Chapter 5 Manifolds

We are now going to begin our study of calculus on *curved* spaces. Everything we have done up to this point has been concerned with what one might call the *flat* Euclidean spaces $\mathbb{R}^n$. The objects that we shall now be investigating are called *manifolds*. Each of them will have a certain *dimension* m. This is a positive integer that tells how many independent "coordinates" are needed to describe the manifold, at least locally. For instance, the surface of the earth is frequently modeled as a sphere, a 2-dimensional manifold, with points located in terms of the two quantities latitude and longitude. (This description clearly holds only locally — for instance, the north pole is described in terms of latitude $= 90°$ and longitude is undefined there. Further, longitude ranges between $-180°$ and $180°$, so there's a discontinuity if one tries to coordinatize the entire sphere.)

We shall thus be concerned with m-dimensional manifolds M which are themselves subsets of the n-dimensional Euclidean space $\mathbb{R}^n$. In almost all cases we consider, $m = 1, 2, \dots,$ or $n-1$. There is a case $m = 0$, but these "manifolds" are zero dimensional and thus are just made up of isolated points. The case $m = n$ is actually of some interest; however, a manifold $M \subset \mathbb{R}^n$ of dimension n is just an open set in $\mathbb{R}^n$ and is therefore essentially flat. M and $\mathbb{R}^n$ are locally the same in this case.

When $M \subset \mathbb{R}^n$ we say that $\mathbb{R}^n$ is the *ambient space* in which M lies.

We usually call 1-dimensional manifolds "curves," and 2-dimensional manifolds "surfaces." But we shall generically use the neutral word "manifold."

A. **Hypermanifolds**

Assume $\mathbb{R}^n$ is the ambient space, and $M \subset \mathbb{R}^n$ the manifold. Given that we are not very interested in the case of n-dimensional M, we distinguish manifolds which have the maximal dimension $n-1$ and we call them *hypermanifolds*.

IMPLICIT DESCRIPTION. Suppose $\mathbb{R}^n \xrightarrow{g} \mathbb{R}$ is a function which is of class C^1. We have already thought about its *level sets*, sets of the form

$$M = \{x \mid x \in \mathbb{R}^n,\ g(x) = c\},$$

where c is a constant; see p. 2–42. The fundamental thinking here is that in $\mathbb{R}^n$ there are n independent coordinates; the restriction $g(x_1, \dots, x_n) = c$ removes one degree of freedom, so that points of M can locally be described in terms of only $n-1$ coordinates. Thus we anticipate that M is a manifold of dimension $n-1$, a hypermanifold.

A very nice example of a hypermanifold is the unit sphere in $\mathbb{R}^n$:

$$S(0,1) = \{x \mid x \in \mathbb{R}^n,\ \|x\| = 1\}.$$

There is a very important restriction we impose on this situation. It is motivated by our recognition from p. 2–43 that $\nabla g(x)$ is a vector which should be orthogonal to M at the point $x \in M$. If M is truly $(n-1)$-dimensional, then this vector $\nabla g(x)$ should probably be nonzero. For this reason we impose the restriction that

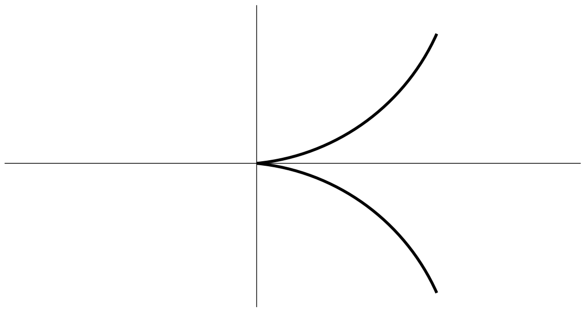
$$\text{for all} \quad x \in M, \quad \nabla g(x) \neq 0.$$

Notice how the unit sphere $S(0,1)$ fits in this scene. If we take $g(x) = \|x\|^2$, then $\nabla g(x) = 2x$. And this vector is not zero for points in the manifold; the fact that $\nabla g(0) = 0$ is irrelevant, as 0 is not a point of the manifold. We could also use $g(x) = \|x\|$, for which $\nabla g(x) = x\|x\|^{-1}$ is never 0. The fact that g is not differentiable at the origin is irrelevant, as the restriction $\|x\| = 1$ excludes the origin.

We now present four examples in $\mathbb{R}^2$ showing what sort of things can go wrong in the absence of this assumption. In all these cases we write $g = g(x,y)$, $c = 0$, and the "bad" point is located at the origin, so $\nabla g(0,0) = 0$.

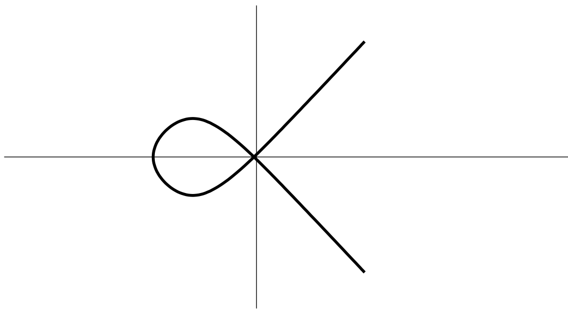
EXAMPLE 1. $x^3 - y^2 = 0$:

$$\nabla g = (3x^2, -2y).$$



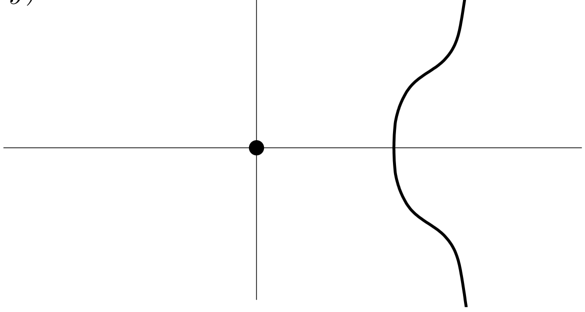
EXAMPLE 2. $x^2 + x^3 - y^2 = 0$:

$$\nabla g = (2x + 3x^2, -2y).$$



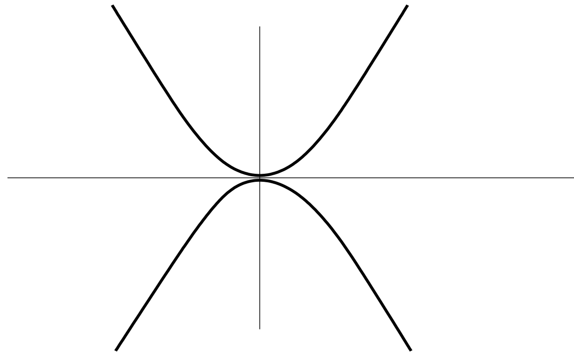
EXAMPLE 3. $-x^2 + x^3 - y^2 = 0$:

$$\nabla g = (-2x + 3x^2, -2y).$$



EXAMPLE 4. $x^4 - y^2 = 0$:

$$\nabla g = (4x^3, -2y).$$



PROBLEM 5–1. This problem generalizes the situation of the sphere $S(0,1)$ described above. Let A be an $n \times n$ real symmetric matrix. Suppose that the set $M = \{x \mid x \in \mathbb{R}^n,\ Ax \bullet x = 1\}$ is not empty. Then M is called a *quadric* in $\mathbb{R}^n$. Prove that it is a hypermanifold. That is, prove that for all $x \in M$ $\nabla(Ax \bullet x) \neq 0$.

PROBLEM 5–2. Manifolds do not necessarily have to be "curved" at all. Thus suppose $h \in \mathbb{R}^n$ is not zero, and suppose $c \in \mathbb{R}$ is fixed. Prove that $\{x \mid x \in \mathbb{R}^n,\ h \bullet x = c\}$ is a hypermanifold.

PROBLEM 5–3. Suppose the set

$$M = \{x \mid x \in \mathbb{R}^n,\ Ax \bullet x + h \bullet x = c\}$$

is not empty. (Here A is an $n \times n$ real symmetric matrix, $h \in \mathbb{R}^n$ is not 0, c is a real number.) Assume that A is invertible and that $A^{-1}h \bullet h \neq -4c$. Prove that M is a hypermanifold.

PROBLEM 5–4. Continuing with the preceding situation, assume A is the identity matrix. Prove that

$$M = \{x \mid x \in \mathbb{R}^n,\ \|x\|^2 + h \bullet x = -\|h\|^2/4\}$$

is not a hypermanifold.

B. **Intrinsic gradient — warm up**

In this section the scenario is that of a hypermanifold $M \subset \mathbb{R}^n$, where M is described implicitly by the level set

$$g(x) = 0.$$

(We can clearly modify g by subtracting a constant from it in order to make M the zero-level set of g.) We assume $\nabla g(x) \neq 0$ for $x \in M$.

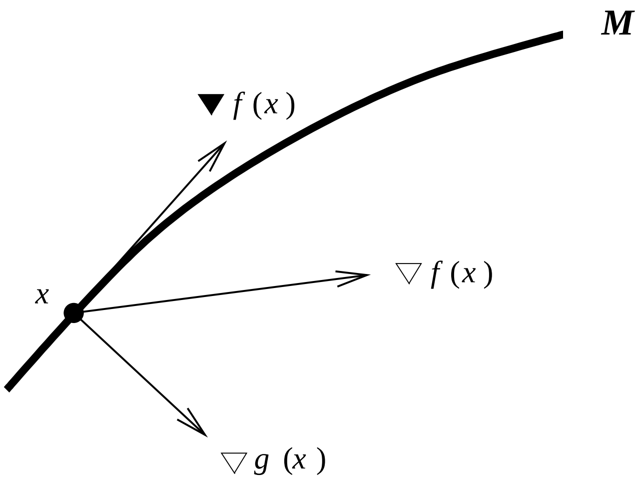
We are definitely thinking that $\nabla g(x)$ represents a vector at $x \in M$ which is *orthogonal* to M. We haven't actually defined this notion yet, but we shall do so in Section F when we talk about the tangent space to M at x. We've discussed this orthogonality on p. 2–43, and it is to our benefit to keep this geometry in mind.

A recurring theme in this chapter is the understanding of the calculus of a function $M \xrightarrow{f} \mathbb{R}$. For such a function we do not have the luxury of knowing that f is defined in the ambient space $\mathbb{R}^n$. As a result, we cannot really talk about partial derivatives $\partial f/\partial x_i$. These are essentially meaningless.

For instance, consider the unit circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$ and the function f defined *only* on the unit circle by $f(x, y) = x^2 - y$. We cannot really say that $\partial f/\partial x = 2x$. For we could also use the formula $f(x, y) = 1 - y^2 - y$ for f, which might lead us to believe $\partial f/\partial x = 0$. And anyway, the notation $\partial f/\partial x$ asks us to "hold y fixed and differentiate with respect to x," but holding y fixed on the unit circle doesn't allow x to vary at all.

Nevertheless, we very much want to have a calculus for functions defined on M, and at the very least we want to be able to define the *gradient* of $M \xrightarrow{f} \mathbb{R}$ in a sensible way. We shall not succeed in accomplishing this right away. In fact, this entire chapter is concerned with giving such a definition, and we shall finish it only when we reach Section F.

However, we want to handle a very interesting special case right away. Namely, we assume $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$, so that f is actually defined in the ambient $\mathbb{R}^n$, and we focus attention on a fixed $x \in M$. As such, $\nabla f(x) \in \mathbb{R}^n$ exists, but this is not what we are interested in. We actually want a vector like $\nabla f(x)$ but which is also *tangent* to M. That is, according to our expectations, which is orthogonal to $\nabla g(x)$. Here's a schematic view:



What we can do is *orthogonally project* $\nabla f(x)$ onto the tangent space to M at x, by the simple device of subtracting from $\nabla f(x)$ the unique scalar multiple of $\nabla g(x)$ which makes the resulting vector orthogonal to $\nabla g(x)$:

$$(\nabla f(x) - c\nabla g(x)) \bullet \nabla g(x) \;=\; 0.$$

Thus

$$c \;=\; \frac{\nabla f(x) \bullet \nabla g(x)}{\|\nabla g(x)\|^2}.$$

You should recognize this procedure; it's exactly the theme of Problem 1–12 (see also pp. 1–14 and 15).

DEFINITION. The *intrinsic gradient* of f in the above situation is the vector

$$\blacktriangledown f(x) \;=\; \nabla f(x) \;-\; \frac{\nabla f(x) \bullet \nabla g(x)}{\|\nabla g(x)\|^2}\, \nabla g(x).$$

An immediate example of interest is the defining function g itself. It clearly satisfies

$$\blacktriangledown g(x) \;=\; 0.$$

PROBLEM 5–5. Prove in fact that if $\mathbb{R} \xrightarrow{\varphi} \mathbb{R}$ is differentiable, then $\blacktriangledown(\varphi \circ g)(x) = 0$.

PROBLEM 5–6. More generally, prove this form of the chain rule:

$$\blacktriangledown(\varphi \circ f)(x) = \varphi'(f(x))\blacktriangledown f(x).$$

Our notation $\blacktriangledown$ displays the difference between the intrinsic gradient and the *ambient* gradient ∇. However, it fails to denote which manifold is under consideration. The intrinsic gradient definitely depends on M. Of course, it must depend on M if for no other reason than we are computing $\blacktriangledown f(x)$ only if x *belongs* to M. We illustrate this dependence with the simple

EXAMPLE. Let M be the sphere $S(0, r)$ in $\mathbb{R}^n$. Then we may use $g(x) = \|x\|^2$ (or $\|x\|^2 - r^2$), so that $\nabla g = 2x$ and we obtain the result

$$\blacktriangledown f(x) \;=\; \nabla f(x) - \frac{x \bullet \nabla f(x)}{r^2}\, x.$$

PROBLEM 5–7. Let $f(x, y) = x^2 - y$ and calculate for the manifold $x^2 + y^2 = 1$

$$\blacktriangledown f = x(2y + 1)(y, -x).$$

Repeat the exercise for the function $h(x, y) = 1 - y^2 - y$, and note that $\blacktriangledown f = \blacktriangledown h$.

PROBLEM 5–8. The intrinsic gradient in the preceding problem is zero at which four points of the unit circle? Describe the nature of each of these "intrinsic critical points" (local maximum, local minimum, saddle point?). Repeat the entire discussion for the function obtained by using the polar angle parameter:

$$f(\cos\theta, \sin\theta) = h(\cos\theta, \sin\theta) = \cos^2\theta - \sin\theta.$$

That is, use single variable calculus and the usual first and second deriative with respect to θ.

PROBLEM 5–9. Let $1 \le i \le n$ be fixed and let M be the hyperplane $\{x \in \mathbb{R}^n \mid x_i = 0\}$. Calculate $\blacktriangledown f$ for this manifold.

PROBLEM 5–10. Show that in general

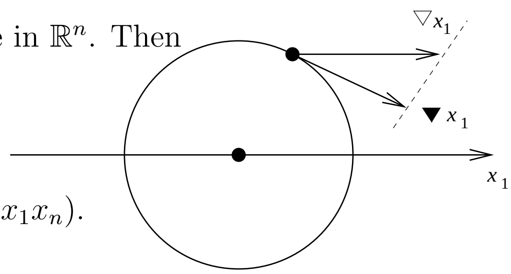
$$\|\blacktriangledown f(x)\| \;=\; \|\nabla f(x)\| \sin\theta,$$

where θ is the angle between the vectors $\nabla f(x)$ and $\nabla g(x)$.

PROBLEM 5–11. Prove the *product rule*:

$$\blacktriangledown(fh) \;=\; f\blacktriangledown h + h\blacktriangledown f.$$

EXAMPLE. Let $f(x) = x_1$ and let M be the unit sphere in $\mathbb{R}^n$. Then



$$\begin{aligned}\blacktriangledown f(x) &= \hat{e}_1 - (x \bullet \hat{e}_1)x \\ &= \hat{e}_1 - x_1 x \\ &= (1 - x_1^2, -x_1x_2, \ldots, -x_1x_n).\end{aligned}$$

Notice that $\blacktriangledown f$ is quite a bit more complicated in form than ∇f.
Also, whereas $\nabla x_1 = \hat{e}_1$ is never 0, it is clear that $\blacktriangledown x_1 = 0$ is a real possibility. In fact, it is zero at the two points $\pm\hat{e}_1$. These happen to be the two points of the sphere where the function x_1 attains its extreme values.

Another way of phrasing the definition of the intrinsic gradient is to let $\hat{N}$ denote a unit vector at x which is *orthogonal* to M. (What this means precisely will be discussed later.) Then we expect that $\nabla g(x) = c\hat{N}$ for some real $c \neq 0$, so that

$$\blacktriangledown f(x) = \nabla f(x) - \nabla f(x) \bullet \hat{N}\hat{N}.$$

PROBLEM 5–12. Let M be a 1-dimensional manifold (a "curve") in $\mathbb{R}^2$. Let $\hat{T}$ be a unit vector tangent to M at x, and prove that $\blacktriangledown f(x)$ is related to the directional derivative of f by the formula

$$\blacktriangledown f(x) = Df(x; \hat{T})\hat{T}.$$

PROBLEM 5–13. Suppose $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$ is homogeneous of degree a. For the manifold which is the sphere $S(0,r)$ of radius r show that

$$\blacktriangledown f = \nabla f - \frac{afx}{r^2}.$$

Another extremely important property is the fact that $\blacktriangledown f$ is truly *intrinsic*: it depends only on the knowledge of f when restricted to the manifold. This is not quite clear at the present stage of our development because we have employed the ambient gradient ∇f in our definition. However, the following argument should serve to make it intuitively clear. In order that $\blacktriangledown f$ just depend on the function f restricted to M, it must be the case that two functions which are equal on M turn out to have the same intrinsic gradient. Equivalently, their difference has zero intrinsic gradient. Equivalently, if $f = 0$ on M, then $\blacktriangledown f = 0$ on M. This seems reasonable, as we believe that if M is a level set of f, then ∇f must be orthogonal to M; that is, ∇f is a scalar multiple of ∇g. But then the orthogonal projection of ∇f orthogonal to ∇g is zero: $\blacktriangledown f = 0$.

Rather than continuing with this discussion at the present time, we instead turn to some wonderful numerical calculations.

C. **Intrinsic critical points**

We continue with the notation and ideas of Section B, so that M is a hypermanifold in $\mathbb{R}^n$ described by an equation $g(x) = 0$. We say that $x \in M$ is an **intrinsic critical point** of f if $\blacktriangledown f(x) = 0$. That is,

$$\nabla f(x) \quad = \quad \frac{\nabla f(x) \bullet \nabla g(x)}{\|\nabla g(x)\|^2}\, \nabla g(x).$$

This is a rather daunting equation to solve for x, but it helps to notice that it says precisely that $\nabla f(x)$ equals a scalar times $\nabla g(x)$; for then taking the inner product with $\nabla g(x)$ shows that the scalar must be the one displayed. This scalar λ is called a *Lagrange multiplier* for the problem. We then are required to solve the equations

$$\begin{cases} \nabla f(x) & = \lambda \nabla g(x), \\ g(x) & = 0 \qquad (\text{as } x \in M). \end{cases}$$

The "unknowns" are *both* x and λ. If we count scalar unknowns and equations, we have $n+1$ unknowns $x_1, \ldots, x_n$, λ, as well as $n+1$ equations. Hopeful!

The equation $g(x) = 0$ is often referred to as the *constraint*.

As is usual in similar situations, setting up the equations is the easy part. Solving them is the hard part, as they are usually nonlinear. Here are some examples.

EXAMPLE. Find the intrinsic critical points of $f(x,y) = x(y-1)$ on the unit circle $x^2+y^2=1$.
Solution. To eliminate some "2's" we can use $g(x,y) = \frac{1}{2}(x^2+y^2-1)$. Then the Lagrange formulation is:

$$\begin{cases} (y-1,x) & = \lambda(x,y), \\ x^2+y^2 & = 1. \end{cases}$$

Thus $y-1=\lambda x$ and $x=\lambda y$. Thus

$$y-1 \;=\; \lambda^2 y,$$

so

$$y(\lambda^2-1) \;=\; -1.$$

Thus

$$y \;=\; \frac{-1}{\lambda^2-1}, \quad x \;=\; \frac{-\lambda}{\lambda^2-1}.$$

Finally, the constraint gives

$$\frac{\lambda^2+1}{(\lambda^2-1)^2} \;=\; 1.$$

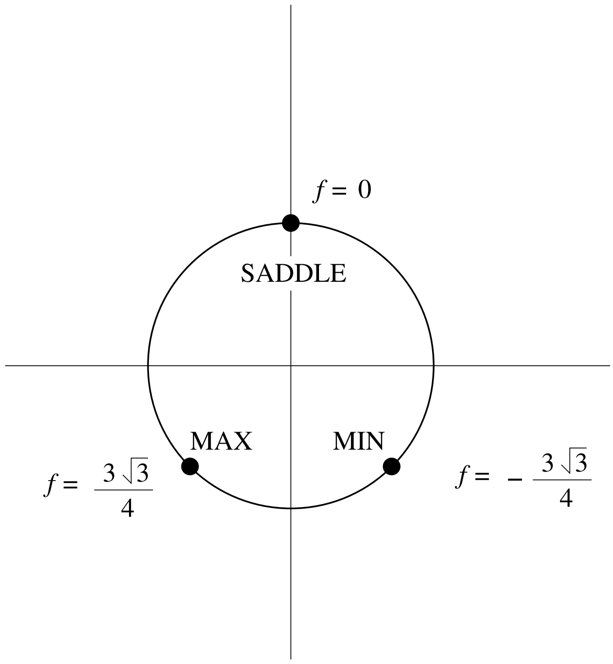
Thus

$$\begin{aligned} \lambda^2+1 &= (\lambda^2-1)^2 \\ &= \lambda^4-2\lambda^2+1; \\ \lambda^4 &= 3\lambda^2. \end{aligned}$$

Thus $\lambda=0$ or $\lambda^2=3$. These give three points:

$$\begin{aligned} \lambda \;=\; 0: &\quad (0,1); \\ \lambda \;=\; \sqrt{3}: &\quad \left(\frac{-\sqrt{3}}{2}, -\frac{1}{2}\right); \\ \lambda \;=\; -\sqrt{3}: &\quad \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right). \end{aligned}$$

Here's a sketch, together with the evident natures of the critical points *relative to M*:



Notice, by the way, that $(0,1)$ is actually an ambient critical point of f: $\nabla f(0,1) = (0,0)$. It is the only one.

PROBLEM 5–14. Just as in Problem 5–8, analyze the function we just studied by examining the function $\cos\theta(\sin\theta - 1)$.

EXAMPLE. Find the intrinsic critical points of $f(x,y) = x^3 + 8y$ on the ellipse $\frac{x^2}{4} + \frac{y^2}{2} = 1$.
Solution. The Lagrange formulation is:

$$\begin{cases} 3x^2 & = \lambda x/2, \\ 8 & = \lambda y, \\ \frac{x^2}{4} + \frac{y^2}{2} & = 1. \end{cases}$$

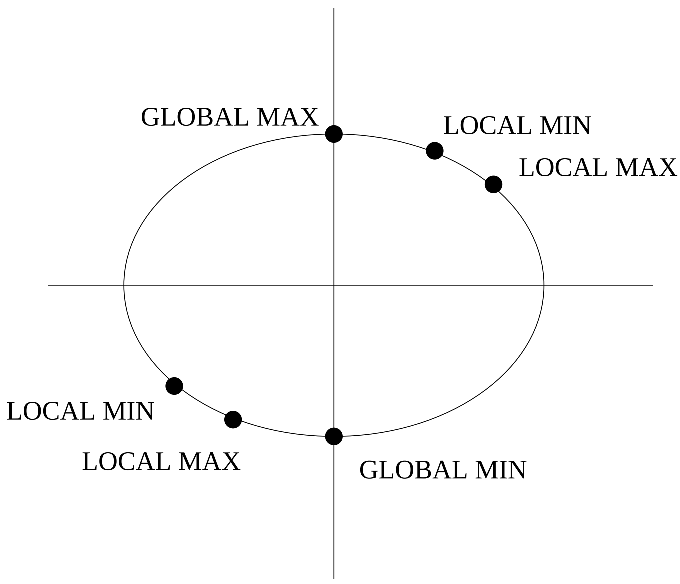
If $x = 0$, we get $y^2 = 2$ and thus two points: $(0, \sqrt{2})$, $(0, -\sqrt{2})$. If $x \neq 0$, then

$$x = \frac{\lambda}{6}, \quad y = \frac{8}{\lambda}.$$

Thus $xy = 4/3$, so the constraint gives

$$\begin{aligned} \frac{x^2}{4} + \frac{1}{2}\left(\frac{4}{3x}\right)^2 &= 1; \\ \frac{x^2}{4} + \frac{8}{9x^2} &= 1; \\ 9x^4 + 32 &= 36x^2; \\ 9x^4 - 36x^2 + 32 &= 0; \\ (3x^2 - 4)(3x^2 - 8) &= 0; \\ x^2 &= 4/3 \text{ or } 8/3. \end{aligned}$$

Thus we find four more points: $\pm\left(\sqrt{\frac{4}{3}}, \sqrt{\frac{4}{3}}\right)$ and $\pm\left(\sqrt{\frac{8}{3}}, \sqrt{\frac{2}{3}}\right)$. Here is a sketch of the six intrinsic critical points:



In both of the above examples we drew conclusions about intrinsic extreme values; that is, maximum and minimum values of the relevant function's restriction to the manifold. We shall prove in Section E that such intrinsic extreme values are indeed intrinsic critical points, just as in the case of functions defined on $\mathbb{R}^n$ (see Section 2G).

It is quite interesting to consider the special case of quadratic forms, as we did in proving the principal axis theorem in Chapter 4. So let A be an $n \times n$ symmetric real matrix. We there were analyzing A by means of the Rayleigh quotient

$$\frac{Ax \bullet x}{\|x\|^2},$$

and we essentially found its critical points in $\mathbb{R}^n$. The homogeneity shows that to be the same as finding the intrinsic critical points of $Ax \bullet x$ *on the unit sphere.* Thus we ask for points x satisfying

$$\blacktriangledown(Ax \bullet x) \;=\; 0 \quad \text{and} \quad \|x\| \;=\; 1.$$

The Lagrange formulation gives

$$\begin{cases} \nabla(Ax \bullet x) & = \lambda \nabla(\|x\|^2), \\ \qquad \|x\| & = 1. \end{cases}$$

That is,

$$\begin{cases} Ax & = \lambda x, \\ \|x\| & = 1. \end{cases}$$

Thus the intrinsic critical points of $Ax \bullet x$ on the unit sphere are precisely the eigenvectors of A! As an example of this procedure, work out the following problem:

PROBLEM 5–15. Find the intrinsic critical points of $(x+y)(y+z)$ on the unit sphere $x^2 + y^2 + z^2 = 1$.

PROBLEM 5–16. For the Rayleigh quotient function $Q(x) = Ax \bullet x / \|x\|^2$ show that the intrinsic gradient on the unit sphere equals

$$\begin{aligned} \blacktriangledown Q &= \nabla Q \\ &= 2Ax - 2Ax \bullet xx. \end{aligned}$$

Here are six more or less routine exercises, followed by six challenging ones.

PROBLEM 5–17. Use the Lagrange technique to find the points on the parabola $y^2 + 2x = 8$ which are closest to the origin.

PROBLEM 5–18. Find the minimum of $x^4 + 4axy + y^4$ on the hyperbola $x^2 - y^2 = 1$.

PROBLEM 5–19. Find the minimum distance from $(9, 12, -5)$ to points on the cone in $\mathbb{R}^3$ given by $4z^2 = x^2 + y^2$.

PROBLEM 5–20. Consider the function $f(x, y) = x$ on the level set $y^2 - x^3 = 0$. Show that f attains its minimum value at the origin only. Show that the Lagrange formulation fails to produce this result. Explain why.

PROBLEM 5–21. Repeat the preceding exercise for the function $f(x, y) = y$ on the set $M : y^3 = x^6 + x^8$. In this case show also that the level set M *is* a *bona fide* 1-dimensional manifold in $\mathbb{R}^2$.

PROBLEM 5–22. Let a, b, c be positive constants. Find the intrinsic critical points of $\frac{a}{x} + \frac{b}{y} + \frac{c}{z}$ on the unit sphere $x^2 + y^2 + z^2 = 1$.

PROBLEM 5–23. Find the intrinsic critical points of $2(x_1 + x_2 + x_3)(x_1 + x_2 + x_4)$ on the unit sphere in $\mathbb{R}^4$.

PROBLEM 5–24. Find all the intrinsic critical points of $f(x) = x_1^3 + x_2^3 + x_3^3 + 2x_1x_2x_3$ on the unit sphere in $\mathbb{R}^3$.

PROBLEM 5–25*. Find all the intrinsic critical points of $f(x) = x_1^3 + x_2^3 + x_3^3 - 3x_1x_2x_3$ on the unit sphere in $\mathbb{R}^3$.

PROBLEM 5–26.** Let a be an arbitrary but fixed real number. Find all the intrinsic critical points of $f(x) = x_1^3 + x_2^3 + x_3^3 + ax_1x_2x_3$ on the unit sphere in $\mathbb{R}^3$, and count how many there are depending on the value of a.

PROBLEM 5–27. Find all the intrinsic critical points of $f(x) = x_1^3 + x_2^2 + x_3$ on the unit sphere in $\mathbb{R}^3$. Determine the maximum and minimum values of f on the sphere.

The next problem is a gem I found in the book *Ideals, Varieties, and Algorithms*, by Cox, Little, and O'Shea. (David Cox is a 1970 BA Rice, majoring in mathematics, and is now on the faculty of the Department of Mathematics, Amherst College).

PROBLEM 5–28*. Find all the intrinsic critical points of $f(x) = x_1^3 + 2x_1x_2x_3 - x_3^2$ on the unit sphere (there are ten), and the maximum and minimum values as well.

PROBLEM 5–29*. Consider the ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ where $0 < b < a$. For any point (x, y) on the ellipse construct the line through that point which is orthogonal to the ellipse. This line intersects the ellipse in another point (x', y'). Let $D(x, y)$ denote the distance between (x, y) and (x', y'). Find the minimum value of this function by first showing that

$$D(x,y) = \frac{2(\frac{x^2}{a^4} + \frac{y^2}{b^4})^{3/2}}{\frac{x^2}{a^6} + \frac{y^2}{b^6}}$$

and then using the Lagrange technique to minimize $D(x, y)$.

(Answer includes $D=\sqrt{27}a^2b^2(a^2+b^2)^{-3/2}$ in case $a\geq\sqrt{2}b$.)

D. **Explicit description of manifolds**

Frequently hypermanifolds in $\mathbb{R}^n$ are described in terms of *graphs* of functions $\mathbb{R}^{n-1} \xrightarrow{\varphi} \mathbb{R}$. This means that on the manifold one of the coordinates in $\mathbb{R}^n$ is expressed "explicitly" as a function of the other $n-1$ coordinates. As there is no particular coordinate to prefer over another, we shall with no loss of generality consider the abstract situation in which M is represented by a formula

$$x_n = \varphi(x_1, \ldots, x_{n-1}).$$

Of course, M may then be thought of as represented *implicitly* in terms of the zero level set of $\mathbb{R}^n \xrightarrow{g} \mathbb{R}$, where we simply define the new function g by

$$g(x) = x_n - \varphi(x_1, \ldots, x_{n-1}).$$

The normal vector is then given by

$$\begin{aligned} \nabla g &= (-D_1\varphi, \ldots, -D_{n-1}\varphi, 1) \\ &= (-\nabla\varphi, 1) \qquad \text{(for short)}. \end{aligned}$$

Notice that ∇g is definitely not zero!

Now we give an interesting calculation to show what $\blacktriangledown f$ looks like in this framework. We shall require *only* the values of the function f on the manifold M. A convenient way to use these values is to define an associated function f_0 on $\mathbb{R}^{n-1}$ by the formula

$$f_0(x_1, \dots, x_{n-1}) = f\big(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1})\big).$$

Notice that the new function f_0 indeed uses only the evaluation of f at points of M.

Note first that the chain rule implies

$$D_k f_0 = D_k f + D_n f D_k \varphi, \quad 1 \le k \le n-1.$$

Using vector notation in $\mathbb{R}^{n-1}$,

$$\begin{aligned} \nabla f_0 &= (D_1 f_0, \dots, D_{n-1} f_0) \\ &= (D_1 f, \dots, D_{n-1} f) + D_n f (D_1 \varphi, \dots, D_{n-1} \varphi). \end{aligned}$$

Now we simply regard ∇f_0 as a vector in $\mathbb{R}^n$ with n^{th} component 0, which we write as

$$0 = D_n f + D_n f(-1).$$

Thus

$$\begin{aligned} \nabla f_0 &= (D_1 f, \dots, D_{n-1} f, D_n f) + D_n f (D_1 \varphi, \dots, D_{n-1} \varphi, -1) \\ &= \nabla f + D_n f (\nabla \varphi, -1) \\ &= \nabla f - D_n f \nabla g. \end{aligned}$$

Now we are all set to compute the intrinsic gradient of f. By definition

$$\begin{aligned} \blacktriangledown f &= \nabla f - \frac{\nabla f \bullet \nabla g}{\|\nabla g\|^2} \nabla g \\ &= \nabla f - \frac{(\nabla f_0 + D_n f \nabla g) \bullet \nabla g}{\|\nabla g\|^2} \nabla g \\ &= \nabla f - \frac{\nabla f_0 \bullet \nabla g}{\|\nabla g\|^2} \nabla g - D_n f \nabla g \\ &= \nabla f_0 - \frac{\nabla f_0 \bullet \nabla g}{\|\nabla g\|^2} \nabla g. \end{aligned}$$

We summarize:

THEOREM. *In the above context, where M is given explicitly*

$$M = \{x \in \mathbb{R}^n \mid x_n = \varphi(x_1, \dots, x_{n-1})\},$$

we define

$$f_0(x_1, \dots, x_{n-1}) = f\big(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1})\big).$$

Then

$$\blacktriangledown f = \nabla f_0 + \frac{\nabla f_0 \bullet \nabla\varphi}{\|\nabla\varphi\|^2 + 1} \quad (-\nabla\varphi, 1).$$

In particular, the intrinsic gradient $\blacktriangledown f$ ***depends* only *on the restriction of*** f ***to*** M***.***

This theorem is of great theoretical importance in that it shows dramatically the intrinsic nature of $\blacktriangledown f$. However, it does not appear to be of any particular use in solving exercises, as the Lagrange formulation of the preceding section is indeed quite applicable.

PROBLEM 5–30. Prove that f has an intrinsic critical point at $x \iff f_0$ has a critical point at the corresponding point.

E. **Implicit function theorem**

We now turn to a theorem of immense importance in the study of manifolds. It actually provides the complete understanding of intrinsic gradients. More than that, it shows that hypermanifolds which are described *implicitly* can also be described *explicitly*. Thus it is well named: THE IMPLICIT FUNCTION THEOREM.

We do not prove this theorem (one of the "hard" ones) in this course. It is commonly proved in beginning courses in mathematical analysis. However, we very much need to understand exactly what it does (and does not) say.

Suppose then that the hypermanifold $M \subset \mathbb{R}^n$ is described implicitly in a neighborhood of $x_0 \in M$ by the equation

$$g(x) = 0.$$

As usual, $\mathbb{R}^n \xrightarrow{g} \mathbb{R}$ is assumed to be of class C^1, and $\nabla g(x) \neq 0$ for all x in a neighborhood of x_0. We want to describe M explicitly near x_0, so what we need to do is *solve* the equation $g(x) = 0$ for one of the variables in terms of the others. Say we succeed in solving for x_n as a function of $x_1, \dots, x_{n-1}$: in a neighborhood of x_0 we then have a situation

$$g(x) = 0 \iff x_n = \varphi(x_1, \dots, x_{n-1}),$$

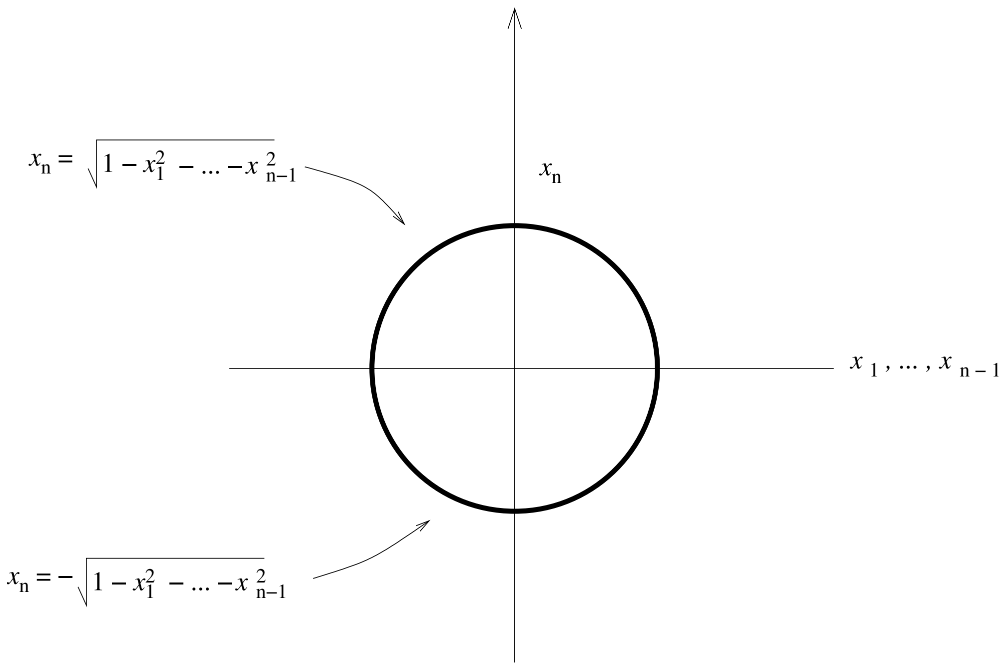
for some function φ defined on a neighborhood in $\mathbb{R}^{n-1}$.

It is surely reasonable to expect that we need a condition like $\partial g/\partial x_n \neq 0$ in order to do this. At the very least, we need $g(x_1, \dots, x_n)$ to involve x_n in a significant way. Furthermore, the manifold $x_n = \varphi(x_1, \dots, x_{n-1})$ has the normal vector $(-\nabla\varphi, 1)$ with nonzero n^{th} coordinate, and thus the normal vector ∇g should also have a nonzero n^{th} coordinate.

As a simple example take $g(x) = \|x\|^2 - 1$, so that we are dealing with the all-important unit sphere as M. Then $\nabla g(x) = 2x$. Solving $g(x) = 0$ goes something like this:

$$\begin{aligned} x_n^2 &= 1 - x_1^2 - \cdots - x_{n-1}^2 \\ x_n &= \pm\sqrt{1 - x_1^2 - \cdots - x_{n-1}^2}. \end{aligned}$$

The choice of sign gives the "upper" or "lower" "hemisphere."



We see one thing immediately: we cannot in general expect the explicit presentation of M to be anything but local. We also are going to want φ to belong to class C^1 so that we can do calculus. In this example this C^1 quality will not happen if $x_1^2 + \cdots + x_{n-1}^2 = 1$ (the "equator" $x_n = 0$).

Now suppose $x_0 \in S(0, 1)$ and its n^{th} coordinate $x_{on} \neq 0$. Then $\nabla g(x_0) = 2x_0 \neq 0$. Furthermore, we can solve for x_n near x_0. If $x_{on} > 0$, we obtain the + sign above, and the reverse if $x_{on} < 0$. Thus in the latter case $x_{on} < 0$, we have

$$x_n = -\sqrt{1 - x_1^2 - \cdots - x_{n-1}^2} \qquad \text{for} \quad x_1^2 + \cdots + x_{n-1}^2 < 1.$$

Of course, the entire sphere can be handled in a similar way since $\nabla g(x_0) = 2x_0 \neq 0$ requires some coordinate x_{oi} of x_0 to be nonzero, and we can then solve for x_i locally:

$$x_i = \pm\sqrt{1 - x_1^2 - \cdots - x_{i-1}^2 - x_{i+1}^2 - \cdots - x_n^2}.$$

The above simple example of the sphere is completely typical. Here is the result.

IMPLICIT FUNCTION THEOREM. *Suppose the hypermanifold $M \subset \mathbb{R}^n$ is described as the level set*

$$g(x) = 0,$$

where $\mathbb{R}^n \xrightarrow{g} \mathbb{R}$ is of class C^1 and $\nabla g \neq 0$ on M. Suppose $x_0 \in M$. Suppose that $\partial g/\partial x_i(x_0) \neq 0$. Then there exists $\mathbb{R}^{n-1} \xrightarrow{\varphi} \mathbb{R}$ of class C^1 such that for all x in a sufficiently small neighborhood of x_0,

$$g(x) = 0 \iff x_i = \varphi(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Though this is an existence theorem and as such does not provide a clue about the actual calculation of φ, the chain rule gives "explicit" formulas for the partial derivatives of φ. For we may start with the functional identity in the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$:

$$g(x_1, \ldots, x_{i-1}, \varphi, x_{i+1}, \ldots, x_n) = 0.$$

Now differentiate this identity with respect to x_j for any $j \neq i$. The chain rule implies

$$D_j g + D_i g D_j \varphi = 0.$$

Therefore we conclude that

$$D_j \varphi = -\frac{D_j g}{D_i g}.$$

On the right side of the latter equation, the partial derivatives of g are evaluated at $x_i = \varphi$. Notice the appearance of the nonzero quantity $D_i g$ in the denominator.

There is a nice moral to get from all of this. Solving the equation $g(x) = 0$ for x_i in terms of the other coordinates is likely to be a very difficult task. But once that has been done and the above function φ has been produced, the calculation of the partial derivatives $D_j \varphi$ is very simple. In fact, it's a *linear* task. You have seen this sort of "implicit differentiation" in your introductory calculus courses. For instance there are lots of exercises of the following nature: the equation

$$\pi e^{xy} + \sin y - y - x^2 = 0$$

is satisfied by a function $y = y(x)$ near $x = 0$, and $y(0) = \pi$. Compute dy/dx. The solution is obtained by performing d/dx:

$$\pi e^{xy}\left(x\frac{dy}{dx} + y\right) + \cos y\frac{dy}{dx} - \frac{dy}{dx} - 2x = 0.$$

Then solve:

$$\frac{dy}{dx} = \frac{2x - \pi y e^{xy}}{\pi x e^{xy} + \cos y - 1}.$$

How simple. (Never mind that we don't really "know" the terms $y = y(x)$ on the right side.) Notice that when $x = 0$ and $y = \pi$, the denominator equals -2 and this is not 0. In particular,

$$\frac{dy}{dx}\bigg|_{x=0} = \frac{\pi^2}{2}.$$

Another nice result of the implicit function theorem is the proof that the intrinsic gradient $\blacktriangledown f$ depends only on M and not on the particular function g whose level set is equal to M. This is clear once we check that ∇g is uniquely determined by M, up to a nonzero scalar multiple; for the formula for $\blacktriangledown f$ shows that the scalar multiple cancels out of the equation. More geometrically, $\blacktriangledown f$ is just the vector ∇f with the correct multiple of ∇g added so that the resulting vector is orthogonal to ∇g; this makes it clear that only the direction of ∇g is needed.

THEOREM. *Given a hypermanifold $M \subset \mathbb{R}^n$, suppose it is described as the level set $\{x \in \mathbb{R}^n \mid g(x) = 0\}$, where $\nabla g \neq 0$. Then ∇g is uniquely determined by M, up to a nonzero scalar multiple (which may be a function of x).*

PROOF. Suppose for instance that $D_n g(x_0) \neq 0$. Then the implicit function theorem yields a function $\mathbb{R}^{n-1} \xrightarrow{\varphi} \mathbb{R}$ such that near x_0 we have

$$x \in M \iff x_n = \varphi(x_1, \dots, x_{n-1}).$$

Then as above we obtain on M

$$D_j g + D_n g D_j \varphi = 0, \qquad 1 \le j \le n-1.$$

We conclude

$$\nabla g = D_n g(-\nabla\varphi, 1).$$

If another function $\tilde{g}$ also gives the manifold, and $\nabla\tilde{g}(x_0) \neq 0$, then $\tilde{g}(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1})) = 0$ so that the chain rule again gives

$$\nabla\tilde{g} = D_n\tilde{g}(-\nabla\varphi, 1).$$

Thus $D_n\tilde{g}(x) \neq 0$ and $\nabla\tilde{g}(x)$ =scalar times $\nabla g(x)$.

QED

We can now clear up an issue that we have been ignoring, thanks to the fact that we understand that the intrinsic gradient $\blacktriangledown f$ is completely determined by the restriction of the function f to the manifold in question. The issue is this: suppose f attains a local maximum or minimum value at $x_0 \in M$ relative to the restriction of f to M. Then we want to know that necessarily x_0 is an intrinsic critical point of f. Here's the result:

THEOREM. *Suppose M is a hypermanifold in $\mathbb{R}^n$ and suppose $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$ is a C^1 function defined in a neighborhood of a point $x_0 \in M$. Suppose that $f(x) \leq f(x_0)$ for all $x \in M$ belonging to some neighborhood of x_0. Then $\blacktriangledown f(x_0) = 0$.*

PROOF. Thanks to the implicit function theorem, we know that M can be represented explicitly near x_0 by an equation of the form

$$x_n = \varphi(x_1, \dots, x_{n-1})$$

(we have named x_n as the distinguished coordinate for simplicity of writing only). We use the function f_0 of Section D,

$$f_0(x_1, \dots, x_{n-1}) = f(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1})).$$

Let $x'_0 = (x_{01}, \dots, x_{0,n-1})$. Then our hypothesis means precisely that

$$f_0(x') \;\leq\; f_0(x'_0)$$

for all $x' \in \mathbb{R}^{n-1}$ sufficiently near x'_0. Thus x'_0 is a critical point for the function f_0, and we conclude that its gradient $\nabla f_0(x'_0) = 0$. But then the theorem on p. 5–15 yields

$$\begin{aligned}\blacktriangledown f(x_0) &= \nabla f_0 + \frac{\nabla f_0 \bullet \nabla\varphi}{\|\nabla\varphi\|^2 + 1} \qquad (-\nabla\varphi, 1) \\ &= 0.\end{aligned}$$

QED

REMARKS. Of course, the conclusion still holds if we are dealing with a local minimum instead: $f(x) \geq f(x_0)$ for all $x \in M$ sufficiently near x_0. In the next section we shall give a somewhat different proof of this result that is even more intrinsic. And in Section G we shall learn that the theorem remains valid for manifolds $M \subset \mathbb{R}^n$ of any dimension, not just $n-1$.

PROBLEM 5–31. THE ARITHMETIC-GEOMETRIC MEAN INEQUALITY. Using the following outline, prove that for any $x_1 \geq 0, \dots, x_n \geq 0$,

$$(x_1 \dots x_n)^{1/n} \leq \frac{x_1 + \cdots + x_n}{n},$$

and that equality holds $\Longleftrightarrow x_1 = \cdots = x_n$.

a. Prove first that you may assume that $x_1 + \cdots + x_n = 1$.

b. Show that the function $f = x_1 \dots x_n$ restricted to the set $x_1 + \cdots + x_n = 1$, $x_i \geq 0$ for all i, attains its maximum value at some x_0.

c. Show that all the coordinates of this point x_0 are positive.

d. Use the Lagrange technique to determine x_0.

PROBLEM 5–32. Find the minimum of $x_1\, x_2 \dots x_n$ subject to the constraint

$$\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n} = 1, \quad \text{all } x_i > 0.$$

F. **The tangent space**

Now we are going to face the problem of actually defining tangent vectors to a manifold. Suppose that $M \subset \mathbb{R}^n$ is a manifold, not necessarily a hypermanifold. Suppose $x_0 \in M$. We want to give a definition of tangent vectors to M at x_0 that is as intrinsic to M as possible (how the "inhabitants" of M view tangent vectors). We shall accomplish this by focusing attention on curves which lie in the manifold.

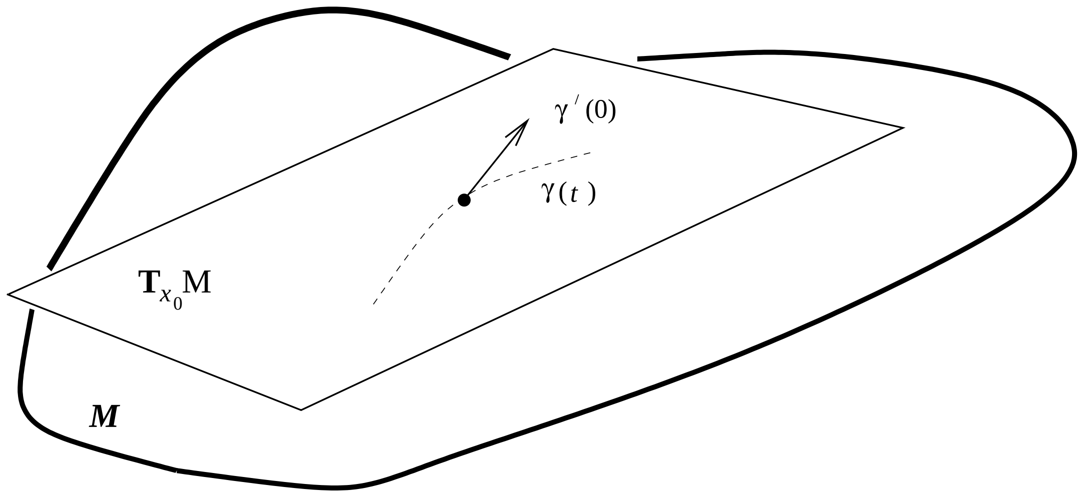
DEFINITION. In the above situation consider all curves (see p. 2–3) γ from $\mathbb{R}$ to $\mathbb{R}^n$ which satisfy the following:

$$\begin{aligned}&\gamma(t) \in M \quad \text{for all} \quad t,\\ &\gamma(0) \;=\; x_0,\\ &\gamma \quad \text{is of class} \quad C^1.\end{aligned}$$

Then the velocity vector $\gamma'(0)$ is called a *tangent vector to* M *at* x_0. Notice that $\gamma'(0) \in \mathbb{R}^n$. The set of all such vectors is called the *tangent space to* M *at* x_0, and is written

$$T_{x_0}M.$$

PROBLEM 5–33. Prove that $0 \in T_{x_0}M$ and that if $h \in T_{x_0}M$ and $c \in \mathbb{R}$, then $ch \in T_{x_0}M$.
(HINT: γ(ct).)



We anticipate that if M is an m-dimensional manifold, then $T_{x_0}M$ is an m-dimensional subspace of $\mathbb{R}^n$. The preceding problem indeed shows that scalar multiples of tangent vectors at x_0 are themselves tangent vectors at x_0, but no such simple technique will handle sums of tangent vectors. We shall prove all these results when we come to Chapter 6. (Though $T_{x_0}M$ is a subspace of $\mathbb{R}^n$ and contains the origin, we always imagine this subspace *attached* to M at the point x_0, with the origin in $T_{x_0}M$ being regarded as located at x_0.)

For the present time we continue to focus our attention on the hypermanifold case with a theorem that reveals all we really need to know about $T_{x_0}M$ in this case.

THEOREM. *Let M be a hypermanifold in $\mathbb{R}^n$, described implicitly as a level set*

$$g(x) \;=\; 0,$$

where g is of class C^1 and $\nabla g \neq 0$ on M. Let $x_0 \in M$. Then

$$T_{x_0}M \;=\; \{h \in \mathbb{R}^n \mid \nabla g(x_0) \bullet h \;=\; 0\}.$$

In particular, $T_{x_0}M$ is an $(n-1)$-dimensional subspace of $\mathbb{R}^n$.

PROOF. First, suppose $h \in T_{x_0}M$. Use a curve γ as in the definition above, with $\gamma'(0) = h$. Then since $\gamma(t) \in M$,

$$g(\gamma(t)) \ = \ 0.$$

Computing the t derivative and using the chain rule,

$$\nabla g(\gamma(t)) \bullet \gamma'(t) \ = \ 0.$$

Setting $t = 0$,

$$\nabla g(x_0) \bullet h \ = \ 0.$$

Conversely, suppose $h \in \mathbb{R}^n$ satisfies

$$\nabla g(x_0) \bullet h \ = \ 0.$$

Now we have to do something quite significant. We are required to produce a curve in M with all the right properties. Since we don't even know how to produce individual points in M, much less a whole curve, we need a theorem of some sort. The *implicit function theorem* serves the purpose perfectly. For ease in writing let us suppose $D_n g(x_0) \neq 0$. Then we know (thanks to the implicit function theorem) that M can be described explicitly as a graph

$$x_n \ = \ \varphi(x_1, \dots, x_{n-1})$$

near x_0. Write $x_0 = (x_{01}, \dots, x_{0n})$. We then define a curve $\gamma(t)$ by making it affine in the "independent" coordinates $x_1, \dots, x_{n-1}$ in the following way:

$$\begin{aligned} \gamma_j(t) &= x_{0j} + h_j t, \quad 1 \leq j \leq n-1, \\ \gamma_n(t) &= \varphi\left(\gamma_1(t), \dots, \gamma_{n-1}(t)\right). \end{aligned}$$

From the chain rule and the formula on p. 5–18 we obtain

$$\begin{aligned} \gamma_n'(0) &= \sum_{j=1}^{n-1} D_j\varphi\left(x_{01}, \dots, x_{0,n-1}\right)\gamma_j'(0) \\ &= \sum_{j=1}^{n-1} -\frac{D_j g(x_0)}{D_n g(x_0)}\, h_j. \end{aligned}$$

But also we have

$$\nabla g(x_0) \bullet h \ = \ 0;$$

that is,

$$\sum_{j=1}^{n-1} D_j g(x_0) h_j + D_n g(x_0) h_n \;=\; 0.$$

Thus

$$\gamma_n'(0) \;=\; h_n.$$

This proves that $\gamma'(0) = h$, as desired.

QED

As a nice bonus, we can now easily give a complete understanding of the intrinsic gradient of a function, relative to the hypermanifold M. Suppose first that $\mathbb{R}^n \xrightarrow{f} \mathbb{R}$ is of class C^1 near a point $x_0 \in M$. Suppose that $h \in T_{x_0}M$. There are now two ways to view this situation.

(1) We use $\nabla g(x_0) \bullet h = 0$ and the definition from p. 5–5,

$$\blacktriangledown f(x_0) \;=\; \nabla f(x_0) - \frac{\nabla f(x_0) \bullet \nabla g(x_0)}{\|\nabla g(x_0)\|^2}\, \nabla g(x_0),$$

to conclude that

$$\begin{aligned} \blacktriangledown f(x_0) \bullet h \;&=\; \nabla f(x_0) \bullet h \\ &=\; Df(x_0; h). \end{aligned}$$

Remember from p. 2–14 that $Df(x_0; h)$ is our notation for the directional derivative of f at x_0 in the direction h. Thus $\blacktriangledown f(x_0)$ is the unique vector in $T_{x_0}M$ whose inner product with every $h \in T_{x_0}M$ equals $Df(x_0; h)$.

(2) Consider an arbitrary curve γ in M such that $\gamma(0) = x_0$ and $\gamma'(0) = h$. Then the chain rule gives

$$\frac{d}{dt} f\left(\gamma(t)\right) \;=\; \nabla f\left(\gamma(t)\right) \bullet \gamma'(t),$$

so that

$$\begin{aligned} \frac{d}{dt} f\left(\gamma(t)\right)\Big|_{t=0} \;&=\; \nabla f(x_0) \bullet h \\ &=\; \blacktriangledown f(x_0) \bullet h. \end{aligned}$$

It is this second relationship that is so intriguing, since the function $f \circ \gamma$ depends on the behavior of f *only on the manifold* M and not on the ambient $\mathbb{R}^n$. We can therefore extend the definition of p. 5–5 as in the theorem we are preparing to consider.

But before we state the theorem, we need to explain part of the hypothesis. Namely, we are going to assume that $M \xrightarrow{f} \mathbb{R}$ is of class C^1. Since f might be defined *only* on M, it is not immediately clear how to define this continuous differentiability. In fact, a moment's thought might lead to two competing ideas:

(1) Representing M in an explicit manner such as

$$x_n = \varphi(x_1, \dots, x_{n-1}),$$

require the resulting function

$$f_0(x_1, \dots, x_{n-1}) = f(x_1, \dots, x_{n-1}, \varphi(x_1, \dots, x_{n-1}))$$

to be of class C^1 on (a neighborhood in) $\mathbb{R}^n$.

(2) Require that there exist a C^1 function $\mathbb{R}^n \xrightarrow{F} \mathbb{R}$ such that $F(x) = f(x)$ for $x \in M$ (in a neighborhood of some point).

PROBLEM 5–34. Prove that these two definitions are equivalent.

THEOREM. *Let M be a hypermanifold in $\mathbb{R}^n$ and $x_0 \in M$. Let $M \xrightarrow{f} \mathbb{R}$ be a C^1 function defined* only *on M. Then there exists a unique vector $\blacktriangledown f(x_0)$ in $T_{x_0}M$ such that for all C^1 curves γ in M such that $\gamma(0) = x_0$,*

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

DEFINITION. The tangent vector $\blacktriangledown f(x_0)$ is called the *intrinsic gradient* of f at x_0. Because of the discussion right before the theorem, it agrees with the definition given on p. 5–5 in case f is defined in a neighborhood of x_0 in $\mathbb{R}^n$.

PROOF. Use the second of the two definitions of C^1 given above. Thus in a neighborhood of x_0 there exists some C^1 function $\mathbb{R}^n \xrightarrow{F} \mathbb{R}$ which agrees with f on M. Then we simply compute

$$\begin{aligned} \frac{d}{dt} f(\gamma(t)) \Big|_{t=0} &= \frac{d}{dt} F(\gamma(t)) \Big|_{t=0} \\ &= \blacktriangledown F(x_0) \bullet \gamma'(0), \end{aligned}$$

thanks to the known properties of the intrinsic gradient $\blacktriangledown F(x_0)$ of the ambient function F. This finishes the *existence* part of the proof, as we may simply define $\blacktriangledown f(x_0) = \blacktriangledown F(x_0)$.

The *uniqueness* is a separate argument. If there were two vectors v and $w \in T_{x_0}M$ fulfilling the conclusion of the theorem, then we would have

$$v \bullet \gamma'(0) \;=\; w \bullet \gamma'(0) \qquad \text{for all curves} \;\; \gamma.$$

Thus,

$$(v - w) \bullet h \;=\; 0 \qquad \text{for all} \;\; h \in T_{x_0}M.$$

Since $v - w \in T_{x_0}M$, we conclude that $v - w = 0$. Thus $v = w$.

QED

REMARK. Given a C^1 function $M \xrightarrow{f} \mathbb{R}$, there are *many* ways to extend it to $\mathbb{R}^n \xrightarrow{F} \mathbb{R}$ in a neighborhood of x_0. Each such F has a gradient $\nabla F(x_0)$, but the *intrinsic* gradient $\blacktriangledown F(x_0)$ is independent of the choice of the extension F. Our results show that each $\blacktriangledown F(x_0)$ is just equal to $\blacktriangledown f(x_0)$. Thus we have a very practical "algorithm" for computing $\blacktriangledown f$. Namely, first extend f to an ambient C^1 function F; second, compute $\blacktriangledown F$. The resulting vector is precisely $\blacktriangledown f$.

The above results are in agreement with what we accomplished in the theorem on p. 5–15, where we noticed that the intrinsic gradient of a function depends only on its restriction to the manifold. The extra information we now have comes from the intrinsic understanding of tangent vectors themselves.

G. **Manifolds that are not hyper**

In this section we want to move away from the restriction that $M \subset \mathbb{R}^n$ has dimension $n - 1$. Thus we shall study manifolds of dimension m contained in $\mathbb{R}^n$, where $1 \le m \le n - 1$. This range of dimension covers all that we are really concerned with in the present chapter, going from $m = 1$ ("curves") to $m = 2$ ("surfaces") on up to $m = n - 1$ (hypermanifolds). We do not deal with $m = 0$, as 0-dimensional "manifolds" would just consist of isolated points in $\mathbb{R}^n$, and no actual calculus could be done. At the other extreme, $m = n$, we would be talking about n-dimensional manifolds contained in $\mathbb{R}^n$ and these are simply open subsets of $\mathbb{R}^n$. This leads essentially to the "flat" calculus we have studied in detail in Chapters 2–4 and presents no new ideas at the present time.

Again we shall use both an *implicit* presentation and an *explicit* presentation of M. In addition we shall also consider a third method, a *parametric* presentation. Here is a preliminary summary:

IMPLICIT M is defined by $n - m$ constraints placed on the points of $\mathbb{R}^n$.

EXPLICIT — M is defined by giving $n - m$ coordinates of $\mathbb{R}^n$ as explicit functions of the other m coordinates.

PARAMETRIC — M is defined by describing its points as a function of m other real variables (called *parameters*).

We devote the rest of this section to the discussion of three examples which happen to be quite interesting manifolds.

EXAMPLE 1. M is the ellipse $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$ in the $x_1 - x_2$ plane, but thought of as lying in $\mathbb{R}^3$. (The semiaxes a and b are arbitrary positive numbers.) As M is surely a 1-dimensional manifold, we have $m = 1$ and $n = 3$.

Implicitly, $M = \{x \in \mathbb{R}^3 \mid \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1, x_3 = 0\}$.

Explicitly, we "solve" the constraint equations near a point of M; for instance, near the point $(-a, 0, 0)$ we can write

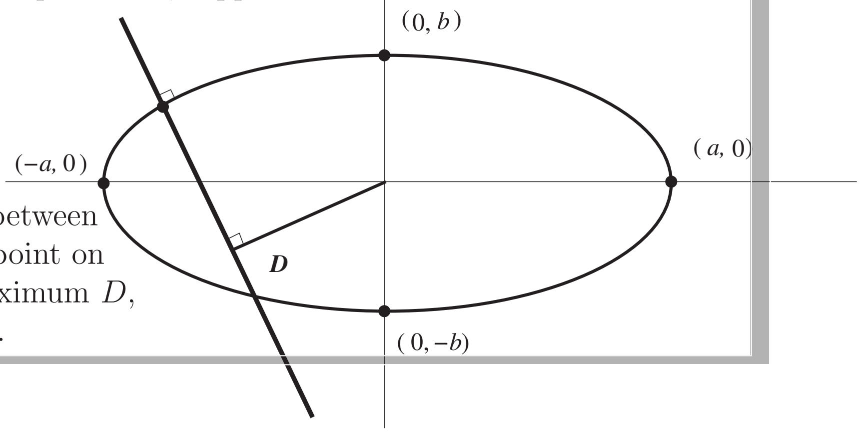
$$x_1 = -a\sqrt{1 - \frac{x_2^2}{b^2}}, \quad x_3 = 0.$$

Parametrically, we can express

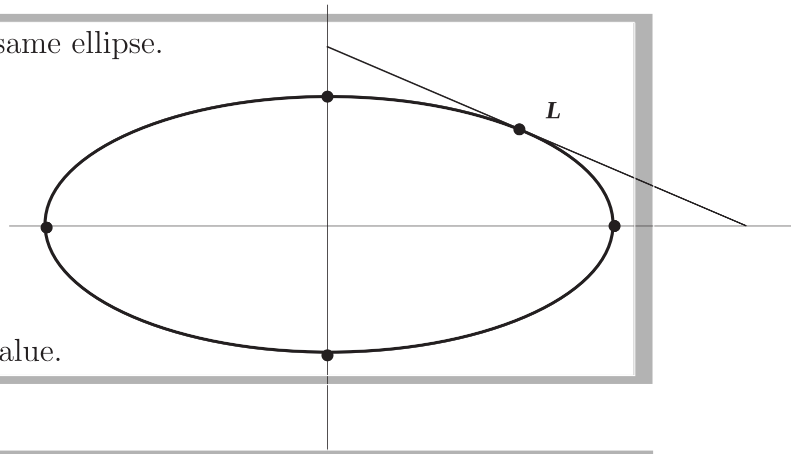
$$M = \{(a\cos t, b\sin t, 0) \mid 0 \le t \le 2\pi\};$$

thus all three coordinates are given as functions of the single parameter t.

PROBLEM 5–35. Given the ellipse above, suppose $0 < b < a$. Through any point on the ellipse draw the straight line which is orthogonal to the ellipse. This straight line contains a point which is closest to the origin; let D be the corresponding distance between this point and the origin. Find a point on the ellipse which produces the maximum D, and compute this maximum value.



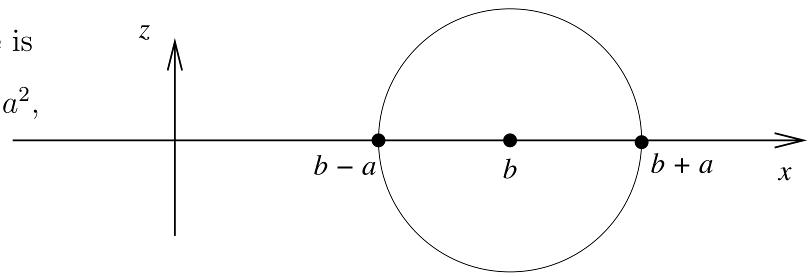
PROBLEM 5–36. Continue with the same ellipse. Through any point on it and in the first quadrant, draw the straight line which is tangent to the ellipse. This line contains a line segment whose end points lie on the coordinate axes. Let L be the length of this segment. Find a point on the ellipse which produces the minimum L, and compute this minimum value.



PROBLEM 5–37. In the preceding problem calculate the lengths of the two indicated line segments which constitute the optimal L, and show that they are a and b. (Do you see a simple geometric reason for this outcome?)

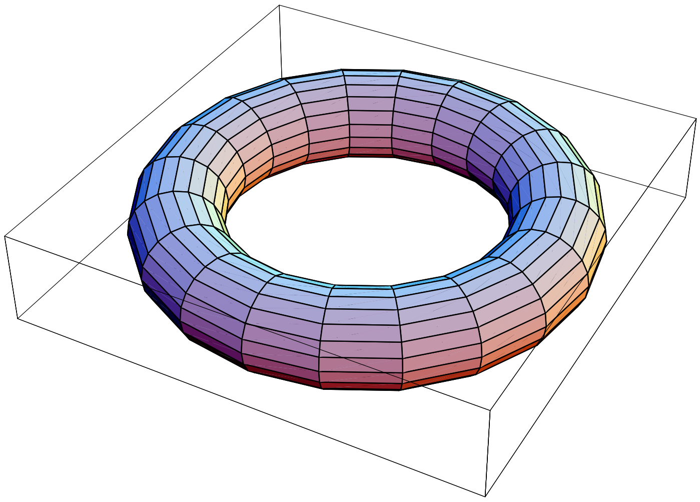
EXAMPLE 2. **Torus of revolution**. This is obtained by revolving a circle in $\mathbb{R}^3$ about an axis disjoint from it. We arrange things as in the illustration:

The circle in the $x - z$ plane is described as $(x-b)^2 + z^2 = a^2$, where $0 < a < b$.



Now revolve this around the z-axis. Starting with a point $(x, 0, z)$ on the circle produces points $(x', y', z) \in \mathbb{R}^3$ for which $\sqrt{x'^2 + y'^2} = x$. Thus the equation of the resulting surface is

$$\left(\sqrt{x^2+y^2} - b\right)^2 + z^2 = a^2. \qquad \textbf{Implicit presentation.}$$



Of course, $m = 2$ and $n = 3$.

We can also give a parametric presentation using two angles. The first angle ψ can parametrize the circle in the $x - z$ plane in the usual polar coordinate way: $x = b + a\cos\psi$, $z = a\sin\psi$. To get the revolved surface, we leave z alone, but use $b + a\cos\psi$ as the distance from the z-axis. If we have revolved through the angle θ, the x, y coordinates are then $x = (b + a\cos\psi)\cos\theta$, $y = (b + a\cos\psi)\sin\theta$. Thus we have

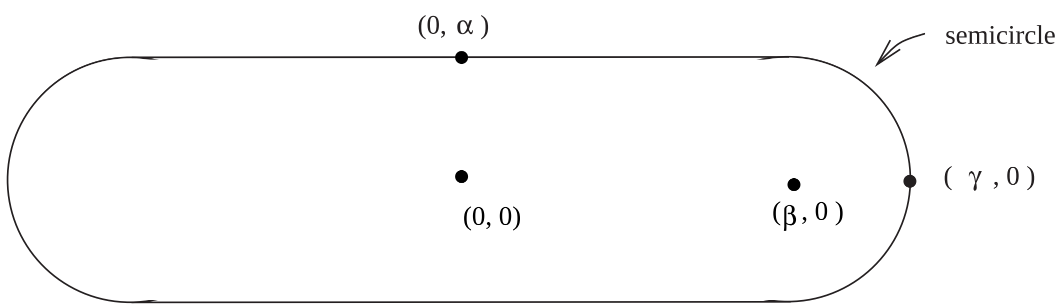
$$\begin{cases} x &= (b + a\cos\psi)\cos\theta, \\ y &= (b + a\cos\psi)\sin\theta, \\ z &= a\sin\psi. \end{cases}$$

Parametric presentation.

PROBLEM 5–38. Consider the given implicit presentation of the torus of revolution. The point $(b+a,0,0)$ lies on this torus. Show that near this point the torus has the **explicit** presentation

$$x = \sqrt{\left(b+\sqrt{a^2-z^2}\right)^2 - y^2}.$$

Show that the maximal region in the $y-z$ plane for which this representation is valid has the shape:



What are α, β, and γ?

In a certain sense this torus of revolution can be thought of as the Cartesian product of two circles, as two independent periodic coordinates ψ, θ are used in its presentation. However, it is not really a Cartesian product. There does exist a very interesting surface, a 2-dimensional manifold, which is actually the Cartesian product of two circles. As each circle is contained in $\mathbb{R}^2$, the Cartesian product we are going to exhibit is contained in $\mathbb{R}^2 \times \mathbb{R}^2 = \mathbb{R}^4$. This manifold is often called a **flat torus**:

EXAMPLE 3. Just as $\mathbb{R}^4$ is the Cartesian product $\mathbb{R}^2 \times \mathbb{R}^2$ of two planes, this 2-dimensional manifold M is literally the Cartesian product of two circles:

$$M = \{(x_1,x_2,x_3,x_4) \mid x_1^2+x_2^2=1, \quad x_3^2+x_4^2=1\}. \quad \textbf{Implicit presentation.}$$

We can use the polar coordinates for the two unit circles to write the points of M in the form

$$\begin{cases} x_1 &= \cos\theta_1, \\ x_2 &= \sin\theta_1, \\ x_3 &= \cos\theta_2, \\ x_4 &= \sin\theta_2. \end{cases} \qquad \textbf{Parametric} \text{ presentation.}$$

In a very definite sense to be explained later, this manifold is *flat*, unlike the torus of revolution.

At any fixed point $x \in M$, we can find two linearly independent vectors orthogonal to M, using the implicit presentation; and two linearly independent vectors tangent to M, using the parametric presentation. Namely, orthogonal vectors can be found by using the gradients of the two defining functions:

$$(x_1, x_2, 0, 0) \quad \text{and} \quad (0, 0, x_3, x_4).$$

And tangent vectors can be found by using the partial derivatives $\partial x/\partial \theta_i$ of the parametrizations:

$$(-\sin\theta_1, \cos\theta_1, 0, 0) \quad \text{and} \quad (0, 0, -\sin\theta_2, \cos\theta_2).$$

Thus we can split $\mathbb{R}^4$ into the 2-dimensional space orthogonal to M at x plus the 2-dimensional space tangent to M at x. We have just found the relevant orthonormal sequence:

$$\left.\begin{matrix}(x_1, x_2, 0, 0)\\(0, 0, x_3, x_4)\end{matrix}\right\} \qquad \text{orthogonal to } M,$$

$$\left.\begin{matrix}(-x_2, x_1, 0, 0)\\(0, 0, -x_4, x_3)\end{matrix}\right\} \qquad \text{tangent to } M.$$

Incidentally, we draw no pictures of the flat torus. It cannot be located in $\mathbb{R}^3$, whereas the torus of revolution is a hypermanifold in $\mathbb{R}^3$. Thus it gives us an interesting example of a 2-dimensional manifold in $\mathbb{R}^4$ which is not a hypersurface in $\mathbb{R}^3$.

These two manifolds are "topologically" indistinguishable. Without pausing to define the adjective, we go ahead and display a function

$$\text{flat torus} \xrightarrow{f} \text{torus of revolution}$$

in a rather obvious fashion. Namely,

$$f(x_1, x_2, x_3, x_4) = \big((b + ax_3)x_1, (b + ax_3)x_2, ax_4\big).$$

PROBLEM 5–39. Prove that f is a continuous bijection of the flat torus onto the torus of revolution. Prove that its inverse is given as

$$f^{-1}(x, y, z) = \left(\frac{x}{\sqrt{x^2 + y^2}}, \frac{y}{\sqrt{x^2 + y^2}}, \frac{\sqrt{x^2 + y^2} - b}{a}, \frac{z}{a}\right)$$

and that f^{-1} is also continuous.

Thus this function f provides a one-to-one correspondence between the points of these two tori, with the property that f and f^{-1} are both continuous. More than that, f and f^{-1} are both infinitely differentiable. Thus the flat torus and the "round" torus cannot be distinguished in the sense of manifolds.

However, these tori are *quite* different *geometrically*. The round one is really a curved surface, and the surface appears quite different at different points. However, though we are as yet not equipped to define the adjective "flat," it is rather clear that all points of the flat torus look alike from a geometric perspective. Thus the two tori are distinct geometric objects.

PROBLEM 5–40. Prove that at any point of the torus of revolution the vectors

$$\begin{aligned}&(-\sin\psi\cos\theta,\ -\sin\psi\sin\theta,\ \cos\psi),\\&(-\sin\theta,\ \cos\theta,\ 0),\\&(\cos\psi\cos\theta,\ \cos\psi\sin\theta,\ \sin\psi),\end{aligned}$$

form an orthonormal basis of $\mathbb{R}^3$, the first two being tangent to the torus and the third orthogonal to the torus.