Topological methods in Euclidean spaces

Gregory L. Naber

Topological methods in Euclidean spaces

Gregory L. Naber

Department of Mathematics
Pennsylvania State University

Cambridge University Press

Cambridge
London New York New Rochelle
Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1980

Printed in the United States of America
Typeset by Progressive Typographers, Inc., Emigsville, Pennsylvania
Printed and bound by Vail-Ballou Press, Inc., Binghamton, New York

Library of Congress Cataloging in Publication Data

Naber, Gregory, 1948–

Topological methods in Euclidean spaces.

Bibliography: p.

Includes index.

1. Topology I. Title.
QA611.N22 514 79-9225
ISBN 0 521 22746 1 hard covers
ISBN 0 521 29632 3 paperback

Preface

Topology evolved as an independent discipline in response to certain rather specific problems in classical analysis. Of course, it is characteristic of any fruitful branch of mathematics that the subject develop and take on a significance independent of those problems from which it arose. In the case of topology, however, this development has been so extensive and so rapid that, unfortunately, its origins and relations to other areas of mathematics are often lost sight of entirely, and, even less desirable, the essential unity of the subject itself is sacrificed to the demands of specialization.

It is the intention of this introduction to the methods of topology in Euclidean spaces to persuade students of mathematics, at the earliest possible point in their studies, that the evolution of topology from analysis and geometry was natural and, indeed, inevitable; that the most fruitful concepts and most interesting problems in the subject are still drawn from independent branches of mathematics; and that, underlying its sometimes overwhelming diversity of ideas and techniques, there is a fundamental unity of purpose. To this end an ambitious agenda of topics from point-set, algebraic, and differential topology has been included, although much of the material familiar from standard introductions to topology is omitted altogether. Indeed, metric space and topological space are never defined. Rather, we restrict attention exclusively to subspaces of Euclidean spaces where geometrical intuition remains strong so that we can avoid the tiresome technicalities inherent in axiomatic treatments. In this way it is possible to go rather far in the development of those techniques that are central to topology itself as well as its applications in other areas of mathematics and the sciences.

A very considerable emphasis has been placed on motivation, which we draw primarily from the student's background in differential equations, linear algebra, modern algebra, and advanced calculus. We assume this background to be rather strong and, in addition, that our readers are possessed of a healthy supply of that elusive quality known as "mathematical maturity." A great many arguments are left to the reader in the form of exercises embedded in the body of the text and no asterisk appears to designate those that are used in the sequel – they are all used and must be worked conscientiously. Of the 214 exercises in the text, 162 are of this variety, while 52 are included in Supplementary Exercises at the ends of chapters; the latter, although no less important, are not specifically called upon in the development. A Guide to Further Study has been included at the end of the book to suggest several directions in which to proceed to obtain a deeper understanding of various aspects of the subject.

Gregory L. Naber

October 1979

Contents

Chapter 1
Point-set topology of Euclidean spaces

1-1 Introduction

Geometry, in the broadest possible sense, emerged before the written and perhaps even the spoken word as a gradual accumulation of subconscious notions about physical space based on the ability of our species to recognize "forms" and compare shapes and sizes. Until approximately 600 B.C. the study of geometrical figures proceeded in the manner of an experimental science in which induction and empirical procedure were the tools of discovery. General properties and relationships were extracted from observations necessitated by the demands of daily life, the result being a rather formidable collection of "laboratory" results on areas, volumes, and relations between various figures. It was left to the Greeks to transform this vast array of empirical data into the very beautiful intellectual discipline we now know as Euclidean geometry. The transformation required approximately three centuries to complete and culminated, around 300 B.C., with the appearance of Euclid's *Elements*. It is difficult indeed to exaggerate the importance of this event for the development of mathematics. So decisive was the influence of Euclid that it was not until the seventeenth century that mathematicians found themselves capable of adopting essentially new attitudes toward their subject. Slowly, at times it seems unwillingly, mathematics began to free itself from the constraints imposed by the strict axiomatic method of the *Elements*. New and remarkably powerful concepts and techniques evolved that eventually led to an expanded and more lucid view of mathematics in general and geometry in particular. The object in this introductory section is to indicate how the subject of interest to us here (topology) arose as a branch of geometry in this expanded sense.

Perhaps the most fundamental concept of the earlier books of Euclid's *Elements* is that of congruence. Intuitively, two plane geometric figures (arbitrary subsets of the plane from our point of view) are congruent if they differ only in the position they occupy in the plane, that is, if they can be made to coincide by the application of some rigid motion in the plane. Somewhat more precisely, two figures F_1 and F_2 are said to be congruent if there is a mapping f of the plane onto itself that leaves invariant the distance between each pair of points (i.e., $d(f(p), f(q)) = d(p, q)$ for all p and q) and carries F_1 onto F_2 (i.e., $f(F_1) = F_2$). A map that preserves the distance between any pair of points is called an *isometry* and is the mathematical analog of a *rigid motion;* the study of congruent figures in the plane is, for this reason, often referred to as *plane Euclidean metric geometry*. If we construct an orthogonal Cartesian coordinate system in the plane, we can show that the isometries of the plane are precisely the maps $(x, y) \rightarrow (x', y')$, where

$$\begin{aligned} x' &= Ax + By + C \\ y' &= \pm(-Bx + Ay) + D, \end{aligned} \tag{1}$$

A, B, C, and D being real constants with $A^2 + B^2 = 1$ (see Gans, p. 65). Observe that the composition of any two isometries is again an isometry and that each isometry has an inverse that is again an isometry. Now, any collection of invertible mappings of a set S onto itself that is closed under the formation of compositions and inverses is called a *group of transformations* on S; the collection of all maps of the form (1) is therefore referred to as the *group of planar isometries*. From the point of view of plane Euclidean metric geometry the only properties of a geometric figure F that are of interest are those that are possessed by all figures congruent to F, that is, those properties that are invariant under the group of planar isometries. Since any map of the form (1) carries straight lines onto straight lines, the property of being a straight line is one such property. Similarly, the property of being a square or, more generally, a polygon of a particular type is invariant under the group of planar isometries, as is the property of being a conic of a particular type. The length of a line segment, area of a polygon, and eccentricity of a conic are likewise all invariants and are thus legitimate objects of study in plane Euclidean metric geometry.

Of course, the point of view of plane Euclidean metric geometry is not the only point of view. Indeed, in Book VI of the *Elements* itself, emphasis shifts from congruent to similar figures. Roughly speaking, two geometric figures are similar if they have the same shape, but not necessarily the same size. In order to formulate a more precise definition, let us refer to a map f of the plane onto itself under which each distance is multiplied by the same positive constant k (i.e., $d(f(p), f(q)) = k\, d(p, q)$ for all p and q) as a *similarity transformation* with *similarity ratio k*. It can be shown that, relative to an orthogonal Cartesian coordinate system, each such map has the form

$$\begin{aligned} x' &= ax + by + m \\ y' &= \pm(-bx + ay) + n, \end{aligned} \tag{2}$$

where $(a^2 + b^2)^{1/2} = k$ (see Gans, p. 77). Two plane geometric figures F_1 and F_2 are then said to be similar if there exists a similarity transformation of the plane onto itself that carries F_1 onto F_2. Again, the set of all similarity transformations is easily seen to be a transformation group, and we might reasonably define *plane Euclidean similarity geometry* as the study of those properties of geometric figures that are invariant under this group, that is, those properties that, if possessed by some figure, are necessarily possessed by all similar figures. Since any isometry is also a similarity transformation (with $k = 1$), any such property is necessarily an invariant of the group of planar isometries; but the converse is false since, for example, the length of a line segment and area of a polygon are not preserved by all similarity transformations.

At this point it is important to observe that, in each of the two geometries discussed thus far, certain properties of geometric figures were of interest while others were not. In plane Euclidean metric geometry we are interested in the shape and size of a given figure, but not in its position or orientation in the plane, while similarity geometry concerns itself only with the shape of the figure. Those properties that we deem important depend entirely on the particular sort of investigation we choose to carry out. Similarity transformations are, of course, capable of "distorting" geometric figures more than isometries, but this additional distortion causes no concern as long as we are interested only in properties that are not effected by such distortions. In other sorts of studies the permissible degree of distortion may be even greater. For example, in the mathematical analysis of perspective it was found that the "interesting" properties of a geometric figure are those that are invariant under a class of maps called *plane projective transformations*, each of which can be represented, relative to an orthogonal Cartesian coordinate system, in the following form (see Gans, p. 174):

$$(3)\qquad \begin{aligned} x' &= \frac{a_1x + a_2y + a_3}{c_1x + c_2y + c_3} \\ y' &= \frac{b_1x + b_2y + b_3}{c_1x + c_2y + c_3} \end{aligned} \qquad \text{where } \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} \neq 0.$$

The collection of all such maps can be shown to form a transformation group, and we define *plane projective geometry* as the study of those properties of geometric figures that are invariant under this group. Two figures are said to be "projectively equivalent" if there is a projective transformation that carries one onto the other. Since any similarity transformation is also a projective transformation, any invariant of the projective group is also an invariant of the similarity group. The converse, however, is false since projective maps are capable of greater distortions of geometric figures than are similarities. For example, two conics are always projectively equivalent, but they are similar only if they have the same eccentricity.

Needless to say, the approach we have taken here to these various geometrical studies is of relatively recent vintage. Indeed, it was Felix Klein, in his famous Erlanger Program of 1872, who first proposed that a "geometry" be defined quite generally as the study of those properties of a set S that are invariant under some specified group of transformations of S. Plane Euclidean metric, similarity, and projective geometries and their obvious generalizations to three and higher dimensional spaces all fit quite nicely into Klein's scheme, as did the various other offshoots of classical Euclidean geometry known at the time. Despite the fact that, during this century, our conception of geometry has expanded still further and now includes studies that cannot properly be considered "geometries" in the Kleinian sense, the influence of the ideas expounded in the Erlanger Program has been great indeed. Even in theoretical physics Klein's emphasis on the study of invariants of transformation groups has had a profound impact. The special theory of relativity, for example, is perhaps best regarded as the invariant theory of the so-called Lorentz group of transformations on Minkowski space.

Based on his appreciation of the importance of Riemann's work in complex function theory, Klein was also able to anticipate the rise of a new branch of geometry that would concern itself with those properties of a geometric figure that remain invariant when the figure is bent, stretched, shrunk or deformed in any way that does not create new points or fuse existing points. Such a deformation is accomplished by any bijective map that, roughly speaking, "sends nearby points to nearby points," that is, a continuous one. In dimension two, then, the relevant group of transformations is the collection of all one-to-one maps of the plane onto itself that are continuous and have continuous inverse; such maps are called *homeomorphisms* or *topological maps* of the plane. Consider, for example, the map f of the plane onto itself, which is given by $f(x, y) = (x, y^3)$. Now, f is continuous and has inverse $f^{-1}(x, y) = (x, y^{1/3})$ that is also continuous, so f is indeed a homeomorphism of the plane. What sort of properties of a plane geometric figure are preserved by f? Certainly, the property of being a straight line is not since, for example, the line given by the equation $y = x$ is mapped by f onto the curve $y = x^3$ (see Figure 1-1 (a)). Similarly, the property of being a conic is not invariant since the circle $x^2 + y^2 = 1$ is carried by f onto the locus of $x^2 + y^{2/3} = 1$, which is shown in Figure 1-1 (b).

Topological transformations are clearly capable of a very great deal of distortion. Indeed, virtually all of the properties the reader is accustomed to associating with plane geometric figures are destroyed by even the rela-

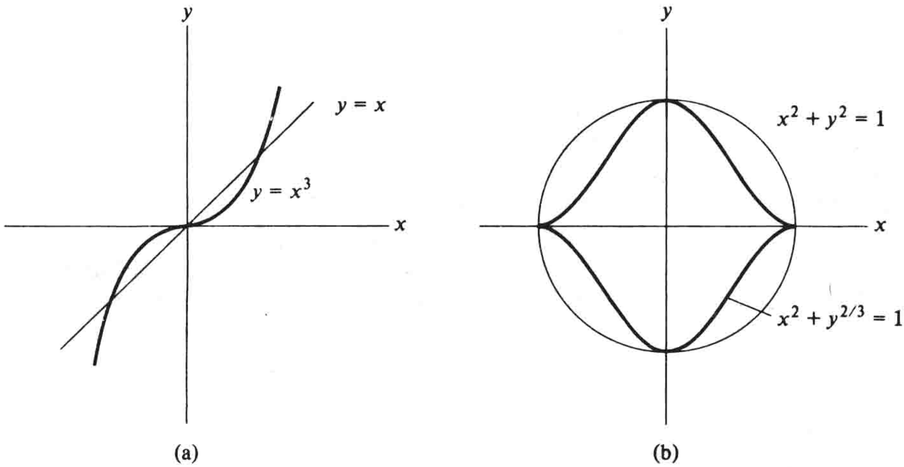


Figure 1-1

tively simple map f. Nevertheless, f does preserve a number of very important, albeit less obvious properties. For example, although a straight line need not be mapped by f onto another straight line, its image must also be "one-dimensional" and consist of one "connected" piece. The image of the circle $x^2 + y^2 = 1$, although not a conic, shares with the circle the property of being a "simple closed curve." Properties of plane geometric figures such as these that are invariant under the group of topological transformations of the plane are called *extrinsic topological properties*.

During the past one hundred years topology has outgrown its geometrical origins and today stands alongside analysis and algebra as one of the most fundamental branches of mathematics. Roughly speaking, topology might now be defined simply as the study of continuity. The approach we take here to this subject, while less general than it might be, is somewhat more general than that just outlined. We observe that the ambient space in which our geometrical figures are thought of as existing is, to a large extent, arbitrary (e.g., any plane figure can also be regarded as a subset of 3-space) and that, by insisting that the topological transformations be defined on this entire space, we have imposed rather unnatural restrictions on our study. We therefore choose to take a broader view of topological maps, allowing them to be defined on the given geometric figure itself without reference to the space in which it happens to be embedded, thus turning our attention from "extrinsic" to "intrinsic" topological properties, that is, properties of the figure itself that do not depend on the particular space in which it happens to reside.

1-2 Preliminaries

We shall denote by $\mathbf{R}$ the set of all real numbers and assume that the reader is familiar with the basic properties of this set (specifically, that under the usual operations $\mathbf{R}$ is a complete ordered field; see Apostol, Sections 1-1 through 1-9, or Buck, Appendix I). Recall that if $A_1, \ldots, A_n$ are arbitrary sets, then the Cartesian product $A_1 \times \cdots \times A_n$ is defined by $A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) : a_i \in A_i \text{ for } i = 1, \ldots, n\}$. Euclidean n-space $\mathbf{R}^n$ is thus the product $\mathbf{R}^n = \mathbf{R} \times \cdots \times \mathbf{R}$ (n factors). As usual we identify $\mathbf{R}^n \times \mathbf{R}^m$ and $\mathbf{R}^{n+m}$ by not distinguishing between the ordered pair $((a_1, \ldots, a_n), (b_1, \ldots, b_m))$ and the $(n + m)$-tuple $(a_1, \ldots, a_n, b_1, \ldots, b_m)$. Thus, if $A \subseteq \mathbf{R}^n$ and $B \subseteq \mathbf{R}^m$, we may regard $A \times B$ as a subset of $\mathbf{R}^{n+m}$.

If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are points of $\mathbf{R}^n$ and a is a real number, we define $x + y = (x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$ and $ax = a(x_1, \ldots, x_n) = (ax_1, \ldots, ax_n)$ and thus endow $\mathbf{R}^n$ with the structure of a real vector space of algebraic dimension n. (We assume the reader to be acquainted with basic linear algebra.) We denote by 0 the additive identity $(0, 0, \ldots, 0, 0)$ in $\mathbf{R}^n$ and let $e_1 = (1, 0, \ldots, 0, 0)$, $e_2 = (0, 1, \ldots, 0, 0)$, $\ldots$, $e_n = (0, 0, \ldots, 0, 1)$ be the standard basis vectors for $\mathbf{R}^n$. A map $S : \mathbf{R}^n \to \mathbf{R}^m$ is said to be *affine* if there is a $y_0 \in \mathbf{R}^m$ and a linear map $T : \mathbf{R}^n \to \mathbf{R}^m$ such that $S(x) = y_0 + T(x)$ for each x in $\mathbf{R}^n$. Since the range of a linear map is a linear subspace, the range of an affine map must be of the form $y_0 + V = \{y_0 + v : v \in V\}$ for some linear subspace V of $\mathbf{R}^m$; such a "translation" of a linear subspace of $\mathbf{R}^m$ is called an *affine subspace* or *hyperplane* in $\mathbf{R}^m$ (see Section 2-2 for more details).

If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are arbitrary points of $\mathbf{R}^n$, then their inner product (or dot product) is defined, as usual, by $x \cdot y = x_1y_1 + \cdots + x_ny_n$. The norm of x, denoted $\|x\|$, is then given by $\|x\| = (x \cdot x)^{1/2}$. Finally, the distance $d(x, y)$ between x and y is defined by $d(x, y) = \|y - x\|$. Standard properties of the inner product and norm (Apostol, Section 3-6, and Buck, Section 1.3) translate immediately to the following result on the "metric function" d.

Theorem 1-1. Let x, y, and z be arbitrary points in $\mathbf{R}^n$. Then
(a) $d(x, y) \geq 0$ and $d(x, y) = 0$ iff $x = y$.
(b) $d(x, y) = d(y, x)$.
(c) $d(x, y) \leq d(x, z) + d(z, y)$.

Now let x_0 be a point in $\mathbf{R}^n$ and $r > 0$ a real number. The *open ball* of radius r about x_0 is defined by $U_r(x_0) = \{x \in \mathbf{R}^n\colon d(x_0, x) < r\}$; the *closed ball* of radius r about x_0 is $B_r(x_0) = \{x \in \mathbf{R}^n\colon d(x_0, x) \leq r\}$. The ball $B_1(0) = \{x \in \mathbf{R}^n\colon \|x\| \leq 1\}$ is called the *closed n-ball* and denoted B^n, while the subset $S^{n-1} = \{x \in \mathbf{R}^n : \|x\| = 1\}$ is called the *$(n-1)$-sphere*. If A is an arbitrary subset of $\mathbf{R}^n$, the *diameter* of A is defined by diam $A = \sup\{d(x, y) : x, y \in A\}$ if $A \neq \emptyset$ and diam $\emptyset = 0$; A is said to be *bounded* if diam A is finite (this is the case iff $A \subseteq B_r(0)$ for some $r > 0$). If B is another subset of $\mathbf{R}^n$, then the *distance* between A and B is defined by $\text{dist}(A, B) = 0$ if $A = \emptyset$ or $B = \emptyset$ and $\text{dist}(A, B) = \inf\{d(x, y) : x \in A, y \in B\}$ if $A \neq \emptyset$ and $B \neq \emptyset$.

If x and y are any two points in $\mathbf{R}^n$, then the *open line segment* joining x and y is denoted (x, y) and defined by $(x, y) = \{tx + (1-t)y : 0 < t < 1\}$; the *closed line segment* joining x and y is $[x, y] = \{tx + (1-t)y : 0 \leq t \leq 1\}$. A subset A of $\mathbf{R}^n$ is *convex* if $[x, y] \subseteq A$ whenever x and y are in A.

Exercise 1-1. Let x_0 be a point in $\mathbf{R}^n$ and $r > 0$ a real number. Show that $U_r(x_0)$ and $B_r(x_0)$ are both convex.

Observe that any intersection of convex sets is also convex and that any subset A of $\mathbf{R}^n$ is contained in a convex set (e.g., $\mathbf{R}^n$ itself). We may therefore define the *convex hull* $H(A)$ of A as the intersection of all convex subsets of $\mathbf{R}^n$ containing A and be assured that $H(A)$ is convex for every A.

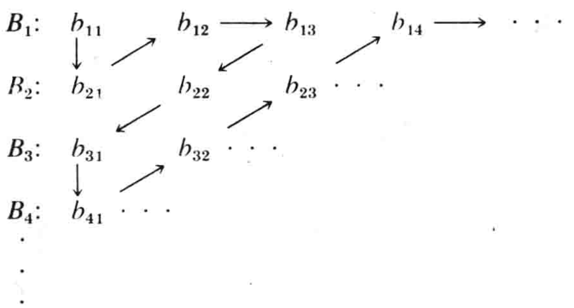
The final preliminary matter we must consider is the distinction, no doubt already familiar to the reader (see Apostol, Section 2-11, or Buck, p. 30), between countable and uncountable sets. Let us say that two nonempty sets S_1 and S_2 are *numerically equivalent*, or of the same *cardinality*, if there is a one-to-one mapping of S_1 onto S_2. A set is *finite* if it is either empty or numerically equivalent to $\{1, \ldots, n\}$ for some positive integer n. A set is *countably infinite* if it is numerically equivalent to the set $\mathbf{N} = \{1, 2, \ldots, n, \ldots\}$ of all positive integers. If a set is either finite or countably infinite we say that it is *countable*. Intuitively, a set is countable if it is either empty or if its elements can be listed in a (perhaps terminating) sequence. Finally, a set that is not countable is *uncountable*.

Lemma 1-2. Every subset A of a countable set S is countable.

Proof: Since every subset of a finite set is finite (and therefore countable), we may assume without loss of generality that S is countably infinite. Let $f : \mathbf{N} \to S$ be a bijection, where $\mathbf{N} = \{1, 2, \ldots, n, \ldots\}$, and define $g : \mathbf{N} \to \mathbf{N}$ inductively as follows: Let $g(1)$ be the least positive integer for which $f(g(1))$ is in A and assume that $g(1), \ldots, g(n-1)$ have been defined. Let $g(n)$ be the least positive integer greater than $g(n-1)$ such that $f(g(n))$ is in A. The composition $f \circ g : \mathbf{N} \to A$ is a bijection, so A is countable. Q.E.D.

Lemma 1-3. The union of countably many countable sets is countable.

Proof: By Lemma 1-2 it will suffice to show that the union of a countably infinite collection $\{A_1, A_2, \ldots, A_n, \ldots\}$ of countably infinite sets is countable. Define $B_1 = A_1$ and, for $n > 1$, let $B_n = A_n - \bigcup_{k=1}^{n-1} A_k$. Then each B_n is countable by Lemma 1-2, $B_i \cap B_j = \varnothing$ if $i \neq j$ and $\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$. Again by Lemma 1-2, we need only consider the case in which each B_n is countably infinite. Thus, we may enumerate the elements of each B_n as indicated:



Now define $f: \mathbf{N} \to \bigcup_{n=1}^{\infty} B_n$ by $f(1) = b_{11}$, $f(2) = b_{21}$, $f(3) = b_{12}$, $f(4) = b_{13}$, $f(5) = b_{22}, \ldots$, and so on, following the scheme indicated by the arrows. Then f is surjective. Moreover, since the B_n are disjoint, f is one-to-one and the result follows. Q.E.D.

Lemma 1-4. Let $S_1, \ldots, S_k$ be countable sets. Then $S_1 \times \cdots \times S_k$ is countable.

Exercise 1-2. Prove Lemma 1-4. Hint: Use Lemma 1-3 and induction.

Example 1-1. Countable and Uncountable Subsets of R. (a) The set Z of integers is countable. This follows immediately from the enumeration indicated:

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ 0 & 1 & -1 & 2 & -2 & 3 & -3 & \cdots \end{array}$$

(b) The set $\mathbf{Q}$ of rational numbers is countable. To see this, write each element of $\mathbf{Q}$ as m/n, where m and n are integers with no common factors and n is positive. The map that carries m/n to the ordered pair (m, n) thus maps $\mathbf{Q}$ bijectively onto a subset of $Z \times \mathbf{N}$. But $Z \times \mathbf{N}$ is countable by (a) and Lemma 1-4, so each of its subsets is countable by Lemma 1-2. It follows that **Q** is countable.

(c) The closed unit interval $I = [0, 1]$ is uncountable. To see this, let $f: \mathbf{N} \rightarrow [0, 1]$ be any one-to-one map. We show that f is not surjective. For each n in **N** let $0.a_{n1}a_{n2}a_{n3} \ldots$ be a decimal expansion for $f(n)$. Define a number $0.b_1b_2b_3 \ldots$ in $[0, 1]$ as follows: $b_k = 5$ if $a_{kk} \neq 5$ and $b_k = 7$ if $a_{kk} = 5$. Then $0.b_1b_2b_3 \ldots$ is not in the image of f since it has a unique decimal expansion that differs from $f(n)$ in the nth place for each n in **N**.

(d) If a and b are real numbers with $a < b$, then the interval $[a, b]$ is uncountable. Since the map $f: [a, b] \rightarrow [0, 1]$ defined by $f(x) = (x - a)/(b - a)$ is bijective, this follows immediately from (c).

(e) From (d) and Lemma 1-2 it follows that any subset of **R** that contains an interval $[a, b]$, where $a < b$, is uncountable. In particular, **R** itself is uncountable. However, an uncountable subset of **R** need not contain an interval, for example, the set **P** of irrational numbers is uncountable since **Q** is countable and $\mathbf{R} = \mathbf{Q} \cup \mathbf{P}$. Another example is constructed in (f).

(f) Recall that for each x in $[0, 1]$ there exists a sequence $s_1, s_2, s_3, \ldots$ with $s_i \in \{0, 1, 2\}$ for each i such that $x = \sum_{i=1}^{\infty} s_i/3^i$. (The procedure for determining the s_i will become clear shortly.) We shall write $x = {:}s_1s_2s_3 \ldots$ and call ${:}s_1s_2s_3 \ldots$ the *triadic expansion* of x. Some numbers have two such expansions. For example, :2000 . . . and :1222 . . . both represent the number 2/3 since $2/3 + 0/3^2 + 0/3^3 + \cdots = 2/3$ and $1/3 + 2/3^2 + 2/3^3 + \cdots = 1/3 + 2 \sum_{i=2}^{\infty} (1/3)^i = 1/3 + 2 [\sum_{i=0}^{\infty} (1/3)^i - 1 - (1/3)] = 1/3 + 2 [(3/2) - 1 - (1/3)] = 2/3$. This situation will occur only when one of the expansions repeats 0's and the other repeats 2's from some point on. We define the *Cantor set* C to be the set of all those x's in $[0, 1]$ that have a triadic expansion in which the digit 1 does not occur. This set has a simple geometrical interpretation that we obtain as follows: Let F_1 denote the closed interval $[0, 1]$. Delete the open interval $(\frac{1}{3}, \frac{2}{3})$ from F_1 to obtain the set $F_2 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ (see Figure 1-2). Note that the "middle third" $(\frac{1}{3}, \frac{2}{3})$ of $[0, 1]$ consists precisely of those x's in $[0, 1]$ whose triadic expansions must have a 1 in the first digit. Thus, F_2 consists of those x's in $[0, 1]$ that have a triadic expansion with $s_1 \neq 1$. Now delete from F_2 the middle thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ of each of the two closed intervals $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ to obtain the set $F_3 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$ (see Figure 1-2).

Observe that $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ consist precisely of those x's in $[0, 1]$ whose triadic expansions must have a 1 in the second digit, but not in the first. Thus, F_3 consists of those x's in $[0, 1]$ that have a triadic expansion ${:}s_1s_2s_3 \ldots$ with $s_1 \neq 1$ and $s_2 \neq 1$. We now continue this process inductively, at each stage deleting the open middle third of each closed interval

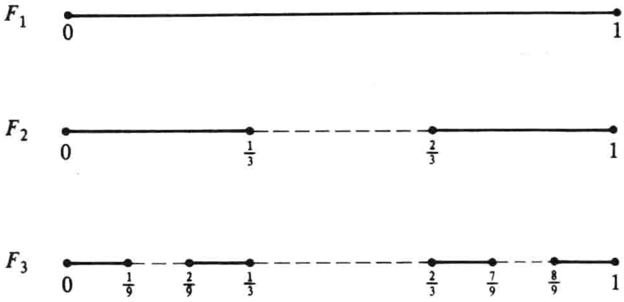


Figure 1-2

remaining from the previous stage. We therefore obtain a descending sequence $F_1 \supseteq F_2 \supseteq F_3 \supseteq \ldots$ of subsets of [0, 1], each of which is a finite union of disjoint closed intervals (e.g., F_{25} consists of 16,777,216 such intervals). The Cantor set C is then $\bigcap_{n=1}^{\infty} F_n$.

Remark: The sum of the lengths of all the open intervals removed from [0, 1] to form C is 1 since $\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = (\frac{1}{3}) \sum_{n=0}^{\infty} (\frac{2}{3})^n = (\frac{1}{3})(3) = 1$. It follows that C cannot contain an interval.

Finally, we show that C is uncountable by exhibiting a bijective map of [0, 1) onto C. For each x in [0, 1) let $x = {:}b_1 b_2 b_3 \ldots$ be a binary expansion for x. Thus, each b_i is either 0 or 1 and $x = \sum_{i=1}^{\infty} b_i/2^i$. Let $s_i = 2b_i$ for each i, and let $f(x)$ be the point in [0, 1] whose triadic expansion is ${:}s_1 s_2 s_3 \ldots$. Then f is one-to-one, $f(x)$ is in C for each x in [0, 1), and, moreover, every element of C is the image under f of some x in [0, 1) so f is surjective. It follows from (e) that C is uncountable.

1-3 Open sets, closed sets, and continuity

You will recall (Apostol, Definition 3-24, or Buck, Section 1.5) that a subset U of $\mathbf{R}^n$ is said to be *open in* $\mathbf{R}^n$ if, for each $x_0 \in U$, there is an $r > 0$ such that the open ball $U_r(x_0)$ is contained entirely in U.

Theorem 1-5. (a) $\varnothing$ and $\mathbf{R}^n$ are open in $\mathbf{R}^n$.
(b) Any union of open subsets of $\mathbf{R}^n$ is open in $\mathbf{R}^n$.
(c) Any finite intersection of open subsets of $\mathbf{R}^n$ is open in $\mathbf{R}^n$.

Exercise 1-3. Prove Theorem 1-5. Q.E.D.

A set C in $\mathbf{R}^n$ is *closed in* $\mathbf{R}^n$ if its complement $\mathbf{R}^n - C$ is open in $\mathbf{R}^n$ (see Apostol, Theorem 3-31, or Buck, Section 1.5).