such that

(1) near x_0, $x \in M \iff g_i(x) = 0$ for $1 \le i \le k$,

and

(2) $\nabla g_1(x_0), \dots, \nabla g_k(x_0)$ are linearly independent.

Then we say that M is *implicitly presented.*

This condition of linear independence can be put in the form of a *Jacobian* matrix, as on p. 2–54. Namely, write the defining functions as a column vector

$$g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_k(x) \end{pmatrix},$$

so that $\mathbb{R}^n \xrightarrow{g} \mathbb{R}^k$. Then

$$Dg(x) = \begin{pmatrix} \partial g_1/\partial x_1 & \dots & \partial g_1/\partial x_n \\ \vdots & & \vdots \\ \partial g_k/\partial x_1 & \dots & \partial g_k/\partial x_n \end{pmatrix}$$

is the Jacobian matrix of g. It is a $k \times n$ matrix, and our condition is that at $x = x_0$ its *rows* are linearly independent.

In the context of matrix algebra, the largest number of linearly independent rows of a matrix A is called the *row rank* of A. Likewise for *column rank*. A relatively simple matrix algebra theorem asserts that always row rank = column rank. This is proved in the next section. Thus we would say that an $m \times n$ matrix has *maximal rank* if its row rank = column rank = $\min(m, n)$.

In these terms our condition is that the row rank of Dg is k. As $k < n$, the condition can be rephrased by saying that

$$Dg \;\text{ has maximal rank.}$$

Moreover, it is not hard to see that if the matrix $Dg(x_0)$ has maximal rank, then so does $Dg(x)$ for all x in some neighborhood of x_0. This is such simple yet important linear algebra that we pause for a discussion of this material.

B. **Rank**

We work with matrices with real entries, although the identical things hold for matrices with complex entries, or rational entries, etc.

DEFINITION. Let A be an $m \times n$ matrix and let the columns of A be $a_1, \dots, a_n \in \mathbb{R}^m$:

$$A = (a_1 \; a_2 \; \dots \; a_n).$$

The *column rank* of A is the dimension of the vector space spanned by the columns of A. This vector space is called the *column space* of A. The *row space* and *row rank* are defined similarly.

We can now immediately prove the theorem alluded to above:

THEOREM. *Row rank = column rank.*

PROOF. Let ℓ denote the column rank of A. Then the column space has dimension ℓ. Therefore there exist ℓ vectors $b_1, \dots, b_\ell \in \mathbb{R}^m$ such that each a_j is a linear combination of the form

$$a_j = \sum_{k=1}^{\ell} c_{kj} b_k, \quad 1 \le j \le n.$$

(By the way, the b_k's are linearly independent and thus the c_{kj}'s are unique, but none of that matters for the proof!) Rewrite these vector equations as equations for their i^{th} components:

$$a_{ij} = \sum_{k=1}^{\ell} b_{ik} c_{kj}, \quad 1 \le i \le m.$$

(We recognize this as a matrix product

$$\underset{m\times n}{A} = \underset{m\times k}{B} \ \underset{k\times n}{C}.)$$

Now we simply reinterpret these equations as representing the *rows* of A:

$$(a_{i1}, a_{i2}, \dots, a_{in}) = \sum_{k=1}^{\ell} b_{ik}(c_{k1}, c_{k2}, \dots, c_{kn}), \qquad 1 \le i \le m.$$

These equations show that the row space of A is spanned by the ℓ vectors $(c_{k1}, \dots, c_{kn})$. Thus the row rank of A is no greater than ℓ.

We have therefore proved that

$$\text{row rank of } A \le \text{column rank of } A.$$

Now just note that replacing A with A^t interchanges row rank and column rank.

QED

DEFINITION. *rank*(A) = column rank of A = row rank of A.

PROBLEM 6–2. Let A be a real $m \times n$ matrix. Prove that A has rank $\ge \ell \iff$ there exist rows numbered $i_1 < i_2 < \dots < i_\ell$ and columns numbered $j_1 < j_2 < \dots < j_\ell$ such that the square matrix B with entries

$$b_{\alpha\beta} = a_{i_\alpha j_\beta}$$

has nonzero determinant.
(HINT: use the row rank to find ℓ rows which are linearly independent. Discard the other rows to get an $\ell \times n$ matrix A' with row rank ℓ. Now use the fact that A' has column rank ℓ.)

PROBLEM 6–3. Prove that if the real $m \times n$ matrix A has rank $\ge \ell$, then all real $m \times n$ matrices whose entries are sufficiently close to those of A also have rank $\ge \ell$. In particular, if A has maximal rank, then all nearby matrices also have maximal rank.

PROBLEM 6–4. In the definition of Section A show that the second condition may be replaced with
(2) $\nabla g_1(x), \dots, \nabla g_k(x)$ are linearly independent for all x in some neighborhood of x_0.

PROBLEM 6–5. What connection is there among the numbers rank(A), rank(B), and rank(AB)?

PROBLEM 6–6. Let A be an $m \times n$ matrix. Prove that the following conditions are equivalent.

a. rank$(A) = 1$.

b. $A = (a_i b_j)$ where not all a_i are zero and not all b_j are zero.

c. $A = BC$ where B is a nonzero $m \times 1$ matrix and C is a nonzero $1 \times n$ matrix.

PROBLEM 6–7. Let a, b be nonzero vectors in $\mathbb{R}^n$, written as $n \times 1$ matrices (column vectors). Prove that the eigenvectors of the rank 1 matrix ab^t are the scalar multiples of a together with the orthogonal complement of b, and that the eigenvalues are $a \bullet b$, $0, \dots, 0$.

PROBLEM 6–8. Continuing with Problem 6–7, prove that the characteristic polynomial of ab^t is $(-1)^n \lambda^{n-1}(\lambda - a \bullet b)$. (Cf. Problem 3–41.)

C. **Implicit function theorem**

We have already discussed a version of this theorem in Section 5E. That version pertained to the case of hypermanifolds, in which $k = 1$ and there is just one function for which the equation $g(x) = 0$ needs to be "solved." The present situation calls for a direct generalization. The hypothesis and conclusion are by now easy to guess, and the easy guess is correct.

THEOREM. *Suppose $1 \le k \le n-1$, and suppose $\mathbb{R}^n \xrightarrow{g} \mathbb{R}^k$ is of class C^1 in a neighborhood of x_0, and*

$$g(x_0) = 0.$$

Suppose the Jacobian matrix $Dg(x_0)$ has maximal rank (equal to k). Then matrix algebra asserts that there are k columns of $Dg(x_0)$ which are linearly independent. For ease in writing, suppose these are the last k columns: columns $m+1$ through n. Write points in $\mathbb{R}^n$ in the symbolic fashion of $\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^k$ as

$$\begin{aligned} x &= (x_1, \dots, x_m;\ x_{m+1}, \dots, x_n) \\ &= (x'; x''). \end{aligned}$$

CONCLUSION: *there are k uniquely determined functions*

$$\mathbb{R}^m \xrightarrow{\varphi_i} \mathbb{R}, \qquad 1 \le i \le k,$$

of class C^1 near the point x_0', such that for all x in some neighborhood of x_0,

$$g(x) \;=\; 0 \Longleftrightarrow x_{m+i} \;=\; \varphi_i(x') \quad \text{for} \;\; 1 \le i \le k.$$

Moreover, the functions φ_i are as differentiable as g: if g is of class C^ℓ, then all φ_i are of class C^ℓ.

The conclusion of this theorem asserts that we have succeeded in “solving” the system of equations $g(x) = 0$ for k of the coordinates in terms of the other m coordinates. It’s again as though we have k equations in k “unknowns” $(x_{m+1}, \dots, x_n)$, which we are able to solve.

As mentioned already on p. 5–16, we do not prove this theorem in this course.

Its relevance is just as before: it is exactly the theorem we need to understand the manifold M determined by the implicit presentation

$$\begin{cases} g_1(x) &= \;0, \\ \quad\vdots & \\ g_k(x) &= \;0. \end{cases}$$

The essential nature of this theorem is that it guarantees that the linear approximation information contained in the Jacobian matrix $Dg(x_0)$ is decisive in describing the nonlinear situation contained in the function g near x_0, provided that the rank condition holds. The prototype is the basic calculus situation $y = f(x)$ in which we want to find an inverse function $x = f^{-1}(y)$. If f is affine it’s elementary provided $f' \ne 0$. And if $f' \ne 0$, then we are guaranteed a local solution $x = f^{-1}(y)$ exists, though we don’t have an explicit formula.

D. **Results**

We suppose that the manifold $M \subset \mathbb{R}^n$ satisfies everything required in Section B. We are now in a position to reap the many wonderful benefits of the implicit function theorem (“manifold” benefits).

TANGENT SPACE. We use the exact definition of Section 5F. Thus $T_{x_0}M$ is the set of all $\gamma'(0)$, for C^1 curves lying in M with $\gamma(0) = x_0$. The same sort of theorem as in Section 5F gives the implicit representation of $T_{x_0}M$. Namely,

$$h \in T_{x_0}M \Longleftrightarrow \nabla g_i(x_0) \bullet h \;=\; 0 \quad \text{for all} \;\; 1 \le i \le k.$$

Since the normal vectors $\nabla g_i(x_0)$ are linearly independent, this shows that $T_{x_0}M$ is a vector space of dimension m, as we would expect and desire.

INTRINSIC GRADIENT. If $M \xrightarrow{f} \mathbb{R}$ is defined *only* on M, then as before we can prove that there is a unique vector $\blacktriangledown f(x_0)$ in the tangent space $T_{x_0}M$ such that for all C^1 curves γ in M with $\gamma(0) = x_0$,

$$\frac{d}{dt} f(\gamma(t)) \;=\; \blacktriangledown f(x_0) \bullet \gamma'(0).$$

The other view point is that if $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$ is defined in the ambient space, then $\blacktriangledown f(x_0)$ is the orthogonal projection of the ambient gradient $\nabla f(x_0)$ onto the tangent space. That is, there are unique scalars $\lambda_1, \dots, \lambda_k$ such that

$$\blacktriangledown f(x_0) \;=\; \nabla f(x_0) - \sum_{i=1}^{k} \lambda_i \nabla g_i(x_0)$$

belongs to $T_{x_0}M$.

Of course we continue to call $\blacktriangledown f(x_0)$ the intrinsic gradient of f with respect to the manifold M.

LAGRANGE MULTIPLIERS. An *intrinsic critical point* of f is a point $x_0 \in M$ such that $\blacktriangledown f(x_0) = 0$. In other words, there exist scalars $\lambda_1, \dots, \lambda_k$, known as *Lagrange multipliers*, such that

$$\nabla f(x_0) \;=\; \sum_{i=1}^{k} \lambda_i \nabla g_i(x_0).$$

In practice this works in the usual way. There are $n+k$ "unknowns": $x_1, \dots, x_n; \lambda_1, \dots, \lambda_k$; and there are $n+k$ equations they have to satisfy — the n equations above together with the k constraints

$$\begin{cases} g_1(x) &= 0, \\ \quad\vdots & \\ g_k(x) &= 0. \end{cases}$$

PROBLEM 6–9. Find the intrinsic critical points of $x^4 + y^4 + z^4$ on the manifold $x^2 + y^2 + z^2 = 1$, $x + y + z = 1$.

(Solution: First, we note that this *is* a 1-manifold: we can recognize it as a "small circle" on the unit sphere containing the points $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$, or we can see that the gradients $(2x, 2y, 2z)$

and $(1,1,1)$ are linearly independent (since no point (s,s,s) belongs to the set). We may therefore use Lagrange multipliers. After adjusting constants, we have

$$\begin{cases} x^3 &= \lambda_1 x + \lambda_2, \\ y^3 &= \lambda_1 y + \lambda_2, \\ z^3 &= \lambda_1 z + \lambda_2, \\ x^2+y^2+z^2 &= 1, \\ x+y+z &= 1. \end{cases}$$

We already know $x=y=z$ to be impossible. Thus one coordinate must differ from the other two. We treat the case $z \neq x$ and $z \neq y$. Then eliminate λ_2 to obtain

$$\begin{cases} x^3 - z^3 &= \lambda_1(x-z), \\ y^3 - z^3 &= \lambda_1(y-z). \end{cases}$$

Divide by $x-z$ and $y-z$, respectively:

$$\begin{cases} x^2 + xz + z^2 &= \lambda_1, \\ y^2 + yz + z^2 &= \lambda_1. \end{cases}$$

Subtract:

$$\begin{aligned} x^2 - y^2 + xz - yz &= 0; \\ (x-y)(x+y+z) &= 0; \\ x-y &= 0. \end{aligned}$$

Thus $x=y$. The constraints imply

$$\begin{cases} 2x^2 + z^2 &= 1, \\ 2x + z &= 1. \end{cases}$$

Thus

$$\begin{aligned} 2x^2 + (1-2x)^2 &= 1; \\ 6x^2 - 4x &= 0; \\ x &= 0 \quad \text{or} \quad 2/3. \end{aligned}$$

We thus find $(0,0,1)$ and $(2/3,2/3,-1/3)$. By symmetry, the function has six intrinsic critical points:

$$\begin{array}{ll} (1,0,0), & (-1/3,2/3,2/3), \\ (0,1,0), & (2/3,-1/3,2/3), \\ (0,0,1), & (2/3,2/3,-1/3), . \end{array}$$

The three on the left are maxima ($f = 1$) and the three on the right are minima ($f = 11/27$).)

PROBLEM 6–10. Using the Lagrange method, find the intrinsic critical points of the function $f(x,y,z) = x$ on the "great circle" $x^2+y^2+z^2 = 1$, $x+y+z = 0$.

(Solution: We have

$$\begin{cases} 1 & = \lambda_1 x + \lambda_2, \\ 0 & = \lambda_1 y + \lambda_2, \\ 0 & = \lambda_1 z + \lambda_2, \\ x^2+y^2+z^2 & = 1, \\ x+y+z & = 0. \end{cases}$$

Thus subtraction gives $1 = \lambda_1(x-y) = \lambda_1(x-z)$. Thus $\lambda_1 \neq 0$ and we conclude $y = z$. The constraints give $x^2+2y^2 = 1$, $x = -2y$. Thus $6y^2 = 1$ and we obtain two critical points:

$$\pm\frac{1}{\sqrt{6}}(-2,1,1).\ )$$

PROBLEM 6–11. Find the extreme values of the distance between the origin $(0,0,0)$ and points of the first octant portion of the curve

$$\begin{cases} x+y+z & = 4, \\ xyz & = 2. \end{cases}$$

PROBLEM 6–12. Consider the ellipse

$$E : \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$

and the straight line

$$L : x_1 + x_2 = 1.$$

Find a necessary and sufficient condition on the semiaxes a and b such that E and L have no points in common. Given this condition, find the minimum distance between points of E and L. Do this by using Lagrange multipliers to find the minimum of the function

$$D(x_1, x_2, y_1, y_2) = (x_1 - y_1)^2 + (x_2 - y_2)^2,$$

where x and y are subject to the appropriate constraints.

PROBLEM 6–13. Continuing with the preceding problem, show that in addition to the minimum, there is another intrinsic critical point of the function D. What is the nature of this other critical point (minimum, local maximum, or saddle point)?

PROBLEM 6–14. Here is a generalization of Problem 6–12. Let A be an $n \times n$ symmetric positive definite matrix and let u be a nonzero vector in $\mathbb{R}^n$. Consider the ellipsoid

$$E = \{x \in \mathbb{R}^n \mid Ax \bullet x = 1\}$$

and the hyperplane

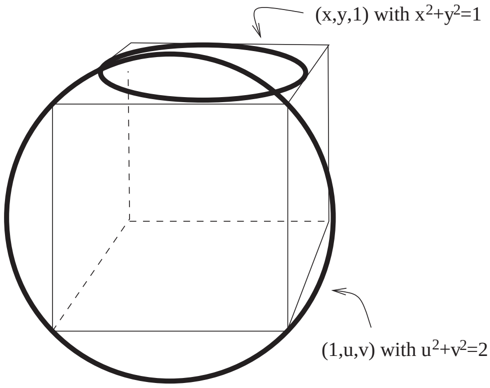
$$L = \{x \in \mathbb{R}^n \mid u \bullet x = 1\}.$$

a. Prove that $E \cap L = \emptyset \iff A^{-1}u \bullet u < 1$.
b. Find the minimum of $\|x - y\|$ for $x \in E$, $y \in L$.
(HINT: principal axis theorem.)

PROBLEM 6–15. Let $n \geq 3$ and find the maximum value of $\sum_{i=1}^n x_i^3$ on the sphere $\|x\| = 1$, $\sum_{i=1}^n x_i = 1$.

PROBLEM 6–16. (This problem is from *Mathematical Intelligencer*, vol 16, no. 4 (1994), p.8.)
A circle is inscribed in a face of a cube. Another circle is circumscribed about a neighboring face of the cube. Find the least and greatest distances between the points of the circles.

Please be kind to the grader standardizing the cube to have length 2 and be centered at the origin with sides parallel to the coordinate axes:



PROBLEM 6–17*. Here is a problem from *American Mathematical Monthly*, Volume 101, Number 6, June - July, 1994, *proposed by Murray S. Klamkin, University of Alberta, Edmonton, Alberta, Canada*:
Determine the extreme values of

$$f(x,y,z,u,v,w) = \frac{1}{1+x+u} + \frac{1}{1+y+v} + \frac{1}{1+z+w}$$

where $xyz = a^3$, $uvw = b^3$, and x, y, z, u, v, $w > 0$.

Clearly, $0 < f < 3$. The function f is defined on a 4-dimensional manifold M contained in $\mathbb{R}^6$. If you let $x \to 0$, $y \to 0$, $u \to \infty$, $v \to \infty$ on this manifold, you see that f assumes values arbitrarily close to 0. Thus f does not *attain* a minimum on M; rather, we say

$$\inf_M f = 0.$$

Also if you let $x \to 0$, $y \to 0$, $u \to 0$, $v \to 0$ on M, you see that f assumes values arbitrarily close to 2. Therefore,

$$2 \le \sup_M f \le 3.$$

Now here is my version of Klamkin's problem:

1. Find all the intrinsic critical points of f on M.
2. Calculate $\sup_M f$.
3. Investigate whether or not f *attains* the value $\sup_M f$ at some point of M.

PROBLEM 6–18. Find all the intrinsic critical points of the function $f(x) = x + y$ on the manifold described intrinsically as $x^2 + y^2 + z^2 = 49$ and $xyz = 24$. Also identify the local nature of each of these intrinsic critical points.

Cartesian product of manifolds. The three preceding problems provide an illustration of a very general construction of manifolds. In Problem 6–14, for instance, we are really dealing with the Cartesian product

$$E \times L = \{(x, y) \mid x \in E, y \in L\} \subset \mathbb{R}^n \times \mathbb{R}^n.$$

Thus $E \times L$ is a manifold of dimension $2n - 2$ contained in $\mathbb{R}^{2n}$, and the problem asks that we consider the function $\mathbb{R}^{2n} \xrightarrow{f} \mathbb{R}$ defined by

$$f(x, y) = \|x - y\|.$$

The general situation is quite easily understood. If $M_i \subset \mathbb{R}^{n_i}$ is a manifold of dimension m_i for $i = 1, 2$, then $M_1 \times M_2$ is a manifold of dimension $m_1 + m_2$ in $\mathbb{R}^{n_1+n_2}$. We saw an excellent example of these ideas in Section 5G where we introduced the flat torus.

E. **Parametric presentation**

We have seen many examples of this way of viewing a manifold. For instance, *curves* in $\mathbb{R}^n$ were actually defined this way on p. 2–3: as continuous functions $\mathbb{R} \xrightarrow{f} \mathbb{R}^n$. The variable t in the expression $x = f(t)$ is often called a *parameter*. In terms of manifolds, we are more interested in the image of f,

$$M = \{f(t) \mid t \in \mathbb{R}\},$$

and we need to make the usual sort of assumption that f is of class C^1 and $f'(t) \neq 0$. Furthermore, f needs to be one-to-one in order to have just one parameter value for each point of the curve. The fact that $t \in \mathbb{R}$ is what gives the image M its dimension of 1.

In the case of the tori of Section 5G, we have seen parametric presentations of each involving two parameters. The geometric idea is quite clear in general. We have a one-to-one function

$$\mathbb{R}^m \xrightarrow{f} \mathbb{R}^n,$$

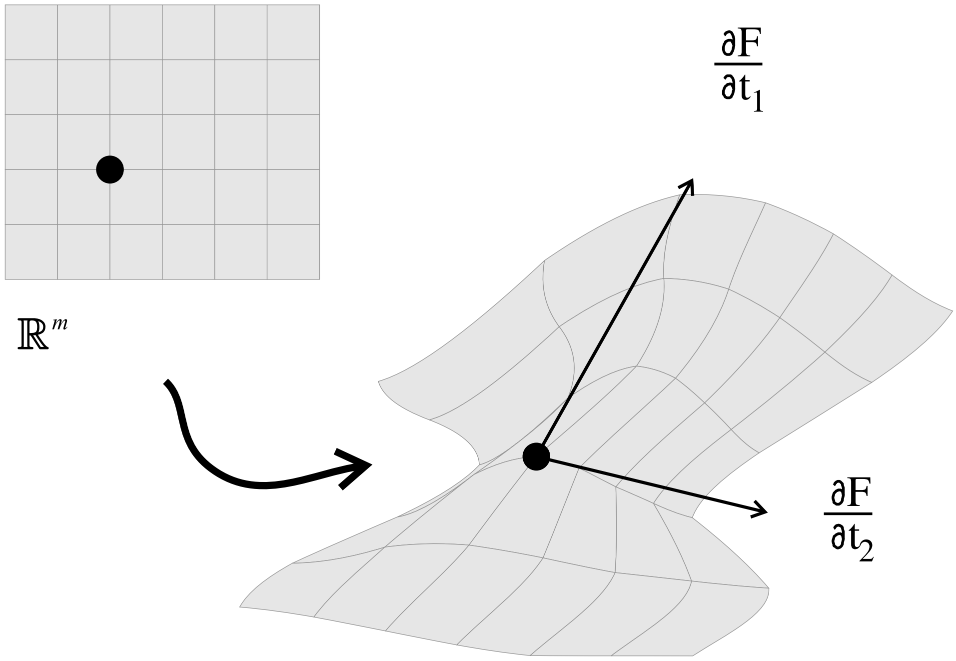
where the functional relation may be expressed

$$x = F(t) = F(t_1, \ldots, t_m).$$

We are anticipating an image

$$M = \{F(t) \mid t \in \mathbb{R}^m\}$$

which is an m-dimensional manifold. The idea is that the function F maps the "flat" parameter space $\mathbb{R}^m$ in a "curvy" manner into $\mathbb{R}^n$. The straight coordinate lines of $\mathbb{R}^m$ themselves are transported to curves which form a kind of "curvilinear" grid on M:



We shall require an important condition to insure that this provides an actual smooth manifold. That is, we require that the vectors in $\mathbb{R}^n$ given as

$$\frac{\partial F}{\partial t_1}, \ldots, \frac{\partial F}{\partial t_m}$$

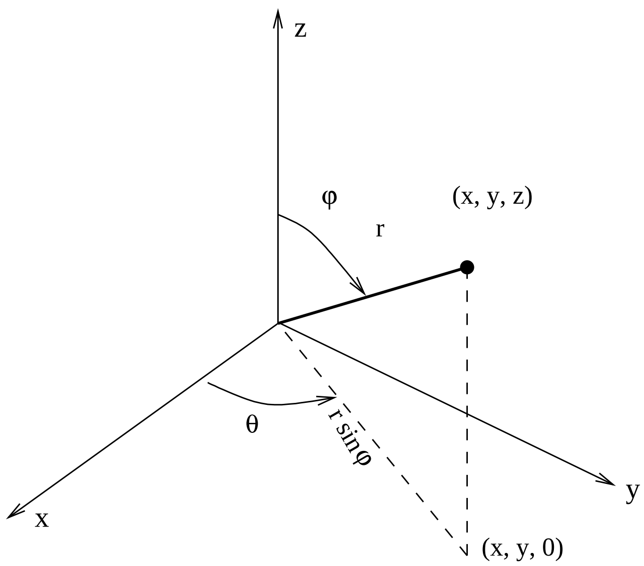
be linearly independent. . . lest the tangent lines in the curvilinear grid be tangent to one another.

The reasoning we have just given seems quite clear. If we imagine t_i as varying while all the other t_j's remain fixed, the function $F(t)$ gives a curve from $\mathbb{R}$ into $\mathbb{R}^n$, and the curve lies in the manifold M. Thus its derivative $\partial F/\partial t_i$ is tangent to M. Therefore if $x_0 = F(t_0)$ is a fixed point of M, then

$$\frac{\partial F}{\partial t_1}(t_0), \ldots, \frac{\partial F}{\partial t_m}(t_0) \in T_{x_0}M.$$

We demand $T_{x_0}M$ be an m-dimensional space, and thus the assumption of linear independence of the vectors $\partial F/\partial t_i$ is quite natural.

EXAMPLE. Spherical coordinates on the unit sphere in $\mathbb{R}^3$. We begin with the usual spherical coordinates for $\mathbb{R}^3$, which we describe with a sketch:



The formulas are

$$\begin{cases} x &= r\sin\varphi\cos\theta, \\ y &= r\sin\varphi\sin\theta, \\ z &= r\cos\varphi. \end{cases}$$

Here we leave out the z-axis in $\mathbb{R}^3$, and r is the norm of the vector (x, y, z); thus $0 < r < \infty$. We stress that $0 < \varphi < \pi$ only, but the usual polar coordinate θ is determined only up to additive integer multiples of 2π.

Warning. Many books denote the norm r by ρ. What is more troublesome is that many books, especially in physics, have the letters φ and θ interchanged. A little care needs to be taken to make sure of the notation used in any material, and this is only a minor inconvenience.

The unit sphere $M \subset \mathbb{R}^3$ is of course described by the restriction $r = 1$. Therefore we obtain spherical coordinates φ, θ for the sphere itself, the formulas being of course

$$\begin{cases} x &= \sin\varphi\cos\theta, \\ y &= \sin\varphi\sin\theta, \\ z &= \cos\varphi. \end{cases}$$

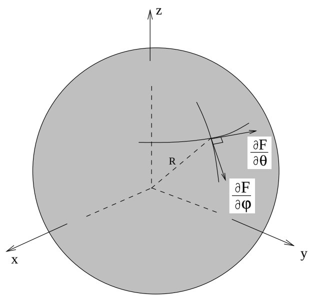
Besides the ambiguity in θ, the restriction $0 < \varphi < \pi$ leaves out the north pole $(0, 0, 1)$ and the south pole $(0, 0, -1)$. The corresponding tangent vectors are the φ and θ partial derivatives, respectively. Writing them as column vectors we have

$$\frac{\partial F}{\partial \varphi} = \begin{pmatrix} \cos\varphi\cos\theta \\ \cos\varphi\sin\theta \\ -\sin\varphi \end{pmatrix} \quad \text{and} \quad \frac{\partial F}{\partial \theta} = \begin{pmatrix} -\sin\varphi\sin\theta \\ \sin\varphi\cos\theta \\ 0 \end{pmatrix}.$$

Notice that these vectors satisfy the relations

$$\left\| \frac{\partial F}{\partial \varphi} \right\| = 1, \quad \left\| \frac{\partial F}{\partial \theta} \right\| = \sin\varphi, \quad \frac{\partial F}{\partial \varphi} \bullet \frac{\partial F}{\partial \theta} = 0.$$

Here's a sketch:



Notice also that these vectors are indeed orthogonal to the normal vector $\begin{pmatrix} \sin\varphi\cos\theta \\ \sin\varphi\sin\theta \\ \cos\varphi \end{pmatrix}$ to the surface (this was guaranteed to occur).

The spherical coordinate situation generalizes to $\mathbb{R}^n$. This is easily accomplished by induction on n, the general step being just like the transition from $\mathbb{R}^2$ to $\mathbb{R}^3$. Thus suppose $x \in \mathbb{R}^n$ is represented as $x = (x', x_n)$, where $x' \in \mathbb{R}^{n-1}$ and $x_n \in \mathbb{R}$. Suppose that $x' \neq 0$ *and* that we already know how to represent x' in spherical coordinates for $\mathbb{R}^{n-1}$. As usual, let

$$r = \|x\|.$$

Then

$$\frac{x_n}{r} = \frac{x_n}{\sqrt{\|x'\|^2 + x_n^2}}$$

and thus this quotient is in the open interval $(-1, 1)$. It is therefore equal to $\cos\varphi_{n-2}$ for a unique $0 < \varphi_{n-2} < \pi$. This is the "new" angle in the spherical coordinate system for $\mathbb{R}^n$. Then

$$\|x'\|^2 = r^2 - x_n^2 = r^2 - r^2\cos^2\varphi_{n-2} = r^2\sin^2\varphi_{n-2},$$

so we have

$$\|x'\| = r\sin\varphi_{n-2} > 0.$$

We then have the relevant angles $\varphi_1, \ldots, \varphi_{n-3}$, θ and formulas for x', using its norm $r\sin\varphi_{n-2}$.

PROBLEM 6–19. Show that points in $\mathbb{R}^4$ can be represented as

$$\begin{cases} x_1 &= r\sin\varphi_2\sin\varphi_1\cos\theta, \\ x_2 &= r\sin\varphi_2\sin\varphi_1\sin\theta, \\ x_3 &= r\sin\varphi_2\cos\varphi_1, \\ x_4 &= r\cos\varphi_2, \end{cases}$$

where $0 < \varphi_1 < \pi$, $0 < \varphi_2 < \pi$, $0 < r < \infty$, and θ is determined up to integer multiples of 2π. Show that the condition for $x \in \mathbb{R}^4$ to be so represented is simply $x_1^2 + x_2^2 \neq 0$. Show also that if $F(r, \varphi_1, \varphi_2, \theta)$ represents the above formulas, then the first order partial derivatives are mutually orthogonal, and

$$\begin{aligned} \left\|\frac{\partial F}{\partial r}\right\| &= 1, \\ \left\|\frac{\partial F}{\partial \varphi_2}\right\| &= r, \\ \left\|\frac{\partial F}{\partial \varphi_1}\right\| &= r\sin\varphi_2, \\ \left\|\frac{\partial F}{\partial \theta}\right\| &= r\sin\varphi_2\sin\varphi_1. \end{aligned}$$

PROBLEM 6–20. Consider a curve $\mathbb{R} \xrightarrow{f} \mathbb{R}^3$ of class C^1, represented in spherical coordinates by naming the spherical coordinates of $f(t)$ as $r(t)$, $\varphi(t)$, $\theta(t)$ (abuse of notation). Prove that

$$\|f'(t)\| = \sqrt{r'^2 + r^2\varphi'^2 + r^2 \sin^2\!\varphi\ \theta'^2}.$$

This relation is often expressed symbolically as

$$(ds)^2 = (dr)^2 + r^2(d\varphi)^2 + r^2 \sin^2\!\varphi\ (d\theta)^2.$$

PROBLEM 6–21. What is the corresponding result for a polar coordinate representation of a curve in $\mathbb{R}^2$?

PROBLEM 6–22. There is a particularly interesting class of curves on a sphere in $\mathbb{R}^3$. Assume you are dealing with the unit sphere $\|x\| = 1$, so that the two angles φ, θ serve as coordinates. Then of course

$$(ds)^2 = (d\varphi)^2 + \sin^2\varphi\ (d\theta)^2.$$

Let $0 < \alpha < \frac{\pi}{2}$ be fixed and consider a curve on the sphere "going north" at a constant angle α with the meridian. That is, if the curve is given as $f(t)$,

$$\frac{f'(t)}{\|f'(t)\|} \bullet (-\cos\varphi\cos\theta, -\cos\varphi\sin\theta, \sin\varphi) = \cos\alpha.$$

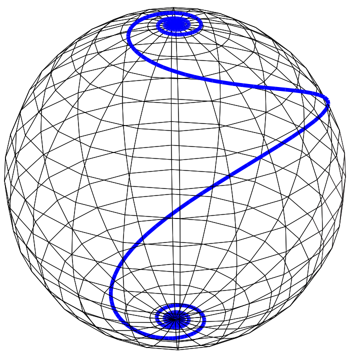
Such a curve is called a *loxodrome.*

a. Show that
$$\varphi' = -\cos\alpha\sqrt{\varphi'^2 + \sin^2\varphi\ \theta'^2}.$$

b. Show that
$$\varphi' = \pm\cot\alpha\sin\varphi\ \theta'.$$

c. Assume that $\varphi' < 0$ (going north) and $\theta' > 0$ and solve the differential equation in (b) to express θ as a function of φ.

d. Show that the loxodrome exists for $-\infty < \theta < \infty$ and that its total length is $\pi\sec\alpha$.



As we mentioned just before the example, the parametric representation of M gives a very nice understanding of the tangent space $T_{x_0}M$. For if $F(t_0) = x_0$, then the vectors $\partial F/\partial t_i(t_0)$ are in $T_{x_0}M$, an m-dimensional space. Thus the linear independence of these m vectors guarantees the situation that every vector in $T_{x_0}M$ is a unique linear combination of $\partial F/\partial t_1, \ldots, \partial F/\partial t_m$.

In Section C we discussed the implicit presentation of a manifold and the crucial role played by the Jacobian matrix of the defining function. The parametric representation we are now considering has a corresponding Jacobian matrix. Namely, in properly understanding the parameterizing function $\mathbb{R}^m \xrightarrow{f} \mathbb{R}^n$ we introduce the Jacobian matrix of F:

$$DF = \begin{pmatrix} \frac{\partial F_1}{\partial t_1} & \dots & \frac{\partial F_1}{\partial t_m} \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial t_1} & \dots & \frac{\partial F_n}{\partial t_m} \end{pmatrix}.$$

It is the m columns of this matrix that are required to be linearly independent. Thus DF has maximal rank.

In the notation of Section A, if M is also represented implicitly by $\mathbb{R}^n \xrightarrow{g} \mathbb{R}^k$ as the set of points satisfying $g(x) = 0$, we have

$$\mathbb{R}^m \xrightarrow{f} \mathbb{R}^n \xrightarrow{g} \mathbb{R}^k,$$
$$g \circ F = 0.$$

Thus the chain rule says exactly that

$$\underset{k\times n}{Dg(x_0)}\,\underset{n\times m}{DF(t_0)} = 0.$$

The implicit function theorem is once again important in the present context. It shows that a parametric presentation of M generates (in theory) an explicit presentation. We still work with the above notation $\mathbb{R}^m \xrightarrow{f} \mathbb{R}^n$.

We then introduce the new function

$$\mathbb{R}^m \times \mathbb{R}^n \xrightarrow{\tilde{F}} \mathbb{R}^n$$

defined by

$$\tilde{F}(t, x) = F(t) - x.$$

Note that $\tilde{F}(t_0, x_0) = 0$. Also note that the Jacobian matrix $D\tilde{F}$ has the $m + n$ columns,

$$\frac{\partial F}{\partial t_1}, \dots, \frac{\partial F}{\partial t_m}; \quad -\hat{e}_1, \dots, -\hat{e}_n.$$

The first m of these are linearly independent, by hypothesis. A standard linear algebra technique allows us to start with the vectors $-\hat{e}_1, \dots, -\hat{e}_n$, and replace some of them one at a time by all of $\partial F/\partial t_1, \dots, \partial F/\partial t_m$, and still have a basis for $\mathbb{R}^n$. For ease in writing, let us say that $-\hat{e}_1, \dots, -\hat{e}_m$ have been replaced. Then the columns

$$\frac{\partial F}{\partial t_1}, \dots, \frac{\partial F}{\partial t_m}; \quad -\hat{e}_{m+1}, \dots, -\hat{e}_n$$

are a basis for $\mathbb{R}^n$. Thus the implicit function theorem of Section C allows us to assert that the equation

$$F(t) - x \;=\; 0$$

can locally be "solved" for $t_1, \dots, t_m, x_{m+1}, \dots, x_n$ as functions of the other variables $x_1, \dots, x_m$. In particular, we obtain near x_0

$$x \in M \Longleftrightarrow x_{m+i} \;=\; \varphi_i(x_1, \dots, x_m) \qquad \text{for } \; 1 \le i \le k.$$

(The theorem also guarantees that the parameters $t_1, \dots, t_m$ are also functions of $x_1, \dots, x_m$, but we don't need that extra information here.) Thus the manifold is again represented explicitly.

Here is how the above analysis works in the case of a parametrized surface $M \subset \mathbb{R}^3$. The points of M might be represented as depending on parameters t_1 and t_2:

$$\begin{cases} x &= f(t_1, t_2), \\ y &= g(t_1, t_2), \\ z &= h(t_1, t_2). \end{cases} \tag{$*$}$$

We would assume that the corresponding Jacobian matrix have rank 2. For instance, suppose that

$$\det \begin{pmatrix} \frac{\partial f}{\partial t_1} & \frac{\partial f}{\partial t_2} \\ \frac{\partial g}{\partial t_1} & \frac{\partial g}{\partial t_2} \end{pmatrix} \neq 0.$$

Then we could in principle locally solve the first two equations in $(*)$ to produce for points in M,

$$\begin{cases} t_1 &= a(x, y), \\ t_2 &= b(x, y). \end{cases}$$

And then we have in turn

$$z = h\left(a(x, y), b(x, y)\right),$$

so that M is locally represented explicitly.

F. **The explicit bridge**

This section consists primarily of observations about what we have already achieved for an m-dimensional manifold $M \subset \mathbb{R}^n$. We have thoroughly discussed both an implicit description and a parametric description of M. A local *explicit* description presents M by giving $n-m$ of the coordinates of $\mathbb{R}^n$ as "explicit" functions of the other m coordinates. For instance, locally we might have the presentation

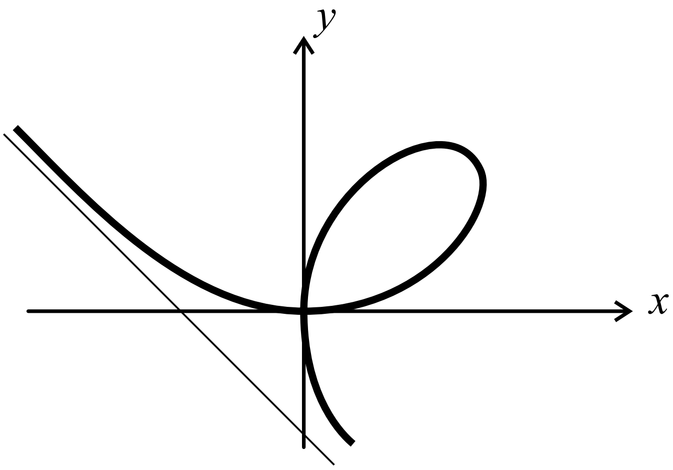
$$x \in M \Longleftrightarrow x_{m+i} \;=\; \varphi_i(x_1, \dots, x_m) \qquad \text{for} \;\; 1 \le i \le k = n-m.$$

This presentation is clearly both implicit and parametric, simultaneously. It is interesting that to go from a general implicit or parametric presentation of M to the opposite kind, we used the implicit function theorem in each case to find an explicit presentation. Thus all three types of presentation of M are equivalent, at least in theory.

The phrase "in theory" is of course quite important, as we may not be able to solve the equations in closed form, or we may not care to do so. The famous *folium of Descartes* is a case in point. This is a one-dimensional manifold (except for one bad point) in $\mathbb{R}^2$. Its natural and famous implicit presentation is this:

$$x^3 + y^3 \;=\; 3xy.$$

Notice the gradient of the defining function is $(3x^2 - 3y, 3y^2 - 3x)$, and this is nonzero on M except at the origin. Here's a plot of the folium:



Notice the singular point (0,0).

In this example a "trick" produces a parameterization. Namely, consider the intersection of the curve with the line $y = tx$ of slope t. Then

$$x^3 + t^3x^3 = 3xtx.$$

Discounting the origin, we can divide by x^2: $x + t^3x = 3t$. Thus

$$\begin{cases} x = \dfrac{3t}{1+t^3}, \\ y = \dfrac{3t^2}{1+t^3}. \end{cases}$$

This formula makes sense for all $t \neq -1$ and produces all points on the folium of Descartes. It is easy to check that the velocity vector $\left(\frac{dx}{dt}, \frac{dy}{dt}\right)$ is never 0. Thus for this manifold we are able to go from implicit to parametric presentation without stopping to find an explicit representation at all. This is quite fortunate, as solving $x^3 + y^3 = 3xy$ for y as a function of x, or solving $x = \frac{3t}{1+t^3}$ for t as a function of x are both tremendously difficult tasks, involving the solution of cubic equations.

PROBLEM 6–23. The sketch shows that the folium of Descartes has a vertical tangent at a point in the first quadrant.

a. Find this point by computing and using dx/dt.
b. Find this point by using Lagrange multipliers to find an intrinsic critical point of x on the folium.

PROBLEM 6–24. Consider the "figure 8 curve" of Problem 2–6, and its polar representation $r^2 = \cos 2\theta$. Show that the length of this curve equals

$$2\int_0^{\pi/2} \frac{dt}{\sqrt{\cos t}}.$$

PROBLEM 6–25. Consider the "cardioid" given in polar coordinates as $r = 1 + \cos\theta$. Sketch this curve in the $x - y$ plane and show that its length equals 8.

G. **The derivative reconsidered**

We can now give a brief discussion of a vast but easy generalization of the concept of derivative. Suppose that we have two manifolds N and M and a function f from N to M:

$$N \text{ is } n\text{-dimensional, } N \subset \mathbb{R}^q;$$
$$M \text{ is } m\text{-dimensional, } M \subset \mathbb{R}^p;$$

$$N \xrightarrow{f} M.$$

As we want only to introduce the concepts, we assume that f is differentiable, without actually giving the precise definition.

Then consider a point $x_0 \in N$ and the image $y_0 = f(x_0)$. We have the two tangent spaces at our disposal, $T_{x_0}N$ and $T_{y_0}M$. If we think of these as tangent vectors of curves, then we can see how f can be used to map $T_{x_0}N$ into $T_{y_0}M$.

Namely, suppose $h \in T_{x_0}N$. Then there is some curve $\mathbb{R} \xrightarrow{\gamma} \mathbb{R}^q$ such that

$$\begin{aligned} \gamma(t) &\in N \quad \text{for all } t, \\ \gamma(0) &= x_0, \\ \gamma'(0) &= h. \end{aligned}$$

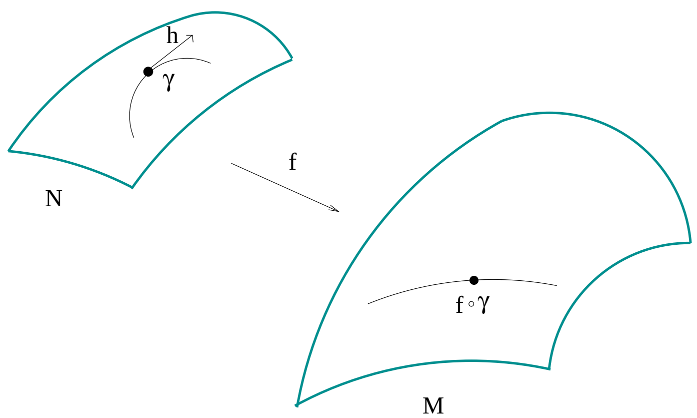
But then the composition $f \circ \gamma$ is a curve into M such that

$$f \circ \gamma(0) = y_0.$$

Thus its tangent vector

$$(f \circ \gamma)'(0)$$

is a vector in $T_{y_0}M$.



DEFINITION. In the above situation, $Df(x_0)$ is the mapping from $T_{x_0}N$ to $T_{y_0}M$ given by the formula

$$Df(x_0)(\gamma'(0)) = (f \circ \gamma)'(0).$$

That's the idea. Of course, we have not made this at all rigorous. We would have to define the concept of differentiability of f and also would have to show that the definition of $Df(x_0)(h)$ is independent of the choice of curve γ with $\gamma'(0) = h$. All of that can be done quite readily, and it also follows that $Df(x_0)$ is a *linear* mapping from $T_{x_0}N$ to $T_{y_0}M$.

Thus in a very precise sense $Df(x_0)$ represents an affine approximation to the original mapping f.

By the way, a more standard notation is $f_*(x_0) = Df(x_0)$, and $f_*(x_0)$ is said to "push forward" tangent vectors.

The next problem is an easy exercise illustrating this concept.

PROBLEM 6–26. Consider a cartographer's attempt to map a portion of the unit sphere $S(0,1) \subset \mathbb{R}^3$ onto a flat piece of paper $\mathbb{R}^2$:

$$S(0,1) \xrightarrow{f} \mathbb{R}^2.$$

Use standard spherical coordinates on $S(0,1)$ to write this in the form

$$f(\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi) = (u(\varphi,\theta), v(\varphi,\theta)).$$

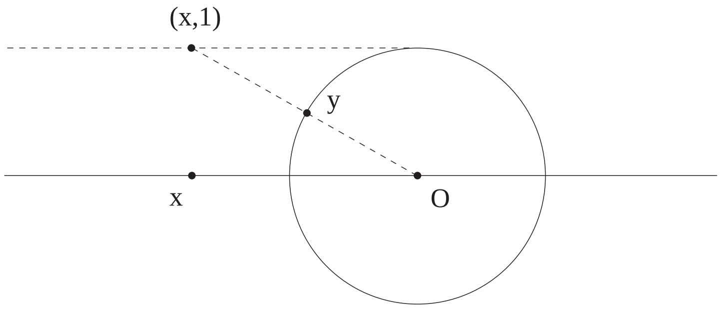
Prove that the linear map $Df(x_0,y_0,z_0)$ maps the orthonormal basis vectors as follows:

$$(\cos\varphi_0\cos\theta_0, \cos\varphi_0\sin\theta_0, -\sin\varphi_0) \quad \text{to} \quad (u_\varphi, v_\varphi)$$

and

$$(-\sin\theta_0, \cos\theta_0, 0) \quad \text{to} \quad \csc\varphi_0(u_\theta, v_\theta).$$

PROBLEM 6–27. An interesting cartographic mapping is the *gnomonic projection.* Here is one way to view it as a function from $\mathbb{R}^2$ to $S(0,1)$:



$$y = \frac{(x,1)}{\|(x,1)\|} = \frac{(x,1)}{\sqrt{\|x\|^2+1}}.$$

a. Show that the inverse of this mapping is given by

$$f(y) = \left(\frac{y_1}{y_3}, \frac{y_2}{y_3}\right).$$

b. Show that the corresponding functions in the preceding problem are

$$u(\varphi,\theta) = \tan\varphi\cos\theta,$$
$$v(\varphi,\theta) = \tan\varphi\sin\theta.$$

c. Prove that every great circle on the sphere gets mapped to a straight line in $\mathbb{R}^2$. (You can therefore imagine such a map might be useful in flying along great circle routes.)

H. **Conformal mapping**

We continue to consider a mapping $N \xrightarrow{f} M$ of one manifold into another, but now we assume that the two manifolds have the same dimension. One thing that could be asked is the question of what effect the mapping f has on *angles.* In particular, whether f *preserves* angles.

This really comes down to a question about the derivative $Df(x_0)$, which is a linear mapping from the tangent space $T_{x_0}N$ to $T_{f(x_0)}M$. We shall assume that $Df(x_0)$ is invertible, so that $h \neq 0 \Longrightarrow Df(x_0)h \neq 0$. Then the fact is that f preserves angles at x_0 if and only if

$$\frac{Df(x_0)h \bullet Df(x_0)h'}{\|Df(x_0)h\| \; \|Df(x_0)h'\|} = \frac{h \bullet h'}{\|h\| \; \|h'\|}$$

for all nonzero vectors h, $h' \in T_{x_0}N$. This is of course due to the fact that the cosines of the relevant angles are measured by the respective expressions.

To say the least, the verification of such an equation appears unwieldy. However some elementary linear algebra comes to our rescue. To see the relevance of the following result, notice that the linear mapping $Df(x_0)$ maps the vector space $T_{x_0}N$ into the vector space $T_{f(x_0)}M$ and these two vector spaces have the same dimension. We can therefore choose orthonormal basis for them and can thus represent $Df(x_0)$ by a square matrix. The following theorem about $n \times n$ matrices is therefore just what we need.

THEOREM. *Let A be a real invertible $n \times n$ matrix. Then the following conditions are equivalent.*

(1) A preserves angles.

(2) A preserves right angles.

(3) A is a positive scalar multiple of an orthogonal matrix.

(4) A maps some orthonormal basis of $\mathbb{R}^n$ to a nonzero scalar multiple of an orthnormal basis.

PROOF. Of course, (1) is exactly the sort of condition we are interested in. In symbols, it says that for any nonzero h and $h' \in \mathbb{R}^n$,

$$\frac{Ah \bullet Ah'}{\|Ah\|\;\|Ah'\|} = \frac{h \bullet h'}{\|h\|\;\|h'\|}.$$

Our strategy is to prove that (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (1). All of these are straightforward except for (2) $\Longrightarrow$ (3).

(1) $\Longrightarrow$ (2) is clear, as (2) means that $h \bullet h' = 0 \Longrightarrow Ah \bullet Ah' = 0$.

(2) $\Longrightarrow$ (3) requires some calculation. If h, h' are arbitrary nonzero vectors, then h is orthogonal to

$$h' - \frac{h' \bullet h}{\|h\|^2} h$$

(remember p. 1–14?), so (2) gives

$$Ah \bullet \left(Ah' - \frac{h' \bullet h}{\|h\|^2} Ah \right) = 0.$$

Thus

$$Ah \bullet Ah' = \frac{h' \bullet h}{\|h\|^2} \|Ah\|^2.$$

Interchange h and h':

$$Ah' \bullet Ah = \frac{h \bullet h'}{\|h'\|^2}\|Ah'\|^2.$$

We conclude that

$$h' \bullet h \left(\frac{\|Ah\|^2}{\|h\|^2} - \frac{\|Ah'\|^2}{\|h'\|^2} \right) = 0.$$

Therefore if $h' \bullet h \neq 0$,

$$\frac{\|Ah\|}{\|h\|} = \frac{\|Ah'\|}{\|h'\|}.$$

In case $h' \bullet h = 0$, then we can use the equation we have just proved for the vectors h and $h + h'$ and then also for the vectors h' and $h + h'$, and conclude that it still holds. Thus, for *all* nonzero h and $h' \in \mathbb{R}^n$,

$$\frac{\|Ah\|}{\|h\|} = \frac{\|Ah'\|}{\|h'\|}.$$

Call this common ratio c. As A is invertible, $c > 0$. Thus

$$\|c^{-1}Ah\| = \|h\| \quad \text{for all} \quad h \in \mathbb{R}^n.$$

According to Problem 4–20f, $c^{-1}A = \Phi \in O(n)$. Thus $A = c\Phi$ and (3) is verified.

(3) $\Longrightarrow$ (4) is clear, for if $\{\hat{\varphi}_i\}$ is an orthonormal basis, then so is $\{\Phi\hat{\varphi}_i\}$ for any $\Phi \in O(n)$.

(4) $\Longrightarrow$ (1). Let $\{\hat{\varphi}_i\}$ be the particular orthonormal basis such that $\{cA\hat{\varphi}_i\}$ is also an orthonormal basis. Let Φ and Ψ be the corresponding orthogonal matrices having those columns:

$$\begin{aligned} \Phi &= (\hat{\varphi}_1 \; \hat{\varphi}_2 \; \ldots \; \hat{\varphi}_n), \\ \Psi &= c(A\hat{\varphi}_1 \; A\hat{\varphi}_2 \; \ldots \; A\hat{\varphi}_n). \end{aligned}$$

Then we have the matrix equation

$$\Psi = cA\Phi,$$

so that

$$A = c^{-1}\Psi\Phi^{-1}.$$

Thus A is a scalar multiple of an orthogonal matrix and we conclude that A preserves angles.

QED

Now we return to our discussion of $N \xrightarrow{f} M$. Our question about the preservation of angles by f now is answered by the condition that at each x_0 the derivative $Df(x_0)$ is a positive scalar multiple of an "orthogonal matrix." That is, according to Problem 4–20f, we need

$$\|Df(x_0)h\| = c(x_0)\|h\| \quad \text{for all} \quad h \in T_{x_0}N.$$

The *local magnification factor* $c(x_0)$ is of course allowed to vary with x_0. Here is the summary

DEFINITION. Suppose $N \xrightarrow{f} M$ is a C^1 map from an n-dimensional manifold N to an n-dimensional manifold M. Then f is a *conformal mapping* if it preserves angles. That is, if for every $x \in N$ there corresponds a *local magnification factor* $c(x) > 0$ such that

$$\|Df(x)h\| = c(x)\|h\| \quad \text{for all} \quad h \in T_x N.$$

For instance, a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ has the form $x \to Ax$ for a certain $n \times n$ matrix A. This mapping is conformal if and only if A is a positive scalar multiple of an orthogonal matrix, and the magnification factor is that scalar.

In the general case the conformal function f will introduce some "distortion" in the way it moves points of N to M. Though f distorts lengths, it leaves all angles exactly unchanged. If it so happens that the local magnification factor $c(x)$ is *constant*, then the mapping introduces no distortion at all. In particular if $c(x) = 1$ for all x, we would even say that f is an *isometry*.

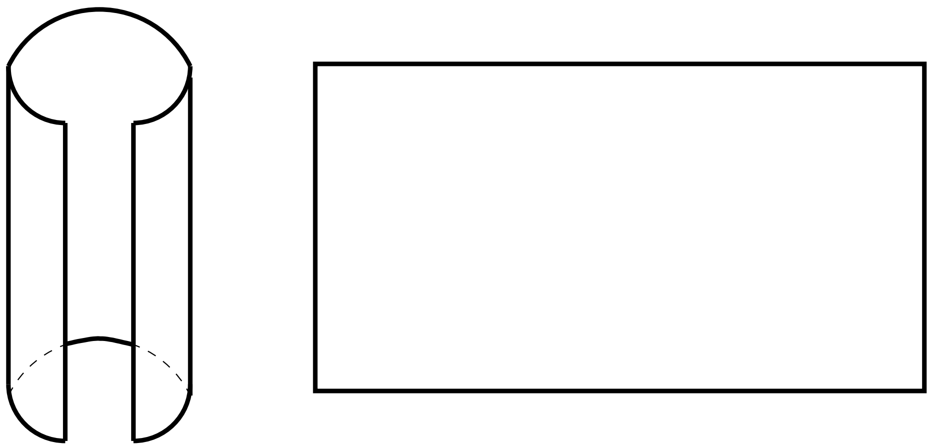
For example, the right circular cylinder $x^2 + y^2 = a^2$ in $\mathbb{R}^3$ is locally isometric to $\mathbb{R}^2$. This can of course be readily seen from a polar coordinate representation of the cylinder by defining

$$f(a\cos\theta, a\sin\theta, z) = (a\theta, z), \quad 0 < \theta < 2\pi, \ -\infty < z < \infty.$$

Then

$$\begin{aligned} Df(a\cos\theta, a\sin\theta, z)(-a\sin\theta, a\cos\theta, 0) &= (a, 0), \\ Df(a\cos\theta, a\sin\theta, z)(0, 0, 1) &= (0, 1). \end{aligned}$$

Thus Df maps the orthonormal basis $(-\sin\theta, \cos\theta, 0)$, $(0, 0, 1)$ (of the tangent space to the cylinder) to the orthonormal basis $\hat{\imath}$, $\hat{\jmath}$ (of $\mathbb{R}^2$). The picture to keep in mind is that of cutting a paper cylinder along a straight line parallel to its axis and unrolling it to make it flat.



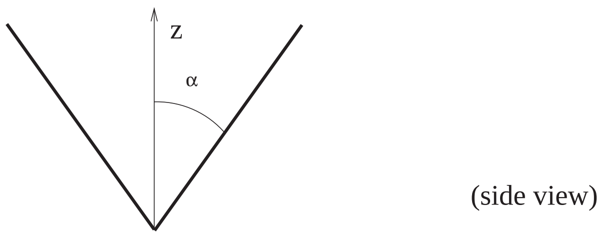
This mapping f is thus an isometry.

PROBLEM 6–28. Any right cylinder, circular or not, can be isometrically flattened. Show this by using a curve γ in $\mathbb{R}^2$ which has unit speed and defining the cylinder to be the set of points of the form

$$C = \{(\gamma(t), z) \mid a < t < b,\ -\infty < z < \infty\}.$$

The choice of $C \xrightarrow{f} \mathbb{R}^2$ should now be clear.

PROBLEM 6–29. A right circular cone can also be flattened isometrically (locally). Show this by using a cone of the form shown here:

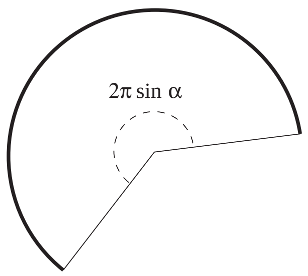


$z = \cot\alpha\sqrt{x^2+y^2}$

Use cylindrical coordinates for $\mathbb{R}^3$ so that the cone has the parametric representation

$$\begin{cases} x &= z\tan\alpha\cos\theta, \\ y &= z\tan\alpha\sin\theta, \\ z &= z. \end{cases}$$

Then "cut" the cone along $\theta = 0$ and flatten it out into the form



with the mapping from the cone to $\mathbb{R}^2$:

$$f(z\tan\alpha\cos\theta, z\tan\alpha\sin\theta, z) = z\sec\alpha(\cos(\theta\sin\alpha), \sin(\theta\sin\alpha)),$$

Prove that f is an isometry.

PROBLEM 6–30. Consider the two points $(3,4,5)$ and $(4,3,5)$ on the cone $z = \sqrt{x^2+y^2}$. What is the smallest possible length of curves joining these two points and lying entirely on the cone? (Use a calculator.)
[Answer: 1.41659... .] Notice that it's greater than $\sqrt{2}$. Comment?

Notice very particularly that this flattening of the cone preserves all angles. We are excluding the vertex, and, indeed, angles seem to be distorted there.

I. **Examples of conformal mappings**

We now present important classical instances of conformal mappings, all of which are extremely important in many contexts.

1. **Holomorphic functions**

Surely the most basic possible situation is that of a function $\mathbb{R}^2 \xrightarrow{f} \mathbb{R}^2$. Both manifolds are the Euclidean plane. Let us denote the Cartesian representation of f as

$$f(x,y) = (u(x,y),\ v(x,y)).$$

Then the derivative $Df(x,y)$ maps the standard orthonormal basis as follows:

$$\begin{aligned} Df(x,y)\hat{\imath} &= (u_x,\ v_x), \\ Df(x,y)\hat{\jmath} &= (u_y,\ v_y). \end{aligned}$$

Thus the condition that f be conformal is that the vectors (u_x, v_x) and (u_y, v_y) have the same (nonzero) norm and be orthogonal:

$$\begin{aligned} u_x^2 + v_x^2 &= u_y^2 + v_y^2 \neq 0, \\ u_x u_y + v_x v_y &= 0. \end{aligned}$$

If for example $u_x \neq 0$, then we have $u_y = -v_x v_y / u_x$, so if we write $t = v_y/u_x$ then we have

$$\begin{aligned} tu_x &= v_y, \\ u_y &= -tv_x. \end{aligned}$$

And then

$$\begin{aligned} u_x^2 + v_x^2 &= t^2 v_x^2 + t^2 u_x^2 \\ &= t^2(u_x^2 + v_x^2). \end{aligned}$$

Thus $1 = t^2$ and we have two cases. If $t = 1$, then the condition is

$$\begin{cases} \frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y}, \\ \frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x}. \end{cases}$$

These equations are known as the ***Cauchy − Riemann equations***. They are surely the most important partial differential equations in all of mathematics. A solution of them is said to be a *holomorphic* function; rather, the corresponding complex-valued function

$$f(x+iy) = u(x,y) + iv(x,y)$$

is said to be holomorphic (or *analytic*). As we have seen, in case $u_x^2 + u_y^2 \neq 0$, f is also conformal.

The other case $t = -1$ is entirely similar, the corresponding equations being $u_x = -v_y$, $u_y = v_x$.

A good example is the function which represents the square of a complex number. As $(x+iy)^2 = x^2 + 2ixy - y^2$, we define

$$f(x,y) = (x^2 - y^2, 2xy).$$

This is seen to be holomorphic, with local magnification factor $2\sqrt{x^2+y^2}$. This mapping is thus conformal except at the origin. Indeed, angles are *doubled* at the origin.

The study of holomorphic functions has been going on for centuries. In fact all universities devote at least one entire semester to this subject, in courses titled "complex analysis," or "complex variables," or "analytic functions," etc. At Rice these come as either MATH 382 or MATH 427. Both are highly recommended.

2. **Inversion in a sphere**

We have already looked at this in Problems 2–79, 80, 81. The function $\mathbb{R}^n \xrightarrow{f} \mathbb{R}^n$ is given by

$$f(x) = \frac{x}{\|x\|^2}.$$

Its derivative is

$$Df(x) = \|x\|^{-2}I - 2\|x\|^{-4}(x_i x_j),$$

and we have the result

$$\|Df(x)h\| = \|x\|^{-2}\|h\|.$$

Thus f is conformal with the local magnification factor $\|x\|^{-2}$.

PROBLEM 6–31. As a possible generalization, let $\mathbb{R}^n \xrightarrow{f} \mathbb{R}^n$ be given in the form

$$f(x) = g(\|x\|)x.$$

Compute $Df(x)$ and show that f is conformal $\iff$ $g'(t) = 0$ or $g'(t) + 2g(t)/t = 0$. Conclude that g is constant or $g(t) = ct^{-2}$.

PROBLEM 6–32. We can perform inversion in any sphere. If the sphere is centered at x_0 and has radius R, the correct formula is

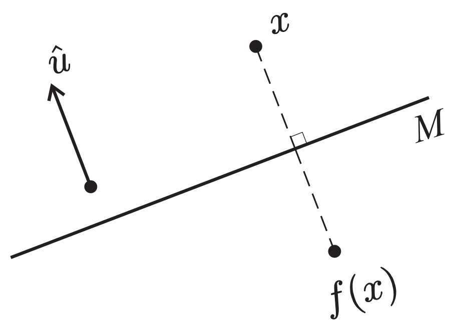
$$f(x) = x_0 + R^2 \frac{x - x_0}{\|x - x_0\|^2}.$$

Prove that the mapping is conformal. What is the local magnification factor?

PROBLEM 6–33. Reflection across a "plane" in $\mathbb{R}^n$ is a somewhat simpler sort of mapping than inversion in a sphere. Suppose the "plane" (really, *hyperplane*) to be described as follows: let $\hat{u}$ be a unit vector in $\mathbb{R}^n$ and $c \in \mathbb{R}$, and define

$$M = \{x \in \mathbb{R}^n | x \bullet \hat{u} = c\}.$$

Schematic figure:



Then we map any $x \in \mathbb{R}^n$ to its "mirror image" $f(x)$ with respect to M. That is, we determine $t \in \mathbb{R}$ such that

$$f(x) = x + t\hat{u}$$
$$\frac{f(x) + x}{2} \in M.$$

a. Prove that $f(x) = x + 2(c - x \bullet \hat{u})\hat{u}$.

b. Prove that
$$Df(x; h) = h - 2h \bullet \hat{u}\hat{u}.$$

c. Prove that f is conformal.

3. **Mercator projection**

This is a cartographic exercise coming from the desire to map a portion of the unit sphere $S(0, 1) \subset \mathbb{R}^3$ in a certain way onto a sheet of paper $\mathbb{R}^2$, so that the meridians $\theta =$ constant get mapped into equally spaced parallel lines and the mapping is conformal. Thus, using standard spherical coordinates for the sphere, we are led to $S(0,1) \xrightarrow{f} \mathbb{R}^2$ of the form

$$f(\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi) = (\theta, g(\varphi)),$$

where g is to be determined. Then we compute that the linear mapping Df has to transform the corresponding unit tangent vectors as follows:

$$\begin{aligned} &(\cos\varphi\cos\theta, \cos\varphi\sin\theta, -\sin\varphi) \longmapsto (0, g'(\varphi)), \\ &(-\sin\theta, \cos\theta, 0) \longmapsto \csc\varphi\ (1,0). \end{aligned}$$

The resulting two vectors are orthogonal, so the condition for conformality is that they have the same norm:

$$|g'(\varphi)| = \csc\varphi.$$

PROBLEM 6–34. Assume that g increases as the north pole is approached (φ decreases to 0) and that the equator $\varphi = \frac{\pi}{2}$ gets mapped to the θ-axis. Then show that

$$g(\varphi) = \log(\csc\varphi + \cot\varphi) = \log\cot\frac{\varphi}{2}.$$

This formula gives the famous *Mercator projection*:

$$f(\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi) = (\theta, \log\cot\frac{\varphi}{2}).$$

PROBLEM 6–35. Show that the local magnification factor of the Mercator projection is $\csc\varphi$.

PROBLEM 6–36. Show that the loxodromes (Problem 6–22) on the sphere, which are the curves that maintain a constant angle with the meridians, correspond to straight lines under the Mercator projection.

This problem illustrates one navigational use of a Mercator projection. If you want to sail from A to B on $S(0,1)$, use your Mercator map to locate A and B, connect them with a straight line, then sail at that constant bearing. The resulting journey will of course be longer than the great circle path.

PROBLEM 6–37. The coordinates of Houston, Texas, are 29.97°N latitude, 95.35°W longitude. The coordinates of Samara, Russia, are 53.23°N and 50.17°E. Assume the earth is a ball with radius 3960 miles.

a. Find the great circle distance between these cities.
b. Find the “loxodromic distance” between them.

PROBLEM 6–38. Consider a “loxodromic” “triangle” on the unit sphere. Explain what this should mean, and prove that the sum of the interior angles of such a “triangle” equals π.

4. **Stereographic projection**

In this cartographic exercise we again map

$$S(0,1) \xrightarrow{f} \mathbb{R}^2,$$

but now the desire is that the meridians $\theta =$ constant near the north pole themselves get mapped to the straight lines $\theta =$ constant in a standard polar coordinate representation of $\mathbb{R}^2$. Thus we have for an as yet unknown polar coordinate $r(\varphi)$,

$$f(\sin\varphi\cos\theta, \sin\varphi\sin\theta, \cos\varphi) = (r(\varphi)\cos\theta, r(\varphi)\sin\theta).$$

As we want the north pole to map to the origin, we assume $r(0) = 0$.

PROBLEM 6–39. Show that this mapping f is conformal $\Longleftrightarrow$

$$\frac{dr}{d\varphi} = r\csc\varphi.$$

Conclude that

$$r(\varphi) = \frac{c}{\csc\varphi + \cot\varphi}$$

for some positive constant c.

PROBLEM 6–40. Take $c = 1$ in the preceding problem. Let points in $S(0,1)$ be represented in Cartesian coordinates as $y = (y_1, y_2, y_3)$, and let

$$f(y) = (x_1, x_2).$$

Show that

$$x_1 = \frac{y_1}{1 + y_3},$$
$$x_2 = \frac{y_2}{1 + y_3}.$$

The function f of this problem is called *stereographic projection* of the unit sphere onto $\mathbb{R}^2$.

PROBLEM 6–41. Stereographic projection provides a nice result concerning loxodromes. Consider a loxodrome as described in Problem 6–22, described by the relation in spherical coordinates,

$$\theta \cot \alpha = \log(\csc \varphi + \cot \varphi).$$

Using stereographic projection, this loxodrome becomes a curve in $\mathbb{R}^2$. Show that the equation of this projected curve in the usual polar coordinates for $\mathbb{R}^2$ is

$$r = e^{-\theta \cot \alpha}.$$

This plane is called a *logarithmic spiral.*

REMARK. Martin Gardner's delightful book, *aha*! *Insight*, Scientific American 1978, has an interesting discussion in the Section "Payoff at the Poles." Here's a quotation from p. 43: "Here is a different navigational problem that involves a fascinating curve on the sphere known as a loxodrome or rhumb-line. ... A loxodrome, plotted on a flat map, has different forms depending on the type of map projection. On the familiar world map called the Mercator projection, it is plotted as a straight line. Indeed, this is why a Mercator map is so useful to navigators. ... When a loxodrome is projected on a plane parallel to the equator and tangent to a pole, it is an equiangular or logarithmic spiral."

Notice that Martin Gardner's discussion should have been more precise at the end, by stating that the projection used is *stereographic.*

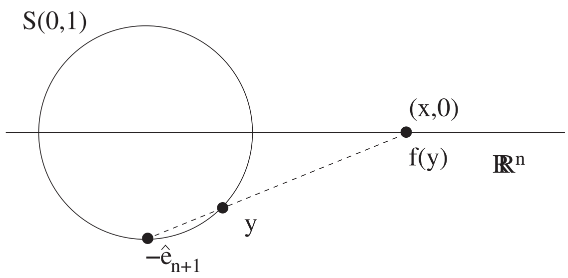
PROBLEM 6–42. Here is a simple geometric way of defining stereographic projection. Work with the unit sphere

$$S(0,1) \subset \mathbb{R}^{n+1}$$

and denote its points with Cartesian coordinates:

$$y = (y_1, \dots, y_n, y_{n+1}) = (y', y_{n+1}),$$

where $y' \in \mathbb{R}^n$. We then set up a function from all of $S(0,1)$ except for the south pole $-\hat{e}_{n+1} = (0,-1)$ to all of $\mathbb{R}^n$ by means of starting with y, constructing the straight line determined by y and $-\hat{e}_{n+1}$, and finding its intersection with the "equatorial plane" $\mathbb{R}^n \times \{0\}$:



a. Prove that

$$x = f(y) = \frac{y'}{1 + y_{n+1}}.$$

(This agrees with what we found earlier in case $n = 2$.

b. Prove that the inverse is

$$y = f^{-1}(x) = \left(\frac{2x}{1 + \|x\|^2}, \frac{1 - \|x\|^2}{1 + \|x\|^2} \right).$$

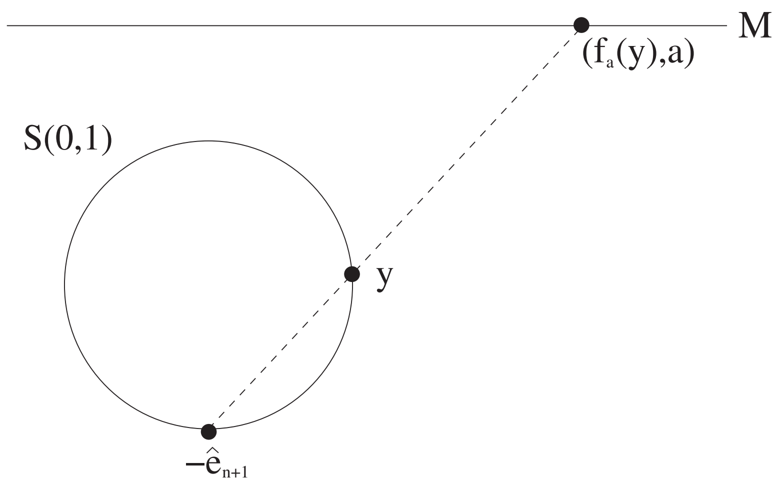
c. Prove that f is conformal with local magnification factor $(1 + y_{n+1})^{-1}$. (In computing $\|Df(y)h\|$ remember that $y \bullet h = y' \bullet h' + y_{n+1}h_{n+1} = 0$.)

PROBLEM 6–43. Here is yet another way to analyze stereographic projection. Consider the inversion mapping of $\mathbb{R}^{n+1}$ with respect to the sphere $S(-\hat{e}_{n+1}, \sqrt{2})$ of radius $\sqrt{2}$ centered at $-\hat{e}_{n+1}$:

$$f(y) = -\hat{e}_{n+1} + 2\frac{y + \hat{e}_{n+1}}{\|y + \hat{e}_{n+1}\|^2}, \; y \in \mathbb{R}^{n+1}.$$

We know that f is a conformal mapping from $\mathbb{R}^{n+1}$ to $\mathbb{R}^{n+1}$. Prove that when restricted to the unit sphere $S(0,1)$ f equals the stereographic projection of Problem 6–42.

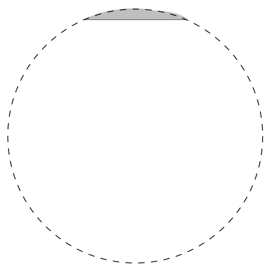
PROBLEM 6–44. Sometimes it is convenient to alter the stereographic projection by projecting onto a hyerplane other than the one we have been using. Continue to project from the south pole $-\hat{e}_{n+1}$, but now project onto any hyerplane orthogonal to the vector $-\hat{e}_{n+1}$. Denote the new hyerplane as $M = \{(x,a) \mid x \in \mathbb{R}^n\}$, where $a \neq -1$. Let f_a denote the corresponding stereographic projection onto $\mathbb{R}^n$.



Prove that f_a is conformal by showing that $f_a = (a+1)f_0$. (The two most common choices for a are 0 and 1.)

For obvious reasons there has been an enormous amount of work dedicated to conformal mappings of portions of $S(0,1) \subset \mathbb{R}^3$ into the Euclidean plane $\mathbb{R}^2$. You should be aware that no such map can exist with no distortion.

PROBLEM 6–45. Prove that there is no isometry of any small region of $S(0,1)$ into $\mathbb{R}^2$. For instance, no matter how small $\epsilon > 0$ is, there is no isometry of $\{(x,y,z) \mid x^2+y^2+z^2=1,\ 1-\epsilon < z \le 1\}$ onto a region in $\mathbb{R}^2$.



(HINT: you might want to think about small circles in $S(0,1)$.)

5. **Lambert conformal conical projection**

PROBLEM 6–46*.

a. Find a map of the USA which displays the Lambert type of projection used.

b. Write an essay about this projection, complete with formulas and proofs.