CHAPTER 1

Manifolds and Vector Fields

Better is the end of a thing than the beginning thereof.

Ecclesiastes 7:8

AS students we learn differential and integral calculus in the context of euclidean space $\mathbb{R}^n$, but it is necessary to apply calculus to problems involving "curved" spaces. Geodesy and cartography, for example, are devoted to the study of the most familiar curved surface of all, the surface of planet Earth. In discussing maps of the Earth, latitude and longitude serve as "coordinates," allowing us to use calculus by considering functions on the Earth's surface (temperature, height above sea level, etc.) as being functions of latitude and longitude. The familiar Mercator's projection, with its stretching of the polar regions, vividly informs us that these coordinates are badly behaved at the poles: that is, that they are not defined everywhere; they are not "global." (We shall refer to such coordinates as being "local," even though they might cover a huge portion of the surface. Precise definitions will be given in Section 1.2.) Of course we may use two sets of "polar" projections to study the Arctic and Antarctic regions. With these three maps we can study the entire surface, provided we know how to relate the Mercator to the polar maps.

We shall soon define a "manifold" to be a space that, like the surface of the Earth, can be covered by a family of local coordinate systems. *A manifold will turn out to be the most general space in which one can use differential and integral calculus with roughly the same facility as in euclidean space.* It should be recalled, though, that calculus in $\mathbb{R}^3$ demands special care when curvilinear coordinates are required.

The most familiar manifold is N-dimensional euclidean space $\mathbb{R}^N$, that is, the space of ordered N tuples $(x^1, \ldots, x^N)$ of real numbers. Before discussing manifolds in general we shall talk about the more familiar (and less abstract) concept of a submanifold of $\mathbb{R}^N$, generalizing the notions of curve and surface in $\mathbb{R}^3$.

1.1. Submanifolds of Euclidean Space

What is the configuration space of a rigid body fixed at one point of $\mathbb{R}^n$?

1.1a. Submanifolds of $\mathbb{R}^N$

Euclidean space, $\mathbb{R}^N$, is endowed with a global coordinate system $(x^1, \ldots, x^N)$ and i
the most important example of a manifold.

In our familiar $\mathbb{R}^3$, with coordinates (x, y, z), a locus $z = F(x, y)$ describes a (2
dimensional) surface, whereas a locus of the form $y = G(x)$, $z = H(x)$, describes
(1-dimensional) curve. We shall need to consider higher-dimensional versions of thes
important notions.

A subset $M = M^n \subset \mathbb{R}^{n+r}$ is said to be an n-dimensional **submanifold** of $\mathbb{R}^{n+r}$
if *locally* M can be described by giving r of the coordinates differentiably in terms o
the n remaining ones. This means that given $p \in M$, a neighborhood of p on M ca
be described in *some* coordinate system $(x, y) = (x^1, \ldots, x^n, y^1, \ldots, y^r)$ of $\mathbb{R}^{n+r}$ b
r differentiable functions

$$y^\alpha = f^\alpha(x^1, \ldots, x^n), \qquad \alpha = 1, \ldots r$$

We abbreviate this by $y = f(x)$, or even $y = y(x)$. We say that $x^1, \ldots, x^n$ are **loca**
(curvilinear) coordinates for M near p.

Examples:

(i) $y^1 = f(x^1, \ldots, x^n)$ describes an n-dimensional submanifold of $\mathbb{R}^{n+1}$.

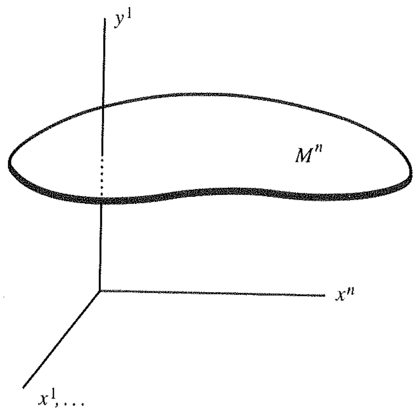

Figure 1.1

In Figure 1.1 we have drawn a portion of the submanifold M. This M is the **graph** of a function $f : \mathbb{R}^n \to \mathbb{R}$, that is, $M = \{(\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid y = f(\mathbf{x})\}$. When $n = 1$, M is a curve; while if $n = 2$, it is a surface.

(ii) The *unit sphere* $x^2 + y^2 + z^2 = 1$ in $\mathbb{R}^3$. Points in the northern hemisphere can be described by $z = F(x, y) = (1 - x^2 - y^2)^{1/2}$ and this function is differentiable everywhere except at the equator $x^2 + y^2 = 1$. Thus x and y are local coordinates for the northern hemisphere except at the equator. For points on the equator one can solve for x or y in terms of the others. If we have solved for x then y and z are the two local coordinates. For points in the southern hemisphere one can use the negative square root for z. The unit sphere in $\mathbb{R}^3$ is a 2-dimensional submanifold of $\mathbb{R}^3$. We note that we have *not* been able to describe the *entire* sphere by expressing one of the coordinates, say z, in terms of the two remaining ones, $z = F(x, y)$. We settle for local coordinates.

More generally, given r functions $F^\alpha(x_1, \dots, x_n, y_1, \dots, y_r)$ of $n + r$ variables, we may consider the locus $M^n \subset \mathbb{R}^{n+r}$ defined by the equations

$$F^\alpha(x, y) = c^\alpha, \qquad (c^1, \dots, c^r) \text{ constants}$$

If the **Jacobian** determinant

$$\left[\frac{\partial(F^1, \dots, F^r)}{\partial(y^1, \dots, y^r)}\right](x_0, y_0)$$

at $(x_0, y_0) \in M$ of the locus is not 0, the **implicit function theorem** assures us that locally, near (x_0, y_0), we may solve $F^\alpha(x, y) = c^\alpha$, $\alpha = 1, \dots, r$, for the y's in terms of the x's

$$y^\alpha = f^\alpha(x^1, \dots, x^n)$$

We may say that "a portion of M^n near (x_0, y_0) is a submanifold of $\mathbb{R}^{n+r}$." If the Jacobian $\neq 0$ *at all points of the locus*, then the entire M^n is a submanifold.

Recall that the Jacobian condition arises as follows. If $F^\alpha(x, y) = c^\alpha$ can be solved for the y's differentiably in terms of the x's, $y^\beta = y^\beta(x)$, then if, for fixed i, we differentiate the identity $F^\alpha(x, y(x)) = c^\alpha$ with respect to x^i, we get

$$\frac{\partial F^\alpha}{\partial x^i} + \sum_\beta \left[\frac{\partial F^\alpha}{\partial y^\beta}\right] \frac{\partial y^\beta}{\partial x^i} = 0$$

and

$$\frac{\partial y^\beta}{\partial x^i} = -\sum_\alpha \left(\left[\frac{\partial F}{\partial y}\right]^{-1}\right)^\beta{}_\alpha \left[\frac{\partial F^\alpha}{\partial x^i}\right]$$

provided the subdeterminant $\partial(F^1, \dots, F^r)/\partial(y^1, \dots, y^r)$ is not zero. (Here $([\partial F/\partial y]^{-1})^\beta{}_\alpha$ is the $\beta\alpha$ entry of the inverse to the matrix $\partial F/\partial y$; we shall use the convention that for matrix indices, the index to the *left* always is the *row* index, whether it is up or down.) This *suggests* that if the indicated Jacobian is nonzero then we might indeed be able to solve for the y's in terms of the x's, and the implicit function theorem confirms this. The (nontrivial) *proof* of the implicit function theorem can be found in most books on real analysis.

Still more generally, suppose that we have r functions of $n+r$ variables, $F^\alpha(x^1, \dots, x^{n+r})$. Consider the locus $F^\alpha(x) = c^\alpha$. Suppose that at each point x_0 of the locus the Jacobian *matrix*

$$\left(\frac{\partial F^\alpha}{\partial x^i}\right) \qquad \alpha = 1, \dots, r \qquad i = 1, \dots, n + r$$

has rank r. Then the equations $F^\alpha = c^\alpha$ define an n-dimensional submanifold of $\mathbb{R}^{n+r}$, since we may locally solve for r of the coordinates in terms of the remaining n.

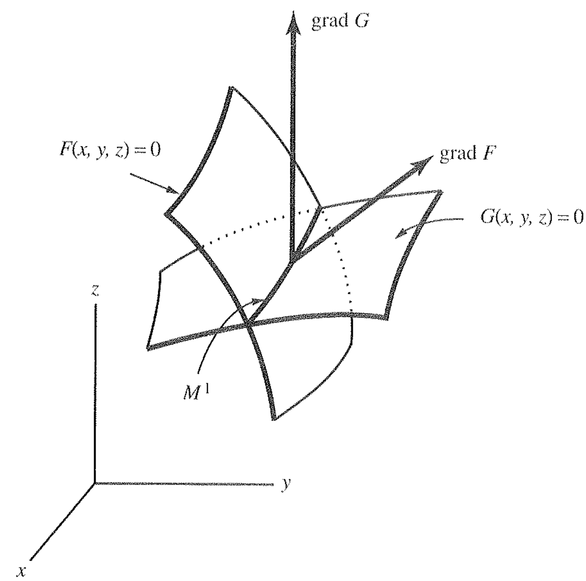

Figure 1.2

In Figure 1.2, two surfaces $F = 0$ and $G = 0$ in $\mathbb{R}^3$ intersect to yield a curve M.

The simplest case is *one* function F of N variables $(x^1, \ldots, x^N)$. If *at each point of the locus* $F = c$ there is always *at least one partial derivative that does not vanish*, then the Jacobian (row) matrix $[\partial F/\partial x^1, \partial F/\partial x^2, \ldots, \partial F/\partial x^N]$ has rank 1 and we may conclude that *this locus is indeed an* $(N-1)$*-dimensional submanifold of* $\mathbb{R}^N$. This criterion is easily verified, for example, in the case of the 2-sphere $F(x, y, z) = x^2 + y^2 + z^2 = 1$ of Example (ii). The column version of this row matrix is called in calculus the gradient vector of F. In $\mathbb{R}^3$ this vector

$$\begin{bmatrix} \frac{\partial F}{\partial x} \\ \frac{\partial F}{\partial y} \\ \frac{\partial F}{\partial z} \end{bmatrix}$$

is orthogonal to the locus $F = 0$, and we may conclude, for example, that if this gradient vector has a nontrivial component in the z direction at a point of $F = 0$, then locally we can solve for $z = z(x, y)$.

A submanifold of dimension $(N-1)$ in $\mathbb{R}^N$, that is, of "**codimension**" 1, is called a **hypersurface**.

(iii) The x axis of the xy plane $\mathbb{R}^2$ can be described (perversely) as the locus of the quadratic $F(x, y) := y^2 = 0$. Both partial derivatives vanish on the locus, the x axis, and our criteria would not allow us to say that the x axis is a 1-dimensional submanifold of $\mathbb{R}^2$. Of course the x axis *is* a submanifold; we should have used the usual description $G(x, y) := y = 0$. Our Jacobian criteria are *sufficient* conditions, not necessary ones.

(iv) The locus $F(x, y) := xy = 0$ in $\mathbb{R}^2$, consisting of the union of the x and y axes, is not a 1-dimensional submanifold of $\mathbb{R}^2$. It seems "clear" (and can be proved) that in a neighborhood of the intersection of the two lines we are not going to be able to describe the locus in the form of $y = f(x)$ or $x = g(y)$, where f, g, are differentiable functions. The best we can say is that this locus *with the origin removed* is a 1-dimensional submanifold.

1.1b. The Geometry of Jacobian Matrices: The "Differential"

The **tangent space** to $\mathbb{R}^n$ at the point x, written here as $\mathbb{R}^n_x$, is by definition the vector space of all vectors in $\mathbb{R}^n$ based at x (i.e., it is a copy of $\mathbb{R}^n$ with origin shifted to x).

Let $x^1, \ldots, x^n$ and $y^1, \ldots, y^r$ be coordinates for $\mathbb{R}^n$ and $\mathbb{R}^r$ respectively. Let $F : \mathbb{R}^n \to \mathbb{R}^r$ be a **smooth** map. ("Smooth" ordinarily means infinitely differentiable. For our purposes, however, it will mean differentiable at least as many times as is necessary in the present context. For example, if F is once continuously differentiable, we may use the chain rule in the argument to follow.) In coordinates, F is described by giving r functions of n variables

$$y^\alpha = F^\alpha(x) \qquad \alpha = 1, \ldots, r$$

or simply $y = F(x)$. We will frequently use the more dangerous notation $y = y(x)$.

Let $y_0 = F(x_0)$; the Jacobian *matrix* $(\partial y^\alpha / \partial x^i)(x_0)$ has the following significance.

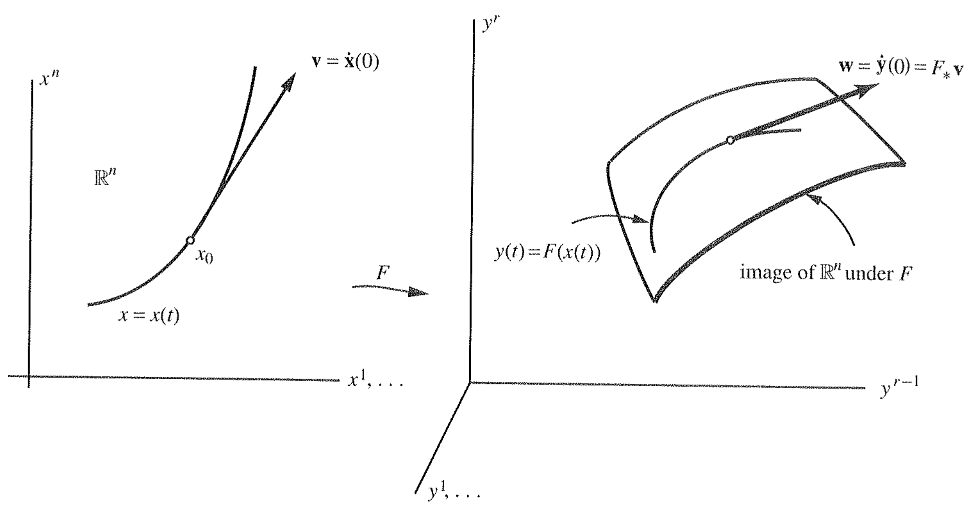

Figure 1.3

Let $\mathbf{v}$ be a tangent vector to $\mathbb{R}^n$ at x_0. Take *any* smooth curve $x(t)$ such that $x(0) = x_0$ and $\dot{x}(0) := (dx/dt)(0) = \mathbf{v}$, for example, the straight line $x(t) = x_0 + t\mathbf{v}$. The image of this curve

$$y(t) = F(x(t))$$

has a tangent vector $\mathbf{w}$ at y_0 given by the chain rule

$$w^\alpha = \dot{y}^\alpha(0) = \sum_{i=1}^{n} \left(\frac{\partial y^\alpha}{\partial x^i}\right)(x_0)\dot{x}^i(0) = \sum_{i=1}^{n} \left(\frac{\partial y^\alpha}{\partial x^i}\right)(x_0)v^i$$

The assignment $\mathbf{v} \mapsto \mathbf{w}$ is, from this expression, independent of the curve $x(t)$ chosen, and defines a *linear transformation*, the **differential** of F at x_0

$$F_* : \mathbb{R}^n_{x_0} \to \mathbb{R}^r_{y_0} \qquad F_*(\mathbf{v}) = \mathbf{w} \tag{1.1}$$

whose matrix is simply the Jacobian matrix $(\partial y^\alpha/\partial x^i)(x_0)$. *This interpretation of the Jacobian matrix, as a linear transformation sending tangents to curves into tangents to the image curves under F, can sometimes be used to replace the direct computation of matrices.* This philosophy will be illustrated in Section 1.1d.

1.1c. The Main Theorem on Submanifolds of $\mathbb{R}^N$

The main theorem is a geometric interpretation of what we have discussed. Note that the statement "F has rank r at x_0," that is, $[\partial y^\alpha/\partial x^i](x_0)$ has rank r, is geometrically the statement that the differential

$$F_* : \mathbb{R}^n_{x_0} \to \mathbb{R}^r_{y_0=F(x_0)}$$

is **onto** or "surjective"; that is, given any vector **w** at y_0 there is at least one vector **v** at x_0 such that $F_*(\mathbf{v}) = \mathbf{w}$. We then have

Theorem (1.2): *Let $F : \mathbb{R}^{r+n} \to \mathbb{R}^r$ and suppose that the locus*

$$F^{-1}(y_0) := \{x \in \mathbb{R}^{r+n} \mid F(x) = y_0\}$$

is not empty. Suppose further that for all $x_0 \in F^{-1}(y_0)$

$$F_* : \mathbb{R}^{n+r}_{x_0} \to \mathbb{R}^r_{y_0}$$

is onto. Then $F^{-1}(y_0)$ is an n-dimensional submanifold of $\mathbb{R}^{n+r}$.

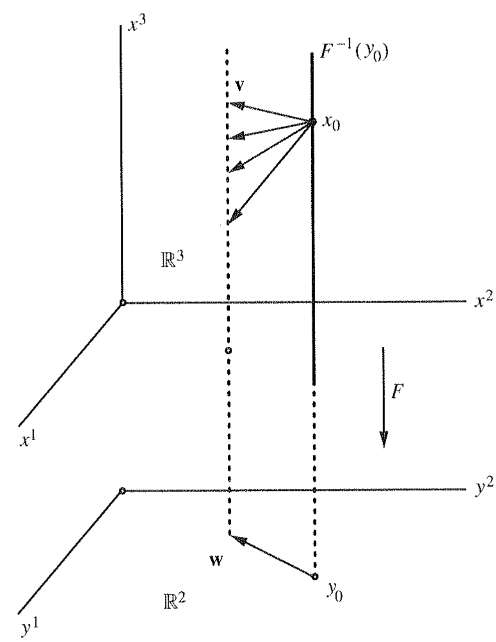

Figure 1.4

The best example to keep in mind is the linear "projection" $F : \mathbb{R}^3 \to \mathbb{R}^2$, $F(x^1, x^2, x^3) = (x^1, x^2)$, that is, $y^1 = x^1$ and $y^2 = x^2$. In this case, x^3 serves as global coordinate for the submanifold $x^1 = y_0^1, x^2 = y_0^2$, that is, the vertical line.

1.1d. A Nontrivial Example: The Configuration Space of a Rigid Body

Assume a rigid body has one point, the origin of $\mathbb{R}^3$, fixed. By comparing a cartesian right-handed system fixed in the body with that of $\mathbb{R}^3$ we see that the configuration of the body at any time is described by the rotation matrix taking us from the basis of $\mathbb{R}^3$ to the basis fixed in the body. The configuration space of the body is then the **rotation group** SO(3), that is, the 3×3 real matrices $x = (x_{ij})$ such that

$$x^T = x^{-1} \quad \text{and} \quad \det x > 0$$

where T denotes transpose. (If we omit the determinant condition, the group is the full **orthogonal** group, O(3).) By assigning (in some fixed order) the nine coordinates $x_{11}, x_{12}, \dots, x_{33}$ to *any* matrix x, we see that the space of all 3×3 real matrices, $M(3 \times 3)$, is the euclidean space $\mathbb{R}^9$. The group O(3) is then the locus in this $\mathbb{R}^9$ defined by the equations $x^T x = I$, that is, by the system of nine quadratic equations (i, k)

$$(i,k) \qquad \sum_{j=1}^{3} x_{ji} x_{jk} = \delta_{ik}$$

We then have the following situation. The configuration of the body at time t can be represented by a point $x(t)$ in $\mathbb{R}^9$, but in fact the point $x(t)$ lies on the locus O(3) in $\mathbb{R}^9$. We shall see shortly that *this locus is in fact a 3-dimensional submanifold* of $\mathbb{R}^9$. As time t evolves, the point $x(t)$ traces out a curve on this 3-dimensional locus. Since O(3) *is* a submanifold, we shall see, in Section 10.2c from the principle of least action, that this path is a very special one, a "geodesic" on the submanifold O(3), and this in turn will yield important information on the existence of periodic motions of the body even when the body is subject to an unusual potential field. All this depends on the fact that O(3) is a submanifold, and we turn now to the proof of this crucial result.

Note first that since $x^T x$ is a symmetric matrix, equation (i, k) is the same as equation (k, i); there are, then, only 6 independent equations. This suggests the following. Let

$$\text{Sym}^6 := \{x \in M(3 \times 3) \mid x^T = x\}$$

be the space of all *symmetric* 3×3 matrices. Since this is defined by the three *linear* equations $x_{ik} - x_{ki} = 0, i \neq k$, we see that Sym^6 is a 6-dimensional linear subspace of $\mathbb{R}^9$; that is, it can be considered as a copy of $\mathbb{R}^6$. To exhibit O(3) as a locus in $\mathbb{R}^9$, we consider the map

$$F : \mathbb{R}^9 \to \mathbb{R}^6 = \text{Sym}^6 \quad \text{defined by } F(x) = x^T x - I$$

O(3) is then the locus $F^{-1}(0)$. Let $x_0 \in F^{-1}(0) = \text{O}(3)$. We shall show that $F_* : \mathbb{R}^9_{x_0} \to \text{Sym}^6$ is onto.

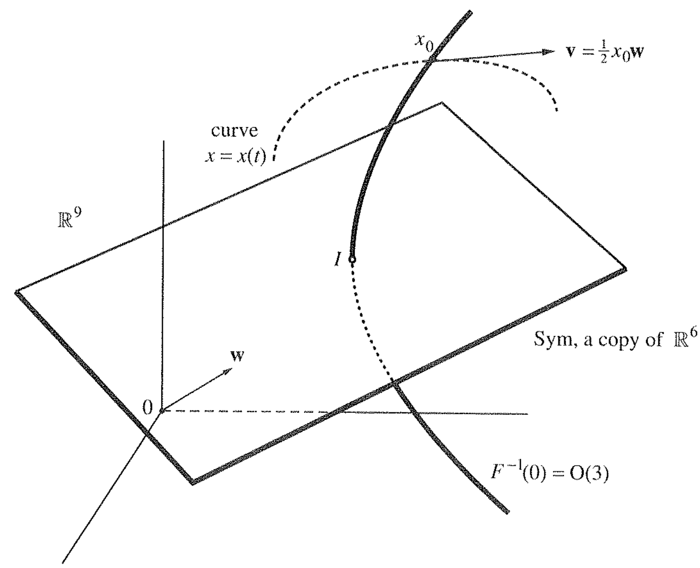

Figure 1.5

Let $\mathbf{w}$ be tangent to Sym^6 at the zero matrix. As usual, we identify a vector at the origin of $\mathbb{R}^n$ with its endpoint. *Then* $\mathbf{w}$ *is itself a symmetric matrix.* We must find $\mathbf{v}$, a tangent vector to $\mathbb{R}^9$ at x_0, such that $F_*\mathbf{v} = \mathbf{w}$. Consider a general curve $x = x(t)$ of matrices such that $x(0) = x_0$; its tangent vector at x_0 is $\dot{x}(0)$. The image curve

$$F(x(t)) = x(t)^T x(t) - I$$

has tangent at $t = 0$ given by

$$\frac{d}{dt}[F(x(t))]_{t=0} = \dot{x}(0)^T x_0 + x_0^T \dot{x}(0)$$

We wish this quantity to be $\mathbf{w}$. You should verify that it is sufficient to satisfy the matrix equation $x_0^T \dot{x}(0) = \mathbf{w}/2$. Since $x_0 \in O(3)$, $x_0^T = x_0^{-1}$ and we have as solution the matrix product $\mathbf{v} = \dot{x} = x_0\mathbf{w}/2$. Thus F_* *is* onto at x_0 and by our main theorem O(3)$= F^{-1}(0)$ is a $(9-6) = 3$-dimensional submanifold of $\mathbb{R}^9$.

What about the subset SO(3) of O(3)? Recall that each orthogonal matrix has determinant ± 1, whereas SO(3) consists of those orthogonal matrices with determinant $+1$. The mapping

$$\det : \mathbb{R}^9 \to \mathbb{R}$$

that sends each matrix x into its determinant is continuous (it is a cubic polynomial function of the coordinates x_{ik}) and consequently the two subsets of O(3) where det is $+1$ and where det is -1 must be separated. This means that SO(3) itself must have the property that it is locally described by giving 6 of the coordinates in terms of the remaining 3, that is, SO(3) is a 3-dimensional submanifold of $\mathbb{R}^9$.

Thus the configuration space of a rigid body with one point fixed is the group SO(3). *This is a 3-dimensional submanifold of* $\mathbb{R}^9$. *Each point of this configuration space lies in some local curvilinear coordinate system.*

In physics books the coordinates in an n-dimensional configuration space are usually labeled $q^1, \dots, q^n$. For SO(3) physicists usually use the three "Euler angles" as coordinates. These coordinates do not cover all of SO(3) in the sense that they become singular at certain points, just as polar coordinates in the plane are singular at the origin.

Problems

1.1(1) Investigate the locus $x^2 + y^2 - z^2 = c$ in $\mathbb{R}^3$, for $c > 0$, $c = 0$, and $c < 0$. Are they submanifolds? What if the origin is omitted? Draw all three loci, for $c = 1, 0, -1$, in one picture.

1.1(2) SO(n) is defined to be the set of all *orthogonal* $n \times n$ matrices x with $\det x = 1$. The preceding discussion of SO(3) extends immediately to SO(n). What is the dimension of SO(n) and in what euclidean space is it a submanifold?

1.1(3) Is the **special linear group**

$$\mathrm{Sl}\,(n) := \{n \times n \text{ real matrices } x \mid \det x = 1\}$$

a submanifold of some $\mathbb{R}^N$? Hint: You will need to know something about $\partial/\partial x_{ij}$ (det x); expand the determinant by the j^{th} column.This is an example where it might be easier to deal directly with the Jacobian matrix rather than the differential.

1.1(4) Show, in $\mathbb{R}^3$, that if the cross product of the gradients of F and G has a nontrivial component in the x direction at a point of the intersection of $F = 0$ and $G = 0$, then x can be used as local coordinate for this curve.

1.2. Manifolds

In learning the sciences examples are of more use than precepts.

Newton, *Arithmetica Universalis* (1707)

The notion of a "topology" will allow us to talk about "continuous" functions and points "neighboring" a given point, in spaces where the notion of distance and metric might be lacking.

The cultivation of an intuitive "feeling" for manifolds is of more importance, at this stage, than concern for topological details, but some basic notions from point set topology are helpful. The reader for whom these notions are new should approach them as one approaches a new language, with some measure of fluency, it is hoped, coming later.

In Section 1.2c we shall give a technical (i.e., complete) definition of a manifold.

1.2a. Some Notions from Point Set Topology

The **open ball** in $\mathbb{R}^n$, of radius ϵ, centered at $\mathbf{a} \in \mathbb{R}^n$ is

$$B_{\mathbf{a}}(\epsilon) = \{\mathbf{x} \in \mathbb{R}^n \mid \| \mathbf{x} - \mathbf{a} \| < \epsilon\}$$

The **closed ball** is defined by

$$\overline{B}_{\mathbf{a}}(\epsilon) = \{\mathbf{x} \in \mathbb{R}^n \mid \| \mathbf{x} - \mathbf{a} \| \leq \epsilon\}$$

that is, the closed ball is the open ball with its edge or boundary included.

A set U *in* $\mathbb{R}^n$ is declared **open** if given any $\mathbf{a} \in U$ there is an open ball of some radius $r > 0$, centered at **a**, that lies entirely in U. Clearly each $B_{\mathbf{b}}(\epsilon)$ is open if $\epsilon > 0$ (take $r = (\epsilon - \| \mathbf{b} - \mathbf{a} \|)/2$), whereas $\overline{B}_{\mathbf{b}}(\epsilon)$ is not open because of its boundary points. $\mathbb{R}^n$ itself is trivially open. The empty set is technically open since there are no points **a** in it.

A set F *in* $\mathbb{R}^n$ is declared **closed** if its complement $\mathbb{R}^n - F$ is open. It is easy to check that each $\overline{B}_{\mathbf{a}}(\epsilon)$ is a closed set, whereas the open ball is not. Note that the entire space $\mathbb{R}^n$ is both open and closed, since its complement is empty.

It is immediate that the *union* of *any* collection of open sets in $\mathbb{R}^n$ is an open set, and it is not difficult to see that the *intersection* of any *finite* number of open sets in $\mathbb{R}^n$ is open.

We have described explicitly the "usual" open sets in euclidean space $\mathbb{R}^n$. What do we mean by an open set in a more general space? We shall define the notion of open set axiomatically.

A **topological space** is a set M with a distinguished collection of subsets, to be called the **open** sets. These open sets must satisfy the following.

1. Both M and the empty set are open.
2. If U and V are open sets, then so is their intersection $U \cap V$.
3. The union of any collection of open sets is open.

These open subsets "define" **the topology** of M. □

(A different collection might define a different topology.) Any such collection of subsets that satisfies 1, 2, and 3 is eligible for defining a topology in M. In our introductory discussion of open balls in $\mathbb{R}^n$ we also defined the collection of open subsets of $\mathbb{R}^n$. These define the topology of $\mathbb{R}^n$, the "usual" topology. An example of a "perverse" topology on $\mathbb{R}^n$ is the **discrete** topology, in which *every* subset of $\mathbb{R}^n$ is declared open! In discussing $\mathbb{R}^n$ in this book we shall always use the usual topology.

A subset of M is **closed** if its complement is open.

Let A be any subset of a topological space M. Define a topology for the space A (the **induced** or **subspace** topology) by declaring $V \subset A$ to be an open subset of A provided V is the intersection of A with some open subset U of M, $V = A \cap U$. These sets *do* define a topology for A. For example, let A be a line in the plane $\mathbb{R}^2$. An open ball in $\mathbb{R}^2$ is simply a disc without its edge. This disc either will not intersect A or will intersect A in an "interval" that does not contain its endpoints. This interval will be an open set in the induced topology on the line A. It can be shown that any open set in A will be a union of such intervals.

Any *open* set in M that contains a point $x \in M$ will be called a **neighborhood** of x.

If $F{:}M \to N$ is a map of a topological space M into a topological space N, we say that F is **continuous** if for every open set $V \subset N$, the **inverse image** $F^{-1}V := \{x \in M \mid F(x) \in V\}$ is open in M. (This reduces to the usual ϵ, δ definition in the case where M and N are euclidean spaces.) The map sending all of $\mathbb{R}^n$ into a single point of $\mathbb{R}^m$ is an example showing that a continuous map need not send open sets into open sets.

If $F:M \to N$ is one to one (1 : 1) and onto, then the inverse *map* $F^{-1}:N \to M$ exists. If further both F and F^{-1} are continuous, we say that F is a **homeomorphism** and that M and N are homeomorphic. A homeomorphism takes open (closed) sets into open (closed) sets. Homeomorphic spaces are to be considered to be "the same" as topological spaces; we say that they are "topologically the same." It can be proved that $\mathbb{R}^n$ and $\mathbb{R}^m$ are homeomorphic if and only if $m = n$.

The technical definition of a manifold requires two more concepts, namely "Hausdorff" and "countable base." We shall not discuss these here since they will not arise *explicitly* in the remainder of the book. The reader is referred to [S] for questions concerning point set topology.

There is one more concept that plays a very important role, though not needed for the definition of a manifold; the reader may prefer to come back to this later on when needed.

A topological space X is called **compact** if from *every* covering of X by open sets one can pick out a *finite* number of the sets that still covers X. For example, the open interval (0,1), considered as a subspace of $\mathbb{R}$, is *not* compact; we cannot extract a finite subcovering from the open covering given by the sets $U_n = \{x \mid 1/n < x < 1\} n = 1, 2, \ldots$.

On the other hand, the closed interval [0,1] is a compact space. In fact, it is shown in every topology book that *any subset X of $\mathbb{R}^n$* (with the induced topology) *is compact if and only if*

1. X is a *closed* subset of $\mathbb{R}^n$,
2. X is a *bounded* subset, that is, $\| \mathbf{x} \| <$ some number c, for all $\mathbf{x} \in X$.

Finally we shall need two properties of continuous maps. First

The continuous image of a compact space is itself compact.

PROOF: If $f : G \to M$ is continuous and if $\{U_i\}$ is an open cover of $f(G) \subset M$, then $\{f^{-1}(U_i)\}$ is an *open* cover of G. Since G is compact we can extract a finite open subcover $\{f^{-1}(U_\alpha)\}$ of G, and then $\{U_\alpha\}$ is a finite subcover of $f(G)$. □

Furthermore

A continuous real-valued function $f : G \to \mathbb{R}$ on a compact space G is bounded.

PROOF: $F(G)$ is a compact subspace of $\mathbb{R}$, and thus is closed and bounded. □

1.2b. The Idea of a Manifold

An n-dimensional (differentiable) manifold M^n (briefly, an n-manifold) is a topological space that is locally $\mathbb{R}^n$ in the following sense. It is covered by a family of local (curvilinear) coordinate systems $\{U; x_U^1, \ldots, x_U^n\}$, consisting of open sets or "patches" U and coordinates x_U in U, such that a point $p \in U \cap V$ that lies in two coordinate patches will have its two sets of coordinates related differentiably

$$x_V^i(p) = f_{VU}^i(x_U^1, \ldots, x_U^n) \qquad i = 1, 2, \ldots, n. \tag{1.3}$$

(If the functions f_{VU} are C^∞, that is, infinitely differentiable, or real analytic, ..., we say that M is C^∞, or real analytic, ...) There are more requirements; for example, we shall demand that each coordinate patch is homeomorphic to some open subset of $\mathbb{R}^n$. Some of these requirements will be mentioned in the following examples, but details will be spelled out in Section 1.2c.

Examples:

(i) $M^n = \mathbb{R}^n$, covered by a single coordinate system. The condition (1.3) is vacuous.

(ii) M^n is an open ball in $\mathbb{R}^n$, again covered by one patch.

(iii) The *closed* ball in $\mathbb{R}^n$ is *not* a manifold. It can be shown that a point on the edge of the ball can never have a neighborhood that is homeomorphic to an *open* subset of $\mathbb{R}^n$. For example, with $n = 1$, a half open interval $0 \le x < 1$ in $\mathbb{R}^1$ can never be homeomorphic to an open interval $0 < x < 1$ in $\mathbb{R}^1$.

(iv) $M^n = S^n$, the unit *sphere* in $\mathbb{R}^{n+1}$. We shall illustrate this with the familiar case $n = 2$. We are dealing with the locus $x^2 + y^2 + z^2 = 1$.

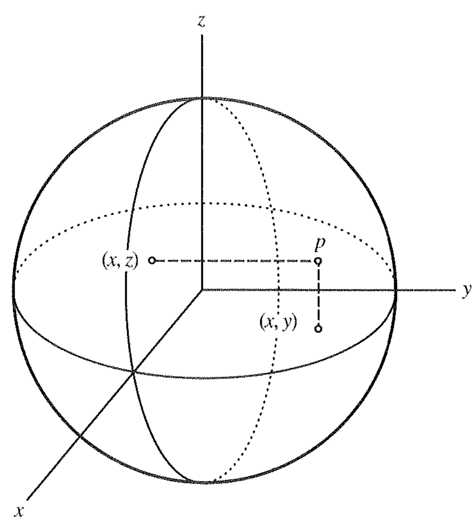

Figure 1.6

Cover S^2 with six "open" subsets (patches)

$$U_x+ = \{p \in S^2 \mid x(p) > 0\} \qquad U_x- = \{p \in S^2 \mid x(p) < 0\}$$
$$U_y+ = \{p \in S^2 \mid y(p) > 0\} \qquad U_y- = \{p \in S^2 \mid y(p) < 0\}$$
$$U_z+ = \{p \in S^2 \mid z(p) > 0\} \qquad U_z- = \{p \in S^2 \mid z(p) < 0\}$$

The point p illustrated sits in $[U_x+] \cap [U_y+] \cap [U_z+]$. Project U_z+into the xy plane; this introduces x and y as curvilinear coordinates in U_z+.

Do similarly for the other patches. For $p \in [U_y+] \cap [U_z+]$, p is assigned the two sets of coordinates $\{(u_1, u_2) = (x, z)\}$ and $\{(v_1, v_2) = (x, y)\}$ arising from the two projections

$$\pi_{xz} : U_y \to xz \text{ plane} \quad \text{and} \quad \pi_{xy} : U_z \to xy \text{ plane}$$

These are related by $v_1 = u_1$ and $v_2 = +[1 - u_1^2 - u_2^2]^{1/2}$; these are differentiable functions provided $u_1^2 + u_2^2 < 1$, and this is satisfied since $p \in U_y+$.

S^2 is "locally $\mathbb{R}^2$." The indicated point p has a neighborhood (in the topology of S^2 induced as a subset of $\mathbb{R}^3$) that is homeomorphic, via the projection π_{xy}, say, to an open subset of $\mathbb{R}^2$ (in this case an open subset of the xy plane). We say that a manifold is **locally euclidean.**

If two sets of coordinates are related differentiably in an overlap we shall say that they are **compatible**. On S^2 we could introduce, in addition to the preceding coordinates, the usual spherical coordinates θ and ϕ, representing colatitude and longitude. They do not work for the entire sphere (e.g., at the poles) but where they do work they are compatible with the original coordinates.

We could also introduce (see Section 1.2d) coordinates on S^2 via stereographic projection onto the planes $z = 1$ and $z = -1$, again failing at the south and north pole, respectively, but otherwise being compatible with the previous coordinates. On a manifold we should allow the use of *all* coordinate systems that are compatible with those that originally were used to define the manifold. Such a collection of compatible coordinate systems is called a **maximal atlas**.

(v) If M^n is a manifold with local coordinates $\{U; x^1, \ldots, x^n\}$ and W^r is a manifold with local coordinates $\{V; y^1, \ldots, y^r\}$, we can form the **product manifold**

$$L^{n+r} = M^n \times W^r = \{(p, q) \mid p \in M^n \text{ and } q \in W^r\}$$

by using $x^1, \ldots, x^n, y^1, \ldots, y^r$ as local coordinates in $U \times V$.

S^1 is simply the unit circle in the plane $\mathbb{R}^2$; it has a local coordinate $\theta = \tan^{-1}(y/x)$, using any branch of the multiple-valued function θ. One must use at least two such coordinates (branches) to cover S^1. "Topologically" S^1 is conveniently represented by an interval on the real line $\mathbb{R}$ with endpoints identified; by this we mean that there is a homeomorphism between these two models. In order to talk about a homeomorphism

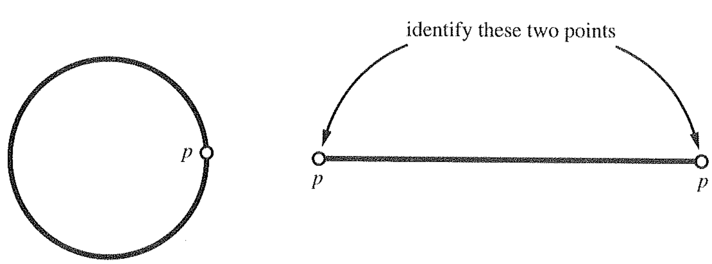

Figure 1.7

we would first have to define the topology in the space consisting of the interval with endpoints identified; it clearly is not the same space as the interval without the identification. To define a topology, we may simply consider the map $F : [0 \le \theta \le 2\pi] \to \mathbb{R}^2 = \mathbb{C}$ defined by $F(\theta) = e^{i\theta}$. It sends the endpoints $\theta = 0$ and $\theta = 2\pi$ to the point $p = 1$ on the unit circle in the complex plane. This map is $1 : 1$ and onto if we identify the endpoints. The unit circle has a topology induced from that of the plane, built up from little curved intervals. We can construct open subsets of the interval by taking the inverse images under F of such sets. (What then is a neighborhood of the endpoint p?) By using this topology we force F to be a homeomorphism.

S^1 is the configuration space for a rigid *pendulum* constrained to oscillate in the plane

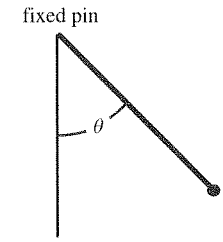

Figure 1.8

The n-dimensional **torus** $T^n := S^1 \times S^1 \times \cdots \times S^1$ has local coordinates given by the n-angular parameters $\theta^1, \ldots, \theta^n$. Topologically it is the n cube (the product of n intervals) with identifications. For $n = 2$

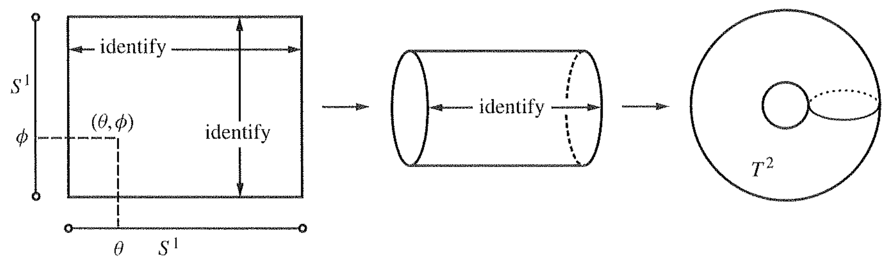

Figure 1.9

T^2 is the configuration space of a planar *double pendulum*. It might be thought that it is simpler to picture the double pendulum itself rather than the seemingly abstract version of a 2-dimensional torus. We shall see in Section 10.2d that this abstract picture allows us to conclude, for example, that *a double pendulum, in an arbitrary potential field, always has periodic motions in which the upper pendulum makes p revolutions while the lower makes q revolutions.*

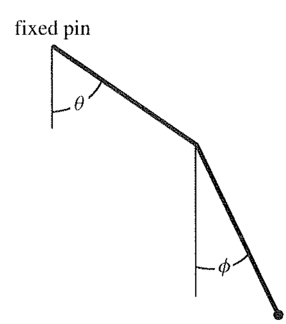

Figure 1.10

(vi) The **real projective** n **space** $\mathbb{R}P^n$ is the space of all *unoriented* lines L through the origin of $\mathbb{R}^{n+1}$. We illustrate with the *projective plane* of lines through the origin of $\mathbb{R}^3$.

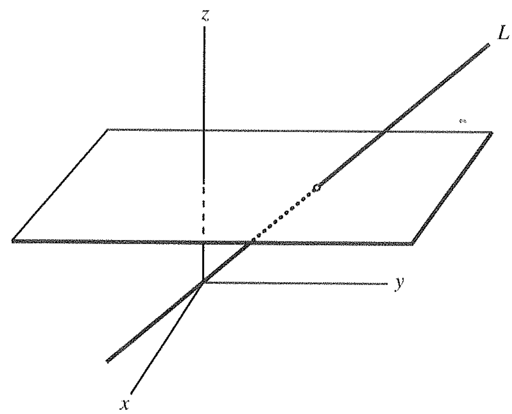

Figure 1.11

Such a line L is completely determined by any point (x, y, z) on the line, other than the origin, but note that (ax, ay, az) represents the same line if $a \neq 0$. We should really use the ratios of coordinates to describe a line. We proceed as follows.

We cover $\mathbb{R}P^2$ by three sets:

$$U_x := \text{those lines not lying in the } yz \text{ plane}$$

$$U_y := \text{those lines not lying in the } xz \text{ plane}$$

$$U_z := \text{those lines not lying in the } xy \text{ plane}$$

Introduce coordinates in the U_z patch; if $L \in U_z$, choose any point (x, y, z) on L other than the origin and define (since $z \neq 0$)

$$u_1 = \frac{x}{z}, \qquad u_2 = \frac{y}{z}$$

Do likewise for the other two patches. In Problem 1.2(1) you are asked to show that these patches make $\mathbb{R}P^2$ into a 2-dimensional manifold.

These coordinates are the most convenient for analytical work. Geometrically, the coordinates u_1 and u_2 are simply the xy coordinates of the point where L intersects the plane $z = 1$.

Consider a point in $\mathbb{R}P^2$; it represents a line through the origin 0. Let (x, y, z) be a point other than the origin that lies on this line. We may represent this line by the triple $[x, y, z]$, called the **homogeneous coordinates** of the point in $\mathbb{R}P^2$ where we must identify $[x, y, z]$ with $[\lambda x, \lambda y, \lambda z]$ for all $\lambda \neq 0$. They are not true coordinates in our sense.

We have suceeded in "parameterizing" the *set* of undirected lines through the origin by means of a manifold, $M^2 = \mathbb{R}P^2$. *A manifold is a generalized parameterization of some set of objects*. $\mathbb{R}P^2$ *is* the set of undirected lines through the origin; each point of $\mathbb{R}P^2$ is an entire line in $\mathbb{R}^3$ and $\mathbb{R}P^2$ is a *global* object. If, however, one insists on describing a particular line L by coordinates, that is, pairs of *numbers* (u, v), then this can, in general, only be done *locally*, by means of the manifold's local coordinates.

Note that if we had been considering *directed* lines, then the manifold in question would have been the sphere S^2, since each directed line $\vec{L}$ could be uniquely defined by the "forward" point where $\vec{L}$ intersects the unit sphere. An *un*directed line meets S^2 in a pair of antipodal points; $\mathbb{R}P^2$ *is topologically* S^2 *with antipodal points identified.*

We can now construct a topological model of $\mathbb{R}P^2$ that will allow us to identify certain spaces we shall meet as projective spaces. Our model will respect the topology; that is, "nearby points" in $\mathbb{R}P^2$ (that is, nearby lines in $\mathbb{R}^3$) will be represented by nearby points in the model, but we won't be concerned with the differentiability of our procedure. Also it will be clear that certain natural "distances" will not be preserved; *in the rigorous definition of manifold,* to be given shortly, *there is no mention of metric notions such as distance or area or angle.*

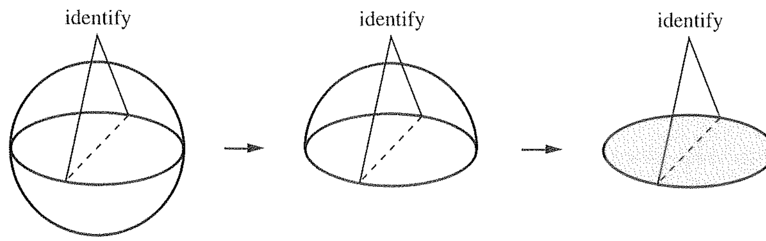

Figure 1.12

In the sphere with antipodal points identified, we may discard the entire southern hemisphere (exclusive of the equator) of redundant points, leaving us with the northern hemisphere, the equator, and with antipodal points only on the equator identified. We may then project this onto the disc in the plane. *Topologically $\mathbb{R}P^2$ is the unit disc in the plane with antipodal points on the unit circle identified.*

Similarly, $\mathbb{R}P^n$ is topologically the unit n sphere S^n in $\mathbb{R}^{n+1}$ with antipodal points identified, and this in turn is the solid n-dimensional unit ball in $\mathbb{R}^n$ with antipodal points on the boundary unit $(n-1)$ sphere identified.

(vii) It is a fact that *every submanifold of $\mathbb{R}^n$ is a manifold.* We verified this in the case of $S^2 \subset \mathbb{R}^3$ in Example (ii). In 1.1d we showed that the rotation group SO(3) is a 3-dimensional submanifold of $\mathbb{R}^9$. A convenient topological model is constructed as follows. Use the "right-hand rule" to associate the endpoint of the vector $\theta\mathbf{r}$ to the rotation through an angle θ (in radians) about an axis descibed by the unit vector $\mathbf{r}$. Note, however, that the rotation $\pi\mathbf{r}$ is exactly the same as the rotation $-\pi\mathbf{r}$ and $(\pi+\alpha)\mathbf{r}$ is the same as $-(\pi-\alpha)\mathbf{r}$. The collection of all rotations then can be represented by the points in the solid ball of radius π in $\mathbb{R}^3$ with antipodal points on the sphere of radius π identified; *$SO(3)$ can be identified with the real projective space $\mathbb{R}P^3$.*

(viii) The *Möbius band Mö* is the space obtained by identifying the left and right hand edges of a sheet of paper after giving it a "half twist"

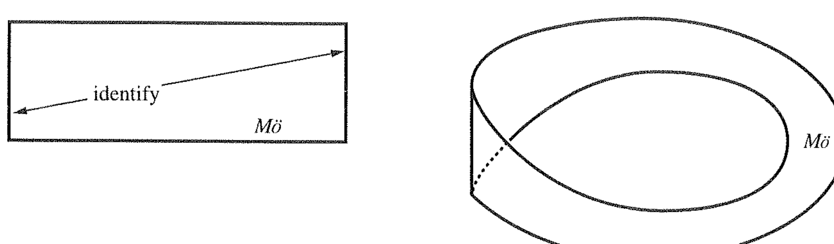

Figure 1.13

If one omits the edge one can see that *Mö* is a 2-dimensional submanifold of $\mathbb{R}^3$ and is therefore a 2-manifold. You should verify (i) that the Möbius band sits naturally as the shaded "half band" in the model of $\mathbb{R}P^2$ consisting of S^2 with antipodal points identified, and (ii) that this half band is the same as the full band. The *edge* of the

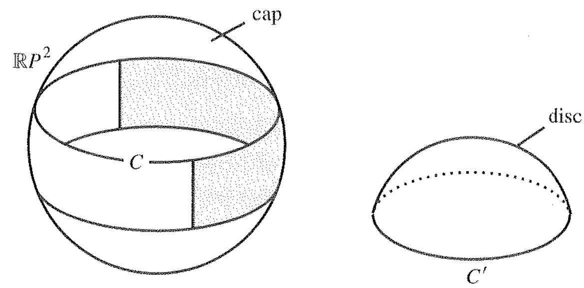

Figure 1.14

Möbius band consists of a *single* closed curve C that can be pictured as the "upper" edge of this full band in $\mathbb{R}P^2$. Note that the indicated "cap" is topologically a 2-dimensional disc with a circular edge C'. If we observe that the lower cap is the same as the upper, we conclude that *if we take a 2-disc and sew its edge to the single edge of a Möbius band, then the resulting space is topologically the projective plane*! We may say that $\mathbb{R}P^2$ is Mö with a 2-disc attached along its boundary. Although the actual sewing, say with cloth, cannot be done in ordinary space $\mathbb{R}^3$ (the cap would have to slice through itself), this sewing *can* be done in $\mathbb{R}^4$, where there is "more room."

1.2c. A Rigorous Definition of a Manifold

Let M be any *set* (without a topology) that has a covering by subsets $M = U \cup V \cup \ldots$, where each subset U is in $1:1$ correspondence $\phi_U : U \to \mathbb{R}^n$ with an *open* subset $\phi_U(U)$ of $\mathbb{R}^n$.

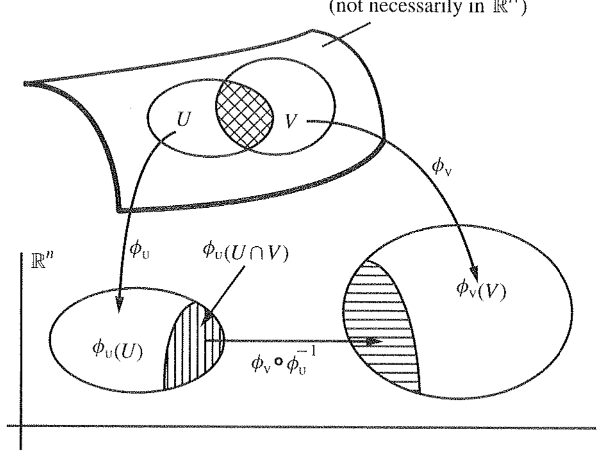

Figure 1.15

We require that each $\phi_U(U \cap V)$ be an open subset of $\mathbb{R}^n$. We require that the overlap maps

$$f_{VU} = \phi_V \circ \phi_U^{-1} : \phi_U(U \cap V) \to \mathbb{R}^n \tag{1.4}$$

that is,

$$\phi_U(U \cap V) \xrightarrow{\phi_U^{-1}} M \xrightarrow{\phi_V} \mathbb{R}^n$$

be differentiable (we know what it means for a map $\phi_V \circ \phi_U^{-1}$ from an open set of $\mathbb{R}^n$ to $\mathbb{R}^n$ to be differentiable). Each pair U, ϕ_U defines a **coordinate patch** on M; to $p \in U \subset M$ we may assign the n coordinates of the point $\phi_U(p)$ in $\mathbb{R}^n$. For this reason we shall call ϕ_U a **coordinate map**.

Take now a maximal atlas of such coordinate patches; see Example (iv). Define a **topology** in the set M by declaring a subset W of M to be open provided that given any $p \in W$ there is a coordinate chart U, ϕ_U such that $p \in U \subset W$. If the resulting topology for M is Hausdorff and has a countable base (see [S] for these technical conditions) we say that M is an n-dimensional differentiable manifold. We say that a *map* $F : \mathbb{R}^p \to \mathbb{R}^q$ is of class C^k if all k^{th} partial derivatives are continuous. It is of class C^∞ if it is of class C^k for all k. We say that a *manifold* M^n is of class C^k if its overlap maps f_{VU} are of class C^k. Likewise we have the notion of a C^∞ manifold. An analytic manifold is one whose overlap functions are analytic, that is, expandable in power series.

Let $F : M^n \to \mathbb{R}$ be a real-valued function on the manifold M. Since M is a topological space we know from 1.2a what it means to say that F is continuous. We say that F is **differentiable** if, when we express F in terms of a local coordinate system (U, x), $F = F_U(x^1, \dots, x^n)$ is a differentiable function of the coordinates x. Technically this means that that when we compose F with the inverse of the coordinate map ϕ_U

$$F_U := F \circ \phi_U^{-1}$$

(recall that ϕ_U is assumed $1 : 1$) we obtain a real-valued function F_U defined on a portion $\phi_U(U)$ of $\mathbb{R}^n$, and we are asking that this function be differentiable. Briefly speaking, *we envision the coordinates x as being engraved on the manifold M*, just as we see lines of latitude and longitude engraved on our globes. A function on the Earth's surface is continuous or differentiable if it is continuous or differentiable when expressed in terms of latitude and longitude, at least if we are away from the poles. Similarly with a manifold.With this understood, *we shall usually omit the process of replacing F by its composition $F \circ \phi_U^{-1}$, thinking of F as directly expressible as a function $F(x)$ of any local coordinates.*

Consider the real projective plane $\mathbb{R}P^2$, Example (vi) of Section 1.2b. In terms of homogeneous coordinates we may define a map $(\mathbb{R}^3 - 0) \to \mathbb{R}P^2$ by

$$(x, y, z) \to [x, y, z]$$

At a point of $\mathbb{R}^3$ where, for example, $z \neq 0$ we may use $u = x/z$ and $v = y/z$ as local coordinates in $\mathbb{R}P^2$, and then our map is given by the two smooth functions $u = f(x, y, z) = x/z$ and $v = g(x, y, z) = y/z$.

1.2d. Complex Manifolds: The Riemann Sphere

A **complex manifold** is a set M together with a covering $M = U \cup V \cup \ldots$, where each subset U is in 1 : 1 correspondence $\phi_U : U \to \mathbb{C}^n$ with an open subset $\phi_U(U)$ of complex n-space $\mathbb{C}^n$. We then require that the overlap maps f_{VU} mapping sets in $\mathbb{C}^n$ into sets in $\mathbb{C}^n$ be *complex analytic*; thus if we write f_{VU} in the form $w^k = w^k(z^1, \ldots, z^n)$ where $z^k = x^k + iy^k$ and $w^k = u^k + iv^k$, then u^k and v^k satisfy the Cauchy–Riemann equations with respect to each pair (x^r, y^r). Briefly speaking, each w^k can be expressed entirely in terms of $z^1, \ldots, z^n$, with no complex conjugates $\bar{z}^r$ appearing. We then proceed as in the real case in 2.3c. The resulting manifold is called an n-dimensional complex manifold, although its topological dimension is $2n$.

Of course the simplest example is $\mathbb{C}^n$ itself. Let us consider the most famous non-trivial example, the **Riemann sphere** M^1.

The complex plane $\mathbb{C}$ (topologically $\mathbb{R}^2$) comes equipped with a global complex coordinate $z = x + iy$. It is a complex 1-dimensional manifold $\mathbb{C}^1$. To study the behavior of functions at "∞" we introduce a point at ∞, to form a new manifold that is topologically the 2-sphere S^2. We do this by means of stereographic projection, as follows.

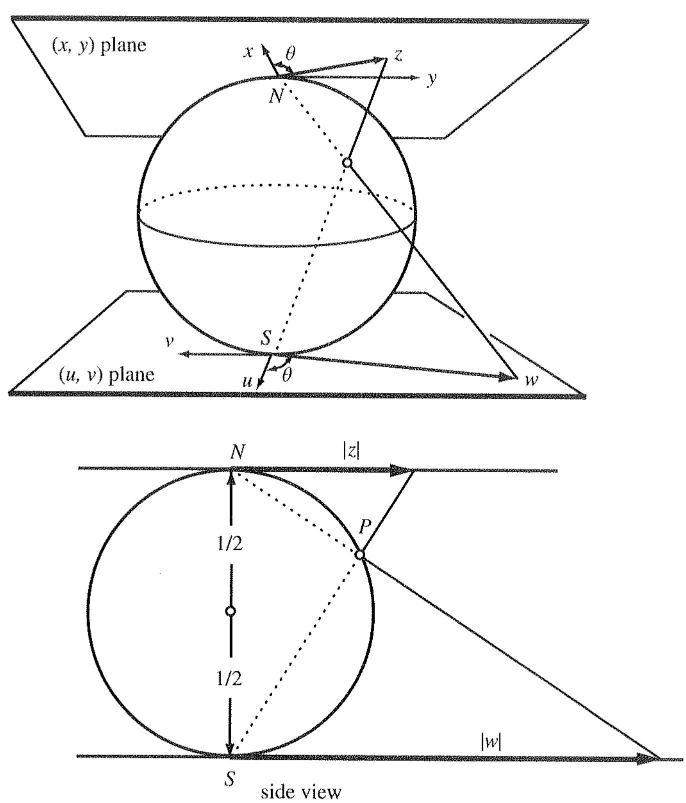

Figure 1.16

In the top part of the figure we have a sphere of radius $1/2$, resting on a $w = u + iv$ plane, with a tangent $z = x + iy$ plane at the north pole. Note that we have oriented these two tangent planes to agree with the usual orientation of S^2 (questions of orientation will be discussed in Section 2.8).

Let U be the subset of S^2 consisting of all points except for the south pole, let V be the points other than the north pole, let ϕ_U and ϕ_V be stereographic projections of U and V from the south and north poles, respectively, onto the z and the w planes. In this way we assign to any point p other than the poles two complex coordinates, $z = |z|e^{i\theta}$ and $w = |w|e^{-i\theta}$. From the bottom of the figure, which depicts the planar section in the plane holding the two poles and the point p, one reads off from elementary geometry that $|w| = 1/|z|$, and consequently

$$w = f_{VU}(z) = \frac{1}{z} \tag{1.5}$$

gives the relation between the two sets of coordinates. Since this is complex analytic in the overlap $U \cap V$, we may consider S^2 as a 1-dimensional complex manifold, the Riemann sphere. The point $w = 0$ (the south pole) represents the point $z = \infty$ that was missing from the original complex plane $\mathbb{C}$.

Note that the two sets of real coordinates (x, y) and (u, v) make S^2 into a real analytic manifold.

Problems

1.2(1) Show that $\mathbb{R}P^2$ is a differentiable 2-manifold by looking at the transition functions.

1.2(2) Give a coordinate covering for $\mathbb{R}P^3$, pick a pair of patches, and show that the overlap map is differentiable.

1.2(3) **Complex projective *n*-space** $\mathbb{C}P^n$ is defined to be the space of *complex* lines through the origin of $\mathbb{C}^{n+1}$. To a point $(z_0, z_1, \ldots, z_n)$ in $(\mathbb{C}^{n+1} - 0)$ we associate the line consisting of all *complex* multiples $\lambda\ (z_0, z_1, \ldots, z_n)$ of this point, $\lambda \in \mathbb{C}$. We call $[z_0, z_1, \ldots, z_n]$ the homogeneous coordinates of this line, that is, of this point in $\mathbb{C}P^n$; thus $[z_0, z_1, \ldots, z_n] = [\mu z_0, \mu z_1, \ldots, \mu z_n]$ for all $\mu \in (\mathbb{C} - 0)$. If $z_p \neq 0$ on this line, we may associate to this point $[z_0, z_1, \ldots, z_n]$ its n complex U_p coordinates $z_0/z_p, z_1/z_p, \ldots, z_n/z_p$, with z_p/z_p omitted.

Show that $\mathbb{C}P^2$ is a complex manifold of complex dimension 2.

Note that $\mathbb{C}P^1$ has complex dimension 1, that is, real dimension 2. For $z_1 \neq 0$ the U_1 coordinate of the point $[z_0, z_1]$ is $z = z_0/z_1$, whereas if $z_0 \neq 0$ the U_0 coordinate is $w = z_1/z_0$. These two patches cover $\mathbb{C}P^1$ and in the intersection of these two patches we have $w = 1/z$. *Thus* $\mathbb{C}P^1$ *is nothing other than the Riemann sphere*!

1.3. Tangent Vectors and Mappings

What do we mean by a "critical point" of a map $F : M^n \to V^r$?

We are all acquainted with vectors in $\mathbb{R}^N$. A tangent vector to a *submanifold* M^n of $\mathbb{R}^N$, at a given point $p \in M^n$, *is simply the usual velocity vector* $\dot{x}$ *to some parameterized* *curve* $x = x(t)$ *of* $\mathbb{R}^N$ *that lies on* M^n. On the other hand, a *manifold* M^n, as defined in the previous section, is a rather abstract object that need not be given as a subset of $\mathbb{R}^N$. For example, the projective plane $\mathbb{R}P^2$ was defined to be the space of lines through the origin of $\mathbb{R}^3$, that is, a point in $\mathbb{R}P^2$ is an entire *line* in $\mathbb{R}^3$; if $\mathbb{R}P^2$ were a submanifold of $\mathbb{R}^3$ we would associate a *point* of $\mathbb{R}^3$ to each point of $\mathbb{R}P^2$. We will be forced to define what we mean by a tangent vector to an abstract manifold. This definition will coincide with the previous notion in the case that M^n is a submanifold of $\mathbb{R}^N$. The fact that we understand tangent vectors to submanifolds is a powerful psychological tool, for it can be shown (though it is not elementary) that *every manifold can be realized as a submanifold of some* $\mathbb{R}^N$. In fact, Hassler Whitney, one of the most important contributors to manifold theory in the twentieth century, has shown that every M^n can be realized as a submanifold of $\mathbb{R}^{2n}$. Thus although we cannot "embed" $\mathbb{R}P^2$ in $\mathbb{R}^3$ (recall that we had a difficulty with sewing in 1.2b, Example (vii)), it can be embedded in $\mathbb{R}^4$. It is surprising, however, that for many purposes it is of little help to use the fact that M^n can be embedded in $\mathbb{R}^N$, and we shall try to give definitions that are "intrinsic," that is, independent of the use of an embedding. Nevertheless, we shall not hesitate to use an embedding for purposes of visualization, and in fact most of our examples will be concerned with submanifolds rather than manifolds.

A good reference for manifolds is [G, P]. The reader should be aware, however, that these authors deal only with manifolds that are *given* as subsets of some euclidean space.

1.3a. Tangent or "Contravariant" Vectors

We motivate the definition of vector as follows. Let $p = p(t)$ be a curve lying on the manifold M^n; thus p is a map of some interval on $\mathbb{R}$ into M^n. In a coordinate system (U, x_U) about the point $p_0 = p(0)$ the curve will be described by n functions $x_U^i = x_U^i(t)$, which will be assumed differentiable. The "velocity vector" $\dot{p}(0)$ was classically described by the n-tuple of real numbers $dx_U^1/dt]_0, \ldots, dx_U^n/dt]_0$. If p_0 also lies in the coordinate patch (V, x_V), then this same velocity vector is described by another n-tuple $dx_V^1/dt]_0, \ldots, dx_V^N/dt]_0$, related to the first set by the chain rule applied to the overlap functions (1.3), $x_V = x_V(x_U)$,

$$\left.\frac{dx_V^i}{dt}\right]_0 = \sum_{j=1}^{n} \left(\frac{\partial x_V^i}{\partial x_U^j}\right)(p_0)\left(\frac{dx_U^j}{dt}\right)_0$$

This suggests the following.

Definition: A **tangent vector,** or **contravariant vector,** or simply a **vector** at $p_0 \in M^n$, call it $\mathbf{X}$, assigns to each coordinate patch (U, x) holding p_0, an n-tuple of real numbers

$$(X_U^i) = (X_U^1, \ldots, X_U^n)$$

such that if $p_0 \in U \cap V$, then

$$X_V^i = \sum_j \left[\frac{\partial x_V^i}{\partial x_U^j}(p_0)\right] X_U^j \tag{1.6}$$

If we let $X_U = (X_U^1, \ldots, X_U^n)^T$ be the column of vector "components" of $\mathbf{X}$, we can write this as a matrix equation

$$X_V = c_{VU} X_U \tag{1.7}$$

where the **transition function** c_{VU} is the $n \times n$ Jacobian matrix evaluated at the point in question.

The term contravariant is traditional and is used throughout physics, and we shall use it even though it conflicts with the modern mathematical terminology of "categories and functors."

1.3b. Vectors as Differential Operators

In euclidean space an important role is played by the notion of differentiating a function f with respect to a vector at the point p

$$D_{\mathbf{v}}(f) = \frac{d}{dt}[f(p + t\mathbf{v})]_{t=0} \tag{1.8}$$

and if (x) is any cartesian coordinate system we have

$$D_{\mathbf{v}}(f) = \sum_j \left[\frac{\partial f}{\partial x^j}\right](p) v^j$$

This is the motivation for a similar operation on functions on any manifold M. A real-valued function f defined on M^n near p can be described in a local coordinate system x in the form $f = f(x^1, \ldots, x^n)$. (Recall, from Section 1.2c, that we are really dealing with the function $f \circ \phi_U^{-1}$ where ϕ_U is a coordinate map.) If $\mathbf{X}$ is a vector at p we define the derivative of f with respect to the vector $\mathbf{X}$ by

$$\mathbf{X}_p(f) := D_{\mathbf{X}}(f) := \sum_j \left[\frac{\partial f}{\partial x^j}\right](p) X^j \tag{1.9}$$

This seems to depend on the coordinates used, although it should be apparent from (1.8) that this is not the case in $\mathbb{R}^n$. We must show that (1.9) defines an operation that is independent of the local coordinates used. Let (U, x_U) and (V, x_V) be two coordinate systems. From the chain rule we see

$$\begin{aligned} D_{\mathbf{X}}^V(f) &= \sum_j \left(\frac{\partial f}{\partial x_V^j}\right) X_V^j = \sum_j \left(\frac{\partial f}{\partial x_V^j}\right) \sum_i \left(\frac{\partial x_V^j}{\partial x_U^i}\right) X_U^i \\ &= \sum_i \left(\frac{\partial f}{\partial x_U^i}\right) X_U^i = D_{\mathbf{X}}^U(f) \end{aligned}$$

This illustrates a basic point. *Whenever we define something by use of local coordinates, if we wish the definition to have intrinsic significance we must check that it has the same meaning in all coordinate systems.*

Note then that there is a 1 : 1 correspondence between tangent vectors $\mathbf{X}$ to M^n at p and first-order differential operators (on differentiable functions defined near p) that take the special form

$$\mathbf{X}_p = \sum_j X^j \frac{\partial}{\partial x^j}\bigg]_p \tag{1.10}$$

in a local coordinate system (x). From now on, *we shall make no distinction between a vector and its associated differential operator*. Each one of the n operators $\partial/\partial x^i$ then defines a vector, written $\boldsymbol{\partial}/\boldsymbol{\partial} x^i$, at each p in the coordinate patch.

The i^{th} component of $\boldsymbol{\partial}/\boldsymbol{\partial} x^\alpha$ is, from (1.9), given by δ^i_α (where the Kronecker δ^i_α is 1 if $i = \alpha$ and 0 if $i \neq \alpha$). On the other hand, consider the α^{th} **coordinate curve** through a point, the curve being parameterized by x^α. This curve is described by $x^i(t) =$ constant for $i \neq \alpha$ and $x^\alpha(t) = t$. The velocity vector for this curve at parameter value t has components $dx^i/dt = \delta^i_\alpha$. *The j^{th} coordinate vector $\boldsymbol{\partial}/\boldsymbol{\partial} x^j$ is the velocity vector to the j^{th} coordinate curve parameterized by x^j!* If $M^n \subset \mathbb{R}^N$, and if $\mathbf{r} = (y^1, \ldots, y^N)^T$ is the usual position vector from the origin, then $\boldsymbol{\partial}/\boldsymbol{\partial} x^j$ would be written *classically* as $\partial \mathbf{r}/\partial x^j$,

$$\frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x^j} = \frac{\partial \mathbf{r}}{\partial x^j} = \left(\frac{\partial y^1}{\partial x^j}, \ldots, \frac{\partial y^N}{\partial x^j}\right)^T \tag{1.11}$$

A familiar example will be given in the next section.

1.3c. The Tangent Space to M^n at a Point

It is evident from (1.6) that the sum of two vectors at a point, defined in terms of their n-tuples, is again a vector at that point, and that the product of a vector by a scalar, that is, a real number, is again a vector.

Definition: The **tangent space** to M^n at the point $p \in M^n$, written M^n_p, is the real vector space consisting of all tangent vectors to M^n at p. If (x) is a coordinate system holding p, then the n vectors

$$\frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x^1}\bigg]_p, \ldots, \frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x^n}\bigg]_p$$

form a basis of this n-dimensional vector space (as is evident from (1.10)) and this basis is called a **coordinate basis** or **coordinate frame**.

If M^n is a submanifold of $\mathbb{R}^N$, then M^n_p is the usual n-dimensional affine subspace of $\mathbb{R}^N$ that is "tangent" to M^n at p, and this is the picture to keep in mind.

A vector **field** on an open set U will be the differentiable assignment of a vector $\mathbf{X}$ to each point of U; in terms of local coordinates

$$\mathbf{X} = \sum_j X^j(x) \frac{\boldsymbol{\partial}}{\boldsymbol{\partial} x^j}$$

where the components X^j are differentiable functions of (x). In particular, each $\boldsymbol{\partial}/\boldsymbol{\partial} x^j$ is a vector field in the coordinate patch.

Example:

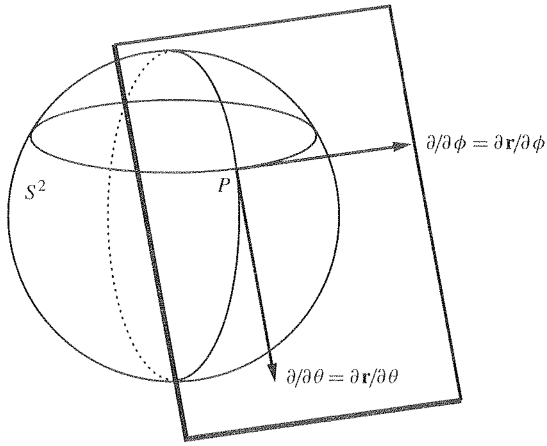

Figure 1.17

We have drawn the unit 2-sphere $M^2 = S^2$ in $\mathbb{R}^3$ with the usual spherical coordinates θ and ϕ (θ is colatitude and $-\phi$ is longitude). The equations defining S^2 are $x = \sin\theta\cos\phi$, $y = \sin\theta\sin\phi$, and $z = \cos\theta$. The coordinate vector $\boldsymbol{\partial}/\boldsymbol{\partial}\theta = \partial\mathbf{r}/\partial\theta$ is the velocity vector to a line of longitude, that is, keep ϕ constant and parameterize the meridian by "time" $t = \theta$. $\boldsymbol{\partial}/\boldsymbol{\partial}\phi$ has a similar description. Note that *these two vectors at p do not live in S^2, but rather in the linear space S_p^2 attached to S^2 at p*. Vectors at $q \neq p$ live in a different vector space S_q^2.

Warning: Because S^2 is a submanifold of $\mathbb{R}^3$ and because $\mathbb{R}^3$ carries a familiar metric, it makes sense to talk about the length of tangent vectors to this particular S^2; for example, we would say that $\| \boldsymbol{\partial}/\boldsymbol{\partial}\theta \| = 1$ and $\| \boldsymbol{\partial}/\boldsymbol{\partial}\phi \| = \sin\theta$. However, the definition of a manifold given in 1.2c does not require that M^n be given as some specific subset of some $\mathbb{R}^N$; *we do not have the notion of length of a tangent vector to a general manifold*. For example, the configuration space of a thermodynamical system might have coordinates given by pressure p, volume v, and temperature T, and the notions of the *lengths* of $\boldsymbol{\partial}/\boldsymbol{\partial}p$, and so on, seem to have no physical significance. If we wish to talk about the "length" of a vector on a manifold we shall be forced to introduce an *additional structure* on the manifold in question. The most common structure so used is called a *Riemannian structure*, or *metric*, which will be introduced in Chapter 2. See Problem 1.3 (1) at this time.

1.3d. Mappings and Submanifolds of Manifolds

Let $F : M^n \to V^r$ be a map from one manifold to another. In terms of local coordinates x near $p \in M^n$ and y near $F(p)$ on V^r F is described by r functions of n variables $y^\alpha = F^\alpha(x^1, \ldots, x^n)$, which can be abbreviated to $y = F(x)$ or $y = y(x)$. If, as we shall assume, the functions F^α are differentiable functions of the x's, we say that F is differentiable. As usual, such functions are, in particular, continuous.

When $n = r$, we say that F is a **diffeomorphism** provided F is 1 : 1, onto, and if, in addition, F^{-1} is also differentiable. Thus such an F is a differentiable homeomorphism (see 1.2a) with a differentiable inverse. (If F^{-1} does exist and the Jaçobian determinant does not vanish, $\partial(y^1, \dots, y^n)/\partial(x^1, \dots, x^n) \neq 0$, then the inverse function theorem of advanced calculus (see 1.3e) would assure us that the inverse *is* differentiable.)

The map $F : \mathbb{R} \to \mathbb{R}$ given by $y = x^3$ is a differentiable homeomorphism, but it is not a diffeomorphism since the inverse $x = y^{1/3}$ is not differentiable at $x = 0$.

We have already discussed submanifolds of $\mathbb{R}^n$ but now we shall need to discuss submanifolds of a manifold. A good example is the equator S^1 of S^2.

Definition: $W^r \subset M^n$ is an **(embedded) submanifold** of the manifold M^n provided W is *locally* described as the common locus

$$F^1(x^1, \dots, x^n) = 0, \dots, F^{n-r}(x^1, \dots, x^n) = 0$$

of $(n - r)$ differentiable functions that are independent in the sense that the Jacobian matrix $[\partial F^\alpha/\partial x^i]$ has rank $(n - r)$ at each point of the locus.

The implicit function theorem assures us that W^r can be locally described (after perhaps permuting some of the x coordinates) as a locus

$$x^{r+1} = f^{r+1}(x^1, \dots x^r), \ \dots, \ x^n = f^n(x^1, \dots, x^r)$$

It is not difficult to see from this (as we saw in the case $S^2 \subset \mathbb{R}^3$) that *every embedded submanifold of M^n is itself a manifold*!

Later on we shall have occasion to discuss submanifolds that are not "embedded," but for the present we shall assume "embedded" without explicit mention.

Definition: The **differential** F_* of the map $F : M^n \to V^r$ has the same meaning as in the case $\mathbb{R}^n \to \mathbb{R}^r$ discussed in 1.1b. $F_* : M^n_p \to V^r_{F(p)}$ is the linear transformation defined as follows. For $\mathbf{X} \in M^n_p$, let $p = p(t)$ be a curve on M with $p(0) = p$ and with velocity vector $\dot{p}(0) = \mathbf{X}$. Then $F_*\mathbf{X}$ is the velocity vector $d/dt\{F(p(t))\}_{t=0}$ of the image curve at $F(p)$ on V. This vector is independent of the curve $p = p(t)$ chosen (as long as $\dot{p}(o) = \mathbf{X}$). The matrix of this linear transformation, in terms of the bases $\boldsymbol{\partial/\partial} x$ at p and $\boldsymbol{\partial/\partial} y$ at $F(p)$, is the Jacobian matrix

$$(F_*)^\alpha{}_i = \frac{\partial F^\alpha}{\partial x^i}(p) = \frac{\partial y^\alpha}{\partial x^i}(p)$$

The main theorem on submanifolds is exactly as in euclidean space (Section 1.1c).

Theorem (1.12): *Let $F : M^n \to V^r$ and suppose that for some $q \in V^r$ the locus $F^{-1}(q) \subset M^n$ is not empty. Suppose further that F_* is onto, that is, F_* is of rank r, at each point of $F^{-1}(q)$. Then $F^{-1}(q)$ is an $(n-r)$-dimensional submanifold of M^n.*

Example: Consider a 2-dimensional torus T^2 (the surface of a doughnut), embedded in $\mathbb{R}^3$.

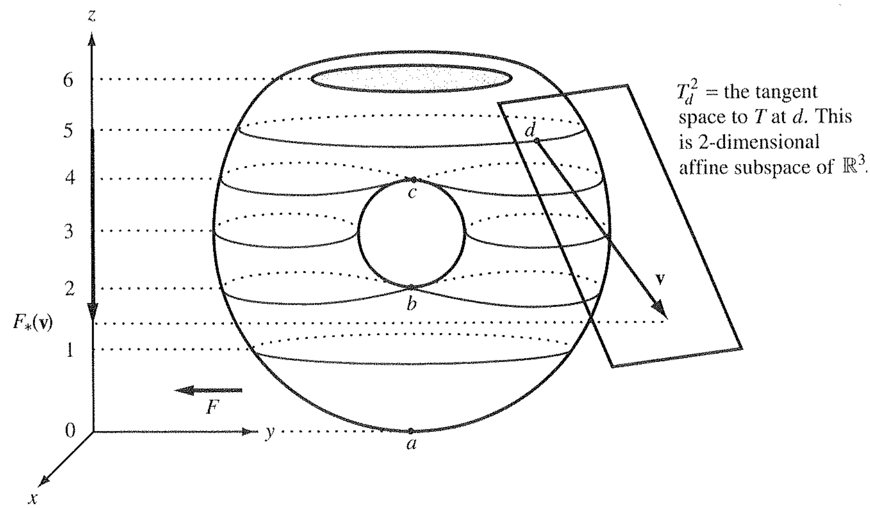

Figure 1.18

We have drawn it smooth with a flat top (which is supposed to join *smoothly* with the rest of the torus). Define a differentiable map (function) $F : T^2 \to \mathbb{R}$ by $F(p) = z$, the height of the point $p \in T^2$ above the z plane ($\mathbb{R}$ is being identified with the z axis). Consider a point $d \in T$ and a tangent vector $\mathbf{v}$ to T at d. Let $p = p(t)$ be a curve on T such that $p(0) = d$ and $\dot{p}(0) = \mathbf{v}$. The image curve in $\mathbb{R}$ is described in the coordinate z for $\mathbb{R}$ by $z(t) = z(p(t))$, and it is clear from the geometry of $T^2 \subset \mathbb{R}^3$ that $\dot{z}(0)$ is simply the z component of the spatial vector $\mathbf{v}$. In other words $F_*(\mathbf{v})$ *is the projection of* $\mathbf{v}$ *onto the* z *axis*. Note then that F_* will be onto at each point $p \in T^2$ for which the tangent plane $T^2(p)$ is *not* horizontal, that is, at all points of T^2 except $a \in F^{-1}(0)$, $b \in F^{-1}(2)$, $c \in F^{-1}(4)$, and the entire flat top $F^{-1}(6)$.

From the main theorem, we may conclude that $F^{-1}(z)$ is a 1-dimensional submanifold of the torus for $0 \leq z \leq 6$ except for $z = 0, 2, 4$, and 6, and this is indeed "verified" in our picture. (We have drawn the inverse images of $z = 0, 1, \ldots, 6$.) Notice that $F^{-1}(2)$, which looks like a figure 8, is *not* a submanifold; a neighborhood of the point b on $F^{-1}(2)$ is topologically a cross $+$ and thus no neighborhood of b is topologically an open interval on $\mathbb{R}$.

Definition: If $F : M^n \to V^r$ is a differentiable map between manifolds, we say that

(i) $x \in M$ is a **regular point** if F_* maps M^n_x onto $V^r_{F(x)}$; otherwise we say that x is a **critical point**.

(ii) $y \in V^r$ is a **regular value** provided *either* $F^{-1}(y)$ is empty, or $F^{-1}(y)$ consists entirely of regular points. Otherwise y is a **critical value**.

Our main theorem on submanifolds can then be stated as follows.

Theorem (1.13): *If $y \in V^r$ is a regular value, then $F^{-1}(y)$ either is empty or is a submanifold of M^n of dimension $(n - r)$.*

Of course, if x is a critical *point* then $F(x)$ is a critical *value*. In our toroidal example, Figure 1.18, all values of z other than 0, 2, 4, and 6 are regular. The critical points on T^2 consist of a, b, c, and the entire flat top of T^2. These latter critical points thus fill up a positive area (in the sense of elementary calculus) on T^2. Note however, that the image of this 2-dimensional set of critical points consists of the single critical value $z = 6$. The following theorem assures us that the critical *values* of a map form a "small" subset of V^r; the critical values cannot fill up any open set in V^r and they will have "measure" 0. We will not be precise in defining "almost all"; roughly speaking we mean, in some sense, "with probability 1."

Sard's Theorem (1.14): *If $F : M^n \to V^r$ is sufficiently differentiable, then almost all values of F are regular values, and thus for almost all points $y \in V^r$, $F^{-1}(y)$ either is empty or is a submanifold of M^n of dimension $(n - r)$.*

By sufficiently differentiable, we mean the following. If $n \leq r$, we demand that F be of differentiability class C^1, whereas if $n - r = k > 0$, we demand that F be of class C^{k+1}. The proof of Sard's theorem is delicate, especially if $n > r$; see, for example, [A, M, R].

1.3e. Change of Coordinates

The inverse function theorem is perhaps the most important theoretical result in all of differential calculus.

The Inverse Function Theorem (1.15): *If $F : M^n \to V^n$ is a differentiable map between manifolds of the same dimension, and if at $x_0 \in M$ the differential F_* is an isomorphism, that is, it is 1 : 1 and onto, then F is a local diffeomorphism near x.*

This means that there is a neighborhood U of x such that $F(U)$ is open in V and $F : U \to F(U)$ is a diffeomorphism. This theorem is a powerful tool for introducing new coordinates in a neighborhood of a point, for it has the following consequence.

Corollary (1.16): *Let $x^1, \ldots, x^n$ be local coordinates in a neighborhood U of the point $p \in M^n$. Let $y^1, \ldots, y^n$ be any differentiable functions of the x's (thus yielding a map:$U \to \mathbb{R}^n$) such that*

$$\frac{\partial(y^1, \ldots, y^n)}{\partial(x^1, \ldots, x^n)}(p) \neq 0$$

Then the y's form a coordinate system in some (perhaps smaller) neighborhood of p.

For example, when we put $x = r\cos\theta$, $y = r\sin\theta$, we have $\partial(x, y)/\partial(r, \theta) = r$, and so $\partial(r, \theta)/\partial(x, y) = 1/r$. This shows that polar coordinates are good coordinates in a neighborhood of any point of the plane other than the origin.

It is important to realize that *this theorem is only local*. Consider the map $F : \mathbb{R}^2 \to \mathbb{R}^2$ given by $u = e^x \cos y$, $v = e^x \sin y$. This is of course the complex analytic map $w = e^z$. The real Jacobian $\partial(u, v)/\partial(x, y)$ never vanishes (this is reflected in the complex Jacobian $dw/dz = e^z$ never vanishing). Thus F is locally $1 : 1$. It is not globally so since $e^{z+2\pi ni} = e^z$ for all integers n. u, v form a coordinate system not in the whole plane but rather in any strip $a \le y < a + 2\pi$.

The inverse function theorem and the implicit function theorem are essentially equivalent, the proof of one following rather easily from that of the other. The proofs are fairly delicate; see for example, $[A, M, R]$.

Problems

1.3(1) What would be wrong in defining $\| \mathbf{X} \|$ in an M^n by

$$\| \mathbf{X} \|^2 = \sum_j (X_U^j)^2 \ ?$$

1.3(2) Lay a 2-dimensional torus flat on a table (the xy plane) rather than standing as in Figure 1.18. By inspection, what are the critical points of the map $T^2 \to \mathbb{R}^2$ projecting T^2 into the xy plane?

1.3(3) Let M^n be a submanifold of $\mathbb{R}^N$ that does not pass through the origin. Look at the critical points of the function $f : M \to \mathbb{R}$ that assigns to each point of M the square of its distance from the origin. Show, using local coordinates $u^1, \dots, u^n$, that a point is a critical point for this distance function iff the position vector to this point is normal to the submanifold.

1.4. Vector Fields and Flows

Can one solve $dx^i/dt = \partial f/\partial x^i$ to find the curves of steepest ascent?

1.4a. Vector Fields and Flows on $\mathbb{R}^n$

A vector field on $\mathbb{R}^n$ assigns in a differentiable manner a vector $\mathbf{v}_p$ to each p in $\mathbb{R}^n$. In terms of cartesian coordinates $x^1, \dots, x^n$

$$\mathbf{v} = \sum_j v^j(x) \frac{\partial}{\partial x^j}$$

where the components v^j are differentiable functions. Classically this would be written simply in terms of the cartesian components $\mathbf{v} = (v^1(x), \dots, v^n(x))^T$.

Given a "stationary" (i.e., time-independent) flow of water in $\mathbb{R}^3$, we can construct the 1-parameter family of maps

$$\phi_t : \mathbb{R}^3 \to \mathbb{R}^3$$

where ϕ_t takes the molecule located at p when $t = 0$ to the position of the same molecule t seconds later. Since the flow is time-independent

$$\phi_s(\phi_t(p)) = \phi_{s+t}(p) = \phi_t(\phi_s(p))$$

and (1.17)

$$\phi_{-t}(\phi_t(p)) = p, \quad \text{i.e., } \phi_{-t} = \phi_t^{-1}$$

We say that this defines a **1-parameter group** of maps. Furthermore, if each ϕ_t is differentiable, then so is each ϕ_t^{-1}, and so *each* ϕ_t *is a diffeomorphism.* We shall call such a family simply a **flow.** Associated with any such flow is a time-independent **velocity field**

$$\mathbf{v}_p := \left.\frac{d\phi_t(p)}{dt}\right]_{t=0}$$

In terms of coordinates we have

$$v^j(p) = \left.\frac{dx^j(\phi_t(p))}{dt}\right]_{t=0}$$

which will usually be written

$$v^j(x) = \frac{dx^j}{dt}$$

Thought of as a differential operator on functions f

$$\mathbf{v}_p(f) = \sum_j v^j(p)\frac{\partial f}{\partial x^j} = \sum_j \frac{dx^j}{dt}\frac{\partial f}{\partial x^j}$$
$$= \left.\frac{d}{dt} f(\phi_t(p))\right]_{t=0}$$

is the derivative of f along the "streamline" through p.

We thus have the almost trivial observation that to each flow $\{\phi_t\}$ we can associate the velocity vector field. The converse result, perhaps the most important theorem relating calculus to science, states, roughly speaking, that to each vector field $\mathbf{v}$ in $\mathbb{R}^n$ one may associate a flow $\{\phi_t\}$ having $\mathbf{v}$ as its velocity field, and that $\phi_t(p)$ can be found by solving the system of ordinary differential equations

$$\frac{dx^j}{dt} = v^j(x^1(t), \ldots, x^n(t)) \tag{1.18}$$

with initial conditions

$$x(0) = p$$

Thus one finds the **integral curves** of the preceding system, and $\phi_t(p)$ says, "Move along the integral curve through p (the 'orbit' of p) for time t." We shall now give a precise statement of this "fundamental theorem" on the existence of solutions of ordinary differential equations. For details one can consult [A, M, R; chap. 4], where this result is proved in the context of Banach spaces rather than $\mathbb{R}^n$. I recommend highly chapters 4 and 5 of Arnold's book [A2].

The Fundamental Theorem on Vector Fields in $\mathbb{R}^n$ (1.19): *Let* $\mathbf{v}$ *be a* C^k *vector field,* $k \geq 1$ *(each component* $v^j(x)$ *is of differentiability class* C^k*) on an open subset* U *of* $\mathbb{R}^n$*. This can be written* $\mathbf{v} : U \to \mathbb{R}^n$ *since* $\mathbf{v}$ *associates to each* $x \in U$ *a point* $\mathbf{v}(x) \in \mathbb{R}^n$*. Then for each* $p \in U$ *there is a curve* γ *mapping an interval* $(-b, b)$ *of the real line into* U

$$\gamma : (-b, b) \to U$$

such that

$$\frac{d\gamma(t)}{dt} = v(\gamma(t)) \quad \text{and} \quad \gamma(0) = p$$

for all $t \in (-b, b)$*. (This says that* γ *is an integral curve of* $\mathbf{v}$ *starting at* p*.) Any two such curves are equal on the intersection of their* t*-domains ("uniqueness"). Moreover, there is a neighborhood* U_p *of* p*, a real number* $\epsilon > 0$*, and a* C^k *map*

$$\Phi : U_p \times (-\epsilon, \epsilon) \to \mathbb{R}^n$$

such that the curve $t \in (-\epsilon, \epsilon) \mapsto \phi_t(q) := \Phi(q, t)$ *satisfies the differential equation*

$$\frac{\partial}{\partial t}\phi_t(q) = \mathbf{v}(\phi_t(q))$$

for all $t \in (-\epsilon, \epsilon)$ *and* $q \in U_p$*. Moreover, if* t*,* s*, and* $t + s$ *are all in* $(-\epsilon, \epsilon)$*, then*

$$\phi_t \circ \phi_s = \phi_{t+s} = \phi_s \circ \phi_t$$

for all $q \in U_p$*, and thus* $\{\phi_t\}$ *defines a local 1-parameter "group" of diffeomorphisms, or* ***local flow****.*

The term *local* refers to the fact that ϕ_t is defined only on a subset $U_p \subset U \subset \mathbb{R}^n$. The word "group" has been put in quotes because this family of maps does not form a group in the usual sense. In general (see Problem 1.4 (1)), the maps ϕ_t are only defined for small t, $-\epsilon < t < \epsilon$; that is, **the integral curve through a point q need only exist for a small time**. Thus, for example, if $\epsilon = 1$, then although $\phi_{1/2}(q)$ exists neither $\phi_1(q)$ nor $\phi_{1/2} \circ \phi_{1/2}$ need exist; the point is that $\phi_{1/2}(q)$ need not be in the set U_p on which $\phi_{1/2}$ is defined.

Example: $\mathbb{R}^n = \mathbb{R}$, the real line, and $v(x) = x\,d/dx$. Thus v has a single component x at the point with coordinate x. Let $U = \mathbb{R}$. To find ϕ_t we simply solve the differential equation

$$\frac{dx}{dt} = x \qquad \text{with initial condition } x(0) = p$$

to get $x(t) = e^t p$, that is, $\phi_t(p) = e^t p$. In this example the map ϕ_t is clearly defined on all of $M^1 = \mathbb{R}$ and for all time t. It can be shown that this is true for any *linear* vector field

$$\frac{dx^j}{dt} = \sum_k a^j_k x^k$$

defined on *all* of $\mathbb{R}^n$.

Note that if we solved the differential equation $dx/dt = 1$ on the real line with the origin deleted, that is, on the *manifold* $M^1 = \mathbb{R} - 0$, then the solution curve starting at $x = -1$ at $t = 0$ would exist for all times less than 1 second, but ϕ_1 would not exist; the solution simply runs "off" the manifold because of the missing point. One might think that if we avoid dealing with pathologies such as digging out a point from $\mathbb{R}^1$, then our solutions would exist for all time, but as you shall verify in Problem 1.4(1) this is not the case. The growth of the vector field can cause a solution curve to "leave" $\mathbb{R}^1$ in a finite amount of time.

We have required that the vector field $\mathbf{v}$ be differentiable. Uniqueness can be lost if the field $\mathbf{v}$ is only continuous. For example, again on the real line, consider the differential equation $dx/dt = 3x^{2/3}$. The usual solutions are of the form $x(t) = (t-c)^3$, but there is also the "singular" solution $x(t) = 0$ identically. This is a reflection of the fact that $x^{2/3}$ is not differentiable when $x = 0$.

1.4b. Vector Fields on Manifolds

If $\mathbf{X}$ is a C^k vector field on an open subset W of a *manifold* M^n then we can again recover a 1-parameter local group ϕ_t of diffeomorphisms for the following reasons. If W is contained in a single coordinate patch (U, x_U) we can proceed just as in the case $\mathbb{R}^n$ earlier since we can use the local coordinates x_U. Suppose that W is not contained in a single patch. Let $p \in W$ be in a coordinate overlap, $p \in U \cap V$. In U we can solve the differential equations

$$\frac{dx_U^j}{dt} = X_U^j(x_U^1, \ldots, x_U^n)$$

as before. In V we solve the equations

$$\frac{dx_V^j}{dt} = X_V^j(x_V^1, \ldots, x_V^n)$$

Because of the transformation rule (1.6), the right-hand side of this last equation is $\sum_k [\partial x_V^j / \partial x_U^k] X_U^k$; the left-hand side is, by the chain rule, $\sum_k [\partial x_V^j / \partial x_U^k] dx_U^k/dt$. Thus, *because of the transformation rule for a contravariant vector, the two differential equations say exactly the same thing*. Using uniqueness, we may then patch together the U and the V solutions to give a local solution in W.

Warning: Let $f : M^n \to \mathbb{R}$ be a differentiable function on M^n. In elementary mathematics it is often said that the n-tuple

$$\left[\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n}\right]^T$$

form the components of a vector field "grad f." However, if we look at the transformation properties in $U \cap V$, by the chain rule

$$\frac{\partial f}{\partial x_V^j} = \sum_k \left[\frac{\partial x_U^k}{\partial x_V^j}\right] \frac{\partial f}{\partial x_U^k}$$

and this is *not* the rule for a contravariant vector. One sees then that a proposed differential equation for "steepest ascent," $dx/dt =$"grad f," that is,

$$\frac{dx_U^j}{dt} = \frac{\partial f}{\partial x_U^j} \quad \text{in } U \quad \text{and} \quad \frac{dx_V^j}{dt} = \frac{\partial f}{\partial x_V^j} \quad \text{in } V$$

would not say the same thing in two overlapping patches, and consequently *would not yield a flow* ϕ_t! In the next chapter we shall see how to deal with n-tuples that transform as "grad f."

1.4c. Straightening Flows

Our version of the fundamental theorem on the existence of solutions of differential equations, as given in the previous section, is not the complete story; see [A, M, R, theorem 4.1.14] or [A2, chap. 4] for details of the following. The map $(p, t) \to \phi_t(p)$ depends smoothly on the initial condition p and on the time of flow t. This has the following consequence. (Since our result will be local, it is no loss of generality to replace M^n by $\mathbb{R}^n$.) *Suppose that the vector field* $\mathbf{v}$ *does not vanish at the point* p. Then of course it doesn't vanish in some neighborhood of p in M^n. Let W^{n-1} be a hypersurface, that is, a submanifold of codimension 1, that passes through p. Assume that W is **transversal** to $\mathbf{v}$, that is, the vector field $\mathbf{v}$ is not tangent to W.

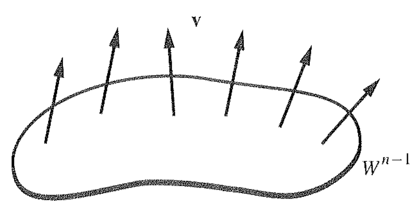

Figure 1.19

Let $u^1, \ldots, u^{n-1}$ be local coordinates for W, and let p_u be the point on W with local coordinates u. Then $\phi_t(p_u)$ is the point t seconds along the orbit of $\mathbf{v}$ through p_u. This point can be described by the n-tuple (u, t). The fundamental theorem states that if W is sufficiently small and if t is also sufficiently small, then (u, t) *can be used as* (curvilinear) *coordinates for some n-dimensional neighborhood of p in* M^n. To see this we shall apply the inverse function theorem. We thus consider the map $L : W^{n-1} \times (-\epsilon, \epsilon) \to M^n$ given by $L(u, t) = \phi_t(p_u)$. We compute the differential of this map *at the origin* $u = 0$ of the coordinates on W^{n-1}. Then by the geometric meaning of L_*, and since $\phi_0(p) = p$

$$L_*\left(\frac{\partial}{\partial u^1}\right) = \frac{\partial}{\partial u}[\phi_0(u, 0, \ldots, 0)]_0 = \left.\frac{\partial p_{(u,0,\ldots,0)}}{\partial u}\right|_{u=0} = \frac{\partial}{\partial u^1}$$

Likewise $L_*(\partial/\partial u^i) = \partial/\partial u^i$, for $i = 1, \ldots, n-1$. Finally

$$L_*(\mathbf{v}) = \frac{\partial}{\partial t}\phi_t(p_0) = \mathbf{v}$$

Thus L_* is the identity linear transformation, and by Corollary (1.16) we may use $u^1, \ldots, u^{n-1}, t$ as local coordinates for M^n near p_0.

It is then clear that *in these new local coordinates* near p, the flow defined by the vector field $\mathbf{v}$ is simply $\phi_S : (u, t) \to (u, s+t)$ and the vector field $\mathbf{v}$ in terms of $\partial/\partial u^1, \ldots, \partial/\partial u^{n-1}, \partial/\partial t$, is simply $\mathbf{v} = \partial/\partial t$. We have "straightened out" the flow!

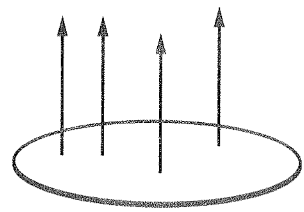

Figure 1.20

This says that near a **nonsingular** point of $\mathbf{v}$, that is, a point where $\mathbf{v} \neq 0$, coordinates $u^1, \ldots, u^n$ can be introduced such that the original system of differential equations $dx^1/dt = v^1(x), \ldots, dx^n/dt = v^n(x)$ becomes

$$\frac{du^1}{dt} = 0, \ldots, \frac{du^{n-1}}{dt} = 0, \qquad \frac{du^n}{dt} = 1 \tag{1.20}$$

Thus all flows near a nonsingular point are qualitatively the same! In a sense this result is of theoretical interest only, for in order to introduce the new coordinates u *one must solve the original system of differential equations*. The theoretical interest is, however, considerable. For example, $u^1 = c_1, \ldots, u^{n-1} = c_{n-1}$, are $(n-1)$ "first integrals," that is, constants of the motion, for the system (1.20). We conclude that near any nonsingular point of any system there are $(n-1)$ first integrals, $u^1(x) = c^1, \ldots, u^{n-1}(x) = c_{n-1}$ (but of course, we might have to solve the original system to write down explicitly the functions u^j in terms of the x's).

Problems

1.4(1) Consider the quadratic vector field problem on $\mathbb{R}$, $v(x) = x^2 d/dx$. You must solve the differential equation

$$\frac{dx}{dt} = x^2 \quad \text{and} \quad x(0) = p$$

Consider, as in the statement of the fundamental theorem, the case when U_p is the set $1/2 < x < 3/2$. Find the largest ϵ so that $\Phi : U_p \times (-\epsilon, \epsilon) \to \mathbb{R}$ is defined; that is, find the largest t for which the integral curve $\phi_t(q)$ will be defined for all $1/2 < q < 3/2$.

1.4(2) In the complex plane we can consider the differential equations $dz/dt = 1$, where t is real. The integral curves are of course lines parallel to the real axis.This can also be considered a differential equation on the z patch of the Riemann sphere of Section 1.2d. Extend this differential equation to the entire sphere by writing out the equivalent equation in the w patch. Write out the general solution $w = w(t)$ in the neighborhood of $w = 0$, and draw in particular the solutions starting at i, ± 1, and $-i$.

CHAPTER 2

Tensors and Exterior Forms

IN Section 1.4b we considered the n-tuple of partial derivatives of a single function $\partial F/\partial x^j$ and we noticed that this n-tuple does not transform in the same way as the n-tuple of components of a vector. These components $\partial F/\partial x^j$ transform as a new type of "vector." In this chapter we shall talk of the general notion of "tensor" that will include both notions of vector and a whole class of objects characterized by a transformation law generalizing 1.6. We shall, however, strive to define these objects and operations on them "intrinsically," that is, in a basis-free fashion. We shall also be very careful in our use of sub- and superscripts when we express components in terms of bases; *the notation is designed to help us recognize intrinsic quantities when they are presented in component form and to help prevent us from making blatant errors.*

2.1. Covectors and Riemannian Metrics

How *do* we find the curves of steepest ascent?

2.1a. Linear Functionals and the Dual Space

Let E be a real vector space. Although for some purposes E may be infinite-dimensional, we are mainly concerned with the finite-dimensional case. Although $\mathbb{R}^n$, as the space of real n-tuples $(x^1, \dots, x^n)$, comes equipped with a distinguished basis $(1, 0, 0, \dots, 0)^T, \dots$, the general n-dimensional vector space E has no basis prescribed.

Choose a basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ for the n-dimensional space E. Then a vector $\mathbf{v} \in E$ has a unique expansion

$$\mathbf{v} = \sum_j \mathbf{e}_j v^j = \sum_j v^j \mathbf{e}_j$$

where the n real numbers v^j are the **components** of $\mathbf{v}$ with respect to the given basis. For algebraic purposes, *we prefer the first presentation*, where we have put the "scalars" v^j to the right of the basis elements. We do this for several reasons, but mainly so that *we can use matrix notation*, as we shall see in the next paragraph. *When dealing* *with calculus, however, this notation is awkward.* For example, in $\mathbb{R}^n$ (thought of as a manifold), we can write the standard basis at the origin as $\mathbf{e}_j = \boldsymbol{\partial}/\boldsymbol{\partial} x^j$ (as in Section 1.3c); then our favored presentation would say $\mathbf{v} = \sum_j \boldsymbol{\partial}/\boldsymbol{\partial} x^j \ v^j$, making it appear, incorrectly, that we are differentiating the components v^j. *We shall employ the bold* $\boldsymbol{\partial}$ *to remind us that we are not differentiating the components in this expression.* Sometimes we will simply use the traditional $\sum_j v^j \mathbf{e}_j$.

We shall use the matrices

$$\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_n) \quad \text{and} \quad v = (v^1, \dots, v^n)^T$$

The first is a symbolic *row* matrix since each entry is a vector rather than a scalar. Note that in the matrix v we are preserving the traditional notation of representing the components of a vector by a *column* matrix. We can then write our preferred representation as a matrix product

$$\mathbf{v} = \mathbf{e}\, v \tag{2.1}$$

where $\mathbf{v}$ is a 1×1 matrix. As usual, we see that the n-dimensional vector space E, *with a choice of basis*, is isomorphic to $\mathbb{R}^n$ under the correspondence $\mathbf{v} \to (v^1, \dots, v^n) \in \mathbb{R}^n$, but that this isomorphism is "unnatural," that is, dependent on the choice of basis.

Definition: A (real) **linear functional** α on E is a real-valued linear function α, that is, a linear transformation $\alpha : E \to \mathbb{R}$ from E to the 1-dimensional vector space $\mathbb{R}$. Thus

$$\alpha(a\mathbf{v} + b\mathbf{w}) = a\alpha(\mathbf{v}) + b\alpha(\mathbf{w})$$

for real numbers a, b, and vectors $\mathbf{v}$, $\mathbf{w}$.

By induction, we have, for any basis $\mathbf{e}$

$$\alpha\left(\sum \mathbf{e}_j v^j\right) = \sum \alpha(\mathbf{e}_j) v^j \tag{2.2}$$

This is simply of the form $\sum a_j v^j$ (where $a_j := \alpha(\mathbf{e}_j)$), and this is a linear function of the components of $\mathbf{v}$. Clearly if $\{a_j\}$ are any real numbers, then $\mathbf{v} \mapsto \sum a_j v^j$ defines a linear functional on all of E. Thus, *after* one has picked a basis, *the most general linear functional on the finite-dimensional vector space E is of the form*

$$\alpha(\mathbf{v}) = \sum a_j v^j \quad \text{where } a_j := \alpha(\mathbf{e}_j) \tag{2.3}$$

Warning: A linear functional α on E is not itself a member of E; that is, α is not to be thought of as a *vector* in E. This is especially obvious in infinite-dimensional cases. For example, let E be the vector space of all continuous real-valued functions $f : \mathbb{R} \to \mathbb{R}$ of a real variable t. The **Dirac functional** δ_0 is the linear functional on E defined by

$$\delta_0(f) = f(0)$$

You should convince yourself that E *is* a vector space and that δ_0 *is* a linear functional on E. *No one would confuse* δ_0*, the Dirac* δ *"function," with a continuous function,* *that is, with an element of E.* In fact δ_0 is not a function on $\mathbb{R}$ at all. Where, then, do the linear functionals live?

Definition: The collection of all linear functionals α on a vector space E form a new vector space E^*, the **dual space** to E, under the operations

$$(\alpha + \beta)(\mathbf{v}) := \alpha(\mathbf{v}) + \beta(\mathbf{v}), \qquad \alpha, \beta \in E^*, \qquad \mathbf{v} \in E$$
$$(c\alpha)(\mathbf{v}) := c\alpha(\mathbf{v}), \qquad c \in \mathbb{R}$$

We shall see in a moment that if E is n-dimensional, then so is E^*.

If $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is a basis of E, we define the **dual basis** $\sigma^1, \ldots, \sigma^n$ of E^* by first putting

$$\sigma^i(\mathbf{e}_j) = \delta^i{}_j$$

and then "extending σ by linearity," that is,

$$\sigma^i\left(\sum_j \mathbf{e}_j v^j\right) = \sum_j \sigma^i(\mathbf{e}_j)v^j = \sum_j \delta^i{}_j v^j = v^i$$

Thus σ^i *is the linear functional that reads off the* i^{th} *component* (with respect to the basis $\mathbf{e}$) *of each vector* $\mathbf{v}$.

Let us verify that the σ's do form a basis. To show linear independence, assume that a linear combination $\sum a_j\sigma^j$ is the 0 functional. Then $0 = \sum_j a_j\sigma^j(\mathbf{e}_k) = \sum_j a_j\delta^j{}_k = a_k$ shows that all the coefficients a_k vanish, as desired. To show that the σ's span E^*, we note that if $\alpha \in E^*$ then

$$\alpha(\mathbf{v}) = \alpha\left(\sum \mathbf{e}_j v^j\right) = \sum \alpha(\mathbf{e}_j)v^j$$
$$= \sum \alpha(\mathbf{e}_j)\sigma^j(\mathbf{v}) = \left(\sum \alpha(\mathbf{e}_j)\sigma^j\right)(\mathbf{v})$$

Thus the two linear functionals α and $\sum \alpha(\mathbf{e}_j)\sigma^j$ must be the same!

$$\alpha = \sum_j \alpha(\mathbf{e}_j)\sigma^j \tag{2.4}$$

This very important equation shows that the σ's do form a basis of E^*.

In (2.3) we introduced the n-tuple $a_j = \alpha(\mathbf{e}_j)$ for each $\alpha \in E^*$. From (2.4) we see $\alpha = \sum a_j\sigma^j$. a_j defines the j^{th} **component** of α.

If we introduce the matrices

$$\sigma = (\sigma^1, \ldots, \sigma^n)^T \quad \text{and} \quad a = (a_1, \ldots, a_n)$$

then we can write

$$\alpha = \sum_j a_j\sigma^j = a\sigma \tag{2.5}$$

Note that the components of a linear functional are written as a *row* matrix a.

Under a change of local coordinates the chain rule yields

$$dx_V{}^i = \sum_j \left(\frac{\partial x_V{}^i}{\partial x_U{}^j}\right) dx_U{}^j \tag{2.9}$$

and for a general covector $\sum_i a^V{}_i dx_V{}^i = \sum_{ij} a^V{}_i (\partial x_V{}^i/\partial x_U{}^j)\, dx_U{}^j$ must be the same as $\sum_j a^U{}_j dx_U{}^j$. We then must have

$$a^U{}_j = \sum_i a^V{}_i \left(\frac{\partial x_V{}^i}{\partial x_U{}^j}\right) \tag{2.10}$$

But $\sum_j (\partial x_V{}^i/\partial x_U{}^j)(\partial x_U{}^j/\partial x_V{}^k) = \partial x_V{}^i/\partial x_V{}^k = \delta^i{}_k$ shows that $\partial x_U/\partial x_V$ is the inverse matrix to $\partial x_V/\partial x_U$. Equation (2.10) is, in matrix form, $a^U = a^V\,(\partial x_V/\partial x_U)$ and this yields $a^V = a^U\,(\partial x_U/\partial x_V)$, or

$$a^V{}_i = \sum_j a^U{}_j \left(\frac{\partial x_U{}^j}{\partial x_V{}^i}\right) \tag{2.11}$$

This is the *transformation rule* for the components of a *covariant* vector, and should be compared with (1.6). In the notation of (1.7) we may write

$$a^V = a^U\, c_{UV} = a^U\, c^{-1}_{VU} \tag{2.12}$$

Warning: Equation (1.6) tells us how the components of a *single* contravariant vector transform under a change of coordinates. Equation (2.11), likewise, tells us how the components of a *single* 1-form α transform under a change of coordinates. This should be compared with (2.9). This latter tells us how the n-coordinate 1-forms $dx_V{}^1, \ldots, dx_V{}^n$ are related to the n-coordinate 1-forms $dx_U{}^1, \ldots, dx_U{}^n$. In a sense we could say that the n-tuple of *covariant vectors* $(dx^1, \ldots, dx^n)$ transforms as do the *components* of a single *contravariant* vector. *We shall never use this terminology.*

See Problem 2.1 (1) at this time.

2.1c. Scalar Products in Linear Algebra

Let E be an n-dimensional vector space with a given **inner** (or **scalar) product** $\langle , \rangle$. Thus, for each pair of vectors $\mathbf{v}, \mathbf{w}$ of E, $\langle \mathbf{v}, \mathbf{w}\rangle$ is a real number, it is linear in each entry when the other is held fixed (i.e., it is *bi*linear), and it is symmetric $\langle \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{v}\rangle$. Furthermore $\langle , \rangle$ is **nondegenerate** in the sense that if $\langle \mathbf{v}, \mathbf{w}\rangle = 0$ for all $\mathbf{w}$ then $\mathbf{v} = \mathbf{0}$; that is, the only vector "orthogonal" to every vector is the zero vector. If, further $\| \mathbf{v} \|^2 := \langle \mathbf{v}, \mathbf{v}\rangle$ is positive when $\mathbf{v} \neq \mathbf{0}$, we say that the inner product is positive definite, but to accommodate relativity we shall *not* always demand this.

If $\mathbf{e}$ is a basis of E, then we may write $\mathbf{v} = \mathbf{e}v$ and $\mathbf{w} = \mathbf{e}w$. Then

$$\begin{aligned}\langle \mathbf{v}, \mathbf{w}\rangle &= \langle \sum_i \mathbf{e}_i v^i, \sum_j \mathbf{e}_j w^j\rangle \\ &= \sum_i v^i \langle \mathbf{e}_i, \sum_j \mathbf{e}_j w^j\rangle = \sum_i \sum_j v^i \langle \mathbf{e}_i, \mathbf{e}_j\rangle w^j\end{aligned}$$

If we define the matrix $G = (g_{ij})$ with entries

$$g_{ij} := \langle \mathbf{e}_i, \mathbf{e}_j \rangle$$

then

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{ij} v^i g_{ij} w^j \tag{2.13}$$

or

$$\langle \mathbf{v}, \mathbf{w} \rangle = v G w$$

The matrix (g_{ij}) is briefly called the **metric tensor**. This nomenclature will be explained in Section 2.3.

Note that when $\mathbf{e}$ is an **orthonormal basis**, that is, when $g_{ij} = \delta^i_j$ is the identity matrix (and this can happen only if the inner product is positive definite), then $\langle \mathbf{v}, \mathbf{w} \rangle = \sum_j v^j w^j$ takes the usual "euclidean" form. If one restricted oneself to the use of orthonormal bases, one would never have to introduce the matrix (g_{ij}), and this is what is done in elementary linear algebra.

By hypothesis, $\langle \mathbf{v}, \mathbf{w} \rangle$ is a linear function of $\mathbf{w}$ when $\mathbf{v}$ is held fixed. Thus if $\mathbf{v} \in E$, the function ν defined by

$$\nu(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle \tag{2.14}$$

is a linear functional, $\nu \in E^*$. Thus to each vector $\mathbf{v}$ in the inner product space E we may associate a covector ν; we shall call ν the **covariant version** of the vector $\mathbf{v}$. In terms of any basis $\mathbf{e}$ of E and the dual basis σ of E^* we have from (2.4)

$$\begin{aligned} \nu &= \sum_j \nu_j \sigma^j = \sum_j \nu(\mathbf{e}_j)\sigma^j \\ &= \sum_j \langle \mathbf{v}, \mathbf{e}_j \rangle \sigma^j \\ &= \sum_j \langle \sum_i \mathbf{e}_i v^i, \mathbf{e}_j \rangle \sigma^j \\ &= \sum_j (\sum_i v^i g_{ij}) \sigma^j \end{aligned}$$

Thus the covariant version of the vector $\mathbf{v}$ has components $\nu_j = \sum_i v^i g_{ij}$ and *it is traditional in "tensor analysis" to use the same letter* v *rather than* ν. Thus we write for the components of the covariant version

$$v_j = \sum_i v^i g_{ij} = \sum_i g_{ji} v^i \tag{2.15}$$

since $g_{ij} = g_{ji}$. The *sub*script j in v_j tells us that we are dealing with the covariant version; in tensor analysis one says that we have "lowered the upper index i, making it a j, by means of the metric tensor g_{ij}." We shall also call the (v_j), with abuse of language, **the covariant components of the contravariant vector v**.

Note that if $\mathbf{e}$ is an orthonormal basis then $v_j = v^j$.

In our finite-dimensional inner product space E, every linear functional ν is t covariant version of some vector $\mathbf{v}$. Given $\nu = \sum_j v_j \sigma^j$ we shall find $\mathbf{v}$ such th $\nu(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{w}$. For this we need only solve (2.15) for v^i in terms of the giv v_j. Since $G = (g_{ij})$ is assumed nondegenerate, the inverse matrix G^{-1} must exist and again symmetric. We shall denote the entries of this inverse matrix by the same lette g but written with superscripts

$$G^{-1} = (g^{ij})$$

Then from (2.15) we have

$$v^i = \sum_j g^{ij} v_j \qquad (2.1$$

yields the contravariant version $\mathbf{v}$ of the covector $\nu = \sum_j v_j \sigma^j$. Again we call (v^i) tl contravariant components of the covector ν.

Let us now compare the contravariant and covariant components of a vector $\mathbf{v}$ in simple case. First of all, we have immediately

$$v_j = \nu(\mathbf{e}_j) = \langle \mathbf{v}, \mathbf{e}_j \rangle \qquad (2.1$$

and then $v^i = \sum_j g^{ij} v_j = \sum_j g^{ij} \langle \mathbf{v}, \mathbf{e}_j \rangle$. Thus although we always have $\mathbf{v} = \sum_i v^i$ e

$$\mathbf{v} = \sum_i \Big(\sum_j g^{ij} \langle \mathbf{v}, \mathbf{e}_j \rangle \Big) \mathbf{e}_i$$

replaces the euclidean $\mathbf{v} = \sum_i \langle \mathbf{v}, \mathbf{e}_i \rangle \mathbf{e}_i$ that holds when the basis is orthonormal. Co sider, for instance, the plane $\mathbb{R}^2$, where we use a basis $\mathbf{e}$ that consists of *unit* but n orthogonal vectors.

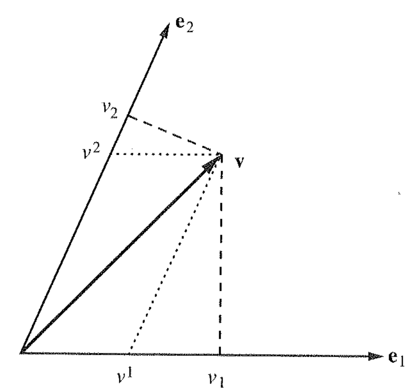

Figure 2.1

We must make some final remarks about linear functionals. It is important to realiz that given an n-dimensional vector space E, whether or not it has an inner produc one can always construct the dual vector space E^*, and the construction has nothin to do with a basis in E. If a basis $\mathbf{e}$ *is* picked for E, *then* the dual basis σ for E^* i determined. There is then an **isomorphism**, that is, a 1:1 correspondence between E^* and E given by $\sum a_j\sigma^j \to \sum a_j\mathbf{e}_j$, but this isomorphism is said to be "unnatural" since if we change the basis in E the correspondence will change. We shall *never* use this correspondence. Suppose now that an inner product has been introduced into E. As we have seen, there is another correspondence $E^* \to E$ that is independent of basis; namely to $\nu \in E^*$ we associate the unique vector $\mathbf{v}$ such that $\nu(\mathbf{w}) = \langle \mathbf{v}, \mathbf{w}\rangle$; we may write $\nu = \langle \mathbf{v}, \bullet\rangle$. In terms of a basis we are associating to $\nu = \sum v_i\sigma^i$ the vector $\sum v^i\mathbf{e}_i$. Then we know that each σ^i can be represented as $\sigma^i = \langle \mathbf{f}_i, \bullet\rangle$; that is, there is a unique vector $\mathbf{f}_i$ such that $\sigma^i(\mathbf{w}) = \langle \mathbf{f}_i, \mathbf{w}\rangle$ for all $\mathbf{w} \in E$. Then $\mathbf{f} = \{\mathbf{f}_i\}$ is a new basis of the original vector space E, sometimes called the basis of E dual to $\mathbf{e}$, and we have $\langle \mathbf{f}_i, \mathbf{e}_j\rangle = \delta^i_j$. Although this new basis *is* used in applied mathematics, *we shall not do so*, for there is a very powerful calculus that has been developed for *co*vectors, a calculus that cannot be applied to vectors!

2.1d. Riemannian Manifolds and the Gradient Vector

A **Riemannian metric** on a manifold M^n assigns, in a differentiable fashion, a positive definite inner product $\langle , \rangle$ in each tangent space M^n_p. If $\langle , \rangle$ is only nondegenerate (i.e., $\langle \mathbf{u}, \mathbf{v}\rangle = 0$ for all $\mathbf{v}$ only if $\mathbf{u} = \mathbf{0}$) rather than positive definite, then we shall call the resulting structure on M^n a **pseudo-Riemannian** metric. A manifold with a (pseudo-) Riemannian metric is called a (pseudo-) **Riemannian manifold**.

In terms of a coordinate basis $\mathbf{e}_i = \partial_i := \partial/\partial x^i$ we then have the differentiable matrices (the "metric tensor")

$$g_{ij}(x) = \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle$$

as in (2.13). In an overlap $U \cap V$ we have

$$g^V_{ij} = \left\langle \frac{\partial}{\partial x_V{}^i}, \frac{\partial}{\partial x_V{}^j} \right\rangle \tag{2.18}$$

$$= \langle \sum_r \left(\frac{\partial x_U{}^r}{\partial x_V{}^i}\right)\partial^U_r, \sum_s \left(\frac{\partial x_U{}^s}{\partial x_V{}^j}\right)\partial^U_s \rangle$$

$$g^V_{ij} = \sum_{rs} \left(\frac{\partial x_U{}^r}{\partial x_V{}^i}\right)\left(\frac{\partial x_U{}^s}{\partial x_V{}^j}\right) g^U_{rs}$$

This is the transformation rule for the components of the metric tensor.

Definition: If M^n is a (pseudo-) Riemannian manifold and f is a differentiable function, the **gradient vector**

$$\text{grad } f = \nabla f$$

is the contravariant vector associated to the covector df

$$df(\mathbf{w}) = \langle \nabla f, \mathbf{w} \rangle \tag{2.19}$$

In coordinates

$$(\nabla f)^i = \sum_j g^{ij} \frac{\partial f}{\partial x^j}$$

Note then that $\| \nabla f \|^2 := \langle \nabla f, \nabla f \rangle = df(\nabla f) = \sum_{ij} (\partial f/\partial x^i) g^{ij} (\partial f/\partial x^j)$. We see that df and ∇f will have the same components if the metric is "euclidean," that is, if the coordinates are such that $g^{ij} = \delta^i_j$.

Example (special relativity): Minkowski space is, as we shall see in Chapter 7, $\mathbb{R}^4$ but endowed with the pseudo-Riemannian metric given in the so-called inertial coordinates $t = x^0,\ x = x^1,\ y = x^2,\ z = x^3$, by

$$\begin{aligned} g_{ij} = \left\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right\rangle &= 1 && \text{if } i = j = 1, 2, \text{ or } 3 \\ &= -c^2 && \text{if } i = j = 0, \quad \text{where } c \text{ is the speed of light} \\ &= 0 && \text{otherwise} \end{aligned}$$

that is, (g_{ij}) is the 4×4 diagonal matrix

$$(g_{ij}) = \text{diag}(-c^2, 1, 1, 1)$$

Then

$$df = \left(\frac{\partial f}{\partial t}\right) dt + \sum_{j=1}^{3} \left(\frac{\partial f}{\partial x^j}\right) dx^j$$

is classically written in terms of components

$$df \sim \left[\frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right]$$

but

$$\nabla f = -\frac{1}{c^2}\left(\frac{\partial f}{\partial t}\right) \boldsymbol{\partial}_t + \sum_{j=1}^{3} \left(\frac{\partial f}{\partial x^j}\right) \boldsymbol{\partial}_j$$

$$\nabla f \sim \left[-\frac{1}{c^2}\frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right]^T$$

(It should be mentioned that the famous **Lorentz transformations** in general are simply *the* changes of coordinates in $\mathbb{R}^4$ that leave the origin fixed and *preserve the form* $-c^2t^2 + x^2 + y^2 + z^2$, just as orthogonal transformations in $\mathbb{R}^3$ are those transformations that preserve $x^2 + y^2 + z^2$!)

2.1e. Curves of Steepest Ascent

The gradient vector in a Riemannian manifold M^n has much the same meaning as in euclidean space. If $\mathbf{v}$ is a unit vector at $p \in M$, then the derivative of f with respect to $\mathbf{v}$ is $\mathbf{v}(f) = \sum (\partial f/\partial x^j) v^j = df(\mathbf{v}) = \langle \nabla f, \mathbf{v} \rangle$. Then Schwarz's inequality (which holds for a positive definite inner product), $|\mathbf{v}(f)| = |\langle \nabla f, \mathbf{v} \rangle| \leq \| \nabla f \| \| \mathbf{v} \| = \| \nabla f \|$, shows that f has a maximum rate of change in the direction of ∇f. If $f(p) = a$, then the **level set** of f through p is the subset defined by

$$M^{n-1}(a) := \{x \in M^n \mid f(x) = a\}$$

A good example to keep in mind is the torus of Figure 1.18. If df does not vanish at p then $M^{n-1}(a)$ is a submanifold in a neighborhood of p. If $x = x(t)$ is a curve in this level set through p then its velocity vector there, dx/dt, is "annihilated" by df; $df(dx/dt) = 0$ since $f(x(t))$ is constant. We are tempted to say that df is "orthogonal" to the tangent space to $M^{n-1}(a)$ at p, but this makes no sense since df is not a vector. *Its contravariant version* ∇f *is, however, orthogonal to this tangent space* since $\langle \nabla f, dx/dt\rangle = df(dx/dt) = 0$ for all tangents to $M^{n-1}(a)$ at p. We say that ∇f is orthogonal to the level sets.

Finally recall that we showed in paragraph 1.4b that one does not get a well-defined flow by considering the local differential equations $dx^i/dt = \partial f/\partial x^i$; one simply cannot equate a contravariant vector dx/dt with a covariant vector df. However it makes good sense to write $dx/dt = \nabla f$; that is, the "correct" differential equations are

$$\frac{dx^i}{dt} = \sum_j g^{ij}\left(\frac{\partial f}{\partial x^j}\right)$$

The integral curves are then tangent to ∇f, and so are orthogonal to the level sets $f =$ constant. How does f change along one of these "curves of steepest ascent"? Well, $df/dt = df(dx/dt) = \langle \nabla f, \nabla f\rangle$. Note then that if we solve *instead* the differential equations

$$\frac{dx}{dt} = \frac{\nabla f}{\|\nabla f\|^2}$$

(i.e., we move along the same curves of steepest ascent but at a different speed) then $df/dt = 1$. *The resulting flow has then the property that in time t it takes the level set* $f = a$ *into the level set* $f = a + t$. Of course this result need only be true locally and for small t (see 1.4a). Such a motion of level sets into level sets is called a **Morse deformation**. For more on such matters see [M, chap. 1].

Problems

2.1(1) If **v** is a vector and α is a covector, compute directly in coordinates that $\sum a_i^V v_V^i = \sum a_j^U v_U^j$. What happens if **w** is another vector and one considers $\sum v^i w^i$?

2.1(2) Let x, y, and z be the usual cartesian coordinates in $\mathbb{R}^3$ and let $u^1 = r$, $u^2 = \theta$ (colatitude), and $u^3 = \phi$ be spherical coordinates.

(i) Compute the metric tensor components for the spherical coordinates

$$g_{r\theta} := g_{12} = \left\langle \frac{\partial}{\partial r}, \frac{\partial}{\partial \theta} \right\rangle \quad \text{etc.}$$

(Note: Don't fiddle with matrices; just use the chain rule $\partial/\partial r = (\partial x/\partial r)\partial/\partial x + \cdots$)

(ii) Compute the coefficients $(\nabla f)^j$ in

$$\nabla f = (\nabla f)^r \frac{\partial}{\partial r} + (\nabla F)^\theta \frac{\partial}{\partial \theta} + (\nabla f)^\phi \frac{\partial}{\partial \phi}$$

(iii) Verify that $\partial/\partial r$, $\partial/\partial\theta$, and $\partial/\partial\phi$ are orthogonal, but that not all are unit vectors. Define the unit vectors $\mathbf{e}'_j = (\partial/\partial u^j)/ \parallel \partial/\partial u^j \parallel$ and write ∇f in terms of this orthonormal set

$$\nabla f = (\nabla f)'^r \mathbf{e}'_r + (\nabla f)'^\theta \mathbf{e}'_\theta + (\nabla f)'^\phi \mathbf{e}'_\phi$$

These new components of grad f are the usual ones found in all physics books (they are called the **physical** components); *but we shall have little use for such components*; df, as given by the simple expression $df = (\partial f/\partial r)\, dr + \cdots$, frequently has all the information one needs!

2.2. The Tangent Bundle

What is the space of velocity vectors to the configuration space of a dynamical system?

2.2a. The Tangent Bundle

The **tangent bundle**, TM^n, to a differentiable manifold M^n is, by definition, the collection of all tangent vectors at all points of M.

Thus a "point" in this new space consists of a pair $(p, \mathbf{v})$, where p is a point of M and $\mathbf{v}$ is a tangent vector to M at the point p, that is, $\mathbf{v} \in M^n_p$. Introduce local coordinates in TM as follows. Let $(p, \mathbf{v}) \in TM^n$. p lies in some local coordinate system U, $x^1, \ldots, x^n$. At p we have the coordinate basis $(\boldsymbol{\partial}_i = \partial/\partial x^i)$ for M^n_x. We may then write $\mathbf{v} = \sum_i v^i \boldsymbol{\partial}_i$. Then $(p, \mathbf{v})$ is completely described by the $2n$-tuple of real numbers

$$x^1(p), \ldots, x^n(p), v^1, \ldots, v^n$$

The $2n$-tuple (x,v) represents the vector $\sum_j v^j \boldsymbol{\partial}_j$ at p. In this manner we associate $2n$ local coordinates to each tangent vector to M^n that is based in the coordinate patch (U, x). Note that the first n-coordinates, the x's, take their values in a portion U of $\mathbb{R}^n$, whereas the second set, the v's, fill out an entire $\mathbb{R}^n$ since there are no restrictions on the components of a vector. This $2n$-dimensional coordinate patch is then of the form $(U \subset \mathbb{R}^n) \times \mathbb{R}^n \subset \mathbb{R}^{2n}$. Suppose now that the point p also lies in the coordinate patch (U', x'). Then the same point $(p, \mathbf{v})$ would be described by the new $2n$-tuple

$$x'^1(p), \ldots, x'^n(p), v'^1, \ldots, v'^n$$

where

$$x'^i = x'^i(x^1, \ldots, x^n) \tag{2.20}$$

and

$$v'^i = \sum_j \left[\frac{\partial x'^i}{\partial x^j}\right](p) v^j$$

We see then that TM^n *is a $2n$-dimensional differentiable manifold*!

We have a mapping

$$\pi : TM \to M \qquad \pi(p, \mathbf{v}) = p$$

called **projection** that assigns to a vector tangent to M the point in M at which the vector sits. In local coordinates,

$$\pi(x^1, \ldots, x^n, v^1, \ldots, v^n) = (x^1, \ldots, x^n)$$

It is clearly differentiable.

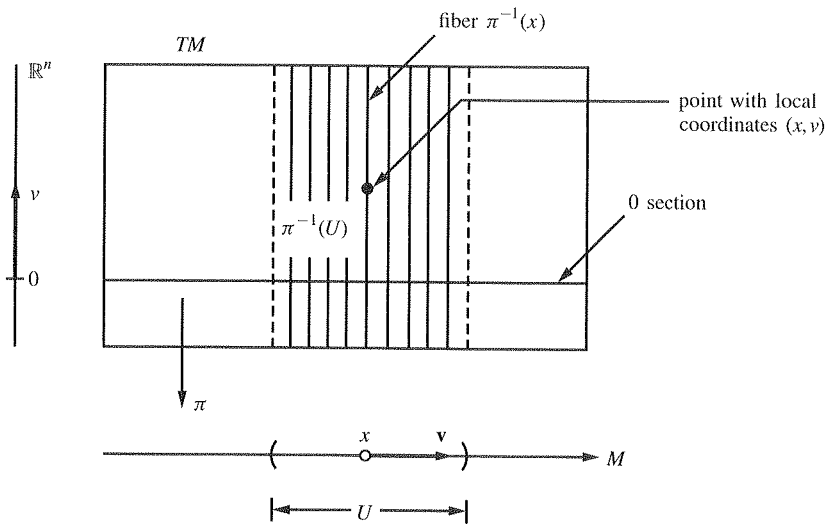

Figure 2.2

We have drawn a schematic diagram of the tangent bundle TM. $\pi^{-1}(x)$ represents all vectors tangent to M at x, and so $\pi^{-1}(x) = M_x^n$ is a copy of the vector space $\mathbb{R}^n$. It is called "the **fiber** over x." Our picture makes it seem that TM is the product space $M \times \mathbb{R}^n$, *but this is not so*! Although we do have a global projection $\pi : TM \to M$, *there is no projection map* $\pi' : TM \to \mathbb{R}^n$.

A point in TM represents a tangent vector to M at a point p but there is no way to read off the components of this vector until a coordinate system (or basis for M_p) has been designated at the point at which the vector is based!

Locally of course we may choose such a projection; if the point is in $\pi^{-1}(U)$ then by using the coordinates in U we may read off the components of the vector. Since $\pi^{-1}(U)$ is topologically $U \times \mathbb{R}^n$ we say that the tangent bundle TM is **locally a product**.

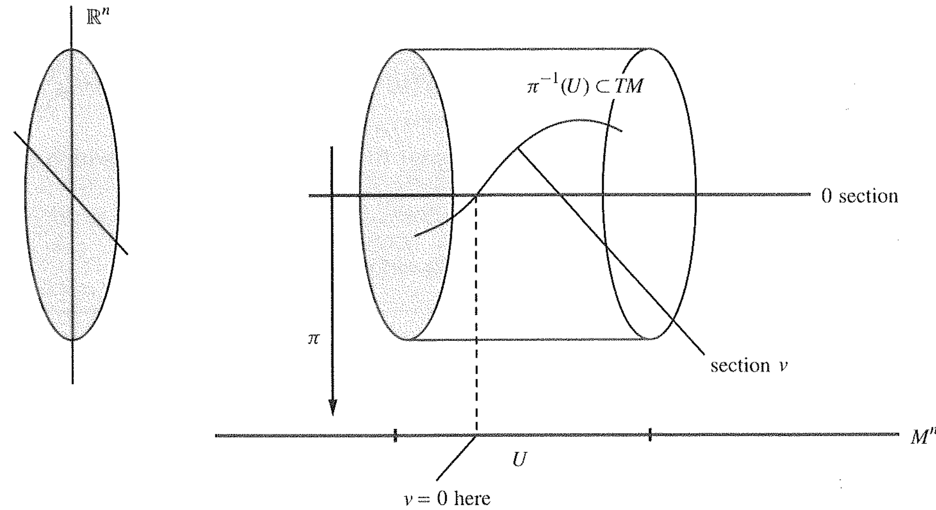

Figure 2.3

A vector field $\mathbf{v}$ on M clearly assigns to each point x in M a point $\mathbf{v}(x)$ in $\pi^{-1}(x) \subset TM$ that "lies over x." Thus a vector field can be considered as a map $v : M \to TM$ such that $\pi \circ v$ is the identity map of M into M. As such it is called a **(cross) section** of the tangent bundle. In a patch $\pi^{-1}(U)$ it is described by $v^i = v^i(x^1, \dots, x^n)$ and the image $v(M)$ is then an n-dimensional submanifold of the $2n$-dimensional manifold TM. A special section, the 0 section (corresponding to the identically 0 vector field), always exists. Although different coordinate systems will yield perhaps different components for a given vector, they will all agree that the 0-vector will have all components 0.

Example: In mechanics, the configuration of a dynamical system with n degrees of freedom is usually described as a point in an n-dimensional manifold, the **configuration space**. The coordinates x are usually called $q^1, \dots, q^n$, the "generalized coordinates." For example, if we are considering the motion of two mass points on the real line $M^2 = \mathbb{R} \times \mathbb{R}$ with coordinates q^1, q^2 (one for each particle). The configuration space need not be euclidean space. For the planar double pendulum of paragraph 1.2b (v) the configuration space is $M^2 = S^1 \times S^1 = T^2$. For the *spatial* single pendulum M^2 is the 2-sphere S^2 (with center at the pin). A tangent vector to the configuration space M^n is thought of, in mechanics, as a velocity vector; its components with respect to the coordinates q are written $\dot{q}_1, \dots, \dot{q}_n$ rather than $v^1, \dots, v^n$. These are the **generalized velocities**. Thus TM is the space of all generalized velocities, but there is no standard name for this space in mechanics (it is *not* the phase space, to be considered shortly).

2.2b. The Unit Tangent Bundle

If M^n is a Riemannian manifold (see 2.1d) then we may consider, in addition to TM the space of all *unit* tangent vectors to M^n. Thus in TM we may restrict ourselves to the subset T_0M of points $(x, \mathbf{v})$ such that $\| \mathbf{v} \|^2 = 1$. If we are in the coordinate patch

$(x^1, \ldots, x^n, v^1, \ldots, v^n)$ of TM, then this **unit tangent bundle** is locally defined by

$$T_0M^n : \sum_{ij} g_{ij}(x)v^i v^j = 1$$

In other words, we are looking at the locus in TM defined locally by putting the single function $f(x, v) = \sum_{ij} g_{ij}(x)v^i v^j$ equal to a constant. The local coordinates in TM are (x, v). Note, using $g_{ij} = g_{ji}$, that

$$\frac{\partial f}{\partial v^k} = 2\sum_j g_{kj}(x)v^j$$

Since $\det(g_{ij}) \neq 0$, we conclude that not all $\partial f/\partial v^k$ can vanish on the subset $v \neq 0$, and thus T_0M^n *is a* $(2n-1)$*-dimensional submanifold of* TM^n! In particular T_0M is itself a manifold.

In the following figure, $\mathbf{v}_0 = \mathbf{v}/\|\mathbf{v}\|$.

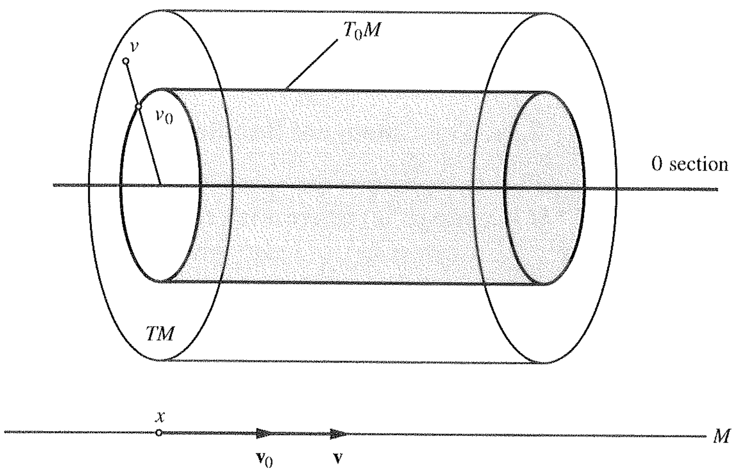

Figure 2.4

Example: T_0S^2 is the space of unit vectors tangent to the unit 2-sphere in $\mathbb{R}^3$.

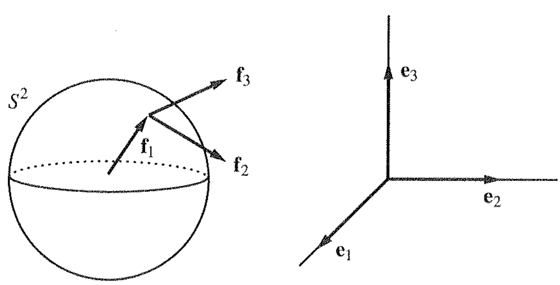

Figure 2.5

Let $\mathbf{v} = \mathbf{f}_2$ be a unit tangent vector to the unit sphere $S^2 \subset \mathbb{R}^3$. It is based at some point on S^2, described by a unit vector $\mathbf{f}_1$. Using the right-hand rule we may put $\mathbf{f}_3 = \mathbf{f}_1 \times \mathbf{f}_2$.

It is clear that by this association, there is a 1:1 correspondence between unit tange vectors $\mathbf{v}$ to S^3 (i.e., to a point in T_0S^2) and such orthonormal triples $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3$. Transla these orthonormal vectors to the origin of $\mathbb{R}^3$ and compare them with a fixed right-hand orthonormal basis $\mathbf{e}$ of $\mathbb{R}^3$. Then $\mathbf{f}_i = \mathbf{e}_j R^j{}_i$ for a unique rotation matrix $R \in SO(3$ In this way we have set up a 1:1 correspondence $T_0S^2 \to SO(3)$. It also seems evide that the topology of T_0S^2 is the same as that of $SO(3)$, meaning roughly that near unit vectors tangent to S^2 will correspond to nearby rotation matrices; precisely, v mean that $T_0S^2 \to SO(3)$ is a diffeomorphism. We have seen in 1.2b(vii) that $SO(3)$ topologically projective space.

The unit tangent bundle T_0S^2 to the 2-sphere is topologically the 3-dimensional real projective 3-space $T_0S^2 \sim \mathbb{R}P^3 \sim SO(3)$.

2.3. The Cotangent Bundle and Phase Space

What *is* phase space?

2.3a. The Cotangent Bundle

The cotangent bundle to M^n is by definition the space T^*M^n of all *covectors* at a points of M. A point in T^*M is a pair (x, α) where α is a covector at the point x. If x in a coordinate patch $U, x^1, \ldots, x^n$, then $dx^1, \ldots, dx^n$, gives a basis for the cotange space M_x^{n*}, and α can be expressed as $\alpha = \sum a_i(x)dx^i$. Then (x, α) is complete described by the $2n$-tuple

$$x^1(x), \ldots, x^n(x), a_1(x), \ldots, a_n(x)$$

The 2n-tuple (x, a) represents the covector $\sum a_i\, dx^i$ at the point x. If the point p als lies in the coordinate patch $U', x'^1, \ldots, x'^n$, then

$$x'^i = x'^i(x^1, \ldots, x^n)$$

and (2.2

$$a'_i = \sum_j \left[\frac{\partial x^j}{\partial x'^i}\right](x)a_j$$

T^*M^n is again a 2n-dimensional manifold. We shall see shortly that the phase space i mechanics is the cotangent bundle to the configuration space.

2.3b. The Pull-Back of a Covector

Recall that the differential ϕ_* of a smooth map $\phi : M^n \to V^r$ has as matrix the Jacobia matrix $\partial y/\partial x$ in terms of local coordinates $(x^1, \ldots, x^n)$ near x and $(y^1, \ldots, y^r)$ ne $y = \phi(x)$. Thus, in terms of the coordinate bases

$$\phi_*\left(\frac{\partial}{\partial x^j}\right) = \sum_R \left(\frac{\partial y^R}{\partial x^j}\right)\frac{\partial}{\partial y^R} \qquad (2.2$$

Note that if we think of vectors as differential operators, then for a function f near y

$$\phi_*\left(\frac{\partial}{\partial x^j}\right)(f) = \sum_R \left(\frac{\partial y^R}{\partial x^j}\right)\left(\frac{\partial f}{\partial y^R}\right)$$

simply says, "*Apply the chain rule to the composite function* $f \circ \phi$, that is, $f(y(x))$."

Definition: Let $\phi : M^n \to V^r$ be a smooth map of manifolds and let $\phi(x) = y$. Let $\phi_* : M_x \to V_y$ be the differential of ϕ. The **pull-back** ϕ^* is the linear transformation taking *covectors* at y into *covectors* at x, $\phi^* : V(y)^* \to M(x)^*$, defined by

$$\phi^*(\beta)(\mathbf{v}) := \beta(\phi_*(\mathbf{v})) \tag{2.23}$$

for all covectors β at y and vectors $\mathbf{v}$ at x.

Let (x^i) and (y^R) be local coordinates near x and y, respectively. The bases for the tangent vector spaces M_x and V_y are given by $(\partial/\partial x^j)$ and $(\partial/\partial y^R)$. Then

$$\begin{aligned}
\phi^*\beta &= \sum_j \phi^*(\beta)\left(\frac{\partial}{\partial x^j}\right)dx^j = \sum_j \beta\left(\phi_*\frac{\partial}{\partial x^j}\right)dx^j \\
&= \sum_j \beta\left(\sum_R \left(\frac{\partial y^R}{\partial x^j}\right)\frac{\partial}{\partial y^R}\right)dx^j \\
&= \sum_{jR} \left(\frac{\partial y^R}{\partial x^j}\right)\beta\left(\frac{\partial}{\partial y^R}\right)dx^j \\
&= \sum_{jR} b_R\left(\frac{\partial y^R}{\partial x^j}\right)dx^j, \qquad \text{where } \beta = \sum_R b_R dy^R
\end{aligned}$$

Thus

$$\phi^*(\beta) = \sum_{jR} b_R\left(\frac{\partial y^R}{\partial x^j}\right)dx^j \tag{2.24}$$

In terms of matrices, the *differential* ϕ_* is given by the Jacobian matrix $\partial y/\partial x$ acting on *columns* v at x from the *left*, whereas the *pull-back* ϕ^* is given by the same matrix acting on *rows* b at y from the *right*. (If we had insisted on writing covectors also as columns, then ϕ^* acting on such columns from the left would be given by the *transpose* of the Jacobian matrix.)

$\phi^*(dy^S)$ is given immediately from (2.24); since $dy^S = \sum_R \delta^S{}_R dy^R$

$$\phi^*(dy^S) = \sum_j \left(\frac{\partial y^S}{\partial x^j}\right)dx^j \tag{2.25}$$

This *is again simply the chain rule applied to the composition* $y^S \circ \phi$!

Warning: Let $\phi : M^n \to V^r$ and let $\mathbf{v}$ be a vector *field* on M. It may very well be that there are two distinct points x and x' that get mapped by ϕ to the same point $y = \phi(x) = \phi(x')$. Usually we shall have $\phi_*(\mathbf{v}(x)) \neq \phi_*(\mathbf{v}(x'))$ since the field $\mathbf{v}$ need have no relation to the map ϕ. In other words, $\phi_*(\mathbf{v})$ does not yield a well defined vector field on V (does one pick $\phi_*(\mathbf{v}(x))$ or $\phi_*(\mathbf{v}(x'))$ at y?). ϕ_* *does not take vector fields* *into vector fields.* (There is an exception if $n = r$ and ϕ is 1:1.) On the other hand, *if β is a covector field on V^r, then $\phi^*\beta$ is always a well-defined covector field on M^n*; $\phi^*(\beta(y))$ yields a definite covector at each point x such that $\phi(x) = y$. As we shall see, this fact makes covector fields easier to deal with than vector fields.

See Problem 2.3 (1) at this time.

2.3c. The Phase Space in Mechanics

In Chapter 4 we shall study Hamiltonian dynamics in a more systematic fashion. For the present we wish merely to draw attention to certain basic aspects that seem mysterious when treated in most physics texts, largely because they draw no distinction there between vectors and covectors.

Let M^n be the configuration space of a dynamical system and let $q^1, \ldots, q^n$ be local generalized coordinates. For simplicity, we shall restrict ourselves to time-independent Lagrangians. The Lagrangian L is then a function of the generalized coordinates q and the generalized velocities $\dot{q}$, $L = L(q, \dot{q})$. It is important to realize that q and $\dot{q}$ are $2n$-independent coordinates. (Of course *if* we consider a specific path $q = q(t)$ in configuration space then the Lagrangian along this evolution of the system is computed by putting $\dot{q} = dq/dt$.) Thus the Lagrangian L is to be considered as a function on the space of generalized velocities, that is, *L is a real-valued function on the tangent bundle to M*,

$$L : TM^n \to \mathbb{R}$$

We shall be concerned here with the transition from the Lagrangian to the Hamiltonian formulation of dynamics. Hamilton was led to define the functions

$$p_i(q, \dot{q}) := \frac{\partial L}{\partial \dot{q}^i} \tag{2.26}$$

We shall only be interested in the case when $\det(\partial p_i/\partial \dot{q}^j) \neq 0$. In many books (2.26) is looked upon merely as a change of coordinates in TM; that is, one switches from coordinates $q, \dot{q}$, to q, p. Although this is technically acceptable, it has the disadvantage that the p's do not have the direct geometrical significance that the coordinates $\dot{q}$ had. Under a change of coordinates, say from q_U to q_V in configuration space, there is an associated change in coordinates in TM

$$q_V = q_V(q_U)$$
$$\dot{q}_U^j = \sum_i \left(\frac{\partial q_U^j}{\partial q_V^i}\right)\dot{q}_V^i \tag{2.27}$$

This is the meaning of the tangent bundle! Let us see now how the p's transform.

$$p_i^V := \frac{\partial L}{\partial \dot{q}_V^i} = \sum_j \left\{\left(\frac{\partial L}{\partial q_U^j}\right)\left(\frac{\partial q_U^j}{\partial \dot{q}_V^i}\right) + \left(\frac{\partial L}{\partial \dot{q}_U^j}\right)\left(\frac{\partial \dot{q}_U^j}{\partial \dot{q}_V^i}\right)\right\}$$

However, q_V does not depend on $\dot{q}_U$; likewise q_U does not depend on $\dot{q}_V$, and therefore the first term in this sum vanishes. Also, from (2.27),

$$\frac{\partial \dot{q}_U^j}{\partial \dot{q}_V^i} = \frac{\partial q_U^j}{\partial q_V^i} \tag{2.28}$$

Thus

$$p_i^V = \sum_j p_j^U \left(\frac{\partial q_U^j}{\partial q_V^i} \right) \tag{2.29}$$

and so the p's represent then not the components of a vector on the configuration space M^n but rather a *covector*. The q's and p's then are to be thought of not as local coordinates in the tangent bundle but as coordinates for the *cotangent* bundle. Equation (2.26) is then to be considered not as a change of coordinates in TM but rather as the local description of a *map*

$$p : TM^n \to T^*M^n \tag{2.30}$$

from the tangent bundle to the cotangent bundle. We shall frequently call $(q^1, \ldots, q^n, p_1, \ldots, p_n)$ the local coordinates for T^*M^n (even when we are not dealing with mechanics). This space T^*M of covectors to the configuration space is called in mechanics the **phase space** of the dynamical system.

Recall that there is no *natural* way to identify vectors on a manifold M^n with covectors on M^n. We have managed to make such an identification, $\sum_j \dot{q}^j \partial/\partial q^j \to \sum_j (\partial L/\partial \dot{q}^j) dq^j$, by introducing an extra structure, a Lagrangian function. TM and T^*M exist as soon as a manifold M is given. We may (locally) identify these spaces by giving a Lagrangian function, but of course the identification changes with a change of L, that is, a change of "dynamics."

Whereas the $\dot{q}$'s of TM are called generalized velocities, the p's are called generalized **momenta.** This terminology is suggested by the following situation. The Lagrangian is frequently of the form

$$L(q, \dot{q}) = T(q, \dot{q}) - V(q)$$

where T is the kinetic energy and V the potential energy. V is usually independent of $\dot{q}$ and T is frequently a positive definite symmetric quadratic form in the velocities

$$T(q, \dot{q}) = \frac{1}{2} \sum_{jk} g_{jk}(q) \dot{q}^j \dot{q}^k \tag{2.31}$$

For example, in the case of two masses m_1 and m_2 moving in one dimension, $M = \mathbb{R}^2$, $TM = \mathbb{R}^4$, and

$$T = \frac{1}{2} m_1 (\dot{q}^1)^2 + \frac{1}{2} m_2 (\dot{q}^2)^2$$

and the "mass matrix" (g_{ij}) is the diagonal matrix $\mathrm{diag}(m_1, \ m_2)$.

In (2.31) we have generalized this simple case, allowing the "mass" terms to depend on the positions. For example, for a single particle of mass m moving in the plane, we have, using cartesian coordinates, $T = (1/2)m[\dot{x}^2 + \dot{y}^2]$, but if polar coordinates are used we have $T = (1/2)m[\dot{r}^2 + r^2\dot{\theta}^2]$ with the resulting mass matrix $\mathrm{diag}(m, \ mr^2)$. In the general case,

$$p_i = \frac{\partial L}{\partial \dot{q}^i} = \frac{\partial T}{\partial \dot{q}^i} = \sum_j g_{ij}(q) \dot{q}^j \tag{2.32}$$

Thus, if we think of $2T$ as defining a *Riemannian metric* on the configuration space M

$$\langle \dot{q}, \dot{q} \rangle = \sum_{ij} g_{ij}(q)\dot{q}^i \dot{q}^j$$

then the kinetic energy represents half the length squared of the velocity vector, an the *momentum* p is by (2.32) simply *the covariant version of the velocity vector* $\dot{q}$. I the case of the two masses on $\mathbb{R}$ we have

$$p_1 = m_1 \dot{q}^1 \quad \text{and} \quad p_2 = m_2 \dot{q}^2$$

are indeed what everyone calls the momenta of the two particles.

The tangent and cotangent bundles, TM and T^*M, exist for any manifold M, *ir dependent of mechanics*. They are distinct geometric objects. If, however, M is a Rie mannian manifold, we may define a diffeomorphism $TM^n \to T^*M^n$ that sends th coordinate patch $(q, \dot{q})$ to the coordinate patch (q, p) by

$$p_i = \sum_j g_{ij} \dot{q}^j$$

with inverse

$$\dot{q}^i = \sum_j g^{ij} p_j$$

We did just this in mechanics, where the metric tensor was chosen to be that define by the kinetic energy quadratic form.

2.3d. The Poincaré 1-Form

Since TM and T^*M are diffeomorphic, it might seem that there is no particular reaso for introducing the more abstract T^*M, but this is not so. *There are certain geometric objects that live naturally on T^*M, not TM.* Of course these objects can be broug back to TM by means of our identifications, but this is not only frequently awkwar it would also depend, say, on the specific Lagrangian or metric tensor employed.

Recall that "1-form" is simply another name for covector. We shall show, wi Poincaré, that there is a well-defined 1-form field on every cotangent bundle T^*M This will be a linear functional defined on each tangent vector to the $2n$-dimension manifold T^*M^n, *not* M.

Theorem (2.33): *There is a globally defined 1-form on every cotangent bundle T^*M^n, the* ***Poincaré 1-form*** λ. *In local coordinates (q, p) it is given by*

$$\lambda = \sum_i p_i dq^i$$

(Note that the most general 1-form on T^*M is locally of the form $\sum_i a_i(q, p)dq^i$ $\sum_i b_i(q, p)dp_i$, and also note that the expression given for λ cannot be considered 1-form on the manifold M since p_i is not a function on M!)

PROOF: We need only show that λ is well defined on an overlap of local coordinate patches of T^*M. Let (q', p') be a second patch. We may restrict ourselves to coordinate changes of the form (2.21), for that is how the cotangent bundle was defined. Then

$$dq'^i = \sum_j \left\{ \left(\frac{\partial q'^i}{\partial q^j}\right) dq^j + \left(\frac{\partial q'^i}{\partial p_j}\right) dp_j \right\}$$

But from (2.21), q' is independent of p, and the second sum vanishes. Thus

$$\sum_i p'_i dq'^i = \sum_i p_i' \sum_j \left(\frac{\partial q'^i}{\partial q^j}\right) dq^j = \sum_j p_j dq^j \quad \square$$

There is a simple *intrinsic* definition of the form λ, that is, a definition not using coordinates. Let A be a point in T^*M; we shall define the 1-form λ at A. A represents a 1-form α at a point $x \in M$. Let $\pi : T^*M^n \to M^n$ be the *projection* that takes a point A in T^*M, to the point x at which the form α is located. Then the pull-back $\pi^*\alpha$ defines a 1-form at each point of $\pi^{-1}(x)$, in particular at A. λ at A is precisely this form $\pi^*\alpha$!

Let us check that these two definitions are indeed the same. In terms of local coordinates (q) for M and (q, p) for T^*M the map π is simply $\pi(q, p) = (q)$. The point A with local coordinates (q, p) represents the form $\sum_j p_j dq^j$ at the point q in M. Compute the pull-back (i.e., use the chain rule)

$$\begin{aligned}
\pi^*\left(\sum_i p_i dq^i\right) &= \sum_i p_i \pi^*(dq^i) \\
&= \sum_i p_i \sum_j \left\{ \left(\frac{\partial q^i}{\partial q^j}\right) dq^j + \left(\frac{\partial q^i}{\partial p_j}\right) dp_j \right\} \\
&= \sum_i p_i \sum_j \delta^i_j dq^j = \sum_i p_i dq^i = \lambda \quad \square
\end{aligned}$$

As we shall see when we discuss mechanics, the presence of the Poincaré 1-form field on T^*M and the capability of pulling back 1-form fields under mappings endow T^*M with a powerful tool that is not available on TM.

Problems

2.3(1) Let $F : M^n \to W^r$ and $G : W^r \to V^s$ be smooth maps. Let x, y, and z be local coordinates near $p \in M$, $F(p) \in W$, and $G(F(p)) \in V$, respectively. We may consider the composite map $G \circ F : M \to V$.

(i) Show, by using bases $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$, that

$$(G \circ F)_* = G_* \circ F_*$$

(ii) Show, by using bases dx, dy, and dz, that

$$(G \circ F)^* = F^* \circ G^*$$

2.3(2) Consider the tangent bundle to a manifold M.

(i) Show that under a change of coordinates in M, $\partial/\partial q$ depends on both $\partial/\partial q'$ and $\partial/\partial \dot{q}'$.

(ii) Is the locally defined vector field $\sum_j \dot{q}^j \partial/\partial q^j$ well defined on all of TM?

(iii) Is $\sum_j q^j \partial/\partial \dot{q}^j$ well defined?

(iv) If any of the above in (ii), (iii) is well defined, can you produce an intrinsic definition?

2.4. Tensors

How does one construct a field strength from a vector potential?

2.4a. Covariant Tensors

In this paragraph we shall again be concerned with linear algebra of a vector space E. Almost all of our applications will involve the vector space $E = M^n_x$ of tangent vectors to a manifold at a point $x \in E$. Consequently we shall denote a basis $\mathbf{e}$ of E by $\partial = (\partial_1, \ldots, \partial_n)$, with dual basis $\sigma = dx = (dx^1, \ldots, dx^n)$. It should be remembered, however, that most of our constructions are simply linear algebra.

Definition: A **covariant tensor of rank r** is a **multilinear** real-valued function

$$Q : E \times E \times \cdots \times E \to \mathbb{R}$$

of r-tuples of vectors, multilinear meaning that the function $Q(\mathbf{v}_1, \ldots, \mathbf{v}_r)$ is linear in each entry provided that the remaining entries are held fixed.

We emphasize that *the values of this function must be independent of the basis in which the components of the vectors are expressed.*

A *covariant vector* is a covariant tensor of rank 1. When $r = 2$, a multilinear function is called bilinear, and so forth. Probably the most important covariant second-rank tensor is the *metric tensor* G, introduced in 2.1c:

$$G(\mathbf{v}, \mathbf{w}) = \langle \mathbf{v}, \mathbf{w} \rangle = \sum_{ij} g_{ij} v^i w^j$$

is clearly bilinear (and is assumed independent of basis).

We need a systematic notation for indices. Instead of writing $i, j, \ldots, k$, we shall write $i_1, \ldots, i_p$.

In components, we have, by multilinearity,

$$\begin{aligned} Q(\mathbf{v}_1, \ldots, \mathbf{v}_r) &= Q\left(\sum_{i_1} v_1^{i_1} \partial_{i_1}, \ldots, \sum_{i_r} v_r^{i_r} \partial_{i_r}\right) \\ &= \sum_{i_1} v_1^{i_1} Q\left(\partial_{i_1}, \ldots, \sum_{i_r} v_r^{i_r} \partial_{i_r}\right) = \ldots \\ &= \sum_{i_1, \ldots, i_r} v_1^{i_1} \ldots v_r^{i_r} Q(\partial_{i_1}, \ldots, \partial_{i_r}) \end{aligned}$$

That is,

$$Q(\mathbf{v}_1, \dots, \mathbf{v}_r) = \sum_{i_1,\dots,i_r} Q_{i_1,\dots,i_r} v_1^{i_1} \dots v_r^{i_r}$$

where (2.34)

$$Q_{i_1,\dots,i_r} := Q(\partial_{i_1}, \dots, \partial_{i_r})$$

We now introduce a very useful notational device, the *Einstein summation convention*. In any single term involving indices, a summation is implied over any index that appears as both an upper (contravariant) and a lower (covariant) index. For example, in a matrix $A = (a^i{}_j)$, $a^i{}_i = \sum_i a^i{}_i$ is the trace of the matrix. With this convention we can write

$$Q(\mathbf{v}_1, \dots, \mathbf{v}_r) = Q_{i_1,\dots,i_r} v_1^{i_1} \dots v_r^{i_r} \tag{2.35}$$

The collection of all covariant tensors of rank r forms a vector space under the usual operations of addition of functions and multiplication of functions by real numbers. These simply correspond to addition of their *components* $Q_{i,\dots,j}$ and multiplication of the components by real numbers. The number of components in such a tensor is clearly n^r. This vector space is the space of covariant r^{th} rank tensors and will be denoted by

$$E^* \otimes E^* \otimes \cdots \otimes E^* = \otimes^r E^*$$

If α and β are covectors, that is, elements of E^*, we can form the second-rank covariant tensor, the **tensor product** of α and β, as follows. We need only tell how $\alpha \otimes \beta : E \times E \to \mathbb{R}$.

$$\alpha \otimes \beta(\mathbf{v}, \mathbf{w}) := \alpha(\mathbf{v})\,\beta(\mathbf{w})$$

In components, $\alpha = a_i dx^i$ and $\beta = b_j dx^j$, and from (2.34)

$$(\alpha \otimes \beta)_{ij} = \alpha \otimes \beta(\partial_i, \partial_j) = \alpha(\partial_i)\beta(\partial_j) = a_i b_j$$

$(a_i b_j)$, where $i, j = 1, \dots, n$, form the components of $\alpha \otimes \beta$. See Problem 2.4 (1) at this time.

2.4b. Contravariant Tensors

Note first that a contravariant vector, that is, an element of E, can be considered as a linear functional on covectors by defining

$$\mathbf{v}(\alpha) := \alpha(\mathbf{v})$$

In components $\mathbf{v}(\alpha) = a_i v^i$ is clearly linear in the components of α.

Definition: A **contravariant tensor of rank** s is a multilinear real valued function T on s-tuples of *covectors*

$$T : E^* \times E^* \times \cdots \times E^* \to \mathbb{R}$$

As for covariant tensors, we can show immediately that for an s-tuple of 1-forms $\alpha_1, \ldots, \alpha_s$

$$T(\alpha_1, \ldots, \alpha_s) = a_{1\,i_1} \ldots a_{s\,i_s} T^{i_1 \ldots i_s}$$

where (2.36)

$$T^{i_1 \ldots i_s} := T(dx^{i_1}, \ldots, dx^{i_s})$$

We write for this space of contravariant tensors

$$E \otimes E \otimes \cdots \otimes E := \otimes^r E$$

Contravariant vectors are of course contravariant tensors of rank 1. An example of a second-rank contravariant tensor is the inverse to the metric tensor G^{-1}, with components (g^{ij}),

$$G^{-1}(\alpha, \beta) = g^{ij}\, a_i b_j$$

(see 2.1c). Does the matrix g^{ij} really define a *tensor* G^{-1}? The local expression for G^{-1} (α, β) given is certainly bilinear, but are the values really independent of the coordinate expressions of α and β? Note that the vector **b** associated to β is coordinate-independent since $\beta(\mathbf{v}) = \langle \mathbf{v}, \mathbf{b} \rangle$, and the metric $\langle , \rangle$ is coordinate-independent. But then $G^{-1}(\alpha, \beta) = g^{ij} a_i b_j = a_i b^i = \alpha(\mathbf{b})$ is indeed independent of coordinates, and G^{-1} is a tensor.

Given a pair $\mathbf{v}$, $\mathbf{w}$ of contravariant vectors, we can form their tensor product $\mathbf{v} \otimes \mathbf{w}$ in the same manner as we did for covariant vectors. It is the second-rank contravariant tensor with components $(\mathbf{v} \otimes \mathbf{w})^{ij} = v^i w^j$. As in Problem 2.4 (1) we may then write

$$G = g_{ij}\, dx^i \otimes dx^j \quad \text{and} \quad G^{-1} = g^{ij} \partial_i \otimes \partial_j \tag{2.37}$$

2.4c. Mixed Tensors

The following definition in fact includes that of covariant and contravariant tensors as special cases when r or $s = 0$.

Definition: A **mixed tensor**, r times covariant and s times contravariant, is a real multilinear function W

$$W : E^* \times E^* \times \cdots \times E^* \times E \times E \times \cdots \times E \to \mathbb{R}$$

on s-tuples of covectors and r-tuples of vectors.

By multilinearity

$$W(\alpha_1, \ldots, \alpha_s, \mathbf{v}_1, \ldots, \mathbf{v}_r) = a_{1\,i_1} \ldots a_{s\,i_s}\, W^{i_1 \ldots i_s}{}_{j_1 \ldots j_r} v_1^{j_1} \ldots v_r^{j_r}$$

where (2.38)

$$W^{i_1 \ldots i_s}{}_{j_1 \ldots j_r} := W(dx^{i_1}, \ldots, \partial_{j_r})$$

A second-rank mixed tensor arises from a *linear transformation* $\mathbf{A} : E \to E$. Define $W_A : E^* \times E \to \mathbb{R}$ by $W_A(\alpha, \mathbf{v}) = \alpha(\mathbf{A}\mathbf{v})$. *Let* $A = (A^i_j)$ be the matrix of $\mathbf{A}$, that is, $\mathbf{A}(\partial_j) = \partial_i A^i_j$. The components of W_A are given by

$$W_A{}^i{}_j = W_A(dx^i, \partial_j) = dx^i(\mathbf{A}(\partial_j)) = dx^i(\partial_k A^k{}_j) = \delta^i_k A^k{}_j = A^i{}_j$$

The matrix of the mixed tensor W_A is simply the matrix of $\mathbf{A}$! Conversely, given a mixed tensor W, once covariant and once contravariant, we can define a linear transformation $\mathbf{A}$ by saying $\mathbf{A}$ is that unique linear transformation such that $W(\alpha, \mathbf{v}) = \alpha(\mathbf{A}\mathbf{v})$. Such an $\mathbf{A}$ exists since $W(\alpha, \mathbf{v})$ *is* linear in $\mathbf{v}$. We shall not distinguish between a linear transformation $\mathbf{A}$ and its associated mixed tensor W_A; a linear transformation $\mathbf{A}$ *is* a mixed tensor with components $(A^i{}_j)$.

Note that in components the bilinear form has a pleasant matrix expression

$$W(\alpha, \mathbf{v}) = a_i A^i{}_j v^j = a A v$$

The tensor product $\mathbf{w} \otimes \beta$ of a vector and a covector is the mixed tensor defined by

$$(\mathbf{w} \otimes \beta)(\alpha, \mathbf{v}) = \alpha(\mathbf{w})\beta(\mathbf{v})$$

As in Problem 2.4 (1)

$$\mathbf{A} = A^i{}_j \partial_i \otimes dx^j = \partial_i \otimes A^i{}_j \, dx^j$$

In particular, the *identity* linear transformation is

$$I = \partial_i \otimes dx^i \tag{2.38}$$

and its components are of course δ^i_j.

Note that we have written matrices A in three different ways, A_{ij}, A^{ij}, and $A^i{}_j$. The first two define bilinear forms (on E and E^*, respectively)

$$A_{ij} v^i w^j \quad \text{and} \quad A^{ij} a_i b_j$$

and only the last is the matrix of a linear transformation $\mathbf{A} : E \to E$. A point of confusion in elementary linear algebra arises since the matrix of a linear transformation there is usually written A_{ij} and they make no distinction between linear transformations and bilinear forms. We *must* make the distinction. In the case of an inner product space E, $\langle , \rangle$ we may relate these different tensors as follows. Given a linear transformation $\mathbf{A} : E \to E$, that is, a mixed tensor, we may associate a covariant bilinear form A' by

$$A'(\mathbf{v}, \mathbf{w}) := \langle \mathbf{v}, \mathbf{A}\mathbf{w} \rangle = v^i \, g_{ij} A^j{}_k w^k$$

Thus $A'_{ik} = g_{ij} A^j{}_k$. Note that we have "lowered the index j, making it a k, by means of the metric tensor." *In tensor analysis one uses the same letter*; that is, instead of A' one merely writes A,

$$A_{ik} := g_{ij} A^j{}_k \tag{2.39}$$

It is clear from the placement of the indices that we now have a covariant tensor. This is the matrix of the covariant bilinear form associated to the linear transformation $\mathbf{A}$. In general its components differ from those of the mixed tensor, *but they coincide when* *the basis is orthonormal*, $g_{ij} = \delta^i_j$. Since orthonormal bases are almost always used in elementary linear algebra, they may dispense with the distinction.

In a similar manner one may associate to the linear transformation **A** a contravariant bilinear form

$$\bar{A}(\alpha, \beta) = a_i A^i{}_j g^{jk} b_k$$

whose matrix of components would be written

$$A^{ik} = A^i{}_j g^{jk}$$

Recall that the components of a second-rank tensor always form a matrix such that the *left*-most index denotes the *row* and the *right*-most index the *column, independen of whether the index is up or down.*

A final remark. The metric tensor $\{g_{ij}\}$, being a covariant tensor, does *not* represent a linear transformation of E into itself. However, it *does* represent a linear transformation from E to E^*, sending the vector with components v^j into the covector with components $g_{ij} v^j$.

2.4d. Transformation Properties of Tensors

As we have seen, a mixed tensor W has components (with respect to a basis $\boldsymbol{\partial}$ of E and the dual basis dx of E^*) given by

$$W^{i\cdots j}{}_{k\cdots l} = W(dx^i, \ldots, dx^j, \boldsymbol{\partial}_k, \ldots, \boldsymbol{\partial}_l).$$

Under a change of bases, $\boldsymbol{\partial}'_l = \boldsymbol{\partial}_s(\partial x^s / \partial x'^l)$ and $dx'^i = (\partial x'^i / \partial x^c)\, dx^c$ we have, by multilinearity,

$$W'^{i\cdots j}{}_{k\cdots l} = W(dx'^i, \ldots, dx'^j, \boldsymbol{\partial}'_k, \ldots, \boldsymbol{\partial}'_l) \tag{2.41a}$$

$$= \left(\frac{\partial x'^i}{\partial x^c}\right) \cdots \left(\frac{\partial x'^j}{\partial x^d}\right)\left(\frac{\partial x^r}{\partial x'^k}\right) \cdots \left(\frac{\partial x^s}{\partial x'^l}\right) W^{c\cdots d}{}_{r\cdots s}$$

Similarly, for covariant Q and contravariant T we have

$$Q'_{i\ldots j} = \left(\frac{\partial x^k}{\partial x'^i}\right) \cdots \left(\frac{\partial x^l}{\partial x'^j}\right) Q_{k\ldots l} \tag{2.41b}$$

and

$$T'^{i\ldots j} = \left(\frac{\partial x'^i}{\partial x^k}\right) \cdots \left(\frac{\partial x'^j}{\partial x^l}\right) T^{k\ldots l} \tag{2.41c}$$

Classical tensor analysts dealt not with multilinear functions, but rather with the components. They would say that a mixed tensor assigns, to each basis of E, a collectio of "components" $W^{i\cdots j}{}_{k\ldots l}$ such that under a change of basis the components transfor by the law (2.41a). This is a convenient terminology generalizing (2.1).

Warning: A linear transformation (mixed tensor) A has eigenvalues λ determine by the equation $A\mathbf{v} = \lambda\mathbf{v}$, that is, $A^i_j v^j = \lambda v^i$, but a covariant second-rank tensor does not. This is evident just from our notation; $Q_{ij}\, v^j = \lambda v^i$ makes no sense sinc i is a covariant index on the left whereas it is a contravariant index on the right. C course we *can* solve the linear equations $Q_{ij}\, v^j = \lambda v^i$ as in linear algebra; that i we solve the secular equation $\det(Q - \lambda I) = 0$, but the point is that *the solutions* *depend on the basis* used. Under a change of basis, the transformation rule (2.41b) says $Q'_{ij} = (\partial x^k/\partial x'^i) Q_{kl} (\partial x^l/\partial x'^j)$. Thus we have

$$Q' = \left(\frac{\partial x}{\partial x'}\right)^T Q \left(\frac{\partial x}{\partial x'}\right)$$

and the solutions of $\det[Q' - \lambda I] = 0$ in general differ from those of $\det[Q - \lambda I] = 0$. (In the case of a *mixed* tensor W, the transpose T is replaced by the inverse, yielding an invariant equation $\det(W' - \lambda I) = \det(W - \lambda I)$.) *It thus makes no intrinsic sense to talk about the eigenvalues or eigenvectors of a quadratic form.* Of course *if we have a metric tensor* g given, to a covariant matrix Q we may form the mixed version $g^{ij} Q_{jk} = W^i{}_k$ and then find the eigenvalues of this W. This is equivalent to solving

$$Q_{ij} v^j = \lambda g_{ij} v^j$$

and this requires

$$\det(Q - \lambda g) = 0$$

It is easy to see that this equation is independent of basis, as is clear also from our notation. We may call these eigenvalues λ **the eigenvalues of the quadratic form with respect to the given metric** g. This situation arises in the problems of *small oscillations of a mechanical system*; see Problem 2.4(4).

2.4e. Tensor Fields on Manifolds

A (differentiable) tensor field on a manifold has components that vary differentiably. A Riemannian metric (g_{ij}) is a very important second-rank covariant tensor field.

Tensors are important on manifolds because we are frequently required to construct expressions by using local coordinates, yet we wish our expressions to have an intrinsic *meaning* that all coordinate systems will agree upon.

Tensors in physics usually describe physical fields. For example, Einstein discovered that the metric tensor (g_{ij}) in 4-dimensional space–time describes the gravitational field, to be discussed in Chapter 11. (This is similar to describing the Newtonian gravitational field by the scalar Newtonian potential function ϕ.) Different observers will usually use different local coordinates in 4-space. By making measurements with "rulers and clocks," each observer can in principle measure the components g_{ij} for their coordinate system. Since the metric of space–time is assumed to have physical significance (Einstein's discovery), although two observers will find different components in their systems, the two sets of components g_{ij} and g'_{ij} will be related by the transformation law for a covariant tensor of the second rank. The observers will then want to describe and *agree* on the *strength* of the gravitational field, and this will involve derivatives of their metric components, just as the Newtonian strength is measured by grad ϕ. By "agree," we mean, presumably, that the strengths will again be components of some tensor, perhaps of higher rank. In the Newtonian case the field is described by a scalar ϕ and the strength is a vector, grad(ϕ). We shall see that this is not at all a trivial task. We shall illustrate this point with a far simpler example; this example will be dealt with more extensively later on, after we have developed the appropriate tools.

Space–time is some manifold M, perhaps not $\mathbb{R}^4$. Electromagnetism is desc locally by a "vector potential," that is, by some vector field. It is not usually cle the texts whether the vector is contravariant or covariant; recall that even in Minko space there are differences in the components of the covariant and contravariant vers of a vector field (see 2.1d). As you will learn in Problem 2.4(3), there is good re to assume that *the vector potential is a covector* $\alpha = A_j dx^j$.

In the following we shall use the popular notations $\partial_i \phi := \partial\phi/\partial x^i$, and ∂'_i $\partial\phi/\partial x'^i$.

The electromagnetic *field strength* will involve derivatives of the A's, but it wi clear from the following calculation that the expressions

$$\partial_i A_j$$

do *not* form the components of a second-rank tensor!

Theorem (2.42): *If A_j are the components of a covariant vector on any manifol then*

$$F_{ij} := \partial_i A_j - \partial_j A_i$$

form the components of a second-rank covariant tensor.

PROOF: We need only verify the transformation law in (2.42). Since $\alpha = A_j dx$ *is* a covector, we have $A'_j = (\partial'_j x^l) A_l$ and so

$$\begin{aligned} F'_{ij} &= \partial'_i A'_j - \partial'_j A'_i = \partial'_i\{(\partial'_j x^l) A_l\} - \partial'_j\{(\partial'_i x^l) A_l\} \\ &= (\partial'_j x^l)(\partial'_i A_l) + [(\partial'_i \partial'_j x^l) A_l] - (\partial'_i x^l)(\partial'_j A_l) - (\partial'_j \partial'_i x^l) A_l \\ &= (\partial'_j x^l)(\partial_r A_l)(\partial'_i x^r) - (\partial'_i x^l)(\partial_r A_l)(\partial'_j x^r) \end{aligned}$$

(and since r and l are dummy summation indices)

$$\begin{aligned} &= (\partial'_i x^l)(\partial'_j x^r)(\partial_l A_r - \partial_r A_l) \\ &= (\partial'_i x^l)(\partial'_j x^r) F_{lr} \quad \square \end{aligned}$$

Note that the term in brackets [ ] is what prevents $\partial_i A_j$ itself from defining a sor. Note also that if our manifold were $\mathbb{R}^n$ and if we restricted ourselves to *li* changes of coordinates, $x'^i = L^i{}_j x^j$, then $\partial_i A_j$ *would* transform as a tensor. One talk about objects that transform as tensors with respect to some restricted clas coordinate systems; a *cartesian* tensor is one based on cartesian coordinate syst that is, on orthogonal changes of coordinates. For the present we shall allow *all* cha of coordinates. In our electromagnetic case, (F_{ij}) *is* the **field strength tensor**.

Our next immediate task will be the construction of a mathematical machine, "exterior calculus," that will allow us systematically to generate "field strengths" eralizing (2.42).

Problems

2.4(1) Show that the second-rank tensor given in components by $a_i b_j dx^i \otimes dx^j$ has the same values as $\alpha \otimes \beta$ on any pair of vectors, and so

$$\alpha \otimes \beta = a_i\, b_j dx^i \otimes dx^j$$

2.4(2) Let $\mathbf{A} : E \to E$ be a linear transformation.

(i) Show by the transformation properties of a mixed tensor that the trace $\mathrm{tr}(\mathbf{A}) = A^i{}_i$ is indeed a scalar, that is, is independent of basis.

(ii) Investigate $\sum_i A_{ii}$.

2.4(3) Let $\mathbf{v} = v^i \partial_i$ be a contravariant vector field on M^n.

(i) Show by the transformation properties that $v_j = g_{ji} v^i$ yields a covariant vector.

For the following you will need to use the chain rule

$$\frac{\partial}{\partial x'^i}\left(\frac{\partial x'^j}{\partial x^k}\right) = \sum_r \left(\frac{\partial^2 x'^j}{\partial x^r \partial x^k}\right)\left(\frac{\partial x^r}{\partial x'^i}\right)$$

(ii) Does $\partial_j\, v^i$ yield a tensor?

(iii) Does $(\partial_i\, v^j - \partial_j\, v^i)$ yield a tensor?

2.4(4) Let $(q = 0, \dot{q} = 0)$ be an equilibrium point for a dynamical system, that is, a solution of Lagrange's equations $d/dt(\partial L/\partial \dot{q}^k) = \partial L/\partial q^k$ for which q and $\dot{q}$ are identically 0. Here $L = T - V$ where $V = V(q)$ and where $2T = g_{ij}(q)\dot{q}^i\dot{q}^j$ is assumed positive definite. Assume that $q = 0$ is a nondegenerate minimum for V; thus $\partial V/\partial q^k = 0$ and the Hessian matrix $Q_{jk} = (\partial^2 V/\partial q^j \partial q^k)(0)$ is positive definite. For an approximation of small motions near the equilibrium point one assumes q and $\dot{q}$ are small and one discards all cubic and higher terms in these quantities.

(i) Using Taylor expansions, show that Lagrange' s equations in our quadratic approximation become

$$g_{kl}(0)\ddot{q}^l = -Q_{kl}q^l$$

One may then find the eigenvalues of Q with respect to the kinetic energy metric g; that is, we may solve $\det(Q - \lambda g) = 0$. Let $y = (y^1, \ldots, y^n)$ be an (constant) eigenvector for eigenvalue λ, and put $q^i(t) := sin(t\sqrt{\lambda})y^i$.

(ii) Show that $q(t)$ satisfies Lagrange's equation in the quadratic approximation, and hence the eigendirection y yields a small harmonic oscillation with frequency $\omega = \sqrt{\lambda}$. The direction y yields a **normal mode** of vibration.

(iii) Consider the double planar pendulum of Figure 1.10, with coordinates $q^1 = \theta$ and $q^2 = \phi$, arm lengths $l_1 = l_2 = l$, and masses $m_1 = 3, m_2 = 1$. Write down T and V and show that in our quadratic approximation we have

$$g = l^2 \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad Q = gl \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

Show that the normal mode frequencies are $\omega_1 = (2g/3l)^{1/2}$ and $\omega_2 = (2g/l)^{1/2}$ with directions $(y^1, y^2) = (\theta, \phi) = (1, 2)$ and $(1, -2)$.