Preface

The purpose of this book is to provide core material in nonlinear analysis for mathematicians, physicists, engineers, and mathematical biologists. The main goal is to provide a working knowledge of manifolds, dynamical systems, tensors and differential forms. Some applications to Hamiltonian mechanics, fluid mechanics, electromagnetism, plasma dynamics and control theory are given in Chapter 8, using both invariant and index notation.

Throughout the text supplementary topics are noted that may be downloaded from the internet from `http://www.cds.caltech.edu/~marsden`. This device enables the reader to skip various topics without disturbing the main flow of the text. Some of these provide additional background material intended for completeness, to minimize the necessity of consulting too many outside references.

Philosophy. We treat finite and infinite-dimensional manifolds simultaneously. This is partly for efficiency of exposition. Without advanced applications, using manifolds of mappings (such as applications to fluid dynamics), the study of infinite-dimensional manifolds can be hard to motivate. Chapter 8 gives an introduction to these applications. Some readers may wish to skip the infinite-dimensional case altogether. To aid in this, we have separated some of the technical points peculiar to the infinite-dimensional case into supplements, either directly in the text or on-line. Our own research interests lean toward physical applications, and the choice of topics is partly shaped by what has been useful to us over the years.

We have tried to be as sympathetic to our readers as possible by providing ample examples, exercises, and applications. When a computation in coordinates is easiest, we give it and do not hide things behind complicated invariant notation. On the other hand, index-free notation sometimes provides valuable geometric and computational insight so we have tried to simultaneously convey this flavor.

Prerequisites and Links. The prerequisites required are solid undergraduate courses in linear algebra and advanced calculus along with the usual *mathematical maturity*. At various points in the text contacts are made with other subjects. This provides a good way for students to link this material with other courses. For example, Chapter 1 links with point-set topology, parts of Chapters 2 and 7 are connected with functional analysis, Section 4.3 relates to ordinary differential equations and dynamical systems, Chapter 3 and Section 7.5 are linked to differential topology and algebraic topology, and Chapter 8 on applications is connected with applied mathematics, physics, and engineering.

Use in Courses. This book is intended to be used in courses as well as for reference. The sections are, as far as possible, lesson sized, if the supplementary material is omitted. For some sections, like 2.5, 4.2, or

7.5, two lecture hours are required if they are to be taught in detail. A standard course for mathematics graduate students could omit Chapter 1 and the supplements entirely and do Chapters 2 through 7 in one semester with the possible exception of Section 7.4. The instructor could then assign certain supplements for reading and choose among the applications of Chapter 8 according to taste.

A shorter course, or a course for advanced undergraduates, probably should omit all supplements, spend about two lectures on Chapter 1 for reviewing background point set topology, and cover Chapters 2 through 7 with the exception of Sections 4.4, 7.4, 7.5 and all the material relevant to volume elements induced by metrics, the Hodge star, and codifferential operators in Sections 6.2, 6.4, 6.5, and 7.2.

A more applications oriented course could skim Chapter 1, review without proofs the material of Chapter 2 and cover Chapters 3 to 8 omitting the supplementary material and Sections 7.4 and 7.5. For such a course the instructor should keep in mind that while Sections 8.1 and 8.2 use only elementary material, Section 8.3 relies heavily on the Hodge star and codifferential operators, and Section 8.4 consists primarily of applications of Frobenius' theorem dealt with in Section 4.4.

The notation in the book is as standard as conflicting usages in the literature allow. We have had to compromise among utility, clarity, clumsiness, and absolute precision. Some possible notations would have required too much interpretation on the part of the novice while others, while precise, would have been so dressed up in symbolic decorations that even an expert in the field would not recognize them.

History and Credits. In a subject as developed and extensive as this one, an accurate history and crediting of theorems is a monumental task, especially when so many results are folklore and reside in private notes. We have indicated some of the important credits where we know of them, but we did not undertake this task systematically. We hope our readers will inform us of these and other shortcomings of the book so that, if necessary, corrected printings will be possible. The reference list at the back of the book is confined to works actually cited in the text. These works are cited by author and year like this: deRham [1955].

Acknowledgements. During the preparation of the book, valuable advice was provided by Malcolm Adams, Morris Hirsch, Sameer Jalnapurkar, Jeff Mess, Charles Pugh, Clancy Rowley, Alan Weinstein, and graduate students in mathematics, physics and engineering at Berkeley, Santa Cruz, Caltech and Lausanne. Our other teachers and collaborators from whom we learned the material and who inspired, directly and indirectely, various portions of the text are too numerous to mention individually, so we hereby thank them all collectively. We have taken the opportunity in this edition to correct some errors kindly pointed out by our readers and to rewrite numerous sections. We thank Connie Calica, Dotty Hollinger, Anne Kao, Marnie MacElhiny and Esther Zack for their excellent typesetting of the book. We also thank Hendra Adiwidjaja, Nawoyuki Gregory Kubota, Robert Kochwalter and Wendy McKay for the typesetting and figures for this third edition.

Jerrold E. Marsden and Tudor S. Ratiu
January, 2001

0 Preface

1 Topology

The purpose of this chapter is to introduce just enough topology for later requirements. It is assumed that the reader has had a course in advanced calculus and so is acquainted with open, closed, compact, and connected sets in Euclidean space (see for example Marsden and Hoffman [1993]). If this background is weak, the reader may find the pace of this chapter too fast. If the background is under control, the chapter should serve to collect, review, and solidify concepts in a more general context. Readers already familiar with point set topology can safely skip this chapter.

A key concept in manifold theory is that of a differentiable map between manifolds. However, manifolds are also topological spaces and differentiable maps are continuous. Topology is the study of continuity in a general context, so it is appropriate to begin with it. Topology often involves interesting excursions into pathological spaces and exotic theorems that can consume lifetimes. Such excursions are deliberately minimized here. The examples will be ones most relevant to later developments, and the main thrust will be to obtain a working knowledge of continuity, connectedness, and compactness. We shall take for granted the usual logical structure of analysis, including properties of the real line and Euclidean space

1.1 Topological Spaces

The notion of a topological space is an abstraction of ideas about open sets in $\mathbb{R}^n$ that are learned in advanced calculus.

1.1.1 Definition. *A **topological space** is a set S together with a collection $\mathcal{O}$ of subsets of S called **open sets** such that*

T1. $\varnothing \in \mathcal{O}$ *and* $S \in \mathcal{O}$;

T2. *if* $U_1, U_2 \in \mathcal{O}$, *then* $U_1 \cap U_2 \in \mathcal{O}$;

T3. *the union of any collection of open sets is open.*

The Real Line and n-space. For the real line with its ***standard topology***, we choose $S = \mathbb{R}$, with $\mathcal{O}$, by definition, consisting of all sets that are unions of open intervals. Here is how to prove that this is a topology: As exceptional cases, the empty set $\varnothing \in \mathcal{O}$ and $\mathbb{R}$ itself belong to $\mathcal{O}$. Thus, **T1** holds. For **T2**, let

U_1 and $U_2 \in \mathcal{O}$; to show that $U_1 \cap U_2 \in \mathcal{O}$, we can suppose that $U_1 \cap U_2 \neq \varnothing$. If $x \in U_1 \cap U_2$, then x lies in an open interval $]a_1, b_1[\subset U_1$ and also in an interval $]a_2, b_2[\subset U_2$. We can write $]a_1, b_1[\cap]a_2, b_2[=]a, b[$ where $a = \max(a_1, a_2)$ and $b = \min(b_1, b_2)$. Thus $x \in]a, b[\subset U_1 \cap U_2$. Hence $U_1 \cap U_2$ is the union of such intervals, so is open. Finally, **T3** is clear by definition.

Similarly, $\mathbb{R}^n$ may be topologized by declaring a set to be open if it is a union of open rectangles. An argument similar to the one just given for $\mathbb{R}$ shows that this is a topology, called the ***standard topology*** on $\mathbb{R}^n$.

The Trivial and Discrete Topologies. The ***trivial topology*** on a set S consists of $\mathcal{O} = \{\varnothing, S\}$. The ***discrete topology*** on S is defined by $\mathcal{O} = \{ A \mid A \subset S \}$; that is, $\mathcal{O}$ consists of all subsets of S.

Closed Sets. Topological spaces are specified by a pair $(S, \mathcal{O})$; we shall, however, simply write S if there is no danger of confusion.

1.1.2 Definition. *Let S be a topological space. A set $A \subset S$ will be called* ***closed*** *if its complement $S \backslash A$ is open. The collection of closed sets is denoted $\mathcal{C}$.*

For example, the closed interval $[0, 1] \subset \mathbb{R}$ is closed because it is the complement of the open set $]-\infty, 0[\cup]1, \infty[$.

1.1.3 Proposition. *The closed sets in a topological space S satisfy:*

C1. *$\varnothing \in \mathcal{C}$ and $S \in \mathcal{C}$;*

C2. *if $A_1, A_2 \in \mathcal{C}$ then $A_1 \cup A_2 \in \mathcal{C}$;*

C3. *the intersection of any collection of closed sets is closed.*

Proof. Condition **C1** follows from **T1** since $\varnothing = S \backslash S$ and $S = S \backslash \varnothing$. The relations

$$S \backslash (A_1 \cup A_2) = (S \backslash A_1) \cap (S \backslash A_2) \quad \text{and} \quad S \backslash \left(\bigcap_{i \in I} B_i \right) = \bigcup_{i \in I} (S \backslash B_i)$$

for $\{B_i\}_{i \in I}$ a family of closed sets show that **C2** and **C3** are equivalent to **T2** and **T3**, respectively. ■

Closed rectangles in $\mathbb{R}^n$ are closed sets, as are closed balls, one-point sets, and spheres. Not every set is either open or closed. For example, the interval $[0, 1[$ is neither an open nor a closed set. In the discrete topology on S, any set $A \subset S$ is both open and closed, whereas in the trivial topology any $A \neq \varnothing$ or S is neither.

Closed sets can be used to introduce a topology just as well as open ones. Thus, if $\mathcal{C}$ is a collection satisfying **C1**–**C3** and $\mathcal{O}$ consists of the complements of sets in $\mathcal{C}$, then $\mathcal{O}$ satisfies **T1**–**T3**.

Neighborhoods. The idea of neighborhoods is to *localize* the topology.

1.1.4 Definition. *An* ***open neighborhood*** *of a point u in a topological space S is an open set U such that $u \in U$. Similarly, for a subset A of S, U is an open neighborhood of A if U is open and $A \subset U$. A* ***neighborhood*** *of a point (or a subset) is a set containing some open neighborhood of the point (or subset).*

Examples of neighborhoods of $x \in \mathbb{R}$ are $]x-1, x+3]$, $]x-\epsilon, x+\epsilon[$ for any $\epsilon > 0$, and $\mathbb{R}$ itself; only the last two are open neighborhoods. The set $[x, x+2[$ contains the point x but is not one of its neighborhoods. In the trivial topology on a set S, there is only one neighborhood of any point, namely S itself. In the discrete topology any subset containing p is a neighborhood of the point $p \in S$, since $\{p\}$ is an open set.

First and Second Countable Spaces.

1.1.5 Definition. *A topological space is called **first countable** if for each $u \in S$ there is a sequence $\{U_1, U_2, \dots\} = \{U_n\}$ of neighborhoods of u such that for any neighborhood U of u, there is an integer n such that $U_n \subset U$. A subset $\mathcal{B}$ of $\mathcal{O}$ is called a **basis** for the topology, if each open set is a union of elements in $\mathcal{B}$. The topology is called **second countable** if it has a countable basis.*

Most topological spaces of interest to us will be second countable. For example $\mathbb{R}^n$ is second countable since it has the countable basis formed by rectangles with rational side length and centered at points all of whose coordinates are rational numbers. Clearly every second-countable space is also first countable, but the converse is false. For example if S is an infinite non-countable set, the discrete topology is not second countable, but S is first countable, since $\{p\}$ is a neighborhood of $p \in S$. The trivial topology on S is second countable (see Exercises 1.1-9 and 1.1-10 for more interesting counter-examples).

1.1.6 Lemma (Lindelöf's Lemma)**.** *Every covering of a set A in a second countable space S by a family of open sets U_a (i.e., $\cup_a U_a \supset A$) contains a countable subcollection also covering A.*

Proof. Let $\mathcal{B} = \{B_n\}$ be a countable basis for the topology of S. For each $p \in A$ there are indices n and α such that $p \in B_n \subset U_\alpha$. Let $B' = \{\, B_n \mid \text{there exists an } \alpha \text{ such that } B_n \subset U_\alpha \,\}$. Now let $U_{\alpha(n)}$ be one of the U_α that includes the element B_n of $\mathcal{B}'$. Since $\mathcal{B}'$ is a covering of A, the countable collection $\{U_{\alpha(n)}\}$ covers A. ■

Closure, Interior, and Boundary.

1.1.7 Definition. *Let S be a topological space and $A \subset S$. The **closure** of A, denoted $\mathrm{cl}(A)$ is the intersection of all closed sets containing A. The **interior** of A, denoted $\mathrm{int}(A)$ is the union of all open sets contained in A. The **boundary** of A, denoted $\mathrm{bd}(A)$ is defined by*

$$\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(S \backslash A).$$

By **C3**, $\mathrm{cl}(A)$ is closed and by **T3**, $\mathrm{int}(A)$ is open. Note that as $\mathrm{bd}(A)$ is the intersection of closed sets, $\mathrm{bd}(A)$ is closed, and $\mathrm{bd}(A) = \mathrm{bd}(S\backslash A)$.

On $\mathbb{R}$, for example,

$$\mathrm{cl}([0,1[) = [0,1], \quad \mathrm{int}([0,1[) =]0,1[, \quad \text{and} \quad \mathrm{bd}([0,1[) = \{0,1\}.$$

The reader is assumed to be familiar with examples of this type from advanced calculus.

1.1.8 Definition. *A subset A of S is called **dense** in S if $\mathrm{cl}(A) = S$, and is called **nowhere dense** if $S \backslash \mathrm{cl}(A)$ is dense in S. The space S is called **separable** if it has a countable dense subset. A point u in S is called an **accumulation point** of the set A if each neighborhood of u contains a point of A other than itself. The set of accumulation points of A is called the **derived set** of A and is denoted by $\mathrm{der}(A)$. A point of A is said to be **isolated** if it has a neighborhood in A containing no other points of A than itself.*

The set $A = [0,1[\cup \{2\}$ in $\mathbb{R}$ has the element 2 as its only isolated point, its interior is $\mathrm{int}(A) =]0,1[$, $\mathrm{cl}(A) = [0,1] \cup \{2\}$, and $\mathrm{der}(A) = [0,1]$. In the discrete topology on a set S, $\mathrm{int}\{p\} = \mathrm{cl}\{p\} = \{p\}$, for any $p \in S$.

Since the set $\mathbb{Q}$ of rational numbers is dense in $\mathbb{R}$ and is countable, $\mathbb{R}$ is separable. Similarly $\mathbb{R}^n$ is separable. A set S with the trivial topology is separable since $\mathrm{cl}\{p\} = S$ for any $p \in S$. But $S = \mathbb{R}$ with the discrete topology is not separable since $\mathrm{cl}(A) = A$ for any $A \subset S$. Any second-countable space is separable, but the converse is false; see Exercises 1.1-9 and 1.1-10.

1.1.9 Proposition. *Let S be a topological space and $A \subset S$. Then*

(i) *$u \in \mathrm{cl}(A)$ iff for every neighborhood U of u, $U \cap A \neq \varnothing$;*

(ii) *$u \in \mathrm{int}(A)$ iff there is a neighborhood U of u such that $U \subset A$;*

(iii) $u \in \mathrm{bd}(A)$ *iff for every neighborhood* U *of* u, $U \cap A \neq \varnothing$ *and* $U \cap (S \backslash A) \neq \varnothing$.

Proof. (i) $u \notin \mathrm{cl}(A)$ iff there exists a closed set $C \supset A$ such that $u \notin C$. But this is equivalent to the existence of a neighborhood of u not intersecting A, namely $S \backslash C$. (ii) and (iii) are proved in a similar way. ■

1.1.10 Proposition. *Let* A, B *and* A_i, $i \in I$ *be subsets of* S. *Then*

(i) $A \subset B$ *implies* $\mathrm{int}(A) \subset \mathrm{int}(B)$, $\mathrm{cl}(A) \subset \mathrm{cl}(B)$, *and* $\mathrm{der}(A) \subset \mathrm{der}(B)$;

(ii) $S \backslash \mathrm{cl}(A) = \mathrm{int}(S \backslash A)$, $S \backslash \mathrm{int}(A) = \mathrm{cl}(S \backslash A)$, *and* $\mathrm{cl}(A) = A \cup \mathrm{der}(A)$;

(iii) $\mathrm{cl}(\varnothing) = \mathrm{int}(\varnothing) = \varnothing$, $\mathrm{cl}(S) = \mathrm{int}(S) = S$, $\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$, *and* $\mathrm{int}(\mathrm{int}(A)) = \mathrm{int}(A)$;

(iv) $\mathrm{cl}(A \cup B) = \mathrm{cl}(A) \cup \mathrm{cl}(B)$, $\mathrm{der}(A \cup B) = \mathrm{der}(A) \cup \mathrm{der}(B)$, *and* $\mathrm{int}(A \cup B) \supset \mathrm{int}(A) \cup \mathrm{int}(B)$;

(v) $\mathrm{cl}(A \cap B) \subset \mathrm{cl}(A) \cap \mathrm{cl}(B)$, $\mathrm{der}(A \cap B) \subset \mathrm{der}(A) \cap \mathrm{der}(B)$, *and* $\mathrm{int}(A \cap B) = \mathrm{int}(A) \cap \mathrm{int}(B)$;

(vi) $\mathrm{cl}(\bigcup_{i \in I} Ai) \supset \bigcup_{i \in I} \mathrm{cl}(A_i)$, $\mathrm{cl}(\bigcap_{i \in I} A_i) \subset \bigcap_{i \in I} \mathrm{cl}(A_i)$,
$\mathrm{int}(\bigcup_{i \in I} A_i) \supset \bigcup_{i \in I} \mathrm{int}(A_i)$, *and* $\mathrm{int}(\bigcap_{i \in I} A_i) \subset \bigcap_{i \in I} \mathrm{int}(A_i)$.

Proof. (i), (ii), and (iii) are consequences of the definition and of Proposition 1.1.9. Since for each $i \in I$, $A_i \subset \bigcup_{i \in I} A_i$, by (i) $\mathrm{cl}(A_i) \subset \mathrm{cl}(\bigcup_{i \in I} A_i)$ and hence $\bigcup_{i \in I} \mathrm{cl}(A_i) \subset \mathrm{cl}(\bigcup_{i \in I} A_i)$. Similarly, since $\bigcap_{i \in I} A_i \subset A_i \subset \mathrm{cl}(A_i)$ for each $i \in I$, it follows that $\bigcap_{i \in I}(A_i)$ is a subset of the closet set $\bigcap_{i \in I} \mathrm{cl}(A_i)$; thus by (i)

$$\mathrm{cl}\left(\bigcap_{i \in I} A_i\right) \subset \mathrm{cl}\left(\bigcap_{i \in I} \mathrm{cl}(A_i)\right) = \bigcap_{i \in I} (\mathrm{cl}(A_i)) \, .$$

The other formulas of (vi) follow from these and (ii). This also proves all the other formulas in (iv) and (v) except the ones with equalities. Since $\mathrm{cl}(A) \cup \mathrm{cl}(B)$ is closed by **C2** and $A \cup B \subset \mathrm{cl}(A) \cup \mathrm{cl}(B)$, it follows by (i) that $\mathrm{cl}(A \cup B) \subset \mathrm{cl}(A) \cup \mathrm{cl}(B)$ and hence equality by (vi). The formula $\mathrm{int}(A \cap B) = \mathrm{int}(A) \cap \mathrm{int}(B)$ is a corollary of the previous formula via (ii). ■

The inclusions in the above proposition can be strict. For example, if we let $A =]0,1[$ and $B = [1,2[$, then one finds that

$$\mathrm{cl}(A) = \mathrm{der}(A) = [0,1], \ \mathrm{cl}(B) = \mathrm{der}(B) = [1,2], \ \mathrm{int}(A) =]0,1[,$$

$$\mathrm{int}(B) =]1,2[, \ A \cup B =]0,2[, \text{ and } A \cap B = \varnothing,$$

and therefore

$$\mathrm{int}(A) \cup \mathrm{int}(B) =]0,1[\cup]1,2[\neq]0,2[= \mathrm{int}(A \cup B),$$

and

$$\mathrm{cl}(A \cap B) = \varnothing \neq \{1\} = \mathrm{cl}(A) \cap \mathrm{cl}(B).$$

Let $A_n =]-1/n, 1/n[$, $n = 1, 2, \ldots$, then

$$\bigcap_{n \geq 1} A_n = \{0\}, \quad \mathrm{int}(A_n) = A_n$$

for all n, and

$$\mathrm{int}\left(\bigcap_{n \geq 1} A_n\right) = \varnothing \neq \{0\} = \bigcap_{n \geq 1} \mathrm{int}(A_n).$$

Dualizing this via (ii) gives

$$\bigcup_{n\geq 1} \mathrm{cl}(\mathbb{R}\backslash A_n) = \mathbb{R}\backslash\{0\} \neq \mathbb{R} = \mathrm{cl}\left(\bigcup_{n\geq 1}(\mathbb{R}\backslash A_n)\right).$$

If $A \subset B$, there is, in general, no relation between the sets $\mathrm{bd}(A)$ and $\mathrm{bd}(B)$. For example, if $A = [0,1]$ and $B = [0,2]$, $A \subset B$, yet we have $\mathrm{bd}(A) = \{0,1\}$ and $\mathrm{bd}(B) = \{0,2\}$.

Convergence and Limit Points. The notion of a convergent sequence carries over from calculus in a straightforward way.

1.1.11 Definition. *Let S be a topological space and $\{u_n\}$ a sequence of points in S. The sequence is said to **converge** if there is a point $u \in S$ such that for every neighborhood U of u, there is an N such that $n \geq N$ implies $u_n \in U$. We say that u_n **converges** to u, or u is a **limit point** of $\{u_n\}$.*

For example, the sequence $\{1/n\} \in \mathbb{R}$ converges to 0. It is obvious that limit points of sequences u_n of distinct points are accumulation points of the set $\{u_n\}$. In a first countable topological space any accumulation point of a set A is a limit of a sequence of elements of A. Indeed, if $\{U_n\}$ denotes the countable collection of neighborhoods of $a \in \mathrm{der}(A)$ given by Definition 1.1.5, then choosing for each n an element $a_n \in U_n \cap A$ such that $a_n \neq a$, we see that $\{a_n\}$ converges to a. We have proved the following.

1.1.12 Proposition. *Let S be a first-countable space and $A \subset S$. Then $u \in \mathrm{cl}(A)$ iff there is a sequence of points of A that converges to u (in the topology of S).*

Separation Axioms. It should be noted that a sequence can be divergent and still have accumulation points. For example $\{2, 0, 3/2, -1/2, 4/3, -2/3, \dots\}$ does not converge but has both 1 and -1 as accumulation points. In arbitrary topological spaces, limit points of sequences are in general *not* unique. For example, in the trivial topology of S any sequence converges to all points of S. In order to avoid such situations several *separation axioms* have been introduced, of which the three most important ones will be mentioned.

1.1.13 Definition. *A topological space S is called **Hausdorff** if each two distinct points have disjoint neighborhoods (i.e., with empty intersection). The space S is called **regular** if it is Hausdorff and if each closed set and point not in this set have disjoint neighborhoods. Similarly, S is called **normal** if it is Hausdorff and if each two disjoint closed sets have disjoint neighborhoods.*

Most standard spaces that we meet in geometry and analysis are normal. The discrete topology on any set is normal, but the trivial topology is not even Hausdorff. It turns out that "Hausdorff" is the necessary and sufficient condition for uniqueness of limit points of sequences in first countable spaces (see Exercise 1.1-5). Since in Hausdorff space single points are closed (Exercise 1.1-6), we have the implications: normal $\Longrightarrow$ regular $\Longrightarrow$ Hausdorff. Counterexamples for each of the converses of these implications are given in Exercises 1.1-9 and 1.1-10.

1.1.14 Proposition. *A regular second-countable space is normal.*

Proof. Let A and B be two disjoint closed sets in S. By regularity, for every point $p \in A$ there are disjoint open neighborhoods U_p of p and U_B of B. Hence $\mathrm{cl}(U_p) \cap B = \varnothing$. Since $\{\, U_p \mid p \in A \,\}$ is an open covering of A, by the Lindelöf lemma 1.1.6, there is a countable collection $\{\, U_k \mid k = 1, 2, \dots \}$ covering A. Thus $\bigcup_{k\geq 1} U_k \supset A$ and $\mathrm{cl}(U_k) \cap B = \varnothing$.

Similarly, find a family $\{V_k\}$ such that $\bigcup_{k\geq 0} V_k \supset B$ and $\mathrm{cl}(V_k) \cup A = \varnothing$. Then the sets G_n defined inductively by $G_0 = \mathcal{U}_0$ and

$$G_{n+1} = U_{n+1} \backslash \bigcup_{k=0,1,\dots,n} \mathrm{cl}(V_k), \quad H_n = V_n \backslash \bigcup_{k=0,1,\dots,n} \mathrm{cl}(U_k)$$

are open and $G = \bigcup_{n\geq 0} G_n \supset A$, $H = \bigcup_{n\geq 0} H_n \supset B$ are also open and disjoint. ■

In the remainder of this book, Euclidean n-space $\mathbb{R}^n$ will be understood to have the standard topology unless explicitly stated to the contrary.

Some Additional Set Theory. For technical completeness we shall present the axiom of choice and an equivalent result. These notions will be used occasionally in the text, but can be skipped on a first reading.

Axiom of choice. *If $\mathfrak{S}$ is a collection of nonempty sets, then there is a function*

$$\chi : \mathfrak{S} \to \bigcup_{S \in \mathfrak{S}} S$$

such that $\chi(S) \in \mathfrak{S}$ *for every* $S \in \mathfrak{S}$.

The function χ, called a ***choice function***, chooses one element from each $S \in \mathfrak{S}$. Even though this statement seems self-evident, it has been shown to be equivalent to a number of nontrivial statements, using other axioms of set theory. To discuss them, we need a few definitions. An ***order*** on a set A is a binary relation, usually denoted by "$\leq$" satisfying the following conditions:

$a \leq a$	(***reflexivity***),
$a \leq b$ and $b \leq a$ implies $a = b$	(***antisymmetry***), and
$a \leq b$ and $b \leq c$ implies $a \leq c$	(***transitivity***).

An ordered set A is called a ***chain*** if for every $a, b \in A$, $a \neq b$ we have $a \leq b$ or $b \leq a$. The set A is said to be ***well ordered*** if it is a chain and every nonempty subset B has a first element; i.e., there exists an element $b \in B$ such that $b \leq x$ for all $x \in B$.

An ***upper bound*** $u \in A$ of a chain $C \subset A$ is an element for which $c \leq u$ for all $c \in C$. A ***maximal element*** m of an ordered set A is an element for which there is no other $a \in A$ such that $m \leq a$, $a \neq m$; in other words $x \leq m$ for all $x \in A$ that are comparable to m.

We state the following without proof.

Theorem. *Given other axioms of set theory, the following statements are equivalent:*
(i) *The axiom of choice.*

(ii) **Product Axiom.** *If* $\{A_i\}_{i \in I}$ *is a collection of nonempty sets then the product space*

$$\prod_{i \in I} A_i = \{\, (x_i) \mid x_i \in A_i \,\}$$

is nonempty.

(iii) **Zermelo's Theorem.** *Any set can be well ordered.*

(iv) **Zorn's Theorem.** *If A is an ordered set for which every chain has an upper bound (i.e., A is inductively ordered), then A has at least one maximal element.*

Exercises

◇ **1.1-1.** Let $A = \{\, (x, y, z) \in \mathbb{R}^3 \mid 0 \leq x < 1 \text{ and } y^2 + z^2 \leq 1 \,\}$. Find $\mathrm{int}(A)$.

◇ **1.1-2.** Show that any finite set in $\mathbb{R}^n$ is closed.

◇ **1.1-3.** Find the closure of the set $\{\, 1/n \mid n = 1, 2, \ldots \,\}$ in $\mathbb{R}$.

◇ **1.1-4.** Let $A \subset \mathbb{R}$. Show that $\sup(A) \in \mathrm{cl}(A)$ where $\sup(A)$ is the supremum (least upper bound) of A.

◇ **1.1-5.** Show that a first countable space is Hausdorff iff all sequences have at most one limit point.

◇ **1.1-6.** (i) Prove that in a Hausdorff space, single points are closed.

(ii) Prove that a topological space is Hausdorff iff the intersection of all closed neighborhoods of a point equals the point itself.

◇ **1.1-7.** Show that in a Hausdorff space S the following are equivalent;

(i) S is regular;

(ii) for every point $p \in S$ and any of its neighborhoods U, there exists a closed neighborhood V of p such that $V \subset U$;

(iii) for any closed set A, the intersection of all of the closed neighborhoods of A equals A.

◇ **1.1-8.** (i) Show that if $\mathcal{V}(p)$ denotes the set of all neighborhoods of a point $p \in S$, a topological space, then the following are satisfied:

V1. if $A \supset U$ and $U \in \mathcal{V}(p)$, then $A \in \mathcal{V}(p)$;

V2. every finite intersection of elements in $\mathcal{V}(p)$ is an element of $\mathcal{V}(p)$;

V3. p belongs to all elements of $\mathcal{V}(p)$;

V4. if $V \in \mathcal{V}(p)$ then there is a set $U \in \mathcal{V}(p)$, $U \subset V$ such that for all $q \in U$, $U \in \mathcal{V}(q)$.

(ii) If for each $p \in S$ there is a family $\mathcal{V}(p)$ of subsets of S satisfying **V1**–**V4**, prove that there is a unique topology $\mathcal{O}$ on S such that for each $p \in S$, the family $\mathcal{V}(p)$ is the set of neighborhoods of p in the topology $\mathcal{O}$.

Hint: Prove uniqueness first and then define elements of $\mathcal{O}$ as being subsets $A \subset S$ satisfying: for each $p \in A$, we have $A \in \mathcal{V}(p)$.

◇ **1.1-9.** Let $S = \{ p = (x,y) \in \mathbb{R}^2 \mid y \geq 0 \}$ and denote the usual ε-disk about p in the plane $\mathbb{R}^2$ by $D_\varepsilon(p) = \{ q \mid \|q - p\| < e \}$. Define

$$B_\varepsilon(p) = \begin{cases} D_\varepsilon(p) \cap S, & \text{if } p = (x,y) \text{ with } y > 0; \\ \{ (x,y) \in D_\varepsilon(p) \mid y > 0 \} \cup \{p\}, & \text{if } p = (x,0). \end{cases}$$

Prove the following:

(i) $\mathcal{V}(p) = \{ U \subset S \mid \text{there exists } B_\varepsilon(p) \subset U \}$ satisfies **V1**–**V4** of Exercise 1.1-8. Thus S becomes a topological space.

(ii) S is first countable.

(iii) S is Hausdorff.

(iv) S is separable.

Hint: The set $\{ (x,y) \in S \mid x, y \in \mathbb{Q},\ y > 0 \}$ is dense in S.

(v) S is not second countable.

Hint: Assume the contrary and get a contradiction by looking at the points $(x,0)$ of S.

(vi) S is not regular.

Hint: Try to separate the point $(x_0,0)$ from the set $\{ (x,0) \mid x \in \mathbb{R} \} \backslash \{(x_0,0)\}$.

◇ **1.1-10.** With the same notations as in the preceding exercise, except changing $B_\varepsilon(p)$ to

$$B_\varepsilon(p) = \begin{cases} D_\varepsilon(p) \cap S, & \text{if } p = (x,y) \text{ with } y > 0; \\ \{ (x,y) \in D_\varepsilon(p) \mid y > 0 \} \cup \{p\}, & \text{if } p = (x,0), \end{cases}$$

show that (i)–(v) of Exercise 1.1-9 remain valid and that

(vi) S is regular;

HINT: Use Exercise 1.1-7.

(vii) S is not normal.

HINT: Try to separate the set $\{ (x, 0) \mid x \in \mathbb{Q} \}$ from the set $\{ (x, 0) \mid x \in \mathbb{R} \backslash \mathbb{Q} \}$.

◇ **1.1-11.** Prove the following properties of the boundary operation and show by example that each inclusion cannot be replaced by equality.

Bd1. $\operatorname{bd}(A) = \operatorname{bd}(S \backslash A)$;

Bd2. $\operatorname{bd}(\operatorname{bd}(A) \subset \operatorname{bd}(A)$;

Bd3. $\operatorname{bd}(A \cup B) \subset \operatorname{bd}(A) \cup \operatorname{bd}(B) \subset \operatorname{bd}(A \cup B) \cup A \cup B$;

Bd4. $\operatorname{bd}(\operatorname{bd}(\operatorname{bd}(A))) = \operatorname{bd}(\operatorname{bd}(A))$.

Properties **Bd1**–**Bd4** may be used to characterize the topology.

◇ **1.1-12.** Let p be a polynomial in n variables $z_1, \ldots, z_n$ with complex coefficients. Show that $p^{-1}(0)$ has open dense complement.
HINT: If p vanishes on an open set of $\mathbb{C}^n$, then all its derivatives also vanish and hence all its coefficients are zero.

◇ **1.1-13.** Show that a subset $\mathcal{B}$ of $\mathcal{O}$ is a basis for the topology of S if and only if the following three conditions hold:

B1. $\varnothing \in \mathcal{B}$;

B2. $\cup_{B \in \mathcal{B}} B = S$;

B3. if $B_1, B_2 \in \mathcal{B}$, then $B_1 \cap B_2$ is a union of elements of $\mathcal{B}$.

1.2 Metric Spaces

One of the common ways to form a topological space is through the use of a distance function, also called a (topological) metric. For example, on $\mathbb{R}^n$ the standard distance

$$d(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}$$

between $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ can be used to construct the open disks and from them the topology. The abstraction of this proceeds as follows.

1.2.1 Definition. *Let M be a set. A **metric** (also called a **topological metric**) on M is a function $d : M \times M \to \mathbb{R}$ such that for all $m_1, m_2, m_3 \in M$,*

M1. $d(m_1, m_2) = 0$ *iff* $m_1 = m_2$ (definiteness);

M2. $d(m_1, m_2) = d(m_2, m_1)$ (symmetry); *and*

M3. $d(m_1, m_3) \leq d(m_1, m_2) + d(m_2, m_3)$ (triangle inequality).

*A **metric space** is the pair (M, d); if there is no danger of confusion, just write M for (M, d).*

Taking $m_1 = m_3$ in **M3** shows that $d(m_1, m_2) \geq 0$. It is proved in advanced calculus courses (and is geometrically clear) that the standard distance on $\mathbb{R}^n$ satisfies **M1**–**M3.**

The Topology on a Metric Space. The topology determined by a metric is defined as follows.

1.2.2 Definition. *For $\varepsilon > 0$ and $m \in M$, the* ***open*** *ε-**ball** (or **disk**) about m is defined by*

$$D_\varepsilon(m) = \{ m' \in M \mid d(m', m) < \varepsilon \},$$

and the ***closed*** *$\varepsilon-$**ball** is defined by*

$$B_\varepsilon(m) = \{ m' \in M \mid d(m', m) \leq \varepsilon \}.$$

*The collection of subsets of M that are unions of open disks defines the **metric topology** of the metric space (M, d).*

*Two metrics on a set are called **equivalent** if they induce the same metric topology.*

1.2.3 Proposition.

(i) *The open sets defined in the preceding definition is a topology.*

(ii) *A set $U \subset M$ is open iff for each $m \in U$ there is an $\varepsilon > 0$ such that $D_\varepsilon(m) \subset U$.*

Proof. (i) **T1** and **T3** are clearly satisfied. To prove **T2**, it suffices to show that the intersection of two disks is a union of disks, which in turn is implied by the fact that any point in the intersection of two disks sits in a smaller disk included in this intersection. To verify this, suppose that $p \in D_\varepsilon(m) \cap D_\delta(n)$ and let $0 < r < \min(\varepsilon - d(p, m), \delta - d(p, n))$. Hence $D_r(p) \subset D_\varepsilon(m) \cap D_\delta(n)$, since for any $x \in D_r(p)$,

$$d(x, m) \leq d(x, p) + d(p, m) < r + d(p, m) < \varepsilon,$$

and similarly $d(x, n) < \delta$.

(ii) By definition of the metric topology, a set V is a neighborhood of $m \in M$ iff there exists a disk $D_\varepsilon(m) \subset V$. Thus the statement in the theorem is equivalent to $U = \operatorname{int}(U)$. ■

Notice that every set M can be made into a metric space by the ***discrete metric*** defined by setting $d(m, n) = 1$ for all $m \neq n$. The metric topology of M is the discrete topology.

Pseudometric Spaces. A ***pseudometric*** on a set M is a function $d : M \times M \to \mathbb{R}$ that satisfies **M2**, **M3**, and

PM1. $d(m, m) = 0$ for all m.

Thus the distance between distinct points can be zero for a pseudometric. The pseudometric topology is defined exactly as the metric space topology. Any set M can be made into a pseudometric space by the ***trivial pseudometric***: $d(m, n) = 0$ for all $m, n \in M$; the pseudometric topology on M is the trivial topology. Note that a pseudometric space is Hausdorff iff it is a metric space.

Metric Spaces are Normal. To show that metric spaces are normal, it will be useful to have the notion of the distance from a point to a set. If M is a metric space (or pseudometric space) and $u \in M$, $A \subset M$, we define

$$d(u, A) = \inf \{ d(u, v) \mid v \in A \}$$

if $A \neq \varnothing$, and $d(u, \varnothing) = \infty$. The ***diameter*** of a set $A \subset M$ is defined by

$$\operatorname{diam}(A) = \sup \{ d(u, v) \mid u, v \in A \}.$$

A set is called ***bounded*** if its diameter is finite.

Clearly metric spaces are first-countable and Hausdorff; in fact:

1.2.4 Proposition. *Every metric space is normal.*

Proof. Let A and B be closed, disjoint subsets of M, and let

$$U = \{\, u \in M \mid d(u, A) < d(u, B) \,\} \quad \text{and} \quad V = \{\, v \in M \mid d(v, A) > d(v, B) \,\}.$$

It is verified that U and V are open, disjoint and $A \subset U$, $B \subset V$. ■

Completeness. We learn in calculus the importance of the notion of completeness of the real line. The general notion of a complete metric space is as follows.

1.2.5 Definition. *Let M be a metric space with metric d and $\{u_n\}$ a sequence in M. Then $\{u_n\}$ is a* ***Cauchy sequence*** *if for all real $\varepsilon > 0$, there is an integer N such that $n, m \geq N$ implies $d(u_n, u_m) < \varepsilon$. The space M is called* ***complete*** *if every Cauchy sequence converges.*

We claim that *a sequence $\{u_n\}$ converges to u iff for every $\varepsilon > 0$ there is an integer N such that $n \geq N$ implies $d(u_n, u) < \varepsilon$*. This follows readily from the Definitions 1.1.11 and 1.2.2.

We also claim that *a convergent sequence $\{u_n\}$ is a Cauchy sequence.* To see this, let $\varepsilon > 0$ be given. Choose N such that $n \geq N$ implies $d(u_n, u) < \varepsilon/2$. Thus, $n, m \geq N$ implies

$$d(u_n, u_m) \leq d(u_n, u) + d(u, u_m) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

by the triangle inequality. Completeness requires that, conversely, every Cauchy sequence converges. A basic fact about $\mathbb{R}^n$ is that with the standard metric, it is complete. The proof is found in any textbook on advanced calculus.

Contraction Maps. A key to many existence theorems in analysis is the following.

1.2.6 Theorem (Contraction Mapping Theorem)**.** *Let M be a complete metric space and $f : M \to M$ a mapping. Assume there is a constant k, where $0 \leq k < 1$ such that*

$$d(f(m), f(n)) \leq k\, d(m, n),$$

for all $m, n \in M$; such an f is called a ***contraction****. Then f has a unique fixed point; that is, there exists a unique $m_* \in M$ such that $f(m_*) = m_*$.*

Proof. Let m_0 be an arbitrary point of M and define recursively $m_{i+1} = f(m_i)$, $i = 0, 1, 2, \ldots$. Induction shows that

$$d(m_i, m_{i+1}) \leq k^i\, d(m_0, m_1),$$

so that for $i < j$,

$$d(m_i, m_j) \leq (k^i + \cdots + k^{j-1})\, d(m_0, m_1).$$

For $0 \leq k < 1$, $1 + k + k^2 + k^3 + \ldots$ is a convergent series, and so

$$k^i + k^{i+1} + \cdots + k^{j-1} \to 0$$

as $i, j \to \infty$. This shows that the sequence $\{m_i\}$ is Cauchy and thus by completeness of M it converges to a point m_*. Since

$$\begin{aligned} d(m_*, f(m_*)) &\leq d(m_*, m_i) + d(m_i, f(m_i)) + d(f(m_i), f(m_*)) \\ &\leq (1 + k)\, d(m_*, m_i) + k^i\, d(m_0, m_1) \end{aligned}$$

is arbitrarily small, it follows that $m_* = f(m_*)$, thus proving the existence of a fixed point of f. If m' is another fixed point of f, then

$$d(m', m_*) = d(f(m'), f(m_*)) \leq k\, d(m', m_*),$$

which, by virtue of $0 \leq k < 1$, implies $d(m', m_*) = 0$, so $m' = m_*$. Thus we have uniqueness. ■

The condition $k < 1$ is necessary, for if $M = \mathbb{R}$ and $f(x) = x + 1$, then $k = 1$, but f has no fixed point (see also Exercise 1.5-5).

At this point the true significance of the contraction mapping theorem cannot be demonstrated. When applied to the right spaces, however, it will yield the inverse function theorem (Chapter 2) and the basic existence theorem for differential equations (Chapter 4). A hint of this is given in Exercise 1.2-9.

Exercises

◇ **1.2-1.** Let $d((x_1, y_1), (x_2, y_2)) = \sup(|x_1 - x_2|, |y_1 - y_2|)$. Show that d is a metric on $\mathbb{R}^2$ and is equivalent to the standard metric.

◇ **1.2-2.** Let $f(x) = \sin(1/x)$, $x > 0$. Find the distance between the graph of f and $(0, 0)$.

◇ **1.2-3.** Show that every separable metric space is second countable.

◇ **1.2-4.** Show that every metric space has an equivalent metric in which the diameter of the space is 1. Hint: Consider the new metric $d_1(m, n) = d(m, n)/[1 + d(m, n)]$.

◇ **1.2-5.** In a metric space M, let $\mathcal{V}(m) = \{ U \subset M \mid \text{there exists } \varepsilon > 0 \text{ such that } D_\varepsilon(m) \subset U \}$. Show that $\mathcal{V}(m)$ satisfies **V1–V4** of Exercise 1.1-8. This shows how the metric topology can be defined in an alternative way starting from neighborhoods.

◇ **1.2-6.** In a metric space show that $\operatorname{cl}(A) = \{ u \in M \mid d(u, A) = 0 \}$.

Exercises 1.2-7–1.2-9 *use the notion of continuity from elementary calculus* (*see Section* 1.3).

◇ **1.2-7.** Let M denote the set of continuous functions $f : [0, 1] \to \mathbb{R}$ on the interval $[0, 1]$. Show that

$$d(f, g) = \int_0^1 |f(x) - g(x)| \, dx$$

is a metric.

◇ **1.2-8.** Let M denote the set of all continuous functions $f : [0, 1] \to \mathbb{R}$. Set

$$d(f, g) = \sup \{ |f(x) - g(x)| \mid 0 \le x \le 1 \}$$

(i) Show that d is a metric on M.

(ii) Show that $f_n \to f$ in M iff f_n converges *uniformly* to f.

(iii) By consulting theorems on uniform convergence from your advanced calculus text, show that M is a complete metric space.

◇ **1.2-9.** Let M be as in the previous exercise and define $T : M \to M$ by

$$T(f)(x) = a + \int_0^x K(x, y) \, f(y) \, dy,$$

where a is a constant and K is a continuous function of two variables. Let

$$k = \sup \left\{ \int_0^x |K(x, y)| dy \;\middle|\; 0 \le x \le 1 \right\}$$

and suppose $k < 1$. Prove the following:

(i) T is a contraction.

(ii) Deduce the existence of a unique solution of the integral equation

$$f(x) = a + \int_0^x K(x, y) \, f(y) \, dy.$$

(iii) Taking a special case of (ii), prove the "existence of e^x."

1.3 Continuity

Definition of Continuity. We learn about continuity in calculus. Its general setting in topological spaces is as follows.

1.3.1 Definition. *Let S and T be topological spaces and $\varphi : S \to T$ be a mapping. We say that φ is* ***continuous at*** *$u \in S$ if for every neighborhood V of $\varphi(u)$ there is a neighborhood U of u such that $\varphi(U) \subset V$. If, for every open set V of T, $\varphi^{-1}(V) = \{ u \in S \mid \varphi(u) \in V \}$ is open in S, φ is* ***continuous****. (Thus, φ is continuous if φ is continuous at each $u \in S$.) If the map $\varphi : S \to T$ is a* ***bijection*** *(i.e., one-to-one and onto), and both φ and φ^{-1} are continuous, φ is called a* ***homeomorphism*** *and S and T are said to be* ***homeomorphic****.*

For example, notice that any map from a discrete topological space to any topological space is continuous. Similarly, any map from an arbitrary topological space to the trivial topological space is continuous. Hence the identity map from the set S topologized with the discrete topology to S with the trivial topology is bijective and continuous, but its inverse is not continuous, hence it is not a homeomorphism.

Properties of Continuous Maps. It follows from Definition 1.3.1, by taking complements and using the set theoretic identity $S \backslash \varphi^{-1}(A) = \varphi^{-1}(T \backslash A)$, that $\varphi : S \to T$ is continuous iff the inverse image of every closed set is closed. Here are additional properties of continuous maps.

1.3.2 Proposition. *Let S, T be topological spaces and $\varphi : S \to T$. The following are equivalent:*

(i) *φ is continuous;*

(ii) *$\varphi(\mathrm{cl}(A)) \subset \mathrm{cl}(\varphi(A))$ for every $A \subset S$;*

(iii) *$\varphi^{-1}(\mathrm{int}(B)) \subset \mathrm{int}(\varphi^{-1}(B))$ for every $B \subset T$.*

Proof. If φ is continuous, then $\varphi^{-1}(\mathrm{cl}(\varphi(A)))$ is closed. But

$$A \subset \varphi^{-1}(\mathrm{cl}(\varphi(A))),$$

and hence

$$\mathrm{cl}(A) \subset \varphi^{-1}(\mathrm{cl}(\varphi(A))),$$

that is, $\varphi(\mathrm{cl}(A)) \subset \mathrm{cl}(\varphi(A))$. Conversely, let $B \subset T$ be closed and $A = \varphi^{-1}(B)$. Then

$$\mathrm{cl}(A) \subset \varphi^{-1}(B) = A,$$

so A is closed. A similar argument shows that (ii) and (iii) are equivalent. ■

This proposition combined with Proposition 1.1.12 (or a direct argument) gives the following.

1.3.3 Corollary. *Let S and T be topological spaces with S first countable and $\varphi : S \to T$. The map φ is continuous iff for every sequence $\{u_n\}$ converging to u, $\{\varphi(u_n)\}$ converges to $\varphi(u)$, for all $u \in S$.*

1.3.4 Proposition. *The composition of two continuous maps is a continuous map.*

Proof. If $\varphi_1 : S_1 \to S_2$ and $\varphi_2 : S_2 \to S_3$ are continuous maps and if U is open in S_3, then $(\varphi_2 \circ \varphi_1)^{-1}(U) = \varphi_1^{-1}(\varphi_2^{-1}(U))$ is open in S_1 since $\varphi_2^{-1}(U)$ is open in S_2 by continuity of φ_2 and hence its inverse image by φ_1 is open in S_1, by continuity of φ_1. ■

1.3.5 Corollary. *The set of all homeomorphisms of a topological space to itself forms a group under composition.*

Proof. Composition of maps is associative and has for identity element the identity mapping. Since the inverse of a homeomorphism is a homeomorphism by definition, and since for any two homeomorphisms φ_1, φ_2 of S to itself, the maps $\varphi_1 \circ \varphi_2$ and $(\varphi_1 \circ \varphi_2)^{-1} = \varphi_2^{-1} \circ \varphi_1^{-1}$ are continuous by Proposition 1.3.4, the corollary follows. ■

1.3.6 Proposition. *The space of continuous maps $f : S \to \mathbb{R}$ forms an **algebra** under pointwise addition and multiplication. That is, if f and g are continuous, then so are $f + g$ and fg.*

Proof. Let $s_0 \in$ S be fixed and $\varepsilon > 0$. By continuity of f and g at s_0, there exists an open set U in S such that

$$|f(s) - f(s_0)| < \frac{\varepsilon}{2}, \quad \text{and} \quad |g(s) - g(s_0)| < \frac{\varepsilon}{2}$$

for all $s \in U$. Then

$$|(f+g)(s) - (f+g)(s_0)| \le |f(s) - f(s_0)| + |g(s) - g(s_0)| < \varepsilon.$$

Similarly, for $\varepsilon > 0$, choose a neighborhood V of s_0 such that

$$|f(s) - f(s_0)| < \delta, \quad |g(s) - g(s_0)| < \delta$$

for all $s \in V$, where δ is any positive number satisfying

$$(\delta + |f(s_0)|)\delta + |g(s_0)|\delta < \varepsilon.$$

Then

$$\begin{aligned} |(fg)(s) - (fg)(s_0)| &\le |(f(s)| \, |g(s) - g(s_0)| + |f(s) - f(s_0)| \, |g(s_0)| \\ &< (\delta + |f(s_0)|)\delta + \delta|g(s_0)| < \varepsilon. \end{aligned}$$

Therefore, $f + g$ and fg are continuous at s_0. ■

Open and Closed Maps. Continuity is defined by requiring that *inverse images* of open (closed) sets are open (closed). In many situations it is important to ask whether the *image* of an open (closed) set is open (closed).

1.3.7 Definition. *A map $\varphi : S \to T$, where S and T are topological spaces, is called **open** (resp., **closed**) if the image of every open (resp., closed) set in S is open (resp., closed) in T.*

Thus, a homeomorphism is a bijective continuous open (closed) map.

An example of an *open map that is not closed* is

$$\varphi :]0,1[\to \mathbb{R}, \quad x \mapsto x,$$

the inclusion map. An example of a *closed map that is not open* is

$$\varphi : \mathbb{R} \to \mathbb{R}, \quad \text{defined by } x \mapsto x^2$$

which maps $]-1,1[$ to $[0,1[$. An example of a map that is neither open nor closed is the map

$$\varphi :]-1,1[\to \mathbb{R}, \quad \text{defined by } x \mapsto x^2.$$

Finally, note that the identity map of a set S topologized with the trivial and discrete topologies on the domain and range, respectively, is not continuous but is both open and closed.

Continuous Maps between Metric Spaces. For these spaces, continuity may be expressed in terms of ε's and δ's familiar from calculus.

1.3.8 Proposition. *Let (M_1, d_1) and (M_2, d_2) be metric spaces, and $\varphi : M_1 \to M_2$ a given mapping. Then φ is continuous at $u_1 \in M_1$ iff for every $\varepsilon > 0$ there is a $\delta > 0$ such that $d_1(u_1, u_1') < \delta$ implies $d_2(\varphi(u_1), \varphi(u_1')) < \varepsilon$.*

Proof. Let φ be continuous at u_1 and consider $D^2_\varepsilon(\varphi(u_1))$, the ε-disk at $\varphi(u_1))$ in M_2. Then there is a δ-disk $D^1_\delta(u_1)$ in M_1 such that

$$\varphi(D^1_\delta(u_1)) \subset D^2_\varepsilon(\varphi(u_1))$$

by Definition 1.3.1; that is, $(u_1, u_1') < \delta$ implies

$$d_2(\varphi(u_1), \varphi(u_1')) < \varepsilon.$$

Conversely, assume this latter condition is satisfied and let V be a neighborhood of $\varphi(u_1)$ in M_2. Choosing an ε-disk $D^2_\varepsilon(\varphi(u_1)) \subset V$ there exists $\delta > 0$ such that $\varphi(D^1_\delta(u_1)) \subset D^2_\varepsilon(\varphi(u_1))$ by the foregoing argument. Thus φ is continuous at u_1. ■

Uniform Continuity and Convergence. In a metric space we also have the notions of uniform continuity and uniform convergence.

1.3.9 Definition. (i) *Let (M_1, d_1) and (M_2, d_2) be metric spaces and $\varphi : M_1 \to M_2$. We say φ is **uniformly continuous** if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $d_1(u, v) < \delta$ implies $d_2(\varphi(u), \varphi(v)) < \varepsilon$.*

(ii) *Let S be a set, M a metric space, $\varphi_n : S \to M$, $n = 1, 2, \ldots$, and $\varphi : S \to M$ be given mappings. We say $\varphi_n \to \varphi$ **uniformly** if for every $\varepsilon > 0$ there is an N such that $d(\varphi_n(u), \varphi(u)) < \varepsilon$ for all $n \geq N$ and all $u \in S$.*

For example, a map satisfying $d(\varphi(u), \varphi(v)) \leq K d(u, v)$ for a constant K is uniformly continuous. Uniform continuity and uniform convergence ideas come up in the construction of a metric on the space of continuous maps. This is considered next.

1.3.10 Proposition. *Let M be a topological space and (N, d) a complete metric space. Then the collection $C(M, N)$ of all bounded continuous maps $\varphi : M \to N$ forms a complete metric space with the metric*

$$d^0(\varphi, \psi) = \sup\{\, d(\varphi(u), \psi(u)) \mid u \in M \,\}.$$

Proof. It is readily verified that d^0 is a metric. Convergence of a sequence $f_n \in C(M, N)$ to $f \in C(M, N)$ in the metric d^0 is the same as ***uniform convergence***, as is readily checked. (See Exercise 1.2-8.) Now, if $\{f_n\}$ is a Cauchy sequence in $C(M, N)$, then $\{f_n(x)\}$ is Cauchy for each $x \in M$ since $d(f_n(x), f_m(x)) \leq d^0(f_n, f_m)$. Thus f_n converges pointwise, defining a function $f(x)$. We must show that $f_n \to f$ uniformly and that f is continuous. First, given $\varepsilon > 0$, choose N such that $d^0(f_n, f_m) < \varepsilon/2$ if $n, m \geq N$. Second, for any $x \in M$, pick $N_x \geq N$ so that

$$d(f_m(x), f(x)) < \frac{\varepsilon}{2}$$

if $m \geq N_x$. Thus with $n \geq N$ and $m \geq N_x$,

$$d(f_n(x), f(x)) \leq d(f_n(x), f_m(x)) + d(f_m(x), f(x)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

so $f_n \to f$ uniformly. The reader can similarly verify that f is continuous (see Exercise 1.3-6; look in any advanced calculus text such as Marsden and Hoffman [1993] for the case of $\mathbb{R}^n$ if you get stuck). ■

Exercises

◇ **1.3-1.** Show that a map $\varphi : S \to T$ between the topological spaces S and T is continuous iff for every set $B \subset T$, $\mathrm{cl}(\varphi^{-1}(B)) \subset \varphi^{-1}(\mathrm{cl}(B))$. Show that continuity of φ does *not* imply any inclusion relations between $\varphi(\mathrm{int}(A))$ and $\mathrm{int}(\varphi(A))$.

◇ **1.3-2.** Show that a map $\varphi : S \to T$ is continuous and closed if for every subset $U \subset S$, $\varphi(\mathrm{cl}(U)) = \mathrm{cl}(\varphi(U))$.

◇ **1.3-3.** Show that compositions of open (closed) mappings are also open (closed) mappings.

◇ **1.3-4.** Show that $\varphi :]0, \infty[\to]0, \infty[$ defined by $\varphi(x) = 1/x$ is continuous but not uniformly continuous.

◇ **1.3-5.** Show that if d is a pseudometric on M, then the map $d(\cdot, A) : M \to \mathbb{R}$, for $A \subset M$ a fixed subset, is continuous.

◇ **1.3-6.** If S is a topological space, T a metric space, and $\varphi_n : S \to T$ a sequence of continuous functions uniformly convergent to a mapping $\varphi : S \to T$, then φ is continuous.

1.4 Subspaces, Products, and Quotients

This section concerns the construction of new topological spaces from old ones.

Subset Topology. The first basic operation of this type we consider is the formation of subset topologies.

1.4.1 Definition. *If A is a subset of a topological space S with topology $\mathcal{O}$, the **relative topology** on A is defined by $\mathcal{O}_A = \{U \cap A \mid U \in \mathcal{O}\}$.*

In other words, the open subsets in A are declared to be those subsets that are intersections of open sets in S with A. The following identities show that $\mathcal{O}_A$ *is indeed a topology*:

(i) $\varnothing \cap A = \varnothing$, $S \cap A = A$;

(ii) $(U_1 \cap A) \cap (U_2 \cap A) = (U_1 \cap U_2) \cap A$; and

(iii) $\bigcup_\alpha (U_\alpha \cap A) = (\bigcup_\alpha U_\alpha) \cap A$.

Example. The topology on the $n-1$-dimensional sphere $S^{n-1} = \{ x \in \mathbb{R}^n \mid d(x, 0) = 1 \}$ is the relative topology induced from $\mathbb{R}^n$; that is, a neighborhood of a point $x \in S^{n-1}$ is a subset of S^{n-1} containing the set $D_\varepsilon(x) \cap S^{n-1}$ for some $\varepsilon > 0$. Note that an open (closed) set in the relative topology of A is in general *not* open (closed) in S. For example, $D_\varepsilon(x) \cap S^{n-1}$ is open in S^{n-1} but it is neither open nor closed in $\mathbb{R}^n$. However, if A is open (closed) in S, then any open (closed) set in the relative topology is also open (closed) in S.

If $\varphi : S \to T$ is a continuous mapping, then the restriction $\varphi|A : A \to T$ is also continuous in the relative topology. The converse is false. For example, the mapping $\varphi : \mathbb{R} \to \mathbb{R}$ defined by $\varphi(x) = 0$ if $x \in \mathbb{Q}$ and $\varphi(x) = 1$ if $x \in \mathbb{R} \backslash \mathbb{Q}$ is discontinuous, but $\varphi|\mathbb{Q} : \mathbb{Q} \to \mathbb{R}$ is a constant mapping and is thus continuous.

Products. We can build up larger spaces by taking products of given ones.

1.4.2 Definition. *Let S and T be topological spaces and*

$$S \times T = \{ (u, v) \mid u \in S \text{ and } v \in T \}.$$

*The **product topology** on $S \times T$ consists of all subsets that are unions of sets which have the form $U \times V$, where U is open in S and V is open in T. Thus, these **open rectangles** form a basis for the topology.*

Products of more than two factors can be considered in a similar way; it is straightforward to verify that the map $((u,v),w) \mapsto (u,(v,w))$ is a homeomorphism of $(S \times T) \times Z$ onto $S \times (T \times Z)$. Similarly, one sees that $S \times T$ is homeomorphic to $T \times S$. Thus one can take products of any number of topological spaces and the factors can be grouped in any order; we simply write $S_1 \times \cdots \times S_n$ for such a finite product. For example, $\mathbb{R}^n$ has the product topology of $\mathbb{R} \times \cdots \times \mathbb{R}$ (n times). Indeed, using the ***maximum metric***

$$d(\mathbf{x}, \mathbf{y}) = \max_{1 \leq i \leq n} (|x^i - y^i|),$$

which is equivalent to the standard one, we see that the ε-disk at $\mathbf{x}$ coincides with the set

$$]x^1 - \varepsilon, x^1 + \varepsilon[\times \cdots \times]x^n - \varepsilon, x^n + \varepsilon[.$$

For generalizations to infinite products see Exercise 1.4-11, and to metric spaces see Exercise 1.4-14.

1.4.3 Proposition. *Let S and T be topological spaces and denote by $p_1 : S \times T \to S$ and $p_2 : S \times T \to T$ the canonical projections: $p_1(s,t) = s$ and $p_2(s,t) = t$. Then*

(i) *p_1 and p_2 are open mappings; and*

(ii) *a mapping $\varphi : X \to S \times T$, where X is a topological space, is continuous iff both the maps $p_1 \circ \varphi : X \to S$ and $p_2 \circ \varphi : X \to T$ are continuous.*

Proof. (i) follows directly from the definitions.

(ii) φ is continuous iff $\varphi^{-1}(U \times V)$ is open in X, for $U \subset S$ and $V \subset T$ open sets. Since

$$\begin{aligned} \varphi^{-1}(U \times V) &= \varphi^{-1}(U \times T) \cap \varphi^{-1}(S \times V) \\ &= (p_1 \circ \varphi)^{-1}(U) \cap (p_2 \circ \varphi)^{-1}(V), \end{aligned}$$

the assertion follows. ■

In general, the maps p_i, $i = 1, 2$, are *not* closed. For example, if $S = T = \mathbb{R}$ the set $A = \{ (x,y) \mid xy = 1,\ x > 0 \}$ is closed in $S \times T = \mathbb{R}^2$, but $p_1(A) =]0, \infty[$ which is not closed in S.

1.4.4 Proposition. *A topological space S is Hausdorff iff the **diagonal** which is defined by $\Delta_S = \{ (s,s) \mid s \in S \} \subset S \times S$ is a closed subspace of $S \times S$, with the product topology.*

Proof. It is enough to remark that S is Hausdorff iff for every two distinct points $p, q \in S$ there exist neighborhoods U_p, U_q of p, q, respectively, such that $(U_p \times U_q) \cap \Delta_S = \varnothing$. ■

Quotient Spaces. In a number of places later in the book we are going to form new topological spaces by collapsing old ones. We define this process now and give some examples.

1.4.5 Definition. *Let S be a set. An **equivalence relation** $\sim$ on S is a binary relation such that for all $u, v, w \in S$,*

(i) *$u \sim u$ (**reflexivity**);*

(ii) *$u \sim v$ iff $v \sim u$ (**symmetry**); and*

(iii) *$u \sim v$ and $v \sim w$ implies $u \sim w$ (**transitivity**).*

*The **equivalence class** containing u, denoted $[u]$, is defined by*

$$[u] = \{ v \in S \mid u \sim v \}.$$

*The set of equivalence classes is denoted $S/\sim$, and the mapping $\pi : S \to S/\sim$ defined by $u \mapsto [u]$ is called the **canonical projection**.*

Note that S is the disjoint union of its equivalence classes. The collection of subsets U of $S/\sim$ such that $\pi^{-1}(U)$ is open in S is a topology because

(i) $\pi^{-1}(\varnothing) = \varnothing$, $\pi^{-1}(S/\sim) = S$;

(ii) $\pi^{-1}(U_1 \cap U_2) = \pi^{-1}(U_1) \cap \pi^{-1}(U_2)$; and

(iii) $\pi^{-1}(\bigcup_\alpha U_\alpha) = \bigcup_\alpha \pi^{-1}(U_\alpha)$.

1.4.6 Definition. *Let S be a topological space and $\sim$ an equivalence relation on S. Then the collection of sets $\{ U \subset S/\sim \mid \pi^{-1}(U)$ is open in $S\}$ is called the **quotient topology** on $S/\sim$.*

1.4.7 Examples.

A. The Torus. Consider $\mathbb{R}^2$ and the relation $\sim$ defined by

$$(a_1, a_2) \sim (b_1, b_2) \quad \text{if } a_1 - b_1 \in \mathbb{Z} \text{ and } a_2 - b_2 \in \mathbb{Z}$$

($\mathbb{Z}$ denotes the integers). Then $\mathbb{T}^2 = \mathbb{R}^2/\sim$ is called the 2-***torus***. In addition to the quotient topology, it inherits a group structure by setting $[(a_1, a_2)] + [(b_1, b_2)] = [(a_1, a_2) + (b_1, b_2)]$. The n-dimensional torus $\mathbb{T}^n$ is defined in a similar manner.

The torus $\mathbb{T}^2$ may be obtained in two other ways. First, let $\square$ be the unit square in $\mathbb{R}^2$ with the subspace topology. Define $\sim$ by $\mathbf{x} \sim \mathbf{y}$ iff any of the following hold:

(i) $\mathbf{x} = \mathbf{y}$;

(ii) $x_1 = y_1$, $x_2 = 0$, $y_2 = 1$;

(iii) $x_1 = y_1$, $x_2 = 1$, $y_2 = 0$;

(iv) $x_2 = y_2$, $x_1 = 0$, $y_1 = 1$; or

(v) $x_2 = y_2$, $x_1 = 1$, $y_1 = 0$,

as indicated in Figure 1.4.1. Then $\mathbb{T}^2 = \square/\sim$. Second, define $\mathbb{T}^2 = S^1 \times S^1$, also shown in Figure 1.4.1.

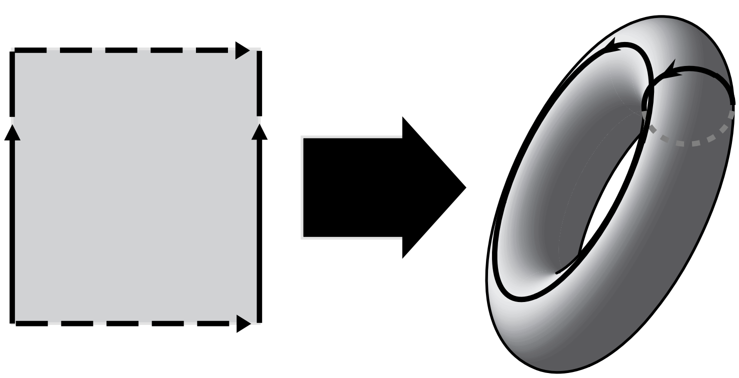

Figure 1.4.1. A torus

B. The Klein bottle. The Klein bottle is obtained by reversing one of the orientations on $\square$, as indicated in Figure 1.4.2. Then $\mathbb{K} = \square/\sim$ (the equivalence relation indicated) is the ***Klein bottle***. Although it is realizable as a subset of $\mathbb{R}^4$, it is convenient to picture it in $\mathbb{R}^3$ as shown. In a sense we will make precise in Chapter 6, one can show that $\mathbb{K}$ is not "orientable." Also note that $\mathbb{K}$ does not inherit a group structure from $\mathbb{R}^2$, as did $\mathbb{T}^2$.

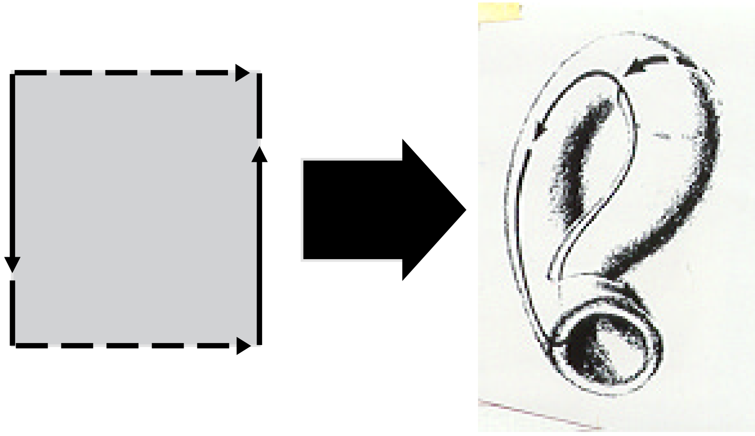

FIGURE 1.4.2. A Klein bottle

C. Projective Space. On $\mathbb{R}^n \backslash \{0\}$ define $\mathbf{x} \sim \mathbf{y}$ if there is a nonzero real constant λ such that $\mathbf{x} = \lambda \mathbf{y}$. Then $(\mathbb{R}^n \backslash \{0\})/\sim$ is called ***real projective*** $(n-1)$-***space*** and is denoted by $\mathbb{RP}^{n-1}$. Alternatively, $\mathbb{RP}^{n-1}$ can be defined as S^{n-1} (the unit sphere in $\mathbb{R}^n$) with antipodal points $\mathbf{x}$ and $-\mathbf{x}$ identified. (It is easy to see that this gives a homeomorphic space.) One defines ***complex projective space*** $\mathbb{CP}^{n-1}$ in an analogous way where now λ is complex. ♦

Continuity of Maps on Quotients. The following is a convenient way to tell when a map on a quotient space is continuous.

1.4.8 Proposition. *Let $\sim$ be an equivalence relation on the topological space S and $\pi : S \to S/\sim$ the canonical projection. A map $\varphi : S/\sim \to T$, where T is another topological space, is continuous iff $\varphi \circ \pi : S \to T$ is continuous.*

Proof. φ is continuous iff for every open set $V \subset T$, $\varphi^{-1}(V)$ is open in $S/\sim$, that is, iff the set $(\varphi \circ \pi)^{-1}(V)$ is open in S. ■

1.4.9 Definition. *The set $\Gamma = \{ (s, s') \mid s \sim s' \} \subset S \times S$ is called the **graph** of the equivalence relation $\sim$. The equivalence relation is called **open** (**closed**) if the canonical projection $\pi : S \to S/\sim$ is open (closed).*

We note that $\sim$ is open (closed) iff for any open (closed) subset A of S the set $\pi^{-1}(\pi(A))$ is open (closed). As in Proposition 1.4.8, for an open (closed) equivalence relation $\sim$ on S, a map $\varphi : S/\sim \to T$ is open (closed) iff $\varphi \circ \pi : S \to T$ is open (closed). In particular, if $\sim$ is an open (closed) equivalence relation on S and $\varphi : S/\sim \to T$ is a bijective continuous map, then φ is a homeomorphism iff $\varphi \circ \pi$ is open (closed).

1.4.10 Proposition. *If $S/\sim$ is Hausdorff, then the graph Γ of $\sim$ is closed in $S \times S$. If the equivalence relation $\sim$ is open and Γ is closed (as a subset of $S \times S$), then $S/\sim$ is Hausdorff.*

Proof. If $S/\sim$ is Hausdorff, then $\Delta_{S/\sim}$ is closed by Proposition 1.4.4 and hence $\Gamma = (\pi \times \pi)^{-1}(\Delta_{S/\sim})$ is closed on $S \times S$, where

$$\pi \times \pi : S \times S \to (S/\sim) \times (S/\sim)$$

is given by $(\pi \times \pi)(x, y) = ([x], [y])$.

Assume that Γ is closed and $\sim$ is open. If $S/\sim$ is not Hausdorff then there are distinct points $[x], [y] \in S/\sim$ such that for any pair of neighborhoods U_x and U_y of $[x]$ and $[y]$, respectively, we have $U_x \cap U_y \neq \varnothing$. Let V_x and V_y be any open neighborhoods of x and y, respectively. Since $\sim$ is an open equivalence relation,

$$\pi(V_x) = U_x \quad \text{and} \quad \pi(V_y) = U_y$$

are open neighborhoods of $[x]$ and $[y]$ in $S/\sim$. Since $U_x \cap U_y \neq \varnothing$, there exist $x' \in V_x$ and $y' \in V_y$ such that $[x'] = [y']$; that is, $(x', y') \in \Gamma$. Thus $(x, y) \in \mathrm{cl}(\Gamma)$ by Proposition 1.1.9(i). As Γ is closed, $(x, y) \in \Gamma$, that is, $[x] = [y]$, a contradiction. ■

Exercises

◇ **1.4-1.** Show that the sequence $x_n = 1/n$ in the topological space $]0,1]$ (with the relative topology from $\mathbb{R}$) does not converge.

◇ **1.4-2.** If $f : S \to T$ is continuous and T is Hausdorff, show that the graph of f, $\Gamma_f = \{ (s, f(s)) \mid s \in S \}$ is closed in $S \times T$.

◇ **1.4-3.** Let X and Y be topological spaces with Y Hausdorff. Show that for any continuous maps $f, g : X \to Y$, the set $\{ x \in X \mid f(x) = g(x) \}$ is closed in X.
Hint: Consider the mapping $x \mapsto (f(x), g(x))$ and use Proposition 1.4.4. Thus, if $f(x) = g(x)$ at all points of a dense subset of X, then $f = g$.

◇ **1.4-4.** Define a ***topological manifold*** to be a space locally homeomorphic to $\mathbb{R}^n$. Find a topological manifold that is not Hausdorff.
Hint: Consider $\mathbb{R}$ with "extra origins."

◇ **1.4-5.** Show that a mapping $\varphi : S \to T$ is continuous iff the mapping $s \mapsto (s, f(s))$ of S to the graph $\Gamma_f = \{ (s, f(s)) \mid s \in S \} \subset S \times S$ is a homeomorphism of S with Γ_f (give Γ_f the subspace topology induced from the product topology of $S \times T$).

◇ **1.4-6.** Show that every subspace of a Hausdorff (resp., regular) space is Hausdorff (resp., regular). Conversely, if each point of a topological space has a closed neighborhood that is Hausdorff (resp., regular) in the subspace topology, then the topological space is Hausdoff (resp., regular).
Hint: use Exercises 1.1-6 and 1.1-7.

◇ **1.4-7.** Show that a product of topological spaces is Hausdorff iff each factor is Hausdorff.

◇ **1.4-8.** Let S, T be topological spaces and $\sim, \approx$ be equivalence relations on S and T, respectively. Let $\varphi : S \to T$ be continuous such that $s_1 \sim s_2$ implies $\varphi(s_1) \approx \varphi(s_2)$. Show that the induced mapping $\hat{\varphi} : S/\sim \to T/\approx$ is continuous.

◇ **1.4-9.** Let S be a Hausdorff space and assume there is a continuous map $\sigma : S/\sim \to S$ such that $\pi \circ \sigma = i_{S/\sim}$, the identity. Show that $S/\sim$ is Hausdorff and $\sigma(S/\sim)$ is closed in S.

◇ **1.4-10.** Let M and N be metric spaces, N complete, and $\varphi : A \to N$ be uniformly continuous (A with the induced metric topology). Show that φ has a unique extension $\varphi : \mathrm{cl}(A) \to N$ that is uniformly continuous.

◇ **1.4-11.** Let S be a set, T_α a family of topological spaces, and $\varphi_\alpha : S \to T_\alpha$ a family of mappings. Let $\mathcal{B}$ be the collection of finite intersections of sets of the form $\varphi_\alpha^{-1}(U_\alpha)$ for U_α open in T_α. The ***initial topology*** on S given by the family $\varphi_\alpha : S \to T_\alpha$ has as basis the collection $\mathcal{B}$. Show that this topology is characterized by the fact that any mapping $\varphi : R \to S$ from a topological space R is continuous iff all $\varphi_\alpha \circ \varphi : R \to T_\alpha$ are continuous. Show that the subspace and product topologies are initial topologies. Define the product of an arbitrary infinite family of topological spaces and describe the topology.

◇ **1.4-12.** Let T be a set and $\varphi_\alpha : S_\alpha \to T$ a family of mappings, S_α topological spaces with topologies $\mathcal{O}_\alpha$. Let $\mathcal{O} = \{ U \subset T \mid \varphi_\alpha^{-1}(U) \in \mathcal{O} \text{ for each } \alpha \}$. Show that $\mathcal{O}$ is a topology on T, called the ***final topology*** on T given by the family $\varphi_\alpha : S_\alpha \to T$. Show that this topology is characterized by the fact that any mapping $\varphi : T \to R$ is continuous iff $\varphi \circ \varphi_\alpha : S_\alpha \to R$ are all continuous. Show that the quotient topology is a final topology.

◇ **1.4-13.** Show that in a complete metric space a subspace is closed iff it is complete.

◇ **1.4-14.** Show that a product of two metric spaces is also a metric space by finding at least three equivalent metrics. Show that the product is complete if each factor is complete.

1.5 Compactness

Some basic theorems of calculus, such as "every real valued continuous function on $[a, b]$ attains its maximum and minimum" implicitly use the fact that $[a, b]$ is compact.

Definition of Compactness. The general definition of compactness is rather unintuitive at the beginning. In fact, the general formulation of compactness and the realization of it as a useful tool is one of the excellent achievements of topology. But one has to be patient to see the rewards of formulating the definition the way it is done.

1.5.1 Definition. *Let S be a topological space. Then S is called **compact** if for every covering of S by open sets U_α (i.e., $\bigcup_\alpha U_\alpha = S$) there is a finite subcollection of the U_α also covering S. A subset $A \subset S$ is called **compact** if A is compact in the relative topology. A subset A is called **relatively compact** if $\mathrm{cl}(A)$ is compact. A space is called **locally compact** if it is Hausdorff and each point has a relatively compact neighborhood.*

Properties of Compactness. We shall soon see the true power of this notion, but lets work up to this with some simple observations.

1.5.2 Proposition.

(i) *If S is compact and $A \subset S$ is closed, then A is compact.*

(ii) *If $\varphi : S \to T$ is continuous and S is compact, then $\varphi(S)$ is compact.*

Proof. (i) Let $\{U_\alpha\}$ be an open covering of A. Then $\{U_\alpha, S\backslash A\}$ is an open covering of S and hence contains a finite subcollection of this covering also covering S. The elements of this collection, except $S\backslash A$, cover A.

(ii) Let $\{U_\alpha\}$ be an open covering of $\varphi(S)$. Then $\{\varphi^{-1}(U_\alpha)\}$ is an open covering of S and thus by compactness of S a finite subcollection $\{\, \varphi^{-1}(U_{\alpha(i)}) \mid i = 1, \ldots, n \,\}$, covers S. But then $\{U_{\alpha(i)}\}$, $i = 1, \ldots, n$ covers $\varphi(S)$ and thus $\varphi(S)$ is compact. ■

In a Hausdorff space, compact subsets are closed (exercise). Thus if S is compact, T is Hausdorff and φ is continuous, then φ is closed; if φ is also bijective, then it is a homeomorphism.

Compactness of Products. It is a basic fact that the product of compact spaces is compact.

1.5.3 Proposition. *A product space $S \times T$ is compact iff both S and T are compact.*

Proof. In view of Proposition 1.5.2 all we have to show is that if S and T are compact, so is $S \times T$. Let $\{A_\alpha\}$ be a covering of $S \times T$ by open sets. Each A_α is the union of sets of the form $U \times V$ with U and V open in S and T, respectively. Let $\{U_\beta \times V_\beta\}$ be a covering of $S \times T$ by open rectangles. If we show that there exists a finite subcollection of $U_\beta \times V_\beta$ covering $S \times T$, then clearly also a finite subcollection of $\{A_\alpha\}$ will cover $S \times T$.

A finite subcollection of $\{U_\beta \times V_\beta\}$ is found in the following way. Fix $s \in S$. Since the set $\{s\} \times T$ is compact, there is a finite collection

$$U_S \times V_{\beta_1}, \ldots, U_S \times V_{\beta_{i(s)}}$$

covering it. If $U_S = \bigcap_{j=1,\ldots,i(s)} U_{\beta_j}$, then U_s is open, contains s, and

$$U_s \times V_{\beta_1}, \ldots, U_S \times V_{\beta_{i(s)}}$$

covers $\{s\} \times T$. Let $W_s = U_s \times T$; then the collection $\{W_s\}$ is an open covering of $S \times T$ and if we show that only a finite number of these W_s cover $S \times T$, then since

$$W_s = \bigcup_{j=1,\ldots,i(s)} (U_s \times V_{\beta_j}),$$

it follows that a finite number of $U_\beta \times V_\beta$ will cover $S \times T$. Now look at $S \times \{t\}$, for $t \in T$ fixed. Since this is compact, a finite subcollection $W_{s_1}, \ldots, W_{s_k}$ covers it. But then

$$\bigcup_{j=1,\ldots,k} W_{s_i} = S \times T, \tag{1.5.1}$$

which proves the result. ■

As we shall see shortly in Theorem 1.5.9, $[-1, 1]$ is compact. Thus $\mathbb{T}^1$ is compact. It follows from Proposition 1.5.3 that the torus $\mathbb{T}^2$, and inductively $\mathbb{T}^n$, are compact. Thus, if $\pi : \mathbb{R}^2 \to \mathbb{T}^2$ is the canonical projection we see that $\mathbb{T}^2$ is compact without $\mathbb{R}^2$ being compact; that is, the converse of Proposition 1.5.2(ii) is false. Nevertheless it sometimes occurs that one does have a converse; this leads to the notion of a proper map discussed in Exercise 1.5-10.

Bolzano–Weierstrass Theorem. This theorem links compactness with convergence of sequences.

1.5.4 Theorem (Bolzano–Weierstrass Theorem). *If S is a compact first countable Hausdorff space, then every sequence has a convergent subsequence.*[1] *The converse is also true in metric and second-countable Hausdorff spaces.*

Proof. Suppose S is compact and $\{u_n\}$ contains no convergent subsequences. We may assume that the points of the sequence are distinct. Then $\mathrm{cl}(\{u_n\}) = \{u_n\}$ is compact and since S is first countable, each u_n has a neighborhood U_n that contains no other u_m, for otherwise u_n would be a limit of a subsequence. Thus $\{U_n\}$ is an open covering of the compact subset $\{u_n\}$ which contains no finite subcovering, a contradiction.

Let S be second countable, Hausdorff, and such that every sequence has a convergent subsequence. If $\{U_\alpha\}$ is an open covering of S, by the Lindelöf lemma there is a countable collection $\{\, U_n \mid n = 1, 2 \ldots \}$ also covering S. Thus we have to show that $\{\, U_n \mid n = 1, 2, \ldots \}$ contains a finite collection covering S. If this is not the case, the family consisting of sets of the form

$$S \backslash \bigcup_{i=1,\ldots,n} U_i$$

consists of closed nonempty sets and has the property that

$$S \backslash \bigcup_{i=1,\ldots,n} U_i \supset S \backslash \bigcup_{i=1,\ldots,m} U_i$$

for $m \geq n$. Choose

$$p_n \in S \backslash \bigcup_{i=1,\ldots,n} U_i.$$

If $\{\, p_n \mid n = 1, 2, \ldots \}$ is infinite, by hypothesis it contains a convergent subsequence; let its limit point be denoted p. Then

$$p \in S \backslash \bigcup_{i=1,\ldots,n} U_i$$

[1] There are compact Hausdorff spaces in which there are sequences with no convergent subsequences. See page 69 of Sims [1976] for more information.

for all n, contradicting the fact that $\{U_n \mid n = 1, 2, \dots\}$ covers S. Thus, $\{p_n \mid n = 1, 2, \dots\}$ must be a finite set; that is, for all $n \geq N$, $p_n = p_N$. But then again

$$p_N \in S \backslash \bigcup_{i=1,\dots,n} U_i$$

for all n, contradicting the fact that $\{U_n \mid n = 1, 2, \dots\}$ covers S. Hence S is compact.

Let S be a metric space such that every sequence has a convergent subsequence. If we show that S is separable, then since S is a metric space it is second countable (Exercise 1.2-3), and by the preceding paragraph, it will be compact. Separability of S is proved in two steps.

First we show that for any $\varepsilon > 0$ there is a *finite* set of points $\{p_1, \dots, p_n\}$ such that $S = \bigcup_{i=1,\dots,n} D_\varepsilon(p_i)$. If this were false, there would exist an $\varepsilon > 0$ such that no finite number of ε-disks cover S. Let $p_1 \in S$ be arbitrary. Since $D_\varepsilon(p_1) \neq S$, there is a point $p_2 \in S \backslash D_\varepsilon(p_1)$. Since

$$D_\varepsilon(p_1) \cup D_\varepsilon(p_2) \neq S,$$

there is also a point

$$p_3 \in S \backslash (D_\varepsilon(p_1) \cup D_\varepsilon(p_2)),$$

etc. The sequence $\{p_n \mid n = 1, 2, \dots\}$ is infinite and $d(p_i, p_j) \geq \varepsilon$. But this sequence has a convergent subsequence by hypothesis, so this subsequence must be Cauchy, contradicting $d(p_i, p_j) \geq \varepsilon$ for all i, j.

Second, we show that the existence for every $\varepsilon > 0$ of a finite set $\{p_1, \dots, p_{n(\varepsilon)}\}$ such that

$$S = \bigcup_{i=1,\dots,n(\varepsilon)} D_\varepsilon(p_i)$$

implies S is separable. Let A_n denote this finite set for $\varepsilon = 1/n$ and let

$$A = \bigcup_{n \geq 0} A_n.$$

Thus A is countable and it is easily verified that $\operatorname{cl}(A) = S$. ∎

Total Boundedness. A property that came up in the preceding proof turns out to be important.

1.5.5 Definition. *Let S be a metric space. A subset $A \subset S$ is called* ***totally bounded*** *if for any $\varepsilon > 0$ there exists a finite set $\{p_1, \dots, p_n\}$ in S such that*

$$A \subset \bigcup_{i=1,\dots,n} D_\varepsilon(p_i).$$

1.5.6 Corollary. *A metric space is compact iff it is complete and totally bounded. A subset of a complete metric space is relatively compact iff it is totally bounded.*

Proof. The previous proof shows that compactness implies total boundedness. As for compactness implying completeness, it is enough to remark that in this context, a Cauchy sequence having a convergent subsequence is itself convergent. Conversely, if S is complete and totally bounded, let $\{p_n \mid n = 1, 2, \dots\}$ be a sequence in S. By total boundedness, this sequence contains a Cauchy subsequence, which by completeness, converges. Thus S is compact by the Bolzano–Weierstrass theorem. The second statement now readily follows. ∎

1.5.7 Proposition. *In a metric space compact sets are closed and bounded.*

Proof. This is a particular case of the previous corollary but can be easily proved directly. If A is compact, it can be finitely covered by ε-disks:

$$A = \bigcup_{i=1,\ldots,n} D_\varepsilon(p_i).$$

Thus,

$$\operatorname{diam}(A) \leq \sum_{i=1}^{n} \operatorname{diam}(D_\varepsilon(p_i)) = 2n\varepsilon. \qquad \blacksquare$$

From Proposition 1.5.2 and Proposition 1.5.7, we conclude that

1.5.8 Corollary. *If S is compact and $\varphi : S \to \mathbb{R}$ is continuous, then φ is bounded and attains its sup and inf.*

Indeed, since S is compact, so is $\varphi(S)$ and so $\varphi(S)$ is closed and bounded. Thus (see Exercise 1.1-4) the inf and sup of this set are finite and are members of this set.

Heine–Borel Theorem. This result makes it easy to spot compactness in Euclidean spaces.

1.5.9 Theorem (Heine–Borel Theorem)**.** *In $\mathbb{R}^n$ a closed and bounded set is compact.*

Proof. By Proposition 1.5.2(i) it is enough to show that closed bounded rectangles are compact in $\mathbb{R}^n$, which in turn is implied via Proposition 1.5.3 by the fact that closed bounded intervals are compact in $\mathbb{R}$. To show that $[-a, a]$, $a > 0$ is compact, it suffices to prove (by Corollary 1.5.6) that for any given $\varepsilon > 0$, $[-a, a]$ can be finitely covered by intervals of the form $]p - \varepsilon, p + \varepsilon[$, since we are accepting completeness of $\mathbb{R}$. Let n be a positive integer such that $a < n\varepsilon$. Let $t \in [-a, a]$ and k be the largest (positive or negative) integer satisfying $k\varepsilon \leq t$. Then $-n \leq k \leq n$ and $k\varepsilon \leq t < (k+1)\varepsilon$. Thus any point $t \in [-a, a]$ belongs to an interval of the form $]k\varepsilon - \varepsilon, k\varepsilon + \varepsilon[$, where $k = -n, \ldots, 0, \ldots, n$ and hence $\{\,]k\varepsilon - \varepsilon, k\varepsilon + \varepsilon[\mid k = 0, \pm 1, \ldots, \pm n \,\}$ is a finite covering of $[-a, a]$. $\blacksquare$

This theorem is also proved in virtually every textbook on advanced calculus.

Uniform Continuity. As is known from calculus, continuity of a function on an interval $[a, b]$ implies uniform continuity. The generalization to metric spaces is the following.

1.5.10 Proposition. *A continuous mapping $\varphi : M_1 \to M_2$, where M_1 and M_2 are metric spaces and M_1 is compact, is uniformly continuous.*

Proof. The metrics on M_1 and M_2 are denoted by d_1 and d_2. Fix $\varepsilon > 0$. Then for each $p \in M_1$, by continuity of φ there exists $\delta_p > 0$ such that if $d_1(p, q) < \delta_p$, then $d_2(\varphi(p), \varphi(q)) < \varepsilon/2$. Let

$$D_{\delta_1/2}(p_1), \ldots, D_{\delta_n/2}(p_n)$$

cover the compact space M_1 and let $\delta = \min\{\delta_1/2, \ldots, \delta_n/2\}$. Then if $p, q \in M_1$ are such that $d_1(p, q) < \delta$, there exists an index i, $1 \leq i \leq n$, such that $d_1(p, p_i) < \delta_i/2$ and thus

$$d_1(p_i, q) \leq d_1(p_i, p) + d_1(p, q) < \frac{\delta_i}{2} + \delta \leq \delta_i.$$

Thus,

$$d_2(\varphi(p), \varphi(q)) \leq d_2(\varphi(p), \varphi(p_i)) + d_2(\varphi(p_i)), \varphi(q)) < \varepsilon. \qquad \blacksquare$$

Equicontinuity. A useful application of Corollary 1.5.6 concerns relatively compact sets in $C(M,N)$, for metric spaces (M,d_M) and (N,d_N) with M compact and N complete. Recall from §1.3 that we put a metric on $C(M,N)$ and that in this metric, convergence is the same as uniform convergence.

1.5.11 Definition. *A subset $\mathcal{F} \subset C(M,N)$ is called **equicontinuous** at $m_0 \in M$, if given $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $d_M(m,m_0) < \delta$, we have $d_N(\varphi(m),\varphi(m_0)) < \varepsilon$ for every $\varphi \in \mathcal{F}$ (δ is independent of φ). $\mathcal{F}$ is called **equicontinuous**, if it is equicontinuous at every point in M.*

1.5.12 Theorem (Arzela–Ascoli Theorem)**.** *Let (M,d_M) and (N,d_N) be metric spaces, and assume that M is compact and N is complete. A set $\mathcal{F} \subset C(M,N)$ is relatively compact iff it is equicontinuous and all the sets $\mathcal{F}(m) = \{\, \varphi(m) \mid \varphi \in \mathcal{F} \,\}$ are relatively compact in N.*

Proof. If $\mathcal{F}$ is relatively compact, it is totally bounded and hence so are all the sets $\mathcal{F}(m)$. Since N is complete, by Corollary 1.5.6 the sets $\mathcal{F}(m)$ are relatively compact. Let $\{\varphi_1,\dots,\varphi_n\}$ be the centers of the ε-disks covering $\mathcal{F}$. Then there exists $\delta > 0$ such that if $d_M(m,m') < \delta$, we have $d_N(\varphi_i(m),\varphi_i(m')) \le \varepsilon/3$, for $i = 1,\dots,n$ and hence if $\varphi \in \mathcal{F}$ is arbitrary, φ lies in one of the ε-disks whose center, say, is φ_i, so that

$$\begin{aligned} d_N(\varphi(m),\varphi(m')) &\le d_N(\varphi(m),\varphi_i(m)) + d_N(\varphi_i(m),\varphi_i(m')) \\ &\quad + d_N(\varphi_i(m'),\varphi(m')) < \varepsilon. \end{aligned}$$

This shows that $\mathcal{F}$ is equicontinuous.

Conversely, since $C(M,N)$ is complete, by Corollary 1.5.6 we need only show that $\mathcal{F}$ is totally bounded. For $\varepsilon > 0$, find a neighborhood U_m of $m \in M$ such that for all $m' \in U_m$, $d_N(\varphi(m),\varphi(m')) < \varepsilon/4$ for all $\varphi \in \mathcal{F}$ (this is possible by equicontinuity). Let $U_{m(1)},\dots,U_{m(n)}$ be a finite collection of these neighborhoods covering the compact space M. By assumption each $\mathcal{F}(m)$ is relatively compact, hence $\mathcal{F}(m(1)) \cup \cdots \cup \mathcal{F}(m(n))$ is also relatively compact, and thus totally bounded. Let $D_{\varepsilon/4}(x_1),\dots,D_{e/4}(x_k)$ cover this union. If $\mathcal{A}$ denotes the set of all mappings $\alpha : \{1,\dots,n\} \to \{1,\dots,k\}$, then $\mathcal{A}$ is finite and

$$\mathcal{F} = \bigcup_{a \in \mathcal{A}} \mathcal{F}_\alpha,$$

where $\mathcal{F}_\alpha = \{\, \varphi \in \mathcal{F} \mid d_N(\varphi(m(i)), x_{\alpha(i)}) < \varepsilon/4 \text{ for all } i = 1,\dots,n \,\}$. But if $\varphi, \psi \in \mathcal{F}_\alpha$ and $m \in M$, then $m \in D_{\varepsilon/4}(x_i)$ for some i, and thus

$$\begin{aligned} d_N(\varphi(m),\psi(m)) &\le d_N(\varphi(m),\varphi(m(i)) + d_N(\varphi(m(i)),x_{\alpha(i)}) \\ &\quad + d_N(x_{\alpha(i)},\psi(m(i))) + d_N(\psi(m(i)),\psi(m)) < \varepsilon; \end{aligned}$$

that is, the diameter of $\mathcal{F}_\alpha$ is $\le \varepsilon$, so $\mathcal{F}$ is totally bounded. ■

Combining this with the Heine–Borel theorem, we get the following.

1.5.13 Corollary. *If M is a compact metric space, a set $\mathcal{F} \subset C(M,\mathbb{R}^n)$ is relatively compact iff it is equicontinuous and uniformly bounded (i.e., $\|\varphi(m)\| \le$ constant for all $\varphi \in \mathcal{F}$ and $m \in M$).*

The following example illustrates one way to use the Arzela–Ascoli theorem.

Example. Let $f_n : [0,1] \to \mathbb{R}$ be continuous and be such that $|f_n(x)| \le 100$ and the derivatives f_n' exist and are uniformly bounded on $]0,1[$. Prove f_n has a uniformly convergent subsequence.

We verify that the set $\{f_n\}$ is equicontinuous and bounded. The hypothesis is that $|f_n'(x)| \le M$ for a constant M. Thus by the mean-value theorem,

$$|f_n(x) - f_n(y)| \le M|x - y|,$$

so given ε we can choose $\delta = \varepsilon/M$, independent of x, y, and n. Thus $\{f_n\}$ is equicontinuous. It is bounded because

$$\|f_n\| = \sup_{0 \le x \le 1} |f_n(x)| \le 100. \qquad ♦$$

Exercises

◇ **1.5-1.** Show that a topological space S is compact iff every family of closed subsets of S whose intersection is empty contains a finite subfamily whose intersection is empty.

◇ **1.5-2.** Show that every compact metric space is separable.
HINT: Use total boundedness.

◇ **1.5-3.** Show that the space of Exercise 1.1-9 is not locally compact.
HINT: Look at the sequence $(1/n, 0)$.

◇ **1.5-4.** (i) Show that every closed subset of a locally compact space is locally compact.

(ii) Show that $S \times T$ is locally compact if both S and T are locally compact.

◇ **1.5-5.** Let M be a compact metric space and $T : M \to M$ a map satisfying $d(T(m_1), T(m_2)) < d(m_1, m_2)$ for $m_1 \neq m_2$. Show that T has a unique fixed point.

◇ **1.5-6.** Let S be a compact topological space and $\sim$ an equivalence relation on S, so that $S/\sim$ is compact. Prove that the following conditions are equivalent (cf. Proposition 1.4.10):

(i) The graph C of $\sim$ is closed in $S \times S$;

(ii) $\sim$ is a closed equivalence relation;

(iii) $S/\sim$ is Hausdorff.

◇ **1.5-7.** Let S be a Hausdorff space that is locally homeomorphic to a locally compact Hausdorff space (i.e., for each $u \in S$, there is a neighborhood of u homeomorphic, in the subspace topology, to an open subset of a locally compact Hausdorff space). Show that S is locally compact. In particular, Hausdorff spaces locally homeomorphic to $\mathbb{R}^n$ are locally compact. Is the conclusion true without the Hausdorff assumption?

◇ **1.5-8.** Let M_3 be the set of all 3×3 matrices with the topology obtained by regarding M_3 as $\mathbb{R}^9$. Let $\mathrm{SO}(3) = \{ A \in M_3 \mid A \text{ is orthogonal and } \det A = 1 \}$.

(i) Show that SO(3) is compact.

(ii) Let $P = \{ Q \in \mathrm{SO}(3) \mid Q \text{ is symmetric} \}$ and let $\varphi : \mathbb{RP}^2 \to \mathrm{SO}(3)$ be given by $\varphi(\ell) =$ the rotation by π about the line $\ell \subset \mathbb{R}^3$. Show that φ maps the space $\mathbb{RP}^2$ homeomorphically onto $P\backslash\{\text{Identity}\}$.

◇ **1.5-9.** Let $f_n : [a, b] \to \mathbb{R}$ be uniformly bounded continuous functions. Set

$$F_n(x) = \int_a^x f_n(t)\, dt, \quad a \le x \le b.$$

Prove that F_n has a uniformly convergent subsequence.

◇ **1.5-10.** Let X and Y be topological spaces, Y be first countable, and $f : X \to Y$ be a continuous map. The map f is called ***proper***, if $f(x_n) \to y$ implies the existence of a convergent subsequence $\{x_{n(i)}\}$, $x_{n(i)} \to x$ such that $f(x) = y$.

(i) Show that f is a closed map.

(ii) Show that if Y is locally compact, f is proper if and only if the inverse image by f of every compact set in Y is a compact set in X.

(iii) Show that if f is proper and Y is locally compact, then X is also locally compact. (We have defined properness only when Y is first countable. The same definition and properties of proper maps hold for general Y if in the definition "sequence" is replaced by "net.")

(iv) Show that the composition of two proper maps is again proper.

(v) Show that any continuous map defined on a compact space is proper.

1.6 Connectedness

Three types of connectedness treated in this section are arcwise connectedness, connectedness, and simple connectedness.

Arcwise Connectedness. We begin with the most intuitive notion of connectedness.

1.6.1 Definition. *Let S be a topological space and $I = [0, 1] \subset \mathbb{R}$. An* ***arc*** *$\varphi$ in S is a continuous mapping $\varphi : I \to S$. If $\varphi(0) = u$, $\varphi(1) = v$, we say φ* ***joins*** *u and v; S is called* ***arcwise connected*** *if every two points in S can be joined by an arc in S. A space S is called* ***locally arcwise connected*** *if for each point $x \in S$ and each neighborhood U of x, there is a neighborhood V of x such that any pair of points in V can be joined by an arc in U.*

For example, $\mathbb{R}^n$ is arcwise and locally arcwise connected: any two points of $\mathbb{R}^n$ can be joined by the straight line segment connecting them. A set $A \subset \mathbb{R}^n$ is called ***convex*** if this property holds for any two of its points. Thus, convex sets in $\mathbb{R}^n$ are arcwise and locally arcwise connected. A set with the trivial topology is arcwise and locally arcwise connected, but in the discrete topology it is neither (unless it has only one point).

Connected Spaces. Less intuitive is the basic notion of connectedness.

1.6.2 Definition. *A topological space S is* ***connected*** *if $\varnothing$ and S are the only subsets of S that are both open and closed. A subset of S is* ***connected*** *if it is connected in the relative topology. A* ***component*** *A of S is a nonempty connected subset of S such that the only connected subset of S containing A is A itself; S is called* ***locally connected*** *if each point has a connected neighborhood. The* ***components*** *of a subset $T \subset S$ are the components of T in the relative topology of T in S.*

For example, $\mathbb{R}^n$ and any convex subset of $\mathbb{R}^n$ are connected and locally connected. The union of two disjoint open convex sets is disconnected but is locally connected; its components are the two convex sets. The trivial topology is connected and locally connected, whereas the discrete topology is neither: its components are all the one-point sets.

Connected spaces are characterized by the following.

1.6.3 Proposition. *The following are equivalent:*

(i) *S is not connected*;

(ii) *there is a nonempty proper subset of S that is both open and closed*;

(iii) *S is the disjoint union of two nonempty open sets*; *and*

(iv) *S is the disjoint union of two nonempty closed sets.*

The sets in (iii) *or* (iv) *are said to* ***disconnect*** *S.*

Proof. To prove that (i) implies (ii), assume there is a nonempty proper set A that is both open and closed. Then $S = A \cup (S \backslash A)$ with A, $S \backslash A$ open and nonempty. Conversely, if $S = A \cup B$ with A, B open and nonempty, then A is also closed, and thus A is a proper nonempty set of S that is both open and closed. The equivalences of the remaining assertions are similarly checked. ■

Behavior under Mappings. Connectedness is preserved by continuous maps, as is shown next.

1.6.4 Proposition. *If $f : S \to T$ is a continuous map of topological spaces and S is connected* (*resp., arcwise connected*) *then so is $f(S)$.*

Proof. Let S be arcwise connected and consider $f(s_1), f(s_2) \in f(S) \subset T$. If $c : I \to S$, $c(0) = s_1$, $c(1) = s_2$ is an arc connecting s_1 to s_2, then clearly $f \circ c : I \to T$ is an arc connecting $f(s_1)$ to $f(s_2)$; that is, $f(S)$ is arcwise connected. Let S be connected and assume $f(S) \subset U \cup V$, where U and V are open and $U \cap V = \varnothing$. Then $f^{-1}(U)$ and $f^{-1}(V)$ are open by continuity of f,

$$f^{-1}(U) \cup f^{-1}(V) = f^{-1}(U \cup V) \supset f^{-1}(f(S)) = S,$$

and $f^{-1}(U) \cap f^{-1}(V) = f^{-1}(\varnothing) = \varnothing$, thus contradicting connectedness of S by Proposition 1.6.3. Hence $f(S)$ is connected. ■

Arcwise Connected Spaces are Connected. We shall use the following.

1.6.5 Lemma. *The only connected sets of $\mathbb{R}$ are the intervals (finite, infinite, open, closed, or half-open).*

Proof. Let us prove that $[a, b[$ is connected; all other possibilities have identical proofs. If not, $[a, b[= U \cup V$ with U, V nonempty disjoint closed sets in $[a, b[$. Assume that $a \in U$. If $x = \sup(U)$, then $x \in U$ since U is closed in $[a, b[$, and $x < b$ since $V \neq \varnothing$. But then $]x, b[\subset V$ and, since V is closed, $x \in V$. Hence $x \in U \cap V$, a contradiction.

Conversely, let A be a connected set of $\mathbb{R}$. We claim that $[x, y] \subset A$ whenever $x, y \in A$, which implies that A is an interval. If not, there exists $z \in [x, y]$ with $z \notin A$. But in this case $]-\infty, z[\cup A$ and $]z, \infty[\cup A$ are open nonempty sets disconnecting A. ■

1.6.6 Proposition. *If S is arcwise connected then it is connected.*

Proof. If not, there are nonempty, disjoint open sets U_0 and U_1 whose union is S. Let $x_0 \in U_0$ and $x_1 \in U_1$ and let φ be an arc joining x_0 to x_1. Then $V_0 = \varphi^{-1}(U_0)$ and $V_1 = \varphi^{-1}(U_1)$ disconnect $[0, 1]$. ■

A standard example of a space that is connected but is not arcwise connected nor locally connected, is

$$\left\{ (x, y) \in \mathbb{R}^2 \mid x > 0 \text{ and } y = \sin(1/x) \right\} \cup \{ (0, y) \mid -1 < y < 1 \}.$$

1.6.7 Proposition. *If a space is connected and locally arcwise connected, it is arcwise connected. In particular, a space locally homeomorphic to $\mathbb{R}^n$ is connected iff it is arcwise connected.*

Proof. Fix $x \in S$. The set

$$A = \{ y \in S \mid y \text{ can be connected to } x \text{ by an arc} \}$$

is nonempty and open since S is locally arcwise connected. For the same reason, $S \backslash A$ is open. Since S is connected we must have $S \backslash A = \varnothing$; thus, $A = S$, that is, S is arcwise connected. ■

Intermediate Value Theorem. Connectedness provides a general context for this theorem learned in calculus.

1.6.8 Theorem (Intermediate Value Theorem). *Let S be a connected space and $f : S \to \mathbb{R}$ be continuous. Then f assumes every value between any two values $f(u)$ and $f(v)$.*

Proof. Suppose $f(u) < a < f(v)$ and f does not assume the value a. Then the set $U = \{ u_0 \mid f(u_0) < a \}$ is both open and closed. ■

An alternative proof uses the fact that $f(S)$ is connected in $\mathbb{R}$ and therefore is an interval.

Miscellaneous Properties of Connectedness.

1.6.9 Proposition. *Let S be a topological space and $B \subset S$ be connected.*

(i) *If $B \subset A \subset \mathrm{cl}(B)$, then A is connected.*

(ii) *If B_α is a family of connected subsets of S and $B_\alpha \cap B \neq \varnothing$, then*

$$B \cup \left(\bigcup_\alpha B_\alpha \right)$$

is connected.

Proof. If A is not connected, A is the disjoint union of $U_1 \cap A$ and $U_2 \cap A$ where U_1 and U_2 are open in S. Then from Proposition 1.1.9(i), $U_1 \cap B \neq \varnothing$ and $U_2 \cap B \neq \varnothing$, so B is not connected. We leave (ii) as an exercise. ■

1.6.10 Corollary. *The components of a topological space are closed. Also, S is the disjoint union of its components. If S is locally connected, the components are open as well as closed.*

1.6.11 Proposition. *Let S be a first countable compact Hausdorff space and $\{A_n\}$ a sequence of closed, connected subsets of S with $A_n \subset A_{n-1}$. Then $A = \bigcap_{n \geq 1} A_n$ is connected.*

Proof. As S is normal, if A is not connected, A lies in two disjoint open subsets U_1 and U_2 of S. If $A_n \cap (S \backslash U_1) \cap (S \backslash U_2) \neq \varnothing$ for all n, then there is a sequence $u_n \in A_n \cap (S \backslash U_1) \cap (S \backslash U_2)$ with a subsequence converging to u. As A_n, $S \backslash U_1$, and $S \backslash U_2$ are closed sets, $u \in A \cap (S \backslash U_1) \cap (S \backslash U_2)$, a contradiction. Hence some A_n is not connected. ■

Simple Connectivity. This notion means, intuitively, that loops can be continuously shrunk to points.

1.6.12 Definition. *Let S be a topological space and $c : [0,1] \to S$ a continuous map such that $c(0) = c(1) = p \in S$. We call c a* ***loop*** *in S based at p. The loop c is called* ***contractible*** *if there is a continuous map $H : [0,1] \times [0,1] \to S$ such that $H(t,0) = c(t)$ and $H(0,s) = H(1,s) = H(t,1) = p$ for all $t \in [0,1]$. (See Figure 1.6.1.)*

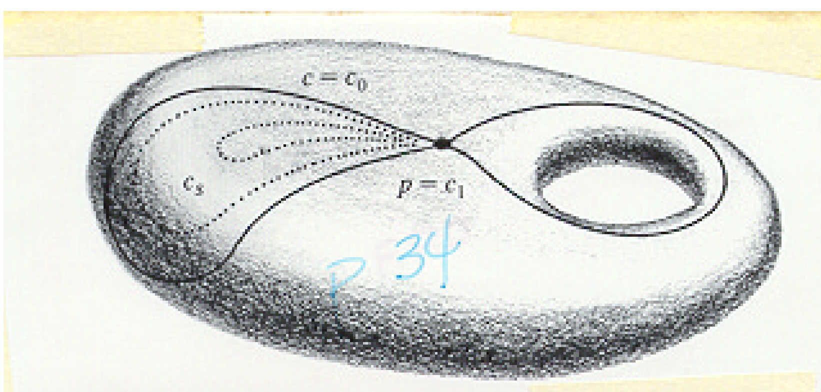

FIGURE 1.6.1. The loop c is contractible

We think of $c_s(t) = H(t,s)$ as a family of arcs connecting $c_0 = c$ to c_1, a constant arc; see Figure 1.6.1. Roughly speaking, a loop is contractible when it can be shrunk continuously to p by loops beginning and ending at p. The study of loops leads naturally to homotopy theory. In fact, the loops at p can, by successively traversing them, be made into a group called the ***fundamental group***; see Exercise 1.6-6.

1.6.13 Definition. *A space S is **simply connected** if S is connected and every loop in S is contractible.*

In the plane $\mathbb{R}^2$ there is an alternative approach to simple connectedness, by way of the Jordan curve theorem; namely, that every simple (nonintersecting) loop in $\mathbb{R}^2$ divides $\mathbb{R}^2$ (*divides* means that its complement has two components). The bounded component of the complement is called the interior, and a subset A of $\mathbb{R}^2$ is simply connected iff the interior of every loop in A lies in A.

Alexandroff's Theorem. We close this section with an optional theorem sometimes used in Riemannian geometry (to show that a Riemannian manifold is second countable) that illustrates the interplay between various notions introduced in this chapter.

1.6.14 Theorem (Alexandroff's Theorem)**.** *An arcwise connected locally compact metric space is separable and hence is second countable.*

Proof (Pfluger [1957])**.** Since the metric space M is locally compact, each $m \in M$ has compact neighborhoods that are disks. Denote by $r(m)$ the least upper bound of the radii of such disks. If $r(m) = \infty$, since every metric space is first countable, M can be written as a countable union of compact disks. But since each compact metric space is separable (Exercise 1.5-2), these disks and also their union will be separable, and so the proposition is proved in this case. If $r(m_0) < \infty$, since $r(m) \leq r(m_0) + d(m, m_0)$, we see that $r(m) < \infty$ for all $m \in M$. By the preceding argument, if we show that M is a countable union of compact sets, the proposition is proved. Then second countability will follow from Exercise 1.2-3.

To show that M is a countable union of compact sets, define the set G_m by

$$G_m = \{\, m' \in M \mid d(m', m) \leq r(m)/2 \,\}.$$

These G_m are compact neighborhoods of m. Fix $m(0) \in M$ and put $A_0 = G_{m(0)}$, and, inductively, define

$$A_{n+1} = \bigcup \{\, G_m \mid m \in A_n \,\}.$$

Since M is arcwise connected, every point $m \in M$ can be connected by an arc to $m(0)$, which in turn is covered by finitely many G_m. This shows that

$$M = \bigcup_{n \geq 0} A_n.$$

Since A_0 is compact, all that remains to be shown is that the other A_n are compact. Assume inductively that A_n is compact and let $\{m(i)\}$ be an infinite sequence of points in A_{n+1}. There exists $m(i)' \in A_n$ such that $m(i) \in G_{m(i)'}$. Since A_n is assumed to be compact there is a subsequence $m(i_k)'$ that converges to a point $m' \in A_n$. But

$$\begin{aligned} d(m(i), m') &\leq d(m(i), m(i)') + d(m(i)', m') \\ &\leq \frac{r(m(i)')}{2} + d(m(i)', m') \\ &\leq \frac{r(m')}{2} + \frac{3d(m(i)', m')}{2}. \end{aligned}$$

Hence for i_k big enough, all $m(i_k)$ are in the compact set

$$\{\, n \in M \mid d(n, m') \leq 3r(m')/2 \,\},$$

so $m(i_k)$ has a subsequence converging to a point m. The preceding inequality shows that $m \in A_{n+1}$. By the Bolzano–Weierstrass theorem, A_{n+1} is compact. ∎

Exercises

◇ **1.6-1.** Let M be a topological space and $H : M \to \mathbb{R}$ continuous. Suppose $e \in \operatorname{Int} H(M)$. Then show $H^{-1}(e)$ divides M; that is, $M \backslash H^{-1}(e)$ has at least two components.

◇ **1.6-2.** Let $\mathcal{O}(3)$ be the set of orthogonal 3×3 matrices. Show that $\mathcal{O}(3)$ is not connected and that it has two components.

◇ **1.6-3.** Show that $S \times T$ is connected (locally connected, arcwise connected, locally arcwise connected) iff both S and T are.
HINT: For connectedness write

$$S \times T = \bigcup_{t \in T} [(S \times \{t\}) \cup (\{s_0\} \times T)]$$

for $s_0 \in S$ fixed and use Proposition 1.6.9(ii).

◇ **1.6-4.** Show that S is locally connected iff every component of an open set is open.

◇ **1.6-5.** Show that the quotient space of a connected (locally connected, arcwise connected) space is also connected (locally connected, arcwise connected).
HINT: For local connectedness use Exercise 1.6-4 and show that the inverse image by π of a component of an open set is a union of components.

◇ **1.6-6.** (i) Let S and T be topological spaces. Two continuous maps $f, g : T \to S$ are called ***homotopic*** if there exists a continuous map $F : [0,1] \times T \to S$ such that $F(0,t) = f(t)$ and $F(1,t) = g(t)$ for all $t \in T$. Show that homotopy is an equivalence relation.

(ii) Show that S is simply connected if and only if any two continuous paths $c_1, c_2 : [0,1] \to S$ satisfying $c_1(0) = c_2(0)$, $c_1(1) = c_2(1)$ are homotopic, via a homotopy which preserves the end points, that is, $F(s,0) = c_1(0) = c_2(0)$ and $F(s,1) = c_1(1) = c_2(1)$.

(iii) Define the ***composition*** $c_1 * c_2$ of two paths $c_1, c_2 : [0,1] \to S$ satisfying $c_1(1) = c_2(0)$ by

$$(c_1 * c_2)(t) = \begin{cases} c_1(2t) & \text{if } t \in [1, 1/2]; \\ c_2(2t-1) & \text{if } t \in [1/2, 1]. \end{cases}$$

Show that this composition, when defined, induces an associative operation on endpoints preserving homotopy classes of paths.

(iv) Fix $s_0 \in S$ and consider the set $\pi_1(S, s_0)$ of endpoint fixing homotopy classes of paths starting and ending at s_0. Show that $\pi_1(S, s_0)$ is a group: the identity element is given by the class of the constant path equal to s_0 and the inverse of c is given by the class of $c(1-t)$.

(v) Show that if S is arcwise connected, then $\pi_1(S, s_0)$ is isomorphic to $\pi_1(S, s)$ for any $s \in S$. $\pi_1(S)$ will denote any of these isomorphic groups.

(vi) Show that if S is arcwise connected, then S is simply connected iff $\pi_1(S) = 0$.

1.7 Baire Spaces

The Baire condition on a topological space is fundamental to the idea of "genericity" in differential topology and dynamical systems; see Kelley [1975] and Choquet [1969] for additional information.

1.7.1 Definition. *Let X be a topological space and $A \subset X$ a subset. Then A is called* ***residual*** *if A is the intersection of a countable family of open dense subsets of X. A space X is called a* ***Baire space*** *if every residual set is dense. A set $B \subset X$ is called a* ***first category*** *set if*

$$B \subset \bigcup_{n \geq 1} C_n,$$

where C_n is closed with $\operatorname{int}(C_n) = \varnothing$. *A* ***second category set*** *is a set which is not of the first category.*

A set $B \subset X$ is called ***nowhere dense*** if $\operatorname{int}(\operatorname{cl}(B)) = \varnothing$, so that $X \backslash A$ is residual iff A is the union of a countable collection of nowhere dense closed sets, that is, iff $X \backslash A$ is of first category. Clearly, a countable intersection of residual sets is residual.

In a Baire space X, if

$$X = \bigcup_{n \geq 1} C_n,$$

where C_n are closed sets, then $\operatorname{int}(C_n) \neq \varnothing$ for some n. For if all $\operatorname{int}(C_n) = \varnothing$, then $O_n = X \backslash C_n$ are open, dense, and we have

$$\bigcap_{n \geq 1} O_n = X \backslash \bigcup_{n \geq 1} C_n = \varnothing$$

contradicting the definition of Baire space. In other words, *Baire spaces are of second category*.

1.7.2 Proposition. *Let X be a* ***locally Baire space****; that is, each point $x \in X$ has a neighborhood U such that* $\operatorname{cl}(U)$ *is a Baire space. Then X is a Baire space.*

Proof. Let $A \subset X$ be residual, $A = \bigcap_{n \geq 1} O_n$, where $\operatorname{cl}(O_n) = X$. Then if U is an open set for which $\operatorname{cl}(U)$ is a Baire space, from the equality $A \cap \operatorname{cl}(U) = \bigcap_{n \geq 1}(O_n \cap \operatorname{cl}(U))$ and the density of $O_n \cap \operatorname{cl}(U)$ in $\operatorname{cl}(U)$ (if $u \in \operatorname{cl}(U)$ and $u \in O$, O open in X, then $O \cap U \neq \varnothing$, and therefore $O \cap U \cap O_n \neq \varnothing$), it follows that $A \cap \operatorname{cl}(U)$ is residual in $\operatorname{cl}(U)$ hence dense in $\operatorname{cl}(U)$, that is, $\operatorname{cl}(A) \cap \operatorname{cl}(U) = \operatorname{cl}(U)$ so that $\operatorname{cl}(U) \subset \operatorname{cl}(A)$. Therefore $X = \operatorname{cl}(A)$. ■

1.7.3 Theorem (Baire Category Theorem). *Complete pseudometric and locally compact spaces are Baire spaces.*

Proof. Let X be a complete pseudometric space. Let $U \subset X$ be open and

$$A = \bigcap_{n \geq 1} O_n$$

be residual. We must show $U \cap A \neq \varnothing$. Since $\operatorname{cl}(O_n) = X$,

$$U \cap O_n \neq \varnothing$$

and so we can choose a disk of diameter less than one, say V_1, such that $\operatorname{cl}(V_1) \subset U \cap O_1$. Proceed inductively to obtain

$$\operatorname{cl}(V_n) \subset U \cap O_n \cap V_{n-1},$$

where V_n has diameter $< 1/n$. Let $x_n \in \operatorname{cl}(V_n)$. Clearly $\{x_n\}$ is a Cauchy sequence, and by completeness has a convergent subsequence with limit point x. Then

$$x \in \bigcap_{n \geq 1} \operatorname{cl}(V_n)$$

and so

$$U \cap \left(\bigcap_{n \geq 1} O_n \right) \neq \varnothing;$$

that is, A is dense in X. If X is a locally compact space the same proof works with the following modifications: V_n are chosen to be relatively compact open sets, and $\{x_n\}$ has a convergent subsequence since it lies in the compact set $\mathrm{cl}(V_1)$. ■

To get a feeling for this theorem, let us prove that the set of rationals $\mathbb{Q}$ cannot be written as a countable intersection of open sets. For suppose $\mathbb{Q} = \bigcap_{n\geq 1} O_n$. Then each O_n is dense in $\mathbb{R}$, since $\mathbb{Q}$ is, and so $C_n = \mathbb{R}\backslash O_n$ is closed and nowhere dense. Since

$$\mathbb{R} \cup \left(\bigcup_{n\geq 1} C_n \right)$$

is a complete metric space (as well as a locally compact space), it is of second category, so $\mathbb{Q}$ or some C_n should have nonempty interior. But this is impossible.

The notion of category can lead to interesting restrictions on a set. For example in a nondiscrete Hausdorff space, any countable set is first category since the one-point set is closed and nowhere dense. Hence in such a space *every second category set is uncountable*. In particular, nonfinite complete pseudometric and locally compact spaces are uncountable.

Exercises

◇ **1.7-1.** Let X be a Baire space. Show that

(i) X is a second category set;

(ii) if $U \subset X$ is open, then U is Baire.

◇ **1.7-2.** Let X be a topological space. A set is called an $\mathcal{F}_\sigma$ if it is a countable union of closed sets, and is called a $\mathcal{G}_\delta$ if it is a countable intersection of open sets. Prove that the following are equivalent:

(i) X is a Baire space;

(ii) any first category set in X has a dense complement;

(iii) the complement of every first category $\mathcal{F}_\sigma$-set is a dense $\mathcal{G}_\delta$-set;

(iv) for any countable family of closed sets $\{C_n\}$ satisfying

$$X = \bigcup_{n\geq 1} C_n,$$

the open set

$$\bigcup_{n\geq 1} \mathrm{int}(C_n)$$

is dense in X.

Hint: First show that (ii) is equivalent to (iv). For (ii) implies (iv), let $U_n = C_n \backslash \mathrm{int}(C_n)$ so that $\bigcup_{n\geq 1} U_n$ is a first category set and therefore $X \backslash \bigcup_{n\geq 1} U_n$ is dense and included in $\bigcup_{n\geq 1} \mathrm{int}(C_n)$. For the converse, assume X is not Baire so that $A = \bigcap_{n\geq 1} U_n$ is not dense, even though all U_n are open and dense. Then

$$X = \mathrm{cl}(A) \cup \{\, X\backslash U_n \mid n = 1, 2, \ldots \}.$$

Put

$$F_0 = \mathrm{cl}(A), \quad F_n = X\backslash U_n,$$

and show that $\mathrm{int}(F_n) = \mathrm{int}(\mathrm{cl}(A))$ which is not dense.

◇ **1.7-3.** Show that there is a residual set E in the metric space $C([0,1],\mathbb{R})$ such that each $f \in E$ is not differentiable at any point. do this by following the steps below.

(i) Let E_ε denote the set of all $f \in C([0,1],\mathbb{R})$ such that for every $x \in [0,1]$,

$$\operatorname{diam}\left\{ \frac{f(x+h)-f(x)}{h} \;\middle|\; \frac{\varepsilon}{2} < |h| < \varepsilon \right\} > 1$$

for $\varepsilon > 0$. Show that E_ε is open and dense in $C([0,1],\mathbb{R})$.

Hint: For any polynomial $p \in C([0,1],\mathbb{R})$, show that $p + \delta\cos(kx) \in E_\varepsilon$ for δ small and δk large.

(ii) Show that $E = \bigcap_{n\geq 1} E_{1/n}$ is dense in $C([0,1],\mathbb{R})$.

Hint: Use the Baire category theorem.

(iii) Show that if $f \in E$, then f has no derivative at any point.

◇ **1.7-4.** Prove the following: In a complete metric space (M,d) with no isolated points, no countable dense set is a G_δ-set.
Hint: Suppose $E = \{x_1, x_2, \dots\}$ is dense in M and is also a G_δ set, that is, $E = \bigcap_{n>0} V_n$ with V_n open, $n = 1, 2, \dots$. Conclude that V_n is dense in M. Let $W_n = V_n \backslash \{x_1, \dots, x_n\}$. Show that W_n is dense in M and that $\bigcap_{n>0} W_n = \varnothing$. This contradicts the Baire property.