# Tensor Analysis on Manifolds

**Richard L. Bishop**
*University of Illinois*

**Samuel I. Goldberg**
*University of Illinois*

Dover Publications, Inc.
New York

# CHAPTER 0

# Set Theory and Topology

## 0.1. SET THEORY

Since we cannot hope to convey the significance of set theory, it is mostly for the sake of logical completeness and to fix our notation that we give the definitions and deduce the facts that follow.

### 0.1.1. Sets

Set theory is concerned with abstract objects and their relation to various collections which contain them. We do not define what a set is but accept it as a primitive notion. We gain an intuitive feeling for the meaning of sets and, consequently, an idea of their usage from merely listing some of the synonyms: class, collection, conglomeration, bunch, aggregate. Similarly, the notion of an object is primitive, with synonyms element and point. Finally, the relation between elements and sets, the idea of an element being in a set, is primitive. We use a special symbol to indicate this relation, $\in$, which is read "is an element of." The negation is written $\notin$, read "is not an element of."

As with all modern mathematics, once the primitive terms have been specified, axioms regarding their usage can be specified, so the set theory can be developed as a sequence of theorems and definitions. (For example, this is done in an appendix to J. Kelly, *General Topology*, Van Nostrand, Princeton, N.J., 1955.) However, the axioms are either very transparent intuitively or highly technical, so we shall use the naïve approach of dependence on intuition, since it is quite natural (deceptively so) and customary.

We do not exclude the possibility that sets are elements of other sets. Thus we may have $x \in A$ and $A \in \tau$, which we interpret as saying that $A$ and $\tau$ are sets, $x$ and $A$ are elements, and that $x$ belongs to $A$ and $A$ belongs to $\tau$. It may also be that $x$ belongs to the set $B$, that $x$ is itself a set, and that $\tau$ is an element

of some set. In fact, in formal set theory no distinction is made between sets and elements.

We specify a set by placing all its elements or a typical element and the condition which defines "typical" within braces, { }. In the latter case we separate the typical element from the condition by a vertical |. For example, the set having the first three odd natural numbers as its only elements is $\{1, 3, 5\}$. If $Z$ is the set of all integers, then the set of odd integers is $\{x \mid \text{there is } n \in Z \text{ such that } x = 2n + 1\}$, or, more simply, $\{x \mid x = 2n + 1,\ n \in Z\}$ or $\{2n + 1 \mid n \in Z\}$.

Set $A$ is a *subset* of set $B$ if every element of $A$ is also an element of $B$. The relation is written $A \subset B$ or $B \supset A$, which can also be read "$A$ is contained in $B$" or "$B$ contains $A$." Although the word "contain" is used for both "$\in$" and "$\subset$," the meaning is different in each case, and which is meant can be determined from the context. To make matters worse, frequently an element $x$ and the single-element set $\{x\}$ (called *singleton* $x$) are not distinguished, which destroys the distinction (notationally) between "$x \subset x$," which is always true, and "$x \in x$," which is usually false.

The sets $A$ and $B$ are *equal*, written $A = B$, if and only if $A \subset B$ and $B \subset A$.

We shall abbreviate the phrase "if and only if" as "iff."

## 0.1.2. Set Operations

For two sets $A$ and $B$, the *intersection* of $A$ and $B$, $A \cap B$, read "$A$ intersect $B$," is the set consisting of those elements which belong to both $A$ and $B$. The *union* of $A$ and $B$, $A \cup B$, consists of those elements which belong to $A$ or $B$ (or both). The operations of union and intersection are easily described in terms of the notation given above:

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\},$$
$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

Note that the use of "or" in mathematics is invariably inclusive, so that "or both" is not needed.

It is sometimes convenient to use the generalization of the operations of union and intersection to more than two sets. To include the infinite cases we start with a collection of sets which are labeled with subscripts ("indexed") from an index set $J$. Thus the collection of sets which we wish to unite or intersect has the form $\{A_\alpha \mid \alpha \in J\}$. The two acceptable notations in each case, with the first the more usual, are

$$\bigcap_{\alpha \in J} A_\alpha = \bigcap \{A_\alpha \mid \alpha \in J\}, \text{ the general intersection,}$$
$$\bigcup_{\alpha \in J} A_\alpha = \bigcup \{A_\alpha \mid \alpha \in J\}, \text{ the general union.}$$



Frequently $J$ will be finite, for example, the first $n$ positive integers, in which case we shall use one of the following forms:

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n,$$

and similarly for union,

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n.$$

In order that the intersection of sets be a set even when they have no common elements, we introduce the *empty set* $\varnothing$, the set which has no elements. For this and other reasons, appearing below, $\varnothing$ is a useful gadget. The empty set is a subset of every set.

The set-theoretic *difference* between two sets $A$ and $B$ is defined by $A - B = \{x \mid x \in A \text{ and } x \notin B\}$. We do not require that $B$ be a subset of $A$ in order for this difference to be formed. If $A \supset B$, then $A - B$ is called the *complement* of $B$ with respect to $A$. Frequently we are concerned primarily with a fixed set $A$ and its subsets, in which case we shall speak of the complement of a subset, omitting the phrase "with respect to $A$."

**Problem 0.1.2.1.** The *disjunctive union* or *symmetric difference* of two sets $A$ and $B$ is $A \triangle B = A \cup B - A \cap B = (A - B) \cup (B - A)$. Observe that $A \triangle B = B \triangle A$. Prove the last equality. A distributive law is true for these set operations: $(A \triangle B) \cap C = A \cap C \triangle B \cap C$. However, $A \triangle A = \varnothing$ for every $A$.

## 0.1.3. Cartesian Products

An *ordered pair* is an object which consists of a pair of elements distinguished as a *first element* and a *second element* of the ordered pair. The ordered pair whose first element is $a \in A$ and second element is $b \in B$ is denoted $(a, b)$. In contrast we may also consider nonordered pairs, sets having two elements, say $a$ and $b$, which would be denoted $\{a, b\}$ in accordance with what we said above. To be called a pair we should have $a \neq b$, and in any case $\{a, b\} = \{b, a\}$. On the other hand, we do consider ordered pairs of the form $(a, a)$, and if $a \neq b$, then $(a, b) \neq (b, a)$. Indeed, $(a, b) = (c, d)$ iff $a = c$ and $b = d$.

The set of ordered pairs of elements from $A$ and $B$, denoted $A \times B$,

$$A \times B = \{(a, b) \mid a \in A, b \in B\},$$

is called the *cartesian product* of $A$ and $B$.

**Problem 0.1.3.1.** Is $A \times B = B \times A$?

The operation of taking cartesian products may be iterated, in which case certain obvious identifications are made. For example, $A \times (B \times C)$ and $(A \times B) \times C$ are both considered the same as the *triple cartesian product*, which is defined to be the set of triplets (3-tuples)

$$A \times B \times C = \{(a, b, c) \mid a \in A, b \in B, c \in C\}.$$

Thus no distinction is made between $((a, b), c)$, $(a, (b, c))$, and $(a, b, c)$. More generally, we only use one $n$-fold cartesian product $A_1 \times A_2 \times \cdots \times A_n$ rather than the many different ones which could be obtained by distributing parentheses so as to take the products two at a time. If the same set is used repeatedly, we generally use exponential notation, so $A \times A \times A$ is denoted $A^3$, etc.

A subset $S$ of $A \times B$ is called a *relation* on $A$ to $B$. An alternative notation for $(a, b) \in S$ is $aSb$, which can be read "$a$ is $S$-related to $b$," although in many common examples it is read as it stands. For example, if $A = B = R$ we have the relation $<$, called "is less than," which formally consists of all those ordered pairs of real numbers $(x, y)$ such that $x$ is less than $y$. A function (see Section 0.1.4) is a special kind of relation.

Of particular importance in analysis and its special topic, tensor analysis, is the *real cartesian n-space* $R^n$, where $R$ is the set of real numbers. In the case when $n = 2$ or 3 this is not quite the same as the analytic euclidean plane or analytic euclidean space in that the word "euclidean" indicates that the additional structure derived from a particular definition of distance is being considered. Moreover, in euclidean space no single point or line has preference over any other, whereas in $R^3$ the point (0, 0, 0) and the coordinate axes are obviously distinguishable from other points and lines in $R^3$.

## 0.1.4. Functions

A *function from A into B*, denoted $f: A \to B$, is a rule which assigns to each $a \in A$ an element $fa = b \in B$. The idea of a "rule" is apparently a primitive notion in this definition, but need not be, since it can be defined in terms of the other notions previously given—"element of" and "cartesian product." This is done by means of the *graph of a function*—the subset

$$\{(a, fa) \mid a \in A\} \text{ of } A \times B.$$

The properties of a subset of $A \times B$ which are necessary and sufficient for the subset to be the graph of a function can be given in purely set-theoretic terms and the function itself can likewise be recaptured from its graph. In fact, it is customary to say that the function *is* its graph, but we shall use the distinction indicated by our phrasing of the definition given above.



Synonyms for "function" are "transformation," "map," "mapping," and "operator." Some authors use the convention that "function" is to be used for real-valued transformations.

We shall avoid the customary parentheses unless they are required to resolve ambiguity. Thus it is customary to write $f(a)$ instead of $fa$, which we used above. Parentheses must be used where $a$ is itself composite; for example, $f(a + b)$ is not the same as $fa + b$. In fact, the latter is meaningless, except that we take it conventionally to be $(fa) + b$, the general rule being that operations such as addition are to be performed after evaluation of functions in the operands.

The *domain* of a function $f: A \to B$ is $A$. The *range* (image, target) of $f$ is $fA = \{fa \mid a \in A\} \subset B$. The set $B$ is called the *range set* of $f$. An element of the range, $b = fa$, is called a *value* of $f$, or the *image of a under f*.

If $fA = B$, then we say that $f$ is *onto*, or that $f$ *maps A onto B* (in contrast to "into" above).

If for every $b \in fA$ there is just one $a \in A$ such that $b = fa$, then $f$ is said to be *one-to-one*, abbreviated 1–1. In this case we can define the *inverse of* $f$, $f^{-1}: fA \to A$, by setting $f^{-1}fa = a$.

If $f: A \to B$ and $C \subset A$, then the *restriction of f to C* is denoted $f|_C: C \to B$. It is frequently unnecessary to distinguish between $f$ and $f|_C$, since they have the same rule, but merely apply to different sets.

If $C \subset A$, then the *inclusion map* $i_C: C \to A$ is defined simply by $i_C c = c$. If $C = A$, then $i_C$ is called the *identity map on C*.

If $f: A \to B$ and $g: C \to D$, then the *composition* of $g$ and $f$, denoted $g \circ f$, is the function obtained by following $f$ by $g$, applied to every $a \in A$ for which this makes sense: $(g \circ f)a = g(fa)$. The domain of $g \circ f$ is thus $E = \{a \mid a \in A \text{ and } fa \in C\}$. (If $C \cap B = \varnothing$, then $g \circ f$ is the *empty function* $\varnothing: \varnothing \to D$.) If $g$ and $f$ are defined by formulas, or sets of formulas, the formula(s) for $g \circ f$ is obtained by *substituting* the formula(s) for $f$ into the formula(s) for $g$.

For any functions, $f, g, h$, composition is associative; that is, $(f \circ g) \circ h = f \circ (g \circ h)$.

**Problem 0.1.4.1.** Let $f: A \to B$. Suppose there is $g: B \to A$ such that $f \circ g = i_B$. Then $f$ is onto, $g$ is 1–1, $h = f|_{gB}$ is 1–1 onto, and $g = i_{gB} \circ h^{-1}$. Show by an example that $f$ need not be 1–1.

**Problem 0.1.4.2.** $f: A \to B$ is 1–1 onto iff there is $g: B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$. This characterizes $g = f^{-1}$.

**Examples.** **(a)** If $N$ is the set consisting of the first $n$ natural numbers, $N = \{z \mid z \in Z, 0 < z < n + 1\}$, then $R^n$ may be considered to be the set of

all functions, $f: N \to R$. For such a function we obtain the $n$-tuple $(f1, f2, \ldots, fn)$, and from this it is obvious how, conversely, we get a function from an $n$-tuple.

(b) The $i$th *coordinate function* $u^i: R^n \to R$, also called the *projection into the* $i$th *factor*, or *cartesian coordinate function*, is defined by $u^i(x^1, \ldots, x^n) = x^i$. If we think of $R^n$ as being functions $f: N \to R$, then we would define $u^i f = fi$.

(c) Using the idea of Example (a), infinite cartesian products may be defined: If $\{A_\alpha \mid \alpha \in J\}$ is a collection of sets, then their cartesian product is

$$\prod_{\alpha \in J} A_\alpha = \{f \mid f: J \to \bigcup_{\alpha \in J} A_\alpha \text{ and } f\alpha \in A_\alpha \text{ for every } \alpha\}.$$

The *projections* or coordinate functions, $u_\alpha: \prod_{\beta \in J} A_\beta \to A_\alpha$, are defined as in Example (b), by setting $u_\alpha f = f\alpha$. Projections are always onto.

## 0.1.5. Functions and Set Operations

If $A$ is a set, we denote by $\mathscr{P}A$ the collection of all subsets of $A$, $\mathscr{P}A = \{C \mid C \subset A\}$. $\mathscr{P}A$ is called the *power set* of $A$.

If $f: A \to B$, then we define the *power map* of $f$, $f: \mathscr{P}A \to \mathscr{P}B$ by $fC = \{fc \mid c \in C\}$ for every $C \in \mathscr{P}A$. In particular, the range of $f$ may still be denoted $fA$.

If $f: A \to B$, we also define the *complete inverse image map of* $f$, $f^{-1}: \mathscr{P}B \to \mathscr{P}A$, by $f^{-1}D = \{a \mid fa \in D\}$, for every $D \in \mathscr{P}B$. If $f$ is 1–1 and onto, then the set map $f^{-1}$ agrees with the power map of the inverse of $f$.

The facts to be established in the following problems show, generally, that the inverse image map is better behaved than the power map with respect to set operations.

**Problem 0.1.5.1.** The map $f$ is onto iff the inverse image map $f^{-1}$ is 1–1.

**Problem 0.1.5.2.** (a) $f^{-1}(D_1 \cap D_2) = (f^{-1}D_1) \cap (f^{-1}D_2)$.
(b) $f^{-1}(D_1 \cup D_2) = (f^{-1}D_1) \cup (f^{-1}D_2)$.
(c) $f(C_1 \cap C_2) \subset (fC_1) \cap (fC_2)$.
(d) $f(C_1 \cup C_2) = (fC_1) \cup (fC_2)$.

**Problem 0.1.5.3.** Find an example of $f, C_1, C_2$ such that $(fC_1) \cap (fC_2) \neq f(C_1 \cap C_2)$.

**Problem 0.1.5.4.** If $C \subset A$, we define the *characteristic function* $\Phi_C: A \to \{0,1\}$ by $\Phi_C a = 0$ if $a \in A - C$ and $\Phi_C a = 1$ if $a \in C$. Denote the set of all functions $f: A \to \{0,1\}$ by $2^A$. Show that the function $\Phi: \mathscr{P}A \to 2^A$ given by $\Phi C = \Phi_C$ is 1–1 and onto, so that $\mathscr{P}A$ and $2^A$ are essentially the same.



**Problem 0.1.5.5.** If $A$ is finite, show that $2^A$ is finite. How many elements does $2^A$ have?

**Problem 0.1.5.6.** If $F: A \to 2^A$, define $f \in 2^A$ by $fa \neq (Fa)a$ for every $a \in A$. This definition of $f$ makes sense because there are only two possibilities for $(Fa)a$. Show that $f$ is not in the range of $F$, so that $F$ cannot be onto. In particular, there can be no 1–1 correspondence between $A$ and $2^A$. This is a precise statement of the intuitively clear contention that $2^A$ is "larger" than $A$.

A set is *countable* if it is either finite or its members can be arranged in an infinite sequence; or, what is the same, there is a 1–1 map from the set into the positive integers. The set of all integers, $Z$, is countable, as can be seen from the sequence $0, 1, -1, 2, -2, 3, -3, \ldots$. The cartesian product of the positive integers with itself is countable, as can be seen from the 1–1 map taking $(m, n)$ into $2^m 3^n$. From this last statement it is easy to conclude that the union of a countable collection of countable sets is countable. It can be shown that the rational numbers are countable.

By Problem 0.1.5.6 we conclude that $2^Z$ is not countable. A similar trick using binary expansions of real numbers shows that the real numbers are not countable.

## 0.1.6. Equivalence Relations

An *equivalence relation* on a set $P$ with elements $m, n, p, \ldots$, is a relation $E$ which satisfies three properties:

(a) *Reflexivity*: For every $m$, $mEm$.
(b) *Symmetry*: If $mEn$, then $nEm$.
(c) *Transitivity*: If $mEn$ and $nEp$, then $mEp$.

($mEn$ can be read "$m$ is $E$-related to $n$.")

For every equivalence relation there is an exhaustive partition of $P$ into disjoint subsets, the *equivalence classes of* $E$, for which the equivalence class to which an arbitrary $m$ belongs is

$$[m] = \{n \mid nEm\}.$$

From (a), (b), and (c) we have

for every $m, m \in [m]$;
if $m \in [n]$, then $n \in [m]$;
if $m \in [n]$ and $n \in [p]$, then $m \in [p]$;

from which it follows that

$$[m] = [n] \quad \text{iff } mEn.$$

Conversely, if we are given an exhaustive partition of $P$ into disjoint subsets, we define two elements of $P$ to be $E$-related if they are in the same subset, and thus obtain an equivalence relation $E$ for which the subsets of the partition are the equivalence classes.

The set of equivalence classes, called the *quotient*, or $P$ *divided by* $E$, is denoted

$$P/E = \{[m] \mid m \in P\}.$$

## 0.2. TOPOLOGY

### 0.2.1. Topologies

We cannot expect to convey here much of the significance of topological spaces. It is mostly for the sake of greater logical completeness that we give the definitions and theorems that follow. An initial study of tensor analysis can almost ignore the topological aspects since the topological assumptions are either very natural (continuity, the Hausdorff property) or highly technical (separability, paracompactness). However, a deeper analysis of many of the existence problems encountered in tensor analysis requires assumption of some of the more difficult-to-use topological properties, such as compactness and paracompactness. For example, the existence of complete integral curves of vector fields (Theorem 3.4.3) and existence of maxima and minima of continuous functions (Proposition 0.2.8.3) both require compactness; existence of riemannian metrics is proved using paracompactness (Section 5.2). Finally, we expect and hope that the extensive theory of algebraic topological invariants (Betti numbers, etc.) will be used a great deal more in applied mathematics and therefore we have included a few examples and remarks hinting of such uses (cf. Morse theory in Section 3.10 and de Rham's theorem in Section 4.5).

A *topology* on a set $X$ is a subset $T$ of $\mathscr{P}X$, $T \subset \mathscr{P}X$, such that

**(a)** If $G_1, G_2 \in T$, then $G_1 \cap G_2 \in T$.

**(b)** If $\{G_\alpha \mid \alpha \in J\} \subset T$, then $\bigcup_{\alpha \in J} G_\alpha \in T$.

**(c)** $\varnothing \in T$ and $X \in T$.

The combination $(X, T)$ is called a *topological space*. The elements of $T$ are called the *open sets* of the topological space. Frequently we shall have a specific topology in mind and then speak of the topological space $X$, with $T$ being understood. The same space, however, can have many different topologies. In particular, there are always the *discrete topology* for which $T = \mathscr{P}X$ and the *concrete topology* for which $T = \{\varnothing, X\}$. These are so trivial as to be practically useless.



**Problem 0.2.1.1.** How many distinct topologies does a finite set having two or three points admit?

The *closed sets* of a topology $T$ on $X$ are the complements of the members of $T$, that is, the sets $X - G$ where $G \in T$. A topology could equally well be defined in terms of closed sets, with axioms corresponding to those above, which we state as theorems.

**Proposition 0.2.1.1.** **(a)** *A finite union of closed sets is a closed set.*

**(b)** *An arbitrary intersection of closed sets is closed.*

**(c)** $\varnothing$ *and* $X$ *are closed sets.*

We emphasize that closedness and openness are not negations of each other or even contrary to each other; a set may be only closed, or only open, or both, or neither.

If $A \subset X$, $X$ a topological space, then the union of all open sets contained in $A$ is the *interior* of $A$, denoted $A^0$. Thus $A^0 = \bigcup \{B \mid B \subset A \text{ and } B \in T\}$. By **(b)**, the interior of $A$ is an open set itself, and is in fact one of the open sets of which we take the union in its definition. It is the largest open subset of $A$.

Just as "open" and "closed," "union" and "intersection" are "dual" notions, the dual notion to "interior" is "closure." The closure of $A \subset X$ is the intersection of all closed sets containing $A$ and is denoted $A^-$. Thus $A^- = \bigcap \{B \mid A \subset B \text{ and } X - B \in T\}$ is closed by **(b)**, and is the smallest closed set containing $A$. The following theorem shows that a complete knowledge of the operations of taking the interior or the closure is adequate to determine the topology.

**Proposition 0.2.1.2.** *A set is open iff the interior of the set equals the set. A set is closed iff the closure of the set equals the set.*

Axioms for the closure operation, which is really a function $^-: \mathscr{P}X \to \mathscr{P}X$, have been formulated by Kuratowski. When they are taken as axioms, Proposition 0.2.1.2 is essentially the definition of a closed set, and the axioms for closed sets, **(a)**, **(b)**, **(c)** of Proposition 0.2.1.1 are then theorems. In our scheme Kuratowski's axioms become theorems, as follows.

**Proposition 0.2.1.3.** *For all subsets* $A$, $B$ *of* $X$:

**(a)** $(A \cup B)^- = A^- \cup B^-$.

**(b)** $A \subset A^-$.

**(c)** $(A^-)^- = A^-$.

**(d)** $\varnothing^- = \varnothing$.

**Problem 0.2.1.2.** Prove Proposition 0.2.1.3 and state and prove the dual proposition for the operation of taking the interior $^0: \mathscr{P}X \to \mathscr{P}X$.

The *boundary* (also called the *frontier*, or the *derived set*) of a set $A \subset X$ is the set $\partial A = A^- - A^0$. The elements of $\partial A$ are called *boundary points* of $A$. Again, it is possible to axiomatize topology by taking $\partial: \mathscr{P}X \to \mathscr{P}X$ as the fundamental concept. For example, if we know all about $\partial$, then open sets may be defined as those $G$ for which $G \cap \partial G = \varnothing$.

A *neighborhood of* $x \in X$ is any $A \subset X$ such that $x \in A^0$. In particular, any open set containing $x$ is a neighborhood of $x$. A *basis of neighborhoods at* $x$ is a collection of neighborhoods of $x$ such that every neighborhood of $x$ contains one of the basis neighborhoods. In particular, the collection of all open sets containing $x$ is a basis of neighborhoods at $x$, but generally there are many other possibilities for bases of neighborhoods. A *basis of neighborhoods of* $X$ is a specification of a basis of neighborhoods for each $x \in X$.

Topologies are frequently defined by the specification of a basis of neighborhoods. The definitive procedure is as follows.

A neighborhood of $x$ is any set which contains a basis neighborhood of $x$. An open set is then any set which is a neighborhood of every one of its points.

It is interesting that closed sets, closure, and boundary points can be defined directly in terms of basis neighborhoods. A set $G$ is closed iff whenever every basis neighborhood of $x$ intersects $G$, then $x \in G$. The closure of $A$ consists of those $x$ such that every basis neighborhood of $x$ intersects $A$. The boundary of $A$ consists of those points $x$ such that every basis neighborhood of $x$ intersects both $A$ and $X - A$.

## 0.2.2. Metric Spaces

Basis neighborhoods, and hence a topology, are frequently defined in turn by means of a *metric* or *distance function*, which is a function $d: X \times X \to R$ satisfying axioms as follows.

(a) For all $x, y \in X$, $d(x, y) \geq 0$ (*positivity*).

(b) If $d(x, y) = 0$, then $x = y$ (*nondegeneracy*).

(c) For all $x, y \in X$, $d(x, y) = d(y, x)$ (*symmetry*).

(d) For all $x, y, z \in X$, $d(x, y) + d(y, z) \geq d(x, z)$ (the *triangle inequality*).

There is no essential change if we also allow $+\infty$ as a value of $d$. A set with a metric function is called a *metric space*.

The *open ball* with center $x$ and radius $r > 0$ with respect to $d$ is defined as $B(x, r) = \{y \mid d(x, y) < r\}$. It can then be demonstrated that such open balls will serve as basis neighborhoods for a topology of $X$, the *metric topology* of $d$. Two metrics are *equivalent* if they give rise to the same topology.

Two metrics $d, d_1: X \times X \to R$ are *strongly equivalent* if there are positive constants $c, c_1$ such that for every $x, y \in X$, $d(x, y) \leq c_1 d_1(x, y)$ and $d_1(x, y) \leq c d(x, y)$. Strongly equivalent metrics are equivalent but not conversely. In fact, a metric $d$ is always equivalent to $d_1 = d/(1 + d)$, but these two are strongly equivalent iff $d$ is *bounded*; that is, there is a constant $k$ such that $d(x, y) \leq k$ for all $x, y$. The metric $d_1 = d/(1 + d)$ is always bounded ($k = 1$) whether $d$ is bounded or not, but a bounded metric cannot be strongly equivalent to an unbounded one.



## 0.2.3. Subspaces

If $A \subset X$ and $X$ has a topology $T$, then we get the *relative*, or *induced*, topology $T_A$ by defining

$$T_A = \{G \cap A \mid G \in T\}.$$

It is easy to verify that $T_A$ actually is a topology on $A$. When $A$ is given this topology, it is said to be a (topogical) *subspace* of $X$. The closed sets of a subspace $A$ are the intersections of closed sets of $X$ with $A$.

## 0.2.4. Product Topologies

If $X$ and $Y$ are topological spaces, then we define a topology on $X \times Y$ by specifying the basis neighborhoods of $(x, y)$ to be $G \times H \subset X \times Y$, where $G$ is a neighborhood of $x$ and $H$ is a neighborhood of $y$. The choices for $G$ and $H$ may be restricted to basis systems and there will be no difference in the resulting topology on $X \times Y$. When $X \times Y$ is provided with this topology it is called the *topological product* of $X$ and $Y$.

If $X$ and $Y$ are metric spaces with metrics $d_X$, $d_Y$, then we define $d_p$, a metric on $X \times Y$, for every $p \geq 1$, by

$$d_p((x, y), (x_1, y_1)) = [d_X(x, x_1)^p + d_Y(y, y_1)^p]^{1/p}$$

The limiting case as $p \to \infty$ is the metric $d_\infty$, given by

$$d_\infty((x, y), (x_1, y_1)) = \max[d_X(x, x_1), d_Y(y, y_1)].$$

Although these metrics are all different, they are all strongly equivalent, so give the same topology on $X \times Y$; in fact, this topology is the product topology. Indeed, the balls with respect to $d_\infty$ are just the products of balls with respect to $d_X$ and $d_Y$ of the same radii.

The *standard topology* on $R$ is that of the metric defined by absolute value of differences, $(x, y) \to |x - y|$. The *standard topology* on $R^n$ is obtained by taking repeated products of the standard topology on $R$. It is thus the topology of any of the metrics, for $p \geq 1$, $x, y \in R^n$,

$$d_p(x, y) = \left[\sum_{i=1}^{n} |u^i x - u^i y|^p\right]^{1/p}$$

$$d_\infty(x, y) = \max[|u^i x - u^i y|, \ i = 1, 2, \ldots, n].$$

Of these, $d_2$ is the usual *euclidean metric on* $R^n$, but they are all strongly equivalent to each other. Unless otherwise specified, we shall assume that a topology on $R^n$ is the standard one.

**Problem 0.2.4.1.** **(a)** For fixed $x,y$ show that $d_p(x, y)$ is a nonincreasing function of $p \geq 1$. (*Hint:* Show that the derivative is $\leq 0$.)

**(b)** For every $x,y \in R^n$, $d_1(x, y) \leq n d_\infty(x, y)$ and $d_\infty(x, y) = \lim_{p\to\infty} d_p(x, y)$.

**(c)** All $d_p$, $1 \leq p \leq \infty$, are strongly equivalent.

## 0.2.5. Hausdorff Spaces

A topological space $X$ is a *Hausdorff space* if for every $x,y \in X$, $x \neq y$, there are neighborhoods $U$, $V$ of $x$, $y$, respectively, such that $U \cap V = \varnothing$.

* In a Hausdorff space the singleton sets $\{x\}$ are closed sets.

A metric topology is always Hausdorff.

**Problem 0.2.5.1.** The product of Hausdorff spaces is a Hausdorff space.

## 0.2.6. Continuity

Let $X$, $Y$ be topological spaces. A function $f: X \to Y$ is *continuous* if for every open set $G$ in $Y$, $f^{-1}G$ is open in $X$. In other words, $f^{-1}: \mathscr{P}Y \to \mathscr{P}X$ maps open sets into open sets.

Since $f^{-1}$ behaves well with respect to set operations and, in particular, preserves complementation, we have immediately that $f$ is continuous iff $f^{-1}$ maps closed sets into closed sets.

The above definition of continuity is the most convenient one for working abstractly with topological spaces. For example, it is trivial to prove

**Proposition 0.2.6.1.** *The composition of continuous functions is continuous.*

However, we can recast this definition into forms which are more directly abstractions from the $\varepsilon = \delta$ definition of continuity of real-valued functions of a real variable. In that definition we first define continuity at $x$, and continuity itself is obtained by requiring it at every $x$. In the definition of continuity of $f: R \to R$ at $x$, where $y = fx$, the $\varepsilon$ served to define a basis neighborhood of $y$, given a priori, and the requirement was that there be a basis neighborhood of $x$ determined by $\delta$, such that $f$ map the $\delta$-neighborhood into the $\varepsilon$-neighborhood. The student should be able to show that this description is the essential content of the customary definition: "For every $\varepsilon > 0$ there is a $\delta > 0$ such that for every $x_1$ for which $|x - x_1| < \delta$ it is true that $|fx_1 - y| < \varepsilon$."



Abstracting the description in terms of neighborhoods is not a great chore. If $f: X \to Y$ we say that $f$ is *continuous at* $x \in X$ if for every neighborhood $V$ ($\leftrightarrow \varepsilon$ neighborhood) of $y = fx$ there is a neighborhood $U$ ($\leftrightarrow \delta$ neighborhood) of $x$ such that $U \subset f^{-1}V$ (or $fU \subset V$).

The following theorem shows that our definition of a continuous function is a correct abstraction of the usual one.

**Proposition 0.2.6.2.** *A function $f: X \to Y$ is continuous iff $f$ is continuous at every $x \in X$.*

**Problem 0.2.6.1.** Show that all functions $f: X \to X$ are continuous in the discrete topology and that the only continuous functions in the concrete topology are the constant functions.

The notion of a limit can also be abstracted. We define that $\lim_{x\to x_0} fx = y$ iff for every neighborhood $V$ of $y$ there is a neighborhood $U$ of $x_0$ such that $(U - \{x_0\}) \subset f^{-1}V$. It follows, as usual, that $f$ is continuous at $x_0$ iff **(a)** $\lim_{x\to x_0} fx = y$ and **(b)** $fx_0 = y$.

A *homeomorphism* $f: X \to Y$ is a 1–1 onto function such that $f$ and $f^{-1}: Y \to X$ are both continuous. If $f: X \to Y$ is 1–1 but not onto, then $f$ is said to be a *homeomorphism into* if $f$ and $f^{-1}: (\text{range} f) \to X$ are both continuous, where range $f$ is given the relative topology from $Y$. A homeomorphism $f$ is also called a *topological equivalence* because $f|_{T_X}$ and $f^{-1}|_{T_Y}$ are then 1–1 onto; that is, they give a 1–1 correspondence between the topologies $T_X$ and $T_Y$ of $X$ and $Y$. A property of a topological space is said to be a *topological property* if every homeomorphic space has the property. A *topological invariant* is a rule which associates to topological spaces an object which is a topological property of the space. The object usually consists of a number or some algebraic system.

**Problem 0.2.6.2.** $\tan: (-\pi/2, \pi/2) \to R$ is a homeomorphism, where tan = sin/cos.

## 0.2.7. Connectedness

A topological space $X$ is *connected* if the only subsets of $X$ which are both open and closed are $\varnothing$ and $X$. Another formulation of the same concept, in terms of its negation, is

**Proposition 0.2.7.1.** *A topological space $X$ is not connected iff there are nonempty open sets $G, H$ such that $G \cap H = \varnothing$, $G \cup H = X$.*

A subset $A$ of $X$ is *connected* if $A$ with the relative topology is connected. The following is not hard to prove.

**Proposition 0.2.7.2.** (Chaining Theorem). *If* $\{A_\alpha \mid \alpha \in J\}$ *is a family of connected subsets of* $X$ *and* $\bigcap_{\alpha\in J} A_\alpha \neq \varnothing$, *then* $\bigcup_{\alpha\in J} A_\alpha$ *is connected.*

A harder theorem is the following.

**Proposition 0.2.7.3.** *If* $A$ *is connected and* $A \subset B$, $B \subset A^-$, *then* $B$ *is connected. In particular,* $A^-$ *is connected.*

The situation for real numbers is particularly simple. An *interval* (general sense) is a subset of $R$ of one of the forms

$$(a, b) = \{x \mid a < x < b\},$$
$$(a, b] = \{x \mid a < x \leq b\},$$
$$[a, b) = \{x \mid a \leq x < b\},$$
$$[a, b] = \{x \mid a \leq x \leq b\},$$

where we allow $a = -\infty$, $b = \infty$ at open ends, with obvious meanings. The connected sets in $R$ are precisely these intervals. In particular, $R$ itself is connected.

**Problem 0.2.7.1.** Connectedness is a topological property; that is, the image of a connected set under a homeomorphism is connected.

**Proposition 0.2.7.4.** *If* $f: X \to Y$ *is continuous, and* $A \subset X$ *is connected, then* $fA$ *is connected. In particular, if* $Y = R$, *then* $fA$ *is an interval. (This is a generalization of the intermediate-value theorem for continuous functions of a real variable defined on an interval.)*

In particular, if $f: [a, b] \to Y$ is continuous, the range of $f$ is connected. Such an $f$ is called a continuous curve in $Y$ from $y_a = fa$ to $y_b = fb$.

A topological space $Y$ is *arcwise connected* if for every $y_1, y_2 \in Y$ there is a continuous curve from $y_1$ to $y_2$. It follows from Propositions 0.2.7.2 and 0.2.7.4 that an arcwise connected space is connected.

The *connected component of* $X$ *containing* $x$ is the union of all the connected subsets of $X$ which contain $x$. By Proposition 0.2.7.2 we see that the component containing $x$ is itself connected and that the components containing two different points are either identical or do not meet. Thus $X$ is split up into a disjoint union of connected sets, the *components* of $X$, each of which is *maximal-connected*, that is, is not contained in a larger connected set. It follows from Proposition 0.2.7.3 that the components of $X$ are closed. The number of components is a topological invariant.

If we substitute "arcwise connected" for "connected" above, we arrive at the notion of the *arc components* of a topological space. The subdivision into arc components is generally finer than the subdivision into components, and



the arc components are not necessarily closed. Both these facts are illustrated by the space $A$ in the following example, since $A$ is connected but has two arc components, only one of which is closed.

**Example.** The subset of $R^2$,

$$A = \{(x, \sin 1/x) \mid 0 < x \leq 1\}^-$$

is connected but not arcwise connected. That it is connected is easy by Proposition 0.2.7.3, since it is the closure of an arcwise connected set $B = \{(x, \sin 1/x) \mid 0 < x \leq 1\}$. However, the points on the boundary $\partial B = \{(0, y) \mid -1 \leq y \leq 1\}$ cannot be joined to those in $B$ by a continuous curve in $A$.

For the open subsets of $R^n$ the notions of connectedness and arcwise connectedness coincide. Indeed, in a connected open set of $R^n$ any two points can be joined by a *polygonal continuous curve*, that is, a continuous curve for which the range consists of a finite number of straight-line segments.

**Problem 0.2.7.2.** **(a)** Show that if $A$ is an open set in $R^n$ and $a \in A$, then the set of points in $A$ which can be joined to $a$ by a polygonal continuous curve is an open subset of $A$.

**(b)** Prove that $A$ is polygonally connected if $A$ is connected.

## 0.2.8. Compactness

If $A \subset X$, a *covering* of $A$ is a family $\{C_\alpha \mid \alpha \in J\}$ in $\mathscr{P}X$ such that $A \subset \bigcup_{\alpha\in J} C_\alpha$. An *open covering* is one for which the family consists of open sets. A *subcovering* of a covering $\{C_\alpha \mid \alpha \in J\}$ is a covering $\{C_\alpha \mid \alpha \in K\}$, where $K \subset J$. A *finite* covering is one for which $J$ is finite.

A subset $A$ of $X$ is *compact* if every open covering of $A$ has a finite subcovering.

**Problem 0.2.8.1.** Compactness is a topological property.

To illustrate how the definition of compactness operates, we prove that a compact subset $A$ of $R$ is bounded. Consider the open covering of $R$ consisting of open intervals of length 2, $\{(n, n + 2) \mid n \in Z\}$. Since this also is an open covering of $A$, there must be a finite subcovering. Among the $(n, n + 2)$'s which occur in the finite subcovering we must have one for which $n$ is greatest, $n = n_1$, and one for which $n$ is least, $n = n_0$. Then clearly $A \subset [n_0, n_1 + 2]$. We shall see below that $A$ is also closed (see Proposition 0.2.8.2).

Conversely, the closed bounded subsets of $R$ are compact, since this is only a restatement of the *Heine-Borel covering theorem*. This result generalizes to $R^n$—the compact subsets of $R^n$ are exactly those which are closed and bounded. A *bounded* set is one which is contained in some ball with respect to one, and hence all, of the metrics $d_p$ given previously. This notion is *not* related to the notion of a boundary of a set.

A dual formulation of compactness is given in terms of closed sets and finite intersections. A family of sets $\{C_\alpha \mid \alpha \in J\}$ has the *finite intersection property* (abbreviated FIP) if for every finite subset $K$ of $J$, $\bigcap_{\alpha\in K} C_\alpha \neq \varnothing$.

**Proposition 0.2.8.1.** *A subset $A$ of $X$ is compact iff for every family of relatively closed subsets $\{C_\alpha \mid \alpha \in J\}$ of $A$ which has the FIP, $\bigcap_{\alpha\in J} C_\alpha \neq \varnothing$.* (FIP means that no finite number of complements $A - C_\alpha$ cover $A$, whereas total intersection being nonempty means that all the complements do not cover $A$.)

**Proposition 0.2.8.2.** (a) *A compact subset of a Hausdorff space is closed.*

(b) *A closed subset in a compact space is also compact.*

*Proof.* For part (b) note that the complement of the closed subset may be added to any open covering of the closed subset so as to obtain an open covering of the containing compact space. A finite subcovering of the whole space exists and the complement of the closed subset may be deleted if it is there, leaving a finite subcovering of the closed subset.

Suppose that $A$ is a compact subset in a Hausdorff space $X$ and $A \neq A^-$, so there is an $x \in A^- - A$. For every $a \in A$ there are open sets $G_a$, $G_a^x$ such that $G_a \cap G_a^x = \varnothing$, $a \in G_a$, and $x \in G_a^x$, because $X$ is Hausdorff. Then $\{G_a \mid a \in A\}$ is an open covering of $A$, so there is a finite subcovering $\{G_a \mid a \in J\}$, where $J$ is a finite subset of $A$. But then $\bigcap_{a\in J} G_a^x$ is a neighborhood of $x$ which does not meet $\bigcup_{a\in J} G_a \supset A$, so $x$ cannot be in $A^-$, a contradiction. ∎

**Proposition 0.2.8.3.** *Let $f\colon X \to Y$ be continuous and $A$ a compact subset of $X$. Then $fA$ is compact. In particular, if $Y = R$, then $f$ has a maximum and a minimum on $A$ (since $fA$ is closed and bounded its supremum exists and is in $fA$); that is, there is an $a_M \in A$ such that for every $a \in A$, $fa \leq fa_M$, and similarly for a minimum.*

The proof is automatic.

**Proposition 0.2.8.4.** *Let $f\colon X \to Y$ be continuous, 1–1, and onto, where $X$ is compact and $Y$ is Hausdorff. Then $f$ is a homeomorphism. In particular, $X$ is Hausdorff.*

*Proof.* The problem is to show that $f^{-1}$ is continuous. We do this in the form $f$ (closed set) is closed. But for $F$ closed in $X$, $F$ is compact [Proposition 0.2.8.2(b)], $fF$ is compact (Proposition 0.2.8.3), so $fF$ is closed [Proposition 0.2.8.2(a)]. The last step uses the Hausdorff property of $Y$. ∎



## 0.2.9. Local Compactness

A topological space $X$ is called *locally compact* if each point of $X$ has a compact neighborhood. Thus a compact space is automatically locally compact.

**Problem 0.2.9.1.** (a) A closed subspace of a locally compact space is locally compact.

(b) A discrete space is locally compact.

(c) $R^n$ is locally compact.

## 0.2.10. Separability

A topological space $X$ is called *separable* if it has a countable basis of neighborhoods.

**Problem 0.2.10.1.** (a) Suppose that the metric space $X$ has a countable subset $A$ such that $A^- = X$. Show that the open balls with centers at points of $A$ and rational radii is a basis of neighborhoods for $X$, and hence that $X$ is separable.

(b) $R^n$ is separable.

**Problem 0.2.10.2.** The product of two separable spaces is separable.

## 0.2.11. Paracompactness

A family of sets $U_\alpha$ of a topological space $X$ is said to be *locally finite* if every point of $X$ has a neighborhood meeting only a finite number of the $U_\alpha$. A covering $V_\beta$ of $X$ is called a *refinement* of a covering $U_\alpha$ of $X$ if for every index $\beta$ there is at least one set $U_\alpha$ such that $V_\beta \subset U_\alpha$. A topological space $X$ is said to be *paracompact* if it is Hausdorff and if every open covering has an open refinement which is locally finite.

**Proposition 0.2.11.1.** *If $X$ is a locally compact separable Hausdorff space, then $X$ is the union of countable family of compact subsets $\{A_i\}$. This sequence of compact subsets may be taken to be increasing; that is, $A_i \subset A_{i+1}$ for every $i$.*

*Proof.* Let $\{U_i\}$, $i = 1, 2, \ldots$ be an open countable basis for $X$. We claim that those $U_i$ such that $U_i^-$ is compact are still a basis. It suffices to show that if a subset $G$ is open, then for every $x \in G$ there is a $U_i \subset G$ such that $U_i^-$ is compact and $x \in U_i$. Since $X$ is locally compact there is a compact neighborhood $V$ of $x$. Then $V^0 \cap G$ is open, so there is $U_i \subset V^0 \cap G$ such that $x \in U_i$.

But then $U_i^- \subset V^{0-} \subset V$, since $V$ is closed by Proposition 0.2.8.2, so $U_i$ is compact by Proposition 0.2.8.2.

Discarding those $U_i$ for which $U_i^-$ is not compact, we have a countable basis whose elements have compact closures, which we again denote $\{U_i\}$. We define a sequence of compact sets with the increasing property by letting $A_i = \bigcup_{j=1}^{i} U_j^-$. ∎

**Lemma.** *If a locally compact Hausdorff space $X$ is the union of a countable family of compact subsets, then it is the union of the interiors of such a family.*

*Proof.* Let $X = \bigcup_{i=1}^{\infty} A_i$, $A_i$ compact. Each $A_i$ can be covered by open neighborhoods having compact closures and hence by a finite number of such neighborhoods. The closures of these neighborhoods, a finite number for each $A_i$, comprise a countable family of compact sets whose interiors cover $X$.

**Proposition 0.2.11.2.** *If a locally compact Hausdorff space $X$ is the countable union of compact sets, then $X$ is paracompact.*

*Proof.* By the lemma we may suppose that $X = \bigcup_{i=1}^{\infty} A_i^0$, where $A_i$ is compact and $A_i \subset A_{i+1}^0$ for every $i$.

Now if $\{W_\alpha\}$ is an open covering of $X$, then for each $i$ the sets $(A_{i+2}^0 - A_{i-1}) \cap W_\alpha$ comprise an open covering of $A_{i+1} - A_i^0$. Therefore, we can choose a finite subcovering $V_{i1}, \ldots, V_{ip_i}$. Since the sets $A_{i+1} - A_i^0$ cover $X$, the $V_{ij}$, $i, j = 1, 2 \ldots$, cover $X$. Moreover, $\{V_{ij}\}$ refines the covering $\{W_\alpha\}$. Now let $x \in A_k$; then $A_{k+1}^0$ is a neighborhood of $x$ which does not intersect any $V_{ij}$ for $i > k + 1$. Thus $\{V_{ij}\}$ is locally finite. ∎

**Example.** In $R^n$ the compact sets are the closed, bounded sets. If we let $A_i$ be the closed ball with radius $i$, $i = 1, 2, \ldots$, and center a fixed $x \in R^n$, then $R^n$ is the union of the increasing sequence of the interiors of the compact sets $A_i$.

**Proposition 0.2.11.3.** *A locally compact separable Hausdorff space is paracompact.*

This follows immediately from the previous two propositions.

**Problem 0.2.11.1.** If a Hausdorff space $X$ is the countable union of subspaces homeomorphic to open subsets of $R^n$, then $X$ is paracompact.

**Problem 0.2.11.2.** The space of rational numbers, with the induced topology from the reals, is paracompact but not locally compact.

*Remark.* A continuous function has as its domain a topological space. To generalize the notion of a differentiable function on $R^n$ we shall require the concept of a differentiable manifold on which it will make sense to speak of differentiable functions.

# CHAPTER 1
# Manifolds

## 1.1. Definition of a Manifold

A manifold, roughly, is a topological space in which some neighborhood of each point admits a coordinate system, consisting of real coordinate functions on the points of the neighborhood, which determine the position of points and the topology of that neighborhood; that is, the space is locally cartesian. Moreover, the passage from one coordinate system to another is smooth in the overlapping region, so that the meaning of "differentiable" curve, function, or map is consistent when referred to either system. A detailed definition will be given below.

The mathematical models for many physical systems have manifolds as the basic objects of study, upon which further *structure* may be defined to obtain whatever system is in question. The concept generalizes and includes the special cases of the cartesian line, plane, space, and the surfaces which are studied in advanced calculus. The theory of these spaces which generalizes to manifolds includes the ideas of differentiable functions, smooth curves, tangent vectors, and vector fields. However, the notions of distance between points and straight lines (or shortest paths) are not part of the idea of a manifold but arise as consequences of additional structure, which may or may not be assumed and in any case is not unique.

A manifold has a dimension. As a model for a physical system this is the number of degrees of freedom. We limit ourselves to the study of finite-dimensional manifolds.

Some preliminary definitions will facilitate the definition of a manifold. If $X$ is a topological space, a *chart* at $p \in X$ is a function $\mu: U \to R^d$, where $U$ is an open set containing $p$ and $\mu$ is a homeomorphism onto an open subset of $R^d$. The *dimension* of the chart $\mu: U \to R^d$ is $d$. The *coordinate functions* of the chart are the real-valued functions on $U$ given by the entries of values of $\mu$;

that is, they are the functions $x^i = u^i \circ \mu: U \to R$, where $u^i: R^d \to R$ are the standard coordinates on $R^d$. [The $u^i$ are defined by $u^i(a^1, \ldots, a^d) = a^i$. The superscripts are not powers, of course, but are merely the customary tensor indexing of coordinates. If powers are needed, extra parentheses may be used, $(x)^3$ instead of $x^3$ for the cube of $x$, but usually the context will contain enough distinction to make such parentheses unnecessary.] Thus for each $q \in U$, $\mu q = (x^1 q, \ldots, x^d q)$, so we shall also write $\mu = (x^1, \ldots, x^d)$. In other terminology we call $\mu$ a *coordinate map*, $U$ the *coordinate neighborhood*, and the collection $(x^1, \ldots, x^d)$ *coordinates* or a *coordinate system at p*.

We shall restrict the symbols "$u^i$" to this usage as standard coordinates on $R^d$. For $R^2$ and $R^3$ we shall also use $x$, $y$, $z$ as coordinates as is customary, except that we shall usually treat them as functions.

A real-valued function $f: V \to R$ is $C^\infty$ (continuous to order $\infty$) if $V$ is an open set in $R^d$ and $f$ has continuous partial derivatives of all orders and types (mixed and not). A function $\varphi: V \to R^e$ is a $C^\infty$ *map* if the components $u^i \circ \varphi: V \to R$ are $C^\infty$, $i = 1, \ldots, e$.

More generally $\varphi$ is $C^k$, $k$ a nonnegative integer, if all partial derivatives up to and including those of order $k$ exist and are continuous. ($C^0$ means merely continuous.) A map $\varphi$ is *analytic* if $u^i \circ \varphi$ are real-analytic, that is, may be expressed in a neighborhood of each point by means of a convergent power series in cartesian coordinates having their origin at the point. Analytic maps are $C^\infty$ but not conversely.

**Problem 1.1.1.** **(a)** Define $f: R \to R$ by

$$fx = \begin{cases} 0 & \text{if } x \le 0, \\ e^{-1/x} & \text{if } x > 0. \end{cases}$$

Show that $f$ is $C^\infty$ and that all the derivatives of $f$ at 0 vanish; that is, $f^{(k)}0 = 0$ for every $k$.

**(b)** If $g: R \to R$ is analytic in a neighborhood of 0, then

$$gx = \sum_{k=0}^{\infty} (g^{(k)}0)x^k/k!$$

for all $x$ in a symmetric interval with center 0. Thus $f$ in part **(a)** cannot be analytic at 0.

**Example.** Letting $z = x + iy$, a complex variable, we define $u(x, y)$ by $u + iv = e^{-1/z^4}$, $u(0, 0) = 0$. Then $u$ is not $C^\infty$, and in fact not even continuous at (0, 0), but the partial derivatives of $u$ of all orders exist everywhere, including (0, 0). Thus the requirements of continuity in the definition of $C^\infty$ is not superfluous. For functions of one variable, it is of course true that differentiable functions are continuous.



Two charts $\mu: U \to R^d$ and $\tau: V \to R^e$ on a topological space $X$ are $C^\infty$-*related* if $d = e$ and either $U \cap V = \varnothing$ (the empty set) or $\mu \circ \tau^{-1}$ and $\tau \circ \mu^{-1}$ are $C^\infty$ maps. The domain of $\mu \circ \tau^{-1}$ is $\tau(U \cap V)$, an open set in $R^d$ (see Figure 1).

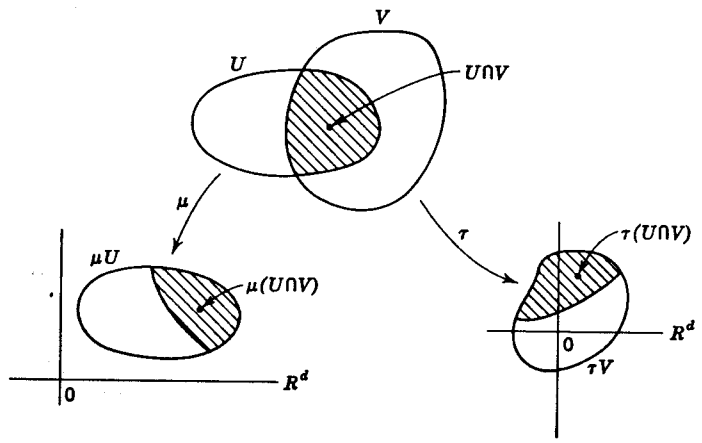


*Figure 1*

Other degrees of relatedness are defined by replacing "$C^\infty$" by "$C^k$" or "analytic." Two charts of the same dimension are always $C^0$-related because coordinate maps are continuous.

A *topological* ($C^0$) *manifold* is a separable Hausdorff space such that there is a $d$-dimensional chart at every point. The *dimension* of the manifold is the same as the dimension of the charts. Thus there is a collection of charts $\{\mu_\alpha: U_\alpha \to R^d \mid \alpha \in I\}$ such that $\{U_\alpha \mid \alpha \in I\}$ is a covering of the space. Such a collection is called an *atlas*. A $C^\infty$ *atlas* is one for which every pair of charts is $C^\infty$-related. A chart is *admissible* to a $C^\infty$ atlas if it is $C^\infty$-related to every chart in the atlas. In particular the members of a $C^\infty$ atlas are themselves admissible.

A $C^\infty$ *manifold* is a topological manifold together with all the admissible charts of some $C^\infty$ atlas. In this book the term "manifold," with no adjective, will always mean "$C^\infty$ manifold." (The reason for including all admissible charts rather than merely those which are in some given atlas is to convey the idea that no particular coordinate systems are to be preferred over any others and also to resolve the logical problem of saying just what a manifold *is*. The source of this logical difficulty is the fact that two different atlases can have the same collection of admissible charts, in which case we should like to say we have only one manifold, not two different manifolds, one for each atlas. On

the other hand, it is almost invariably the case that a manifold is specified by giving just one atlas, not the whole collection of admissible charts.)

The $C^k$ *manifolds* and *real-analytic manifolds* are defined by replacing "$C^\infty$" by "$C^k$" and "analytic," respectively, throughout the above chain of definitions. It should be clear that a $C^\infty$ manifold becomes a $C^k$ manifold simply by enlarging the collection of admissible charts to include all the $C^k$-related ones, and, similarly, a real-analytic manifold becomes a $C^\infty$ manifold. Conversely, a $C^1$ manifold becomes a real-analytic (and hence $C^\infty$) manifold, in many ways, by discarding a suitable collection of $C^1$ admissible charts so as to leave only charts which are mutually analytically related, but this result is not at all obvious, being a very difficult theorem of Whitney. That a $C^0$ manifold may fail to become a $C^1$ manifold is known, and even more difficult to prove.

*Remark.* In the definition of a coordinate system we have required that the coordinate neighborhood and the range in $R^d$ be open sets. This is contrary to popular usage, or at least more specific than the usage of curvilinear coordinates in advanced calculus. For example, spherical coordinates are used even along points of the $z$ axis where they are not even 1–1. The reasons for the restriction to open sets are that it forces a uniformity in the local structure which simplifies analysis on a manifold (there are no "edge points") and, even if local uniformity were forced in some other way, it avoids the problem of spelling out what we mean by differentiability at boundary points of the coordinate neighborhood; that is, one-sided derivatives need not be mentioned. On the other hand, in applications, boundary value problems frequently arise, the setting for which is a *manifold with boundary*. These spaces are more general than manifolds and the extra generality arises from allowing a *boundary manifold* of one dimension less. The points of the boundary manifold have a coordinate neighborhood in the boundary manifold which is attached to a coordinate neighborhood of the interior in much the same way as a face of a cube is attached to the interior. Just as the study of boundary value problems is more difficult than the study of spatial problems, the study of manifolds with boundary is more difficult than that of mere manifolds, so we shall limit ourselves to the latter.

## 1.2. Examples of Manifolds

**(a)** CARTESIAN SPACES. We define a manifold structure on $R^d$ in the most obvious way by taking as atlas the single chart $I$: $R^d \to R^d$, the identity map. The coordinate functions of this chart are thus the standard (cartesian) coordinates $u^i$. When we speak of $R^d$ as a manifold we shall intend this standard structure, unless otherwise stated.



A $C^\infty$ admissible coordinate map on $R^d$ is a 1–1 $C^\infty$ map $\mu$: $U \to R^d$, where $U$ is an open set and the jacobian determinant $|\partial x^i/\partial u^j| \neq 0$, where $x^i = u^i \circ \mu$ are the coordinate functions. Nonvanishing of the jacobian determinant is just another way of requiring the map $\mu^{-1}$ to be $C^\infty$.

If $f^i$, $i = 1, \ldots, d$, are real-valued $C^\infty$ functions on some open set of $R^d$ and at some $p \in R^d$ we have $|\partial f^i/\partial u^j| \neq 0$, then the *inverse function* theorem states that there is a neighborhood $U$ of $p$ and a neighborhood $V$ of $(f^1 p, \ldots, f^d p)$ such that the map $\mu = (f^1, \ldots, f^d)$ takes $U$ onto $V$, is 1–1, and has a $C^\infty$ inverse. This gives an effective means of obtaining admissible coordinates. In particular, polar coordinates, cylindrical coordinates, spherical coordinates, and the other customary curvilinear coordinates are admissible coordinates for $R^2$ and $R^3$ provided they are suitably restricted so as to be 1–1 and have nonzero jacobian determinant.

**Example.** Let $\mu = (x^2 + 2y^2, 3xy)$: $R^2 \to R^2$, $u = x^2 + 2y^2$, $v = 3xy$. The jacobian determinant is $(\partial u/\partial x)(\partial v/dy) - (\partial u/\partial y)(\partial v/\partial x) = 6(x^2 - 2y^2)$, which is nonzero except on the two lines $y = x/\sqrt{2}$, $y = -x/\sqrt{2}$ of *singular points*. For every point except those on these lines there is some neighborhood on which $\mu$ is an admissible coordinate map. To find what these neighborhoods might be requires a more detailed analysis. By eliminating $x$ and $y$ from $u = x^2 + 2y^2$, $v = 3xy$, $y = x/\sqrt{2}$, we obtain $v = 3u/2\sqrt{2}$, and we note that $u \geq 0$. Thus the line of singular points $y = x/\sqrt{2}$ is mapped into the half-line $v = 3u/2\sqrt{2}$, $u \geq 0$; similarly, we find that $y = -x/\sqrt{2}$ is mapped into $v = -3u/2\sqrt{2}$, $u \geq 0$. Letting $x = c$ and eliminating $y$ we get a parabola $u = c^2 + 2v^2/9c^2$ which is found to be tangent to the two half-lines just found and, except for the tangent points, lying in the open angle region $V$ between the two half-lines (see Figure 2). Each of the four connected regions of nonsingular points is mapped by $\mu$ 1–1 onto $V$, so for any nonsingular point $p$ the one of these four regions which contains $p$, or any smaller neighborhood of $p$,

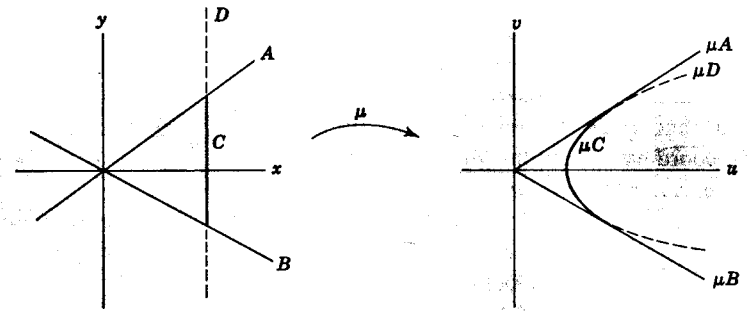


*Figure 2*

may be taken as the neighborhood $U$ asserted to exist by the inverse function theorem. No neighborhood of a singular point is mapped 1–1 by $\mu$; such neighborhoods are folded over onto themselves, and neighborhoods of (0, 0) are folded twice, so that $\mu$ is generally 4–1 in neighborhoods of (0, 0).

**Problem 1.2.1.** What restrictions on the domains and/or ranges of spherical and cylindrical coordinates can be imposed so as to make them admissible $C^\infty$ coordinates for $R^3$? Show that all points (but not all simultaneously for one system), except those where the cylindrical radius $r = 0$, may be included in domains of systems of both types.

**Problem 1.2.2.** If $u: R \to R$ is the identity map, then its cube $u^3: R \to R$ is also 1–1, continuous and has continuous inverse $u^{1/3}: R \to R$. If we take $\{u^3: R \to R\}$ as an atlas for $R$, this defines a manifold structure on $R$ with a single chart. Show that this is not the standard manifold structure since $u^3: R \to R$ is not an admissible chart in the standard structure.

**(b) OPEN SUBMANIFOLDS.** If $M$ is a manifold and $N$ is any open subset of $M$, then $N$ inherits a manifold structure by restricting the topology and coordinate maps of $M$ to $N$. We call $N$ an *open submanifold* of $M$. (A general submanifold may have a smaller dimension and will be defined in Section 1.4.) In particular, any open subset of $R^d$ is a $d$-dimensional manifold.

**Problem 1.2.3.** Show that a manifold may be considered as an open submanifold of $R^d$ iff the manifold has an atlas with only one chart.

**(c) PRODUCT MANIFOLDS.** If $M$ and $N$ are manifolds of dimensions $d$ and $e$, respectively, then $M \times N$ is given a manifold structure by taking the product topology as its topology (basic neighborhoods are products of those in $M$ and $N$) and as atlas the products of charts from atlases for $M$ and $N$. If $\mu: U \to R^d$ is a chart on $M$, and $\varphi: V \to R^e$ is a chart on $N$, their product is $(\mu, \varphi): U \times V \to R^{d+e}$, which is defined by $(\mu, \varphi)(m, n) = (\mu m, \varphi n)$. If $x^i$ are the coordinate functions of $\mu$ and $y^j$ are the coordinate functions of $\varphi$, then the coordinates of $(m, n)$ in the product chart are $(x^1 m, \ldots, x^d m, y^1 n, \ldots, y^e n)$. Thus if $p: M \times N \to M$ and $q: M \times N \to N$ are the projections, $p(m, n) = m$, $q(m, n) = n$, the coordinate functions on $U \times V$ are $z^1 = x^1 \circ p, \ldots, z^d = x^d \circ p, z^{d+1} = y^1 \circ q, \ldots, z^{d+e} = y^e \circ q$.

This product operation can obviously be iterated, and we may take different copies of the same manifold as factors. Thus even as a manifold $R^d = R \times R \times \cdots \times R$ ($d$ factors). It is easy to see that a circle $S^1$ (the curve) is a one-dimensional manifold. Picturing $S^1$ as a part of $R^2$ we see that a cylinder (the surface) is the manifold $S^1 \times R$ and may be pictured in $R^3 = R^2 \times R$.



We may consider $S^1 \times S^1$ as a union, $\{\{p\} \times S^1 \mid p \in S^1\}$, of circles $\{p\} \times S^1$, one for each $p \in S^1$. Now if we picture the first factor as being in the $xy$ plane of $R^3$, satisfying the equations $x^2 + y^2 = 1$, $z = 0$, and for each $p$ in the first factor picture $\{p\} \times S^1$ as being a smaller circle with center $p$ and diameters perpendicular to the first circle at $p$, then the union $S^1 \times S^1$ is the surface of revolution of the small circle about the $z$ axis—a torus (see Figure 3). It is not difficult to see that the topology induced from $R^3$ on the torus is the product topology.

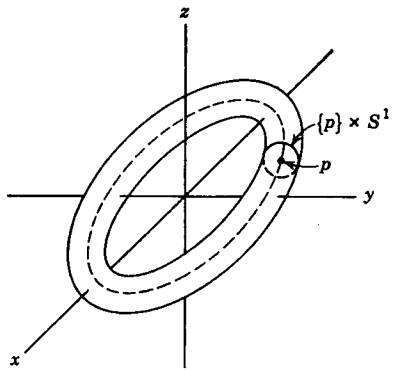


***Figure 3***

The torus is the underlying manifold which models the set of positions (the *configuration space*) of a double pendulum. We are thinking of a mechanical system consisting of two rods, the first of which is free to rotate in a plane about a fixed axis and the second of which rotates about an axis in a plane which is fixed relative to the first rod—usually, but not necessarily, the plane of the first rod. The angles these rods make with a coordinate axis in their planes may be matched with the angles $u$, $v$ which occur in the parametrization of the torus given below, giving a 1–1 correspondence between the positions of the double pendulum and the torus. The linkage must be arranged so that each rod is free to make a complete circuit about its axis, or else only a part of the torus is the model. In fact, if the second rod is blocked by the axis of the first, so that $v$ is restricted to $0 < \varepsilon < v < 2\pi$, then the model is a cylinder rather than a torus.

By adding more rods we obtain physical systems for which the model is the product of more copies of $S^1$. If the linkage is arranged so that the rod is free to move in space rather than in a plane, then some factors $S^2$ (see below) may be needed. Finally, if one end of the first rod is not fixed at all but is allowed to move freely in space (or a plane), then a factor $R^3$ (or $R^2$) may be needed.

More generally, if a physical system is a composite of two systems, each of which can assume all its positions independently of the other, then the composite system has as its manifold of positions the product of the manifolds of positions of the two component systems. This is so even though there is some dynamic linkage (e.g., gravitational or elastic) between the components.

**Problem 1.2.4.** Consider a spring with a weight attached to each end which is allowed to move freely in space except that the length $L$ of the spring is restricted to $L_1 < L < L_2$. Describe the configuration space as a triple product of $R^3$ and two other manifolds.

**(d)** LOW DIMENSIONS. A manifold of dimension 0 is a set of isolated points, that is, a set with discrete topology.

A manifold of dimension 1 which is connected is either $R$ or $S^1$. (This is not obvious, but a proof will not be given here.) The other manifolds of dimension 1 consist of disjoint unions of copies of $R$ and $S^1$. The number of copies of each must be finite or countably infinite in order that the manifold have a countable basis of neighborhoods.

**Problem 1.2.5.** Let $M = S^1 = \{(a, b) \mid a^2 + b^2 = 1, a, b \in R\}$. As topology on $M$ we take the induced topology from $R^2$. The following conditions define a unique $f: M \to R$: ($\alpha$) For every $p \in S^1$, $0 \le fp < 2\pi$; and ($\beta$) if $p = (a, b) \in S^1$, then $a = \cos fp$, $b = \sin fp$.

**(a)** Of the properties of a coordinate map listed, which does $f$ satisfy?

**(1)** A coordinate map has open domain.

**(2)** A coordinate map is 1–1.

**(3)** A coordinate map has open range.

**(4)** A coordinate map is continuous.

**(5)** The inverse of a coordinate map is continuous.

**(b)** What is the largest set to which $f$ can be restricted so as to be a coordinate map $f^-$?

**(c)** Let $g$ be defined in the same way as $f^-$ except that the range of $g$ is a different (and open) interval in $R$. For some specific choice of interval show that $\{f^-: U \to R, g: V \to R\}$ is an analytic atlas for $S^1$.

A manifold of dimension 2 may reasonably be called a surface, although there are such manifolds which cannot be placed in $R^3$. (See Problems 1.2.13 and 1.2.14.) Also, to make what are usually called surfaces in $R^3$ into manifolds, it is necessary to eliminate singular points, but these singular points cannot be handled by the usual methods of analysis from advanced calculus



anyway. (For example, the tangent plane is customarily defined only at nonsingular points.) To see that surfaces in $R^3$ are manifolds we examine how they usually arise.

**(1)** If the surface is the level surface of a $C^\infty$ function $f: R^3 \to R$, then the singular points are those at which $df = 0$, that is, at which all three partial derivatives of $f$ vanish. At a nonsingular point $p = (x_0, y_0, z_0)$, at which say $\partial f/\partial y(p) \neq 0$, there is an open neighborhood $U$ of $(x_0, z_0)$ in $R^2$ such that the equation $f(x, y, z) = c$ has a unique $C^\infty$ solution $y = g(x, z)$ with $y_0 = g(x_0, z_0)$, where $c = fp$. This follows from the implicit function theorem. Then

$$V = \{(x, g(x, z), z) \mid (x, z) \in U\}$$

is an open subset of the surface with respect to the induced topology from $R^3$, and the projection from $V$ to the $xz$ plane,

$$\mu: V \to U, \text{ given by } \mu(x, g(x, z), z) = (x, z),$$

is a coordinate map on $V$.

We can form an atlas for the nonsingular part of the surface $f^{-1}c$ from such maps. If $\partial f/\partial z(q) \neq 0$ and

$$\varphi: W \to X \text{ is given by } \varphi(x, y, h(x, y)) = (x, y)$$

in a neighborhood $W$ of $q$, where $z = h(x, y)$ is the $C^\infty$ solution of $f(x, y, z) = c$ for $z$ on $X$ such that $q = (x_1, y_1, h(x_1, y_1))$, then on the overlap of $W$ and $V$,

$$\begin{aligned} \mu \circ \varphi^{-1}(x, y) &= \mu(x, y, h(x, y)) \\ &= (x, h(x, y)), \end{aligned}$$

and similarly,

$$\varphi \circ \mu^{-1}(x, z) = (x, g(x, z)).$$

Since $h$ and $g$ are $C^\infty$ functions, and $x$ is a $C^\infty$ function of either $(x, y)$ or $(x, z)$, the maps $\mu \circ \varphi^{-1}$ and $\varphi \circ \mu^{-1}$ are $C^\infty$. This shows that the described atlas is $C^\infty$-related and that the nonsingular points on the surface form a $C^\infty$ manifold.

More specifically, if we take $f = x^2 + y^2 + z^2$ and $c = 1$, then the set of solutions to $f = 1$ is a sphere, $S^2$, and since $df = 0$ only at $(0, 0, 0)$, all points of the sphere are nonsingular.

The equation $x^2 + y^2 + z^2 = 1$ has two analytic solutions for $z$ in the open disk $U_z = \{(x, y) \mid x^2 + y^2 < 1\}$, namely, $z = \sqrt{(1 - x^2 - y^2)}$ and $z = -\sqrt{(1 - x^2 - y^2)}$. The corresponding charts on $S^2$ are $\mu_z^+: U_z^+ \to U_z$, $\mu_z^-: U_z^- \to U_z$, where $U_z^+$ is the open upper hemisphere and $U_z^-$ is the open lower hemisphere, and $\mu_z^+(x, y, z) = (x, y)$ for $(x, y, z) \in U_z^+$. The other map, $\mu_z^-$, has the same formula as $\mu_z^+$, but it is defined on $U_z^-$, where the third

coordinate is negative rather than positive. In the same way we get charts $\mu_y^+ : U_y^+ \to U_y$, $\mu_y^- : U_y^- \to U_y$, $\mu_x^+ : U_x^+ \to U_x$, $\mu_x^- : U_x^- \to U_x$ on the left, right, front, and back open hemispheres. These six charts form an analytic atlas for $S^2$, so $S^2$ is an analytic manifold.

**(2)** Surfaces are sometimes given parametrically. That is, three $C^\infty$ functions $x = f(u, v)$, $y = g(u, v)$, $z = h(u, v)$ are defined in some open region in the $uv$ plane. The singular points are those for which the two triples of partial derivatives $(\partial f/\partial u, \partial g/\partial u, \partial h/\partial u)$ and $(\partial f/\partial v, \partial g/\partial v, \partial h/\partial v)$, are proportional (including one or both having all three entries = 0).

At nonsingular points these two triples will be direction numbers for two nonparallel lines which determine the tangent plane, but at singular points the two lines unite or are indeterminate (one or both all 0) and the tangent plane may not exist.

If $(u_0, v_0)$ are the parameters of a nonsingular point, then there is an open neighborhood $U$ of $(u_0, v_0)$ in $R^2$ on which the parametrization is 1–1 onto an open set $V$ in the surface. Indeed, nonsingularity implies that one of the jacobian determinants

$$\begin{vmatrix} \partial f/\partial u & \partial f/\partial v \\ \partial g/\partial u & \partial g/\partial v \end{vmatrix}, \quad \begin{vmatrix} \partial f/\partial u & \partial f/\partial v \\ \partial h/\partial u & \partial h/\partial v \end{vmatrix}, \quad \begin{vmatrix} \partial g/\partial u & \partial g/\partial v \\ \partial h/\partial u & \partial h/\partial v \end{vmatrix},$$

is nonzero, say the first one, in which case there is an open neighborhood $U$ of $(u_0, v_0)$ such that $(u, v) \to (f(u, v), g(u, v))$ is 1–1 with a $C^\infty$ inverse on $U$, so certainly $(u, v) \to (f(u, v), g(u, v), h(u, v))$ is also 1–1 on $U$. The inverse of this map $U \to V$ is then a coordinate map $\mu: V \to U$, the parameters $u$, $v$ themselves being the coordinate functions. The projection into $R^2, \varphi: V \to W$, $\varphi(x, y, z) = (x, y)$ is also 1–1 and $C^\infty$-related to $\mu$, so it can serve as an alternative coordinate map. However, the parametrization is usually 1–1 on a larger neighborhood than $U$ on which one of the three jacobians is nonzero, so that $\mu$ may be extended to a more inclusive coordinate map and is thus usually to be preferred over $\varphi$.

The complete parametrization map $(u, v) \to (x, y, z)$ may not be 1–1 even on the nonsingular part, but may cover the same part of the surface with several different regions of the $uv$ plane. Thus there can be nonidentical coordinate transformations from the $uv$ plane into itself. These will be $C^\infty$ at nonsingular points, so the set of nonsingular points forms a two-dimensional manifold.

In a neighborhood of a nonsingular point a normal vector can be chosen to vary as a $C^\infty$ function of $(u, v)$. Letting $f$ be the directed distance to the surface, with the direction determined by the chosen normal field, we get the surface locally as the solutions of $f = 0$, where $f$ is a $C^\infty$ function. Thus nonsingular



level surfaces are locally parametrized surfaces (the coordinates are parameters) and nonsingular parametrized surfaces are locally level surfaces. Methods **(1)** and **(2)** of specifying surfaces are locally equivalent. However, they are not globally equivalent, since nonsingular level surfaces are always orientable (two-sided, having a global continuous nonzero normal field), whereas nonsingular parametrized surfaces may be nonorientable (one-sided). In fact, the gradient of $f$ is a normal field to the surface $f = c$, and it is not difficult to realize the Möbius band, which is nonorientable, as a parametrized surface.

The singularities of a parametrization may be either an unavoidable consequence of the shape of the surface (it may have a cusp or a corner at which no tangent space can be defined) or it may be an accident of the parametrization itself. An example of the latter is the standard spherical coordinate parametrization of the unit sphere,

$$\begin{aligned} x &= \sin u \cos v, \\ y &= \sin u \sin v, \\ z &= \cos u, \end{aligned}$$

for which the points $(0, 0, 1)$ and $(0, 0, -1)$ are singular points. For this parametrization the $uv$ coordinate transformations assume one of two forms:

$$\begin{aligned} u_\alpha &= u_\beta + 2p\pi, \\ v_\alpha &= v_\beta + 2r\pi, \end{aligned}$$

or

$$\begin{aligned} u_\alpha &= u_\gamma + (2q + 1)\pi, \\ v_\alpha &= -v_\gamma + 2s\pi, \end{aligned}$$

where $p$, $q$, $r$, and $s$ are integers and the three coordinate maps $\mu_\alpha = (u_\alpha, v_\alpha)$, $\mu_\beta = (u_\beta, v_\beta)$, $\mu_\gamma = (u_\gamma, v_\gamma)$ are related.

**Problem 1.2.6.** Show that $S^2$ has an atlas with two charts.

The torus in $R^3$ may be parametrized without singularities:

$$\begin{aligned} x &= (a + b \sin v) \cos u, \\ y &= (a + b \sin v) \sin u, \\ z &= b \cos v, \end{aligned}$$

where $a$ is the radius of the first circle $S^1$ in the $xy$ plane and $b$ is the radius of the small second circles having their diameters perpendicular to the first circle, as in the above description of the torus as a product $S^1 \times S^1$. The parameters $u$ and $v$ measure the angles around the first and second circles.

The possible $uv$ coordinate transformations are of the form

$$u_\alpha = u_\beta + 2p\pi,$$
$$v_\alpha = v_\beta + 2q\pi,$$

where $p$ and $q$ are integers.

**Problem 1.2.7.** Show that the parametrization of the torus given above may be inverted on three different domains so as to obtain an atlas of three charts for the torus.

(e) HYPERSURFACES. The idea of a surface may be generalized to higher dimensions. In a manner analogous to that for surfaces, we may show that the nonsingular points of a *level hypersurface*,

$$M = \{m \mid fm = c, df_m \neq 0\},$$

where $f: R^d \to R$ is a $C^\infty$ function and $c$ is a constant, form a manifold of dimension $d - 1$. Local coordinates are obtained by projections into the $(d-1)$-dimensional coordinate hyperplanes and are shown to be $C^\infty$-related by means of the implicit function theorem. Alternatively, we may consider parametric manifolds in $R^d$, with the number of parameters any number less than $d$, in particular, $d - 1$ parameters for a hypersurface. Nonsingularity is defined in terms of rank of jacobian matrices.

In particular, we define the *d-dimensional sphere* to be

$$S^d = \left\{ p \in R^{d+1} \mid \sum_{i=1}^{d+1} (u^i p)^2 = 1 \right\}.$$

In analogy with $S^2$, the projections which kill one component $u^i p$ give $2(d + 1)$ coordinate maps on the hemispheres for which a given $u^i p$ is constant in sign.

**Problem 1.2.8.** An open subset of $R^d$ is not compact (cf. Section 0.2.8). Show that a compact manifold (e.g., $S^d$, which is a closed bounded subset of $R^{d+1}$) cannot have an atlas consisting of just one chart (cf. Problem 1.2.3).

**Problem 1.2.9.** Consider a rod of length $L$ in space $R^3$. Letting the standard coordinates of one end be $u^1$, $u^2$, $u^3$ and of the other end be $u^4$, $u^5$, $u^6$, the collection of positions of this rod can be viewed as the hypersurface in $R^6$ given by the equation

$$(u^1 - u^4)^2 + (u^2 - u^5)^2 + (u^3 - u^6)^2 = L^2.$$

Show how this manifold is also the same as $R^3 \times S^2$.

The manner in which $S^1 \times S^1$ is placed in $R^3$ to get a torus may be generalized to an imbedding of $S^d \times S^e$ in $R^{d+e+1}$ as a hypersurface; that is, a small copy of $S^e$ is placed in an $R^{e+1}$ perpendicular to $S^d$ at each point of $S^d$ as it is contained in $R^{d+1} = R^{d+1} \times \{0\} \subset R^{d+e+1}$.



(f) MANIFOLDS PATCHED TOGETHER. A manifold can be given by specifying the coordinate ranges of an atlas, the images in those coordinate ranges of the overlapping parts of the coordinate domains, and the coordinate transformations for each of those overlapping domains. When a manifold is specified in this way, a rather tricky condition on the specifications is needed to give the Hausdorff property, but otherwise the topology can be defined completely by simply requiring the coordinate maps to be homeomorphisms. Two examples follow.

(1) Let there be two charts $\mu: U \to S$, $\varphi: V \to S$ such that the range of each is the rectangular strip

$$S = \{(a, b) \mid -5 < a < 5, -1 < b < 1\}.$$

The overlapping domain $U \cap V$ corresponds to the union of two end rectangles under both $\mu$ and $\varphi$,

$$\begin{aligned} T &= \mu(U \cap V) \\ &= \varphi(U \cap V) \\ &= \{(a, b) \mid -5 < a < -4 \text{ or } 4 < a < 5 \text{ and } -1 < b < 1\}. \end{aligned}$$

It remains to define $\mu \circ \varphi^{-1}$ (or $\varphi \circ \mu^{-1}$) on $T$, which we do by the formula

$$\mu \circ \varphi^{-1}(a, b) = \begin{cases} (a + 9, b) & \text{if } -5 < a < -4, \\ (a - 9, -b) & \text{if } 4 < a < 5. \end{cases}$$

The reader should paste two strips of paper together in accordance with this formula (at least mentally) if he wishes to see what this manifold represents. Since the formula components represent rigid euclidean transformations, the paper need not be torn or stretched.

To obtain the manifold more specifically as a set of elements with topology, etc., we take disjoint copies of the ranges of the coordinate maps and "identify" points in these ranges which correspond under the overlap formulas. The precise meaning of "identification" comes from the idea of an "equivalence relation," which is a modification of the idea of equality in sets to mean something other than "identically the same." The idea is not new, since it is necessary to give precise meaning to such things as $4/6 = 6/9$.

In the case at hand the coordinate ranges are not already disjoint, so we manufacture disjoint copies of their common range $S$ by tagging the elements of $S$ with a 1 or a 2:

$$S_\alpha = \{(s, \alpha) \mid s \in S\},$$

where $\alpha = 1$ or 2 and let $P = S_1 \cup S_2$. We define an equivalence relation on $P$ in accordance with a desire to identify a member of $S_1$ with a member of $S_2$

if they are connected by the coordinate transformation $F = \mu \circ \varphi^{-1}$, but otherwise to make no identifications between members of $P$: For all $s, t \in S$,

$$(s, 1)E(t, 1) \text{ iff } s = t,$$
$$(s, 2)E(t, 2) \text{ iff } s = t,$$
$$(s, 1)E(t, 2) \text{ iff } t \in T \text{ and } s = Ft,$$
$$(s, 2)E(t, 1) \text{ iff } s \in T \text{ and } t = Fs.$$

In this case the equivalence classes have only one or two elements: If $s \notin T$ and $t \in T$, then

$$[s, \alpha] = \{(s, \alpha)\}, \quad \text{where } \alpha = 1 \text{ or } 2,$$
$$[t, 1] = \{(t, 1), (F^{-1}t, 2)\},$$
$$[t, 2] = \{(Ft, 1), (t, 2)\}.$$

We unify the definitions of the coordinate maps $\mu$ and $\varphi$ by calling them $\mu_\alpha$, $\alpha = 1$ or 2. Their domains are $U_\alpha = [S_\alpha]$, the collection of all equivalence classes of members of $S_\alpha$. The maps are given simply by

$$\mu_\alpha[s, \alpha] = s.$$

Since these $\mu_\alpha$ are to be the coordinate maps on $M = P/E$, the topology on $M$ must be defined in such a way that they are homeomorphisms. Accordingly we define the open sets of $M$ to be of three types: (**a**) A subset of $U_1$ is open iff it corresponds under $\mu_1$ to an open set in $S$; (**b**) a subset of $U_2$ is open iff it corresponds under $\mu_2$ to an open set of $S$; (**c**) a subset of $M$ which is neither a subset of $U_1$ nor a subset of $U_2$ is open iff the intersections of the subset with $U_1$ and $U_2$ are both open according to (**a**) and (**b**).

**Problem 1.2.10.** Complete the demonstration that the $M$ defined above is an analytic manifold, including the proof that it is a Hausdorff space.

**Problem 1.2.11.** By extending $S$ and the formula to include points where $b = \pm 1$, a boundary manifold is attached to this $M$. What is this boundary manifold intrinsically?

**(2)** In this example there are three coordinate systems in the given atlas, all with $R^2$ as their range. Let them be $\mu_1 = (x^1, x^2)$, $\mu_2 = (y^1, y^2)$, $\mu_3 = (z^1, z^2)$. The overlapping domains correspond to as much of $R^2$ in each case as makes sense in the following formulas.

$$x^1 = 1/y^2, \quad x^2 = y^1/y^2,$$
$$y^1 = 1/z^2, \quad y^2 = z^1/z^2,$$
$$z^1 = 1/x^2, \quad z^2 = x^1/x^2.$$

We could proceed as in (**1**) to manufacture the manifold by taking three copies of $R^2$ and defining an equivalence relation corresponding to these



formulas. The manifold defined by these coordinate transformations admits a more concrete interpretation. Let $S^2$ be the unit sphere in $R^3$ with center at the origin. We define two opposite points of $S^2$ to be equivalent and $M$ to be the set of equivalence classes; thus an element of $M$ is a nonordered pair $\{p, -p\}$, where $p \in S^2$. If $p = (a, b, c)$ we have written $-p$ for $(-a, -b, -c)$. We could also consider the elements of $M$ to be the lines through the origin in $R^3$, where the line through $p$ and $-p$ corresponds to $\{p, -p\}$. The name for $M$ is the *analytic real projective plane*.

If $x, y, z$ are the cartesian coordinates on $R^3$, then the ratios, $x/y$, $x/z$, $y/z$, etc., have the same values on $p$ and $-p$, so they are well-defined functions on the subsets of $M$ on which the denominators are nonzero. We obtain the coordinate maps on $M$ from pairs of these ratios.

$$\mu_1 = (y/x, z/x) = (x^1, x^2),$$
$$\mu_2 = (z/y, x/y) = (y^1, y^2),$$
$$\mu_3 = (x/z, y/z) = (z^1, z^2).$$

The corresponding coordinate domains are those $\{p, -p\}$ for which $xp \neq 0$, $yp \neq 0$, and $zp \neq 0$, respectively.

Projective spaces of higher dimension can be defined analogously as opposite pairs on higher-dimensional spheres.

**Problem 1.2.12.** Just as the circle may be thought of as a half-closed interval $[0, 2\pi)$ with the end 0 bent around to fill the hole at $2\pi$, the torus may be considered to be the "half-closed" square $[0, 2\pi) \times [0, 2\pi)$ with the closed sides folded over to fill the opposite open side in the *same* direction (see

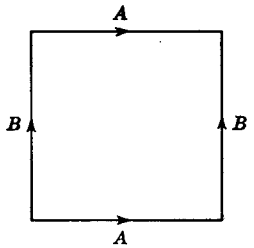


***Figure 4***

Figure 4). ("Direction" refers to direction in the plane $R^2$, not cyclic direction around the square.)

**Problem 1.2.13.** Show that the projective plane may be formed by folding the square so that the closed sides fill the open sides in the *opposite* direction (see

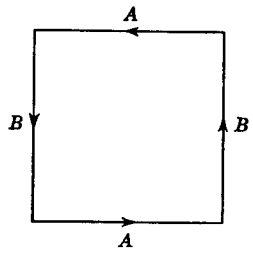


*Figure 5*

Figure 5). Another corner must be provided. To make the correspondence, stretch the square over a hemisphere with the edges laid along the bounding circle so that the corners divide the circle into four equal arcs. Since that stretching cannot be done so that the map at the corners is $C^\infty$, this identification is only intended to be topological.

**Problem 1.2.14.** By identifying one pair of opposite sides of the square in the same direction and the other in the opposite direction we get a two-dimensional manifold known as a *Klein bottle* (see Figure 6). The identification can be done

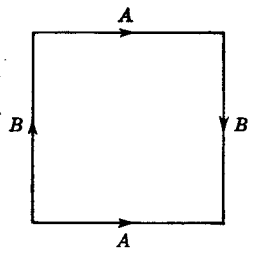


*Figure 6*

differentiably since the four corners of the square fit together nicely. Give an analytic definition of the Klein bottle in the form of (1) and (2) above, which has four charts pictured as having centers at the center of the original square, the corner of the original square, and the centers of the two sides.

The Klein bottle can be realized as a parametric surface in $R^4$ in much the same way as the torus in $R^3$. At each point of the circle of radius $a$ in the $xy$ plane there is now available a three-dimensional hyperplane in $R^4$ perpendicular to the circle. A smaller circle of radius $b < a$ can be rotated about a



diameter at half the rate of revolution about the circle of radius $a$, giving a Klein bottle. The parametrization is given analytically as follows:

$$x = (a + b \sin v) \cos u,$$
$$y = (a + b \sin v) \sin u,$$
$$z = b \cos v \cos u/2,$$
$$w = b \cos v \sin u/2.$$

Points in the $uv$ plane which are identified as indicated in Figure 6 are mapped into the same points in $R^4$ by these equations.

*Remarks.* The projective plane and the Klein bottle cannot be faithfully represented as surfaces in $R^3$ without "self-intersections." To describe what self-intersections are, we give the example of the disconnected manifold consisting of two copies of $R^2$ pictured in $R^3$ as two *intersecting* planes. The points along the line of intersection have a dual role, each being considered as two points, one in each copy of $R^2$. (This is the reason the planes form a disconnected manifold.) When such duplications are allowed, we say that the manifold is *immersed* rather than *imbedded* in $R^3$. In this sense the projective plane (*Boy's surface*) and the Klein bottle can be immersed as surfaces in $R^3$.

The three-dimensional projective space, $RP^3$, is the same, insofar as its manifold structure is concerned, as the set of all orthogonal matrices of order 3 having determinant $+1$. Since an orthogonal matrix of order 3 having determinant $+1$ is equivalent to a rotation of $R^3$ about the origin, projective 3-space is in turn the same manifold as the configuration space of an object in $R^3$ which has one fixed point but is otherwise free to rotate about any axis through the fixed point.

If an object is free to move in any way in space, we may determine its position by choosing a point in the object and specifying both where that point is placed in $R^3$ and how the object is rotated about that point relative to some initial position. Since these specifications are independent, the manifold of positions of a rigid object in space is $R^3 \times RP^3$.

## 1.3. Differentiable Maps

If $F: M \to N$, where $M$ and $N$ are $C^\infty$ manifolds, then we call $F$ a $C^\infty$ map if the coordinate expression for $F$ consists of $C^\infty$ maps on cartesian spaces. We now elaborate this statement into a complete definition, in particular making clear what is meant by "coordinate expressions."

Let $\mu_1: U \to R^d$ and $\mu_2: V \to R^e$ be $C^\infty$ charts on $M$ and $N$, so that $U$ and $V$ are open subsets of $M$ and $N$, respectively. Assume that $F: M \to N$ is a continuous map, so that $W = F^{-1}V$ is an open subset of $M$ (see Figure 7). Let $W_1 = \mu_1 W$, so that $W_1$ is an open set in $R^d$. The $\mu_1$-$\mu_2$ *coordinate expression*