This Book Is Part of the
Allyn and Bacon Series in Advanced Mathematics
Consulting Editor: Irving Kaplansky
University of Chicago

Other books in this series are

The Theory of Groups: An Introduction, Second Edition

Joseph J. Rotman

TOPOLOGY

James Dugundji

Professor of Mathematics
University of Southern California
Los Angeles

ALLYN AND BACON, INC.
BOSTON • LONDON • SYDNEY • TORONTO

Library of Congress Catalog Card Number: 66-10940

Printed in the United States of America

ISBN: 0-205-00271-4

16 15 14 86 85 84 83 82 81

Preface

In this book, my aim is to provide the reader with a foundation in general topology that will adequately prepare him for further work in a broad variety of mathematical disciplines. The arrangement of the material is such that the book can also serve as a reference for the more advanced mathematician.

The reader is assumed to have a background of at least one semester of rigorous analysis; for such persons, the treatment herein is self-contained and the material can easily be covered in a two-semester course.

Chapters I–II provide a short introduction to the axiomatic foundations of set theory. Chapters III–VI are devoted to general topological structures; the emphasis is on the mapping, and an extended treatment of identifications is given. Separation axioms are introduced in Chapter VII; after that, with only very few and explicitly noted exceptions, all spaces in the book are assumed to be Hausdorff. By imposing increasingly more severe conditions on the topology, the development proceeds down the hierarchy of topological spaces to the metric spaces; then convergence, compactness, function spaces, and completeness are taken up. The use of homotopy methods in general topology is started in Chapter XV. In the usual manner, these methods lead to "elementary" proofs of the classical results in Euclidean n-space, such as Brouwer's fixed-point and domain-invariance theorems; a proof of the complete Jordan curve theorem is also given. After discussing the homotopy classification of spaces and some of its features, the entry of algebra into topology (via the fundamental and higher homotopy groups) is presented in Chapter XIX; this chapter is part of some work that was done jointly with W. Hurewicz. The last chapter is devoted to the covering homotopy theorem in fiber spaces and illustrates an interplay between many of the concepts discussed in the book. Two appendices, one on linear topological spaces and the other on limit spaces, are included.

Nearly every definition is followed by examples illustrating the use of the abstract concept in some fairly concrete situations. This device makes the book suitable for self-study. It also enables the instructor who uses the book as a text to proceed rapidly to those parts of the subject that he deems of greater importance.

Remarks, in small type, call attention either to further developments, or to direct applications in other branches of mathematics.

The problems, which are given at the end of each chapter, can all be solved by the methods developed in the book. Moreover, no proof in the text relies on the solution of any problem. Some of the problems are routine. Others are important theorems that complement the material in the text; these are accompanied by hints for their solution.

The notation of symbolic logic used throughout the book is given immediately after the table of contents.

I wish to thank E. A. Michael and Ky Fan, who read the original manuscript, for their valuable suggestions; and H.-J. Groh and P. A. White, for their help with the proofreading. I am particularly indebted to H. Salzmann for his constant willingness to discuss points of detail and content: his imaginative and penetrating criticisms and suggestions have led to many improvements.

I also wish to thank Mrs. L. Syfritt, for her tidy and meticulous typing work; and the members of Allyn and Bacon who were involved with this book, for their patience and cooperation. Finally, I gratefully acknowledge the support given me by the National Science Foundation during the period that this book was being written.

James Dugundji
The University of Southern California

Contents

III. Topological Spaces 62

IV. Cartesian Products 98

V. Connectedness 107

VI. Identification Topology; Weak Topology 120

Basic Notation

Throughout this book, the notation of symbolic logic will be used to shorten statements. If p and q are propositions, then:

$p \vee q$ (read: p or q) denotes the disjunction of p and q. The assertion "$p \vee q$" is true whenever at least one of p, q is true.

$p \wedge q$ (read: p and q) denotes the conjunction of p and q. The assertion "$p \wedge q$" is true only in case both p and q are true.

$\neg q$ (read: not q) denotes the negation of q. The assertion "$\neg q$" is true only if q is false.

$p \Rightarrow q$ is read: p implies q. By definition, "$p \Rightarrow q$" denotes "$(\neg p) \vee q$". In particular, the statement "$p \Rightarrow q$" is true if and only if the statement "$(\neg q) \Rightarrow (\neg p)$" is true.

$p \Leftrightarrow q$ is read: p is logically equivalent to q. By definition, "$p \Leftrightarrow q$" denotes "$(p \Rightarrow q) \wedge (q \Rightarrow p)$".

An expression $p(x)$ that becomes a proposition whenever values from a specified domain of discourse are substituted for x is called a propositional function or, equivalently, a condition on x; and p is called a property, or predicate. The assertion "y has property p" means that "$p(y)$" is true. Thus, if $p(x)$ is the propositional function "x is an integer," then p denotes the property "is an integer," and "$p(2)$" is true whereas "$p(1/2)$" is false.

The quantifier "there exists" is denoted by $\exists$, and the quantifier "for each" is denoted by $\forall$. The assertion "$\forall x \; \exists y \; \forall z : p(x, y, z)$" reads "For each x there exists a y such that for each z, $p(x, y, z)$ is true"; its negation is obtained mechanically by changing the sense of each quantifier (preserving the given order of the variables!) and negating the proposition: thus, "$\exists x \; \forall y \; \exists z : \neg p(x, y, z)$".

Elementary Set Theory

I

In this chapter, we shall first discuss a part of set theory informally and then give an axiomatization adequate to support both this and the subsequent development.

1. Sets

Intuitively, we think of a set as something made up by all the objects that satisfy some given condition, such as the set of prime numbers, the set of points on a line, or the set of objects named in a given list. The objects making up the set are called the elements, or members, of the set and may themselves be sets, as in the set of all lines in the plane.

Rigorously, the word *set* is an undefined term in mathematics, so that definite axioms are required to govern the use of this term; one such system is given in **8.** Although we shall deal with sets on an intuitive basis until then, whenever we apply the label *set* to something, we shall later prove this usage to have been formally justified.

The membership relation is denoted by $\in$ and sets are generally indicated by capital letters: "$a \in A$" is read "a belongs to (is a member, element, point of) the set A"; $\neg(a \in A)$ is written $a \bar{\in} A$. The notation $a = b$ will mean that the objects a and b are the same, and $a \neq b$ denotes $\neg(a = b)$. If A, B are sets, then "$A = B$" will indicate that A and B

have the same elements; that is, $\forall\, x$: $(x \in A) \Leftrightarrow (x \in B)$; $\neg(A = B)$ is written $A \neq B$.

$A \subset B$ (or $B \supset A$), read "A is a subset of (is contained in) B," signifies that each element of A is an element of B, that is, $\forall\, x$: $(x \in A) \Rightarrow (x \in B)$; equality is *not* excluded—we call A a *proper* subset of B whenever $(A \subset B) \wedge (A \neq B)$. The following statements are evident:

1.1 $A \subset A$ for each set A.

1.2 If $A \subset B$ and $B \subset C$, then $A \subset C$ (that is, "$\subset$" is transitive).

1.3 $A = B$ if and only if both $A \subset B$ and $B \subset A$.

Of these, **1.3** is very important: the equality of two sets is usually proved by showing each of the two inclusions valid.

The axioms of set theory allow only two methods for forming subsets of a given set. One of these is by appeal to the axiom of choice, and will be discussed later. The other is by use of the following principle: If A is a set and p is a property that each element of A either has or does not have, then all the $x \in A$ having the property p form a set. This subset of A is denoted by $\{x \in A \mid p(x)\}$; it is uniquely determined by the property p. Clearly, $\{x \in A \mid p(x)\} \subset \{x \in A \mid q(x)\}$ if and only if $\forall\, x \in A$: $p(x) \Rightarrow q(x)$; thus two properties determine the same subset of A whenever each object in A having one of them also has the other.

Ex. 1 If A is the real line, the closed unit interval is $\{x \in A \mid 0 \leqslant x \leqslant 1\}$.

Ex. 2 If A is the real line, $\{x \in A \mid x^2 = 1\} = \{x \in A \mid x^4 = 1\}$ even though the defining properties are different. Note that if A were the set of complex numbers, these two properties would not determine the same subset.

Ex. 3 For each set A, $\{x \in A \mid x = x\} = \{x \in A \mid x \in A\} = A$.

For each set A, the null subset $\varnothing_A \subset A$ is $\{x \in A \mid x \neq x\}$; it has no members, since each $x \in A$ satisfies $x = x$.

1.4 All null subsets are equal. Thus there is one and only one null set, $\varnothing$, and it is contained in every set: $\varnothing \subset A$ for every set A.

Proof: Let A, B be any two sets. If $\varnothing_A \subset \varnothing_B$ were false, there would be at least one element a in $\varnothing_A$ not in $\varnothing_B$; in particular, we would then have an $a \in A$ such that $a \neq a$, and this is impossible. In the same way, $\varnothing_B \subset \varnothing_A$; therefore, by **1.3**, $\varnothing_A = \varnothing_B$ and all null sets are equal.

If A has only the finitely many members $a_1, \cdots, a_n$, we say that A is a finite set and write $A = \{a_1, \cdots, a_n\}$. Observe that for any set A (A may even be the null set), $\{A\}$ is a set containing exactly *one* element, and $A \in \{A\}$.

Because of their frequent occurrence, the following special symbols will be used for certain sets:

Z = set of all positive and negative integers, and zero.
$Z^+ = \{n \in Z \mid n > 0\}$.
$N = \{n \in Z \mid n \geqslant 0\}$.
Q = set of all rational numbers, and $Q^+ = \{q \in Q \mid q > 0\}$.
E^1 = Euclidean line. The closed interval $\{x \in E^1 \mid a \leqslant x \leqslant b\}$ is written $[a, b]$; the open interval $\{x \in E^1 \mid a < x < b\} =]a, b[$, and $]a, b]$ ($[a, b[$) is the left (right) half-open interval. The unit interval $[0, 1]$ is denoted simply by I.
E^n = Euclidean n-space. We shall use vector notation: If

$$x = (x_1, \cdots, x_n) \in E^n \text{ and } y = (y_1, \cdots, y_n) \in E^n,$$

then $x + y = (x_1 + y_1, \cdots, x_n + y_n)$, and for any real λ, $\lambda x = (\lambda x_1, \cdots, \lambda x_n)$. We write $|x| = \sqrt{\sum_1^n x_i^2}$, so that the distance between x and y is $|x - y|$. The unit n-cube I^n is the subset $\{(x_1, \cdots, x_n) \in E^n \mid 0 \leqslant x_i \leqslant 1, i = 1, \cdots, n\}$.
S^n = the unit n-sphere in $E^{n+1} = \{x \in E^{n+1} \mid |x| = 1\}$. Thus S^1 is the circumference of the unit circle in E^2; observe that $S^0 = \{x \in E^1 \mid x = \pm 1\}$.
V^n = the unit n-ball in $E^n = \{x \in E^n \mid |x| \leqslant 1\}$; its boundary is S^{n-1}.

2. Boolean Algebra

The elementary set-theoretic operations are defined in this section, and a list of standard formulas, which are convenient in symbolic work, will be indicated.

2.1 Definition Let Γ be a given set, and A, B two subsets. The *union*, $A \cup B$, of A and B is $\{x \in \Gamma \mid x$ belongs to at least one of $A, B\}$. The *intersection*, $A \cap B$, of A and B is $\{x \in \Gamma \mid x$ belongs to both A and $B\}$.

Ex. 1 Let $A = [0, 1] \subset E^1$ and $B =]1, 2]$; then $A \cup B = [0, 2]$ and $A \cap B = \varnothing$. Note that it is precisely because of the null set that the intersection of two sets is always a well-defined set.

Ex. 2 For every set A, $A \cup \varnothing = A$ and $A \cap \varnothing = \varnothing$.

According to the definition, a necessary and sufficient condition for two sets A, B to have elements in common is that $A \cap B \neq \varnothing$; if $A \cap B = \varnothing$, the sets A and B are called *disjoint*. The following two statements are immediate consequences of **2.1**:

2.2 For any two sets A, B, always $A \cap B \subset A \subset A \cup B$.

2.3 If $A \subset C$ and $B \subset D$, then $A \cup B \subset C \cup D$ and $A \cap B \subset C \cap D$.

The formal properties of the operations $\cup$ and $\cap$ are given in

2.4 Theorem Each of the operations $\cup$ and $\cap$ is:

(1). Idempotent: $\forall A\colon A \cup A = A = A \cap A$.
(2). Associative: $A \cup (B \cup C) = (A \cup B) \cup C$ and

$$A \cap (B \cap C) = (A \cap B) \cap C.$$

(3). Commutative: $A \cup B = B \cup A$ and $A \cap B = B \cap A$.

Furthermore, $\cap$ distributes over $\cup$ and $\cup$ distributes over $\cap$:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Proof: Verification of (1)–(3) is trivial. To give an example of a set-theoretic proof, we establish distributivity of $\cap$ over $\cup$. Using **1.3**, we find that the proof decomposes into two parts:

(a). Left side $\subset$ right side:

$$\begin{aligned} x \in A \cap (B \cup C) &\Rightarrow (x \in A) \wedge [(x \in B) \vee (x \in C)] \\ &\Rightarrow (x \in A \cap B) \vee (x \in A \cap C) \\ &\Rightarrow x \in (A \cap B) \cup (A \cap C). \end{aligned}$$

(b). Right side $\subset$ left side: All implications in the above string are reversible.

Because of associativity, we can designate $A \cup (B \cup C)$ simply by $A \cup B \cup C$. Similarly, a union (or intersection) of four sets, say $(A \cup B) \cup (D \cup C)$, can be written $A \cup B \cup C \cup D$ because, by associativity, the distribution of parentheses is irrelevant, and by commutativity, the order of the terms plays no role. By induction, the same remarks apply to the union (or intersection) of any finite number of sets. The union of n sets, $A_1, \cdots, A_n$, is written $\bigcup_1^n A_i$, and their intersection is $\bigcap_1^n A_i$.

The relation between $\cap$, $\cup$, and $\subset$ is given in

2.5 The statements (1) $A \subset B$, (2) $A = A \cap B$, and (3) $B = A \cup B$ are equivalent.

Proof: As an example in the use of **2.2–2.4**, we prove only (1) $\Leftrightarrow$ (2), leaving the rest for the reader. If (1), then we have

$$A = A \cap A \subset A \cap B \subset A,$$

which proves (2). Conversely, if (2), then $A = A \cap B \subset B$, establishing (1).

2.6 Definition The difference $A - B$ of two sets is $\{x \in A \mid x \notin B\}$.

Ex. 1 If $A = [0, 1]$ and $B =]1, 2]$, then $A - B = A$.
Ex. 2 $A - \varnothing = A$ and $A - B = A - (A \cap B)$.

The difference operation does not have formal properties so simple as those of $\cup$ and $\cap$: for example, since $(A \cup A) - A \neq A \cup (A - A)$, the location of parentheses in $A \cup A - A$ is important. To construct a suitable calculus involving the difference operation, we introduce the complementation operation:

2.7 Definition If $B \subset A$, the complement $\mathscr{C}_A B$ of B with respect to A is $A - B$.

Note that the complementation operation is defined *only* when one set is contained in the other, whereas the difference operation does not have such a restriction. The relation between **2.6** and **2.7** is given in

2.8 For any two sets A, B, if the complement is taken with respect to any set E containing $A \cup B$, then $A - B = A \cap \mathscr{C}_E B$.

Proof: Since $A \cup B \subset E$, we have

$$\begin{aligned} A - B &= \{x \in E \mid (x \in A) \wedge (x \notin B)\} \\ &= \{x \in E \mid x \in A\} \cap \{x \in E \mid x \notin B\} = A \cap \mathscr{C}_E B. \end{aligned}$$

The following properties of complementation are immediate:

2.9 If E is any set containing $A \cup B$, then:

(1). $A \cap \mathscr{C}_E A = \varnothing$, $\quad A \cup \mathscr{C}_E A = E$.
(2). $\mathscr{C}_E(\mathscr{C}_E A) = A$.
(3). $\mathscr{C}_E \varnothing = E$, $\quad \mathscr{C}_E E = \varnothing$.
(4). $A \subset B$ if and only if $\mathscr{C}_E B \subset \mathscr{C}_E A$.

We write $\mathscr{C}$ instead of $\mathscr{C}_E$ whenever the set E has been specified and is to be kept fixed. The basic relation between $\cup$, $\cap$, and $\mathscr{C}$ is

2.10 Theorem (De Morgan) If complements be taken with respect to any set E containing $A \cup B$, then:

(1). $\mathscr{C}(A \cup B) = (\mathscr{C}A) \cap (\mathscr{C}B)$.
(2). $\mathscr{C}(A \cap B) = (\mathscr{C}A) \cup (\mathscr{C}B)$.

Proof: We have

$$x \in \mathscr{C}(A \cup B) \Leftrightarrow x \bar{\in} A \cup B \Leftrightarrow (x \bar{\in} A) \wedge (x \bar{\in} B)$$
$$\Leftrightarrow x \in (\mathscr{C}A) \cap (\mathscr{C}B)$$

and this establishes (1). The proof of (2) is similar; it can, however, be deduced from (1) by "complements":

$$\mathscr{C}[\mathscr{C}A \cup \mathscr{C}B] = \mathscr{C}\mathscr{C}A \cap \mathscr{C}\mathscr{C}B = A \cap B,$$

and apply **2.9**(2).

2.11 Remark The formulas in **2.2–2.5** and **2.8–2.10** comprise a short list of results useful for formal calculations with sets; the role of **2.8** is to change differences to complements. As examples, we prove:

a. $(A - B) \cup (B - A) = (A \cup B) - (A \cap B)$. For taking complements with respect to $E = A \cup B$, and using **2.4**, **2.5**, **2.9**, **2.10**, gives

$$\begin{aligned}(A - B) \cup (B - A) & \\ &= (A \cap \mathscr{C}B) \cup (B \cap \mathscr{C}A) \\ &= (A \cup B) \cap (A \cup \mathscr{C}A) \cap (\mathscr{C}B \cup B) \cap (\mathscr{C}B \cup \mathscr{C}A) \\ &= A \cup B \cap \mathscr{C}(B \cap A) = (A \cup B) - (A \cap B).\end{aligned}$$

b. If $A \cup X = E$ and $A \cap X = \varnothing$, then $X = \mathscr{C}_E A$. For,

$$\begin{aligned}X &= E \cap X = (A \cup \mathscr{C}_E A) \cap X = X \cap \mathscr{C}_E A \\ &= (A \cap \mathscr{C}_E A) \cup (X \cap \mathscr{C}_E A) = (A \cup X) \cap \mathscr{C}_E A \\ &= E \cap \mathscr{C}_E A = \mathscr{C}_E A.\end{aligned}$$

2.12 Remark A Boolean algebra B is a set together with two binary operations $+$, $\cdot$, and a unary operation $'$, satisfying the following axioms:

1. Each operation $+$, $\cdot$ is commutative.
2. There exist elements 0, 1 with $a + 0 = a$, $a \cdot 1 = a$ for every $a \in B$.
3. The distributive laws

$$a \cdot (b + c) = a \cdot b + a \cdot c,$$
$$a + (b \cdot c) = (a + b) \cdot (a + c)$$

hold.

4. $a \cdot a' = 0$ and $a + a' = 1$ for each $a \in B$.

(It is not necessary to postulate associativity of the $+$ and $\cdot$ operations; this is a consequence of the axioms.) The collection of all subsets of a fixed set E, with $+$, $\cdot$, $'$, 0, 1 interpreted as $\cup$, $\cap$, $\mathscr{C}_E$, $\varnothing$, E, respectively, evidently forms a Boolean algebra. By observing that the systematic interchange of $+$ with $\cdot$ and 0 with 1 in the axioms simply gives the same set of axioms, we obtain the duality principle: For each formula true in a Boolean algebra, there is a true "dual" formula obtained by replacing each occurrence of $+$, $\cdot$, 0, 1 with $\cdot$, $+$, 1, 0, respectively. This is the "method of complements"; observe that each one of De Morgan's rules follows from the other by duality.

The theory of Boolean algebras is equivalent to that of commutative rings with unit, in which each element is idempotent, $a \cdot a = a$, (that is, Boolean rings). Indeed, given a Boolean algebra B, define operations $\oplus$, $\bigcirc$ by $a \oplus b = (a \cdot b') + (a' \cdot b)$, $a \bigcirc b = a \cdot b$ [*cf.* **2.11**(a)]; with $\oplus$, $\bigcirc$, B is a Boolean ring, $r(B)$. Conversely, from a Boolean ring R, one obtains a Boolean algebra $b(R)$ by using the operations $a + b = a \oplus b - (a \bigcirc b)$, $a \cdot b = a \bigcirc b$ in R. These transformations are inverses in that $b[r(B)] = B$ and $r[b(R)] = R$.

3. Cartesian Product

The cartesian product is one of the most important constructions of set theory: it enables us to express many concepts in terms of sets.

With each two objects a, b, there corresponds a new object (a, b), called their ordered pair. Ordered pairs are subject to the one condition: $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$; in particular, $(a, b) = (b, a)$ if and only if $a = b$. The first (second) element of an ordered pair is called its first (second) coordinate.

3.1 Remark The concept of an ordered pair can be expressed in terms of sets by defining $(a, b) = \{\{a\}, \{a, b\}\}$; the reader can easily show that $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$ if and only if $a = c$ and $b = d$. For the sequel, we need only know that ordered pairs are uniquely determined by their first and second coordinates; the means for accomplishing this are immaterial.

3.2 Definition Let A, B be two sets, distinct or not. Their cartesian product, $A \times B$, is the set of all ordered couples $\{(a, b) \mid a \in A, b \in B\}$.

Ex. 1 Let $A = B = Z$; then $A \times B$ can be represented as the set of lattice points in E^2.

Ex. 2 Let $A = B = E^1$; then $A \times B$, being the set of all ordered couples of reals, can be represented as the points in E^2.

Ex. 3 Let $A = S^1$, $B = [0, 1]$. $A \times B$ can be regarded as those points of E^3 lying on a cylinder of altitude 1. Similarly, $S^1 \times S^1$ is represented as the points on a torus.

We have the basic

3.3 $A \times B = \varnothing \Leftrightarrow [A = \varnothing] \vee [B = \varnothing]$.

Proof: If $A \times B \neq \varnothing$, we can display some $(a, b) \in A \times B$; then $a \in A$, $b \in B$ shows $A \neq \varnothing$ and $B \neq \varnothing$. Conversely, if $A \neq \varnothing$ and $B \neq \varnothing$, we can display some $a \in A$, $b \in B$; by displaying (a, b), we show $A \times B \neq \varnothing$.

In similar fashion, the reader can show

3.4 If $C \times D \neq \varnothing$, then $C \times D \subset A \times B$ if and only if

$$[C \subset A] \wedge [D \subset B].$$

It follows at once from this and **1.3** that for nonempty sets A, B, $A \times B = B \times A$ if and only if $A = B$; the operation $A \times B$ is therefore not commutative.

The relation of $\times$ to $\cup$ and $\cap$ is summarized in the following trivial

3.5 Theorem $\times$ distributes over $\cup$, $\cap$, and $-$:

$$\begin{aligned} A \times (B \cup C) &= A \times B \cup A \times C, \\ A \times (B \cap C) &= A \times B \cap A \times C, \\ A \times (B - C) &= A \times B - A \times C. \end{aligned}$$

The cartesian product of three sets A, B, C is defined by $A \times B \times C = (A \times B) \times C$, and that of n sets by induction: $A_1 \times \cdots \times A_n = (A_1 \times \cdots \times A_{n-1}) \times A_n$; an element of $A_1 \times \cdots \times A_n$ is written $(a_1, \cdots, a_n)$, and a_i is called its ith coordinate. One should note that the operation "$\times$" is not associative: $(A \times B) \times C \neq A \times (B \times C)$ in general.

4. Families of Sets

If to each element α of some set $\mathscr{A} \neq \varnothing$ there corresponds a set A_α, then the collection of sets $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is called a *family* of sets, and $\mathscr{A}$ is called an indexing set for the family. It is not required that sets with distinct indices be different. Observe that any set $\mathscr{F}$ of sets can be converted to a family of sets by "self-indexing": one uses the set $\mathscr{F}$ itself as indexing set and assigns to each member of the set $\mathscr{F}$ the set it represents. In this section we extend the notions of union and intersection to families of sets; it should be noted that this is not done by any limiting process, but rather by independent definitions that reduce to the previous ones whenever the family is finite.

4.1 Definition Let Γ be a given set, and $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ a family of subsets of Γ. The union $\bigcup_\alpha A_\alpha$ of this family is the set

$$\{x \in \Gamma \mid \exists\ \alpha \in \mathscr{A} : x \in A_\alpha\},$$

and the intersection $\bigcap_\alpha A_\alpha$ is the set

$$\{x \in \Gamma \mid \forall\ \alpha \in \mathscr{A} : x \in A_\alpha\}.$$

We frequently denote $\bigcup_\alpha A_\alpha$ by $\bigcup_{\alpha \in \mathscr{A}} A_\alpha$ and by $\cup\{A_\alpha \mid \alpha \in \mathscr{A}\}$; similarly for $\bigcap_\alpha A_\alpha$.

Ex. 1 For any set X, $X = \cup\{\{x\} \mid x \in X\}$.

Ex. 2 In the case of infinite intersections and unions, some nonintuitive situations can arise:

a. Let $A_k = \{n \in Z \mid n \geqslant k\}$, $k = 1, 2, \cdots$. Note that $A_1 \supset A_2 \supset \cdots$ is a descending sequence of nonempty sets, each containing its successor; yet $\bigcap_n A_n = \varnothing$.

b. For each $n \in Z^+$, let $A_n = [0, 1 - 2^{-n}]$ and $B_n = [0, 1 - 3^{-n}]$. Note that each A_n is a proper subset of B_n, and that no A_n equals any B_k. However, $\bigcup_n A_n = \bigcup_n B_n = [0, 1[$.

From the definitions, it is evident that the union and intersection of a family of sets does not depend on how the family is indexed: This is expressed by saying $\bigcup_\alpha$, $\bigcap_\alpha$ are unrestrictedly commutative. Similarly, if the indexing set itself is expressed as a union, $\mathscr{A} = \bigcup_{\lambda \in \mathscr{L}} \mathscr{A}_\lambda$, it is immediate that $\bigcup\{A_\alpha \mid \alpha \in \mathscr{A}\} = \bigcup_{\lambda \in \mathscr{L}} [\cup\{A_\alpha \mid \alpha \in \mathscr{A}_\lambda\}]$ and

$$\bigcap\{A_\alpha \mid \alpha \in \mathscr{A}\} = \bigcap_{\lambda \in \mathscr{L}} [\cap\{A_\alpha \mid \alpha \in \mathscr{A}_\lambda\}],$$

which expresses that $\bigcup_\alpha$, $\bigcap_\alpha$ are unrestrictedly associative.

For these extended notions,

4.2 Theorem (1). $\bigcup_\alpha$ distributes over $\cap$ and $\bigcap_\alpha$ distributes over $\cup$:

$$[\bigcup\{A_\alpha \mid \alpha \in \mathscr{A}\}] \cap [\bigcup\{B_\beta \mid \beta \in \mathscr{B}\}] = \bigcup\{A_\alpha \cap B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}.$$

$$[\bigcap\{A_\alpha \mid \alpha \in \mathscr{A}\}] \cup [\bigcap\{B_\beta \mid \beta \in \mathscr{B}\}] = \bigcap\{A_\alpha \cup B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}.$$

(2). If complements be taken with respect to Γ, then

$$\mathscr{C}(\bigcup_\alpha A_\alpha) = \bigcap_\alpha \mathscr{C}A_\alpha \text{ and } \mathscr{C}(\bigcap_\alpha A_\alpha) = \bigcup_\alpha \mathscr{C}A_\alpha.$$

(3). $\bigcup_\alpha$ and $\bigcap_\alpha$ distribute over the cartesian product:

$$\bigcup\{A_\alpha \mid \alpha \in \mathscr{A}\} \times \bigcup\{B_\beta \mid \beta \in \mathscr{B}\} = \bigcup\{A_\alpha \times B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}.$$

$$\bigcap\{A_\alpha \mid \alpha \in \mathscr{A}\} \times \bigcap\{B_\beta \mid \beta \in \mathscr{B}\} = \bigcap\{A_\alpha \times B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}.$$

The proofs are all trivial and are therefore omitted.

The general distributive law of $\bigcap_\alpha$ over $\bigcup_\beta$ is more delicate, and will be considered in **9.7** and **9.8**.

So far, we have assumed that $\mathscr{A} \neq \varnothing$; there are formal advantages in allowing the indexing set to be the null set. If $\mathscr{A} = \varnothing$, the definition in **4.1** gives that $\bigcup_\alpha \{A_\alpha \mid \alpha \in \mathscr{A}\} = \varnothing$, since no $x \in \Gamma$ satisfies the condition $\exists\, \alpha \in \mathscr{A}\colon x \in A_\alpha$; similarly, we find that $\bigcap_\alpha \{A_\alpha \mid \alpha \in \mathscr{A}\} = \Gamma$, where Γ is the specified domain of discourse. In the future, we will call a family $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ *nonempty* whenever we wish to emphasize that the indexing set $\mathscr{A} \neq \varnothing$; this, of course, does not exclude the possibility that some $A_\alpha = \varnothing$.

5. Power Set

5.1 Definition Let A be any set. Its power set $\mathscr{P}(A)$ is the set of all subsets of A.

Ex. 1 For any set A, $\varnothing \in \mathscr{P}(A)$ and $A \in \mathscr{P}(A)$.
Ex. 2 $B \subset A$ is equivalent to $B \in \mathscr{P}(A)$; $a \in A$ equivalent to $\{a\} \in \mathscr{P}(A)$.

Its relation to the set operations is

5.2 Theorem $\bigcap_\alpha$ and $\mathscr{P}$ commute: $\bigcap_\alpha \mathscr{P}(A_\alpha) = \mathscr{P}(\bigcap_\alpha A_\alpha)$. Though $\bigcup_\alpha$ and $\mathscr{P}$ do not commute, $\bigcup_\alpha \mathscr{P}(A_\alpha) \subset \mathscr{P}(\bigcup_\alpha A_\alpha)$.

The simple proof of the first part is left for the reader. To see that $\bigcup_\alpha$ and $\mathscr{P}$ do not generally commute, let $A_1 = \{1\}$, $A_2 = \{2\}$; then $\bigcup_\alpha \mathscr{P}(A_\alpha)$ has three elements, whereas $\mathscr{P}(\bigcup_\alpha A_\alpha)$ has four.

6. Functions, or Maps

The notion of a map (or function) is basic in all mathematics; it is here defined in terms of the primitive concept "set" by identifying functions with their graphs.

A map of a set X into a set Y is determined by sending each $x \in X$ to some $y \in Y$; since to say "x is sent to y" can be expressed simply by "(x, y) is an ordered couple," a map is determined by specifying what the ordered couples are, that is, by specifying a subset of $X \times Y$.

6.1 Definition Let X and Y be two sets. A map $f\colon X \to Y$ (or function with domain X and range Y) is a subset $f \subset X \times Y$ with the property: for each $x \in X$, there is one, and only one, $y \in Y$ satisfying $(x, y) \in f$.

To denote $(x, y) \in f$, we write $y = f(x)$ and say that y is the value f assumes (takes on) at x, or that y is the image of x under f, or that f sends x to y.

Since maps are sets (of ordered couples), two maps f, g are equal if and only if they have the same domain X and $f(x) = g(x)$ for each $x \in X$. The usual way to define a map is to specify its domain X and value at each $x \in X$: if the value at $x \in X$ is $r(x) \in Y$, the map

$$\{(x, r(x)) \mid x \in X\}$$

is written $x \to r(x)$.

Ex. 1 A map $f: X \to Y$ sending all $x \in X$ to a single point, $b \in Y$, is called a constant map. Note that a map need not send distinct points of X to distinct points of Y, nor need it take on all values in its range.

Ex. 2 The map $x \to x$ of X onto itself is called the identity map of X, and is written $1: X \to X$, or 1_X. If $A \subset X$, the map $i: A \to X$ given by $a \to a$ is called the inclusion map of A into X.

Ex. 3 Assume that at least one of X, Y is $\varnothing$, so that $X \times Y = \varnothing$ by **3.3**. If $X = \varnothing$, there is exactly *one* map $\varnothing \to Y$, since $\varnothing = \varnothing \times Y$ is a subset as required by **6.1** (the assertion that $\varnothing$ is not a map is false). However, if $X \neq \varnothing$ and $Y = \varnothing$, then $\varnothing$ is *not* a map (there can be no $y \in \varnothing$).

Ex. 4 For any sets X, Y, the map $p_1: X \times Y \to X$ determined by $(x, y) \to x$ is called "projection onto the first coordinate"; $p_2: X \times Y \to Y$, where $p_2(x, y) = y$ is "projection onto the second coordinate."

6.2 Definition Let $f: X \to Y$. Then:

(1). For each $A \subset X$, $f(A) = \{f(x) \mid x \in A\} \subset Y$ is called the image of A in Y under f.

(2). For each $B \subset Y$, $f^{-1}(B) = \{x \mid f(x) \in B\} \subset X$ is called the inverse image of B in X under f.

Ex. 5 Let $X = [-1, 1]$, $Y = [0, 2]$, and $f: X \to Y$ be $x \to x^2$. Then $f[0, \frac{1}{2}] = [0, \frac{1}{4}]$ and $f^{-1}[0, \frac{1}{4}] = [-\frac{1}{2}, \frac{1}{2}]$.

Ex. 6 If $X = \varnothing$, the unique map $\varnothing \to Y$ has image $\varnothing$.

Ex. 7 Let $f: X \to Y$. If p_1, p_2 are the projections in Ex. 4, we have $f(A) = p_2[f \cap (A \times Y)]$ and $f^{-1}(B) = p_1[f \cap (X \times B)]$.

Let $f: X \to Y$ be given. Then f *induces* a map (still denoted by f), $f: \mathscr{P}(X) \to \mathscr{P}(Y)$ by $A \to f(A)$. The map $f: X \to Y$ also induces a map $f^{-1}: \mathscr{P}(Y) \to \mathscr{P}(X)$ by $B \to f^{-1}(B)$. Of these two induced maps, f^{-1} is the most important because of

6.3 Theorem Let $f: X \to Y$. Then the induced $f^{-1}: \mathscr{P}(Y) \to \mathscr{P}(X)$ preserves the elementary set operations. Precisely,

(1). $$f^{-1}\left(\bigcup_\alpha B_\alpha\right) = \bigcup_\alpha f^{-1}(B_\alpha).$$

(2). $$f^{-1}\left(\bigcap_\alpha B_\alpha\right) = \bigcap_\alpha f^{-1}(B_\alpha).$$

(3). $$f^{-1}(B_1 - B_2) = f^{-1}(B_1) - f^{-1}(B_2).$$

Proof: We establish only (2); the proofs of (1) and (3) are similar. We have

$$x \in f^{-1}(\bigcap_\alpha B_\alpha) \Leftrightarrow f(x) \in \bigcap_\alpha B_\alpha \Leftrightarrow \forall\alpha: f(x) \in B_\alpha$$
$$\Leftrightarrow \forall\alpha: x \in f^{-1}(B_\alpha)$$
$$\Leftrightarrow x \in \bigcap_\alpha f^{-1}(B_\alpha).$$

In contrast to **6.3**, the induced map $f: \mathscr{P}(X) \to \mathscr{P}(Y)$ behaves less satisfactorily. Though it preserves unions, it does not in general preserve intersections:

Ex. 8 Let $f: E^1 \to E^1$ be the constant map $x \to 1$. Then, with $A = [0, 1]$, $B = [2, 3]$, we obtain $\varnothing = f(A \cap B) \neq f(A) \cap f(B) = \{1\}$.

The reader can easily prove

6.4 If $f: X \to Y$, then for the induced map $f: \mathscr{P}(X) \to \mathscr{P}(Y)$:

(1). $f(\bigcup_\alpha A_\alpha) = \bigcup_\alpha f(A_\alpha)$.
(2). $f(\bigcap_\alpha A_\alpha) \subset \bigcap_\alpha f(A_\alpha)$.

For the combined action of f and f^{-1}, it is simple to verify

6.5 If $f: X \to Y$, then:

(1). For each $A \subset X$, $f^{-1}[f(A)] \supset A$.
(2). For each $A \subset X$ and $B \subset Y$, $f[f^{-1}(B) \cap A] = B \cap f(A)$; in particular, $f[f^{-1}(B)] = B \cap f(X)$.

Given $f: X \to Y$ and $g: Y \to Z$, their composition $g \circ f: X \to Z$ is defined as the map $x \to g(f(x))$. We can clearly compose the induced maps f^{-1}, g^{-1} and we have

6.6 Theorem Let $f: X \to Y$ and $g: Y \to Z$. Then $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

Proof:

$$x \in (g \circ f)^{-1} C \Leftrightarrow g \circ f(x) \in C \Leftrightarrow f(x) \in g^{-1}(C)$$
$$\Leftrightarrow x \in f^{-1}[g^{-1}(C)] \Leftrightarrow x \in f^{-1} \circ g^{-1}(C),$$

Given an $f: X \to Y$ and a subset $A \subset X$, the map f considered *only* on A is called the *restriction* of f to A, is written $f \mid A$, and can alternatively be defined as $f \mid A = f \cap (A \times Y)$. In the reverse direction, if $A \subset X$

and $g: A \to Y$ is a given map, a map $G: X \to Y$ coinciding with g on A (that is, satisfying $G \mid A = g$) is called an *extension* of g over X relative to Y. The following result is very useful:

6.7 Theorem Let X be any set, and $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ any family of subsets with $\bigcup_\alpha A_\alpha = X$ (a "covering" of X). For each $\alpha \in \mathscr{A}$, let an $f_\alpha: A_\alpha \to Y$ be given, and assume that $f_\alpha \mid A_\alpha \cap A_\beta = f_\beta \mid A_\alpha \cap A_\beta$ for each $(\alpha, \beta) \in \mathscr{A} \times \mathscr{A}$. Then there exists one, and only one, $f: X \to Y$ which is an extension of each f_α; that is, $\forall \alpha \in \mathscr{A}: f \mid A_\alpha = f_\alpha$.

Proof: For each $x \in X$, define $f(x) = f_\alpha(x)$, where α is any index for which $x \in A_\alpha$; this definition is unique because, if also $x \in A_\beta$, the hypothesis $f_\alpha \mid A_\alpha \cap A_\beta = f_\beta \mid A_\alpha \cap A_\beta$ gives $f(x) = f_\alpha(x) = f_\beta(x)$. $x \to f(x)$ is therefore a map $f: X \to Y$; it is evidently an extension of each f_α over X; and it is unique because each x belongs to some A_α, so any map that satisfies the requirements will have to assume the value $f_\alpha(x)$ at x (that is, the same value as f).

If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a covering of X, and if $A_\alpha \cap A_\beta = \varnothing$ whenever $\alpha \neq \beta$, then the family $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is called a *partition of* X. We obtain at once the

6.8 Corollary If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a partition of X and if for each $\alpha \in \mathscr{A}$ there is given an $f_\alpha: A_\alpha \to Y$, then there exists a unique $f: X \to Y$ which is an extension of each f_α.

Proof: The requirement $f_\alpha \mid A_\alpha \cap A_\beta = f_\beta \mid A_\alpha \cap A_\beta$ of **6.7** is evidently satisfied.

If $f: X \to Y$ takes on every value in its range, f is called *surjective* (or a surjection; or "onto"). Observe that for surjective f, **6.5**(2) takes the simpler form: $\forall B: B \subset Y \Rightarrow f[f^{-1}(B)] = B$.

If f sends distinct elements of X to distinct elements of Y, f is called *injective* (or an injection; or one-to-one). Evidently f is injective if and only if $[x \neq x'] \Rightarrow [f(x) \neq f(x')]$, or equivalently, $[f(x) = f(x')] \Rightarrow [x = x']$. The restriction of an injection to any subset is also an injection.

If f is both injective and surjective, f is called *bijective* (or a bijection; or a one-to-one onto map).

Ex. 9 The map $F: A \times B \to B \times A$ given by $(a, b) \to (b, a)$ is bijective. The map $p_1: X \times Y \to X$ is surjective. The map f of Ex. 5 is neither surjective nor injective; however, $f \mid [0, 1]$ is injective.

Clearly, $f: X \to Y$ is bijective if and only if $\forall y \in Y: f^{-1}\{y\}$ is a single point. Thus, with each bijection $f: X \to Y$, we have also a *map* $f^{-1}: Y \to X$ determined by $y \to f^{-1}(\{y\})$ [this differs from $f^{-1}: \mathscr{P}(Y) \to \mathscr{P}(X)$, since the domains are not the same]; it is evident that $f^{-1}: Y \to X$ is also bijective and that $(f^{-1})^{-1} = f$.

The following proposition indicates a simple method for establishing that a given map f(resp. g) is injective (resp. surjective).

6.9 Let $f: X \to Y$ and $g: Y \to X$ satisfy $g \circ f = 1_X$. Then f is injective and g is surjective.

Proof: The map f is injective, since $[f(x) = f(x')] \Rightarrow x = g \circ f(x) = g \circ f(x') = x'$. g is surjective, since for any $x_0 \in X$, $x_0 = g[f(x_0)]$.

Ex. 10 As a trivial example illustrating the use of **6.9**, we show that for any map $h: X \to Y$, the map $f: X \to X \times Y$ defined by $x \to (x, h(x))$ is injective. Let $p_1: X \times Y \to X$ be projection on the first coordinate; since it is immediate that $p_1 \circ f: X \to X$ is 1_X, it follows from **6.9** that f is injective.

7. Binary Relations; Equivalence Relations

A binary relation R in a set A is, intuitively, a proposition such that for each ordered couple (a, b) of elements of A, one can determine whether $a\,R\,b$ ("a is in relation R to b") is or is not true. We state this formally in terms of the set concept.

7.1 Definition A binary relation R in a set A is a subset $R \subset A \times A$. $(a, b) \in R$ is written $a\,R\,b$.

Ex. 1 For any set A, the diagonal $\Delta = \{(a, a) \mid a \in A\} \subset A \times A$ is the relation of equality. Note that in binary relations, each pair of elements need not be related: If $a \neq b$, neither (a, b) nor (b, a) is in Δ.

Ex. 2 $R = (A \times A) - \Delta$ is the relation of inequality.

Ex. 3 The relation $\leqslant$ between real numbers is the set

$$\{(x, y) \mid x \leqslant y\} \subset E^1 \times E^1.$$

Ex. 4 The relation of inclusion in $\mathscr{P}(A)$ is $\{(A, B) \mid A \subset B\} \subset \mathscr{P}(A) \times \mathscr{P}(A)$.

Ex. 5 An evident generalization of **7.1** is to define any subset of $X \times Y$ to be a binary relation between the elements of X and those of Y; thus a map $f: X \to Y$ would be a special type of binary relation.

Let A be a set with a binary relation R. If $B \subset A$, then **(3.4)** $B \times B \subset A \times A$, and therefore $R \cap (B \times B)$ is a binary relation on B [corresponding to the statement "$(x \in B) \wedge (y \in B) \wedge (x\,R\,y)$"]; we call

$R \cap (B \times B)$ the relation induced by R in B. In particular, a binary relation in A induces a definite binary relation in every subset of A.

Equivalence relations are very frequently used in all mathematics; they arise whenever one desires to regard all those members of a set that have some preassigned characteristic as a single entity.

7.2 Definition A binary relation R in A is called an *equivalence* relation if:

(1). $\forall\, a \in A \colon a\, R\, a$ (reflexive).
(2). $(a\, R\, b) \Rightarrow (b\, R\, a)$ (symmetric).
(3). $(a\, R\, b) \wedge (b\, R\, c) \Rightarrow (a\, R\, c)$ (transitive).

If $a\, R\, b$, we say that a and b are equivalent.

Ex. 6 The relation of Ex. 1 is an equivalence relation.

Ex. 7 In N, $\{(x, y) \mid x \equiv y \bmod 2\}$ is an equivalence relation.

Ex. 8 In the set of vectors $\subset E^3$, let $a\, R\, b$ denote that a and b have the same direction. Then R is an equivalence relation.

Ex. 9 Let $f \colon X \to Y$ be a map. The relation $\{(x, x') \mid f(x) = f(x')\}$ is an equivalence relation in X.

Ex. 10 Each of the following relations satisfy exactly two of the requirements of **7.2**. (a) In E^1, the relation $[0, 1] \times [0, 1]$ is not reflexive. (b) The relation of Ex. 3 is not symmetric. (c) In E^1, the relation $\{(x, y) \mid |x - y| \leqslant 1\}$ is not transitive.

Ex. 11 Any symmetric and transitive relation R that satisfies the condition $\forall\, a\, \exists\, b \colon a\, R\, b$ is in fact an equivalence relation, as the reader will immediately verify.

Let R be an equivalence relation in A. For each $a \in A$, the subset $Ra = \{b \in A \mid b\, R\, a\}$ is called the *equivalence class of a*. The fundamental theorem on equivalence relations is a consequence of

7.3 Lemma Let R be an equivalence relation in A. Then:

(1). $\bigcup\{R\, a \mid a \in A\} = A$.
(2). If $a\, R\, b$, then $R\, a = R\, b$.
(3). If $\neg(a\, R\, b)$, then $R\, a \cap R\, b = \varnothing$.

Proof: Ad (1). Because R is reflexive, we have $a \in Ra$, and therefore $\cup\, \{Ra \mid a \in A\} = A$.

Ad (2). Let $x \in R\, a$; using transitivity of R gives

$$x \in R\, a \Leftrightarrow x\, R\, a \Rightarrow x\, R\, b \Leftrightarrow x \in R\, b$$

showing $R\, a \subset R\, b$. For the converse inclusion, observe that because R is symmetric, we have $b\, R\, a$; consequently, $x\, R\, b \Rightarrow x\, R\, a$ by transitivity, proving that $R\, b \subset R\, a$.

Ad (3). Assume $R\,a \cap R\,b \neq \varnothing$. Choosing $\xi \in R\,a \cap R\,b$, we have $(\xi R\,a) \wedge (\xi R\,b)$, so that symmetry and transitivity of R give $a\,R\,b$.

7.4 Theorem Let A have an equivalence relation R. Then the collection of distinct equivalence classes partitions A into mutually disjoint sets, called R-equivalence classes, such that any two elements of A belong to a common R-equivalence class if, and only if, they are equivalent.

Proof: Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be the collection of R-equivalence classes. Since each A_α is some one of the sets $R\,a$, the partition property of $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is clear from **7.3**. Again by **7.3**, observe that if $A_\alpha = R\,a$, then also $A_\alpha = R\,b$ for each b satisfying $b\,R\,a$, and $R\,b \cap A_\alpha = \varnothing$ if $\neg(b\,R\,a)$; the rest of the theorem is now immediate.

Each element a of an R-equivalence class A_α is called a *representative* of A_α; observe that $a \in A_\alpha \Leftrightarrow A_\alpha = R\,a$.

Ex. 12 In Ex. 1, each R-equivalence class contains exactly one element. In Ex. 7, there are exactly two equivalence classes, the set of even integers, and the set of odd integers: 0 and 1 are representatives of these classes. In Ex. 8, there is one R-equivalence class for each direction in E^3; the unit vector in each class can be taken as a representative for that class. In Ex. 9, the R-equivalence classes are the sets $\{f^{-1}(y) \mid y \in f(X)\}$.

7.5 Remark The reader can prove the converse and extension of **7.4**: R is an equivalence relation in A if and only if there is a partition $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ of A such that $R = \cup\{A_\alpha \times A_\alpha \mid \alpha \in \mathscr{A}\}$. Furthermore, the sets A_α are precisely the R-equivalence classes.

With each equivalence relation R in A, we construct a new set according to the following

7.6 Definition Let A have an equivalence relation R. The set whose *elements* are the R-equivalence classes is called the quotient set of A by R and is written A/R. The map $p_A\colon A \to A/R$ given by $a \to R\,a$ is called the projection of A onto A/R.

Clearly, p_A is surjective, but not in general injective (since $R\,a = R\,b = A_\alpha$ whenever $a\,R\,b$); note also that $A/R \subset \mathscr{P}(A)$. We omit the subscript on p_A when no confusion arises. A set of elements, one from each equivalence class, is called a *system of representatives for A/R*.

Let A, B be two sets with equivalence relations R, S, respectively. A map $f\colon A \to B$ is called *relation-preserving* if $a\,R\,a' \Rightarrow f(a) S f(a')$.

7.7 Theorem Let $f: A \to B$ be relation-preserving. Then there is one, and only one, map $f_*: A/R \to B/S$ such that the diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle p_A}\downarrow & & \downarrow{\scriptstyle p_B} \\ A/R & \xrightarrow[f_*]{} & B/S \end{array}$$

commutes (that is, $p_B \circ f = f_* \circ p_A$). f_* is called the map induced by f in "passing to the quotient." Conversely, if for *any* two maps f, f_* the above diagram commutes, then f is necessarily relation-preserving, and f_* is the map induced by f.

Proof: For each R-equivalence class $A_\alpha = R\,a$, define f_* by $R\,a \to S\,f(a)$. The S-equivalence class $S\,f(a)$ is independent of the representative $a \in A_\alpha$ selected: for if also $a' \in A_\alpha$, then $R\,a = R\,a'$ so, by **7.3**, $a\,R\,a'$; since f is relation-preserving, we find $f(a)\,S\,f(a')$; consequently $S\,f(a) = S\,f(a')$. Thus, f_* is uniquely defined. The commutativity follows by observing

$$p_B \circ f(a) = S\,f(a) = f_*\,R\,a = f_* \circ p_A(a).$$

Finally, f_* is unique because p_A is surjective: if g_* were another map for which the diagram commutes, then $g_*(A_\alpha) \neq f_*(A_\alpha)$ for at least one A_α; since p_A is surjective, this would require $g_* \circ p_A(a) \neq f_* \circ p_A(a)$ for some $a \in A$, which is impossible because of commutativity.

To prove the converse, assume that maps f, f_* give a commutative diagram. Let $a\,R\,a'$; then $p_A(a) = p_A(a')$ and therefore, by commutativity, we have $p_B \circ f(a) = p_B \circ f(a')$. This shows $f(a)\,S\,f(a')$ and proves that f is relation-preserving.

Ex. 12 Let $A = B = Z$, with $R = \{(a, a') \mid a \equiv a' \bmod 4\}$, $S = \{(b, b') \mid b \equiv b' \bmod 2\}$. The $f: A \to B$ given by $n \to n$ is relation-preserving, and passing to the quotient gives $f_*: A/R \to B/S$. By using the representatives 0, 1, 2, 3 for A/R and 0, 1 for B/S, the reader can verify $f_*(0) = f_*(2) = 0, f_*(1) = f_*(3) = 1$.

8. Axiomatics

Though the intuitive idea of calling any collection of objects a set will suffice for most purposes, an exposition of general Set Theory requires more precision, for without explicit axioms telling how the term *set* can be used, and to what collections it can be applied, various contradictions arise. There are several different axiomatic set theories, each having technical advantages and shortcomings; we present here a version based on the Bernays–Gödel–von Neumann axiomatics. The treatment is not intended to be either complete or formal, nor is the system of axioms asserted to be independent; these matters properly belong to the domain of Logic. It is desired to indicate only a framework within which we will work, which avoids the known antinomies and which, at least until now, has not led to any contradictions.

Ideally, we would like to have associated with each property p a set $E(p)$ consisting of all objects having property p. The assumption that this is true leads at once to the Russell antinomy of the set of all sets not members of themselves: assuming that the property $p(x) = (x \text{ is a set}) \wedge (x \,\bar{\in}\, x)$ determines some *set* $\mathscr{R}(p)$, we must conclude that $[\mathscr{R}(p) \,\bar{\in}\, \mathscr{R}(p)] \Leftrightarrow [\mathscr{R}(p) \in \mathscr{R}(p)]$. To block this contradiction, we adopt the attitude that it is not the conversion of properties to collections that is at fault, but rather the assumption that $\mathscr{R}(p)$ is a *set* and is therefore eligible for membership in the collection determined by p. The basic idea of this approach is, then, that there are two types of collections—classes and sets: any collection of objects specified by some property is a class, whereas only those classes that can be *members* of a class are *sets*. Heuristically, a set is a class that can be regarded as a single entity.

The undefined terms in the axiomatic development are "class" and a dyadic relation $\in$ between classes; all variables, such as $\mathscr{A}$, A, $x, \cdots$, represent classes, and for any two classes, the statement $\mathscr{A} \in \mathscr{B}$ is either true or false. A property p will mean a formula built up from statements "$\mathscr{A} \in \mathscr{B}$" by negation, conjunction, disjunction, and quantification of class variables by means of the predicate calculus.

We begin by defining classes to be equal if they have the same members; formally,

8.1 Definition

$$(\mathscr{A} \subset \mathscr{B}) \Leftrightarrow (\forall\, x: x \in \mathscr{A} \Rightarrow x \in \mathscr{B}),$$

and

$$(\mathscr{A} = \mathscr{B}) \Leftrightarrow (\mathscr{A} \subset \mathscr{B}) \wedge (\mathscr{B} \subset \mathscr{A}).$$

This definition permits substitution with respect to the second class variable in the relation $x \in \mathscr{A}$; that is, $(x \in \mathscr{A}) \wedge (\mathscr{A} = \mathscr{B}) \Rightarrow (x \in \mathscr{B})$; to obtain it also for the first requires

I Axiom (of Individuality) $(x \in \mathscr{A}) \wedge (x = y) \Rightarrow (y \in \mathscr{A})$.

Next we distinguish between classes and sets by

8.2 Definition The class A is called a set if there is a class $\mathscr{A}$ such that $A \in \mathscr{A}$.

We now wish to postulate that any collection specified by a property that characterizes its members is a *class*. However, since nonsets cannot be members of anything, the members of a class must be *sets*. We formulate this by

II Axiom (of Class Formation) For each property p in which only set variables are quantified and in which the class variable $\mathscr{A}$ does not appear, there is a class $\mathscr{A}$ whose members are just those *sets* having property p; in symbols, $(x \in \mathscr{A}) \Leftrightarrow (x \text{ is a set}) \wedge p(x)$.

Because of Axiom I, the class $\mathscr{A}$ is uniquely determined by its defining property; we will denote $\mathscr{A}$ by the notation $\{x \mid (x \text{ is a set}) \wedge p(x)\}$ and sometimes by $\mathscr{A}(p)$. Observe that with this terminology, the Russell antinomy becomes the harmless statement

8.3 The Russell class $\mathscr{R}(p)$ is not a set.

Using Axiom II, the Boolean operations $\mathscr{A} \cup \mathscr{B} = \{x \mid (x \in \mathscr{A}) \vee (x \in \mathscr{B})\}$ and $\mathscr{A} \cap \mathscr{B} = \{x \mid (x \in \mathscr{A}) \wedge (x \in \mathscr{B})\}$ with classes, as well as the cartesian product $\mathscr{A} \times \mathscr{B}$ of classes, are defined and are classes. The universal class is $\{x \mid (x \text{ is a set}) \wedge (x = x)\}$, and the null class $\varnothing$ is $\{x \mid (x \text{ is a set}) \wedge (x \neq x)\}$; as in **1.4**, $\varnothing$ is unique and a subclass of every class. Equivalence relations in classes can be defined as in **7**, leading to

8.4 An equivalence relation in a class $\mathscr{A}$ partitions $\mathscr{A}$ into pairwise disjoint subclasses.

The next string of axioms guarantees at least one *set* and postulates that certain constructions using sets will yield sets.

III Axiom (of Null Set) $\varnothing$ is a set.

IV Axiom (of Pairing) If A, B are distinct sets, then $\mathscr{A} = \{x \mid (x = A) \vee (x = B)\}$ is a *set* (which contains exactly two elements). It is denoted by $\{A, B\}$.

V Axiom (of Union) If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a family of sets (recall that, as defined in **4**, this means that $\mathscr{A}$ and each A_α are *sets*), then $\bigcup \{A_\alpha \mid \alpha \in \mathscr{A}\} = \{x \mid \exists\, \alpha \in \mathscr{A} : x \in A_\alpha\}$ is a set.

VI Axiom (of Replacement) If A is a set and if $f: A \rightarrow \mathscr{A}$ is a map, then $f(A)$ is a set.

The next axiom deals with subset formation.

VII Axiom (of Sifting) If A is a set, then for any class $\mathscr{A}$, $A \cap \mathscr{A}$ is a set.

In particular,

8.5 If A is a set and p is a property in which only set variables are quantified, then $\{x \mid (x \in A) \wedge p(x)\}$ (which we will write as $\{x \in A \mid p(x)\}$) is a set.

For, if $\mathscr{A}$ is the class determined by p, we have $A \cap \mathscr{A} = \{x \mid (x \in A) \wedge (x \text{ is a set}) \wedge p(x)\}$, and the requirement $x \in A$ makes the stipulation "(x is a set)" redundant.

Since members of classes must necessarily be sets, the precise definition of the power class $\mathscr{P}(\mathscr{A})$ of a class $\mathscr{A}$ is $\mathscr{P}(\mathscr{A}) = \{\mathscr{B} \mid (\mathscr{B} \text{ is a set}) \wedge \mathscr{B} \subset \mathscr{A}\}$; thus, even if A is a set, $\mathscr{P}(A)$ has for members only those subclasses of A known to be sets.

VIII Axiom (of Power Set) If A is a set, then $\mathscr{P}(A)$ is also a set.

To indicate how these axioms are used, we establish that some frequently occurring constructions with sets will yield sets.

8.6 If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a family of sets, then $\bigcap \{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a set.

According to Axiom V, $S = \bigcup \{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a set; letting $p(x) = [\forall \alpha \in \mathscr{A} : x \in A_\alpha]$, in which only the set variable α is quantified, **8.5** shows that $\{x \in S \mid p(x)\}$, which is precisely $\bigcap \{A_\alpha \mid \alpha \in \mathscr{A}\}$, is a set.

8.7 If A is a set, then so also is $\{A\}$.

If $A = \varnothing$, then III and VIII give that $\{\varnothing\}$ is a set. If $A \neq \varnothing$, then $\{A, \varnothing\}$ is a set because of IV. Letting $\mathscr{A}$ be the class determined by the property $p(x) = (x = A)$, we find from **8.5** that $\{A, \varnothing\} \cap \mathscr{A} = \{A\}$ is a set.

8.8 The cartesian product of two sets is a set.

Let A, B be sets, and for each $a_0 \in A$ define a map $B \to A \times B$ by $b \to (a_0, b)$; according to VI, the image, $\{a_0\} \times B$, is a set; since $A \times B = \bigcup \{\{a_0\} \times B \mid a_0 \in A\}$, V shows that $A \times B$ is a set.

8.9 If A and B are sets, then the class of all maps $A \to B$ is a set.

We have just seen that $A \times B$ is a set so, by VIII, $\mathscr{P}(A \times B)$ is also a set. Since a map is a subclass of $A \times B$ specified by some property, **8.5** shows that each map is a member of the set $\mathscr{P}(A \times B)$. Using now the property m expressed in **6.1**, the class of all maps $A \to B$ is $\{x \in \mathscr{P}(A \times B) \mid x \text{ has property } m\}$ so, again by **8.5**, it is a set.

8.10 The class of all sets is not a set.

Let $\mathscr{R}(p)$ be the Russell class. If the class A of all sets were a set, then VII would imply that $A \cap \mathscr{R}(p)$ is a set; since $A \cap \mathscr{R}(p) = \{x \mid (x \text{ is a set}) \wedge (x \,\bar{\in}\, x)\} = \mathscr{R}(p)$, this contradicts **8.3**.

We now add

IX Axiom (of Foundation) In each nonempty set A there is a $u \in A$ such that $u \cap A = \varnothing$ (that is, $\forall\, x\colon x \in A \Rightarrow \neg\, (x \in u)$).

Loosely speaking, this axiom asserts that each nonempty set must contain "atoms" u, which form its "foundation." Its use is shown in

8.11 (1) No nonempty set can be a member of itself.

(2). If A, B are nonempty sets, then it is not possible that both $A \in B$ and $B \in A$ are true.

Ad (1). Assume there were a nonempty set A such that $A \in A$; by **8.7**, $\{A\}$ would also be a set, and because A is also the only member of $\{A\}$, the set $\{A\}$ would not have a foundation.

Ad (2). Consider the set (*cf.* IV) $\{A, B\}$ in an analogous way.

We now provide for the existence of infinite sets by

X Axiom (of Infinity) There exists a *set* A with the properties: (i) $\varnothing \in A$, and (ii) if $a \in A$, then $a \cup \{a\} \in A$.

As an application, we have

8.12 The class of nonnegative integers is a set.

Let A be any set having the two properties listed in Axiom X, and let $\mathscr{B} \subset \mathscr{P}(A)$ be defined by $\mathscr{B} = \{B \in \mathscr{P}(A) \mid B \text{ has the two properties in Axiom X}\}$. Each B is a set, and by **8.5** and VIII, so also is $\mathscr{B}$; it therefore follows from **8.6** that $N = \bigcap \{B \mid B \in \mathscr{B}\}$ is also a set. Because each B has the properties (i) and (ii) of Axiom X, it is evident that N has them also. Referring now to the Peano axioms for the integers, and calling $x \cup \{x\} \in N$ the successor of $x \in N$, it can be easily verified that all the Peano axioms are satisfied by N [the principle of mathematical induction is valid because, by the definition of N, N has *no* proper subset that satisfies both (i) and (ii)]. Since the Peano axioms are categorical, it follows that N can be regarded as the set of nonnegative integers. We denote $\varnothing$ by "0," $\{\varnothing\}$ by "1," $\{\varnothing, \{\varnothing\}\}$ by "2," and so on.

An easy consequence from this, **8.8** and VI, is that the class Q of rationals is a set; we will see later that there is a bijection of $\mathscr{P}(N)$ onto the reals; therefore, by VI and VIII, E^1 is also a set.

According to this axiomatization, the only general method for producing subsets of a given set is that given in **8.5**. To see that there are subsets that we would like to consider but that cannot be described by any property, consider the following example of Russell: Let A be an infinite collection of pairs of shoes; we can define a subset consisting of exactly *one* shoe from each pair by the property "right shoe." If now A is an infinite collection of pairs of stockings, analogy would indicate that a subset consisting of exactly *one* stocking from each pair could be formed; but because the stockings are identical in all respects, there can be no property that characterizes exactly one of each pair; in particular, we are not allowed to call such a collection a *subset* of A. Analogous situations arise frequently in mathematics; to give the broadest scope to mathematical considerations, we adopt as another method for producing subsets,

XI Axiom (of Choice) Given any nonempty family $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ of nonempty pairwise disjoint *sets*, there exists a *set* S consisting of exactly *one* element from each A_α.

This is the only *existential* axiom: in contrast to all the others, a set obtained by application of this axiom is *not*, in general, uniquely determined by the given conditions. It has been shown (1938) by K. Gödel that if the set theory based on I–X is consistent, then the set theory based on I–XI is also consistent. Gödel's result obviously leaves open the possibility that XI is derivable from the other axioms, and, in 1963, P. J. Cohen proved that it is not. Thus, the axiom of choice is in fact an independent axiom.

Remark Note that appeal to XI is not necessary if $\mathscr{A}$ is a *finite* set. Indeed, if $A_1, \cdots, A_n$ are the sets, then each A_i contains some element a_i so, by taking $p_i(x)$ to be the property $(x = a_i)$, we obtain $\{x \in \bigcup_1^n A_i \mid p_1(x) \vee \cdots \vee p_n(x)\}$ as a set satisfying the requirement of Axiom XI. However, if $\mathscr{A}$ is *infinite* (even though, as in Russell's example, each A_α is finite), some principle such as XI appears to be necessary. A property such as "contains one element from each A_α" is illegitimate because properties carve out *unique* subsets and it is evident that if there is one collection satisfying the proposed predicate, then there are others also (unless each A_α consists of a single element). Taking $p_i(x) = (x = a_i)$, the procedure used above for finite $\mathscr{A}$ cannot be emulated, since an infinite "or" chain is *not* a proposition. And to use a predicate such as $\exists\, i: p_i(x)$ is inadequate: for, assuming about x that $\neg\, p_1(x), \neg p_2(x), \cdots$, one cannot in general conclude $\forall\, i: \neg p_i(x)$, that is, $\neg\, \exists\, i: p_i(x)$ [or, that x does not belong to the set determined by $\exists\, i: p_i(x)$], without tampering with the rules of logic in the predicate calculus; from this viewpoint, the need for XI is related to the ω-incompleteness of consistent systems.

9. General Cartesian Products

In this section, the concept of cartesian product is extended to any family of sets. This extension is based on the observation that the elements of $A_1 \times A_2$ can be considered to be those maps f of the index set $\{1, 2\}$ into $A_1 \cup A_2$ having the property $f(1) \in A_1, f(2) \in A_2$.

9.1 Definition Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of sets. The cartesian product $\prod_\alpha A_\alpha$ is the set of all maps $c: \mathscr{A} \to \bigcup_\alpha A_\alpha$ having the property $\forall \alpha \in \mathscr{A}: c(\alpha) \in A_\alpha$.

That $\prod_\alpha A_\alpha$ is indeed a *set* follows from **8.5** and **8.9**. The notations $\prod_\alpha A_\alpha$ and $\prod\{A_\alpha \mid \alpha \in \mathscr{A}\}$ are used interchangeably. An element $c \in \prod_\alpha A_\alpha$ is generally written $\{a_\alpha\}$, indicating that $c(\alpha) = a_\alpha$ for each α; with this notation, $a_\alpha \in A_\alpha$ is called the αth coordinate of $\{a_\alpha\}$. The set A_α is called the αth factor of $\prod_\alpha A_\alpha$; for each $\beta \in \mathscr{A}$, the map

$$p_\beta: \prod_\alpha A_\alpha \to A_\beta$$

given by $\{a_\alpha\} \to a_\beta$ [or, equivalently, by $c \to c(\beta)$] is termed "projection onto the βth factor."

Ex. 1 If each A_α has exactly one element, $\prod_\alpha A_\alpha$ consists of a single element. If $\mathscr{A} = \varnothing$, then again $\prod_\alpha A_\alpha$ has exactly one element, the null set (*cf.* **6**, Ex. 3). If $\mathscr{A} \neq \varnothing$ and some one $A_\alpha = \varnothing$, then $\prod_\alpha A_\alpha = \varnothing$.

Ex. 2 If each $A_\alpha = A$, a fixed set, then $\prod_\alpha A_\alpha$ is simply the set of all maps $\mathscr{A} \to A$.

Ex. 3 Let $A_i = \{0, 2\}$ for each $i \in Z^+$; $\prod_n A_n$ is then the set of all sequences of 0's and 2's: $\{\{n_i\} \mid n_i = 0 \text{ or } 2;\ i = 1, 2, \cdots\}$. The map $f: \prod_n A_n \to [0, 1] \subset E^1$, defined by

$$f(\{n_i\}) = \sum_{i=1}^{\infty} \frac{n_i}{3^i}$$

is easily seen to be injective; the image is called the Cantor set, and can be described geometrically as follows: Divide [0, 1] into three equal parts and remove the middle-third open interval $]\frac{1}{3}, \frac{2}{3}[$; this removes all real numbers in [0, 1] that require $n_1 = 1$ in their triadic expansion. At the second stage, remove the middle third of each of the two remaining intervals, $[0, \frac{1}{3}]$, $[\frac{2}{3}, 1]$, thus eliminating all real numbers in [0, 1] requiring $n_2 = 1$ in their triadic expansion. Proceeding analogously, removing at the nth stage the union M_n of the middle thirds of the 2^{n-1} intervals present, $C = [0, 1] - \bigcup_1^\infty M_n$ is the Cantor set. It consists of all real numbers in [0, 1] that do not require the use of "1" in their triadic expansion. Since the expansion (using no 1's) of each member of C is unique, f is a bijection of $\prod_n A_n$ on C. In view of Ex. 2, there is a bijection of the set of all maps $Z^+ \to \{0, 2\}$ onto the Cantor set.

Ex. 3 is analogous to the Russell example of pairs of shoes, in that each A_n has distinctive elements. In the general case where the A_α are abstractly given sets, the possibility remains open that, even though each $A_\alpha \neq \varnothing$ still $\prod_\alpha A_\alpha = \varnothing$; to show that it is not empty requires that we exhibit a $c: \mathscr{A} \to \bigcup_\alpha A_\alpha$ having the property required in **9.1**, and this requires appeal to the axiom of choice. In fact,

9.2 Theorem The following three properties are equivalent:

(1). Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a nonempty family of sets. If each $A_\alpha \neq \varnothing$, then $\prod \{A_\alpha \mid \alpha \in \mathscr{A}\} \neq \varnothing$.
(2). The axiom of choice.
(3). If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a nonempty family of nonempty sets (not necessarily pairwise disjoint!), then there exists a map $c: \mathscr{A} \to \bigcup_\alpha A_\alpha$ such that $\forall \alpha \in \mathscr{A}: c(\alpha) \in A_\alpha$. ($c$ is called a "choice function" for the family $\{A_\alpha \mid \alpha \in \mathscr{A}\}$.)

Proof: (1) $\Rightarrow$ (2). Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of nonempty pairwise disjoint sets. Since $\prod_\alpha A_\alpha \neq \varnothing$, we can exhibit an element $c = \{a_\alpha\}$; then $S = c(\mathscr{A})$ is a set satisfying the requirement in the axiom of choice.

(2) $\Rightarrow$ (3). For each $\alpha \in \mathscr{A}$, let $A'_\alpha = \{\alpha\} \times A_\alpha$; each A'_α is a nonempty set (*cf.* **3.3** and **8.8**) and the family $\{A'_\alpha \mid \alpha \in \mathscr{A}\}$ is pairwise disjoint. By the axiom of choice, there is a set S consisting of exactly one member from each A'_α; that is, for each α there is a unique $(\alpha, a_\alpha) \in S$ with $a_\alpha \in A_\alpha$. Since

$$S \subset \bigcup_\alpha (\{\alpha\} \times A_\alpha) \subset \bigcup_\alpha (\mathscr{A} \times A_\alpha) = \mathscr{A} \times \bigcup_\alpha A_\alpha,$$

S is indeed a map $\mathscr{A} \to \bigcup_\alpha A_\alpha$ as required. Observe that the "chosen" elements may be the same for distinct α.

(3) $\Rightarrow$ (1). If $c: \mathscr{A} \to \bigcup_\alpha A_\alpha$ is a choice function, it is an element of $\prod_\alpha A_\alpha$.

We can now derive some further properties of cartesian products.

9.3 Theorem Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of nonempty sets, let $\mathscr{B} \subset \mathscr{A}$, and define $P: \prod_\alpha \{A_\alpha \mid \alpha \in \mathscr{A}\} \to \prod_\alpha \{A_\alpha \mid \alpha \in \mathscr{B}\}$ by $P(c) = c \mid \mathscr{B}$. Then P is surjective; in particular, each projection $p_\beta: \prod_\alpha A_\alpha \to A_\beta$ is surjective.

Proof: Let $f \in \prod\{A_\alpha \mid \alpha \in \mathscr{B}\}$ be any given element; we are to find a $c \in \prod\{A_\alpha \mid \alpha \in \mathscr{A}\}$ with $P(c) = f$. By **9.2**(3) there is a choice function $\bar{c}: \mathscr{A} - \mathscr{B} \to \cup \{A_\alpha \mid \alpha \in \mathscr{A} - \mathscr{B}\}$; then the map $c: \mathscr{A} \to \cup \{A_\alpha \mid \alpha \in \mathscr{A}\}$ given by $c \mid \mathscr{B} = f$, $c \mid \mathscr{A} - \mathscr{B} = \bar{c}$ (*cf.* **6.8**) is an element of

$$\prod\{A_\alpha \mid \alpha \in \mathscr{A}\}$$

and $P(c) = c|\mathscr{B} = f$. If $\mathscr{B}$ consists of the single element $\beta \in \mathscr{A}$, it is clear that the map P will be p_β, which proves the second part.

9.4 Corollary If $A_\alpha \subset B_\alpha$, for each $\alpha \in \mathscr{A}$, then $\prod_\alpha A_\alpha \subset \prod_\alpha B_\alpha$. Conversely, if each $A_\alpha \neq \varnothing$ and $\prod_\alpha A_\alpha \subset \prod_\alpha B_\alpha$, then $A_\alpha \subset B_\alpha$ for each α.

Proof: The first assertion is trivial; the second follows because, since each p_β is surjective,

$$A_\beta = p_\beta \prod_\alpha A_\alpha \subset p_\beta \prod_\alpha B_\alpha = B_\beta.$$

For the Boolean algebra of cartesian products, we have the following theorem, whose proof is left for the reader:

9.5 Theorem Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of nonempty sets; for each $\alpha \in \mathscr{A}$, let A_α, B_α be subsets of Y_α. Then

(1). $\prod_\alpha A_\alpha \cap \prod_\alpha B_\alpha = \prod_\alpha (A_\alpha \cap B_\alpha)$.

(2). $\prod_\alpha A_\alpha \cup \prod_\alpha B_\alpha \subset \prod_\alpha (A_\alpha \cup B_\alpha)$.

For a given α and $C_\alpha \subset Y_\alpha$, we denote $p_\alpha^{-1}(C_\alpha)$ by $\langle C_\alpha \rangle$; this is the "slab" in $\prod_\alpha Y_\alpha$ where each factor is Y_α except the αth, which is C_α. Similarly, for finitely many indices $\alpha_1, \cdots, \alpha_n$ and sets

$$C_{\alpha_1} \subset Y_{\alpha_1}, \cdots, C_{\alpha_n} \subset Y_{\alpha_n},$$

the subset $\langle C_{\alpha_1} \rangle \cap \cdots \cap \langle C_{\alpha_n} \rangle = p_{\alpha_1}^{-1}(C_{\alpha_1}) \cap \cdots \cap p_{\alpha_n}^{-1}(C_{\alpha_n})$ is denoted by $\langle C_{\alpha_1}, \cdots, C_{\alpha_n} \rangle$.

9.6 Corollary In $\prod\{Y_\alpha \mid \alpha \in \mathscr{A}\}$,

(1). $\prod_\alpha C_\alpha = \bigcap_\alpha \langle C_\alpha \rangle$,

(2). $\mathscr{C}\langle C_\alpha \rangle = \langle \mathscr{C} C_\alpha \rangle$,

(3). $\mathscr{C} \prod_\alpha C_\alpha = \bigcup_\alpha \langle \mathscr{C} C_\alpha \rangle$.

Proof:

(1). $c \in \prod_\alpha C_\alpha \Leftrightarrow \forall \alpha: p_\alpha(c) \in C_\alpha \Leftrightarrow \forall \alpha: c \in p_\alpha^{-1}(C_\alpha) \Leftrightarrow c \in \bigcap_\alpha \langle C_\alpha \rangle$.

(2) is proved similarly, and (3) follows from (1) and (2), using **4.2**(2).

We now establish the general distributive law of $\bigcap_\alpha$ over $\bigcup_\beta$:

9.7 Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of sets, and assume that $\{\mathscr{A}_\lambda \mid \lambda \in \mathscr{L}\}$ is a partition of $\mathscr{A}$ with each $\mathscr{A}_\lambda \neq \varnothing$. Let $T = \prod \{\mathscr{A}_\lambda \mid \lambda \in \mathscr{L}\}$. Then

$$\bigcap_{\lambda \in \mathscr{L}} [\cup \{A_\alpha \mid \alpha \in \mathscr{A}_\lambda\}] = \bigcup_{t \in T} [\cap \{A_{t(\lambda)} \mid \lambda \in \mathscr{L}\}].$$

Proof: We have

$$\begin{aligned} x \in \bigcup_{t \in T} [\cap \{A_{t(\lambda)} \mid \lambda \in \mathscr{L}\}] &\Leftrightarrow \exists\, t \in T: x \in \cap \{A_{t(\lambda)} \mid \lambda \in \mathscr{L}\} \\ &\Leftrightarrow \exists\, t \in T\ \forall\, \lambda \in \mathscr{L}: x \in A_{t(\lambda)} \\ &\Rightarrow \forall\, \lambda \in \mathscr{L}\ \exists\, \alpha \in \mathscr{A}_\lambda: x \in A_\alpha \\ &\Leftrightarrow \forall\, \lambda \in \mathscr{L}: x \in \cup \{A_\alpha \mid \alpha \in \mathscr{A}_\lambda\} \\ &\Leftrightarrow x \in \bigcap_\lambda [\cup \{A_\alpha \mid \alpha \in \mathscr{A}_\lambda\}] \end{aligned}$$

To establish the converse inclusion, we need only reverse the single implication. Now, from $\forall\, \lambda \in \mathscr{L}\ \exists\, \alpha \in \mathscr{A}_\lambda: x \in A_\alpha$, we find that the set

$$\mathscr{B}_\lambda = \{\alpha \in \mathscr{A}_\lambda \mid x \in A_\alpha\} \neq \varnothing$$

for each $\lambda \in \mathscr{L}$, so by **9.2**(1) there is a $t \in \prod_\lambda \mathscr{B}_\lambda \subset \prod_\lambda \mathscr{A}_\lambda$; this shows that $\exists\, t \in T\ \forall\, \lambda \in \mathscr{L}: x \in A_{t(\lambda)}$ and the proof is complete. By "complements," we obtain

$$\bigcup_\lambda [\cap \{A_\alpha \mid \alpha \in \mathscr{A}_\lambda\}] = \bigcap_{t \in T} [\cup \{A_{t(\lambda)} \mid \lambda \in \mathscr{L}\}].$$

9.8 Remark We have seen that **9.2**(1) implies **9.7**; we now show that, conversely, **9.7** implies **9.2**(1), so that the validity of the general distributive law is in fact equivalent to that of the axiom of choice.

Let $\{\mathscr{A}_\lambda \mid \lambda \in \mathscr{L}\}$ be a nonempty family of nonempty sets, and let $T = \prod \{\mathscr{A}_\lambda \mid \lambda \in \mathscr{L}\}$. For each $\alpha \in \bigcup_\lambda \mathscr{A}_\lambda$, let $A_\alpha = \{\varnothing\} \neq \varnothing$; it is clear that $\bigcap_\lambda [\cup \{A_\alpha \mid \alpha \in \mathscr{A}_\lambda\}] = \{\varnothing\}$ so that by **9.7**, also $\bigcup_{t \in T} [\cap \{A_{t(\lambda)} \mid \lambda \in \mathscr{L}\}] \neq \varnothing$, and it follows easily that therefore $T \neq \varnothing$.

Problems

Section 1

1. Prove: $\{a\} = \{b, c\}$ if and only if $a = b = c$.
2. If a, b, c, d are any objects, show that $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$ if and only if both $a = c$ and $b = d$.
3. Show that $A \subset \{A\}$ if and only if $A = \varnothing$.

4. Though the relation "$\subset$" is transitive, give an example to show that "$\in$" is not transitive.
5. Let $A = \{a_1, \cdots, a_n\}$. Show that A has 2^n subsets (do not forget to count the null subset).

Section 2

1. Let $A_q = \{n \in N \mid n \text{ is divisible by } q\}$. What is $A_q \cup A_r$, $A_q \cap A_r$?
2. Let A, B be subsets of E. Show:
   a. $A \cap B = \varnothing \Leftrightarrow A \subset \mathscr{C}_E B \Leftrightarrow B \subset \mathscr{C}_E A$.
   b. $A \cup B = E \Leftrightarrow \mathscr{C}_E B \subset A \Leftrightarrow \mathscr{C}_E A \subset B$.
3. For any two sets A, B, show:
   a. $A = (A \cap B) \cup (A - B)$ is a representation of A as a disjoint union.
   b. $A \cup B = A \cup (B - A)$ is a representation of $A \cup B$ as a disjoint union.

Verify the following formulas:

4. $(A - C) - (B - C) = (A - B) - C$.
5. $(A - C) \cup (B - C) = (A \cup B) - C$.
6. $(A - C) \cap (B - C) = (A \cap B) - C$.
7. $(A - B) - (A - C) = A \cap (C - B)$.
8. $(A - B) \cup (A - C) = A - (B \cap C)$.
9. $(A - B) \cap (A - C) = A - (B \cup C)$.
10. $A_1 \cup \cdots \cup A_n = (A_1 - A_2) \cup \cdots \cup (A_{n-1} - A_n) \cup (A_n - A_1) \cup (\bigcap_1^n A_i)$.
11. Prove that the system of equations $A \cup X = A \cup B$, $A \cap X = \varnothing$ has at most one solution for X.
12. The set $(A - B) \cup (B - A)$ is called the symmetric difference, or discrepancy, of A and B. Give a geometric interpretation of this set.

Section 3

1. Prove: If A, B are nonempty sets and $(A \times B) \cup (B \times A) = C \times C$, then $A = B = C$.
2. Let $A, B \subset X$ and $C, D \subset Y$. Prove:
   a. $(A \times C) \cap (B \times D) = (A \cap B) \times (C \cap D)$.
   b. $(A \times C) \cup (B \times D) \subset (A \cup B) \times (C \cup D)$; show that, in general, equality does not hold, by verifying
   $(A \cup B) \times (C \cup D) = (A \times C) \cup (B \times D) \cup (A \times D) \cup (B \times C)$.
   c. $\mathscr{C}_{X \times Y}(B \times D) = (\mathscr{C}_X B) \times Y \cup X \times \mathscr{C}_Y D$.

Section 4

1. Show that $\bigcap_1^\infty]-(1/n), 1 + (1/n)[= \bigcap_1^\infty [-(1/n), 1 + (1/n)] = [0, 1]$.
2. Let $\{A_n \mid n \in N\}$ be a family of sets. Let $S_k = \bigcup_0^k A_i$, $k = 0, 1, \cdots$. Show $\bigcup_0^\infty A_n = A_0 \cup (A_1 - S_0) \cup \cdots \cup (A_n - S_{n-1}) \cup \cdots$ and that this is a pairwise disjoint union.

3. Let $\{A_n \mid n \in N\}$ be a family of subsets of a set Γ. Define

$$\text{Lim sup } A_n = \bigcap_{n=0}^{\infty} \Big(\bigcup_{k=0}^{\infty} A_{n+k} \Big),$$

$$\text{Lim inf } A_n = \bigcup_{n=0}^{\infty} \Big(\bigcap_{k=0}^{\infty} A_{n+k} \Big).$$

Prove:

a. $\text{Lim sup } A_n = \{x \in \Gamma \mid x \text{ belongs to infinitely many } A_i\}$.
b. $\text{Lim inf } A_n = \{x \in \Gamma \mid x \text{ belongs to all but at most finitely many } A_i\}$.
c. $\bigcap_1^{\infty} A_\iota \subset \text{Lim inf } A_n \subset \text{Lim sup } A_n \subset \bigcup_1^{\infty} A_\iota$.
d. $\text{Lim inf } \mathscr{C}A_n = \mathscr{C}[\text{Lim sup } A_n]$, complement being with respect to any set containing all A_ι.
e. $\text{Lim inf } A_n \cup \text{Lim inf } B_n \subset \text{Lim inf } (A_n \cup B_n)$, and equality holds if $\cup$ is everywhere replaced by $\cap$.
f. $\text{Lim sup } (A_n \cap B_n) \subset \text{Lim sup } A_n \cap \text{Lim sup } B_n$, and equality holds if $\cap$ is everywhere replaced by $\cup$.
g. If $A_1 \subset A_2 \subset \cdots$ or $A_1 \supset A_2 \supset \cdots$, then $\text{Lim sup } A_n = \text{Lim inf } A_n$.

4. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of sets in Γ. Prove: $\bigcup_\alpha A_\alpha = \varnothing$ if and only if either each $A_\alpha = \varnothing$ or $\mathscr{A} = \varnothing$.

Section 5

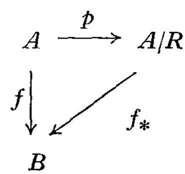
1. Prove: $A \subset B \Rightarrow \mathscr{P}(A) \subset \mathscr{P}(B)$.
2. Prove: $\cap \{A \mid A \in \mathscr{P}(E)\} = \varnothing$.

Section 6

1. Let $A \subset X$ and $f\colon X \to Y$. Let $i\colon A \to X$ be the map $a \to a$. Show:
   a. $f \mid A = f \circ i$.
   b. Writing $g = f \mid A$, $g^{-1}(B) = A \cap f^{-1}(B)$.
2. Show $f(A \cap B) = f(A) \cap f(B)$ for all $A, B \subset X$ if and only if f is injective.
3. Let $f\colon X \to Y$. Show:
   a. $A \subset B \Rightarrow f(A) \subset f(B)$.
   b. $f^{-1}(\mathscr{C}D) = \mathscr{C}f^{-1}(D)$.
4. Let $f\colon X \to Y$. Prove:
   a. f is injective $\Leftrightarrow \forall y \in Y\colon f^{-1}(y) = \varnothing$ or a single point $\Leftrightarrow \forall\, A\colon f(\mathscr{C}A) \subset \mathscr{C}f(A)$.
   b. f is surjective $\Leftrightarrow \forall y \in Y\colon f^{-1}(y) \neq \varnothing \Leftrightarrow \forall\, A\colon f(\mathscr{C}A) \supset \mathscr{C}f(A)$.
5. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$, $\{B_\beta \mid \beta \in \mathscr{B}\}$ be two coverings (partitions) of X. Show that $\{A_\alpha \cap B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$ is also a covering (partition) of X.
6. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$, $\{B_\beta \mid \beta \in \mathscr{B}\}$ be coverings (partitions) of sets X and Y, respectively. Show $\{A_\alpha \times B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$ is a covering (partition) of $X \times Y$.
7. Let $f\colon X \to Y$ and $g\colon Y \to X$ be any two maps. Show that X and Y can each be expressed as *disjoint* unions: $X = X_1 \cup X_2$, $Y = Y_1 \cup Y_2$, such that $f(X_1) = Y_1$ and $g(Y_2) = X_2$.
(*Hint:* For each $E \subset X$, let $Q(E) = X - g[Y - f(E)]$, and take $X_1 = \bigcap \{Q(E) \mid Q(E) \subset E\}$.)

Section 7

1. For relations R, S in A, define $R \circ S$ by $a\, R \circ S\, b \Leftrightarrow \exists\, c\colon (a\, R\, c) \wedge (c\, S\, b)$. Show that $R \circ (S \circ T) = (R \circ S) \circ T$. If each of R, S are equivalence relations, is $R \circ S$ an equivalence relation?
2. For any given R, define R^{-1} by $a\, R^{-1}\, b \Leftrightarrow b\, R\, a$. Show that a reflexive R is an equivalence relation if and only if $R \circ R = R$ and $R = R^{-1}$.
3. If R is any reflexive and transitive relation, show that $R \cap R^{-1}$ is an equivalence relation.
4. For relations R, S in A, B, respectively, define $R \times S$ in $A \times B$ by
$$(a, b)\, R \times S\, (c, d) \Leftrightarrow (a\, R\, c) \wedge (b\, S\, d).$$
If R, S are equivalence relations, show that $R \times S$ is an equivalence relation.
5. Let $f\colon A \to B$. Show that $a\, R\, b \Leftrightarrow f(a) = f(b)$ is an equivalence relation on A and that there is an $f_*\colon A/R \to B$ such that the diagram



commutes.
6. Let S, R be two equivalence relations in A, with $S \subset R$. Let $1_*\colon A/S \to A/R$ be the map induced by the relation-preserving map 1_A. Define $(S\, a) R/S (S\, b)$ if $1_*(S\, a) = 1_*(S\, b)$. Show that R/S is an equivalence relation and that there is a bijection of $(A/S)/(R/S)$ onto A/R.

Section 9

1. Prove the extended version of **9.5(1)**:
$$\bigcap_{\rho} \Big(\prod_{\alpha} A_{\alpha,\rho}\Big) = \prod_{\alpha} \Big(\bigcap_{\rho} A_{\alpha,\rho}\Big).$$
2. Prove $\prod_{\alpha} A_\alpha - \prod_{\alpha} B_\alpha = \bigcup_{\alpha} Q_\alpha$, where in each Q_β, each factor $\alpha \neq \beta$ is A_α, and the βth factor is $A_\beta - B_\beta$.
3. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of nonempty sets, and let $\mathscr{A} = \cup\{\mathscr{A}_\beta \mid \beta \in \mathscr{B}\}$ be a partition of $\mathscr{A}$. Construct a bijective map of $\prod\{A_\alpha \mid \alpha \in \mathscr{A}\}$ onto
$$\prod_{\beta} \{\prod \{A_\alpha \mid \alpha \in \mathscr{A}_\beta\}\}.$$

Ordinals and Cardinals

II

1. Orderings

Certain types of binary relations are called orderings.

1.1 Definition A binary relation R in a set A is called a *preorder* if it is reflexive and transitive; that is, if

(1). $\forall a: a\, R\, a.$
(2). $(a\, R\, b) \wedge (b\, R\, c) \Rightarrow a\, R\, c.$

A set together with a definite preorder is called a preordered set.

We denote a preorder R by $\prec$; "$a \prec b$" is read "a precedes b," or "b follows a." A preordered set A will also be written $(A, \prec)$ whenever it is necessary to indicate explicitly the preorder being used. If A is preordered and $B \subset A$, the induced relation on B is clearly a preordering; when we consider B itself as a preordered set, it will always be with the induced preordering.

Ex. 1 In any set A, the relation $\{(a, b) \mid a = b\}$ is a preordering. Note that we do *not* require each pair of elements $a, b \in A$ to be related; that is, to satisfy either $a \prec b$ or $b \prec a$.

Ex. 2 In E^1, the relation $\{(x, y) \mid x \leqslant y\}$ is a preordering, whereas $\{(x, y) \mid x < y\}$ is not.

Ex. 3 Let C be the set of complex numbers, and define $z_1 \prec z_2$ if and only if $|z_1| \leqslant |z_2|$. This is a preordering on C. Observe that we do *not* require that $(a \prec b) \wedge (b \prec a)$ imply $a = b$.

Ex. 4 In $\mathscr{P}(X)$, the relation $A \prec B$, defined by $A \prec B$ if and only if $A \subset B$, is a preordering. More generally, any family of sets preordered in this manner is said to be *preordered by inclusion.*

There is a standard terminology pertaining to preordered sets:

1.2 Definition Let $(A, \prec)$ be preordered.

(1). $m \in A$ is called a maximal element in A if $\forall a: m \prec a \Rightarrow a \prec m$; that is, if either no $a \in A$ follows m or each a that follows m also precedes m.

(2). $a_0 \in A$ is called an upper bound for a subset $B \subset A$ if $\forall b \in B: b \prec a_0$.

(3). $B \subset A$ is called a *chain* in A if each two elements in B are related.

Ex. 5 In Ex. 1, each element is maximal. No subset of A containing at least two elements has an upper bound. Thus a maximal element in A need not be an upper bound for A.

Ex. 6 In Ex. 2, there is no maximal element. Every bounded set has many upper bounds.

Ex. 7 In Ex. 3, $a_0 = 1$ is an upper bound for $B = \{z \mid |z| \leqslant 1\} \subset C$, but is not maximal in C. Note also that although a_0 is an upper bound for B, it is possible for some $b \in B$ also to satisfy $a_0 \prec b$. In the set B (with the induced preorder!) each z with $|z| = 1$ is both maximal and an upper bound for B.

Placing additional requirements on preorderings gives other types of ordering relations.

1.3 Definition A preordering in A with the additional property

$$(a \prec b) \wedge (b \prec a) \Rightarrow (a = b) \qquad \text{(antisymmetry)}$$

is called a *partial ordering.* A set together with a definite partial ordering is called a partially ordered set. A partially ordered set that is also a chain is called a *totally ordered set.*

It is evident that partial (total) orders induce partial (total) orders on subsets.

Ex. 8 Ordering by inclusion in $\mathscr{P}(X)$, or any family of sets, is always a partial ordering (but, clearly, need not be a total order).

Ex. 9 Let $(A, \prec)$ be a preordered set, and define a relation S in A by $a\, S\, b \Leftrightarrow (a \prec b) \wedge (b \prec a)$. It is easy to verify that S is an equivalence relation and that A/S is *partially* ordered by $S\,a \prec S\,b \Leftrightarrow a \prec b$.

Note that in a partially ordered A, the statement "m is maximal" is equivalent to "each $a \in A$ is either not related to m or satisfies only $a \prec m$," and also that if a_0 is an upper bound for $B \subset A$, then there can be no $b \in B - \{a_0\}$ with $a_0 \prec b$.

Total orderings of the type in the following definition are very important, as we shall see.

1.4 Definition A partially ordered set W is called well-ordered (or an ordinal) if each nonempty subset $B \subset W$ has a first element; that is, for each $B \neq \varnothing$, there exists a $b_0 \in B$ satisfying $b_0 \prec b$ for each $b \in B$.

Every well-ordered set W is in fact totally ordered, since each subset $\{a, b\} \subset W$ has a first element; furthermore, the induced order on a subset of a well-ordered set is a well-order on that subset.

Ex. 10 $\varnothing$ is a well-ordered set. In any set $\{a\}$ containing exactly one element, $a \prec a$ is a well-ordering. The partial ordering by inclusion in $\mathscr{P}(X)$ is not a well-ordering if X has more than one element.

Ex. 11 The nonnegative integers are well-ordered: this is one of the ways to state the principle of mathematical induction. This well-ordered set is denoted by ω; that is, $\omega = (N, \leqslant)$.

Let W be well-ordered, and $q \bar{\in} W$; in $W \cup \{q\}$ define an order that coincides with the given one on W, satisfies $q \prec q$, and $\forall\, w$: $(w \in W) \Rightarrow w \prec q$. Then $W \cup \{q\}$ is well-ordered, since for each nonempty $E \subset W \cup \{q\}$, either $E = \{q\}$ or $E \cap W \neq \varnothing$, and in the latter case, the first element in $E \cap W$ is the first in $E \subset W \cup \{q\}$. We say $W \cup \{q\}$ is formed from W by adjoining q as last element.

Each element w of a well-ordered set that has a successor in the set, has an immediate successor; that is, we can find an $s \neq w$ satisfying $w \prec s$ and such that no $c \neq s, w$ satisfies $w \prec c \prec s$: we need only choose $s =$ first element in the nonempty set $\{x \in W \mid (w \prec x) \wedge (w \neq x)\}$. However, an element w need *not* have an immediate predecessor: for each $b \prec w$ there may always be some $c \neq b, w$ with $b \prec c \prec w$. In fact, adjoining to ω a last element q, we note that q has no immediate predecessor.

2. Zorn's Lemma; Zermelo's Theorem

Our objective in this section is to prove the fundamental

2.1 Theorem The following three statements are equivalent:

(1). The axiom of choice.

(2). Zorn's lemma: Let X be a preordered set. If each chain in X has an upper bound, then X has at least one maximal element.

(3). Zermelo's theorem: Every set can be well ordered.

Before we begin to prove this theorem, we must first study the auxiliary concept of a "tower."

2.2 Definition Let X be a set, let $\mathscr{F} \subset \mathscr{P}(X)$ be any nonempty family of sets, and let $\varphi: \mathscr{F} \to X$ be a fixed map. The family $\mathscr{F}$ is called a φ-tower whenever:

(a). $\varnothing \in \mathscr{F}$.

(b). If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is any totally ordered (by inclusion) family of sets in $\mathscr{F}$, then $\bigcup_\alpha A_\alpha \in \mathscr{F}$.

(c). If $A \in \mathscr{F}$, then also $A \cup \{\varphi(A)\} \in \mathscr{F}$.

Observe that if $\{\mathscr{F}_\beta \mid \beta \in \mathscr{B}\}$ is any family of φ-towers in $\mathscr{P}(X)$, then $\bigcap_\beta \mathscr{F}_\beta$ is also a φ-tower. Indeed, (a) $\varnothing \in \bigcap_\beta \mathscr{F}_\beta$ is clear; (b) any chain $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ in $\bigcap_\beta \mathscr{F}_\beta$, by virtue of its members belonging to each $\mathscr{F}_\beta$, has its union in each $\mathscr{F}_\beta$, so $\bigcup_\alpha A_\alpha \in \bigcap_\beta \mathscr{F}_\beta$; and (c) is verified analogously. In particular, then, the intersection $\mathscr{M}$ of all sub φ-towers of a given φ-tower is itself a φ-tower. $\mathscr{M}$ is minimal in the sense that it contains no proper sub φ-tower.

The proof of **2.1** will lean heavily on the

2.3 Lemma If $\mathscr{F}$ is a φ-tower, then there exists an $A \in \mathscr{F}$ such that $\varphi(A) \in A$.

Proof of Lemma: Let $\mathscr{M} \subset \mathscr{F}$ be the minimal φ-tower. We wish to show that $\mathscr{M}$ is totally ordered by inclusion. For, by defining

$$A = \cup \{M \mid M \in \mathscr{M}\},$$

we would then have $A \in \mathscr{M}$ and also $A \cup \{\varphi(A)\} \in \mathscr{M}$, which would show $A \cup \{\varphi(A)\} \subset A$ and consequently that $\varphi(A) \in A$, as required. To prove that $\mathscr{M}$ is totally ordered, we proceed formally: calling $M \in \mathscr{M}$ medial whenever, for each $M_\alpha \in \mathscr{M}$, either $M_\alpha \subset M$ or $M \subset M_\alpha$ is true, we are to show that each $M \in \mathscr{M}$ is medial. To do this, we first establish an important property of medial sets:

I. Let $M \in \mathscr{M}$ be medial. Then, for each $M_\alpha \in \mathscr{M}$, either $M_\alpha \subset M$, or $M \cup \{\varphi(M)\} \subset M_\alpha$; that is, any $M_\alpha \in \mathscr{M}$ either lies in M or contains $M \cup \{\varphi(M)\}$.

Proof of I: Let $\mathscr{G}_M = \{M_\alpha \in \mathscr{M} \mid M_\alpha \subset M \text{ or } M \cup \{\varphi(M)\} \subset M_\alpha\}$. To show $\mathscr{G}_M = \mathscr{M}$, it suffices to prove that $\mathscr{G}_M$ is a φ-tower, since $\mathscr{G}_M \subset \mathscr{M}$ and $\mathscr{M}$ contains no proper sub φ-tower.

(a). $\varnothing \in \mathscr{G}_M$, since $\varnothing \subset M$; in particular, $\mathscr{G}_M$ is not empty.

(b). Let $\{M_\alpha \mid \alpha \in \mathscr{A}\}$ be a chain in $\mathscr{G}_M$. Since each $M_\alpha \in \mathscr{G}_M$, there are two cases:

(i). $\exists\, \alpha \in \mathscr{A}: M \cup \{\varphi(M)\} \subset M_\alpha$.

(ii). $\forall\, \alpha \in \mathscr{A}: M_\alpha \subset M$.

In case (i), $M \cup \{\varphi(M)\} \subset \bigcup_\alpha M_\alpha$; and in case (ii), $\bigcup_\alpha M_\alpha \subset M$, so always $\bigcup_\alpha M_\alpha \in \mathscr{G}_M$.

(c). Let $M_\alpha \in \mathscr{G}_M$. We must show $M_\alpha \cup \{\varphi(M_\alpha)\} \in \mathscr{G}_M$; that is, either $M_\alpha \cup \{\varphi(M_\alpha)\} \subset M$ or $M \cup \{\varphi(M)\} \subset M_\alpha \cup \{\varphi(M_\alpha)\}$. Because $M_\alpha \in \mathscr{G}_M$, there are two cases to consider:

(i). $M \cup \{\varphi(M)\} \subset M_\alpha$.

(ii). $M_\alpha \subset M$.

In case (i), $M \cup \{\varphi(M)\} \subset M_\alpha \subset M_\alpha \cup \{\varphi(M_\alpha)\}$; therefore one of the desired alternatives is true. In case (ii), because M is medial and $M_\alpha \cup \{\varphi(M_\alpha)\} \in \mathscr{M}$, we have in addition to (ii) that either $M_\alpha \cup \{\varphi(M_\alpha)\} \subset M$ or $M \subset M_\alpha \cup \{\varphi(M_\alpha)\}$. In the first of these cases, we are through; in the second case, (ii) gives $M_\alpha \subset M \subset M_\alpha \cup \{\varphi(M_\alpha)\}$ and because $\varphi(M_\alpha)$ is a single point, we conclude either that $M_\alpha = M$ or that $M = M_\alpha \cup \{\varphi(M_\alpha)\}$; each possibility shows $M_\alpha \cup \{\varphi(M_\alpha)\} \in \mathscr{G}_M$.

Thus $\mathscr{G}_M$ is a φ-tower, and I has been proved.

We finally show that $\mathscr{M}$ is totally ordered:

II. Each $M \in \mathscr{M}$ is medial.

Proof of II: Let $\mathscr{G} = \{M \in \mathscr{M} \mid M \text{ is medial}\}$; again we need only prove that $\mathscr{G}$ is a φ-tower, to assure $\mathscr{G} = \mathscr{M}$.

(a). $\varnothing \in \mathscr{G}$, as is clear; in particular, $\mathscr{G}$ is not empty.

(b). Let $\{M_\alpha \mid \alpha \in \mathscr{A}\}$ be a chain in $\mathscr{G}$. If M is any member of $\mathscr{M}$, then because each M_α is medial, either $\forall\, \alpha \in \mathscr{A}: M_\alpha \subset M$ or $\exists\, \alpha \in \mathscr{A}: M \subset M_\alpha$; thus either $\bigcup_\alpha M_\alpha \subset M$ or $M \subset \bigcup_\alpha M_\alpha$; that is, $\bigcup_\alpha M_\alpha \in \mathscr{G}$.

(c). Let $M \in \mathscr{G}$; to show that $M \cup \{\varphi(M)\}$ is medial, choose any $M_\alpha \in \mathscr{M}$. Because M is medial, (I) says that either $M_\alpha \subset M$ (so that $M_\alpha \subset M \cup \{\varphi(M)\}$) or $M \cup \{\varphi(M)\} \subset M_\alpha$; this shows $M \cup \{\varphi(M)\}$ is medial.

Thus $\mathscr{G} = \mathscr{M}$, and **2.3** has been proved.

We now prove the theorem

Proof of Theorem 2.1: (1) $\Rightarrow$ (2). Let $\mathscr{F} \subset \mathscr{P}(X)$ be the set of all chains in X; for each nonempty $A \in \mathscr{F}$, let a_A be an upper bound and define $T_A = \{x \in X \mid (a_A \prec x) \wedge \neg(x \prec a_A)\}$. To prove (2), it suffices to show some $T_A = \varnothing$, for this means that either there is no x with $a_A \prec x$ or that each x satisfying $a_A \prec x$ also satisfies $x \prec a_A$, that is, a_A is maximal.

We argue by contradiction. Assume all $T_A \neq \varnothing$; then, by I, **9.2**, there is a choice function $c\colon \mathscr{F} \to \bigcup \{T_A \mid A \in \mathscr{F}\}$ with $c(A) \in T_A$ for each $A \in \mathscr{F}$. Defining $c(\varnothing) = x_0$ arbitrarily, we will now prove that $\mathscr{F}' = \mathscr{F} \cup \{\varnothing\}$ is a c-tower.

(a). $\varnothing \in \mathscr{F}'$.

(b). If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is any chain in $\mathscr{F}'$, then $\bigcup_\alpha A_\alpha \in \mathscr{F}'$; that is, any two elements $a, b \in \bigcup_\alpha A_\alpha$ are related. In fact, we can find two sets A_α, A_β with $a \in A_\alpha$, $b \in A_\beta$; since $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a chain, one of these (say, A_α) contains the other, so that a, b are contained in the single set A_α and consequently are related. Thus $\bigcup_\alpha A_\alpha \in \mathscr{F}'$.

(c). Let $A \in \mathscr{F}'$; then $A \cup \{c(A)\} \in \mathscr{F}'$. We need show only that each $a \in A$ is related to $c(A)$. For this, we have $a \prec a_A$ (since a_A is an upper bound for A) and $a_A \prec c(A)$ since $c(A) \in T_A$.

Applying Lemma **2.3**, there is some $A \in \mathscr{F}'$ with $c(A) \in A$. This is the desired contradiction: for, since a_A is an upper bound for A, necessarily $c(A) \prec a_A$, whereas $c(A) \in T_A$ requires $\neg(c(A) \prec a_A)$. Thus some $T_A = \varnothing$, and therefore a_A is maximal in X.

(2) $\Rightarrow$ (3). Let E be any set. Clearly, there are subsets of E that can be well-ordered ($\varnothing$, for example). We are to show that E itself is such a set. To this end, we work with the family $\mathscr{F}$ of all ordered couples $\{(A, \prec_A) \mid (A \subset E) \wedge (\prec_A \text{ is a well-ordering of } A)\}$.

Preorder $\mathscr{F}$ by "propagation," defining $(A, \prec_A) \prec (B, \prec_B)$ if:

(a). $A \subset B$.
(b). $\prec_B$ induces $\prec_A$ on A.
(c). $(y \in B - A) \wedge (x \in A) \Rightarrow x \prec_B y$.

Transitivity of $\prec$ is obvious, and because of (a), (b), $(\mathscr{F}, \prec)$, is in fact a *partially ordered* set. We first prove that $(\mathscr{F}, \prec)$ satisfies the requirements of Zorn's lemma.

Let $\{(A_\alpha, \prec_\alpha) \mid \alpha \in \mathscr{A}\}$ be any chain in $(\mathscr{F}, \prec)$; we will show $\bigcup_\alpha A_\alpha$ has a well-order $\prec_U$ such that $(\bigcup_\alpha A, \prec_U)$ is an upper bound for this chain. Define $\prec_U$ as follows: Given two elements $a, b \in \bigcup_\alpha A_\alpha$, observe that because of (a), $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is a chain under set-inclusion so that

a, b are contained in a common A_α. Selecting any such A_α, define $a \prec_U b$ if and only if $a \prec_\alpha b$. This definition is independent of the A_α used because of (b) and is evidently a partial order on $\bigcup_\alpha A_\alpha$. To prove that it is a well-ordering, we note first that because of (a) and (c), $\prec_U$ also has the property $(y \in A_\alpha) \wedge (x \prec_U y) \Rightarrow (x \in A_\alpha)$. Thus, if $Q \subset \bigcup_\alpha A_\alpha$ is nonempty, there is an index α with $Q \cap A_\alpha \neq \varnothing$, and the first element in $Q \cap A_\alpha$ is first in Q. Clearly, each $(A_\alpha, \prec_\alpha) \prec (\bigcup_\alpha A_\alpha, \prec_U)$, so that we have an upper bound.

By Zorn's lemma, there is a maximal $(A, \prec_A)$ in $(\mathscr{F}, \prec)$. We assert that $A = E$. Otherwise, there is an $a_0 \in E - A$, which we could adjoin to A as last element, getting $(A, \prec_A) \prec (A \cup \{a_0\}, \prec')$; and, because $(\mathscr{F}, \prec)$ is *partially* ordered, this would contradict the maximality of $(A, \prec_A)$.

(3) $\Rightarrow$ (1). Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of nonempty sets. Well-order $\bigcup_\alpha A_\alpha$ and define $c(\alpha)$ = first element in A_α, to obtain a choice function.

2.4 Remarks Though Zermelo's theorem assures that every set can be well-ordered, no specific construction for well-ordering any uncountable set (say, the real numbers) is known. Furthermore, there are sets for which no specific construction of a total order (let alone a well-order) is known, for example, the set of real-valued functions of one real variable. Note, also, that a well-ordering guaranteed by Zermelo's theorem is obviously not unique, and is not stated to have any relation to any given structure on the set. For example, a well-ordering of the reals cannot coincide with the usual ordering.

2.5 Applications 1. Zorn's lemma is a particularly useful version of the axiom of choice: It is applicable for existence theorems whenever the underlying set is partially ordered and the required object is characterized by maximality [*cf.* the proof of (2) $\Rightarrow$ (3)]. As a simple example of its use, we prove the existence of a Hamel basis B for the real numbers.

A subset $B = \{b_\alpha \mid \alpha \in \mathscr{A}\} \subset E^1$ is a Hamel basis if (1) each real x can be written as a finite sum

$$x = \sum_1^n r_{\alpha_i} b_{\alpha_i}$$

with rational r_{α_i}; and (2) the set $\{b_\alpha \mid \alpha \in \mathscr{A}\}$ is rationally independent, that is,

$$\sum_1^n r_{\alpha_i} b_{\alpha_i} = 0 \Leftrightarrow r_{\alpha_i} = 0 \qquad \text{for each } i = 1, \cdots, n.$$

The rational independence assures that each x can be written as required by (1) in exactly one way.

To prove existence of B, let $\mathscr{A}$ be the family of all rationally independent sets of reals. $\mathscr{A} \neq \varnothing$, since, say, $\{1\} \in \mathscr{A}$. Partially order $\mathscr{A}$ by inclusion: Then any chain $\{A_\beta \mid \beta \in \mathscr{B}\}$ has an upper bound, $\bigcup_\beta A_\beta$, since any finite collection of elements in $\bigcup_\beta A_\beta$ lies in some one A_β and so is rationally independent. Thus there is a maximal $B \in \mathscr{A}$; B is a Hamel basis, since for each x, $B \cup \{x\}$ is not rationally independent and therefore there is a relation $rx + r_{\alpha_1}b_{\alpha_1} + \cdots r_{\alpha_n}b_{\alpha_n} = 0$, which necessarily involves x and has $r \neq 0$ because B is rationally independent, so that

$$x = -\sum_1^n \frac{r_{\alpha_i}}{r} b_{\alpha_i}$$

as required.

Hamel bases arise in several connections: If $b_1 \in B$, then the set of reals generated by $B - \{b_1\}$ is not a Lebesgue measurable set. Similarly, the functional equation $f(x + y) = f(x) + f(y)$, which has $f(x) = cx$ as its only continuous solutions, has others also, none of which is a Lebesgue measurable function.

2. A proof analogous to the preceding one shows that every vector space has a basis.

3. In a commutative ring R with unit, every ideal $\mathscr{I} \neq R$ is contained in a maximal ideal (Krull's theorem). The proof is immediate by verifying that the set of all ideals containing $\mathscr{I}$, and not containing 1, satisfies the requirement of Zorn's lemma.

Since in R every maximal ideal is a prime ideal, this has as immediate consequence the following result: Let X be an infinite set. Then there exists a μ: $\mathscr{P}(X) \to \{0\} \cup \{1\}$ such that $\mu(F) = 0$ if F is finite, $\mu(X) = 1$, and $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever A, B, are disjoint. [In the Boolean ring, $\mathscr{P}(X)$ (*cf.* I, **2.12**), let $\mathscr{F}$ be a maximal ideal containing the ideal of all finite subsets, and set $\mu(A) = 0$ if $A \in \mathscr{F}$, and $\mu(A) = 1$ otherwise.]

3. Ordinals

Since every set can be well-ordered, we study ordinals in greater detail.

3.1 Definition Let W be a well-ordered set.

(1). $S \subset W$ is an ideal in W if $\forall x$: $(x \in S) \wedge (y \prec x) \Rightarrow y \in S$.
(2). For each $a \in W$, the set $W(a) = \{x \in W \mid (x \prec a) \wedge (x \neq a)\}$ is the initial interval determined by a.

Clearly, W and $\varnothing$ are ideals in W; $\varnothing$ is also an initial interval, but W is not. For the properties and interrelation of these concepts we have

3.2 (a). Every intersection, and every union, of ideals in W is itself an ideal in W.

(b). Let $I(W)$ be the set of all ideals in W and $J(W)$ the set of all initial intervals in W. Then $J(W) = I(W) - \{W\}$: the ideals $\neq W$ are the initial intervals.

Proof: (a). $(x \in \bigcap_\alpha S_\alpha) \wedge (y \prec x) \Rightarrow \forall\, \alpha: (x \in S_\alpha) \wedge (y \prec x) \Rightarrow \forall\, \alpha: y \in S_\alpha \Rightarrow y \in \bigcap_\alpha S_\alpha$, and similarly for union.

(b). Each initial interval is obviously an ideal. Conversely, let $S \neq W$ be an ideal; then $W - S \neq \varnothing$, so it has a first element a. We prove $S = W(a)$.

(i). $x \in W(a) \Rightarrow x \in S$, since a is the first element of W not in S.

(ii). $x \,\bar{\in}\, W(a) \Leftrightarrow a \prec x \Rightarrow x \,\bar{\in}\, S$, since otherwise, because S is an ideal, we would have $a \in S$.

A map f of a well-ordered set $(W, \prec)$ into a well-ordered $(X, \prec')$ is called a *monomorphism* if it is an order-preserving injection [that is, $a \prec b \Rightarrow f(a) \prec' f(b)$]; f is an *isomorphism* if it is a bijective monomorphism. Clearly, the composition of two monomorphisms is also a monomorphism. It is easy to see that if $f: W \to A$ is an order-preserving bijection and if W is well-ordered, then the order in A is also a well-ordering and f is an isomorphism.

3.3 Theorem (1). The set $I(W)$ of all ideals of a well-ordered set is well-ordered by inclusion.

(2). The map $a \to W(a)$ is an isomorphism of W onto the set $J(W)$ of its initial intervals [$\varnothing$ included in $J(W) \subset I(W)$].

Proof: (2). Clearly, $a \prec b \Rightarrow W(a) \subset W(b)$ and $a \neq b \Rightarrow W(a) \neq W(b)$; thus $a \to W(a)$ is bijective and [using order by inclusion in $J(W)$] order-preserving. It follows at once that $J(W)$ is well-ordered by inclusion and that $a \to W(a)$ is an isomorphism.

(1). Since

$$I(W) = J(W) \cup \{W\}$$

and since the ordering by inclusion in $I(W)$ is determined by adjoining $\{W\}$ to $J(W)$ as last element, $I(W)$ is also well-ordered.

One consequence is the extremely useful

3.4 Theorem Let W be well-ordered, and $\Sigma \subset I(W)$ any family with the following properties:

(a). Each union of members of Σ belongs to Σ.
(b). If $W(a) \in \Sigma$, then also $W(a) \cup \{a\} \in \Sigma$.

Then $\Sigma = I(W)$ and, in particular, $W \in \Sigma$.

Proof: Assume $\Sigma \neq I(W)$; by **3.3** there is a smallest ideal $S \mathrel{\bar{\in}} \Sigma$. Either S has a last element, or it does not.

(i). If S has a last element, b, then $S = W(b) \cup \{b\}$; because $W(b) \subset S$, we would then have $W(b) \in \Sigma$ and, by using (b), that $S \in \Sigma$, which contradicts the definition of S.

(ii). If S does not have a last element, then $S = \bigcup \{W(a) \mid W(a) \subset S\}$; as before, each $W(a) \in \Sigma$, so by using (a), we conclude that $S \in \Sigma$, again contradicting the definition of S.

Thus, the assumption $\Sigma \neq I(W)$ is false, and the theorem has been proved.

4. Comparability of Ordinals

In this section we are going to show that given any two ordinals, one of them must be isomorphic to an ideal of the other.

4.1 Lemma Let W, X be well-ordered sets, and let $\varphi\colon W \to X$ be an isomorphism onto an *ideal* of X. Then any monomorphism $f\colon W \to X$ must satisfy the condition $\forall w\colon \varphi(w) \prec f(w)$. In particular, there can be at most one isomorphism between an ideal $S \subset W$ and an ideal $T \subset X$.

Proof: We show that the assumption

$$\{w \in W \mid (f(w) \prec \varphi(w)) \wedge (f(w) \neq \varphi(w))\} \neq \varnothing$$

implies that $\varphi(W)$ is not an ideal in X. For there would then be a smallest w_0 with $[f(w_0) \neq \varphi(w_0)] \wedge [f(w_0) \prec \varphi(w_0)]$, and φ cannot take on the value $f(w_0)$: indeed, if $w \prec w_0$, then $\varphi(w) \prec f(w) \prec f(w_0)$; and if $w_0 \prec w$, then $f(w_0) \prec \varphi(w_0) \prec \varphi(w)$. The second part is immediate.

4.2 Theorem (Comparability of Ordinals) Let W and X be well-ordered sets. Then one, and only one, of the following three statements is true:

(1). There is a unique isomorphism of W onto X.
(2). There is a unique isomorphism of W onto an initial interval of X.
(3). There is a unique isomorphism of X onto an initial interval of W.

Proof: We first show these three possibilities to be mutually exclusive. (1) $\wedge$ (2) [or (1) $\wedge$ (3)] cannot occur because this would contradict the lemma, whereas (2) $\wedge$ (3) cannot occur either, since two such maps

$g\colon X \to W(w_0)$ and $h\colon W \to X(x_0)$ would compose to yield a monomorphism $g \circ h\colon W \to W(w_0)$, which satisfies $g \circ h(w_0) \prec w_0$ and contradicts the lemma on comparison with the isomorphism $1\colon W \to W$. We now prove that one of these possibilities always holds by showing that one of X, W is isomorphic to an ideal of the other.

Let $\Sigma \subset I(W)$ be the set of all ideals for which there are isomorphisms onto ideals of X. We verify that Σ has the property (a) of **3.4**: Let $\bigcup_\alpha \{S_\alpha \mid \alpha \in \mathscr{A}\}$ be a union of members of Σ, and for each α, let

$$\varphi_\alpha\colon S_\alpha \to X$$

be the isomorphism onto an ideal of X. By **3.2**(a) and the lemma **4.1**, $\varphi_\alpha \mid S_\alpha \cap S_\beta = \varphi_\beta \mid S_\alpha \cap S_\beta$ for all $(\alpha, \beta) \in \mathscr{A} \times \mathscr{A}$, so (*cf.* I, **6.7**), there is a unique map $\varphi\colon \bigcup_\alpha S_\alpha \to \bigcup_\alpha \varphi_\alpha(S_\alpha)$, which is easily verified to be an isomorphism. Since by **3.2**(a) both $\bigcup_\alpha S_\alpha$ and $\bigcup_\alpha \varphi_\alpha(S_\alpha)$ are ideals, we find that $\bigcup_\alpha S_\alpha \in \Sigma$.

If $W \in \Sigma$, we have an isomorphism as required (that is, alternative 1 or 2 of the theorem). If $W \notin \Sigma$, then by **3.4** there must be some $S_\alpha = W(w_0)$ with $W(w_0) \cup \{w_0\} \notin \Sigma$. In this case, we must have $\varphi_\alpha(S_\alpha) = X$ (that is, alternative (3) of the theorem), for otherwise, $\varphi_\alpha(S_\alpha) = X(x_0)$ and extending φ_α by $\varphi(w_0) = x_0$ would give the contradiction $W(w_0) \cup \{w_0\} \in \Sigma$.
The uniqueness of the isomorphisms follows from the lemma.

For subsets of a well-ordered set, the following consequence is of importance:

4.3 Corollary (a). Any *subset* A of a well-ordered set W is isomorphic either to W or to an initial interval of W.
(b). No *initial interval* of W can be isomorphic to W.

Proof: (a). Regarding A as a well-ordered set, we show the possibility that W is isomorphic to an initial interval of A cannot occur. If $g\colon W \to A(a_0)$ were such an isomorphism, then, using the identity map $i\colon A \to W$, we would have a monomorphism $g \circ i\colon A \to A$ satisfying $g \circ i\,(a_0) \prec a_0$ and contradicting **4.1** on comparison with $1\colon A \to A$. (b) is a direct consequence of **4.2**.

Ex. 1 It is important to note that $A \subset W$ may be isomorphic to W: Taking $W = \omega$ and A as the subset of even integers, $2n \to n$ is such an isomorphism. However, no *initial interval* of W can be isomorphic to W.

4.4 Corollary The class of all well-ordered sets is well ordered, if we define $W \leqslant X$ to mean that W is isomorphic to an ideal of X (and $W = X$ means that W is isomorphic to X).

Proof: Clearly, $W \leqslant X$ is a preordering. To see that it is a partial ordering, observe that $(X \leqslant W) \wedge (W \leqslant X) \Rightarrow W = X$, since by **4.2**, one set isomorphic to an initial interval of another excludes the possibility of the second set being isomorphic to an ideal of the first. To show that this is a well-ordering, let $\mathscr{B}$ be any nonempty class and $W \in \mathscr{B}$; because each $X \in \mathscr{B}$ that precedes W is isomorphic to an ideal of W, and because $I(W)$ is well ordered, there is a first element in $\mathscr{B}$.

5. Transfinite Induction and Construction

There are two induction principles associated with ordinals.

5.1 Theorem (Transfinite Induction) Let W be a well-ordered set, and let $Q \subset W$. If $[W(x) \subset Q] \Rightarrow [x \in Q]$ for each $x \in W$, then $Q = W$.

Proof: The first element, 0, of W is in Q, since $\varnothing = W(0) \subset Q$, and evidently there can be no first among the elements not in Q.

Remark 1: Theorem **5.1** is frequently stated in the following form:

(a) Let $\{P(x) \mid x \in W\}$ be a set of propositions. Assume both (i) $P(0)$ is true, and (ii) for each x, the hypothesis that every $P(\alpha)$, $\alpha \in W(x)$, is true implies $P(x)$ is also true. Then every $P(x)$ is true.

This is evidently **5.1**, with $Q = \{x \mid P(x) \text{ is true}\}$.

Remark 2: On ω, the induction principle is equivalent to:

(b) Let $\{P(i) \mid i \in \omega\}$ be a set of propositions. Assume both (i) $P(0)$ is true, and (ii) for each n, the hypothesis that $P(n-1)$ is true implies $P(n)$ is also true. Then every $P(n)$ is true.

The equivalence of (a) and (b), *on* ω, follows precisely because each element of ω has an immediate predecessor; and since, as we have seen, well-ordered sets do not always have this property, the analogue of (b) is not generally true. For example, on $\omega \cup \{q\}$, where q is adjoined as the last element, the form (b) does *not* establish that $P(q)$ is true.

The second induction principle is the important

5.2 Theorem (Transfinite Construction) Let W be a well-ordered set, and E an arbitrary class. Assume:

For each $x \in W$, there is given a rule R_x that associates with each $\varphi: W(x) \to E$, a unique $R_x(\varphi) \in E$.

Then there is one, and only one, $F: W \to E$ such that $F(x) = R_x[F \mid W(x)]$ for each $x \in W$.

Proof: There is at most one such F. For, if F, G were two distinct such maps, there would be a first x with $F(x) \neq G(x)$; then, since $F \mid W(x) = G \mid W(x)$, we would have $F(x) = G(x)$, contradicting the choice of x. We now use **3.4** to show that a map F, as required, exists. Let Σ be the set of those ideals S of W for which there is a φ_S: $S \to E$ satisfying the stated condition. By what we have already proved, φ_S is unique on each S, and therefore, on (the ideal) $S \cap S'$ we have $\varphi_S \mid S \cap S' = \varphi_{S'} \mid S \cap S'$; as in **4.2**, we conclude that any union of members of Σ belongs to Σ. Let $S = W(x) \in \Sigma$; then $W(x) \cup \{x\} \in \Sigma$, since we can extend φ_S by setting $\varphi(x) = R_x(\varphi_S \mid W(x))$. It now follows from **3.4** that $W \in \Sigma$, proving the theorem.

Remark 3: Writing 0 for the first element of W, note that $W(0) = \varnothing$; since there is only one map, $\varnothing \to E$, the hypothesis of **5.2** tacitly assumes that $\varphi(0)$ is uniquely defined. Thus, in greater detail, **5.2** can be stated as follows: Let W be well-ordered, E an arbitrary class, and $e \in E$ a given element. Assume that for each $x \neq 0$, there is given a rule R_x that assigns to each φ: $W(x) \to E$ an element $R_x(\varphi) \in E$. Then there is exactly one F: $W \to E$ such that $F(0) = e$ and $F(x) = R_x[F \mid W(x)]$ for each $x \neq 0$.

Remark 4: In ω, **5.2** can be given a simpler form because each element has an immediate predecessor: Let E be an arbitrary class and $e \in E$ a given element. Assume that for each n there is given a map R_n: $E \to E$. Then there exists one, and only one, F: $\omega \to E$ such that $F(0) = e$ and $F(n + 1) = R_{n+1}[F(n)]$ for each $n \in \omega$.

Remark 5: The reader should observe that **5.2** cannot be proved by simply defining F: $W \to E$ by the formula $F(x) = R_x[F \mid W(x)]$: such a definition would be circular in that it defines F in terms of itself. The whole purpose of **5.2** is to show that there exists a map F: $W \to E$ having this property.

6. Ordinal Numbers

In the class of all ordinals, define $W = X$ if W is isomorphic to X. This is evidently an equivalence relation, so it divides the class of all ordinals into mutually exclusive subclasses. We wish to attach to each ordinal an object, called its *ordinal number* (underscored in the subsequent text), so that two ordinals have the same ordinal number if and only if they are isomorphic. Following Frege, we could define the ordinal number of an ordinal to be the equivalence *class* of that ordinal. Though this definition is adequate for most mathematical purposes, it has the disadvantage that ordinal numbers are not *sets*: Without separate axiomatics for them, we could not, for example, legitimately consider any collection of ordinal numbers. In this section, we will define "ordinal number" within the framework of set theory. It will be shown that there exists a uniquely defined well-ordered class $\mathscr{Z}$ such that each well-ordered set is isomorphic to some initial interval of $\mathscr{Z}$. The desired objective is attained by calling the members of $\mathscr{Z}$ ordinal numbers and assigning to each ordinal $\overline{W}$ the $\underset{\sim}{\alpha} \in \mathscr{Z}$ for which $\overline{W} = \mathscr{Z}(\underset{\sim}{\alpha})$.

The basic idea is to use the sets postulated by the axiom of infinity (I, **8**, Axiom X): each $\underset{\sim}{\alpha} \in \mathscr{Z}$ will be a set whose elements are all the sets in $\mathscr{Z}$ that precede it; in other words, each $\underset{\sim}{\alpha} \in \mathscr{Z}$ will be simply the initial interval $\mathscr{Z}(\underset{\sim}{\alpha})$ in $\mathscr{Z}$. To illustrate the mechanics, we write down the first few members of $\mathscr{Z}$:

$$\varnothing; \{\varnothing\}; \{\varnothing, \{\varnothing\}\}; \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}.$$

The ordinal number of $\{1, 2, 3\}$, in its natural order, is then $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$.

6.1 Definition An ordinal number is a set $\underset{\sim}{\alpha}$ with the properties

(1). $(x \in \underset{\sim}{\alpha}) \wedge (y \in \underset{\sim}{\alpha}) \Rightarrow (x \in y) \vee (y \in x) \vee (y = x)$.
(2). $(x \in y) \wedge (y \in \underset{\sim}{\alpha}) \Rightarrow (x \in \underset{\sim}{\alpha})$.

We say that "x precedes y" in $\underset{\sim}{\alpha}$ if "$x \in y$"; observe that $\in$ is not an ordering, since by I, **8.11**, $x \in x$ is not true. The $\in$-relation in ordinal numbers has the properties

6.2 (a). In each nonempty set $A \subset \underset{\sim}{\alpha}$, there is a unique $a \in A$ (called the first element of A) such that $(a \in x) \vee (a = x)$ for each $x \in A$.
(b). The first member in $\underset{\sim}{\alpha}$ is $\varnothing$.
(c). If $z \in \underset{\sim}{\alpha}$, then z is also an ordinal number.

Proof: (a). By the axiom of foundation (I, **8**, Axiom IX), there is an $a \in A$ with $a \cap A = \varnothing$ [that is, $(x \in A) \Rightarrow \neg(x \in a)$]; in view of **6.1**(1), the element a has the required property. If there were another element $b \in A$ with this property, we would have $a \in b$ and $b \in a$, which contradicts I, **8.11**(2).

(b). If a is the first member of $\underset{\sim}{\alpha}$, there can be no $x \in a$ because of **6.1**(2).

(c). Let $x, y \in z$; since $(x, y \in z) \wedge (z \in \underset{\sim}{\alpha}) \Rightarrow (x, y \in \underset{\sim}{\alpha})$, we find that **6.1**(1) is valid for the members of z. To verify **6.1**(2), assume that $(x \in y) \wedge (y \in z)$; by what we have just proved, we must have one of $(x \in z)$, $(z \in x)$, $(z = x)$, and we need show only that the last two possibilities cannot be true.

(i). If $z \in x$, then the subset $A = \{x, y, z\} \subset \underset{\sim}{\alpha}$ does not have a first element, as required by (a).

(ii). If $z = x$, we would have $(x \in y) \wedge (y \in x)$ which violates the axiom of foundation [*cf.* I, **8.11**(2).]

According to (c), ordinal numbers can be treated as sets or as elements of sets. We next establish the essential uniqueness of ordinal numbers.

6.3 (a). If $\underline{\alpha}, \underline{\beta}$ are ordinal numbers, and $\underline{\alpha} \neq \underline{\beta}$, then $\underline{\alpha} \subset \underline{\beta}$ if and only if $\underline{\alpha} \in \underline{\beta}$ (that is, an ordinal number consists of all those ordinal numbers that are proper subsets of it).

(b). If $\underline{\alpha}$ and $\underline{\beta}$ are any two ordinal numbers, then either $\underline{\alpha} \subset \underline{\beta}$ or $\underline{\beta} \subset \underline{\alpha}$.

Proof: (a). If $\underline{\alpha} \in \underline{\beta}$, then by **6.1**(2), $x \in \underline{\alpha} \Rightarrow x \in \underline{\beta}$, so that $\underline{\alpha} \subset \underline{\beta}$. Conversely, assume $\underline{\alpha} \subset \underline{\beta}$ and let $x_0 \in \underline{\beta}$ be the first element of $\underline{\beta} - \underline{\alpha}$. The definition of x_0 shows that $y \in x_0 \Rightarrow y \in \underline{\alpha}$, so that $x_0 \subset \underline{\alpha}$. Now assume that $y \in \underline{\alpha}$; since the possibilities $x_0 \in y$, $x_0 = y$ are excluded because each implies $x_0 \in \underline{\alpha}$, we find that $y \in \underline{\alpha} \Rightarrow y \in x_0$, and so $\underline{\alpha} \subset x_0$. Thus $\underline{\alpha} = x_0 \in \underline{\beta}$.

(b). It is immediate that $\underline{\alpha} \cap \underline{\beta}$ (that is, $\{x \mid x \in \underline{\alpha} \wedge x \in \underline{\beta}\}$) is an ordinal number; we need show only that it is one of $\underline{\alpha}, \underline{\beta}$. If it were neither, then $\underline{\alpha} \cap \underline{\beta}$ would be a proper subset of both $\underline{\alpha}$ and $\underline{\beta}$, so, by (a), we would have $(\underline{\alpha} \cap \underline{\beta} \in \underline{\alpha}) \wedge (\underline{\alpha} \cap \underline{\beta} \in \underline{\beta})$, which would then imply $\underline{\alpha} \cap \underline{\beta} \in \underline{\alpha} \cap \underline{\beta}$, in contradiction to I, **8.11**.

We now have the comprehensive

6.4 Theorem Let $\mathscr{X}$ be the class of all ordinal numbers. Define $\underline{\alpha} \leqslant \underline{\beta}$ if and only if $\underline{\alpha} \subset \underline{\beta}$. Then:

(1). $\mathscr{X}$ is well-ordered by $\leqslant$.
(2). $\mathscr{X}$ is not a set.
(3). For each $\underline{\alpha} \in \mathscr{X}$, the initial interval $\mathscr{X}(\underline{\alpha}) = \underline{\alpha}$ and, in particular, is a set.
(4). If E is any *set* of ordinal numbers, there is an ordinal number greater than all the ordinal numbers in E (and, in fact, a smallest such ordinal number).
(5). Every nonincreasing sequence of ordinal numbers is necessarily finite (that is, if $\underline{\alpha}_0 \geqslant \underline{\alpha}_1 \geqslant \cdots$, then there is an n such that $\underline{\alpha}_i = \underline{\alpha}_n$ for all $i \geqslant n$).
(6). Each well-ordered set W is isomorphic to a suitable $\mathscr{X}(\underline{\alpha})$. $\underline{\alpha}$ is called the ordinal number of W and is denoted by "ord W."

Proof: (1). $\leqslant$ is clearly a partial ordering. To see that it is a well-ordering, let $E \subset \mathscr{X}$ be a nonempty set; we are to show that E has a first element. Choose $\underline{\alpha}_0 \in E$ and define

$$A = \underline{\alpha}_0 \cap E = \{x \mid (x \in \underline{\alpha}_0) \wedge (x \in E)\}.$$

If $A = \varnothing$, then by **6.3**(a) and (b), we find that

$$x \in E \Rightarrow \neg(x \subset \underline{\alpha}_0) \Rightarrow \underline{\alpha}_0 \subset x$$

so that $\underline{\alpha}_0$ is first in E.

If $A \neq \varnothing$, then by **6.2**(a), there is an $a \in A$ with $(a \in x) \vee (a = x)$ for each $x \in A$. We now have $a \in E$ and $a \subset x$ for each $x \in \underline{\alpha}_0 \cap E$; but since $a \subset \underline{\alpha}_0$ and $\underline{\alpha}_0 \subset y$ for each $y \in E - (\underline{\alpha}_0 \cap E)$, we also have $a \subset y$ for each $y \in E - (\underline{\alpha}_0 \cap E)$, so that a is first in E.

(2). If $\mathscr{Z}$ were a set, we verify immediately that $\mathscr{Z}$ would then be an ordinal number, and this would imply that $\mathscr{Z} \in \mathscr{Z}$, which is impossible (I, **8.11**).

(3). Given $\underline{\alpha} \in \mathscr{Z}$, $\mathscr{Z}(\underline{\alpha}) = \{\underline{\beta} \mid \underline{\beta} \in \mathscr{Z} \wedge \underline{\beta} \in \underline{\alpha}\}$; by **6.2**(c), the condition $\underline{\beta} \in \mathscr{Z}$ is redundant, so $\mathscr{Z}(\underline{\alpha}) = \{\underline{\beta} \mid \underline{\beta} \in \underline{\alpha}\} = \underline{\alpha}$.

(4). Let E be any *set* of ordinal numbers. Then $\underline{\beta} = \bigcup \{\underline{\alpha} \mid \underline{\alpha} \in E\}$ is a set, easily verified to be an ordinal number. Since we have $\underline{\alpha} \subset \underline{\beta}$ for each $\underline{\alpha} \in E$, it follows that $\underline{\alpha} \leqslant \underline{\beta}$ in $\mathscr{Z}$. Noting that $\underline{\beta} \cup \{\underline{\beta}\}$ is also an ordinal number and that $\underline{\beta}$ is a proper subset of $\underline{\beta} \cup \{\underline{\beta}\}$, we find $\underline{\beta} \cup \{\underline{\beta}\}$ larger than each $\underline{\alpha} \in E$. The second part follows from (1).

(5). Since $\mathscr{Z}$ is well-ordered, the set $\{\underline{\alpha}_i \mid i \in N\}$ has a first element, which is some one of its members, say, $\underline{\alpha}_n$. Then, for all $i \geqslant n$, we have $\underline{\alpha}_i = \underline{\alpha}_n$.

(6). We will use transfinite construction. For each $x \in W$ and each $\varphi\colon W(x) \to \mathscr{Z}$ define $R_x(\varphi)$ to be the smallest ordinal number greater than all the ordinal numbers in $\varphi(W(x))$; this definition is legitimate, since by the axiom I, **8**, VI of replacement, $\varphi(W(x)) \subset \mathscr{Z}$ is a set and (4) applies. By **5.2**, there exists a map $F\colon W \to \mathscr{Z}$ such that $F(x) = R_x[F \mid W(x)]$ for each $x \in W$, and it is easy to see that F is a monomorphism. Since $F(W) \subset \mathscr{Z}$ is a set, we have by (4) that $F(W) \subset \mathscr{Z}(\underline{\beta})$ for some $\underline{\beta}$, and from **4.3** that W is isomorphic to some $\mathscr{Z}(\underline{\alpha})$. [The reader may show that, in fact, $F(W) = \mathscr{Z}(\underline{\alpha})$.]

According to **6.2**(b), the smallest ordinal number is $\varnothing$, the null set; it is denoted by "$\underline{0}$." The next one, as indicated before, is $\{\varnothing\}$, whose only member is the null set, and is denoted by "$\underline{1}$." The ordinal number $\{\underline{0}, \underline{1}\}$ is denoted by $\underline{2}$, and in general, the ordinal number with the n elements $\{\underline{0}, \underline{1}, \cdots, \underline{n-1}\}$ is denoted by $\underline{n}$. If $\underline{\alpha}$ is any ordinal number, $\underline{\alpha} \cup \{\underline{\alpha}\}$ is also an ordinal number, written $\underline{\alpha} + \underline{1}$. Note, however, that the axiom of infinity (I, **8**, X) states the existence of ordinal numbers not of the form $\underline{\alpha} \cup \{\underline{\alpha}\}$; these are called *limit* ordinal numbers and have no immediate predecessor in $\mathscr{Z}$.

A well-ordered set W is called *finite* if ord $W = \underline{n}$ for some $\underline{n}$. The ordinal number of the well-ordered set $\{1, 2, \cdots\}$ in its natural order is denoted by $\underline{\omega}$. Taking the notation for intervals in E^1, we denote $\mathscr{Z}(\underline{\alpha})$ by $[\underline{0}, \underline{\alpha}[$; obviously, $\text{ord}[\underline{0}, \underline{\alpha}[= \underline{\alpha}$.

6.5 Remark We have used the ordinal numbers as "counters"—attention has been directed to the initial intervals of $\mathscr{X}$. However, if in any ordinal number $\underline{\alpha}$ we define $x \prec y$ by $(x \in y) \vee (x = y)$, **6.2**(c) and (a) show that $\prec$ is a well-ordering of $\underline{\alpha}$. Denoting $\underline{\alpha}$ together with this well-ordering by $\underline{\alpha}(w)$, it follows at once from **6.3**(a) and **6.4**(3) that the identity map of $\underline{\alpha}(w)$ onto $\mathscr{X}(\underline{\alpha})$ is in fact an isomorphism of well-ordered sets, so that $\text{ord}\ \underline{\alpha}(w) = \underline{\alpha}$. Thus we can (and shall) equally well regard $\mathscr{X}$ as a well-ordered class of well-ordered sets, containing exactly *one* representative from each equivalence class of isomorphic ordinals.

In the future, we will denote ordinal numbers by a Greek letter without underscoring whenever there is no danger of confusion.

7. Cardinals

The ordinals are associated with counting: to count, one counts some elements first and thus tacitly induces a well-ordering. The concept of cardinal is related simply to size: we wish to determine if one of two given sets has more members than the other. Counting is not needed for this purpose: we need only pair off each member of one set with a member of the other and see if any elements are left over. "Same size" is thus formalized in

7.1 Definition Two sets, X and Y, are equipotent (or have the same cardinal) if a bijective map of X onto Y exists. We denote "X equipotent to Y" by "$\text{card}\ X = \text{card}\ Y$."

It is evident that equipotence is an equivalence relation in the class of all sets, so it decomposes this class into mutually exclusive subclasses, called *equipotence classes.*

Ex. 1 Let $X = \{2n \mid n \in N\} \subset N$. Then $\text{card}\ X = \text{card}\ N$, since $n \to 2n$ is a bijection of N onto X. Note that a set may be equipotent with a proper subset.

Ex. 2 Any open interval $]a, b[\subset E^1$ is equipotent to $\mathscr{J} =]-1, +1[$, since $x \to \frac{b-a}{2} x + \frac{b+a}{2}$ is a bijection $\mathscr{J} \to]a, b[$. Furthermore, $\mathscr{J}$ is equipotent with E^1, as $x \to \frac{x}{1+|x|}$ shows. Thus, by transitivity, $\text{card}\ ]a, b[= \text{card}\ E^1$: each open interval in E^1 has "just as many points" as E^1 itself.

Ex. 3 Distinct ordinals may be equipotent. The ordinals $[0, \omega[$ and $[0, \omega]$ in $\mathscr{X}$ are not isomorphic [*cf.* **4.3**(b)], but $\varphi(\omega) = 0$, $\varphi(n) = n + 1$ is a bijection of $[0, \omega]$ onto $[0, \omega[$. This indicates an important distinction between finite and infinite ordinals: For infinite sets, the ordinal number depends *both* on the size of the set *and* the manner in which the set is counted. Thus the decomposition of the class of all ordinals according to *equipotence* does not coincide with that according to *isomorphism*: each isomorphism class lies in one equipotence class, ' ut an equipotence class generally contains many isomorphism classes.

Let X be any set. The set of all maps $X \to \{0, 1\}$ ($\{0, 1\}$ is the set consisting of the two integers 0, 1) is denoted by 2^X. For any set $A \subset X$, the map $c_A \in 2^X$ defined by

$$c_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}$$

is called the *characteristic function of A*. The following simple result has important consequences.

7.2 Card 2^X = card $\mathscr{P}(X)$.

Proof: Define $f: 2^X \to \mathscr{P}(X)$ by $c \to c^{-1}(1)$; f is clearly surjective, since $f(c_A) = A$. But f is also injective: $c \neq d \Rightarrow \exists\, x\colon\ c(x) \neq d(x)$ $\Rightarrow x$ belongs to exactly *one* of $c^{-1}(1)$, $d^{-1}(1) \Rightarrow f(c) \neq f(d)$.

To compare the size of sets, we make the

7.3 Definition For two sets, X, Y, we write card $X \leqslant$ card Y if an injection $X \to Y$ exists.

Note that we use "$\leqslant$" rather than "smaller": The existence of an injection $X \to Y$ does not exclude the possibility that there is also a bijection $X \to Y$, as Ex. 1 and 2 show.

7.4 (a). Card $A \leqslant$ card X for any subset $A \subset X$.
(b). If there exists a surjection $f: X \to Y$, then card $Y \leqslant$ card X.

Proof: (a). This proof is trivial. (b). Let c be a choice function for the family $\mathscr{P}(X)$; since $\{f^{-1}(y) \mid y \in Y\}$ is a partition of X, it follows that $y \to c[f^{-1}(y)]$ is an injection $Y \to X$.

We wish to attach a symbol to each set X, called its *cardinal number*, in such a way that two sets have the same cardinal number if and only if they are equipotent. For this purpose, note that because [**2.1**(3)] every set can be well-ordered, it follows from **6.4**(6) and **6.5** that each equipotence class contains at least one ordinal number and, since $\mathscr{Z}$ is well-ordered, a smallest such ordinal number. This uniquely determined representative is called the *initial ordinal number* of the equipotence class, and is denoted generically by the Hebrew letter $\aleph$ (aleph).

7.5 Definition The cardinal number of X, denoted by $\aleph(X)$, is the initial ordinal number of its equipotence class.

Thus a cardinal number is an ordinal number that is not equipotent to any smaller ordinal number. The notation gives (a) the smallest ordinal number to which X is equipotent (that is, the "most economical" way to count X) and (b) a unique standard set equipotent to X. It is clear that $\aleph(\aleph(X)) = \aleph(X)$ and also that $\aleph(W) \leqslant \text{ord } W$, for any ordinal W (in Ex. 3, we have $\aleph[\underline{0}, \underline{\omega}] < \text{ord } [\underline{0}, \underline{\omega}]$).

Certain sets occur frequently and have special symbols: $\aleph(\varnothing) = \underline{0}$, and $\aleph(\{1, \cdots, n\}) = \underline{n}$; $\aleph(N) = \aleph_0$ (rather than ω, to indicate that size but not order is the property being considered). A set X is called *finite* if $\aleph(X) = \underline{n}$ for some n; otherwise, it is called *infinite*, and its cardinal number is a *transfinite* cardinal number. Any set X with $\aleph(X) \leqslant \aleph_0$ is called *countable*; otherwise, it is called uncountable.

The statement "card $\aleph(X) \leqslant$ card $\aleph(Y)$" means that there is an injection of the *set* $\aleph(X)$ into $\aleph(Y)$; on the other hand "$\aleph(X) \leqslant \aleph(Y)$" means that the ordinal $\aleph(X)$ is *isomorphic* to an ideal of $\aleph(Y)$. The following fundamental theorem shows that the comparison of the sizes of X, Y is identical with the comparison of the ordinal numbers $\aleph(X)$, $\aleph(Y)$. Precisely,

7.6 Theorem Card $X \leqslant$ card Y if and only if $\aleph(X) \leqslant \aleph(Y)$. In particular, card $X =$ card Y if and only if $\aleph(X) = \aleph(Y)$.

Proof: Because of transitivity and equipotence of X (resp. Y) to $\aleph(X)$ [resp. $\aleph(Y)$], it is evident that $\aleph(X) \leqslant \aleph(Y) \Rightarrow \text{card } X \leqslant \text{card } Y$. Conversely, assume that card $X \leqslant$ card Y; then there exists an injection $X \to \aleph(Y)$ so we can regard X as a subset of $\aleph(Y)$ and then conclude, from **4.3**(a), that ord $X \leqslant \aleph(Y)$. Since $\aleph(X) \leqslant \text{ord } X$ is always true, this proves the theorem.

One of the most important consequences of **7.6** is

7.7 Corollary (Bernstein-Schröder) If there exists an injection $X \to Y$ and also an injection $Y \to X$, then there exists a bijection $X \to Y$. In particular, card $X <$ card Y (that is, card $X \leqslant$ card Y and card $X \neq$ card Y) if and only if (a) there is an injection $X \to Y$ and (b) there exists *no* injection $Y \to X$.

Proof: According to **4.2**, the ordinal inequalities $\aleph(X) \leqslant \aleph(Y)$ and $\aleph(Y) \leqslant \aleph(X)$ imply that $\aleph(X)$ is isomorphic (and so, in particular, equipotent) to $\aleph(Y)$.

Ex. 4 The cardinal numbers are characterized as those ordinal numbers $\underline{\beta}$ such that card $\underline{\alpha} <$ card $\underline{\beta}$ for all $\underline{\alpha} < \underline{\beta}$.

Ex. 5 Card $(N \times N) =$ card N; more generally, for *any* finite number of factors, card $(N \times \cdots \times N) = \text{card } N$: the set of all ordered k-uples of non-negative integers has cardinal number $\aleph_0$; that is, it is countable. For, $n \to (n, 0, \cdots 0)$ is an injection $N \to N \times \cdots \times N$, and if $2, 3, \cdots p_k$ are the first k

primes, the fundamental theorem of arithmetic shows that $(n_1, \cdots n_k) \to 2^{n_1} \cdots p_k^{n_k}$ is an injection $N \times \cdots \times N \to N$. It is immediate that the set of all ordered k-uples of nonnegative rationals is countable.

Ex. 6 For any nondegenerate interval (open, closed, half-open) $K \subset E^1$, card K = card E^1. For open $\mathcal{J}$, Ex. 2 shows $\aleph(\mathcal{J}) = \aleph(E^1)$. For arbitrary K, find open $\mathcal{J}, \mathcal{J}'$ with $\mathcal{J} \subset K \subset \mathcal{J}'$; then by **7.4**(a) and **7.6**,

$$\aleph(\mathcal{J}) \leqslant \aleph(K) \leqslant \aleph(\mathcal{J}') = \aleph(\mathcal{J}).$$

Ex. 7 Card $\mathscr{P}(N)$ = card 2^N = card E^1. We first observe that each real number in [0, 1] has a dyadic expansion $\sum_1^\infty 2^{-n}a_n$, with each $a_n = 0$ or 1. The expansion is not necessarily unique: expansions ending with all 1's can be converted to expansions ending in all 0's; the convention that if any number has two expansions, we always use the expansion with infinitely many 1's, gives uniqueness. Now let $A \subset 2^N$ consist of all characteristic functions taking the value 1 infinitely often. Define $\varphi\colon 2^N \to E^1$ by

$$\begin{aligned} \varphi(c) &= \sum_1^\infty \frac{c(n)}{2^n} \qquad \text{if } c \in A, \\ &= 2 + \sum_1^\infty \frac{c(n)}{2^n} \qquad \text{if } c \notin A. \end{aligned}$$

By our remarks, φ is injective. Since $]0, 1] \subset \varphi(2^N) \subset E^1$, the assertion follows as in Ex. 6.

We now establish the comprehensive

7.8 Theorem Let $\mathscr{K}$ be the class of all cardinal numbers.

(1). $\mathscr{K}$ is well-ordered.
(2). For any given cardinal number, there exists a larger one. In fact, for each set X, card $\mathscr{P}(X)$ > card X.
(3). $\underline{0}$ is the smallest cardinal number, and $\aleph_0$ is the smallest transfinite cardinal number; that is, every infinite set contains a countably infinite subset.
(4). The class $\mathscr{K}$ is not a set.

Proof: (1) is trivial, since $\mathscr{K} \subset \mathscr{Z}$.

(2). We replace $\mathscr{P}(X)$ by 2^X. Then:

(a). There is an injection $X \to 2^X$: define $x \to c_{\{x\}}$, the characteristic function of $\{x\}$.

(b). There is no injection $2^X \to X$. We argue by contradiction. If there were such an injection, then (a) and **7.7** would yield a bijection $\varphi\colon X \to 2^X$. Denote the value of φ at x by $c^{(x)}$, some characteristic function, and define $c \in 2^X$ by $c(x) = 1 - c^{(x)}(x)$. This differs from each $c^{(x)}$ and is therefore not in $\varphi(X)$, contradicting that φ is bijective.

(3). The first part is trivial. For the second, we need show only that for any infinite X, there is an injection $N \rightarrow X$. We proceed by inductive construction on the well-ordered N. Let c be a choice function for the family $\mathscr{P}(X)$; define $\varphi(1) = c(X)$, and for each $\varphi: N(n) \rightarrow X$, let $R_n(\varphi) = c(X - \varphi[N(n)])$. $X - \varphi[N(n)]$ cannot be empty for any n, since then $\varphi: N(n) \rightarrow X$ is surjective and by **7.4**(b), $\aleph(X) \leqslant \aleph(N(n)) = \underline{n}$. Applying **5.2** yields an $F: N \rightarrow X$ with $F(n) = R_n[F \mid N(n)]$. F is clearly injective, so $\aleph_0 \leqslant \aleph(X)$, and (3) is proved.

(4). Assume that $\mathscr{K}$ is a set. Then $X = \cup \{\aleph \mid \aleph \in \mathscr{K}\}$ is also a set, and so 2^X is a set; therefore $\aleph(2^X) \subset X$, and, with **7.4**(a), this gives $\aleph(\aleph(2^X)) \leqslant \aleph(X)$, that is, $\aleph(2^X) \leqslant \aleph(X)$. This is impossible in view of (2).

Ex. 8 Ex. 7 and **7.8**(2) yield the following important result: Card $N <$ card E^1 (there are "more" real numbers than there are rational numbers). The cardinal number of E^1 is denoted by $\mathfrak{c}$ and is called the *cardinal of the continuum.*

7.9 Remark In view of **7.8**(2), the question arises: For an infinite set X, is there any cardinal between $\aleph(X)$ and $\aleph(\mathscr{P}(X))$? The hypothesis that there is no such cardinal is called the *generalized continuum hypothesis* (the "continuum hypothesis" is that there does not exist any cardinal between $\aleph_0$ and $\mathfrak{c}$). Under the assumption that the Axioms I–XI of I, **8** are consistent, K. Gödel proved that the generalized continuum hypothesis is consistent with these axioms, and P. J. Cohen proved that the *negation* of the continuum hypothesis is also consistent with the given axioms. Thus, both the continuum and generalized continuum hypotheses are *independent* of the other axioms of set theory: neither hypothesis can be proved or disproved on the basis of I–XI alone. However, if the generalized continuum hypothesis be taken as an additional axiom for set theory, W. Sierpinski has shown that the Axiom XI of choice becomes redundant: it can be derived from the generalized continuum hypothesis and axioms I–X.

7.10 Remark Using **7.8**(1), the aleph notation for cardinal numbers is supplemented with ordinals as follows: For any cardinal number $\bar{\aleph}$, the set $\{\aleph \mid \aleph_0 \leqslant \aleph \leqslant \bar{\aleph}\}$ is well-ordered, so it is isomorphic to an ideal $\mathscr{L}(\alpha)$ in $\mathscr{L}$. It is then conventional to write $\bar{\aleph}$ as $\aleph_\alpha$, when used to indicate size, and as ω_α when it is to be considered as an ordinal number. If α has no immediate predecessor, $\aleph_\alpha$ is called an *inaccessible* cardinal number. In this notation, the continuum hypothesis is that $\aleph_1 = \mathfrak{c}$; the generalized continuum hypothesis is that for each ordinal number α, $\aleph(\mathscr{P}(\aleph_\alpha)) = \aleph_{\alpha+1}$.

8. Cardinal Arithmetic

In this section, operations on cardinal numbers that extend those of ordinary arithmetic will be defined, and the fundamental theorem of this arithmetic will be proved. There is a more interesting ordinal number arithmetic, differing from the cardinal arithmetic, but we shall not discuss it here.

8.1 Definition Let $\{\aleph_\mu \mid \mu \in \mathscr{M}\}$ be any set of cardinal numbers. Their sum $\sum_\mu \aleph_\mu$ is determined by selecting a pairwise disjoint family $\{A_\mu \mid \mu \in \mathscr{M}\}$ of sets with $\aleph(A_\mu) = \aleph_\mu$ for each μ, and setting $\sum_\mu \aleph_\mu = \aleph(\bigcup_\mu A_\mu)$. Their product, $\prod_\mu \aleph_\mu$ is $\aleph(\prod_\mu A_\mu)$.

Note first that for any family $\{A_\mu \mid \mu \in \mathscr{M}\}$ of sets, we can always find a pairwise disjoint family $\{A'_\mu \mid \mu \in \mathscr{M}\}$ with $\aleph(A_\mu) = \aleph(A'_\mu)$ for each μ, by setting $A'_\mu = \{\mu\} \times A_\mu$; the operations in **8.1** are consequently always defined. Furthermore, the definitions are unambiguous: If $\{B_\mu \mid \mu \in \mathscr{M}\}$ is any other family of pairwise disjoint sets with $\aleph(A_\mu) = \aleph(B_\mu)$ for each μ, to verify that $\aleph(\bigcup_\mu A_\mu) = \aleph(\bigcup_\mu B_\mu)$ and $\aleph(\prod_\mu A_\mu) = \aleph(\prod_\mu B_\mu)$ is trivial. In fact, observe that $\aleph(\prod_\mu A_\mu)$ is the same whether or not the family $\{A_\mu \mid \mu \in \mathscr{M}\}$ is pairwise disjoint; though this is not true for sums, we have

8.2 If $\{A_\mu \mid \mu \in \mathscr{M}\}$ is any family of sets, pairwise disjoint or not, then $\aleph(\bigcup_\mu A_\mu) \leqslant \sum_\mu \aleph(A_\mu)$.

Proof: Using the pairwise disjoint family $\{\mu\} \times A_\mu = A'_\mu$, there is clearly a surjection $\bigcup_\mu A'_\mu \to \bigcup_\mu A_\mu$ given by $(\mu, a_\mu) \to a_\mu$; from **7.4**(b) we find $\aleph(\bigcup_\mu A_\mu) \leqslant \aleph(\bigcup_\mu A'_\mu) = \sum_\mu \aleph(A_\mu)$.

The Boolean operations and I, **4.2**, immediately give that addition (and multiplication) is unrestrictedly commutative and associative, and that $\aleph \cdot \sum_\mu \aleph_\mu = \sum (\aleph \cdot \aleph_\mu)$. It is simple to verify monotoneity (that is if $\aleph_\mu \leqslant \aleph'_\mu$ for each μ, then $\sum_\mu \aleph_\mu \leqslant \sum_\mu \aleph'_\mu$ and $\prod_\mu \aleph_\mu \leqslant \prod_\mu \aleph'_\mu$) and that $\underline{0} \cdot \aleph = \underline{0}$, $\underline{0} + \aleph = \aleph = \underline{1} \cdot \aleph$. For finite sets and sums, these operations reduce to the usual ones, in that $\underline{n} + \underline{m} = \underline{n+m}$, $\underline{n} \cdot \underline{m} = \underline{n \cdot m}$.

Starting from $\underline{2} \cdot \aleph = (\underline{1} + \underline{1}) \cdot \aleph = \underline{1} \cdot \aleph + \underline{1} \cdot \aleph$, we get by induction that the sum of n terms $\aleph$ is $\underline{n} \cdot \aleph$. This extends even to infinitely many terms $\aleph$,

8.3 (a). If the cardinal number of the indexing set is $\aleph'$, and if each $\aleph_\mu = \aleph$, then $\sum_\mu \aleph_\mu = \aleph' \cdot \aleph$.

(b). The union of a countable collection of countable sets is countable.

Proof: (a). Letting $\{A_\mu \mid \mu \in \mathscr{M}\}$ be a pairwise disjoint family of sets with $\aleph(A_\mu) = \aleph$ for each μ, and choosing one of them (say, A_0), we have for each $\mu \in \mathscr{M}$ a bijection $f_\mu: A_0 \to A_\mu$. Since the map $\mathscr{M} \times A_0 \to \bigcup_\mu A_\mu$ given by $(\mu, a) \to f_\mu(a)$ is easily seen to be bijective, this proves (a).

(b). Since $\aleph(A_\iota) \leqslant \aleph_0$ for each $i \in N$, use **8.2**, monotoneity, and (a) to find

$$\aleph(\bigcup_\iota A_\iota) \leqslant \sum_1^\infty \aleph(A_\iota) \leqslant \sum_1^\infty \aleph_0 = \aleph_0 \cdot \aleph_0.$$

According to **7**, Ex. 5, we have card $(N \times N) =$ card N, that is, $\aleph_0 \cdot \aleph_0 = \aleph_0$, which establishes (b).

Ex. 1 With **8.3**(b) we can extend **7**, Ex. 5; the cardinal number of all rationals in E^1 (positive, negative, and zero) is $\aleph_0$. More generally, the set of all points in E^n with rational coordinates is countable.

Ex. 2 Transcendental numbers exist. Indeed, for each n, the cardinal number of the set of all polynomials of degree n with integer coefficients is $\aleph_0$, according to Ex. 1, so by **8.3**(b) the set of their roots is also countable. From **8.3**(b), again, it then follows that the set of *all* algebraic numbers is countable. But $\aleph(E^1) = \mathfrak{c} > \aleph_0$.

8.4 Definition Let $\bar{\aleph}$, $\aleph$ be two cardinal numbers. The exponentiation $\bar{\aleph}^{\aleph}$ is the cardinal number of the set of all maps $\aleph \to \bar{\aleph}$.

This notion of exponentiation reduces to the usual concept for finite cardinals: $\underline{n}^{\underline{m}} = \underline{n^m}$. Note that $2^\aleph$ is simply the cardinal number of the set of all characteristic functions on $\aleph$ so, by **7.8**(2), it is always true that $\aleph < 2^\aleph$ and, by **7**, Ex. 7, that $2^{\aleph_0} = \mathfrak{c}$.

If X, $\bar{X}$ are any two sets, let $\bar{X}^X$ denote the set of all maps $X \to \bar{X}$. It is trivial to verify that, if $\aleph(X) = \aleph$ and $\aleph(\bar{X}) = \bar{\aleph}$, then $\aleph(\bar{X}^X) = \bar{\aleph}^\aleph$.

The usual rules for exponentiation apply: $(\bar{\aleph})^{\aleph + \aleph'} = \bar{\aleph}^\aleph \cdot \bar{\aleph}^{\aleph'}$, $(\aleph \cdot \bar{\aleph})^{\aleph'} = \aleph^{\aleph'} \cdot \bar{\aleph}^{\aleph'}$, and $\bar{\aleph}^{\aleph \cdot \aleph'} = (\bar{\aleph}^\aleph)^{\aleph'}$. We indicate a proof of the last equality; the others are proved similarly. Choosing X, Y, Z, such that $\aleph(X) = \bar{\aleph}$, $\aleph(Y) = \aleph$, and $\aleph(Z) = \aleph'$, associate with each $f \in X^{Y \times Z}$ the map $\varphi_f: Z \to X^Y$ given by $z_0 \to f(y, z_0)$; then $f \to \varphi_f$ is easily verified to be bijective.

Ex. 3 $\aleph \cdot \aleph = \aleph^{\underline{2}}$, since $\aleph^{\underline{1}} = \aleph$ and $\aleph^{\underline{1}+\underline{1}} = \aleph^{\underline{2}}$; also $\aleph^{\underline{0}} = \underline{1}$ from I, **6**, Ex. 3; thus, with finite cardinals as exponents, exponentiation has the usual interpretation. Also from I, **6**, Ex. 3, $\underline{0}^\aleph = \underline{0}$.

Ex. 4 As with sums, Ex. 3 extends to infinitely many factors: If $\aleph_\mu = \aleph$ for each $\mu \in \mathscr{M}$, and if $\aleph(\mathscr{M}) = \aleph'$, then $\prod_\mu \aleph_\mu = \aleph^{\aleph'}$. For $\prod_\mu \aleph_\mu$ is the cardinal number of $\prod_\mu X_\mu$, where each factor X_μ is the same set X having $\aleph(X) = \aleph$ and, by I, **9**, Ex. 2, $\prod_\mu X_\mu$ is the set of all maps $\mathscr{M} \to X$. In particular, the Cantor set has cardinal number $2^{\aleph_0} = \mathfrak{c}$.

We have seen that $\aleph_0 \cdot \aleph_0 = \aleph_0$; the fundamental theorem asserts that this is true for all transfinite cardinal numbers:

8.5 Theorem $\aleph \cdot \aleph = \aleph$ for each $\aleph \geqslant \aleph_0$.

Lemma If $\aleph \cdot \aleph = \aleph$ for any $\aleph$, then also $\aleph = \underline{2} \cdot \aleph = \underline{3} \cdot \aleph = \aleph \cdot \aleph$.

Proof of Lemma: We have

$$\aleph = \aleph + \underline{0} \leqslant \aleph + \aleph = \underline{2} \cdot \aleph \leqslant \underline{3} \cdot \aleph \leqslant \aleph \cdot \aleph = \aleph.$$

Proof of Theorem: It suffices to prove that for some set M with $\aleph(M) = \aleph$, there is a bijection $M \to M \times M$. We use Zorn's lemma, **2.1**(2).

Let X be any set with $\aleph(X) = \aleph$, and let $\mathscr{A}$ be the set of all couples (A, φ_A), where $A \subset X$ and $\varphi_A: A \to A \times A$ is a bijection. Then $\mathscr{A} \neq \varnothing$: since X is infinite, **7.8**(3) shows that X contains a subset A with $\aleph(A) = \aleph_0$, and we know that $\aleph_0 \cdot \aleph_0 = \aleph_0$.

Partially order $\mathscr{A}$ by $(A, \varphi_A) \prec (B, \varphi_B)$ if both $A \subset B$ and $\varphi_B \mid A = \varphi_A$. Each chain $\{(A_\beta, \varphi_\beta) \mid \beta \in \mathscr{B}\}$ has an upper bound $(\bigcup_\beta A_\beta, \varphi)$ because of I, **6.7**, and consequently there is a maximal (M, φ_M); since $\varphi_M: M \to M \times M$ is bijective, we have $\aleph(M) \cdot \aleph(M) = \aleph(M)$, and the argument reduces to showing that $\aleph(M) = \aleph$.

Since $M \subset X$, we have $\aleph(M) \leqslant \aleph(X) = \aleph$; we shall prove that $\aleph(M) < \aleph(X)$ is impossible because it would contradict that (M, φ_M) is maximal in $\mathscr{A}$. Thus, assume that $\aleph(M) < \aleph(X)$. Then we must have

$$(1). \qquad \aleph(M) < \aleph(X - M).$$

For, if $\aleph(X - M) \leqslant \aleph(M)$, then because $X = M \cup (X - M)$ is a disjoint union, using the lemma would yield

$$\aleph(X) = \aleph(M) + \aleph(X - M) \leqslant \aleph(M) + \aleph(M) = \underline{2}\aleph(M) = \aleph(M),$$

which contradicts the assumed $\aleph(M) < \aleph(X)$.

Because of (1), there is a bijective map of M onto a subset $Y \subset X - M$, and we have $\aleph(Y) = \aleph(M)$, $Y \cap M = \varnothing$. To extend φ_M to a bijection

$$\begin{aligned}\varphi: (M \cup Y) \to & (M \cup Y) \times (M \cup Y) \\ & = (M \times M) \cup (M \times Y) \cup (Y \times M) \cup (Y \times Y),\end{aligned}$$

we need show (I, **6.8**) only that a bijection

$$Y \to (M \times Y) \cup (Y \times M) \cup (Y \times Y)$$

exists. This is, however, apparent, since on the right side the sets are pairwise disjoint, so from $\aleph(Y)\cdot\aleph(Y) = \aleph(Y)$ and the lemma, the right side has cardinal number $\underline{3}\cdot\aleph(Y) = \aleph(Y)$. Since

$$(M, \varphi_M) \prec (M \cup Y, \varphi),$$

this contradicts the maximality of (M, φ_M). Thus $\aleph(M) < \aleph(X)$ is impossible, and as remarked, this proves the theorem.

The basic facts of cardinal arithmetic follow from **8.5**. They show that the only way to obtain larger cardinal numbers is through exponentiation.

8.6 Corollary Let $\aleph$, $\overline{\aleph}$ be two cardinal numbers, at least one of which is transfinite. Then:

(a). $\aleph + \overline{\aleph} = \aleph\cdot\overline{\aleph} = \max(\aleph, \overline{\aleph})$.

(b). If $\aleph < \overline{\aleph}$, then $\overline{\aleph} - \aleph = \overline{\aleph}$; that is, removal of a set having a *smaller* cardinal number does not reduce the cardinal number of the given set.

(c). If $\underline{2} \leqslant \aleph \leqslant \overline{\aleph}$, then $\aleph^{\overline{\aleph}} = \underline{2}^{\overline{\aleph}}$.

(d). $\aleph^{\underline{n}} = \aleph$ for each finite $\underline{n}$ and transfinite $\aleph$.

Proof: (a). Assume that $\aleph \leqslant \overline{\aleph}$. Then:

$$\overline{\aleph} = \underline{1}\cdot\overline{\aleph} \leqslant \aleph\cdot\overline{\aleph} \leqslant \overline{\aleph}\cdot\overline{\aleph} = \overline{\aleph},$$

$$\overline{\aleph} \leqslant \overline{\aleph} + \aleph \leqslant \overline{\aleph} + \overline{\aleph} = \underline{2}\cdot\overline{\aleph} \leqslant \overline{\aleph}\cdot\overline{\aleph} = \overline{\aleph}.$$

(b). Let $A \subset X$ and $\aleph(A) < \aleph(X)$. Since $X = A \cup (X - A)$ is a disjoint union, $\aleph(X) = \aleph(A) + \aleph(X - A) = \max[\aleph(A), \aleph(X - A)]$; since $\aleph(A) < \aleph(X)$, we obtain $\aleph(X) = \aleph(X - A)$.

(c). By **7.8**(2), always $\aleph < \underline{2}^{\aleph}$; thus:

$$\underline{2}^{\overline{\aleph}} \leqslant \aleph^{\overline{\aleph}} \leqslant (\underline{2}^{\aleph})^{\overline{\aleph}} = \underline{2}^{\aleph\cdot\overline{\aleph}} = \underline{2}^{\overline{\aleph}}.$$

(d). This follows by induction, starting from $\aleph^{\underline{2}} = \aleph\cdot\aleph = \aleph$.

Ex. 5 Card E^n = card E^1; that is, "E^1 has just as many points as E^n." For, $E^n = E^1 \times \cdots \times E^1$ (n-factors) and **8.6**(d) applies.

Ex. 6 There are uncountably many transcendental numbers. For, by Ex. 2, there are $\aleph_0$ many algebraic numbers, and **8.6**(b) shows that the cardinal number of the set of transcendental numbers is $\mathfrak{c}$.

Ex. 7 The cardinal number of the set of all continuous real-valued functions on E^1 is $\mathfrak{c}$. For, a continuous function is completely determined by its values on the (countable) set of rationals; therefore it can be identified with a map $Q \to E^1$. Thus their cardinal number does not exceed $\mathfrak{c}^{\aleph_0} = (\underline{2}^{\aleph_0})^{\aleph_0} = \underline{2}^{\aleph_0} = \mathfrak{c}$, and since the subset of constant maps has cardinal number $\mathfrak{c}$, the assertion follows. Note, however, that the set of *all* real-valued functions on E^1 has cardinal number $\mathfrak{c}^{\mathfrak{c}} = 2^{\mathfrak{c}} > \mathfrak{c}$.

Ex. 8 For any ordinal number α, let ω_α be the ordinal number $\aleph_\alpha$ (*cf.* **7.10**). Then, in $\mathscr{Z}$, $\aleph([\omega_\alpha, \omega_{\alpha+1}[) = \aleph_{\alpha+1}$. Indeed, since (by definition) $\aleph_\alpha$ is the first ordinal in its equipotence class, we have $\aleph([0, \gamma[) \leqslant \aleph_\alpha$ for each $\gamma < \omega_{\alpha+1}$ and $\aleph([0, \omega_{\alpha+1}[) = \aleph_{\alpha+1}$, so we apply **8.6**(b). In particular, given any countably infinite X, there are exactly $\aleph_1$ ways in which X can be well-ordered.

Ex. 9 The theorem makes no assertion for exponentials $\aleph^{\bar{\aleph}}$ when $\bar{\aleph} < \aleph$ and $\bar{\aleph} \geqslant \aleph_0$. Assuming that $\bar{\aleph} \geqslant \aleph_0$, we show:

(i). If $\bar{\aleph} < \aleph$, then $\aleph \leqslant \aleph^{\bar{\aleph}} \leqslant 2^{\aleph}$, and
(ii). If $\aleph = 2^{\bar{\aleph}}$, then $\aleph^{\bar{\aleph}} = \aleph$.

(i) follows from $\aleph \leqslant \aleph^{\bar{\aleph}} \leqslant \aleph^{\aleph} = 2^{\aleph}$, and (ii) from $\aleph^{\bar{\aleph}} = 2^{\bar{\aleph}\bar{\aleph}} = 2^{\bar{\aleph}} = \aleph$. Observe that if the generalized continuum hypothesis be taken as an axiom, then (i) and **8.6**(c) show that the only infinite cardinals of form $\aleph^{\bar{\aleph}}$ are those of form $2^{\aleph'}$.

We will need two further consequences of **8.6**, the first of which elaborates on **7.4**(b).

8.7 Let Y be an infinite set, and $f: X \to Y$ surjective. If $\operatorname{card} f^{-1}(y) \leqslant \operatorname{card} Y$ for each $y \in Y$, then $\operatorname{card} X = \operatorname{card} Y$.

Proof: Since the sets $\{f^{-1}(y) \mid y \in Y\}$ partition X, the hypothesis gives [**8.3**(a)] $\aleph(X) \leqslant \aleph(Y)\cdot\aleph(Y) = \aleph(Y)$; but, from **7.4**(b), we have $\aleph(Y) \leqslant \aleph(X)$ also.

8.8 Let Y be an infinite set, and $\mathscr{A} \subset \mathscr{P}(Y)$ the set of all its *finite* subsets. Then $\operatorname{card} \mathscr{A} = \operatorname{card} Y$.

Proof: For each n, let $\mathscr{A}_n$ be the set of all n-element subsets of Y, and let Y^n be the cartesian product of n copies of Y. It is clear that there is an injection $\mathscr{A}_n \to Y^n$ consequently $\aleph(\mathscr{A}_n) \leqslant \aleph(Y^n) = \aleph(Y)^n = \aleph(Y)$ and, by **8.2**, **8.3**(a), we find that $\aleph(\mathscr{A}) = \aleph(\bigcup_n \mathscr{A}_n) \leqslant \sum_n \aleph(\mathscr{A}_n) \leqslant \sum_n \aleph(Y) = \aleph_0 \cdot \aleph(Y) = \aleph(Y)$. On the other hand, the map $y \to \{y\}$ of Y into $\mathscr{A}$ is injective, so we also have $\aleph(Y) \leqslant \aleph(\mathscr{A})$ and the proof is complete.

9. The Ordinal Number Ω

Let $\aleph_1$ be the first cardinal number larger than $\aleph_0$. When treated as an ordinal number, $\aleph_1$ is denoted by Ω and called the *first uncountable ordinal*. We derive two properties of $\mathscr{Z}(\Omega)$ that will be used later.

The initial interval $\mathscr{Z}(\omega) = [0, \omega[$ has the property that every *finite* subset has an upper bound in $\mathscr{Z}(\omega)$. This generalizes to

9.1 Theorem Each countable subset of $[0, \Omega[$ has an upper bound in $[0, \Omega[$.

Proof: Let $A \subset \mathscr{L}(\Omega)$ be countable, and let S be the ideal

$$\cup \{\mathscr{L}(\alpha) \mid \alpha \in A\} \subset [0, \Omega[.$$

Since Ω is not equipotent to any smaller ordinal number, $\aleph(\mathscr{L}(\alpha)) \leqslant \aleph_0$ for each $\alpha < \Omega$, so that, as a countable union of countable sets, **8.3**(b) shows that $\aleph(S) \leqslant \aleph_0 < \aleph_1 = \aleph(\mathscr{L}(\Omega))$, and consequently that S cannot be isomorphic to $[0, \Omega[$. By **3.2**(b), $S = [0, \beta[$ for some $\beta < \Omega$ and β is the least upper bound of A.

9.2 Theorem Let f: $[0, \Omega[\to [0, \Omega[$ be any map such that $f(\alpha) < \alpha$ for all $\alpha \geqslant$ some α_0. Then there exists a $\beta_0 < \Omega$ with the following property: As α increases, its image $f(\alpha)$ repeatedly returns to value below β_0. In symbols: $\exists\, \beta_0\, \forall\, \beta\, \exists\, \alpha \geqslant \beta : f(\alpha) \leqslant \beta_0$.

Proof: Assume that the conclusion is false. Then

$$\forall\, \beta_0\, \exists\, \beta\, \forall\, \alpha \geqslant \beta : f(\alpha) > \beta_0,$$

and so we could define a map R: $\mathscr{L}(\Omega) \to \mathscr{L}(\Omega)$ by sending each $\beta_0 \in \mathscr{L}(\Omega)$ to the smallest β satisfying this condition. By inductive construction, **5**, Remark 4, there would exist a map φ: $[0, \omega[\to [0, \Omega[$ with $\varphi(0) = \alpha_0$ and $\varphi(n + 1) = R[\varphi(n)]$ for each n. Thus, writing $\alpha_n = \varphi(n)$, we would obtain a countable subset $\{\alpha_n \mid n \in N\} \subset \mathscr{L}(\Omega)$ such that for each n, $(\alpha \geqslant \alpha_{n+1}) \Rightarrow (f(\alpha) > \alpha_n)$. By **9.1**, $\{\alpha_n \mid n \in N\}$ would have a *least* upper bound $\bar{\alpha} < \Omega$, and we would have the contradictory statements:

(1). $f(\bar{\alpha}) < \bar{\alpha}$, since $\bar{\alpha} \geqslant \varphi(0) = \alpha_0$.

(2). $f(\bar{\alpha}) \geqslant \bar{\alpha}$ because $\bar{\alpha} \geqslant \alpha_{n+1}$ for each n, so that also $f(\bar{\alpha}) > \alpha_n$, showing that $f(\bar{\alpha})$ is an upper bound for $\{\alpha_n \mid n \in N\}$.

This proves the theorem.

9.3 Remark The ordinal Ω intervenes in any construction utilizing operations that involve at most countably many objects at a time. The general pattern and result is as follows:

9.4 Let X be a set, and $\mathscr{A}$ any family of maps $\prod_1^\infty X \to X$ (that is, all factors in the countable cartesian product are X). For any $A \subset X$, write

$$\mathscr{A}(A) = \bigcup \{f\left(\prod_1^\infty A\right) \mid f \in \mathscr{A}\}.$$

Then, given any M_0 such that $M_0 \subset \mathscr{A}(M_0)$, there exists an $M_\Omega \supset M_0$ such that

(1). $\mathscr{A}(M_\Omega) = M_\Omega$.

(2). M_Ω is the smallest set $\supset M_0$ and satisfying (1): that is, if $M \supset M_0$ and satisfies (1), then $M_\Omega \subset M$. In particular, M_Ω is uniquely determined by M_0.

(3). If $\max [\aleph(\mathscr{A}), \aleph(M_0)] \geqslant 2$, then $\aleph(M_\Omega) \leqslant [\aleph(\mathscr{A}) \cdot \aleph(M_0))]^{\aleph_0}$.

Proof: Ad (1). Let $R: \mathscr{P}(X) \to \mathscr{P}(X)$ be the map $A \to \mathscr{A}(A)$. By transfinite construction (**5.2**), there exists a map $\varphi: \mathscr{Z}(\Omega) \to \mathscr{P}(X)$ with $\varphi(0) = M_0$ and $\varphi(\alpha) = R[\bigcup \{\varphi(\beta) \mid \beta < \alpha\}]$ for each $\alpha < \Omega$. Writing $\varphi(\beta) = M_\beta \subset X$, let $M_\Omega = \bigcup \{M_\beta \mid \beta < \Omega\}$. Then:

(a). $M_\Omega \subset \mathscr{A}(M_\Omega)$, since for each $x \in M_\Omega$ there is a first β in $[1, \Omega[$ with $x \in M_\beta$, so that $x \in \mathscr{A}(\cup \{M_\alpha \mid \alpha < \beta\}) \subset \mathscr{A}(M_\Omega)$.

(b). $\mathscr{A}(M_\Omega) \subset M_\Omega$: If $\{x_i\} \in \prod_1^\infty M_\Omega$, then each coordinate $x_i \in \varphi(\alpha_i)$ for some $\alpha_i < \Omega$; letting $\beta < \Omega$ be an upper bound of $\{\alpha_i \mid i \in N\}$, we have $\bigcup \{x_i\} \subset \bigcup \{M_\alpha \mid \alpha \leqslant \beta\}$, so that $\mathscr{A}(\{x_i\}) \in M_{\beta+1} \subset M_\Omega$.

Ad (2). We proceed by transfinite induction. Given M, let $P(\alpha)$ be the proposition "$\bigcup \{M_\beta \mid \beta \leqslant \alpha\} \subset M$." $P(0)$ is clearly true by hypothesis, and assuming that $P(\beta)$ is true for all $\beta < \alpha$, it follows that $M_\alpha = \mathscr{A}(\bigcup \{M_\beta \mid \beta < \alpha\}) \subset \mathscr{A}(M) = M$, so that $P(\alpha)$ will be true also.

Ad (3). Again we use transfinite induction. Let $P(\alpha)$ be the proposition "$\aleph(\bigcup \{M_\beta \mid \beta \leqslant \alpha\}) \leqslant \aleph(\mathscr{A})^{\aleph_0} \cdot \aleph(M_0)^{\aleph_0}$." $P(0)$ is obviously true. Assuming $P(\beta)$ to be true for all $\beta < \alpha$, we have [using **8.2**, **8.3**(a), and **8**, Ex. 4] that

$$\aleph(M_\alpha) \leqslant \aleph(\mathscr{A}) \cdot [\aleph(\mathscr{A})^{\aleph_0} \aleph(M_0)^{\aleph_0}]^{\aleph_0} = \aleph(\mathscr{A})^{\aleph_0} \cdot \aleph(M_0)^{\aleph_0},$$

and consequently, by **8.2** again,

$$\begin{aligned}\aleph(\bigcup \{M_\beta \mid \beta \leqslant \alpha\}) &\leqslant \aleph_1 \cdot \aleph(\mathscr{A})^{\aleph_0} \cdot \aleph(M_0)^{\aleph_0} \\ &= \aleph(\mathscr{A})^{\aleph_0} \aleph(M_0)^{\aleph_0},\end{aligned}$$

the latter equality holding because one of $\aleph(\mathscr{A})$, $\aleph(M_0) \geqslant 2$ and $\aleph_1 \leqslant 2^{\aleph_0}$. Thus $P(\alpha)$ is true, proving (3).

Ex. 1 A family of sets $\mathscr{B} \subset \mathscr{P}(X)$ is called a *Borel family* if, whenever $X_i \in \mathscr{B}$ for all $i \in N$, then $\bigcup_1^\infty X_i \in \mathscr{B}$ and $\bigcap_1^\infty X_i \in \mathscr{B}$ also. Letting $\mathscr{M}$ be *any* family of sets, and letting $\mathscr{A}$ consist of the two maps $\cap, \cup: \prod_1^\infty \mathscr{P}(X) \to \mathscr{P}(X)$ defined by $\{X_i\} \to \bigcap_1^\infty X_i$, $\{X_i\} \to \bigcup_1^\infty X_i$, respectively, **9.4** asserts that there is a unique smallest Borel family $\mathscr{B}(\mathscr{M}) \supset \mathscr{M}$, and that in fact, $\aleph(\mathscr{B}(\mathscr{M})) \leqslant \aleph(\mathscr{M})^{\aleph_0}$. The existence of $\mathscr{B}(\mathscr{M})$ can also be established more directly by observing (i) that $\mathscr{P}(X)$ is a Borel family, and (ii) that any intersection $\bigcap_\alpha \mathscr{B}_\alpha$ of Borel families is also a Borel family. Then $\mathscr{B}(\mathscr{M})$ is simply the intersection of all Borel families containing $\mathscr{M}$. However, this approach gives no information about the cardinality of the family $\mathscr{B}(\mathscr{M})$.

Ex. 2 The existence of M_Ω in **9.4** can also be established directly, as follows: Let $B = \bigcup \{M \mid M \subset \mathscr{A}(M) \wedge (M_0 \subset M\}$. Then $B \subset \mathscr{A}(B)$, since $[x \in B] \Rightarrow [x$ belongs to some $M] \Rightarrow [x \in \mathscr{A}(M) \subset \mathscr{A}(B)]$. On the other hand, $B \subset \mathscr{A}(B)$ implies that $\mathscr{A}(B) \subset \mathscr{A}(\mathscr{A}(B))$; that is, $\mathscr{A}(B)$ is one of the sets M. Thus $\mathscr{A}(B) \subset B$, showing that $\mathscr{A}(B) = B$. Again no information about the cardinality of B is obtained.

Problems

Section 1

1. Let Δ be the diagonal in $A \times A$. Show that $R \subset A \times A$ is a preorder if and only if $\Delta \subset R$ and $R \circ R = R$.
2. In Z^+, define $m \prec n$ if n divides m. Show that this is a partial ordering, that every chain has an upper bound, and determine the set of maximal elements.
3. Let $\mathscr{F}$ be the set of all real-valued functions of a real variable. Show that by defining $f \prec g$ to mean "$\forall x: f(x) \leqslant g(x)$," $(\mathscr{F}, \prec)$ is a partially ordered set. If $f \prec' g$ denotes
$$(f = g) \quad \text{or} \quad \left(\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0\right),$$
is $(\mathscr{F}, \prec')$ partially ordered?
4. A partially ordered set is a lattice if each pair of its elements has a least upper bound and a greatest lower bound. Is $(\mathscr{F}, \prec)$ in Problem 3 a lattice? Is $(\mathscr{P}(X), \subset)$ a lattice? In $(\mathscr{P}(A \times A), \subset)$, is the set of all transitive relations a lattice? Determine also if the set of all partial orders, preorders, and well-orders are each lattices.
5. Let A be the set of all infinite sequences of real numbers. Order A lexicographically; that is, $(a_1, a_2, \cdots) \prec (b_1, b_2, \cdots)$ if either $a_i = b_i$ for all i, or $a_n < b_n$ at the first place n where they differ. Show that this is a total ordering in A. Is the conventional ordering of the rationals the same as the lexicographic ordering of their decimal expansions?
6. Let X be a totally ordered set. A pair of subsets A, B satisfying (a) $A \cup B = X$, (b) $A \cap B = \varnothing$, and (c) $(a \in A) \wedge (b \in B) \Rightarrow a \prec b$ is called a cut in X. If A, B, and A', B' are cuts, show that $(A \subset A') \vee (A' \subset A)$.
7. Let $R \subset A \times A$ be a well-order. Show that unless A is a finite set, R^{-1} is not a well-order.
8. Show that if A is a finite set, each total ordering is a well-ordering.
9. Let A, B be well-ordered. Show that lexicographic ordering in $A \times B$ is also a well-ordering.
10. Let A be well-ordered. Show that there does not exist any sequence $\{a_n \mid n \in N\}$ with $(a_{n+1} \prec a_n) \wedge (a_{n+1} \neq a_n)$ for each n.
11. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of well-ordered sets, and assume that $\mathscr{A}$ is also well-ordered. Order $\prod_\alpha A_\alpha$ lexicographically (as in Problem 5, this means: if β is the first element in $\{\alpha \in \mathscr{A} \mid p_\alpha(x) \neq p_\alpha(y)\}$, then $x \prec y$ if and only if $p_{\beta_0}(x) \prec p_{\beta_0}(y)$). Using Problem 10, show that this a well-ordering in $\prod_\alpha A_\alpha$ if and only if $\mathscr{A}$ is a finite set.

Section 2

1. Prove the following extension of Zorn's lemma: If X is preordered and if each chain in X has an upper bound, then for each given $x_0 \in X$ there exists a maximal m with $x_0 \prec m$.

2. Prove the following equivalent to the axiom of choice.
   (a). [H. Kneser's form of Zorn's Lemma.] If X is a preordered set, and if each well-ordered subset has an upper bound, then X has at least one maximal element [*Hint:* One way to prove this is to require that a φ-tower be well-ordered by inclusion, rather than merely totally ordered, and then modify the proofs in **2.1**, **2.3** accordingly. Another proof first establishes the result in Problem 7 for this case].
   (b) If $F: X \to \mathscr{P}(Y)$ is any mapping such that $F(x) \neq \varnothing$ for each $x \in X$, then there exists an $f: X \to Y$ such that $f(x) \in F(x)$ for each $x \in X$.
   (c). If X is a partially ordered set, then each chain in X is contained in a maximal chain. [M is a maximal chain if $(M \subset C) \Rightarrow (M = C)$ for each chain $C \subset X$.]
   (d). If $\mathscr{P}(X)$ is partially ordered by inclusion, then each subset $P \subset \mathscr{P}(X)$ of finite character contains a maximal element. [$P \subset \mathscr{P}(X)$ is of finite character if $(A \in P) \Leftrightarrow$ (every finite subset of A belongs to P)].
3. Let $\mathscr{P}(X)$ be partially ordered by inclusion, and let $\mathscr{A} \subset \mathscr{P}(X)$ be a maximal chain. In X, define $x \prec y$ if either $x = y$ or $\exists A \in \mathscr{A}: (x \in A) \wedge (y \mathrel{\bar{\in}} A)$. Show that this is a total order in X. If, in addition, for each $\mathscr{L} \subset \mathscr{A}$, it is true that
$$\bigcap \{A \mid A \in \mathscr{L}\} \in \mathscr{A},$$
show that $\prec$ is a well-order.
4. Let X be partially ordered by $R \subset X \times X$. Show that there exists a total order $R' \supset R$.
5. Let A be partially ordered. A set $B \subset A$ is called *cofinal* in A if
$$\forall a \in A \; \exists b \in B: a \prec b.$$
Prove that every totally ordered set has a cofinal well-ordered subset.
6. A partially ordered set is "of type ω" if each chain C has a cofinal *sequence* $\{c_i \mid i \in N\}$. Prove that if, in a partially ordered set of type ω, each ascending *sequence* has an upper bound, then there is a maximal element.
7. Let X be a preordered set, and assume that each chain in X has an upper bound. Let $f: X \to X$ satisfy $x \prec f(x)$ for each $x \in X$. Prove: there exists at least one x_0 such that $f(x_0) \prec x_0$ also.
8. Let Z be the additive group of reals, and C the additive group of complex numbers. Prove that the groups Z and C are isomorphic.

Section 3

1. Prove **3.4** directly from the concept of φ-tower.
2. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of well-ordered sets. Assume that for each $(\alpha, \beta) \in \mathscr{A} \times \mathscr{A}$, one of A_α, A_β is an ideal of the other. Prove that in $\bigcup_\alpha A_\alpha$, there is exactly one well-ordering that coincides with that on each A_α; also prove that with this well-ordering, each initial interval of an A_α is also an initial interval in $\bigcup_\alpha A_\alpha$.
3. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a pairwise disjoint family of well-ordered sets, and assume that $\mathscr{A}$ is also well-ordered. In $\bigcup_\alpha A_\alpha$, define $\prec$ by preserving the given ordering on each A_α and by setting $a \prec b$ whenever $a \in A_\alpha$, $b \in A_\beta$, and $\alpha \prec \beta$. Show that this is a well-ordering in $\bigcup_\alpha A_\alpha$.

Section 4

1. Prove **4.2**, using Zorn's lemma.
2. Each of the sets $\{1, 2, \cdots, \omega\}$, $\{\omega, 1, 2, \cdots\}$, and $\{1, 3, 5, \cdots; 2, 4, 6 \cdots\}$, ordered as written, is well-ordered. Show that no two are isomorphic, and arrange them in order of magnitude.
3. In the well-ordering of **3**, Problem 3, show that $A_\alpha \leqslant \bigcup_\alpha A_\alpha$ for each $\alpha \in \mathscr{A}$.
4. Let A, B be well-ordered. Order $A \times B$ and $B \times A$ lexicographically. Are the sets $A \times B$ and $B \times A$ isomorphic?

Section 5

1. Prove **4.2** by transfinite induction.
2. If W is any ordinal with infinitely many members, show that $\omega \leqslant W$.
3. Show that a set contains infinitely many members if and only if there is a bijective map of it onto a proper subset.
4. Let E be totally ordered. Assume $Q \subset E$ satisfies
   a. There is an ideal $S \neq \varnothing$ with $S \subset Q$.
   b. For each ideal $S_1 \subset Q$ with $S_1 \neq E$, there exists an ideal $S_2 \neq S_1$ with $S_1 \subset S_2 \subset Q$.

   Prove that $Q = E$.
5. Let E be totally ordered, and have a first element. Show that E is well-ordered if and only if the transfinite induction principle **5.1** is true.
6. By ω^* denote the ordered set of nonpositive integers in their natural order. Prove: A necessary and sufficient condition that a totally ordered set be well-ordered is that it contain no subset isomorphic to ω^*.

Section 6

1. Given two ordinal numbers $\underset{\sim}{\alpha}$, $\underset{\sim}{\beta}$, the sets $A = \underset{\sim}{\alpha}$, $B = \{\varnothing\} \times \underset{\sim}{\beta}$ (ordered in obvious fashion) are disjoint. Well-order $A \cup B$ by preserving the order on each and having each element of A precede all of B. Well-order $A \times B$ by $(a, b) \prec (\tilde{a}, \bar{b})$ if $(b, a) \prec (\bar{b}, \tilde{a})$ in the lexicographic ordering of $B \times A$. The sum $\underset{\sim}{\alpha} + \underset{\sim}{\beta} = \operatorname{ord}(A \cup B)$ and product $\alpha \cdot \beta = \operatorname{ord}(A \times B)$. Prove:
   a. $\underset{\sim}{\alpha} + \underset{\sim}{\beta} \neq \underset{\sim}{\beta} + \underset{\sim}{\alpha}$ in general.
   b. $\underset{\sim}{2} \cdot \underset{\sim}{\omega} \neq \underset{\sim}{\omega} \cdot \underset{\sim}{2}$.
   c. $\underset{\sim}{2} \cdot \underset{\sim}{\omega} = \underset{\sim}{\omega}$.
   d. $\underset{\sim}{\alpha} \cdot (\underset{\sim}{\beta} + \underset{\sim}{\gamma}) = \underset{\sim}{\alpha} \cdot \underset{\sim}{\beta} + \underset{\sim}{\alpha} \cdot \underset{\sim}{\gamma}$, but $(\underset{\sim}{\beta} + \underset{\sim}{\gamma}) \cdot \underset{\sim}{\alpha} \neq \underset{\sim}{\beta} \cdot \underset{\sim}{\alpha} + \underset{\sim}{\gamma} \cdot \underset{\sim}{\alpha}$ in general.
   e. $\underset{\sim}{\alpha} < \underset{\sim}{\beta}$ if and only if there is a $\underset{\sim}{\gamma} > \underset{\sim}{0}$ with $\underset{\sim}{\alpha} + \underset{\sim}{\gamma} = \underset{\sim}{\beta}$ (and show that $\underset{\sim}{\gamma}$ is uniquely determined by $\underset{\sim}{\alpha}$ and $\underset{\sim}{\beta}$).
2. Prove that every ordinal number is of form $\underset{\sim}{\alpha} + \underset{\sim}{n}$, where $\underset{\sim}{\alpha}$ is a limit ordinal number and $\underset{\sim}{n}$ is an integer.

Section 7

1. Prove: A set is finite if and only if every total ordering is a well-ordering.
2. Show that the set of all *countable* subsets of E^1 has cardinal number $\mathfrak{c}$.
3. Prove: If every countable subset of a totally ordered set is well-ordered, then the entire set is well ordered.

4. Prove: For each ordinal number $\underset{\sim}{\alpha}$, there exists a cardinal number $\aleph_{\underset{\sim}{\alpha}}$.
5. Prove: Each infinite cardinal number is a limit ordinal number.
6. Let X be a set. For the characteristic functions on X prove:

$$c_{A\cap B} = c_A \cdot c_B; \qquad c_{A\cup B} = c_A + c_B - c_{A\cap B};$$

$$c_{\bigcup_1^n A_i} = 1 - \prod_1^n (1 - c_{A_i});$$

if $Q = (A - B) \cup (B - A)$, then

$$c_Q = c_A + c_B.$$

7. Prove the Bernstein-Schröder theorem directly; that is, that

$$(\text{card } A \leqslant \text{card } B) \wedge (\text{card } B \leqslant \text{card } A) \Rightarrow (\text{card } A = \text{card } B)$$

by using Problem 7 in I, **6**.
8. Let $f: Z^+ \times Z^+ \to Z^+$ be given by $(p, q) \to [p + (1/2)(p + q - 1)(p + q - 2)]$. Prove that f is bijective. (f gives the "diagonal enumeration" of $Z^+ \times Z^+$).

Section 8

1. Theorem **8.5** is not true for finite cardinal numbers. Where does the proof break down?
2. Show that in **8.6**(b), $\overline{\aleph} - \aleph$ is indeterminate if $\overline{\aleph} = \aleph$, and may be any cardinal $\leqslant \overline{\aleph}$.
3. Prove: Any set of nondegenerate disjoint intervals in E^1 is countable. The same result holds for balls in E^n.
4. Let $E^1 \times E^1$ be ordered lexicographically. Show that there does not exist any $f: E^1 \times E^1 \to E^1$ such that:

$$[(x, y) \prec (x', y')] \wedge [(x, y) \neq (x', y')] \Rightarrow f(x, y) < f(x', y').$$

5. Show that with any uncountable set of real numbers, one can always construct a divergent series.
6. Prove: If $\aleph(E) = \aleph$, then $E = \bigcup \{A_\alpha \mid \alpha \in \mathscr{A}\}$, where $\aleph(\mathscr{A}) = \aleph$ and $\aleph(A_\alpha) = \aleph$ for each $\alpha \in \mathscr{A}$.
7. Show that $\aleph_2^{\aleph_0} > \aleph_1^{\aleph_0}$ if and only if $\aleph_1 = 2^{\aleph_0}$.
8. What is the cardinal number of the set of all injections $A \to A$?
9. Let $\aleph(X) > \aleph_0$ and let Y be an arbitrary set. Let f_i, $i \in N$ be maps $X \to Y$. Let $g: X \to Y$ be such that:

$$\forall\, i: \aleph(\{x \mid g(x) = f_i(x)\}) \leqslant \aleph_0.$$

Show that there exists an $x_0 \in X$ such that $\forall\, i: g(x_0) \neq f_i(x_0)$.
10. Let E be an infinite set. Prove: There exists a partition $\{E_i \mid i = 0, 1, \cdots\}$ of E such that $\aleph(E_i) = \aleph(E)$ for each $i = 0, 1, \cdots$.
11. Let E be an infinite set, and $\{f_n \mid n \in Z^+\}$ any family of mappings of E into itself. Prove: There exist two mappings $\varphi, \psi: E \to E$ such that $f_n = \varphi^2 \circ \psi^n \circ \varphi \circ \psi$ for each $n \in Z^+$ (here, $g^n = g \circ g \circ \cdots \circ g$ n times). [*Hint:* Let $\{E_i \mid i = 0, 1, \cdots\}$ be a partition of E as in Problem 10, and choose any $\psi: E \to E$ such that $\psi \mid E_i: E_i \to E_{i+1}$ is bijective for each $i = 0, 1, \cdots$. Now let $\{E_{0,\,n} \mid n \in Z^+\}$ be a partition of E_0 as in Problem 10, and define $\varphi: E - E_0 \to E_0$ so that $\varphi \mid E_n: E_n \to E_{0,\,n}$ is bijective for each $n \in Z^+$. To define φ on E_0, observe that $\theta_n = \varphi \circ \psi^n \circ \varphi \circ \psi: E \to E_{0,\,n}$, and set $\varphi(x) = f_n \circ \theta_n^{-1}(x)$ for $x \in E_{0,\,n}$. This proof is due to S. Banach.]

Section 9

1. Prove an analogue of **9.1** for accessible cardinal numbers and explain the difficulty when the cardinal number is inaccessible.
2. Let $]0, \Omega[= \cup \{]\alpha_\mu, \beta_\mu] \mid \mu \in \mathscr{M}\}$, where each $]\alpha_\mu, \beta_\mu] \subset]0, \Omega[$. Now prove that $\aleph(\mathscr{M}) = \aleph_1$, and that there is a $\beta_0 < \Omega$ that lies in at least $\aleph_0$ intervals $]\alpha_\mu, \beta_\mu]$.
3. Let $\mathscr{M} \subset \mathscr{P}(E^n)$ be any family of sets with $\aleph(\mathscr{M}) \leqslant 2^{\aleph_0}$, and let $\mathscr{B}(\mathscr{M})$ be the smallest Borel family containing $\mathscr{M}$. Show $\mathscr{B}(\mathscr{M}) \neq \mathscr{P}(E^n)$.
4. A family $\mathscr{A} \subset \mathscr{P}(X)$ is called *additive* if $X, Y \in \mathscr{A} \rightarrow X \cup Y \in \mathscr{A}$; it is called *multiplicative* if $X, Y \in \mathscr{A} \Rightarrow X \cap Y \in \mathscr{A}$. Prove: Given any family $\mathscr{M} \subset \mathscr{P}(X)$, there exists a smallest additive family $a(\mathscr{M}) \supset \mathscr{M}$ and a smallest multiplicative family $m(\mathscr{M}) \supset \mathscr{M}$. Furthermore, $\aleph(a(\mathscr{M})) = \aleph(m(\mathscr{M})) \leqslant \aleph_0 \cdot \aleph(\mathscr{M})$.
5. $\mathscr{A} \subset \mathscr{P}(X)$ is called a finite Borel family if

$$X_i \in \mathscr{A},\ i = 1, \cdots, n \Rightarrow \bigcup_1^n X_i,\ \bigcap_1^n X_i \in \mathscr{A}$$

for all finite n. Given any family of sets $\mathscr{M}$, show that there is a smallest finite Borel family $\mathscr{B}_F(\mathscr{M}) \supset \mathscr{M}$, and estimate its cardinality. Show that $\mathscr{B}_F(\mathscr{M})$ is the smallest of all families containing $\mathscr{M}$ that are both additive and multiplicative.

Topological Spaces

III

Though E^1 and E^n are obviously different, this is not due to one having more points than the other, since we know that card E^1 = card E^n. Geometrically, it is evident that the points are arranged differently, so that different subsets are "close together." To detect inherent differences of this sort, we study sets in which a notion of "nearness" is specified, that is, in which a topology is specified.

1. Topological Spaces

1.1 Definition Let X be a set. A topology (or topological structure) in X is a family $\mathscr{T}$ of subsets of X that satisfies:

(1). Each union of members of $\mathscr{T}$ is also a member of $\mathscr{T}$.
(2). Each *finite* intersection of members of $\mathscr{T}$ is also a member of $\mathscr{T}$.
(3). $\varnothing$ and X are members of $\mathscr{T}$.

1.2 Definition A couple $(X, \mathscr{T})$ consisting of a set X and a topology $\mathscr{T}$ in X is called a *topological space.*

Instead of "topological space $(X, \mathscr{T})$," we also say "$\mathscr{T}$ is the topology of the space X," or "X carries topology $\mathscr{T}$." When it is not necessary to specify $\mathscr{T}$ explicitly, we simply say, "X is a *space*" (to distinguish from "X is a *set*").

Elements of topological spaces are called *points*. The members of $\mathscr{T}$ are called the "open sets" of the topological space $(X, \mathscr{T})$ (or of the topology $\mathscr{T}$). There is no preconceived idea of what "open" means, other than that the sets called "open" in any discussion satisfy the axioms **1.1** (1–3). Observe that since the union (resp. intersection) of an empty family of sets in X is $\varnothing$ (resp. X), Axiom (3) of **1.1** is actually redundant.

1.3 Definition Let $(X, \mathscr{T})$ be a space. By a neighborhood (written nbd) of an $x \in X$ is meant any open set (that is, member of $\mathscr{T}$) containing x.

"U is a nbd of x" is written "nbd $U(x)$"; the points of $U(x)$ are "U-close" to x, so that $\mathscr{T}$ organizes X into chunks of "nearby" points.

Ex. 1 Let X be any set; $\mathscr{I} = \{\varnothing, X\}$. This topology, in which no set other than $\varnothing$ and X is open, is called the *indiscrete topology* $\mathscr{I}$: There are no "small" nbds.

Ex. 2 In the set X, let $\mathscr{T} = \mathscr{P}(X)$. This is called the *discrete topology* $\mathscr{D}$: every set is an open set. Comparing with Ex. 1 indicates the sense in which different topologies in a set X give different organizations of the points of X.

Ex. 3 The topological space consisting of the two points $\{0, 1\}$ with the discrete topology is denoted by 2. The same set, with the topology $\mathscr{T} = \{\varnothing, 0, X\}$ is called the Sierpinski space $\mathscr{S}$, and will be used frequently in the sequel. In contrast to the space 2, $\{1\}$ has no "small" nbds in $\mathscr{S}$.

Ex. 4 Let E^1 be the set of real numbers. In the usual introduction to analysis, a subset $G \subset E^1$ is called "open" if for each $x \in G$ there is an $r > 0$ such that the symmetric open interval $B(x; r) = \{y \mid |y - x| < r\} \subset G$. We verify that the family $\mathscr{T}$ of sets declared "open" by this criterion actually is a topology in the set E^1.

Ad (1). If each member of $\{G_\alpha \mid \alpha \in \mathscr{A}\}$ is "open," so also is $\bigcup_\alpha G_\alpha$, since $x \in \bigcup_\alpha G_\alpha \Rightarrow (\exists\, \alpha : x \in G_\alpha) \Rightarrow (\exists\, r > 0 : B(x; r) \subset G_\alpha \subset \bigcup_\alpha G_\alpha)$.

Ad (2). If $G_1, \cdots, G_n$ are "open," so also is $\bigcap_1^n G_i$ because $x \in \bigcap_1^n G_i \Rightarrow$ $(\forall\, i : x \in G_i) \Rightarrow (\forall\, i\, \exists\, r_i > 0 : B(x; r_i) \subset G_i) \Rightarrow [B(x; \min(r_1, \cdots, r_n)) \subset \bigcap_1^n G_i]$.

(3) is trivial.

This topology, $\mathscr{T}$, is called the *Euclidean topology* of E^1; the topological space $(E^1, \mathscr{T})$ is called the Euclidean 1-space. Note that $\mathscr{T}$ is *not* the indiscrete topology in the set E^1, that is, $\mathscr{T}$ does not consist only of $\varnothing$ and E^1: indeed, each $B(y; r)$ belongs to $\mathscr{T}$, since, given any $x \in B(y; r)$, we have $d = |y - x| < r$, and therefore $x \in B(x; r - d) \subset B(y; r)$. It is simple to see that $\mathscr{T}$ can also be described more directly as the family of all unions of open intervals.

Ex. 5 Let E^n be the set of all ordered n-uples of real numbers. Using vector notation (*cf.* I, **1**) call "ball of center x and radius r" the set $B(x; r) = \{y \mid |y - x| < r\}$. The *Euclidean topology* in E^n is determined by calling $G \subset E^n$ "open" if for each $x \in G$ there is some $r > 0$ such that $B(x; r) \subset G$. The verification that this criterion describes a topology $\mathscr{T}$, and that in fact each $B(x; r)$ belongs to $\mathscr{T}$, is contained in Ex. 4, since only reinterpretation of $B(x; r)$ is involved. With this topology, E^n is called *Euclidean n-space*. As before, the open sets in E^n can be described equally as the arbitrary unions of balls.

As the examples show, a set X may have many topologies; with each, it is a distinct topological space. By regarding each topology as a subset of $\mathscr{P}(X)$, the topologies in X are partially ordered by inclusion; clearly, $\mathscr{I} \subset \mathscr{T} \subset \mathscr{D}$ for each topology $\mathscr{T}$. We call $\mathscr{T}_1$ larger (or "with more open sets," or "finer") than $\mathscr{T}_0$ whenever $\mathscr{T}_0 \subset \mathscr{T}_1$. It is trivial to verify

1.4 Let $\{\mathscr{T}_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of topologies in X. Then $\bigcap_\alpha \mathscr{T}_\alpha = \{U \mid \forall\, \alpha \in \mathscr{A}: U \in \mathscr{T}_\alpha\}$ is also a topology in X; however, $\bigcup_\alpha \mathscr{T}_\alpha$ need not be a topology.

2. Basis for a Given Topology

The task of specifying a topology is simplified by giving only enough open sets to "generate" all the open sets.

2.1 Definition Let $(X, \mathscr{T})$ be a topological space. A family $\mathscr{B} \subset \mathscr{T}$ is called a basis for $\mathscr{T}$ if each open set (that is, member of $\mathscr{T}$) is the *union* of members of $\mathscr{B}$.

$\mathscr{B}$ is also called a "basis for the space X," and its members the "basic open sets of the topology $\mathscr{T}$." Not only is each member of $\mathscr{T}$ the union of members of $\mathscr{B}$, but also, because of $\mathscr{B} \subset \mathscr{T}$ and **1.1**, each union of members of $\mathscr{B}$ belongs to $\mathscr{T}$; thus a basis for $\mathscr{T}$ completely determines $\mathscr{T}$.

Ex. 1 $\mathscr{T}$ is a basis for $\mathscr{T}$.

Ex. 2 Let $\mathscr{D}$ be the discrete topology on X. Then $\mathscr{B} = \{\{x\} \mid x \in X\}$ is a basis for $\mathscr{D}$.

In view of Exs. 1 and 2, a given $\mathscr{T}$ may have many bases. The families $\mathscr{B} \subset \mathscr{T}$ that can serve as a basis are characterized by

2.2 Theorem Let $\mathscr{B} \subset \mathscr{T}$. The following two properties of $\mathscr{B}$ are equivalent:

(1). $\mathscr{B}$ is a basis for $\mathscr{T}$.

(2). For each $G \in \mathscr{T}$ and each $x \in G$ there is a $U \in \mathscr{B}$ with $x \in U \subset G$.

Proof: (1) $\Rightarrow$ (2). Let $x \in G$; since $G \in \mathcal{T}$ and $\mathcal{B}$ is a basis, $G = \bigcup_\alpha U_\alpha$, where each $U_\alpha \in \mathcal{B}$. Thus there is at least one $U_\alpha \in \mathcal{B}$ with $x \in U_\alpha \subset G$.

(2) $\Rightarrow$ (1). Let $G \in \mathcal{T}$; for each $x \in G$, find $U_x \in \mathcal{B}$ with $x \in U_x \subset G$; then $G = \bigcup \{U_x \mid x \in G\}$.

Ex. 3 In each E^n, $n \geqslant 1$, $\mathcal{B} = \{B(x; r) \mid x \in E^n, r > 0\}$ is a basis for the Euclidean topology, as the descriptions in **I**, Exs. 4 and 5 show.

Ex. 4 *E^n has a countable basis*: the family $\mathcal{B} = \{B(\xi, r) \mid \xi$ has all coordinates rational, and $r > 0$ is rational$\}$. For, let G be any open set, and $x \in G$. By **I**, Exs. 4 and 5, there is a $B(x, r) \subset G$, and we can clearly assume that r is rational. The reader will easily verify that there is a point ξ, with all coordinates rational, within a distance $r/3$ of x; then $x \in B(\xi, r/2) \subset B(x, r) \subset G$, as required by **2.2.** Immediate consequences are:

(a). Each set open in E^n is the union of at most countably many balls.

(b). The cardinal number of the *Euclidean topology* of E^n is $2^{\aleph_0}$.

By specifying a basis for $\mathcal{T}$, all the open sets are generated as unions. However, there is a more convenient way to describe the open sets:

2.3 Theorem Let $\mathcal{B} \subset \mathcal{T}$ be a basis for $\mathcal{T}$. Then A is open (that is, is in $\mathcal{T}$) if and only if for each $x \in A$ there is a $U \in \mathcal{B}$ with $x \in U \subset A$.

Proof: If A is open, the condition follows from **2.2.** Conversely, if the condition holds, then (as in **2.2**) we find $A = \bigcup \{U_a \mid a \in A\}$, where each $U_a \in \mathcal{B} \subset \mathcal{T}$; from **1.1** it follows that A is open.

In particular, since $\mathcal{T}$ is a basis for $\mathcal{T}$, **2.3** provides a very useful method for showing that a given set A is open.

3. Topologizing of Sets

In this section, two general methods for introducing topologies in sets will be given.

The first, and most popular, starts from *any* given family $\Sigma \subset \mathcal{P}(X)$ and leads to a unique topology containing Σ; thus we can to an extent preassign the notion of nearness desired.

3.1 Theorem Given *any* family $\Sigma = \{A_\alpha \mid \alpha \in \mathcal{A}\}$ of subsets of X, there always exists a unique, smallest topology $\mathcal{T}(\Sigma) \supset \Sigma$. The family $\mathcal{T}(\Sigma)$ can be described as follows: It consists of $\varnothing$, X, all finite intersections of the A_α, and all arbitrary unions of these finite intersections. Σ is called a subbasis for $\mathcal{T}(\Sigma)$, and $\mathcal{T}(\Sigma)$ is said to be generated by Σ.

Proof: Let $\mathscr{T}(\Sigma)$ be the intersection of all topologies containing Σ; such topologies exist, since $\mathscr{P}(X)$ is one such. By **1.4**, $\mathscr{T}(\Sigma)$ is a topology; it evidently satisfies the requirements "unique" and "smallest." To verify that the members of $\mathscr{T}(\Sigma)$ are as described, note that since $\Sigma \subset \mathscr{T}(\Sigma)$, it follows from **1.1** that $\mathscr{T}(\Sigma)$ must contain all the sets listed. Conversely, because $\bigcup_\alpha$ distributes over $\cap$, the sets listed actually do form a topology containing Σ, and which therefore contains $\mathscr{T}(\Sigma)$.

In **3.1**, we started from Σ and obtained a topology $\mathscr{T}(\Sigma) \supset \Sigma$. If, conversely, we are given a topology $\mathscr{T}$, a family $\Sigma \subset \mathscr{T}$ is called a subbasis for $\mathscr{T}$ whenever $\mathscr{T} = \mathscr{T}(\Sigma)$.

Ex. 1 For any topology $\mathscr{T}$, $\mathscr{T}$ is a subbasis for $\mathscr{T}$.

Ex. 2 The *finite* intersections of members of Σ are a *basis* for $\mathscr{T}(\Sigma)$.

Ex. 3 In the set E^1 of all real numbers, let Σ be all sets of form $\{x \mid x > a\}$ and $\{x \mid x < b\}$. Then $\mathscr{T}(\Sigma)$ is precisely the Euclidean topology: Each finite open interval, being an intersection of two subbasic open sets, belongs to $\mathscr{T}(\Sigma)$, and it is evident from the description of $\mathscr{T}(\Sigma)$ in **3.1** that the family $\mathscr{B}$ of all these finite intervals forms a basis for $\mathscr{T}(\Sigma)$. But, by **2**, Ex. 3, $\mathscr{B}$ is a basis for the Euclidean topology.

Ex. 4 In the set E^1 of all real numbers, let Σ be all sets of the form $\{x \mid x > a\}$ and $\{x \mid x \leqslant b\}$. The topology $\mathscr{T}(\Sigma)$ is called the *upper limit topology* and has the sets $]a, b]$ as basis. $\mathscr{T}(\Sigma)$ is not the Euclidean topology of E^1, since the sets $]a, b]$ do not belong to the Euclidean topology. The reader can verify that the Euclidean topology is smaller than $\mathscr{T}(\Sigma)$. The topological space $(E^1, \mathscr{T}(\Sigma))$ is denoted by E^1_u.

Ex. 5 Let Γ be any ordinal number and in $[0, \Gamma]$ use the topology generated by all sets of form $\{x \mid x > \alpha\}$ and $\{x \mid x < \beta\}$. We call this topological space the *ordinal space* $[0, \Gamma]$. Observe that the sets $]\alpha, \beta] = \{x \mid x > \alpha\} \cap \{x \mid x < \beta + 1\}$ are a *basis* for the topology. Note also that $[\alpha, \beta[$ is open if and only if $\alpha = 0$ or if α has an immediate predecessor.

Ex. 6 Let $(X_1, \mathscr{T}_1), \cdots, (X_n, \mathscr{T}_n)$ be spaces. The *cartesian product* topology in $X_1 \times \cdots \times X_n$ is that having as subbasis all sets (*cf.* I, **9.6**) of form $\langle A_i \rangle$, where $A_i \in \mathscr{T}_i$, $i = 1, \cdots, n$. In view of I **9.6**(1), a basis for this topology is simply the family of cartesian products of open sets of the spaces X_i.

The construction of a topology from a subbasis loses some control over the open sets; they build up from the finite intersections of the A_α rather than from the A_α themselves. In the second general method for topologizing a set, which we will now describe, the open sets are constructed only by union from the given family; that is, by specifying a family to be used as a *basis* for constructing the topology. Since intersections are involved in topologies but no intersections are involved in forming open sets from a basis, it is to be expected that not every family can serve as the basis for *some* topology.

3.2 Theorem Let $\mathscr{B} = \{U_\mu \mid \mu \in \mathscr{M}\}$ be any family of subsets of X that satisfies the following condition:

For each $(\mu, \lambda) \in \mathscr{M} \times \mathscr{M}$ and each $x \in U_\mu \cap U_\lambda$, there exists some U_α with $x \in U_\alpha \subset U_\mu \cap U_\lambda$.

Then the family $\mathscr{T}(\mathscr{B})$ consisting of $\varnothing$, X, and all unions of members of $\mathscr{B}$, is a topology for X; that is, $\mathscr{B} \cup \{\varnothing\} \cup \{X\}$ is a basis for *some* topology. $\mathscr{T}(\mathscr{B})$ is unique and is the smallest topology containing $\mathscr{B}$.

Proof: Using **3.1**, we obtain a topology $\mathscr{T}(\mathscr{B})$ having $\mathscr{B}$ as *subbasis.* To see that $\mathscr{T}(\mathscr{B})$ actually has $\mathscr{B}$ as a *basis*, we need show only that each finite intersection of members of $\mathscr{B}$ is in fact a union of members of $\mathscr{B}$; and, as in **2.2** ($2 \Rightarrow 1$), it suffices to show that for each $x \in U_1 \cap \cdots \cap U_n$, there is a $U \in \mathscr{B}$ with $x \in U \subset U_1 \cap \cdots \cap U_n$. We proceed by induction, the assertion being true (by the hypothesis) for $n = 2$. If it is true for $(n - 1)$, then writing $x \in U_1 \cap \cdots \cap U_n = (U_1 \cap \cdots \cap U_{n-1}) \cap U_n$, the inductive hypothesis gives $x \in U \cap U_n$ for some

$$U \subset U_1 \cap \cdots \cap U_{n-1}$$

in $\mathscr{B}$ so, by the case $n = 2$, we find

$$x \in U' \subset U \cap U_n \subset U_1 \cap \cdots \cap U_n$$

for some $U' \in \mathscr{B}$, completing the induction and the proof of the theorem.

The specification of a topology by giving a basis is generally accomplished by specifying for each $x \in X$ a family of nbds $\{U_\alpha(x) \mid \alpha \in \mathscr{A}(x)\}$ (called a "basis at x" or a "complete system of nbds at x") and verifying that the family $\mathscr{B} = \{U_\alpha(x) \mid \alpha \in \mathscr{A}(x), x \in X\}$ satisfies the requirement of **3.2.**

Ex. 7 Let C be the set of all continuous real-valued functions on $[0, 1]$. For each $f \in C$ and $\varepsilon > 0$, define

$$M(f, \varepsilon) = \{g \mid \int_0^1 |f - g| < \varepsilon\}.$$

The family $\{M(f, \varepsilon) \mid f \in C, \varepsilon > 0\}$ is a *basis* for some topology $\mathscr{M}$ in C. For, if $h \in M(f, \varepsilon) \cap M(g, \eta)$, let

$$r_f = \int_0^1 |f - h| \quad \text{and} \quad r_g = \int_0^1 |g - h|,$$

and let $\zeta = \min [\varepsilon - r_f, \eta - r_g]$; then $\zeta > 0$ and $M(h, \zeta) \subset M(f, \varepsilon) \cap M(g, \eta)$, since

$$\varphi \in M(h, \zeta) \Rightarrow \int_0^1 |f - \varphi| \leqslant \int_0^1 |f - h| + \int_0^1 |h - \varphi| < r_f + (\varepsilon - r_f) = \varepsilon,$$

so that $\varphi \in M(f, \varepsilon)$, and similarly, $\varphi \in M(g, \eta)$. Another topology $\mathscr{U}$ in C can be defined by using as a basis the sets

$$U(f, \varepsilon) = \{g \mid \sup_x | f(x) - g(x)| < \varepsilon\}.$$

Though each basis in X gives a unique topology, it is obvious that distinct bases may give the same topology: For example, $\mathscr{T}(\mathscr{B})$ has $\mathscr{B}$ and $\mathscr{T}(\mathscr{B})$ as bases. We now determine when this will occur.

3.3 Definition Two bases $\mathscr{B}$, $\mathscr{B}'$ in X are *equivalent* if $\mathscr{T}(\mathscr{B}) = \mathscr{T}(\mathscr{B}')$.

3.4 Theorem A necessary and sufficient condition that two bases $\mathscr{B}$, $\mathscr{B}'$ in X be equivalent is that *both* the following conditions hold:

(1). For each $U \in \mathscr{B}$ and each $x \in U$, there is a $U' \in \mathscr{B}'$ with $x \in U' \subset U$.

(2). For each $U' \in \mathscr{B}'$ and each $x \in U'$, there is a $U \in \mathscr{B}$ with $x \in U \subset U'$.

If only condition (1) holds, then $\mathscr{T}(\mathscr{B})$ is a proper subset of $\mathscr{T}(\mathscr{B}')$.

Proof: Assume $\mathscr{T}(\mathscr{B}) = \mathscr{T}(\mathscr{B}')$. Since each $U \in \mathscr{B} \subset \mathscr{T}(\mathscr{B})$ and $\mathscr{T}(\mathscr{B})$ has $\mathscr{B}'$ as basis, (1) follows from **2.2**; similarly, (2) is true.

For the converse, assume (1) is true. Since each $V \in \mathscr{T}(\mathscr{B})$ is a union of sets belonging to $\mathscr{B}$, it follows from **2.3** that $V \in \mathscr{T}(\mathscr{B}')$, showing that $\mathscr{T}(\mathscr{B}) \subset \mathscr{T}(\mathscr{B}')$. If, in addition, (2) is true, we find $\mathscr{T}(\mathscr{B}') \subset \mathscr{T}(\mathscr{B})$, completing the proof.

Ex. 8 According to **I**, Ex. 4, the Euclidean topology in E^2 is that which has for basis all balls $\{B(x; r) \mid x \in E^2, r > 0\}$. On the other hand, $E^2 = E^1 \times E^1$, and according to Ex. 6, the cartesian product topology in E^2 is that which has for basis the boxes $U \times U'$ where U, U' are open sets. *These two bases determine the same topology*, since in each ball containing an $x \in E^2$, one can inscribe a box containing x, and conversely. Similarly, the Euclidean topology in E^n is the same as the cartesian product topology $E^1 \times \cdots \times E^1$ (n-factors).

Ex. 9 The topologies $\mathscr{M}$ and $\mathscr{U}$ on C in Ex. 7 are *not* the same: we have $\mathscr{M} \subset \mathscr{U}$. In fact, given any $\varphi \in M(f, \varepsilon)$, we let $r = \int_0^1 |f - \varphi|$, and find $U(\varphi, \varepsilon - r) \subset M(f, \varepsilon)$. However, consider the constant map $f(x) \equiv 0$: We show that there can be no $M(f, \varepsilon) \subset U(f, 1)$. For, letting $f_n \in C$ be the map $x \to x^n$, we have $\sup_x |f_n(x) - f(x)| = 1$ so no f_n belongs to $U(f, 1)$; on the other hand, $\int_0^1 |f_n - f| = 1/(n+1)$ so that all f_n with sufficiently large n must belong to $M(f, \varepsilon)$. Thus (2) of **3.4** does not hold.

4. Elementary Concepts

Throughout this section, we consider a fixed *space* X, and give some definitions, all of which have familiar meaning when specialized to E^1. There are no genuine theorems. Verification of the examples is left for the reader.

4.1 Definition $A \subset X$ is called *closed* if $\mathscr{C}_X A$ is an open set.

Ex. 1 In E^1, $[a, b]$ is closed according to **4.1**, in conformity with the usual terminology. $N \subset E^1$ is also a closed set.

Ex. 2 The concepts "closed" and "open" are neither exclusive nor exhaustive: In any space X, X and $\varnothing$ are both open and closed, while in E^1, $[0, 1[$ is neither open nor closed. Similarly (*cf.* **1**, Exs. 1 and 2), in $(X, \mathscr{D})$, every set is both open and closed, and in $(X, \mathscr{I})$ points are neither open nor closed.

Ex. 3 In E^2, $\{(x, 0) \mid 0 < x < 1\}$ is neither open nor closed; however, we verify that $\{(x, 0) \mid 0 \leqslant x \leqslant 1\}$ is closed in E^2.

4.2 (a). The intersection of any family of closed sets is a closed set.
(b). The union of *finitely* many closed sets is a closed set.

Proof: These follow by DeMorgan's rules; we prove only (a). To prove $\bigcap_\alpha A_\alpha$ closed, we are to show that $\mathscr{C}\bigcap_\alpha A_\alpha = \bigcup_\alpha \mathscr{C}A_\alpha$ is open; however, since each $\mathscr{C}A_\alpha$ is open, so also is (by **1.1**) the union.

Ex. 4 (b) does not extend to infinite unions. In E^1,

$$\bigcup_1^\infty [0, 1 - 1/n] = [0, 1[.$$

The definition of closed set is not intrinsic, since to decide the question one considers the complement rather than the set. We will obtain some intrinsic formulations by using other concepts.

4.3 Definition Let $A \subset X$. A point $x \in X$ is adherent to A if each nbd of x contains at least one point of A (which may be x itself). The set $\bar{A} = \{x \in X \mid \forall\, U(x)\colon U(x) \cap A \neq \varnothing\}$ of all points in X adherent to A is called the closure of A.

Ex. 5 In E^1, let $A =]0, 1]$, $B = \{1/n \mid n \in Z^+\}$. Then $\bar{A} = [0, 1]$, and $\bar{B} = B \cup \{0\}$.

Ex. 6 In E^2, let $A = \{(x, y) \mid y = \sin 1/x,\ 0 < x \leqslant 1\}$. Then

$$\bar{A} = A \cup \{(0, y) \mid -1 \leqslant y \leqslant 1\},$$

since each ball centered at a point of the latter intersects A.

Ex. 7 In E^1_u (*cf.* **3**, Ex. 4), let $A =]a, b]$. Then $\bar{A} =]a, b]$.

Ex. 8 In Sierpinski space $\mathscr{S}$ (*cf.* **1**, Ex. 3), $\bar{0} = \mathscr{S}$, $\bar{1} = 1$.

4.4 (a). $A \subset \bar{A}$ for every set A.
(b). A is closed if and only if $A = \bar{A}$.

Proof: (a) is immediate from **4.3**.

(b). (A closed) $\Rightarrow$ ($A = \bar{A}$): For, A closed $\Rightarrow$ $\mathscr{C}A$ is open, so each $x \,\bar{\in}\, A$ has a nbd (namely, $\mathscr{C}A$) not meeting A and therefore does not belong to $\bar{A}$. Thus $\bar{A} \subset A$ and with (a), $A = \bar{A}$.

($A = \bar{A}$) $\Rightarrow$ (A closed): For ($A = \bar{A}$) $\Leftrightarrow$ each $x \,\bar{\in}\, A$ has a nbd $U(x)$ not meeting A; by **2.3**, this means that $\mathscr{C}A$ is open.

An alternative characterization of $\overline{A}$ and the algebraic properties of the closure operation are contained in

4.5 $\overline{A}$ is the smallest closed set containing A; that is,

$$\overline{A} = \bigcap \{F \mid (F \text{ closed}) \wedge (F \supset A)\}.$$

Furthermore,

(1). $A \subset B \Rightarrow \overline{A} \subset \overline{B}$.

(2). $\overline{\overline{A}} = \overline{A}$; that is, $\overline{A}$ is closed.

(3). $\overline{A \cup B} = \overline{A} \cup \overline{B}$.

(4). $\overline{\varnothing} = \varnothing$.

Proof: We evidently have $A \subset \bigcap F$.

(a). $x \in \overline{A} \Rightarrow x \in \bigcap F$. Since by **4.2** $\bigcap F$ is closed, $x \notin \bigcap F \Rightarrow x$ has a nbd, $\mathscr{C} \bigcap F$, not meeting $\bigcap F$, and so $x \notin \overline{A}$.

(b). $x \in \bigcap F \Rightarrow x \in \overline{A}$. For $x \notin \overline{A} \Leftrightarrow$ there is a nbd U of x not meeting A; since $\mathscr{C}U$ is closed and contains A, $\mathscr{C}U$ is some set F, and because $x \notin \mathscr{C}U = F$, also $x \notin \bigcap F$.

(1)–(4) follow easily from this characterization:

(1). Any closed set containing B also contains A.

(2). $(\overline{A} = \bigcap F) \Rightarrow \overline{A}$ is closed, by **4.2**, and so **4.4**(b) shows $\overline{\overline{A}} = \overline{A}$.

(3). $\overline{A \cup B}$ is closed, and [**4.4**(a)] contains each of A, B, so that $\overline{A} \cup \overline{B} \subset \overline{A \cup B}$; but, $\overline{A} \cup \overline{B}$ is closed (**4.2**) and contains $A \cup B$, so that also $\overline{A \cup B} \subset \overline{A} \cup \overline{B}$.

(4). $\varnothing$ is a closed set.

Ex. 9 **4.5**(3) holds, by induction, for any finite number of factors; it is not true in general for infinitely many factors, as Ex. 4 shows. However $\bigcup_\alpha \overline{A_\alpha} \subset \overline{\bigcup_\alpha A_\alpha}$ is always true, as follows from $\forall \alpha\colon A_\alpha \subset \bigcup_\alpha A_\alpha \subset \overline{\bigcup_\alpha A_\alpha}$ and application of **4.5**(1), (2).

Ex. 10 The formula $\overline{A \cap B} = \overline{A} \cap \overline{B}$ is not true in general, as $A = [0, 1[$, $B = [1, 2]$ in E^1, shows. As in Ex. 9, it is easy to show that $\overline{\bigcap_\alpha A_\alpha} \subset \bigcap_\alpha \overline{A_\alpha}$ is always true.

Another way to describe closed sets is through cluster points.

4.6 Definition Let $A \subset X$. A point $x \in X$ is called a cluster point of A if each nbd of x contains at least one point of A *distinct* from x. The set $A' = \{x \in X \mid \forall\, U(x)\colon U(x) \cap (A - x) \neq \varnothing\}$ of all cluster points of A is called the *derived* set of A.

Ex. 5 $A' = [0, 1]$; $B' = \{0\}$.
Ex. 6 $A' = \overline{A}$.

Ex. 7 $A' = A$.
Ex. 8 $0' = 1, \quad 1' = \varnothing$.

4.7 $\overline{A} = A \cup A'$. In particular, A is closed if and only if $A' \subset A$; that is, A contains all its cluster points.

Proof: Since $A' \subset \overline{A}$ from **4.3** and **4.6**, and always $A \subset \overline{A}$, we find $A \cup A' \subset \overline{A}$. For the converse inclusion, let $x \in \overline{A}$; if $x \in A$, we are through, and if $x \bar{\in} A$, then each nbd $U(x)$ intersects A at a point necessarily distinct from x, so $x \in A'$. Thus $\overline{A} \subset A \cup A'$, completing the proof. The second part now follows by using **4.4**(b).

Ex. 11 It is evident from **4.6** that $(x \in A') \Leftrightarrow x \in \overline{(A - x)}$. The term *cluster point* arises from the easily proved fact that in E^1, $x \in A'$ if and only if each nbd $U(x)$ contains *infinitely* many points of A.

4.8 Definition Let $A \subset X$. The interior Int (A) of A is the largest open set contained in A; that is, $\text{Int}(A) = \bigcup \{U \mid (U \text{ open}) \wedge (U \subset A)\}$.

Ex. 5 $\text{Int}(A) =]0, 1[$; $\quad \text{Int}(B) = \varnothing$.
Ex. 6 $\text{Int}(A) = \varnothing$.
Ex. 7 $\text{Int}(A) = A$.
Ex. 8 $\text{Int}(0) = 0, \quad \text{Int}(1) = \varnothing$.

4.9 $\text{Int}(A) = \mathscr{C}(\overline{\mathscr{C}A})$ for any set A. In particular, A is open if and only if $A = \text{Int}(A)$.

Proof: Since $E \subset A \Leftrightarrow \mathscr{C}A \subset \mathscr{C}E$, we note that the *open* sets $E \subset A$ are precisely the complements of closed sets $F \supset \mathscr{C}A$. Thus

$$\begin{aligned}\text{Int}(A) &= \bigcup \{\mathscr{C}F \mid (F \text{ is closed}) \wedge (F \supset \mathscr{C}A)\}\\ &= \mathscr{C} \bigcap \{F \mid (F \text{ is closed}) \wedge (F \supset \mathscr{C}A)\}\\ &= \mathscr{C}(\overline{\mathscr{C}A})\end{aligned}$$

by **4.5**. The second part is now trivial.

Ex. 12 It is not generally true that $\text{Int}(A \cup B) = \text{Int}(A) \cup \text{Int}(B)$. Indeed, the union of two sets, each with empty interior, may have an interior, as the example A = rationals, B = irrationals, in E^1 shows.

4.10 Definition Let $A \subset X$. The boundary Fr (A) of A is $\overline{A} \cap \overline{\mathscr{C}A}$.

Ex. 5 $\text{Fr}(A) = \{0, 1\}$; $\text{Fr}(B) = B \cup \{0\}$. Thus the entire set may be contained in its own boundary.
Ex. 6 $\text{Fr}(A) = \overline{A}$.
Ex. 7 $\text{Fr}(A) = \varnothing$.
Ex. 8 $\text{Fr}(0) = 1, \ \text{Fr}(1) = 1$.

The boundary of A is a closed set, and both A, $\mathscr{C}A$ have the same boundary. The boundary can be described as the part of the closure not in the interior:

4.11 Let $A \subset X$. Then:
(1). $\mathrm{Fr}(A) = \bar{A} - \mathrm{Int}(A)$;
(2). $\mathrm{Fr}(A) \cap \mathrm{Int}(A) = \varnothing$;
(3). $\bar{A} = \mathrm{Int}(A) \cup \mathrm{Fr}(A)$;
(4). $X = \mathrm{Int}(A) \cup \mathrm{Fr}(A) \cup \mathrm{Int}(\mathscr{C}A)$ is a pairwise disjoint union.

Proof: (1). $\mathrm{Fr}\,(A) = \bar{A} \cap \mathscr{C}[\mathscr{C}\overline{\mathscr{C}A}] = \bar{A} - \mathrm{Int}\,(A)$. The proofs of the remaining assertions are entirely similar to this, and are left for the reader.

4.12 Definition $D \subset X$ is dense in X if $\bar{D} = X$.

Clearly, X is dense in X, and in fact X is the only *closed* set dense in X.

Ex. 13 The rationals are dense in E^1 and in E^1_u; more generally, the points of E^n with all coordinates rational are dense in E^n.
Ex. 8 0 is dense in $\mathscr{S}$.

4.13 The following four statements are equivalent:
(1). D is dense in X.
(2). If F is any closed set in X and $D \subset F$, then $F = X$.
(3). Each nonempty basic open set in X contains an element of D.
(4). The complement of D has empty interior.

Proof: (1) $\Rightarrow$ (2). For $D \subset F \Rightarrow X = \bar{D} \subset \bar{F} = F$.
(2) $\Rightarrow$ (3). Let U be open, nonempty, with $U \cap D = \varnothing$; then $D \subset \mathscr{C}U \neq X$, which contradicts (2), since $\mathscr{C}U$ is closed.
(3) $\Rightarrow$ (4). Assume $\mathrm{Int}\,(\mathscr{C}D) \neq \varnothing$; since $\mathrm{Int}\,(\mathscr{C}D)$ is open, there is (**2.3**) a nonempty basic $U \subset \mathrm{Int}\,(\mathscr{C}D)$, and since $\mathrm{Int}\,(\mathscr{C}D) \subset \mathscr{C}D$, U contains no points of D.
(4) $\Rightarrow$ (1). $\mathrm{Int}\,(\mathscr{C}D) = \mathscr{C}[\overline{\mathscr{C}\mathscr{C}D}] = \mathscr{C}\bar{D} = \varnothing$, so that $\bar{D} = X$.

5. Topologizing with Preassigned Elementary Operations

Each of the concepts discussed in **4** (except density) can be used as the primitive concept for introducing a topology in a set: If in a set X one preassigns what (say) the closure of each set is to be, then provided the assignment $A \to \bar{A}$ is not "unreasonable," there is indeed a topology for X in which the closure of each set is exactly the predetermined set.

5.1 Theorem Let X be a set, and $u: \mathscr{P}(X) \to \mathscr{P}(X)$ a map with the properties:

(1). $u(\varnothing) = \varnothing$.
(2). $A \subset u(A)$ for each A.
(3). $u \circ u(A) = u(A)$ for each A.
(4). $u(A \cup B) = u(A) \cup u(B)$ for each A, B.

Then the family $\mathscr{T} = \{\mathscr{C}u(A) \mid A \in \mathscr{P}(X)\}$ is a topology, and with $\mathscr{T}$, $\overline{A} = u(A)$ for each A.

Proof: We first establish the consequence

(i). $$(A \subset B) \Rightarrow (u(A) \subset u(B)).$$

For, $(A \subset B) \Leftrightarrow (B = A \cup B)$ which, with postulate (4), gives $u(B) = u(A) \cup u(B)$ and proves (i).

We now verify that $\mathscr{T}$ is a topology:

(a). $X \in \mathscr{T}$, since $\mathscr{C}u(\varnothing) = X$, and $\varnothing \in \mathscr{T}$ because [by (2)] $X \subset u(X)$ gives $u(X) = X$.

(b). $\mathscr{C}u(A) \cap \mathscr{C}u(B) = \mathscr{C}[u(A) \cup u(B)] = \mathscr{C}u(A \cup B)$; by induction it follows that the intersection of any finite family of sets of $\mathscr{T}$ is also a member of $\mathscr{T}$.

(c). Let $S = \bigcup_\alpha \mathscr{C}u(A_\alpha)$; we are to show that $S = \mathscr{C}u(Q)$ for some $Q \subset X$. Because $S = \mathscr{C} \bigcap_\alpha u(A_\alpha)$, we find $\mathscr{C}S = \bigcap_\alpha u(A_\alpha) \subset u(A_\alpha)$ for each α so, by (i) and (3), $u(\mathscr{C}S) \subset u(A_\alpha)$ for each α, and therefore $u(\mathscr{C}S) \subset \bigcap_\alpha u(A_\alpha) = \mathscr{C}S$. Use of (2) shows $u(\mathscr{C}S) = \mathscr{C}S$, and so, $S = \mathscr{C}u(\mathscr{C}S)$, as required.

Using the topology $\mathscr{T}$, we now show that $\overline{A} = u(A)$. First, $\overline{A} \subset u(A)$: For, since each $u(A)$ is closed in $\mathscr{T}$, from $A \subset u(A)$ we find that $\overline{A} \subset \overline{u(A)} = u(A)$. Finally, $u(A) \subset \overline{A}$: Since $\mathscr{C}\overline{A}$ is open in $\mathscr{T}$, we have $\overline{A} = u(B)$ for some B; using (i) and $A \subset \overline{A}$, we find that

$$u(A) \subset u(\overline{A}) = u \circ u(B) = u(B) = \overline{A}.$$

The proofs of the following propositions are left for the reader; each one indicates the Boolean operations that completely characterize the various concepts.

5.2 Let X be a set, and $\gamma: \mathscr{P}(X) \to \mathscr{P}(X)$ a map such that:

(1). $\gamma(\varnothing) = \varnothing$.
(2). $\gamma \circ \gamma(A) \subset A \cup \gamma(A)$.
(3). $\gamma(A \cup B) = \gamma(A) \cup \gamma(B)$.
(4). For each $x \in X$, $x \mathrel{\bar{\in}} \gamma(x)$.

Then $\mathscr{T} = \{\mathscr{C}(A \cup \gamma(A)) \mid A \in \mathscr{P}(X)\}$ is a topology, and $A' = \gamma(A)$.

5.3 Let X be a set and $\eta: \mathscr{P}(X) \to \mathscr{P}(X)$ a map such that:

(1). $\eta(X) = X.$
(2). $\eta(A) \subset A.$
(3). $\eta \circ \eta(A) = \eta(A).$
(4). $\eta(A \cap B) = \eta(A) \cap \eta(B).$

Then $\mathscr{T} = \{\eta(A) \mid A \in \mathscr{P}(X)\}$ is a topology, and Int $(A) = \eta(A)$.

5.4 Let X be a set, and $\beta: \mathscr{P}(X) \to \mathscr{P}(X)$ a map such that:

(1). $\beta(\varnothing) = \varnothing.$
(2). $\beta(A) = \beta(\mathscr{C}A).$
(3). $\beta \circ \beta(A) \subset \beta(A).$
(4). $A \cap B \cap \beta(A \cap B) = A \cap B \cap [\beta(A) \cup \beta(B)].$

Then $\mathscr{T} = \{\mathscr{C}(A \cup \beta(A)) \mid A \in \mathscr{P}(X)\}$ is a topology, and $\text{Fr}(A) = \beta(A)$.

5.5 Let X be a set, and $\mathscr{A} \subset \mathscr{P}(X)$ a family with the properties:

(1). $\varnothing$, X belong to $\mathscr{A}$.
(2). The intersection of any family of members of $\mathscr{A}$ is also a member of $\mathscr{A}$.
(3). The finite union of members of $\mathscr{A}$ belongs to $\mathscr{A}$.

Then $\mathscr{T} = \{\mathscr{C}A \mid A \in \mathscr{A}\}$ is a topology, in which $\mathscr{A}$ is the complete family of closed sets.

6. G_δ, F_σ, and Borel Sets

Though the countable union of closed sets need not be closed, and the countable intersection of open sets need not be open, such sets occur frequently in analysis.

6.1 Definition A set F is called an F_σ-set (or an F_σ) if it is the union of at most countably many closed sets. A set G is called a G_δ if it is the intersection of at most countably many open sets.

Ex. 1 Each closed set is an F_σ and each open set is a G_δ.

Ex. 2 The concepts F_σ and G_δ are neither exclusive nor exhaustive. In E^1, the closed interval $[a, b]$ is evidently an F_σ and it is also a G_δ because

$$[a, b] = \bigcap_1^\infty \,]a - 1/n, b + 1/n[.$$

In $(X, \mathscr{I})$ (*cf.* **1**, Ex. 1), any proper subset of X is neither an F_σ nor a G_δ.

Ex. 3 In E^1, the set Q of rationals is an F_σ, since Q is countable and each point is a closed set. We shall see much later (XI, **10**, Ex. 2) that Q is not a G_δ.

Ex. 4 In the ordinal space $[0, \Omega]$, the point Ω is an F_σ, being a closed set. However, it is not a G_δ. For, if $\{G_i \mid i \in N\}$ is any countable collection of open sets containing Ω, then because the sets $]\alpha, \beta]$ are a basis, $\forall\, i\ \exists \alpha_i < \Omega$: $]\alpha_i, \Omega] \subset G_i$. Being countable, the collection $\{\alpha_i \mid i \in N\}$ has an upper bound $\beta < \Omega$, so

$$\bigcap_0^\infty G_i \supset]\beta, \Omega] \neq \{\Omega\}$$

It is frequently convenient to express F_σ and G_δ sets in a standard manner:

6.2 (1). If F is an F_σ-set, then there is a nondecreasing sequence $F_1 \subset F_2 \subset \cdots$ of closed sets with $F = \bigcup_1^\infty F_i$.
(2). If G is a G_δ-set, then there is a nonincreasing sequence $G_1 \supset G_2 \supset \cdots$ of open sets with $G = \bigcap_1^\infty G_i$.

Proof: We prove only (2); (1) is analogous. Since G is a G_δ, $G = \bigcap_1^\infty W_i$, where the W_i are open. Define $G_n = W_1 \cap \cdots \cap W_n$; then each G_n is open, $G_n \supset G_{n+1}$ for each n, and $\bigcap_1^\infty W_i = \bigcap_1^\infty G_i$.

For the set operations,

6.3 (a). The countable union and finite intersection of F_σ-sets is an F_σ.
(b). The countable intersection and finite union of G_δ-sets is a G_δ.
(c). The complement of an F_σ is a G_δ, and conversely.

Proof: Ad (a). Since

$$\bigcup_{i=0}^\infty \left[\bigcup_{j=0}^\infty F_{i,j}\right] = \bigcup \{F_{i,j} \mid (i,j) \in N \times N\}$$

and $N \times N$ is countable, this is an F_σ. Similarly,

$$\bigcap_{i=0}^n \left[\bigcup_{j=0}^\infty F_{i,j}\right] = \bigcup \{F_{0,j_0} \cap \cdots \cap F_{n,j_n} \mid (j_0, \cdots, j_n) \in N \times \cdots \times N\}$$

by distributivity; since each $F_{0,j_0} \cap \cdots \cap F_{n,j_n}$ is closed, and $N \times \cdots \times N$ is countable, this also is an F_σ. (b) is proved analogously, and (c) involves only DeMorgan's rules.

As **6.3** indicates, the properties F_σ, G_δ are generally not preserved under *all* countable set operations. We now define a family which is in fact preserved under these operations.

6.4 Definition A nonempty family $\Sigma \subset \mathscr{P}(X)$ is called a σ-ring if

(1). $A \in \Sigma \Rightarrow \mathscr{C}A \in \Sigma$,

(2). $A_i \in \Sigma$ for $i = 1, 2, \cdots \Rightarrow \bigcup_1^\infty A_i \in \Sigma$.

6.5 There always exists a unique smallest σ-ring $\mathscr{B}$ containing the topology $\mathscr{T}$ of X. $\mathscr{B}$ is called the family of Borel sets in X, and $\aleph(\mathscr{B}) \leqslant \aleph(\mathscr{T})^{\aleph_0}$. Furthermore:

(1). The countable union, countable intersection, and the difference of Borel sets is a Borel set.

(2). Each F_σ and each G_δ is a Borel set.

Proof: Observing that $\mathscr{P}(X)$ is a σ-ring containing $\mathscr{T}$, and that the intersection of any family of σ-rings is also a σ-ring, we define $\mathscr{B}$ to be the intersection of all σ-rings containing $\mathscr{T}$. Since only two operations are involved, the estimate of $\aleph(\mathscr{B})$ follows from II, **9.4**. To establish (1), we need only verify preservation under intersection, and this follows from

$$\bigcap_1^\infty B_i = \mathscr{C} \bigcup_1^\infty \mathscr{C}B_i,$$

where $B_i \in \mathscr{B}$, $i \in Z^+$. (2) is trivial.

6.6 Remark For the special case of E^n, we can establish more:

(1). There are sets that are not Borel sets.

(2). Each open set (and each closed set) is both an F_σ and a G_δ.

(3). The family of Borel sets can be described (*cf.* II, **9**, Ex. 1) as the smallest family $\mathscr{B}$ containing $\mathscr{T}$ and satisfying

$$B_i \in \mathscr{B}, i \in Z^+ \Rightarrow \bigcup_1^\infty B_i \in \mathscr{B} \quad \text{and} \quad \bigcap_1^\infty B_i \in \mathscr{B}$$

Ad (1). By **2**, Ex. 4, $\aleph(\mathscr{T}) = 2^{\aleph_0}$, hence $\aleph(\mathscr{B}) \leqslant 2^{\aleph_0}$; since $\aleph(\mathscr{P}(E^n)) = 2^{\mathfrak{c}}$, this proves (1).

Ad (2). Each ball $B(x; r)$ is an F_σ, since $B(x; r) = \bigcup_1^\infty \overline{B(x; r - n^{-1})}$.

By **2**, Ex. 4, each open set is a union of at most countably many balls (that is, an F_σ-set), so **6.3**(a) shows that each open set is an F_σ.

Ad (3). We first show that $\mathscr{B}$ as defined here is a σ-ring. Let $\mathscr{F}^\circ = \{\mathscr{C}B \mid B \in \mathscr{B}\}$. Then $\mathscr{T} \subset \mathscr{F}^\circ$, since because of (2),

$$G \in \mathscr{T} \Rightarrow G \text{ is an } F_\sigma \Rightarrow G = \bigcup_1^\infty F_i = \bigcup_1^\infty \mathscr{C}G_i = \mathscr{C} \bigcap_1^\infty G_i$$

and we have $\bigcap_1^\infty G_i \in \mathscr{B}$. Noting that $\mathscr{F}^\circ$ also satisfies the condition on $\mathscr{B}$ (De Morgan's rules), it follows that $\mathscr{B} \subset \mathscr{F}^\circ$. This says that $A \in \mathscr{B} \Rightarrow A \in \mathscr{F}^\circ \Rightarrow A = \mathscr{C}B$ for some $B \in \mathscr{B} \Rightarrow \mathscr{C}A \in \mathscr{B}$, showing that $\mathscr{B}$ is a σ-ring. Since, conversely, any σ-ring containing $\mathscr{T}$ satisfies the conditions required of $\mathscr{B}$, the descriptions here and in **6.4** coincide.

7. Relativization

Let $Y \subset X$. If X carries a topology, we will define a topology for the *set* Y, called the *relative* (or induced) topology on Y. Its importance lies in this: To determine what any concept defined for topological spaces becomes when discussion is restricted to $Y \subset X$, we simply regard Y as a space with the *induced* topology and carry over the discussion *verbatim*.

7.1 Definition Let $(X, \mathscr{T})$ be a topological space, and $Y \subset X$. The induced topology $\mathscr{T}_Y$ on Y is $\{Y \cap U \mid U \in \mathscr{T}\}$. $(Y, \mathscr{T}_Y)$ is called a sub*space* of $(X, \mathscr{T})$.

To verify $\mathscr{T}_Y$ is actually a topology on Y is trivial.

Ex. 1 In the subspace $Y = [0, 1[\cup \{2\}$ of E^1, the following sets are open: (a) $\{2\}$; (b) all open intervals contained in $[0, 1[$; (c) all intervals of form $[0, a[$ (these are nbds of 0 in Y). In Y, the subset $[0, 1[$ is both open and closed. Note that sets open (or closed) in Y need not be open (or closed) in the containing space E^1.

Ex. 2 To the set R of all real numbers, adjoin two points, $\{+\infty\}$, $\{-\infty\}$, and let $\mathscr{T}$ be the topology generated by all sets of form $\{+\infty\} \cup \{x \mid x > a\}$, $\{-\infty\} \cup \{x \mid x < a\}$. $(R \cup \{-\infty, +\infty\}, \mathscr{T})$ is called the *extended real line*, denoted by $\tilde{E}^1$. With the induced topology, $R \subset \tilde{E}^1$ is the Euclidean space E^1.

Note: (a). By defining $-\infty \leqslant x \leqslant +\infty$, $\tilde{E}^1$ is a totally ordered set. (b). For any set $A \subset \tilde{E}^1$, sup A (= least upper bound of A) and $\inf A = -\sup\{-x \mid x \in A\}$ always exist (and may be $\pm\infty$). (c). The sets $\{+\infty\} \cup \{x \mid x > a\}$ are nbds of $+\infty$.

Let Y be a subspace of X, and $A \subset Y$. Since $(Y, \mathscr{T}_Y)$ is a space, we can form the closure of A, using $\mathscr{T}_Y$, to obtain $\overline{A}_Y$; but also $A \subset X$, so we can form $\overline{A}$, using $\mathscr{T}$. We now determine the relation between $\overline{A}$ and $\overline{A}_Y$, as well as that for the other operations in **4**.

7.2 Theorem Let $(X, \mathscr{T})$ be a space and $(Y, \mathscr{T}_Y)$ a subspace. Then:

(1). If $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ is a basis (subbasis) for $\mathscr{T}$, $\{Y \cap U_\alpha \mid \alpha \in \mathscr{A}\}$ is a basis (subbasis) for $\mathscr{T}_Y$.

(2). Let $A \subset Y$. Then A is $\mathscr{T}_Y$-closed if and only if $A = Y \cap F$, where F is $\mathscr{T}$-closed (that is, the closed sets in Y are the intersections of Y with sets closed in X).

(3). $\overline{A}_Y = Y \cap \overline{A}$; $A'_Y = Y \cap A'$; $Y \cap \text{Int}(A) \subset \text{Int}_Y(A)$; $\text{Fr}_Y(A) \subset Y \cap \text{Fr}(A)$.

Proof: (1) is trivial.

(2). Let A be closed in Y; then $A = Y - W$, where W is open in Y, and since $W = Y \cap V$ where $V \in \mathscr{T}$, we find $A = Y - Y \cap V = Y \cap \mathscr{C}V$. Conversely, if $A = Y \cap F$, F closed in X, then $Y - A = Y \cap \mathscr{C}F$, showing that A is closed in Y.

(3). $y \in \overline{A} \cap Y \Rightarrow \forall\, U(y)\colon U(y) \cap A \neq \varnothing$, and since $A \subset Y$, it follows that $\forall\, U(y)\colon (Y \cap U(y)) \cap A \neq \varnothing$, which shows that $y \in \overline{A}_Y$; the implications all reverse. The second statement is proved similarly, and the remaining inclusions are trivial.

Ex. 3 In Euclidean 2-space, $E^2 = E^1 \times E^1$, the *set* E^1 can be identified with the subset $E^1 \times 0 \subset E^2$. Since the topology of E^2 has as basis the open boxes $J \times J'$, we find from **7.2**(1) that the *relative* topology in $E^1 \times 0$ is precisely *the Euclidean topology* of E^1. This generalizes easily: Writing $E^n = E^s \times E^t$, $s + t = n$, the relative topology on $E^s \times 0 \subset E^n$ is the Euclidean topology of $E^s \times 0$.

We have seen in Ex. 1 that sets open in a subspace need not be open in the entire space; the following theorem gives a simple but useful case where this cannot occur.

7.3 Theorem Let Y be a subspace of X. If $A \subset Y$ is closed (open) in Y, and Y is closed (open) in X, then A is closed (open) in X.

Proof: For $A = Y \cap K$, and since Y and K are each closed (open) in X, so also is the intersection.

Ex. 4 A subspace $Y \subset X$ is called a *discrete subspace* of X whenever $\mathscr{T}_Y$ is the discrete topology. Observe that if Y is a discrete subspace of X, then Y need not be open or closed in X, as $Y = \{1/n \mid n \in Z^+\} \subset E^1$ shows.

7.4 (Transitivity) A subspace of a subspace is a subspace of the entire space.

Proof: Let $Z \subset Y \subset X$, and let $\mathscr{T}_{YZ}$ be the topology of Z as a subspace of Y; we are to show that $\mathscr{T}_{YZ} = \mathscr{T}_Z$. Let $W \in \mathscr{T}_{YZ}$ so that $W = Z \cap V$, where $V \in \mathscr{T}_Y$. Since $V = Y \cap U$, $U \in \mathscr{T}$, we find $W = Z \cap Y \cap U = Z \cap U$, showing that $W \in \mathscr{T}_Z$. The converse inclusion is trivial.

8. Continuous Maps

We have been considering topologies on one given set; we now want to relate different topological spaces. Given $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$, note that a map $f\colon X \to Y$ relates the sets and also induces two maps $f\colon \mathscr{P}(X) \to \mathscr{P}(Y)$, $f^{-1}\colon \mathscr{P}(Y) \to \mathscr{P}(X)$. Of these, f^{-1} should be used to relate the topologies, since it is the only one that preserves the Boolean operations involved in the definition of a topology. Thus the suitable maps $f\colon X \to Y$ are those for which simultaneously $f^{-1}\colon \mathscr{T}_Y \to \mathscr{T}_X$. Formally stated,

8.1 Definition Let $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ be spaces. A map $f\colon X \to Y$ is called *continuous* if the inverse image of each set open in Y is open in X [that is, if f^{-1} maps $\mathscr{T}_Y$ into $\mathscr{T}_X$].

Ex. 1 A constant map $f\colon X \to Y$ is always continuous: The inverse image of any set U open in Y is either $\varnothing$ or X, which are open.

Ex. 2 Let X be any set, $\mathscr{T}_1$, $\mathscr{T}_2$, two topologies on X. The *bijective* map $1\colon (X, \mathscr{T}_1) \to (X, \mathscr{T}_2)$ is continuous if and only if $\mathscr{T}_2 \subset \mathscr{T}_1$. Note that a continuous map need not send open sets to open sets, and also that increasing the topology $\mathscr{T}_1$ preserves continuity.

Ex. 3 A map sending open sets to open sets is called an *open map*. An open map need not be continuous. $1\colon (X, \mathscr{T}_1) \to (X, \mathscr{T}_2)$ is open if and only if $\mathscr{T}_1 \subset \mathscr{T}_2$, but it is not continuous whenever $\mathscr{T}_1 \neq \mathscr{T}_2$.

Ex. 4 Let $Y \subset X$. The relative topology $\mathscr{T}_Y$ can be characterized as the *smallest* topology on Y for which the inclusion map $i\colon Y \to X$ is continuous. For, if $U \in \mathscr{T}$, the continuity of i requires $i^{-1}(U) = U \cap Y$ to be open in Y, so that any topology for which i is continuous must contain $\mathscr{T}_Y$.

The elementary properties are

8.2 (1). (Composition.) If $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous, so also is $g \circ f\colon X \to Z$.

(2). (Restriction of domain.) If $f\colon X \to Y$ is continuous and $A \subset X$ is taken with the subspace topology, then $f \mid A\colon A \to Y$ is continuous.

(3). (Restriction of range.) If $f\colon X \to Y$ is continuous and $f(X)$ is taken with the subspace topology, then $f\colon X \to f(X)$ is continuous.

Proof: (a). We have $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$; since $g^{-1}\colon \mathscr{T}_Z \to \mathscr{T}_Y$ and $f^{-1}\colon \mathscr{T}_Y \to \mathscr{T}_X$, the continuity of $g \circ f$ follows.

(b). Note $f \mid A = f \circ i$, where $i\colon A \to X$; thus, apply Ex. 4 and (a).

(c). $f^{-1}(U \cap f(X)) = f^{-1}(U) \cap f^{-1}f(X) = f^{-1}(U)$, and the statement is proved.

The basic theorem on continuity is:

8.3 Theorem Let X, Y be topological spaces, and $f\colon X \to Y$ a map. The following statements are equivalent:

(1). f is continuous.

(2). The inverse image of each closed set in Y is closed in X.

(3). The inverse image of each member of a subbasis (basis) for Y is open in X (not necessarily a member of a subbasis, or basis for X!).
(4). For each $x \in X$ and each nbd $W(f(x))$ in Y, there exists a nbd $V(x)$ in X such that $f(V(x)) \subset W(f(x))$.
(5). $f(\overline{A}) \subset \overline{f(A)}$ for every $A \subset X$.
(6). $\overline{f^{-1}(B)} \subset f^{-1}(\overline{B})$ for every $B \subset Y$.

Proof: (1) $\Leftrightarrow$ (2), since $f^{-1}(Y - E) = X - f^{-1}(E)$ for any $E \subset X$.
(1) $\Leftrightarrow$ (3). Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be a subbasis for Y. If f is continuous, each $f^{-1}(U_\alpha)$ is open. Conversely, if each $f^{-1}(U_\alpha)$ is open, then because any open $U \subset Y$ can be written

$$U = \bigcup \{U_{\alpha_1} \cap \cdots \cap U_{\alpha_n} \mid \{\alpha_1, \cdots, \alpha_n\} \subset \mathscr{A}\},$$

we have that

$$f^{-1}(U) = \bigcup \{f^{-1}(U_{\alpha_1}) \cap \cdots \cap f^{-1}(U_{\alpha_n})\}$$

is a union of open sets and so is open.

(1) $\Rightarrow$ (4). Since $f^{-1}(W(x))$ is open, we can use it for $V(x)$.
(4) $\Rightarrow$ (5). Let $A \subset X$ and $b \in \overline{A}$; we show $f(b) \in \overline{f(A)}$ by proving each $W(f(b))$ intersects $f(A)$. For, finding $V(b)$ with $f(V(b)) \subset W(f(b))$,

$$\begin{aligned} b \in \overline{A} &\Rightarrow \varnothing \neq V(b) \cap A \\ &\Rightarrow \varnothing \neq f(V(b) \cap A) \subset f(V(b)) \cap f(A) \subset W(f(b)) \cap f(A). \end{aligned}$$

(5) $\Rightarrow$ (6). Let $A = f^{-1}(B)$; then $f(\overline{A}) \subset \overline{f(A)} = \overline{f[f^{-1}(B)]} = \overline{B \cap f(X)} \subset \overline{B}$, so that $\overline{A} \subset f^{-1}(\overline{B})$, as required.
(6) $\Rightarrow$ (2). Let $B \subset Y$ be closed; then $\overline{f^{-1}(B)} \subset f^{-1}(B)$, and since always $f^{-1}(B) \subset \overline{f^{-1}(B)}$, this shows [**4.4**(b)] that $f^{-1}(B)$ is closed.

The formulation (4) of **8.3** shows that continuity is a "local" matter, a fact having many applications. Precisely,

8.4 Definition An $f: X \to Y$ is *continuous at* $x_0 \in X$ if **8.3**(4) is satisfied at x_0.

From this viewpoint, the equivalence of (1) and (4) in **8.3** asserts: f is continuous according to **8.1**, if and only if it is continuous *at each point of* X.

Ex. 5 Let $f: E^n \to E^1$; that is, a real-valued function of n real variables. The continuity of f at $x_0 \in E^n$ is simply the usual notion encountered in analysis. For, by **8.4**, f is continuous at x_0 if for each open interval $]f(x_0) - \varepsilon, f(x_0) + \varepsilon[$, there is a ball $B(x_0; \delta)$ mapped into it by f; that is,

$$\forall \varepsilon > 0 \; \exists \delta > 0 : |x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \varepsilon.$$

Ex. 6 Let X be any space, and for each pair of maps $f, g: X \to Y$, define $\varphi: X \to Y \times Y$ by $x \to (f(x), g(x))$. Then φ is continuous if and only if both f and g are continuous. For, note that $\varphi^{-1}(U \times V) = f^{-1}(U) \cap g^{-1}(V)$; since a subbasis for $Y \times Y$ is all sets of form $U \times Y$, $Y \times V$, where U, V belong to a subbasis for Y, **8.3**(2) gives the result. Thus, for example, the map $E^1 \to E^2$ given by $x \to (\cos x, \sin x)$ is continuous.

Ex. 7 (Vickery). Let Ω be the first uncountable ordinal number, and let $[0, \Omega[$ be the subspace of the ordinal space $[0, \Omega]$ (cf. **3**, Ex. 5). Then each continuous $\varphi: [0, \Omega[\to E^1$ must be constant on a tail $[\beta, \Omega[$.

We first assert that $\forall n \in Z^+ \; \exists \alpha_n < \Omega \; \forall \xi > \alpha_n: |\varphi(\xi) - \varphi(\alpha_n)| < 1/n$. For, if this were not true, then $\exists n_0 \; \forall \alpha < \Omega \; \exists \xi > \alpha: |\varphi(\xi) - \varphi(\alpha)| \geqslant 1/n_0$. We could then use induction to construct a sequence $\{\xi_i \mid i \in Z^+\}$ such that both $\xi_i < \xi_{i+1}$ and $|\varphi(\xi_i) - \varphi(\xi_{i+1})| \geqslant 1/n_0$ for each i: choosing $\xi_1 = 0$ and assuming $\xi_1, \cdots, \xi_k$ defined, let ξ_{k+1} be the first element in $[\xi_k, \Omega[$ satisfying the hypothesis. The $\{\xi_i\}$ thus found would then have a least upper bound $\gamma < \Omega$ and then φ would not be continuous at γ: any basic nbd $]\eta, \gamma]$ contains some ξ_i and therefore all ξ_k, $k > i$, so cannot have image contained in the nbd $]\varphi(\gamma) - 1/3n_0, \varphi(\gamma) + 1/3n_0[$ of $\varphi(\gamma)$. Our assertion is therefore established. Now let β be an upper bound of the $\{\alpha_n\}$; then $\beta < \Omega$ and φ is constant on $[\beta, \Omega[$: if $\zeta \in [\beta, \Omega[$, then we have both $|\varphi(\zeta) - \varphi(\alpha_n)| < 1/n$ and $|\varphi(\beta) - \varphi(\alpha_n)| < 1/n$ for every n, so $|\varphi(\zeta) - \varphi(\beta)| < 2/n$ for all n, and therefore $\varphi(\zeta) = \varphi(\beta)$.

9. Piecewise Definition of Maps

In analysis, continuous functions are frequently defined piecewise: For example, a continuous function on $[0, n] \subset E^1$ may be constructed by using different formulas on each $[i, i + 1]$, adjacent functions agreeing on the common end point. We give here the general formulation of this process.

9.1 Definition A family $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ of sets in a space X is called nbd-finite if each point of X has a nbd V such that $V \cap A_\alpha \neq \varnothing$ for at most finitely many indices α.

This concept is related only to the "position" of the A_α in X, and is completely unrelated to the intersections that occur among the A_α themselves.

Ex. 1 $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ may be nbd-finite, even though each A_α intersects infinitely many other A_β: in E^1, take the family $\{A_n \mid n \in N\}$ with $A_n = \{x \mid x > n\}$. The family $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ may not be nbd-finite, even though no A_α intersects any other A_β: in E^1, take the family $\{A_x \mid x \in E^1\}$, where $A_x = \{x\}$.

The main property of nbd-finite families is

9.2 Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite family in X. Then:

(1). $\{\overline{A}_\alpha \mid \alpha \in \mathscr{A}\}$ is also nbd-finite.
(2). For each $\mathscr{B} \subset \mathscr{A}$, $\bigcup \{\overline{A}_\beta \mid \beta \in \mathscr{B}\}$ is *closed* in X.

Proof: Ad (1). Given x, there is a nbd $U(x)$ such that $A_\alpha \cap U(x) = \varnothing$ for all but at most finitely many α; since

$$A_\alpha \cap U(x) = \varnothing \Rightarrow A_\alpha \subset \mathscr{C}U(x) \Rightarrow \overline{A}_\alpha \subset \mathscr{C}U(x) \Rightarrow \overline{A}_\alpha \cap U(x) = \varnothing,$$

this proves (1).

Ad (2). Let $B = \bigcup \overline{A}_\beta$. For each $x \bar{\in} B$, there is, by (1), a nbd U meeting at most finitely many $\overline{A}_\beta$ say, $\overline{A}_{\beta_1}, \cdots, \overline{A}_{\beta_n}$; then $U \cap \bigcap_1^n \mathscr{C}\overline{A}_{\beta_i}$ is a nbd of x not meeting B so, by **2.3**, $\mathscr{C}B$ is open.

The following simple result on coverings is very important; we shall derive several of its consequences as we proceed.

9.3 Theorem Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of sets that cover the space X; that is, $X = \bigcup_\alpha A_\alpha$. Assume that either:

(1). All the A_α are open,
or (2). All the A_α are closed, and form a nbd-finite family.

Then $B \subset X$ is open (resp. closed) if and only if each $B \cap A_\alpha$ is open (resp. closed) in the subspace A_α.

Proof: The necessity of the condition is clear, from definition of the relative topology. For the sufficiency:

Case (1): Assume that each $B \cap A_\alpha$ is open in the open A_α; **7.4** shows $B \cap A_\alpha$ open in X, so that $B = B \cap X = B \cap \bigcup_\alpha A_\alpha = \bigcup_\alpha B \cap A_\alpha$ is also open in X. If each $B \cap A_\alpha$ is closed in the open A_α, then

$$A_\alpha - (B \cap A_\alpha) = A_\alpha \cap \mathscr{C}B$$

is open in A_α, and we find that $\mathscr{C}B$ is open in X.

Case (2): Assume that each $B \cap A_\alpha$ is closed in the closed A_α; **7.4** shows $B \cap A_\alpha$ closed in X. Since $\{A_\alpha\}$ is nbd-finite, $\{B \cap A_\alpha\}$ is also, so by **9.2,** $B = \cup \{B \cap A_\alpha\}$ is closed in X. The remaining case, where each $B \cap A_\alpha$ is open in the closed A_α, is treated as above.

The general formulation of piecewise definition of maps is

9.4 Theorem Let X be a space, and $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ a covering of X such that either:

(1). The sets A_α are all open,

or (2). The sets A_α are all closed, and form a nbd-finite family.

For each $\alpha \in \mathscr{A}$, let $f_\alpha: A_\alpha \to Y$ be continuous and assume that $f_\alpha \mid A_\alpha \cap A_\beta = f_\beta \mid A_\alpha \cap A_\beta$ for each $(\alpha, \beta) \in \mathscr{A} \times \mathscr{A}$. Then there exists a unique *continuous* map $f: X \to Y$, which is an extension of each f_α; that is, $\forall\, \alpha: f \mid A_\alpha = f_\alpha$.

Proof: The existence of f comes from I, **6.7**. To show continuity, let $U \subset Y$ be open; then $f^{-1}(U) \cap A_\alpha = f_\alpha^{-1}(U)$ and so is open in A_α for each α. According to **9.3**, this means that $f^{-1}(U)$ is open in X, as required.

9.5 Remark In the case that the A_α are closed, it is evident that some restriction on their position is required: In E^1, $[a, b] = \cup \{x \mid x \in [a, b]\}$, and if the closed sets $\{x\}$ are used, it is clear that **9.4** cannot be true.

9.6 Remark Theorems **9.3** and **9.4** have wide applicability. In many cases, it is possible to find a nbd-finite covering of a space X satisfying **9.3**, by sets $\{A_\alpha\}$ that have "nice properties"; in such cases, the topology of X can be expressed by using the sets $\{A_\alpha\}$ alone, and the continuity of an $f: X \to Y$ follows from its continuity on each A_α (see **VI** 8).

10. Continuous Maps into E^1

Let X be an arbitrary space, and $f, g: X \to E^1$. Because E^1 has an algebraic structure, maps can be combined by performing algebraic operations on their values at each point. Thus we can form

$$x \to f(x) + g(x),\quad x \to f(x)\cdot g(x),\quad x \to c\cdot f(x) \ (c \text{ a real constant});$$

these are denoted by $f + g$, $f \cdot g$, $c \cdot f$, respectively, to indicate the operations performed. In this section, we consider the continuity of these maps.

10.1 $f: X \to E^1$ is continuous if and only if for each real b, *both* the sets $\{x \mid f(x) > b\}$ and $\{x \mid f(x) < b\}$ are open.

Proof: The sets involved are the inverse images of, respectively, the subbasic open sets $\{y \mid y > b\}$ and $\{y \mid y < b\}$ for the Euclidean topology of E^1, so **8.3**(3) applies.

Ex. 1 The requirement that *both* types of sets be open cannot be relaxed: If $A =]0, 1[\subset E^1$ and $c_A: E^1 \to E^1$ is its characteristic function, c_A is not continuous [$\{x \mid c_A(x) < 1\}$ is not open], yet all sets of type $\{x \mid c_A(x) > b\}$ are open.

Proposition **10.1** indicates that a splitting of the notion "continuity" for maps $X \to E^1$ is appropriate.

10.2 Definition An $f: X \to E^1$ is *upper* semicontinuous if for each real b, $\{x \mid f(x) < b\}$ is open; it is *lower* semicontinuous if for each real b, $\{x \mid f(x) > b\}$ is open.

Clearly, f is continuous if and only if it is both upper and lower semicontinuous; equivalently, by observing that $\{x \mid f(x) > b\} = \{x \mid -f(x) < -b\}$, f is continuous if and only if both f and $-f$ are upper (or lower) semicontinuous.

The proof of the following theorem is not the simplest, but it does generalize to other situations in analysis (for example, measure theory).

10.3 Theorem Continuity is preserved under the usual operations of analysis. Precisely, let $f, g: X \to E^1$ be continuous. Then:

(1). $|f|^\alpha$ is continuous for each $\alpha \geqslant 0$.
(2). $af + bg$ is continuous for each pair of real constants, a, b.
(3). $f \cdot g$ is continuous.
(4). If $f(x) \neq 0$ on X, then $1/f$ is continuous (that is, $1/f$ is continuous wherever it is defined).

Proof: Ad (1). For $b < 0$, we have $\{x \mid |f(x)|^\alpha < b\} = \varnothing$ and

$$\{x \mid |f(x)|^\alpha > b\} = X.$$

For $b \geqslant 0$, we find

$$\{x \mid |f(x)|^\alpha > b\} = \{x \mid f(x) > b^{1/\alpha}\} \cup \{x \mid f(x) < -b^{1/\alpha}\},$$
$$\{x \mid |f(x)|^\alpha < b\} = \{x \mid f(x) < b^{1/\alpha}\} \cap \{x \mid f(x) > -b^{1/\alpha}\},$$

which, by continuity of f and **10.1**, are therefore open sets.

Ad (2). We first show that for any real a, af is continuous: Note that

$$\{x \mid af(x) > b\} = \{x \mid f(x) > b/a\} \quad \text{if } a > 0$$
$$= \{x \mid f(x) < b/a\} \quad \text{if } a < 0;$$

this shows that (f continuous) $\Rightarrow$ (af lower semicontinuous). In particular, $-(a \cdot f) = (-a) \cdot f$ is lower semicontinuous also, so that af is continuous. Now we need show only that $f + g$ is continuous; but

$$\{x \mid f(x) + g(x) > b\} = \bigcup \{\{x \mid f(x) > b - \lambda\} \cap \{x \mid g(x) > \lambda\} \mid \lambda \text{ real}\}$$

is a union of open sets, so $f + g$ is lower semicontinuous; similarly, since $-f$ and $-g$ are continuous, $-(f + g) = (-f) + (-g)$ is lower semicontinuous also.

Ad (3). We have $f \cdot g = \frac{1}{4}[|f+g|^2 - |f-g|^2]$, so use (2) and (1).
Ad (4). Since $f(x) \neq 0$ on X, we have $\{x \mid 1/f(x) > b\}$ equal to

$$[\{x \mid f(x) > 0\} \cap \{x \mid bf(x) < 1\}] \cup [\{x \mid f(x) < 0\} \cap \{x \mid bf(x) > 1\}];$$

since f and bf are continuous, this shows $1/f$ is lower semicontinuous. Similarly, $-(1/f)$ is lower semicontinuous.

For maps into the extended real line $\tilde{E}^1$ (**7**, Ex. 2), the criterion for continuity is evidently formally the same as **10.1**, so we can speak of upper and of lower semicontinuous maps. Furthermore, $\tilde{E}^1$ has the feature that for any $A \subset \tilde{E}^1$, sup A and inf A always exist.

10.4 Corollary Let $\{f_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of continuous maps $f_\alpha: X \to \tilde{E}^1$. Then:

(a). $M(x) = \sup\{f_\alpha(x) \mid \alpha \in \mathscr{A}\}$ is lower semicontinuous.
(b). $m(x) = \inf\{f_\alpha(x) \mid \alpha \in \mathscr{A}\}$ is upper semicontinuous.

Furthermore, if $\mathscr{A}$ is *finite*, both M and m are continuous.

Proof: Since $M(x) > b$ if and only if at least one $f_\alpha(x) > b$, we have the identity $\{x \mid M(x) > b\} = \bigcup_\alpha \{x \mid f_\alpha(x) > b\}$, which proves (a). Noting that $m(x) = -\sup_\alpha\{-f_\alpha(x)\}$ proves (b). Furthermore, whenever $\aleph(\mathscr{A})$ is finite, then $M(x) < b$ if and only if all $f_\alpha(x) < b$, so that $\{x \mid M(x) < b\} = \bigcap_\alpha \{x \mid f_\alpha(x) < b\}$; being a finite intersection, this set is certainly open and proves continuity of $M(x)$ in this case; that of $m(x)$ follows as before.

We will need later the analogue of the Weierstrass M-test,

10.5 Let $f_i: X \to E^1$, $i = 1, 2, \cdots$ be a sequence of continuous maps such that $|f_i(x)| \leqslant M_i$ for each i, where $\sum_{i=1}^{\infty} M_i$ is a convergent series of reals. Then $f(x) = \sum_1^\infty f_i(x)$ exists and is a continuous map $f: X \to E^1$.

Proof: Let $s_n(x) = \sum_1^n f_i(x)$; from domination by $\sum_1^\infty M_i$ we obtain $\forall\, \varepsilon > 0\ \exists\, n_0\ \forall\, n > n_0\ \forall\, x: |s_n(x) - s_{n_0}(x)| < \varepsilon$, so that s_n converges to f uniformly on X. The continuity of f at each x_0 (*cf.* **8.4**) follows from

$$|f(x) - f(x_0)| \leqslant |f(x) - s_n(x)| + |s_n(x) - s_n(x_0)| + |s_n(x_0) - f(x_0)|,$$

which holds for each n: given any $\varepsilon > 0$, first choose an n_0 so large that the two extreme terms on the right are $< \varepsilon/3$ for all x, and then, using the continuity of $s_{n_0}(x)$, choose a nbd $U(x_0)$ to have the middle term $< \varepsilon/3$.

11. Open Maps and Closed Maps

11.1 Definition A map $f: X \to Y$ is called open (resp. closed) if the image of each set open (resp. closed) in X is open (resp. closed) in Y.

We have already seen (**8**, Ex. 2) that a continuous map need not be an open map, and (**8**, Ex. 3) that an open map need not be continuous. The following example shows that, in general, an open map need not be a closed map (even though it is continuous); the concepts "open map," "closed map," and "continuous map" are therefore independent.

Ex. 1 Let $A \subset X$ and let $i: A \to X$ be the inclusion map $a \to a$. By **8**, Ex. 4, i is continuous. Furthermore, i is open (resp. closed) if and only if A is open (resp. closed) in X. Proof for "open": If A is open, and $U \subset A$ open in A, then by **7.3**, $i(U) = U$ is open in X. Conversely, if i is an open map, then because A is open in A, $i(A) = A$ is open in X. The proof for "closed" is analogous.

Ex. 2 If $f: X \to Y$ is *bijective*, then the conditions "f closed" and "f open" are in fact equivalent. For, if f is open and $A \subset X$ is closed, then $A = X - U$ and $f(A) = f(X) - f(U) = Y - f(U)$, so $f(A)$ is also closed. As **8**, Exs. 2 and 3, show, "bijective open" and "bijective continuous" are still distinct notions.

The behavior of inverse images further emphasizes the distinction between open maps and closed maps:

11.2 Theorem (1). Let $p: X \to Y$ be a closed map. Given any subset $S \subset Y$ and any open U containing $p^{-1}(S)$, there exists an open $V \supset S$ such that $p^{-1}(V) \subset U$.

(2). Let $p: X \to Y$ be an open map. Given any subset $S \subset Y$, and any closed A containing $p^{-1}(S)$, there exists a closed $B \supset S$ such that $p^{-1}(B) \subset A$.

Proof: We prove (1) only, since the proof of (2) is similar. Let $V = Y - p(X - U)$; since $p^{-1}(S) \subset U$, it follows that $S \subset V$, and because p is closed, V is open in Y. Observing that

$$p^{-1}(V) = X - p^{-1}[p(X - U)] \subset X - [X - U] = U$$

completes the proof.

Theorem **11.2**(1) is particularly important and has significant consequences; its most frequently occurring form is with S a single point.

We now give some characterizations of open maps and of closed maps.

11.3 Theorem The following four properties of a map $f: X \to Y$ are equivalent:

(1). f is an open map.

(2). $f[\text{Int}(A)] \subset \text{Int}[f(A)]$ for each $A \subset X$.

(3). f sends each member of a basis for X to an open set in Y.

(4). For each $x \in X$ and nbd $U \ni x$, there exists a nbd W in Y such that $f(x) \in W \subset f(U)$.

Proof: (1) $\Rightarrow$ (2). Since $\text{Int}(A) \subset A$, we have $f[\text{Int}(A)] \subset f(A)$; by hypothesis, $f[\text{Int}(A)]$ is open, and because $\text{Int}[f(A)]$ is the largest open set in $f(A)$, we must have $f[\text{Int}(A)] \subset \text{Int}[f(A)]$.

(2) $\Rightarrow$ (3). Let U be a member of a basis. Being open, $U = \text{Int}(U)$ and so $f(U) = f[\text{Int}(U)] \subset \text{Int}[f(U)] \subset f(U)$; thus, $f(U) = \text{Int}[f(U)]$ and therefore $f(U)$ is open.

(3) $\Rightarrow$ (4). Given x and $U \ni x$, find a member V of the basis for X such that $x \in V \subset U$ (*cf.* **2.3**) and let $W = f(V)$.

(4) $\Rightarrow$ (1). Let U be open in X. By hypothesis, each $y \in f(U)$ has a nbd $W(y) \subset f(U)$ so that $f(U) = \cup \{W(y) \mid y \in f(U)\}$ shows that $f(U)$ is open.

11.4 Theorem $p: X \to Y$ is a closed map if and only if $\overline{p(A)} \subset p(\overline{A})$ for each set $A \subset X$.

Proof: If p is closed, then by **4.4**(b), $p(\overline{A})$ is closed; since $p(A) \subset p(\overline{A})$, we obtain $\overline{p(A)} \subset \overline{p(\overline{A})} = p(\overline{A})$ as required. Conversely, if the condition holds and A is closed, then $p(A) \subset \overline{p(A)} \subset p(\overline{A}) = p(A)$ shows that $\overline{p(A)} = p(A)$, so that $p(A)$ is closed.

12. Homeomorphism

12.1 Definition A continuous bijective map $f: X \to Y$, such that $f^{-1}: Y \to X$ is also continuous, is called a homeomorphism (or a bicontinuous bijection) and denoted by $f: X \cong Y$. Two spaces X, Y are homeomorphic, written $X \cong Y$, if there is a homeomorphism $f: X \cong Y$.

Ex. 1 The map $x \to x/(1 + |x|)$ is a homeomorphism of E^1 and $]-1, +1[$. Interpreting x as a vector in E^n, this map shows that E^n is homeomorphic to its unit ball $B(0; 1)$.

Ex. 2 The extended real line $\tilde{E}^1$ is homeomorphic to $[-1, +1]$, since the map $x \to x/(1 + |x|) \mid (x \in E^1)$, $\pm\infty \to \pm 1$ is a homeomorphism.

Ex. 3 Let $p = (0, 0, 1)$ be the north pole of the sphere S^2; then $S^2 - \{p\} \cong E^2$, since the stereographic projection from p, which sends

$$(x_1, x_2, x_3) \in S^2 - \{p\} \quad \text{to} \quad \left(\frac{x_1}{1 - x_3}, \frac{x_2}{1 - x_3}, 0\right) \in E^2 \subset E^3$$

is easily verified to be a homeomorphism. This is the familiar process in complex analysis, which completes the complex numbers (geometrically, E^2) by adding a "point at infinity" to get S^2. In similar fashion, we have $S^n - \{(0, \cdots, 0, 1)\} \cong E^n$.

The importance of homeomorphisms results from the observation that a homeomorphism is also an open map; for, it then follows at once that a homeomorphism $f: X \cong Y$ provides *simultaneously* a bijection for the underlying spaces *and* for the topologies; that is, both $f: X \to Y$ and the induced $f \mid \mathscr{T}(X): \mathscr{T}(X) \to \mathscr{T}(Y)$ are bijective. Thus their significance is this: Any assertion about X *as a topological space* is also valid for each homeomorph of X; more precisely, every property of X expressed entirely in terms of set operations and open sets (that is, any topological property of X) is also possessed by each space homeomorphic to X.

Somewhat more generally, we call any property of spaces a *topological invariant* if whenever it is true for one space X, it is also true for every space homeomorphic to X; trivial examples are cardinal of point set, and cardinal of topology. With this terminology, every topological property of a space is a topological invariant, homeomorphic spaces have the same topological invariants, and Topology can be described as the study of topological invariants.

This description of Topology can be expressed more formally: Observe that homeomorphism is an equivalence relation in the class of all topological spaces, since (a) $1: X \cong X$; (b) $[f: X \cong Y] \Rightarrow [f^{-1}: Y \cong X]$; and (c) $[f: X \cong Y] \wedge [g: Y \cong Z] \Rightarrow [g \circ f: X \cong Z]$ as the reader can easily verify. Consequently, the relation of homeomorphism decomposes the class of all topological spaces into mutually exclusive classes, called *homeomorphism types*. In these terms, Topology studies invariants of homeomorphism types.

Homeomorphism frequently allows the reduction of a given problem to a simpler one: A space that is given, or constructed, in some complicated manner may possibly be shown homeomorphic to something more familiar, and its topological properties thereby more easily determined. For example, it is known that the Riemann surface of an algebraic function is homeomorphic to a sphere S^2 having suitably many attached handles. Unfortunately, to show that two given spaces are homeomorphic is usually difficult, with construction of a homeomorphism being the only general method. In some special cases, such as two-dimensional manifolds, other (algebraic) techniques have been devised.

It is frequently important to know that two spaces are *not* homeomorphic, as, for example, E^1 and E^2. This problem is somewhat simpler than the former; it is generally solved by displaying a topological invariant possessed by only one of the two spaces. Topological invariants not possessed by all spaces are therefore important; in succeeding chapters, we shall consider many such invariants.

We now give some characterizations and properties of homeomorphisms.

12.2 Theorem Let $f: X \to Y$ be *bijective*. The following properties of f are equivalent:

(1). f is a homeomorphism.
(2). f is continuous and open.
(3). f is continuous and closed.
(4). $f(\overline{A}) = \overline{f(A)}$ for each $A \subset X$.

Proof: (1) $\Leftrightarrow$ (2). The requirement that the map $f^{-1}: Y \to X$ be continuous is equivalent to the stipulation that for each open $U \subset X$, the set $(f^{-1})^{-1}(U) = f(U)$ be open in Y.

(2) $\Leftrightarrow$ (3). This is **11**, Ex. 2.

(3) $\Leftrightarrow$ (4). Continuity of f yields $f(\overline{A}) \subset \overline{f(A)}$, and because f is closed, **11.4** shows that also $\overline{f(A)} \subset f(\overline{A})$.

One frequently used technique for establishing that a given $f: X \to Y$ is a homeomorphism is simply to exhibit a continuous $g: Y \to X$ in accordance with

12.3 Theorem Let $f: X \to Y$ and $g: Y \to X$ be continuous and such that both $g \circ f = 1_X$ and $f \circ g = 1_Y$. Then f is a homeomorphism, and in fact, $g = f^{-1}$.

Proof: We know (I, **6.9**) that both f, g are bijective, and it is trivial to see that $g = f^{-1}$; since both f and g are continuous, the proof is complete.

For subspaces,

12.4 Theorem Let $f: X \cong Y$ and $A \subset X$. Then $f \mid A: A \cong f(A)$ and $f \mid X - A: X - A \cong Y - f(A)$.

Proof: Let $g = f^{-1} \mid f(A)$; then g is continuous, by **8.2**(b), and the pair of maps $f \mid A$, g satisfies **12.3**. The second part is proved in the same way.

If Z is any space and $f: X \to Z$ is a map establishing $X \cong f(X) \subset Z$, then f is called an *embedding map* of X into Z.

Problems

Section 1

1. a. Let X be an infinite set. Show that $\mathscr{A}_0 = \{\varnothing\} \cup \{A \mid \mathscr{C}A \text{ is finite}\}$ is a topology.
   b. Let $\aleph(X) \geqslant \aleph_0$. Show that $\mathscr{A}_1 = \{\varnothing\} \cup \{A \mid \aleph(\mathscr{C}A) < \aleph(X)\}$ is a topology.
2. How many distinct topologies can a set of three elements have? What is their partial ordering?
3. Let $\mathscr{T}_X, \mathscr{T}_Y$ be topologies in X, Y, respectively. Is
$$\mathscr{T} = \{A \times B \mid A \in \mathscr{T}_X, B \in \mathscr{T}_Y\}$$
a topology in $X \times Y$?
4. Let X be a partially ordered set. Define $U \subset X$ to be open if it satisfies the condition: $(x \in U) \wedge (y \prec x) \Rightarrow y \in U$. Show that $\{U \mid U \text{ is open}\}$ is a topology.
5. In Z^+, define $U \subset Z^+$ to be open if it satisfies the condition: $n \in U \Rightarrow$ every divisor of n belongs to U. Show that this is a topology in Z^+ and that it is not the discrete topology.
6. Prove: $\mathscr{T}$ is the discrete topology in X if and only if every point is an open set.

Section 3

1. Use **3.2** to verify the final statements of **1**, Exs. 4 and 5.
2. If, in the plane, all straight lines are taken as subbasis, what is the topology?
3. Describe the open sets if all straight lines in the plane parallel to the x-axis are used for subbasis.
4. Let X be the set of all $(n \times n)$ matrices of real numbers. For each $a = (a_{ij})$ and $r > 0$, let $U_r(a) = \{(b_{ij}) \mid \forall\, i, j\colon |a_{ij} - b_{ij}| < r\}$. Show that these sets are the basis for a topology in X.
5. Let C be the set of all continuous real-valued functions on $[0, 1]$. For each $f \in C$, each finite set $x_1, \cdots, x_n \in [0, 1]$, and $\varepsilon > 0$, let $U_{(x_1,\cdots,x_n,\varepsilon)}(f) = \{g \mid |g(x_i) - f(x_i)| < \varepsilon,\ i = 1, \cdots, n\}$. Show (a) that these nbds form a basis for some topology $\mathscr{L}$; (b) that $\mathscr{L} \subset \mathscr{U}$; and (c) that $\mathscr{L}$ and $\mathscr{M}$ are not related in the partial ordering of topologies in C.
6. Let X be a partially ordered set. Let $U_L(x) = \{y \mid y \prec x\}$ and $U_R(x) = \{y \mid x \prec y\}$. Show:
   a. The families $\{U_L(x)\}$, $\{U_R(x)\}$ form bases for topologies $\mathscr{T}_L$, $\mathscr{T}_R$ in X.
   b. $G \in \mathscr{T}_L$ if and only if it satisfies the condition $x \in G \Rightarrow U_L(x) \subset G$.
   c. In $\mathscr{T}_L$, the arbitrary intersection of open sets is an open set.
   d. The discrete topology is the only one larger than $\mathscr{T}_L$ and larger than $\mathscr{T}_R$.
   e. $\mathscr{T}_L$ and $\mathscr{T}_R$ are not related in the partial ordering of the topologies on X.
7. In Z, let p be a fixed prime. For each integer $a > 0$, define $U_a(n) = \{n + \lambda p^a \mid \lambda \in Z\}$. Show that $\{U_a(n)\}$ is a basis for some topology.
8. Let $\{\mathscr{T}_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of topologies on X. Define $\bigvee_\alpha \mathscr{T}_\alpha$ to be the topology having $\bigcup_\alpha \mathscr{T}_\alpha$ as subbasis.

a. Prove: $\bigvee_\alpha \mathscr{T}_\alpha$ is the smallest of the topologies on X larger than every $\mathscr{T}_\alpha$.

b. Show that with the operations $\mathscr{T}_1 \vee \mathscr{T}_2$ and $\mathscr{T}_1 \cap \mathscr{T}_2$ that the topologies on X form a complete lattice. (*Cf.* Problem II 1(4): a lattice is called complete if *every* subset has a least upper, and a greatest lower, bound.)

c. Is $\mathscr{T}_1 \cap (\mathscr{T}_2 \vee \mathscr{T}_3) = (\mathscr{T}_1 \cap \mathscr{T}_2) \vee (\mathscr{T}_1 \cap \mathscr{T}_3)$ always true?

d. What is $\mathscr{T}_L \vee \mathscr{T}_R$ in Problem 6?

9. In the set R of all real numbers, let Σ consist of all sets of form $\{x \mid x > r\}$, $\{x \mid x < s\}$, where r, s are rational. Show that $\mathscr{T}(\Sigma)$ is the Euclidean topology of R. Is this still true if r, s are restricted to be numbers of the form $k/2^n$ (k and n arbitrary)?

Section 4

1. Determine the closure, derived set, interior, and boundary, of the following sets: (a) The rationals in E^1; (b) the Cantor set in E^1; (c) the set $\{(r_1, r_2) \mid r_1, r_2 \text{ rational}\} \subset E^2$; (d) $\{(x, 0) \mid 0 < x < 1\} \subset E^2$.
2. Show that $E^s \times 0 \subset E^s \times E^t = E^{s+t}$ is closed.
3. Let $A \subset E^1$ be a bounded set. Show $\sup A \in \bar{A}$. Under what conditions is $\sup A \in A'$?
4. Prove: G is open in X if and only if $\overline{G \cap \bar{A}} = \overline{G \cap A}$ for every $A \subset X$.
5. Show that $(A' = \varnothing) \Rightarrow (A \text{ is closed})$.
6. Let $A = \{1/m + 1/n \mid m, n \in Z^+\} \subset E^1$. Show that $A' = \{1/n \mid n \in Z^+\} \cup \{0\}$, $A'' = \{0\}$.
7. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of sets in X. Assume that $\bigcup_\alpha \bar{A}_\alpha$ is closed. Prove $\bigcup_\alpha \bar{A}_\alpha = \overline{\bigcup_\alpha A_\alpha}$.
8. Prove: $\text{Fr}(A) = \varnothing$ if and only if A is both open and closed.
9. Prove the following formulas:
   (a). $\text{Fr}[\text{Fr}\{\text{Fr}(A)\}] = \text{Fr}[\text{Fr}(A)]$;
   (b). $\text{Fr}[\text{Int}(A)] \subset \text{Fr}(A)$;
   (c). $\text{Int}(A - B) \subset \text{Int}(A) - \text{Int}(B)$.
10. Assume that $\text{Fr}(A) \cap \text{Fr}(B) = \varnothing$. Prove: $\text{Int}(A \cup B) = \text{Int}(A) \cup \text{Int}(B)$ and $\text{Fr}(A \cap B) = [\bar{A} \cap \text{Fr}(B)] \cup [\text{Fr}(A) \cap \bar{B}]$.
11. For what spaces X is the only dense set X itself?
12. Let E and G be dense in X. Prove: If E and G are open, then $E \cap G$ is also dense in X.
13. Let D be dense in X. Prove: $\overline{D \cap G} = \bar{G}$ for every open $G \subset X$.
14. Let $\mathscr{B}$ be a subbasis for X, and $D \subset X$ such that $U \cap D \neq \varnothing$ for each $U \in \mathscr{B}$. Does this imply that D is dense in X?
15. In **3**, Problem 6, what is the closure of $\{x\}$ in $\mathscr{T}_L$? For what points is $\{x\}$ a closed set? An open set?
16. Assume $(X, \mathscr{T})$ a space with the properties: (a) The intersection of any family of open sets is open; (b) if $x \neq y$, then there is at least one open set containing some one of x, y, and not the other. Define $x \leqslant y$ if $x \in \bar{y}$. Show that this is a partial ordering and that the topology $\mathscr{T}_R$ (**3**, Problem 6) is precisely $\mathscr{T}$.
17. In $X \times Y$, show that $\overline{A \times B} = \bar{A} \times \bar{B}$; $\text{Int}(A \times B) = \text{Int}(A) \times \text{Int}(B)$; $\text{Fr}(A \times B) = (\text{Fr}(A) \times \bar{B}) \cup (\bar{A} \times \text{Fr}(B))$.

18. The exterior, $\operatorname{Ext}(A)$, of a set $A \subset X$ is defined by $\operatorname{Ext}(A) = \operatorname{Int}(\mathscr{C}A)$. Prove: (a). $\operatorname{Ext}(A \cup B) = \operatorname{Ext}(A) \cap \operatorname{Ext}(B)$; (b). $A \cap \operatorname{Ext}(A) = \varnothing$; (c). $X = \operatorname{Ext}(\varnothing)$; (d). $\operatorname{Ext}[\mathscr{C}\operatorname{Ext}(A)] = \operatorname{Ext}(A)$.
19. If every countable subset of a space is closed, is the topology necessarily discrete?
20. A point $a \in A$ is called isolated whenever $a \in A - A'$. A set is called perfect if it is closed and has no isolated points. Prove: (1) If A has no isolated points, then $\bar{A}$ is perfect. (2) If a space X has no isolated points, then every open set and every dense set in X also have no isolated points.
21. Call a set residual if its complement is dense, and call it nowhere dense if its closure has empty interior. Prove: (1) A nowhere dense set is a residual set. (2) A is nowhere dense if and only if $A \subset \overline{\mathscr{C}\bar{A}}$. (3) The union of a residual and a nowhere dense set is a residual set. (4) The boundary of a closed (or open) set is nowhere dense. (5) For any set A, both $A \cap \overline{\mathscr{C}A}$ and $\bar{A} \cap \mathscr{C}A$ are residual. (6) The boundary of any set is the union of two residual sets.
22. An open set U is called regular if $U = \operatorname{Int}(\bar{U})$; a closed set A is called regular if $A = \overline{\operatorname{Int}(A)}$. Prove:
    a. If A is closed, then $\operatorname{Int}(A)$ is a regular open set.
    b. If U is open, then $\bar{U}$ is a regular closed set.
    c. The complement of a regular open (closed) set is a regular closed (open) set.
    d. If U, V are regular open sets, then $U \subset V$ if and only if $\bar{U} \subset \bar{V}$.
    e. If A, B are regular closed sets, then $A \subset B$ if and only if $\operatorname{Int}(A) \subset \operatorname{Int}(B)$.
    f. If A, B are regular closed sets, so also is $A \cup B$.
    g. If U, V are regular open sets, so also is $U \cap V$.

Section 5

1. Let $A \to u(A)$ and $A \to v(A)$ be two closure operations. Assume that $v \circ u(A)$ is u-closed. Prove that $A \to v \circ u(A)$ is a closure operation and that $v \circ u(A)$ is in fact the intersection of all sets containing A that are closed in both u and v. Finally, show that $u \circ v(A) \subset v \circ u(A)$.
2. Let $\varphi: X \to \mathscr{P}(Y)$ be a map of a *set* X into the *set* $\mathscr{P}(Y)$. For $A \subset X$, let $\varphi(A) = \bigcup \{\varphi(x) \mid x \in A\}$; and for $B \subset Y$, let $\varphi^{-1}(B) = \{x \mid (\varphi(x) \subset B) \wedge (\varphi(x) \neq \varnothing)\}$. Show that $u(A) = \varphi^{-1} \circ \varphi(A)$ satisfies (1)–(3) and (i) of **5.1**.
3. Let X be a space, and let τ be an operation associating with each pair of subsets A and B a set $\tau(A, B) \subset X$ subject to the conditions:
    a. $\tau(A, B \cup C) \cup \tau(B, C \cup A) = \tau(A \cup B, C) \cup \tau(A, B)$.
    b. $\tau(\varnothing, X) = \varnothing$.
    c. $\tau(\bar{A}, \overline{\mathscr{C}A}) \subset \bar{A}$.
    d. $\tau(A, B) \subset A \cup B$.

    Show that, necessarily, $\tau(A, B) = (A \cap \bar{B}) \cup (\bar{A} \cap B)$.

Section 6

1. A set is bivalent if it is both an F_σ and a G_δ. Show that the complements, finite unions, and finite intersections of bivalent sets are bivalent.

For the remaining problems, we assume that all closed sets are G_δ-sets.

2. Prove: Every F_σ is the disjoint union of bivalent sets. (Write $F = \bigcup F_i = F_1 \cup [F_2 - F_1] \cup \cdots [F_n - (F_{n-1} \cup \cdots \cup F_1)] \cup \cdots$ and use Problem 1.)
3. Prove: For every sequence $\{F_i\}$ of F_σ-sets, there exists a pairwise disjoint sequence $\{H_i\}$ of F_σ-sets with $H_i \subset F_i$ for every i and $\cup H_i = \cup F_i$. [Write $F_i = \bigcup_j F_{ij}$, where the F_{ij} are bivalent and disjoint; arrange the pairs (i, j) in a linear order, and repeat Problem 2.]
4. Prove: If $\{G_i\}$ is any sequence of G_δ-sets with $\bigcap_1^\infty G_i = \varnothing$, there exists a sequence of bivalent sets $\{B_i\}$ with $G_i \subset B_i$ and $\bigcap_1^\infty B_i = \varnothing$. (Set $F_i = \mathscr{C}G_i$ in Problem 3 and take $B_i = \mathscr{C}H_i$.)
5. Prove: (a) If G, H are disjoint G_δ-sets, there exists a bivalent set B with $H \subset B$ and $B \cap G = \varnothing$.

   (b) If G is a G_δ and F an F_σ such that $G \subset F$, there exists a bivalent B with $G \subset B \subset F$.

Section 7

1. In $\tilde{E}^1$, show that for any set A, $\sup A \in \overline{A}$.
2. Describe the relative topology of $\{z \mid |z| = 1\}$ as a subspace of E^2.
3. Show that the rationals, as a subspace of E^1, do not have the discrete topology.
4. Let $K \subset E^1$ be the set of irrationals in $]0, 1[$. Expressing each $x, y \in K$ as decimals, define $d(x, y) = 1/n$ if x and y have their first $(n - 1)$ digits identical, and their nth digits different. Let $B(x; 1/n) = \{y \in K \mid d(x, y) < 1/n\}$. Show that the $\{B(x, 1/n) \mid x \in K, n \in Z^+\}$ is a basis for a topology $\mathscr{T}$ on K, and prove that $\mathscr{T}$ coincides with the induced topology of K as a subspace of E^1.
5. Let $A \subset B$ be open in B. Show: For any set S, $A \cap S$ is open in $B \cap S$.
6. Let Y_1, Y_2 be (not necessarily disjoint) subspaces of X, and $A \subset Y_1 \cap Y_2$. Assume that A is open (closed) in Y_1 and open (closed) in Y_2. Prove: A is open (closed) in $Y_1 \cup Y_2$.
7. Let $A \subset X$ be closed and $U \subset A$ open in A. Let V be any set open in X with $U \subset V$. Prove: $U \cup (V - A)$ is open in X.
8. a. Let D be dense in X. Give an example to show $D \cap A$ need not be dense in A.

   b. Prove: If A is dense in $B \subset X$, then A is dense in $\overline{B}$.
9. Let X be the set of **3**, Problem 6, with the topology $\mathscr{T}_R$. Let $A \subset X$, and using the induced ordering on A, obtain $\mathscr{T}_R(A)$. Show that $\mathscr{T}_R(A)$ coincides with the induced topology on A as a subspace of X.
10. For any linearly ordered set X, let $\mathscr{T}_0(X)$ be the topology with subbasis all sets of form $\{x \mid x > a\}$, $\{x \mid x < b\}$. In E^1, $\mathscr{T}_0(E^1)$ is the Euclidean topology. Let $A = \{0\} \cup \{x \mid |x| > 1\}$. Show that its topology $\mathscr{T}_0(A)$ as a linearly ordered set does *not* coincide with its topology as a subspace of E^1.
11. Let X be any space, and let Y be a closed subspace. Let $A \subset X$ be any set, and let H be a nbd of $A \cap Y$ in Y. Prove: $A \cap \overline{(Y - H)} = \varnothing$.
12. Let X be any space, and assume that $X = E_1 \cup E_2$, where E_1 and E_2 are closed in X. Let $B \subset E_1$ be such that $B \cap E_2 \subset Q$ where Q is open in E_2. Prove: $B \subset \text{Int}(E_1 \cup Q)$.

Section 8

1. Let $\mathscr{S}$ be Sierpinski space and let 2 be the discrete space $\{0, 1\}$ (*cf.* **1**, Ex. 3). Let $f: \mathscr{S} \to 2$ be the identity map. Show that f is not continuous, but that $f^{-1}: 2 \to \mathscr{S}$ is.
2. Let c be the characteristic function of $[0, 1[\subset E^1$. Is $c \mid [0, 1]$ continuous at $x = 0$? At $x = 1$? On $[0, 1[$?
3. Let X be any space and c_A the characteristic function of $A \subset X$. Show that $c_A: X \to E^1$ is continuous if and only if A is both open and closed in X.
4. Let C be the set of **3**, Ex. 7. (a) Define $\varphi: C \to E^1$ by $\varphi(f) = f(1)$. Show that φ is continuous in the $\mathscr{U}$ topology, but is *not* continuous in the $\mathscr{M}$ topology. (b) Define $\psi: C \to E^1$ by

$$\psi(f) = \int_0^1 f(x)\, dx.$$

Show that ψ is continuous in both the $\mathscr{M}$ and $\mathscr{U}$ topologies. (c) Are either of these maps continuous in the $\mathscr{L}$ topology (**3**, Problem 5)?
5. Under what conditions is the bijective map $1: (X, \mathscr{T}_1) \to (X, \mathscr{T}_2)$ not continuous, and with inverse not continuous.
6. Let X be the space in **1**, Problem 1(a). State a necessary and sufficient condition that $f: X \to X$ be continuous.
7. Let X be the space in **1**, Problem 4. Show that $f: X \to X$ is continuous if and only if it is order-preserving.
8. Let Z^+ be taken with the topology of **1**, Problem 5. Show that $f: Z^+ \to Z^+$ is continuous if and only if (m divides n) $\Rightarrow$ ($f(m)$ divides $f(n)$).
9. Let X be the set of **3**, Problem 6, with topology $\mathscr{T}_R$. Derive a necessary and sufficient condition for continuity of $f: X \to X$.
10. Prove that the following three statements are equivalent:
    a. $f: X \to Y$ is continuous.
    b. $f(A') \subset \overline{f(A)}$ for each $A \subset X$.
    c. $\text{Fr}[f^{-1}(B)] \subset f^{-1}[\text{Fr}(B)]$ for each $B \subset Y$.
11. Let $X = X_1 \cup X_2$ and $f: X \to Y$. Assume $f \mid X_1$ and $f \mid X_2$ to be continuous at $x \in X_1 \cap X_2$. Show that f is continuous at x.
12. Let $f: X \to Y$ be a map and $A \subset X$. Give an example showing $f \mid A$ continuous, although f is not continuous at any point of A.
13. Let $f: X \to Y$ be continuous. If $B \subset Y$ is a G_δ(resp. F_σ), show that $f^{-1}(B)$ is also a G_δ(resp. F_σ).
14. Construct an example of a map $f: X \times X \to E^1$ continuous in each variable separately but not continuous on $X \times X$ (use $X = E^1$ for simplicity).

Section 9

1. Let $\{A_\lambda\}$ be a nbd-finite closed covering of X. Let $x_0 \in X$, and for each set A_{λ_i} ($i = 1, \cdots, n$) containing x_0, let V_i be a nbd of x_0 in A_{λ_i}. Show that $\bigcup_1^n V_i$ contains a nbd of x_0 in X.
2. Let X be any space, and $\{A_n \mid n \in Z^+\}$ a family of sets in X such that $A_n \supset A_{n+1}$ for each $n \in Z^+$. Prove that, if $\bigcap_1^\infty \overline{A}_n = \varnothing$, then the family $\{A_n \mid n \in Z^+\}$ is nbd-finite.

3. Prove: a. Any open set in E^1 is the union of a pairwise disjoint family of open intervals. Give an example to show that this is not true for E^n, $n > 1$, if "open intervals" is replaced by "open cubes."

   b. Any open set in E^n is the union of at most countably many nonoverlapping closed cubes [*Hint:* Generalize the following procedure for E^2: For each $n \in Z^+$, let H_n be the network of closed squares formed by all the lines $x = p/2^n$, $y = q/2^n$, $p, q \in Z$. Let S_1 be the set of all the closed squares of H_1 contained in the open set G, and proceeding inductively, let S_{n+1} be the set of all the closed squares of H_{n+1} contained in $G - (S_1 \cup \cdots \cup S_n)$. Then $G = \bigcup_1^\infty S_n$].

   c. Show that a non-empty open set in E^n cannot be the union of a nbd-finite family of nonoverlapping closed cubes.

Section 10

1. Let $f: X \to E^1$ be such that for each rational r, $\{x \mid f(x) > r\}$ is open. Show that f is lower semicontinuous.
2. Let $\mathscr{M}$ be a family of sets in X closed under countable unions and finite intersections. Call $f: X \to E^1$ an $\mathscr{M}$-function if $f^{-1}(U) \in \mathscr{M}$ for each subbasic U of E^1. Show that **10.3** is true, with "continuous" replaced by "$\mathscr{M}$-function."
3. Let E be the space of real numbers with subbasis all sets of form $\{x \mid x > b\}$. Prove $f: X \to E$ is continuous if and only if f is lower semicontinuous, as a map into E^1.
4. Show that the sum, inf of finitely many, and sup of an arbitrary family, of lower semicontinuous functions is lower semicontinuous. Show that if f, g are $\geqslant 0$ and lower semicontinuous, so also is $f \cdot g$.
5. Prove: f is lower semicontinuous if and only if:
$$\forall \varepsilon > 0 \, \forall x_0 \, \exists U(x_0): x \in U(x_0) \Rightarrow f(x) > f(x_0) - \varepsilon$$
   f is upper semicontinuous if and only if:
$$\forall \varepsilon > 0 \, \forall x_0 \, \exists U(x_0): x \in U(x_0) \Rightarrow f(x) < f(x_0) + \varepsilon$$
   Equivalently, f is lower (upper) semicontinuous if and only if for each real b, $\{x \mid f(x) \leqslant b\}$ ($\{x \mid f(x) \geqslant b\}$) is closed.
6. Let C^1 be the set of all continuous real-valued functions on [0, 1] that have a continuous derivative on [0, 1]. Let $\mathscr{T}$ be the topology having
$$U_\varepsilon(f) = \{g \mid \forall x: |g(x) - f(x)| < \varepsilon\}$$
   as a basis. Define the length
$$l(f) = \int_0^1 \sqrt{1 + f'(x)^2}\, dx.$$
   Show that $l: C^1 \to E^1$ is lower semicontinuous.
7. Let $\{f_n\}$ be a sequence of continuous real-valued functions on a space X. Show that the set of points $x \in X$ for which the sequence $\{f_n(x)\}$ converges is an $F_{\sigma\delta}$ (that is, a countable intersection of F_σ-sets).

Section 11

1. Let $p: E^1 \times E^1 \to E^1$ be the projection $(x, y) \to x$. Show: (a) that p is an open mapping; (b) that p is not a closed map. (Consider $p\{(x, y) \mid xy = 1\}$.) Thus "bijective" in Ex. 2 cannot be replaced by "surjective."

2. Let $p(x)$ be a polynomial on E^1. Show that $x \to p(x)$ is a closed map of E^1.
3. Show that a continuous bijective map $f: E^1 \to E^1$ is an open map.
4. Give an example to show that a continuous open map need not map the interior of a set onto the interior of the image.
5. Let X be any space, and $f: X \to E^n$ an open mapping. Denote the distance of $f(x) \in E^n$ to the origin by $|f(x)|$. Let $A \subset X$ be any set. Prove that $|f(a)| < \sup\{|f(x)| \mid x \in A\}$ for every $a \in \text{Int}(A)$. [*Hint:* Given $a \in \text{Int}(A)$, there is, by **11.3**(4), a ball $B(f(a); \epsilon) \subset f(A)$]. Since a nonconstant analytic function of a complex variable is an open map on its domain of definition, this result implies "maximum modulus theorem" of complex analysis.
6. Show that $f: X \to Y$ is open if and only if $f^{-1}[\text{Fr}(B)] \subset \text{Fr}[f^{-1}(B)]$ for each $B \subset Y$.
7. Let $f: X \to Y$ and $g: Y \to Z$. Prove:
   a. If f, g are open (closed), so also is $g \circ f$.
   b. If $g \circ f$ is open (closed) and f is continuous, surjective, then g is open (closed).
   c. If $g \circ f$ is open (closed) and g is continuous, injective, then f is open (closed).
8. Let $f: X \to Y$ be open (closed). Show that if $A = f^{-1}(B)$ for some $B \subset Y$, then $f \mid A$ is also an open (closed) map into B.
9. Let $\{B_\alpha \mid \alpha \in \mathscr{A}\}$ be an open, or nbd-finite closed, covering of Y. Let $f: X \to Y$ be such that for each $\alpha \in \mathscr{A}$, $f \mid f^{-1}(B_\alpha)$ is an open (closed) map of $f^{-1}(B_\alpha)$ into B_α. Show that f is an open (closed) map.
10. Prove: A surjective map $p: X \to Y$ is a closed map if and only if it satisfies the condition of **11.2**(1).
11. Let $p: X \to Y$ be a closed surjection. Prove: If $U \subset X$ is open, then $\text{Fr}[p(\overline{U})] \subset p(\overline{U}) \cap p(X - U)$.
12. Let $p: X \to Y$ be closed. Let $U \subset X$ be open, and $p^{-1}(y) \subset U$. Show that $y \in \text{Int}[p(\overline{U})]$.
13. Let $p: X \to Y$ be closed. Prove: If $A \subset X$ is closed, then $p \mid A$ is a closed mapping of A into $p(A)$.
14. Prove that the following three statements are equivalent:
    a. $p: X \to Y$ is a closed map.
    b. If $U \subset X$ is open, then $\{y \mid p^{-1}(y) \subset U\}$ is open in Y.
    c. If $A \subset X$ is closed, then $\{y \mid p^{-1}(y) \cap A \neq \varnothing]$ is closed in Y.
15. Let $c: I \to 2$ be the characteristic function of $[0, \frac{1}{2}] \subset I$. Show that c is surjective, open, closed, but not continuous.
16. Let $p: X \to Y$ be an open and closed map. Let $\varphi: X \to I$ be continuous, and for each $y \in Y$, let $\hat{\varphi}(y) = \sup [\varphi \mid p^{-1}(y)]$. Prove: $\hat{\varphi}: Y \to I$ is continuous.

Section 12

1. In E^2, let $V^2 = \overline{B(0, 1)}$. Show $V^2 \cong I \times I$.
2. Show that the space of $n \times n$ matrices in **3**, Problem 4, is homeomorphic to E^{n^2}.
3. Show: a. $(X \times Y) \times Z \cong X \times (Y \times Z)$,
   b. $X \times Y \cong Y \times X$.

4. Let $f: X \to Y$ be a map, and $G(f) \subset X \times Y$ the subspace $\{(x, f(x)) \mid x \in X\}$. Show that $x \to (x, f(x))$ is a homeomorphism if and only if f is continuous.
5. Let A, B be countable dense subsets of E^1. Show $A \cong B$. [There is even a homeomorphism of E^1 carrying A onto B.]
6. Let $f: X \to Y$ and $g: Y \to X$ be embeddings. Show that (see II, **7**, Problem 7) X and Y can be written as disjoint unions: $X = X_1 \cup X_2$, $Y = Y_1 \cup Y_2$ such that $f \mid X_1: X_1 \cong Y_1$ and $g \mid Y_2: Y_2 \cong X_2$.
7. Let $\{B_\alpha \mid \alpha \in \mathscr{A}\}$ be an open, or nbd-finite closed, covering of Y. Let $f: X \to Y$ be continuous, and assume that for each $\alpha \in \mathscr{A}$, $f \mid f^{-1}(B_\alpha)$ is a homeomorphism of $f^{-1}(B_\alpha)$ and B_α. Prove that $f: X \cong Y$.
8. Show that the map $x \to e^{ix}$ of $[0, 2\pi[$ onto S^1 is not a homeomorphism.

Cartesian Products

IV

In this chapter, we define the cartesian product topology for an arbitrary family of spaces and derive some of its basic properties.

1. Cartesian Product Topology

1.1 Definition Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of topological spaces. For each $\alpha \in \mathscr{A}$, let $\mathscr{T}_\alpha$ be the topology for Y_α. The cartesian product topology in $\prod_\alpha Y_\alpha$ is that having for subbasis all sets $\langle U_\beta \rangle = p_\beta^{-1}(U_\beta)$, where U_β ranges over all members of $\mathscr{T}_\beta$ and β over all elements of $\mathscr{A}$.

If any one factor $Y_\alpha = \varnothing$, then the cartesian product is also empty and the topologies of the non-empty factors then play no role in determining that of the product; we wish to exclude such behavior, so we shall assume throughout this book that each factor of any cartesian product of *topological spaces* is non-empty.

Ex. 1 If $\mathscr{A}$ is a finite set, the cartesian product topology of $\prod_{\alpha=1}^{n} Y_\alpha$ coincides with that previously defined in III, **3**, Ex. 6.

Ex. 2 Because of I, **9.6**, it follows at once that the basic open sets are all those sets of form

$$U_{\alpha_1} \times \cdots \times U_{\alpha_n} \times \prod \{Y_\beta \mid \beta \neq \alpha_1, \cdots, \alpha_n\} = \bigcap_1^n p_{\alpha_i}^{-1}(U_{\alpha_i}),$$

where n is finite and each U_{α_i} is an open set in Y_{α_i}; we will denote this set by $\langle U_{\alpha_1}, \cdots U_{\alpha_n} \rangle$. Thus, whenever $\aleph(\mathscr{A}) = n < \aleph_0$, a *basis* for the topology is simply the family of all "boxes" $\prod_{i=1}^{n} U_i$, where each U_i is an open set in Y_i. However, whenever $\aleph(\mathscr{A}) \geqslant \aleph_0$, this is no longer true: In fact, a set of form $\prod U_\alpha$, where each $U_\alpha \neq Y_\alpha$ and is open in Y_α, is *never* an open set in the cartesian product: because each basic open set restricts only finitely many coordinates, no basic open set can be contained in $\prod_\alpha U_\alpha$, so (III, **2.3**) $\prod_\alpha U_\alpha$ is not open.

Ex. 3 If $\aleph(\mathscr{A}) \geqslant \aleph_0$, and each Y_α is a discrete space having more than one point, $\prod_\alpha Y_\alpha$ is never a discrete space, for since each $a_\alpha \in Y_\alpha$ is an open set, it follows from Ex. 2 that no point $\{a_\alpha\} \in \prod_\alpha Y_\alpha$ is an open set. The cartesian product topology is frequently used in this way, to create a nontrivial topology in infinite families of objects, starting from the discrete topology in each.

Ex. 4 It is useful to observe that if V is any set open in $\prod_\alpha Y_\alpha$, then $p_\alpha(V) = Y_\alpha$ for all but at most finitely many α: any such V contains a basic $U = \langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle$ and $p_\alpha(U) \subset p_\alpha(V)$ for all α.

As Ex. 2 indicates, results for finite cartesian products may not carry over formally to infinite cartesian products, so these two cases will be treated separately.

1.2 Let $\aleph(\mathscr{A})$ be arbitrary. Then, in the space $\prod \{Y_\alpha \mid \alpha \in \mathscr{A}\}$:

(1). For each $\alpha \in \mathscr{A}$, let Σ_α be a *subbasis* for the topology $\mathscr{T}_\alpha$ of Y_α. Then the family $\{\langle V_\beta\rangle \mid \text{all } V_\beta \in \Sigma_\beta;\ \text{all } \beta \in \mathscr{A}\}$ is also a subbasis for the cartesian product topology in $\prod_\alpha Y_\alpha$.

(2). If $A_\alpha \subset Y_\alpha$ for each $\alpha \in \mathscr{A}$, then $\overline{\prod_\alpha A_\alpha} = \prod_\alpha \overline{A}_\alpha$ (that is, in contrast to Ex. 2, the cartesian product of *closed* sets is always closed).

(3). If $A_\alpha \subset Y_\alpha$ for each $\alpha \in \mathscr{A}$, then $\prod_\alpha A_\alpha$ as a cartesian product of *spaces* (each having the induced topology) has the same topology that it receives as a subspace of $\prod_\alpha Y_\alpha$.

(4). Let $y^\circ = \{y_\alpha^\circ\}$ be a fixed element in $\prod_\alpha Y_\alpha$, and let

$$D = \{\{x_\alpha\} \mid \{x_\alpha\} \text{ and } \{y_\alpha^\circ\} \text{ differ in } \textit{at most} \text{ finitely many coordinates}\}.$$

Then D is dense in $\prod_\alpha Y_\alpha$.

Proof: (1). This is immediate from I, **9.5**, and is left for the reader.

(2). Assume $\{y_\alpha\} \in \overline{\prod_\alpha A_\alpha}$; we show that $\forall\, \alpha: y_\alpha \in \overline{A}_\alpha$, that is, $\{y_\alpha\} \in \prod_\alpha \overline{A}_\alpha$. Let $y_\alpha \in U_\alpha$, where U_α is open in Y_α; since $\{y_\alpha\} \in \langle U_\alpha\rangle$, we must have (III, **4.3**)

$$\varnothing \neq \langle U_\alpha\rangle \cap \prod A_\alpha = (U_\alpha \cap A_\alpha) \times \prod \{A_\beta \mid \beta \neq \alpha\}$$

and so by I, **9**, Ex. 1, we find $U_\alpha \cap A_\alpha \neq \varnothing$. This proves $y_\alpha \in \overline{A}_\alpha$. The converse inclusion is established by reversing these steps: If $\{y_\alpha\} \in \prod_\alpha \overline{A}_\alpha$, then for any nbd $\{y_\alpha\} \in \langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle$, each $U_{\alpha_i} \cap A_{\alpha_i} \neq \varnothing$ so that $\langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle \cap \prod_\alpha A_\alpha \neq \varnothing$.

(3). This is left for the reader.

(4). Let $\langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle$ be any basic open set; since all but the $\alpha_1, \cdots, \alpha_n$ coordinates are unrestricted, this set contains a point having the coordinates y°_β for each $\beta \neq \alpha_1, \cdots, \alpha_n$. By III, **4.13**(3), D is dense.

For the case $\aleph(\mathscr{A}) < \aleph_0$, we have

1.3 In the space $\prod_{\iota=1}^{n} Y_\iota$:

$$(1).\quad \left(\prod_1^n A_\iota\right)' = (A_1' \times \overline{A}_2 \times \cdots \times \overline{A}_n) \cup (\overline{A}_1 \times A_2' \times \overline{A}_3 \times \cdots \times \overline{A}_\iota) \cup \cdots \cup (\overline{A}_1 \times \overline{A}_2 \times \cdots \times A_n').$$

$$(2).\quad \operatorname{Int}\left(\prod_1^n A_\iota\right) = \prod_1^n \operatorname{Int}(A_\iota).$$

$$(3).\quad \operatorname{Fr}\left(\prod_1^n A_\iota\right) = [\operatorname{Fr}(A_\iota) \times \overline{A}_2 \times \cdots \overline{A}_n] \cup [\overline{A}_1 \times \operatorname{Fr}(A_2) \times \cdots \times \overline{A}_n] \cup \cdots \cup [\overline{A}_1 \times \overline{A}_2 \times \cdots \times \operatorname{Fr}(A_n)].$$

Proof: (1) is proved by induction, from the result for two factors:

$$\begin{aligned}(a, b) \in (A \times B)' \Leftrightarrow (a, b) &\in \overline{(A \times B) - (a, b)}\\ &= \overline{[(A - a) \times B] \cup [A \times (B - b)]}\\ &= [\overline{(A - a)} \times \overline{B}] \cup [\overline{A} \times \overline{(B - b)}]\\ \Leftrightarrow (a, b) &\in [A' \times \overline{B}] \cup [\overline{A} \times B'],\end{aligned}$$

where we have used **1.2**(3) and III, **4**, Ex. 11.

(2). From I, **9.6** and **1.2**,

$$\begin{aligned}\operatorname{Int}\left(\prod_1^n A_\iota\right) &= \mathscr{C}[\overline{\mathscr{C}\prod A_\iota}] = \mathscr{C}\left[\overline{\bigcup_1^n \langle \mathscr{C}A_\iota\rangle}\right]\\ &= \mathscr{C}\bigcup_1^n \langle \overline{\mathscr{C}A_\iota}\rangle = \bigcap_1^n \langle \mathscr{C}\overline{\mathscr{C}A_\iota}\rangle\\ &= \prod_1^n \operatorname{Int}(A_\iota).\end{aligned}$$

(3). This is proved by induction and is left for the reader.

2. Continuity of Maps

2.1 Theorem Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces. Then for each fixed $\beta \in \mathscr{A}$, the projection $p_\beta\colon \prod_\alpha Y_\alpha \to Y_\beta$ is a continuous open surjection.

Proof: We have already seen (I, **9.3**) that p_β is surjective. If U is open in Y_β, the formula $\langle U \rangle = p_\beta^{-1}(U)$ shows that $p_\beta^{-1}(U)$ is open in $\prod_\alpha Y_\alpha$ and proves that p_β is continuous. Since $p_\beta\langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle = Y_\beta$ (if $\beta \neq \alpha_1, \cdots, \alpha_n$) or U_{α_k} (if $\beta = \alpha_k$), the image of any basic open set is open so, by III, **11.3**, p_β is open.

Ex. 1 p_β is not in general closed, even if $\aleph(\mathscr{A})$ is finite. In $E^2 = E^1 \times E^1$, the set $A = \{(x, y) \mid xy = 1\}$ is closed, but under projection on the first factor, $p_1(A) = \{x \mid x \neq 0\}$.

2.2 Theorem Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces, and $f\colon X \to \prod_\alpha Y_\alpha$ a map. Then f is continuous if and only if $p_\beta \circ f$ is continuous for each $\beta \in \mathscr{A}$.

Proof: Let f be continuous; since p_β is continuous for each β, so also (III, **8.2**) is $p_\beta \circ f$. Conversely, assume each $p_\beta \circ f$ continuous; then for each subbasic $\langle U_\beta \rangle$ in $\prod_\alpha Y_\alpha$, $f^{-1}\langle U_\beta \rangle$ is open because $f^{-1}\langle U_\beta \rangle = f^{-1}[p_\beta^{-1}(U_\beta)] = (p_\beta \circ f)^{-1}(U_\beta)$, and so (III, **8.3**) f is continuous.

For an $f\colon X \to \prod_\alpha Y_\alpha$, the map $p_\beta \circ f\colon X \to Y_\beta$ is called the βth coordinate function of f; in terms of its coordinate functions, f is the mapping $x \to \{p_\alpha \circ f(x)\}$. The importance of **2.2** is that it tells how to construct continuous maps into cartesian products.

2.3 Corollary Let X be a fixed space, and $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ any family of spaces. Assume that for each $\alpha \in \mathscr{A}$, a map $f_\alpha\colon X \to Y_\alpha$ is given. Then, defining $f\colon X \to \prod_\alpha Y_\alpha$ by $x \to \{f_\alpha(x)\}$, f is continuous if and only if each given f_α is continuous.

Proof: This is immediate from **2.2**, since $p_\beta \circ f = f_\beta$ for each $\beta \in \mathscr{A}$.

Ex. 2 The usual notion "parametric representation of curves" is essentially an application of **2.3**. Calling *curve* any continuous image of a closed interval in E^1, to say that the system $x = x(t)$, $y = y(t)$, $(x(t), y(t)$ continuous) represents a curve in E^2, we rely on **2.3** to assure that $t \to (x(t), y(t))$ is continuous.

Ex. 3 By **2.3** it is easy to establish III, **12**, Problem 4: Let $g: X \to G(f)$ be the map $x \to (x, f(x))$, and p_X, p_Y the projections of $X \times Y$ onto X, Y, respectively. If g is a homeomorphism, $f = p_Y \circ g$ and **2.1** show that f is continuous. Conversely, if f is continuous, by **2.3** the map $g: X \to G(f) \subset X \times Y$ is continuous. Defining $p = p_X \mid G(f)$, **2.1** shows that p is continuous, and since $p \circ g = 1$, $g \circ p = 1$, III **12.3** shows that g is a homeomorphism.

2.4 (Unrestricted Associativity) Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces and $\{\mathscr{A}_\mu \mid \mu \in \mathscr{M}\}$ a *partition* of $\mathscr{A}$. For each μ, let $Z_\mu = \prod \{Y_\alpha \mid \alpha \in \mathscr{A}_\mu\}$. Then $\prod_\alpha Y_\alpha \cong \prod_\mu Z_\mu$.

Proof: For each $\mu \in \mathscr{M}$, define $q_\mu: \prod_\alpha Y_\alpha \to Z_\mu$ to be the map $\{y_\alpha\} \to \{y_\alpha \mid \alpha \in \mathscr{A}_\mu\}$; by I, **9.3**, each q_μ is surjective. For each $\beta \in \mathscr{A}_\mu$, the map $p_\beta \circ q_\mu$ is the projection of $\prod_\alpha Y_\alpha$ onto its βth factor, so from **2.1** and **2.2**, each q_μ is also continuous.

Now define $q: \prod_\alpha Y_\alpha \to \prod_\mu Z_\mu$ by $y \to \{q_\mu(y)\}$; then, from **2.3**, q is continuous. The map q is also open: given any basic $\langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle$ in $\prod_\alpha Y_\alpha$, we have

$$q_\mu \langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle = \begin{cases} \langle U_{\alpha_j}, \cdots, U_{\alpha_k}\rangle & \text{if } \{\alpha_1, \cdots, \alpha_n\} \cap \mathscr{A}_\mu = \{\alpha_j, \cdots, \alpha_k\} \neq \varnothing \\ Z_\mu & \text{if } \{\alpha_1, \cdots, \alpha_n\} \cap \mathscr{A}_\mu = \varnothing \end{cases}$$

and from this it follows that $q\langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle$ is open in $\prod_\mu Z_\mu$. Finally, q is bijective because $\{\mathscr{A}_\mu \mid \mu \in \mathscr{M}\}$ is a partition of $\mathscr{A}$. By III, **12.2**, q is therefore a homeomorphism, and the proof is complete.

2.5 Theorem Let $\aleph(\mathscr{A})$ be arbitrary. For each $\alpha \in \mathscr{A}$, let $f_\alpha: X_\alpha \to Y_\alpha$ be a map. Define $\prod f_\alpha: \prod_\alpha X_\alpha \to \prod_\alpha Y_\alpha$ by $\{x_\alpha\} \to \{f_\alpha(x_\alpha)\}$. Then:

(1). If each f_α is continuous, so also is $\prod f_\alpha$.
(2). If each f_α is an open map, and all but at most finitely many are surjective, then $\prod f_\alpha$ is also an open map.

Proof: Ad (1). Let $\langle V_\alpha\rangle$ be a subbasic open set in $\prod_\alpha Y_\alpha$; then $(\prod f_\alpha)^{-1} \langle V_\alpha\rangle = \langle f_\alpha^{-1}(V_\alpha)\rangle$, which is open because f_α is continuous. By III, **8.3**, $\prod f_\alpha$ is continuous.

Ad (2). Each basic $\langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle$ maps to

$$f_{\alpha_1}(U_{\alpha_1}) \times \cdots \times f_{\alpha_n}(U_{\alpha_n}) \times \prod \{f_\beta(X_\beta) \mid \beta \neq \alpha_1, \cdots, \alpha_n\}.$$

Since all but at most finitely many f_β are surjective, all but at most finitely many $f_\beta(X_\beta) = Y_\beta$, and the rest are open sets. Thus the image of $\langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle$ is an open set, which by III, **11.3**, proves that $\prod f_\alpha$ is open.

2.6 Corollary Let $\{X_\alpha \mid \alpha \in \mathscr{A}\}$ and $\{Y_\beta \mid \beta \in \mathscr{B}\}$ be two families of spaces, and $\varphi: \mathscr{A} \to \mathscr{B}$ a bijection. If for each α, $X_\alpha \cong Y_{\varphi(\alpha)}$, then $\prod_\alpha X_\alpha \cong \prod_\beta Y_\beta$. In particular, $\prod_\alpha X_\alpha$ is unrestrictedly commutative.

Proof: This is immediate from **2.5** and III, **12.2.**

2.7 Let Y be a fixed space, $\mathscr{A}$ any indexing set with $\aleph(\mathscr{A}) \geqslant \aleph_0$. Let $Y_\alpha = Y$ for each α, and let $Z = \prod_\alpha Y_\alpha$. Then each cartesian product $Z \times Z \times \cdots \times Z$ consisting of $\leqslant \aleph(\mathscr{A})$ factors Z is homeomorphic to Z.

Proof: If there are $\aleph \leqslant \aleph(\mathscr{A})$ factors Z, the cardinal number of the factors Y present in $Z \times \cdots \times Z$ is $\aleph \cdot \aleph(\mathscr{A}) = \aleph(\mathscr{A})$, so **2.6** applies.

3. Slices in Cartesian Products

In $E^2 = E^1 \times E^1$, we can identify a line parallel to the x-axis with E^1; this idea extends to arbitrary cartesian products.

3.1 Definition Let $\prod_\alpha Y_\alpha$ be an arbitrary cartesian product, and $y^\circ = \{y^\circ_\alpha\}$ a given point. For each index β, the set

$$S(y^\circ; \beta) = Y_\beta \times \prod \{y^\circ_\alpha \mid \alpha \neq \beta\} \subset \prod_\alpha Y_\alpha$$

is called the slice in $\prod_\alpha Y_\alpha$ through y° parallel to Y_β.

Ex. 1 In $E^3 = E^1 \times E^1 \times E^1$, with $y^\circ = (x^\circ_1, x^\circ_2, x^\circ_3)$, the slice $S(y^\circ; 1) = \{(x, x^\circ_2, x^\circ_3) \mid x \in E^1\}$, and so is a line parallel to the x-axis going through y°.

3.2 Theorem The map $s_\beta: Y_\beta \to S(y^\circ; \beta)$ given by

$$y_\beta \to y_\beta \times \prod \{y^\circ_\alpha \mid \alpha \neq \beta\}$$

is a homeomorphism of Y_β with the subspace $S(y^\circ; \beta) = s$.

Proof: We first note that, because of I, **9.6**, the open sets

$$\langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle \cap S$$

of the subspace S are $\varnothing$, S, or $S \cap \langle U_\beta \rangle$, where U_β is open in Y_β. Thus $s_\beta^{-1}(\langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle \cap S) = \varnothing$, Y_β, or U_β, hence it is always open, and therefore s_β is continuous. Since the projection $p_\beta\colon \prod_\alpha Y_\alpha \to Y_\beta$ is continuous, so also is $p = p_\beta \mid S$; and because $p \circ s_\beta = 1$, $s_\beta \circ p = 1$, an application of III, **12.3** shows that both s_β and p are homeomorphisms.

4. Peano Curves

As one simple application of the results in this chapter, we prove the existence of space-filling curves.

For each $n = 1, 2, \cdots$, let A_n be the discrete space $\{0, 2\}$; in I, **9**, Ex. 3, we have seen that if C is the Cantor set, the map $\varphi\colon \prod_n A_n \to C$ given by

$$\{a_i\} \to \sum_1^\infty \frac{a_i}{3^i}$$

is a bijection of the set $\prod A_n$ onto the set C.

4.1 Taking $\prod A_n$ with the cartesian product topology, and $C \subset E^1$ with the subspace topology, $\varphi\colon \prod A_n \cong C$.

Proof: φ is continuous: Given any $c = \sum_i a_i 3^{-i} \in C$, and nbd $B(c; \varepsilon) \cap C$, choose N so large that $\sum_N^\infty 2 \cdot 3^{-i} < \varepsilon$; then $\varphi(\langle a_1, \cdots, a_N \rangle) \subset B(c; \varepsilon) \cap C$.

φ is open: Given any $x = \{a_i\}$ and nbd $\langle a_{i_1}, \cdots, a_{i_k} \rangle$ containing x, let $N = \max\{i_1, \cdots, i_k\}$ and $W = B(\varphi(x); 1/3^{N+1}) \cap C$; then

$$W \subset \varphi(\langle a_{i_1}, \cdots, a_{i_k} \rangle)$$

so III, **11.3**, applies. Being bijective, φ is a homeomorphism.

4.2 Let $I = [0, 1] \subset E^1$, and let $\psi\colon \prod A_n \to I$ be the map

$$\{a_i\} \to \sum_1^\infty \frac{a_i}{2^{i+1}}.$$

Then ψ is a continuous surjection.

Proof: Continuity follows as in the preceding proof and, by using dyadic expansions, the surjectivity is obvious.

The n-cube $I^n \subset E^n$ is $\prod_1^n I$; the space $I^\infty = \prod_1^\infty I$ is called the *Hilbert cube*. We now have

4.3 For each $k \leqslant \infty$, there exists a continuous surjection of the Cantor set on I^k.

Proof: Let $k \leqslant \infty$ be given. We have $\varphi^{-1}: C \cong \prod A_n$, and by **2.7**, an $h: \prod A_n \cong \prod A_n \times \cdots \times \prod A_n$ ($k \leqslant \infty$ factors). Using **2.5** and **4.2**, we obtain a continuous surjection:

$$\prod_1^k \psi: \prod A_n \times \cdots \times \prod A_n \to I \times \cdots \times I = I^k,$$

so that

$$\left(\prod_1^k \psi\right) \circ h \circ \varphi^{-1}: C \to I^k$$

is the desired map.

By a "curve" in a space Y is meant the image $f(I)$ of a continuous map $f: I \to Y$. From **4.3**, the existence of space-filling curves (that is, Peano curves) follows easily:

4.4 Theorem (Generalized Peano) For each $k \leqslant \infty$, there exists a continuous surjection $I \to I^k$ (that is, a curve going through each point of I^k).

Proof: Let $f: C \to I^k = I \times \cdots \times I$ be the continuous surjection of **4.3**, and let $p_i \circ f: C \to I$ be the coordinate functions. Recall that $C \subset I$ is obtained by successively dropping out middle thirds. Extend each $p_i \circ f: C \to I$ to a continuous $f_i: I \to I$ by defining f_i to be linear on each omitted interval. Then, by **2.3**, $F(t) = \{f_i(t)\}$ is a continuous map $I \to I^k$, and since $F \mid C = f$, it is surjective.

Problems

Section 1

1. Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces. Assume that each Y_α has a basis of cardinal number $\leqslant \aleph$. What is the cardinal of a basis for $\prod_\alpha Y_\alpha$?
2. Let $\aleph(\mathscr{A})$ be arbitrary and $\prod A_\alpha \subset \prod Y_\alpha$. If all but at most finitely many factors $A_\alpha = Y_\alpha$, prove $\text{Int}\,(\prod A_\alpha) = \prod\,[\text{Int}(A_\alpha)]$.
3. Let R be the real numbers with the upper-limit topology (III, **3**, Ex. 4). Show that $R \times R$ is not a discrete space, but that $A = \{(x, y) \mid x + y = 1\}$, as a subspace of $R \times R$, has the discrete topology.
4. Prove: $\prod A_\alpha$ is dense in $\prod Y_\alpha$ if and only if each $A_\alpha \subset Y_\alpha$ is dense.

Section 2

1. Prove: The cartesian product topology in $\prod_{\alpha} Y_\alpha$ is the smallest topology for which all projections $p_\beta : \prod_{\alpha} Y_\alpha \to Y_\beta$ are continuous.
2. Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces. For each $\mathscr{B} \subset \mathscr{A}$, let

$$p_{\mathscr{B}} : \prod \{Y_\alpha \mid \alpha \in \mathscr{A}\} \to \prod \{Y_\beta \mid \beta \in \mathscr{B}\}$$

be the projection. Let $A \subset \prod_{\alpha} Y_\alpha$ be closed. Prove:

$$A = \bigcap_{\mathscr{B}\text{ finite}} p_{\mathscr{B}}^{-1}[p_{\mathscr{B}}(A)].$$

Section 3

1. Let $\mathscr{S}$ be the Sierpinski space, and $\mathscr{S} \times \{0\}$ the slice in $\mathscr{S} \times \mathscr{S}$ parallel to the first factor. Is $\mathscr{S} \times \{0\}$ closed in $\mathscr{S} \times \mathscr{S}$?

Connectedness

V

Intuitively, a space is connected if it does not consist of two separate pieces. This simple idea has had important consequences in topology and has led to highly sophisticated algebraic techniques for distinguishing between spaces. In this chapter, we obtain the basic properties of this concept and those of some of its modifications.

1. Connectedness

1.1 Definition A space Y is connected if it is not the union of two nonempty disjoint open sets. A subset $B \subset Y$ is connected if it is connected as a sub*space* of Y.

Ex. 1 Sierpinski space $\mathscr{S}$ is connected: The only possible decomposition is 0, 1, and 1 is not open. The discrete space 2 is not connected.

Ex. 2 The real number system with the upper-limit topology (III, 3, Ex. 4) is not a connected space, since $\{x \mid x > a\}$ and $\{x \mid x \leqslant a\}$ are both open sets.

Ex. 3 The rationals $Q \subset E^1$ are not connected, since $\{x \mid x > \sqrt{2}\} \cap Q$, $\{x \mid x < \sqrt{2}\} \cap Q$ is a decomposition as required.

1.2 Theorem The only connected subsets of E^1 having more than one point are E^1 and the intervals (open, closed, or half-open).

Proof: Y connected implies that Y is an interval: For if Y is not an interval, then by definition there must be $a, b \in Y$, $c \bar{\in} Y$ with $a < c < b$, so $Y \cap \{x \mid x > c\}$, $Y \cap \{x \mid x < c\}$ is a decomposition of Y.

Y is an interval implies that Y is connected: If Y were not connected, then $Y = A \cup B$, where A, B are disjoint nonempty open sets, and there would be an $a \in A$, $b \in B$ that (relabeling if necessary) satisfy $a < b$. Define $\alpha = \sup \{x \mid [a, x[\subset A\}$; then $\alpha \leqslant b$, and because Y is an interval, $\alpha \in Y$. Clearly, $\alpha \in \bar{A}_Y$; noting that $A = \mathscr{C}_Y B$ is closed in Y, we must have $\alpha \in A$. However, A is also open in Y, and since Y is an interval, an application of III, **2.3** shows there must be a nondegenerate $]\alpha - h, \alpha + h[\subset A$, which contradicts the definition of α.

The definition **1.1** can be formulated in handier fashion:

1.3 The following three properties are equivalent:

(1). Y is connected.
(2). The only subsets of Y both open and closed are $\varnothing$ and Y.
(3). No continuous $f: Y \to 2$ is surjective.

Proof: (1) $\Rightarrow$ (2). If $G \subset Y$ is both open and closed, and $G \neq \varnothing, Y$, then $Y = G \cup \mathscr{C}G$ shows that Y is not connected.

(2) $\Rightarrow$ (3). If $f: Y \to 2$ were a continuous surjection, then $f^{-1}(0) \neq \varnothing, Y$, and because 0 is open and closed in 2, $f^{-1}(0)$ is open and closed in Y.

(3) $\Rightarrow$ (1). If $Y = A \cup B$, A, B disjoint nonempty open sets, then A, B are also closed, and the characteristic function $c_A: Y \to 2$ is a continuous surjection.

Connectedness is clearly a topological invariant; even more,

1.4 Theorem The continuous image of a connected set is connected. That is, if X is connected and $f: X \to Y$ is continuous, then $f(X)$ is connected.

Proof: The map $f: X \to f(X)$ is continuous; if $f(X)$ were not connected, there would be, by **1.3**, a continuous surjection $g: f(X) \to 2$, and then $g \circ f: X \to 2$ would also be a continuous surjection, contradicting the connectedness of X.

1.5 Theorem Let Y be any space. The union of any family of connected subsets having at least one point in common is also connected.

Proof: Let $C = \bigcup_\alpha A_\alpha$, $y_0 \in \bigcap_\alpha A_\alpha$, and $f: C \to 2$ continuous. Since each A_α is connected, no $f \mid A_\alpha$ is surjective, and because $y_0 \in A_\alpha$ for each α, $f(y) = f(y_0)$ for all $y \in A_\alpha$ and all α. Thus f cannot be surjective.

Ex. 4 In contrast, the intersection of even two connected sets need not be connected. Furthermore, if all the $A_i, i \in Z^+$ are connected, and $A_1 \supset A_2, \cdots$, still $\bigcap_n A_n$ need not be connected: let $Y = I^2 - \{(x, 0) \mid \frac{1}{3} \leqslant x \leqslant \frac{2}{3}\}$ and $A_n = \{(x, y) \in Y \mid y \leqslant 1/n\}$.

1.6 Theorem Let $A \subset Y$ be connected. Then any set B satisfying $A \subset B \subset \overline{A}$ is also connected. In particular, the closure of a connected set is connected.

Proof: Let f: $B \to 2$ be continuous; since A is connected, $f \mid A$ is not surjective. Noting that $B = \overline{A} \cap B = \overline{A}_B$, the continuity of f on B shows $f(B) = f(\overline{A}_B) \subset \overline{f(A)} = f(A)$, so that f cannot be surjective.

Ex. 5 Since $Y = \{(x, y) \mid y = \sin 1/x,\ 0 < x \leqslant 1\} \subset E^2$ is a continuous image of $]0, 1]$, it follows from **1.4** and **1.6** that $\overline{Y} = Y \cup \{(0, y) \mid -1 \leqslant y \leqslant 1\}$ is connected. Observe that even with omission of any subset of $\{(0, y) \mid -1 \leqslant y \leqslant 1\}$, the resulting set is still connected.

1.7 Theorem Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces. $\prod_\alpha Y_\alpha$ is connected if and only if each Y_α is connected.

Proof: The connectedness of $\prod_\alpha Y_\alpha$ implies that of each Y_α, since each projection $p_\beta: \prod_\alpha Y_\alpha \to Y_\beta$ is a continuous surjection and **1.4** applies. For the converse, let $y^\circ \in \prod_\alpha Y_\alpha$. We first show that

(1). If $y^\circ_{(n)}$ and y° differ by at most $n < \infty$ coordinates, then $y^\circ_{(n)}$ and y° lie in a connected set.

The proof is by induction on the number n of differing coordinates. Assertion (1) is true for $n = 1$, because if $y^\circ_{(1)}$ and y° differ in the αth coordinate, the slice $S(y^\circ; \alpha)$ through y° parallel to the αth factor, being homeomorphic to Y_α, is a connected set containing y° and $y^\circ_{(1)}$. Now, assume (1) to be true for all $y^\circ_{(n-1)}$. Then given any $y^\circ_{(n)}$, find a $y^\circ_{(n-1)}$ that differs from it by one coordinate. By the case $n = 1$, $y^\circ_{(n)}$ and $y^\circ_{(n-1)}$ lie in a connected subset C, and by the inductive hypothesis, $y^\circ_{(n-1)}$ and y° lie in a connected C_1; since $C \cap C_1$ contains $y^\circ_{(n-1)}$, $C \cup C_1$ is connected, completing the inductive step.

Now let A be the union of all connected subsets of $\prod_\alpha Y_\alpha$ containing y°; by **1.5**, A is connected and, by (1), contains $D = \{y \mid y \text{ and } y^\circ \text{ differ in at most finitely many coordinates}\}$. But [IV, **1.2**(c)] D is dense in $\prod_\alpha Y_\alpha$; since $\overline{D} \subset \overline{A}$, and, by **1.6**, $\overline{A}$ is connected, the theorem is proved.

2. Applications

We obtain a generalization of the "intermediate value theorem" of analysis.

2.1 Theorem Each continuous real-valued function on a connected space X takes on all values between any two that it assumes.

Proof: Since $f: X \to E^1$ is continuous, $f(X) \subset E^1$ is connected according to **1.4**, so by **1.2**, $f(X)$ is an interval. Thus, if $f(x) = a$, $f(x') = b$, we have $[a, b] \subset f(X)$, and therefore for each c such that $a \leqslant c \leqslant b$, there is an x'' with $f(x'') = c$.

From **1.2** and **1.7** follows that E^n, I^n, and I^∞ are connected; even more,

2.2 Theorem Let $n > 1$, and $B \subset E^n$ be countable. Then $E^n - B$ is connected.

Proof: We can assume that $0 \bar{\in} B$, otherwise we move the origin. According to **1.5**, it suffices to show that the origin and each $x \in E^n - B$ are contained in a connected set lying in $E^n - B$. Draw $\overrightarrow{0x}$ and let l be any line segment (say, of length 1) intersecting $\overrightarrow{0x}$ at exactly one point, distinct from 0 and x. For each $z \in l$, let $l_z = \overrightarrow{0z} \cup \overrightarrow{zx}$; each l_z is a connected set, and any two have only 0 and x in common. At least one l_z must lie in $E^n - B$: for if $l_z \cap B \neq \varnothing$ for each $z \in l$, then since the points of intersection for differing z are necessarily distinct, we would find that B has a subset in 1-to-1 correspondence with the points of l and consequently B would not be countable.

The usual technique for distinguishing between spaces stems from the observation: If $h: X \cong Y$, then by removing a set A of prescribed topological type from X, the spaces $X - A$ and $Y - h(A)$ are also homeomorphic, so that they must have the same topological invariants.

2.3 Theorem E^1 and E^n, $n > 1$, are not homeomorphic.

Proof: Assume that $h: E^n \cong E^1$; removing one point $a \in E^n$, we must have $h: E^n - a \cong E^1 - h(a)$, by III, **12.4**. However, this is impossible since, by **2.2**, $E^n - a$ is connected whereas $E^1 - h(a)$ is not.

The theorem that E^n is not homeomorphic to E^m for $n \neq m$ is much deeper, involving more delicate topological invariants (although the technique is the same). Concerning I^n ($n > 1$) and I^1, a proof similar to **2.3** shows that they are not homeomorphic. Thus, though there is a bijective map of the *set* I^1 onto the set I^n, there is no bicontinuous bijection and, as we shall see later, not even a continuous bijection (XI, **2**, Ex. 4).

2.4 Theorem In E^1 each closed interval is homeomorphic to $[-1, +1]$, each open interval to $]-1, +1[$, and each half-open interval to $]-1, +1]$. Furthermore, no two of these three intervals are homeomorphic.

Proof: Given an interval with end points a, b, a suitable one of the maps $x \to \dfrac{b+a}{2} \pm \dfrac{b-a}{2}x$ exhibits a homeomorphism. To see that none of the three standard intervals are homeomorphic, note that we can remove 2, 0, 1 (respectively) points without destroying the connectedness.

3. Components

A disconnected space can be decomposed uniquely into connected "components"; the number of components provides a rough indication of how "disconnected" a space is.

3.1 Definition Let Y be a space, and $y \in Y$. The component $C(y)$ of y in Y is the union of all connected subsets of Y containing y.

It is evident from **1.5** that $C(y)$ is connected.

Ex. 1 Let $Q \subset E^1$ be the subspace of rationals. The component of each $y \in Q$ is the point y itself. Thus, even though Y does not have the discrete topology, the components may reduce to points. Y is called *totally disconnected* if $C(y) = y$ for each $y \in Y$.

Ex. 2 Let $Y \subset E^2$ be the subspace consisting of the segments joining the origin 0 to the points $\{(1, 1/n) \mid n \in Z^+\}$, together with the segment $]\frac{1}{2}, 1]$ on the x-axis. As in **1**, Ex. 5, Y is connected, but $Y - \{0\}$ is not: in $Y - \{0\}$ the component of each point is the ray containing it.

Ex. 3 In $\prod_\alpha Y_\alpha$, the component of $y = \{y_\alpha\}$ is $C(y) = \prod_\alpha C(y_\alpha)$, where $C(y_\alpha)$ is the component of the coordinate y_α in Y_α. Indeed, by **1.7**, $C(y)$ is connected, so it is contained in the component K of y. If $K - C(y) \neq \varnothing$, select $y^\circ \in K - C(y)$; then some coordinate $y^\circ_\alpha \bar{\in} C(y_\alpha)$, so that $p_\alpha(K) \subset Y_\alpha$ is connected and contains $y^\circ_\alpha \bar{\in} C(y_\alpha)$, in contradiction to $C(y_\alpha)$ being the component of y_α. In particular, though the infinite cartesian product of discrete spaces is never discrete (IV, **1**, Ex. 3) it is always totally disconnected.

3.2 Theorem (1). Each component $C(y)$ is a maximal connected set in Y: there is no connected subset of Y that properly contains $C(y)$.

(2). The set of all distinct components in Y forms a *partition* of Y.

(3). Each $C(y)$ is closed in Y.

Proof: (1) follows from the definition.

(2). If $C(y) \cap C(y') \neq \varnothing$, then by **1.5**, $C(y) \cup C(y')$ is connected, contradicting the maximality of $C(y)$.

(3). Since $C(y)$ is connected, so also (**1.6**) is $\overline{C(y)}$; by the maximality of $C(y)$, we must have $\overline{C(y)} \subset C(y)$, so that $C(y)$ is closed.

Ex. 4 Components need *not* be sets *open* in Y, as Ex. 1 and Ex. 3 show. In particular, a splitting of a disconnected space into disjoint nonempty open sets is not generally accomplished simply by taking any one component as one of the "pieces."

The number and structure of each component of a space Y is a topological invariant:

3.3 Theorem Let $f: X \to Y$ be continuous. Then the image of each component of X must lie in a component of Y. Furthermore, if $h: X \cong Y$, then h induces a 1-to-1 correspondence between the components of X and those of Y, corresponding ones being homeomorphic.

Proof: If f is continuous, then $f(C(x)) \subset C(f(x))$ follows from **1.4**, since $f(C(x))$ is a connected set in Y containing $f(x)$. If $h: X \cong Y$, then because h is bicontinuous and bijective, we have both $h(C(x)) \subset C(h(x))$ and $h^{-1}(C(h(x))) \subset C(x)$, which shows that $h(C(x)) = C(h(x))$. The rest of the proof is trivial.

Ex. 5 The Cantor set is totally disconnected, by Ex. 3 and **3.3**.

Ex. 6 Components are also used to decide questions of nonhomeomorphism. As an application, we show that a strong version of the Bernstein-Schröder theorem is not valid for topological spaces: There may be a continuous bijection $f: X \to Y$ and a continuous bijection $g: Y \to X$; yet, X and Y need not be homeomorphic (C. Kuratowski). In E^1, let

$$X =]0, 1[\cup \{2\} \cup]3, 4[\cup \{5\} \cup \cdots \cup]3n, 3n + 1[\cup \{3n + 2\} \cup \cdots,$$
$$Y =]0, 1] \cup]3, 4[\cup \{5\} \cup \cdots \cup]3n, 3n + 1[\cup \{3n + 2\} \cup \cdots.$$

These spaces cannot be homeomorphic, since the component $]0, 1]$ is not homeomorphic to any component of X. However, $f(x) = x$ $(x \neq 2)$, $f(2) = 1$ is a continuous bijection $X \to Y$, and $g: Y \to X$, where

$$g(x) = \begin{cases} x/2 & \text{if } x \in]0, 1] \\ (x - 2)/2 & \text{if } x \in]3, 4[. \\ x - 3 & \text{otherwise} \end{cases}$$

is also a continuous bijection.

4. Local Connectedness

4.1 Definition A space Y is locally connected if it has a basis consisting of connected (open) sets.

Ex. 1 E^n is locally connected, since each ball $B(x; r)$ is connected. Furthermore, each interval in E^1 is locally connected. For each $n \geq 0$, S^n is locally connected.

Ex. 2 A space may be locally connected, but not connected, as the discrete space 2 shows.

Ex. 3 A space may be connected, but not locally connected. Let Y be the space of **3**, Ex. 2, $y = (\frac{3}{4}, 0)$ and $U = B(y; \frac{1}{2}) \cap Y$. Then U, and any nbd $V(y) \subset U$, is not connected: For V must intersect a ray joining 0 to some $(1, 1/n)$ and it is trivial to verify that this intersection is both open and closed in V. Thus no basis for Y can consist only of connected sets.

4.2 Theorem Y is locally connected if and only if the components of each *open* set are open sets.

Proof: Let $G \subset Y$ be open, C a component of G, and $\{U\}$ a basis consisting of connected open sets. Given $y \in C$, then because $y \in G$, there is a U with $y \in U \subset G$; but since C is the component of y and U is connected, $y \in U \subset C$, showing that C is open (III, **2.3**). For the converse, note that the family of all components of all open sets in Y is a basis as required.

Ex. 4. Observe that **4.2** need not be true for nonopen sets, as $\{0\} \cup \{1\} \subset E^1$ shows.

Ex. 5. Let Y be the space of **1**, Ex. 5, and $Z = Y \cup \{0\}$. Then Z is not locally connected, since the components of $Z \cap \{(x, y) \mid y < \frac{1}{2}\}$ are not open in Z.

4.3 A cartesian product $\prod_{\alpha} Y_\alpha$ is locally connected if and only if all the Y_α are locally connected, and all but at most finitely many are also connected.

Proof: Assume that $\prod_{\alpha} Y_\alpha$ is locally connected. Then: (1) All but at most finitely many Y_α are connected: if V is any connected open set in $\prod_{\alpha} Y_\alpha$, we have (IV, **1**, Ex. 4) that $p_\alpha(V) = Y_\alpha$ for all but at most finitely many α, and projections are continuous. (2) Each Y_β is locally connected: given $y_\beta \in Y_\beta$ and an open $U \subset Y_\beta$ containing y_β, then $\langle U \rangle$ is a nbd of some y having βth coordinate y_β, so there is an open connected V with $y \in V \subset \langle U \rangle$; since $p_\beta(V)$ is an open connected set, and $y_\beta \in p_\beta(V) \subset U$, this shows that Y_β has a basis consisting of connected sets.

To prove the converse, let $\mathscr{B} \subset \mathscr{A}$ be the finite set of indices for which the space Y_α is not connected. Now let $x \in \prod_\alpha Y_\alpha$ and let V be any open set containing x. Then $x \in \langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle \subset V$ for suitable $\alpha_1, \cdots, \alpha_n$. For each α_i let V_{α_i} be a connected nbd of $p_{\alpha_i}(x)$ such that $V_{\alpha_i} \subset U_{\alpha_i}$; and for each $\beta_1, \cdots, \beta_k \in \mathscr{B} - \{\alpha_1, \cdots, \alpha_n\}$, let V_{β_i} be a connected nbd of $p_{\beta_i}(x)$. Then $\langle V_{\alpha_1}, \cdots, V_{\alpha_n}, V_{\beta_1}, \cdots, V_{\beta_k} \rangle$ is a connected (*cf.* **1.7**) nbd of x and is contained in $\langle U_{\alpha_1}, \cdots U_{\alpha_n} \rangle$. Thus, $\prod_\alpha Y_\alpha$ has a basis consisting of connected open sets, so it is locally connected.

Ex. 6 The hypothesis that all but at most finitely many Y_α be connected is essential: If $A_n = \{0, 2\}$, we have seen $\prod_n A_n$ is totally disconnected. But though each A_n is locally connected, $\prod_n A_n$ is not: Its components are its points, and since $\prod_n A_n$ is not discrete, none is an open set.

Ex. 7 Local connectedness is evidently a topological invariant, and therefore it can be used in questions of nonhomeomorphism. Thus the space Z of Ex. 5 cannot be homeomorphic to any interval in E^1.

Ex. 8 Local connectedness is not invariant under continuous maps. Let X be the discrete space $\{0, 1, 2, \cdots\}$, Y the subspace $0 \cup \{1/n \mid n = 1, 2, \cdots\}$ of E^1 and $f: X \to Y$ the map $f(0) = 0$, $f(n) = 1/n$. Then X is locally connected and f is a continuous bijection; but Y is not locally connected.

5. Path-Connectedness

For most purposes of analysis, the natural notion of connectedness is joining by paths.

In **IV**(4), we have defined a curve in a space Y to be a continuous image of the unit interval I. A *path* in Y is a continuous mapping $f: I \to Y$, rather than the image $f(I)$ in Y. Thus, a path is a continuous function, whereas a curve is a subset of Y; we shall see later (Chapter XIX) one reason for this distinction between paths and curves. If $f: I \to Y$ is a path in Y, we call $f(0) \in Y$ the initial (or starting) point, and $f(1) \in Y$ the terminal (or end) point, of the path f, and say that f runs from $f(0)$ to $f(1)$, or joins $f(0)$ to $f(1)$. If f runs from $f(0)$ to $f(1)$, it is clear that the mapping $t \to f(1 - t)$, $t \in I$, is a path in Y running from $f(1)$ to $f(0)$.

5.1 Definition A space Y is path-connected (or: pathwise connected) if each pair of its points can be joined by a path.

Ex. 1 E^n and $S^n (n \geqslant 1)$ are path-connected. For any countable $B \subset E^n$, $E^n - B$ is also path-connected.

Ex. 2 Sierpinski space $\mathscr{S}$ is path-connected: The characteristic function of $1 \in I$, regarded as a map $I \to \mathscr{S}$, is a path joining 0 to 1.

Ex. 3 A discrete space having more than one point is never path-connected. Every indiscrete space is path-connected.

A trivial, but useful, reformulation of **5.1** is given in

5.2 Let Y be a space, and $y_0 \in Y$ any element. Y is path-connected if and only if each $y \in Y$ can be joined to y_0 by a path.

Proof: If Y is path-connected, the condition is trivially true. Conversely, assume that the condition is satisfied and that $y, y' \in Y$ have been given. Let $f: I \to Y$ run from y to y_0 and $g: I \to Y$ from y_0 to y'; then

$$\varphi(t) = \begin{cases} f(2t) & 0 \leqslant t \leqslant \frac{1}{2} \\ g(2t-1) & \frac{1}{2} \leqslant t \leqslant 1 \end{cases}$$

is continuous [because at $t = \frac{1}{2}$, we have $f(1) = g(0) = y_0$ (cf. III, **9.4**)] and is a path running from y to y'.

The general relation of connectedness and path-connectedness is

5.3 Theorem Each path-connected space is connected. But a connected space need not be path-connected.

Proof: Since the continuous image of I is connected, the assertion follows from **1.5** and **5.2**. The following example shows that the converse is not generally true.

Ex. 4 Let Y be the space of **1**, Ex. 5; we have seen that $\overline{Y}$ is connected. However, $\overline{Y}$ is not path-connected: there is no path joining $(0, 0)$ to the point $(1/\pi, 0)$: For, note first that by IV, **2.2**, an $f: I \to \overline{Y} \subset E^2$ is continuous if and only if both $p_1 \circ f$ and $p_2 \circ f$ are. If $p_1 \circ f$ is continuous, it must take on all the values $1/n\pi$, $n = 1, 2, \cdots$, so that $p_2 \circ f$ must assume the values $+1$, -1 is each nbd of $0 \in I$; thus there can be no nbd $[0, \delta[$ mapped into $]-\frac{1}{2}, \frac{1}{2}[$ by $p_2 \circ f$, so $p_2 \circ f$ cannot be continuous.

It is evident that path-connectedness is a topological invariant: Indeed, the continuous image of a path-connected space is path-connected. Furthermore, the union of any family of path-connected spaces having a point in common is, by **5.2**, also path-connected. However, the *closure* of a path-connected set need not be path-connected: in Ex. 4, Y is path-connected.

Because of the property of unions, we can define path-connected components as maximal path-connected subsets; as before, the path components partition the space; indeed, from **5.3**, the path components partition the components. However, the path components need not be *closed* subsets of the space: in Ex. 4, Y is a path component of $\overline{Y}$.

To determine when path-connectedness and connectedness are equivalent, we need

5.4 The following two properties of a space Y are equivalent:

(1). Each path component is open (and therefore also closed).
(2). Each point of Y has a path-connected nbd.

Proof: (1) $\Rightarrow$ (2) is clear, using the path component containing the given point.

(2) $\Rightarrow$ (1). Let K be any path component, and let $x \in K$. Since x has a path-connected nbd U, and since K is a maximal path-connected set containing x, $x \in U \subset K$, proving that K is open. Noting that $\mathscr{C}K$ is the union of the remaining (open) path components, K is also closed.

5.5 Theorem Y is path-connected if and only if it is connected, and each $y \in Y$ has a path-connected nbd.

Proof: Since path-connectedness implies connectedness, and Y is a path-connected nbd of each point, only the converse requires proof. For this, we find from **5.4**(1) that each path component is both open and closed in Y; since Y is connected, this path component must therefore be Y.

This yields the important

5.6 Corollary An open set in E^n (or S^n) is connected if and only if it is path-connected.

Proof: If $U \subset E^n$ is open, each point x of the *space* U has a nbd $B(x; r) \subset U$, and $B(x; r)$ is path-connected. The proof for S^n is similar.

Of course, as Ex. 4 shows, **5.6** need not be true for nonopen subsets of E^n.

Problems

Section 1

1. Show that a discrete space having more than one point is never connected and that a space having indiscrete topology is always connected.
2. Show that the extended real line $\tilde{E}^1$ is connected.

3. Let X be an infinite set, with topology $\mathscr{T} = \{\varnothing\} \cup \{A \mid \mathscr{C}A$ is finite$\}$. Show that X is connected.
4. Let $(X, \mathscr{T})$ be connected, and $\mathscr{T}_1 \subset \mathscr{T}$. Prove that $(X, \mathscr{T}_1)$ is connected.
5. Let $\{A_i \mid i \in Z^+\}$ be connected sets in Y, with $A_i \cap A_{i+1} \neq \varnothing$ for each i. Prove: $\bigcup_i A_i$ is connected.
6. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of connected subsets of Y, and assume that there exists a connected set A with $A \cap A_\alpha \neq \varnothing$ for each A_α. Show that $A \cup \bigcup_\alpha A_\alpha$ is connected.
7. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of connected sets. Assume that any two of them have nonempty intersection. Prove that $\bigcup_\alpha A_\alpha$ is connected.
8. Prove:
   a. Y is connected if and only if every open covering $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ of Y has the following property: For each pair of sets $U_{\alpha_1}, U_{\alpha_n}$, there are finitely many $U_{\alpha_2}, \cdots, U_{\alpha_{n-1}}$ such that $U_{\alpha_i} \cap U_{\alpha_{i+1}} \neq \varnothing$, $i = 1, \cdots, n-1$.
   b. Y is connected if and only if every nbd-finite closed covering has the same property as in (a).
9. (a) Let Y be a space and $A \subset Y$ any subset. Let $C \subset Y$ be connected, containing points of A and points not in A. Prove: C must contain points of the boundary of A. [*Hint*: Recall that $Y = \text{Int}(A) \cup \text{Fr}(A) \cup \text{Int}(\mathscr{C}A)$.]
   (b) Why is
   $$A = \{(x, y, 0) \in E^3 \mid x^2 + y^2 \leqslant 1\}$$
   and
   $$C = \{(0, 0, z) \mid -1 \leqslant z \leqslant 1\} \subset E^3$$
   not a counterexample to this result?
10. For each pair of positive integers a, b, let $U(a, b) = \{an + b \mid n \in Z\} \cap Z^+$. Prove that
   $\{U(a, b) \mid$ all (a, b) such that a is relatively prime to $b\}$ is a basis for a topology $\mathscr{T}$ in Z^+. Using this topology, show:
   a. For each prime p, the set $\{kp \mid k \in Z^+\}$ is closed in Z^+.
   b. If P is the set of all primes, then $\text{Int}(P) = \varnothing$.
   c. $(Z^+, \mathscr{T})$ is connected. [*Hint:* Show that if W is open and if $W \cap U(a, b) = \varnothing$, then no multiple of a can belong to W.]

Section 2

1. Prove: S^n is connected for all $n \geqslant 1$.
2. Prove: I is not homeomorphic to S^1; and also $[0, 2\pi[$ is not homeomorphic to S^1.
3. Show that E^1 is not homeomorphic to $\tilde{E}^1$.
4. Prove that S^n and S^1 are not homeomorphic.
5. Let $n \geqslant 2$, and in E^n, show:
   a. If $A_1 = \{x \in E^n \mid$ all coordinates of x are rational$\}$, then $E^n - A_1$ is connected.
   b. If $A_2 = \{x \in E^n \mid$ all coordinates of x are irrational$\}$, then A_2 is not connected.
   c. If $A_3 = \{x \in E^n$ at least one coordinate is irrational$\}$, then A_3 is connected.

6. Show: If $A \subset E^n$ is connected and non-empty then $\aleph(A) = \mathfrak{c}$ or 1.
7. Let $\mathscr{A} \subset \mathscr{P}(E^2)$ be the family of all connected sets. Find $\aleph(\mathscr{A})$.
8. In **1**, Problem 3, show that the only continuous real-valued functions on X are constant functions.
9. Show that the analogue of the Bernstein-Schröder theorem for topological spaces is not valid, by exhibiting two spaces X, Y such that X is homeomorphic to a subset of Y and Y is homeomorphic to a subset of X, although X and Y are not homeomorphic. (*Hint*: Use **2.4.**)

Section 3

1. In a space X, define $x \sim y$ if x and y are contained in a connected set. Show that this is an equivalence relation. What are the equivalence classes?
2. Let $A \subset Y$, where both A and Y are connected. Let U be any set both open and closed in $Y - A$. Prove that $A \cup U$ is connected.
3. Let $A \subset Y$, where both A and Y are connected. Let C be any component of $Y - A$. Show that $Y - C$ is connected.
4. Let $B \subset Y$ be a connected set both open and closed in Y. Prove: B is a component of Y.
5. In a space X, define $x \sim y$ if there is no decomposition of X into two disjoint open sets, one of which contains x, and the other y.
   a. Prove that this is an equivalence relation in X. The equivalence classes are called the quasi-components of X.
   b. Prove each quasi-component is the intersection of all the open closed sets containing a given element.
   c. Prove: Each component is contained in a quasi-component.
   d. In E^2 let L_1 be the line $x = 1$ and L_2 the line $x = -1$. For each $n \in Z^+$, let R_n be the rectangle $\{(x, y) \mid |x| \leqslant n/(n+1), |y| \leqslant n\}$. Finally, let Y be the subspace $L_1 \cup L_2 \cup \bigcup_n \mathrm{Fr}(R_n)$ of E^2. Show: The component of $(1, 0)$ is L_1 and the quasi-component of $(1, 0)$ is $L_1 \cup L_2$.
   e. Show that if x_1, x_2 (resp. y_1, y_2) belong to a quasi-component of X (resp. Y), then (x_1, y_1) and (x_2, y_2) belong to a quasi-component of $X \times Y$.

Section 4

1. Prove that an open set in E^n can have at most countably many components. Give an example to show that this is no longer true for closed sets.
2. Let Y be locally connected and let U be a component of the open $G \subset Y$. Show $G \cap \mathrm{Fr}(U) = \varnothing$.
3. Let Y be locally connected and $A \subset Y$ arbitrary. Let C be a component of A. Prove:
   a. $\mathrm{Int}(C) = C \cap \mathrm{Int}(A)$,
   b. $\mathrm{Fr}(C) \subset \mathrm{Fr}(A)$,
   c. If A is closed, then $\mathrm{Fr}(C) = C \cap \mathrm{Fr}(A)$.
4. Let Y be locally connected, and $A \subset Y$ arbitrary. If $\mathrm{Fr}(A)$ is locally connected, prove that $\bar{A}$ is locally connected.
5. Let Y be locally connected, $Y = A \cup B$, where A, B are closed and $A \cap B$ is locally connected. Prove that both A and B are locally connected.

6. Let Y be locally connected, $A \subset Y$ arbitrary. Let $S \subset A$ be connected and open in A. Show $S = A \cap U$, where U is connected and open in Y.
7. Let Y be locally connected, but not connected. Show that a decomposition of Y into two nonempty disjoint open sets can always be accomplished by taking any component as one of the sets, and all the rest as the other set.

Section 5

1. Show: If any one of the conditions in **5.4** holds, then the path components of Y coincide with the components of Y.
2. Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces. Show $\prod_\alpha Y_\alpha$ is path-connected if and only if each Y_α is path-connected.
3. Let X be the connected set in **I**, Problem 3. Show that X is totally pathwise disconnected. (By definition, X is called totally pathwise disconnected if the only continuous maps $f: I \to X$ are the constant maps.)

Identification Topology; Weak Topology

VI

In this chapter we consider two methods for topologizing sets that play an increasingly important rôle in topology. The first uses a map of a space into a set to topologize the set: it makes precise numerous constructions, such as identifying sets to points. The second constructs a space by "pasting" given spaces together along preassigned subsets. Finally, we obtain a relation between these two processes.

1. Identification Topology

1.1 Definition Let Y be an arbitrary set, X a topological space, and $p: X \to Y$ a surjection. The identification topology in Y determined by p is $\mathscr{T}(p) = \{U \subset Y \mid p^{-1}(U) \text{ is open in } X\}$.

$\mathscr{T}(p)$ is indeed a topology, since p^{-1} preserves set operations; and since p^{-1} preserves complementation, a $B \subset Y$ is closed if and only if $p^{-1}(B)$ is closed in X.

Ex. 1 Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces. Using the cartesian product topology in $\prod_\alpha Y_\alpha$, the identification topology $\mathscr{T}(p_\beta)$ in the *set* Y_β, determined by the projection $p_\beta\colon \prod_\alpha Y_\alpha \to Y_\beta$, is precisely the original topology in Y_β, since $p_\beta^{-1}(U)$ is open if and only if U is open in Y_β.

Ex. 2 Let $p\colon I \to \{0\} \cup \{1\}$ be the characteristic function of $[\frac{1}{2}, 1]$. The identification topology in $\{0\} \cup \{1\}$ makes it into Sierpinski space. Observe that with the topology $\mathscr{T}(p)$ in Y, the map p need *not* be either an open or a closed map.

Ex. 3 The requirement that p be surjective is not essential. However, note that if $Y - p(X) \neq \varnothing$, then $Y - p(X)$ receives the discrete topology, since $Q \subset Y - p(X) \Rightarrow p^{-1}(Q) = \varnothing \Rightarrow Q$ is both open and closed.

1.2 $\mathscr{T}(p)$ is the largest topology in Y for which $p\colon X \to Y$ is continuous.

Proof: Continuity of p when $\mathscr{T}(p)$ is used is immediate from the definition; if $\mathscr{T}$ is any other such topology, then $U \in \mathscr{T} \Rightarrow p^{-1}(U)$ is open in $X \Rightarrow U \in \mathscr{T}(p)$.

For ease of reference to the situation in **1.1**,

1.3 Definition Let X and Y be two spaces. A continuous surjection $p\colon X \to Y$ is called an *identification map* (or identification) whenever the topology in Y is exactly $\mathscr{T}(p)$, that is, the sets U open in Y are all those, and only those, for which $p^{-1}(U)$ is open in X.

Clearly, not every continuous surjection is an identification: indeed, the continuous bijection $i\colon (X, \mathscr{T}_1) \to (X, \mathscr{T}_2)$ is an identification if and only if $\mathscr{T}_1 = \mathscr{T}_2$. However, the concept of identification map contains that of open and of closed continuous surjections:

1.4 If $p\colon X \to Y$ is a continuous open (or closed) surjection, then p is an identification.

Proof: Let $\mathscr{T}$ be the topology in Y; since p is continuous, **1.2** shows that $\mathscr{T} \subset \mathscr{T}(p)$; for the converse inclusion in case p is an open map, let $U \in \mathscr{T}(p)$; then $p^{-1}(U)$ is open in X, and since p is open and surjective, the formula $U = p(p^{-1}(U))$ shows that $U \in \mathscr{T}$. For closed maps, the proof is similar.

Another frequently used sufficient condition is the simple

1.5 Let $p\colon X \to Y$ be continuous. If there is a continuous $s\colon Y \to X$ such that $p \circ s = 1_Y$, then p is an identification.

Proof: Clearly, p is surjective (I, **6.9**). Let $A \subset Y$ be any set for which $p^{-1}(A)$ is open; then $s^{-1}p^{-1}(A)$ is open in Y; since $s^{-1}p^{-1} = (p \circ s)^{-1} = 1_Y$, this shows that A is open in Y, and so p is an identification.

It is of importance to determine when an identification $p\colon X \to Y$ is an open (or closed) map. To this end, a set $A \subset X$ is called p-saturated whenever it is the complete inverse image of some set in Y. Note that A is p-saturated if and only if $A = p^{-1}pA$: for if the formula holds, then A is the inverse image of pA; and if $A = p^{-1}(B)$ for some $B \subset Y$, then because p is surjective, $p^{-1}p(A) = p^{-1}[pp^{-1}(B)] = p^{-1}(B) = A$.

Define the p-load of any $A \subset X$ to be the p-saturated set $p^{-1}p(A) \supset A$. As Ex. 2 shows, the p-load of an open set in X need not be open. We now have

1.6 Let $p\colon X \to Y$ be an identification. Then p is open (closed) if and only if the p-load of each set open (closed) in X is also open (closed) in X.

Proof: If p is open, then (U open in X) $\Rightarrow$ ($p(U)$ open in Y) $\Rightarrow$ ($p^{-1}p(U)$ open in X), as required. Conversely, if (U open in X) $\Rightarrow$ ($p^{-1}p(U)$ open in X), this says that $p(U)$ is open in Y, that is, p is an open map. The proof for closed maps is similar.

2. Subspaces

The identification topology does not behave well for subspaces of Y. Let $p\colon X \to Y$ be an identification map, and let $F \subset Y$. Then F can receive two topologies: (1) $\mathscr{T}(F)$, that as a subspace of Y; and (2) the identification topology $\mathscr{T}(p, F)$ determined by the surjection $p\colon p^{-1}(F) \to F$. Since $p\colon p^{-1}(F) \to F$ is continuous when F carries $\mathscr{T}(F)$, it follows from **1.2** that we always have $\mathscr{T}(F) \subset \mathscr{T}(p, F)$; however, these two topologies generally do not coincide.

Ex. 1 Let $F \subset I$ be the set of irrationals, and let $Y = \{1\} \cup F$. Give Y the identification topology determined by $p\colon I \to Y$, where $p(x) = x$ if $x \in F$ and $p(x) = 1$ otherwise; it is easy to verify that the only nonempty open sets in Y are those sets of form $p(W)$, where $W \subset I$ is open and contains $I - F$ [$\mathscr{T}(p)$ is not the indiscrete topology!]. Thus the set $F \cap]0, \frac{1}{2}[$ does not belong to $\mathscr{T}(F)$; however, $F \cap]0, \frac{1}{2}[$ evidently belongs to the topology $\mathscr{T}(p; F)$.

2.1 Theorem Let $p\colon X \to Y$ be an identification, and $F \subset Y$. If either:

(1). F is open or closed in Y (no restriction on p),

or (2). p is an open or closed map (no restriction on F),

then $\mathscr{T}(F) = \mathscr{T}(p; F)$.

Proof: Since $\mathscr{T}(F) \subset \mathscr{T}(p, F)$, only the converse inclusion requires proof.

Ad (1). Assume that F is open. Then $U \in \mathscr{T}(p, F)$ implies that $p^{-1}(U)$ is open in the open $p^{-1}(F)$ which implies that $p^{-1}(U)$ is open in X (by III, **7.3**) and therefore $U \in \mathscr{T}(F)$. The proof for closed F is similar.

Ad (2). Assume that p is an open map. Then $U \in \mathscr{T}(p, F)$ implies that $p^{-1}(U)$ is open in $p^{-1}(F)$, so that $p^{-1}(U) = p^{-1}(F) \cap V$, where V is open in X. By I, **6.5**(b), $U = F \cap p(V)$, and since p is an open map, this shows that U is open in $\mathscr{T}(F)$. The proof for closed maps is similar.

Ex. 2 The requirement that the complete inverse image $p^{-1}(F)$ be used to determine $\mathscr{T}(p, F)$ is essential. Let $p: E^2 \to E^1$ be the projection, so that p is an open map. Let $F = E^1$ (which is open in E^1), and instead of $p^{-1}(F)$, use $M = \{(x, y) \mid y = x + 1 \ (x \geqslant 0), y = x - 1 \ (x < 0)\}$. Then $p: M \to F$ is surjective, and the identification topology determined by $p|M$ has all sets $[0, x[$ open. Thus it does not coincide with the Euclidean topology.

3. General Theorems

If $p: X \to Y$ is a continuous map, the continuity of any $g: Y \to Z$ implies that of $g \circ p$; the fundamental (in fact, characterizing) property of identification maps p is that the converse is also true.

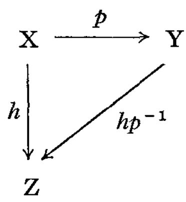
3.1 Theorem Let $p: X \to Y$ be a continuous surjection. p is an identification if and only if: for each space Z and each map $g: Y \to Z$, the continuity of $g \circ p$ implies that of g.

Proof: "Only if": Assume that p is an identification and that $g \circ p$ is continuous. Then, for each open $U \subset Z$, $(g \circ p)^{-1}(U) = p^{-1}(g^{-1}(U))$ is open in X, so that $g^{-1}(U)$ is open in Y; g is therefore continuous. "If": Assume that the condition holds. Let Y' be the *set* Y with topology $\mathscr{T}(p)$; to prevent confusion, let $p': X \to Y'$ be the identification map. Let $i: Y \to Y'$ be the identity map. Since $i \circ p = p'$ is continuous, the condition assures that i is continuous. Since $i^{-1} \circ p' = p$ is continuous, and p' is an identification, the "only if" part of the theorem shows that i^{-1} is continuous. Thus $i: Y \cong Y'$, so p is an identification.

The most frequently used consequence is

3.2 Theorem (Transgression) Let $p: X \to Y$ be an identification, and $h: X \to Z$ be continuous. Assume that hp^{-1} is single-valued [that is, h is constant on each fiber $p^{-1}(y)$]. Then:

(1). $hp^{-1}: Y \to Z$ is continuous, and in addition, the diagram



is commutative.

(2). $hp^{-1}: Y \to Z$ is an open (closed) map if and only if $h(U)$ is open (closed) whenever U is an open (closed) set satisfying $U = p^{-1}pU$ [that is, whenever U is a p-saturated open (closed) set].

Proof: Ad (1). Since for each $x \in X$, $p^{-1}p(x)$ is the fiber containing x, and since h is constant on each fiber, we have

$$h(x) = hp^{-1}p(x) = (hp^{-1}) \circ p(x),$$

that is, $h = hp^{-1} \circ p$. Because h is continuous, **3.1** shows that the map hp^{-1} is continuous.

Ad (2). Let $V \subset Y$ be open; then $U = p^{-1}(V)$ is open in X, and, being p-saturated, satisfies $p^{-1}pU = U$; thus $h(U) = hp^{-1}p(U) = hp^{-1}(V)$, from which the conclusion follows at once.

3.3 Theorem (Transitivity) Let $p: X \to Y$ be an identification, Z a set, and $g: Y \to Z$ a surjection. Then $\mathscr{T}(g \circ p) = \mathscr{T}(g)$; in particular, $g \circ p$ is an identification if and only if g is.

Proof: $g \circ p$ is continuous if and only if g is, so that $(g \circ p)^{-1}(U)$ is open if and only if $g^{-1}(U)$ is open.

Whenever $p: X \to Y$ is an identification, the results on connectedness and local connectedness can be strengthened.

3.4 Let $p: X \to Y$ be an identification, and assume that $p^{-1}(y)$ is connected for each $y \in Y$. Then an open (or closed) $F \subset Y$ is connected if and only if $p^{-1}(F)$ is connected.

Proof: It is necessary to prove only that the connectedness of F implies that of $p^{-1}(F)$. By **2.1**, F has the identification topology determined by $q = p \mid p^{-1}(F)$. If $p^{-1}(F)$ is not connected, there is a continuous surjection $h: p^{-1}(F) \to 2$; h is constant on each fiber $q^{-1}(y)$, since $q^{-1}(y)$ is connected, so from **3.2** it follows that $hq^{-1}: F \to 2$ is continuous. Since hq^{-1} is evidently also surjective, F is not connected.

3.5 Let $p: X \to Y$ be an identification. If X is locally connected, so also is Y.

Proof: Let $U \subset Y$ be open and K a component of U; by V, **4.2**, it suffices to show that K is open, that is, $p^{-1}(K)$ is open in X. Let $x \in p^{-1}(K)$, and let $C(x)$ be its component in the open set $p^{-1}(U)$; since $p(C(x))$ is connected, contains $p(x) \in K$, and lies in U, we have $p(C(x)) \subset K$, so that $x \in C(x) \subset p^{-1}(K)$. Because X is locally connected, $C(x)$ is open, and from III, **2.3**, it follows that $p^{-1}(K)$ is open.

4. Spaces with Equivalence Relations

Let X be a space, R an equivalence relation in X, and X/R the quotient set; recall (I, **7.6**) the projection $p: X \to X/R$ is given by $x \to Rx$.

4.1 Definition The set X/R with the identification topology determined by the projection $p: X \to X/R$ is called the *quotient space* of X by R.

Since R "identifies points," this makes precise the topological structure of the resulting space.

Ex. 1 For $A \subset X$, let R_A be the equivalence relation $(A \times A) \cup \{(x, x) \mid x \in X\}$. The quotient space X/R_A is the space X with A identified to a point, $[A]$, and is written X/A. Observe that if A is either open or closed, then the complement C of $[A]$ is homeomorphic to $X - A: p \mid X - A$ is bijective, and by **2.1**, the topology of $C \subset X/A$ is the identification topology determined by $p \mid X - A$.

Ex. 2 For any space X, let R_C be the equivalence relation determined by its components: $x \sim x'$ if and only if x and x' lie in a common component. Then X/R_C is totally disconnected: since each fiber $p^{-1}(\bar{x})$ is a component, it is connected so, by **3.4**, an $F \subset X/R_C$ is connected if and only if $p^{-1}(F)$ is connected. Thus F must consist of a single point.

4.2 (1). Let $p: X \to X/R$ be the projection. If $B \subset X/R$ is either open or closed, then B is homeomorphic to the space $p^{-1}(B)/R_0$, where R_0 is the relation on $p^{-1}(B)$ induced by R.

(2). $p: X \to X/R$ is open (closed) if and only if

$$R(U) = \bigcup \{Ru \mid u \in U\}$$

is open (closed) in X for each open (closed) $U \subset X$.

(3). If $h: X \to Z$ is continuous and if hp^{-1} is single-valued, then $hp^{-1}: X/R \to Z$ is continuous; and it is open (closed) if and only if $h(U)$ is open (closed) for each open (closed) $U \subset X$ satisfying $U = R(U)$.

Proof: (1). By **2.1**, B has the identification topology determined by $s = p \mid p^{-1}(B)$. Let $q: p^{-1}(B) \to p^{-1}(B)/R_0$ be the projection. Then $sq^{-1}: p^{-1}(B)/R_0 \to B$ is single-valued, and by **3.2**, is continuous. Similarly, $qs^{-1}: B \to p^{-1}(B)/R_0$ is continuous, and since qs^{-1}, sq^{-1} are inverses of one another, each is a homeomorphism (III, **12.3**). (2) follows from **1.6**, since $R(U)$ is the p-load of U. (3) is simply **3.2** in this special case.

Ex. 3 Let I be the unit interval and R the equivalence relation $0 \sim 1$, $x \sim x$ $(x \neq 0, 1)$, which "identifies" 0 and 1. Then $I/R \cong S^1$. For, let $p: I \to I/R$ be the identification and $h: I \to S^1$ be the map $x \to e^{2\pi i x}$. Since hp^{-1} is single-valued, it is continuous and also bijective. To see that it is an open map, note that no open subset of I having form $[0, b[$, or $]a, 1]$ satisfies $U = R(U)$; therefore such intervals must always occur in pairs for sets that satisfy this condition. It follows that each p-saturated open set is mapped by h onto an open set in E^1, so hp^{-1} is open.

Ex. 4 If $p: X \to X/R$ is an open (closed) map, R is called an open (closed) relation. This is *not* related to the behavior of R as a subset of $X \times X$: In I, define $x \sim y$ by "$|x - y|$ is rational"; then $p(]a, b[) = I/R$ for each open interval so that $\mathscr{T}(p)$ is indiscrete. Thus p is an open map, but R is not open in $I \times I$.

4.3 Theorem Let X, Y be spaces with equivalence relations R, S, respectively, and let $f: X \to Y$ be a relation-preserving, continuous map. Then, passing to the quotient (I, **7.7**), the map $f_*: X/R \to Y/S$ is also continuous. Furthermore, f_* is an identification whenever f is an identification.

Proof: We have the commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle q} \\ X/R & \xrightarrow[f_*]{} & Y/S \end{array}$$

Since $q \circ f$ is continuous, and $f_* \circ p = q \circ f$, **3.2** shows that f_* is continuous. Now assume that f is an identification; then **3.3** shows first that $f_* \circ p$ $(= q \circ f)$ is an identification and then that f_* is also.

5. Cones and Suspensions

5.1 ***Definition*** For any space X, the cone TX over X is the quotient space $(X \times I)/R$, where R is the equivalence relation $(x, 1) \sim (x', 1)$ for all $x, x' \in X$.

Equivalently, $TX = (X \times I)/(X \times 1)$; intuitively, TX is obtained from $X \times I$ by pinching $X \times 1$ to a single point. The elements of TX are denoted by $\langle x, t\rangle$. It is trivial to verify that the map $x \to \langle x, 0\rangle$ is a homeomorphism, so we can identify X with the subspace $\{\langle x, 0\rangle \mid x \in X\} \subset TX$.

Ex. 1 Though TX is, geometrically, a cone, it may have more open sets than a cone formed by possibly more elementary methods. In E^2, let $X = \{(n, 0) \mid n \in Z^+\}$, and let CX be the subspace of E^2 obtained by joining each $x \in X$ to $w_0 = (0, 1)$ by a segment. It is easy to verify that there is a continuous bijection $TX \to CX$; but the spaces are not homeomorphic. Indeed, let V_n = all points on the segment from w_0 to $(n, 0)$ within $1/n$ of w_0, and $V = \bigcup_n V_n$; then V is evidently not open in CX, but it is open in TX since $p^{-1}(V)$ is open in $X \times I$.

5.2 A continuous $f: X \to Y$ induces a continuous map $Tf: TX \to TY$ by $\langle x, t\rangle \to \langle f(x), t\rangle$.

Proof: Define $\bar{f}: X \times I \to Y \times I$ by $(x, t) \to (f(x), t)$. Noting that $\bar{f}$ is relation-preserving, we pass to the quotient and use **4.3**.

5.3 Definition Let $\mathcal{J}$ be the interval $[-1, +1]$. For any space X, the suspension SX of X is the quotient space $(X \times \mathcal{J})/R$, where R is the equivalence relation $(x, 1) \sim (x', 1)$, $(x, -1) \sim (x', -1)$ for all $x, x' \in X$.

Intuitively, SX is obtained from $X \times \mathcal{J}$ by pinching to points each of the sets $X \times 1$, $X \times (-1)$. The elements of SX are denoted by $\langle x, t\rangle$ also.

Ex. 2 It is easy to verify $S\,S^0 \cong S^1$; we will see later, once compactness has been discussed, that $S\,S^n \cong S^{n+1}$ for all n (*cf.* XI, **2**, Ex. 5).

5.4 (1). $SX \cong TX/X$.
(2). TX is homeomorphic to the subspace $\{\langle x, t\rangle \in SX \mid t \geqslant 0\}$.
(3). A continuous $f: X \to Y$ induces a continuous map $Sf: SX \to SY$ by $\langle x, t\rangle \to \langle f(x), t\rangle$.

Proof: (1) This follows easily by using the map $u: TX \to SX$ sending $\langle x, t\rangle \to \langle x, 2t - 1\rangle$ and passing to the quotient. (2) is straightforward, and (3) follows as in **5.2**.

6. Attaching of Spaces

The process of attaching a space X to a space Y by a map f has great importance in modern topology; it contains as special cases the cone and suspension constructions and also the identification of closed sets to points.

The *free* union $X + Y$ of disjoint spaces X, Y is the set $X \cup Y$ with topology: $U \subset X + Y$ is open if and only if $U \cap X$ is open in X and $U \cap Y$ is open in Y. Since $X \cap Y = \varnothing$, X and Y keep their own topologies and are disjoint open sets in $X + Y$. Clearly, $B \subset X + Y$ is closed if and only if both $B \cap X$ and $B \cap Y$ are closed.

6.1 Definition Let X, Y be two disjoint spaces, $A \subset X$ a *closed* subset, and $f: A \to Y$ continuous. In $X + Y$, generate an equivalence relation R by $a \sim f(a)$ for each $a \in A$. The quotient space $(X + Y)/R$ is said to be "X attached to Y by f," and is written $X \cup_f Y$; f is called the attaching map.

In intuitive terminology, we "identify each $a \in A$ with its image $f(a) \in Y$."

Ex. 1 Let $A \subset X$ be closed, and attach X to a point y_0 by $f(A) = y_0$; then $X \cup_f y_0 \cong X/A$. For, defining $\varphi: X \to X + y_0$ by $\varphi(x) = x$, and $\psi: X + y_0 \to X$ by $\psi(x) = x$, $\psi(y_0) \in A$, both φ and ψ are continuous and relation-preserving; passing to the quotient gives continuous maps $\varphi_*: X/A \to X \cup_f y_0$, $\psi_*: X \cup_f y_0 \to X/A$, which are inverses. In particular, attaching I to y_0 by $f(0) = f(1) = y_0$ gives S^1.

Ex. 2 TX is obtained by attaching $X \times I$ to a point p° by $f(X \times 1) = p^\circ$, since $TX = (X \times I)/(X \times 1)$; the suspension SX is obtained by attaching $X \times \mathcal{J}$ to $p^\circ \cup p^-$ by $f(X \times 1) = p^\circ, f(X \times (-1)) = p^-$.

Ex. 3 Attaching $X \times I \times Y$ to the free union $X + Y$ by $(x, 0, y) \to x$, $(x, 1, y) \to y$ gives a space called the *join* $X * Y$ of X and Y. This consists of the spaces X, Y together with line segments joining each $x \in X$ to each $y \in Y$, where no two of the segments have interior points in common. The reader can verify $X * y_0 \cong TX$, and $X * S^0 \cong SX$.

The construction of closed sets in $X \cup_f Y$ is generally based on

6.2 Let $p: X + Y \to X \cup_f Y$ be the projection, and let $C \subset X + Y$ be such that $C \cap X$ is closed in X. Then $p(C)$ is closed in $X \cup_f Y$ if and only if $(C \cap Y) \bigcup f(C \cap A)$ is closed in Y.

Proof: It is elementary to verify that for any $C \subset X + Y$, $p^{-1}p(C) = C \bigcup f(C \cap A) \bigcup f^{-1}[f(C \cap A)] \bigcup f^{-1}(C \cap Y)$; consequently,

$$p^{-1}[p(C)] \cap Y = (C \cap Y) \bigcup f(C \cap A),$$
$$p^{-1}[p(C)] \cap X = (C \cap X) \bigcup f^{-1}[p^{-1}p(C) \cap Y].$$

Now assume that $C \cap X$ is closed. Then, because A is closed in X and f is continuous, $p^{-1}p(C)$ is closed in $X + Y$ if and only if $(C \cap Y) \bigcup f(C \cap A)$ is closed in Y.

6.3 Theorem Let $p: X + Y \to X \cup_f Y$ be the projection. Then:

(1). Y is embedded as a closed set, homeomorphic to Y, and $p \mid Y$ is a homeomorphism.

(2). $X - A$ is embedded homeomorphically as an open set, and $p \mid X - A$ is a homeomorphism.

Proof: (1) $p \mid Y$ is continuous, and evidently bijective; by **6.2** it is also a closed map and hence is a homeomorphism. (2) is proved similarly.

For subspaces,

6.4 Let X be attached to Y by $f: A \to Y$. Let $X_1 \subset X$ and $Y_1 \subset Y$ be closed subsets such that $f(A \cap X_1) \subset Y_1$, and attach X_1 to Y_1 by $f_1 = f \mid A \cap X_1$. Then $X_1 \cup_{f_1} Y_1$ is homeomorphic to a closed subset of $X \cup_f Y$.

Proof: Let $i: X_1 + Y_1 \to X + Y$ be the identity map; since $p \circ i \circ p_1^{-1}: X_1 \cup_{f_1} Y_1 \to X \cup_f Y$ is single-valued, it is continuous. It is clearly also injective. We now show that it is a closed map by proving (**3.2**) that $p \circ i(C_1)$ is closed for each closed $C_1 \subset X_1 + Y_1$ satisfying $p_1^{-1}p_1(C_1) = C_1$. Since $C_1 \cap X = C_1 \cap X_1$ is closed in the closed X_1, it is closed in X. Furthermore, by **6.2**, $(C_1 \cap Y_1) \bigcup f_1(C_1 \cap A)$ is closed in the closed Y_1 and therefore also in Y; since $f_1 = f \mid A \cap X_1$ and $C_1 \subset X_1 + Y_1$, this set is $(C_1 \cap Y) \bigcup f(C_1 \cap A)$, and so, by **6.2**, we have $p(C_1) = p \circ i(C_1)$ closed in $X \cup_f Y$, completing the proof.

The construction of continuous maps $X \cup_f Y \to Z$ is of frequent occurrence and is based on the important

6.5 Theorem Let X be attached to Y by $f: A \to Y$, and let $p: X + Y \to X \cup_f Y$ be the identification map. Let $\varphi: X \to Z$ and $\psi: Y \to Z$ be a pair of continuous maps, and let $(\varphi, \psi): X + Y \to Z$ be their unique common extension. If φ and ψ are "consistent" (that is, if $\varphi(a) = \psi[f(a)]$ for each $a \in A$) then $(\varphi, \psi)p^{-1}: X \cup_f Y \to Z$ is continuous.

Proof: The transgressive map $(\varphi, \psi)p^{-1}$ is single-valued, so **3.2** applies.

7. The Relation K(f) for Continuous Maps

If $f: X \to Y$ is a homomorphism of groups, we can construct $f(X)$ from X and the kernel of f. We will see that a similar process for continuous maps of spaces is possible if and only if f is an identification.

Let $f: X \to Y$ be any continuous map. We define the relation $K(f)$ in X by $x \sim x'$ if $f(x) = f(x')$. This is clearly an equivalence relation in X, and therefore we have the identification map $p: X \to X/K(f)$; $X/K(f)$ is called the decomposition space of f. Since fp^{-1} is single-valued, it is a continuous (in fact, injective) map $X/K(f) \to Y$. Using **3.2**, we have thus proved the first part of

7.1 (1). Let $f: X \to Y$ be continuous. Then f can always be factored as $X \xrightarrow{p} X/K(f) \xrightarrow{g} Y$ (that is, $f = g \circ p$), where p and g $(= fp^{-1})$ are continuous, p is *surjective*, and g is *injective*.

(2.) If

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \varphi \downarrow & & \downarrow \psi \\ Z & \xrightarrow[h]{} & W \end{array}$$

is a commutative diagram of continuous maps, there is a continuous λ: $X/K(f) \to Z/K(h)$ such that the diagram

$$\begin{array}{ccccc} X & \xrightarrow{p} & X/K(f) & \xrightarrow{g} & Y \\ \varphi\downarrow & & \downarrow\lambda & & \downarrow\psi \\ Z & \xrightarrow[p']{} & Z/K(h) & \xrightarrow[g']{} & W \end{array} \qquad (g \circ p = f;\ g' \circ p' = h)$$

commutes in each square.

Proof: (2) Because the given diagram commutes, it is trivial to verify that φ: $X \to Z$ is relation-preserving; λ is the map φ_*, obtained from φ by passing to the quotient (**4.3**). The verification of commutativity in the resulting diagram is trivial.

If f: $X \to Y$ is surjective, then fp^{-1}: $X/K(f) \to Y$ is bijective and continuous. It need not be a homeomorphism:

7.2 Theorem Let f: $X \to Y$ be a continuous surjection. Then

$$fp^{-1}: X/K(f) \cong Y$$

if and only if f is an identification.

Proof: Assume that f is an identification. We need show only that fp^{-1} is an open map. Using **3.2**, if $U = p^{-1}p(U)$ for some open set $U \subset X$, then since $p^{-1}p(U) = f^{-1}f(U)$, we find that $f(U)$ is open, as required. Conversely, if fp^{-1} is a homeomorphism, it is an identification, and by **3.2** and **3.3**, so also is $fp^{-1} \circ p = f$.

Let $\mathscr{A}$ be any set, and for each $\alpha \in \mathscr{A}$, let p_α: $X_\alpha \to Y_\alpha$ be continuous. The relation $K(\prod p_\alpha)$ for $\prod p_\alpha$: $\prod X_\alpha \to \prod Y_\alpha$ is evidently the same as $\prod_\alpha K(p_\alpha)$.

7.3 If each p_α: $X_\alpha \to Y_\alpha$ is a continuous open surjection, then

$$\frac{\prod_\alpha X_\alpha}{\prod_\alpha K(p_\alpha)} \cong \prod_\alpha Y_\alpha.$$

Proof: By IV, **2.5**, $\prod p_\alpha$ is an open map, and therefore is an identification.

Ex. 1 Proposition **7.3** is *not* necessarily true if any one p_α is not open. Let Q be the set of rationals, R the identity relation, and S the relation identifying all the integers. Then p: $Q \to Q/R$ is an open surjection, but q: $Q \to Q/S$ is not (q is in fact a closed map). It is not difficult to verify that $(Q \times Q)/(R \times S)$ is *not* homeomorphic to $Q/R \times Q/S$. This example, due to J. Dieudonné, also shows (in view of **7.2**) that the cartesian product of identification maps need not be an identification map. This question will be treated later, after local compactness has been discussed (*cf.* XII, **4**).

8. Weak Topologies

In this section, we shall study a method for topologizing a set that, under certain simple conditions, preserves some predetermined topology on a given family of subsets. We shall also show that this process is closely related to identifications.

8.1 Definition Let X be a set, and let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of subsets of X, with each A_α having a topology. Assume that for each $(\alpha, \beta) \in \mathscr{A} \times \mathscr{A}$, both

(1). The topologies of A_α and A_β agree on $A_\alpha \cap A_\beta$.
(2). Either (a) each $A_\alpha \cap A_\beta$ is open in A_α and in A_β or (b) each $A_\alpha \cap A_\beta$ is closed in A_α and in A_β.

The weak topology in X determined (or induced) by $\mathfrak{A}$ is $\mathscr{T}(\mathfrak{A}) = \{U \subset X \mid \forall\, \alpha\colon U \cap A_\alpha \text{ is open in } A_\alpha\}$.

It is evident that $\mathscr{T}(\mathfrak{A})$ is indeed a topology; furthermore, $B \subset X$ is closed in the space X if and only if its intersection with each A_α is closed in A_α. The force of condition (2) is

8.2 Each A_α, as a subspace of $(X, \mathscr{T}(\mathfrak{A}))$, retains its original topology. In case (a), it is an open subset of space X; in case (b), it is a closed subset of the space X.

Proof: We prove this for case (a); that for (b) is similar. Let $B \subset A_\alpha$ be open in A_α; then for any β, $B \cap A_\beta$ is open in $A_\alpha \cap A_\beta$, so by **8.1**(2) and III, **7.3**, $B \cap A_\beta$ is open in A_β; by **8.1**, this means that B is open in X, as required. In particular, taking $B = A_\alpha$, we find that A_α is open in X.

Ex. 1 $\mathscr{T}(\mathfrak{A})$ is the largest topology in X that preserves the given topology of each A_α: for, in any other such topology $\mathscr{T}$, $U \in \mathscr{T} \Rightarrow U \cap A_\alpha$ is open in A_α for each $\alpha \Rightarrow U \in \mathscr{T}(\mathfrak{A})$.

Ex. 2 If $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is an open covering or a nbd-finite closed covering, of a *space* X then the topology of X is exactly the weak topology $\mathscr{T}(\mathfrak{A})$ (*cf.* III, **9.3**).

Ex. 3 Let R be a real linear vector space (*cf.* Appendix, I, **4.2**). The weak topology $\mathscr{T}(w)$ in R is that determined by the Euclidean topology on each of its finite-dimensional flats. This is called the *finite topology of* R and illustrates a typical way of using **8.1**.

Ex. 4 If $X - \bigcup_\alpha A_\alpha \neq \varnothing$ in **8.1**, then the *subspace* $X - \bigcup_\alpha A_\alpha$ of the space X is discrete.

For spaces with weak topology, the continuity of maps is easily tested:

8.3 If X is a space with weak topology determined by $\{A_\alpha \mid \alpha \in \mathscr{A}\}$, then an $f: X \to Y$ is continuous if and only if each $f \mid A_\alpha: A_\alpha \to Y$ is continuous.

Proof: The only proof required is that the continuity of each $f_\alpha = f \mid A_\alpha$ implies the continuity of f. Let $U \subset Y$ be open; then, $f^{-1}(U) \cap A_\alpha = f_\alpha^{-1}(U)$ is open in A_α for each α, and by **8.1**, $f^{-1}(U)$ is therefore open in X.

Remark: In view of Ex. 2, note that III, **9.4** is a special case of **8.3.**

We now describe weak topologies in terms of identifications; it will be seen that spaces with weak topology are always obtainable by "pasting" spaces together along specified subsets.

8.4 Definition Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces. For each α, let Y'_α be the space $\{\alpha\} \times Y_\alpha$, so that $Y'_\alpha \cong Y_\alpha$ and the family $\{Y'_\alpha \mid \alpha \in \mathscr{A}\}$ is pairwise disjoint. The *free union* of the given family $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ is the set $\bigcup_\alpha Y'_\alpha$ with the weak topology determined by the spaces Y'_α; this space is denoted by $\sum_\alpha Y'_\alpha$.

8.5 Theorem Let $(X, \mathscr{T})$ be a space with weak topology determined by the *covering* $\{A_\alpha \mid \alpha \in \mathscr{A}\}$. Let $A = \sum_\alpha A'_\alpha$ be the free union of $\{A_\alpha \mid \alpha \in \mathscr{A}\}$, and for each α let $h_\alpha: A'_\alpha \to A_\alpha \subset X$ be the homeomorphism $(\alpha, a) \to a$. Define $h: \sum_\alpha A'_\alpha \to X$ by $h \mid A'_\alpha = h_\alpha$. Then h is continuous and $A/K(h) \cong X$.

Proof: The continuity of h is evident from **8.3**, and its surjectivity is obvious. To prove the theorem, we need show (*cf.* **7.2**) only that h is an identification. For this, let $U \subset X$ be such that $h^{-1}(U)$ is open in $\sum_\alpha A'_\alpha$; then, $h^{-1}(U) \cap A_\alpha = h_\alpha^{-1}(U \cap A_\alpha)$ is open in A'_α for each α, and, since h_α is a homeomorphism, $U \cap A_\alpha$ is open in A_α. Thus U is open in X, completing the proof.

Ex. 5 If X has the weak topology determined by $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ and Y that determined by $\{B_\beta \mid \beta \in \mathscr{B}\}$, the cartesian product topology $X \times Y$ may *not* be the weak topology *induced* by $\{A_\alpha \times B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$. The following example is due to C. H. Dowker:

Let $\mathscr{M}$ be the set of all maps of Z^+ into itself. Let X be a real vector space having a basis $\{u_\varphi \mid \varphi \in \mathscr{M}\}$ in 1-to-1 correspondence with the set $\mathscr{M}$, and let Y be a real vector space with a basis $\{v_n \mid n \in Z^+\}$. We take both X and Y with the finite topology of Ex. 3. The *induced* weak topology in $X \times Y$ is simply the finite topology of the vector space $X \times Y$; we will show that this is not the cartesian product topology of the spaces X and Y.

Let $P \subset X \times Y$ be the set

$$\left\{\left(\frac{1}{\varphi(n)}\, u_\varphi, \frac{1}{\varphi(n)}\, v_n\right) \middle| \varphi \in \mathscr{M},\ n \in Z^+\right\}.$$

We observe that P is closed in the *finite topology* of $X \times Y$, since its intersection with any finite-dimensional flat is a finite set and so is closed in the Euclidean topology of that flat. However, P is *not* closed in the *cartesian product* topology of the spaces X and Y. For, if it were, $\mathscr{C}P$ would be open, and since the origin $0 \in \mathscr{C}P$, there would be a basic nbd $U \times V$ with $0 \in U \times V \subset \mathscr{C}P$. Since U, V are open in X, Y, respectively, then for each φ and each n, there would be an a_φ and a_n such that

$$\{\lambda u_\varphi \mid 0 \leqslant \lambda < a_\varphi\} \subset U, \qquad \{\mu v_n \mid 0 \leqslant \mu < a_n\} \subset V.$$

Let $\bar{\varphi} \in \mathscr{M}$ be the map

$$\bar{\varphi}(n) = \max\left[n, \frac{1}{a_n}\right] + 1$$

and find $\bar{n}$ with $\bar{\varphi}(\bar{n}) > 1/a_{\bar{\varphi}}$. Then

$$\left(\frac{1}{\bar{\varphi}(\bar{n})}\, u_{\bar{\varphi}}, \frac{1}{\bar{\varphi}(\bar{n})}\, v_{\bar{n}}\right) \in U \times V,$$

but is not in $\mathscr{C}P$, a contradiction.

[The reader will note (*cf.* **7**, Ex. 1) that this example shows again that $(X \times Y)/(R \times S)$ is in general not homeomorphic to $(X/R) \times (Y/S)$.] We shall see later that if one of the spaces X, Y is locally compact, the cartesian and the induced weak topology in $X \times Y$ do in fact agree (*cf.* XII, **4.4**).

Problems

Section 1

1. Reversing the situation treated in the text, let X be a *set*, $(Y, \mathscr{T})$ a *space*, and $p\colon X \to Y$ a subjective map. Prove:

   a. $\mathscr{T}_X = \{p^{-1}(U) \mid U \text{ open in } Y\}$ is a topology in X.

   b. $p\colon (X, \mathscr{T}_X) \to (Y, \mathscr{T})$ is continuous, open, and closed.

2. For each $\alpha \in \mathscr{A}$, let $p_\alpha\colon X_\alpha \to Y_\alpha$ be a continuous, open surjection. Show $\prod_\alpha p_\alpha\colon \prod_\alpha X_\alpha \to \prod_\alpha Y_\alpha$ is an identification.

3. Let X be a space and $A \subset X$ a subspace. Assume that there exists a continuous $r\colon X \to A$ such that $r \mid A = 1_A$ (such a map is called a *retraction* of X onto A). Show that r is an identification.

4. Let X be any set. Given any family $\{(Y_\alpha, \mathscr{T}_\alpha), f_\alpha \mid \alpha \in \mathscr{A}\}$ of spaces and maps $f_\alpha: X \to Y_\alpha$, the "projective limit topology of X determined by this family" is $\bigvee_\alpha f_\alpha^{-1}(\mathscr{T}_\alpha)$ (*cf.* III, **3**, Problem 8). Prove:
   a. If $j: X \to \prod_\alpha Y_\alpha$ is the map $j(x) = \{f_\alpha(x)\}$, then $\bigvee_\alpha f_\alpha^{-1}(\mathscr{T}_\alpha)$ is the topology in X determined by j as in Problem 1.
   b. If whenever $x \neq x'$, there is some index α such that $f_\alpha(x) \neq f_\alpha(x')$, then j is an embedding.

Section 2

1. Let X have the projective limit topology (**1**, Problem 4) determined by
$$\{Y_\alpha, f_\alpha \mid \alpha \in \mathscr{A}\}$$
and let $A \subset X$. Prove: The subspace topology of A is the projective limit topology determined by the maps $f_\alpha \mid A$.

Section 3

1. Let $p: X \to Y$ be a continuous open surjection, and assume that each fiber $p^{-1}(y)$ is connected. For any $F \subset Y$, show that F is connected if and only if $p^{-1}(F)$ is connected.
2. Let X have the projective limit topology $\mathscr{T}$ determined by the family
$$\{(Y_\alpha, \mathscr{T}_\alpha), f_\alpha \mid \alpha \in \mathscr{A}\}.$$
Assume that each $\mathscr{T}_\alpha$ is the projective limit topology determined by a family $\{(Z_{\alpha,\beta}, \mathscr{T}_{\alpha,\beta}), g_{\alpha,\beta} \mid \beta \in \mathscr{B}\}$. Prove: $\mathscr{T}$ is the projective limit topology determined by
$$\{(Z_{\alpha,\beta}, \mathscr{T}_{\alpha,\beta}), g_{\alpha,\beta} \circ f_\alpha \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}.$$
3. Let X have the projective limit topology determined by $\{Y_\alpha, f_\alpha \mid \alpha \in \mathscr{A}\}$. Prove: $f: Z \to X$ is continuous if and only if each $f_\alpha \circ f$ is continuous.

Section 4

1. Let $p: X \to X/R$ be an open (or closed) map, and $B \subset X/R$ any subset. Show that B is homeomorphic to $p^{-1}(B)/R_0$.
2. Give an example showing that if $A \subset X$ is not open or closed, then $X - A$ need not be homeomorphic to the complement of $[A]$ in X/A (*cf.* **2**, Ex. 1).
3. These problems will be much easier after studying compactness. (Use now III, **9**, Problem 1.)
   a. In I^2, let $(0, y) \sim (1, y)$. Show $I^2/R \cong$ the cylinder $S^1 \times I$.
   b. In I^2, let $(0, y) \sim (1, 1 - y)$. Show $I^2/R \cong$ Moebius band.
   c. In I^2, let $\mathrm{Fr}(I^2) \sim (0, 0)$. Show $I^2/R \cong S^2$.
   d. In I^2, let $(0, y) \sim (1, y), (x, 0) \sim (x, 1)$. Show $I^2/R \cong$ the torus $S^1 \times S^1$.
4. Let R be an equivalence relation in X. For each $A \subset X$ define $C(A) = \{x \in X \mid \exists\, a \in A: x\,R\,a\}$. Show that $p(U)$ is open in X/R if and only if $C(U)$ is open in X.
5. Let R, S be two equivalence relations in X, and such that $S \subset R$ (*cf.* I, **7**, Problem 6). Prove $(A/S)/(R/S) \cong A/R$.

6. Let 0 be the origin in E^3. In $E^3 - 0$, define $x\,R\,y$ if x and y lie on a line through the origin. Show that R is an equivalence relation; $(E^3 - 0)/R$ is called the projective plane P^2. Call *line* in P^2 any set A such that $p^{-1}(A)$ is a plane in $E^3 - 0$ going through the origin. Show that a line in P^2 is homeomorphic to S^1.
7. Let $V^2 = \{x \in E^2 |\ |x| \leqslant 1\}$. Generate an equivalence relation by $x\,R\,y$ if $|x| = |y| = 1$, and x, y are diametrically opposite. Show V^2/R is homeomorphic to P^2.

Section 5

1. If $A \subset X$ is closed, prove that TA is homeomorphic to a closed subspace of TX.
2. Let i: $\text{Int}(I) \to I$ be the inclusion map. Show that the map Ti: $T[\text{Int}(I)] \to TI$ is not an embedding.

Section 6

1. Let $A \subset X$ and $B \subset Y$ be closed. Show that $A * B$ can be identified with a closed subspace of $X * Y$ (it is a convention that $\varnothing * Y = Y, X * \varnothing = X$).
2. Prove: $X * Y \cong Y * X$.
3. Use **6.4** to prove that TX can be considered a closed subspace of SX.
4. Attach TX to TX by $f\langle x, 0\rangle = \langle x, 0\rangle$. Show that $TX \cup_f TX \cong SX$.
5. Let X be attached to Y by f: $A \to Y$. Let $V \subset Y$ be open, and let $U \subset X$ be open, such that $f^{-1}(V) = U \cap A$. Prove: $p(V \cup U)$ is open in $X \cup_f Y$.
6. Let $F \subset X \cup_f Y$ be a given set. Assume that $V \subset Y$, $U \subset X$ are open and satisfy: (a) $Y \cap p^{-1}(F) \subset V$; (b) $X \cap p^{-1}[F \cup p(V)] \subset U$. Show that $p[(U - A) \cup V]$ is open in $X \cup_f Y$ and contains F.

Section 7

1. Show that I^n is a quotient space of the Cantor set.
2. If Y is a quotient space of X, and Z is a quotient space of Y, prove that Z is homeomorphic to a quotient space of X.
3. Let R_1, R_2 be two relations in a space X such that $x\,R_1\,x' \Rightarrow x\,R_2\,x'$ for every pair x, x'. Show that X/R_2 is a quotient space of X/R_1.
4. Let $A \subset X$ be a retract of X, and let r: $X \to A$ be the map in **I**, Problem 3. Show $A \cong X/K(r)$.

Section 8

1. If we drop both requirements in **8.1**, does the process still determine a topology?
2. Let both X and Y have weak topology, with that for X determined by $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ and that for Y determined by $\{B_\beta \mid \beta \in \mathscr{B}\}$. Let $X \times Y$ be the cartesian product topology, and let $(X \times Y, \mathscr{T})$ be the weak topology in $X \times Y$ determined by $\{A_\alpha \times B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$. Show that the map l: $(X \times Y; \mathscr{T}) \to X \times Y$ is continuous.
3. In Problem 2, let both $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ and $\{B_\beta \mid \beta \in \mathscr{B}\}$ be open coverings or nbd-finite closed coverings. Show that l is then always a homeomorphism.

4. In Problem 2, assume that each $x \in X$ is in the interior of some A_α and also that each $y \in Y$ is in the interior of some B_β. Show that the map $l: (X \times Y; \mathscr{T}) \to X \times Y$ is a homeomorphism.
5. Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces, and $\{b_\alpha^0\}$ a fixed point in $\prod_\alpha Y_\alpha$. Let PY_α be the subset of all points in the set $\prod_\alpha Y_\alpha$ having at most finitely many coordinates different from $\{b_\alpha^0\}$. For each finite $\mathscr{F} \subset \mathscr{A}$, let $s(\mathscr{F})$ be the slice through $\{b_\alpha^0\}$ parallel to $\prod\{Y_\alpha \mid \alpha \in \mathscr{F}\}$. Take each $s(\mathscr{F})$ with the cartesian product topology, and let PY_α be given the weak topology determined by the family $\{s(\mathscr{F})\}$. Prove: (1) Each projection $p_\beta \mid PY_\alpha : PY_\alpha \to Y_\beta$ is a continuous open map. (2) If $s(b; \alpha)$ is the slice in $\prod_\alpha Y_\alpha$ through $\{b_\alpha\}$ parallel to Y_α, the map $s_\alpha: Y_\alpha \to s(b; \alpha)$ is continuous. (3) This type of cartesian product is in general not associative (*hint*: use Dowker's example).

Separation Axioms

VII

So far, our only requirement for a topology has been that it satisfy the axioms. From now on, we will impose increasingly more severe additional conditions on it. With each new condition, we will determine the invariance properties of the resulting topology: by this we mean (1) whether the topology is invariant under open or closed maps rather than only homeomorphisms; (2) whether the additional properties are inherited by each subspace topology; and (3) whether the additional properties are transmitted to cartesian products. We will also give various desirable features that each such topology has, and more important, we will determine to an extent its behavior under quotient-space formation.

In this chapter, we will require of a topology that it "separate" varying types of subsets.

1. Hausdorff Spaces

Hausdorff topologies have the weakest kind of separation that we will consider in this book; *after this section, all spaces will be assumed to be Hausdorff,* unless specifically stated otherwise.

1.1 Definition A space Y is Hausdorff (or separated) if each two distinct points have nonintersecting nbds, that is, whenever $p \neq q$ there are nbds $U(p)$, $V(q)$ such that $U \cap V = \varnothing$.

Ex. 1 E^n and discrete spaces are Hausdorff. In any set Y, a topology larger than a Hausdorff topology is also Hausdorff.

Ex. 2 Sierpinski space and indiscrete spaces having more than one point are not Hausdorff.

Ex. 3 Hausdorff spaces are also called T_2-spaces in the literature. There are two weaker separation axioms. T_0—for each pair of distinct points, *at least one* has a nbd not containing the other, and T_1—for each pair of distinct points, *each one* has a nbd not containing the other. Sierpinski space is T_0 but not T_1; an infinite set with topology $\mathscr{T} = \{\varnothing\} \cup \{U \mid \mathscr{C}U \text{ is finite}\}$ is a T_1-space that is not Hausdorff.

The definition has several equivalent formulations:

1.2 The following four properties are equivalent:

(1). Y is Hausdorff.

(2). Let $p \in Y$. For each $q \neq p$, there is a nbd $U(p)$ such that $q \,\overline{\in}\, \overline{U(p)}$.

(3). For each $p \in Y$, $\bigcap \{\overline{U} \mid U \text{ is a nbd of } p\} = p$.

(4). The diagonal $\Delta = \{(y, y) \mid y \in Y\}$ is closed in $Y \times Y$.

Proof: (1) ⇒ (2). Given $q \neq p$, there are disjoint $U(p)$, $U(q)$, which says that $q \,\overline{\in}\, \overline{U(p)}$.

(2) ⇒ (3). If $q \neq p$, there is a nbd $U(p)$ with $q \,\overline{\in}\, \overline{U(p)}$, so that $q \,\overline{\in} \bigcap \{\overline{U} \mid U \text{ is a nbd of } p\}$.

To prove the remaining implications, observe that the statement "$U \cap V = \varnothing$" is equivalent to "$(U \times V) \cap \Delta = \varnothing$," since

$$\begin{aligned}(U \times V) \cap \Delta \neq \varnothing &\Leftrightarrow \exists\, p\colon (p, p) \in U \times V \\ &\Leftrightarrow (p \in U) \wedge (p \in V) \\ &\Leftrightarrow U \cap V \neq \varnothing.\end{aligned}$$

(3) ⇒ (4). We show that $\mathscr{C}\Delta$ is open. Let $(p, q) \,\overline{\in}\, \Delta$; then $p \neq q$, and since $p = \bigcap \{\overline{U} \mid U \text{ is a nbd of } p\}$, there is some U with $p \in U$, $q \,\overline{\in}\, \overline{U}$. Since $U \cap \mathscr{C}\overline{U} = \varnothing$, $U \times \mathscr{C}\overline{U}$ is a nbd of (p, q) in $\mathscr{C}\Delta$.

(4) ⇒ (1). If $p \neq q$, then $(p, q) \,\overline{\in}\, \Delta$; therefore (p, q) has a nbd $U \times V$ not meeting Δ, that is, $p \in U$, $q \in V$ and $U \cap V = \varnothing$.

The invariance properties of Hausdorff topologies are

1.3 Theorem (1). Hausdorff topologies are invariant under closed bijections.

(2). Each subspace of a Hausdorff space is also a Hausdorff space.

(3). The cartesian product $\prod \{Y_\alpha \mid \alpha \in \mathscr{A}\}$ is Hausdorff if and only if each Y_α is Hausdorff.

Proof: (1). Since a closed bijection is also an open map, the images of disjoint nbds are disjoint nbds, and the result follows.

(2). Let $A \subset Y$ and $p, q \in A$; since there are disjoint nbds $U(p)$, $U(q)$ in Y, the nbds $U(p) \cap A$ and $U(q) \cap A$ in A are also disjoint.

(3). Assume that each Y_α is Hausdorff and that $\{p_\alpha\} \neq \{q_\alpha\}$; then $p_\alpha \neq q_\alpha$ for some α, so choosing disjoint nbds $U(p_\alpha)$, $U(q_\alpha)$ gives the required disjoint nbds $\langle U(p_\alpha)\rangle$, $\langle U(q_\alpha)\rangle$ in $\prod_\alpha Y_\alpha$. Conversely, if $\prod_\alpha Y_\alpha$ is Hausdorff, then (IV, **3.2**) each Y_α is homeomorphic to some slice in $\prod_\alpha Y_\alpha$, so by (2) (since the Hausdorff property is a topological invariant), Y_α is Hausdorff.

Ex. 4 In contrast to connectedness, the Hausdorff property is not preserved under continuous maps, nor even under continuous open maps, as V, **5**, Ex. 2, shows.

Hausdorff topologies have the following special features, which serve to minimize acutely pathological behavior.

1.4 In Hausdorff spaces:

(1). Each finite set is a closed set.

(2). y is a cluster point of $A \subset Y$ if and only if each nbd of y contains infinitely many points of A.

(3). Each slice $S(y^\circ, \beta) \subset \prod_\alpha Y_\alpha$ is closed in $\prod_\alpha Y_\alpha$.

(4). Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite family of nonempty sets in Y. If $\aleph(\mathscr{A}) \geqslant \aleph_0$, then there exists a discrete closed subspace D such that $\aleph(D) = \aleph(\mathscr{A})$ and $D \subset \bigcup_\alpha A_\alpha$.

Proof: (1). $A = \{y_1, \cdots, y_n\}$ is the union of its finitely many points, and by **1.2**(3), each point (being the intersection of closed sets) is closed.

(2). If the condition holds, $y \in A'$ is clear. If the condition does not hold, there is a nbd $U(y)$ with $U \cap (A - \{y\}) = \{y_1, \cdots, y_n\}$; by (1), $U \cap \mathscr{C}\{y_1, \cdots, y_n\}$ is a nbd of y not intersecting $A - \{y\}$, so $y \bar{\in} A'$.

(3). To see that $\mathscr{C}S(y^\circ, \beta)$ is open, observe that any $y \bar{\in} S(y^\circ, \beta)$ has some coordinate $y_\gamma \neq y_\gamma^\circ$, $(\gamma \neq \beta)$; because Y_γ is Hausdorff, there is a nbd $U(y_\gamma)$ not containing y_γ°, and then $y \in \langle U(y_\gamma)\rangle \subset \mathscr{C}S(y^\circ; \beta)$.

(4). Let Γ be the initial ordinal of cardinal $\aleph(\mathscr{A})$; we will define an injection $\varphi\colon [0, \Gamma[\to \bigcup_\alpha A_\alpha$ by transfinite construction. Let $\beta < \Gamma$ and assume $\varphi(\gamma) = y_\gamma \in Y$ defined for all $\gamma < \beta$. Let

$$\mathscr{B}(y_\gamma) = \{\alpha \in \mathscr{A} \mid y_\gamma \in A_\alpha\}$$

and let $\mathscr{B}_\beta = \cup \{\mathscr{B}(y_\gamma) \mid \gamma < \beta\}$. Each $\aleph[\mathscr{B}(y_\gamma)] < \aleph_0$ because the family is nbd-finite, so since $\aleph([0, \beta[) < \aleph(\mathscr{A})$, we find (II, **8.3**) that

$\aleph(\mathscr{B}_\beta) < \aleph(\mathscr{A})$. Thus, $\mathscr{A} - \mathscr{B}_\beta \neq \varnothing$, and we define $\varphi(\beta)$ to be the first element in $\cup \{A_\mu \mid \mu \in \mathscr{A} - \mathscr{B}_\beta\}$ in some fixed well-ordering of Y. This completes the inductive step.

Now let $D = \varphi([0, \Gamma[)$; because φ is injective, we have $\aleph(D) = \aleph(\mathscr{A})$. Furthermore, D is clearly a nbd-finite system of points in Y; since Y is Hausdorff, so that points are closed sets, III, **9.2** shows that D and each subset of D is closed in Y, consequently D is closed and discrete.

For continuous maps *into* Hausdorff spaces,

1.5 Let X be arbitrary, Y be Hausdorff, and $f, g: X \to Y$ be continuous. Then:
(1). $\{x \mid f(x) = g(x)\}$ is closed in X.
(2). If $D \subset X$ is dense, and $f \mid D = g \mid D$, then $f = g$ on X.
(3). The graph of the continuous $f: X \to Y$ is closed in $X \times Y$.
(4). If f is injective and continuous, then X is Hausdorff.

Proof: (1). $\{x \mid f(x) = g(x)\}$ is the inverse image of the closed $\Delta \subset Y \times Y$ under the continuous map $x \to (f(x), g(x))$ of $X \to Y \times Y$.

(2). The set on which f and g agree is a closed set containing the dense set D; by III, **4.13**, this must be X.

(3). The graph of f is the inverse image of the closed $\Delta \subset Y \times Y$ under the continuous map $(x, y) \to (f(x), y)$ of $X \times Y \to Y \times Y$.

(4). The inverse map $f^{-1}: f(X) \to X$ is a closed bijection of the Hausdorff space $f(X)$ onto X.

We now consider conditions under which an identification topology is Hausdorff. Given $p: X \to Y$ observe that $\mathscr{T}(p)$ is Hausdorff if and only if distinct fibers are contained in disjoint p-saturated open sets. Since this is a condition on both p and the topology in X, it cannot be expected that a simple requirement only on either p or the topology of X will suffice to assure $\mathscr{T}(p)$ is Hausdorff. One simple sufficient condition that frequently appears is

1.6 Let X be arbitrary, let R be an equivalence relation in X, and let $p: X \to X/R$ be the identification map. If both
(1). $R \subset X \times X$ is closed in $X \times X$,
(2). p is an open map,
then X/R is Hausdorff.

Proof: Let $p(x)$, $p(y)$ be distinct members of X/R; since x and y are not related, and $R \subset X \times X$ is closed, there is a nbd $U \times V$ of (x, y) such that $U \times V \subset \mathscr{C}R$. Thus, $p(U)$, $p(V)$ are disjoint, and since p is an open map, are open in X/R.

Ex. 5 The hypothesis (1) is essential: If X/R is Hausdorff, for any space X, then R *must* be closed in $X \times X$, since R is the inverse image of $\Delta \subset X/R \times X/R$ under the continuous map $(x, x') \to (p(x), p(x'))$ of $X \times X \to X/R \times X/R$.

One other particularly simple condition that assures a quotient space is Hausdorff is

1.7 Let Y be Hausdorff, X be arbitrary, and $f: X \to Y$ any continuous map. Then $X/K(f)$ is Hausdorff.

Proof: Since the map

$$g = fp^{-1}: X/K(f) \to Y$$

is continuous and injective, this follows from **1.5**(4).

2. Regular Spaces

A separation condition stronger than Hausdorff is obtained by replacing one of the points in **1.1** by a closed set:

2.1 ***Definition*** A Hausdorff space is regular (or: T_3) if each $y \in Y$ and closed set A not containing y have disjoint nbds; that is, if A is closed and $y \bar{\in} A$, then there is a nbd U of y and an open $V \supset A$ such that $U \cap V = \varnothing$.

Ex. 1 Discrete spaces, and E^n, are regular (the regularity of E^n is easily seen from **2.2** below).

Ex. 2 Every regular space is a Hausdorff space, but not conversely. Let E be the set of all real numbers, and $\mathscr{T}$ the topology having the open intervals and the set Q of rationals as subbasis. Since $\mathscr{T}$ is larger than the Euclidean topology of E, $(E, \mathscr{T})$ is a Hausdorff space. However, $\mathscr{T}$ is not regular: $\mathscr{C}Q$ is a closed set, but 1 and $\mathscr{C}Q$ do not have disjoint nbds. Observe also that, since $\mathscr{T}$ is larger than the Euclidean topology, this example shows that a topology larger than a regular topology need not be regular.

The definition has several equivalent formulations:

2.2 The following three properties are equivalent:

(1). Y is regular.

(2). For each $y \in Y$ and nbd U of y, there exists a nbd V of y with $y \in V \subset \overline{V} \subset U$.

(3). For each $y \in Y$ and closed A not containing y, there is a nbd V of y with $\overline{V} \cap A = \varnothing$.

Proof: (1) ⇒ (2). Given U, then y and the closed $\mathscr{C}U$ have nbds $V \supset y$, $W \supset \mathscr{C}U$ with $V \cap W = \varnothing$; thus $V \subset \mathscr{C}W$, so that $\overline{V} \subset \mathscr{C}W$ also, and from $\overline{V} \cap \mathscr{C}U \subset \overline{V} \cap W = \varnothing$, we find $\overline{V} \subset U$.

(2) ⇒ (3). Using y and its nbd $\mathscr{C}A$, find V satisfying $y \in V \subset \overline{V} \subset \mathscr{C}A$; then $\overline{V} \cap A = \varnothing$.

(3) ⇒ (1). Let A be closed, $y \bar{\in} A$. Choose a nbd $V \supset y$ such that $\overline{V} \cap A = \varnothing$; then $A \subset \mathscr{C}\overline{V}$, and $V \cap \mathscr{C}\overline{V} = \varnothing$.

Ex. 3 If the condition **2.2**(2) is known to hold only for all *subbasic* open U, the space is still regular: for if G is any given nbd of y, there is a finite intersection $\bigcap_1^n U_i$ of subbasic open U_i with $y \in \bigcap_1^n U_i \subset G$; from $y \in V_i \subset \overline{V}_i \subset U_i$ for each i, we find

$$y \in \bigcap_1^n V_i \subset \bigcap_1^n \overline{V}_i \subset \bigcap_1^n U_i \subset G.$$

For invariance properties we have

2.3 Theorem (1). Every subspace of a regular space is regular.
(2). $\prod \{Y_\alpha \mid \alpha \in \mathscr{A}\}$ is regular if and only if each Y_α is regular.

Proof: (1). Given $X \subset Y$, let $B \subset X$ be closed in X and $x_0 \in X - B$. Then $B = X \cap A$, where A is closed in Y, and since A does not contain x_0, there are disjoint open $U \supset x_0$, $V \supset A$. Then $U \cap X$ and $V \cap X$ are the required disjoint nbds of x_0 and B in X.

(2). As in **1.3**, it follows immediately from (1) that if $\prod Y_\alpha$ is regular, then each Y_α is regular. For the converse, let $\{y_\alpha\}$ be given and $\langle U_\alpha \rangle$ be any sub-basic nbd; choosing V_α so that $y_\alpha \in V_\alpha \subset \overline{V}_\alpha \subset U_\alpha$ gives

$$\{y_\alpha\} \in \langle V_\alpha \rangle \subset \overline{\langle V_\alpha \rangle} = \langle \overline{V}_\alpha \rangle \subset \langle U_\alpha \rangle$$

(*cf.* IV, **1.2**), and therefore, by Ex. 3, the regularity of $\prod_\alpha Y_\alpha$ follows.

The following special property of regular spaces is frequently useful:

2.4 Let Y be regular and $A \subset Y$ any infinite subset. Then there exists a family $\{U_n \mid n \geqslant 0\}$ of open sets whose closures are pairwise disjoint and such that $A \cap U_n \neq \varnothing$ for each $n \geqslant 1$.

Proof: We proceed by induction, taking $U_0 = \varnothing$. Assume that $U_0, \cdots, U_n$ have been defined so that $\overline{U}_0, \cdots, \overline{U}_n$ are pairwise disjoint, $A \cap U_k \neq \varnothing$ for $1 \leqslant k \leqslant n$, and $A_n = A - \bigcup_0^n \overline{U}_i$ is an infinite set. Choose $a, b \in A_n$; since Y is regular, we find first an open V such that

$a \in V \subset \overline{V} \subset Y - [\overline{U}_0 \cup \cdots \cup \overline{U}_n \cup \{b\}]$ and then an open W such that $b \in W \subset \overline{W} \subset Y - [\overline{U}_0 \cup \cdots \cup \overline{U}_n \cup \overline{V}]$. Now define $U_{n+1} = V$ if $\overline{V} \cap A$ is finite, and $U_{n+1} = W$ otherwise. Then $\overline{U}_0, \cdots, \overline{U}_{n+1}$ are pairwise disjoint, $A - \bigcup_0^{n+1} \overline{U}_i$ is infinite, and $A \cap U_k \neq \varnothing$ for $1 \leqslant k \leqslant n + 1$, so the inductive step is complete.

We now consider quotient spaces of a regular space X. Since the unit interval is regular, VI, **4**, Ex. 4, shows that if the projection map $p: X \to X/R$ is open, then $R \subset X \times X$ need not be open; however, for *closed* maps,

2.5 Let X be regular and let the projection $p: X \to X/R$ be a closed map. Then $R \subset X \times X$ is closed.

Proof: Let $(x, y) \bar{\in} R$; we are to find a nbd $(x, y) \in U \times W \subset \mathscr{C}R$, that is, open sets such that $p(U) \cap p(W) = \varnothing$. To this end, note that $(x, y) \bar{\in} R \Rightarrow p(x) \neq p(y) \Rightarrow x \bar{\in} p^{-1}p(y)$; since (**1.4**) y is a closed set and p is a closed map, $p^{-1}p(y)$ is closed, so there are disjoint open sets U, V with $x \in U$, and $p^{-1}p(y) \subset V$. Since p is a closed map, III, **11.2**, gives a nbd W of $p(y)$ with $p^{-1}p(y) \subset p^{-1}(W) \subset V$, and so $U \times p^{-1}(W)$ is the desired nbd of (x, y).

Ex. 4 The converse of **2.5** is *not* true: In E^1, define $x \sim 1/x$ $(x \neq 0)$ and $0 \sim 0$; then $R \subset X \times X$ is closed; yet $p: X \to X/R$ is not a closed map, since the p-load of the closed $\{x \mid |x| \geqslant 1\}$ is not closed.

2.6 Let X be regular and $p: X \to X/R$ be a closed and open map. Then X/R is *Hausdorff*.

Proof: Immediate from **1.6** and **2.5.**

For the identification of closed subsets of regular spaces to points we have the useful

2.7 Let X be regular and $A \subset X$ be closed. Then X/A is *Hausdorff*.

Proof: Let ξ, ξ' be two elements of X/A; if neither one is the element $[A]$, the existence of disjoint nbds follows by noting $X - A$ is Hausdorff. If $\xi' = [A]$, then $p^{-1}(\xi)$ is a single point and $p^{-1}(\xi) \bar{\in} A$; consequently, there are disjoint nbds of $p^{-1}(\xi)$ and A, the images of which are evidently open and provide the disjoint nbds of ξ and $[A]$.

3. Normal Spaces

Separation stronger than regularity is given by

3.1 Definition A Hausdorff space is normal (or: T_4) if each pair of disjoint closed sets have disjoint nbds.

Ex. 1 Discrete spaces, and E^n, are normal (the normality of E^n follows more simply later). We shall see later that the cartesian product of any family of closed intervals is normal (XI, **1**, Ex. 1).

Ex. 2 Let Γ be any ordinal number. With the topology in III, **3**, Ex. 5, $[0, \Gamma[$ is normal: Let A, B be disjoint closed sets. For each $\alpha \in A$, the set $\{\beta < \alpha \mid \beta \in B\}$ has a supremum b_α (II, **6.4**), which necessarily belongs to $\bar{B} = B$; note that $]b_\alpha, \alpha]$ is an open set containing no points of B. We thus get an open $U = \bigcup \{]b_\alpha, \alpha] \mid \alpha \in A\} \supset A$, and similarly, an open $V = \bigcup \{]a_\beta, \beta] \mid \beta \in B\} \supset B$. Now, U and V are disjoint: For, if $U \cap V \neq \varnothing$, then some $]b_\alpha, \alpha] \cap]a_\beta, \beta] \neq \varnothing$; assuming, say, that $\beta < \alpha$, this gives $\beta \in]b_\alpha, \alpha]$, which is impossible. Similarly, $[0, \Gamma]$, and E^1_u, the reals with the upper limit topology (III, **3**, Ex. 4) are normal spaces.

Ex. 3 Clearly, every normal space is regular; but the converse is not true. Since E^1_u is normal, and therefore regular, $E^1_u \times E^1_u$ is certainly regular; we will show that it is *not* normal. To do this, we first establish the following result of F. B. Jones: If a space Y contains a dense set D and a closed discrete subspace S with $\aleph(S) \geqslant 2^{\aleph(D)}$ then Y is not normal. Indeed, assume that Y were normal. Since every subset of S is closed in Y, we could find for each $A \subset S$ an open $U(A) \supset A$ and an open $V(S - A) \supset S - A$ such that $U \cap V = \varnothing$. We would then have $D \cap U(A) \neq D \cap U(B)$ whenever $A - B \neq \varnothing$: for, because D is dense in Y, the set $D \cap U(A) \cap V(S - B)$ is not empty, is contained in $D \cap U(A)$, and does not meet $U(B)$. This shows that the map $\mathscr{P}(S) \to \mathscr{P}(D)$ given by $A \to D \cap U(A)$ would be injective; but this is impossible, since $\aleph[\mathscr{P}(S)] > \aleph[\mathscr{P}(D)]$. With this result established, to see that $E^1_u \times E^1_u$ is not normal, we need observe only that $D = \{(r, s) \mid r, s, \text{ rational}\}$ is a countable dense subset, and that

$$S = \{(x, -x) \mid x \text{ irrational}\}$$

is a closed discrete subspace of cardinal $2^{\aleph_0}$.

The definition has several equivalent formulations:

3.2 The following four properties are equivalent:

(1). Y is normal.

(2). For each closed A and open $U \supset A$ there is an open V with $A \subset V \subset \bar{V} \subset U$.

(3). For each pair of disjoint closed sets A, B, there is an open U with $A \subset U$ and $\overline{U} \cap B = \varnothing$.
(4). Each pair of disjoint closed sets have nbds whose closures do not intersect.

The straightforward proofs are left for the reader.

In the invariance properties, we meet a situation different from those met before.

3.3 Theorem (1). Normality is invariant under continuous closed surjections.
(2). A subspace of a normal space need *not* be normal. However, a *closed* subspace is normal.
(3). The cartesian product of normal spaces need *not* be normal. However, if the product is normal, each factor must be normal.

Proof: (1). Let Y be normal and $p: Y \to Z$ be closed and continuous. Given disjoint closed A, B in Z, the normality of Y gives disjoint open sets with $p^{-1}(A) \subset U$, $p^{-1}(B) \subset V$. Because p is closed, III, **11.2**, assures that there exist open $U_A \supset A$, $V_B \supset B$ such that $p^{-1}(U_A) \subset U$, $p^{-1}(V_B) \subset V$ and U_A, V_B are evidently the required disjoint nbds of A and B.

(2). We give an example below; the proof of the second assertion is immediate from the observation that a set closed in a closed subspace is also closed in the entire space.

(3). Ex. 3 shows that the cartesian product of normal spaces need not be normal; we remark that it is in fact an open problem in general topology whether $Y \times I$ is normal whenever Y is normal. The second part of (3) follows from (2) and (1), as in **1.3**(3) because of **1.4**(3).

Ex. 4 Let Ω be the first uncountable ordinal, ω the first infinite ordinal, and take both $[0, \Omega]$ and $[0, \omega]$ with the topology in III, **3**, Ex. 5. It is not difficult to verify directly that $[0, \Omega] \times [0, \omega]$ is in fact normal, but we shall do this more simply later (XI, **6**, Ex. 2). Let T be the subspace $[0, \Omega] \times [0, \omega] - (\Omega, \omega)$; we show that T is *not* normal by proving the disjoint closed *in* T sets $A = \{(\Omega, n) \mid 0 \leqslant n < \omega\}$, $B = \{(\xi, \omega) \mid 0 \leqslant \xi < \Omega\}$ cannot be separated. Indeed, let U be any nbd of A; since for each fixed n the point $(\Omega, n) \in U$, there is an ordinal $\alpha_n < \Omega$ such that $]\alpha_n, \Omega] \times n \subset U$. By II, **9.1**, the countable collection $\{\alpha_n\}$ has an upper bound $\alpha_0 < \Omega$ so that the "tube" $]\alpha_0, \Omega] \times [0, \omega[\subset U$. It follows that any nbd of $(\alpha_0 + 1, \omega) \in B$ must contain points of U; therefore each $V \supset B$ will intersect U.

For identification topologies we have

3.4 Let X, Y be normal, $A \subset X$ closed, and $f: A \to Y$ continuous. Then $X \cup_f Y$ is normal.

Proof: Let $F_1, F_2 \subset X \cup_f Y$ be disjoint closed sets. For $i = 1, 2$, there are open $V_i \subset Y$ such that $Y \cap p^{-1}(F_i) \subset V_i$ and $\overline{V}_1 \cap \overline{V}_2 = \varnothing$. Since $p(\overline{V}_i)$ is closed in $X \cup_f Y$ (VI, **6.3**), there are also disjoint open $U_i \subset X$ with $X \cap p^{-1}[F_i \cup p(\overline{V}_i)] \subset U_i$. The sets $p[(U_i - A) \cup V_i]$ are disjoint and $p[(U_i - A) \cup V_i] \supset F_i$; we will show them open in $X \cup_f Y$. Noting that $Y \cap p^{-1}p[(U_i - A) \cup V_i] = V_i$, and

$$X \cap p^{-1}p[(U_i - A) \cup V_i] = (U_i - A) \cup f^{-1}(V_i),$$

we need prove only that $(U_i - A) \cup f^{-1}(V_i)$ is open in X. To this end, observe $X \cap p^{-1}p(V_i) \subset U_i$ so that $f^{-1}(V_i) \subset U_i$ and, being open in A, $f^{-1}(V_i) = W_i \cap A$. Thus

$$\begin{aligned}(U_i - A) \cup f^{-1}(V_i) &= (U_i - A) \cup (U_i \cap W_i \cap A)\\ &= U_i \cap [W_i \cup \mathscr{C}A],\end{aligned}$$

and so is open in X. That points are closed sets (so that in particular $X \cup_f Y$ is Hausdorff) is trivial.

3.5 Let X be normal, and $A \subset X$ closed. Then X/A is normal.

Proof: Immediate from **3.4**, since X/A is obtained by attaching X to a point q by the constant map $f: A \to q$.

There is a separation property stronger than normality: A Hausdorff Y is *completely normal* if every pair of sets A, B satisfying $\overline{A} \cap B = A \cap \overline{B} = \varnothing$ can be separated. It is easy to see from the proof given in Ex. 2 that $[0, \Omega]$, $[0, \Omega[$, and E^1_u are indeed completely normal. Example 4 can also be regarded as stating that the normal space $[0, \Omega] \times [0, \omega]$ is not completely normal. We shall not deal extensively with complete normality.

4. Urysohn's Characterization of Normality

It is not true that a nonconstant continuous real-valued function can be defined on any given space. For example, since the spaces in **1**, Ex. 3, are connected and countable, each continuous $f: X \to E^1$ must be a constant map; there are examples of regular spaces for which this occurs. One reason for the importance of normal spaces is that one can be sure nonconstant continuous real-valued functions exist on them.

4.1 Theorem (P. Urysohn) Let Y be Hausdorff. The following two properties are equivalent:

(1). Y is normal.

(2). For each pair of disjoint closed sets, A, B in Y, there exists a continuous $f: Y \to E^1$, called a Urysohn function for A, B, such that:

(a). $0 \leqslant f(y) \leqslant 1$ for all $y \in Y$.
(b). $f(a) = 0$ for all $a \in A$.
(c). $f(b) = 1$ for all $b \in B$.

Proof: (2) $\Rightarrow$ (1). Let A, $B \subset Y$ be disjoint closed sets. Using a Urysohn function f for this pair, $U = \{y \mid f(y) < \frac{1}{2}\}$, $V = \{y \mid f(y) > \frac{1}{2}\}$ are disjoint open sets with $U \supset A$, $V \supset B$.

(1) $\Rightarrow$ (2). Let R be the set of all rationals r of form $k/2^n$, $0 \leqslant k/2^n \leqslant 1$. We first show that with $r \in R$, we can associate an open $U(r) \subset Y$ such that:

(i). $A \subset U(r)$ and $U(r) \cap B = \varnothing$.
(ii). $r < r' \Rightarrow \overline{U(r)} \subset U(r')$; that is, the ordering in R is identical to set inclusion (with closure).

We proceed by induction on the exponent of the dyadic fractions, letting

$$D_m = \left\{U\left(\frac{k}{2^m}\right) \,\middle|\, k = 0, 1, \cdots, 2^m\right\}.$$

D_0 consists of $U(1) = \mathscr{C}B$ and $U(0) =$ some open set satisfying

$$A \subset U(0) \subset \overline{U(0)} \subset \mathscr{C}B,$$

which exists by normality. Assume D_{m-1} constructed, and note that only $U(k/2^m)$ for *odd* k requires definition; for each odd k, we have from D_{m-1} that $\overline{U((k-1)/2^m)} \subset U((k+1)/2^m)$, so we define $U(k/2^m)$ to be an open set U satisfying

$$\overline{U\left(\frac{k-1}{2^m}\right)} \subset U \subset \overline{U} \subset U\left(\frac{k+1}{2^m}\right).$$

This completes the inductive step.

To define f, replace $U(1)$ by Y, let $D = \bigcup_m D_m$, and let

$$f(y) = \inf\{r \mid y \in U(r)\}.$$

Clearly, $0 \leqslant f(y) \leqslant 1$. Furthermore $f(B) = 1$, since each $b \in U(1) = Y$ only; and $f(A) = 0$, since each $a \in U(r)$ for all r. It remains to prove continuity. Let $f(y_0) = r_0$, and choose any $W =]r_0 - \varepsilon, r_0 + \varepsilon[$. If $r_0 \neq 0$ or 1, choose also $\bar{r}$, $\underline{r}$ such that $r_0 - \varepsilon < \underline{r} < r_0 < \bar{r} < r_0 + \varepsilon$; then $U = U(\bar{r}) - \overline{U(\underline{r})}$ is a nbd of y_0 and $f(U) \subset W$, since $y \in U(\bar{r}) \Rightarrow f(y) \leqslant \bar{r}$, and $y \mathrel{\bar{\in}} \overline{U(\underline{r})} \Rightarrow f(y) \geqslant \underline{r}$. If $r_0 = 0$ (or 1) then the nbd $U(\bar{r})$ (or $\mathscr{C}\overline{U(\underline{r})}$) of y_0 alone suffices. This completes the proof.

Remark: The pair of numbers 0 and 1 can be obviously replaced by any (not necessarily nonnegative) pair $\alpha < \beta$, since the continuity of f implies that of $\alpha + (\beta - \alpha)f$.

The Urysohn function f in **4.1** evidently satisfies $A \subset f^{-1}(0)$. The theorem does *not* assert that $A = f^{-1}(0)$; in fact, this is possible only for certain types of closed sets:

4.2 Corollary A necessary and sufficient condition for the existence of a Urysohn function satisfying $A = f^{-1}(0)$ is that A be a G_δ.

Proof: If there is such a function, $A = \bigcap_n \{y \mid f(y) < (1/n)\}$ shows A to be a G_δ. Conversely, assume $A = \bigcap_1^\infty U_i$; it is no restriction (III, **6.2**) to assume that $U_1 \supset U_2 \supset \cdots$ and that $U_1 \cap B = \varnothing$. Let f_n be a Urysohn function for the closed sets $A, \mathscr{C}U_n$, where $f_n(A) = 0$. By III, **10.5**, the function $f(y) = \sum_1^\infty \left(\frac{1}{2^n}\right) f_n(y)$ is continuous on Y and is evidently a Urysohn function for A, B. Since

$$y \,\bar{\in}\, A \Rightarrow y \,\bar{\in} \bigcap_1^\infty U_i \Rightarrow \exists\, n_0 : y \,\bar{\in}\, U_{n_0} \Rightarrow f_{n_0}(y) = 1 \Rightarrow f(y) \geqslant \frac{1}{2^{n_0}},$$

we have $f^{-1}(0) = A$.

4.3 Corollary A necessary and sufficient condition that there be a Urysohn function f with $A = f^{-1}(0)$, $B = f^{-1}(1)$ is that both A and B be G_δ-sets.

Proof: If both A and B are G_δ-sets, find Urysohn functions f, g for the pair A, B such that $f^{-1}(0) = A$, $g^{-1}(0) = B$; then

$$\varphi(x) = \frac{f(x)}{f(x) + g(x)}$$

is the required function. Because $A \cap B = \varnothing$, the denominator never vanishes, so φ is indeed continuous. The converse is trivial.

A normal topological space in which each closed set is a G_δ is called *perfectly normal.* The ordinal space $[0, \Omega]$ is an example of a completely normal space that is not perfectly normal: as we have seen in III, **6**, Ex. 4, the closed set $\{\Omega\}$ is not a G_δ. However,

4.4 Every perfectly normal space is completely normal.

Proof: Let Y be perfectly normal, and let A, B be subsets with $\bar{A} \cap B = A \cap \bar{B} = \varnothing$. Find $f, g \colon Y \to I$ vanishing only on $\bar{A}$, $\bar{B}$, respectively, and set $U = \{y \mid f(y) < g(y)\}$, $V = \{y \mid g(y) < f(y)\}$. Clearly, U, V are open sets, and $U \cap V = \varnothing$. Furthermore, $A \subset U$, since if $y \in A$, then $f(y) = 0$, and, because $\bar{B} \cap A = \varnothing$, we have $g(y) > 0$; similarly, $B \subset V$, completing the proof.

5. Tietze's Characterization of Normality

Let X, Y be two spaces, $A \subset X$ closed, and $f \colon A \to Y$ continuous. A continuous $F \colon X \to Y$ such that $F \mid A = f$ is called an extension of f over X relative to Y. A "general extension theorem for maps of closed subsets of X into Y" is a statement giving conditions on X and Y, under which it is true that for every closed $A \subset X$, each continuous $f \colon A \to Y$ is extendable over X relative to Y. General extension theorems are rare and usually have interesting topological consequences.

It is important to realize that the requirement that Y *not* be enlarged is essential. Without some such restriction, there is no problem, since by simply attaching X to Y by f, the map $F = p \mid X \colon X \to X \cup_f Y$ is an extension of f over X (relative to $X \cup_f Y$!). However, once Y has been fixed, a genuine question arises: It is generally difficult to decide if a given map $A \to Y$ is extendable over X and, indeed, some may not be. For example, if $A \subset I$ is $\{0, 1\}$, the map $f(0) = 0, f(1) = 1$ of $A \to 2$ is obviously not extendable over I (relative to 2).

It is easy to see that if Y is Hausdorff (with at least two points) and if there is a general extension theorem for maps of closed subsets of X into Y, then X must be normal. For, given disjoint closed A, $B \subset X$, the map $f \colon A \cup B \to Y$ sending A to y_0 and B to $y_1 \neq y_0$ is, by hypothesis, extendable to a continuous $F \colon X \to Y$, and if U, V are disjoint nbds of y_0, y_1, then $F^{-1}(U)$, $F^{-1}(V)$ are disjoint nbds of A, B.

Urysohn's theorem can be regarded as a special case of the following general extension theorem.

5.1 Theorem (H. Tietze) Let X be Hausdorff. The following two properties are equivalent:

(1). X is normal.
(2). For every closed $A \subset X$, each continuous $f \colon A \to E^1$ has a continuous extension $F \colon X \to E^1$. Furthermore, if $|f(a)| < c$ on A, then F can be chosen so that $|F(x)| < c$ on X.

Proof: (2) $\Rightarrow$ (1). This conclusion follows from the preceding remarks, since E^1 is Hausdorff.

(1) ⇒ (2). To keep the discussion clear, we state as a lemma the main tool used in the construction of F:

Lemma: Let $g: A \to E^1$ be continuous, $|g(a)| \leqslant c$ for all $a \in A$. Then there exists a continuous $h: X \to E^1$ such that
(1). $|h(x)| \leqslant \frac{1}{3}c$ for all $x \in X$.
(2). $|g(a) - h(a)| \leqslant \frac{2}{3}c$ for all $a \in A$.

Proof of Lemma: Let

$$A_+ = \{a \in A \mid g(a) \geqslant \tfrac{1}{3}c\}, \quad A_- = \{a \in A \mid g(a) \leqslant -\tfrac{1}{3}c\};$$

these sets are clearly disjoint and, being closed in the closed set $A \subset X$, are also closed in X. Since X is normal, a Urysohn function $h: X \to E^1$ having value $\frac{1}{3}c$ on A_+, $-\frac{1}{3}c$ on A_- and satisfying $-\frac{1}{3}c \leqslant h(x) \leqslant \frac{1}{3}c$ on X, evidently fulfills the requirements.

The proof of (1) ⇒ (2) is accomplished in three steps.

(a). $|f(a)| \leqslant c$ on A. Using f as the given g in the lemma, find the function $h_0: X \to E^1$. On A we thus have $|f(a) - h_0(a)| \leqslant \frac{2}{3}c$. Apply the lemma again, this time to the function $f - h_0$ defined *on* A, to get $h_1: X \to E^1$ satisfying

$$|h_1(x)| \leqslant \tfrac{1}{3}\cdot\tfrac{2}{3}c \qquad x \in X,$$
$$|f(a) - h_0(a) - h_1(a)| \leqslant \tfrac{2}{3}\cdot\tfrac{2}{3}c \qquad a \in A.$$

Proceeding by induction, assume $h_0, \cdots, h_n$ to be defined; apply the lemma to the function $g = f - h_0 - \cdots - h_n$ *on* A to get $h_{n+1}: X \to E^1$ satisfying

$$|h_{n+1}(x)| \leqslant \tfrac{1}{3}(\tfrac{2}{3})^n c \qquad x \in X,$$
$$|f(a) - h_0(a) - \cdots - h_{n+1}(a)| \leqslant \tfrac{2}{3}(\tfrac{2}{3})^n c \qquad a \in A.$$

We thus have a function $h_n: X \to E^1$ for each $n \in N$. By III, **10.5**, the function $F(x) = \sum_0^\infty h_n(x)$ is continuous on X; the second inequality shows $F(a) = f(a)$ for each $a \in A$, and the first shows that, on X, $|F(x)| \leqslant \frac{1}{3}c\cdot\sum_0^\infty (\frac{2}{3})^n = c$.

(b). $|f(a)| < c$ on A. The extension F constructed in (a) satisfies $|F(x)| \leqslant c$. Let $A_0 = \{x \mid |F(x)| = c\}$; then A_0 is closed in X, and $A_0 \cap A = \varnothing$. Therefore there exists a Urysohn function $\varphi: X \to E^1$ having value 1 on A and 0 on A_0, with $0 \leqslant \varphi(x) \leqslant 1$ on X. Define $G(x) = \varphi(x)\cdot F(x)$; this is continuous (III, **10.3**), and since $\varphi(a) = 1$ for each $a \in A$, we have $G(a) = F(a) = f(a)$ for each $a \in A$, showing that G is also an extension of f over X relative to E^1. Furthermore, $|G(x)| < c$ on X: if $x \in A_0$, then $G(x) = 0$, and if $x \in X - A_0$, then $|\varphi(x)| \leqslant 1$ and $|F(x)| < c$.

(c). f is not necessarily bounded. Let $h: E^1 \to]-1, +1[$ be the homeomorphism $x \to x/(1 + |x|)$. By (b), the map $h \circ f: A \to]-1, +1[$ has an extension $F: X \to]-1, +1[$, and then $h^{-1} \circ F$ is an extension of f, since $h^{-1} \circ F(a) = h^{-1} \circ h \circ f(a) = f(a)$, for each $a \in A$. The theorem has been proved.

Remark: It is frequently important to know that if $f(A) \subset]a, b[$ an extension F can be found so that $F(X) \subset]a, b[$ also. This follows from the theorem by translating the origin in E^1 to $(a + b)/2$.

A space that can be substituted for E^1 in Tietze's theorem is called an absolute retract for normal spaces, abbreviated as "AR (normal)". Formally, Y is an AR (normal) if for every normal X and closed $A \subset X$, each continuous $f: A \to Y$ has an extension $F: X \to Y$. There are many AR (normal):

5.2 Corollary Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces. $\prod_\alpha Y_\alpha$ is an AR (normal) if and only if each Y_α is an AR (normal). In particular, E^n, I^n, and I^∞ are AR (normal).

Proof: Let X be normal and $A \subset X$ closed. Assume that each Y_α is an AR (normal) and $f: A \to \prod_\alpha Y_\alpha$ continuous; then (IV, **2.2**) each $p_\alpha \circ f: A \to Y_\alpha$ is continuous, so is extendable to a continuous $F_\alpha: X \to Y_\alpha$, and $x \to \{F_\alpha(x)\}$ is an extension of f over X. Conversely, assume $\prod_\alpha Y_\alpha$ is an AR (normal) and $f: A \to Y_\beta$ is continuous. Let $s_\beta: Y_\beta \to \prod_\alpha Y_\alpha$ be a homeomorphism onto a slice $S(y^\circ; \beta) \subset \prod_\alpha Y_\alpha$ (IV, **3.2**); then $s_\beta \circ f: A \to \prod_\alpha Y_\alpha$ is extendable to $F: X \to \prod_\alpha Y_\alpha$, and because $p_\beta \circ s_\beta = 1$, it follows that $p_\beta \circ F$ is an extension of f over X relative to Y_β.

If we replace E^n by the n-sphere S^n, the conclusion of **5.1** is different:

5.3 Corollary Let X be normal, $A \subset X$ closed, and $f: A \to S^n$ continuous. Then there exists a nbd $U \supset A$ (U depends on f) over which f can be extended (relative to S^n!).

Proof: We consider $f: A \to S^n \subset E^{n+1}$; by **5.2**, f has an extension $F: X \to E^{n+1}$. Let $U = \{x \mid F(x) \neq 0\}$, which is open and contains A. On U, define

$$\hat{F}(x) = \frac{F(x)}{|F(x)|};$$

then $\hat{F}: U \to S^n$ is clearly an extension of f.

No improvement can be made on **5.3** unless restrictions are placed on X or on f; the map f cannot in general be extended over X. For $n = 0$ and connected X, this is obvious, and we shall see later that if V^{n+1} is the ball bounded by S^n, the identity map $1: S^n \to S^n$ cannot be extended to a map $V^{n+1} \to S^n$ (this is equivalent to Brouwer's fixed-point theorem).

Ex. 1 This illustrates a restriction on f that allows an extension over X in **5.3**: If $f(A) \neq S^n$, then f is extendable over X, and an extension F can be chosen so that $F(X) \neq S^n$ also. For, selecting $p \in S^n - f(A)$, let $\mu: S^n - p \cong E^n$ be a stereographic projection from p; then $\mu \circ f: A \to E^n$ is extendable to $F: X \to E^n$, and $\mu^{-1} \circ F: X \to S^n$ is an extension of f.

Any space that can be used for S^n in **5.3** is called an absolute nbd retract for normal spaces, written ANR (normal). Formally, Y is an ANR (normal) if for every normal space X and closed $A \subset X$, each continuous $f: A \to Y$ can be extended over a nbd $U \supset A$ (relative to Y!). Clearly, every AR (normal) is an ANR (normal), but not conversely.

5.4 Corollary Let $\{Y_i \mid i = 1, \cdots, n\}$ be any *finite* family of spaces. Then $\prod_1^n Y_i$ is an ANR (normal) if and only if each Y_i is an ANR (normal).

Proof: Let X be normal and $A \subset X$ be closed. Assume that each Y_i is an ANR (normal) and $f: A \to \prod_1^n Y_i$; since $p_i \circ f: A \to Y_i$ has an extension F_i over some $U_i \supset A$, $x \to \{F_i(x)\}$ is an extension of f over $\bigcap_1^n U_i$. The converse follows as in **5.2**.

6. Covering Characterization of Normality

A covering $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ of a space Y is called *point-finite* if for each $y \in Y$ there are at most finitely many indices $\alpha \in \mathscr{A}$ such that $y \in A_\alpha$. The normal spaces are characterized by the "shrinkability" of such *open* coverings:

6.1 Theorem The following two properties are equivalent:

(1). Y is normal.

(2). If $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ is any point-finite covering of Y by open sets, there exists a covering $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ of Y by open sets such that $\overline{V}_\alpha \subset U_\alpha$ for each $\alpha \in \mathscr{A}$, and $V_\alpha \neq \varnothing$ whenever $U_\alpha \neq \varnothing$.

Proof: (1) $\Rightarrow$ (2). Well-order the indexing set $\mathscr{A}$ and for each $y \in Y$, set $h(y) = \sup\{\alpha \mid y \in U_\alpha\}$; $h(y) \in \mathscr{A}$ is well-defined, since each

Well-order $\mathscr{P}(Y)$; we will define a map $\varphi: \mathscr{A} \to \mathscr{P}(Y)$ by transfinite construction, so that $\varphi(\alpha) = V_\alpha$ is an open set for each α and

(a). $\overline{V}_\alpha \subset U_\alpha$, $V_\alpha \neq \varnothing$ whenever $U_\alpha \neq \varnothing$.

(b). $\{V_\beta \mid \beta \leqslant \alpha\} \cup \{U_\gamma \mid \gamma > \alpha\}$ is a covering of Y.

Assume that $\varphi(\beta)$ is defined for all $\beta < \alpha$. Note first that

$$\{V_\beta \mid \beta < \alpha\} \cup \{U_\gamma \mid \gamma \geqslant \alpha\}$$

is also a covering of Y: indeed, given $y \in Y$, if $h(y) \geqslant \alpha$, then $y \in \bigcup \{U_\gamma \mid \gamma \geqslant \alpha\}$, and if $h(y) < \alpha$, then $y \,\overline{\in}\, \bigcup \{U_\gamma \mid \gamma > h(y) > \alpha\}$ and (b) shows $y \in V_\tau$ for some $\tau \leqslant h(y) < \alpha$. From this observation follows that

$$F = Y - [\bigcup_{\beta < \alpha} V_\beta \cup \bigcup_{\gamma > \alpha} U_\gamma] \subset U_\alpha$$

and, since F is closed, there is an open V with $F \subset V \subset \overline{V} \subset U_\alpha$ (if $F = \varnothing$, replace F by a point in U_α); letting $\varphi(\alpha) = V_\alpha$ be the first such V in the well-ordering of $\mathscr{P}(Y)$, the conditions (a) and (b) are evidently satisfied by the new family. Thus, according to II, **5.2**, we have a uniquely defined map $\varphi: \mathscr{A} \to \mathscr{P}(Y)$ with $\overline{V}_\alpha \subset U_\alpha$ for each $\alpha \in \mathscr{A}$. It remains to show that $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ is a covering of Y: given $y \in Y$, we have $y \,\overline{\in}\, \bigcup \{U_\gamma \mid \gamma > h(y)\}$ therefore, and because of (b), y belongs to some V_β with $\beta \leqslant h(y)$.

(2) $\Rightarrow$ (1). Let A, B be disjoint closed sets in Y. Then

$$\{Y - A, Y - B\}$$

is evidently a point-finite open covering and hence can be shrunk. Since $\overline{V}_1 \subset Y - A$, $\overline{V}_2 \subset Y - B$, we find that $\mathscr{C}\overline{V}_1$, $\mathscr{C}\overline{V}_2$ are nbds of A, B, respectively, and $\mathscr{C}\overline{V}_1 \cap \mathscr{C}\overline{V}_2 = \mathscr{C}(\overline{V}_1 \cup \overline{V}_2) = \mathscr{C}Y = \varnothing$. Thus Y is normal.

7. Completely Regular Spaces

Since subspaces of normal spaces need not be normal, the question arises whether they can be directly characterized by some topological property. We will see in this section that they do in fact form a well-defined class, between the regular and the normal spaces, characterized by the following weaker form of **4.1**:

7.1 Definition A Hausdorff space Y is completely regular (or Tychonoff) if for each point $p \in Y$ and closed A not containing p, there is a continuous $\varphi: Y \to I$ such that $\varphi(p) = 1$ and $\varphi(a) = 0$ for each $a \in A$.

Clearly, the definition can be equally well formulated in terms of p and nbds $U \supset p$ by taking $A = \mathscr{C}U$.

Ex. 1 Every subspace of a normal space is completely regular. For, if Y is normal and $B \subset Y$ is a subspace, then for a given b and closed C in B not containing applies to the closed sets b, F. This result and **3**, Ex. 4, also shows that a completely regular space need not be normal.

Ex. 2 Every completely regular space is regular. For, given p and a nbd $U \supset p$, find the function φ for p and $\mathscr{C}U$; since $V = \{y \mid \varphi(y) > \frac{1}{2}\}$ satisfies $p \in V \subset \bar{V} \subset U$ **2.2**(2) gives the result.

Ex. 3 A regular space may not be completely regular. Let (*cf.* **3**, Ex. 4) $T = [0, \Omega] \times [0, \omega] - (\Omega, \omega)$ and for each $n \in Z$ let $T_n = T \times \{n\}$; denote the elements of T_n by $(\alpha, x; n)$. In the free union (VI, **8.4**) $\sum_{-\infty}^{\infty} T_n$, make the identifications $(\Omega, x; 2k + 1) \sim (\Omega, x; 2k + 2)$ and $(\alpha, \omega; 2k + 1) \sim (\alpha, \omega; 2k)$ for each $k \in Z$, α, and x, to get a "spiral staircase" space S. Let $Y = S \cup a \cup b$, where a, b, are two objects not in S, and take the nbds of a (resp. b) to be all sets of form $\{a\} \cup \mathscr{C} \bigcup \{T_n \subset S \mid n \leqslant N\}$ (resp. $\{b\} \cup \mathscr{C} \bigcup \{T_n \subset S \mid n \geqslant N\}$). Clearly, Y is a regular space; we will show that Y is not completely regular.

First note that for a continuous $f: T \to E^1$,

(i) If $f(\alpha, \omega) \geqslant r > s$ for all large α, then $f(\Omega, x) > s$ for all large x.
(ii) If $f(\Omega, x) \geqslant r > s$ for all large x, then $f(\alpha, \omega) > s$ for all large α.

Indeed, one easily verifies (*cf.* XI, **3**, Ex. 2) that f has a continuous extension $F: [0, \Omega] \times [0, \omega] \to E^1$; the hypothesis of either (i) or (ii) gives $F(\Omega, \omega) \geqslant r$ since F takes values $\geqslant r$ in each nbd of (Ω, ω); and because (Ω, ω) has a nbd mapped by F into $]s, \infty[$, the conclusions follow.

Now let $f: Y \to E^1$ be any continuous function with $f(a) = 1$; we will prove that $f(b) \geqslant 0$. For, f is $\geqslant \frac{1}{2}$ on some nbd of a, and therefore also on some T_{2k+1}. Thus, $f(\alpha, \omega; 2k + 1) = f(\alpha, \omega; 2k) \geqslant \frac{1}{2}$; using (i) gives $f(\Omega, x; 2k) = f(\Omega, x; 2k - 1) > \frac{1}{3}$ for all large x; then (ii) gives $f(\alpha, \omega; 2k - 1) = f(\alpha, \omega; 2k - 2) > \frac{1}{4}$ for all large α. By induction, $f(\alpha, \omega; 2k - N) > 1/2N$ for each $N \in Z^+$ and all sufficiently large α; thus, f is positive at points in each nbd of b, so $f(b) \geqslant 0$. This result implies that Y is not completely regular: for, if it were, a continuous $F: Y \to E^1$ with $F(a) = 1$, $F(b) = 0$ would exist, and then $f = 2F - 1$ would be a continuous function on Y with $f(a) = 1$, $f(b) = -1$. [More generally, $f(a) = f(b)$ for each continuous $f: Y \to E^1$.]

For invariance properties, we have

7.2 Theorem (1). Every subspace of a completely regular space is completely regular.
(2). $\prod \{Y_\alpha \mid \alpha \in \mathscr{A}\}$ is completely regular if and only if each Y_α is completely regular.

Proof: Ad (1). Let $B \subset Y$ be a subspace and $p \in V$, where V is open in B. Since $V = B \cap U$, where U is open in Y, then letting $\varphi: Y \to I$ satisfy **7.1** for p and $\mathscr{C}U$, it is clear that $\varphi \mid B$ satisfies the requirement for p and $\mathscr{C}_B V$.

Ad (2). That the complete regularity of $\prod_\alpha Y_\alpha$ implies that of each Y_α follows in the usual manner from (1), since complete regularity is a topological invariant. For the converse, let $y = \{y_\alpha\}$ and $\langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle$ be any nbd of y. For each $i = 1, 2, \cdots, n$, let $\varphi_i: Y \to I$ be a map with

$$g(y) = \min\{\varphi_i \circ p_{\alpha_i}(y) \mid i = 1, \cdots, n\}$$

which, according to III, **10.4**, is continuous. Then $g(y) = 1$ and it is evident that g vanishes outside $\langle U_{\alpha_1}, \cdots, U_{\alpha_n}\rangle$.

Ex. 4 Any cartesian product of unit intervals is called a *parallelotope*. By **3**, Ex. 1 and **7.2**(1), (2), every parallelotope and all subsets of parallelotopes are completely regular (we will prove later that all parallelotopes are in fact normal).

For any space Y, let I^Y denote the set of all continuous maps $f\colon Y \to I$. Let $\{I_f \mid f \in I^Y\}$ be a family of unit intervals indexed by I^Y, and let P^Y be the parallelotope $\prod \{I_f \mid f \in I^Y\}$; the points of P^Y are denoted by $\{t_f\}$.

The following theorem is a converse of Ex. 4.

7.3 Theorem If Y is completely regular, then it can be embedded in a parallelotope. Precisely, the map $\rho\colon Y \to P^Y$, defined by $\rho(y) = \{f(y)_f\}$ is a homeomorphism of Y and $\rho(Y) \subset P^Y$.

Proof: ρ is injective, for if $x \neq y$, there is a nbd containing x but not y, so by complete regularity, there is an $f \in I^Y$ with $f(x) = 1, f(y) = 0$; then $\rho(x) \neq \rho(y)$, since they differ at the fth coordinate.

ρ is continuous: for the projection of ρ on the fth axis is $p_f \circ \rho(y) = f(y)$, which is continuous.

$\rho\colon Y \to \rho(Y)$ is an open map. First observe that the family of open sets $\{V_f \mid V_f = f^{-1}]\,0, 1]\}$ in Y form a basis: For any open U and $p \in U$ there is, by regularity, a V with $p \in V \subset \overline{V} \subset U$ and therefore an $f\colon Y \to I$ with $f(p) = 1$, $f(\mathscr{C}V) = 0$ so that $p \in V_f \subset U$ (*cf.* III, **2.2**). Since for any basic set V_{f_0}, we have $\rho(V_{f_0}) = \{\{t_f\} \mid t_{f_0} > 0\} \cap \rho(Y)$; this shows that $\rho(V_{f_0})$ is open in $\rho(Y)$, and, by III, **11.3**, proves that ρ is an open map.

Since Y and $\rho(Y)$ are homeomorphic, it is evident that for any space Z, each continuous $h\colon Y \to Z$ is uniquely of the form $g \circ \rho$, where $g\colon \rho(Y) \to Z$ is continuous. If Z is also completely regular, we have

7.4 Corollary Let Y, Z be completely regular spaces, and let $h\colon Y \to Z$ be continuous. Then there exists a continuous $H\colon P^Y \to P^Z$ (H is defined on the entire parallelotope P^Y) such that the diagram

$$\begin{array}{ccc} Y & \xrightarrow{\ h\ } & Z \\ {\scriptstyle\rho}\big\downarrow & & \big\downarrow{\scriptstyle\rho_1} \\ P^Y & \xrightarrow[\ H\ ]{} & P^Z \end{array}$$

Proof: Since for each $g \in I^Z$ we have $g \circ h \in I^Y$, define $h_g: P^Y \to I_g$ by $h_g\{t_f\} = t_{g \circ h}$ (that is, project P^Y onto the $g \circ h$-axis and identify $I_{g \circ h}$ with I_g). Since each h_g is continuous, we obtain a continuous map $H: P^Y \to P^Z$ by $H\{t_f\} = \{h_g\{t_f\}\}$. From the definition of H we have that

$$H \circ \rho(y) = H(\{f(y)_f\}) = \{h_g\{f(y)_f\}\} = \{(g \circ h(y))_g\},$$

and from **7.3** that

$$\rho_1 \circ h(y) = \rho_1(h(y)) = \{g(h(y))_g\}$$

so that the diagram is indeed commutative. Since the commutativity implies that $H(\rho(Y)) \subset \rho_1(Z)$, the continuity of H on P^Y shows that $H(\overline{\rho(Y)}) \subset \overline{H(\rho(Y))} \subset \overline{\rho_1(Z)}$, and the proof is complete.

Problems

Section 1

1. Let Y be Hausdorff. Prove:
   a. $\bigcap \{F \mid (p \in F) \wedge (F \text{ is closed})\} = p$.
   b. $\bigcap \{U \mid (p \in U) \wedge (U \text{ is open})\} = p$.

   Give examples to show that *neither* of these two properties is equivalent to "Hausdorff."
2. Let $\{x_1, \cdots, x_n\}$ be a finite subset of a Hausdorff space. Show that there exist pairwise disjoint nbds $U(x_1), \cdots, U(x_n)$.
3. Let X be a finite set. Prove that the only Hausdorff topology $\mathscr{T}$ in X is the discrete topology.
4. Prove that in Hausdorff spaces: (a) A' is always closed; (b) $(A')' \subset A'$; and (c) $(\bar{A})' = A'$.
5. Let $f: X \to Y$, $g: Y \to X$ be continuous, with $g \circ f = 1_X$. Prove: If Y is Hausdorff so also is X, and $f(X)$ is closed in Y.
6. Let $Y = I \cup \{\xi\}$, where $\xi \bar{\in} I$, with the identification topology determined by $p: [-1, +1] \to Y$, where $x \to |x|$, $x \neq -1$, and $p(-1) = \xi$. Show that (a) p is an open map; (b) Y is not Hausdorff; (c) the relation $K(p)$ is not closed in $X \times X$.
7. Prove: A necessary and sufficient condition that points be closed sets is that the topology be T_1.
8. Prove: X/R is T_1 if and only if each equivalence class is closed in X.
9. A point y_0 of a connected Hausdorff Y is called a *dispersion point* if $Y - \{y_0\}$ is totally disconnected. Prove: Y can have at most one dispersion point.
10. Let X be Hausdorff, let $f: X \to Y$ be continuous, and let $D \subset X$ be dense. Assume $f \mid D$ is a homeomorphism into Y. Prove: $f(X - D) \subset Y - f(D)$.
11. Let X be Hausdorff and $f: X \to X$ be continuous. Prove: $\{x \mid f(x) = x\}$ is closed in X.
12. Prove: Every infinite Hausdorff space contains a countably infinite discrete subspace.

13. Let D be a dense subset in each of the two Hausdorff spaces X and Y. Let $1: D \to D$ be extendable to a continuous $f: X \to Y$ and also to a continuous $g: Y \to X$. Prove that f and g are homeomorphisms and that $f = g^{-1}$.
14. Let Y be Hausdorff. Prove that the cone TY is Hausdorff.
15. Show that the space $(Z^+, \mathscr{T})$ of **V**, Problem 1. 10 is a Hausdorff space. [This example of a countable connected Hausdorff space is due to Morton Brown.]

Section 2

1. Let Y be Hausdorff, and assume that each $y \in Y$ has a nbd V such that $\bar{V}$ is regular. Prove: Y is regular.
2. Prove: If Y is regular, each pair of distinct points have nbds whose closures do not intersect.
3. Retopologize the real line by taking as complete system of nbds at each x the sets $U_n(x) = \{x\} \cup \{y \mid (y \text{ is rational}) \wedge (|y - x| < 1/n)\}$, $n \in Z^+$. Show that this space is not regular, but that each pair of distinct points have nbds whose closures do not intersect. Thus, the converse of Problem 2 is false.
4. Show that the space $(Z^+, \mathscr{T})$ of Problem 1. 15 is not regular. [*Hint:* The closed sets $\{2n \mid n \in Z^+\}$ and $\{1\}$ do not have disjoint nbds.]
5. Let X be regular, and $A \subset X$ closed. Show that
$$A = \bigcap \{U \mid [U \text{ is open}] \wedge [U \supset A]\}.$$

Section 3

1. Show that if "Hausdorff" is omitted in the Definition **3.1**, then the indiscrete spaces and Sierpinski space are normal.
2. Let X be normal and $p: X \to X/R$ be a closed and open map. Show that X/R is normal.
3 (a). Let X be the set of irrationals in E^1 and let $u: X \to E^1$ be a map that is always positive. For each $n \in Z^+$ let $H_n = \{x \mid u(x) \geqslant 1/n\}$. Prove: There exists an m and an open interval $\mathscr{J}_0$ such that $\mathscr{J} \cap H_m \neq \varnothing$ for every open interval $\mathscr{J} \subset \mathscr{J}_0$. [*Hint:* Assume this assertion were false; letting $\{r_n \mid n \in Z^+\}$ be an enumeration of the rationals, we could then find a sequence $\{J_n\}$ of intervals, each having rational end points such that for each n, (1) $r_n \,\bar{\in}\, J_n$; (2) $J_n \cap H_n = \varnothing$; (3) $J_{n+1} \subset J_n$; (4) $0 <$ length $J_n \leqslant 1/n$. By Cantor's definition of real numbers $\bigcap_n J_n$ would be some real number ξ. But ξ cannot be rational because of (1), and it cannot be irrational either since $\forall n$: $\xi \,\bar{\in}\, H_n$, so $u(\xi) = 0$, an impossibility.]

(b). Use part (a) to show that in $E^1_u \times E^1_u$ the disjoint closed sets $A = \{(r, -r) \mid r \text{ rational}\}$ and $B = \{(x, -x) \mid x \text{ irrational}\}$ cannot be separated. [*Hint:* Let $U \supset B$ be any open set. For each irrational x, let
$$u(x) = \sup \{\lambda \mid]x - \lambda, x] \times]-x - \lambda, -x] \subset U\};$$
$u(x)$ is never zero. Letting $r \in \mathscr{J}_0$, every nbd of $(r, -r)$ intersects U.]

4. Let T be the space of Ex. 4 and A, B be the subsets indicated there.
   a. Attach T to a single point p_A by the map $f(A) = p_A$. Show that $T \cup_f p_A$ is Hausdorff, but not regular.
   b. Attach T to two points, $p_A \cup p_B$ by the map $f(A) = p_A$, $f(B) = p_B$. Show that $T \cup_f (p_A \cup p_B)$ is not Hausdorff.

5. Let X be the upper half of the Euclidean plane E^2, bounded by the x-axis. Use the Euclidean topology on $\{(x, y) \mid y > 0\}$, but define nbds of the points $(x, 0)$ to be $[(x, 0)] \cup$ [open disc in $\{(x, y) \mid y > 0\}$ tangent to the x-axis at $(x, 0)$]. Prove that this space is not normal.
6. Show that the unit interval I is completely normal.
7. Let I be the unit interval, and $\mathscr{T}$ its Euclidean topology. Let $Q \subset I$ be the set of irrationals in I. Define a topology $\mathscr{T}_1$ in the set I by taking $\mathscr{T} \cup \mathscr{P}(Q)$ as subbasis. Prove that $(I, \mathscr{T}_1)$ is normal.

Section 4

1. Let Y be a connected normal space having more than one point. Determine a lower bound for $\aleph(Y)$.
2. Let X be normal. Prove: Every F_σ-set in X is also normal.
3. Let X be perfectly normal. Prove that every subspace of X is also perfectly normal.
4. Let X be normal, $A \subset X$ closed, and U an open set containing A. Prove: There exists an open F_σ-set V such that $A \subset V \subset U$.

Section 5

1. Show $[a, b[\subset E^1$ is an AR (normal).
2. Prove: If Y is an ANR (normal), then every open subspace of Y is also an ANR (normal). Give an example to show that this need not be true for closed subsets.
3. A normal space Y is said to have property Q if whenever Z is normal and $Y \subset Z$ is closed, there is a continuous $r: Z \to Y$ such that $r(y) = y$ for each $y \in Y$. Prove: Y has property Q if and only if it is an AR (normal). Formulate a similar criterion for ANR (normal).

Section 6

1. Let Y be normal, and $F_1, \cdots, F_n$ closed subsets such that $\bigcap_1^n F_i = \varnothing$. Prove: There exist open sets $V_i \supset F_i$ such that $\bar{V}_1 \cap \bar{V}_2 \cap \cdots \cap \bar{V}_n = \varnothing$ also.
2. Prove: Y is normal if and only if for each finite covering $U_1, \cdots, U_n$ of Y by open sets, there exist n continuous maps $f_i: Y \to I$ such that $\sum_1^n f_i(y) \equiv 1$ and each $f_i(y) = 0$ for $y \in Y - U_i$.
3. Let Y be a space having weak topology with respect to a countable closed covering $\{A_n \mid n \in Z\}$. Prove: If each A_n is a normal space, then Y is normal.
4. Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite covering of the normal space Y by open sets. For each $(\alpha, \beta) \in \mathscr{A} \times \mathscr{A}$, and each $y \in U_\alpha \cap U_\beta$, let $V_{\alpha,\beta}(y) \subset U_\alpha \cap U_\beta$ be a given nbd of y. Prove that one can assign a nbd $V(y)$ to each $y \in Y$ such that:
   a. $y \in U_\alpha \cap U_\beta \Rightarrow V(y) \subset V_{\alpha,\beta}(y)$.
   b. If $V(y) \cap V(x) \neq \varnothing$, there exists an α such that $V(y) \cup V(x) \subset U_\alpha$.

Section 7

1. Let Y be a T_0 space satisfying the condition: each $y \in Y$ and closed A not containing y have disjoint nbds. Prove: Y is a regular (Hausdorff) space.
2. Let Y be completely regular, V open, and $p \in V$. Prove: A necessary and sufficient condition that there be a continuous $\varphi: Y \to I$ such that $\varphi^{-1}(1) = p$, $\varphi(\mathscr{C}V) = 0$, is that p be a G_δ.
3. Prove: If Y is a connected and completely regular space with more than one point, then $\aleph(Y) \geqslant \mathfrak{c}$.
4. Let X be completely regular. Show that the topology of X is precisely the projective limit topology determined by the family of all continuous maps $f: X \to I$ (*cf.* VI, **1**, Problem 4).
5. Let X be completely regular, and $C(X)$ the set of all bounded continuous real-valued functions on X. For each x, f, ϵ let $U(x, f, \epsilon) = \{y \mid |f(x) - f(y)| < \epsilon\}$. Prove: $\{U(x, f, \epsilon) \mid \text{all } f, x, \text{ and } \epsilon > 0\}$ is a subbasis for the topology of X.
6. Let Y be completely regular, and $f: Y \to E^1$ lower semicontinuous. Show that $f = \sup f_\alpha$ for a suitable family $\{f_\alpha \mid \alpha \in \mathscr{A}\}$ of continuous functions. Conversely, show that if Y is any space such that each lower-semicontinuous function is the sup of continuous functions, then Y is completely regular.

Covering Axioms

VIII

In VII, **6.1**, a normal space was characterized by a property of certain types of coverings by open sets. In this chapter, we study some other important classes of spaces that are determined by some requirement on the behavior of their coverings.

1. Coverings of Spaces

In this section, we shall be concerned primarily with finite, nbd-finite, and point-finite coverings of a space Y by arbitrary sets. It is clear that any finite covering is nbd-finite and any nbd-finite covering is point-finite; however, as III, **9**, Ex. 1, shows, a point-finite covering need not be nbd-finite. By a subcovering of the covering $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ of Y is meant any subfamily $\{A_\alpha \mid \alpha \in \mathscr{B}\}$, $\mathscr{B} \subset \mathscr{A}$, that is also a covering of Y.

1.1 Theorem Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a point-finite covering of Y. Then there exists an irreducible subcovering, that is, a subcovering that, when any single set is removed, is no longer a covering of Y.

Proof: Call a family $R \subset \{A_\alpha\}$ removable if $\{A_\alpha\} - R$ still covers Y. Partially order the set $\mathscr{R}$ of all removable families by inclusion. If $\{R_\mu\}$ is any chain in $\mathscr{R}$, it has an upper bound, $R = \bigcup_\mu R_\mu$ in $\mathscr{R}$: for, if $R \bar{\in} \mathscr{R}$, there would be some $y \in Y$ such that the finitely many sets

$A_{\alpha_1}, \cdots, A_{\alpha_n}$ containing y are in R, and since $\{R_\mu\}$ is a chain, these A_{α_i} would all lie in some one $R_{\mu'}$, which would contradict $R_{\mu'} \in \mathscr{R}$. By Zorn's lemma, there is a maximal removable family R_0, and so

$$\{A_\alpha \mid \alpha \in \mathscr{A}\} - R_0$$

is irreducible.

Ex. 1 The hypothesis of point-finiteness is essential: The covering of E^1 by the sets $A_n =]-n, n[$, $n = 1, 2, \cdots$, has no irreducible subcovering.

The family of all coverings of a given space has a natural preorder:

1.2 Definition Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ and $\{B_\beta \mid \beta \in \mathscr{B}\}$ be two coverings of a space Y. $\{A_\alpha\}$ is said to refine (or be a refinement of) $\{B_\beta\}$ if for each A_α there is some B_β with $A_\alpha \subset B_\beta$. We write $\{A_\alpha\} \prec \{B_\beta\}$.

It is clear that any subcovering of a given covering is a refinement of that covering. If $\{C_\gamma\} \prec \{A_\alpha\}$ and $\{C_\gamma\} \prec \{B_\beta\}$, then $\{C_\gamma\}$ is called a common refinement of $\{A_\alpha\}$ and $\{B_\beta\}$.

Ex. 2 The relation $\prec$ of refinement is easily seen to be a preordering in the set of all coverings of Y. It is *not* a partial ordering: In E^1, each of the two coverings $A_n = \{x \mid x < n\}$, $n = 1, 2, \cdots$ and $B_n = \{x \mid x < n + \frac{1}{2}\}$, $n = 1, 2, \cdots$ is a refinement of the other.

1.3 Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ and $\{B_\beta \mid \beta \in \mathscr{B}\}$ be two coverings of Y. Then:

(1). $\{A_\alpha \cap B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$ is a covering of Y, refining both $\{A_\alpha\}$ and $\{B_\beta\}$. Furthermore, if both $\{A_\alpha\}$ and $\{B_\beta\}$ are nbd-finite (point-finite), so also is $\{A_\alpha \cap B_\beta\}$.

(2). Any common refinement of $\{A_\alpha\}$ and $\{B_\beta\}$ is also a refinement of $\{A_\alpha \cap B_\beta\}$.

Proof: (1). Each $y \in Y$ belongs to some $A_\alpha \cap B_\beta$, since $(\exists\, \alpha: y \in A_\alpha) \wedge (\exists\, \beta: y \in B_\beta)$; thus $\{A_\alpha \cap B_\beta\}$ is a covering, and it is evidently a common refinement. If both $\{A_\alpha\}$ and $\{B_\beta\}$ are nbd-finite, each $y \in Y$ has a nbd U intersecting at most finitely many A_α, and a nbd V intersecting at most finitely many B_β, so $U \cap V$ is a nbd of y intersecting at most finitely many $\{A_\alpha \cap B_\beta\}$. The proof of the point-finite case is similar.

(2). If the covering $\{C_\gamma \mid \gamma \in \mathscr{V}\}$ refines both $\{A_\alpha\}$ and $\{B_\beta\}$, then for each C_γ, we have $(\exists\, A_\alpha: C_\gamma \subset A_\alpha) \wedge (\exists\, B_\beta \mid C_\gamma \subset B_\beta)$, so that $C_\gamma \subset A_\alpha \cap B_\beta$ and $\{C_\gamma\} \prec \{A_\alpha \cap B_\beta\}$.

A refinement of a covering may contain more sets than the given covering; a refinement $\{B_\beta \mid \beta \in \mathscr{B}\}$ of $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is called *precise* if $\mathscr{B} = \mathscr{A}$ and $B_\alpha \subset A_\alpha$ for each α.

1.4 Theorem If the covering $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ of Y has a point (nbd)-finite refinement $\{B_\beta \mid \beta \in \mathscr{B}\}$, then it also has a precise point (nbd)-finite refinement $\{C_\alpha \mid \alpha \in \mathscr{A}\}$. Furthermore, if each B_β is an open set, then each C_α can be chosen to be an open set also.

Proof: Define a map $\varphi: \mathscr{B} \to \mathscr{A}$ by assigning to each $\beta \in \mathscr{B}$ some $\alpha \in \mathscr{A}$ such that $B_\beta \subset A_\alpha$. For each α, let $C_\alpha = \bigcup \{B_\beta \mid \varphi(\beta) = \alpha\}$; some C_α may be empty. Clearly, $C_\alpha \subset A_\alpha$ for each α, and also $\{C_\alpha \mid \alpha \in \mathscr{A}\}$ is a covering because each B_β appears somewhere; C_α is evidently open whenever each B_β is open. If B_β is point (nbd)-finite, then each (some nbd of) $y \in Y$ lies in at most finitely many B_β and therefore cannot meet more than that number of C_α; $\{C_\alpha\}$ is therefore point (nbd)-finite.

A covering of a space by open (closed) sets is called an open (closed) covering. We have the important "expansion"

1.5 Theorem Let $\{E_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of sets (not necessarily a covering!) in a space Y, and let $\{B_\beta \mid \beta \in \mathscr{B}\}$ be any nbd-finite *closed* covering of Y. Assume that each B_β intersects at most finitely many sets E_α. Then each E_α can be embedded in an open set $U(E_\alpha)$ such that the family $\{U(E_\alpha) \mid \alpha \in \mathscr{A}\}$ is nbd-finite.

Proof: For each α, define $U(E_\alpha) = Y - \bigcup \{B_\beta \mid B_\beta \cap E_\alpha = \varnothing.\}$ Each $U(E_\alpha)$ is open, because $\{B_\beta\}$ is a nbd-finite family of closed sets and III, **9.2** applies. $\{U(E_\alpha) \mid \alpha \in \mathscr{A}\}$ is nbd-finite: Any given $y \in Y$ has some nbd lying in a finite union $\bigcup_1^n B_{\beta_i}$; since $B_\beta \cap U(E_\alpha) \neq \varnothing$ if and only if $B_\beta \cap E_\alpha \neq \varnothing$, and since each B_{β_i} intersects at most finitely many E_α, $\bigcup_1^n B_{\beta_i}$ intersects at most finitely many $U(E_\alpha)$. Finally, $E_\alpha \subset U(E_\alpha)$ is evident, and the proof is complete.

2. Paracompact Spaces

2.1 Definition A Hausdorff space Y is paracompact if each open covering of Y has an open nbd-finite refinement.

Ex. 1 Any discrete space is paracompact; we will see in Ex. 5 that E^n is paracompact.

Ex. 2 The ordinal space $[0, \Omega]$ is paracompact. Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be any open covering. Since the sets $]\lambda, \mu]$ form a basis, define $\varphi: [0, \Omega] \to [0, \Omega]$ by associating with each $\beta \neq 0$ a $\varphi(\beta) < \beta$ such that $]\varphi(\beta), \beta] \subset$ some U_α, and setting $\varphi(0) = 0$. By induction, construct a sequence $\beta_0 = \Omega$, $\beta_1 = \varphi(\Omega), \cdots$, $\beta_n = \varphi(\beta_{n-1}), \cdots$; then $\beta_0 > \beta_1 > \cdots$ and, since every descending sequence of ordinals is finite (II, **6.4**), this terminates with some β_n. Because the process cannot be continued, $\beta_n = 0$, and so $]0, \Omega] \subset \bigcup_{i=1}^{n}]\beta_{i-1}, \beta_i]$. Choosing a U_{α_i} containing each $]\beta_{i-1}, \beta_i]$ and some $U_{\alpha_0} \ni 0$, we have a finite subcovering of $\{U_\alpha \mid \alpha \in \mathscr{A}\}$, which is consequently an open nbd-finite refinement.

Ex. 3 $[0, \Omega[$ is not paracompact: the open covering by the sets $[0,\alpha[$, $0 < \alpha < \Omega$, has no open nbd-finite refinement. For, given any open refinement $\{U_\alpha\}$, define $\varphi: [0, \Omega[\to [0, \Omega[$ as in Ex. 2; because of II, **9.2**, there must be some β_0 such that $\forall \gamma \, \exists \beta > \gamma: \varphi(\beta) \leqslant \beta_0$, and it follows easily that $\beta_0 + 1$ is contained in infinitely many sets U_α.

The relation of paracompact spaces to those we know is

2.2 Theorem Every paracompact space is normal.

Proof: We first show that the paracompact Y is regular. Let a closed $A \subset Y$ and a $y \,\bar{\in}\, A$ be given. Since Y is Hausdorff, we find by VII, **1.2**, that each $a \in A$ has a nbd U_a with $y \,\bar{\in}\, \overline{U}_a$. Since $\{U_a \mid a \in A\} \cup \{\mathscr{C}A\}$ is an open covering of Y, we use paracompactness and **1.4** to get a precise nbd-finite open refinement $\{V_a \mid a \in A\} \cup G$. Then $W = \bigcup \{V_a \mid a \in A\}$ is open and contains A; furthermore, because $\{V_a\}$ is nbd-finite, III, **9.2**, shows that $\overline{W} = \cup \{\overline{V}_a \mid a \in A\}$, and since $y \,\bar{\in}\, \overline{U}_a \supset \overline{V}_a$ for each $a \in A$, we find $y \,\bar{\in}\, \overline{W}$; thus $\mathscr{C}\overline{W}$ and W are disjoint nbds of y and A, as required. We now show Y to be normal. Given disjoint closed A, B, the regularity of Y gives for each $a \in A$ a nbd U_a with $\overline{U}_a \cap B = \varnothing$; reasoning as before (with y replaced by B) gives disjoint nbds of A and B.

Ex. 4 As Ex. 3 shows, not every normal (even completely normal) space is paracompact.

For regular spaces Y, Definition **2.1** has several equivalent formulations, wherein the nature of the sets making up the refinement is varied.

2.3 Theorem (E. Michael) Let Y be a regular space. The following four properties are equivalent:

(1). Y is paracompact.

(2). Each open covering of Y has an open refinement that can be decomposed into an at most countable collection of nbd-finite families of open sets.

(3). Each open covering of Y has a nbd-finite refinement, consisting of sets not necessarily either open or closed.

(4). Each open covering of Y has a closed nbd-finite refinement.

Proof: That (1) $\Rightarrow$ (2) is clear.

(2) ⇒ (3). Let $\{U_\beta \mid \beta \in \mathscr{B}\}$ be any open covering of Y; by (2) there is an open refinement $\{V_{n,\alpha} \mid (n, \alpha) \in Z^+ \times \mathscr{A}\}$, where for each fixed n_0, the family $\{V_{n_0,\alpha} \mid \alpha \in \mathscr{A}\}$ is nbd-finite (not necessarily a covering!). For each n, let $W_n = \bigcup_\alpha V_{n,\alpha}$; then $\{W_n \mid n \in Z^+\}$ is an open covering of Y. For each $i = 1, 2, \cdots$ define $A_i = W_i - \bigcup_{j<i} W_j$. We observe that $\{A_i\}$ is a refinement of $\{W_n \mid n \in Z^+\}$; it is a covering, since each $y \in A_{n(y)}$ where $n(y)$ is the first i for which $y \in W_i$, and it is also nbd-finite, since the nbd $W_{n(y)}$ of y does not intersect any A_i for $i > n(y)$. Now $\{A_n \cap V_{n,\alpha}\}$ is a refinement of $\{U_\beta\}$; it is nbd-finite, since each $y \in Y$ has a nbd intersecting at most finitely many A_n, and for each such n, the point y has a nbd intersecting at most finitely many $V_{n,\alpha}$.

(3) ⇒ (4). Let $\{U\}$ be an open covering. With each $y \in Y$, associate a definite $U_y \in \{U\}$ containing it, and then, since Y is regular, find a V_y with $y \in V_y \subset \overline{V}_y \subset U_y$. The family $\{V_y \mid y \in Y\}$ is an open covering, so it has a precise nbd-finite refinement $\{A_y \mid y \in Y\}$. Since by III, **9.2**, $\{\overline{A}_y \mid y \in Y\}$ is also nbd-finite, and since $\overline{A}_y \subset \overline{V}_y \subset U_y$, for each y, $\{\overline{A}_y \mid y \in Y\}$ is the desired refinement.

(4) ⇒ (1). Let $\{U\}$ be any open covering and $\{E\}$ be a closed nbd-finite refinement. Then each $y \in Y$ has a nbd V_y meeting at most finitely many sets E. Using $\{V_y \mid y \in Y\}$, find a closed nbd-finite refinement $\{B\}$; since each B intersects at most finitely many sets E, it follows from **1.5** that we can enlarge each E to an open $G(E)$ such that $\{G(E)\}$ is nbd-finite. Associating with each E a single set $U(E) \in \{U\}$ containing E, it is evident that $\{G(E) \cap U(E)\}$ is an open nbd-finite refinement of $\{U\}$.

Ex. 5 E^n and any subspace of E^n (or, more generally, any regular space with countable basis) is paracompact. For, given any open covering $\{U\}$, express each U as the union of sets V belonging to the countable basis. The sets V used form an open refinement of $\{U\}$, and since there are at most countably many distinct sets V, the conditions of **2.3**(2), with each family consisting of a single set V, are satisfied.

Ex. 6 E^1_u is paracompact. Let $\{U\}$ be any open covering; we will show that $\{U\}$ has a countable subcovering. Since E^1_u is normal, by VII, **3**, Ex. 2, its paracompactness will then follow from **2.3**(2). Each $x \in E^1_u$ is the end point of an interval $I_x =]a_x, x] \subset$ some U. For each rational r_i in the interior of some I_x, let $b_i = \sup \{b \mid]r_i, b] \subset \text{some } U\}$, and let $U_i \in \{U\}$ be a set with $]r_i, b_i] \subset \overline{U}_i$. The countable family $\{U_i\}$ covers $Y = E^1_u - \bigcup_1^\infty b_i - \bigcup_1^\infty r_i$: each $x \in Y$ lies in some U, so there is an r_n with $]r_n, x] \subset$ some U; and, since $x \,\overline{\in}\, \bigcup_1^\infty b_i$, we have $b_n > x$, showing that $x \in U_n$. Adding to the $\{U_n\}$ one set containing each b_i and one containing each r_i gives the required countable subcovering.

The invariance properties are similar to those for normality; they are summarized here for convenience, and the proofs are given separately, as indicated.

2.4 Theorem (1). (**E. Michael**) Paracompactness is invariant under continuous closed surjections (*cf.* **2.6**).
(2). A subspace of a paracompact space need not be paracompact. However, a closed subspace is paracompact (*cf.* **2.5**).
(3). The cartesian product of paracompact spaces need not be paracompact. However, if the product is paracompact, each factor is paracompact.

Proof of (3): In Ex. 6, we have seen that E_u^1 is paracompact. Since by VII, **3**, Ex. 3, $E_u^1 \times E_u^1$ is not normal, it cannot be paracompact. The second assertion follows in the usual manner from (2).

For (2), we prove, more generally,

2.5 (a). (**E. Michael**) Each F_σ-set in a paracompact space is paracompact.
(b). If each open set in a paracompact space is paracompact, then every subspace is paracompact.

Proof: (a). Let $F = \bigcup_1^\infty F_i$, where each F_i is closed, and let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be any open covering of F; each $U_\alpha = F \cap V_\alpha$, where V_α is open in Y. For each fixed n, $\{V_\alpha\} \cup \{\mathscr{C}F_n\}$ is an open covering of Y and so has an open nbd-finite refinement $\{W_{\alpha,n}\}$. For each n, let $\mathscr{B}_n = \{W_{\alpha,n} \cap F \mid W_{\alpha,n} \cap F_n \neq \varnothing\}$; then each $\mathscr{B}_n$ is a nbd-finite family of sets open in F, and $\bigcup_n \mathscr{B}_n$ is clearly an open covering of F refining $\{U_\alpha \mid \alpha \in \mathscr{A}\}$; by **2.3**(2), F is therefore paracompact.

(b). Given any $B \subset Y$ and any open covering $\{W \cap B\}$, where each W is open in Y, then $\bigcup W$ is an open set, and so is paracompact by hypothesis. If $\{V\}$ is an open nbd-finite refinement of the covering $\{W\}$ of $\bigcup W$, then $\{V \cap B\}$ is a nbd-finite refinement of $\{W \cap B\}$.

2.6 Let X be paracompact, and $p: X \to Y$ a continuous closed surjection. Then Y is paracompact.

Proof: Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be any open covering of Y. Since by VII, **3.3**, Y is normal, **2.3**(2) says that it suffices to show $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ has an open refinement which can be decomposed into at most countably many nbd-finite families.

We assume $\mathscr{A}$ is well-ordered and begin by constructing an open covering $\{V_{\alpha,n} \mid (\alpha, n) \in \mathscr{A} \times Z^+\}$ of X such that:

(1). For each n, $\{\overline{V}_{\alpha,n} \mid \alpha \in \mathscr{A}\}$ is a covering of X and a precise nbd-finite refinement of $\{p^{-1}(U_\alpha) \mid \alpha \in \mathscr{A}\}$.

(2). If $\beta > \alpha$, then $p(\overline{V}_{\beta,n+1}) \cap p(\overline{V}_{\alpha,n}) = \varnothing$.

Proceeding by induction, we take a precise open nbd-finite refinement of $\{p^{-1}(U_\alpha)\}$ and shrink it (VII, **6.1**) to get $\{\overline{V}_{\alpha,1}\}$. Assuming $\{V_{\alpha,i}\}$ to be defined for all $i \leqslant n$, let $W_{\alpha,n+1} = p^{-1}(U_\alpha) - p^{-1}p(\bigcup_{\lambda<\alpha} \overline{V}_{\lambda,n})$. Each $W_{\alpha,n+1}$ is open, since by nbd-finiteness $\bigcup_{\lambda<\alpha} \overline{V}_{\lambda,n}$ is closed and p is a closed map. Furthermore, $\{W_{\alpha,n+1} \mid \alpha \in \mathscr{A}\}$ is a covering of X: given $x \in X$, let α_0 be the first index for which $x \in p^{-1}(U_\alpha)$; then $x \in W_{\alpha_0,n+1}$, since $p^{-1}p(\overline{V}_{\lambda,n}) \subset p^{-1}(U_\lambda)$ for each λ. Taking a precise, open nbd-finite refinement of $\{W_{\alpha,n+1} \mid \alpha \in \mathscr{A}\}$, shrink it to get $\{\overline{V}_{\alpha,n+1}\}$. Clearly, condition (1) holds, and since $\overline{V}_{\beta,n+1}$ is not in the inverse image of any $p(\overline{V}_{\alpha,n})$ for $\alpha < \beta$, condition (2) is also satisfied.

For each n and α, let $H_{\alpha,n} = Y - p(\bigcup_{\beta\neq\alpha} V_{\beta,n})$ which is an open set. We have

(a). $\qquad H_{\alpha,n} \subset p(\overline{V}_{\alpha,n}) \subset U_\alpha$ for each n and α.

Indeed,

$$p^{-1}(H_{\alpha,n}) = X - p^{-1}p\Big(\bigcup_{\beta\neq\alpha} \overline{V}_{\alpha,n}\Big) \subset X - p^{-1}p(X - \overline{V}_{\alpha,n})$$
$$\subset \overline{V}_{\alpha,n} \subset p^{-1}(U_\alpha).$$

(b). $\qquad H_{\alpha,n} \cap H_{\beta,n} = \varnothing$ for each n whenever $\alpha \neq \beta$.

In fact, $y \in H_{\alpha,n} \Rightarrow y \in p(\overline{V}_{\alpha,n})$ and is in no other $p(\overline{V}_{\beta,n})$.

(c). $\qquad \{H_{\alpha,n} \mid (\alpha, n) \in \mathscr{A} \times Z^+\}$ is an open covering of Y.

Let $y \in Y$ be given; for each fixed n there is, because of (1), a first α_n with $y \in p(\overline{V}_{\alpha_n,n})$; choosing now

$$\alpha_k = \min\{\alpha_n \mid n \in Z^+\},$$

we have $y \in p(\overline{V}_{\alpha_k,k})$. If $\beta < \alpha_k$, then the definition of α_k shows $y \notin p(\overline{V}_{\beta,k+1})$; if $\beta > \alpha_k$, then by (2), we find that $y \notin p(\overline{V}_{\beta,k+1})$; therefore we conclude that $y \in H_{\alpha_k,k+1}$.

To complete the proof, we need only modify the $H_{\alpha,n}$ slightly to assure nbd-finiteness for each n. Choose a precise open nbd-finite refinement of $\{p^{-1}(H_{\alpha,n}) \mid (\alpha, n) \in \mathscr{A} \times Z^+\}$, and shrink it to get an open nbd-finite covering $\{K_{\alpha,n}\}$ satisfying $p(\overline{K}_{\alpha,n}) \subset H_{\alpha,n}$. For each n, let $S_n = \{y \mid \text{some nbd of } y \text{ intersects at most one } H_{\alpha,n}\}$; S_n is open and contains the closed $\bigcup_\alpha p(\overline{K}_{\alpha,n}) = p(\bigcup_\alpha \overline{K}_{\alpha,n})$, so by normality of Y we find an open G_n with $\bigcup_\alpha p(\overline{K}_{\alpha,n}) \subset G_n \subset \overline{G}_n \subset S_n$. The open covering $\{G_n \cap H_{\alpha,n} \mid (\alpha, n) \in \mathscr{A} \times Z^+\}$, with the decomposition $\{G_n \cap H_{\alpha,n} \mid \alpha \in \mathscr{A}\}$ for $n = 1, 2, \cdots$ satisfies the conditions of **2.3**(2) for the given $\{U_\alpha\}$.

3. Types of Refinements

In this section, we obtain characterizations of paracompact spaces by means of refinements that are not necessarily nbd-finite.

Let $\mathfrak{U} = \{U_\alpha \mid \alpha \in \mathscr{A}\}$ be a covering of a space Y. For any $B \subset Y$, the set $\bigcup \{U_\alpha \mid B \cap U_\alpha \neq \varnothing\}$ is called the *star* of B with respect to $\mathfrak{U}$, and is denoted by $\text{St}(B, \mathfrak{U})$.

3.1 Definition A covering $\mathfrak{V}$ is called a barycentric refinement of a covering $\mathfrak{U}$ whenever the covering $\{\text{St}(y, \mathfrak{V}) \mid y \in Y\}$ refines $\mathfrak{U}$.

3.2 Let Y be normal, and $\mathfrak{U} = \{U_\alpha \mid \alpha \in \mathscr{A}\}$ an nbd-finite open covering. Then $\mathfrak{U}$ has an open barycentric refinement.

Proof: Shrink $\mathfrak{U}$ to an open covering $\mathfrak{V} = \{V_\alpha \mid \alpha \in \mathscr{A}\}$ such that $\overline{V}_\alpha \subset U_\alpha$ for each α; clearly, $\mathfrak{V}$ is also nbd-finite. For each $y \in Y$, define

$$W(y) = \bigcap \{U_\alpha \mid y \in \overline{V}_\alpha\} \cap \bigcap \{\mathscr{C}\overline{V}_\beta \mid y \bar{\in} \overline{V}_\beta\}.$$

We show that $\mathfrak{W} = \{W(y) \mid y \in Y\}$ is the required open covering. Note first that each $W(y)$ is open: the nbd-finiteness of $\mathfrak{V}$ assures that the first term is a finite intersection, and that the last term, $\mathscr{C} \bigcup \overline{V}_\beta$, is an open set (III, **9.2**). Next, $\mathfrak{W}$ is a covering, since $y \in W(y)$ for each $y \in Y$. Finally, fix any $y_0 \in Y$ and choose a $\overline{V}_\alpha$ containing y_0. Now, for each y such that $y_0 \in W(y)$, we must have $y \in \overline{V}_\alpha$ also, otherwise $W(y) \subset \mathscr{C}\overline{V}_\alpha$; and because $y \in \overline{V}_\alpha$, we conclude that $W(y) \subset U_\alpha$. Thus, $\text{St}(y_0, \mathfrak{W}) \subset U_\alpha$, and the proof is complete.

3.3 Definition A covering $\mathfrak{V} = \{V_\beta \mid \beta \in \mathscr{B}\}$ is called a star refinement of the covering $\mathfrak{U}$ whenever the covering $\{\text{St}(V_\beta, \mathfrak{V} \mid \beta \in \mathscr{B}\}$ refines $\mathfrak{U}$.

3.4 A barycentric refinement $\mathfrak{W}$ of a barycentric refinement $\mathfrak{V}$ of $\mathfrak{U}$ is a star refinement of $\mathfrak{U}$.

Proof: Given $W_0 \in \mathfrak{W}$, choose a fixed $y_0 \in W_0$. For each $W \in \mathfrak{W}$ such that $W \cap W_0 \neq \varnothing$, choose a $z \in W \cap W_0$; then $W \cup W_0 \subset \text{St}(z, \mathfrak{W}) \subset$ some $V \in \mathfrak{V}$. Because each such V contains y_0, we conclude that $\text{St}(W_0, \mathfrak{W}) \subset \text{St}(y_0, \mathfrak{V}) \subset$ some $U \in \mathfrak{U}$.

Since it is clear that a barycentric refinement of *any* refinement of $\mathfrak{U}$ is also a barycentric refinement of $\mathfrak{U}$, it follows from **3.2** and **3.4** that each open covering of a paracompact space has an open barycentric, and an open star, refinement. Much more important, however, is that this property characterizes the paracompact spaces, not only among the Hausdorff spaces, but in fact also among the T_1 spaces:

3.5 Theorem **(A. H. Stone)** A T_1 space Y is paracompact if and only if each open covering has an open barycentric refinement.

Proof: Only the sufficiency requires proof. We first show that any open covering $\mathfrak{U} = \{U_\alpha \mid \alpha \in \mathscr{A}\}$ has a refinement as required in **2.3**(2).

Let $\mathfrak{U}^*$ be an open star refinement of $\mathfrak{U}$ (*cf.* **3.4**) and let $\{\mathfrak{U}_n \mid n \geqslant 0\}$ be a sequence of open coverings, where each $\mathfrak{U}_{n+1}$ star refines $\mathfrak{U}_n$ and $\mathfrak{U}_0$ star refines $\mathfrak{U}^*$. Define a sequence of coverings inductively by

$$\mathfrak{V}_1 = \mathfrak{U}_1, \ \mathfrak{V}_2 = \{\mathrm{St}(V, \mathfrak{U}_2) \mid V \in \mathfrak{V}_1\}, \cdots, \quad \mathfrak{V}_n = \{\mathrm{St}(V, \mathfrak{U}_n) \mid V \in \mathfrak{V}_{n-1}\}, \cdots$$

Each $\mathfrak{V}_n$ is an open refinement of $\mathfrak{U}_0$; in fact, each covering $\{\mathrm{St}(V, \mathfrak{U}_n) \mid V \in \mathfrak{V}_n\}$ refines $\mathfrak{U}_0$: this is true for $n = 1$ and, proceeding by induction, if it is true for $n = k - 1$, its truth for $n = k$ follows by noting that whenever $V = \mathrm{St}(V_0, \mathfrak{U}_k)$ for some $V_0 \in \mathfrak{V}_{k-1}$, then $\mathrm{St}(V, \mathfrak{U}_k) = \mathrm{St}[\mathrm{St}(V_0, \mathfrak{U}_k), \mathfrak{U}_k] \subset \mathrm{St}(V_0, \mathfrak{U}_{k-1})$ because $\mathfrak{U}_k$ is a star refinement of $\mathfrak{U}_{k-1}$.

Now well-order Y and for each $(n, y) \in Z^+ \times Y$ define

$$E_n(y) = \mathrm{St}(y, \mathfrak{V}_n) - \bigcup \{\mathrm{St}(z, \mathfrak{V}_{n+1}) \mid z \text{ precedes } y\}$$

Then $\mathfrak{E} = \{E_n(y) \mid (n, y) \in Z^+ \times Y\}$ is a covering: given $p \in Y$, the set $A = \{z \mid p \in \bigcup_{i=1}^{\infty} \mathrm{St}(z, \mathfrak{V}_i)\}$ is not empty, since $p \in A$; if y is the first member of A, then $p \in \mathrm{St}(y, \mathfrak{V}_n)$ for some $n \in Z^+$ and $p \bar{\in} \mathrm{St}(z, \mathfrak{V}_{n+1})$ for all z preceding y, so $p \in E_n(y)$. Moreover, since $\mathfrak{V}_n$ refines $\mathfrak{U}_0$, we find that $\mathfrak{E}$ refines $\mathfrak{U}^*$.

Each $U \in \mathfrak{U}_{n+1}$ can meet at most one $E_n(y)$: for, if $U \cap E_n(y) \neq \varnothing$, then there is a $V \in \mathfrak{V}_n$ with $y \in V$ and $V \cap U \neq \varnothing$, so $y \in V \cup U \subset V_0 \in \mathfrak{V}_{n+1}$ and $U \subset \mathrm{St}(y, \mathfrak{V}_{n+1})$. Thus, if $E_n(y)$ is the first set U meets, it cannot meet any $E_n(p)$ for p following y.

Now let $W_n(y) = \mathrm{St}(E_n(y), \mathfrak{U}_{n+2})$. Then

$$\mathfrak{W} = \{W_n(y) \mid (n, y) \in Z^+ \times Y\}$$

is clearly an open covering of Y. Furthermore, $\mathfrak{W}$ refines $\mathfrak{U}$ because $\mathfrak{E}$ refines $\mathfrak{U}^*$. Finally, for each fixed $n \in Z^+$, the family $\{W_n(y) \mid y \in Y\}$ is nbd-finite: indeed, each $U \in \mathfrak{U}_{n+2}$ can meet at most one $W_n(y)$, because

$U \cap W_n(y) \neq \varnothing$ if, and only if, $E_n(y) \cap \text{St}(U, \mathfrak{U}_{n+2}) \neq \varnothing$ and $\text{St}(U, \mathfrak{U}_{n+2})$ is contained in some $U_0 \in \mathfrak{U}_{n+1}$ which we know can meet at most one $E_n(y)$.

The theorem will follow from **2.3**(2), once we show that Y is regular. To this end, let $B \subset Y$ be closed and $y \notin B$. Since in a T_1 space each point is a closed set, $\mathfrak{U} = \{Y - y, \mathscr{C}B\}$ is an open covering. Let $\mathfrak{V}$ be an open star refinement. Then $\text{St}(y, \mathfrak{V})$ and $\text{St}(B, \mathfrak{V})$ are the required disjoint nbds of y and B: for if there were a V containing y and a V' meeting B such that $V \cap V' \neq \varnothing$, then $\text{St}(V, \mathfrak{V})$ would contain y and points of B, which is impossible. The theorem is proved.

This leads to still another characterization of paracompactness, this time based on a sequence of open coverings.

3.6 Definition Let $\mathfrak{U} = \{U_\alpha \mid \alpha \in \mathscr{A}\}$ be an open covering of Y. A sequence $\{\mathfrak{U}_n \mid n \in Z^+\}$ of open coverings is called *locally starring* for $\mathfrak{U}$ if for each $y \in Y$ there exists an nbd $V(y)$ and an $n \in Z^+$ such that $\text{St}(V, \mathfrak{U}_n) \subset$ some U_α.

3.7 Theorem (A. Arhangel'skii) A T_1 space is paracompact if and only if for each open covering $\mathfrak{U}$ there exists a sequence $\{\mathfrak{U}_n \mid n \in Z^+\}$ of open coverings that is locally starring for $\mathfrak{U}$.

Proof: "Only if" is trivial. "If": We can assume that $\mathfrak{U}_{n+1} \prec \mathfrak{U}_n$ for each $n \in Z^+$. Let

$$\mathfrak{V} = \{V \text{ open in } Y \mid \exists\, n\colon [V \subset U \in \mathfrak{U}_n] \wedge [\text{St}(V, \mathfrak{U}_n) \subset \text{some } U_\alpha]\}.$$

For each $V \in \mathfrak{V}$, let $n(V)$ be the smallest integer satisfying the condition. Because $\{\mathfrak{U}_n \mid n \in Z^+\}$ is locally starring for $\mathfrak{U}$, it follows that $\mathfrak{V}$ is an open covering; we will show that $\mathfrak{V}$ is in fact a barycentric refinement of $\mathfrak{U}$.

Let $y \in Y$ be fixed, let $n(y) = \min\{n(V) \mid (y \in V) \wedge (V \in \mathfrak{V})\}$, and let $V_0 \in \mathfrak{V}$ be a set containing y such that $n(V_0) = n(y)$. For any $V \in \mathfrak{V}$ containing y, we have $n(V) \geqslant n(y)$, and consequently

$$\text{St}(y, \mathfrak{V}) \subset \bigcup \{\text{St}(y, \mathfrak{U}_i) \mid i \geqslant n(y)\}.$$

Since $\mathfrak{U}_{i+1} \prec \mathfrak{U}_i$ for each i, this shows

$$\text{St}(y, \mathfrak{V}) \subset \text{St}(y, \mathfrak{U}_{n(y)}) \subset \text{St}(V_0, \mathfrak{U}_{n(V_0)}) \subset \text{some } U_\alpha.$$

By Stone's theorem, Y is therefore paracompact.

4. Partitions of Unity

Partitions of unity play an important role in modern topology; one of the reasons that paracompact spaces are useful is that "arbitrarily fine" such partitions exist on them.

For any space Y, the *support* of a map $f\colon Y \to E^1$ is the closed set

$\overline{\{y \mid f(y) \neq 0\}}$; observe that y is not in the support of f if and only if y has a nbd on which f vanishes identically.

4.1 Definition Let Y be a Hausdorff space. A family $\{\kappa_\alpha \mid \alpha \in \mathscr{A}\}$ of continuous maps $\kappa_\alpha \colon Y \to I$ is called a partition of unity on Y if:

(1). The supports of the κ_α form a nbd-finite closed covering of Y.

(2). $\sum_\alpha \kappa_\alpha(y) = 1$ for each $y \in Y$ (this sum is well-defined because each y lies in the support of at most finitely many κ_α).

If $\{U_\beta \mid \beta \in \mathscr{B}\}$ is a given open covering of Y, we say that a partition $\{\kappa_\beta \mid \beta \in \mathscr{B}\}$ of unity is subordinated to $\{U_\beta\}$ if the support of each κ_β lies in the corresponding U_β.

Clearly, every space has a partition of unity subordinated to the covering by the single set itself.

4.2 Theorem Let Y be paracompact. Then for each open covering $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ of Y there is a partition of unity subordinated to $\{U_\alpha\}$.

Proof: Shrink a precise nbd-finite refinement of $\{U_\alpha\}$ to get a nbd-finite open covering $\{V_\alpha\}$ with $\overline{V}_\alpha \subset U_\alpha$ for each α. Now shrink $\{V_\alpha\}$ to get a nbd-finite open covering $\{W_\alpha\}$ satisfying $\overline{W}_\alpha \subset V_\alpha$. For each $\alpha \in \mathscr{A}$, VII, **4.1**, gives a continuous $g_\alpha \colon Y \to I$, which is identically 1 on $\overline{W}_\alpha$ and vanishes on $\mathscr{C}V_\alpha$ (we take $g_\alpha \equiv 0$ if $V_\alpha = \varnothing$); each g_α has its support in U_α. Since $\{\overline{W}_\alpha\}$ is a nbd-finite covering, it follows that for each $y \in Y$ at least one, and at most finitely many, g_α are not zero, consequently $\sum_\alpha g_\alpha$ is a well-defined real-valued function on Y and is never zero. $\sum_\alpha g_\alpha$ is continuous on Y: every point has a nbd on which all but at most finitely many g_α vanish identically, so the continuity of $\sum g_\alpha$ on this nbd follows from that of each g_α, and by III, **9.4**, $\sum g_\alpha$ is therefore continuous on Y. The required partition of unity is given by the family of functions $\{\kappa_\alpha \mid \alpha \in \mathscr{A}\}$, where

$$\kappa_\alpha(y) = \frac{g_\alpha(y)}{\sum_\alpha g_\alpha(y)}.$$

We remark that in a normal space Y, the proof shows that a partition of unity subordinated to a given *nbd-finite* open cover exists; C. H. Dowker has shown that their existence for *each* open cover is equivalent to paracompactness of Y [*cf.* **5.5**(2)].

To give an application of **4.2**, note that if $\{\kappa_\alpha \mid \alpha \in \mathscr{A}\}$ is a partition of unity on Y, and if $\{\varphi_\alpha \mid \alpha \in \mathscr{A}\}$ is any family of continuous maps $\varphi_\alpha\colon Y \to E^1$, then the map $Y \to E^1$ given by $y \to \sum_\alpha \varphi_\alpha(y)\kappa_\alpha(y)$ is also continuous.

4.3 (C. H. Dowker) Let Y be paracompact. Assume that g is a lower, and G an upper, semicontinuous real-valued function on Y such that $G(y) < g(y)$ for each $y \in Y$. Then there exists a *continuous* $\varphi\colon Y \to E^1$ such that $G(y) < \varphi(y) < g(y)$ for each $y \in Y$.

Proof: For each rational r, let $U_r = \{y \mid G(y) < r\} \cap \{y \mid g(y) > r\}$; due to the semicontinuities, this is open; and because for each y there is some rational $\bar{r}$ with $G(y) < \bar{r} < g(y)$, the family $\{U_r\}$ is in fact an open covering of Y. Let $\{\kappa_r\}$ be a partition of unity subordinated to $\{U_r\}$; the required continuous function is $\varphi(y) = \sum_r r \cdot \kappa_r(y)$. For, let $y \in Y$ be given, and let $\kappa_{r_1}, \cdots, \kappa_{r_n}$ be all those functions whose support contains y; then $y \in U_{r_1} \cap \cdots \cap U_{r_n}$ so that $G(y) < r_i < g(y)$ for each $i = 1, \cdots, n$, and therefore

$$G(y) = G(y)\cdot \sum \kappa_{r_i}(y) < \sum r_i\kappa_{r_i}(y) = \varphi(y) < g(y)\cdot \sum \kappa_{r_i}(y) = g(y).$$

5. Complexes; Nerves of Coverings

The concept of a partition of unity subordinated to a given open covering has an alternative, more geometrical, interpretation. To develop this, we need two preliminary notions.

(1). Let $\mathscr{A}$ be any set. By an n-simplex σ^n in $\mathscr{A}$ is meant a set $(\alpha_0, \cdots, \alpha_n)$ of $n + 1$ distinct elements of $\mathscr{A}$; $\alpha_0, \cdots, \alpha_n$ are called the vertices of σ^n, and any $\sigma^q \subset \sigma^n$ is termed a q-face of σ^n.

5.1 Definition An abstract simplicial complex $\mathscr{K}$ over $\mathscr{A}$ is a set of simplexes in $\mathscr{A}$ with the property that each face of a $\sigma \in \mathscr{K}$ also belongs to $\mathscr{K}$.

With each abstract simplicial complex we will associate a standard topological space. For this we need

(2). Given $(n + 1)$ independent points $p_0, \cdots, p_n$ in an affine space, the open geometric n-simplex σ^n spanned by $p_0, \cdots, p_n$ is

$$\left\{\sum_0^n \lambda_i p_i \;\middle|\; \sum_0^n \lambda_i = 1, \quad 0 < \lambda_i \leqslant 1, \quad i = 0, \cdots, n\right\};$$

it is denoted by $(p_0, \cdots, p_n)$.

σ^n is the interior of the convex hull of $\{p_0, \cdots, p_n\}$ in the n-dimensional Euclidean space that these vertices span; for example, (p_0, p_1) is a segment without its end points, and (p_0, p_1, p_2) is a triangle without its boundary. The λ_i, $i = 0,\cdots,n$, are called the barycentric coordinates of

$$x = \sum_0^n \lambda_i p_i;$$

the closed geometric n-simplex $\bar{\sigma}^n = \overline{(p_0,\cdots,p_n)}$ consists of σ^n with its boundary, and is obtained by allowing $0 \leqslant \lambda_i \leqslant 1$ for $i = 0,\cdots,n$.

5.2 Definition Given any set $\mathscr{A}$, let $L(\mathscr{A})$ be a real vector space with finite topology, having a basis $\{b_\alpha\}$ in fixed 1-to-1 correspondence $b_\alpha : \alpha$ with the elements of $\mathscr{A}$, and let u_α be the unit point on the vector b_α. Given any complex $\mathscr{K}$ over $\mathscr{A}$, let $K \subset L(\mathscr{A})$ be the union of all open geometric simplexes $(u_{\alpha_0},\cdots,u_{\alpha_n})$ for which $(\alpha_0,\cdots,\alpha_n)$ is a simplex in $\mathscr{K}$. The subspace $K \subset L(\mathscr{A})$ is called a polytope with vertex scheme $\mathscr{K}$ (or a standard geometrical realization of $\mathscr{K}$).

It is evident that the space K has the weak topology determined by the Euclidean topology on its closed simplexes, so that an $f\colon K \to Y$ is continuous if and only if it is so on each $\bar{\sigma}^n$. This implies that any two standard geometrical realizations K_1, K_2 of a given $\mathscr{K}$ are homeomorphic: for, to each $\sigma^n \in \mathscr{K}$ there correspond unique $\sigma_1^n = (p_0^1,\dots,p_n^1)$ and $\sigma_2^n = (p_0^2,\cdots p_n^2)$ in K_1, K_2, respectively, and by barycentrically mapping each σ_1^n on the corresponding σ_2^n (that is $\sum_0^n \lambda_i p_i^1 \to \sum_0^n \lambda_i p_i^2$), the desired homeomorphism is obtained. Thus we can speak of *the* geometric realization of $\mathscr{K}$.

In a polytype, the star, $\operatorname{St} u_0$, of a vertex u_0 is the set of all open geometric simplexes having u_0 as vertex. It is important to note that $\operatorname{St} u_0$ is an open set in K: given any closed $\bar{\sigma} = \overline{(u_{\alpha_0},\cdots,u_{\alpha_n})}$, its intersection with $K - \operatorname{St} u_0$ is either $\bar{\sigma}$ if no $u_{\alpha_i} = u_0$ or a face of σ if some $u_{\alpha_i} = u_0$; in either case, this intersection is closed in $\bar{\sigma}$, so $K - \operatorname{St} u_0$ is closed in K.

The process of associating with each open covering of a space a complex called its *nerve* is very important because it is one method for relating the topological to the algebraic properties of spaces; intuitively geometric realizations of nerves approximate the space with the finer covering giving the better approximation.

5.3 Definition Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be any covering of a space. Define a complex $\mathscr{N}$ over $\mathscr{A}$ by the following condition: $(\alpha_0,\cdots,\alpha_n)$ is a simplex of $\mathscr{N}$ if and only if $U_{\alpha_0} \cap \cdots \cap U_{\alpha_n} \neq \varnothing$. It is evident that $\mathscr{N}$ is indeed a complex, called the *nerve* of $\{U_\alpha \mid \alpha \in \mathscr{A}\}$. The standard geometric realization of $\mathscr{N}$ is called the geometric nerve of $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ and is denoted by $N(U_\alpha)$.

The vertex of $N(U_\alpha)$ corresponding to the set U_α is denoted by u_α.

5.4 Theorem Let Y be any space and $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be an open covering. Then for each partition of unity subordinated to $\{U_\alpha\}$ there exists a continuous $\kappa\colon Y \to N(U_\alpha)$ such that $\kappa^{-1}(\operatorname{St} u_\alpha) \subset U_\alpha$ for each α.

Proof: Let $\{\kappa_\alpha \mid \alpha \in \mathscr{A}\}$ be a partition of unity subordinated to $\{U_\alpha\}$, and define

$\kappa: Y \to N(U_\alpha)$ by $\kappa(y) = \sum_\alpha \kappa_\alpha(y) u_\alpha$. This is continuous: each $y \in Y$ has a nbd on which all but at most finitely many κ_α vanish, and since this nbd is mapped into a finite-dimensional flat in $L(\mathscr{A})$, the addition is continuous (*cf.* Appendix I, **4**), so κ is continuous on that nbd and its continuity on Y results from III, **8.3**. Since $\sum_\alpha \kappa_\alpha(y) = 1$, $\kappa(y)$ is in fact a point of the closed geometric simplex spanned by $\{u_\alpha \mid \kappa_\alpha(y) \neq 0\}$. The inverse image of $\mathrm{St}\, u_{\alpha_0}$ consists of all y for which $\kappa_{\alpha_0}(y) \neq 0$, and because the support of κ_{α_0} is in U_{α_0}, we have

$$\kappa^{-1}(\mathrm{St}\, u_{\alpha_0}) \subset U_{\alpha_0}$$

as required.

It should be observed that if $V \subset Y$ is an open set intersecting the supports of only the finitely many $\kappa_{\alpha_1}, \cdots, \kappa_{\alpha_n}$, then $\kappa(V) \subset \overline{(u_{\alpha_0}, \cdots, u_{\alpha_n})}$.

5.5 Remark It is known (*cf.* Appendix I, **5.2**) that the geometric nerve $N(U_\alpha)$ is always a paracompact space. Using this fact, we can prove

(1). A continuous $\kappa: Y \to N(U_\alpha)$ satisfying $\kappa^{-1}(\mathrm{St}\, u_\alpha) \subset U_\alpha$ for each α exists if and only if there is a partition of unity subordinated to $\{U_\alpha\}$.

The "if" is **5.4**; the "only if" follows by finding a partition of unity $\{\kappa_\alpha \mid \alpha \in \mathscr{A}\}$ subordinated to the open cover $\{\mathrm{St}\, u_\alpha \mid \alpha \in \mathscr{A}\}$ of the paracompact $N(U_\alpha)$ and defining $\lambda_\alpha: Y \to I$ by $\lambda_\alpha = \kappa_\alpha \circ \kappa$.

(2). Y is paracompact if and only if for each open covering $\{U_\alpha\}$ there is a subordinated partition of unity.

The "only if" is **4.2**; the "if" follows by finding a nbd-finite refinement $\{V_\beta \mid \beta \in \mathscr{B}\}$ of the open covering $\{\mathrm{St}\, u_\alpha \mid \alpha \in \mathscr{A}\}$ in $N(U_\alpha)$; then

$$\{\kappa^{-1}(V_\beta) \mid \beta \in \mathscr{B}\}$$

is the desired nbd-finite refinement of $\{U_\alpha\}$. There is a simpler proof of "if" which uses the geometry, rather than the paracompactness, of $N(U_\alpha)$: letting N' be the barycentric subdivision of $N(U_\alpha)$ [*cf.* Appendix I, **5**] and using stars in N', we have that $\{\kappa^{-1}(\mathrm{St}\, p') \mid p' \text{ a vertex of } N'\}$ is a barycentric refinement of $\{U_\alpha \mid \alpha \in \mathscr{A}\}$. This indicates the origin of the term *barycentric refinement.*

6. Second-countable Spaces; Lindelöf Spaces

In this section, we study two properties of spaces related to the behavior of their open coverings; it turns out that when any one of them is present, weak separation properties become very strong.

6.1 Definition A Hausdorff space is 2° countable (or, satisfies the second axiom of countability) if it has a countable basis.

In recent literature, the least cardinal of a basis for a space X is called the weight of X; thus, X is 2° countable if it has weight $\leqslant \aleph_0$.

Ex. 1 E^n is 2° countable, as seen in III, **2**, Ex. 3. A countable discrete space is

2° countable, whereas an uncountable discrete space is not.

Ex. 2 The space E^1_u is not 2° countable: If $\{]a_n, b_n] \mid n \in Z\}$ is any countable collection of open sets, then by choosing $a, b \neq$ any b_n, the open set $]a, b]$ can evidently not be exhibited as a union of sets $]a_n, b_n]$.

6.2 Theorem (1). 2° countability is invariant under continuous open surjections.

(2). Every subspace of a 2° countable space is 2° countable.

(3). $\prod\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ is 2° countable if and only if each Y_α is second countable and all but at most $\aleph_0$ are spaces consisting of a single point.

Proof: (1). If $\{U_n \mid n \in Z\}$ is a basis for X and $p: X \to Y$ is a continuous open surjection, $\{p(U_n) \mid n \in Z\}$ is easily verified to be a basis for Y.

(2). This is trivial.

(3). If $P = \prod Y_\alpha$ is 2° countable, then from (1), each Y_α is 2° countable. Let $\mathscr{B}$ be a countable basis for P and note that, if $\aleph(Y_\alpha) \geqslant 2$, then there is some $B \in \mathscr{B}$ with $\varnothing \neq p_\alpha(B) \neq Y_\alpha$ (choose a B contained in an open $\langle U_\alpha \rangle$ where $\varnothing \neq U_\alpha \neq Y_\alpha$). But (IV, **1**, Ex. 4) $\aleph\{\alpha \mid p_\alpha(B) \neq Y_\alpha\} < \aleph_0$ for each $B \in \mathscr{B}$ and $\aleph(B) \leqslant \aleph_0$, so there are, at most, countably many Y_α with $\aleph(Y_\alpha) \geqslant 2$.

Now assume each Y_α is 2° countable and $\aleph(Y_\alpha) \geqslant 2$ for each α; the cardinal of the subbasis $\mathscr{S}$ for P is $\leqslant \aleph_0 \cdot \aleph(\mathscr{A})$ so, by II, **8.8**, the cardinal of a basis (that is, all finite subsets of $\mathscr{S}$) is also $\leqslant \aleph_0 \cdot \aleph(\mathscr{A})$; thus P is 2° countable whenever $\aleph(\mathscr{A}) \leqslant \aleph_0$.

The main property of 2° countable spaces is

6.3 Theorem (Lindelöf) If Y is 2° countable, then every open covering $\{U_\alpha\}$ has a countable subcovering.

Proof: Let $\{V_i \mid i \in Z\}$ be a countable basis for Y. Since each U_α is a union of the V_i, there is a covering $\{V_{i_j} \mid j \in Z\}$ refining $\{U_\alpha\}$; choosing one $U_{\alpha_j} \supset V_{i_j}$ for each j gives the required countable subcovering $\{U_{\alpha_j}\}$.

Abstracting the property of **6.3** leads to

6.4 Definition A Hausdorff space Y is Lindelöf if each open covering contains a countable subcovering.

Ex. 3 All 2° countable spaces are Lindelöf. The converse is not true: From **2**, Ex. 6, it follows that E^1_u is Lindelöf but, by Ex. 2, it is not 2° countable.

6.5 Theorem (K. Morita) In Lindelöf spaces, the concepts of regularity and paracompactness are equivalent.

Proof: Paracompactness always implies regularity. For the converse, we shall show that **2.3**(2) holds: Given any open covering $\{U_\alpha\}$, extract a countable subcovering $\{U_{\alpha_i}\}$; then $\{U_{\alpha_i}\} \prec \{U_\alpha\}$ and $\{U_{\alpha_i}\}$ decomposes into countably many nbd-finite families, each consisting of one set U_{α_i}.

6.6 Theorem (1). The Lindelöf property is invariant under continuous surjections.

(2). A subspace of a Lindelöf space need not be Lindelöf. However, a closed subspace is Lindelöf.

(3). The cartesian product of Lindelöf spaces need not be Lindelöf. However, if the product is Lindelöf, each factor must be Lindelöf.

Proof: (1). Given $f: X \to Y$ and an open covering $\{U_\alpha\}$ of Y, find a countable subcovering $\{f^{-1}(U_{\alpha_i})\}$ of $\{f^{-1}(U_\alpha)\}$; then $\{U_{\alpha_i}\}$ is a countable subcovering of $\{U_\alpha\}$.

(2). It is shown in **2**, Ex. 2, that $[0, \Omega]$ is in fact Lindelöf; since we know that the subspace $[0, \Omega[$ is regular and not paracompact, it cannot be Lindelöf. The remaining part of (2) is trivial.

(3). E^1_u is Lindelöf. If $E^1_u \times E^1_u$ were Lindelöf, then the subspace $C = \{(x, -x) \mid x \text{ is irrational}\}$, being closed in $E^1_u \times E^1_u$, would also be Lindelöf; but this is impossible, since C is uncountable and discrete. The second assertion follows from (2).

7. Separability

7.1 Definition A Hausdorff space is separable if it contains a countable dense set.

Ex. 1 E^n and countable discrete spaces are separable; uncountable discrete spaces are not separable.

Ex. 2 E^1_u is separable: the rationals are a countable dense set.

For the invariance properties, we have

7.2 Theorem (1). The continuous image of a separable space is separable.

(2). A subspace of a separable space need not be separable. However, an open subspace is separable.

(3). $\prod\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ is separable if and only if each Y_α is separable, and all but at most $2^{\aleph_0}$ are spaces consisting of a single point

Proof: (1). Let $f: X \to Y$ be continuous and $D \subset X$ be dense; since $Y = f(X) = f(\bar{D}) \subset \overline{f(D)}$, $f(D)$ is dense in Y.

(2). It is trivial to verify that if G is open, $G \cap D$ is dense in G; an example for a subspace of a separable space that is not separable, and even closed, is given in Ex. 3 below.

(3). "Only if": Because each projection is continuous, the separability of each factor follows from (1). Now, let $\mathscr{B} = \{\alpha \in \mathscr{A} \mid \aleph(Y_\alpha) \geqslant 2\}$, and for each $\beta \in \mathscr{B}$, let U_β, V_β be nonempty disjoint open sets in Y_β. Let D be a countable dense set in $\prod\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ and for each $\beta \in \mathscr{B}$, let $D^\beta = D \cap \langle U_\beta \rangle$. Observe that if $\beta \neq \gamma$, then $D^\beta \neq D^\gamma$: for, since D is dense in $\prod Y_\alpha$, there is some $d \in D \cap \langle U_\beta, V_\gamma \rangle = D \cap \langle U_\beta \rangle \cap \langle V_\gamma \rangle$, and it is clear that $d \in D^\beta$ but $d \bar{\in} D^\gamma$. Thus, the map $\beta \to D^\beta$ of $\mathscr{B}$ into $\mathscr{P}(D)$ is injective, and therefore $\aleph(\mathscr{B}) \leqslant \aleph(\mathscr{P}(D)) \leqslant 2^{\aleph_0}$

"If": There is no loss in generality to assume that $\aleph(Y_\alpha) \geqslant 2$ for each $\alpha \in \mathscr{A}$. In each Y_α, let $\{y_\alpha(n) \mid n = 0, 1, 2, \cdots\}$ be a countable dense set. Since $\aleph(\mathscr{A}) \leqslant \mathfrak{c}$, there is a bijection φ of $\mathscr{A}$ onto a subset K of the unit interval I; we shall assume that $\aleph(\mathscr{A}) = \mathfrak{c}$ and $K = I$ because in all cases the details are similar. For each finite pairwise disjoint family $J_1, \cdots, J_k$ of closed intervals in I with rational endpoints, and each finite set $n_1, \cdots, n_k$ of nonnegative integers, let $p(J_1, \cdots, J_k; n_1, \cdots, n_k)$ be the point $\{y_\alpha(s_\alpha)\} \in \prod Y_\alpha$, where for each α, we take $s_\alpha = n_i$ if $\varphi(\alpha) \in J_i$ and $s_\alpha = 0$ otherwise. It is clear that the set

$$\{p(J_1, \cdots, J_k; n_1, \cdots, n_k) \mid \text{all } (J_1, \cdots, J_k; n_1, \cdots, n_k), \text{all } k\}$$

is countable. It is also dense: given any $\langle U_{\alpha_1}, \cdots, U_{\alpha_k} \rangle$, find pairwise disjoint $J_i \supset \varphi(\alpha_i)$, and for each i find an n_i such that $y_{\alpha_i}(n_i) \in U_{\alpha_i}$; then

$$p(J_1, \cdots, J_k; n_1, \cdots, n_k) \in \langle U_{\alpha_1}, \cdots, U_{\alpha_k} \rangle.$$

Ex. 3 $E^1_u \times E^1_u$ is separable, since E^1_u is separable. The subspace $\{(x, -x) \mid x$ is irrational$\}$ is closed and uncountable; being discrete, it is not separable.

Ex. 4 If a Hausdorff space X is separable, then $\aleph(X) \leqslant 2^{\mathfrak{c}}$. To see this, let D be a countable dense set, and for each $x \in X$ define a map $\varphi_x: \mathscr{P}(D) \to 2$ as follows: $\varphi_x(A) = 1$ if $A = D \cap U$ for some nbd U of x, and $\varphi_x(A) = 0$ otherwise. The map $x \to \varphi_x$ of X into $2^{\mathscr{P}(D)}$ is injective: if $x \neq y$, choose disjoint nbds $U(x)$, $U(y)$ and let $A = D \cap U(x)$; then $\varphi_x(A) = 1$, but $\varphi_y(A) = 0$ because every nbd of y will contain points of the dense set D which are not in $U(x)$. Since $x \to \varphi_x$ is injective and $\aleph[\mathscr{P}(D)] \leqslant \mathfrak{c}$, the desired conclusion follows. Observe that by taking X to be the cartesian product of $2^{\aleph_0}$ discrete spaces $\{0, 1\}$, Theorem **7.2**(3), shows that this estimate of $\aleph(X)$ cannot be improved.

The relation of separability to the previous concepts is

7.3 Theorem If Y is 2° countable, then every subspace of Y is separable.

Proof: Letting $\{U_i\}$ be a basis for Y, choose a $y_i \in U_i$ for each i; $\{y_i\}$ is countable; it is dense because each open U is a union of the U_i.

Ex. 5 The converse of **7.3** is not true: Every subspace of E_u^1 is separable, yet it is not 2° countable.

It is important to observe that separability and Lindelöf are not related: *$[0, \Omega]$ is Lindelöf, as shown in* **2**, Ex. 2, *but it is not separable* (since any countable set of ordinals $< \Omega$ has an upper bound $< \Omega$). *On the other hand,* $E_u^1 \times E_u^1$ *is separable, but it is not Lindelöf.* However,

7.4 *In paracompact spaces, separability implies Lindelöf.*

Proof: Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be a given open covering of the separable paracompact Y, and let $\{V_\beta \mid \beta \in \mathscr{B}\}$ be an open nbd-finite refinement; we can assume that each $V_\beta \neq \varnothing$. Because Y has a countable dense set $\{d_i\}$, the family $\{V_\beta\}$ is at most countable: otherwise, since each V_β contains at least one d_i, there would be some d_n contained in uncountably many V_β, which would contradict nbd-finiteness. Choosing for each V_β a $U_\alpha \supset V_\beta$ gives the required countable subcovering.

As $[0, \Omega]$ shows, a paracompact Lindelöf space need not be separable.

Problems

Section 1

1. Call a covering of a space *star-finite* if each set of the covering intersects at most finitely many others. Show that the concepts of star-finiteness and nbd-finiteness are independent, but that for open coverings, star-finiteness implies nbd-finiteness.
2. Let X be Hausdorff, Y arbitrary, and $f: X \to Y$ surjective. Prove that f is a bijective open map if and only if for each open covering $\{U\}$ of X there exists an open covering $\{V\}$ of Y such that $\{f^{-1}(V)\}$ refines $\{U\}$.
3. Let Y be any space, and $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ a covering of Y having a nbd-finite closed refinement. Show that $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ has a precise nbd-finite closed refinement.
4. Let Y be any space, let $\Delta = \{(y, y) \mid y \in Y\}$ be the diagonal in $Y \times Y$, and let $U \supset \Delta$ be any set open in $Y \times Y$. For each $y \in Y$, let $U[y] = \{z \mid (y, z) \in U\}$. Prove:
   a. $\Delta(U) = \{U[y] \mid y \in Y\}$ is an open covering of Y.
   b. If $\mathfrak{B} = \{V_\alpha \mid \alpha \in \mathscr{A}\}$ is an open covering of Y that has a nbd-finite closed refinement, then there is a nbd U of the diagonal Δ such that $\Delta(U)$ refines $\mathfrak{B}$.

   [*Hint:* Let $\{F_\alpha \mid \alpha \in \mathscr{A}\}$ be a precise closed nbd-finite refinement of $\mathfrak{B}$. For each $\alpha \in \mathscr{A}$, let $W_\alpha = (V_\alpha \times V_\alpha) \cup (\mathscr{C}F_\alpha \times \mathscr{C}F_\alpha)$; then $U = \operatorname{Int} \bigcap \{W_\alpha \mid \alpha \in \mathscr{A}\}$ is not empty, and is the required nbd of Δ.]

5. Let $\{F_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite closed covering of a space Y. Prove: $\{\overline{\mathrm{Int}(F_\alpha)} \mid \alpha \in \mathscr{A}\}$ is also a nbd-finite closed covering of Y. [*Hint:* Well-order the indexing set $\mathscr{A}$, and use transfinite induction to show that each family $\mathfrak{F}_\beta = \{\overline{\mathrm{Int}(F_\alpha)} \mid \alpha \leqslant \beta\} \cup \{F_\gamma \mid \gamma > \beta\}$ is a covering.]
6. A grating of a space Y is a family $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ of pairwise disjoint open sets such that $\{\bar{V}_\alpha \mid \alpha \in \mathscr{A}\}$ is a covering of Y and $\mathrm{Int}(\bar{V}_\alpha) = V_\alpha$ for each $\alpha \in \mathscr{A}$. Let $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ and $\{W_\beta \mid \beta \in \mathscr{B}\}$ be two gratings of Y. Prove:

$$\{V_\alpha \cap W_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$$

is a grating of Y. [*Cf.* Problems III, **4.22.**]
7. Let $\{F_\alpha \mid \alpha \in \mathscr{A}\}$ be an nbd-finite closed covering of a space Y. Prove that there exists a grating $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ of Y such that $V_\alpha \subset F_\alpha$ for each $\alpha \in \mathscr{A}$. [*Hint:* Well-order $\mathscr{A}$, and let

$$V_\alpha = \mathrm{Int}(F_\alpha) - \bigcup \{\overline{\mathrm{Int}(F_\beta)} \mid \beta \text{ precedes } \alpha\}].$$

Section 2

1. Let Y be paracompact and let $p: X \to Y$ be an identification. Show: each covering of X by p-saturated open sets has a nbd-finite open refinement.
2. Let X be paracompact and $A \subset X$ closed. Show that X/A is paracompact.
3. Let $A \subset X$ be closed and $f: A \to Y$ be continuous. Assume that X, Y are paracompact and that f is a closed map. Prove that $X \cup_f Y$ is paracompact. (The conclusion is also true without the assumption that f is a closed map).
4. Let Y be paracompact, and let $A \subset Y$ be a subset with the following property: for each open $U \supset A$, there is an F_σ-set F with $A \subset F \subset U$. Show that A is paracompact.
5. Let Y be regular, and let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite family of closed sets in Y. Assume that each A_α is paracompact. Prove: $\bigcup_\alpha A_\alpha$ is paracompact.
6. Prove: Each subset of a perfectly normal paracompact space is paracompact.
7. Prove: If Y is paracompact, then for every nbd U of the diagonal in $Y \times Y$ (*cf.* **1**, Problem 4) there exists an nbd V of the diagonal such that

$$\bigcup \{V[y] \times V[y] \mid y \in Y\} \subset U.$$

8. A space is called countably paracompact if each countable open covering has an open nbd-finite refinement. Prove: Y is countably paracompact if and only if for every descending sequence $F_1 \supset F_2 \supset \cdots$ of nonempty closed sets such that $\bigcap F_i = \varnothing$, there exists a sequence of open sets $G_i \supset F_i$ such that $\bigcap \bar{G}_i = \varnothing$. [*Hint:* "Only if": take $\mathfrak{V}$ to be a precise open nbd-finite refinement of $\{Y - F_n \mid n \in Z^+\}$ and let $G_n = \bigcup_{n+1}^{\infty} V_i$. "If": Given an open covering $\{U_n \mid n \in Z^+\}$, let $F_n = Y - \bigcup_1^n U_i$, and find a family $\{G_n\}$; then $\{G_n \cap U_{n+1} \mid n \in Z^+\}$ is the required refinement.]

Section 3

1. For any space Y and any covering $\mathfrak{U}$, let $\mathrm{St}(\mathfrak{U})$ be the covering $\{\mathrm{St}(y, \mathfrak{U}) \mid y \in Y\}$. Prove $\mathrm{St}[\mathrm{St}(M, \mathfrak{U}), \mathfrak{U}] = \mathrm{St}[M, \mathrm{St}(\mathfrak{U})]$ for every $M \subset Y$.

2. Let $\mathfrak{F} = \{F_\alpha \mid \alpha \in \mathscr{A}\}$ be an nbd-finite closed covering of a space Y. Prove $\{\text{Int}[\text{St}(y, \mathfrak{F})] \mid y \in Y\}$ is an open covering of Y.
3. Show that the space $(I, \mathscr{T}_1)$ of VII, **3**, Problem 7 is paracompact.
4. Supply the details in the following outline of Stone's proof of his Theorem 3.5 (it is in this proof that the important idea of decomposing an open covering into countably many nbd-finite families was first presented). Let $\mathfrak{U}_0 = \{U_\alpha \mid \alpha \in \mathscr{A}\}$ be the given open covering, and let $\{\mathfrak{U}_n \mid n \in Z^+\}$ be a sequence of open coverings where each $\mathfrak{U}_n$ is a barycentric refinement of $\mathfrak{U}_{n-1}$.

   For each $A \subset Y$, let $(A, -n) = \{y \mid \text{St}(y, \mathfrak{U}_n) \subset A\}$ and $(A, n) = \text{St}(A, \mathfrak{U}_n)$.

   For each $\alpha \in \mathscr{A}$, let $V_\alpha^1 = (U_\alpha, -1)$, $V_\alpha^2 = (V_\alpha^1, 2), \cdots, V_\alpha^n = (V_\alpha^{n-1}, n), \cdots$.

   Prove: (i). $V_\alpha = \bigcup_n V_\alpha^n \subset U_\alpha$ [Show by induction that in fact $(V_\alpha^n, n) \subset U_\alpha$].

   a. $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ is an open covering of Y and has the property: If $x \in V_\alpha$, then $(x, n) \subset V_\alpha$ for suitable n.

   Now, well-order $\mathscr{A}$ and for each $n \in Z^+$ define a transfinite sequence of sets by

   $$H_{n,1} = (V_1, -n)$$
   $$\cdots$$
   $$H_{n,\alpha} = (V_\alpha - \bigcup \{H_{n,\beta} \mid \beta < \alpha\}, -n)$$
   $$\cdots$$

   Prove:

   b. $\{H_{n,\alpha} \mid (n, \alpha) \in Z^+ \times \mathscr{A}\}$ is a covering [Given $y \in Y$, let $\beta = \min \{\alpha \mid y \in V_\alpha\}$ and choose n so that $(y, n) \subset V_\beta$; then $y \in H_{n,\beta}$].

   c. Each $U \in \mathfrak{U}_{n+2}$ can intersect at most one $H_{n,\alpha}$ [If $U \cap H_{n,\alpha} \neq \varnothing$, then $U \subset V_\alpha - \bigcup_{\beta < \alpha} H_{n,\beta}$].

   d. Let $G_{n,\alpha} = \text{St}(H_{n,\alpha}, \mathfrak{U}_{n+2})$. Then $\{G_{n,\alpha} \mid (n, \alpha) \in Z^+ \times \mathscr{A}\}$ is an open covering, and for each n, a set $U \in \mathfrak{U}_{n+2}$ can intersect at most one $G_{n,\alpha}$.

   e. Define $F_n = \bigcup_\alpha \overline{\text{St}(H_{n,\alpha}, \mathfrak{U}_{n+3})}$. Then each F_n is closed.

   f. Let $W_{1,\alpha} = G_{1,\alpha}$ and for each $n > 1$ let $W_{n,\alpha} = G_{n,\alpha} - (F_1 \cup \ldots \cup F_{n-1})$. Then $\{W_{n,\alpha} \mid (n, \alpha) \in Z^+ \times \mathscr{A}\}$ is a nbd-finite open covering and $W_{n,\alpha} \subset U_\alpha$ for each (n, α). [Each $\text{St}(y, \mathfrak{U}_{n+3})$ meets no $W_{k,\beta}$ for $k > n$ and at most one $W_{l,\beta}$ for each $l \leqslant n$.]
5. Let X be paracompact, and $A \subset X$ closed. Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite open covering of A. Show that there exists a nbd-finite open covering $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ of X such that $U_\alpha = A \cap V_\alpha$ for each $\alpha \in \mathscr{A}$. [Hint: Let $\mathfrak{W}$ be an open star-refinement of an nbd-finite open covering in which each set meets at most finitely many of the U_α. Choose an index α_0, and let

   $V_{\alpha_0} = U_{\alpha_0} \cup (X - A)$, $V_\alpha = U_\alpha \cup [\text{St}(U_\alpha, \mathfrak{W}) - A]$ if $\alpha \neq \alpha_0$.]

Section 4

1. Let Y be paracompact and let $\{U_\alpha\}$ be a nbd-finite open covering. Assume that for each U_α, there is given a real number $i_\alpha > 0$. Show that there exists a continuous $\varphi\colon Y \to E^1$ such that $\varphi(y) > 0$ for all $y \in Y$, and

   $$\varphi(y) \leqslant \sup\{i_\alpha \mid y \in U_\alpha\}$$

   for each y.

Section 5

1. Let $\{U_\alpha\}$ be an open covering of Y, and let $\kappa: Y \to N(U_\alpha)$ be the map of **5.4**. Let $f: Y \to Y$ be continuous, and such that $f(U_\alpha) \cap U_\alpha = \varnothing$, for each α. Show $\kappa \circ f(y) \neq \kappa(y)$ for each $y \in Y$.

Section 6

1. Prove $[0, \Omega]$ is Lindelöf, but not second countable.
2. A point y in a space Y is called a condensation point if each open set containing y contains uncountably many elements of Y. Prove: If Y is second countable, and $\aleph(Y) > \aleph_0$, then Y has at least one condensation point.
3. Let Y be second countable. Prove:
   a. From any family of open sets in Y (not necessarily a covering of Y), one can always extract a countable subfamily having exactly the same union.
   b. Any family of pairwise disjoint open sets is necessarily countable.
   c. Any chain of open sets, well-ordered by inclusion, is necessarily countable. Conclude that any increasing, well-ordered sequence of real numbers is necessarily countable.
   d. $\aleph(Y) \leqslant 2^{\mathfrak{c}}$ and the cardinal of the topology is $\leqslant \mathfrak{c}$.
4. Let Y be 2° countable. Call a property P of subsets of Y *inductive* if for each descending sequence $A_1 \supset A_2 \supset \cdots$ of closed sets, whenever each A_i has P, then so also does $\bigcap_1^\infty A_i$. The set $A \subset Y$ is called P-irreducible if no proper closed subset has property P. Prove: If P is inductive and some closed set has property P, then there is a P-irreducible closed subset of Y having property P.
5. State and prove the analogs of **6.2**, **6.3** for spaces of weight $\aleph$.
6. Let X, Y, be Lindelöf, and $A \subset X$ a closed set. Let $f: A \to Y$ be continuous. Prove: $X \cup_f Y$ is Lindelöf.
7. If each open subset of a regular space Y is Lindelöf, prove that Y is perfectly normal (and paracompact). (*Hint:* Use VII, 5, Problem 2.)
8. Prove: Y is Lindelöf if and only if every open covering has a countable refinement.

Section 7

1. Let $A \subset E^1$ be the subspace of irrationals. Is this space separable?
2. Let Y be any space, and let D be an infinite dense subset. Show that for any point-finite covering $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ of Y by nonempty open sets, card $\{U_\alpha\} \leqslant$ card D.

Metric Spaces

IX

In this chapter, we will study spaces in which the topology is derived from a notion of distance.

1. Metrics on Sets

1.1 Definition A metric (or distance function) on a *set* Y is a map $d\colon Y \times Y \to E^1$ with the properties:

(1). $d(x, y) \geqslant 0$ for each pair x, y.
(2). $d(x, y) = 0$ if and only if $x = y$.
(3). $d(x, y) = d(y, x)$ for all x, y (symmetry).
(4). $d(x, z) \leqslant d(x, y) + d(y, z)$ for each triple of points (triangle inequality).

$d(x, y)$ is called the distance between x and y.

Ex. 1 If Y is any set, defining $d(x, y) = 1$ if $x \neq y$ and $d(x, y) = 0$ if $x = y$ gives a metric on Y.

Ex. 2 In the *set* of all real numbers E^1, $d(x, y) = |x - y|$ is a metric. More generally, in the *set* E^n, $d_0(x, y) = \max\{|x_i - y_i| \mid 1 \leqslant i \leqslant n\}$ is a metric: the triangle inequality follows because

$$|x_i - z_i| \leqslant |x_i - y_i| + |y_i - z_i| \leqslant d_0(x, y) + d_0(y, z)$$

for each $1 \leqslant i \leqslant n$.

Ex. 3 In the *set* E^n,

$$d_p(x, y) = \sqrt[p]{\sum_1^n |x_i - y_i|^p}$$

is a metric for each $p \geq 1$. To verify the triangle inequality, note that for $p \geq 1$, the function $f(x) = x^p$ satisfies $f''(x) \geq 0$ on $\{x \in E^1 \mid x \geq 0\}$ and so is convex, that is, it satisfies $f(\lambda x + [1 - \lambda]y) \leq \lambda f(x) + [1 - \lambda] f(y)$ for $x, y \geq 0$. Thus, for any set of real numbers $a_1, \cdots, a_n, b_1, \cdots, b_n$, letting

$$A = \sqrt[p]{\sum_1^n |a_i|^p}, \qquad B = \sqrt[p]{\sum_1^n |b_i|^p}$$

and expressing the convexity of x^p for each $1 \leq i \leq n$ with

$$x = \frac{a_i}{A}, \qquad y = \frac{b_i}{B}, \qquad \lambda = \frac{A}{A + B},$$

addition of these n inequalities gives Minkowski's inequality:

$$\sqrt[p]{\sum_1^n |a_i + b_i|^p} \leq \sqrt[p]{\sum_1^n |a_i|^p} + \sqrt[p]{\sum_1^n |b_i|^p};$$

the triangle inequality follows by taking $a_i = x_i - y_i$, $b_i = y_i - z_i$, for $1 \leq i \leq n$.

Ex. 4 Let Y be any set and $C(Y) = \{f \mid f\colon Y \to E^1$ and f is bounded$\}$. Then $d(f, g) = \sup\{|f(y) - g(y)| \mid y \in Y\}$ is a metric in $C(Y)$. Note that if some f_0 were unbounded, d would *not* be a metric: using $g(y) \equiv 0$, the function d is not defined at the point $(f_0, g) \in C(Y) \times C(Y)$.

If axiom (2) is present only in the weaker form (2′): $(x = y) \Rightarrow (d(x, y) = 0)$, d is called an écart or a pseudometric in Y. An écart ρ in Y always gives a metric in a suitable quotient set: By defining xRy to mean $\rho(x, y) = 0$, R is easily seen to be an equivalence relation in Y, and $d(Rx, Ry) = \rho(x, y)$ is a metric in the quotient set Y/R.

The absence of only axiom (3) is easily remedied:

$$d'(x, y) = d(x, y) + d(y, x)$$

gives a metric d' in Y itself.

2. Topology Induced by a Metric

With each metric d in a set Y, we are going to associate a definite topology $\mathscr{T}(d)$ in Y.

The set $B_d(a, r) = \{y \mid d(y, a) < r\}$ is called the d-ball of radius r and center a. Clearly, $B_d(a, r) \subset B_d(a, r')$ if $r \leq r'$ and $B_d(a, 0) = \varnothing$. In the future, we will omit the distinguishing d whenever the metric is clear from the context.

2.1 The family $\{B_d(y, r) \mid y \in Y, r > 0\}$ of all d-balls in Y can serve as the basis for a topology.

Proof: According to III, **3.2**, we need verify only that if

$$a \in B(x_0, r_x) \cap B(y_0, r_y),$$

then a belongs to some ball lying in this intersection. Let

$$r = \min[r_x - d(a, x_0), r_y - d(a, y_0)];$$

then $r > 0$, since the statements "$a \in B(\xi, r_\xi)$" and "$d(a, \xi) < r_\xi$" are equivalent, so $a \in B(a, r)$. Furthermore, $B(a, r) \subset B(x_0, r_x) \cap B(y_0, r_y)$ because if $x \in B(a, r)$, then $d(x, x_0) \leqslant d(x, a) + d(a, x_0) < r + d(a, x_0) \leqslant [r_x - d(a, x_0)] + d(a, x_0) = r_x$; that is, $x \in B(x_0, r_x)$ and, similarly, $x \in B(y_0, r_y)$.

Because of **2.1**, the following definition is therefore legitimate:

2.2 Definition Let Y be a set and d be a metric in Y. The topology $\mathscr{T}(d)$, having for basis the family $\{B_d(y, r) \mid y \in Y, r > 0\}$ of all d-balls in Y, is called the topology in Y induced (or determined) by the metric d.

Ex. 1 In the *set* E^n, let d be the metric of **1**, Ex. 1; then $\mathscr{T}(d)$ is the discrete topology, since $B_d(x, \frac{1}{2}) = x$ for each $x \in E^n$.

Ex. 2 In the *set* E^n, let d_0 be the metric of **1**, Ex. 2; then $\mathscr{T}(d_0)$ is the Euclidean topology, since the d_0-balls are precisely the open cubes in E^n, which we have seen (III, **3**, Ex. 8) are a basis for the Euclidean topology.

Ex. 3 The family $\{B_d(y, r) \mid (y \in Y) \wedge (r \text{ is rational})\}$ is also a basis for $\mathscr{T}(d)$, as is immediate from III, **2.2**.

The converse question arises: Given a topological space $(Y, \mathscr{T})$, is there a metric d in Y such that $\mathscr{T} = \mathscr{T}(d)$? The answer is "no" in general. For example, it is easy to see that no metric can induce the topology in Sierpinski space.

2.3 Definition A topological *space* $(Y, \mathscr{T})$ is called a metric (or metrizable) space if its topology is that induced by a metric in Y. A metric for a *space* Y is one that induces its topology.

With this terminology, the Euclidean space E^n is a metric space, and d_0 is a metric for this space. In metric spaces, topological concepts can be phrased in the ε, δ terms of classical analysis. For example,

2.4 Let X have topology $\mathscr{T}(d)$ and Y have topology $\mathscr{T}(\rho)$. An $f: X \to Y$ is continuous if

$$\forall x \, \forall \varepsilon > 0 \, \exists \, \delta(\varepsilon, x) > 0: d(\xi, x) < \delta \Rightarrow \rho(f(\xi), f(x)) < \varepsilon;$$

that is, if $f[B_d(x, \delta)] \subset B_\rho[f(x), \varepsilon]$.

Ex. 4 It is immediate from the triangle inequality that the formula

$$| d(x, y) - d(x', y') | \leqslant d(x, x') + d(y, y')$$

is valid for any metric d. Using this, the reader can easily prove that if $(Y, \mathscr{T}(d))$ is any metric space and if the cartesian product topology is used in $Y \times Y$, then the map $d: Y \times Y \to E^1$ is continuous (and, in fact, an identification.)

3. Equivalent Metrics

In this section we give a criterion for determining in advance whether two different metrics in a set Y will induce the same topology.

3.1 Definition Two metrics d, ρ, in a set Y are called equivalent, $d \sim \rho$, if $\mathscr{T}(d) = \mathscr{T}(\rho)$.

This is clearly an equivalence relation in the set of all metrics on Y.

3.2 Theorem Let ρ, d, be two metrics on Y. A necessary and sufficient condition that $\rho \sim d$ is that for each $a \in Y$ and $\varepsilon > 0$, the following two conditions hold:

(1). $\exists\ \delta_1 = \delta_1(a, \varepsilon): \rho(a, y) < \delta_1 \Rightarrow d(a, y) < \varepsilon$.
(2). $\exists\ \delta_2 = \delta_2(a, \varepsilon): d(a, y) < \delta_2 \Rightarrow \rho(a, y) < \varepsilon$.

Proof: This is simply III, **3.4**, or equivalently, metric statements (*cf.* **2.4**) that $1: (Y, \mathscr{T}(d)) \to (Y, \mathscr{T}(\rho))$ is a homeomorphism.

Ex. 1 In the set E^n, all the metrics d_p, $p \geqslant 1$ in **1**, Ex. 3, are equivalent to the d_0 in **1**, Ex. 2, and therefore all metrize the Euclidean topology: this follows from **2**, Ex. 2 and the observations that, for each $p \geqslant 1$,

(a). $d_p(x, y) \leqslant \sqrt[p]{n} \cdot d_0(x, y)$,

and

(b). $d_0(x, y) \leqslant d_p(x, y)$.

3.3 Corollary Let $(Y, \mathscr{T}(d))$ be a metric space. Then for each $M > 0$ there is a metric $\rho_M \sim d$ such that $\rho_M(x, y) \leq M$ for all (x, y). Equivalently, each metric space is homeomorphic to a bounded metric space.

Proof: Given M, define $\rho_M(x, y) = \min[M, d(x, y)]$; it is trivial to verify that ρ_M is indeed a metric for Y and, using **3.2**, that $d \sim \rho_M$.

4. Continuity of the Distance

4.1 Definition In a metric space Y, with metric d,

(1). The distance of a point y_0 to a nonempty set A is

$$d(y_0, A) = \inf\{d(y_0, a) \mid a \in A\}.$$

(2). The distance between nonempty sets A and B is

$$d(A, B) = \inf\{d(a, b) \mid a \in A, b \in B\} = \inf\{d(a, B) \mid a \in A\}.$$

(3). The diameter of a nonempty set A is

$$\delta(A) = \sup\{d(x, y) \mid x \in A, y \in A\}.$$

Ex. 1 $\delta(B_d(a, r)) \leqslant 2r$, since if $x, y \in B_d(a, r)$, then

$$d(x, y) \leqslant d(x, a) + d(a, y) < 2r;$$

as **2**, Ex. 1, shows, we may have $\delta(B(a, r))$ actually smaller than r.

Ex. 2 By convention, $\delta(\varnothing) = 0$. A set A is called bounded if $\delta(A) < \infty$; d is a *bounded* metric if $\delta(Y) < \infty$.

Ex. 3 $d(A, B) \neq 0 \Rightarrow A \cap B = \varnothing$; the converse implication need not be true, *even though both A and B are closed:* In E^1, let

$$A = \{n \mid n \in Z^+\} \quad \text{and} \quad B = \{n + (1/2n) \mid n \in Z^+\}.$$

To obtain this converse implication at least in the trivial cases that $A = \varnothing$ or $B = \varnothing$, we define $d(A, \varnothing) = d(\varnothing, B) = \infty$.

4.2 (a). $d(y, A) = 0$ if and only if $y \in \overline{A}$; thus $\overline{A} = \{y \mid d(y, A) = 0\}$.
(b). $\rho \sim d$ if and only if for each $A \subset Y$, $d(y, A) = 0 \Leftrightarrow \rho(y, A) = 0$.

Proof: (a). We have $y \in \overline{A} \Leftrightarrow \forall\, B(y, r)\colon A \cap B(y, r) \neq \varnothing \Leftrightarrow \forall\, r > 0\, \exists\, a_r \in A\colon d(y, a_r) < r \Leftrightarrow d(y, A) = 0$.

(b). By what we have just proved, $\mathscr{T}(\rho)$ and $\mathscr{T}(d)$ give the same closure operation in Y so (III, **5.1**) $\mathscr{T}(\rho) = \mathscr{T}(d)$.

4.3 Theorem Let Y be a metric space, with metric d, and let A be any subset of Y. Then the map $f\colon Y \to E^1$ defined by $y \to d(y, A)$ is continuous.

Proof: Let x, y be any two elements of Y. Then, for each $a \in A$, we have $d(x, a) \leqslant d(x, y) + d(y, a)$, so that

$$d(x, A) = \inf_a d(x, a) \leqslant d(x, y) + \inf_a d(y, a) = d(x, y) + d(y, A),$$

which shows that $d(x, A) - d(y, A) \leqslant d(x, y)$. Interchanging the roles of x and y, we obtain

$$|d(x, A) - d(y, A)| \leqslant d(x, y),$$

which clearly gives the continuity of f (*cf.* **2.4**).

5. Properties of Metric Topologies

In this section, we derive the main properties of metric topologies, but defer the study of cartesian products to a later section.

5.1 Theorem (1). Metrizability is a topological invariant.

(2). Every subspace of a metric space is also a metric space. Precisely: Given $(Y, \mathscr{T}(d))$ and $A \subset Y$, the topology induced by $\mathscr{T}(d)$ on A is exactly the topology derived from the metric $d \mid A \times A: A \times A \to E^1$.

Proof: (1). Let $h: (X, \mathscr{T}) \to (Y, \mathscr{T}(d))$ be a homeomorphism, and define $\rho(x, x') = d(h(x), h(x'))$; then ρ is clearly a metric on the set X, and since h is evidently a homeomorphism (*cf.* **2.4**) of $(X, \mathscr{T}(\rho))$ with $(Y, \mathscr{T}(d))$, we have $\mathscr{T} = \mathscr{T}(\rho)$. The proof of (2) is obvious.

The position of metric topologies is

5.2 Every metric space is perfectly normal.

Proof: For any closed $A \subset Y$, the continuous map $f: Y \to E^1$ given by $y \to d(y, A)$ vanishes only on A because of **4.2**(a).

Much more important is

5.3 Theorem (A. H. Stone) Every metric space is paracompact.

Proof: We will show that the requirement of VIII, **3.7**, is satisfied; in fact, there is *one* sequence $\{\mathfrak{B}_n \mid n \in Z^+\}$ that is locally starring for *every* open covering. Let d be a metric for the space X, and for each $n \in Z^+$ let $\mathfrak{B}_n = \{B(x, 1/n) \mid x \in X\}$. Given any open covering $\{U_\alpha \mid \alpha \in \mathscr{A}\}$, and any $x \in X$, choose a U_α containing x; then $d(x, \mathscr{C}U_\alpha) \geqslant 1/n > 0$ for a suitable n, and taking $V(x) = B(x, 1/3n)$, it is evident that $\mathrm{St}(V, \mathfrak{B}_{3n}) \subset U_\alpha$.

Ex. 1 $[0, \Omega]$ is not perfectly normal (the closed set $\{\Omega\}$ is not a G_δ); therefore it is not metrizable. In particular, not every paracompact space is metrizable.

Ex. 2 $[0, \Omega[$ is not metrizable because it is not paracompact (VIII, **2**, Ex. 3).

Metric spaces have another important property that we introduce with the general

5.4 Definition A space Y is 1° countable (or satisfies the first axiom of countability) if with each $y \in Y$ there is given an at most countable family $\{U_n(y) \mid n \in Z^+\}$ of nbds with the property: For each open $G \supset y$, there is some $U_n(y) \subset G$ (expressed briefly: if Y has a countable basis at each point).

Ex. 3 Clearly, 2° countability $\Rightarrow$ 1° countability, but not conversely: any uncountable discrete space is 1° countable, but not 2° countable.

Ex. 4 E_u^1 is 1° countable: assign to each $x \in E_u^1$ the family of all $]r_n, x]$, where $r_n < x$ is rational. Recall that E_u^1 is not second countable (VIII, **6**, Ex. 2).

Ex. 5 $[0, \Omega]$ is not 1° countable, though its subspace $[0, \Omega[$ is.

5.5 Every metric space is 1° countable.

Proof: By the definition of the topology $\mathscr{T}(d)$, it is evident that $\{B_d(y, r) \mid r \text{ rational}\}$ is a countable basis at y.

Ex. 6 A metric space need not be 2° countable, as uncountable discrete spaces (which are metrizable as in **1**, Ex. 1) show.

We have seen that 2° countability implies both separability and Lindelöf, but that even for paracompact spaces, neither of these implications is reversible (VIII, **6**, Ex. 3; **7**). However,

5.6 Theorem In metric spaces, the concepts of 2° countability, separability, and Lindelöf are all equivalent.

Proof: Separability $\Rightarrow$ 2° countability. Let $\{y_i\}$ be a countable dense set in the metric space Y; the family $\{B(y_i, r) \mid r \text{ rational}, i = 1, 2, \cdots\}$ is countable, and we show it is a basis by verifying that III, **2.2**, is true. Let G be open, and $y \in G$; there is some $B(y, r) \subset G$, r rational, and because $\{y_i\}$ is dense, we can find some $y_n \in B(y, r/3)$; then

$$y \in B\left(y_n, \frac{2r}{3}\right) \subset B(y, r) \subset G,$$

completing the proof.

By VIII, **6.3**, we find that 2° countability $\Rightarrow$ Lindelöf is always true.

Lindelöf $\Rightarrow$ separability. We will say that a set $A \subset X$ is ε-dense if each point of X is at a distance $< \varepsilon$ from A. For each $\varepsilon > 0$, there exists a countable ε-dense set $A(\varepsilon)$: From the open covering $\{B(x, \varepsilon) \mid x \in X\}$, we extract a countable covering $\{B(x_i, \varepsilon) \mid i \in Z\}$ and let $A(\varepsilon) = \{x_i \mid i \in Z\}$. Defining $D = \bigcup_1^\infty A(1/n)$ gives a countable dense set in the space.

Ex. 7 E_u^1 is not metrizable: it is separable, but not 2° countable.

6. Maps of Metric Spaces into Affine Spaces

The Tietze extension theorem (VII, **5.1**) for normal spaces can be improved if we use the richer structure of metric spaces.

In the Appendix (1, **4.4**) to this book, we have defined an affine space L to be of type m if, for every metric space X and every continuous $f: X \to L$, the following is true: For each $x \in X$ and nbd $W \supset f(x)$, there is a nbd $U \supset x$ and some convex set $C \subset L$ such that $f(U) \subset C \subset W$. This class of affine spaces includes all locally convex linear topological spaces (for example, Euclidean spaces, or more generally, Banach spaces) and also all real vector spaces with the finite topology (which, as shown in the Appendix, need not be locally convex nor even linear topological spaces).

6.1 Theorem (J. Dugundji) Let X be an arbitrary metric space, $A \subset X$ a closed subset, and L an affine space of type m. Then each continuous $f: A \to L$ has a continuous extension $F: X \to L$, and in fact $F(X) \subset$ [convex hull of $f(A)$].

Proof: For each $x \in X - A$, let B_x be an open ball centered at x with radius $< \frac{1}{2}d(x, A)$, where d is a metric for X. The family

$$\{B_x \mid x \in X - A\}$$

is an open covering of the paracompact $X - A$, so it has a nbd-finite refinement $\{U\}$. Let $B(A, \eta) = \{y \mid d(y, A) < \eta\}$, and observe that a ball B_x centered outside $B(A, 2\eta)$ cannot intersect $B(A, \eta)$; consequently, any $U \in \{U\}$ that intersects $B(A, \eta)$ is contained in a B_x centered within $B(A, 2\eta)$ and so has diameter $\delta(U) \leqslant 2\eta$.

With each (nonempty) $U \in \{U\}$, associate a point $a_U \in A$ as follows: choose an $x_U \in U$ and find $a_U \in A$ with $d(x_U, a_U) < 2d(x_U, A)$. The fundamental property of the sets $\{U\}$ and points $\{a_U\}$ is:

(*) For each $a \in A$ and nbd $W(a)$ in X, there is a nbd $V(a) \subset W(a)$ with the property: $U \cap V(a) \neq \varnothing \Rightarrow [U \subset W(a)] \wedge [a_U \in A \cap W(a)]$.

Indeed, we can assume $W(a) = B(a, \varepsilon)$; taking $V(a) = B(a, \varepsilon/12)$, any U intersecting $V(a)$ has diameter $\leqslant \varepsilon/6$, so that it is completely within $B(a, \varepsilon/4)$. For any such U, $d(x_U, a) < \varepsilon/4$, so that $d(x_U, A) < \varepsilon/4$ and also $d(a_U, a) \leqslant d(a_U, x_U) + d(x_U, a) \leqslant 2d(x_U, A_U) + d(x_U, A) < \frac{3}{4}\varepsilon$; that is, $a_U \in W(a)$.

Now let $\{\kappa_U\}$ be a partition of unity on $X - A$ subordinated to $\{U\}$ and define $F: X \to L$ by

$$F(x) = \begin{cases} f(x) & x \in A \\ \sum_U \kappa_U(x) \cdot f(a_U) & x \in X - A \end{cases}$$

As in VIII, **5.4**, F is continuous at each point of $X - A$, so only its continuity at each point of A need be proved. Let $a \in A$ and let W be a nbd of $F(a) = f(a)$. Since L is of type m, and f is continuous, there is a nbd $W(a)$ in X such that $f(W(a) \cap A) \subset C \subset W$ for some convex C

in L. Find $V(a) \subset W(a)$ satisfying the condition in (*); we will show $F(V(a)) \subset W$. Clearly, if $x \in A \cap V(a)$, then $F(x) \in C \subset W$. If $x \in V(a) - A$, then x belongs to at most finitely many $U_1, \cdots, U_n$, so that at x, only $\kappa_{U_1}, \cdots, \kappa_{U_n}$ are not zero; since each U_i intersects $V(a)$, the corresponding a_{U_i} all lie in $A \cap W(a)$, so that the $f(a_{U_i})$ are all elements of C; and since $F(x)$ is in the convex hull of the points $f(a_{U_1}), \cdots, f(a_{U_n})$, we find $F(x) \in C$ also. Thus $F(V(a)) \subset W$, and F is continuous at a. Since F is continuous at each point of X, the map $F: X \to L$ is continuous; the formula shows that F is an extension of f and that $F(X) \subset$ [convex hull of $f(A)$].

6.2 Corollary The theorem is valid if we replace L by any convex subset $K \subset L$.

Remark: Observe that the points $\{a_U\}$ and functions $\{\kappa_U\}$ are independent of the function f. It follows that we can assign to each $f: A \to L$ an extension $e(f): X \to L$ in such a way that $e(f + g) = e(f) + e(g)$ for all f, g (and, if L is normed, also $\|f\| = \|e(f)\|$).

7. Cartesian Products of Metric Spaces

In this section, we prove that a cartesian product is metrizable if and only if the number of nontrivial factors is at most countable and each is metrizable. First,

7.1 Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces, each of which has more than one point. If $\prod_\alpha Y_\alpha$ is metrizable, then each Y_α is metrizable and $\aleph(\mathscr{A}) \leqslant \aleph_0$.

Proof: Since each Y_α is homeomorphic to a slice in $\prod_\alpha Y_\alpha$, its metrizability follows from **5.1**. We establish $\aleph(\mathscr{A}) \leqslant \aleph_0$ by proving $[\aleph(\mathscr{A}) > \aleph_0] \Rightarrow [\prod_\alpha Y_\alpha$ is not 1° countable]. Indeed, let $\{y_\alpha^\circ\} \in \prod Y_\alpha$ be given, and let $\{V_n \mid n \in Z\}$ be any countable family of its nbds. With each V_n, associate a finite set $R(n) = \{\alpha_1, \cdots, \alpha_i\} \subset \mathscr{A}$ such that $\{y_\alpha^\circ\} \in \langle U_{\alpha_1}, \cdots, U_{\alpha_i}\rangle \subset V_n$. Then $\bigcup_n R(n)$ is at most countable and, since $\aleph(\mathscr{A}) > \aleph_0$, there is a $\beta \in \mathscr{A}$ not in $\bigcup_n R(n)$, that is, no V_n restricts the β-coordinate. Choosing a nbd $U_\beta \supset y_\beta^\circ$ with $U_\beta \neq Y_\beta$, which is possible because Y_β has more than one point, there is no V_n contained in the nbd $\langle U_\beta \rangle$ of $\{y_\alpha^\circ\}$.

The converse of **7.1** will follow from

7.2 Theorem Let $\{Y_n \mid n \in Z^+\}$ be a countably infinite family of metric spaces. In each Y_n, choose a metric d_n giving its topology, and let $\delta_n(Y_n)$ be the diameter of Y_n according to d_n. For $x = \{x_n\}, y = \{y_n\}$, define $\rho(x, y) = \sup\{d_n(x_n, y_n) \mid n \in Z^+\}$. Then:

(1). ρ is a metric in the *set* $\prod_n Y_n$ whenever the $\delta_n(Y_n)$ are uniformly bounded for all large n.

(2). ρ metrizes the cartesian product topology of the *space* $\prod_n Y_n$ if and only if $\delta_n(Y_n) \to 0$.

Proof: Ad (1). Since $\delta_n(Y_n) \leqslant M < \infty$ for some fixed constant M and all $n \geqslant$ some n_0, it follows that ρ is in fact defined on $\prod Y_n \times \prod Y_n$; it clearly has the properties (1) through (3) of **1.1**, and the triangle inequality results from

$$\begin{aligned}\rho(x, z) &= \sup_n d_n(x_n, z_n) \leqslant \sup_n[d_n(x_n, y_n) + d_n(y_n, z_n)]\\ &\leqslant \sup_n d_n(x_n, y_n) + \sup_n d_n(y_n, z_n)\\ &= \rho(x, y) + \rho(y, z).\end{aligned}$$

Ad (2). Assume $\delta_n(Y_n) \to 0$; we show that the basis for $\mathscr{T}(\rho)$ is equivalent to that of the cartesian product topology:

(a). Let $x \in \langle B(x_1, r_1), \cdots, B(x_n, r_n)\rangle = U$.
Choose $\mu = \min\{r_1, \cdots, r_n\}$; then $\mu > 0$ and we have $x \in B_\rho(x, \mu) \subset U$, since

$$\begin{aligned}y \in B_\rho(x, \mu) \Rightarrow \rho(x, y) < \mu &\Rightarrow \sup_n d_n(x_n, y_n) < \mu\\ &\Rightarrow \forall\, n\colon d_n(x_n, y_n) < r_n\\ &\Rightarrow \forall\, n\colon y_n \in B(x_n, r_n) \Rightarrow y \in U.\end{aligned}$$

(b). Let $x \in B_\rho(x, \eta)$. Since $\delta_n(Y_n) \to 0$, there is an n_0 with $\delta_n(Y_n) < \eta/2$ for all $n \geqslant n_0$; then

$$x \in U = \langle B(x_1, \eta/2), \cdots, B(x_{n_0}, \eta/2)\rangle \subset B_\rho(x, \eta).$$

For, if $y \in U$ then $d_i(x_i, y_i) < \eta/2$ for $1 \leqslant i \leqslant n_0$, and since this holds also for $i \geqslant n_0$, we have

$$y \in U \Rightarrow \rho(x, y) \leqslant \eta/2 \Rightarrow y \in B_\rho(x, \eta).$$

(a) and (b) show that the metric and cartesian product topologies are the same.

Now assume that $\delta_n(Y_n) \nrightarrow 0$; then we can find an $\varepsilon > 0$ and infinitely many indices i_n for which there are points $x_{i_n}^\circ, y_{i_n}^\circ \in Y_{i_n}$ that satisfy

$d_{i_n}(x^\circ_{i_n}, y^\circ_{i_n}) \geqslant \varepsilon$. If we select any $x^\circ \in \prod Y_n$ having *all* the $x^\circ_{i_n}$ among its coordinates, there can be no cartesian product nbd U satisfying $x^\circ \in U \subset B_\rho(x^\circ, \varepsilon)$: since U restricts only finitely many coordinates, at least one i_n is not restricted, so U contains the point z having coordinates $z_i = x^\circ_i$ $(i \neq i_n)$, $z_{i_n} = y^\circ_{i_n}$, and therefore $\rho(x^\circ, z) \geqslant \varepsilon$.

7.3 Corollary Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of metrizable spaces. If $\aleph(\mathscr{A}) \leqslant \aleph_0$, then $\prod_\alpha Y_\alpha$ is metrizable.

Proof: Case $\aleph(\mathscr{A}) = n < \aleph_0$: The proof of **7.2**(2) given above applies to show that if we choose any (not necessarily bounded) metric d_i for each Y_i, then $\rho(x, y) = \max\{d_i(x_i, y_i) \mid 1 \leqslant i \leqslant n\}$ metrizes the cartesian product topology of $\prod_1^n Y_i$.

Case $\aleph(\mathscr{A}) = \aleph_0$: By **3.3**, we can choose a metric d_i for each Y_i so that the requirement of **7.2**(2) is satisfied.

It should be noted that although in each case the metric ρ depends on the choices of the d_i, the topology $\mathscr{T}(\rho)$ does not.

Remark. We have seen that the cartesian product of paracompact spaces need not be normal. In view of **7.3** it is natural to ask about the nature of a cartesian product of a paracompact and a metric space. This question has recently been settled by E. A. Michael, who has shown that if X is paracompact and Y is metric, then again $X \times Y$ need not be normal, even though Y is separable metric. His example is the following: Let Y be the metric space of all irrationals, and let X be the paracompact space of VII, **3**, Problem 7. Let R be the subset of rationals in X, and T the irrationals. Then the closed sets $A = R \times Y$ and $B = \{(z, z) \mid z \in T\}$ in $X \times Y$ cannot be separated. The reader may construct a proof by making suitable modifications of that given in VII, **3**, problem 3.

8. The Space $l^2(\mathscr{A})$; Hilbert Cube

8.1 Definition Let $\mathscr{A}$ be any set, of arbitrary cardinal, and let $E^{\aleph(\mathscr{A})}$ be the cartesian product of $\aleph(\mathscr{A})$ factors E^1. The metric space $l^2(\mathscr{A})$ is that having for:

Elements: every $x = \{x_\alpha\} \in E^{\aleph(\mathscr{A})}$ such that $x_\alpha = 0$ for all but at most countably many $\alpha \in \mathscr{A}$, and $\sum x_\alpha^2$ converges.

Topology: that induced by the metric $d(x, y) = \sqrt{\sum_\alpha (x_\alpha - y_\alpha)^2}$.

It must be verified that $d(x, y)$ is indeed a metric. Its finiteness and the triangle inequality follow from Minkowski's inequality: Given $x, y \in l^2(\mathscr{A})$, then summing over the at most countably many elements of

$\mathscr{A}$ for which at least one of x, y has a nonzero coordinate, we find for each finite N that

$$\sqrt{\sum_1^N (x_{\alpha_i} - y_{\alpha_i})^2} \leqslant \sqrt{\sum_1^N x_{\alpha_i}^2} + \sqrt{\sum_1^N y_{\alpha_i}^2} \leqslant \sqrt{\sum_\alpha x_\alpha^2} + \sqrt{\sum_\alpha y_\alpha^2} < \infty.$$

The partial sums being bounded, this monotone nondecreasing sequence must therefore converge, and in fact

$$\sqrt{\sum_\alpha (x_\alpha - y_\alpha)^2} \leqslant \sqrt{\sum_\alpha x_\alpha^2} + \sqrt{\sum_\alpha y_\alpha^2},$$

from which both conclusions follow. We remark that $l^2(\mathscr{A})$ is an example of a class of linear spaces called Hilbert spaces.

It is trivial to verify that $l^2(\mathscr{A}) \cong l^2(\mathscr{B})$ if and only if $\aleph(\mathscr{A}) = \aleph(\mathscr{B})$; whenever $\aleph(\mathscr{A}) = n < \aleph_0$, it is clear that $l^2(\mathscr{A}) \cong E^n$.

Ex. 1 For any $\aleph$ let $\varphi\colon l^2(\aleph) \to E^\aleph$ be the identity map. Then (a) φ is always continuous, (b) φ is surjective if and only if $\aleph < \aleph_0$, and (c) φ is a homeomorphism of $l^2(\aleph)$ and $\varphi[l^2(\aleph)]$ if and only if $\aleph < \aleph_0$: indeed, if $\aleph \geqslant \aleph_0$, there can be no nbd of $0 = \{0, \cdots, 0, \cdots\}$ in $E^\aleph \cap \varphi(l^2)$ mapped by φ^{-1} into the nbd $B(0, \frac{1}{2})$, since any cartesian product nbd of 0 contains a point of $\varphi(l^2)$ having all coordinates zero except one, which is 1. We remark that there can be no homeomorphism of $l^2(\aleph)$ and $E^\aleph$ whenever $\aleph > \aleph_0$ because $E^\aleph$ is not then metrizable; however, R. D. Anderson has shown recently that $E^{\aleph_0}$ is in fact homeomorphic to $l^2(\aleph_0)$.

Ex. 2 Let X be any space and for each $\alpha \in \aleph$ let $f_\alpha\colon X \to E^1$ be a continuous map. If $\aleph \geqslant \aleph_0$ then, unlike the situation with cartesian products, (a) the correspondence $x \to \{f_\alpha(x)\}$ need not determine a map into $l^2(\aleph)$, and (b) even though $x \to \{f_\alpha(x)\}$ is a map into $l^2(\aleph)$, it need not be continuous: for each $n \in Z^+$ let $f_n\colon E^1 \to E^1$ be a continuous function with value 1 at $x = 2/n$ and zero outside $[1/n, 3/n]$; then $x \to \{f_n(x)\}$ defines a map into $l^2(\aleph_0)$, but it is not continuous at $x = 0$.

8.2 $l^2(\mathscr{A})$ is separable (and so 2° countable) if and only if $\aleph(\mathscr{A}) \leqslant \aleph_0$.

Proof: It is separable if $\aleph(\mathscr{A}) \leqslant \aleph_0$: The set of all points in $l^2(\mathscr{A})$ having all but at most finitely many coordinates rational and the rest 0 is easily verified to be dense in $l^2(\mathscr{A})$. Now assume that $\aleph(\mathscr{A}) > \aleph_0$: Then, in any countable collection of elements, we find nonzero entries for at most countably many coordinates; if β is an index for which all members of the countable set have coordinate 0, the point x with $x_\beta = 1$ and $x_\alpha = 0$ $(\alpha \neq \beta)$ has nbd $B(x, \frac{1}{2})$ not containing any member of the countable set.

8.3 Definition The subspace $\{\{x_n\} \in l^2(\aleph_0) \mid |x_n| \leqslant 1/n\}$ is called the Hilbert cube and is denoted by I^∞.

8.4 The Hilbert cube I^∞ is homeomorphic to the cartesian product $\prod_1^\infty I$ of countably many unit intervals.

Proof: Since $I \cong J = [-1, +1]$, we have $\prod_1^\infty I \cong \prod_1^\infty J$, so we prove $I^\infty \cong \prod_1^\infty J$. Define $\varphi: I^\infty \to \prod_1^\infty J$ by $\varphi(x_1, x_2, \cdots) = (x_1, 2x_2, 3x_3 \cdots)$; clearly, this is bijective; and φ is continuous, since the projection on each factor is continuous. To see that φ^{-1} is continuous, metrize $\prod_1^\infty J$ as in **7.2**, using

$$d_n(x_n, y_n) = \frac{|x_n - y_n|}{n^{1/4}};$$

if $\rho(x, y) < \delta$ in $\prod_1^\infty J$, we then have

$$d(\varphi^{-1}(x), \varphi^{-1}(y)) = \sqrt{\sum_1^\infty \left(\frac{x_n}{n} - \frac{y_n}{n}\right)^2}$$

$$= \sqrt{\sum_1^\infty \left|\frac{x_n - y_n}{n^{1/4}}\right|^2 \cdot \frac{1}{n^{3/2}}} < \delta \cdot \sqrt{\sum_1^\infty \frac{1}{n^{3/2}}}$$

and, since the series converges, the continuity of φ^{-1} follows.

8.5 I^∞ has no interior in $l^2(\aleph_0)$, so that the complement of I^∞ is dense.

Proof: Let $x_0 = (x_1, x_2, \cdots) \in I^\infty$, and let $B(x_0, 2\varepsilon)$ be any nbd in $l^2(\aleph_0)$. All the points $\xi_n = (x_1, \cdots, x_n + \varepsilon, x_{n+1}, \cdots)$ lie in $B(x_0, 2\varepsilon)$, but not all can belong to I^∞, since this requires $|x_n + \varepsilon| \leqslant 1/n$ for each n, which is clearly impossible unless $\varepsilon = 0$. Thus, no set open in $l^2(\aleph_0)$ lies in I^∞.

9. Metrization of Topological Spaces

Let Y be a metric space, with metric d; being paracompact, Y is regular. Furthermore, for each $n = 1, 2, \cdots$, let $\{U_{n,\alpha}\}$ be a nbd-finite open refinement of the open covering $\{B_d(y, 1/n) \mid y \in Y\}$; then the family of all the open sets $\{U_{n,\alpha}\}$ is (1) nbd-finite for each n, and more importantly, (2) constitutes a basis for Y: letting U be any open set and $y_0 \in U$, choose n_0 so large that $1/n_0 < \frac{1}{2}d(y_0, Y - U)$; then any $U_{n_0,\alpha}$ containing y_0 satisfies $y_0 \in U_{n_0,\alpha} \subset U$, since it lies in some $1/n_0$ ball containing y_0 and any such ball must lie completely in U. These remarks constitute the "necessary" part of

9.1 Theorem (J. Nagata and Yu. M. Smirnov) A topological space Y is metrizable if and only if it is regular and has a *basis* that can be decomposed into an at most countable collection of nbd-finite families.

Proof: The precise condition is that there be a basis $\mathfrak{U}$ such that $\mathfrak{U} = \bigcup_n \mathfrak{U}_n$, where for each n, $\mathfrak{U}_n = \{U_{n,\alpha} \mid \alpha \in \mathscr{A}_n\}$ is a nbd-finite family of open sets. Such a space is evidently paracompact: any open covering can be refined by a covering consisting of basic open sets only, and the assumed regularity of Y allows application of VIII, **2.3**(2).

We next show that Y is also perfectly normal. In fact, if U is any open set, then for each $y \in U$, regularity gives a basic $U_{n(y),\alpha(y)}$ such that $y \in U_{n(y),\alpha(y)} \subset \overline{U}_{n(y),\alpha(y)} \subset U$. For each k, define

$$F_k = \cup \{\overline{U}_{n(y),\alpha(y)} \mid n(y) = k\}.$$

F_k is closed in Y, since $\{U_{k,\alpha}\}$ is a nbd-finite family; and because $U = \bigcup_k F_k$, this shows that U is an F_σ-set as required.

Since Y is perfectly normal, for each pair (n, α) there is a continuous $\varphi_{n,\alpha}: Y \to [0, 1]$ with $\varphi_{n,\alpha}^{-1}(0) = Y - U_{n,\alpha}$. Because for each n_0 the family $\{U_{n_0,\alpha}\}$ is nbd-finite, each point of Y has a nbd meeting at most finitely many $U_{n_0,\alpha}$, so the sum $\sum_\alpha \varphi_{n_0,\alpha}(y)$ is finite and, continuity being a local matter, is a continuous function on Y. We are now going to show that the map f of Y into the Hilbert space $l^2(\mathfrak{U})$, given by the coordinate functions

$$f_{n,\alpha}(y) = \frac{1}{n} \cdot \frac{\varphi_{n,\alpha}(y)}{\sqrt{1 + \sum_\beta \varphi_{n,\beta}^2(y)}},$$

is a homeomorphism onto a subspace. The verification is routine:

(1). For each y, $\{f_{n,\alpha}(y)\}$ is indeed a point of $l^2(\mathfrak{U})$: Since for each n_0 we have $\varphi_{n_0,\alpha}(y) = 0$ for all but at most finitely many α, it follows that all but at most countably many $f_{n,\alpha}(y)$ are zero, and then that

$$\sum_{n,\alpha} f_{n,\alpha}^2(y) = \sum_n \sum_\alpha f_{n,\alpha}^2(y) \leqslant \sum_n \frac{1}{n^2} < \infty.$$

(2). Each $f_{n,\alpha}$ is continuous, since the denominator never vanishes.

(3). f is injective: If $x \neq y$, there exists some $U_{n,\alpha}$ with $x \in U_{n,\alpha}$ and $y \,\overline{\in}\, U_{n,\alpha}$; thus $\varphi_{n,\alpha}(x) > 0$, $\varphi_{n,\alpha}(y) = 0$, and so $f(x) \neq f(y)$.

(4). f is a closed map of Y onto $f(Y)$: For, if $A \subset Y$ is closed and $y \,\overline{\in}\, A$, then for some (n_0, α_0), we have $y \in U_{n_0,\alpha_0} \subset Y - A$; because $\varphi_{\alpha_0,n_0}(a) = 0$ for each $a \in A$, whereas $\varphi_{n_0,\alpha_0}(y) = k > 0$, we find $d(f(y), f(A)) \geqslant k \cdot 1/n_0$, so that $f(y) \,\overline{\in}\, \overline{f(A)}$.

(5). f is continuous. Let $y_0 \in Y$ and $\varepsilon > 0$ be given; choose n_0 so that

$$\sum_{n > n_0} \frac{1}{n^2} < \frac{\varepsilon^2}{4}$$

and let $W(y_0)$ be a nbd that intersects at most finitely many sets $U_{n,\alpha}$ for $n \leqslant n_0$. It follows that among all the $f_{n,\alpha}$ with first index $n \leqslant n_0$, there are only finitely many, N, that do not vanish identically on $W(y_0)$. Choose a nbd $V(y_0) \subset W(y_0)$ on which each of these N continuous functions satisfy

$$|f_{n,\alpha}(y) - f_{n,\alpha}(y_0)| \leqslant \frac{\varepsilon}{\sqrt{2N}};$$

then, for any $y \in V(y_0)$, we find

$$\sum_{\substack{n,\alpha \\ n \leq n_0}} |f_{n,\alpha}(y) - f_{n,\alpha}(y_0)|^2 \leqslant \frac{\varepsilon^2}{2}$$

and

$$\sum_{\substack{n,\alpha \\ n > n_0}} |f_{n,\alpha}(y) - f_{n,\alpha}(y_0)|^2 \leqslant 2 \sum_{n > n_0} \frac{1}{n^2} < \frac{\varepsilon^2}{2},$$

which shows that $f(V(y_0)) \subset B(f(y_0), \varepsilon)$, as required.

Therefore f is a homeomorphism; since Y is embedded as a subspace of the metric space $l^2(\mathfrak{U})$, Y is metrizable and the theorem has been proved.

This criterion for the metrizability of a regular space should be clearly distinguished from that in VIII, **2.3** (2) for paracompactness: in **9.1** we require that the covering by a *basis* be decomposable in a certain way.

We have seen (VIII, **6.5**) that in Lindelöf spaces, regularity is equivalent to paracompactness; a stronger result is true for 2° countable spaces:

9.2 Corollary (P. Urysohn) In 2° countable spaces, regularity is equivalent to *metrizability*. In fact, each 2° countable regular space is homeomorphic to a subset of the Hilbert cube I^∞.

Proof: Since a 2° countable space has a countable basis $\mathscr{B} = \{U_n \mid n \in Z\}$, the decomposition $\mathscr{B} = \bigcup_n \mathscr{B}_n$, where each $\mathscr{B}_n$ is a single set U_n, satisfies the requirements of **9.1** and gives metrizability. Observe that the embedding of **9.1** is into $l^2(\mathscr{B})$, where $\aleph(\mathscr{B})$ = cardinal of the basis; for 2° countable spaces, the embedding is therefore into $l^2(\aleph_0)$. The formula for the coordinate functions f_n shows $|f_n(y)| \leqslant 1/n$ for each n, so the image lies in I^∞.

Remarks. The proof of **9.1** gives somewhat more information in case each $\{U_{n,\alpha} \mid \alpha \in \mathscr{A}_n\}$ is an open *covering* of Y:

9.3 Let Y be regular and let $\mathfrak{U}_n = \{U_{n,\alpha} \mid \alpha \in \mathscr{A}_n\}$ be a sequence of nbd-finite open coverings such that $\mathfrak{U} = \bigcup_1^\infty \mathfrak{U}_n$ is a basis. Then Y is metrizable, and a metric ρ can be selected so that $B_\rho(y, 1/n) \subset \mathrm{St}(y, \mathfrak{U}_n)$ for each n and $y \in Y$.

Proof: We can now use the coordinate functions

$$f_{n,\alpha} = n^{-1}\varphi_{n,\alpha}/\sqrt{\sum_\beta \varphi_{n,\beta}^2}$$

to get an embedding in $l^2(\mathfrak{U})$, since the "1" in the denominator is no longer necessary to assure that the denominator never vanishes. Let ρ be the metric obtained by this embedding, and let y, n, be given. Note that if $U_{n,\alpha_1}, \cdots, U_{n,\alpha_k}$ are all the sets in $\mathrm{St}(y, \mathfrak{U}_n)$, then $f_{n,\alpha}(y) \neq 0$ if and only if $\alpha \in \{\alpha_1, \cdots, \alpha_k\}$; consequently $f_{n\,\alpha_1}^2(y) + \cdots + f_{n,\alpha_k}^2(y) = 1/n^2$. Now, if $x \bar{\in} \mathrm{St}(y, \mathfrak{U}_n)$, then $f_{n,\alpha}(x) = 0$ for $\alpha \in \{\alpha_1, \cdots, \alpha_k\}$, so

$$[\rho(x, y)]^2 = \sum_{n,\alpha} | f_{n,\alpha}(x) - f_{n,\alpha}(y) |^2 \geqslant f_{n,\alpha_1}^2(y) + \cdots + f_{n,\alpha_k}^2(y) = 1/n^2$$

and therefore $\rho(x, y) \geqslant 1/n$. The assertion is proved.

This observation leads to

9.4 Let (Y, d) be a metric space, and $\{\mathfrak{V}_n \mid n \in Z^+\}$ any sequence of open coverings. Then there exists an equivalent metric ρ such that for each n, $\{B_\rho(y, 1/n) \mid y \in Y)$ refines $\mathfrak{V}_n$.

Proof: For each $n \in Z^+$ let $\mathfrak{W}_n$ be a common refinement of the coverings $\mathfrak{V}_n$ and $\{B_d(y, 1/n \mid y \in Y\}$, and let $\mathfrak{U}_n$ be a nbd-finite open refinement of a barycentric open refinement of $\mathfrak{W}_n$. It is evident that $\{\mathfrak{U}_n \mid n \in Z^+\}$ is a basis, so we obtain an equivalent metric ρ by embedding as in **9.3**. Then for any $y \in Y$, we have $B_\rho(y, 1/n,) \subset \mathrm{St}(y, \mathfrak{U}_n) \subset$ some $V \in \mathfrak{V}_n$.

The Nagata-Smirnov theorem characterizes those regular spaces that are metrizable, by imposing a condition on the covering by a basis; by using other types of coverings, one can in fact characterize the T_1 (and even the T_0) spaces that are metrizable.

9.5 Theorem Let Y be an arbitrary space. The following statements are equivalent:

(1). Y is metrizable.

(2). (**K. Morita**) Y is a T_0 space, and there exists a sequence $\{\mathfrak{F}_n \mid n \in Z^+\}$ of nbd-finite closed coverings with the property: for each $y \in Y$ and nbd $W(y)$ there is an n such that $\mathrm{St}(y, \mathfrak{F}_n) \subset W$.

(3). (**A. H. Stone**) Y is a T_0 space and there exists a sequence $\{\mathfrak{U}_n \mid n \in Z^+\}$ of open coverings with the property: for each $y \in Y$ and nbd $W(y)$ there is a nbd $V(y)$ and an n such that $\mathrm{St}(V, \mathfrak{U}_n) \subset W$.

(4). (**A. Arhangel'skii**) Y is a T_1 space and there exists one sequence $\{\mathfrak{U}_n \mid n \in Z^+\}$ of open coverings that is locally starring for every open covering.

Proof: (1) $\Rightarrow$ (2) is trivial, since metric spaces are paracompact.

(2) $\Rightarrow$ (3). It is no restriction to assume that $\mathfrak{F}_{n+1} \prec \mathfrak{F}_n$ for each n. We are going to show that

$$\mathfrak{U}_n = \{\text{Int}[\text{St}(y, \mathfrak{F}_n)] \mid y \in Y\}, \quad n \in Z^+,$$

is a sequence of open coverings having the property required in (3).

First, each $\mathfrak{U}_n$ is an open covering. Indeed, for each $y \in Y$, let

$$V_n(y) = Y - \bigcup \{F \in \mathfrak{F}_n \mid y \bar{\in} F\};$$

each $V_n(y)$ is an open set, and $y \in V_n(y) \subset \text{St}(y, \mathfrak{F}_n)$; since $\{V_n(y) \mid y \in Y\}$ is evidently a covering, therefore so also is $\mathfrak{U}_n$.

We note that $F \cap V_n(y) \neq \varnothing$ if and only if $y \in F$; it follows from this that: (a) if $x \in V_n(y)$, then $\text{St}(x, \mathfrak{F}_n) \subset \text{St}(y, \mathfrak{F}_n)$; and (b) if $\text{St}(z, \mathfrak{F}_n) \cap V_n(y) \neq \varnothing$ for some $z \in Y$, then $z \in \text{St}(y, \mathfrak{F}_n)$.

Now, to see that $\{\mathfrak{U}_n \mid n \in Z^+\}$ has the property in (3), let $W(y)$ be given; determine k and $n > k$ so that

$$W(y) \supset \text{St}(y, \mathfrak{F}_k) \supset V_k(y) \supset \text{St}(y, \mathfrak{F}_n) \supset V_n(y);$$

then $\text{St}(V_n(y), \mathfrak{U}_n) \subset W(y)$: for, whenever $\text{St}(z, \mathfrak{F}_n) \cap V_n(y) \neq \varnothing$, we find from (b) that $z \in \text{St}(y, \mathfrak{F}_n) \subset V_k(y)$, then from (a) that $\text{St}(z, \mathfrak{F}_k) \subset \text{St}(y, \mathfrak{F}_k) \subset W(y)$, and finally from $\mathfrak{F}_n \prec \mathfrak{F}_k$ that $\text{St}(z, \mathfrak{F}_n) \subset \text{St}(z, \mathfrak{F}_k) \subset W(y)$; since $\text{Int St}(z, \mathfrak{F}_n) \subset \text{St}(z, \mathfrak{F}_n)$, the desired conclusion follows, and the proof is complete.

(3) $\Rightarrow$ (4). We need only show that Y is T_1. Given two points x, y, assume that x is that one having a nbd $W(x)$ not containing y. There is by hypothesis a nbd $V(x)$ and an n such that $\text{St}(V, \mathfrak{U}_n) \subset W$. Since $\mathfrak{U}_n$ is an open covering, there is some $U \in \mathfrak{U}_n$ containing y, and clearly $U \cap V = \varnothing$; thus, Y is T_1 (even Hausdorff).

(4) $\Rightarrow$ (1). Y is paracompact, by VIII, **3.7**. For each n, let $\mathfrak{B}_n$ be a nbd-finite refinement of $\mathfrak{U}_n$; we will show that $\{\mathfrak{B}_n \mid n \in Z^+\}$ is a basis. Let W be open, and $y \in W$; find an open G such that $y \in G \subset \overline{G} \subset W$ and form the open covering $\mathfrak{U} = \{W, Y - \overline{G}\}$. Since $\{\mathfrak{U}_n \mid n \in Z^+\}$ is locally starring for $\mathfrak{U}$, and since y belongs only to W, there is a nbd $V(y)$ and an n such that $\text{St}(V, \mathfrak{U}_n) \subset W$. Since $\mathfrak{B}_n \prec \mathfrak{U}_n$, we therefore have

$$y \in \text{St}(y, \mathfrak{B}_n) \subset \text{St}(y, \mathfrak{U}_n) \subset \text{St}(V, \mathfrak{U}_n) \subset W.$$

Thus, $\{\mathfrak{B}_n \mid n \in Z^+\}$ is a basis, and by the Nagata-Smirnov theorem, Y is therefore metrizable.

The reader should contrast **9.5** (4) with VIII, **3.7**, to emphasize the distinction between metrizability and paracompactness.

10. Gauge Spaces

In this section, we consider spaces in which the topology is induced by a pseudometric or, more generally, by a family of pseudometrics. It will be seen that this property characterizes the completely regular spaces.

The technique of using a family of pseudometrics to express the topology of a completely regular space permits the extension of many purely metric notions (such as completeness) to this wider class of spaces.

In order to make the terminology and notation precise, we repeat here the definition of pseudometric that has been given in Section **1**:

10.1 Definition Let Y be any set. A map $d: Y \times Y \to E^1$ is called a pseudometric (or, an écart; or, a gauge) in Y whenever

(1). $d(x, y) \geqslant 0$ for all x, y.
(2). If $x = y$, then $d(x, y) = 0$.
(3). $d(x, y) = d(y, x)$ for all x, y.
(4). $d(x, z) \leqslant d(x, y) + d(y, z)$ for every triple of points.

The d-ball of radius ε centered at y is the set

$$B(y; d, \varepsilon) = \{x \mid d(x, y) < \varepsilon\}.$$

We prefer to use the term "gauge" rather than the more usual "écart" and "pseudometric," and will do so hereafter.

Ex. 1 Every metric is a gauge, but not conversely: $d(x, y) = 0$ need not imply that $x = y$.

Ex. 2 For any set Y and any $f: Y \to E^1$, the map $d_f: Y \times Y \to E^1$ defined by $d_f(x, y) = |f(x) - f(y)|$ is a gauge in Y, called the gauge derived from f. Not all gauges are derived from maps: the Euclidean metric on I^2 is not derivable from any $f: I^2 \to E^1$, since such a map would then be a continuous injection, which (*cf.* XI, **2**, Ex. 4) is impossible.

Ex. 3 A gauge d_β on any one factor Y_β of a cartesian product $\prod Y_\beta$ induces a gauge $\hat{d}$ on $\prod Y_\alpha$ by $\hat{d}(x, y) = d_\beta[p_\beta(x), p_\beta(y)]$.

Ex. 4 If $\{d_\alpha \mid \alpha \in \mathscr{A}\}$ is any uniformly bounded family of gauges on Y, it is simple to verify that $d(x, y) = \sup \{d_\alpha(x, y) \mid \alpha \in \mathscr{A}\}$ is also a gauge on Y. Furthermore, for any gauge d and any constant $M > 0$, the function $\min(M, d)$ is also a gauge.

A family $\mathscr{D} = \{d_\alpha \mid \alpha \in \mathscr{A}\}$ of gauges on Y is called *separating* if for each pair of points $x \neq y$ there is a $d_\alpha \in \mathscr{D}$ such that $d_\alpha(x, y) \neq 0$.

10.2 Definition Let Y be a set and $\mathscr{D} = \{d_\alpha \mid \alpha \in \mathscr{A}\}$ a separating family of gauges on Y. The topology $\mathscr{T}(\mathscr{D})$ having for a subbasis the family $\mathfrak{B}(\mathscr{D}) = \{B(y; d_\alpha, \varepsilon) \mid y \in Y, d_\alpha \in \mathscr{D}, \varepsilon > 0\}$ of balls is called the topology in Y induced by the family $\mathscr{D}$.

Because we require $\mathscr{D}$ to be separating, it follows that (a): the topology $\mathscr{T}(\mathscr{D})$ is always Hausdorff, and (b): if $\mathscr{D}$ consists of one gauge alone, then that gauge must be a metric and $\mathscr{T}(\mathscr{D})$ is the topology induced by that metric. As metric spaces show, distinct families of gauges in Y may lead to the same topology.

10.3 Definition A gauge structure for a topological space $(Y, \mathscr{T})$ is a separating family $\mathscr{D}$ of gauges such that $\mathscr{T} = \mathscr{T}(\mathscr{D})$. A topological space that admits a gauge structure is called a *gauge space.*

If Y has the gauge structure $\mathscr{D}$, a basis for its topology can be obtained by enlarging $\mathscr{D}$:

10.4 Let $\mathscr{D} = \{d_\alpha \mid \alpha \in \mathscr{A}\}$ be a separating family of gauges in Y. Let $\mathscr{D}^+$ be the family of gauges

$$\{\max(d_{\alpha_1}, \cdots, d_{\alpha_n}) \mid \text{ all finite subsets } \{\alpha_1, \cdots, \alpha_n\} \subset \mathscr{A}\}.$$

Then the family $\mathfrak{B}(\mathscr{D}^+)$ of all balls is a *basis* for $\mathscr{T}(\mathscr{D})$.

Proof: It is simple to verify that if $d^+ = \max(d_{\alpha_1}, \cdots, d_{\alpha_n})$, then $B(y; d^+, \varepsilon) = \bigcap_1^n B(y; d_{\alpha_i}, \varepsilon)$ for each $\varepsilon > 0$; thus, each member of $\mathfrak{B}(\mathscr{D}^+)$ is open in $\mathscr{T}(\mathscr{D})$. The proof that $\mathfrak{B}(\mathscr{D}^+)$ is a basis for $\mathscr{T}(\mathscr{D})$ is now entirely similar to that given in **2.1**.

Ex. 5 In a gauge space $(Y, \mathscr{T}(\mathscr{D}))$, the expression of any given topological concept in terms of the gauges uses the basis $\mathfrak{B}(\mathscr{D}^+)$ [rather than the subbasis $\mathfrak{B}(\mathscr{D})$] and is similar to that given in terms of the metric in metric spaces: calling the ball $B(y; d^+, \epsilon)$ a (d^+, ϵ)-nbd of y, the idea of an ϵ-nbd in metric spaces is replaced by that of a (d^+, ϵ)-nbd in gauge spaces. Thus, $y \in \bar{A}$ if and only if each (d^+, ϵ)-nbd of y meets A. Similarly, an $f: Y \to Z$ is continuous if for each $y \in Y$ and open $W \supset f(y)$ there is a $d^+ \in \mathscr{D}^+$ and an $\epsilon > 0$ such that $f[B(y; d^+, \epsilon)] \subset W$.

Ex. 6 The formula $| d(x, y) - d(x', y') | \leq d(x, x') + d(y, y')$ is valid for any gauge; it shows that in any gauge space $(Y, \mathscr{T}(\mathscr{D}))$ each $d^+ \in \mathscr{D}^+$ is a continuous map of $Y \times Y$ into E^1.

In addition to showing that subspaces and cartesian products of gauge spaces are gauge spaces, the following theorem also gives the construction of a gauge structure for such spaces starting from one in each factor:

10.5 Theorem (1). Let the space Y have the gauge structure $\mathscr{D}$ and let A be a subspace of Y. Let $\mathscr{D}_A$ be the family of gauges in $\mathscr{D}$, each restricted to $A \times A$. Then $\mathscr{D}_A$ is a gauge structure for the subspace A.

(2). Let $\{(Y_\beta, \mathscr{T}(\mathscr{D}_\beta)) \mid \beta \in \mathscr{B}\}$ be any family of gauge spaces. For each $\beta \in \mathscr{B}$, let $\hat{\mathscr{D}}_\beta$ be the family of gauges induced on $\prod Y_\beta$ by the members of $\mathscr{D}_\beta$. Then the family $\{\hat{\mathscr{D}}_\beta \mid \beta \in \mathscr{B}\}$ of gauges is a gauge structure for the cartesian product topology of the gauge spaces.

Proof: (1) is left for the reader. (2). We need only observe that the sets $\langle B(y; \hat{d}_\alpha, \varepsilon)\rangle \subset \prod Y_\beta$ are a subbasis for both topologies.

The position of gauge space topologies is

10.6 Theorem A space Y is a gauge space if and only if it is completely regular.

Proof: Assume that Y has the topology $\mathscr{T}(\mathscr{D})$ for some separating family $\mathscr{D}$ of gauges. Let $y_0 \in U$, where U is an open set. By **10.4**, there is a $d^+ \in \mathscr{D}^+$ such that $B(y_0; d^+, \varepsilon) \subset U$; then $f(y) = \min[1, \varepsilon^{-1}d^+(y, y_0)]$ is a continuous map of Y into E^1 such that $f(y_0) = 0, f(\mathscr{C}U) = 1$. Thus, Y is completely regular.

Conversely, assume that Y is completely regular. Then Y can be embedded (VII, **7.3**) in a suitable cartesian product P^Y of unit intervals I. Since I is a gauge space, so also (**10.5**) is P^Y and therefore also Y.

10.7 Corollary A completely regular space is metrizable if and only if it admits a countable gauge structure.

Proof: Only the sufficiency requires proof. Letting $\mathscr{D} = \{d_n \mid n \in Z^+\}$ be a countable gauge structure for the space, the reader can verify that

$$d(x, y) = \sum_1^\infty \min[d_n(x, y), 2^{-n}]$$

is a metric that induces the topology $\mathscr{T}(\mathscr{D})$.

11. Uniform Spaces

A metric d in a space Y can be regarded as providing a measure of nearness that is applicable throughout the space: for each $\varepsilon > 0$, we may consider all the sets $B(y; \varepsilon)$ as being equally small. This notion of uniform smallness is not a topological concept: equivalent metrics specify different sets as being equally small. In this section, we consider the idea of uniformity in general topology due to A. Weil; it has many important applications, particularly in the study of topological groups.

The notation of binary relations in a set Y will be used. Given $U, V \subset Y \times Y$, we define $U^{-1} = \{(y, x) \mid (x, y) \in U\}$ and $U \circ V = \{(x, z) \mid \exists y: [(x, y) \in V] \wedge [(y, z) \in U]\}$. Furthermore, for each $y \in Y$, we let $U[y] = \{z \mid (y, z) \in U\}$; and $\Delta \subset Y \times Y$ will denote the diagonal.

11.1 Definition A uniform structure in a set Y is a family $\mathfrak{F}$ of subsets of $Y \times Y$ such that

(1). If $V \in \mathfrak{F}$, then $\Delta \subset V$.
(2). If $V_1, V_2 \in \mathfrak{F}$, then there is a $W \in \mathfrak{F}$ such that $W \subset V_1 \cap V_2$.
(3). If $V \in \mathfrak{F}$, then there is a $W \in \mathfrak{F}$ such that $W \circ W^{-1} \subset V$.
(4). If $V \in \mathfrak{F}$ and $V \subset W$, then $W \in \mathfrak{F}$.

The uniform structure is called separating if $\bigcap \{V \mid V \in \mathfrak{F}\} = \Delta$.

A family satisfying only (1), (2), and (3) is called a base for a uniform structure or, more simply, a uniformity. Since $\{A \in \mathscr{P}(Y \times Y) \mid \Delta \subset A\}$ is a uniform structure, and since the intersection of uniform structures is also a uniform structure, each uniformity generates a unique smallest uniform structure containing it. Two uniformities $\mathfrak{F}_1$, $\mathfrak{F}_2$, are equivalent if they generate the same uniform structure. It is simple to see that $\mathfrak{F}_1$ is equivalent to $\mathfrak{F}_2$ if and only if for each $V_1 \in \mathfrak{F}_1$ there is a $V_2 \in \mathfrak{F}_2$ such that $V_2 \subset V_1$, and also for each $W_2 \in \mathfrak{F}_2$ there is a $W_1 \in \mathfrak{F}_1$ such that $W_1 \subset W_2$.

For an interpretation of a uniformity $\mathfrak{F}$, call two points V-close whenever $(x, y) \in V$ or, equivalently, $y \in V[x]$; in this terminology, (3) says that for each $V \in \mathfrak{F}$, there is a $W \in \mathfrak{F}$ such that whenever x is W-close to y and y is W-close to z, then x is V-close to z. Thus, each member of the family $\mathfrak{F}$ can be regarded as a relation of uniform nearness defined over the entire set Y.

Ex. 1 Let Y be a metric space, and d a metric for Y. For each $\epsilon > 0$, let $V(\epsilon) = \{(x, y) \in Y \times Y \mid d(x, y) < \epsilon\}$; then the family $\{V(\epsilon) \mid \epsilon > 0\}$ is a uniformity in Y, called the uniformity determined by d. Observe that equivalent metrics in a space Y need not determine equivalent uniformities: for example, in the space $\{x \mid x \geqslant 1\} \subset E^1$, the uniformities determined by the equivalent metrics $d(x, y) = |x - y|$ and $d'(x, y) = |x^{-1} - y^{-1}|$ are not equivalent.

Ex. 2 If $(Y, \mathscr{T}(\mathscr{D}))$ is any gauge space, then the family of sets

$$\{(x, y) \subset Y \times Y \mid d(x, y) < \epsilon\}$$

for all $d \in \mathscr{D}$ and $\epsilon > 0$ is a uniformity in Y, called the uniformity determined by $\mathscr{D}$.

Ex. 3 The idea of uniform continuity has its natural expression in terms of uniformities: if X, Y are sets with uniformities $\mathfrak{F}$, $\mathfrak{G}$, respectively, a map $f: X \to Y$ is called uniform with respect to the given uniformities whenever for each $V \in \mathfrak{G}$ there is a $U \in \mathfrak{F}$ such that if $(x, y) \in U$, then $(f(x), f(y)) \in V$. This leads to the idea of a uniformly continuous map of a gauge space $(X, \mathscr{T}(\mathscr{D}))$ into another, $(Y, \mathscr{T}(\mathscr{D}'))$: it is a continuous map of X into Y that is uniform with respect to the uniformities determined by $\mathscr{D}$, $\mathscr{D}'$. Stated directly in terms of the gauges: $f: X \to Y$ is uniformly continuous if for each $d' \in \mathscr{D}'$ and $\epsilon > 0$, there exists a $d \in \mathscr{D}$ and a $\delta > 0$ such that if $d(x, y) < \delta$, then $d'(f(x), f(y)) < \epsilon$. Notice that even in metric spaces, a continuous map may be uniformly continuous if one pair of metrics is used, but not uniformly continuous when another pair of equivalent metrics is used; uniform continuity is therefore not a topological concept.

The concept of a uniformity can be expressed equivalently by coverings of the set Y:

11.2 Definition A family $\mathbf{G} = \{\mathfrak{U}_\alpha \mid \alpha \in \mathscr{A}\}$ of coverings of a set Y is called a uniformizing family if

(1). Each pair of coverings $\mathfrak{U}_\alpha$, $\mathfrak{U}_\beta \in \mathbf{G}$ have a common refinement $\mathfrak{U}_\gamma \in \mathbf{G}$.

(2). For each $\mathfrak{U}_\alpha \in \mathbf{G}$ there is a barycentric refinement $\mathfrak{U}_\beta \in \mathbf{G}$.

The family is separating if for each pair $x \neq y$ of elements, there is some $\mathfrak{U}_\alpha$ such that $y \bar{\in} \mathrm{St}(x, \mathfrak{U}_\alpha)$.

The intuitive idea is that sets belonging to the same covering have the same "size". The precise relation between **11.1** and **11.2** is

11.3 (a). Let $\mathbf{G} = \{\mathfrak{U}_\alpha \mid \alpha \in \mathscr{A}\}$ be a uniformizing family. For each $\alpha \in \mathscr{A}$, let $V_\alpha = \bigcup \{U \times U \mid U \in \mathfrak{U}_\alpha\}$. Then $u(\mathbf{G}) = \{V_\alpha \mid \alpha \in \mathscr{A}\}$ is a uniformity.

(b). Let $\mathfrak{F}$ be a uniformity. For each $V \in \mathfrak{F}$, let $\mathfrak{U}(V)$ be the covering $\{V[y] \mid y \in Y\}$. Then the family $\mathbf{c}(\mathfrak{F}) = \{\mathfrak{U}(V) \mid V \in \mathfrak{F}\}$ is a uniformizing family. Furthermore, $u[\mathbf{c}(\mathfrak{F})]$ is a uniformity equivalent to $\mathfrak{F}$.

Proof: The proof of (a) is straightforward, and is omitted. The proof for the first part of (b) depends on the observation that, whenever $W \circ W^{-1} \subset V$, then $\mathrm{St}(y, \mathfrak{U}(W)) \subset V[y]$ for each $y \in Y$. For the second part, one first establishes that $V_{\mathfrak{U}(V)} = V \circ V^{-1}$, and then notes that $V \subset V \circ V^{-1}$.

We now use a uniformity to derive a topology. Any uniformity $\mathfrak{F}$ in Y gives a topology $\mathscr{T}(\mathfrak{F})$ in Y by taking the family $\{V[y] \mid V \in \mathfrak{F}, y \in Y\}$ as basis. With this topology in Y, it is easy to verify that in the *space* $Y \times Y$, the family $\{\mathrm{Int}\, V \mid V \in \mathfrak{F}\}$ is a uniformity equivalent to $\mathfrak{F}$ and inducing the same topology; in particular, when considering the space $(Y, \mathscr{T}(\mathfrak{F}))$ there is no loss in generality to assume that each $V \in \mathfrak{F}$ is open in $Y \times Y$. If one starts with a uniformizing family $\mathbf{G} = \{\mathfrak{U}_\alpha \mid \alpha \in \mathscr{A}\}$ rather than with $\mathfrak{F}$, the induced topology in Y is that having the family $\{\mathrm{St}(y, \mathfrak{U}_\alpha) \mid y \in Y, \alpha \in \mathscr{A}\}$ as basis; then for each $\alpha \in \mathscr{A}$, each $U \in \mathfrak{U}_\alpha$ is open and, furthermore, the uniformity $u(\mathbf{G})$ gives the same topology.

A space with a topology derived from a uniformity is called a *uniform space*; it is Hausdorff if and only if the uniform structure (or uniformizing family) is separating.

Now let $(Y, \mathscr{T})$ be an arbitrary topological *space*. A uniformity $\mathfrak{F}$ in Y is called *compatible* with the given topology $\mathscr{T}$ whenever $\mathscr{T}(\mathfrak{F}) = \mathscr{T}$. Assume, for example, that Y is a metric space, and $\mathfrak{F}$ is a uniformity compatible with the topology. We have seen that equivalent metrics do not necessarily give the same uniformity, so the question arises: is there a metric that gives the topology of the space and also determines the given uniformity? The general answer to this question, which accounts for the importance of gauge spaces, is

11.4 Theorem Let $(Y, \mathscr{T})$ be a topological space, and $\mathfrak{F}$ a separating uniformity compatible with the topology. Then Y must be a gauge space, and there exists a family $\mathscr{D}$ of gauges such that

(1). The topology $\mathscr{T}$ is precisely the topology $\mathscr{T}(\mathscr{D})$.

(2). The uniformity determined by $\mathscr{D}$ is equivalent to the given uniformity $\mathfrak{F}$.

Proof: We may assume that the uniformity is given by a uniformizing family $\mathbf{G} = \{\mathfrak{U}_\alpha \mid \alpha \in \mathscr{A}\}$ of open coverings. For each open covering $\mathfrak{U}_\alpha \in \mathbf{G}$, we first construct a gauge d_α as follows:

Let $\mathfrak{V}_0 = \{Y\}$, $\mathfrak{V}_1 = \mathfrak{U}_\alpha$, and define a sequence $\{\mathfrak{V}_n \mid \mathfrak{V}_n \in \mathbf{G}\}$ of open coverings inductively so that each $\mathfrak{V}_{n+1}$ is a star refinement of $\mathfrak{V}_n$ (*cf.* VIII, **3.4**). For each pair of points $x, y \in Y$, let

$$\begin{aligned}\lambda(x, y) &= \inf\{2^{-n} \mid x \in \operatorname{St}(y, \mathfrak{V}_n)\},\\ \mu(x_0, \cdots, x_k) &= \lambda(x_0, x_1) + \cdots + \lambda(x_{k-1}, x_k),\\ d_\alpha(x, y) &= \inf\{\mu(x_0, \cdots, x_k) \mid \text{all finite sets } \{x_0, \cdots, x_k\} \subset Y\\ &\qquad \text{such that } x_0 = x, x_k = y\}.\end{aligned}$$

Then d_α is a gauge: $d_\alpha(x, y) = 0$ whenever $x = y$ and $d_\alpha(x, y) = d_\alpha(y, x)$ for each pair of points x, y. The triangle inequality is immediate from the definition of $d_\alpha(x, y)$ as an infimum.

We now determine the size of the d_α-balls. It is evident that $d_\alpha(x, y) \leqslant \lambda(x, y)$ so that $\operatorname{St}(y, \mathfrak{V}_n) \subset B(y; d_\alpha, 1/2^{n-1})$. We will show that also $B(y; d_\alpha, 1/2^n) \subset \operatorname{St}(y, \mathfrak{V}_{n-1})$; this will follow once we prove that

$$\lambda(x_0, x_k) \leqslant 2\mu(x_0, \cdots, x_k)$$

because then we would have $\lambda(x, y) \leqslant 2d_\alpha(x, y)$ so that $d_\alpha(x, y) < 2^{-n}$ gives $\lambda(x, y) < 2^{-n+1}$ and therefore that $x \in \operatorname{St}(y, \mathfrak{V}_{n-1})$.

The proposed inequality is proved by induction on k, the assertion being true for $k = 1$. Assume it is true for all $k \leqslant r$, and let $\mu(x_0, \cdots, x_r, x_{r+1}) = a$. Let s be the largest index for which $\mu(x_0, \cdots, x_s) \leqslant a/2$; then $\mu(x_{s+1}, \cdots, x_{r+1}) \leqslant a/2$ also, and by the induction hypothesis we find $\lambda(x_o, x_s) \leqslant 2a/2$, $\lambda(x_{s+1}, x_{r+1}) \leqslant a$, and certainly $\lambda(x_s, x_{s+1}) \leqslant a$.

Now choose the largest $2^{-m} \leqslant a$. By the definition of λ we must have $\lambda(x_0, x_s) \leqslant 2^{-m}$, $\lambda(x_s, x_{s+1}) \leqslant 2^{-m}$, and $\lambda(x_{s+1}, x_{r+1}) \leqslant 2^{-m}$. It follows from this that there is a $V \in \mathfrak{V}_m$ containing $\{x_s, x_{s+1}\}$ and that x_0 and x_{r+1} belong to $\mathrm{St}(V, \mathfrak{V}_m)$, which is contained in some $U \in \mathfrak{V}_{m-1}$. Thus, $\lambda(x_0, x_{r+1}) \leqslant 2^{-m+1} \leqslant 2a$ and the inductive step is complete.

We now complete the proof of the theorem rapidly. Let $\mathscr{D} = \{d_\alpha \mid \alpha \in \mathscr{A}\}$ be the family of gauges obtained in the manner indicated above, one for each $\mathfrak{U}_\alpha$. From what we have proved, the family $\mathfrak{B}(\mathscr{D})$ of balls is a basis for the topology $\mathscr{T}(\mathfrak{F})$ so that $\mathscr{T}(\mathscr{D}) = \mathscr{T}$. Verification that the uniformity determined by $\mathscr{D}$ is equivalent to $\mathfrak{F}$ being routine, the proof is complete.

It is not hard to verify that the family $\mathscr{D}$ of gauges determining the topology $\mathscr{T}(\mathfrak{F})$ and giving the uniformity $\mathfrak{F}$ can be described alternatively as follows: $\mathscr{D}$ consists of all gauges uniformly continuous when the uniform structure $\mathfrak{F}$ is used in Y and that determined by the metric $d(x, y) = |x - y|$ is used in E^1.

Problems

Section 1

1. Let $d: Y \times Y \to E^1$ satisfy: (1) $d(x, y) = 0$ if and only if $x = y$, and (2) $d(x, y) \leqslant d(z, x) + d(z, y)$ for all x, y, z. Prove: d is a metric.
2. Let d_i, $i = 1, 2, \cdots, n$ be n metrics in a set Y. For any constants $a_i \geqslant 0$, not all zero, show that $\sum_1^n a_i d_i(x, y)$ is a metric.

Section 2

1. Give an example to show that in a metric space $(X, \mathscr{T}(d))$, it is *not* necessarily true that (a) $\overline{B_d(a, r)} = \{x \mid d(x, a) \leqslant r\}$, and (b) $\mathrm{Fr}[B(a, r)] = \{x \mid d(x, a) = r\}$.
2. Prove that a pseudometric d in Y induces a topology that has $\{B_d(y, r) \mid y \in Y, r > 0\}$ as basis.

Section 3

1. Let $(Y, \mathscr{T}(d))$ be a metric space. Show that
$$\rho(x, y) = \frac{d(x, y)}{1 + d(x, y)}$$
is a metric in Y and that $\rho \sim d$.
2. Let $Y_1, \cdots, Y_n$ be metric spaces, and $d_1, \cdots, d_n$ be metrics for these spaces. Show that
$$d[(x_1, \cdots, x_n), (y_1, \cdots, y_n)] = \max\{d_i(x_i, y_i) \mid i = 1, \cdots, n\}$$
is a metric for the space $\prod_1^n Y_i$.

3. Let $X =]0, 1[\subset E^1$. Show:
   a. $d(x, y) = |x^{-1} - y^{-1}|$ is a metric on X.
   b. d is equivalent to the usual metric $d_0(x, y) = |x - y|$ on X.
   c. There exists no metric on E^1 that coincides with metric d on X.

Section 4

1. Let $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite family of sets in a metric space Y. Let d be a metric for Y. Prove that the map $y \to \sup\{d(y, \mathscr{C}A_\alpha) \mid a \in \mathscr{A}\}$ of Y into $\tilde{E}^1$ is continuous.
2. Prove:
   a. $\delta(A) = 0 \Leftrightarrow (A = \varnothing) \vee (A \text{ is a single point})$.
   b. $\delta(\bar{A}) = \delta(A)$.
   c. $A \subset B \Rightarrow \delta(A) \leqslant \delta(B)$.
   d. If $A \cap B \neq \varnothing$, then $\delta(A \cup B) \leqslant \delta(A) + \delta(B)$.
3. Prove:
   a. $d(\bar{A}, \bar{B}) = d(A, B)$.
   b. $d(A, C) \leqslant d(A, B) + d(B, C) + \delta(B)$.
   c. $d(A, C) \leqslant d(A, B) + d(A \cup B, C) + \delta(B)$.
   d. $d(A, B \cup C) = \min[d(A, B), d(A, C)]$.
   e. $(A \subset B \subset \bar{A}) \Rightarrow (d(p, A) = d(p, B) = d(p, \bar{A}))$.
4. Let S be the set of all *closed* non-empty subsets of a metric space Y, and let p be a fixed point of Y. In the set S, define
   $$d_p(A, B) = \sup\{|d(y, A) - d(y, B)|\, e^{-d(y,p)} \mid y \in Y\}.$$
   Show that d_p is a metric in the set S and that $d_p \sim d_q$ for any two $p, q \in Y$.
5. In the complex plane, let $Y = \{z \mid |z| < 1\}$, and define $\rho(x, y)$ in $Y \times Y$ as follows: Extend the line joining x, y so that it intersects the circumference at points v, u, and take the order of the points to be v, x, y, u. Let d be ordinary distance measured along this line and let
   $$\rho(x, y) = \log\left[\frac{d(x, u)}{d(x, v)} \cdot \frac{d(y, v)}{d(y, u)}\right].$$
   Prove that ρ is a metric in Y.
6. Prove: A connected metric space with more than one point must contain at least $2^{\aleph_0}$ points.
7. Let d be a pseudometric in Y, and let R be the equivalence relation $xRy \Leftrightarrow d(x, y) = 0$. Then Y/R is a metric space (see **1**). Prove that, if the pseudometric topology of **2**, Problem 2, is used in Y, then the projection $p\colon Y \to Y/R$ is a continuous open and closed surjection.
8. Let (X, d) be a metric space, and $\mathscr{F}(X)$ the set of all nonempty bounded closed sets in X. For $(A, B) \in \mathscr{F}(X) \times \mathscr{F}(X)$ define $h(A, B) = \sup\{d(x, B) \mid x \in A\}$ and let $\rho(A, B) = \max[h(A, B), h(B, A)]$. Prove:
   a. ρ is a metric (called the Hausdorff metric) in $\mathscr{F}(X)$.
   b. $\rho(\{a\}, \{b\}) = d(a, b)$.
   c. $\rho(A, B) \leqslant \epsilon$ if and only if $[A \subset U(B, \epsilon)] \wedge [B \subset U(A, \epsilon)]$.
   d. $\rho(A, B) \leqslant \epsilon$ if and only if $B \subset U(A, \epsilon)$ and $\forall a \in A\colon B(a, \epsilon) \cap B \neq \varnothing$.

Section 5

1. Let $(Y, \mathscr{T}(d))$ be a path-connected metric space, and $f: I \to Y$ be a path in Y. For each finite subdivision $t_0 = 0 < t_1 < \cdots < t_n = 1$ of I, let
$$L(t_0, \cdots, t_n) = \sum_1^n d[f(t_{i-1}), f(t_i)];$$
if $\lambda(f) = \sup\{L(t_0, \cdots, t_n) \mid \text{all finite subdivisions of } I\}$ is finite, then f is called rectifiable, and $\lambda(f)$ is its length. Now, let $A \subset Y$ also be path-connected; for each $a, b \in A$, define $\rho(a, b) = \inf\{\lambda(f) \mid \text{all rectifiable paths } in\ A \text{ joining } a, b\}$. Prove: ρ is a metric on A (ρ is called the "geodesic distance" on A). Show that in E^2, starting from the metric of **I**, Ex. 3, the geodesic distance on $A = S^1$ is the length of the shortest of the two arcs on S^1 joining the points. Finally, let C be a closed curve in E^2 enclosing the origin and having no rectifiable subarc, and let A be the cone having $(0, 0, 1) = v_0$ as vertex and C as base; prove that the geodesic distance between two distinct points $c, c' \in C$ is the length of the broken straight line $\overline{cv_0} \cup \overline{v_0c'}$.
2. Let Y be a separable metric space, $\{U_i \mid i \in Z^+\}$ a basis. For any set $A \subset Y$, define
$$\mu_n(A) = 1 \quad \text{if} \quad U_n \cap A \neq \varnothing,$$
$$\mu_n(A) = 0 \quad \text{if} \quad U_n \cap A = \varnothing,$$
and set
$$\mu(A) = \sum_1^\infty 2^{-n}\mu_n(A).$$
Prove:
   a. If $A \subset B$, then $\mu(A) \leqslant \mu(B)$.
   b. $\mu(A) = \mu(\bar{A})$.
   c. $\mu(A \cup B) \leqslant \mu(A) + \mu(B)$.
   d. If A is closed and a proper subset of B, then $\mu(A) \neq \mu(B)$.
3. Prove: (a): The continuous open image of a $1°$ countable space is also $1°$ countable, and (b): $X \times Y$ is $1°$ countable if and only if both X and Y are $1°$ countable.
4. Let Y be a topological space. A subset $\mathscr{G} \subset \mathscr{P}(Y)$ is called local if each $y \in Y$ has a nbd $U(y) \in \mathscr{G}$. A subset $\mathscr{H} \subset \mathscr{P}(Y)$ is called G-hereditary whenever it satisfies the following three conditions:
   (a). If $A \in \mathscr{H}$ and if $W \subset A$ is open in A, then $W \in \mathscr{H}$.
   (b). If U, V are open, and if $U, V \in \mathscr{H}$, then $U \cup V \in \mathscr{H}$.
   (c). If $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ is any family of pairwise disjoint open sets in Y, and if each $U_\alpha \in \mathscr{H}$, then $\bigcup \{U_\alpha \mid \alpha \in \mathscr{A}\} \in \mathscr{H}$.

   Prove: If Y is a metric space, and if $\mathscr{H} \subset \mathscr{P}(Y)$ is both G-hereditary and local, then $Y \in \mathscr{H}$.

   [*Hint:* Let $\mathfrak{U} = \{U(y) \mid y \in Y\}$ be an open covering of Y with each $U(y) \in \mathscr{H}$. Show first that $\mathfrak{U}$ has a refinement $\mathfrak{B}$ that can be decomposed as $\mathfrak{B} = \{\mathscr{V}_i \mid i \in Z^+\}$, where each $\mathscr{V}_i$ is a pairwise disjoint family of open sets (*cf.* VIII, 3.5). Then prove successively that each of the following sets belong to $\mathscr{H}$: (1). Each $Q_i = \bigcup \{V \mid V \in \mathscr{V}_i\}, i \geqslant 1$; (2). Each $G_n = \bigcup_1^n Q_i, n \geqslant 1$; (3). Each $W_n = \{y \mid d(y, Y - G_n) > 1/n\}, n \geqslant 1$; (4). Each $H_n = W_n - W_{n-3}, n \geqslant 1$, where $W_i = \varnothing$ for $i \leqslant 0$; (5). $Y = \bigcup_1^\infty H_{4n+1} \cup \bigcup_1^\infty H_{4n-1}$. This result is due to E. A. Michael, and is also true for paracompact spaces Y.]

5. Use the result of Problem 4 to prove D. Montgomery's theorem: Let Y be a metric space, and let $B \subset Y$. If each $b \in B$ has a nbd $U(b)$ in B that is a G_δ in Y, then B is a G_δ in Y.

Section 7

1. Let $\{Y_n \mid n \in Z^+)$ be metric spaces, and choose a metric d_i for each Y_i. Let $k_i > 0$ be constants such that $\sum_N^\infty k_i\delta_i(Y_i) < \infty$ for large N. Show that $\rho(x, y) = \Sigma k_i d_i(x_i, y_i)$ metrizes the topology of $\prod_i Y_i$.
2. Let $\{Y_i \mid i \in Z^+\}$ be metric spaces and choose metrics $\{d_i\}$ such that $\delta_i(Y_i) \leqslant M/i$ for each i. Show that

$$d(x, y) = \sqrt[p]{\sum_1^\infty [d_i(x_i, y_i)]^p} \qquad (p > 1)$$

metrizes the cartesian product topology of $\prod Y_n$.

Section 8

1. Let $\mathfrak{c} = 2^{\aleph_0}$. Prove that there can be no continuous surjection of $E^\mathfrak{c}$ onto $l^2(\mathfrak{c})$.
2. Show that I^∞ has no interior in $E^{\aleph_0}$.

Section 9

1. Let Y be regular. Prove that Y is metrizable if and only if it has a sequence $\{\mathfrak{U}_n \mid n \in Z^+\}$ of open coverings such that each *finite* open covering of Y is refined by some $\mathfrak{A}_n$.
2. Let X be a T_0-space, and assume that there is a countable family $\{\mathfrak{U}_n \mid n \in Z^+\}$ of open coverings such that the sets $\{\text{St}[\text{St}(x, \mathfrak{U}_n), \mathfrak{U}_n] \mid n \in Z^+\}$ form a nbd basis at x for each $x \in X$. Prove that X is metrizable.
3. Prove that a Hausdorff space X is metrizable if and only if for each $x \in X$ there exists a countable family $\{V_n(x) \mid n \in Z^+\}$ of open sets such that
   a. $\{V_n(x) \mid n \in Z^+\}$ forms a basis at x.
   b. $\forall\, V_n(x)\ \exists\, m = m(x, n) > n$: $V_m(y) \cap V_m(x) \neq \varnothing \Rightarrow V_m(y) \subset V_n(x)$.
4. Prove: If a T_0-space Y is the union of a nbd-finite family of closed metrizable subspaces, then Y is also metrizable.

Section 10

1. If d is any gauge in a set Y, then for any $A \subset Y$ we define $d(x, A) = \inf\{d(x, a) \mid a \in A\}$. Now let $(Y, \mathscr{T}(\mathscr{D}))$ be a gauge space, and let $A \subset Y$ be a fixed set. Prove:
   a. $\bar{A} = \bigcap_{d^+ \in \mathscr{D}^+} \{y \mid d^+(y, A) = 0\}$.
   b. The map $y \to d^+(y, A)$ is continuous for each $d^+ \in \mathscr{D}^+$.
2. Let $(Y, \mathscr{T}(\mathscr{D}))$ be a gauge space, and let $A \subset Y$. Prove:

$$\text{Int}(A) = \{y \mid \exists\, d^+ \in \mathscr{D}^+\ \exists\, \epsilon > 0: B(y; d^+, \epsilon) \subset A\}.$$

3. Show that the class of completely regular spaces is the smallest class of topological spaces that contains the separable metric spaces and is closed under formation of subspaces and cartesian products.

Section 11

1. Let Y be any space, and let $\{\mathfrak{U}_i \mid i \geqslant 0\}$ be a sequence of open coverings such that each $\mathfrak{U}_{i+1}$ is a star-refinement of $\mathfrak{U}_i$. Prove: There exists a metric space M and a continuous surjection $p\colon Y \to M$ such that the open covering $\{p^{-1}[B(m,1)] \mid m \in M\}$ refines $\mathfrak{U}_0$. [*Hint:* Let d be a gauge in Y determined as in **11.4**, and let M be the quotient space of Y by the relation $xRy \Leftrightarrow d(x,y) = 0$.]

Convergence

X

The concept of convergence in general topology will be developed in this chapter. This concept includes the usual ones encountered in analysis, and applies equally well even to multiple-valued functions. The fundamental idea is that of a filterbase in a topological space.

1. Sequences and Nets

In this section we give explicitly the definition of convergence for sequences and nets in general topological spaces, and indicate how these lead to the notion of a filterbase.

1.1 Definition Let Z^+ be the set of positive integers. A *sequence* in a space Y is a map $\varphi: Z^+ \to Y$. We say:

(1). φ *converges* to y_0 (written $\varphi \to y_0$) if:

$$\forall\, U(y_0)\ \exists\, N\ \forall\, n \geqslant N\ : \varphi(n) \in U.$$

(2). φ *accumulates* at y_0 (written $\varphi \succ y_0$) if:

$$\forall\, U(y_0)\ \forall\, N\ \exists\, n \geqslant N\ : \varphi(n) \in U.$$

A sequence φ in Y is also denoted by $\{y_n\}$ to indicate that $\varphi(n) = y_n$; in this case, $\varphi \to y_0$ is written $y_n \to y_0$, or $\lim y_n = y_0$, and y_0 is called a limit point of the sequence $\{y_n\}$. For each monotone increasing map $\mu: Z^+ \to Z^+$ and each sequence $\varphi: Z^+ \to Y$, the map $n \to \varphi \circ \mu(n)$ is called a subsequence of φ.

Ex. 1 For sequences of real numbers, the Definitions in **1.1** reduce to the usual ones of elementary analysis. For example, $\varphi \to y_0$ becomes

$$\forall\ \varepsilon > 0\ \exists\ N\ \forall\ n \geqslant N: \varphi(n) \in B(y_0, \varepsilon).$$

Ex. 2 The concepts of limit and accumulation point should be delineated from each other, and from "cluster point of the set $\varphi(Z^+)$". Clearly, $[\varphi \to y_0] \Rightarrow [\varphi \succ y_0]$, and in Hausdorff spaces, $[y_0 \in \varphi(Z^+)'] \Rightarrow [\varphi \succ y_0]$ (*cf.* VII, **1.4**). However, none of these implications is reversible: in E^1, the sequence $\varphi(n) = n + (-1)^n n$ has exactly one accumulation point, 0; yet φ does not converge, nor does $\varphi(Z^+)$ have any cluster point.

That sequences are not sufficient for the purposes of analysis is already evident in the classical integration theory. There, with each one of the $2^{\aleph_0}$ finite partitions of an interval $[a, b]$, one associates a Riemann sum (in fact, $2^{\aleph_0}$ such sums, depending on the points where [*cf.* **2**, Ex. 7] the function to be integrated is evaluated), and a "limit" of these sums in a "direction," specified by refinement in the set of partitions, is required. Nets provide a generalization of sequences that can handle this case.

1.2 Definition A *directed set* D is a preordered set with the following property: For each $a, b \in D$, there exists a $c \in D$ such that $a \prec c$ and $b \prec c$. A net in a space Y is a map $\varphi: D \to Y$ of some directed set D, and

(1). $\varphi \to y_0$ if $\forall\ U(y_0)\ \exists\ a\ \forall\ b \succ a\ : \varphi(b) \in U$.
(2). $\varphi \succ y_0$ if $\forall\ U(y_0)\ \forall\ a\ \exists\ b \succ a\ : \varphi(b) \in U$.

Ex. 3 With the natural ordering $\leqslant$, Z^+ is a directed set. A net with domain Z^+ is simply a sequence, and the definitions $\varphi \to y_0$, $\varphi \succ y_0$ in **1.2** coincide with those in **1.1**.

Ex. 4 The set of all open coverings of a space Y, with $\{U\} < \{V\}$ meaning $\{V\}$ refines $\{U\}$, is a directed set, since any two coverings have a common refinement. Similarly, the set of all partitions of Y, preordered in the same way, is a directed set.

Definitions **1.2** (1), (2) can be stated more directly by first introducing

1.3 Definition Let D be a directed set. The terminal set T_a determined by an $a \in D$ is $\{b \in D \mid a \prec b\}$.

Then, for a net $\varphi: D \to Y$, we have:

$$\varphi \to y_0 \quad \text{if} \quad \forall\ U(y_0)\ \exists\ T_a\ : \varphi(T_a) \subset U,$$
$$\varphi \succ y_0 \quad \text{if} \quad \forall\ U(y_0)\ \forall\ T_a\ : \varphi(T_a) \cap U \neq \varnothing.$$

This formulation of the definitions shows clearly that the convergence of φ is not determined by the set $\varphi(D) \subset Y$, but rather by the behavior of the family of sets $\{\varphi(T_\alpha) \mid a \in D\}$ in Y. Now, if we regard the map φ as serving simply to define the family $\{\varphi(T_a)\}$ of sets, then we can think of convergence as a notion that involves only the space Y and pertains to any family of subsets of Y having properties similar to those of the family $\{\varphi(T_a)\}$. Such families of sets are called *filterbases*.

2. Filterbases in Spaces

2.1 Definition Let Y be a space. A filterbase $\mathfrak{A}$ in Y is a family $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ of subsets of Y having the two properties:

(1). $\forall\, \alpha \in \mathscr{A} \; : A_\alpha \neq \varnothing$,
(2). $\forall\, \alpha \; \forall\, \beta \; \exists\, \gamma \; : A_\gamma \subset A_\alpha \cap A_\beta$.

Ex. 1 Let $\mathfrak{A}$ consist of one nonempty set; then $\mathfrak{A}$ is a filterbase. If Y is an infinite set, $\{A \subset Y \mid \aleph(\mathscr{C}A) < \aleph_0\}$ is a filterbase in Y.

Ex. 2 Let $\varphi\colon D \to Y$ be a net. Then the family $\mathfrak{A}(\varphi) = \{\varphi(T_a) \mid a \in D\}$ is a filterbase in Y. For, given $\varphi(T_a)$ and $\varphi(T_b)$, first find a $c \in D$ such that $a \prec c$, $b \prec c$, and then observe that $T_c \subset T_a \cap T_b$ because $\prec$ is transitive. $\mathfrak{A}(\varphi)$ is called the filterbase *determined by the net* φ.

Ex. 3 Let Y be a space and $y_0 \in Y$. The family $\mathfrak{U}(y_0)$ of all nbds of y_0 is clearly a filterbase, called the *nbd filterbase of* y_0.

The Boolean algebra of filterbases is given in

2.2 Let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ and $\mathfrak{B} = \{B_\beta \mid \beta \in \mathscr{B}\}$ be two filterbases in Y. Then:

(1). $\mathfrak{A} \cup \mathfrak{B} = \{A_\alpha \cup B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$ is a filterbase.
(2). If each $A_\alpha \cap B_\beta \neq \varnothing$, then

$$\mathfrak{A} \cap \mathfrak{B} = \{A_\alpha \cap B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$$

is a filterbase.
(3). For each finite family $\{A_{\alpha_1}, \cdots, A_{\alpha_n}\} \subset \mathfrak{A}$, there is an $A_\gamma \in \mathfrak{A}$ such that $A_\gamma \subset A_{\alpha_1} \cap \cdots \cap A_{\alpha_n}$. In particular, each finite intersection of members of a filterbase is not empty.

Proof: (1) and (2) are trivial, since $\cup$ distributes over $\cap$, and conversely. We prove (3) by induction. According to **2.1**, the assertion is true for $n = 2$. Assuming its truth for $n = k$, we prove it for $n = k + 1$ by first finding an $A_{\gamma_1} \subset A_{\alpha_1} \cap \cdots \cap A_{\alpha_k}$ and then an $A_\gamma \subset A_{\gamma_1} \cap A_{\alpha_{k+1}}$.

The convergence concepts for filterbases are as follows:

2.3 Definition Let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a filterbase in Y. Then:

(1). $\mathfrak{A}$ converges to y_0 (written $\mathfrak{A} \to y_0$) if: $\forall\, U(y_0) \exists\, A_\alpha : A_\alpha \subset U$.

(2). $\mathfrak{A}$ accumulates at y_0 (written $\mathfrak{A} \succ y_0$) if:

$$\forall\, U(y_0) \forall\, A_\alpha \colon A_\alpha \cap U \neq \varnothing.$$

It is evident from the definition (*cf.* III, **4.3**) that $\mathfrak{A} \succ y_0$ if and only if $y_0 \in \bigcap_\alpha \bar{A}_\alpha$; this alternative characterization of the accumulation points of a filterbase will be used very frequently.

Ex. 4 Let $\mathfrak{A}$ consist of a single set, A. If A has only one point, then $\mathfrak{A} \to a$; if A has more than one point, $\mathfrak{A}$ accumulates at each point of $\bar{A}$. If Y is an infinite discrete space, the filterbase $\{A \mid \aleph(\mathscr{C}A) < \aleph_0\}$ does not converge and has no accumulation point.

Ex. 5 Let $\mathfrak{A}(\varphi)$ be the filterbase determined by the net $\varphi\colon D \to Y$. Then $\varphi \to y_0$ $[\varphi \succ y_0]$ if and only if $\mathfrak{A}(\varphi) \to y_0$ $[\mathfrak{A}(\varphi) \succ y_0]$.

Ex. 6 Let $\mathfrak{U}(y_0)$ be the nbd filterbase at y_0. Then $\mathfrak{U}(y_0) \to y_0$; it may also converge to other points, as in Sierpinski space, where $\mathfrak{U}(0) \to 0$ and also to 1.

Ex. 7 Let $g\colon [a, b] \to E^1$ be given, and let $\{P\}$ be the directed set of all partitions of $[a, b]$ *cf.* 1, Ex. 4. For each partition P, let $a = x_0 < x_1 < \cdots < x_n = b$ be its end points, and let R_P be the set of all real numbers

$$\sum_{i=1}^{n} g(\xi_i)[x_i - x_{i-1}],$$

for all choices $\xi_i \in [x_{i-1}, x_i]$. Define $A_P = \bigcup \{R_Q \mid P < Q\}$; then $\mathfrak{A} = \{A_P\}$ is a filterbase in E^1, and g is Riemann integrable on $[a, b]$ if and only if $\mathfrak{A}$ converges.

The idea of "subsequence" is given by

2.4 Definition Let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ and $\mathfrak{B} = \{B_\beta \mid \beta \in \mathscr{B}\}$ be two filterbases on Y. $\mathfrak{B}$ is *subordinate to* $\mathfrak{A}$, written $\mathfrak{B} \vdash \mathfrak{A}$, if

$$\forall\, A_\alpha\ \exists\, B_\beta \colon B_\beta \subset A_\alpha.$$

Ex. 8 Let $\mathfrak{A}(\varphi)$ be the filterbase in Y determined by a sequence φ. If $\psi\colon Z^+ \to Y$ is a subsequence of φ, then $\mathfrak{A}(\psi) \vdash \mathfrak{A}(\varphi)$ follows at once. However, it is not true that if $\mathfrak{B} \vdash \mathfrak{A}(\varphi)$, then $\mathfrak{B}$ is a filterbase determined by some subsequence of φ, or even that $\mathfrak{B}$ has at most countably many members. For example, if Y is an uncountable set, the sets of $\mathfrak{A}(\varphi)$ together with $\{A \mid \aleph(\mathscr{C}A) < \aleph_0\}$ form a filterbase $\mathfrak{B} \vdash \mathfrak{A}(\varphi)$.

Ex. 9 The relation $\vdash$ is a preordering in the family of all filterbases in Y, but $\vdash$ is not a partial ordering: In the space $Y = Z^+$, the two distinct filterbases $\mathfrak{B} = \{T_{2n} \mid n \in Z^+\}$ and $\mathfrak{A} = \{T_n \mid n \in Z^+\}$ satisfy $\mathfrak{A} \vdash \mathfrak{B}$ and $\mathfrak{B} \vdash \mathfrak{A}$.

The subordination relation has the useful properties:

2.5 (1). If $\mathfrak{A} \subset \mathfrak{B}$, then $\mathfrak{B} \vdash \mathfrak{A}$.
(2). If $\mathfrak{B} \vdash \mathfrak{A}$, then each member of $\mathfrak{B}$ meets every member of $\mathfrak{A}$.
(3). $\mathfrak{A} \to y_0$ if and only if $\mathfrak{A} \vdash \mathfrak{U}(y_0)$.

Proof: (1) is obvious. (2). Assume $\exists B_\beta \; \exists A_\alpha: A_\alpha \cap B_\beta = \varnothing$; since $\mathfrak{B} \vdash \mathfrak{A}$, for this A_α we can find a $B_\gamma \subset A_\alpha$, and then $B_\gamma \cap B_\beta = \varnothing$ contradicts that $\mathfrak{B}$ is a filterbase. (3). Compare the definition of $\mathfrak{A} \vdash \mathfrak{U}(y_0)$ with **2.3**(1).

Remark 1: In any space Y, the concepts of convergence based on filterbases and on nets are "equivalent" in the sense that first: (*cf.* Ex. 5) each net φ determines a filterbase $\mathfrak{A}(\varphi)$ such that $\varphi \to y_0$ according to **1.2**, if and only if $\mathfrak{A}(\varphi) \to y_0$ according to **2.3**; and second: (G. Bruns and J. Schmidt) each filterbase $\mathfrak{A}$ determines a net $\varphi: D \to Y$ such that $\mathfrak{A} \to y_0$ if and only if $\varphi \to y_0$. To prove the second statement, it suffices to show that φ and D can be selected so that for each $A \in \mathfrak{A}$ there is a $d \in D$ such that $\varphi(T_d) = A$. Let D be the set of all ordered couples (a, A), where $a \in A \in \mathfrak{A}$, and preorder D by setting $(a, A) \prec (b, B)$ whenever $B \subset A$; then D is directed because $\mathfrak{A}$ is a filterbase. Defining the net $\varphi: D \to Y$ to be the map $(a, A) \to a$, it is trivial to verify that if T is the terminal segment determined by any given (b, B), then $\varphi(T) = B$.

It frequently happens, in any given instance, that one of these two methods for expressing convergence is more convenient than the other.

Remark 2: In the Boolean algebra $\mathscr{P}(Y)$, a family of sets $\mathscr{F}$ with the properties (1) $[A, B \in \mathscr{F}] \Rightarrow [A \cap B \in \mathscr{F}]$, and (2) $(A \in \mathscr{F}) \wedge (A \subset B) \Rightarrow (B \in \mathscr{F})$, is called a *dual ideal*; it is a *proper* dual ideal, or *filter*, if also (3) $\varnothing \mathrel{\overline{\in}} \mathscr{F}$. It is easy to see that a family $\{A_\alpha \mid \alpha \in \mathscr{A}\}$ is contained in a filter if and only if $\{A_\alpha\}$ is a filterbase; $\{A_\alpha\}$ is then called a base for the smallest filter containing it.

3. Convergence Properties of Filterbases

Though a filterbase may converge to more than one point (*cf.* **2**, Ex. 6), this can occur only in non-Hausdorff spaces.

3.1 Theorem Y is Hausdorff if and only if each convergent filterbase in Y converges to exactly one point.

Proof: Assume that Y is Hausdorff and that $\mathfrak{A} \to y_0$. For any $y_1 \neq y_0$, find disjoint nbds $U(y_0)$, $U(y_1)$; since by hypothesis there is some $A_\alpha \subset U(y_0)$ and since any two A_α, A_β have nonempty intersection, there can be no $A_\beta \subset U(y_1)$; thus, $\mathfrak{A}$ cannot converge to $y_1 \neq y_0$. Conversely, assume that Y is not Hausdorff. Then there must exist y_0, y_1 such that $\forall\, U(y_0) \; \forall\, U(y_1) : U(y_0) \cap U(y_1) \neq \varnothing$; by **2.2**, $\mathfrak{B} = \mathfrak{U}(y_0) \cap \mathfrak{U}(y_1)$ is therefore a filterbase, and evidently $\mathfrak{B} \to y_0$, $\mathfrak{B} \to y_1$.

The usual relations between the convergence properties of sequences and their subsequences extend to filterbases in the form

3.2 Theorem (1). If $\mathfrak{A}$ converges to y_0, then $\mathfrak{A}$ accumulates at y_0 and, in Hausdorff spaces, at no point other than y_0.

(2). Let $\mathfrak{B} \vdash \mathfrak{A}$. Then;

(a). $[\mathfrak{A} \to y_0] \Rightarrow [\mathfrak{B} \to y_0]$.

(b). $[\mathfrak{B} \succ y_0] \Rightarrow [\mathfrak{A} \succ y_0]$.

Proof: (1). Given $U(y_0)$, there is some $A_\alpha \subset U(y_0)$; since each A_β must intersect A_α, it follows that $\forall A_\beta : A_\beta \cap U(y_0) \neq \varnothing$, so $\mathfrak{A} \succ y_0$. Now let Y be Hausdorff, and let $y_1 \neq y_0$; choosing disjoint nbds $U(y_0)$, $U(y_1)$, there must be some $A_\alpha \in \mathfrak{A}$ contained in $U(y_0)$; then $A_\alpha \cap U(y_1) = \varnothing$, and so $\mathfrak{A}$ cannot accumulate at y_1.

(2a). $\forall U(y_0)\ \exists A_\alpha : A_\alpha \subset U(y_0)$; since $\mathfrak{B} \vdash \mathfrak{A}$, there is a $B_\beta \subset A_\alpha$, so $\mathfrak{B} \to y_0$ also.

(2b). Given $U(y_0)$ and A_α, there is some $B_\beta \subset A_\alpha$, and since $\forall B_\beta : B_\beta \cap U(y_0) \neq \varnothing$, we find $\forall A_\alpha : A_\alpha \cap U(y_0) \neq \varnothing$, which proves $\mathfrak{A} \succ y$.

3.3 Corollary (1). $\mathfrak{A} \to y_0$ if and only if $\forall \mathfrak{B} \vdash \mathfrak{A}\ \exists \mathfrak{C} \vdash \mathfrak{B} : \mathfrak{C} \to y_0$.

(2). $\mathfrak{A} \succ y_0$ if and only if $\exists \mathfrak{B} \vdash \mathfrak{A} : \mathfrak{B} \to y_0$.

Proof: Ad (1). The "only if" is trivial. "If": Assume that $\mathfrak{A}$ does not converge to y_0, so that $\exists U(y_0)\ \forall A_\alpha : A_\alpha \not\subset U(y_0)$. It then follows from **2.2** that $\mathfrak{B} = \mathfrak{A} \cap \mathscr{C}U$ is a filterbase, and clearly $\mathfrak{B} \vdash \mathfrak{A}$. Since y_0 is not an accumulation point of $\mathfrak{B}$ we find from **3.2**(1), **3.2**(2b) that no filterbase subordinated to $\mathfrak{B}$ can converge to y_0.

Ad (2). If $\mathfrak{B} \vdash \mathfrak{A}$ and $\mathfrak{B} \to y_0$, then by **3.2**, $\mathfrak{B} \succ y_0$, so that $\mathfrak{A} \succ y_0$. Conversely, assume $\mathfrak{A} \succ y_0$ and let $\mathfrak{B} = \mathfrak{A} \cap \mathfrak{U}(y_0)$; $\mathfrak{B}$ is a filterbase because all $A_\alpha \cap U(y_0) \neq \varnothing$, and we evidently have $\mathfrak{B} \vdash \mathfrak{A}$, and $\mathfrak{B} \to y_0$.

Restricting $\mathfrak{A}$ and $\mathfrak{B}$ to be *sequences*, so that $\mathfrak{B} \vdash \mathfrak{A}$ means $\mathfrak{B}$ is a *subsequence* of $\mathfrak{A}$, **3.2** and **3.3** become the usual statements involving sequences and subsequences in the spaces of elementary analysis. Considering now these *sequential* statements in general topological spaces, it is clear that **3.2** is also true, and we will prove in **6.1**, that **3.3**(1) is valid too; however, **3.3**(2) fails: that is, a *sequence* may accumulate to a point y_0 and yet have no *subsequence* converging to y_0. This pathological behavior is one reason that sequences alone are unsatisfactory for the purposes of general topology.

Ex. 1 (R. Arens). Let $Y = [Z^+ \times Z^+] \cup \{(0, 0)\}$ with the following (Hausdorff) topology:

(a). Discrete topology on $Z^+ \times Z^+$.

(b). Nbds of $(0, 0)$ are all sets U containing $(0, 0)$ and satisfying the condition: $\exists N \ \forall n \geqslant N$: U contains all but at most finitely many points of $n \times Z^+$.

The sequence giving the diagonal enumeration of $Z^+ \times Z^+$ accumulates at $(0, 0)$; but it is easy to see that no *subsequence* can converge to $(0, 0)$.

4. Closure in Terms of Filterbases

In this section, we show that all the basic topological concepts can be expressed by filterbases; for this, it suffices to characterize the closure operation.

4.1 Theorem Let Y be a space, and $A \subset Y$. Then $y \in \overline{A}$ if and only if there is a filterbase *on* A converging to y.

Proof: Let $y \in \overline{A}$; then $U(y) \cap A \neq \varnothing$ for each $U \in \mathfrak{U}(y)$, so that $\mathfrak{B} = A \cap \mathfrak{U}(y)$ is a filterbase on A, and clearly $\mathfrak{B} \to y$. Conversely, assume that $\mathfrak{B}$ is a filterbase on A converging to y; then

$$\forall \ U(y) \ \exists \ B_\beta \ : B_\beta \subset U,$$

so that every $U(y)$ contains points of A, that is, $y \in \overline{A}$.

Ex. 1 Theorem **4.1** is not true if we restrict attention *only* to *sequences*. In the ordinal space $[0, \Omega]$ we have $\Omega \in \overline{[0, \Omega[}$, but there can be no *sequence* in $[0, \Omega[$ that converges to Ω (II, **9.1**). Thus, beside the failure of **3.3**(2), sequences alone are generally incapable of expressing all topological concepts.

4.2 (a). $A \subset Y$ is closed if and only if the accumulation points of each filterbase on A all lie in A.

(b). $a \in A'$ if and only if there is a filterbase on $A - \{a\}$ that converges to a.

The simple proofs are omitted.

5. Continuity; Convergence in Cartesian Products

By using filterbases, the continuity of a map can be expressed in a manner analogous to that in elementary analysis. We begin with the simple observation that if $\mathfrak{A}$ is a filterbase in X and $f\colon X \to Y$ is any map, then (1): $f(\mathfrak{A}) = \{f(A) \mid A \in \mathfrak{A}\}$ is a filterbase in Y, and (2): if $\mathfrak{A} \vdash \mathfrak{B}$, then $f(\mathfrak{A}) \vdash f(\mathfrak{B})$.

5.1 Theorem Let $f\colon X \to Y$. Then f is continuous at $x_0 \in X$ if and only if the filterbase $f(\mathfrak{U}(x_0))$ converges to $f(x_0)$.

Proof: The statement that f is continuous at x_0 is

$$\forall\, W(f(x_0))\ \exists\, U(x_0)\ : f(U) \subset W.$$

This is exactly the statement that the filterbase $f(\mathfrak{U}(x_0))$ converges to $f(x_0)$.

5.2 Corollary $f: X \to Y$ is continuous on X if and only if $f(\mathfrak{A}) \to f(x)$ for each $x \in X$ and each filterbase $\mathfrak{A} \to x$.

Proof: Assume f continuous on X, and let $\mathfrak{A} \to x$; then [**2.5**(c)] $\mathfrak{A} \vdash \mathfrak{U}(x)$ so that $f(\mathfrak{A}) \vdash f(\mathfrak{U}(x))$, and since $f(\mathfrak{U}(x)) \to f(x)$, so also does $f(\mathfrak{A})$. Conversely, assume the condition holds; we prove that $f(\overline{A}) \subset \overline{f(A)}$ for all $A \subset X$. Let $a \in \overline{A}$; then there is a filterbase $\mathfrak{A}$ on A with $\mathfrak{A} \to a$, so that $f(\mathfrak{A}) \to f(a)$; since $f(\mathfrak{A})$ is on $f(A)$, this shows that $f(a) \in \overline{f(A)}$.

For maps of arbitrary spaces into regular spaces, this formulation of continuity gives rise to a process called *extension by continuity*:

5.3 Theorem Let D be a dense subset of X, let Y be a regular space, and let $f: D \to Y$ be continuous. Then f has a continuous extension $F: X \to Y$ if and only if the filterbase $f(D \cap \mathfrak{U}(x))$ converges for each $x \in X$. If F exists, then F is unique.

Proof: *Sufficiency:* Because D is dense, we have $D \cap U \neq \varnothing$ for each fixed $x_0 \in X$ and each $U \in \mathfrak{U}(x_0)$; consequently $D \cap \mathfrak{U}(x_0)$ is indeed a filterbase in X. For each $x \in X$, define $F(x)$ to be the limit of $f(D \cap \mathfrak{U}(x))$; $F(x)$ is uniquely defined because Y is Hausdorff. To prove that F is continuous at x, let W be any nbd of $F(x)$; because Y is regular, there is an open V with $F(x) \in V \subset \overline{V} \subset W$, and since $f(D \cap \mathfrak{U}(x)) \to F(x)$ there is a $U = U(x)$ with $f(U \cap D) \subset V$. Now, $\mathfrak{B}_z = V \cap f(D \cap \mathfrak{U}(z))$ is a filterbase for each $z \in U$: Indeed, each $U \cap U(z)$ is open, so that $D \cap U \cap U(z) \neq \varnothing$, and thus

$$f(D \cap U \cap U(z)) \subset V \cap f(D \cap U(z)).$$

Because $\mathfrak{B}_z \vdash f(D \cap \mathfrak{U}(z))$, we find that $\mathfrak{B}_z \to F(z)$, and since $\mathfrak{B}_z$ is in V, that $F(z) \in \overline{V}$. Thus $F(U) \subset \overline{V} \subset W$, and continuity has been proved.

Necessity: This is immediate from **5.1**, since $D \cap \mathfrak{U}(x) \vdash \mathfrak{U}(x)$, for each x, and $F \mid D = f$. That F is uniquely determined by f follows from VII, **1.5**.

For filterbases in arbitrary cartesian products,

5.4 Theorem A filterbase $\mathfrak{A}$ on $\prod_\gamma Y_\gamma$ converges to $\{y_\gamma^\circ\}$ if and only if $p_\gamma(\mathfrak{A}) \to y_\gamma^\circ$ for each fixed γ.

Proof: Necessity is clear, since each projection map p_γ is continuous. Conversely, let $\langle U_{\gamma_1}, \cdots, U_{\gamma_n} \rangle$ be any basic nbd of $\{y_\gamma^\circ\}$. For each $i = 1, \cdots, n$ the convergence of $p_{\gamma_i}(\mathfrak{A})$ to $y_{\gamma_i}^\circ$ yields an $A_i \in \mathfrak{A}$ such that $p_{\gamma_i}(A_i) \subset U_{\gamma_i}$; by **2.2**(3), there is an $A \in \mathfrak{A}$ such that $A \subset \bigcap_1^n A_i$, and clearly, $A \subset \langle U_{\gamma_1}, \cdots, U_{\gamma_n} \rangle$.

Ex. 1 The statement of **5.4** for sequences is the most commonly used version: explicitly, a sequence $\{y_\gamma^n\}$ in $\prod_\gamma Y_\gamma$ converges to $\{y_\gamma^\circ\}$ if and only if each coordinate $y_\gamma^n \to y_\gamma^\circ$.

6. Adequacy of Sequences

We now show that, in 1° countable spaces, sequences and subsequences not only behave properly, but also are adequate to express all topological concepts; this accounts for their great utility in the metric spaces of elementary analysis.

6.1 Let $\varphi: Z^+ \to X$ be a sequence. Then:

(1). $\varphi \to y_0$ if and only if each subsequence φ' of φ contains a subsequence φ'' such that $\varphi'' \to y_0$.

(2). Let X be 1° countable. Then $\varphi \succ x_0$ if and only if there is some subsequence $\varphi' \to x_0$.

Proof: (1). The "only if" is trivial. "If": Assume that $\varphi \not\to x_0$; then $\exists\, U(x_0)\ \forall\, T_n$: $\varphi(T_n) \not\subset U$. Proceeding by induction, let $n_1 \geqslant 1$ be the first integer in T_1 such that $\varphi(n_1) \,\bar{\in}\, U$, and assuming that $n_1 < \cdots < n_k$ have been obtained, let n_{k+1} be the first integer in T_{n_k+1} such that $\varphi(n_{k+1}) \,\bar{\in}\, U$. Let φ' be the subsequence defined by $\varphi'(k) = \varphi(n_k)$. Since $\varphi'(T_n) \subset \mathscr{C}U$ for every n, no subsequence of φ' can converge to y_0.

(2). Only the existence of φ' requires proof. Let $U_1 \supset U_2 \supset \cdots$ be a countable basis at x_0 and assume that $\varphi \succ x_0$; then $\forall\, U_i\ \forall\, T_m$: $\varphi(T_m) \cap U_i \neq \varnothing$. Proceeding by induction, let $n_1 \geqslant 1$ be the first integer in T_1 with $\varphi(n_1) \in U_1$, and assuming that $n_1 < \cdots < n_k$ have been defined, let n_{k+1} be the first integer in T_{n_k+1} with $\varphi(n_{k+1}) \in U_{k+1}$. The subsequence $\varphi'(k) = \varphi(n_k)$ evidently converges to x_0.

According to **6.1**(2), sequences in 1° countable spaces cannot display the pathological behavior of **3**, Ex. 1; that they are in fact capable of expressing all topological concepts in such spaces follows from

6.2 Theorem Let X be 1° countable and let $A \subset X$. Then $x \in \overline{A}$ if and only if there is a *sequence* on A converging to x.

Proof: Because of **4.1**, it is enough to prove: For each $x_0 \in X$ and each filterbase $\mathfrak{B} \to x_0$, there is a sequence φ lying on $\bigcup\{B \mid B \in \mathfrak{B}\}$ such that $\varphi \to x_0$. To this end, let $U_1 \supset U_2 \supset \cdots$ be a countable basis at x_0. For each U_ι, find a $B_{\alpha_\iota} \subset U_\iota$ and choose a $b_{\alpha_\iota} \in B_{\alpha_\iota}$; then the sequence φ defined by $\varphi(i) = b_{\alpha_\iota}$ evidently converges to x_0, and the proof is complete.

For maps of 1° countable spaces, we have

6.3 Let X be 1° countable. Then for Y arbitrary, and $f: X \to Y$:

(1). $f(\mathfrak{U}(x))$ converges to y_0 if and only if $f(x_n) \to y_0$ for each sequence $x_n \to x$.

(2). f is continuous at x_0 if and only if $f(x_n) \to f(x_0)$ for each sequence $x_n \to x_0$.

(3). Let Y be regular, $D \subset X$ dense, and $g: D \to Y$ continuous. Then g is extendable to a continuous $G: X \to Y$ if and only if for each $x \in X$ and all sequences $\{d_n\} \subset D$ converging to x, the sequences $\{f(d_n)\}$ all converge and to the same limit.

Proof: (1). Assume $f(\mathfrak{U}(x)) \to y_0$ and let the sequence $\varphi \to x$; since $\mathfrak{A}(\varphi) \vdash \mathfrak{U}(x)$, we find $f(\mathfrak{A}(\varphi)) \to y_0$ as required. For the converse, first note that if $\mathfrak{B}: U_1 \supset U_2 \supset \cdots$ is a countable basis at x, then $\mathfrak{B}$ is a filterbase and $\mathfrak{B} \vdash \mathfrak{U}(x) \vdash \mathfrak{B}$, so that $f(\mathfrak{U}(x))$ converges if and only if $f(\mathfrak{B})$ does. Now assume that $f(\mathfrak{B})$ does not converge to y_0; this means that $\exists\, W(y_0)\, \forall\, U_\iota: f(U_\iota) \not\subset W$; choosing $x_\iota \in U_\iota$ so that $f(x_\iota) \bar{\in} W$, we find a sequence $x_n \to x_0$ such that $f(x_n)$ does not converge to y_0, completing the proof. Proofs of (2) and (3) follow at once from (1) and from **5.1** and **5.3** respectively.

7. Maximal Filterbases

7.1 Definition A filterbase $\mathfrak{M}$ in Y is called maximal (or an ultrafilter base) if it has no properly subordinated filterbase; that is, if

$$\forall\, \mathfrak{A} : \mathfrak{A} \vdash \mathfrak{M} \Rightarrow \mathfrak{M} \vdash \mathfrak{A}.$$

Maximal filterbases are characterized in

7.2 The filterbase $\mathfrak{M}$ in Y is maximal if and only if for each $A \subset Y$, one of the two sets A, $\mathscr{C}A$, contains a member of $\mathfrak{M}$.

Proof: Assume $\mathfrak{M} = \{M_\beta \mid \beta \in \mathscr{B}\}$ is a maximal filterbase. Clearly, we cannot have an $M_\beta \subset A$ and an $M_\gamma \subset \mathscr{C}A$, since then $M_\beta \cap M_\gamma = \varnothing$. Assume now that $\forall\, M_\beta: M_\beta \not\subset A$; then all $M_\beta \cap \mathscr{C}A \neq \varnothing$, so $\mathfrak{A} = \mathfrak{M} \cap \mathscr{C}A$ is a filterbase. Since $\mathfrak{A} \vdash \mathfrak{M}$, also $\mathfrak{M} \vdash \mathfrak{A}$; consequently, using any $M_\beta \cap \mathscr{C}A$, there is an $M_\gamma \subset M_\beta \cap \mathscr{C}A \subset \mathscr{C}A$. Conversely, assume that the condition is satisfied and that $\mathfrak{A} \vdash \mathfrak{M}$; given any $A_\alpha \in \mathfrak{A}$, the condition assures that either there is an $M_\beta \subset A_\alpha$ or an $M_\beta \subset \mathscr{C}A_\alpha$; the latter possibility is excluded, since the assumption $\mathfrak{A} \vdash \mathfrak{M}$ implies [**2.5**(2)] that all $M_\beta \cap A_\alpha \neq \varnothing$. Thus $\mathfrak{M} \vdash \mathfrak{A}$ and $\mathfrak{M}$ is maximal.

Maximal filterbases exist: In any space Y, choose $y_0 \in Y$; then the filterbase $\mathfrak{M} = \{M \subset Y \mid y_0 \in M\}$ is maximal, since it satisfies **7.2.** More important, a maximal filterbase subordinate to any given filterbase can always be found:

7.3 Theorem Let $\mathfrak{B}$ be any filterbase in Y. Then there exists a maximal filterbase $\mathfrak{M} \vdash \mathfrak{B}$.

Proof: Let $\mathscr{A}$ be the family of all filterbases $\mathfrak{A} \vdash \mathfrak{B}$; $\mathscr{A}$ is not empty, since $\mathfrak{B} \in \mathscr{A}$. Preorder $\mathscr{A}$ by $\mathfrak{A}' \prec \mathfrak{A}$ if $\mathfrak{A} \vdash \mathfrak{A}'$. To prepare for an application of Zorn's lemma, consider any chain $\{\mathfrak{A}_\mu\}$ in $\mathscr{A}$ and let $\mathfrak{A} = \bigcup_\mu \mathfrak{A}_\mu$. We first note that $\mathfrak{A}$ is a filterbase: if $A, C \in \mathfrak{A}$ are given, say $A \in \mathfrak{A}_\mu$, $C \in \mathfrak{A}_\lambda$ and $\mathfrak{A}_\mu \vdash \mathfrak{A}_\lambda$, then $\exists\, B \in \mathfrak{A}_\mu: B \subset C$ and so there is an $E \in \mathfrak{A}_\mu \subset \mathfrak{A}$ such that $E \subset A \cap B \subset A \cap C$. Next, by using **2.5**(1), it is evident that $\mathfrak{A} \vdash \mathfrak{A}_\mu \vdash \mathfrak{B}$ for each $\mathfrak{A}_\mu$; this shows that $\mathfrak{A} \in \mathscr{A}$ and that $\mathfrak{A}$ is an upper bound for $\{\mathfrak{A}_\mu\}$. Thus, the use of Zorn's lemma is legitimate, and we conclude that there exists a maximal $\mathfrak{M} \in \mathscr{A}$; the filterbase $\mathfrak{M}$ clearly satisfies the requirements of the theorem.

For convergence properties,

7.4 Let $\mathfrak{M}$ be a maximal filterbase in Y. Then $\mathfrak{M} \succ y_0$ if and only if $\mathfrak{M} \to y_0$.

Proof: Only the implication $(\mathfrak{M} \succ y_0) \Rightarrow (\mathfrak{M} \to y_0)$ need be proved. Given $U(y_0)$, there is an $M_\alpha \subset U$ or an $M_\alpha \subset \mathscr{C}U$; since $\mathfrak{M} \succ y_0$, so that $\forall\, M_\alpha: M_\alpha \cap U \neq \varnothing$, the latter possibility is excluded, and therefore $\mathfrak{M} \to y_0$.

Maximality is preserved under arbitrary maps:

7.5 Let $\mathfrak{M}$ be a maximal filterbase on X and $p: X \to Y$ any map. Then $p(\mathfrak{M})$ is a maximal filterbase in Y.

Proof: We show that **7.2** holds for $p(\mathfrak{M})$. Given $A \subset Y$, we consider $p^{-1}(A)$ and $p^{-1}(\mathscr{C}A) = \mathscr{C}p^{-1}(A)$; since $\mathfrak{M}$ is maximal, there is either an $M_\alpha \subset p^{-1}(A)$ or an $M_\alpha \subset \mathscr{C}p^{-1}(A)$, and the conclusion follows.

Problems

Section 1

1. Let (Y, d) be a metric space, and $\{y_n\}$ a sequence in Y. Prove: $y_n \to y_0$ if and only if $d(y_n, y_0) \to 0$.

Section 2

1. Let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a filterbase in X, and $\mathfrak{B} = \{B_\beta \mid \beta \in \mathscr{B}\}$ a filterbase in Y. Show $\mathfrak{A} \times \mathfrak{B} = \{A_\alpha \times B_\beta \mid (\alpha, \beta) \in \mathscr{A} \times \mathscr{B}\}$ is a filterbase in $X \times Y$.
2. Let $f: X \to Y$ be a map, and $\mathfrak{B}$ a filterbase in Y. Prove that
$$f^{-1}(\mathfrak{B}) = \{f^{-1}(B) \mid B \in \mathfrak{B}\}$$
is a filterbase in X if and only if $f^{-1}(B) \neq \varnothing$ for each $B \in \mathfrak{B}$.
3. Let $f: X \to Y$ be a map and $\mathfrak{A}$ a filterbase in X. Prove: $f(\mathfrak{A}) = \{f(A) \mid A \in \mathfrak{A}\}$ is a filterbase in Y.
4. Let $f: X \to Y$ be a map, $\mathfrak{A}$ a filterbase in X and $\mathfrak{B}$ a filterbase in Y. Prove:
   a. $f^{-1} \circ f(\mathfrak{A})$ is a filterbase in X and $\mathfrak{A} \vdash f^{-1}f(\mathfrak{A})$.
   b. If $f^{-1}(\mathfrak{B})$ is a filterbase in X, then $f \circ f^{-1}(\mathfrak{B}) \vdash \mathfrak{B}$.
5. Prove that the set of accumulation points of any filterbase is always a closed (possibly empty) set.

Section 3

1. Let Y be Hausdorff, let $\mathfrak{A} \to y_0$ and let $\mathfrak{B} \vdash \mathfrak{A}$. Prove: If $\mathfrak{B}$ accumulates at any point y, then $y = y_0$.
2. If the filterbase $\mathfrak{A}$ does not converge to y_0, show that there is a filterbase $\mathfrak{B} \vdash \mathfrak{A}$ such that no filterbase $\mathfrak{C} \vdash \mathfrak{B}$ converges to y_0.

Section 4

1. Prove the following analog of the usual diagonal process: Let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a filterbase on Y, with $\mathfrak{A} \to y_0$. Assume that for each $\alpha \in \mathscr{A}$, there is a filterbase $\mathfrak{B}_\alpha = \{B_{\alpha,\beta} \mid \beta \in \mathscr{B}_\alpha\}$ with $\mathfrak{B}_\alpha \succ a_\alpha \in A_\alpha$. Then there is a filterbase on $\bigcup_{\alpha,\beta} B_{\alpha,\beta}$ converging to y_0.
2. Prove that $A \subset Y$ is closed if and only if each convergent filterbase on A converges to a point of A.

Section 5

1. Let $f: X \to Y$ be an open surjection. Prove: For each $x \in X$ and filterbase $\mathfrak{B} \to f(x)$, there is a filterbase $\mathfrak{A} \to x$ such that $f(\mathfrak{A}) \vdash \mathfrak{B}$ and $\mathfrak{B} \vdash f(\mathfrak{A})$.

Section 6

1. Let Y be 1° countable. Prove: $A \subset Y$ is closed if and only if the accumulation points of each sequence on A all lie in A.

Section 7

1. Let $\mathfrak{M}$ be a maximal filterbase in Y, and let $A, B \subset Y$ be disjoint and such that $A \cup B \in \mathfrak{M}$. Prove: Either $A \in \mathfrak{M}$ or $B \in \mathfrak{M}$.

Compactness

XI

In this chapter, we consider spaces having a strengthened version of the Lindelöf property; such spaces play an important role in all branches of mathematics.

1. Compact Spaces

1.1 Definition A Hausdorff space Y is compact if each open covering has a finite subcovering.

Ex. 1 A discrete space is compact if and only if it is finite. The proof in VIII, **2**, Ex. 2, shows the ordinal space $[0, \Omega]$ is compact; note that a compact space need not be even 1° countable.

Ex. 2 In any space X, all finite subsets, and $\varnothing$, are compact subsets. If $\varphi: Z^+ \to X$ is a sequence convergent to x_0, then $A = x_0 \cup \varphi(Z^+)$ is a compact subset of X: any set of an open covering of A that contains x_0 contains all but at most finitely many elements of A. Observe that an infinite subset of A that does not contain x_0 is not a compact set.

Ex. 3 E^1 is not compact, since the open covering $]-n, n[$, $n = 1, 2, \cdots$ has no finite subcovering. However, each closed finite interval $[a, b]$ is compact: In fact, given any open covering $\{U_\alpha\}$ of $[a, b]$, let $c = \sup \{x \mid [a, x]$ can be covered by finitely many $U_\alpha\}$; if $c < b$, we derive a contradiction to the definition of c by choosing any $U_\alpha \supset c$, observing that there is a $B(c, r) \subset U_\alpha$, and that since $[a, c - (r/2)]$ can be covered by finitely many sets $U_{\alpha_1}, \cdots, U_{\alpha_n}$, these sets together with U_α are a finite open covering of $[a, c + (r/2)]$.

The position of compact spaces is

1.2 Theorem Every compact space is paracompact (hence, also normal).

Proof: A finite subcovering of a covering is evidently a nbd-finite refinement of the given covering.

Ex. 4 The converse of **1.2** is false, as an uncountable discrete space shows.

Ex. 5 Though metric spaces are also among the paracompact spaces, compactness and metrizability are not related: E^1 is metrizable, but not compact, and $[0, \Omega]$ is compact, but not metrizable.

Ex. 6 To emphasize the difference between paracompact and compact spaces, we have: Y is compact if and only if each open covering has a nbd-finite *subcovering* (rather than nbd-finite *refinement*). If Y is compact, the condition is evident; conversely, given any open covering $\{U_\alpha\}$, choose $U_{\alpha_0} \neq \varnothing$ and consider the open covering by the sets $V_\alpha = U_{\alpha_0} \cup U_\alpha$; a nbd-finite subcovering of $\{V_\alpha\}$ must evidently be finite.

The definition has several equivalent formulations:

1.3 Theorem The following four properties are equivalent:

(1). Y is compact.
(2). The finite intersection property: For each family $\{F_\alpha \mid \alpha \in \mathscr{A}\}$ of closed sets in Y satisfying $\bigcap_\alpha F_\alpha = \varnothing$, there is a *finite* subfamily $F_{\alpha_1}, \cdots, F_{\alpha_n}$ with $\bigcap_1^n F_{\alpha_i} = \varnothing$.
(3). Each filterbase in Y has at least one accumulation point.
(4). Each maximal filterbase in Y converges.

Proof: (1) $\Leftrightarrow$ (2). These two statements are de Morgan duals: The assertion that an intersection of closed sets, $\bigcap_\beta B_\beta$ is empty is equivalent to the assertion that $\bigcup_\beta \mathscr{C}B_\beta = Y$, that is, that the complementary open sets cover Y.

(2) $\Rightarrow$ (3). Let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a filterbase in Y; since all finite intersections of the A_α are nonempty, so also are all finite intersections of the $\bar{A}_\alpha$, and therefore (2) assures $\bigcap_\alpha \bar{A}_\alpha \neq \varnothing$. Thus, the set of accumulation points of $\mathfrak{A}$ is not empty.

(3) $\Rightarrow$ (4). Since $\mathfrak{M} \succ y_0$, and $\mathfrak{M}$ is a maximal filterbase, $\mathfrak{M} \to y_0$.

(4) $\Rightarrow$ (3). Given $\mathfrak{A}$, there is a maximal filterbase $\mathfrak{M} \vdash \mathfrak{A}$; since $\mathfrak{M} \to y_0$, we find $\mathfrak{A} \succ y_0$.

(3) $\Rightarrow$ (2). Let $\{F_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of closed sets, and assume that each finite intersection is nonempty. The sets F_α, together with

their finite intersections, then form a filterbase $\mathfrak{A}$ in Y. Since $\mathfrak{A}$ accumulates at some $y_0 \in Y$, we have $y_0 \in \bigcap_\alpha \overline{F}_\alpha = \bigcap_\alpha F_\alpha$, and therefore the intersection is not empty.

Ex. 7 E^1 is not compact. Observe that the sets $F_n = \{x \mid x \geqslant n\}$ are closed and each finite family has a nonempty intersection; yet, $\bigcap_n F_n = \varnothing$.

For invariance properties, we have

1.4 Theorem (1). The continuous image of a compact set is compact.

(2). A compact subset A of a Hausdorff space X is closed in X; indeed, for each $x \bar{\in} A$, there are nonintersecting nbds $U(A)$, $U(x)$.

(3). A subspace of a compact space is compact if and only if it is closed.

(4). **(A. Tychonoff)** Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces. Then $\prod_\alpha Y_\alpha$ is compact if and only if each Y_α is compact.

Proof: (1). Let Y be compact, and $f: Y \to Z$ continuous. Let $\{U_\alpha\}$ be any open covering of $f(Y)$; then $\{f^{-1}(U_\alpha)\}$ is an open covering of Y and so can be reduced to a finite covering, $f^{-1}(U_{\alpha_1}), \cdots, f^{-1}(U_{\alpha_n})$; it is evident that $U_{\alpha_1}, \cdots, U_{\alpha_n}$ is a finite covering of $f(Y)$.

(2). To show that $\mathscr{C}A$ open, we prove that each fixed $x_0 \in \mathscr{C}A$ has a nbd lying in $\mathscr{C}A$. For each $a \in A$, find disjoint nbds $U(a)$, $U_a(x_0)$. Since $\{U(a) \cap A \mid a \in A\}$ is an open covering of A, reduce it to a finite covering $U(a_1) \cap A, \cdots, U(a_n) \cap A$; then $U(A) = \bigcup_1^n U(a_i)$ and $U(x_0) = \bigcap_1^n U_{a_i}(x_0)$ are disjoint open sets.

(3). Because of (2), we need show only that if Y is compact and $A \subset Y$ is closed, then A is compact. Let $\{B_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of sets closed in A with $\bigcap_\alpha B_\alpha = \varnothing$; since A is closed in Y, the B_α are also closed in Y, so $\bigcap_1^n B_{\alpha_i} = \varnothing$ for some finite subfamily, and by **1.3**(2), A is therefore compact.

(4). If $\prod_\alpha Y_\alpha$ is compact, then because each projection $p_\beta: \prod Y_\alpha \to Y_\beta$ is a continuous surjection, we find from (1) that each Y_α is compact. Conversely, assume that each Y_α is compact, and let $\mathfrak{M}$ be a maximal filterbase in $\prod_\alpha Y_\alpha$; the image of a maximal filterbase being also a maximal filterbase, each $p_\alpha(\mathfrak{M})$ converges to some y_α; by X, **5.4**, we find that $\mathfrak{M} \to \{y_\alpha\}$; and so, by **1.3**(4), $\prod_\alpha Y_\alpha$ is compact.

Ex. 8 The extended real line $\tilde{E}^1$ (III, **7**, Ex. 2) is compact, since it is homeomorphic to $[-1, +1]$.

Ex. 9 If X is not Hausdorff, **1.4**(2) may be false: In Sierpinski space, the subspace $\{0\}$ is compact, but it is not closed.

Ex. 10 The Cantor set (I, **9**, Ex. 3) and, more generally, any cartesian product of finite discrete spaces is compact.

Ex. 11 Since the cartesian product of any family of closed unit intervals is compact, we have that (a) the Hilbert cube I^∞ is compact and (b) a space is completely regular if and only if it is a subset of a compact Hausdorff (hence, normal) space (*cf.* VII, **7.3**).

In (Hausdorff!) spaces, the compact subsets behave as points do and have the same separation properties:

1.5 (a). A finite union of compact subsets of a Hausdorff space is compact.
(b). Two disjoint compact subspaces of a Hausdorff space have disjoint nbds.
(c). If A is a compact subset of a regular space, then for each open $U \supset A$, there is an open V with $A \subset V \subset \bar{V} \subset U$.
(d). Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be a nbd-finite open covering of a space X. If $A \subset X$ is compact, then A has a nbd meeting at most finitely many sets U_α (in precise terms: there exists a nbd U of A such that $U \cap U_\alpha \neq \varnothing$ for at most finitely many indices α).

Proof: (a) is trivial.

(b). Let A, B be compact subsets of X, and let $A \cap B = \varnothing$. From **1.4**(2), for each $b \in B$ there are disjoint nbds $U_b(A)$, $U(b)$. From the open covering $\{U(b) \cap B \mid b \in B\}$ of B, extract a finite subcovering $U(b_1) \cap B, \cdots, U(b_n) \cap B$; then $\bigcup_1^n U(b_i)$, $\bigcap_1^n U_{b_i}(A)$ are the required nbds.

(c). For each $a \in A$ there is a nbd $V(a)$ such that $\overline{V(a)} \subset U$; extracting a finite subcovering gives

$$A \subset \bigcup_1^n V(a_i) \subset \bigcup_1^n \overline{V(a_i)} \subset U.$$

(d). Each $a \in A$ has a nbd $U(a)$ meeting at most finitely many U_α; extract a finite subcovering $U(a_1), \cdots, U(a_n)$; since each $U(a_i)$ meets only finitely many U_α, so also does $\bigcup_1^n U(a_i)$.

Families of compact subsets of a space have the useful property:

1.6 Let X be Hausdorff, let ω_α be an initial ordinal, and let $\{F_\mu \mid \mu < \omega_\alpha\}$ be a descending family of nonempty compact subsets; that is, $\mu < \nu \Rightarrow F_\mu \supset F_\nu$. Then:

(a). $\bigcap_{\mu} F_\mu = C$ is compact and not empty.

(b). Given any open $U \supset C$, there is a $\mu_0 < \omega_\alpha$ such that $F_\mu \subset U$ for all $\mu \geqslant \mu_0$.

Proof: For the proofs of (a) and (b) we can assume that X is compact, otherwise we work in F_1.

(a). Since each finite intersection of the F_μ is nonempty, so also is $\bigcap_{\mu} F_\mu = C$; since each F_μ is closed in the closed F_1, C is closed in X.

(b). From $\bigcap_{\mu} F_\mu \subset U$, we find $\mathscr{C}U \subset \bigcup_{\mu} \mathscr{C}F_\mu$; since $\mathscr{C}U$ is closed, and hence is compact, extract a finite covering $\bigcup_{1}^{n} \mathscr{C}F_{\mu_i}$; then, for any $\mu > \mu_0 = \max(\mu_1, \cdots, \mu_n)$, we have

$$F_\mu \subset F_{\mu_0} = \bigcap_{1}^{n} F_{\mu_i} \subset U.$$

2. Special Properties of Compact Spaces

Some properties of compact spaces that are frequently used, and which contribute to their importance, are given in this section.

We have seen that a continuous bijection need not be a homeomorphism; one of the important features for maps of compact spaces is

2.1 Theorem Let Y be compact, Z be Hausdorff, and $f\colon Y \to Z$ continuous. Then:

(1). f is a closed map.

(2). If f is a continuous bijection, then f is a homeomorphism.

Proof: (1). Let $A \subset Y$ be closed; it is compact and consequently so is $f(A)$. Since Z is Hausdorff, $f(A)$ is closed in Z. (2) is an immediate consequence [*cf.* III, **12.2**].

Ex. 1 A continuous map of a compact space into a Hausdorff space need not be *open*, even if it is injective. Let $Y = [0, 1] \cup \{2\}$, $Z = [0, 1]$, and $f\colon Y \to Z$ be the map $y \to \frac{1}{2}y$. Y is a closed subset of $[0, 2]$ and thus is compact; f is not open, since the open set $\{2\}$ does not have open image.

Ex. 2 The hypothesis that Z be Hausdorff is essential, as $1\colon 2 \to \mathscr{S}$, the Sierpinski space, shows.

Ex. 3 The delicate position of a compact Hausdorff topology $\mathscr{T}$ in a set Y is shown by its two properties:

(a). If $\mathscr{T}$ is a proper subset of $\mathscr{T}_+$, then $(Y, \mathscr{T}_+)$ is not compact.

(b). If $\mathscr{T}_-$ is a proper subset of $\mathscr{T}$, then $(Y, \mathscr{T}_-)$ is not Hausdorff.

For, in case (a), the continuous $1\colon (Y, \mathscr{T}_+) \to (Y, \mathscr{T})$ would be a homeomorphism if $\mathscr{T}_+$ were compact; in case (b), the continuous $1\colon (Y, \mathscr{T}) \to (Y, \mathscr{T}_-)$ would be a homeomorphism if $\mathscr{T}_-$ were Hausdorff.

Ex. 4 We have seen (IV, **4.4**) that there is a continuous surjection $I \to I^n$, $2 \leqslant n \leqslant \infty$; according to **2.1**, there can be no continuous bijection $I \to I^n$, since we know that I and I^n are not homeomorphic.

2.2 Corollary Let $p: X \to Y$ be an identification, and let $h: X \to Z$ be continuous. If X is compact, if Z is Hausdorff, and if $hp^{-1}: Y \to Z$ is bijective, then hp^{-1} is a homeomorphism.

Proof: According to VI, **3.2**, hp^{-1} is a continuous bijection, so we need to show only that Y is compact. Now, because Z is Hausdorff, VII, **1.5** (4) shows that Y is also Hausdorff and therefore, since p is surjective, Y is indeed compact.

Ex. 5 The cone TS^n over S^n (VI, **5.1**) is homeomorphic to the ball V^{n+1}. For, let $p: S^n \times I \to TS^n$ be the identification map, and $h: S^n \times I \to V^{n+1}$ the map $h(x, t) = (1 - t)x$; then hp^{-1} is bijective, and since $S^n \times I$ is compact, **2.2** applies. Similarly, the suspension of S^n is homeomorphic to S^{n+1}.

The next general feature is

2.3 Theorem Let Y be compact and $f: Y \to E^1$ be continuous. Then f attains its supremum and its infimum, and both are finite. Precisely, there is at least one $y_0 \in Y$ with $f(y_0) = \sup \{f(y) \mid y \in Y\}$ and at least one $y_1 \in Y$ with $f(y_1) = \inf \{f(y) \mid y \in Y\}$.

This is a consequence of the more general

2.4 Let Y be compact, and $f: Y \to E^1$ a map. If f is lower (upper) semicontinuous, then it attains its infimum (supremum).

Proof: Assume f to be lower semicontinuous, and let $m = \inf f(Y)$. For each $q > m$, the set $F_q = \{y \mid f(y) \leqslant q\}$ is closed because f is lower semicontinuous; and it is not empty, by the definition of m. Since the intersection of any finite family of the sets F_q is not empty, the compactness of Y assures $\bigcap_{q>m} F_q \neq \varnothing$; and for $y_0 \in \bigcap_{q>m} F_q$, we clearly have $f(y_0) = m$. The result for upper semicontinuous f follows from this by noting that $-f$ is then lower semicontinuous.

In cartesian products, we know that each projection is an open map and that it need not be a closed map. We show now that projections "parallel to compact factors" are also closed maps.

2.5 Theorem Let Y be compact, Z Hausdorff, and $p: Y \times Z \to Z$ the projection "parallel to the compact factor Y." Then p is a closed map.

Proof: Let $F \subset Y \times Z$ be closed; we show that $Z - p(F)$ is open. Let $z_0 \in Z - p(F)$; then $(Y \times z_0) \cap F = \varnothing$, so that each point (y, z_0) has a nbd $U(y) \times U_y(z_0)$ not intersecting F. From this open covering of the compact $Y \times z_0$, extract a finite covering $U(y_i) \times U_{y_i}(z_0)$, $i = 1, \cdots, n$; then $\bigcap_1^n U_{y_i}(z_0)$ is a nbd of z_0 that does not meet $p(F)$.

The most frequently used version of this result is

2.6 Corollary Let $A \subset X$ be arbitrary and let Y be compact. Let U be a nbd of $A \times Y$ in $X \times Y$. Then there is a nbd $V \supset A$ such that the tube $V \times Y \subset U$.

Proof: Let $p: X \times Y \to X$ be the projection; since p is closed and $p^{-1}(A) = A \times Y \subset U$, the result follows immediately from III, **11.2**(1).

We have seen that a continuous map into a Hausdorff space has a closed graph; for maps into compact spaces,

2.7 Theorem Let X be Hausdorff and Y be compact. Then $f: X \to Y$ is continuous if and only if its graph $G(f)$ is closed in $X \times Y$.

Proof: We need show only that if $G(f)$ is closed in $X \times Y$, then $f: X \to Y$ is continuous. Let $B \subset Y$ be closed; then $p_Y^{-1}(B) \cap G(f)$ is closed in the closed $G(f)$ and therefore in $X \times Y$. The projection p_X parallel to the compact factor Y is a closed map, and since

$$p_X[p_Y^{-1}(B) \cap G(f)] = f^{-1}(B),$$

this establishes the continuity of f.

3. Countable Compactness

A generalization of the compactness concept will be considered in this section.

3.1 Definition A Hausdorff space is countably compact if every *countable* open covering has a finite subcovering.

Ex. 1 Every compact space is evidently countably compact, but the converse is not true. The ordinal space $[0, \Omega[$ is not compact, since it is not a closed subset of $[0, \Omega]$ (or, alternatively, since it is not paracompact). We now show that $[0, \Omega[$ is in fact countably compact. Let $\{U_n \mid n \in Z^+\}$ be any open covering of $[0, \Omega[$; observe that for each $\beta < \Omega$, the interval $[0, \beta]$ is compact, since it is closed in $[0, \Omega]$, and therefore can be covered by finitely many of the U_i. Now, if there were no finite subcovering for $[0, \Omega[$, then for each $n \in Z^+$ we could find an element $\alpha_n \notin U_1 \cup \cdots \cup U_n$; since the α_n have an upper bound $\alpha_0 < \Omega$, we would have an interval $[0, \alpha_0]$ that cannot be covered by finitely many of the U_i.

Countable compactness is characterized by the behavior of sequences only, rather than arbitrary filterbases, as the following equivalent formulations show:

3.2 Theorem The following three properties are equivalent:

(1). Y is countably compact.

(2). (**Bolzano-Weierstrass property.**) Every countably infinite subset of Y has at least one cluster point.

(3). Every sequence in Y has an accumulation point.

Proof: (1) $\Rightarrow$ (2). Assume $A \subset Y$ were a countably infinite subset with no cluster point. Then A would be a closed discrete set, and so each point $a_i \in A$ would have an nbd U_i not containing any other member of A. Thus, $\{U_i \mid i \in Z^+\} \cup \{\mathscr{C}A\}$ would be a countable open covering of Y that has no finite subcovering.

(2) $\Rightarrow$ (3). Let $\varphi: Z^+ \to Y$ be a sequence in Y. If $\aleph[\varphi(Z^+)] = \aleph_0$, then the set $\varphi(Z^+)$ has a cluster point y_0; since each nbd $U(y_0)$ contains infinitely many points of $\varphi(Z^+)$, we have $U(y_0) \cap \varphi(T_m) \neq \varnothing$ for each $m \in Z^+$, so $\varphi \succ y_0$. If $\aleph[\varphi(Z^+)] < \aleph_0$, then there is some y_0 such that $\varphi(n) = y_0$ for infinitely many n, so again $\varphi \succ y_0$.

(3) $\Rightarrow$ (1). Assume that Y is not countably compact. Then there is a countable open covering $\{U_i \mid i \in Z^+\}$ that cannot be reduced to a finite covering. For each $n \in Z^+$ we can therefore find some $y_n \in Y - \bigcup_1^n U_i$; then $n \to y_n$ defines a sequence φ in Y that does not have an accumulation point: each $y \in Y$ belongs to some $U_{n(y)}$ and $\varphi(T_{n(y)}) \cap U_{n(y)} = \varnothing$.

Countable compactness reduces to compactness in a class of spaces that includes the paracompact spaces. Calling a space *metacompact* if each open covering has a point-finite open *refinement*, we have

3.3 Theorem (R. Arens; J. Dugundji) A space is compact if and only if it is both countably compact and metacompact.

Proof: We need prove only that if Y is countably compact and metacompact, then Y is compact. Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be any open covering of Y. By the metacompactness, $\{U_\alpha\}$ has a point-finite open refinement $\{V_\beta \mid \beta \in \mathscr{B}\}$ and, by VIII, **1.1**, $\{V_\beta\}$ has an irreducible subcovering $\{V_\gamma \mid \gamma \in \mathscr{G}\}$. This minimal covering must be finite: for we can find in each V_γ a point y_γ belonging to no set other than V_γ and, if $\{V_\gamma \mid \gamma \in \mathscr{G}\}$

were not finite, then $\{y_\gamma \mid \gamma \in \mathscr{G}\}$ would be an infinite set in Y with no accumulation point. Thus, by choosing for each V_γ some set $U_{\alpha(\gamma)} \supset V_\gamma$, we reduce the covering $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ to a finite covering.

3.4 Corollary Countable compactness is equivalent to compactness in (a): paracompact spaces and in (b): arbitrary Lindelöf spaces.

Proof: (a) is immediate from **3.3**. (b) Given any open covering, the Lindelöf property gives a countable subcovering, and then the countable compactness permits a further reduction to a finite subcovering.

We now study the topological properties of countably compact spaces. Since countable compactness is based entirely on sequences, we can expect it to impose some topological restrictions in the 1° countable spaces. Thus, although countably compact nonregular (Hausdorff) spaces exist, we do have

3.5 A countably compact 1° countable space Y is regular.

Proof: Let $y \in Y$ and a nbd V of y be given, and let $V_1 \supset V_2 \supset \cdots$ be a countable nbd basis at y. From VII, **1.2**(3), it follows that $\bigcap_1^\infty \bar{V}_n = \{y\}$, consequently $\{V\} \cup \{\mathscr{C}\bar{V}_n \mid n \in Z^+\}$ is a countable open covering of Y. Because this covering has a finite subcovering, there is some integer n such that $Y = V \cup \bigcup_1^n \mathscr{C}\bar{V}_\iota = V \cup \mathscr{C} \bigcap_1^n \bar{V}_\iota$; therefore $\bar{V}_n = \bigcap_1^n \bar{V}_\iota \subset V$, and the proof is complete.

For the invariance properties we have

3.6 Theorem (1). The continuous image of a countably compact space is countably compact.

(2). A closed subspace of a countably compact space is countably compact.

(3). If X is 1° countable, and $A \subset X$ is countably compact, then A is closed in X.

(4). The cartesian product of two countably compact spaces need not be countably compact. However, if each X_ι, $\iota \in Z^+$, is 1° countable, then $\prod_1^\infty X_\iota$ is countably compact if and only if each X_ι is countably compact.

Proof: The proofs of (1) and (2) are left for the reader.

(3). Let $x \in \bar{A}$; because X is 1° countable, there is a sequence φ lying on A and converging to x. Since A is countably compact, φ has an accumulation point in A, and since x is the only accumulation point of φ, we have $x \in A$. Thus, A is closed.

(4). We will give an example in **8** of a separable countably compact completely regular space Y such that $Y \times Y$ is not countably compact. To prove the second part, note first that because of (1), the countable compactness of $\prod_1^\infty X_\iota$ implies that of each X_ι. For the converse, we will use the Cantor "diagonal process" to show that each sequence φ in $\prod_1^\infty X_\iota$ has an accumulation point. Let p_n denote the projection onto the nth factor. Now, $p_1 \circ \varphi$ is a sequence in X_1, and since X_1 is 1° countable and countably compact, we can (X, **6.1**) extract a subsequence φ_1 of φ such that $p_1 \circ \varphi_1$ converges to some $\mathring{x}_1 \in X_1$. Now consider $p_2 \circ \varphi_1$; for the same reason as before, we can extract a subsequence φ_2 of φ_1 such that $p_2 \circ \varphi_2$ converges to some $\mathring{x}_2 \in X_2$. Proceeding by induction, we obtain a family $\{\varphi_n \mid n \in Z^+\}$ of subsequences of φ such that φ_{n+1} is a subsequence of φ_n for each $n \in Z^+$, and each sequence $p_n \circ \varphi_n$ converges to some $\mathring{x}_n \in X_n$. Now let $\hat{\varphi}$ be the subsequence $n \to \varphi_n(n)$ of φ; then, for each fixed k, we have $\{\hat{\varphi}(s) \mid s \geqslant k\} \subset \{\varphi_k(s) \mid s \geqslant k\}$, so $\hat{\varphi} \vdash \varphi_k$ and [X, **3.2** (2a)] therefore $p_k \circ \hat{\varphi} \to \mathring{x}_k$. By X, **5.4**, we find that $\hat{\varphi} \to \{\mathring{x}_n\}$ and, because $\hat{\varphi} \vdash \varphi$, it follows that $\varphi \succ \{\mathring{x}_n\}$. The proof is complete.

It is clear that each continuous real-valued function f on a countably compact space Y is bounded, because the countable open covering $\{y \mid |f(y)| < n\}$, $n \in Z^+$, has a finite subcovering. This behavior motivates an extension of the notion of countable compactness.

3.7 Definition A Hausdorff space Y is pseudocompact if every continuous real-valued function on Y is bounded.

Ex. 2 Every countably compact space is pseudocompact, but the converse is not true. Let $Y = [0, \Omega] \times [0, \omega] - (\Omega, \omega)$; since the set $\{(\Omega, n) \mid n \in Z^+\}$ has no cluster point in Y, this space is not countably compact. To prove that Y is pseudocompact, it suffices to show that each continuous $f\colon Y \to E^1$ can be extended over the compact $[0, \Omega] \times [0, \omega]$, since **2.3** then applies. Now, by III, **8**, Ex. 7, for each $n \leqslant \omega$ there exists an $\alpha_n \in \Omega$ and a constant c_n such that $f(\alpha, n) \equiv c_n$ for all $\alpha \geqslant \alpha_n$; letting $\alpha_0 = \sup \alpha_n$, we find $\alpha_0 < \Omega$ and that f is constant on each strip $\{(\alpha, n) \mid \alpha \geqslant \alpha_0\}$. The reader can verify that by defining $f(\Omega, \omega) = c_\omega$, the extended function is continuous at (Ω, ω).

Pseudocompactness may be expected to be significant in completely regular spaces where there are sufficiently many nonconstant maps $f\colon Y \to E^1$. For such spaces, we have the equivalent formulations

3.8 Theorem Let Y be completely regular. The following three properties are equivalent:

(1). Y is pseudocompact.

(2). If $V_1 \supset V_2 \supset \cdots$ is any descending sequence of nonempty open sets, then $\bigcap_1^\infty \overline{V}_n \neq \varnothing$.

(3). Each countable open covering of Y has a finite subfamily whose *closures* cover Y.

Proof: (1) $\Rightarrow$ (2). Assume that $V_1 \supset V_2 \supset \cdots$ and $\bigcap_1^\infty \overline{V}_n = \varnothing$. Then $\{V_n \mid n \in Z^+\}$ must be a nbd-finite family: for if each nbd of some given $y \in Y$ meets infinitely many V_i, then because the V_i descend, each nbd of y would meet all the V_i and therefore y would belong to $\bigcap_1^\infty \overline{V}_n$. Now, for each $n \in Z^+$, let $g_n: Y \to E^1$ be a continuous map such that $g_n(y_n) = n$ for some $y_n \in V_n$ and $g_n(Y - V_n) = 0$. Because $\{V_n \mid n \in Z^+\}$ is nbd-finite, the function $\sum_1^\infty g_n$ is well-defined and continuous; it clearly is not bounded, so Y is not pseudocompact.

(2) $\Rightarrow$ (3). Let $\{U_n \mid n \in Z^+\}$ be a given countable open covering, and let $V_n = Y - \bigcup_1^n \overline{U}_i$ for each $n \in Z^+$; we note that $\overline{V}_n \subset \bigcap_1^n \overline{\mathscr{C}\overline{U}_i} \subset \bigcap_1^n \overline{\mathscr{C}U_i} = \bigcap_1^n \mathscr{C}U_i$. Now, if no V_n were empty, then we would have $\varnothing \neq \bigcap_1^\infty \overline{V}_n \subset \bigcap_1^\infty \mathscr{C}U_i$; thus $Y \neq \bigcup_1^\infty U_i$, that is, $\{U_n \mid n \in Z^+\}$ would not be a covering.

(3) $\Rightarrow$ (1). Let $f: Y \to E^1$ be continuous. Setting $U_n = \{y \mid |f(y)| < n\}$ for each $n \in Z^+$, we obtain a countable open covering $\{U_n \mid n \in Z^+\}$. Since there is an integer m such that $\overline{U}_1 \cup \cdots \cup \overline{U}_m = Y$, the function is bounded.

Pseudocompactness reduces to countable compactness in a class of spaces that includes the normal spaces. A completely regular space is called *weakly normal* if each two disjoint closed sets, one of which is countable, have disjoint nbds. This is actually weaker than normality: $[0, \Omega] \times [0, \Omega] - (\Omega, \Omega)$ is not normal, but it is weakly normal.

3.9 In weakly normal spaces, pseudocompactness is equivalent to countable compactness.

Proof: Let Y be weakly normal. If Y is not countably compact, then there is a countably infinite discrete closed set $D \subset Y$. Now, according to VII, **2.4**, we can find a system $\{U_n \mid n \in Z^+\}$ of open sets whose closures are pairwise disjoint, and such that $D \cap U_n \neq \varnothing$ for each $n \in Z^+$. Choose $y_n \in D \cap U_n$; since $E = \{y_n \mid n \in Z^+\}$ is a countable closed set, we can find disjoint nbds $W \supset E$ and $V \supset \mathscr{C} \bigcup_1^\infty U_n$. Then the covering $\{V\} \cup \{U_n \mid n \in Z^+\}$ has no finite subfamily whose closures cover Y: each y_n belongs only to $\overline{U}_n$ and no y_n belongs to $\overline{V}$. Thus, Y is not pseudocompact.

For the invariance properties of psuedocompactness, we have

(a). The continuous image of a pseudocompact space is pseudocompact.

(b). Although a closed subspace of a pseudocompact space need not be pseudocompact, the reader can verify that if the pseudocompact space Y is completely regular, then the closure of each open subset of Y is pseudocompact. The example to be given in **8** will show that the cartesian product of pseudocompact spaces need not be pseudocompact.

4. Compactness in Metric Spaces

In metric spaces, there is no distinction between countable compactness and compactness, since metric spaces are paracompact and **3.4** applies; this serves to explain the importance of the Bolzano-Weierstrass property in metric spaces. Furthermore, compact metric spaces are always 2° countable; in fact,

4.1 Theorem A countably compact space Y is metrizable if and only if it is 2° countable.

Proof: Sufficiency: 2° countability assures first (**3.5**) that Y is regular and then, by Urysohn's theorem (IX, **9.2**), that Y is metrizable. *Necessity:* By **3.4**, a countably compact metric space Y is compact, and since a compact space is Lindelöf, IX, **5.6** shows that Y is 2° countable.

Ex. 1 Since $[0, \Omega]$ is compact, but has no countable dense set, this is another way to see that $[0, \Omega]$ is not metrizable.

For subsets of metric spaces,

4.2 Let (Y, d) be a metric space, and $A \subset Y$ compact. Then A is closed and bounded.

Proof: Due to **1.4**, only the boundedness of A requires proof. Choose $a_0 \in A$, and define $f: A \to E^1$ by $a \to d(a, a_0)$; being continuous, f attains a finite maximum m and clearly $\delta(A) \leqslant 2m$.

The converse of **4.2** is not true, as an infinite discrete space (metrized so that distinct points have distance 1 from each other) shows. However, in one very important case, the properties are characteristic:

4.3 Theorem In E^n, a set is compact if and only if it is closed and bounded.

Proof: Since A is bounded, $A \subset \prod_1^n I_i$, where each I_i is some finite closed interval $[a_i, b_i]$. By **1.4**(4) and **1**, Ex. 3, $\prod_1^n I_i$ is compact; since A is closed in E^n, A is closed in $\prod_1^n I_i$, and is therefore compact.

We have seen (IX, **4**, Ex. 3) that in metric spaces, two disjoint closed sets may have distance zero; we now show

4.4 Let X be a metric space, $A \subset X$ closed, and $C \subset X$ compact. If $C \cap A = \varnothing$, then $d(A, C) > 0$.

Proof: Since $d(c, A)$ is a continuous real-valued function on C, it attains its minimum at some point c_0, and $d(c_0, A) = \inf\{d(c, A) \mid c \in C\} = d(C, A)$; since $c_0 \not\in \bar{A} = A$, we have $d(c_0, A) > 0$.

Ex. 2 It need not be true that there are points $a \in A$, $c \in C$ with $d(A, C) = d(a, c)$. In $X = E^1 - \{0\}$, let $A = \{-1/n \mid n \in Z^+\}$ and $C = [1, 2]$.

Open coverings of compact metric spaces have the important property

4.5 Theorem Let (Y, d) be a compact metric space and $\{U_\alpha\}$ be an open covering of Y. Then there exists a positive number $\lambda(\{U_\alpha\})$, called a Lebesgue number of the covering, with the following property: Each ball $B(y, \lambda)$ is contained in at least one U_α.

Proof: For each $y \in Y$, choose $r(y) > 0$ so that $B(y, r(y)) \subset$ some U_α; and from the open covering $\left\{B\left(y, \frac{r(y)}{2}\right)\right\}$ extract a finite subcovering $\left\{B\left((y_i, \frac{r(y_i)}{2}\right) \middle| i = 1, \cdots, n\right\}$. Let

$$\lambda = \min\left[\frac{r(y_1)}{2}, \ldots, \frac{r(y_n)}{2}\right];$$

then $\lambda > 0$ is a Lebesgue number. For, given any $B(y, \lambda)$, we have $y \in B\left(y_i, \frac{r(y_i)}{2}\right)$ for some i, and so for any $z \in B(y, \lambda)$,

$$d(z, y_i) \leqslant d(z, y) + d(y, y_i) < \lambda + \frac{r(y_i)}{2} \leqslant r(y_i);$$

that is, $B(y, \lambda) \subset B(y_i, r(y_i))$. Since the latter set lies in some U_α, the result follows.

Ex. 3 Theorem **4.5** frequently occurs in the dual form: If Y is compact metric and $\{F_\alpha\}$ is a family of closed sets with $\bigcap_\alpha F_\alpha = \varnothing$, there is a $\lambda > 0$ with the following property: If $M \subset Y$ is any set with $\delta(M) < \lambda$, then M does not meet at least one F_α.

A basic application of **4.5** generalizes the classical theorem on the uniformity of continuity:

4.6 Theorem Let Y be a compact metric space, Z an arbitrary metric space, and $f\colon Y \to Z$ continuous. Then for each $\varepsilon > 0$, there is a $\delta(\varepsilon) > 0$, depending *only* on ε, such that $f(B(y, \delta)) \subset B(f(y), \varepsilon)$ for every $y \in Y$ (that is, f is *uniformly continuous*).

Proof: Cover Z by the balls $\{B(z, \varepsilon/2)\}$ and let δ be a Lebesgue number of the open covering $\{f^{-1}(B(z, \varepsilon/2))\}$ of Y. Since each $B(y, \delta)$ lies in one of these sets, $f(B(y, \delta)) \subset B(z, \varepsilon/2)$ for some z, and because $f(y) \in B(z, \varepsilon/2)$, we find for any $\xi \in B(y, \delta)$ that

$$d(f(\xi), f(y)) \leqslant d(f(\xi), z) + d(z, f(y)) < \varepsilon/2 + \varepsilon/2 = \varepsilon,$$

that is, $f(B(y, \delta)) \subset B(f(y), \varepsilon)$.

Remark: Let Y be a compact metric space. Because of **3.2**(3) and X, **6.1**(2), from each sequence in Y we can extract a convergent subsequence. However, it is possible to define a notion of convergence in compact metric spaces according to which every sequence converges to a unique limit; such a generalized convergence is used to define a Haar measure in locally compact topological groups.

In Z^+, introduce one of the finitely additive measures $\mu\colon \mathscr{P}(Z^+) \to 2$ of II, **2.5**(3); recall that $\mu(A) = 0$ if A is finite, and $\mu(Z^+) = 1$. Given a sequence $\varphi\colon Z^+ \to Y$, Y Hausdorff, define $\operatorname{Lim} \varphi = y_0$ if $\forall\, U(y_0)\colon \mu\{n \mid \varphi(n) \in U\} = 1$; it is trivial to verify that (1) $\operatorname{Lim} \varphi$, if it exists, is unique (depending on the μ used); (2) if $\varphi \to y_0$, then $\operatorname{Lim} \varphi = y_0$, independently of the μ used. In addition, if Y is compact metric, every sequence converges: covering Y by finitely many closed balls $\overline{B(y, 1)}$, at least one satisfies $\mu[\varphi^{-1}(\overline{B(y, 1)})] = 1$; covering that one by closed balls of radius $\frac{1}{2}$, and repeating the process, **1.6** shows $\operatorname{Lim} \varphi$ exists. In the particular case that $Y = E^1$, we find: (a) Every bounded sequence converges; (b) $\underline{\lim}\, x_n \leqslant \operatorname{Lim} x_n \leqslant \overline{\lim}\, x_n$ (its value depending on the μ used if $\{x_n\}$ does not converge); (c) $\operatorname{Lim}(x_n + y_n) = \operatorname{Lim} x_n + \operatorname{Lim} y_n$; and (d) $\operatorname{Lim}(x_n \cdot y_n) = \operatorname{Lim} x_n \cdot \operatorname{Lim} y_n$. However, a subsequence of a given sequence need *not* have the same Lim.

5. Perfect Maps

We have seen that normality and paracompactness are preserved under continuous closed maps. In this section, we consider a type of map under which the image generally inherits all the properties of the mapped space.

5.1 Definition A map $p\colon X \to Y$ is called perfect (or proper) if it is a continuous closed surjection and each fiber $p^{-1}(y)$, $(y \in Y)$ is compact.

5.2 Theorem Let $p\colon X \to Y$ be a perfect map. Then:

(1). If X is Hausdorff, so also is Y.
(2). If X is regular, so also is Y.
(3). (**S. Hanai**) If X is metrizable, so also is Y.
(4). If X is 2° countable, so also is Y.

Proof: Ad (1). Let y_1, y_2 be distinct points of Y; then $p^{-1}(y_1)$, $p^{-1}(y_2)$ are disjoint compact sets in X and therefore have disjoint nbds U_1, U_2. Since p is closed, there are (III, **11.2**) open $V_i \supset y_i$ with $p^{-1}(y_i) \subset p^{-1}(V_i) \subset U_i$, $i = 1, 2$, and then V_1, V_2 are disjoint nbds of y_1, y_2.

Ad (2). Given $y \in U$ in Y, there is by **1.5**(b) an open $V \subset X$ with $p^{-1}(y) \subset V \subset \overline{V} \subset p^{-1}(U)$. Since p is a closed map, we find a nbd $W \supset y$ with $p^{-1}(y) \subset p^{-1}(W) \subset V$; then $W \subset p(\overline{V}) \subset U$, and since $p(\overline{V})$ is closed, $y \in W \subset \overline{W} \subset p(\overline{V}) \subset U$.

Ad (3). We will show that Y satisfies the metrizability condition in IX, **9.5**(2). Let d be a metric for X, and for each $n \in Z^+$ let $\mathfrak{F}_n$ be a nbd-finite closed refinement of the covering $\{B(x, 1/n) \mid x \in X\}$. Define $\mathfrak{G}_n = \{p(F) \mid F \in \mathfrak{F}_n\}$; each $\mathfrak{G}_n$ is a closed covering of Y, and we first prove that each $\mathfrak{G}_n$ is also nbd-finite: Let $y \in Y$ and n be fixed. Since $p^{-1}(y)$ is compact, there is a nbd $U \supset p^{-1}(y)$ that intersects only finitely many members of $\mathfrak{F}_n$, and since p is closed, there is a nbd $W(y)$ such that $p^{-1}(W) \subset U$; then W clearly meets at most finitely many members of $\mathfrak{G}_n$.

Now let U be any nbd of $y \in Y$. Because $p^{-1}(y)$ is compact, we have $d(p^{-1}(y), \mathscr{C}p^{-1}(U)) \geqslant 1/n > 0$ for a suitable n; then $\mathrm{St}(p^{-1}(y), \mathfrak{F}_{2n}) \subset p^{-1}(U)$ and therefore $\mathrm{St}(y, \mathfrak{G}_{2n}) = p\,\mathrm{St}(p^{-1}(y), \mathfrak{F}_{2n}) \subset U$. Since Y is obviously T_0, the space Y is metrizable.

Ad (4). Let $\{U_i\}$ be a countable basis for X, and let $\{V_i\}$ be the family of all finite unions of the U_i; by II, **8.8**, the family $\{V_i\}$ is countable, and we show that the open sets $W_i = Y - p[X - V_i]$ are a basis for Y. Let $y \in W$; then $p^{-1}(y) \subset p^{-1}(W)$, and since $p^{-1}(y)$ is compact, there are finitely many sets $U_1, \cdots, U_n$, such that $p^{-1}(y) \subset \bigcup_1^n U_1 \subset p^{-1}(W)$; letting $V_k = \bigcup_1^n U_i$, it follows that $y \in W_k \subset W$, completing the proof.

Perfect maps also preserve certain properties under inverse images; in fact they certainly preserve those properties determined by the behavior of open coverings:

5.3 Theorem Let $p: X \to Y$ be a perfect map. Then:

(1). If Y is paracompact, so also is X.
(2). If Y is compact, so also is X.
(3). If Y is Lindelöf, so also is X.
(4). If Y is countably compact, so also is X.

Proof: Ad (1). Let $\{W_\alpha \mid \alpha \in \mathscr{A}\}$ be any open covering of X. For each $y \in Y$, extract a finite covering $W_{\alpha_i(y)}$, $i = 1, 2, \cdots, n(y)$ for the compact fiber $p^{-1}(y)$ and, since p is a closed map, find a nbd $V(y)$ of y such that $p^{-1}(V(y)) \subset \bigcup_1^{n(y)} W_{\alpha_i(y)}$. Now let $\{U(y)\}$ be a precise nbd-finite open refinement of the open covering $\{V(y) \mid y \in Y\}$, and for each $y \in Y$, $i = 1, \cdots, n(y)$, let $W(y, \alpha_i(y)) = p^{-1}[U(y)] \cap W_{\alpha_i(y)}$. The family $\{W(y, \alpha_i(y))\}$ is an open covering of X that clearly refines $\{W_\alpha\}$; and it is nbd-finite: For, given any $x_0 \in X$, there is a nbd W of $p(x_0)$ intersecting at most finitely many sets $U(y)$ and the nbd $p^{-1}(W)$ of x_0 then intersects at most finitely many $W(y, \alpha_i(y))$. The proofs of (2) and (3) are entirely analogous and are left for the reader.

Ad (4). Let $\{W_n \mid n \in Z^+\}$ be any countable open covering of X. The open sets $V_n = Y - p[X - \bigcup_1^n W_i]$, $n \in Z^+$, cover Y: given any $y \in Y$, the compactness of $p^{-1}(y)$ assures that $p^{-1}(y) \subset \bigcup_1^m W_i$ for some m, and therefore that $y \,\bar{\in}\, p[X - \bigcup_1^m W_i]$. Since there is a finite subcovering $V_1, \cdots, V_k$ for Y, and since $p^{-1}(V_i) \subset W_i$ for each i, the sets $W_1, \cdots, W_k$ are a finite subcovering for X.

As an application of **5.3**,

5.4 Let X be compact. If Y is paracompact (resp. Lindelöf, resp. countably compact), then $X \times Y$ is also paracompact (resp. Lindelöf, resp. countably compact).

Proof: Because X is compact, it follows at once from **2.5** that the projection map $p: X \times Y \to Y$ is perfect.

This result should be contrasted with the known fact that (VIII, **2.4**(3)) the cartesian product of paracompact spaces need not be paracompact.

6. Local Compactness

Many of the important spaces occurring in analysis are not compact, but have instead a local version of compactness. Calling a subset A of a space *relatively* compact whenever its closure $\bar{A}$ is compact, this local property is formalized in

6.1 Definition A Hausdorff space is locally compact if each point has a relatively compact nbd.

Ex. 1 A compact space is locally compact. E^n, and any infinite discrete space is locally compact, but not compact. The set of rationals in E^1 is not a locally compact space. A Hilbert space $l^2(\aleph)$, with $\aleph \geqslant \aleph_0$, is not locally compact, as IX, **8**, Ex. 1, shows.

Equivalent formulations are given in

6.2 Theorem The following four properties are equivalent

(1). X is locally compact.
(2). For each $x \in X$ and each nbd $U(x)$, there is a relatively compact open V with $x \in V \subset \overline{V} \subset U$.
(3). For each compact C and open $U \supset C$, there is a relatively compact open V with $C \subset V \subset \overline{V} \subset U$.
(4). X has a basis consisting of relatively compact open sets.

Proof: (1) $\Rightarrow$ (2). There is some open W with $x \in W \subset \overline{W}$ and $\overline{W}$ compact. Since $\overline{W}$ is therefore a regular space, and $\overline{W} \cap U$ is a nbd of x in $\overline{W}$, there is a set G open in $\overline{W}$ such that $x \in G \subset \overline{G}_{\overline{W}} \subset \overline{W} \cap U$. Now $G = E \cap \overline{W}$, where E is open in X, and the desired nbd of x in X is $V = E \cap W$.

(2) $\Rightarrow$ (3). For each $c \in C$, find a relatively compact nbd $V(c)$ with $\overline{V(c)} \subset U$; since C is compact, finitely many of these nbds cover C, and by **1.5**(a), this union has compact closure.

(3) $\Rightarrow$ (4). Let $\mathscr{B}$ be the family of all relatively compact open sets in X; since each $x \in X$ is compact, (3) asserts that $\mathscr{B}$ is a basis.

(4) $\Rightarrow$ (1) is trivial.

The reader should observe that the relative compactness of V distinguishes (3) from **1.5**(c). Furthermore, the basis specified in (4) is very large; we obtain better information in the useful

6.3 Let X be 2° countable. If X is locally compact, it has a *countable* basis consisting of relatively compact open sets.

Proof: Let $\{U_n \mid n \in Z^+\}$ be a basis for X. For each fixed n, cover U_n by relatively compact open sets $\{V(y) \mid y \in U_n\}$ such that each $\overline{V(y)} \subset U_n$. Since a subspace of a 2° countable space is 2° countable, we can extract a countable subcovering $\{V_{n,i} \mid i \in Z^+\}$ of U_n. Repeating for each n, the family of sets $\{V_{n,i} \mid (n, i) \in Z^+ \times Z^+\}$ forms the required basis.

The position of locally compact spaces is given in

6.4 Theorem Every locally compact space is completely regular.

Proof: Let $p \in X$ and let A be a closed set not containing p. Repeated applications of **6.2**(3) give relatively compact open V_1, V_2, such that $p \in V_1 \subset \overline{V}_1 \subset V_2 \subset \overline{V}_2 \subset \mathscr{C}A$. Since $\overline{V}_2$ is compact and therefore normal, there is a continuous $f: \overline{V}_2 \to I$ having value 1 at p and 0 on $\overline{V}_2 - V_1$. Let $F: X \to I$ be the map coinciding with f on $\overline{V}_2$ and identically zero on $\mathscr{C}\overline{V}_2$. Since each of the two functions $F \mid \overline{V}_2$, $F \mid \mathscr{C}V_1$ are continuous and both coincide on $\overline{V}_2 \cap \mathscr{C}V_1$, application of III, **9.4** shows F to be continuous; thus X is completely regular.

Ex. 2 For an example of a nonnormal locally compact space, note first that because of **6.2**(2), an open subset of a compact space is always locally compact. Next note that $[0, \omega]$ being closed in $[0, \Omega]$ is compact, so that the space $[0, \omega] \times [0, \Omega]$ is also compact (and, in particular, normal). The open subspace $[0, \omega] \times [0, \Omega] - \{(\omega, \Omega)\}$ is consequently locally compact, but we have seen in VII, **3**, Ex. 4, that it is not normal.

For invariance properties, we have

6.5 Theorem (1). Local compactness is invariant under continuous open mappings.

(2). A locally compact subset A of a Hausdorff space Y is of the form $V \cap F$, where V is open and F is closed in Y.

(3). A subspace of a locally compact space is locally compact if and only if it is of the form $V \cap F$, where V is open and F is closed.

(4). $\prod\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ is locally compact if and only if all the Y_α are locally compact and at most finitely many are not compact.

Proof: (1). Let $f: X \to Y$ be continuous and open, and let $y \in Y$ be given. Choose $x \in X$ so that $f(x) = y$ and choose a relatively compact nbd $U(x)$. Because f is an open map, $f(U)$ is a nbd of y, and because $f(\overline{U})$ is compact, we find from $\overline{f(U)} \subset \overline{f(\overline{U})} = f(\overline{U})$ that $\overline{f(U)}$ is compact.

(2). Assume that A is locally compact. Each $a \in A$ has a nbd $V(a)$ in Y such that $\overline{V(a)} \cap A$ is compact, and therefore closed, in Y. Define $V = \cup\{V(a) \mid a \in A\}$; then V is open in Y and contains A; furthermore, the formula $V(a) \cap A = V(a) \cap [\overline{V(a)} \cap A]$ shows that each $V(a) \cap A$ is closed in $V(a)$, so by III, **9.3** we conclude that A is in fact closed in V. Thus $A = V \cap F$ for some closed F in Y.

(3). Because of (2) we need show only that, if $A = V \cap F$, then A is locally compact. Given $a \in A$, choose any relatively compact open U in Y satisfying $a \in U \subset \overline{U} \subset V$ and consider the nbd $U \cap A$ of a in A. The closure of this nbd in A is $\overline{U} \cap A = \overline{U} \cap (V \cap F) = \overline{U} \cap F$, which is a set closed in $\overline{U}$, and consequently is compact.

(4). Assume the condition holds. Given $\{y_\alpha\} \in \prod_\alpha Y_\alpha$, for each of the at most finitely many indices for which Y_α is not compact, choose a relatively compact nbd $V_{\alpha_i}(y_{\alpha_i})$; then $\langle V_{\alpha_1}, \cdots, V_{\alpha_n} \rangle$ is a nbd of $\{y_\alpha\}$ and $\overline{\langle V_{\alpha_1}, \cdots, V_{\alpha_n} \rangle} = \langle \overline{V}_{\alpha_1}, \cdots, \overline{V}_{\alpha_n} \rangle$ is compact. Conversely, assume $\prod_\alpha Y_\alpha$ to be locally compact; since each projection p_α is a continuous open surjection, each Y_α is certainly locally compact. But also, choosing any relatively compact open $V \subset \prod_\alpha Y_\alpha$, each $p_\alpha(\overline{V})$ is compact, and since $p_\alpha(\overline{V}) = Y_\alpha$ for all but at most finitely many indices α, the result follows.

We also have

6.6 Let $p: X \to Y$ be a perfect map. Then X is locally compact if and only if Y is locally compact.

Proof: Assume that X is locally compact. Given $y \in Y$, the compact fiber $p^{-1}(y)$ has [**6.2**(3)] a relatively compact nbd U; finding a nbd $V \supset y$ such that $p^{-1}(V) \subset U$, we have (III, **11.4**) that $\overline{V} \subset \overline{p(U)} \subset p(\overline{U})$, and therefore V is a relatively compact nbd of y. Thus Y is locally compact.

Conversely, assume that Y is locally compact. Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be an open covering of Y by relatively compact open sets. For each α, $f \mid f^{-1}(\overline{U}_\alpha): f^{-1}(\overline{U}_\alpha) \to \overline{U}_\alpha$ is evidently a perfect map, so by **5.3**, each $f^{-1}(\overline{U}_\alpha)$ is compact. Since $\{f^{-1}(U_\alpha)\}$ is an open covering of X, and since $\overline{f^{-1}(U_\alpha)} \subset f^{-1}(\overline{U}_\alpha)$, each $f^{-1}(U_\alpha)$ is relatively compact, and therefore X is locally compact.

7. σ-Compact Spaces

As the ordinal space $[0, \Omega[$ shows, locally compact spaces need not be paracompact. In this section, we show that the paracompact ones (and, in particular, the locally compact metric spaces) are characterized by a simple representation property.

7.1 Definition A locally compact space is σ-compact if it can be expressed as the union of at most countably many compact spaces.

Ex. 1 E^n, and any countably infinite discrete space, is σ-compact; an uncountable discrete space is locally compact, but not σ-compact.

Ex. 2 It need not be true that a space which is the countable union of compact spaces is σ-compact, as the rationals in E^1 show; local compactness is an essential part of the definition.

The definition has several equivalent formulations:

7.2 Theorem The following three properties are equivalent:

(1). Y is σ-compact.

(2). Y can be represented as $Y = \bigcup_1^\infty U_i$, where each U_i is a relatively compact open set, and $\overline{U}_i \subset U_{i+1}$ for each $i \in Z^+$.

(3). Y is a Lindelöf locally compact space.

Proof: (1) ⇒ (2). We have $Y = \bigcup_1^\infty C_i$ where each C_i is compact Since Y is locally compact, there is [**6.2**(3)] a relatively compact open $U_1 \supset C_1$ and, proceeding inductively, we choose U_n to be a relatively compact open set containing the compact $\overline{U}_{n-1} \cup C_n$. The sets $\{U_n \mid n \in Z^+\}$ evidently satisfy the requirements.

(2) ⇒ (3). Let $\{V_\alpha \mid \alpha \in \mathscr{A}\}$ be any open covering of Y. For each $i \in Z^+$, extract finitely many sets $\{V_{ij} \mid 1 \leqslant j \leqslant n(i)\}$ to cover $\overline{U}_i$; then the family $\{V_{ij} \mid 1 \leqslant j \leqslant n(i), i \in Z^+\}$ is a countable subcovering.

(3) ⇒ (1). Let $\{V(y) \mid y \in Y\}$ be a covering by relatively compact nbds, and extract a countable subcovering.

Ex. 3 It is useful to observe that, in the representation of **7.2**(2), each compact subset of Y is contained in some U_n: we need only note that the open covering $\{U_n \cap C \mid n \in Z^+\}$ of any compact $C \subset Y$ can be reduced to a finite covering.

Because a σ-compact space is Lindelöf, its regularity implies that it is paracompact (VIII, **6.5**). Thus, we can obtain paracompact locally compact spaces by forming free unions of σ-compact spaces (such spaces will not be Lindelöf if there are uncountably many summands). This is the only way to obtain such spaces, since

7.3 Theorem The paracompact locally compact spaces are precisely those spaces that are free unions of σ-compact spaces.

Proof: In view of the preceding remarks, we need prove only that a paracompact locally compact Y is the free union of σ-compact spaces. Cover Y by relatively compact nbds $\{U(y) \mid y \in Y\}$ and let $\{V(y) \mid y \in Y\}$ be a precise nbd-finite refinement. Since each $\overline{V(y)}$ is compact, it meets [**1.5**(d)] at most finitely many other $V(y')$. To get the summands of the free union, we introduce a relation R in Y as follows: $(x, y) \in R$ if there is a finite family of sets $V(y_1), \cdots V(y_n)$ such that $x \in V(y_1), y \in V(y_n)$ and $V(y_i) \cap V(y_{i+1}) \neq \varnothing$ for $i = 1, \cdots, n-1$. This is clearly an equivalence relation in Y, and we let $\{X_\alpha \mid \alpha \in \mathscr{A}\}$ be the family of equivalence classes. Then the X_α cover Y and are pairwise disjoint; furthermore, each X_α is open because it is a union of sets $V(y)$, and therefore each X_α is also locally compact. We complete the proof by showing that each X_α is σ-compact. Choose any set $V(y_0) \subset X_\alpha$; let Q_2 be the union of all sets $V(y_1)$ meeting $V(y_0)$ and, inductively, let $Q_n = \bigcup \{V(y) \mid V(y) \cap Q_{n-1} \neq \varnothing\}$. Because each $V(y)$ meets at most finitely many other $V(y')$, each Q_n consists of at most finitely many sets $V(y)$ and therefore each $\bar{Q}_n$ is compact. Since it is evident that $X_\alpha = \bigcup_n Q_n$, the proof is complete.

An open covering $\{U_\alpha\}$ of a space Y is called *star-finite* if each U_α intersects at most finitely many other U_β. Clearly, any open star-finite covering is nbd-finite, but the converse is not true; to any star-finite open covering having infinitely many members, we need only adjoin the single set Y to destroy its star-finiteness, but not its nbd-finiteness.

7.4 Corollary Let Y be a σ-compact. Then every open covering has a star-finite open refinement.

Proof: Given $\{U_\alpha\}$, cover Y by relatively compact open sets

$$\{V(y) \mid y \in Y\},$$

where each $\overline{V(y)} \subset$ some U_α. Since Y is paracompact, $\{V(y)\}$ has a nbd-finite open refinement $\{W_\beta\}$, and since each $\bar{W}_\beta$ is compact, it can meet at most finitely many other sets of the covering.

8. Compactification

E^1 is a noncompact completely regular space. It can be embedded in a compact space by at least two distinct methods:

(1). Identifying E^1 with $]-1, 1[\subset [-1, 1]$ by $x \to x/(1 + |x|)$.
(2). Identifying E^1 with S^1 − (north pole) by stereographic projection.

The first process can be regarded as compactifying E^1 by the addition of two new points, whereas the second does so by adding only one. We abstract this situation, and prevent indiscriminate enlargements, in

8.1 Definition A compactification of a space X is a pair $(\hat{X}, h)$ consisting of a compact Hausdorff $\hat{X}$ and a homeomorphism h of X onto a *dense* subset of $\hat{X}$.

We frequently identify X with $h(X) \subset \hat{X}$ and say simply that $\hat{X}$ is a compactification of X. Clearly, only completely regular spaces can be compactified, since a subset of a compact Hausdorff space is necessarily completely regular. In this section, we consider two compactification methods, the first of which applies to completely regular spaces and the second only to locally compact ones.

If X is completely regular, then (VII, **7.3**) there is an embedding $\rho: X \to P^X$ where P^X, being a cartesian product of closed unit intervals, is compact. The Stone-Čech compactification of X is $(\beta(X), \rho)$, where $\beta(X) = \overline{\rho(X)}$, and has the properties

8.2 Theorem (M. Stone; E. Čech) (1). For each compact space Y and each continuous $f: X \to Y$, there exists a unique continuous $F: \beta(X) \to Y$ such that $f = F \circ \rho$.

(2). (Uniqueness.) Any compactification $(\hat{X}, h)$ of X having the property (1) is homeomorphic to $\beta(X)$; indeed there is a homeomorphism $\hat{X} \cong \beta(X)$ that is the identity map on X.

(3). $\beta(X)$ is the "largest" compactification of X: if $\hat{X}$ is any compactification of X, then $\hat{X}$ is a quotient space of $\beta(X)$.

Proof: (1). By VII, **7.4**, we have the commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \rho\downarrow & & \downarrow\rho_0 \\ \beta(X) & \xrightarrow[\varphi]{} & \beta(Y) \end{array}$$

Since Y is compact, we have $\rho_0: Y \cong \beta(Y)$, and we set $F = \rho_0^{-1} \circ \varphi$. F is unique because X is dense in $\beta(X)$ (VII, **1.5**).

(2). We consider X as a subset of $\hat{X}$ and of $\beta(X)$, and let $i: X \to X$ be the identity map. By (1), there is an $F: \beta(X) \to \hat{X}$ with $F \mid X = i$, and also a $G: \hat{X} \to \beta(X)$ with $G \mid X = i^{-1}$. Since X is dense in both spaces and $F \circ G \mid X$, $G \circ F \mid X$ are identity maps of X, we have both $F \circ G = 1_{\hat{X}}$ and $G \circ F = 1_{\beta(X)}$, and consequently (III, **12.3**), both F and G are homeomorphisms.

(3). By (1), there is a continuous $F: \beta(X) \to \hat{X}$ extending the identity map $i: X \to X$; since $\beta(X)$ is compact, $F(\beta(X))$ is a closed set containing the dense subset $X \subset \hat{X}$ and consequently (III, **4.13**), F is surjective. Because F is a closed map, VI, **7.2**, applies, and

$$\hat{X} \cong \beta(X)/K(F).$$

Observe that the method of proof used for (2) can be applied to show that if X and Y are homeomorphic completely regular spaces, then each homeomorphism $h: X \cong Y$ can be extended to a homeomorphism $H: \beta(X) \cong \beta(Y)$; in particular, homeomorphic spaces have homeomorphic Stone-Čech compactifications.

Ex. 1 The Stone-Čech compactification of a space is fairly complicated. $[0, 1]$ is *not* the Stone-Čech compactification of $]0, 1]$, since the continuous map $x \to \sin 1/x$ into the compact space $Y = [-1, +1]$ does not have a continuous extension over $[0, 1]$. Similarly, $[-1, +1]$ is not the Stone-Čech compactification of $]-1, +1[$; and the function $x \to \arctan x$ shows that S^1 is not the Stone-Čech compactification of E^1.

To give some idea about the nature of the Stone-Čech compactification for even such a simple space as Z^+, we will show

(a). (**B. Popišil**) $\aleph(\beta Z^+) = 2^{\mathfrak{c}}$.

(b). (**J. Novàk**) If A is any infinite closed subset of βZ^+, then $\aleph(A) = 2^{\mathfrak{c}}$.

Ad (a). Let $I^{\mathfrak{c}}$ be the cartesian product of $2^{\aleph_0}$ copies of I; then $\aleph(I^{\mathfrak{c}}) = \mathfrak{c}^{\mathfrak{c}} = 2^{\aleph_0 \mathfrak{c}} = 2^{\mathfrak{c}}$, and furthermore (VIII, **7.2**) $I^{\mathfrak{c}}$ has a countable dense set D. Let $\varphi: Z^+ \to D$ be a surjection; φ is continuous, so is extendable to a $\hat{\varphi}: \beta Z^+ \to I^{\mathfrak{c}}$. Since βZ^+ is compact, $\hat{\varphi}(\beta Z^+)$ is a closed set containing D; consequently $\hat{\varphi}$ is also surjective and therefore (II, **7.4**) we have $\aleph(\beta Z^+) \geqslant 2^{\mathfrak{c}}$. However, there are $\mathfrak{c}^{\aleph_0} = \mathfrak{c}$ maps of Z^+ into I, and βZ^+ is a subspace of $I^{\mathfrak{c}}$; thus $\aleph(\beta Z^+) \leqslant 2^{\mathfrak{c}}$ and (a) is proved. This proof is due to J. Mrówka.

Ad (b). By VII, **2.4**, there exists a family $\{V_n \mid n \in Z^+\}$ of pairwise disjoint open sets in βZ^+ such that $A \cap V_n \neq \varnothing$ for each n. Choose any $a_n \in A \cap V_n$ and define $A_0 = \{a_n \mid n \in Z^+\}$. It is clear that $A_0 \cong Z^+$ so that $\beta A_0 \cong \beta Z^+$; we wish to show that $\beta A_0 \cong \bar{A}_0$, since from $A_0 \subset A$ we must have $\bar{A}_0 \subset A$ also, and (a) applies.

Let Y be any compact space, and $f: A_0 \to Y$ any continuous map. Define $g: Z^+ \to Y$ by

$$g(x) = \begin{cases} f(a_n) & \text{if} \quad x \in V_n \cap Z^+ \\ y_0 & \text{if} \quad x \in Z^+ - \bigcup_1^\infty V_n, \end{cases}$$

where y_0 is any point of Y. This map is extendable to a continuous $G: \beta Z^+ \to Y$; we assert that G is also an extension of f. Indeed, since Z^+ is dense in βZ^+, we have $a_n \in \bar{V}_n = \overline{V_n \cap Z^+}$ for each n, so

$$G(a_n) \in G(\overline{V_n \cap Z^+}) \subset \overline{G(V_n \cap Z^+)} = \overline{g(V_n \cap Z^+)} = f(a_n).$$

This shows, in particular, that f is extendable over $\bar{A}_0$; since A_0 is dense in $\bar{A}_0$, we have [**8.2**(2)] that $\bar{A}_0 \cong \beta A_0$ as required.

The space βZ^+ also has the properties

(c). Let $E \subset Z^+$ be the even integers, and $N \subset Z^+$ the odd integers. Then $\bar{E} \cong \bar{N} \cong \beta Z^+, \beta Z^+ = \bar{E} \cup \bar{N}$, and $\bar{E} \cap \bar{N} = \varnothing$.

(d). There is a homeomorphism $h: \beta Z^+ \to \beta Z^+$ that sends Z^+ onto itself, that satisfies $h \circ h = 1$, and is such that $h(x) \neq x$ for each $x \in \beta Z^+$.

Ad (c). Since any continuous map of E into a compact space Y is extendable over βZ^+ (first extend it arbitrarily over Z^+), we find as before that $\bar{E} \cong \beta E \cong \beta Z^+$ and, similarly, that $\bar{N} \cong \beta N \cong \beta Z^+$. Furthermore, $\bar{E}$ and $\bar{N}$ are disjoint: letting

$f: Z^+ \to I$ be the map sending E to 0 and N to 1, its extension $F: \beta Z^+ \to I$ shows that $\overline{E} \subset F^{-1}(0)$, $\overline{N} \subset F^{-1}(1)$, and so $\overline{E} \cap \overline{N} = \varnothing$. Finally, from $Z^+ = E \cup N$, we find $\beta Z^+ = \overline{Z^+} = \overline{E \cup N} = \overline{E} \cup \overline{N}$.

Ad (d). Because of (c), the homeomorphism $2n \to 2n - 1$ of E onto N extends to a homeomorphism $q: \overline{E} \cong \overline{N}$. The map $h: \beta Z^+ \to \beta Z^+$, defined by

$$h(x) = \begin{cases} q(x) & x \in \overline{E} \\ q^{-1}(x) & x \in \overline{N}, \end{cases}$$

has the required properties.

We will use (b) and (d) to construct an example, due to J. Novàk, of a countably compact completely regular space K such that $K \times K$ is not countably compact, nor even pseudocompact.

Let $\mathscr{F}$ be the family of all countably infinite subsets of βZ^+; since $\aleph(\beta Z^+) = 2^{\mathfrak{c}}$, we have $\aleph(\mathscr{F}) = 2^{\mathfrak{c}\aleph_0} = 2^{\mathfrak{c}}$. Well-order $\mathscr{F}$ according to the smallest ordinal Γ of cardinal $2^{\mathfrak{c}}$. We will use induction to construct a set $P \subset \beta Z^+$ containing a cluster point of each member of $\mathscr{F}$ and such that $h(P) \cap P = \varnothing$.

Let $B \in \mathscr{F}$ and assume that P_A has been defined for all $A < B$, that $\aleph(P_A) < 2^{\mathfrak{c}}$, that $P_D \subset P_A$ whenever $D < A$, and that $h(P_A) \cap P_A = \varnothing$ for each A. Let $Q_B = \bigcup \{P_A \mid A < B\}$. Since (II, **8.2**) we have $\aleph(Q_B) < 2^{\mathfrak{c}}$, whereas $\aleph(\overline{B} - B) = 2^{\mathfrak{c}}$, there is an $x_B \in \overline{B} - B \subset B'$ such that $x_B \bar{\in} h(Q_B)$. We define $P_B = Q_B \cup \{x_B\}$; then $\aleph(P_B) < 2^{\mathfrak{c}}$ and, because $h \circ h = 1$ and h has no fixed points, we also have $h(P_B) \cap P_B = \varnothing$. This completes the inductive step. The desired set is $P = \bigcup \{P_B \mid B \in \mathscr{F}\}$.

Let $K = P \cup Z^+$; this subspace of βZ^+ is certainly countably compact, since it contains a cluster point of each countably infinite subset even of βZ^+. Now consider $K \times K$ and let $\Delta = \{(n, h(n)) \mid n \in Z^+\} \subset K \times K$. Then Δ is closed in $K \times K$: in fact, observing that if $x \in K - Z^+$, then $h(x) \bar{\in} P$, it follows that Δ is the intersection of the graph of $h: \beta Z^+ \cong \beta Z^+$ with $K \times K$, and the graph of a continuous map is closed. Furthermore, Δ has no cluster points: each $n \in Z^+ \subset \beta Z^+$ has a nbd not containing any $n' \neq n$ so, since h maps Z^+ onto itself, each $(n, h(n))$ has a nbd not containing any other element of Δ. Thus, Δ is an infinite discrete closed subset of $K \times K$; consequently $K \times K$ is not countably compact nor, as is easy to see, even pseudocompact.

In particular, K is not compact. If we notice that Z^+ is dense in K, the space K is also an example of a separable completely regular countably compact space that is not compact.

Among the completely regular spaces, the locally compact ones are characterized by the position they must have in each compactification:

8.3 X is locally compact if and only if in any given compactification $(\hat{X}, h)$, $h(X)$ is open in $\hat{X}$, or equivalently, $h: X \to \hat{X}$ is an open mapping.

Proof: If X is open in $\hat{X}$, then by **6.5**(2), X is locally compact. Conversely, if X is locally compact, then $X = U \cap A$, where U is open and A is closed in $\hat{X}$; since the closed A contains the dense subset X, we have $A = \hat{X}$, so $X = U$ as required. As in III, 11, Ex. 1, this is equivalent to h being an open mapping.

Because of **8.3**, a nonlocally compact completely regular space can never be compactified by the addition of a single point. We will now show that every locally compact space has such a compactification; the construction generalizes the familiar one-point compactification of function theory, which adds a "point at ∞" to the plane.

8.4 Theorem (P. Alexandroff) (1). Any locally compact space X can be embedded in a compact space $\hat{X}$ so that $\hat{X} - X$ is a single point.

(2). (Uniqueness) Any two compact spaces $\hat{X}$, $\hat{Y}$ having the property (1) are homeomorphic; indeed, there is a homeomorphism $\hat{X} \cong \hat{Y}$, which is the identity map on X.

Proof: (1). Let ∞ be an object not in X, and let $\hat{X}$ be the set $X \cup \{\infty\}$ with the following topology: all open sets in X and the complement in $\hat{X}$ of each compact subset of X. It is easy to see that this is actually a topology and that the subspace $X \subset \hat{X}$ is homeomorphic to X. To show $\hat{X}$ Hausdorff, only the separation of each point $x \in X$ from ∞ need be checked, and this follows because $x \in X$ has a relatively compact nbd. Finally, $\hat{X}$ is compact: given any open covering $\{U_\alpha\}$, choose one set, U_{α_0}, containing ∞; being a nbd of ∞, U_{α_0} covers all $\hat{X}$ except for some compact $C \subset X$; by reducing the open covering $\{U_\alpha\}$ of C to a finite one, we obtain the required reduction of $\{U_\alpha\}$.

(2). If $\hat{Y} - X = \infty_Y$, we show that the bijection $f: \hat{X} \to \hat{Y}$ given by $f(x) = x$ when $x \in X$, and $f(\infty) = \infty_Y$ is a homeomorphism. Because of the symmetric roles of f, f^{-1}, we need show only that f is an open map, and this in turn requires only that the images of nbds of ∞ be shown open in $\hat{Y}$. If $\hat{X} - C$ is a nbd of ∞, then $C \subset X$ is compact, so that $f(C)$ is also. Because f is bijective, $f(\hat{X} - C) = \hat{Y} - f(C)$, so that the image of $\hat{X} - C$ is open in $\hat{Y}$.

Ex. 2 $[0, 1]$ is the one-point compactification of $]0, 1]$; this follows from the uniqueness statement, **8.4**(2). Similarly, E^n can be embedded in S^n − (north pole) by a stereographic projection; by uniqueness again, S^n is the one-point compactification of E^n. Note the equivalence of the statements "$A \subset E^n$ is compact" and "A is a closed set on S^n not containing the north pole," which is frequently used in analysis.

For the extension of maps defined on X over $\hat{X}$, we have

8.5 Corollary Let Y be a Hausdorff space. A necessary and sufficient condition that a continuous $f: X \to Y$ be extendable to a continuous $F: \hat{X} \to Y$ is that the filterbase $\{f(\mathscr{C}C) \mid C \subset X \text{ is compact}\}$ converge.

Proof: We need note only that $\mathfrak{A} = \{\mathscr{C}C \mid C \subset X \text{ is compact}\}$ is the nbd filterbase of ∞, and use X, **5.1**.

Ex. 3 If $Y = E^1$, the condition of **8.5** is equivalent to the following statement: For each $\varepsilon > 0$, there exists a compact $C(\varepsilon) \subset X$ such that $|f(x) - f(y)| < \varepsilon$, for all $x, y \in \mathscr{C}C(\varepsilon)$.

It is frequently important to know whether the one-point compactification of a space is metrizable. For this we have

8.6 Theorem Let X be locally compact. Then its one-point compactification $\hat{X}$ is metrizable if and only if X is 2° countable.

Proof: "Only if" is immediate from **4.1** and VIII, **6.2**. "If": By **7.2**(3), X is σ-compact, so X can be written $X = \bigcup_1^\infty U_n$, as in **7.2**(2). Because of **7** Ex. 3, the nbds $\{X - \bar{U}_n\}$ form a countable basis at ∞, and it follows at once that the countably many sets $\{V_\iota, \hat{X} - \bar{U}_n\}$ form in fact a basis for $\hat{X}$. By **4.1**, $\hat{X}$ is metrizable.

As the proof of **8.6** indicates, the one-point compactification is useful in extending results to locally compact spaces that are valid for normal spaces. As another indication of this technique, we prove a result useful in measure theory and integration:

8.7 Let X be locally compact and $C \subset X$ be compact. If U is a nbd of C there exists a continuous function $f: X \to I$, which is identically 1 on C and identically 0 on $\mathscr{C}U$.

Proof: Let $\hat{X}$ be the one-point compactification of X; because of **8.3**, C and $\hat{X} - U$ are disjoint closed sets, and because $\hat{X}$ is normal, there is an Urysohn function f for C and $\hat{X} - U$; the required function is $f \mid X$.

9. k-Spaces

A generalization of local compactness results from the proposition

9.1 Let X be locally compact. An $A \subset X$ is open if and only if its intersection with each compact $C \subset X$ is open in C.

Proof: Assume $A \cap C$ open in C for each compact C. Given $a \in A$, choose a relatively compact nbd $V(a)$; since $A \cap \overline{V(a)}$ is open in $\overline{V(a)}$, it follows at once that $A \cap V(a)$ is open in $V(a)$, and therefore $A \cap V(a)$ is open in X. Thus, each $a \in A$ has a nbd contained in A, and so A is open in X. The converse is trivial.

From VI, **8.1** we conclude that a locally compact space has the weak topology determined by the family of its compact subsets. We now reverse this procedure by the

9.2 Definition A Hausdorff space X is called a k-space if it has the weak topology determined by the family of its compact subspaces.

The class of k-spaces is larger than that of the locally compact spaces; for example, it contains also all the metric spaces, since

9.3 Every locally compact and every 1° countable space is a k-space.

Proof: Only the case that X is 1° countable requires proof. Let $A \subset X$ be such that $A \cap C$ is closed in C for each compact C; then $A \cap C$ is closed in X for each compact C, and we shall show $\bar{A} \subset A$. Let $x \in \bar{A}$; by X, **6.2**, there is a sequence $\{a_n \mid n \in Z^+\} \subset A$ with $a_n \to x$; since (XI, **1**, Ex. 2) $\{a_n\} \cup x$ is compact, so also is the closed $A \cap \{a_n\} \cup x$, and this intersection, being an infinite subset of $\{a_n\} \cup x$, must contain x. Thus $x \in A$, and so A is closed.

The relation of k-spaces and local compactness is

9.4 Theorem (D. E. Cohen) Let X be Hausdorff. Then X is a k-space if and only if it is a quotient space of a locally compact space.

Proof: Assume that X is a k-space. By VI, **8.5**, we know that X is a quotient space of the free union of its compact subspaces, and since the free union of compact spaces is evidently locally compact, the necessity of the condition follows. For sufficiency, let $p: Y \to X$ be the identification map, where Y is locally compact, and let $U \subset X$ be such that $U \cap C$ is open in C for each compact C; we are to show that U is open in X. For each relatively compact open $V \subset Y$, we have $U \cap p(\bar{V})$ open in the compact $p(\bar{V})$, that is, $U \cap p(\bar{V}) = p(\bar{V}) \cap G$, where G is open in X. Since $p^{-1}(U) \cap p^{-1}p(\bar{V}) = p^{-1}p(\bar{V}) \cap p^{-1}(G)$, we find by intersecting with V, that $p^{-1}(U) \cap V = V \cap p^{-1}(G)$, therefore $p^{-1}(U) \cap V$ is open in Y. Since there is a covering $Y = \bigcup_\alpha V_\alpha$ by relatively compact open sets, the formula $p^{-1}(U) = \bigcup_\alpha p^{-1}(U) \cap V_\alpha$ shows $p^{-1}(U)$ open in Y, so U is open in X.

9.5 Corollary If X is a k-space and $p: X \to Z$ is an identification, then Z is also a k-space.

Proof: Let Y be locally compact and $g: Y \to X$ an identification. Then $p \circ g$ is an identification, by VI, **3.3**, so by the theorem, Z is a k-space.

Ex. 1 The cartesian product of two k-spaces may not be a k-space, as VI, **8**, Ex. 5 shows; however it will be a k-space whenever either (a) both factors are 1° countable, or (b) one factor is locally compact. (a) is clear, since the product is then 1° countable. Though a direct proof of (b) is easy, we will derive it later (XII, **4.4**) from a more inclusive result.

10. Baire Spaces; Category

Locally compact spaces have the following useful property:

10.1 Theorem (R. Baire) The intersection of any countable family of open dense sets in a locally compact space Y is dense.

Proof: Let $D_1, D_2, \cdots$ be open dense sets; we must show that $U \cap \bigcap_1^\infty D_i \neq \varnothing$ for each open $U \subset Y$. Start with U and D_1; since $U \cap D_1 \neq \varnothing$ because D_1 is dense, **6.2** shows that there is a nonempty relatively compact open B_1 such that $\bar{B}_1 \subset U \cap D_1$. Now using B_1 and D_2, we find for the same reason that there is a nonempty relatively compact open B_2 such that $\bar{B}_2 \subset B_1 \cap D_2$, and proceeding by induction, we obtain a sequence $\{B_n\}$ of nonempty relatively compact open sets such that $\bar{B}_n \subset B_{n-1} \cap D_n$ for each n. The $\bar{B}_n$ are closed in the compact $\bar{B}_1$, and evidently have the finite intersection property, so $\bigcap_1^\infty \bar{B}_n \neq \varnothing$. Since $\bigcap_1^\infty \bar{B}_n \subset U \cap \bigcap_1^\infty D_n$, because of $\bar{B}_1 \subset U \cap D_1$ and $\bar{B}_n \subset D_n$ for each n, the proof is complete.

Ex. 1 The hypotheses "open" and "countable" are essential: Without "countable," the family $\{E^1 - \{x\} \mid x \in E^1\}$ of open dense sets in E^1 contradicts the conclusion; and by omitting "open," the dense sets {rationals}, {irrationals} in E^1 show the conclusion false.

Ex. 2 As one immediate use of **10.1**, we prove, as announced in III, **6**, Ex. 3, that the set of rationals $Q \subset E^1$ is not a G_δ-set. For, if $Q = \bigcap_1^\infty G_i$, then each open G_i is also dense, and we could enlarge the family $\{G_i\}$ by the countably many sets $E^1 - \{r\}$, $r \in Q$, to obtain a countable family of open dense sets in E^1 with a vacuous intersection.

The property in **10.1** is not confined only to locally compact spaces; we will see later that another broad class of spaces, which includes the irrationals in E^1, also possesses this property. Since this condition, independently of any other, means that a space has a feature important in topology and in analysis, we abstract and study it separately.

10.2 Definition A space Y is a Baire space if the intersection of each countable family of open dense sets in Y is dense.

With this terminology, every locally compact space is a Baire space. Somewhat more generally, every locally countably compact regular space is a Baire space, since the proof given in **10.1** is clearly applicable to this case also. The significant feature of Baire spaces is

10.3 Theorem (R. Baire) Let Y be a Baire space. If $\{A_n \mid n \in Z^+\}$ is any *countable* closed covering of Y, then at least one A_n must contain an open set, that is, $\text{Int}(A_n) \neq \varnothing$ for at least one n.

Proof: From $Y = \bigcup_1^\infty A_n$ follows $\bigcap_1^\infty \mathscr{C}A_n = \varnothing$, so since Y is a Baire space, not every $\mathscr{C}A_n$ can be dense. Consequently, $\overline{\mathscr{C}A_n} \neq Y$ for at least one n; that is, $\text{Int}(A_n) = \mathscr{C}\overline{\mathscr{C}A_n} \neq \varnothing$.

Ex. 3 This theorem yields the important conclusion that E^n, or any locally compact space, cannot be decomposed into countably many closed sets, each of which has no interior.

Ex. 4 The subspace Q of rationals in E^1 is not a Baire space. For, $\{r \mid r \in Q\}$ is a countable closed covering of Q, and since Q is not discrete, no $\{r\}$ is an open set.

Remark: Definition **10.2** and Theorem **10.3** are equivalent, as the reader can easily show.

The requirement "closed covering" in Theorem **10.3** can be relaxed by imposing a more stringent condition on the sets. To do this, we first require

10.4 Definition In any space Y,

(1). A set $B \subset Y$ is called *nowhere dense* (or: a sieve) if its closure has no interior (that is, $\mathscr{C}\overline{\mathscr{C}\overline{B}} = \varnothing$, or alternatively, if $\mathscr{C}\overline{B}$ is dense).

(2). Any countable union of nowhere dense sets is called a set of the first category (or meager).

(3). Any set not of the first category is said to be of the second category (or nonmeager).

In these terms, the extended version of **10.3** is

10.5 Theorem (R. Baire) In a Baire space, a set of the first category has no interior.

Proof: Let $\{B_n \mid n \in Z^+\}$ be a countable family of nowhere dense sets, and let U be any open set such that $U \subset \bigcup_1^\infty B_n$. We are to prove $U = \varnothing$. From $U \subset \bigcup_1^\infty B_\iota \subset \bigcup_1^\infty \overline{B}_\iota$, we find $\bigcap_1^\infty \mathscr{C}\overline{B}_\iota \subset \mathscr{C}U$; since the

$\mathscr{C}\bar{B}_i$ are open dense sets, the closed set $\mathscr{C}U$ is dense in Y, so that $Y = \mathscr{C}U$ and $U = \varnothing$.

Ex. 5 Though **10.5** assures that a 1° category set C in a Baire space cannot fill out any open set, the *closure* of C may have an interior (that is, $\bar{C}$ may not be nowhere dense). For example, the set Q of rationals in E^1 is evidently a first category set; yet, $\bar{Q} = E^1$.

Ex. 6 In a Baire space, the complement of a 1° category set is necessarily a second category set: otherwise the entire space would be a countable union of nowhere dense sets, contradicting **10.5**. Thus, in E^1, the irrationals are of the second category. Observe that a 2° category set also may have no interior.

Problems

Section 1

1. Let Y be compact. Prove: A filterbase $\mathfrak{A}$ on Y converges if and only if it has exactly one accumulation point.

2. Show that the Hilbert space $l^2(\aleph_0)$ is not compact.

3. Let Y be compact and $\mathscr{F}$ be a family of continuous real-valued functions on Y such that:

   a. If $f, g \in \mathscr{F}$, then their product $f \cdot g \in \mathscr{F}$.
   b. For each $y \in Y$ there is some nbd $U(y)$ and some $f \in \mathscr{F}$ vanishing identically on U.

   Prove: $\mathscr{F}$ contains the function $f \equiv 0$.

4. Let Y be a regular space and $Y = \bigcup_1^\infty C_i$, where each C_i is compact. Prove that Y is paracompact.

5. Let Y be compact. Show that the component containing a $y \in Y$ is the intersection of all closed open sets containing y, that is, the components are the quasicomponents.

6. Let X be compact and connected, and let $A \subset X$ be closed. Prove that there exists a closed connected subset $B \supset A$ such that no closed connected proper subset of B contains A.

7. Let $\{a_n^{(i)} \mid n \in Z^+; i = 1, \cdots, N\}$ be a finite number of sequences in a compact space. Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be any open covering. Prove: There exist integers n_0 and k_0 such that for each $i = 1, \cdots, N$, each pair of elements $a_{n_0}^{(i)}, a_{k_0}^{(i)}$ lie in a set U_{α_i} of the covering.

8. Prove that the countable connected Hausdorff space $(Z^+, \mathscr{T})$ (see V, Problem 1.10) is not compact. [*Hint*: Consider the open covering by the sets $\{U(p, 1) \mid \text{all primes } p\}$.]

9. Prove that X is compact if and only if each open covering of X contains an irreducible subcovering. [*Hint:* Let $\mathfrak{A}$ be a subcovering of minimal cardinality $\aleph_\mu$; write $\mathfrak{A} = \{U_\alpha \mid \alpha < \omega_\mu\}$ and consider $\{V_\beta \mid \beta < \omega_\mu\}$, where

$$V_\beta = \bigcup \{U_\alpha \mid \alpha < \beta\}].$$

10. Let Y be Hausdorff, and let $\{C_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of closed subsets such that $\bigcap_\alpha C_\alpha \neq \varnothing$. Let U be any open set containing $\bigcap_\alpha C_\alpha$. Prove: for each compact C_{α_0} there exist finitely many sets $C_{\alpha_1}, \cdots, C_{\alpha_n}$ such that $C_{\alpha_0} \cap \cdots \cap C_{\alpha_n} \subset U$.
11. Let $\aleph(Y) = \aleph_\alpha$. Show that if $\bigcap_\mu F_\mu \neq \varnothing$ for each descending family $\{F_\mu \mid \mu < \omega_\alpha\}$ of nonempty closed subsets, then Y is compact.
12. Show that the cartesian product of uncountably many copies of I is not perfectly normal (more particularly, no point is a G_δ).

Section 2

1. Prove that there is no continuous bijection of S^1 onto any subset of E^1, and no continuous surjection of S^1 onto E^1.
2. Prove: The suspension of S^n is homeomorphic to S^{n+1}.
3. Prove: I^n, with its boundary identified to a point, is homeomorphic to S^n.
4. Prove: If Y is compact, Z Hausdorff, and $p: Y \to Z$ a continuous surjection, then p is an identification.
5. Let X, Y, Z be Hausdorff spaces, and $f: X \times Y \to Z$ continuous. Let $A \subset X$, $B \subset Y$ be compact, and $G \subset Z$ open such that $f(A \times B) \subset G$. Prove: There exist open $U \supset A$, $V \supset B$ such that $f(U \times V) \subset G$.
6. Let X be compact. Define an equivalence relation R as follows: $x\,R\,y$ if for every continuous $f: X \to E^1$ such that $f(x) = 0, f(y) = 1$, there is a $u \in X$ with $f(u) = \frac{1}{2}$. Show that the equivalence classes are the components of X.
7. Let X be compact and $\{f_n\}$ be a sequence of continuous real-valued functions on X such that $f_n(x) \leqslant f_{n+1}(x)$ for every n and every x. Assume that there is a continuous $g: X \to E^1$ such that $f_n(x) \to g(x)$ for each x. Prove: f_n converges uniformly to g.
8. Let $f: X \to Y$ be continuous, and let $\{F_n \mid n \in Z^+\}$ be a descending sequence of compact subsets of X. Prove that $f(\bigcap_1^\infty F_\iota) = \bigcap_1^\infty f(F_\iota$.
9. Let X be Hausdorff, Y compact, and $f: X \times Y \to \tilde{E}^1$ continuous, where $\tilde{E}^1$ is the extended real line. Prove that $M(x) = \sup\{f(x, y) \mid y \in Y\}$ and $m(x) = \inf\{f(x, y) \mid y \in Y\}$ are continuous. [*Hint*: For any real a, let $Q = \{(x, y) \mid f(x, y) \geqslant a\}$ and use **2.5**.]
10. Let Y be compact and $f: Y \to Y$ continuous. Prove that there exists a nonempty closed set $A \subset Y$ such that $A = f(A)$.
11. Let Y be completely regular, $C \subset Y$ compact, and U a nbd of C. Prove: there exists a continuous $f: Y \to I$ such that $f(c) = 0$ for all $c \in C$ and $f(x) = 1$ for all $x \in Y - U$. [*Hint*: For each $c \in C$, let f_c be such that $f_c(c) = 0$ and $f_c(\mathscr{C}U) = 1$; cover C by finitely many sets $\{f_{c_i}^{-1}[0, \frac{1}{2}[\mid i = 1, \cdots, n\}$ and consider the product $f_{c_1}(x) \cdot f_{c_2}(x) \cdots f_{c_n}(x)$.]

Section 3

1. Prove: X is countably compact if and only if every family of closed subsets having the finite intersection property also has the countable intersection property.

2. Prove: X is countably compact if and only if: whenever $\{F_i\}$ is a descending sequence of closed nonempty sets, then $\bigcap_1^\infty F_i \neq \varnothing$.
3. Prove: If Y is 1° countable and X is countably compact, then a continuous bijection $f: X \to Y$ is a homeomorphism.
4. Prove: A regular space Y is countably compact if and only if every point-finite open covering has a finite subcovering. [**Remark**: This is no longer true if Y is assumed to be only Hausdorff.]
5. Show that a sequence in a 1° countable countably compact space converges if and only if it has a single accumulation point.
6. Prove: A space Y is pseudocompact if and only if each continuous $f: Y \to E^1$ attains its maximum and its minimum.
7. Show that a completely regular space Y is pseudocompact if and only if each nbd-finite system of nonempty open sets in Y is finite.
8. Prove: A completely regular space is pseudocompact if and only if each nbd-finite countable open covering has a *proper* subcovering.
9. Let $Q \subset I$ be the set of rationals. Retopologize I by taking the sets

$$x \cup (]x - \varepsilon, x + \varepsilon[\cap Q), \text{ all } \varepsilon > 0,$$

as the complete system of nbds at each x. Show that this space is pseudocompact, but not countably compact.

Section 4

1. Let X be compact metric and $T: X \to X$ be a map such that $d(x, y) = d(Tx, Ty)$ for all x, y. Prove: T is surjective. [*Hint*: If it is not surjective, select a $y \bar{\in} T(X)$ and consider the sequence $y, Ty, T(Ty), \cdots$.]
2. Let $\mathscr{F}$ be the family of all continuous maps $I \to I$ such that $|f(s) - f(t)| \leqslant |s - t|$. Define

$$d^+(f, g) = \max_{0 \leqslant t \leqslant 1} |f(t) - g(t)|.$$

Prove: $\mathscr{F}$ is compact.
3. Prove: A metric space is compact if and only if every continuous real-valued function on it is bounded.
4. Let X be compact. Prove that X is metrizable if and only if the diagonal $\Delta \subset X \times X$ is a G_δ. [*Hint*: Let $\Delta = \bigcap_1^\infty U_i$; for each i, find finitely many open V_{ji} such that $\Delta \subset \bigcup_j V_{ji} \times V_{ji} \subset U_i$, and show that the family of all V_{ji} is a basis for X.]
5. Let X be compact. Assume that there exists a continuous $f: X \times X \to E^1$ such that $f(x, y) = 0$ if and only if $x = y$. Prove: X is metrizable.
6. Let (X, d) be a compact metric space, and $\mathscr{F}(X)$ the space of all nonempty closed sets with the Hausdorff metric (IX, Problem 4.8) Prove:
   a. For each $E \subset X$, the set $\{A \in \mathscr{F}(X) \mid E \subset A\}$ is closed.
   Now, for any $E \subset X$, let $I(E) = \{A \in \mathscr{F}(X) \mid A \subset E\}$ and $J(E) = \{A \in \mathscr{F}(X) \mid A \cap E \neq \varnothing\}$.
   b. If E is open (closed), then both $I(E)$ and $J(E)$ are open (closed).
   c. The map $A \to \delta(A)$ of $\mathscr{F}(X)$ into E^1 is continuous.
   d. $d(A, B)$ is a continuous function on $\mathscr{F}(X) \times \mathscr{F}(X)$.
   e. $\mathscr{F}(X)$ is compact.

Section 5

1. Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces. Prove: If each $f_\alpha: X \to Y_\alpha$ is perfect, then so also is $\{f_\alpha\}: X \to \prod_\alpha Y_\alpha$.
2. Let $\mathscr{A}$ be an arbitrary set, and for each $\alpha \in \mathscr{A}$, let $f_\alpha: X_\alpha \to Y_\alpha$ be a perfect map. Prove: $\prod_\alpha f_\alpha: \prod X_\alpha \to \prod Y_\alpha$ is also perfect.
3. Prove that the following two conditions are equivalent:
   a. $f: X \to Y$ is perfect.
   b. If $\mathfrak{A}$ is a filterbase in X, and if $f(\mathfrak{A}) \succ y_0$, then there is an $x_0 \in X$ such that $\mathfrak{A} \succ x_0$ and $f(x_0) = y_0$.
4. Let Y be Hausdorff and $p: Y \to Z$ be perfect. Prove: If Z is metacompact, so also is Y.
5. Let X be compact, and $p: X \to X/R$. Prove X/R is Hausdorff if and only if p is closed.
6. Prove: A quotient space of a compact metric space is metrizable if and only if it is Hausdorff.
7. Let $f: X \to Y$ be a continuous map of a compact metric space into a Hausdorff space. Prove: $f(X)$ is metrizable.
8. Let Y be regular, let $p: Y \to Y/R$ be closed, and let each equivalence class be compact. Prove: Y is paracompact if and only if Y/R is paracompact.
9. Let $p: X \to Y$ be continuous and $D \subset X$ a dense proper subset. Assume that $p^{-1}(y) \cap D$ is compact for each $y \in Y$. Prove: $p \mid D$ is not a closed map.
10. Let X, Y, Z, be Hausdorff and $f: X \to Y$, $g: Y \to Z$ be continuous. Prove:
    a. If f and g are perfect, then so also is $g \circ f$.
    b. If $g \circ f$ is perfect and g is injective, then f is perfect.
11. Let $p: X \to Y$ be perfect, and let X have weight $\leqslant \aleph$. Prove that if $\aleph \geqslant \aleph_0$, then Y also has weight $\leqslant \aleph$.
12. Let $p: X \to Y$ be an open perfect map. Show that if X is completely regular, then so also is Y. [*Hint*: Use III, Problem 11.16.]
13. Prove $[0, \Omega] \times [0, \Omega[$ is not normal. Thus, if $p: X \to Y$ is perfect and Y is normal, then X need not be normal.
14. Let X be normal, let Y be $1°$ countable, and let $f: X \to Y$ be a continuous closed map. Prove that $\mathrm{Fr}[f^{-1}(y)]$ is countably compact for each $y \in Y$.
15. Prove: If X is a metric space, and if $f: X \to Y$ is a continuous closed surjection, then Y is metrizable if and only if $\mathrm{Fr}[f^{-1}(y)]$ is compact for each $y \in Y$. [*Hint: Necessity* follows from Problem 14. *Sufficiency*: For each $y \in Y$, define a set $C(y) \subset X$ as follows: If $\mathrm{Fr}[f^{-1}(y)] \neq \varnothing$, let $C(y) = \mathrm{Int}[f^{-1}(y)]$, and if $\mathrm{Fr}[f^{-1}(y)] = \varnothing$, let $C(y) = f^{-1}(y) - x_y$, where x_y is any point of $f^{-1}(y)$. Let $X_0 = X - \bigcup \{C(y) \mid y \in Y\}$, and show that $f \mid X_0$ is a perfect map.]

Section 6

1. For any space X, prove: If the cone TX over X is locally compact, then TX is in fact compact.
2. Prove: For every two points in a connected, locally connected, locally compact Hausdorff space, there exists a compact connected set containing them.

3. Let X be locally compact and $f: X \to Y$ be a continuous open surjection. Prove: For each compact $K \subset Y$ there exists a compact $C \subset X$ such that $f(C) = K$.
4. In an arbitrary space X, an $A \subset X$ is called locally closed if each $a \in A$ has a nbd $U(a)$ such that $U \cap A$ is closed in U. Prove: A is locally closed if and only if $A = B \cap V$, where B is closed and V is open in X.
5. Let X be Hausdorff and Y locally compact. Prove: $f: X \to Y$ is perfect if and only if $f^{-1}(C)$ is compact for each compact $C \subset Y$.
6. Let (Y, d) be a locally compact metric space. Let $p: Y \to Y/R$ be a closed map such that each equivalence class is compact. Prove: Y/R is a locally compact metrizable space.
7. Let X be Hausdorff and $D \subset X$ dense. Prove that every relatively compact nbd in D of a point $b \in D$ is a nbd of b in X.
8. Prove that the space $(Z^+, \mathscr{T})$ of V, Problem 1.10, is not locally compact. [*Hint*: See VII, Problem 2.3.]

Section 7

1. Give examples to show (a) a star-finite *closed* covering of a space need not be nbd-finite; and (b) a nbd-finite closed covering of a space need not be star-finite.
2. Let $\{U_\alpha \mid \alpha \in \mathscr{A}\}$ be a star-finite open covering of a connected space Y. If each $U_\alpha \neq \varnothing$, prove that $\aleph(\mathscr{A}) \leqslant \aleph_0$. [*Hint:* Proceed as in the proof of 7.3.]
3. Prove: Every countable point-finite open covering of a normal space Y has a countable star-finite open refinement. [*Hint:* Shrink the given $\{U_i \mid i \in Z^+\}$ to a closed covering $\{F_i \mid i \in Z^+\}$ such that $F_i \subset U_i$ for every i. For each fixed i, let $\{W_i(n) \mid n \in Z^+\}$ be a sequence of open sets such that $F_i \subset W_i(n) \subset U_i$ and $\overline{W_i(n)} \subset W_i(n+1)$ for every n. For each fixed $n \geqslant 1$, let
$$V_n = \bigcup \{W_i(n) \mid 1 \leqslant i \leqslant n\}.$$
Defining $V_n = \varnothing$ for $n \leqslant 0$, the open covering
$$\{(V_n - \overline{V}_{n-2}) \cap W_i(n) \mid 1 \leqslant i \leqslant n, \quad n = 1, 2, \cdots\}$$
is the required refinement.]
4. Prove the following theorem of K. Morita: Let Y be a connected space. Then every open covering of Y has an open star-finite refinement if and only if Y is regular and Lindelöf.

Section 8

1. Prove: $[0, \Omega]$ is the Stone-Čech compactification of $[0, \Omega[$.
2. Let X be a discrete space. Prove that the Stone-Čech compactification $\beta(X)$ has the following property: The closure of every open set is open.
3. Let X be normal, $A \subset X$ closed, and $f: A \to Y$, where Y is completely regular. Prove that f can be extended to a continuous $F: X \to \overline{Y}$, where $\overline{Y}$ is a compact space containing Y.
4. Prove that the Stone-Čech compactification βX of X is characterized by the following property: each continuous map f of X into the *unit interval* I has a unique continuous extension $F: \beta X \to I$.
5. Prove: $\beta(X)$ is connected if and only if X is connected.

6. Prove: For every lower semicontinuous map $f\colon Y \to E^1$, there exists one and only one lower semicontinuous $F\colon \beta Y \to E^1$ such that $F \mid Y = f$.
7. Let X be completely regular and f be a homeomorphism of X into a compact space Y. Prove: The Stone-Čech extension $F\colon \beta X \to Y$ sends $(\beta X) - X$ into $Y - f(X)$.
8. Show that the Stone-Čech compactification of $[0, \Omega[\times [0, \Omega[$ is $[0, \Omega] \times [0, \Omega]$, and that the Stone-Čech compactification of $[0, \Omega[\times [0, \omega]$ is $[0, \Omega] \times [0, \omega]$.
9. Show that βZ^+ is the union of a countably infinite family of pairwise disjoint subsets, each homeomorphic to βZ^+.
10. Show that $(\beta Z^+) - Z^+$ contains a copy of βZ^+.
11. Let $Q \subset E^1$ be the rationals. Show that $\aleph(\beta Q) = 2^{\mathfrak{c}}$.
12. Let X, X' be locally compact and $\hat{X}$, $\hat{X}'$ their one-point compactifications. Prove: In order that an $f\colon \hat{X} \to \hat{X}'$ be perfect, it is necessary and sufficient that its extension $F\colon \hat{X} \to \hat{X}'$, given by $F(x) = f(x)$ $(x \in X)$, $F(\infty) = \infty'$, be continuous.

Section 10

1. Prove: Baire spaces are invariant under continuous open surjections.
2. Prove: An open subset of a Baire space is also a Baire space.
3. Using Problem 2, prove that **10.2** and **10.3** are equivalent.
4. Let Y be a space with the following property: Each point of Y has a nbd that is a Baire space. Show that Y is a Baire space.
5. Prove: E^1_u is a Baire space.
6. Prove that the intersection of countably many open dense sets in a Baire space is a set of the second category.
7. If Y is a Baire space, and Y is not discrete, prove that $\aleph(Y) \geqslant \aleph_0$.
8. Give an example to show that in a Baire space the complement of a second category set need not be a first category set.
9. Let A be a set of the first category in a Baire space Y, and let $\{H_n\}$ be any countable family of sets such that $\mathscr{C}A \subset \bigcup_1^\infty H_n$. Prove: There exists an n_0 and an open set U such that $U' \cap H_{n_0} \neq \varnothing$ for each open $U' \subset U$. (Compare with the auxiliary proposition in VII, **3**, Problem 3).
10. Show that a completely regular pseudocompact space is a Baire space.
11. In the metric space (Z, d), let $A(z_0; \varepsilon)$ denote the closed ball $\{z \mid d(z, z_0) \leqslant \varepsilon\}$. Now let X be an arbitrary space, let Y be a metric space, and let $f\colon X \times Y \to Z$ be continuous in each variable separately. Let $\varepsilon > 0$ and $y_0 \in Y$ be kept fixed, and for each $x \in X$ define
$$d(x) = \sup\{r \mid f[x, B(y_0, r)] \subset A[f(x, y_0); \varepsilon]\}.$$
Prove: a. $x \to d(x)$ is an upper semicontinuous map of X into E^1.
b. If X is a Baire space, then there exists an $x_0 \in X$ such that $d[F(x, y), F(x_0, y_0)] \leqslant 2\varepsilon$ on some nbd $U(x_0) \times V(y_0)$.

Function Spaces

XII

The idea of topologizing the set of all continuous maps of one space into another plays an important role in modern topology. Of the many possible topologies, we study here one that stems from the classical function-theoretic concept of uniform convergence on every compact set.

1. The Compact-open Topology

Let X, Y be topological spaces, and denote by Y^X the set of *all* continuous maps of X into Y. We wish to define *nearness* of two maps by their "nearness" on the compact subsets of X: if f satisfies a condition $f(A) \subset V$, where $A \subset X$ is compact and $V \subset Y$ is open, then maps "near" to f are required to satisfy the same condition. The smallest topology in Y^X compatible with this requirement is called the compact-open topology (abbreviated: c-topology).

1.1 Definition For each pair of sets $A \subset X$, $B \subset Y$, let $(A, B) = \{f \in Y^X \mid f(A) \subset B\}$. The c-topology in Y^X is that having as subbasis all sets (A, V), where $A \subset X$ is compact and $V \subset Y$ is open.

It is evident that the space Y^X is a joint topological invariant of X and Y, that is, if $X \cong L$ and $Y \cong M$, then $Y^X \cong M^L$: we need only observe first that the *set* Y^X is completely determined by the spaces X, Y, and then that its c-topology is specified by using only topological invariants of X and Y, so that it, too, is uniquely determined by the spaces X, Y.

Ex. 1 If X is discrete, then Y^X is homeomorphic to $\prod\{Y_x \mid x \in X\}$ where each Y_x is a copy of Y.

Ex. 2 Since the basic open sets in Y^X are the finite intersections of subbasic ones, it is useful to observe:

$$\bigcap_1^n (A_i, W) = \left(\bigcup_1^n A_i, W\right); \qquad \bigcap_1^n (A, W_i) = \left(A, \bigcap_1^n W_i\right),$$

and

$$\bigcap_1^n (A_i, W_i) \subset \left(\bigcup_1^n A_i, \bigcup_1^n W_i\right).$$

Ex. 3 For closure, we have the formula $\overline{(A, W)} \subset (A, \overline{W})$: For, if $g \overline{\in} (A, \overline{W})$, then $g \in (a, Y - \overline{W})$ for some $a \in A$; since the single point a is always compact, $(a, Y - \overline{W})$ is a nbd of g in the c-topology, and because $(A, W) \cap (a, Y - \overline{W}) = \varnothing$, we find that $g \overline{\in} \overline{(A, W)}$.

The space Y^X contains some useful subspaces:

1.2 (a). For each $y_0 \in Y$, let $c_{y_0}\colon X \to Y$ be the constant map $x \to y_0$. The map $j\colon Y \to Y^X$ given by $y \to c_y$ is a homeomorphism of Y onto a subspace of Y^X; thus Y can always be embedded in Y^X.

(b). Let $Y_0 \subset Y$ be a subspace of Y; then Y_0^X is homeomorphic to the subspace $S_0 = \{f \in Y^X \mid f(X) \subset Y_0\} \subset Y^X$.

Proof: (a). Since the statements $c_y \in (A, V)$ and $y \in V$ are equivalent, it is evident that the injection j is a homeomorphism.

(b). Let $\varphi\colon Y_0^X \to S_0$ be the identity map; if $W \subset Y$ is open, and $V = W \cap Y_0$, the formula $\varphi[(A, V)] = (A, W) \cap S_0$ shows that φ is a homeomorphism.

Ex. 4 The analogue of **1.2**(b) for $X_0 \subset X$ is false: a map $X_0 \to Y$ need not be extendable over X, so that Y^{X}_0 is generally not a subspace of Y^X.

Some of the separation properties of Y^X are determined by those of Y:

1.3 (a). Y^X is Hausdorff if and only if Y is Hausdorff.

(b). Y^X is regular if and only if Y is regular.

Proof: If Y^X is Hausdorff, or regular, then since (**1.2**) Y can be regarded as a subspace of Y^X, Y must have these properties also. Thus, only the converses require proof.

(a). Let $f \neq g$; then $f(x_0) \neq g(x_0)$ for some $x_0 \in X$. Since Y is Hausdorff, find disjoint nbds $U(f(x_0))$, $V(g(x_0))$; then (x_0, U), (x_0, V) are disjoint nbds of f, g.

(b). Given $f \in (A, V)$, then because $f(A) \subset V$ is compact and Y is regular, there is (XI, **1.5**) an open W with $f(A) \subset W \subset \overline{W} \subset V$; by Ex. 3 we conclude that $f \in (A, W) \subset \overline{(A, W)} \subset (A, \overline{W}) \subset (A, V)$, establishing regularity of Y^X.

Ex. 5 In view of Ex. 1, if Y is normal, then Y^X need not be normal.

Since we consider only compact subsets of X in defining the topology of Y^X, it is to be expected that the c-topology will have particular importance whenever X contains "enough" compact subsets to define its topology.

2. Continuity of Composition; the Evaluation Map

Let X, Y, Z be three spaces. For $f \in Y^X$ and $g \in Z^Y$, the composition $g \circ f \in Z^X$, so that $T(f, g) = g \circ f$ defines a map $Y^X \times Z^Y \to Z^X$. We investigate the continuity of T.

2.1 T is always continuous in each argument separately. That is,

(1). $g \to g \circ f_1$ is a continuous map $Z^Y \to Z^X$ for each fixed f_1.
(2). $f \to g_1 \circ f$ is a continuous map $Y^X \to Z^X$ for each fixed g_1.

Proof: (1). Let (A, V) be any subbasic nbd of $g \circ f_1$. Observe that $g \circ f_1 \in (A, V)$ if and only if $g \in (f_1(A), V)$ and that, since $f_1(A)$ is compact, $(f_1(A), V)$ is in fact a nbd of g; thus $T[f_1, (f_1(A), V)] = (A, V)$, establishing the continuity.

(2). This is proved similarly, noting that the statements $g_1 \circ f \in (A, V)$ and $f \in (A, g_1^{-1}(V))$ are equivalent.

This result is summarized by using the following terminology: For each fixed continuous $f: X \to Y$, the map $f^+: Z^Y \to Z^X$, given by $f^+(g) = g \circ f$, is called the map *induced* by f; similarly, each fixed continuous $g: Y \to Z$ *induces* a map $g_+: Y^X \to Z^X$ by $g_+(f) = g \circ f$. The statement of **2.1** is then simply that a map of function spaces induced by a continuous map of one of the factors is always continuous.

In general, T is not continuous in both variables, as we shall see below. However,

2.2 Theorem Let X, Z be Hausdorff and Y locally compact. Then the map $T: Y^X \times Z^Y \to Z^X$ is continuous.

Proof: Let $f_1 \in Y^X$, $g_1 \in Z^Y$ and a nbd (A, W) of $g_1 \circ f_1$ be given. Since $g_1^{-1}(W) \subset Y$ is open, and $f_1(A) \subset g_1^{-1}(W)$ is compact, the local compactness of Y assures (XI, **6.2**) that there is a relatively compact open V with $f_1(A) \subset V \subset \bar{V} \subset g_1^{-1}(W)$. Thus we have nbds $f_1 \in (A, V)$, $g_1 \in (\bar{V}, W)$, and clearly, $T[(A, V), (\bar{V}, W)] \subset (A, W)$.

The evaluation map plays an important role in function spaces:

2.3 Definition For any two spaces Y, Z, the map $\omega: Z^Y \times Y \to Z$ defined by $(f, y) \to f(y)$ is called the evaluation map of Z^Y.

2.4 Theorem (1). For each fixed y_0, the map $\omega_{y_0}: Z^Y \to Z$, given by $f \to \omega(f, y_0) = f(y_0)$, is continuous.

(2). If Y is locally compact, then $\omega: Z^Y \times Y \to Z$ is continuous.

Proof: Because ω is precisely the composition map

$$T: Y^X \times Z^Y \to Z^X$$

whenever X is a space consisting of a single point, the results follow from **2.1**(1) and **2.2**.

Ex. 1 That the local compactness of Y is essential in **2.4**(2) is shown by the following example. Let $Y = Q \subset E^1$ be the rationals with the induced topology (we require only that Q be a nonlocally compact completely regular space), let I be the closed unit interval, and let $\omega: I^Q \times Q \to I$ be the evaluation map. Let $f_0: Q \to I$ be the map $f_0(Q) = 0$, and let $q_0 \in Q$ be any point; we prove ω not continuous at (f_0, q_0) by showing that this point has no nbd mapping into $W = \{t \mid t < 1\}$. Let U be any nbd of q_0, and $V = \bigcap_1^n (A_i, V_i)$ be any nbd of f_0; since $\bar{U}$ is not compact, whereas each A_i is, U cannot be a subset of $\bigcup_1^n A_i$, so that there is a $\tilde{q} \in U - \bigcup_1^n A_i$. Since Q is completely regular, there is an $\tilde{f}: Q \to I$ with $\tilde{f}(\tilde{q}) = 1$, and $\tilde{f}\left(\bigcup_1^n A_i\right) = 0$. Noting that $0 \in \bigcap_1^n V_i$, we conclude that $\tilde{f} \in V$, $\tilde{q} \in U$, and yet $\omega(\tilde{f}, \tilde{q}) \notin W$. Thus, ω is not continuous. This example also shows that the local compactness of Y is essential in **2.2**: if we take the space X of **2.2** to consist of a single point, then T becomes simply the evaluation map. We further remark that R. Arens has shown the evaluation map $\omega: I^Y \times Y \to I$ to be continuous if and *only if* Y is locally compact.

3. Cartesian Products

Given three spaces X, Y, Z, a function $\alpha(x, y) = z$ can be regarded as a map $X \times Y \to Z$ or as a family of maps $Y \to Z$ with X as the parameter space. In this section, we consider the effect that shifting from one point of view to the other has on the continuity of the maps.

For notation, let $\alpha: X \times Y \to Z$ be continuous in y for each fixed x; the formula

$$[\hat{\alpha}(x)](y) = \alpha(x, y)$$

defines, for each fixed x, an $\hat{\alpha}(x): Y \to Z$, and so $x \to \hat{\alpha}(x)$ is a map $\hat{\alpha}: X \to Z^Y$. Conversely, given an $\hat{\alpha}: X \to Z^Y$, the formula defines an $\alpha: X \times Y \to Z$ continuous in y for each fixed x. Two maps

$$\alpha: X \times Y \to Z \quad \text{and} \quad \hat{\alpha}: X \to Z^Y$$

related by the above formula are called *associates.*

The most important feature of the c-topology is

3.1 Theorem (1). If $\alpha: X \times Y \to Z$ is continuous then $\hat{\alpha}: X \to Z^Y$ is also continuous.

(2). If $\hat{\alpha}: X \to Z^Y$ is continuous, *and if Y is locally compact*, then $\alpha: X \times Y \to Z$ is also continuous.

Proof: (1). It is enough to show that for each x_0 and each subbasic (A, V) satisfying $\hat{\alpha}(x_0) \in (A, V)$, there is a nbd $U(x_0)$ with $\hat{\alpha}(U) \subset (A, V)$. Equivalently, we must show that, if $\alpha(x_0, A) \subset V$, then there is an open $U(x_0)$ with $\alpha(U \times A) \subset V$. To this end, note that $x_0 \times A \subset \alpha^{-1}(V)$ and that because α is continuous, $\alpha^{-1}(V)$ is open; since A is compact, XI, **2.6**, applies to give nbd $U(x_0)$ with $U \times A \subset \alpha^{-1}(V)$, completing the proof.

(2). In the sequence $X \times Y \xrightarrow{\hat{\alpha} \times 1} Z^Y \times Y \xrightarrow{\omega} Z$, the first map is continuous by IV, **2.5**, and because Y is locally compact, so also is the second. The combined map is therefore continuous, and this is

$$(x, y) \to \omega(\hat{\alpha}(x), y) = \hat{\alpha}(x)(y) = \alpha(x, y).$$

Note that although Z determines the separation properties of Z^Y it plays no role in the general continuity considerations of the theorem. Note also that (2) is true for *every* space X whenever Y is locally compact. Now, by imposing conditions on $X \times Y$, we can relax the requirement of local compactness on Y:

3.2 Corollary If $\hat{\alpha}: X \to Z^Y$ is continuous, and if $X \times Y$ is a k-space, then $\alpha: X \times Y \to Z$ is continuous.

Proof: With no assumptions on X or Y, we can still prove that α is continuous on all sets $X \times B$, where $B \subset Y$ is compact: For, using the continuous map $i^+: Z^Y \to Z^B$ induced by the inclusion $i: B \to Y$, each map in the sequence $X \times B \xrightarrow{\hat{\alpha} \times 1} Z^Y \times B \xrightarrow{i^+ \times 1} Z^B \times B \xrightarrow{\omega} Z$ is continuous, so that the combined map is also, and this is simply $\alpha \mid X \times B$.

It now follows that α is continuous on every compact $C \subset X \times Y$: for if $p: X \times Y \to Y$ is the projection, then $p(C)$ is compact and $C \subset X \times p(C)$. By VI, **8.3**, the map α is therefore continuous.

4. Application to Identification Topologies

Theorem **3.1** is very powerful; we use it to derive the fundamental theorem for identification topologies in cartesian products.

4.1 Theorem (J. H. C. Whitehead) Let $p: X \to R$ be an identification, and let Y be locally compact. Then the map

$$p \times 1: X \times Y \to R \times Y$$

is an identification.

Proof: Let Z be any space and $g: R \times Y \to Z$ be any map such that $g \circ (p \times 1): X \times Y \to Z$ is continuous; according to VI, **3.1**, the theorem will be proved if we can show that g is continuous. To this end, note first that since $\alpha = g \circ (p \times 1)$ is continuous, the associated $\hat{\alpha}: X \to Z^Y$ is also continuous. Furthermore, $\hat{\alpha}p^{-1}: R \to Z^Y$ is single-valued: indeed, $\hat{\alpha}p^{-1} = \hat{g}$, since

$$\begin{aligned}[\hat{\alpha}p^{-1}(r)](y) &= \alpha(p^{-1}(r), y) = g(pp^{-1}(r), y) = g(r, y) \\ &= [\hat{g}(r)](y).\end{aligned}$$

By VI, **3.2**, we conclude that $\hat{\alpha}p^{-1} = \hat{g}: R \to Z^Y$ is continuous, and since Y is locally compact, that $g: R \times Y \to Z$ is continuous.

This theorem permits the extension of VI, **3.2**, to cartesian products: With the notation of **4.1**, if $f: X \times Y \to Z$ is continuous, then $(r, y) \to f(p^{-1}(r), y)$ is a continuous map $R \times Y \to Z$ whenever it is single-valued. Of particular importance for our purposes later is the

4.2 Corollary Let $A \subset X$ be closed, and attach X to Y by a continuous $f: A \to Y$. Let $p: X + Y \to X \cup_f Y$ be the identification map. Let K be locally compact and $\varphi: X \times K \to Z$, $\psi: Y \times K \to Z$ continuous maps. If $\varphi(a, k) = \psi[f(a), k]$ for each $k \in K$ and $a \in A$, (that is, φ and ψ are "consistent"), then the map

$$(\varphi, \psi)[p \times 1]^{-1}: (X \cup_f Y) \times K \to Z$$

is continuous.

Proof: $p \times 1$ is an identification, and the consistency condition asserts that $(\varphi, \psi)[p \times 1]^{-1}$ is single-valued.

Theorem **4.1** also has the consequence:

4.3 Corollary Let $f: P \to X$, $g: R \to Y$ be identifications. If the range of one map and the domain of the other is locally compact, then $f \times g: P \times R \to X \times Y$ is an identification.

Proof: Assume that R, X are locally compact. Then

$$f \times 1_R: P \times R \to X \times R$$

and $1_X \times g: X \times R \to X \times Y$ are both identifications; by transitivity (VI, **3.3**) $(1_X \times g) \circ (f \times 1_R) = f \times g$ is therefore also an identification.

As another application of **3.1**, we have

4.4 Theorem Let X be a k-space and Y a locally compact space. Then $X \times Y$ is a k-space.

Proof: We first observe that if P is any k-space, and if R is any locally compact space, then an $f: P \times R \to Z$ is continuous if and only if $f \mid C \times R$ is continuous for each compact $C \subset P$. Indeed, since R is locally compact, the continuity of f is equivalent to that of $\hat{f}: P \to Z^R$ and, since P has the weak topology determined by its compact subsets, VI, **8.3**, shows that $\hat{f}$ is continuous if and only if $\hat{f} \mid C$ is continuous for each compact C.

We now prove the theorem. By the definition of weak topology, every set open in the cartesian product topology $\mathscr{T}(c)$ of $X \times Y$ is open in the k-topology $\mathscr{T}(k)$ of $X \times Y$, so we need prove only that

$$1: (X \times Y, \mathscr{T}(c)) \to (X \times Y, \mathscr{T}(\mathrm{k}))$$

is continuous. For compact $C \subset X$, $C' \subset Y$, the compactness of $C \times C'$ and XI, **9.1**, assures that $1 \mid C \times C'$ is continuous. Now, keeping any compact C fixed, and recalling (XI, **9.3**) that Y is a k-space, our observation above shows that $1 \mid C \times Y$ is continuous; and applying our observation again, since X is a k-space, we find that 1 is in fact continuous.

5. Basis for Z^Y

In the subbasic open sets (A, V) for the topology of Z^Y, the open sets $V \subset Z$ and the compact sets $A \subset Y$ can be restricted to range over certain subfamilies, and the resulting collection will still be a subbasis for the c-topology.

5.1 (a). Let $\mathscr{B} = \{W_\alpha \mid \alpha \in \mathscr{A}\}$ be a subbasis for Z. Then the family $\{(A, W) \mid A \subset Y \text{ is compact}, W \in \mathscr{B}\}$ is also a subbasis for Z^Y.

(b). Let $\mathscr{F} = \{C_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of compact sets in Y with the following property: For each compact A and open $U \supset A$, there are finitely many $C_i \in \mathscr{F}$ with $A \subset \bigcup_1^n C_i \subset U$. Then the family $\{(C, W) \mid C \in \mathscr{F}, W \in \mathscr{B}\}$ is also a subbasis for Z^Y.

Proof: (a). It is enough to show that, given $f \in (A, V)$, there are finitely many (A_i, W_i), $W_i \in \mathscr{B}$, with $f \in \bigcap_1^n (A_i, W_i) \subset (A, V)$. We start by observing that because V is open and $\mathscr{B}$ is a subbasis, it follows that $V = \bigcup_\beta V_\beta$, where each V_β is a finite intersection $\bigcap_{j=1}^{k(\beta)} W_{\beta,j}$. Now, since $f(A) \subset V$, the sets $\{f^{-1}(V_\beta) \cap A\}$ are an open (in A) covering of the compact A, so we can extract a finite covering $\{f^{-1}(V_i) \cap A\}$, $i = 1, \cdots, n$, which we shrink (since A is normal) to obtain compact $A_1, \cdots, A_n$ with $A = \bigcup_1^n A_i$, and $A_i \subset f^{-1}(V_i)$, for each $i = 1, \cdots, n$. Since

$$f \in (A_i, V_i) = \left(A_i, \bigcap_{j=1}^{k(i)} W_{i,j}\right) = \bigcap_{j=1}^{k(i)} (A_i, W_{i,j}),$$

for each i, we find

$$f \in \bigcap_{i=1}^{n} \bigcap_{j=1}^{k(i)} (A_i, W_{i,j}) = \bigcap_{i=1}^{n} (A_i, V_i) \subset (A, \bigcup_1^n V_i) \subset (A, V).$$

as required.

(b). Given $f \in (A, W)$, choose finitely many $C_i \in \mathscr{F}$ with

$$A \subset \bigcup_1^n C_i \subset f^{-1}(W);$$

then $f \in (C_i, W)$ for $i = 1, \cdots, n$, so that

$$f \in \bigcap_1^n (C_i, W) = \left(\bigcup_1^n C_i, W\right) \subset (A, W)$$

and, as in (a), the proof is complete.

Ex. 1 In $Z^{X \times Y}$, the sets $(A \times B, V)$, where $A \subset X$, $B \subset Y$ are compact and $V \subset Z$ is open, form a subbasis for the c-topology. For, let $C \subset X \times Y$ be compact, and let W be a nbd of C. If A, B are the projections of C on X, Y, respectively, then $A \times B$ is compact (hence, regular) and contains C; therefore each $c \in C$ has a nbd $U_c \times V_c$ in $A \times B$ with $\overline{U}_c \times \overline{V}_c \subset (A \times B) \cap W$. Since C is compact, extract a finite covering $\{U_{c_i} \times V_{c_i}\}$; then $C \subset \bigcup \overline{U}_{c_i} \times \overline{V}_{c_i} \subset W$ and because the $\overline{U}_{c_i}, \overline{V}_{c_i}$ are compact, the condition in **5.1**(b) is satisfied.

5.2 Theorem Let Y be locally compact and $\aleph_Y$ be the cardinal of a basis $\{U\}$ consisting of relatively compact open sets. Let Z be arbitrary and $\aleph_Z$ be the cardinal of a basis $\{V\}$. If at least one of $\aleph_Y$, $\aleph_Z$ is not finite, Z^Y has a basis of cardinal $\leqslant \aleph_Y \cdot \aleph_Z$. In particular, Z^Y is 2° countable if both Z and Y are 2° countable.

Proof: We first note that the family $\mathscr{A} = \{(\bar{U}, V)\}$ is a subbasis: For, because of **5.1**(a), the requirement that V lie in the specified basis is no restriction, and since $\{U\}$ is a basis, the condition in **5.1**(b) is also fulfilled. It is evident that $\aleph(\mathscr{A}) = \aleph_Y \cdot \aleph_Z$. Since the finite intersections of the members of $\mathscr{A}$ give a basis, this basis has cardinal no larger than that of the set of all finite subsets of $\mathscr{A}$, and since $\aleph(\mathscr{A}) \geqslant \aleph_0$, this basis has (II, **8.8**) cardinal $\aleph(\mathscr{A})$.

Ex. 2 If both $\aleph_Y$, $\aleph_Z$ are finite, the theorem is false. Let Z be a discrete space with n points and Y a discrete space with m points; then Z^Y has n^m elements and, being Hausdorff, is also discrete.

Ex. 3 The hypothesis that Y be locally compact is essential. Consider the space I^Q of **2**, Ex. 1; though both I and Q are 2° countable, I^Q is not even 1° countable: Indeed, because the identity map $1: I^Q \to I^Q$ is the associate of the evaluation map $\omega: I^Q \times Q \to I$, and because Q is second countable, the 1° countability of I^Q together with **3.2** would yield that ω is continuous, a contradiction to the result of **2**, Ex. 1.

We also use **5.1** to establish

5.3 Theorem Let Y be locally compact, and X, Z arbitrary. The map $\alpha \to \hat{\alpha}$ establishes a homeomorphism of $Z^{X \times Y}$ and $(Z^Y)^X$. (This is also true if $X \times Y$ is a k-space.)

Proof: By **3.1**, $\alpha \to \hat{\alpha}$ is bijective. A subbasis for $(Z^Y)^X$ consists of all pairs $(A, (B, W))$, where $A \subset X$, $B \subset Y$ are compact and $W \subset Z$ is open. By Ex. 1, the sets $(A \times B, W)$ form a subbasis for $Z^{X \times Y}$; since $\alpha \in (A \times B, W)$ if and only if $\hat{\alpha} \in (A, (B, W))$, the theorem is proved. For the parenthetical remark, we require **3.2** to show that $\alpha \to \hat{\alpha}$ is bijective.

6. Compact Subsets of Z^Y

For the spaces Z^Y previously considered, the only restriction on Z was that it be Hausdorff; from now on, we will assume that Z is a metric space.

The fact that some given subset of Z^Y is compact usually has important applications in analysis; in this section, we give a readily verifiable criterion for this.

6.1 Definition Let (Z, d) be a metric space and Y an arbitrary space. A subset $\mathscr{F} \subset Z^Y$ is equicontinuous at $y_0 \in Y$ if

$$\forall \varepsilon > 0 \;\exists\; U(y_0) \;\forall f \in \mathscr{F}: f(U(y_0)) \subset B(f(y_0), \varepsilon).$$

$U(y_0)$ is called a nbd of ε-equicontinuity for $\mathscr{F}$. We say that $\mathscr{F}$ is equicontinuous on Y whenever $\mathscr{F}$ is equicontinuous at each point of Y.

Ex. 1 Any finite subset of Z^Y is equicontinuous on Y.

Ex. 2 Let $Y = [a, b]$, $Z = E^1$, and $\mathscr{F} = \{f \in Z^Y \mid |f'(y)| \leqslant M \text{ on }]a, b[\}$. Then $\mathscr{F}$ is equicontinuous on $[a, b]$, since for each $f \in \mathscr{F}$ we have $|f(\xi) - f(\eta)| \leqslant M|\xi - \eta|$, so that $B(y_0, \varepsilon/M)$ is a nbd of ε-equicontinuity for $\mathscr{F}$.

6.2 Lemma Let Z^Y have the c-topology and $\mathscr{F} \subset Z^Y$ be equicontinuous on Y. Then its closure $\overline{\mathscr{F}}$ is also equicontinuous on Y.

Proof: This will be a simple consequence of

6.3 Let $\mathscr{F}$ be equicontinuous on Y, and for each $y \in Y$, let $\omega_y: Z^Y \to Z$ be the evaluation map $f \to f(y)$. Let $\mathfrak{A}$ be a filterbase on $\mathscr{F}$ such that $\omega_y(\mathfrak{A})$ converges to some $\varphi(y) \in Z$ for each y. Then:

(1). φ is continuous and $\mathfrak{A} \to \varphi$.

(2). If $U(y_0)$ is a nbd of ε-equicontinuity for $\mathscr{F}$, then
$$\varphi(U(y_0)) \subset B(\varphi(y_0), 3\varepsilon).$$

Proof of **6.3**: Let $y_0 \in Y$ be fixed. Given $\varepsilon > 0$, let $U(y_0)$ be a nbd of ε-equicontinuity, so that for each $f \in \mathscr{F}$, $f(U(y_0)) \subset B(f(y_0), \varepsilon)$. Since $\omega_{y_0}\mathfrak{A} \to \varphi(y_0)$, there is some $A_\alpha \in \mathfrak{A}$ with $\omega_{y_0}(A_\alpha) \subset B(\varphi(y_0), \varepsilon)$; that is, for each $f \in A_\alpha$, $f(y_0) \in B(\varphi(y_0), \varepsilon)$. It follows that $\forall\, y \in U(y_0)\; \forall f \in A_\alpha$: $d(f(y), \varphi(y_0)) \leqslant d(f(y), f(y_0)) + d(f(y_0), \varphi(y_0)) < 2\varepsilon$, so we have the following result:

(a). $$\forall\, y \in U(y_0): \omega_y A_\alpha \subset B(\varphi(y_0), 2\varepsilon).$$

Because of (a), we must have $\varphi(y) \in \overline{B(\varphi(y_0), 2\varepsilon)}$ for each $y \in U(y_0)$: indeed, since $\omega_y\mathfrak{A} \to \varphi(y)$, each nbd of $\varphi(y)$ must intersect every $\omega_y A_\beta$, and in particular, the $\omega_y A_\alpha \subset B(\varphi(y_0), 2\varepsilon)$ found in (a). Thus,

$$d(\varphi(y), \varphi(y_0)) < 3\varepsilon$$

for each $y \in U(y_0)$, and because ε and y_0 are arbitrary, this establishes the continuity of φ at each y_0 and also the assertion **6.3** (2).

To prove $\mathfrak{A} \to \varphi$, it suffices to show that if $\varphi \in (C, V)$, then there is some $A_\alpha \in \mathfrak{A}$ with $A_\alpha \subset (C, V)$. To this end, note that because $\varphi(C)$ is compact and $\varphi(C) \cap [Y - V] = \varnothing$, we have $d(\varphi(C), Y - V) = 3\varepsilon$ for some $\varepsilon > 0$. For each $c \in C$, let $U(c)$ be a nbd of ε-equicontinuity for $\mathscr{F}$ and extract a finite covering $U(c_1), \cdots, U(c_n)$ for the compact C. According to (a) there are $A_\iota \in \mathfrak{A}$ such that $\omega_y A_\iota \subset B(\varphi(c_\iota), 2\varepsilon)$ for each $y \in U(c_\iota)$; choosing an $A_\alpha \in \mathfrak{A}$ such that $A_\alpha \subset \bigcap_1^n A_\iota$, we find that whenever $f \in A_\alpha$, then $d(f(c), \varphi(c)) < 3\varepsilon$ for each $c \in C$, and consequently, that $A_\alpha \subset (C, V)$, as required.

Proof of **6.2**: For each $\varphi \in \overline{\mathscr{F}}$, there is a filterbase $\mathfrak{A}$ on $\mathscr{F}$ with $\mathfrak{A} \to \varphi$; because $\omega_y\colon Z^Y \to Z$ is continuous [**2.4**(1)], we have

$$\omega_y \mathfrak{A} \to \varphi(y)$$

for each y, consequently **6.3** applies and **6.3**(2) yields the result.

This leads to

6.4 Theorem (Arzela-Ascoli) Let (Z, d) be a metric space and Y an arbitrary space. Assume $\mathscr{F} \subset Z^Y$ satisfies:

(1). $\mathscr{F}$ is equicontinuous on Y.

(2). $\overline{\omega_y \mathscr{F}}$ is compact for each $y \in Y$ (that is, $\{f(y) \mid f \in \mathscr{F}\}$ has compact closure for each y).

Then $\overline{\mathscr{F}}$ is compact (and equicontinuous) on Y.

Proof: By **6.2**, $\overline{\mathscr{F}}$ is equicontinuous on Y. To prove $\overline{\mathscr{F}}$ compact, let $\mathfrak{M}$ be a maximal filterbase on $\overline{\mathscr{F}}$; note that for each $y \in Y$, the map $\omega_y\colon Z^Y \to Z$ is continuous, so we have $\omega_y \overline{\mathscr{F}} \subset \overline{\omega_y \mathscr{F}}$. Consequently, for each $y \in Y$, $\omega_y \mathfrak{M}$ is a maximal filterbase (X, **7.5**) on the compact $\overline{\omega_y \mathscr{F}}$, and therefore converges to some $\varphi(y) \in Z$. Using **6.3**, we conclude that there is a continuous $\varphi \in \overline{\overline{\mathscr{F}}} = \overline{\mathscr{F}}$, with $\mathfrak{M} \to \varphi$, so by XI, **1.3**, $\overline{\mathscr{F}}$ is compact.

7. Sequential Convergence in the c-Topology

We shall show in this section that whenever Z is a metric space, sequential convergence in Z^Y is precisely the classical function-theoretic uniform convergence on every compact subset.

7.1 Definition Let (Z, d) be a metric space and Y an arbitrary space. A sequence $\{f_n\}$ in Z^Y is said to converge to an $f \in Z^Y$ uniformly on every compact subset if for each compact $C \subset Y$ and each $\varepsilon > 0$ there is an integer $N = N(C, \varepsilon)$ such that $d(f(c), f_n(c)) < \varepsilon$ for every $n \geqslant N$ and every $c \in C$.

7.2 Theorem Let (Z, d) be a metric space and Y an arbitrary space. A sequence $\{f_n\}$ in Z^Y converges to an $f \in Z^Y$ uniformly on every compact subset if and only if $f_n \to f$ in the c-topology of Z^Y.

Proof: Assume that $f_n \to f$ according to **7.1**, and let $f \in (C, V)$. Since C is compact, we have $d(f(C), Z - V) = \varepsilon > 0$. Let now $N = N(C, \varepsilon)$; then for $n \geqslant N$, we have $d(f_n(c), f(c)) < \varepsilon$ for all $c \in C$, and therefore $f_n \in (C, V)$ for all $n \geqslant N$. Thus $f_n \to f$ in the c-topology. Conversely, let $f_n \to f$ in the c-topology. Given any compact C and $\varepsilon > 0$, each $c \in C$ has a nbd $U(c)$ with $f(U(c)) \subset B(f(c), \varepsilon)$; covering C with finitely many $U(c_1) \cap C, \cdots, U(c_n) \cap C$ and shrinking, we obtain compact $C_1, \cdots, C_n$, with $C_\iota \subset U(c_\iota)$, $C = \bigcup_1^n C_\iota$. Letting W be the nbd $\bigcap_1^n (C_\iota, B(f(c_\iota), \varepsilon))$ of f, we find that if $f_n \in W$, then $d(f_n(c), f(c)) < 2\varepsilon$ for all $c \in C$.

In **7.2**, we have assumed that the limit function is continuous; if we drop this continuity requirement, we have

7.3 Let $\{f_n\}$ be a sequence in Z^Y converging uniformly on every compact subset to some function f. Then $f \mid C$ is continuous for each compact $C \subset Y$. In particular, if Y is a k-space, then the limit of a sequence of continuous functions converging uniformly on every compact set is always continuous.

Proof: Let C be compact, and $f_n \to f$ as in **7.1**; we prove that f is continuous at each $c_0 \in C$. Given $\varepsilon > 0$, there is an integer n with $d(f(c), f_n(c)) < \varepsilon$ for all $c \in C$; from

$$d(f(c), f(c_0)) \leqslant d(f(c), f_n(c)) + d(f_n(c), f_n(c_0)) + d(f_n(c_0), f(c_0))$$

we find $d(f(c), f(c_0)) < 3\varepsilon$ on the nbd $C \cap f_n^{-1}[B(f_n(c_0), \varepsilon)]$ of c_0, completing the proof.

Another type of convergence, closely related to that in **7.1**, is given in

7.4 Definition A sequence $\{f_n\}$ in Z^Y converges continuously to an $f \in Z^Y$ if $f_n(y_n) \to f(y)$ for each $y \in Y$ and sequence $y_n \to y$.

7.5 Let (Z, d) be a metric space, and let Y be $1°$ countable. Then, continuous convergence in Z^Y is equivalent to uniform convergence on every compact subset.

Proof: Assume $f_n \to f$ continuously, but not uniformly on some compact set $C \subset Y$. Then there is an $\varepsilon > 0$, a sequence of integers $n_1 < n_2 < \cdots$, and a sequence of points $c_k \in C$ such that $d(f_{n_k}(c_k), f(c_k)) \geqslant \varepsilon$ for each $k \geqslant 1$. Since C is compact and $1°$ countable, we can assume that $c_k \to c_0$; otherwise, we use a subsequence and X, **6.1**(2). Due to continuous convergence, it follows that $f_{n_k}(c_k) \to f(c_0)$; and, by the continuity of f, we find that $d(f_{n_k}(c_k), f(c_k)) \leqslant d(f_{n_k}(c_k), f(c_0)) + d(f(c_0), f(c_k)) \to 0$, which is a contradiction. Conversely, assume $f_n \to f$ uniformly on every compact set, and let $y_n \to y_0$; then $C = \{y_n\} \cup y_0$ is compact, so that $f_n|C \to f|C$ uniformly; thus, from

$$d(f_n(y_n), f(y_0) \leqslant d(f_n(y_n), f(y_n)) + d(f(y_n), f(y_0))$$

and the continuity of f, we find that $f_n(y_n) \to f(y_0)$.

8. Metric Topologies; Relation to the c-Topology

If Z is a metric space, each of its metrics can be used to impose a metric topology on a suitable subset of the *set* Z^Y.

8.1 Definition Let (Z, d) be a metric space and Y an arbitrary space. Let $C(Y, Z; d)$ be the set of all d-bounded continuous maps of Y into Z, that is $C(Y, Z; d) = \{f \in Z^Y \mid \delta f(Y) < \infty\}$. The metric topology in $C(Y, Z; d)$ is that determined by the metric

$$d^+(f, g) = \sup\{d(f(y), g(y)) \mid y \in Y\}.$$

Verification that d^+ is in fact a metric is routine, once it has been established that $d^+(f, g)$ is always finite, and this follows by noting that if $y_0 \in Y$ be fixed, then for each $y \in Y$,

$$\begin{aligned} d(f(y), g(y)) &\leqslant d(f(y), f(y_0)) + d(f(y_0), g(y_0)) + d(g(y_0), g(y)) \\ &\leqslant \delta f(Y) + d(f(y_0), g(y_0)) + \delta g(Y), \end{aligned}$$

a fixed finite constant.

It is important to observe that the metric d in Z plays a dual role: It determines the *set* $C(Y, Z; d)$ and imposes a metric topology on this set. Equivalent metrics in Z may determine distinct sets. For example, with $Y = Z = E^1$, if d_e is the Euclidean metric, then $1: E^1 \to E^1$ does not belong to $C(Y, Z; d_e)$, whereas it does belong to $C(Y, Z; d_1)$ for any equivalent bounded metric d_1 in E^1.

Since $C(Y, Z; d) \subset Z^Y$, the set $C(Y, Z; d)$ inherits two topologies: the metric topology determined by d^+ and the subspace c-topology. The relation between these topologies is summarized in

8.2 (1). If neither Z nor Y is compact, then for any bounded metric d in Z, the *set* $C(Y, Z; d) = Z^Y$. However, the metric topology in Z^Y depends not only on the topologies of Y and Z, but also on the particular bounded metric used in Z, and may be different for *equivalent* bounded metrics.

(2). If Z alone is compact, then for any metric d giving the topology of Z, the *set* $C(Y, Z; d) = Z^Y$. The metric topology in Z^Y is a joint topological invariant of the spaces Y and Z (in particular, it is independent of the metric used to give the topology of Z); but it may *not* coincide with the c-topology of Z^Y.

(3). If Y is compact, then for any metric space (Z, d), the *set* $C(Y, Z; d) = Z^Y$. The metric topology in Z^Y is a joint topological invariant of the spaces Y and Z and always coincides with the c-topology of Z^Y.

Proofs and Examples: (1). To see that the set $C(Y, Z; d) = Z^Y$ if d is a bounded metric for Z is trivial. The following example will show that equivalent bounded metrics in Z may determine distinct metric topologies in the set Z^Y. Let $Y = Z = E^1$, and in Z take the two bounded metrics

$$d_1(x, z) = \min\{1, |x - z|\}, \quad d_2(x, z) = \left|\frac{x}{1 + |x|} - \frac{z}{1 + |z|}\right|$$

These are equivalent, since each is equivalent to the Euclidean metric in E^1. Let $f: Y \to Z$ be the identity map $f(y) = y$, and for each $n = 1, 2, \cdots$, let $f_n: Y \to Z$ be the (continuous) map:

$$f_n(y) = \begin{cases} y & y < n \\ n & y \geq n. \end{cases}$$

Then, clearly, $d_1^+(f, f_n) = 1$ for each n; but since $d_2^+(f, f_n) < 1/n$, the d_1^+-nbd of radius 1 of f can contain no d_2^+-nbd of f.

(2). Since Z is compact, any metric for Z is necessarily bounded, so that $C(Y, Z; d) = Z^Y$. To show that the metric space $C(Y, Z; d)$ is a joint topological invariant of Y and Z, we prove that if $\varphi: Y \cong Y_1$ and $\psi: Z \cong Z_1$, then the map $F: C(Y, Z; d) \to C(Y_1, Z_1; d_1)$, given by $F(f) = \psi \circ f \circ \varphi^{-1}$, is a homeomorphism. Define

$$F_1: C(Y_1, Z_1 : d_1) \to C(Y, Z; d)$$

by $F_1(f_1) = \psi^{-1} \circ f_1 \circ \varphi$; clearly, F, F_1 are inverses, so only (III, **12.3**) the continuity of each need be established, and by symmetry, only that of F. Let $B(F(f_0), \varepsilon)$ be given. Because Z is compact, ψ is uniformly continuous, and therefore there is a $\delta > 0$ such that

$$\psi(B(z, \delta)) \subset B_1(\psi(z), \varepsilon)$$

for each $z \in Z$. Then $F(B(f_0, \delta)) \subset B_1(F(f_0), \varepsilon)$, as required: For, $d^+(f, f_0) < \delta$ implies that $d(f(y), f_0(y)) < \delta$ for every $y \in Y$, and consequently that $d_1(\psi \circ f_0 \circ \varphi^{-1}(y_1), \ \psi \circ f \circ \varphi^{-1}(y_1)) < \varepsilon$ for every $y_1 \in Y_1$. To see that the metric topology need not coincide with the

c-topology, let Z be the discrete space 2 and Y the discrete space Z^+; the metric topology in Z^Y is evidently discrete and, having $2^{\aleph_0}$ elements, is not 2° countable. However, by **5.2**, the c-topology in Z^Y is in fact 2° countable.

(3). Since the set $f(Y)$ is compact for each continuous $f: Y \to Z$, it is (XI, **4.2**) bounded in Z, and consequently the set $C(Y, Z; d) = Z^Y$. We now show that the metric topology in Z^Y coincides with the c-topology of Z^Y.

(a). Each ball $B(f, \varepsilon)$ contains a c-nbd of f. For, the compact Y can be covered by finitely many open sets $U_1, \cdots, U_n$ such that $\delta f(\overline{U}_i) < \varepsilon/4$ for each $i = 1, \cdots, n$. Letting V_i be an $\varepsilon/4$-nbd of the compact $f(\overline{U}_i)$, we have $\delta V_i < \frac{3}{4}\varepsilon$, and so $f \in \bigcap_1^n (\overline{U}_i, V_i) \subset B(f, \varepsilon)$.

(b). Each c-nbd (A, V) of f contains a ball $B(f, \varepsilon)$. For, since $f(A)$ is compact, $\varepsilon = d(f(A), Y - V)$ is positive because

$$f(A) \cap [Y - V] = \varnothing,$$

and therefore $f \in B(f, \varepsilon/2) \subset (A, V)$.

Unlike the c-topology, in *any* metric space $C(Y, Z; d)$ we have

8.3 The evaluation map $\omega: C(Y, Z; d) \times Y \to Z$ is always continuous.

Proof: Given $U = B(f_0(y_0), \varepsilon) \subset Z$, let $V = f_0^{-1}(U)$, which is an open set containing y_0, and let $W = B(f_0, \varepsilon) \subset C(Y, Z; d)$; then $\omega(W, V) \subset B(f_0(y_0), 2\varepsilon)$, since $[f \in W] \Rightarrow [d(f(y), f_0(y)) < \varepsilon$ for all $y \in Y]$ and $[y \in V] \Rightarrow [d(f_0(y), f_0(y_0)) < \varepsilon]$.

In the c-topology, sequential convergence is equivalent to uniform convergence on every compact set; in any metric topology, we have

8.4 In any metric space $C(Y, Z; d)$:

(1). A sequence $\{f_n\} \subset C(Y, Z; d)$ converges to an $f \in C(Y, Z; d)$ in the metric topology if and only if $f_n \to f$ uniformly *on* Y.

(2). If a sequence $\{f_n\} \subset C(Y, Z; d)$ converges to some function f uniformly *on* Y, then f is continuous and belongs to $C(Y, Z; d)$.

Proof: (1). We need observe only that always $d(f(y), f_n(y)) \leqslant d^+(f, f_n)$.

(2). The proof that f is continuous is identical to that of **7.3**; to see that $f \in C(Y, Z; d)$, note that $d(f(y), f_n(y)) < 1$ for all $y \in Y$ and some sufficiently large n, so that $\delta f(Y) \leqslant \delta f_n(Y) + 2$.

Remark: The result **8.2**(3) can be regarded as stating that if Y is compact and Z is metrizable, then the c-topology in Z^Y is also metrizable. Metrization of the c-topology is also possible whenever Y is σ-compact, although in this case the metric cannot be simply that imposed by a metric for Z: in the example of **8.2**(2), the c-topology in Z^Y is 2° countable and regular and, consequently, it is metrizable; but there is no metric d in Z such that d^+ metrizes the c-topology.

8.5 Let Y be σ-compact and write $Y = \bigcup_1^\infty U_i$, where the U_i are open, $\overline{U}_i \subset U_{i+1}$ and $\overline{U}_i$ is compact for each i. Let Z be a metrizable space and d any metric that gives its topology. For $f, g \in Z^Y$, and each $n = 1, 2, \cdots$, define

$$d_n(f, g) = \min\left[\frac{1}{n}, \sup\{d(f(y), g(y)) \mid y \in \overline{U}_n\}\right].$$

Then $\rho(f, g) = \sup_n d_n(f, g)$ metrizes the c-topology in Z^Y.

The proof, which is entirely analogous to that in **8.2**(3), is left for the reader.

9. Pointwise Convergence

We have seen that the c-topology describes uniform convergence on every compact set. Another type of convergence frequently used in analysis is given in

9.1 Definition A sequence $\{f_n\} \subset Z^Y$ converges pointwise to an $f \in Z^Y$ if $f_n(y) \to f(y)$ for each fixed $y \in Y$.

Clearly, uniform convergence on every compact set implies pointwise convergence; the usual examples in analysis show (1) that the converse is not true and (2) that if a sequence $\{f_n\} \subset Z^Y$ converges pointwise to some function f, then f need not be continuous.

To obtain a topology for the set Z^Y in which sequential convergence is equivalent to pointwise convergence, we make the

9.2 Definition In the set Z^Y, the p-topology is that having the family $\{(y, V) \mid y \in Y, \text{all open } V \subset Z\}$ as subbasis. This space is denoted by $Z^Y(p)$.

The p-topology is evidently contained in the c-topology, since the compact sets used in defining the subbasis for the c-topology are replaced by single points. The verification that sequential convergence in $Z^Y(p)$ is equivalent to pointwise convergence is immediate, once we establish the following equivalent description of $Z^Y(p)$:

9.3 $Z^Y(p)$ is homeomorphic to a subspace of the cartesian product $\prod\{Z_y \mid y \in Y\}$, where each Z_y is a copy of Z. Precisely: If for each $f \in Z^Y$, we define $\mu(f)$ = point in $\prod_y Z_y$ with yth coordinate $f(y)$, then $Z^Y(p)$ is homeomorphic to $\mu(Z^Y)$.

Proof: Clearly, μ is injective. Let $\bigcap_1^n (y_\iota, U_\iota)$ be any basic open set in $Z^Y(p)$; defining $U_{y_i} \subset Z_{y_i}$ to be the set U_ι, we have $\mu[\bigcap_1^n (y_\iota, U_\iota)] = \langle U_{y_1}, \cdots, U_{y_n}\rangle \cap \mu(Z^Y)$, which establishes that μ is a homeomorphism.

Since convergence in cartesian products is equivalent to coordinate convergence, it is evident from **9.3** that convergence in $Z^Y(p)$ is equivalent to pointwise convergence.

9.4 (a). For each fixed $y_0 \in Y$, the evaluation map $\omega_{y_0}: Z^Y(p) \to Z$ is continuous.

(b). If $f: X \to Y$ is continuous, the induced map $Z^Y(p) \to Z^X(p)$ is also continuous.

Proof: (a). We need note only that $\omega_{y_0} = p_{y_0} \circ \mu$ where

$$p_{y_0}: \prod Z_y \to Z_{y_0} = Z$$

is the projection onto the y_0th factor. The proof of (b) is left for the reader.

Let Z be a metric space and Y any space. If we choose a bounded metric d in Z, then sequential convergence in the metric space (Z^Y, d^+) is equivalent to uniform convergence on Y; alternatively stated: the concept of uniform convergence on Y is metrizable. The question arises if pointwise convergence can also be metrized in some manner. If Y is discrete, then because $Z^Y(p)$ is homeomorphic to $\prod\{Z_y \mid y \in Y\}$, metrization is possible if and only if $\aleph(Y) \leqslant \aleph_0$. However, in the general case, the possibility remains that although $\prod Z_y$ is not metrizable, the subspace $\mu(Z^Y)$ is. For example, if Y is any connected space and Z is any totally disconnected metric space, then Z^Y consists only of constant maps and $\mu(Z^Y)$, being homeomorphic to Z, is therefore metrizable.

9.5 Let Y be completely regular and Z metric. Assume that $\aleph(Y) > \aleph_0$ and that Z is not totally pathwise disconnected. Then pointwise convergence is *not* metrizable; indeed, there does not exist any real-valued function ρ on $Z^Y \times Z^Y$ having just the two properties:

(1). $\rho(f, \varphi) \geqslant 0$.

(2). $\rho(f_n, \varphi) \to 0$ if and only if $f_n \to \varphi$ pointwise.

Proof: We need some preliminary remarks. Let d be a metric for Z. By hypothesis, there is a nonconstant continuous $p: I \to Z$; let

$$\sup\{d[p(0), p(t)] \mid t \in I\} = \mu,$$

which is positive. The existence of p assures that for each $b \leqslant \mu$, each $y \in Y$, and each open $U \supset \{y\}$, there is a continuous $f: Y \to Z$ such that $f(Y - U) = p(0)$ and $d(p(0), f(y)) = b$: for, since $d[p(0), p(t)]$ is continuous on the connected I, there is at least one $t_0 \in I$ such that $d(p(0), p(t_0)) = b$; complete regularity gives a continuous $g: Y \to I$ such that $g(y) = t_0$, $g(Y - U) = 0$, and then $f = p \circ g$ is the required map.

We are now ready to prove that the existence of ρ leads to a contradiction. Let $\varphi: Y \to Z$ be the constant map $y \to p(0) \in Z$. For each fixed $y_0 \in Y$ and each $n \in Z^+$, let

$$\Delta(y_0, n) = \sup \{d[f(y_0), \varphi(y_0)] \mid f \in Z^Y, \rho(f, \varphi) < 1/n\}.$$

Then for each fixed y_0 we must have $\Delta(y_0, n) \to 0$ as $n \to \infty$: otherwise there would be arbitrarily large n_k with $\Delta(y_0, n_k) \geqslant \epsilon > 0$ for some fixed $\epsilon > 0$, so we would have a sequence $\{f_{n_k}\}$ such that $\rho(f_{n_k}, \varphi) < 1/n_k$, whereas

$$d[f_{n_k}(y_0), \varphi(y_0)] \geqslant \epsilon/2,$$

in contradiction to the condition (2) on ρ.

Now let $E_n = \{y \in Y \mid \Delta(y, n) < \mu/2\}$ where μ is the constant in the introductory remarks. We have $\bigcup_n E_n = Y$ by what we have already proved, and since $\aleph(Y) > \aleph_0$, at least one E_n, say E_{n_0}, must be infinite. By VII, **2.4**, we can find a family $\{U_i \mid i \in Z^+\}$ of open sets in Y, whose closures are pairwise disjoint, and such that $U_i \cap E_{n_0} \neq \varnothing$ for every $i \in Z^+$. For each i, choose a $y_i \in U_i \cap E_{n_0}$ and let $f_i: Y \to Z$ be a continuous map such that $f_i(Y - U_i) = p(0)$ and $d[p(0), f_i(y_i)] = \Delta(y_i, n_0) + \mu/2$. Since the $\overline{U}_i$ are pairwise disjoint, the sequence $\{f_i\}$ converges pointwise to φ: but since $d[f_i(y_i), \varphi(y_i)] > \Delta(y_i, n_0)$, it must be true that $\rho(f_i, \varphi) \geqslant 1/n_0$ for every $i \in Z^+$, and this is the required contradiction.

10. Comparison of Topologies in Z^Y

Let Y, Z be arbitrary spaces. We take the properties in **3.1** as a basis for comparing topologies in the *set* Z^Y. To indicate that a topology t is used in Z^Y, we write $Z^Y(t)$.

10.1 Definition A topology t in Z^Y is *splitting* (or proper) if for every space X, the continuity of $\alpha: X \times Y \to Z$ implies that of its associate $\hat{\alpha}: X \to Z^Y(t)$. A topology t in Z^Y is *conjoining* (or admissible) if for every space X, the continuity of $\hat{\alpha}: X \to Z^Y(t)$ implies that of $\alpha: X \times Y \to Z$.

10.2 The topology t is conjoining if and only if the evaluation map $\omega: Z^Y(t) \times Y \to Z$ is continuous.

Proof: If ω is continuous, it follows as in the proof of **3.1**.(2) that t is conjoining. For the converse, $\hat{\omega} = 1: Z^Y(t) \to Z^Y(t)$ is obviously continuous and therefore ω is also continuous.

Ex. 1 The discrete topology in Z^Y is conjoining, and by **8.3**, so also is any metric topology in Z^Y. The indiscrete topology is splitting. The c-topology is always splitting, and for locally compact Y, also conjoining.

Ex. 2 The significance of the requirement for conjoining topologies (that for *every* space X the continuity of $\hat{\alpha}$ implies that of α) is illustrated by the space I^Q of **2**, Ex. 1: the c-topology in I^Q is *not* conjoining, but for any $1°$ *countable* X, the continuity of $\hat{\alpha}: X \to I^Q(c)$ does imply that of $\alpha: X \times Q \to I$ (*cf.* **3.2**).

10.3 (a). A topology larger than a conjoining topology is also conjoining.
(b). A topology smaller than a splitting topology is also splitting.
(c). Any conjoining topology is larger than any splitting topology.
(d). There is always a unique largest splitting topology.

Proof: (a). Let t be conjoining and $t \subset u$; since $1: Z^Y(u) \to Z^Y(t)$ is continuous, the conjoining property of t shows that $\omega: Z^Y(u) \times Y \to Z$ is continuous and, by **10.2**, that u is conjoining. The proof of (b) is similar.

(c). Let t be conjoining and u be splitting. Then $\omega: Z^Y(t) \times Y \to Z$ is continuous, and since u is splitting, we find that $1: Z^Y(t) \to Z^Y(u)$ is continuous, which shows that $u \subset t$.

(d). Let $\{t_\alpha\}$ be the family of all splitting topologies in Z^Y, and let t be that topology having the members of $\bigcup_\alpha t_\alpha$ as subbasis; we need prove only that t is splitting. Let $\alpha: X \times Y \to Z$ be continuous; since any subbasic $U \in t$ belongs to some splitting t_α, we must have $\hat{\alpha}^{-1}(U)$ open in X, consequently $\hat{\alpha}: X \to Z^Y(t)$ is continuous.

There is no known characterization of the largest splitting topology directly in terms of the topological structures of Y and Z, unless Y is locally compact.

10.4 A function space Z^Y can have at most one topology that is both conjoining and splitting. Such a topology is necessarily the largest splitting topology and the smallest conjoining topology. If Y is locally compact, this topology exists and is the c-topology.

Proof: The necessary uniqueness of such a topology follows from **10.3**(c) and (d) and the rest from **3.1**.

If Y is not locally compact, a topology in Z^Y that is both conjoining and splitting need not exist. For example, R. H. Fox and R. Arens have shown that the set I^Q of **2**, Ex. 3, has no smallest conjoining topology, so that it cannot have a topology that is both conjoining and splitting.

Problems

Section 1

1. Prove in detail that if $X \cong L$, and $Y \cong M$, then $Y^X \cong M^L$.
2. If Y is a Hausdorff space, prove that the embedding $j: Y \to Y^X$ is a closed mapping, that is, that $j(Y)$ is a closed subspace of Y^X.

3. Let Z be Hausdorff. Prove that if the c-topology is used in Z^Y, then
$$\{(f, y, z) \mid f(y) = z\} \subset Z^Y \times Y \times Z$$
is closed.

Section 2

1. Let $f: X \to Y$ be surjective. Prove: $f^+ : Z^Y \to Z^X$ is injective for any space Z.
2. Let Y be locally compact. Prove: If $F \subset Z$ is a fixed closed set, then
$$\{(f, y) \mid f(y) \in F\}$$
is closed in $Z^Y \times Y$.
3. Let Y be compact, and G open in Y. Show that $\{(f, z) \mid f^{-1}(z) \subset G\}$ is open in $Z^Y \times Z$.
4. Let Y be compact, and $F \subset Z$ closed, and $G \subset Y$ open. Prove that
$$\{f \mid f^{-1}(F) \subset G\}$$
is open in Z^Y.

Section 3

1. Prove: The evaluation map $\omega: Z^Y \times Y \to Z$ is continuous if and only if for every space X, the continuity of $\hat{\alpha}: X \to Z^Y$ implies that of $\alpha: X \times Y \to Z$.
2. Show that the identity map $1: Z^Y \to Z^Y$ is the associate of the evaluation map $\omega: Z^Y \times Y \to Z$.

Section 4

1. Let $p: X \to R$ be an identification. Prove that if $Y \times R$ is a k-space, then $p \times 1: X \times Y \to R \times Y$ is an identification.
2. Let $f: P \to X$, $g: R \to Y$ be identifications. Show that if X, R, Y are 1° countable, then $f \times g: P \times R \to X \times Y$ is an identification.
3. Let X be paracompact, $A \subset X$ closed. Let Y be compact, and let $f: A \to I^Y$ be continuous. Show that f has a continuous extension $F: X \to I^Y$.
4. Let A_i, $i = 1, 2, \cdots, n$ be compact subsets of a space X, and let Y be any compact space. Show that the join $[\bigcup_1^n A_i] * Y \cong \bigcup_1^n [A_i * Y]$, and that $[\bigcap_1^n A_i] * Y \cong \bigcap_1^n [A_i * Y]$.

Section 5

1. Let X, Y, Z be arbitrary spaces. Prove: $(Y \times Z)^X \cong Y^X \times Z^X$.
2. Let Y and Z be 2° countable and Y compact. Prove: Z^Y is metrizable if and only if Z is regular.

Section 6

1. Let (Z, d) be a metric space, and let Y be locally compact. Assume that $\overline{\mathscr{F}}$ is compact. Prove:
   a. $\mathscr{F}$ is equicontinuous on Y.
   b. $\overline{\omega_y \mathscr{F}}$ is compact for each $y \in Y$.
2. Show that any equicontinuous family of continuous maps of a compact space into itself has compact closure.

Section 7

1. Let Z be a metric space and Y a k-space. Let $\{f_n\} \subset Z^Y$ be an equicontinuous sequence of functions that converges for each fixed $y \in Y$. Prove: The sequence $\{f_n\}$ converges uniformly on every compact subset of Y to a continuous f.
2. Let $\mathscr{F}$ be a uniformly bounded equicontinuous family of continuous real-valued functions on E^n. Show that one can always extract a sequence $\{f_n\} \subset \mathscr{F}$ that converges uniformly on every compact set to a continuous $f\colon E^n \to E^1$.

Section 8

1. Let $Y = Z = E^1$. For each integer n and $f, g \in Z^Y$, let
$$M_n = \sup\{|f(t) - g(t)| \mid -n \leqslant t \leqslant n\},$$
and define
$$\rho(f, g) = \sum_1^\infty \frac{1}{2^n} \cdot \frac{M_n}{1 + M_n}.$$
Show that ρ metrizes the c-topology of Z^Y.
2. Let X be a compact metric space, let Y be an arbitrary metric space, and let $f\colon X \to Y$ be continuous. Referring to the spaces defined in IX, Problem 4.8, prove that the map $f_*\colon \mathscr{F}(X) \to \mathscr{F}(Y)$ defined by $A \to f(A)$ is continuous.
3. Let Y be a compact metric space, X an arbitrary space, and $A \subset X$. For each $f \in Y^X$ define $\alpha(f) = \overline{f(A)}$. Prove that if Y^X is taken with the metric topology, then $\alpha\colon Y^X \to \mathscr{F}(Y)$ is continuous.

Section 9

1. Let X and Y be arbitrary spaces, and let $\{f_n \mid n \in Z^+\}$ be a sequence of continuous maps $f_n\colon X \to Y$. Assume that the sequence converges pointwise to some function f. Show that if Y is 2° countable, then there exists a nowhere dense set $N \subset X$ such that $f \mid X - N$ is continuous.
2. Let Z be Hausdorff (regular, completely regular). Prove: $Z^Y(p)$ is also Hausdorff (regular, completely regular) for any space Y.

Section 10

1. Show that the p-topology is splitting in Z^Y. Show that if Y is completely regular and contains a homeomorph of I, if Z is arbitrary and also contains a homeomorph of I, then the p-topology in Z^Y is never conjoining.

The Spaces C(Y)

XIII

In this chapter, we study the set of continuous real-valued functions on a space Y; in particular, we obtain the algebraico-topological Stone-Weierstrass theorem, which is one of the fundamental facts of modern analysis.

1. Continuity of the Algebraic Operations

1.1 Definition The set $(E^1)^Y$ of all continuous real-valued functions on Y, with the c-topology, is denoted by $\hat{C}(Y; c)$.

1.2 In $\hat{C}(Y; c)$ the addition, multiplication, and scalar multiplication are continuous operations; that is, the maps

$$(f, g) \to f + g,\ (f, g) \to f \cdot g \text{ of } \hat{C}(Y; c) \times \hat{C}(Y; c) \to \hat{C}(Y; c),$$

and the map $(\lambda, f) \to \lambda f$ of $E^1 \times \hat{C}(Y; c) \to \hat{C}(Y; c)$ are each continuous (in both variables).

Proof: Let $\alpha: \hat{C}(Y; c) \times \hat{C}(Y; c) \to \hat{C}(Y; c)$ be the map

$$(f, g) \to f + g$$

and let (A, W) be a given nbd of $f + g$. Since the addition operation $(\lambda, \mu) \to \lambda + \mu$ for real numbers is continuous, we find that for each $a \in A$

there are nbds $U(f(a))$, $V(g(a))$ such that $U(f(a)) + V(g(a)) \subset W$, and, since f, g are continuous, that there is a nbd $U = U(a)$ with $f(U) \subset U(f(a))$, $g(U) \subset V(g(a))$. From the open covering

$$\{U(a) \cap A \mid a \in A\}$$

of the compact A, extract a finite covering $U(a_1) \cap A, \cdots, U(a_n) \cap A$ and shrink it to obtain compact sets $A_1, \cdots, A_n$ satisfying $A_\iota \subset U(a_\iota)$, $\bigcup_1^n A_\iota = A$. Then $U_0 = \bigcap_1^n (A_\iota, U(f(a_\iota)))$ is a nbd of f, $V_0 = \bigcap_1^n (A_\iota, V(g(a_\iota)))$ is a nbd of g, and $\alpha(U_0, V_0) \subset (A, W)$. The map α is therefore continuous. The continuity of multiplication, and of scalar multiplication, is proved analogously.

By referring to Appendix I, proposition **1.2** implies that $\hat{C}(Y; c)$ is a linear topological space. Even more,

1.3 Theorem $\hat{C}(Y; c)$ is a locally convex linear topological space.

Proof: Since the open intervals $]c, d[$ are a basis for E^1, XII, **5.1**, assures that the subbasic sets of form $(A,]c, d[)$ themselves form a subbasis for $\hat{C}(Y; c)$. Now, each $(A,]c, d[)$ is convex: For, if $f, g \in (A,]c, d[)$, then for each $a \in A$ and $0 \leqslant \lambda \leqslant 1$, the point $\lambda f(a) + (1 - \lambda)g(a)$ on the line segment joining $f(a)$ and $g(a)$ also lies in $]c, d[$ so

$$\lambda f + (1 - \lambda)g \in (A,]c, d[)$$

also. Since the intersection of convex sets is convex, $\hat{C}(Y; c)$ therefore has a basis consisting of convex sets, and so $\hat{C}(Y; c)$ is locally convex.

2. Algebras in $\hat{C}(Y; c)$

2.1 Definition A subset $A \subset \hat{C}(Y; c)$ is called an algebra if whenever f, g belong to A, so also do $f + g$, $f \cdot g$, and λf for each real λ. A is a *unitary* algebra if it also contains the constant function $\delta(y) \equiv 1$.

Ex. 1 $\hat{C}(Y; c)$ is itself a unitary algebra.

Ex. 2 Every algebra contains the constant function $f(x) \equiv 0$; an algebra is unitary if and only if it contains a nonzero constant function.

Ex. 3 In $\hat{C}(I; c)$, the set of all polynomials without constant term is a non-unitary algebra.

Ex. 4 If A is a unitary algebra and $f \in A$, then for each polynomial $p(x)$ in a single variable x, the continuous function $p(f)$ belongs to A also. More generally, if $p(x_1, \cdots, x_n)$ is any polynomial in n variables, and if $f_1, \cdots, f_n \in A$, then $p(f_1, \cdots, f_n)$ belongs to A also.

2.2 If $A \subset \hat{C}(Y; c)$ is an algebra, so also is its closure $\bar{A}$ in the space $\hat{C}(Y; c)$.

Proof: Let $\alpha: \hat{C}(Y; c) \times \hat{C}(Y; c) \to \hat{C}(Y; c)$ be the map

$$(f, g) \to f + g.$$

Then $\alpha(A \times A) \subset A$ because A is an algebra, and

$$\alpha(\overline{A \times A}) \subset \overline{\alpha(A \times A)}$$

because α is continuous; thus (IV, **1.2**) $\alpha(\bar{A} \times \bar{A}) \subset \bar{A}$ and the sum of two functions in $\bar{A}$ also belongs to $\bar{A}$. Similarly, the continuity of $(f, g) \to f \cdot g$ and $(\lambda, f) \to \lambda f$ show that the remaining requirements for an algebra are satisfied by $\bar{A}$.

The main result of this section is

2.3 Theorem Let $A \subset \hat{C}(Y; c)$ be a unitary algebra. If $f \in A$, then $|f| \in \bar{A}$.

This will follow from the

Lemma There exists a sequence $\{p_n(t)\}$ of polynomials that converges uniformly on $[0, 1]$ to the function $\varphi(t) = \sqrt{t}$.

Proof of Lemma: The binomial theorem assures that

$$1 - \sqrt{1 - x} = \sum_1^\infty \left| \binom{\frac{1}{2}}{n} \right| x^n$$

is certainly valid on $[0, 1[$. We first show that the series converges also for $x = 1$. Indeed, because all terms of the series are nonnegative, each partial sum s_n satisfies $0 \leqslant s_n(x) \leqslant 1 - \sqrt{1 - x} \leqslant 1$ on $[0, 1[$; since each s_n is continuous at $x = 1$, we have $s_n(1) \leqslant 1$ also, and since it is a bounded monotone sequence, $\{s_n(1)\}$ therefore converges. Next, the convergence of the series is uniform on $[0, 1]$: For, if $R_n(x)$ is the remainder after n terms, we need observe only that $0 \leqslant R_n(x) \leqslant R_n(1)$ because all terms of the series are monotone increasing functions. Now, letting $x = 1 - t$, we conclude that $s_n(1 - t) \to 1 - \sqrt{t}$ uniformly on $[0, 1]$, so the polynomials $p_n(t) = 1 - s_n(1 - t)$ fulfill the requirement of the lemma.

Proof of Theorem: Let $f \in A$. We are to prove that each nbd $\bigcap_1^n (C_i, V_i)$ of $|f|$ contains some $g \in A$. Letting $d_e(x, y)$ be the metric $|x - y|$ in E^1, and letting ε be the positive number

$$\min \{d_e[|f|(C_i), Y - V_i] \mid i = 1, \cdots, n\},$$

it suffices to find a $g \in A$ such that $||f|(c) - g(c)| < \varepsilon$ for each

$$c \in C = \bigcup_1^n C_i.$$

Since C is compact, we have $|f(c)| \leqslant B$ for some constant $B < \infty$ and all $c \in C$, so using the polynomials p_n of the lemma, we find $p_n(f^2/B^2) \to \sqrt{f^2/B^2} = |f|/B$ uniformly on C. Observing now that each $p_n(f^2/B^2) \in A$, the proof is complete.

2.4 Corollary Let A be a unitary algebra in $\hat{C}(Y; c)$. If $f, g \in \bar{A}$, then the functions $\max[f, g]$ and $\min[f, g]$ also belong to $\bar{A}$.

Proof: We have

$$\max(f, g) = \frac{f + g}{2} + \frac{|f - g|}{2};$$

since $f, g \in \bar{A}$, **2.2** shows $f \pm g \in \bar{A}$ and **2.3** shows $|f - g| \in \bar{\bar{A}} = \bar{A}$. Thus, by **2.2** again, $\max(f, g) \in \bar{A}$. The result for $\min(f, g)$ follows in the same way from the formula

$$\min(f, g) = \frac{f + g}{2} - \frac{|f - g|}{2}.$$

Since

$$\max(a, b, c) = \max(\max(a, b), c) \text{ and } \min(a, b, c) = \min(\min(a, b), c),$$

2.4 and induction show that if $f_i \in \bar{A}$ for $i = 1, \cdots, n$, then

$$\max(f_1, \cdots, f_n) \quad \text{and} \quad \min(f_1, \cdots, f_n)$$

also belong to $\bar{A}$.

3. Stone-Weierstrass Theorem

Let D be any subset of $\hat{C}(Y; c)$; then there is a unique smallest algebra $A(D)$ containing D. Indeed, $\hat{C}(Y; c)$ is an algebra containing D, and it is trivial to verify that the intersection $A(D)$ of all algebras containing D is also an algebra. $A(D)$ is called the algebra generated by D; its description directly in terms of D itself is

3.1 $A(D)$ is the set of all functions of form $p(f_1, \cdots, f_n)$, where the f_i belong to D, and p ranges over all polynomials in $n \geqslant 1$ indeterminates with no constant term. $A(D)$ is unitary if D contains a nonzero constant function.

Proof: If A is any algebra containing D, then by **2**, Ex. 4, A must certainly contain all the functions described above. However, the described set is clearly itself an algebra, so it must be $A(D)$.

The Stone-Weierstrass theorem states that if D satisfies a fairly simple condition, then the generated algebra $A(D)$ is dense in $\hat{C}(Y; c)$.

3.2 Definition A set $D \subset \hat{C}(Y; c)$ is called *separating* if for each pair of points $y \neq y'$ in Y there is an $f \in D$ such that $f(y) \neq f(y')$.

3.3 Theorem (M. Stone) Let Y be an arbitrary space, and $D \subset \hat{C}(Y; c)$ a family that contains a nonzero constant function and is separating. Then $A(D)$ is dense in $\hat{C}(Y; c)$.

We give some examples before entering the proof.

Ex. 1 The Stone-Weierstrass theorem contains the classical Weierstrass theorem. Let $Y = I$, and $D = \{\delta(y) \equiv 1, c(y) = y\}$; this family is separating, so that $A(D)$, which (**3.1**) consists of all polynomials in x, is dense in $\hat{C}(I; c)$. Because I is compact, $\hat{C}(I; c)$ is (XII, **8.2**) homeomorphic to $C(I, E^1; d_e)$ with the metric topology induced by d_e^+, therefore sequences can be used to describe closure (X, **6.2**) and convergence means uniform convergence on I. Thus, for each $f \in \hat{C}(I; c)$ and each $\varepsilon > 0$, there is a polynomial $p_\varepsilon(x)$ such that $|f(x) - p_\varepsilon(x)| < \varepsilon$ for all $x \in I$.

Ex. 2 It contains the n-dimensional form of the Weierstrass theorem: For each $f \in \hat{C}(I^n; c)$ and each $\varepsilon > 0$, there is a polynomial $p_\varepsilon(x_1, \cdots, x_n)$ such that $|f(x_1, \cdots, x_n) - p_\varepsilon(x_1, \cdots, x_n)| < \varepsilon$ for all $(x_1, \cdots, x_n) \in I^n$. We need note only that the functions $f(x_1, \cdots, x_n) \equiv 1, x_1, \cdots, x_n$ satisfy the requirement of **3.3**.

Ex. 3 Let $Y = [0, \infty[$; then $D = \{\delta(x) = 1, f(x) = e^{-x}\}$ separates points; since Y is 2° countable, $\hat{C}(Y; c)$ is metrizable (IX, **9.2**, and XII, **1.3**) so sequences can be used to describe the closure of $A(D)$. Thus: for each continuous real-valued function f on the positive x-axis, there exists a sequence

$$p_k(x) = \sum_{n=0}^{n_k} a_n e^{-nx}$$

such that $p_k \to f$ uniformly on every compact subset (XII, **7.2**).

Proof of Theorem: The proof breaks into three parts:

(1). For any pair $x \neq y$ of points in Y and any constants a, b, there is an $f \in A(D)$ such that $f(x) = a$ and $f(y) = b$.

Indeed, since D is separating, there is a $g \in D$ with $g(x) = \alpha \neq \beta = g(y)$; since D contains a constant function, $A(D)$ contains all constant functions, and so

$$f = a + \frac{(b - a)}{(\beta - \alpha)} [g - \alpha]$$

is the required function in $A(D)$.

To prove the theorem, it evidently suffices to show that $\overline{A(D)}$ is dense in $\hat{C}(Y; c)$; that is, given f and any nbd $\bigcap_1^n (A_i, V_i)$ of f, there is a $g \in \overline{A(D)}$ in this nbd. Letting $\varepsilon = \min\{d_e(f(A_i), Y - V_i) \mid i = 1, \cdots, n\}$, which is positive (XI, **4.4**), this will follow if we can find a $g \in \overline{A(D)}$ such that $|f(r) - g(r)| < \varepsilon$ for all r in the compact $R = \bigcup_1^n A_i$. We produce this g in two stages:

(2). For each fixed $r_0 \in R$ there is a $g \in \overline{A(D)}$ such that:

(a). $g(r_0) = f(r_0)$,

(b). $g(r) < f(r) + \varepsilon$ for all $r \in R$.

In fact, by (1), for each $r \in R$ there is some $h_r \in A(D)$ such that $h_r(r_0) = f(r_0)$ and $h_r(r) < f(r) + \varepsilon/2$. Since h_r is continuous on Y, there is a nbd $V(r)$ such that $y \in V(r) \Rightarrow h_r(y) < f(y) + \varepsilon$; covering the compact R by finitely many sets $V(r_1), \cdots, V(r_n)$ and defining $g = \min[h_{r_1}, \cdots, h_{r_n}]$, we find $g \in \overline{A(D)}$ by **2.4**. Furthermore, since each $r \in R$ belongs to some $V(r_i)$, it follows that $g(r) \leqslant h_{r_i}(r) < f(r) + \varepsilon$, as required.

(3). There is a $g \in \overline{A(D)}$ with $|f(r) - g(r)| < \varepsilon$ for all $r \in R$.

For each r_0, let g_{r_0} be the function obtained in (2), and find a nbd $V(r_0)$ such that $y \in V(r_0) \Rightarrow g_{r_0}(y) > f(y) - \varepsilon$; cover R with finitely many such nbds $V(r_1), \cdots, V(r_n)$, and define

$$g = \max[g_{r_1}, \cdots, g_{r_n}],$$

which belongs to $\overline{A(D)}$. Since each $r \in R$ belongs to some $V(r_i)$, we have $g(r) \geqslant g_{r_i}(r) > f(r) - \varepsilon$; and, by (2), we have $g_{r_i}(r) < f(r) + \varepsilon$, for every $i = 1, \cdots n$, so that $g(r) < f(r) + \varepsilon$ also; consequently, $|f(r) - g(r)| < \varepsilon$ on R, and the proof of **3.3** is complete.

The Stone-Weierstrass theorem need not be true for complex-valued functions. For example, with $Y = \{z \mid |z| < 1\}$, the unit disc in the complex plane, and $Z =$ the complex plane, the family $D \subset C(Y, Z; d_2)$ consisting of $\{\delta(z) = 1, f(z) = z\}$ satisfies the requirement of **3.3**, but any sequence of polynomials converging uniformly on every compact subset of Y necessarily has an *analytic* function for limit. The theorem does, however, extend to complex-valued functions by placing on D the following additional requirement: If f belongs to D, so also does its complex conjugate f^+. For, with this additional requirement, it is easy to see that the family $\{f + f^+, i(f - f^+) \mid f \in D\}$ of real-valued functions is separating; therefore it can be used to approximate separately the real and imaginary parts of any given complex-valued function.

Somewhat surprisingly, Theorem **3.3** (that is, with no additional requirement on D) also holds for quaternion-valued functions (J. C. Holladay). A proof, which will be left for the reader, can be based on the following simple identity: if $q = a + bi + cj + dk$, then $4a = q - iqi - jqj - kqk$.

4. The Metric Space $C(Y)$

Let d_e be the Euclidean metric $d_e(x, y) = |x - y|$ in E^1. In this section, we consider two topologies on the set $C(Y, E^1; d_e)$ of all *bounded* continuous real-valued functions on Y, and investigate their interrelation.

4.1 Definition The set $C(Y, E^1; d_e)$, as a subspace of $\hat{C}(Y; c)$, is denoted by $C(Y; c)$. The same set, with the metric topology induced by the supremum metric d_e^+, is denoted by $C(Y)$.

If Y is compact, then [XII, **8.2**(3)] we know that $C(Y) \cong C(Y; c)$. However, this is not true in general.

Ex. 1 Let $Q \subset E^1$ be the subspace of rationals. Using the c-topology, XII, **1.2**, shows that I^Q can be regarded as a subspace of $(E^1)^Q$, and it is evident that $I^Q \subset C(Q; c) \subset (E^1)^Q$. If the space $C(Q; c)$ were metrizable, so also would be I^Q, and this is impossible because (XII, **5**, Ex. 3) I^Q is not first countable.

For the metric space $C(Y)$ and its relation to $C(Y; c)$, we have

4.2 $C(Y)$ is a locally convex linear topological space. Furthermore, the identity map $1: C(Y) \to C(Y; c)$ is continuous.

Proof: Since $C(Y)$ is a metric space, sequences can be used. Because convergence means uniform convergence on Y, we find from $f_n \to f$, $g_n \to g$, and real numbers $\lambda_n \to \lambda$ that $f_n \pm g_n \to f \pm g$, $f_n \cdot g_n \to f \cdot g$ and $\lambda_n f_n \to \lambda f$, and therefore the algebraic operations are continuous. For local convexity, it is simple to verify that if $f, g \in B(f_0, \varepsilon)$, then also $\lambda f + (1 - \lambda) g \in B(f_0, \varepsilon)$ for all $0 \leqslant \lambda \leqslant 1$. [We remark that these results follow much more simply from the observation that $C(Y)$ is a normed space (*cf.* Appendix I).] To see that $1: C(Y) \to C(Y; c)$ is continuous, note that (XII, **8.3**) the metric topology is conjoining; since the c-topology is splitting, it is smaller than the metric topology (XII, **10.3**).

It is clear that the *set* $C(Y)$ can be described in purely topological terms as $\{f: Y \to E^1 \mid \overline{f(Y)} \text{ is compact}\}$, and because compact sets are bounded, it therefore has the minimal property: $C(Y) \subset C(Y, E^1; d)$ for every metric d equivalent to d_e. This shows that for each d equivalent to d_e, we can use d^+ to impose a topology on $C(Y)$. We now inquire whether all these metrics will yield the same topology on the set $C(Y)$ [or, what is the same, whether the topology of the space $C(Y)$ depends only on the *topology* of E^1 and not on the *metric* d_e]. Although we know [XII, **8.2**(1)] that the metric spaces $C(Y, E^1; d)$ need not be homeomorphic for different equivalent d, the possibility remains that all the metrics d^+ induce the same topology on their common subset $C(Y)$.

4.3 The metric space $C(Y)$ is always a topological invariant of Y. If Y is completely regular, then $C(Y)$ is a *joint* topological invariant of Y and E^1; in particular, the metric topology in $C(Y)$ is then the same as that imposed by any metric in E^1 equivalent to d_e.

Proof: The first assertion is trivial: if $H: Y \cong Y'$, then the formula $d_e^+(f, g) = d_e^+[f \circ H, g \circ H]$, which is valid for all $f, g \in C(Y')$, shows that the bijective map $f \to f \circ H$ is a homeomorphism. Now let Y be completely regular, and let $\rho: Y \to \beta(Y)$ be its Stone-Čech compactification. Since each $f \in C(Y)$ is bounded, there exists (XI, **8.2**) for each f a unique $F: \beta Y \to E^1$ such that $f = F \circ \rho$, and consequently the map $C(\beta Y) \to C(Y)$ defined by $F \to F \circ \rho$ is bijective. Since Y is dense in $\beta(Y)$, we have $d_e^+(F, F') = d_e^+(F \circ \rho, F' \circ \rho)$ so that $F \to F \circ \rho$ is a homeomorphism of the metric spaces $C(\beta(Y))$ and $C(Y)$. Since $\beta(Y)$ is compact, $C(\beta(Y)) \cong C(\beta(Y); c) \cong (E^1)^{\beta(Y)}$; and because (XI, **8.2**) $\beta(Y)$ is a topological invariant of Y, the theorem has been proved.

As Ex. 1 shows, in completely regular spaces Y the spaces $C(Y)$ and $C(Y; c)$ are, in general, distinct joint topological invariants of Y and E^1.

5. Embedding of Y in $C(Y)$

5.1 Definition Let (Y, d) and (Z, d') be metric spaces. A map

$$\varphi: Y \to Z$$

is called an *isometry* if $d'(\varphi(x), \varphi(y)) = d(x, y)$ for all

$$(x, y) \in Y \times Y.$$

It is immediate that an isometry is always injective, is always uniformly continuous, and is a homeomorphism of Y and $\varphi(Y)$; a surjective isometry is called an isomorphism. Note that an isometry is always relative to specified metrics in the two spaces.

Ex. 1 Referring to the proof of **4.3**, we find: If $Y \cong Y'$, then $C(Y)$ is isomorphic to $C(Y')$; if Y is completely regular, then $C(Y)$ is isomorphic to $C(\beta(Y))$.

5.2 Theorem The metric space (Y, d) can always be isometrically embedded in the metric space $(C(Y), d_e^+)$. Furthermore, its image $Y' \subset C(Y)$ is closed in its convex hull.

Proof: Choose a point $p \in Y$, which will remain fixed throughout the discussion. Let φ: $Y \to C(Y)$ be the map $a \to f_a$, where

$$f_a(y) = d(y, a) - d(y, p).$$

We verify $f_a \in C(Y)$ by noting that $|f_a(y)| \leqslant d(a, p)$ and that $d(a, p)$ is a constant independent of y. To prove that φ is an isometry (that is,

$$d_e^+(\varphi(a), \varphi(b)) = d(a, b)),$$

we must show $\sup_y |f_a(y) - f_b(y)| = d(a, b)$. Since

$$|f_a(y) - f_b(y)| \equiv |d(y, a) - d(y, b)| \leqslant d(a, b)$$

for each $y \in Y$, the sup cannot exceed $d(a, b)$. Selecting $y = b$ shows the value $d(a, b)$ is in fact attained, so that $d_e^+(f_a, f_b) = d(a, b)$, as asserted.

Let $H(Y')$ be the convex hull of $\varphi(Y) = Y'$ in $C(Y)$. We are to show that Y' is closed in $H(Y')$. Let $f \in H(Y') - Y'$; then

$$f = \sum_1^n \lambda_i f_{a_i},$$

where $a_i \in Y$, $n \geqslant 2$, $\sum_1^n \lambda_i = 1$, and all $\lambda_i > 0$. Now, for each $a \in Y$, we find

$$f(y) - f_a(y) = \left(\sum_1^n \lambda_i\, d(y, a_i)\right) - d(y, a).$$

Since $n \geqslant 2$, this gives

$$\begin{aligned} f(y) - f_a(y) &\geqslant [\min(\lambda_1, \lambda_2)]\cdot[d(y, a_1) + d(y, a_2)] - d(y, a) \\ &\geqslant [\min(\lambda_1, \lambda_2)]\cdot[d(a_1, a_2)] - d(y, a). \end{aligned}$$

Using $y = a$, we conclude that $d_e^+(f, f_a) \geqslant [\min(\lambda_1, \lambda_2)][d(a_1, a_2)] > 0$ for all $a \in Y$, consequently f is not in the closure of Y'. This proves the theorem.

As an immediate application, we show that if X is attached to Y by f: $A \to Y$, then whenever X, Y are metric spaces, the resulting *set* $X \cup_f Y$ can be given a metric topology (in general, distinct from its identification space topology).

5.3 (C. Kuratowski, F. Hausdorff) Let X, Y be metric spaces, $A \subset X$ closed, and f: $A \to Y$ continuous. Then there exists a metric space $Z \supset Y$ such that:

1. Y is closed in Z.
2. f has a continuous extension F: $X \to Z$.
3. $F | X - A$ is a homeomorphism of $X - A$ and $Z - Y$.

Proof: Let $H(Y)$ be the convex hull of Y in $C(Y)$; since (X, d) is a metric space and $C(Y)$ is a locally convex linear space, then by IX, **6.1**, there is an extension $f^+: X \to H(Y)$ of f. Let $\varphi: X \to C(X)$ be the embedding of **5.2**, and define

$$F: X \to H(Y) \times E^1 \times C(X)$$

by

$$F(x) = [f^+(x), d(x, A), d(x, A) \cdot \varphi(x)].$$

F is evidently continuous (IV, **2.2**). Furthermore, $F \mid X - A$ is injective: if $F(x) = F(x')$, then $d(x, A) = d(x', A)$ and $\varphi(x) \cdot d(x, A) = \varphi(x') \cdot d(x', A)$; since $d(x, A) \neq 0$, we have $\varphi(x) = \varphi(x')$, and since φ is injective, $x = x'$. To prove that $F \mid X - A$ is a homeomorphism, we show its inverse continuous by proving $[F(x_n) \to F(x_0)] \Rightarrow [x_n \to x_0]$. Since $F(x_n) \to F(x_0)$, we find $d(x_n, A) \to d(x_0, A)$ and $\varphi(x_n) \cdot d(x_n, A) \to \varphi(x_0) \cdot d(x_0, A)$; because $d(x_0, A) \neq 0$, we have $d(x_n, A) \neq 0$ for all large n, so that

$$\varphi(x_n) = \frac{\varphi(x_n) \cdot d(x_n, A)}{d(x_n, A)} \to \varphi(x_0)$$

and consequently, $x_n \to x_0$. The space Z is evidently the subspace $Y \cup F(X)$.

6. The Ring $\hat{C}(Y)$

With the operations of pointwise addition and pointwise multiplication, the *set* $\hat{C}(Y)$ of all continuous real-valued functions on Y becomes a commutative ring with unit, the unit being the map $\delta(y) \equiv 1$. It is clear that the ring $\hat{C}(Y)$ is completely determined by the space Y; the objective in this section is to prove the converse result, due to I. Gelfand and A. Kolmogoroff, that whenever Y is compact, the ring $\hat{C}(Y)$ determines the space Y up to a homeomorphism.

We use the following approach: Taking $\hat{C}(Y)$ with the discrete topology, let $\mathscr{H}(Y)$ be the set of all nonzero homomorphisms of $\hat{C}(Y)$ into E^1 and give $\mathscr{H}(Y)$ the function-space topology of pointwise convergence. The space $\mathscr{H}(Y)$ is evidently completely determined by the ring $\hat{C}(Y)$. For each fixed $y \in Y$, observe that the evaluation map $\omega_y: \hat{C}(Y) \to E^1$ belongs to $\mathscr{H}(Y)$, so the correspondence $y \to \omega_y$ defines a map $\mu: Y \to \mathscr{H}(Y)$. To establish the Gelfand–Kolmogoroff result, it is evidently enough to prove that μ is a homeomorphism of Y and $\mathscr{H}(Y)$ whenever Y is compact. In order to carry out this program, we need some elementary facts from algebra; these will first be briefly summarized for convenience.

Each ring R, R' in this discussion is assumed to be commutative and to have a unit. A map $\varphi: R \to R'$ is called a (ring) homomorphism if it satisfies both $\varphi(a + b) = \varphi(a) + \varphi(b)$ and $\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$ for all $a, b \in R$; its kernel, Ker φ, is $\{r \in R \mid \varphi(r) = 0\}$. φ is termed a nonzero homomorphism whenever Ker $\varphi \neq R$, and an isomorphism whenever it is bijective.

A subset $\mathscr{J} \subset R$ with the properties (1) $a, b \in \mathscr{J} \Rightarrow a - b \in \mathscr{J}$, and (2) $a \in \mathscr{J}$, $r \in R \Rightarrow a \cdot r \in \mathscr{J}$ is called an ideal. R is an ideal, and because of (2), it is the only ideal containing the unit $1 \in R$. An ideal $\mathscr{J} \subset R$ is proper if $\mathscr{J} \neq R$; a proper ideal is called maximal if it is not a proper subset of any proper ideal.

If $\varphi: R \to R'$ is any ring homomorphism, it is easy to verify that $\text{Ker}\,\varphi$ is an ideal in R; more important for our purposes is this fact: If $\varphi(R)$ is a field, then $\text{Ker}\,\varphi$ is a maximal ideal. [Usual proof: Let P be any ideal properly containing $\text{Ker}\,\varphi$; we show $1 \in P$. Choose any $p \in P - \text{Ker}\,\varphi$; because $\varphi(R)$ is a field and $\varphi(p) \neq 0$, there is at least one $x \in R$ such that $\varphi(p)\varphi(x) = \varphi(1)$, so that $(px - 1) \in \text{Ker}\,\varphi \subset P$. Since $px \in P$, we conclude that $1 = px - (px - 1) \in P$.]

To determine the nature of an $\omega \in \mathscr{H}(Y)$, we begin with

6.1 Each nonzero homomorphism $\omega: \hat{C}(Y) \to E^1$ is surjective and satisfies $\omega(c\delta) = c$ for each real number c.

Proof: By choosing a $g \in \hat{C}(Y)$ such that $\omega(g) \neq 0$, we find from $\omega(g) = \omega(g \cdot \delta) = \omega(g) \cdot \omega(\delta)$ that $\omega(\delta) = 1$. Using induction, we conclude that $\omega(n\delta) = n$ for each $n \in Z$, and because $n = \omega(n\delta) = \omega(m\delta) \cdot \omega(n\delta/m\delta)$, that $\omega(r\delta) = r$ for each rational r. To complete the proof, it clearly suffices to show that for any two real numbers c, d, $[c \leqslant d] \Rightarrow [\omega(c\delta) \leqslant \omega(d\delta)]$; this follows by noting that $d - c = a^2$ for some real a, so that $\omega(d\delta) - \omega(c\delta) = \omega(a^2\delta) = [\omega(a\delta)]^2$ is consequently nonnegative.

To obtain more information about ω we must study its kernel, which (as follows from **6.1**) is a maximal ideal in $\hat{C}(Y)$.

6.2 Let Y be compact. Then:

(1). If $\mathscr{J}$ is a proper ideal in $\hat{C}(Y)$, there exists a $y_0 \in Y$ such that $f(y_0) = 0$ for every $f \in \mathscr{J}$.

(2). If $\mathscr{J}$ is a maximal ideal, there exists a y_0 such that $\mathscr{J} = \{f \in \hat{C}(Y) \mid f(y_0) = 0\}$.

Proof: Ad (1). We first note that for each $f \in \mathscr{J}$, the closed set $f^{-1}(0) \neq \varnothing$. Indeed, if $f^{-1}(0) = \varnothing$ for some $f \in \mathscr{J}$, then because $1/f \in \hat{C}(Y)$, we would have the unit $\delta = f \cdot 1/f \in \hat{C}(Y)$ so that $\mathscr{J}$ would not be a proper ideal. We next show that the family $\{f^{-1}(0) \mid f \in \mathscr{J}\}$ of closed sets has the finite intersection property: if $f_1, \cdots, f_n \in \mathscr{J}$, then because $f = \sum_1^n f_i^2 \in \mathscr{J}$, we have $\bigcap_1^n f_i^{-1}(0) = f^{-1}(0) \neq \varnothing$. Since Y is compact, this implies that $\bigcap \{f^{-1}(0) \mid f \in \mathscr{J}\} \neq \varnothing$, completing the proof.

Ad (2). Let $\mathscr{J}$ be a maximal ideal. By (1), there is a y_0 such that $\mathscr{J} \subset \mathscr{J}(y_0) = \{f \in \hat{C}(Y) \mid f(y_0) = 0\}$. Since $\mathscr{J}(y_0)$ is clearly a proper ideal, the maximality of $\mathscr{J}$ requires $\mathscr{J}(y_0) \subset \mathscr{J}$, so $\mathscr{J} = \mathscr{J}(y_0)$.

6.3 Let Y be compact. Then the only nonzero homomorphisms $\omega: \hat{C}(Y) \to E^1$ are the evaluation maps ω_y.

Proof: Because E^1 is a field, **6.1** shows Ker ω is a maximal ideal, so that Ker $\omega = \mathscr{J}(y_0)$ for some $y_0 \in Y$. Let $f \in \hat{C}(Y)$ and $f(y_0) = c$; since $f - c\cdot\delta$ vanishes at y_0, we have $f - c\cdot\delta \in$ Ker ω. Thus $\omega(f - c\cdot\delta) = 0$; that is, $\omega(f) = \omega(c\cdot\delta) = c = f(y_0)$ and $\omega = \omega_{y_0}$.

We can now obtain the Gelfand–Kolmogoroff result,

6.4 Theorem Let Y be compact and $\mu: Y \to \mathscr{H}(Y)$ be the map $y \to \omega_y$. Then μ is a homeomorphism of Y and $\mathscr{H}(Y)$.

Proof According to **6.3**, μ is surjective. If $y_0 \neq y_1$, then the complete regularity of Y gives an $f \in \hat{C}(Y)$ such that both $f(y_0) = 0$ and $f(y_1) = 1$, so that $\omega_{y_0} \neq \omega_{y_1}$ and μ is therefore bijective. To prove continuity, let $W = (f, V)$ be a subbasic nbd of $\mu(y_0)$; then

$$\mu^{-1}(W) = \{y \mid \omega_y(f) \in V\} = \{y \mid f(y) \in V\} = f^{-1}(V),$$

which is an open set. Finally, because Y is compact, μ is (XI, **2.1**) a homeomorphism.

We relate the ring structure of $\hat{C}(Y)$ more closely to the topological structure of Y by the following considerations. A continuous map $\varphi: X \to Y$ induces a map $\varphi^+: \hat{C}(Y) \to \hat{C}(X)$ by setting $\varphi^+(g) = g \circ \varphi$ for each $g \in \hat{C}(Y)$; note that $\varphi^+(\delta) = \delta$. Now let φ^{++} be the map of $\mathscr{H}(X)$ induced by φ^+; that is, $\varphi^{++}(\omega) = \omega \circ \varphi^+$. Then each $\varphi^{++}(\omega)$ is a nonzero homomorphism $\hat{C}(Y) \to E^1$, since $\omega \circ \varphi^+(\delta) = \omega(\delta) = 1$, and therefore φ^{++} is a map of $\mathscr{H}(X)$ into $\mathscr{H}(Y)$. Furthermore, because the rings $\hat{C}(Y)$ are taken with the discrete topology, φ^+ is continuous and consequently the induced map $\varphi^{++}: \mathscr{H}(X) \to \mathscr{H}(Y)$ is also continuous (XII, **9.4**). It is trivial to verify that the diagram

$$\begin{array}{ccc} X & \xrightarrow{\varphi} & Y \\ {\scriptstyle \mu_X}\downarrow & & \downarrow{\scriptstyle \mu_Y} \\ \mathscr{H}(X) & \xrightarrow[\varphi^{++}]{} & \mathscr{H}(Y) \end{array}$$

is commutative. This leads to the comprehensive

6.5 Theorem Let X, Y be compact and $h: \hat{C}(Y) \to \hat{C}(X)$ be any ring homomorphism. Then:

(1). If h satisfies $h(\delta) = \delta$, then there exists a unique continuous $\lambda: X \to Y$ such that $\lambda^+ = h$.

(2). If h is an isomorphism, then λ is a homeomorphism.

Proof: Ad (1). As the above considerations show, the induced map h^+ of $\mathscr{H}(X)$ is a continuous map $h^+ : \mathscr{H}(X) \to \mathscr{H}(Y)$; the desired $\lambda : X \to Y$ is given by the formula $\lambda = \mu_Y^{-1} \cdot h^+ \circ \mu_X$, and is clearly continuous; note that $\lambda(x) = y$ if and only if $h(g)[x] = g[y]$ for each $g \in \hat{C}(Y)$. Now let $\lambda^+ : \hat{C}(Y) \to \hat{C}(X)$ be the induced map; then

$$\lambda^+(g)[x] = g \circ \lambda[x] = g[y] = h(g)[x]$$

shows $\lambda^+(g) = h(g)$ for each $g \in \hat{C}(Y)$, as required.

Uniqueness of λ: If λ, φ are such that $\lambda(x) \neq \varphi(x)$ for some $x \in X$, then complete regularity of Y gives an $f \in \hat{C}(Y)$ such that $f[\varphi(x)] \neq f(\lambda(x))$; thus, $\varphi^+ \neq \lambda^+$.

Ad (2). We first note that any homomorphism $h : \hat{C}(Y) \to \hat{C}(X)$ whose image contains δ necessarily satisfies $h(\delta) = \delta$: for, if $h(g) = \delta$, then $h(\delta) = h(\delta) \cdot h(g) = h(\delta \cdot g) = h(g) = \delta$. In particular, every isomorphism satisfies the condition required in (1), and is therefore induced by $\lambda = \mu_Y^{-1} \circ h^+ \circ \mu_X$. To complete the proof, the reader can easily verify that whenever h is an isomorphism, then h^+ is bijective.

Remark: For noncompact spaces, the algebraic structure of $\hat{C}(Y)$ is, in general, not strong enough to determine the topology of Y. For example, using III, **8**, Ex. 7, the reader can easily show that $\hat{C}([0, \Omega])$ and $\hat{C}([0, \Omega[)$ are isomorphic as algebraic structures, though of course $[0, \Omega]$ and $[0, \Omega[$ are not homeomorphic.

Problems

Section 1

1. Let Y be an arbitrary space. Prove: $Z^Y(c)$ is a linear topological space if and only if Z is a linear topological space, and $Z^Y(c)$ is locally convex if and only if Z is locally convex.

Section 3

1. Prove: For each continuous periodic function of period 2π, and each $\varepsilon > 0$, there is a trigonometric function

$$t_\varepsilon(x) = a_0 + \sum_1^N (a_n \cos nx + b_n \sin nx)$$

such that $|f(x) - t_\varepsilon(x)| < \varepsilon$ for all x.

2. Let X, Y be compact spaces and $f \in C(X \times Y; c)$. Prove: For each $\varepsilon > 0$ there are finitely many functions $u_i(x)$, $v_i(y)$, $i = 1, \cdots, n$, such that

$$\left| f(x, y) - \sum_1^n u_i(x) v_i(y) \right| < \varepsilon$$

for all $(x, y) \in X \times Y$.

3. Let Y be compact. Prove that the following three properties are equivalent:

   a. Y is metrizable.
   b. Y has a countable separating family of functions.
   c. $C(Y; c)$ is metrizable.

Section 5

1. Let Y be completely regular and L the set of all linear functionals on $C(Y)$. Taking L with the p-topology, show that Y can be embedded in L.
2. Let X be a metric space and $A \subset X$ closed. Show that the set X/A can be given a metric topology and that, with this topology, there is a continuous surjection $p: X \to X/A$.

Section 6

1. Let φ be a homomorphism $\hat{C}(Y) \to \hat{C}(X)$ whose image contains $C(X)$. Prove: φ sends $C(Y)$ into $C(X)$.
2. Let X be an arbitrary space. A subset $A \subset X$ is called a zero set if there is a continuous $f: X \to E^1$ such that $A = f^{-1}(0)$. Prove: Any countable intersection of zero sets is also a zero set.

Complete Spaces

XIV

Let Y be a metric space or, more generally, a gauge space. In this chapter, we study a certain property, completeness, that some of the gauge structures for Y may have; although this is not a topological concept, it is related to certain topological properties of Y that are of importance in analysis and in topology. We will first discuss complete metrics and then, in Section 9, we consider complete gauge structures.

1. Cauchy Sequences

1.1 Definition Let (Y, d) be a metric space. A sequence $\{y_n\}$ in Y is called a d-Cauchy sequence if

$$\forall\, \varepsilon > 0 \;\exists\, N(\varepsilon) \;\forall\, n, m \geqslant N : d(y_n, y_m) < \varepsilon.$$

Letting T_n be the terminal segment $\{i \in Z^+ \mid i \geqslant n\}$, this definition can be restated: A sequence $\varphi: Z^+ \to Y$ is d-Cauchy if $\forall\, \varepsilon > 0 \;\exists\, n$: d-diameter $\delta[\varphi(T_n)] < \varepsilon$.

The notion of a d-Cauchy sequence is dependent on the particular metric used: The same sequence can be Cauchy for one metric, but not Cauchy for an equivalent metric.

Ex. 1 In E^1, the Euclidean metric $d_e(x, y) = |x - y|$ is equivalent to the metric

$$d_\varphi(x, y) = \left|\frac{x}{1 + |x|} - \frac{y}{1 + |y|}\right|$$

since the latter is derived from the homeomorphism $x \to x/(1 + |x|)$ of E^1 and $]-1, +1[$. The sequence $\{n \mid n = 1, 2, \cdots\}$ in E^1 is not d_e-Cauchy, whereas it obviously is d_φ-Cauchy.

The main properties of Cauchy sequences are summarized in

1.2 Theorem Let (Y, d) be a metric space.

(1). Every convergent sequence in Y is a d-Cauchy sequence.
(2). Every subsequence of a d-Cauchy sequence is also a d-Cauchy sequence.
(3). If a d-Cauchy sequence accumulates at y_0, then it converges to y_0. In particular, a d-Cauchy sequence either converges or has no convergent subsequence.

Proof. (1). Assume $\varphi \to y_0$; then for each $\varepsilon > 0$ there is an $n = n(\varepsilon)$ such that $\varphi(T_n) \subset B(y_0, \varepsilon/2)$ and therefore $\delta[\varphi(T_n)] < \varepsilon$.

(2) is trivial.

(3). Assume that φ is d-Cauchy and that $\varphi \succ y_0$. Given any nbd $B(y_0, \varepsilon)$ of y_0, choose N so large that $\delta[\varphi(T_N)] < \varepsilon/2$; since

$$y_0 \in \bigcap_1^\infty \overline{\varphi(T_n)},$$

it follows that $\varphi(T_N) \subset B(y_0, \varepsilon)$, and we have proved that $\varphi \to y_0$. The second part is an immediate consequence: if a subsequence $\varphi' \to y_0$, then because $\varphi' \vdash \varphi$, it would follow that $\varphi \succ y_0$.

Ex. 2 The converse of **1.2**(1) is false in general: In the space $Y =]0, 1]$ with the Euclidean metric d_e, the sequence $\{1/n\}$ is d_e-Cauchy, yet it does not converge to any $y_0 \in Y$.

2. Complete Metrics and Complete Spaces

2.1 Definition Let Y be a metrizable space. A metric d for Y (that is, one that metrizes the given topology of Y) is called complete if every d-Cauchy sequence in Y converges.

It must be emphasized that completeness is a property of metrics: One metric for Y may be complete, whereas another (equivalent!) metric may not.

Ex. 1 Referring to **1**, Ex. 1, we shall see shortly that the Euclidean metric d_e for E^1 is complete; however, the equivalent metric d_φ for E^1 is not complete, since the sequence $\{n\}$ is d_φ-Cauchy, but does not converge to any point of E^1.

A given metrizable space Y may not have any complete metric; to denote those that do, we have

2.2 Definition A metric space Y is called topologically complete (or briefly, complete) if a complete metric for Y exists. To indicate that d is a complete metric for Y, we say that Y is d-complete.

Ex. 2 We shall see later that the subspace of rationals in I is *not* topologically complete. In contrast, the subspace $\mathscr{J}$ of irrationals in I is topologically complete, though the Euclidean metric d_e is evidently not complete for the space $\mathscr{J}$. To determine a complete metric, express each irrational as an infinite continued fraction $a = \frac{1|}{|a_1} + \frac{1|}{|a_2} + \cdots$ and define $d(a, b) = 1/n$, where n is the first integer for which $a_n \neq b_n$. It is simple to verify that d is a metric (called the Baire metric) for $\mathscr{J}$, that d is equivalent to the Euclidean metric for $\mathscr{J}$, and that d is complete for $\mathscr{J}$.

The question whether a given metric d for a space Y is complete arises in all considerations of limits. For, to determine that a given sequence in Y converges, the definition X, **1.1** of convergence requires that we produce the limit point y_0; that is, we must find a point $y_0 \in Y$ that satisfies the requirements of that definition. On the other hand, if it is known that d is complete, then we can determine that a given sequence converges by simply verifying it to be d-Cauchy, that is, without having to produce the limit point or using any data other than the given sequence. These considerations indicate that for the purposes of analysis, it is generally not so important to know that a space is topologically complete as it is to know what a complete metric for that space is.

We now determine some conditions that assure that a metric, or a space, is complete.

2.3 Theorem Let (Y, d) be a metric space, and assume that d has the property: $\exists\, \varepsilon > 0\ \forall\, y \in Y : \overline{B_d(y, \varepsilon)}$ is compact. Then d is complete.

Proof: Let φ be a d-Cauchy sequence in Y, and choose n so large that $\delta[\varphi(T_n)] < \varepsilon/2$; then $\varphi(T_n) \subset \overline{B_d[\varphi(n), \varepsilon]}$ and therefore [XI, **1.3**(3)] φ has an accumulation point y_0. It now follows from **1.2**(3) that $\varphi \to y_0$ and this completes the proof.

2.4 Corollary Every locally compact metric space Y is topologically complete. Furthermore, if Y is compact, then every metric d for Y is complete.

Proof: If Y is locally compact, it has a covering $\{U\}$ consisting of relatively compact open sets; by IX, **9.4**, there is a metric d for Y such that each ball $B_d(y, 1)$ is contained in some set U, and by **2.3**, d is therefore complete. For compact Y, the entire proposition follows from **2.3**.

Ex. 3 Any metric for a closed bounded set in E^n and any metric for the Hilbert cube I^∞ is complete.

Ex. 4 It follows directly from **2.3** that the Euclidean metric for E^n is complete.

Ex. 5 The converse of **2.4** is false: We have seen in Ex. 2 that the space $\mathscr{J}$ of irrationals in I is topologically complete, but it is not locally compact.

For the invariance properties of topological completeness we find

2.5 Theorem (1). Topological completeness is a topological invariant.

(2). If Y is topologically complete, then any closed subspace A of Y is also topologically complete. Moreover, if d is a complete metric for Y, then $d_A = d \mid A \times A$ is a complete metric for A.

(3). If (Y, d) is any metric space (not necessarily d-complete!) and if $A \subset Y$ is d_A-complete, then A is closed in Y.

(4). A (countable) cartesian product $\prod_1^\infty Y_i$ is topologically complete if and only if each factor is topologically complete.

Proof: (1). Let $\varphi\colon Y \cong Z$, and let Y be d-complete. Define a metric d' in Z by $d'(z_0, z_1) = d(\varphi^{-1}(z_0), \varphi^{-1}(z_1))$; it is straightforward to verify that d' metrizes the space Z and that Z is d'-complete.

(2). Let $\{a_n\}$ be a d-Cauchy sequence in A; since d is complete, $a_n \to y_0 \in Y$, and because A is closed, $y_0 \in A$.

(3). Let $y_0 \in \bar{A}$; by X, **6.2**, there is a sequence $\{a_n\}$ in A such that $a_n \to y_0$. By **1.2**(1), $\{a_n\}$ is d-Cauchy, and since A is d-complete, $\{a_n\}$ must converge to some $a_0 \in A$. Since limits of sequences are unique in Hausdorff spaces, we find $y_0 = a_0 \in A$, so A is closed.

(4). If $\prod_1^\infty Y_i$ is topologically complete, then because each Y_i is homeomorphic to a closed subset of $\prod_1^\infty Y_i$, (2) assures that each Y_i is topologically complete. For the converse, assume that Y_i is d_i-complete for each i. Let $\{k_i\}$ be any sequence of positive constants such that $k_i \to 0$, and for each i define $d_i'(x_i, y_i) = \min[k_i, d_i(x_i, y_i)]$; it is trivial to verify that d_i' is also a complete metric for Y_i. By IX, **7.2**, $\rho(\{x_i\}, \{y_i\}) = \sup_i d_i'(x_i, y_i)$ is a metric for the space $\prod_1^\infty Y_i$; we show that ρ is complete.

Let $\{x^{(n)}\}$ be any ρ-Cauchy sequence in $\prod_1^\infty Y_i$; since

$$d_i'(x_i^{(n)}, y_i^{(n)}) \leqslant \rho(x^{(n)}, y^{(n)}),$$

it follows that for each i, the sequence $\{x_i^{(n)}\}$ is d_i'-Cauchy, consequently it converges to some $x_i^{(0)} \in Y_i$. Since convergence in cartesian products is equivalent to coordinatewise convergence, $x^{(n)} \to x^{(0)}$, concluding the proof.

For function spaces with the sup metric,

2.6 Theorem Let X be an arbitrary space, and let Y be d-complete. Then $C(X, Y; d)$ is d^+-complete.

Proof: Let $\{f_n\}$ be any d^+-Cauchy sequence, so that

$$\forall\, \varepsilon > 0\ \exists\, N(\varepsilon)\ \forall\, n, m \geqslant N(\varepsilon)\ : d^+(f_n, f_m) < \varepsilon.$$

Since $d(f_n(x), f_m(x)) \leqslant d^+(f_n, f_m)$, it follows that $\{f_n(x)\}$ is a d-Cauchy sequence in Y for each x, and therefore converges to some element, which we denote by $F(x)$. Furthermore, we have $f_n(x) \in B(f_m(x), \varepsilon)$ for all x and $n, m \geqslant N(\varepsilon)$, consequently $F(x) \in \overline{B(f_m(x), \varepsilon)}$ for each x and all $m \geqslant N(\varepsilon)$, which shows that the sequence $\{f_n\}$ converges to the function F uniformly on Y. By XII, **8.4**, we find that F is continuous and $F \in C(X, Y; d)$; since $f_n \to F$, this concludes the proof.

3. Cauchy Filterbases; Total Boundedness

In this section, we express the completeness of a metric by the behavior of filterbases.

3.1 Definition A filterbase $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ in a metric space (Y, d) is called a d-Cauchy filterbase if for each $\varepsilon > 0$ there is some A_α having d-diameter $\delta(A_\alpha) < \varepsilon$.

It is evident that a convergent filterbase is necessarily d-Cauchy; in particular, the nbd-filterbase $\mathfrak{U}(y)$ of any $y \in Y$ is a d-Cauchy filterbase. Furthermore, the filterbase $\mathfrak{A}(\varphi)$ determined by a d-Cauchy sequence φ is also clearly a d-Cauchy filterbase.

It is obvious that if each d-Cauchy filterbase in (Y, d) converges, then Y is d-complete. The following theorem goes in the opposite direction:

3.2 Theorem (G. Cantor) Let (Y, d) be a metric space. If d is complete, then every d-Cauchy filterbase in Y converges.

Proof: Let $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ be a d-Cauchy filterbase in Y. For each integer $n \geqslant 1$, find an A_{α_n} such that $\delta(A_{\alpha_n}) < 1/n$, and define a filterbase $\mathfrak{B} = \{B_\iota \mid i \in Z^+\}$ by setting $B_n = \bigcap_1^n A_{\alpha_\iota}$. It is evident that $\mathfrak{A} \vdash \mathfrak{B}$, so we need prove only that $\mathfrak{B}$ is convergent. To this end, for each n choose a $b_n \in B_n$; since $\delta(B_\iota) \to 0$ and $b_n \in B_\iota$ for all $n \geqslant i$, we conclude that $\{b_n\}$ is a d-Cauchy sequence and consequently converges to some $y_0 \in Y$. It follows that $\mathfrak{B} \to y_0$: Given any $U = B(y_0, \varepsilon)$, we need only choose n so large that $b_n \in B(y_0, \varepsilon/2)$, and $1/n < \varepsilon/2$ to assure that $B_n \subset U$.

3.3 Corollary Let (Y, d) be a metric space. The following two properties are equivalent:

(1). d is complete.

(2). If $\{F_\alpha \mid \alpha \in \mathscr{A}\}$ is any family of closed sets in Y such that each finite subfamily has nonempty intersection, and if $\inf \{\delta(F_\alpha) \mid \alpha \in \mathscr{A}\} = 0$, then $\bigcap_\alpha F_\alpha$ is nonempty, and consists of a single point.

Proof: (1) $\Rightarrow$ (2). The family $\mathfrak{A}$, consisting of all the F_α together with their finite intersections, is a filterbase in Y and, because

$$\inf \{\delta(F_\alpha) \mid \alpha \in \mathscr{A}\} = 0,$$

it is a d-Cauchy filterbase. Thus, $\mathfrak{A}$ converges to some y_0 and therefore $y_0 \in \bigcap_\alpha \overline{F}_\alpha = \bigcap_\alpha F_\alpha$. There can be no $y \neq y_0$ in this intersection, since this would require that $\delta(F_\alpha) \geqslant d(y, y_0) > 0$ for all α.

(2) $\Rightarrow$ (1). Let φ be a d-Cauchy sequence in Y. The family

$$\{\overline{\varphi(T_n)} \mid n \in Z^+\}$$

evidently satisfies the requirements in (2), so there is some $y_0 \in \bigcap_n \overline{\varphi(T_n)}$. This means that $\varphi \succ y_0$, and therefore that $\varphi \to y_0$.

The condition $\inf_\alpha \delta(F_\alpha) = 0$ is essential: in the d_e-complete space E^1, this condition is violated by the family $F_n = \{x \mid x \geqslant n\}$, $n \in Z^+$, and the conclusion is false.

Since **3.3** differs from the analogous [XI, **1.3**(2)] characterization of compactness only in the additional requirement that $\inf_\alpha \delta(F_\alpha) = 0$, it suggests that the topological concept of compactness in metric spaces can be expressed entirely by suitable properties of a metric. To obtain such a characterization, we start with the

3.4 Definition A metric d for a metrizable space Y is called totally bounded (or precompact) if, for each $\varepsilon > 0$, the open covering $\{B(y, \varepsilon) \mid y \in Y\}$ of Y has a finite subcovering. To indicate that d is a totally bounded metric for Y, we say that Y is totally d-bounded (or d-precompact).

Ex. 1 A bounded metric need not be totally bounded: in E^1, the bounded metric $d(x, y) = \min[1, d_e(x, y)]$ is not totally bounded.

Ex. 2 Total boundedness is a property of metrics: Referring to **I**, Ex. 1, the Euclidean metric d_e for E^1 is not totally bounded, whereas the (equivalent) metric d_φ is. This also shows that the existence of a totally bounded metric for a space does not imply that the space is compact.

Ex. 3 A metric space Y has a totally bounded metric if and only if it is 2° countable. For, if Y has a totally bounded metric, then Y has a finite $(1/n)$-dense set for each $n \in Z^+$, consequently Y is separable and IX, **5.6** applies. Conversely, if Y is 2° countable, then Y can be embedded in the Hilbert cube I^∞ and, since I^∞ is compact, the induced metric on Y is totally bounded.

3.5 Theorem A metrizable space Y is compact if and only if it has a metric d that is both complete and totally bounded.

Proof: Any metric for a compact space is complete, and by the definition of compactness, also totally bounded. Conversely, let d be a totally bounded and complete metric for Y; we show that Y is compact by proving that each maximal filterbase $\mathfrak{M}$ in Y converges. Since d is complete, it suffices to show that $\mathfrak{M}$ is d-Cauchy. To this end, let $\varepsilon > 0$ be given; by total boundedness of d, there is a finite covering $U_i = B_d(y_i, \varepsilon)$, $i = 1, \cdots, n$ of Y, and we need prove only that at least one U_i contains a member of $\mathfrak{M}$. If none did, then the maximality of $\mathfrak{M}$ would (X, **7.2**) give an $M_{\alpha_i} \subset \mathscr{C}U_i$ for each $i = 1, 2, \cdots, n$ and result in the contradiction that there is a nonempty

$$M_\beta \subset \bigcap_1^n M_{\alpha_i} \subset \bigcap_1^n \mathscr{C}U_i = \varnothing .$$

The following modification of **3.5** is frequently used in analysis.

3.6 Corollary Let Y be d-complete. Then a subset $A \subset Y$ has compact closure if and only if A is totally d-bounded.

Proof: By **2.5**, $\overline{A}$ is d-complete, so in view of **3.5**, we need prove only the general proposition: $\overline{A}$ is totally d-bounded if and only if A is totally d-bounded. Since the total d-boundedness of $\overline{A}$ obviously implies that of A, only the converse requires proof. Let $\varepsilon > 0$ be given; since A is totally d-bounded, there is a finite open covering

$$\{B(a_i, \varepsilon/2) \cap A \mid i = 1, \cdots, n\}$$

of A. In Y, we therefore have

$$\bar{A} \subset \overline{\bigcup_1^n B\left(a_i, \frac{\varepsilon}{2}\right)} = \bigcup_1^n \overline{B\left(a_i, \frac{\varepsilon}{2}\right)} \subset \bigcup_1^n B(a_i, \varepsilon),$$

so that $\{\bar{A} \cap B(a_i, \varepsilon) \mid i = 1, \cdots, n\}$ is a finite covering of $\bar{A}$ by balls of radius ε. Thus $\bar{A}$ is totally d-bounded.

4. Baire's Theorem for Complete Metric Spaces

Theorem **2.4** gives a sufficient condition for topological completeness; the following theorem is a necessary condition. Because completeness is more prevalent than local compactness, this result is one of the most important and useful theorems in topology, and has extensive applications in analysis.

4.1 Theorem (R. Baire) Any topologically complete space is a Baire space.

Proof: The proof is similar to that of XI, **10.1**. Let $D_1, D_2, \cdots$ be open dense sets in the topologically complete space Y; we show

$$U \cap \bigcap_1^\infty D_i \neq \varnothing$$

for each open $U \neq \varnothing$. Choose a complete metric d for Y, and start with U and D_1. Since $U \cap D_1 \neq \varnothing$, there is an open d-ball B_1 such that $\bar{B}_1 \subset U \cap D_1$ and $\delta(\bar{B}_1) \leqslant 1$. We proceed by induction, as before, but this time we determine a sequence $\{B_n\}$ of open balls such that $\bar{B}_n \subset B_{n-1} \cap D_n$ and $\delta(\bar{B}_n) \leqslant 1/n$ for each n. We have

$$\bigcap_1^\infty \bar{B}_n \subset U \cap \bigcap_1^\infty D_n,$$

and since the $\{\bar{B}_n\}$ satisfy the hypotheses of **3.3**, also $\bigcap_1^\infty \bar{B}_n \neq \varnothing$, so the proof is complete.

Ex. 1 More generally, any cartesian product $\prod\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ of topologically complete spaces is a Baire space. For, in the preceding proof, take each B_n to be a basic open set such that each factor not the whole space has diameter $\leqslant 1/n$; observing that each B_n is of the form $C_n \times \prod\{Y_\alpha \mid \alpha \in \mathscr{A} - \mathscr{B}\}$, where $\mathscr{B} \subset \mathscr{A}$ is some fixed set with $\aleph(\mathscr{B}) \leqslant \aleph_0$ and $C_n \subset \prod\{Y_\alpha \mid \alpha \in \mathscr{B}\}$, the conclusion follows easily from **2.5**(4). By using the topologically complete space of irrationals in E^1, we find that nonlocally compact and also nonmetrizable Baire spaces exist.

Ex. 2 The subspace Q of rationals in E^1 is not topologically complete. For, by XI, **10**, Ex. 4, Q is not a Baire space.

Ex. 3 As an application of **4.1** (valid, in fact, in any Baire space, Y), we prove: If $\{f_\alpha \mid \alpha \in \mathscr{A}\}$ is any family of continuous real-valued functions on Y, and if $M(y) = \sup_\alpha \{f_\alpha(y)\}$ is finite for each $y \in Y$, then there is an open set U on which M is *uniformly* bounded. For, M is lower semicontinuous (III, **10.4**) and consequently $A_n = \{y \mid M(y) \leqslant n\}$, $n \in Z^+$, is a closed countable covering of Y; since Y is a Baire space, some A_n must contain an open $U \subset Y$.

The conclusion XI, **10.5**, that a first category set in a topologically complete space has no interior is frequently used in analysis to establish general existence theorems. We give two fairly typical illustrations of its use.

4.2 There exist continuous real-valued functions on I having no derivative at any point. In fact, the functions in $C(I)$ that have infinite right derivatives at each point of I form a set of the second category in $C(I)$.

Proof: For each $n = 1, 2, \cdots$ define $N_n \subset C(I)$ by

$$N_n = \left\{f \in C(I) \mid \exists\, x \in \left[0, 1 - \frac{1}{n}\right] \forall\, h \in \left]0, \frac{1}{n}\right]: \left|\frac{f(x+h) - f(x)}{h}\right| \leqslant n\right\}.$$

The only functions qualified to have a derivative at even one point are those belonging to some N_n and therefore we wish to prove that

$$C(I) - \bigcup_1^\infty N_n \neq \varnothing.$$

It is at this point that Baire's theorem enters. For, noting that $C(I)$ is complete (*cf.* **2.6**), we need show only that each N_n is nowhere dense in $C(I)$ to be assured that $\bigcup_1^\infty N_n$ does not even fill out an open set and, indeed, that the set $C(I) - \bigcup_1^\infty N_n$ of nowhere differentiable functions is of the second category.

We now establish that each N_n is nowhere dense. In fact:

(1). Each N_n is closed.

The evaluation map $\omega: C(I) \times I \to E^1$ being continuous, it follows that for each fixed $h_0 \in]0, 1/n]$ the maps $C(I) \times [0, 1 - 1/n] \to E^1$ given by

$$(f, x) \to (1/h_0)\cdot f(x + h_0)$$

and $(f, x) \to (1/h_0)f(x)$ are also continuous; thus, so also is

$$(f, x) \to \left|\frac{f(x + h_0) - f(x)}{h_0}\right|$$

and therefore

$$\left\{(f, x) \mid \left|\frac{f(x + h_0) - f(x)}{h_0}\right| \leqslant n\right\}$$

is closed in $C(I) \times [0, 1 - 1/n]$. The projection of this closed set parallel to the compact axis $[0, 1 - 1/n]$ is therefore closed in $C(I)$ and is

$$N(h_0) = \left\{f \mid \exists\, x \in \left[0, 1 - \frac{1}{n}\right]: \left|\frac{f(x + h_0) - f(x)}{h_0}\right| \leqslant n\right\}.$$

Since $N_n = \bigcap \{N(h_0) \mid h_0 \in]0, 1/n]\}$, this establishes (1).

(2). Each N_n has no interior.

Let $f \in N_n$; we show that in each ball $B(f; \varepsilon)$, there is a $g \in C(I) - N_n$; that is, a g such that:

(a). $d_e^+(f, g) < \varepsilon$.

(b). $\forall x \in \left[0, 1 - \frac{1}{n}\right] \exists h_x \in \left]0, \frac{1}{n}\right]: \left|\frac{g(x + h_x) - g(x)}{h_x}\right| > n.$

By the Weierstrass theorem (XIII, **3**, Ex. 1), there is a polynomial p with $d_e^+(f, p) < \varepsilon/3$, so we require a function within $\varepsilon/3$ of p and having property (b). Since the derivative p' is continuous on I, let

$$M = \max \{|p'(x)| \mid 0 \leqslant x \leqslant 1\}$$

and define $s(x)$ to be a continuous nonnegative function on I having a graph consisting of straight line segments, the absolute value of the slope of each segment being $M + n + 1$, and the graph never rising more than $\varepsilon/3$ units above the x-axis. Then $d_e^+(p + s, p) < \varepsilon/3$, and $p + s$ is the function we seek, since

$$\left|\frac{p(x + h) + s(x + h) - p(x) - s(x)}{h}\right| \geqslant \left|\left|\frac{s(x + h) - s(x)}{h}\right| - \left|\frac{p(x + h) - p(x)}{h}\right|\right|,$$

and for each $x \in [0,1 - 1/n]$, we can evidently find an $h \in]0, 1/n]$ such that the right side is $\geqslant M + n + 1 - M = n + 1$.

To give another application of this technique, let C^∞ be the set of all functions in $C(I)$ that have continuous derivatives of all orders. Metrizing C^∞ by $d(f, g) = \sum_0^\infty \min [2^{-n}, d_e^+(f^{(n)}, g^{(n)})]$, it follows from **2.6** and standard theorems of analysis that C^∞ is d-complete. An $f \in C^\infty$ is called analytic at $a \in I$ if its Taylor series $\sum_0^\infty (f^{(n)}(a)/n!)(x - a)^n$ at a converges to $f(x)$ for each x in some nbd of a.

4.3 (**D. Morgenstern**) The set of functions in C^∞ that are nowhere analytic is of the second category.

Proof: In order that $f \in C^\infty$ be analytic at $a \in I$, we must have

$$\sup \{\sqrt[k]{|f^{(k)}(a)|/k!} \mid k \in Z^+\} = c < \infty.$$

Letting

$$T(a; c) = \{f \in C^\infty \mid \forall k \in Z^+: |f^{(k)}(a)| \leqslant k!\, c^k\},$$

it follows that if f is analytic at a, then f belongs to $T(a; c)$ for some $c \geqslant 0$. Since analyticity at a point implies analyticity in some nbd of that point, we find that the functions analytic anywhere in I are contained in

$$\bigcup \{T(a; c) \mid a \text{ rational};\ c \in Z^+\}.$$

We will now prove this union to be a set of the first category, by showing that each $T(a; c)$ is nowhere dense.

Each $T(a; c)$ is closed: Indeed, $\{f \in C^{\infty} \mid |f^{(k)}(a)| \leqslant k!\, c^k\}$ is clearly closed for each k, and therefore so also is their intersection.

Each $T(a; c)$ has empty interior. Given $f \in T(a; c)$ and any nbd $B(f, 2\varepsilon)$ of f, choose n so large that $\sum_{n}^{\infty} 2^{-i} < \varepsilon$, and then select a $b > 2$ so that $\varepsilon b^n > (2n)!\, c^{2n}$. Let

$$s(x) = f(x) + \varepsilon b^{-n} \cos b(x - a);$$

then $s \in C^{\infty}$ and $d_e^+(f^{(k)}, s^{(k)}) \leqslant \varepsilon b^{k-n} < \varepsilon \cdot 2^{k-n}$ for each $k < n$, consequently $s \in B(f, 2\varepsilon)$. However, $|s^{(2n)}(a) - f^{(2n)}(a)| = \varepsilon b^n > (2n)!\, c^{2n}$, so that $s \notin T(a; c)$. This proof is due to H. Salzmann and K. Zeller.

5. Extension of Uniformly Continuous Maps

Let X, Y be metric spaces and $f: X \to Y$ continuous. Although the image $f(\mathfrak{A})$ of a convergent filterbase $\mathfrak{A}$ is convergent (X, **5.2**) the image of a Cauchy filterbase in X need not be Cauchy in Y, as **1**, Ex. 1, shows. However,

5.1 If $f: (X, d') \to (Y, d)$ is *uniformly* continuous, then the image of a d'-Cauchy filterbase is a d-Cauchy filterbase.

Proof: Let $\mathfrak{A}$ be d'-Cauchy. Given any $\varepsilon > 0$, let $\delta(\varepsilon) > 0$ be such that $\forall\, x \in X: f(B_{d'}(x, \delta)) \subset B_d(f(x), \varepsilon)$; finding an $A_\alpha \in \mathfrak{A}$ with $\delta(A_\alpha) < \delta$, we have $\delta[f(A_\alpha)] < \varepsilon$, as required.

Ex. 1 The uniformly continuous image of a complete space need not, however, be complete. The bijection $f: (E^1, d_e) \to (]-1, +1[, d_e)$ given by $x \to x/(1 + |x|)$ is uniformly continuous, and even a homeomorphism, yet the image is evidently not complete.

Because of **5.1**, *uniformly* continuous maps into complete spaces can be expected to have special significance. The following fundamental extension theorem has many applications; a standard one is to define the function a^x, x real, $(a > 0)$ from a knowledge of a^r, r rational.

5.2 Theorem Let (X, d') be an arbitrary metric space, and $A \subset X$ a dense subset. Let Y be d-complete and $f: A \to Y$ uniformly continuous. Then there exists one, and only one, continuous extension $F: X \to Y$ of f, and F is also uniformly continuous.

Proof: For each $x \in X$, the nbd filterbase $\mathfrak{U}(x) \cap A$ is certainly d'-Cauchy, so by **5.1**, the filterbase $f(\mathfrak{U}(x) \cap A)$ is d-Cauchy and consequently convergent. By X, **5.3**, f can be extended by continuity to a unique continuous $F: X \to Y$. We now prove that F is uniformly continuous. Given $\varepsilon > 0$, let $\delta > 0$ be such that $d(f(a_1), f(a_2)) < \varepsilon$ whenever $d'(a_1, a_2) < \delta$; we will show that $d(F(x_1), F(x_2)) < 3\varepsilon$

whenever $d'(x_1, x_2) < \delta/3$. In fact, since F is continuous, find nbds $U(x_1)$, $U(x_2)$ such that $F(U(x_i)) \subset B(F(x_i), \varepsilon)$, $i = 1, 2$, and let

$$W(x_i) = U(x_i) \cap B(x_i, \delta/3).$$

Now $A \cap W(x_i) \neq \varnothing$, since A is dense, and we choose $a_i \in A \cap W(x_i)$. Then, because

$$d'(a_1, a_2) \leqslant d'(a_1, x_1) + d'(x_1, x_2) + d'(x_2, a_2) < 3\delta/3,$$

we find that

$$d(F(x_1), F(x_2)) \leqslant d(F(x_1), f(a_1)) + d(f(a_1), f(a_2)) + d(f(a_2), F(x_2)) < 3\varepsilon$$

and this completes the proof.

Ex. 2 The hypothesis that f be uniformly continuous is essential. Let $A = E^1 - \{0\} \subset E^1$ and let $Y = E^1$. The continuous map $f(x) = x/|x|$ of $A \to E^1$ cannot be extended over E^1.

Ex. 3 The hypothesis that Y be d-complete is essential. Let $X = E^1$, $A \subset X$ be the subspace of rationals and $Y = A$. The identity map $f\colon A \to A$ is uniformly continuous, but cannot be extended to a continuous $F\colon E^1 \to A$, since E^1 is connected and F cannot be constant.

If the map f in **5.2** is a homeomorphism, it is not in general true that its extension F is also a homeomorphism. For example, let $X = I$, let $A = \text{Int}(I)$, and let Y be the unit circle S^1. Using the Euclidean metrics, the map $x \to e^{2\pi i x}$ of $\text{Int}(I)$ onto $S^1 - \{(1, 0)\}$ is a uniformly continuous homeomorphism, yet its extension over I is not bijective. However, there is one type of homeomorphism for which such behavior cannot occur.

A homeomorphism $h\colon (X, d) \cong (Y, d')$ is called a uniform isomorphism of X and Y whenever both h and h^{-1} are uniformly continuous. A surjective isometry is a particularly important example of a uniform isomorphism, but the notion is more general: the map

$$1\colon (E^1, d_e) \to (E^1, d'),$$

where $d' = \min(1, d_e)$, is a uniform isomorphism, although it is not an isometry.

5.3 Corollary Let Y be d-complete, let Y' be d'-complete, and let $A \subset Y$, $A' \subset Y'$ be dense. Then each uniform isomorphism $h\colon A \cong A'$ has an extension $H\colon Y \cong Y'$ that is also a uniform isomorphism. Furthermore, if h is an isometry, then so also is H.

Proof: Since h is uniformly continuous, it is extendable to a uniformly continuous $H\colon Y \to Y'$. Since $g = h^{-1}$ is also uniformly continuous, it also has a uniformly continuous extension $G\colon Y' \to Y$. Because $G \circ H \mid A = 1_A$ and A is dense in Y, we have (VII, **1.5**) that $G \circ H = 1_Y$ and, similarly, that $H \circ G = 1_{Y'}$. It now follows from

III, **12.3**, that H is a uniform isomorphism. The second part is immediate from the manner in which the extension H is defined.

6. Completion of a Metric Space

Let Y be a metrizable space and d a given metric for Y. If Y is not d-complete, then some d-Cauchy sequences in Y do not converge. By emulating the Cantor process for getting the reals from the rationals, we successively adjoin ideal elements to Y, which act as limits for the nonconvergent d-Cauchy sequences, to get ultimately an enlarged space $\hat{Y}$ in which Y is dense, and an extension of d to a complete metric $\hat{d}$ for $\hat{Y}$. Observe that this process depends on the metric d with which we start: Different metrics for the same space Y generally result in distinct enlargements.

The Cantor process can be described in a simpler manner, as the proof of the following theorem shows.

6.1 Theorem Let Y be a metrizable space and d a given metric for Y. Then Y can be isometrically embedded as a dense subset of a complete space $(\hat{Y}, \hat{d})$. $\hat{Y}$ and $\hat{d}$ are unique up to an isometry: if (Y, d) is isometrically embedded as a dense subset of any d_0-complete Y_0, then $(\hat{Y}, \hat{d})$ and (Y_0, d_0) are isometric.

Proof: The uniqueness (up to isometry) of $\hat{Y}$ results from **5.3.**
Existence: Given (Y, d), there is by XIII, **5.2**, an isometric embedding i of Y into $(C(Y), d_e^+)$. Define $\hat{Y} = \overline{i(Y)}$; then Y is dense in $\overline{i(Y)}$, and since $(C(Y), d_e^+)$ is complete (**2.6**), the closed subspace $\hat{Y}$ is also.

The metric space $(\hat{Y}, \hat{d})$ is called the completion of the metric space (Y, d). Relating the completion process with uniformly continuous maps, we have

6.2 Corollary Let (Y, d), (Y_0, d_0) be metric spaces, and $f\colon Y \to Y_0$ a uniformly continuous map. Then there exists a unique uniformly continuous $\hat{f}\colon \hat{Y} \to \hat{Y}_0$ such that the diagram

$$\begin{array}{ccc} Y & \xrightarrow{f} & Y_0 \\ {\scriptstyle i}\downarrow & & \downarrow{\scriptstyle i_0} \\ \hat{Y} & \xrightarrow[\hat{f}]{} & \hat{Y}_0 \end{array}$$

is commutative.

Proof: The map $i_0 \circ f \circ i^{-1}: i(Y) \to \hat{Y}_0$ is evidently uniformly continuous; since $\hat{Y}_0$ is complete and $i(Y)$ is dense in $\hat{Y}$, **5.2** gives a uniformly continuous extension $\hat{f}: \hat{Y} \to \hat{Y}_0$, and $\hat{f}$ satisfies the requirements.

7. Fixed-Point Theorem for Complete Spaces

7.1 Definition Let $T: Y \to Y$ be a map of a space Y into itself. A point $y_0 \in Y$ is called a fixed point for T if $T(y_0) = y_0$.

Not every map has a fixed point; for example the map $T: E^1 \to E^1$ given by $x \to x + 1$ has no fixed point. Theorems asserting the existence of a fixed point for certain types of maps usually have important applications, as we shall see.

A map $T: (Y, d) \to (Y, d)$ of a metric space into itself is called d-contractive if there exists an $\alpha < 1$ such that $d(Tx, Ty) \leqslant \alpha d(x, y)$ for all $(x, y) \in Y \times Y$. The following theorem is a topological version of the Picard "successive approximation" process in analysis:

7.2 Theorem (S. Banach) Let Y be d-complete, and let $T: Y \to Y$ be d-contractive. Then T is continuous and has exactly one fixed point.

Proof: The continuity of T is obvious. Furthermore, T cannot have more than one fixed point: $T(x_0) = x_0$, $T(y_0) = y_0$, and $d(x_0, y_0) > 0$ yield the contradiction $d(x_0, y_0) = d(Tx_0, Ty_0) \leqslant \alpha\, d(x_0, y_0) < d(x_0, y_0)$.

To prove that T does have a fixed point, we first choose any $y \in Y$ and show that the sequence $y, Ty, T(Ty) = T^2y, \cdots$ of iterates is a d-Cauchy sequence. In fact, note that

$$\begin{aligned} d(Ty, T^2y) &\leqslant \alpha\, d(y, Ty), \\ d(T^2y, T^3y) &\leqslant \alpha\, d(Ty, T^2y) \\ &\leqslant \alpha^2\, d(y, Ty), \end{aligned}$$

so that, by induction, $d(T^n y, T^{n+1}y) \leqslant \alpha^n\, d(y, Ty)$. It follows that for any n and any $m > n$ we have

$$\begin{aligned} d(T^n y, T^m y) &\leqslant d(T^n y, T^{n+1}y) + \cdots + d(T^{m-1}y, T^m y) \\ &\leqslant (\alpha^n + \cdots + \alpha^{m-1})\, d(y, Ty) \end{aligned}$$

Now, because $\alpha < 1$, this formula shows that

$$d(T^n y, T^m y) \leqslant \frac{\alpha^n}{1 - \alpha}\, d(y, Ty)$$

for all $m > n$; therefore, since $\alpha^n \to 0$ as $n \to \infty$, we conclude that the sequence $\{T^n y\}$ is in fact d-Cauchy.

Since Y is d-complete, the sequence $\{T^n y\}$ converges to some $y_0 \in Y$. The point y_0 is the fixed point of T. For, on the one hand $T^n y \to y_0$ and the continuity of T imply that $T(T^n y) \to Ty_0$; on the other hand, the sequence $\{T(T^n y)\} = \{T^{n+1}y\}$ is a subsequence of the d-Cauchy sequence $\{T^n y\}$ and therefore must converge to y_0. Thus, $y_0 = Ty_0$, and the theorem is proved.

Ex. 1 In analysis, the elements $T^n y$ are called the successive approximations to y_0; note that $T^n y \to y_0$, regardless of which $y \in Y$ is used, and that the "error" $d(y_0, T^n y)$ of the nth approximation is

$$\leqslant \frac{\alpha^n}{1 - \alpha} d(y, Ty).$$

Ex. 2 While the condition $d(Tx, Ty) < d(x, y)$ is sufficient to assure that T has no more than one fixed point, it is too weak to guarantee the existence of one. The map $T: E^1 \to E^1$ defined by $T(x) = \ln(1 + e^x)$ satisfies this weaker condition, since $|T'(x)| < 1$ for all x, but it has no fixed point.

Ex. 3 The essential feature of **7.2** and of various generalizations is that some sequence $\{T^n y\}$ of iterates either converges or has a subsequence that converges. In **7.2**, this has been forced by completeness of Y plus a uniform Lipschitz condition on T. It is not difficult to verify that completeness of Y plus the weaker condition $d(Tx, Ty) \leqslant \alpha(d(x, y))\cdot d(x, y)$, where $\alpha(\xi)$ is a function on the positive real line such that $0 \leqslant \alpha(\xi) < 1$, and is either monotone nonincreasing or monotone nondecreasing also suffices to have each sequence of iterates convergent.

As a typical application illustrating the use of **7.2**, we prove a version of the classical implicit function theorem:

Let $F(x, y)$ be a continuous real-valued function defined on the rectangle $I_a \times I_b \subset E^2$, where $I_a = \{x \mid |x - x_0| \leqslant a\}$ and $I_b = \{y \mid |y - y_0| \leqslant b\}$. Assume that $F(x_0, y_0) = 0$ and that there is a $k < 1$ such that $|F(x, y) - F(x, y')| \leqslant k|y - y'|$ for all $x \in I_a$, $y, y' \in I_b$. Then there exists a positive $s \leqslant a$ and a unique continuous $h: I_s \to I_b$ such that $h(x_0) = y_0$ and $h(x) \equiv y_0 + F(x, h(x))$ on I_s.

Proof: For any fixed positive $\gamma \leqslant a$, consider the space $C(I_\gamma, I_b; d_e)$, which is (**2.6**) d_e^+-complete, and let C_γ be the subspace $\{\varphi \mid \varphi(x_0) = y_0\}$; C_γ is closed, since it is the inverse image of y_0 under the evaluation map ω_{x_0}, and so each C_γ is d_e^+-complete. For $\varphi \in C_\gamma$, define $T(\varphi)$ to be the function $T(\varphi)(x) = y_0 + F(x, \varphi(x))$ on I_γ; then always $T(\varphi)(x_0) = y_0 + F(x_0, \varphi(x_0)) = y_0$, and the problem reduces to showing that in a suitable C_γ, there is an h such that $T(h) = h$. To apply **7.2**, we must first determine a C_γ that is mapped by T into itself; that is, for each φ satisfying $|y_0 - \varphi(x)| \leqslant b$ on I_γ, $T(\varphi)$ satisfies the same condition. Now,

$$\begin{aligned}
|y_0 - T(\varphi)(x)| &= |F(x, \varphi(x))| \\
&\leqslant |F(x, \varphi(x)) - F(x, y_0)| + |F(x, y_0)| \\
&\leqslant k|\varphi(x) - y_0| + |F(x, y_0)| \\
&\leqslant kb + |F(x, y_0)|
\end{aligned}$$

Since $F(x_0, y_0) = 0$ and F is continuous, by choosing $\gamma = s$ so small that $|F(x, y_0)| \leqslant b(1 - k)$ for all $x \in I_s$, we shall indeed have that T maps C_s into itself. Next, for $\varphi, \psi \in C_s$, we have

$$|T(\varphi)(x) - T(\psi)(x)| = |F(x, \varphi(x)) - F(x, \psi(x))|$$
$$\leqslant k|\varphi(x) - \psi(x)|,$$

so that $d_e^+(T\varphi, T\psi) \leqslant kd_e^+(\varphi, \psi)$; since $k < 1$, $T: C_s \to C_s$ is contractive and has a unique fixed point h.

8. Complete Subspaces of Complete Spaces

Let Y be d-complete. It follows from **2.5** that the only *d-complete* subspaces of Y are the closed subspaces. However, if we ask for the subspaces of Y that are *topologically* complete, the answer is quite different; for example, the set of irrationals in (I, d_e) is not a closed subspace, yet (**2**, Ex. 2) it is a topologically complete space. We are going to show that the topologically complete subspaces of a complete space are precisely the G_δ-sets; this is one reason for the importance of G_δ-sets in analysis.

We first modify **5.2** by removing the requirement of uniformity in the continuity, to obtain the weaker

8.1 Let X be an arbitrary metric space, and $A \subset X$ an arbitrary subset. Let Y be complete and $f: A \to Y$ a continuous map. Then f has a continuous extension over a G_δ-set $G \supset A$.

Proof: Let $\mathfrak{U}(x)$ be the nbd filterbase of x. Because A is dense in $\bar{A}$, we know (X, **5.3**) that f has an extension by continuity over the set

$$G = \{x \in \bar{A} \mid f(A \cap \mathfrak{U}(x)) \text{ converges}\}.$$

We will show that G is a G_δ-set in X.

Let d be a complete metric for Y, and for each $n \in Z^+$ let

$$A_n = \{x \in \bar{A} \mid \exists\ U(x) \in \mathfrak{U}(x): \delta[f(A \cap U(x))] < 1/n\}.$$

Since a filterbase on Y converges if and only if it is d-Cauchy, we have $G = \bigcap_n A_n$. Now, each A_n is open in $\bar{A}$, because if $x \in A_n$, then each $y \in \bar{A} \cap U(x)$ also belongs to A_n. Thus, $A_n = \bar{A} \cap U_n$ for some set U_n open in X, and therefore $G = \bar{A} \cap \bigcap_n U_n$. By recalling that every closed set in a metric space is a G_δ, it follows that G is a G_δ, and the theorem is proved.

8.2 Lemma Let Y be complete and let $A \subset Y$ be a topologically complete subset. Then A is a G_δ-set in Y.

Proof: Let d_A be a complete metric for A, and d a complete metric for Y. The identity map $i: (A, d) \to (A, d_A)$ is continuous, so by **8.1**, the map i extends to a map $i^+: G \to (A, d_A)$, where $G \supset A$ is the G_δ-set $\{y \in \bar{A} \mid A \cap \mathfrak{U}(y) \text{ is } d_A\text{-Cauchy}\}$. We need show only that $G \subset A$, to prove the lemma. Let $y \in G$, and note that the filterbase $A \cap \mathfrak{U}(y)$ always converges to y; therefore the filterbase $A \cap \mathfrak{U}(y)$ is d_A-Cauchy and, if d_A is complete, we must have $y \in A$.

8.3 Theorem (S. Mazurkiewicz) Let Y be complete. Then $A \subset Y$ is topologically complete if and only if it is a G_δ-set in Y.

Proof: "Only if" is **8.2**. For "if", let $A = \bigcap_1^\infty U_i$, where the U_i are open in Y, and $U_1 \supset U_2 \supset \cdots$. We begin with the observation that A can be embedded as a closed subspace in the cartesian product $\prod_1^\infty U_n$. Indeed, for each $n \in Z^+$ let $Y_n = Y$ and let $\mu_n: Y \to Y_n$ be the identity map. Define $\mu: Y \to \prod_1^\infty Y_n$ by $y \to \{\mu_n(y)\}$; it is clear that $\mu: Y \cong \mu(Y)$ and, as in VII, **1.2** (4), we find that $\mu(Y)$ is closed in $\prod Y_n$. Now let λ be the mapping of A into the subspace $\prod U_n \subset \prod Y_n$ which agrees with $\mu \mid A$; then λ is an embedding of A and, since $\lambda(A) = \mu(Y) \cap \prod U_n$, the image of A is closed in $\prod U_n$. Our observation established, **2.5** (2) and (4) show that we need prove only that an open set U in a topologically complete space Y is topologically complete.

We may obviously assume that $U \neq Y$, and shall again construct a suitable embedding. Let d be a metric for Y, and define $f: U \to E^1$ by $f(u) = 1/d(u, Y - U)$; since f is continuous, the graph $G \subset U \times E^1$ of f is clearly homeomorphic to U. Let $j: U \times E^1 \to Y \times E^1$ be the inclusion map; then $j(G) \cong G$. Moreoever, $j(G)$ is closed in $Y \times E^1$: for if $h: Y \times E^1 \to E^1$ is the continuous map $(y, t) \to t \cdot d(y, Y - U)$, then $j(G)$ is the closed set $h^{-1}(1)$. Thus, U is homeomorphic to a closed subset of the complete space $Y \times E^1$ so, by 2.5 (2), U is topologically complete. The theorem is proved.

9. Complete Gauge Structures

In this section, we will extend the notion of completeness to arbitrary gauge structures; since gauge spaces need not be 1° countable, this is done by working with filterbases rather than with sequences.

Let d be a gauge in a space Y. The d-diameter of a set $A \subset Y$ is defined, as for metrics, to be $\sup\{d(x, y) \mid x, y \in A\}$. We next define d-Cauchy filterbase in Y exactly as in **3.1**, and finally make the

9.1 Definition A filterbase $\mathfrak{A} = \{A_\alpha \mid \alpha \in \mathscr{A}\}$ in a gauge space $(Y, \mathscr{T}(\mathscr{D}))$ is called a $\mathscr{D}$-Cauchy filterbase if it is d-Cauchy for each $d \in \mathscr{D}$.

The $\mathscr{D}$-Cauchy filterbases have properties analogous to those summarized in **1.2**:

9.2 Theorem Let $(Y, \mathscr{T}(\mathscr{D}))$ be a gauge space. Then

(1). Every convergent filterbase is $\mathscr{D}$-Cauchy.
(2). If $\mathfrak{A}$ is $\mathscr{D}$-Cauchy and if $\mathfrak{B} \vdash \mathfrak{A}$, then $\mathfrak{B}$ is $\mathscr{D}$-Cauchy.
(3). If $\mathfrak{A}$ is $\mathscr{D}$-Cauchy, and if $\mathfrak{A} \succ y_0$, then $\mathfrak{A} \to y_0$.

The proofs are formally the same as those given in **1.2**; in particular, it follows from (3) that a $\mathscr{D}$-Cauchy filterbase either converges or has no convergent subordinated filterbase.

9.3 Definition A gauge structure $\mathscr{D}$ for a space Y is called complete if every $\mathscr{D}$-Cauchy filterbase in Y converges. A completely regular space having a complete gauge structure $\mathscr{D}$ is called $\mathscr{D}$-complete.

From **9.2**(3) and XI, **1.3**(3), we find that every gauge structure for a compact space Y is complete; moreover, because of **3.2**, every d-complete metric space is a d-complete gauge space. The invariance properties for $\mathscr{D}$-complete gauge structures differ from those of their metric analog (**2.5**) in the behavior of cartesian products:

9.4 Theorem (1). If Y is $\mathscr{D}$-complete and $A \subset Y$ is closed, then A is $\mathscr{D}_A$-complete.

(2). If $(Y, \mathscr{T}(\mathscr{D}))$ is any gauge space, and if $A \subset Y$ is $\mathscr{D}_A$-complete, then A is closed in Y.

(3). Let $\{(Y_\beta, \mathscr{T}(\mathscr{D}_\beta)) \mid \beta \in \mathscr{B}\}$ be any family of gauge spaces, and let $\hat{\mathscr{D}}$ be the induced gauge structure of $\prod Y_\beta$. Then $\hat{\mathscr{D}}$ is complete if and only if each $\mathscr{D}_\beta$ is complete.

The proofs are entirely similar to those given in **2.5**, with that for (3) being immediate from the observation that the projection of a $\hat{\mathscr{D}}$-Cauchy filterbase on each factor Y_β is $\mathscr{D}_\beta$-Cauchy.

We next consider total boundedness. Again, the definition of a totally bounded gauge is formally the same as that for a totally bounded metric, and we make the

9.5 Definition A gauge structure $\mathscr{D}$ for a space Y is called totally bounded if each $d \in \mathscr{D}$ is totally bounded.

There is an important difference between the notion of total boundedness using gauges and that of its purely metric analog: because a completely regular space can be embedded in a parallelotope, it follows easily from **9.4**(3) that every completely regular space has a totally bounded gauge structure. In particular, any nonseparable metric space has a totally bounded gauge structure, although (3, Ex. 3) it does not have any totally bounded metric. However, in exactly the same way as in **3.5**, it follows that

9.6 A completely regular space is compact if and only if it has a gauge structure that is both complete and totally bounded.

We now take up the completion of a given gauge structure. Since we have a meaning (IX, **11**, Ex. 3) for "uniformly continuous map of one gauge space into another," we can introduce the notion of uniform isomorphism as in Section **5**; from this viewpoint, the notion has the following significance: the identity map $1: (Y, \mathscr{T}(\mathscr{D})) \to (Y, \mathscr{T}(\mathscr{D}'))$ is a uniform isomorphism if and only if the uniform structures determined by $\mathscr{D}$ and $\mathscr{D}'$ are equivalent.

The proofs of the evident analogs of **5.2** and **5.3** are practically the same as before; in particular, we obtain the following analog of **5.3**:

9.7 Let Y be $\mathscr{D}$-complete, let Y' be $\mathscr{D}'$-complete, and let $A \subset Y$, $A' \subset Y'$ be dense. Then any uniform isomorphism $h: A \cong A'$ can be extended to a uniform isomorphism $H: Y \cong Y'$.

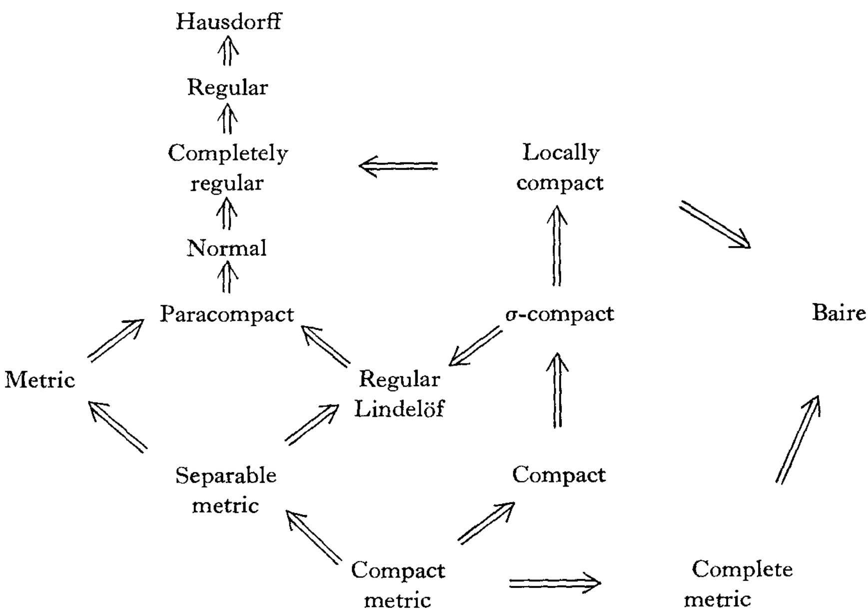
We now prove:

9.8 Theorem Each gauge space $(Y, \mathscr{T}(\mathscr{D}))$ is uniformly isomorphic to a dense subspace of a complete gauge space $(\hat{Y}, \mathscr{T}(\hat{\mathscr{D}}))$. Furthermore, $(\hat{Y}, \mathscr{T}(\hat{\mathscr{D}}))$ is unique, up to a uniform isomorphism.

Proof: The uniqueness comes from **9.7**; we now prove the existence. Let $\{C_d(Y) \mid d \in \mathscr{D}\}$ be a family of copies of the metric space $(C(Y), d_e^+)$ indexed by the gauges $d \in \mathscr{D}$. The gauge structure for the space $\prod \{C_d(Y) \mid d \in \mathscr{D}\}$ derived by using the metric d_e^+ in each factor is, by **9.4** and **2.6**, a complete gauge structure. For each $d \in \mathscr{D}$ and $y \in Y$, let $y_d: Y \to E^1$ be the continuous function $y_d(z) = d(z, y) - d(z, p)$, where p is a fixed element of Y, and let $j: Y \to \prod \{C_d(Y) \mid d \in \mathscr{D}\}$ be the map given by $j(y) = \{y_d\}$, that is, the dth coordinate of $j(y)$ is y_d. Because $\mathscr{D}$ is separating, the map j is readily verified to be injective, and because (as in XIII, **5.2**) we have $d_e^+(y_d, x_d) = d(y, x)$, it follows that j is a uniform isomorphism of $(Y, \mathscr{T}(\mathscr{D}))$ onto $(j(Y), \mathscr{T}(\mathscr{D}_e))$. Thus, $\overline{j(Y)}$ is, by (**9.4**), the required complete gauge space.

We remark that, in this broad extension of the concept of completeness from metric spaces to completely regular spaces, one of the most important consequences of metric completeness is significantly weakened: there seems to be no generalization of Baire's theorem to gauge spaces that does not involve some fairly restrictive hypotheses on the gauge structure (*cf.*, **4**, Ex. 1, for example).

Diagram of the Main Classes of Topological Spaces Discussed in This Book



Without additional hypotheses, none of the implications is reversible.

Problems

Section 1

1. Let d be the Euclidean metric in E^n. Show that a sequence in E^n is d-Cauchy if and only if each coordinate is d_e-Cauchy.
2. Prove: A Cauchy sequence can converge to at most one point.
3. Using the fact that every bounded monotone sequence in E^1 converges, show that a d_e-Cauchy sequence $\{a_n\}$ in E^1 converges to $\sup_m \left\{ \inf_{i \geq m} a_i \right\}$.

Section 2

1. Let (Y, d) be a metric space. Prove: Each arbitrary intersection and each finite union of d-complete subspaces is also d-complete.
2. Let $l^2(\aleph_0)$ be Hilbert space and d its metric. Show that d is complete (use the diagonal process).
3. Let (Y, d), (Z, d') be metric spaces and $f: Y \cong Z$ a homeomorphism such that $d(y, y') \leqslant \alpha\, d'(f(y), f(y'))$ for all $y, y' \in Y$ and some fixed $\alpha > 0$. Prove: If d is complete, so also is d'.
4. Prove that every discrete space is topologically complete.

Section 3

1. Prove:
   a. Any subordinate of a d-Cauchy filterbase is also a d-Cauchy filterbase.
   b. If a d-Cauchy filterbase has a convergent subordinate, then it itself converges.
2. Prove that Y is d-complete if and only if: For every descending sequence $A_1 \supset A_2 \supset \cdots$ of nonempty closed sets having d-diameter $\delta(A_i) \to 0$, the intersection $\bigcap_i A_i$ is not empty.
3. Let Y be d-complete. Let $\{A_i \mid i = 1, 2, \cdots\}$ be a countable family of arbitrary sets such that $A_i \cap A_j \neq \varnothing$ for all (i, j), and such that $\sum_1^\infty \delta(A_i) < \infty$. Show that there exists a $y_0 \in Y$ such that each nbd $U(y_0)$ contains all but at most finitely many A_i.
4. Let Y be d-complete, let Z be an arbitrary (Hausdorff!) space, and let $f: Y \to Z$ be continuous. Assume $A_1 \supset A_2 \supset \cdots$ is a descending sequence of closed nonempty sets and $\delta(A_i) \to 0$. Prove:
$$f\left(\bigcap_1^\infty A_i\right) = \bigcap_1^\infty f(A_i).$$
5. Let Y be totally d-bounded. Show that d is a bounded metric for Y.
6. Prove: Y is totally d-bounded if and only if each infinite sequence in Y contains a d-Cauchy subsequence.
7. Let Y be d-complete, and let Z be any metric space. Let $f: Y \to Z$ be a continuous map with the property:
$$\forall r > 0 \;\exists\, \rho = \rho(r) > 0 \;\forall\, y \in Y \;: \overline{f[B(y, r)]} \supset B[f(y), \rho].$$
Prove that f is an open mapping. [*Hint:* Show that $f[B(y, r + \varepsilon)] \supset B[f(y), \rho]$ for each $\varepsilon > 0$.]

Section 4

1. Prove: Any open and any closed subset of a topologically complete space is a Baire space.
2. Let Y be complete, and let each G_i be a G_δ-set in Y. Assume each G_i to be dense in Y; show that $\bigcap_1^\infty G_i$ is dense in Y.
3. Prove that the Cantor set is nowhere dense in E^1.
4. Show that if Y is topologically complete and has no isolated points, then $\aleph(Y) > \aleph_0$.

Section 5

1. Show that a d-complete space may be isometric with a proper closed subset of itself. [*Hint:* Let f: $]0, \infty[\to]0, \infty[$ be the map $f(x) = x + 1$.]
2. Let (X, d) be a compact metric space, and let $A \subset X$ be dense. Let Y be d'-complete. Prove: A map f: $A \to Y$ has a continuous extension over X if and only if it is uniformly continuous.
3. Show that the homeomorphism $x \to x^3$ of (E^1, d_e) on itself preserves d_e-Cauchy sequences, but is not a uniform isomorphism.

Section 6

1. Let X, Y be complete, and f: $X \to Y$ uniformly continuous. Prove: If A is relatively compact, so also is $f(A)$.
2. Let (Y, d) be a metric space, and $A \subset Y$ any subset. Let i: $Y \to \hat{Y}$ be the embedding of Y in its completion $\hat{Y}$. Let j: $A \to Y$ be the inclusion map, and $\hat{A}$ the completion of A (with metric $d \mid A$). Show: $\hat{j}$: $\hat{A} \to \hat{Y}$ is an isomorphism onto the closure of $i(A)$ in $\hat{Y}$.
3. Let $\{(X_i, d_i) \mid i \in Z^+\}$ be a family of metric spaces, and for each $i \in Z^+$ let $\hat{X}_i$ be the completion of (X_i, d_i). Prove: The completion of $\prod_1^\infty X_i$ is uniformly isomorphic to $\prod_1^\infty \hat{X}_i$.
4. Let (Y, d) be a metric space, and let C be the set of all d-Cauchy sequences in Y. Prove: (1) If $\{x_n\}, \{z_n\} \in C$, then $\lim d(x_n, z_n)$ exists. (2) The map $\hat{d}$: $C \times C \to E^1$ given by $\hat{d}(\{x_n\}, \{z_n\}) = \lim d(x_n, z_n)$ is a gauge on C. (3) If R is the equivalence relation $xRz \Leftrightarrow \hat{d}(x, z) = 0$ in C, and if ρ is the induced metric on C/R, then $(C/R, \rho)$ is isometric to the completion of (Y, d). [*Hint:* Observe that the map $Y \to C/R$ given by $y \to R\{y_n\}$, where $y_n = y$ for all n, is an isometry onto a dense subset.]

Section 7

1. Let $G(x, y)$ be a continuous real-valued function on $I_a \times I_b$. Assume that $G(x_0, y_0) = 0$ and $G_y(x_0, y_0) \neq 0$. Prove: There exists a positive s and a unique continuous h: $I_s \to I_b$ such that $h(x_0) = y_0$ and $G(x, h(x)) \equiv 0$ on I_s. [*Hint:* Define
$$F(x, y) = y - y_0 - \frac{G(x, y)}{G_y(x_0, y_0)}$$
and use the theorem in the text.]
2. Let $f(x, y)$ be continuous on $I_a \times I_b$ and satisfy some Lipschitz condition $|f(x, y) - f(x, \bar{y})| \leqslant k|y - \bar{y}|$ on $I_a \times I_b$. Prove: If $\bar{a}$ is chosen so that $\bar{a} \cdot \max|f| \leqslant b$ and $\bar{a} \cdot k < 1$, then the differential equation $y' = f(x, y)$ has a unique solution h: $I_{\bar{a}} \to I_b$ satisfying $h(x_0) = y_0$. [*Hint:* Consider the map
$$T(\varphi)(x) = y_0 + \int_{x_0}^{x} f(\xi, \varphi(\xi))\, d\xi.]$$

3. Let $K(x, y, z)$ be defined in $R = \{(x, y, z) \mid 0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ |z| \leqslant 1\}$. Assume $K(x, y, z) > 0$ on R, and $|K(x, y, z) - K(x, y, z')| \leqslant c\,|z - z'|$ on R. Prove: There exists a $\delta > 0$ such that for each λ, $0 \leqslant \lambda \leqslant \delta$, the integral equation

$$u(x) = \lambda \int_0^1 K(x, \xi, u(\xi))\, d\xi$$

has solutions.

4. Let X be compact metric, and $T\colon X \to \mathscr{P}(X)$ a map with the following property: $d(a, b) \geqslant d(x, y)$ for each x, y and each $a \in f(x)$, $b \in f(y)$. Prove: T is single-valued, and an isometry. [*Hint:* Starting with $a_0 = a$, $b_0 = b$, define

$$a_n \in T(a_{n-1}), \qquad b_n = T(b_{n-1}).$$

Show: $\forall\, \varepsilon > 0\ \exists\, n\ \exists\, k\colon [d(a_n, a_{n+k}) < \varepsilon] \wedge [d(b_n, b_{n+k}) < \varepsilon]$. Finally, note that $d(a_0, b_0) \leqslant d(a_1, b_1) \leqslant d(a_k, b_k) \leqslant d(a_0, b_0) + 2\varepsilon$.] Conclude:

   a. A compact metric space is not isometric to a proper subset.
   b. If X, Y are compact, if X is isometric to a subset of Y, and if Y is isometric to a subset of X, then X is isometric to Y.
   c. If X is compact, and if $T\colon X \to X$ is a surjection satisfying $d(Tx, Ty) \leqslant d(x, y)$ for all x, y, then T is an isometry.

These results are due to W. Hurewicz and H. Freudenthal.

Section 8

1. Let X, Y be topologically complete, separable metric spaces. Let $A \subset X$ be an arbitrary subset and $f\colon A \to Y$ such that $f\colon A \to f(A)$ is a continuous open mapping. Show that f extends over a G_δ-set $G \supset A$ to a map F such that $F\colon G \to F(G)$ is also a continuous open mapping.
2. (Lavrentieff's theorem.) Let X, Y be topologically complete spaces. Let $A \subset X$, $B \subset Y$ be arbitrary subsets and assume that there is a homeomorphism $h\colon A \cong B$. Prove: There exists an extension of h to a *homeomorphism* $H\colon G \cong G'$, where G, G' are G_δ-sets and $G \supset A$, $G' \supset B$.
3. Show that the topologically complete, separable metric spaces are the G_δ-sets in the Hilbert cube I^∞.
4. Prove: Every metric space that is locally topologically complete is topologically complete. [*Hint:* Use IX, 5, Problem 5.]

Homotopy

XV

Homotopy of maps plays an important role in modern topology, basically because most of the known algebraic invariants of spaces are homotopy invariants. In this chapter, we determine some elementary properties of this relation.

1. Homotopy

1.1 Definition Let X, Y be two spaces and I be the unit interval $\{t \mid 0 \leqslant t \leqslant 1\}$. Two maps $f, g: X \to Y$ are called homotopic (written: $f \simeq g$) if there exists a continuous $\Phi: X \times I \to Y$ such that $\Phi(x, 0) = f(x)$ and $\Phi(x, 1) = g(x)$ for each $x \in X$.

Intuitively, regarding t as a time parameter, Φ represents a continuous deformation of the map f to the map g, with $\Phi \mid X \times t_0$ being the stage of the deformation at instant t_0. Alternatively, $f \simeq g$ if there exists a continuous one-parameter family of maps $f_t: X \to Y$, $0 \leqslant t \leqslant 1$, starting with f and ending with g (remembering that continuity of the family is with respect to the two variables x, t jointly). We write $\Phi: f \simeq g$ to mean "Φ establishes a homotopy of f to g." An $f: X \to Y$ homotopic to a constant map is called nullhomotopic, written $f \simeq 0$.

Ex. 1 Let Y be a convex subset of E^n or, in fact, any linear topological space, and let X be an arbitrary space. Then each $f: X \to Y$ is nullhomotopic. Indeed, choose any $y_0 \in Y$ and define $\Phi(x, t) = ty_0 + (1 - t)f(x)$, which sends each $x \in X$ to the point on the line segment joining y_0 to $f(x)$ that divides it in the ratio $(1 - t)/t$. Then $\Phi: X \times I \to Y$ is continuous, since addition and scalar multiplication are continuous operations in linear spaces, and is a homotopy of f to a constant map because $\Phi(x, 0) = f(x)$, $\Phi(x, 1) = y_0$. More generally, call a space Y *contractible* if $1: Y \to Y$ is nullhomotopic; then, for every space X, each continuous $f: X \to Y$ is nullhomotopic: if $\Phi: Y \times I \to Y$ is a nullhomotopy of the identity map of Y, then $\Psi(x, t) = \Phi[f(x), t]$ is a nullhomotopy of f.

Ex. 2 Reversing the situation of Ex. 1, let X be a convex subset of a linear topological space and Y be an arbitrary space; then, again, each $f: X \to Y$ is nullhomotopic. For, fixing any $x_0 \in X$ and defining $\Phi(x, t) = f(tx_0 + (1 - t)x)$, we have that Φ is continuous and shrinks the image of X to $f(x_0)$. More generally, if X is contractible, then for any space Y, each continuous $f: X \to Y$ is nullhomotopic: if $\Phi: X \times I \to X$ is a nullhomotopy of the identity map $1: X \to X$, then $\Psi(x, t) = f(\Phi(x, t))$ is a nullhomotopy of f.

Ex. 3 Nullhomotopic maps need not be homotopic; indeed, constant maps into a space need not be homotopic. Let X be connected, Y not connected, and y_0, y_1 points in distinct components of Y; then the constant maps $x \to y_0$ and $x \to y_1$ are not homotopic, since $X \times I$ is connected.

Ex. 4 It is important to note that, in considering the homotopy of two maps $f, g: X \to Y$, the space Y in which the deformation is to take place must be specified in advance and kept fixed in the discussion. If it is permitted to enlarge Y arbitrarily, then any two given $f, g: X \to Y$ are homotopic: Simply attach $X \times I$ to Y by $F: (X \times 0) \cup (X \times 1) \to Y$, where $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$, to get $\hat{Y}$; if $p: (X \times I) + Y \to \hat{Y}$ is the identification map then $p \mid X \times I$ is a homotopy of f to g *in* $\hat{Y}$.

By Ex. 1 and Ex. 2, all maps of E^n and all maps into E^n, are nullhomotopic. Replacing E^n by the slightly more complicated space S^n makes both statements false: in the simple case $n = 0$, it is clear that the identity map $1: S^0 \to S^0$ is not nullhomotopic. However, the following proposition is very useful.

1.2 (1). If X is any space and $f, g: X \to S^n$ two maps such for each $x \in X, f(x)$ and $g(x)$ are not antipodal, then $f \simeq g$. In particular, a nonsurjective $f: X \to S^n$ is always nullhomotopic.

(2). Let Y be any space and $f: S^n \to Y$. Then $f \simeq 0$ if and only if f has a continuous extension $F: V^{n+1} \to Y$.

Proof: (1). The deformation is described by pushing f to g along the (unique!) shortest arc of great circle joining each $f(x)$ to $g(x)$. Explicitly, let

$$\Phi(x, t) = \frac{t \cdot g(x) + (1 - t) \cdot f(x)}{|t \cdot g(x) + (1 - t) \cdot f(x)|};$$

because $f(x)$, $g(x)$ are never antipodal, the denominator is never zero, so Φ is continuous and shows $f \simeq g$. For the second part, choose any $s_0 \in S^n - f(X)$; then f and the constant map $X \to -s_0$ are never antipodal.

(2). If $F: V^{n+1} \to Y$ is a continuous extension of $f: S^n \to Y$, then defining $\Phi: S^n \times I \to Y$ by

$$\Phi(x, t) = F([1 - t]x),$$

Φ is a homotopy of f to the constant map $S^n \to F(0)$. Conversely, if Φ is a homotopy of f to a constant map 0, then the same formula shows that F is uniquely defined at the origin (since $F(0) = \Phi(x, 1) =$ constant), and that it is indeed a continuous extension $F: V^{n+1} \to Y$ of f.

The question whether or not two maps $f, g: X \to Y$ are homotopic is essentially an extension problem. In fact, defining a map

$$\varphi: (X \times 0) \cup (X \times 1) \to Y$$

by $\varphi(x, 0) = f(x)$, $\varphi(x, 1) = g(x)$, we have φ a continuous map of the closed subset $(X \times 0) \cup (X \times 1) \subset X \times I$ into Y, and $f \simeq g$ if and only if φ can be extended to a continuous $\Phi: X \times I \to Y$. The problem of whether a given f is *nullhomotopic* can, additionally, be reduced to another extension problem:

1.3 Let X be any space, and TX the cone over X. An $f: X \to Y$ is nullhomotopic if and only if f has a continuous extension

$$F: TX \to Y.$$

Proof: Let $p: X \times I \to TX$ be the identification map; since $p \mid X \times 0$ is a homeomorphism, we regard X and $p(X \times 0)$ as identical. If $\Phi: f \simeq 0$, then Φp^{-1} is single-valued and so is a continuous map $TX \to Y$ extending f. Conversely, if f is extendable to $F: TX \to Y$, then $F \circ p: X \times I \to Y$ shows $f \simeq 0$.

Observe that **1.2**(2) is a special case of **1.3** because (XI, **2**, Ex. 5) we know that $TS^n \cong V^{n+1}$.

2. Homotopy Classes

2.1 Let Y, Z be two spaces. The relation of homotopy is an equivalence relation in the set Z^Y and so decomposes Z^Y into mutually exclusive classes (called homotopy classes), two maps being in the same class if and only if they are homotopic. The set of homotopy classes is denoted by $[Y, Z]$, and the homotopy class of an $f \in Z^Y$ by $[f]$.

Proof: The relation of homotopy is reflexive: $\Phi(y, t) \equiv f(y)$ shows $f \simeq f$. It is symmetric: If $\Phi: f \simeq g$, then $\Psi(y, t) = \Phi(y, 1-t)$ shows $g \simeq f$. It is transitive: If $\Phi: f \simeq g$ and $\Psi: g \simeq h$, then

$$\Delta(y, t) = \begin{cases} \Phi(y, 2t) & 0 \leqslant t \leqslant \frac{1}{2} \\ \Psi(y, 2t - 1) & \frac{1}{2} \leqslant t \leqslant 1 \end{cases}$$

is continuous, since (III, **9.4**) its two definitions agree on $X \times \frac{1}{2}$, and $\Delta: f \simeq h$.

Ex. 1 Two *constant* maps $f, g\colon Y \to Z$ are homotopic if and only if the points $f(Y) = z_f$ and $g(Y) = z_g$ belong to a common path-component of Z. For, if $\Phi\colon f \simeq g$, then choosing any $y_0 \in Y$, $\alpha(t) = \Phi(y_0, t)$ is a path joining z_f to z_g, and conversely, if α is such a path, then $\Phi(y, t) = \alpha(t)$ shows $f \simeq g$. Thus it is only whenever Z is path-connected that all nullhomotopic maps are homotopic (that is, belong to the same homotopy class) and the concept of nullhomotopy is independent of the point to which the map is contracted.

For elementary properties of the homotopy relation,

2.2 Theorem (1). (Composition.) Let $f, f'\colon X \to Y$ and $g, g'\colon Y \to Z$. If $f \simeq f'$ and $g \simeq g'$, then $g \circ f \simeq g' \circ f'$.

(2). (Restriction.) If $f, g\colon X \to Y$ are homotopic, then

$$f|A \simeq g|A$$

for any $A \subset X$.

(3). (Cartesian product.) $f, g\colon X \to \prod_\alpha Y_\alpha$ are homotopic if and only if $p_\alpha \circ f \simeq p_\alpha \circ g$ for each α.

(4). (Attaching.) Let X be attached to Y by $f\colon A \to Y$, where $A \subset X$ is closed, and let $p\colon X + Y \to X \cup_f Y$ be the identification map. Let $g_0, g_1\colon X \to Z$ and $h_0, h_1\colon Y \to Z$. If $\Phi\colon g_0 \simeq g_1$ and $\Psi\colon h_0 \simeq h_1$ are "consistent" (that is, if $\Phi(a, t) = \Psi(f(a), t)$ for all $(a, t) \in A \times I$), then $(g_0, h_0)p^{-1}$ and $(g_1, h_1)p^{-1}$ are homotopic maps of $X \cup_f Y$ into Z.

Proof: (1). Let $\Phi: f \simeq f'$ and $\Psi: g \simeq g'$; then

$$\Delta(x, t) = \begin{cases} g\Phi(x, 2t) & 0 \leqslant t \leqslant \frac{1}{2} \\ \Psi[f'(x), 2t - 1] & \frac{1}{2} \leqslant t \leqslant 1 \end{cases}$$

is clearly a continuous map $X \times I \to Z$ and establishes the required homotopy.

(2). If $i\colon A \to X$ is the inclusion map, then $f | A = f \circ i$, so this result is a special case of (1).

(3). If $\Psi: f \simeq g$, then by (1) we have $p_\alpha \circ f \simeq p_\alpha \circ g$ for each $\alpha \in \mathscr{A}$. Conversely, if $\Phi_\alpha: p_\alpha \circ f \simeq p_\alpha \circ g$ for each α, then the map $(x, t) \to \{\Phi_\alpha(x, t)\}$ of $X \times I \to \prod_\alpha Y_\alpha$ is (IV, **2.3**) continuous and shows $f \simeq g$.

(4). According to XII, **4.2**, the map $\Delta = (\Phi, \Psi)(p \times 1)^{-1}$ of $(X \cup_f Y) \times I$ into Z is continuous, and Δ provides the required homotopy.

For behavior of $[Y, Z]$ under continuous maps,

2.3 Let X, Y be two spaces and $\varphi: X \to Y$ continuous. Then:

(a). For each Z, φ induces a map $\varphi^\#: [Y, Z] \to [X, Z]$, by $\varphi^\#[f] = [f \circ \varphi]$.
(b). For each Z, φ induces a map $\varphi_\#: [Z, X] \to [Z, Y]$ by $\varphi_\#[g] = [\varphi \circ g]$.
(c). If $\varphi \simeq \psi$, then the maps induced by φ, ψ are the same.
(d). If $\psi: Y \to S$, then $(\psi \circ \varphi)^\# = \varphi^\# \circ \psi^\#$ and $(\varphi \circ \psi)_\# = \varphi_\# \circ \psi_\#$.

Proof: These proofs involve straightforward applications of **2.2**(1), and are left for the reader.

It follows directly from **2.3**(d) that $[Y, Z]$ is a joint topological invariant of Y and Z; intuitively, $[Y, Z]$ measures the number of "essentially different" ways that Y can be mapped into Z, and so provides information on the "complexity" of the topological structure of Z relative to that of Y. For example, $\aleph([E^1, Z])$ is the number of path components of Z (as follows from Ex. 1 and **I**, Ex. 2). The explicit determination of $[Y, Z]$ in terms of calculable algebraic invariants of Y and Z is one of the main problems in homotopy. Only particular cases have been settled, even for such simple spaces as spheres; for example, if $n \geqslant 4$, it is known that $[S^n, S^2]$ is a finite set, but the exact number of elements is not yet known for any large value of n (say, $n \geqslant 50$).

3. Homotopy and Function Spaces

A topology in the *set* Z^Y decomposes it into path components. On the other hand, the relation of homotopy decomposes the set Z^Y into homotopy classes. One advantage of the c-topology in Z^Y is that under mild restrictions on Y, the homotopy classes are exactly the path components. This follows from

3.1 Theorem Let Z^Y be given the c-topology. Then:

(1). If $f \simeq g$, then f and g lie in a common path component.
(2). If f and g lie in a common path component of Z^Y, then $f \simeq g$, *provided* Y is a k-space.

Proof: (1). Let $\Phi: f \simeq g$; because $\Phi: Y \times I \to Z$ is continuous, XII, **3.1**, assures that the associated map $\hat{\Phi}: I \to Z^Y$ is also continuous, and $\hat{\Phi}$ is evidently a path in Z^Y joining f to g.

(2). Let $\hat{\Phi}: I \to Z^Y$ be a path joining f to g. Since (XII, **4.4**) $Y \times I$ is a k-space, XII, **3.2** states that the associated $\Phi: Y \times I \to Z$ is continuous, and clearly, $\Phi: f \simeq g$.

A further advantage of the c-topology is that homotopic maps induce homotopic maps of function spaces, again under mild restrictions.

3.2 Theorem In all function spaces, use the c-topology and let Z be an arbitrary space.

(1). If $f_0, f_1: X \to Y$ are homotopic, and if Y is locally compact or if X is a k-space, then the induced maps $f_0^+, f_1^+: Z^Y \to Z^X$ are homotopic.
(2). If $g^0, g^1: Y \to Z$ are homotopic, and if Y is locally compact or if X is a k-space, then the induced maps $g_+^0, g_+^1: Y^X \to Z^X$ are homotopic.

Proof: In case Y is locally compact, the proof of (1) and (2) depends on XII, **2.2**, that the composition $T: Y^X \times Z^Y \to Z^X$ is continuous.

Ad (1). Let $\Phi: f_0 \simeq f_1$; because (XII, **3.1**) the associated $\hat{\Phi}: I \to Y^X$ is continuous, so also (IV, **2.5**) is $\hat{\Phi} \times 1: I \times Z^Y \to Y^X \times Z^Y$; thus $T \circ (\hat{\Phi} \times 1)$ is continuous and shows that $f_0^+ \simeq f_1^+$. The proof of (2) is similar.

In case X is a k-space, the proof depends on XII, **5.3** and **4.4**, that $Z^{X \times I} \cong (Z^X)^I \cong (Z^I)^X$.

Ad (1). Let $\Phi: f_0 \simeq f_1$; by XII, **2.1**, the induced

$$\Phi^+: Z^Y \to Z^{X \times I} \cong (Z^X)^I$$

is continuous, and [XII, **3.1**(2)] the associated continuous $Z^Y \times I \to Z^X$ provides the required homotopy.

Ad (2). Let $\Phi: g^0 \simeq g^1$; then the associated $\hat{\Phi}: Y \to Z^I$ is continuous, so that the induced $(\hat{\Phi})^+: Y^X \to (Z^I)^X \cong (Z^X)^I$ is also continuous; and the associated continuous $Y^X \times I \to Z^X$ shows $g_+^0 \simeq g_+^1$.

4. Relative Homotopy

In many cases, attention is restricted only to maps $f\colon Y \to Z$ having special properties, and two such maps are called homotopic only if one can be deformed to the other in a certain way. The usual pattern for this notion is

4.1 Definition Let $\mathscr{H} \subset Z^Y$ be some subset, and call its members $\mathscr{H}$-maps. Two $\mathscr{H}$-maps are $\mathscr{H}$-homotopic if there is a homotopy $\Phi\colon Y \times I \to Z$ such that $\Phi \mid Y \times t$ is an $\mathscr{H}$-map for each $t \in I$.

Ex. 1 Let $f \in Z^Y$ be a given continuous map, let $A \subset Y$ be any subset, and define $\mathscr{H} = \{g \in Z^Y \mid g|A = f|A\}$. Then two $\mathscr{H}$-maps g, h are $\mathscr{H}$-homotopic only if there is a homotopy of g to h such that the image of each point $a \in A$ remains fixed [at $f(a)$] throughout the entire deformation. Such homotopies are frequently considered; to denote that g, f are homotopic in this way, we shall in the future write "$g \simeq f$ rel A."

Ex. 2 Let $A \subset Y$, $B \subset Z$, and define $\mathscr{H} = \{f \in Z^Y \mid f(A) \subset B\}$; since this set $\mathscr{H}$ occurs frequently, an $\mathscr{H}$-map will be denoted briefly by $f\colon (Y; A) \to (Z; B)$. An $\mathscr{H}$-homotopy of $\mathscr{H}$-maps is now one in which the image of A remains in B (not necessarily pointwise fixed!) during the entire deformation.

Ex. 3 It is evident that $\mathscr{H}$-homotopy of $\mathscr{H}$-maps is more restrictive than ordinary homotopy, and the following trivial example shows this. Let $\mathscr{H} \subset I^I$ be $\{f \mid f\colon (I; \operatorname{Fr} I) \to (I; \operatorname{Fr} I)\}$; then, although the $\mathscr{H}$-maps $f(t) = t$, $g(t) = 1 - t$ are homotopic, it is obvious that they are not $\mathscr{H}$-homotopic.

Let $\mathscr{H} \subset Z^Y$. As before, $\mathscr{H}$-homotopy of $\mathscr{H}$-maps is an equivalence relation in $\mathscr{H}$; the set of $\mathscr{H}$-homotopy classes is denoted by $[Y, Z; \mathscr{H}]$. Regarding $\mathscr{H}$ as a subspace of the c-topologized Z^Y, then whenever Y is a k-space the $\mathscr{H}$-homotopy classes are precisely the path components of $\mathscr{H}$.

As an illustration of these ideas, we prove

4.2 (J. W. Alexander) Let $h\colon V^n \to V^n$ be a homeomorphism such that $h \mid S^{n-1} = 1$. Then $h \simeq 1$ rel S^{n-1}, with each stage of the deformation being a homeomorphism. Alternatively stated: If $\mathscr{H}$ is the set of all homeomorphisms $V^n \to V^n$ coinciding with the identity map on S^{n-1}, then $\mathscr{H}$ is path-connected.

Proof: For each $t \in \,]0, 1]$ define

$$\Phi(x, t) = \begin{cases} x & |x| \geqslant t \\ th(x/t) & |x| \leqslant t, \end{cases}$$

which keeps all points outside the ball V_t^n of radius t fixed and duplicates on $V_t^n \to V_t^n$ the given transformation h on a diminished scale. Φ is not defined for $t = 0$; however, we set $\Phi(x, 0) = x$. It is readily seen that $\Phi\colon V^n \times I \to V^n$ is continuous, that $\Phi\colon 1 \simeq h \operatorname{rel} S^{n-1}$, and that $\Phi \mid V^n \times t$ is a homeomorphism for each t.

5. Retracts and Extendability

Let X be a given space. In this and the next section we will be concerned with characterizing those subsets $A \subset X$ having the two properties: For all spaces Z, (a) each $f: A \to Z$ is extendable over X; and (b) two maps $F, G: X \to Z$ are homotopic whenever only $F \mid A \simeq G \mid A$. To see the importance of such sets, note that since the map $i^{\#}: [X, Z] \to [A, Z]$ induced by the inclusion map $i: A \to X$ is exactly $[F] \to [F \mid A]$, it follows from (a) that $i^{\#}$ is surjective, and from (b) that $i^{\#}$ is injective; thus, for any space Z, the calculation of $[X, Z]$ can always be reduced to the (perhaps easier, or known) determination of $[A, Z]$.

We begin with the extension problem. Conditions for extending a given $f: A \to Z$ over X generally involve Z also. For example, if X is a connected normal space and A is the union of two disjoint closed subsets, every continuous $f: A \to E^1$ is extendable over X (cf. VII, **5.1**), whereas this is not true for continuous maps $f: A \to 2$. However, there is an important case where such conditions do not involve Z.

5.1 Definition Let X be a space and $A \subset X$. A is a retract of X if the identity map $1: A \to A$ is extendable to a continuous $r: X \to A$; such an extension is called a retraction.

Equivalently, A is a retract of X if there exists an $r: X \to A$ such that $r(a) = a$ for each $a \in A$, or alternatively, if $r: X \to A$ is surjective and $r \circ r(x) = r(x)$ for each $x \in X$. The concept of retract is evidently topologically invariant: if $h: X \cong X'$, then $h(A)$ is a retract of X' if and only if A is a retract of X.

Ex. 1 Let X be any space. Then X and each $x_0 \in X$ are retracts of X.

Ex. 2 The unit ball V^n is a retract of E^n, as the map $r(x) = x/|x|$ (if $|x| \geqslant 1$ and $r(x) = x$ otherwise) shows. Further, S^{n-1} is a retract of $E^n - \{0\}$.

Ex. 3 Let $\prod_\alpha Y_\alpha$ be any cartesian product, and for each α, let $B_\alpha \subset Y_\alpha$ be a retract. Then $\prod_\alpha B_\alpha$ is a retract of $\prod_\alpha Y_\alpha$; for, if each $r_\alpha: Y_\alpha \to B_\alpha$ is a retraction, so also is $\prod_\alpha r_\alpha: \prod_\alpha Y_\alpha \to \prod_\alpha B_\alpha$. In particular, $X \times 0$ is a retract of $X \times I$.

Ex. 4 The Hilbert cube I^∞ is a retract of the Hilbert space $l^2(\aleph_0)$. In fact, for each $n = 1, 2, \cdots$, let $r_n: E^1 \to [-1/n, 1/n]$ be a retraction; then the map $r: l^2(\aleph_0) \to I^\infty$, given by $\{x_n\} \to \{r_n(x_n)\}$, is continuous and is a retraction.

Ex. 5 If Z^Y be given the c-topology, then (XII, **1.2**) there is an embedding $h: Z \to Z^Y$, where $h(z) =$ constant map sending Y to the point z. $h(Z)$ is a retract of Z^Y: Choosing any $y_0 \in Y$, the evaluation map $\omega_{y_0}: Z^Y \to Z$ is continuous and $h \circ \omega_y$ is clearly a retraction.

5.2 If X is Hausdorff, and $A \subset X$ a retract of X, then A is closed in X.

Proof: If there were some $x \in \bar{A} - A$, then because $r(x) \neq x$ and X is Hausdorff, there would exist disjoint nbds $U \supset \{x\}$, and $V \supset \{r(x)\}$ such that $r(U) \subset V$; however, since $x \in \bar{A}$, there must be some $a \in A$ in U, and since $a = r(a) \in V$, this contradicts the disjointness of U and V.

Ex. 6 If X is not Hausdorff, **5.2** need not be true, as Sierpinski space shows (*cf.* Ex. 1).

We now show that any continuous map of a retract of a space can always be extended over that space; in fact

5.3 Theorem Let X be any space and $A \subset X$. A necessary and sufficient condition that A be a retract of X is that for every space Z, each continuous $f: A \to Z$ is extendable over X.

Proof: If A is a retract of X and $r: X \to A$ a retraction, then $f \circ r$ is an extension of any given f over X. Conversely, if the condition is satisfied, A is a retract of X: Choose $Z = A$ and $f: A \to A$ to be the identity map.

Remark: Any given extension question can always be formulated as a retraction problem. Let A, X, Z be fixed and $f: A \to Z$; then f is extendable over X if and only if Z is a retract of $X \cup_f Z$. For, if $p: X + Z \to X \cup_f Z$ is the identification map and r is a retraction, $r \circ p \mid X$ is an extension of f. Conversely, if an extension F of f is given, define $\rho: X + Z \to Z$ by $\rho(x) = F(x)$ if $x \in X$ and $\rho(z) = z$ if $z \in Z$; then $\rho p^{-1}: X \cup_f Z \to Z$ is single-valued, so it is continuous and is a retraction of $X \cup_f Z$ onto Z. In particular, A is a retract of X if and only if under any attachment to any space Z by any continuous $f: A \to Z$, the subspace Z is a retract of $X \cup_f Z$.

6. Deformation Retraction and Homotopy

In this section, we find conditions on A which assure, for all spaces Z and any two continuous $F, G: X \to Z$, that $F \simeq G$ whenever

$$F \mid A \simeq G \mid A.$$

Such conditions generally involve Z also.

Ex. 1 Let $X = Z = [0, 1] \cup [2, 3]$ and $A = [0, 1]$. Then the inclusion map $f: A \to Z$ extends to the identity map F, and also to the map G, where $G(x) = x$ for $x \in [0, 1]$ and $G(x) = x - 2$ otherwise. Clearly F and G are not homotopic; they would be if Z were taken as $[0, 3]$.

Again there is an important instance where the homotopy question does not involve Z.

6.1 Definition Let A, B be any two subsets of X. B is said to be deformable into A over X if the identity map $1: B \to B$ is homotopic in X to a map of B into A.

That is, we require a $\Phi: B \times I \to X$ such that $\Phi(b, 0) = b$ for each $b \in B$ and $\Phi(B \times 1) \subset A$. If we have $B = X$ in the definition, we omit "over X" and say simply that X is deformable into A; it is important to observe that we do not require that the image of A remain in A during the deformation.

This concept is clearly topologically invariant: if $h: X \cong X'$, then $h(B)$ is deformable into $h(A)$ over X' if and only if B is deformable into A over X.

Ex. 2 $E^n - \{0\}$ is deformable into S^{n-1}, as $\Phi(x, t) = (1 - t)x + t \cdot x/|x|$ shows.

Ex. 3 S^{n-1} is deformable over V^n to $\{0\}$, by $\Phi(x, t) = (1 - t) \cdot x$; similarly, V^n is deformable into $\{0\}$.

Ex. 4 If each Y_α is deformable into B_α, then $\prod_\alpha Y_\alpha$ is deformable into $\prod_\alpha B_\alpha$ [see **2.2**(3)].

Ex. 5 A set $A \subset X$ may be a retract of X, and yet X may not be deformable into A, as any point x_0 of a discrete space (having more than one point) shows.

Ex. 6 X may be deformable into A, yet A may not be a retract of X. For example, I is deformable into Fr (I), say by $\varphi(x, t) = t \cdot x$; but Fr (I) is evidently not a retract of I.

That this concept allows reduction of the homotopy question is

6.2 Theorem Let B be deformable into A over X. Let Z be any space and $f_0, f_1: X \to Z$ any two continuous maps. Then $f_0 \mid A \simeq f_1 \mid A$ implies $f_0 \mid B \simeq f_1 \mid B$. In particular, if X is deformable into A, then $f_0 \simeq f_1$ if and only if $f_0 \mid A \simeq f_1 \mid A$.

Proof: We prove only the first statement; because of **2.2**(2), the second is an immediate consequence. Let $\Phi: B \times I \to X$ be the deformation, and define $\varphi: B \to A$ by $\varphi(b) = \Phi(b, 1)$. Since $f_0 \mid A \simeq f_1 \mid A$, it follows from **2.2**(1) that $f_0 \circ \varphi \simeq f_1 \circ \varphi$. Since $\Psi(b, t) = f_0 \circ \Phi(b, t)$ shows $f_0 \mid B \simeq f_0 \circ \varphi$ and, in a similar manner, also, $f_1 \mid B \simeq f_1 \circ \varphi$, the proof is complete.

In view of **5.3** and **6.2**, we obtain subsets for which all maps are extendable, with preservation of homotopy relations, by combining the concepts of deformation and retraction.

6.3 Definition Let $A \subset B \subset X$. We call A a deformation retract of B over X if the identity map $1_B: B \to B$ is homotopic in X to a retraction $r: B \to A$. The set A is called a *strong* deformation retract of B over X if $r \simeq 1_B$ rel A, that is, keeping the points of A fixed throughout the entire deformation of B into A.

Ex. 7 In Ex. 2 and 3, all the deformations are strong deformation retractions.

Ex. 8 Let $X \subset E^2$ be the subspace $(0 \times I) \cup (I \times 0) \cup \{1/n \times I \mid n = 1, 2, \cdots\}$. Then $0 \times I$ is a deformation retract of X; but it evidently is not a strong deformation retract of X.

Ex. 9 Let V^n be the unit n-ball in E^n; then $(V^n \times 0) \cup (S^{n-1} \times I)$ is a strong deformation retract of $V^n \times I$. A deformation retraction is obtained by projecting $V^n \times I$ onto $(V^n \times 0) \cup (S^{n-1} \times I)$ from $(0, \cdots, 0, 2) \in E^n \times E^1$; in formulas,

$$\begin{aligned} \Phi(x, t) &= \left(\frac{x}{|x|}, 2 - \frac{2-t}{|x|}\right) & |x| &\geqslant \frac{2-t}{2} \\ &= \left(\frac{2x}{2-t}, 0\right) & |x| &\leqslant \frac{2-t}{2}. \end{aligned}$$

For the characterization question in **5**, we now have

6.4 Theorem Let $A \subset X$. Then A is a deformation retract of X if and only if it has the following two properties:

(1). For every space Z, each continuous $f: A \to Z$ is extendable over X, and

(2). For every space Z and each $F, G: X \to Z$, F is homotopic to G whenever $F \mid A \simeq G \mid A$.

Proof: Necessity is clear, from **6.2** and **5.3**. Sufficiency: from (1) and **5.3**, there is a retraction $r: X \to A$. Taking $Z = X$ and noting that the maps $r, 1_X: X \to X$ satisfy $r \mid A = 1_X \mid A$, we find from (2) that $r \simeq 1_X$.

As remarked, **6.4** implies: If A is a deformation retract of X, then $i^{\#}: [X, Z] \to [A, Z]$ is bijective for every Z. Thus, for example, the homotopy classes of maps of the punctured ball $V^n - \{0\}$ into any space are the same as those of S^{n-1} into the same space. Observe that **6.4**(1) is much stronger than required to assure that $i^{\#}$ is surjective: for this purpose, it would suffice to know for each $f: A \to Z$ simply that it is *homotopic* to a map that is extendable.

We now examine the concept of a deformation retract more closely. Definition **6.3** combines both retraction and deformation into one condition. That this concept is simply the conjunction of the two properties follows from

6.5 Theorem A is a deformation retract of B over X if and only if A is a retract of B and B is deformable into A over X.

Proof: Necessity is clear. For sufficiency, let $r: B \to A$ be a retraction, and $\Phi: B \times I \to X$ a deformation of B into A over X, where $\Phi(b, 0) = b$. Then

$$\begin{aligned} \Delta(b, t) &= \Phi(b, 2t) & 0 &\leqslant t \leqslant \tfrac{1}{2} \\ &= r \circ \Phi(b, 2 - 2t) & \tfrac{1}{2} &\leqslant t \leqslant 1 \end{aligned}$$

is continuous because $\Phi(b, 1) = r \circ \Phi(b, 1)$, and shows $1: B \to B$ homotopic to $r: B \to A$.

"Transitivity" takes the form

6.6 Let A, B, C be subsets of X such that $A \subset B \cap C$. Assume that A is a (strong) deformation retract of B over X. If C is deformable over X into B (rel A), then A is a (strong) deformation retract of C over X.

Proof: Let Φ be a deformation of C into B, and Ψ a deformation retraction of B onto A. The map $\Delta: C \times I \to X$ given by

$$\begin{aligned} \Delta(c, t) &= \Phi(c, 2t) & 0 \leqslant t \leqslant \tfrac{1}{2} \\ &= \Psi[\Phi(c, 1), 2t - 1] & \tfrac{1}{2} \leqslant t \leqslant 1 \end{aligned}$$

is the required deformation retraction and keeps A pointwise fixed whenever both Φ and Ψ do so.

7. Homotopy and Extendability

Let $A \subset X$ and Z be given. In general, a given $f: A \to Z$ may be homotopic to a map that is extendable over X, and yet f itself may not be extendable.

Ex. 1 Let X and Z be the space of **6**, Ex. 8, and let A be the subset

$$\{(a, 0) \mid a = 0, 1, \tfrac{1}{2}, \cdots\}.$$

Let $f: A \to Z$ be the inclusion map, and $g: A \to Z$ the map $g(a, 0) = (a, 1)$. Then f and g are homotopic—even nullhomotopic—and f is extendable over X, but clearly g is not extendable.

We now study a property that implies extendability depends only on the homotopy class of the given map.

7.1 Definition Let $A \subset X$ be given. We say that A has the absolute homotopy extension property ($AHEP$) in X if for each space Z and each continuous $F: X \to Z$, every homotopy of $F \mid A$ is extendable to a homotopy of F.

Whenever A has the $AHEP$ in X, then for any space Z, each continuous $g: A \to Z$ homotopic to an extendable $f: A \to Z$ is itself also extendable and, indeed, has an extension homotopic to any preassigned extension of f: For, if $F: X \to Z$ is any extension of f, then **7.1** says that the homotopy $\varphi: F \mid A \simeq g$ is extendable to a $\varphi: X \times I \to Z$ with $\Phi \mid X \times 0 = F$, so that $x \to \Phi(x, 1)$ is an extension of g over X homotopic to F. The extension found for g clearly depends on the extension of φ, so we have a related question: If F, $G: X \to Z$ are homotopic, and a specified homotopy of $F \mid A$ to $G \mid A$ is given, is this *homotopy* extendable to a homotopy of F to G?

Ex. 2 In Ex. 1, let $F = G =$ identity map of X, and let the given homotopy $F \mid A$ to $G \mid A$ be that in which the image of each $a \in A$ runs up to the top and then back down the spike containing it. This homotopy is clearly not extendable to one of F to G.

7.2 Theorem (1). A has the *AHEP* in X if and only if

$$(X \times 0) \cup (A \times I)$$

is a retract of $X \times I$.

(2). $(X \times 0) \cup (A \times I) \cup (X \times 1)$ is a retract of $X \times I$ if and only if for each space Z and each pair of maps $F, G: X \to Z$, any given homotopy of $F \mid A$ to $G \mid A$ can be extended to a homotopy of F to G.

Proof: (1). Assume that $(X \times 0) \cup (A \times I)$ is a retract of $X \times I$. A given $F: X \to Z$ and a given homotopy $\Phi: A \times I \to Z$ of $F \mid A$ determine a continuous $\psi: (X \times 0) \cup (A \times I) \to Z$ by setting

$$\psi(x, 0) = F(x), \psi(a, t) = \Phi(a, t).$$

According to **5.3**, ψ extends to a $\Psi: X \times I \to Z$ and Ψ is thus a homotopy of F extending Φ. Conversely, assume that A has the *AHEP* in X. Choose

$$Z = (X \times 0) \cup (A \times I),$$

let $F: X \to Z$ be the map $x \to (x, 0)$, and let the homotopy of $F \mid A$ be $\varphi(a, t) = (a, t)$. This extends to a homotopy Φ of F, and

$$\Phi: X \times I \to Z$$

is the desired retraction. The proof of (2) is similar.

In the remainder of this section, we are going to characterize those subsets $A \subset X$ for which $(X \times 0) \cup (A \times I)$ is a retract of $X \times I$; to do this, we will need the idea of a zero-set, so well as that of a halo.

A set C in a space X is called a zero-set in X if there exists a continuous $p: X \to I$ such that $p^{-1}(0) = A$, that is, such that p vanishes on, and only on, A. Clearly, such sets are closed G_δ-sets; and if X is normal, the zero-sets in X are exactly the closed G_δ-sets. Note that in any space Y, not necessarily Hausdorff, $Y \times 0$ is a zero-set in $Y \times I$, as the projection $p: Y \times I \to I$ shows.

The second of the required notions is given in

7.3 Definition Let X be a space and let $A \subset X$. An open set $U \supset A$ is called a halo of A if there exists a continuous $h: X \to I$ such that $A \subset h^{-1}(0)$ and $\{x \mid h(x) < 1\} \subset U$.

Clearly, X is always a halo of any subset $A \subset X$. In normal spaces X, every nbd of a closed $A \subset X$ is a halo of A; but in non-normal spaces, a closed $A \subset Y$ may have nbds that are not halos: in VII, **7**, Ex. 3, each nbd of the point a that excludes the point b is not a halo of a.

The desired characterization is given in

7.4 Theorem (D. Puppe) Let X be a (Hausdorff) space, and $A \subset X$ closed. The following four statements are equivalent:

(1). $(X \times 0) \cup (A \times I)$ is a retract of $X \times I$.
(2). A is a zero-set in X, and also a strong deformation retract over X of a halo $U \supset A$.
(3). There exists a continuous $\lambda: X \to E^1$ with $A \subset \lambda^{-1}(0)$ and a continuous $\Delta: X \times I \to X$ with $\Delta(x, 0) = x$ for each $x \in X$, such that $\Delta[x, \lambda(x)] \in A$ whenever $\lambda(x) \leqslant 1$.
(4). $(X \times 0) \cup (A \times I)$ is a zero-set in $X \times I$ and a strong deformation retract of $X \times I$.

Proof: (1) $\Rightarrow$ (2). Let $r: X \times I \to (X \times 0) \cup (A \times I)$ be a retraction, and let $J =]0, 1]$. Since $A \times J$ is open in $(X \times 0) \cup (A \times I)$, we have $r^{-1}(A \times J)$ open in $X \times I$ and therefore $(X \times 1) \cap r^{-1}(A \times J) = U \times 1$, where $U \subset X$ is an open set that necessarily contains A. Let $p_X: X \times I \to X$ be the projection and define $\Phi: U \times I \to X$ by $\Phi(u, t) = p_X \circ r(u, t)$; then Φ is a strong deformation retraction of U over X into A.

We now show that U is a halo of A. Let $w(x) = p_I \circ r(x, 1)$, where $p_I: X \times I \to I$ is the projection. For $u \in U$, we have $r(u, 1) \in A \times J$, so $w(u) > 0$; for $x \in U$ we must have $r(x, 1) \in X \times 0$, so $w(x) = 0$. Since $w(a) = 1$ for all $a \in A$, the function $h(x) = 1 - w(x)$ shows that U is a halo of A.

To show that A is a zero-set, consider the continuous function $F(x, t) = t - p_I \circ r(x, t)$, and define $p(x) = \sup\{F(x, t) \mid t \in I\}$. Then $p: X \to I$ is also continuous: its lower semi-continuity follows from III, **10.4** (a); and to see that p is also upper semi-continuous, observe that because I is compact, $p(x_0) \geqslant b$ if and only if $F(x_0, t) \geqslant b$ for some t; thus $\{x \mid p(x) \geqslant b\} = p_X\{(x, t) \mid F(x, t) \geqslant b\}$ and, by XI, **2.5**, this set is closed. We now prove that $A = p^{-1}(0)$. If $a \in A$, then $r(a, t) = t$ for each t, so $p(a) = 0$. Conversely, if $p(x) = 0$, then $p_I \circ r(x, t) \geqslant t$ for each $t > 0$ so that, in particular, $p_I \circ r(x, t) > 0$ for each $t > 0$; thus $r(x, t) \in A \times I$ for each $t > 0$ and, because $A \times I$ is closed, we must have $(x, 0) = r(x, 0) \in A \times I$ also, therefore $x \in A$.

(2) $\Rightarrow$ (3). Let $h: X \to I$ show that U is a halo of A, and let

$$\Phi: U \times I \to X$$

be a strong deformation retraction of U onto A. Let $p: X \to I$ show that

A is a zero-set, and define $\lambda(x) = 3\cdot\max[p(x), h(x)]$. Then $\lambda^{-1}(0) = A$, and $\{x \mid \lambda(x) \leqslant 2\} \subset U$. Define $H: X \times I \to X$ by

$$\begin{aligned} H(x, t) &= \Phi(x, t\cdot\min[2 - \lambda(x), 1]) && \text{if } \lambda(x) \leqslant 2 \\ &= x && \text{if } \lambda(x) \geqslant 2 \end{aligned}$$

This is continuous, since if $\lambda(x) = 2$, then $x \in U$ and the two definitions of H at (x, t) coincide. Moreover, $H(x, 0) = x$ for each $x \in X$, $H(a, t) = \Phi(a, t) = a$ for each $(a, t) \in A \times I$, and $H(x, 1) \in A$ whenever $\lambda(x) \leqslant 1$. Now define $\Delta: X \times I \to X$ by

$$\begin{aligned} \Delta(x, t) &= H(x, 0) && \text{if } \lambda(x) = 0 \\ &= H(x, \min[1, t/\lambda(x)]) && \text{if } \lambda(x) \neq 0 \end{aligned}$$

This map is actually continuous on $X \times I$. Indeed, it is clearly continuous at each (x, t) with $\lambda(x) \neq 0$, so we need only prove its continuity at each $(a, t) \in A \times I$. Let W be any nbd of $\Delta(a, t) = H(a, 0) = a$; since $H(a, I) = a$ and H is continuous, there is (XI, **2.6**) a nbd V of a such that $H(V, I) \subset W$; thus, $\Delta(a, t) \in W$ whenever simply $x \in V$ and therefore Δ is continuous at (a, t).

It is now immediate that $\Delta(x, 0) = x$ for each $x \in X$, and that $\Delta(x, \lambda(x)) = H(x, 1) \in A$ whenever $\lambda(x) \leqslant 1$.

(3) $\Rightarrow$ (1). Define $r: X \times I \to (X \times 0) \cup (A \times I)$ by

$$\begin{aligned} r(x, t) &= (\Delta(x, t), 0) && \text{if } t \leqslant \lambda(x) \\ &= (\Delta(x, \lambda(x)), t - \lambda(x)) && \text{if } t \geqslant \lambda(x) \end{aligned}$$

It is clear that r is continuous, and it is trivial to verify that r is a retraction.

(1) $\Rightarrow$ (4). Let $r: X \times I \to (X \times 0) \cup (A \times I)$ be a retraction, and define $\Gamma: (X \times I) \times I \to (X \times I)$ by

$$\Gamma[(x, t), s] = [p_X \circ r(x, st), (1 - s)t + sp_I \circ r(x, t)]$$

Then $\Gamma: 1 \simeq r$ is the desired deformation retraction. Moreover, because (1) $\Leftrightarrow$ (2), there is a $p: X \to I$ showing A is a zero-set in X; then $P(x, t) = t\,p(x)$ shows $(X \times 0) \cup (A \times I)$ is a zero-set in $X \times I$.

(4) $\Rightarrow$ (1) is trivial.

Ex. 3 The requirement in **7.4**(2) that A be a zero-set is essential: even in normal spaces (where every nbd of a closed set is a halo) a closed set may be a strong nbd deformation retract without being a zero-set. For example, let I^c be the cartesian product of $2^{\aleph_0}$ unit intervals, and let 0^c be the origin; then 0^c is a strong deformation retract of I^c, but 0^c is not a G_δ, so it is not a zero-set in I^c. In particular, $(I^c \times 0) \cup (0^c \times I)$ is not a retract of $I^c \times I$.

Ex. 4 We call $A \subset X$ a strong halo deformation retract in X if there is a halo $U \supset A$ and a strong deformation retraction $\Phi: U \times I \to X$ of U onto A. With this terminology, it follows from **7.2** that A has the *AHEP* in X if and only if A is a zero set in X and a strong halo deformation retract in X.

Ex. 5 Let X be any (Hausdorff) space and let TX be the cone over X. Then $(TX \times 0) \cup (X \times I)$ is a strong deformation retract of $TX \times I$. For, letting $V = \{\langle x, t\rangle \mid t < \frac{1}{2}\}$ we find that $X \subset TX$ is a strong deformation retract of V, and the continuous function $p\langle x, t\rangle = 2t$ shows both that V is a halo and that X is a zero-set. Observe that, because $TS^{n-1} \cong V^n$ (cf. XI, **2**, Ex. 5) this result contains that of 6, Ex. 9.

Remark The equivalence of (2) and (3) in **7.4** is frequently useful in showing that some set is a nbd deformation retract. We illustrate its use to prove

7.5 Let $A \subset X$ and $B \subset Y$ be zero-sets. Assume both are strong halo deformation retracts. Then $(A \times Y) \cup (X \times B)$ is a zero-set and a strong halo deformation retract in $X \times Y$.

Proof: Let Δ, λ and Γ, μ be functions satisfying the requirements in **7.4**(3) for A, B respectively. Define $H: X \times Y \times I \to X \times Y$ by $H(x, y, t) = (\Delta(x, t), \Gamma(y, t))$ and let $\sigma: X \times Y \to E^1$ be the map $\sigma(x, y) = \min[\lambda(x), \mu(y)]$; then H, σ satisfy the requirements of (3) for $(A \times Y) \cup (X \times B)$, and the proof is complete.

Whereas the retraction property of $X \times I$ onto $(X \times 0) \cup (A \times I)$ results from a nbd retraction property of A in X, the retraction property of $(X \times 0) \cup (A \times I) \cup (X \times 1)$ is the same as that of A in X:

7.6 Theorem Let X be a Hausdorff space. If $A \subset X$ is a zero-set and a strong [halo] deformation retract in X, then

$$D = (X \times 0) \cup (A \times I) \cup (X \times 1)$$

is a zero-set and a strong [halo] deformation retract in $X \times I$.

Proof: For "[halo]": This follows from **7.5**, since

$$D = (X \times \{0, 1\}) \cup (A \times I)$$

and $\{0, 1\}$ is clearly a zero-set and a strong halo deformation retract in I.

For the remaining case: We note that D is a strong deformation retract of $K = (X \times I) - (X - A) \times \frac{1}{2}$ since we apply **7.4** twice, first to $X \times [0, \frac{1}{2}]$, then to $X \times [\frac{1}{2}, 1]$, and finally remove the discontinuities, $(X - A) \times \frac{1}{2}$. According to **6.6**, the set D will be a deformation retract of $X \times I$ if we can deform $X \times I$ into K keeping D fixed throughout the deformation. Letting $\Phi: X \times I \to X$ be a strong deformation retraction onto A, this is accomplished by the deformation

$$\Gamma: (X \times I) \times I \to X \times I$$

given by

$$\begin{aligned} \Gamma[(x, t), s] &= [\Phi(x, 2st), t] & 0 \leq t \leq \tfrac{1}{2} \\ &= [\Phi(x, 2s(1 - t)), t] & \tfrac{1}{2} \leq t \leq 1 \end{aligned}$$

It remains to show that D is a zero-set: if $p: X \to I$ has A for zero-set, then $P(x, t) = t(1 - t) \cdot p(x)$ has D as zero-set.

8 Applications

In this section, we give some immediate applications of the main results, **7.4** and **7.6**.

As a first application, we obtain a condition under which the concepts "deformation retract" and "strong deformation retract" coincide:

8.1 Let X be a Hausdorff space and $A \subset X$ a zero-set. The A is a strong deformation retract of X if and only if A is both a deformation retract of X and a strong halo deformation retract in X.

Proof: "Only if" is trivial, so we need show only that if A is both a deformation retract of X and a strong halo deformation retract in X, then it is a strong deformation retract of X. Let $\Phi: X \times I \to X$ be a deformation such that $\Phi: 1_X \simeq r$, where $r: X \to A$ is a retraction. Let

$$D = (X \times 0) \cup (A \times I) \cup (X \times 1)$$

and define $\Omega: D \times I \to X$ by

$$\begin{aligned} \Omega_\lambda(x, 0) &= x \\ \Omega_\lambda(a, t) &= \Phi[a, (1 - \lambda)t] \\ \Omega_\lambda(x, 1) &= \Phi[r(x), 1 - \lambda] \end{aligned}$$

Since Ω_λ is consistently defined, it is continuous. The theorem will be proved if we can extend $\Omega_1: D \times 1 \to X$ over $(X \times I) \times 1$.

To this end, we observe that Ω_0 is extendable to a map $\bar{\Omega}_0: X \times I \to X$: $\bar{\Omega}_0(x, t) = \Phi(x, t)$ is the required extension, since $\bar{\Omega}_0$ clearly coincides with Ω_0 on $(X \times 0) \cup (A \times I)$ and on $X \times 1$ we have $\Omega_0(x, 1) = \Phi(r(x), 1) = r \circ r(x) = r(x) = \Phi(x, 1) = \bar{\Omega}_0(x, 1)$. By using $\bar{\Omega}_0$, we have a map $\bar{\Omega}: [(X \times I) \times 0] \cup [D \times I] \to X$. Now, by **7.6**, we find D is a zero set and a strong halo deformation retract of $X \times I$ so, by **7.4**, we have $[(X \times I) \times 0] \cup [D \times I]$ is a retract of $(X \times I) \times I$. Thus, $\bar{\Omega}$ extends to an $\Omega': (X \times I) \times I \to X$ and $\Psi(x, t) = \Omega'(x, t, 1)$ is the required strong deformation retraction of X onto A.

This has the curious consequence:

8.2 Let $A \subset X$ be a zero-set and a strong halo deformation retract in X. Then A is a strong deformation retract of X if and only if X can be deformed into A in such a way that the points of A remain in A during the entire deformation.

Proof: Since X is deformable into A, we need show only that A is a retract of X; then **6.5** and **8.1** yield the desired conclusion. To this end, let $\Phi: X \times I \to X$ be the described deformation, where $\Phi(x, 1) \equiv x$. Let $\Omega = \Phi \mid (X \times 0) \cup (A \times I)$; then Ω maps $(X \times 0) \cup (A \times I)$ into A and is the identity on $A \times 1$. By **5.3** and **7.4**, Ω extends to $\bar{\Omega}: X \times I \to A$, and $\bar{\Omega}(x, 1)$ is the required retraction of X onto A.

As another application, we use **7.4** and **7.6** in another way, to obtain homotopies behaving in a given way on specified subsets:

8.3 Let $f_0, f_1: X \to Y$ be homotopic, and $\Phi: f_0 \simeq f_1$. Let $A \subset X$ and let $\Psi: \Phi \mid A \times I \simeq \Phi_1$ be a given homotopy of $\Phi \mid A \times I$ rel $(A \times 0) \cup (A \times 1)$. If A is a zero-set in X and a strong halo deformation retract in X, then there is a homotopy $\Delta: f_0 \simeq f_1$ such that $\Delta \mid A \times I = \Phi_1$.

Proof: By **7.6**, $D = (X \times 0) \cup (A \times I) \cup (X \times 1)$ is a zero-set and a strong halo deformation retract of $X \times I$, consequently (**7.4**)

$$T = [(X \times I) \times 0] \cup [D \times I]$$

is a retract of $(X \times I) \times I$. Defining $\Delta: T \to Y$ by

$$\begin{aligned} \Delta(x, t, 0) &= \Phi(x, t) \\ \Delta(x, t, \lambda) &= f_0(x) \\ \Delta(x, 1, \lambda) &= f_1(x) \\ \Delta(a, t, \lambda) &= \Psi[(a, t), \lambda], \qquad a \in A, \end{aligned}$$

then Δ is continuous, is extendable over $(X \times I) \times I$, and $\Delta(x, t, 1)$ is the required homotopy.

Problems

Section 1

1. Let $\{I_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of unit intervals. Prove $\prod_\alpha I_\alpha$ is contractible.

Section 2

1. Prove: $\varphi: X \to Y$ is nullhomotopic if and only if for every space Z, the induced map $\varphi^\#: [Y, Z] \to [X, Z]$ is such that $\varphi^\#[f]$ is nullhomotopic for each f.

2. Let $\varphi: X \to Y$ be continuous. Prove: $\varphi^\#: [Y, Z] \to [X, Z]$ is surjective for *every* Z if and only if there is a continuous $\lambda: Y \to X$ such that $\lambda \circ \varphi \simeq 1_X$.

3. Let $\varphi: X \to Y$ be continuous. Prove: $\varphi^\#: [Y, Z] \to [X, Z]$ is injective for *every* Z if and only if there is a continuous $\rho: Y \to X$ such that $\varphi \circ \rho \simeq 1_Y$.

Section 3

1. Prove the following generalization of **3.1**:
   a. If t is a conjoining topology in Z^Y, then each path component is contained in a homotopy class.
   b. If t is a splitting topology in Z^Y, each homotopy class is contained in a path component of Z^Y.

Section 4

1. Let $\mathscr{A} \subset \mathscr{P}(Y)$ and $\mathscr{B} \in \mathscr{P}(Z)$ be families of sets. An inclusion-preserving map $\varphi: \mathscr{A} \to \mathscr{B}$ [that is, $\varphi(A) \subset \varphi(A')$ whenever $A \subset A'$] is called a *carrier*. For each carrier φ, let $\mathscr{H}_\varphi = \{f \in Z^Y \mid f(A) \subset \varphi(A) \text{ for each } A \in \mathscr{A}\}$. Given φ, define $\psi: \mathscr{P}(Y) \to \mathscr{P}(Z)$ by $\psi(C) = \bigcap \{\varphi(A) \mid A \supset C\}$ if such A exist; otherwise set $\psi(C) = Y$. Prove:
   a. ψ is a carrier.
   b. $[Y, Z; \mathscr{H}_\varphi] = [Y, Z; \mathscr{H}_\psi]$.

2. Let $\mathscr{H} = \{f \mid f: (V^{n+1}, S^n) \to (V^{n+1}, S^n)\}$. Let $f, g \in \mathscr{H}$ be such that $f|S^n = g|S^n$. Prove: f is $\mathscr{H}$-homotopic to g. More generally, prove that if $f|S^n$ and $g|S^n$ are homotopic as maps of $S^n \to S^n$, then f is $\mathscr{H}$-homotopic to g.

Section 5

1. Let $A \subset B \subset X$. Assume that B is a retract of X and A is a retract of B. Prove: A is a retract of X.
2. Let $B \subset Z$ be a retract of Z. For any space X, show that B^X is a retract of Z^X (c-topology in the function space).
3. Prove: Any closed convex set of a generalized Hilbert space $l^2(\aleph)$ is a retract of $l^2(\aleph)$.
4. Show: $\{0\} \cup \{1\}$ is not a retract of I.
5. Let A be a retract of the locally compact X. Prove that A is also locally compact.
6. Let $r: X \to A$ be a retraction. Show that r is an identification.
7. Let X be locally compact. Prove: Y is an AR(normal) if and only if Y^X is an AR(normal).
8. Let X be compact. Prove: Y is an ANR(normal) if and only if Y^X is an ANR(normal).

Section 6

1. Let A be a strong deformation retract of X. Prove that for each A and each $f: A \to Z$, the subspace Z is a deformation retract of $X \cup_f Z$.
2. Let $r: X \to A$ be such that $r \mid A \simeq 1$. Prove that if X is contractible, then A is contractible over itself.

Section 7

1. For each $n = 1, 2, \cdots$, let $C_n \subset E^k$ be the sphere $\text{Fr}\, B(0, 1/n)$. Show that $[(E^k - \{0\}) \times 0] \cup \left[\bigcup_1^\infty C_n \times I\right]$ is a strong deformation retract of $(E^k - \{0\}) \times I$.
2. Let X, Y be arbitrary spaces and attach $X \times I$ to Y by $f: X \times 0 \to Y$. Let $C = (X \times I) \cup_f Y$, and let $X = X \times 1 \subset C$. Prove: $(C \times 0) \cup (X \times I)$ is a strong deformation retract of $C \times I$.
3. Prove: $A \times 1$ is a retract of $(X \times 0) \cup (A \times I)$ if and only if there exists a surjective $r: X \to A$ such that $r \circ i \simeq 1$, where $i: A \to X$ is the inclusion map.
4. Let X be normal, Y arbitrary, $A \subset X$ closed, and $U \subset Y$ open. Let
$$h: (X \times 0) \cup (A \times I) \to U$$
be extendable to a continuous $H: X \times I \to Y$. Prove: h has an extension $G: X \times I \to U$. [*Hint:* Observe that $B = p_X[H^{-1}(Y - U)]$ is closed and that $B \cap A = \varnothing$.]

Section 8

1. Let $f: S^n \to Y$ be continuous, and let s_0 be any point of S^n. Let $\alpha: I \to Y$ be any path in Y with $\alpha(0) = f(s_0)$. Show that there is a homotopy Φ of f such that $\Phi(s_0, t) = \alpha(t)$ for each $t \in I$.
2. Let $f, g: S^n \to Y$ be homotopic maps, with $\Phi: f \simeq g$. Let $s_0 \in S^n$ be any point, and let $\beta(t) = \Phi(s_0, t)$ be the path traced by the image of s_0 during the deformation. Let $\alpha: I \to Y$ be any other path in Y such that $\alpha \simeq \beta$ rel $\text{Fr}\, I$. Show that $f \simeq g$ in such a way that the image of s_0 traces the path α during the deformation.
3. Let $f, g: V^n \to Y$ be any two maps and $\alpha: I \to Y$ be a path with $\alpha(0) = f(0)$, $\alpha(1) = g(0)$. Show that $f \simeq g$ in such a way that the origin O traces the path α during the deformation.

4. A space Y is called equiconnected if there exists a continuous
$$\lambda: Y \times Y \times I \to Y$$
such that $\lambda(a, b, 0) = a$, $\lambda(a, b, 1) = b$, and $\lambda(a, a, t) = a$ for all
$$(a, b, t) \in Y \times Y \times I.$$
The space Y is called locally equiconnected if there exists a nbd U of the diagonal $\Delta \subset Y \times Y$ and a continuous $\lambda: U \times I \to Y$ satisfying the above conditions for $(a, b, t) \in U \times I$. Prove:
   a. Y is equiconnected if and only if Δ is a strong deformation retract of $Y \times Y$.
   b. Y is locally equiconnected if and only if Δ is a strong nbd deformation retract in $Y \times Y$.

Maps into Spheres

XVI

Each map of any space into E^{n+1} is nullhomotopic. We are going to show that this is not true for maps into $S^n \subset E^{n+1}$, by proving Brouwer's theorem that the identity map $1: S^n \to S^n$ is not nullhomotopic (that is, that S^n is not contractible over itself). The general fact that different maps of a given space into S^n may not be homotopic has many important consequences, some of which will be derived in this chapter and in the next one.

1. Degree of a Map $S^n \to S^n$

There are several elementary proofs of Brouwer's theorem. The approach used here consists in obtaining for each $f: S^n \to S^n$ an integer (positive, negative, or zero) called its degree, showing that the degree is a homotopy class invariant, and finally that the identity map and a constant map have different degrees. The reader acquainted with homology theory will recognize the proof given here as a direct application (using the simple geometry of S^n) of standard techniques.

In the case $n = 1$, the degree of an $f: S^1 \to S^1$ is simply the number of times, and sense, that the image point $f(z)$ rotates around S^1 when z performs one oriented rotation of S^1. This number can be determined as follows: Orient $S^1 = \{z \mid |z| = 1\}$ so that there is a definite sense of rotation, and subdivide S^1 into arcs so small that the image under f of each arc does not contain antipodal points (say, has diameter < 1). Let

$z_1, \cdots, z_n, z_{n+1} = z_1$ be the subdivision points in order of positive rotation; from now on we work only with the points $f(z_1), \cdots, f(z_n)$. For each $i = 1, \cdots, n$ let α_i be the unique shortest arc running from $f(z_i)$ to $f(z_{i+1})$, and call α_i positive if it runs in the direction of the oriented S^1; otherwise call it negative. Finally, choose a $\zeta \in S^1$ not one of the points $f(z_i)$ and let $p(\zeta)$ be the number of positive α_i, $n(\zeta)$ the number of negative α_i, that contain ζ. The number $p(\zeta) - n(\zeta)$, which can be shown to depend only on f (that is, to be independent of ζ and of the subdivision points $z_1, \cdots, z_n$ used) is called the degree of f. In particular, for each $n = 0, \pm 1, \pm 2, \cdots$, the map $z \to z^n$ has degree n.

The definition of the degree of $f\colon S^n \to S^n$ for $n > 1$ is a straightforward generalization of the above procedure and requires only some preliminary notions from linear algebra.

A. Let E^{n+1} be referred to a fixed coordinate system. If $\{p_0, \cdots, p_{n+1}\}$ is any set of $(n + 2)$ points in E^{n+1}, its convex hull is called a geometric $(n + 1)$-simplex (*cf.* VIII, **5**) and is written $\sigma = (p_0, \cdots, p_{n+1})$. σ is *degenerate* if and only if its $(n + 2)$ vertices lie on an n-hyperplane; a necessary and sufficient condition for nondegeneracy is that $(p_0, \cdots, p_{n+1})$ have nonzero volume, so that if $(x_i^1, \cdots, x_i^{n+1})$ are the coordinates of p_i, this condition is

$$\det(p_0 \cdots p_{n+1}) = \begin{vmatrix} x_0^1, & \cdots, & x_0^{n+1}, & 1 \\ \cdots & \cdots & \cdots & \cdots \\ x_{n+1}^1, & \cdots, & x_{n+1}^{n+1}, & 1 \end{vmatrix} \neq 0.$$

An *ordered* $(n + 1)$-simplex is an $(n + 1)$-simplex together with a definite total ordering of its vertices; the simplex $\sigma = (p_0, \cdots, p_{n+1})$ with the ordering $p_0 < \cdots < p_{n+1}$ will be written $[\sigma] = [p_0, \cdots, p_{n+1}]$. The *sign* of the ordered simplex $[p_0, \cdots, p_{n+1}]$ is that of $\det(p_0, \cdots, p_{n+1})$; a degenerate ordered simplex has no sign. An even permutation of the vertices of $[\sigma]$ clearly does not change its sign.

Lemma 1 Let $[\sigma] = [p_0, p_1, \cdots, p_{n+1}]$ and $[\sigma'] = [p_0', p_1, \cdots, p_{n+1}]$ be two nondegenerate ordered $(n + 1)$-simplexes having a common face $(p_1, \cdots, p_{n+1})$, and let L be the n-hyperplane containing this face. Then p_0, p_0' are on the same side of L (that is, can be joined by a straight line segment not intersecting L) if and only if $[\sigma]$ and $[\sigma']$ have the same sign.

Proof: Since $\lambda p_0 + (1 - \lambda)p_0'$, $0 \leqslant \lambda \leqslant 1$, is the straight line joining p_0 and p_0', we need only remark that (a) $\det[\lambda p_0 + (1 - \lambda)p_0', p_1, \cdots, p_{n+1}] = \lambda \det \sigma + (1 - \lambda) \det \sigma'$, so that the values lie in the interval $[\det \sigma, \det \sigma'] \subset E^1$, and (b) $\det(q, p_1, \cdots, p_{n+1}) = 0$ if and only if $q \in L$.

B. Let $S^n \subset E^{n+1}$. If $\{p_0, \cdots, p_n\}$ is any set of $(n + 1)$ points on S^n having diameter < 1, its convex hull does not contain the origin 0, and so it can be projected from there into S^n to give the *spherical n-simplex* $\sigma = (p_0, \cdots, p_n)$; s is *degenerate* if and only if the $(n + 1)$ vertices lie on an n-hyperplane passing through the origin, that is, if and only if the $(n + 1)$-simplex $(p_0, \cdots, p_n, 0)$ in E^{n+1} is degenerate. As before, an *ordered* spherical n-simplex is a spherical n-simplex together with a definite ordering of its vertices; the *sign* of the ordered spherical n-simplex $[p_0, \cdots, p_n]$ is defined to be that of the ordered $(n + 1)$ simplex $[p_0, \cdots, p_n, 0]$ in E^{n+1}.

By a *triangulation* of S^n is meant a decomposition of S^n into finitely many nonoverlapping nondegenerate spherical n-simplexes such that each $(n - 1)$-face of an n-simplex is the common face of exactly two n-simplexes.

Let S^n, Σ^n be two n-spheres (we use different symbols to keep the concepts clear) and let T be a triangulation of S^n. A *proper vertex-map* $\varphi: T \to \Sigma^n$ is a map defined only on the vertices of T and having the following property: Whenever $p_0, \cdots, p_n$ are vertices of a simplex of T, the set $\{\varphi(p_0), \cdots, \varphi(p_n)\} \subset \Sigma^n$ has diameter < 1. It follows at once that to each simplex $\sigma \in T$, there corresponds a unique simplex $\varphi(\sigma)$ lying on Σ^n; to the *ordered* spherical simplex $[\sigma] = [p_0, \cdots, p_n]$ corresponds the *ordered* spherical n-simplex $\varphi[\sigma] = [\varphi(p_0), \cdots, \varphi(p_n)]$ on Σ^n. Though the ordering of $[\sigma]$ determines that of $\varphi[\sigma]$, it is obvious that the *sign* of $[\sigma]$ may differ from that of $\varphi[\sigma]$.

Ex. 1 Assume that φ sends each vertex to its antipode; if $[\sigma] = [p_0, \cdots, p_n]$, then $\varphi[\sigma] = [-p_0, \cdots, -p_n]$ so that

$$\det \varphi[\sigma] = \begin{vmatrix} -x_0^1, & \cdots, & -x_0^{n+1}, & 1 \\ -x_n^1, & \cdots, & -x_n^{n+1}, & 1 \\ 0, & \cdots, & 0, & 1 \end{vmatrix} = (-1)^{n+1} \det[\sigma].$$

The family of sets $\{\varphi(\sigma) \mid \sigma \in T\}$ need not form a triangulation of Σ^n, may have overlapping simplexes, and may have degenerate simplexes. However, this family has the fundamental property

Lemma 2 Let T be a triangulation of S^n, and $\varphi: T \to \Sigma^n$ a proper vertex map. Order each n-simplex of T positively, and let ξ be any point not on the boundary of any set $\varphi(\sigma)$. Let $p(\xi, T, \varphi)$ be the number of positive $\varphi[\sigma]$ and $n(\xi, T, \varphi)$ the number of negative $\varphi[\sigma]$ containing ξ. Then

$$D(\xi, T, \varphi) = p(\xi, T, \varphi) - n(\xi, T, \varphi)$$

is the same for all $\xi \in \Sigma^n$ not on the boundary of any $\varphi(\sigma)$.

Proof: Case (a). *No* $\varphi(\sigma)$ *is degenerate.* Let $\zeta \in \Sigma^n$ be any other point not on the boundary of any $\varphi(\sigma)$. Join ξ to ζ by a smooth curve on Σ^n, not passing through any face of dimension $<(n-1)$ of any $\varphi(\sigma)$, and let ξ move to ζ along this curve. Clearly, $D(\xi, T, \varphi)$ can change only when ξ crosses an $(n-1)$-face of some $\varphi(\sigma)$. We observe that each $(n-1)$-simplex $(p_1, \cdots, p_n)$ in T corresponding to this face is the common face of exactly two n-simplexes, $\sigma = (p_0, p_1, \cdots, p_n)$ and $\sigma' = (p_0', p_1, \cdots, p_n)$ in T and (more important) that $[\sigma]$ and $[\sigma']$ are of *opposite* sign because T is a triangulation and lemma 1 applies. Now let L be the hyperplane spanned by 0 and the face $(\varphi(p_1), \cdots, \varphi(p_n))$ being crossed by ξ; the argument depends on the positions of $\varphi(p_0)$ and $\varphi(p_0')$ relative to L.

(i) $\varphi(p_0)$ and $\varphi(p_0')$ are on the same side of L. Then ξ leaves (or enters) both $\varphi(\sigma)$ and $\varphi(\sigma')$ as it crosses L. According to lemma 1, $\varphi[\sigma]$ and $\varphi[\sigma']$ have the same sign, but since the sign of each $\varphi[\sigma]$ is determined by using *positive* simplexes of T, it follows from our remarks above that, in this case, ξ loses (or gains) one positive and one negative simplex, consequently $D(\xi, T, \varphi)$ is unchanged.

(ii) $\varphi(p_0)$ and $\varphi(p_0')$ are on opposite sides of L. Then ξ leaves (say) $\varphi(\sigma)$ and enters $\varphi(\sigma')$. Reasoning as in (i), it follows this time that ξ exchanges a simplex of one sign for another of the same sign, so again $D(\xi, T, \varphi)$ does not change.

We conclude in case (a) that $D(\xi, T, \varphi) = D(\zeta, T, \varphi)$, as required.

Case (b). *There are degenerate* $\varphi(\sigma)$. Fixing ξ, ζ, it is clear that we can find an $\varepsilon > 0$ and a proper vertex map $\varphi' \colon T \to \Sigma^n$ such that (1) no $\varphi'(\sigma)$ is degenerate, and $|\varphi(p) - \varphi'(p)| < \varepsilon$ for each vertex p in T, (2) whenever $\varphi(\sigma)$ is nondegenerate, $\varphi[\sigma]$ and $\varphi'[\sigma]$ have the same sign; and (3) ξ (resp. ζ) lies in the interior of $\varphi'(\sigma)$ if and only if it lies in the interior of $\varphi(\sigma)$. Thus, using case (a) for φ', we find

$$\begin{aligned} D(\xi, T, \varphi) &= D(\xi, T, \varphi') = D(\zeta, T, \varphi') \\ &= D(\zeta, T, \varphi), \end{aligned}$$

completing the proof of the lemma.

Note that the statements in the proof of case (b) of the lemma yield the more general

Lemma 3 Given any T, φ, ξ, there is an $\varepsilon > 0$ such that whenever a proper vertex map $\varphi' \colon T \to \Sigma^n$ satisfies $|\varphi(p) - \varphi'(p)| < \varepsilon$ for every vertex p, then $D(\xi, T, \varphi) = D(\xi, T, \varphi')$.

From now on, the common value $D(\xi, T, \varphi)$ will be denoted simply by $D(T, \varphi)$.

C. Let $f\colon S^n \to \Sigma^n$ be a continuous map. Since S^n is compact, we can find a triangulation T of S^n such that $\delta f(\sigma) < 1$ for each $\sigma \in T$. Replacing f by the proper vertex map $\varphi_f\colon T \to \Sigma^n$, given by setting $\varphi_f(p) = f(p)$ for each vertex p of T, we have

Lemma 4 The number $D(T, \varphi_f)$ is independent of the triangulation T of S^n (for which the associated φ_f is a proper vertex map).

Proof: Let T, T' be two triangulations, and let φ_f, φ_f' be the associated proper vertex maps. Since T, T' have a common triangulation T'', it suffices to show that both $D(\varphi_f, T)$ and $D(\varphi_f', T')$ are equal to $D(\varphi_f'', T'')$. This will follow by repetition if it is shown that introducing *one* new vertex to (say) T does not alter the value, and this is trivially true, since we may count at a point of Σ^n lying in an unaltered $\varphi(\sigma)$.

The common value $D(T, \varphi_f)$, which therefore depends only on f, is called the *degree of* f, and is written $D(f)$.

Ex. 2 Let $1\colon S^n \to S^n$ be the identity map; it is clear that $D(1) = 1$.

Ex. 3 Let $f\colon S^n \to S^n$ be a constant map; computing $D(\xi, T, \varphi_f)$ at a point ξ not in the image, we find $D(f) = 0$.

Ex. 4 Let $\alpha\colon S^n \to S^n$ be the antipodal map $\alpha(x) = -x$. It follows from Ex. 1 that $D(\alpha) = (-1)^{n+1}$.

Ex. 5 For maps $S^0 \to S^0$, a separate definition of degree is required, since the one given above cannot be used. We extend the results in the above three examples to this case and define $D(1) = 1$, $D(\alpha) = -1$, and the other two maps to have degree 0.

We now show that $D(f)$ is a homotopy class invariant.

1.1 Theorem Let $n \geqslant 0$. If $f, g\colon S^n \to \Sigma^n$ are homotopic, then $D(f) = D(g)$.

Proof: For $n = 0$, the theorem is evident. We thus assume $n \geqslant 1$. Let $\Phi\colon S^n \times I \to \Sigma^n$ be a homotopy of f and write $\Phi(x, t) = f_t(x)$. Since $S^n \times I$ is compact, the map Φ is uniformly continuous, and therefore there is a $\delta > 0$ such that $d(x, x') < \delta \Rightarrow |f_t(x) - f_t(x')| < 1$ for every $t \in I$. Thus there is a single triangulation T of S^n such that $\delta f_t(\sigma) < 1$ for each $\sigma \in T$ and $t \in I$; from now on we work with T. Given t_0 and fixing any $\xi \in \Sigma^n$, lemma 3 gives an $\varepsilon > 0$ for which any ε-variation of the vertices $\{\varphi_{f_{t_0}}(p)\}$ does not change $D(\xi, T, \varphi_{f_{t_0}})$. By uniform continuity there is a $\delta(\varepsilon) > 0$ such that

$$|t - t_0| < \delta \Rightarrow |f_t(x) - f_{t_0}(x)| < \varepsilon$$

for every $x \in S^n$, and consequently we have $D(f_t) = D(f_{t_0})$ for all $|t - t_0| < \delta$. This says that the integer-valued function $D(f_t)$ is continuous at each point $t_0 \in I$, and therefore $D(f_t)$ is constant on I. The theorem has been proved.

The converse of this theorem will be considered later in **7**.

1.2 Remark A map $f: (V^{n+1}; S^n) \to (V^{n+1}; S^n)$ is called a *regular* map of V^{n+1}. Since V^{n+1} has triangulations T into $(n+1)$-simplexes such that each n-face not on S^n of an $(n+1)$-simplex of T is the face of exactly two $(n+1)$-simplexes, we can define "degree" for regular maps in a manner analogous to that for maps $S^n \to S^n$. Indeed, given T, call a vertex map $\varphi: T \to V^{n+1}$ *regular* if (1) each vertex of S^n maps to a point on S^n and (2) $\varphi \mid S^n$ is a proper vertex map. Then the degree $D_r(f)$ of a regular map f is determined by choosing an associated regular vertex map $\varphi_f: T \to V^{n+1}$, a $\xi \in V^{n+1} - S^n$ not on the boundary of any $\varphi_f(\sigma)$, and letting $D_r(f) =$ (number of positive $\varphi_f[\sigma]$ containing ξ) $-$ (number of negative $\varphi_f[\sigma]$ containing ξ) the sign of $\varphi[\sigma]$ being determined by taking $[\sigma]$ positive. The arguments in lemmas 1–4 and in **1.1** apply *verbatim* to show that $D_r(f)$ depends only on f and that two regular maps have the same degree whenever they are homotopic in such a way that the image of S^n remains on S^n during the entire deformation. A useful consequence is

1.3 Let $g: (V^{n+1}; S^n) \to (V^{n+1}; S^n)$ be a regular map, and let $f = g \mid S^n: S^n \to S^n$. Then $D(f) = D_r(g)$.

Proof: We first extend f to another regular map F of V^{n+1} by sending each radius $\overrightarrow{0x}$ of V^{n+1} linearly onto the radius $\overrightarrow{0f(x)}$. It is obvious that $D(f) = D_r(F)$, by taking a triangulation of V^{n+1} using only simplexes of form $(p_0, \cdots, p_n, 0)$. Furthermore, F is homotopic to g in such a way that the image of S^n remains in fact fixed during the entire deformation, as the homotopy $\Phi: V^{n+1} \times I \to V^{n+1}$ given by $(v, t) \to tF(v) + (1 - t)g(v)$ shows. Thus $D_r(F) = D_r(g)$, and the proof is complete.

We note that **1.3** holds also for $n = 0$ if the definitions in Ex. 5 are used.

2. Brouwer's Theorem

Brouwer's theorem plays a fundamental role in the topology of E^n, as we shall see in the next chapter.

2.1 Theorem (L. E. J. Brouwer) The identity map $1: S^n \to S^n$ is not nullhomotopic.

Proof: We have seen in **1**, Exs. 2 and 3, that $D(1) = 1$, and $D(0) = 0$; by **1.1**, the identity map cannot be nullhomotopic.

2.2 Corollary Brouwer's theorem is equivalent to each of the following two statements:

(1). There exists *no* continuous map $F\colon V^{n+1} \to S^n$ keeping the boundary points fixed (that is, S^n is not a retract of V^{n+1}).

(2). (Brouwer's fixed-point theorem) Every continuous map $f\colon V^{n+1} \to V^{n+1}$ has a fixed point.

Proof: (**2.1**) $\Rightarrow$ (1). Assume a retraction $F\colon V^{n+1} \to S^n$ did exist. By XV, **1.2**, this would imply that the identity map $F \mid S^n\colon S^n \to S^n$ is nullhomotopic, contradicting **2.1**.

(1) $\Rightarrow$ (2). Assume that there were some $f\colon V^{n+1} \to V^{n+1}$ such that $f(x) \neq x$ for each $x \in V^{n+1}$. The map $F\colon V^{n+1} \to S^n$, defined by

$$F(x) = \text{point of } S^n \text{ that lies on the directed ray } \overrightarrow{f(x)x},$$

would evidently be a (continuous) retraction $V^{n+1} \to S^n$.

(2) $\Rightarrow$ (**2.1**). Assume the identity map were nullhomotopic. By XV, **1.2**, the map would extend to an $F\colon V^{n+1} \to S^n$, and then $x \to -F(x)$ would be a map $V^{n+1} \to V^{n+1}$ with no fixed point.

Clearlv, Brouwer's fixed-point theorem is valid in any space homeomorphic to V^{n+1}; furthermore, we can extend **2.2**(1) by

2.3 If $U \subset E^{n+1}$ is any bounded open set, then $\text{Fr}(U)$ is not a retract of $\overline{U}$.

Proof: Assume that there were a retraction $r\colon \overline{U} \to \text{Fr}(U)$. We can assume that $0 \in U$, and we can find a ball $B(0, N) \supset \overline{U}$. Define

$$f\colon \overline{B(0, N)} \to \overline{B(0, N)}$$

by

$$f(x) = \begin{cases} Nr(x)/|r(x)| & x \in \overline{U} \\ Nx/|x| & x \in \overline{B(0, N)} - U. \end{cases}$$

According to III, **9.4**, f is continuous: the intersection of the two closed sets on which f is defined is $\text{Fr}(U)$, and the two definitions agree on $\text{Fr}(U)$. But we now have a contradiction to **2.2**(1), since f is a retraction of $\overline{B(0, N)}$ onto $\text{Fr}[\overline{B(0, N)}]$.

3. Further Applications of the Degree of a Map

We give here only two of the most immediate applications.

(a) The following contains the fundamental theorem of algebra:

3.1 Let f be a continuous complex-valued function defined on the finite complex plane. Assume that $\lim_{z \to \infty} \frac{f(z)}{z^n} = c \neq 0$ for some nonzero positive or negative integer n. Then the equation $f(z) = 0$ has at least one solution.

Proof: We can evidently assume that $c = 1$; with this modification, the hypothesis is that $f(z) = z^n(1 + \eta(z))$, where $\eta(z) \to 0$ as $z \to \infty$. We argue by contradiction, so assume that $f(z) \neq 0$ for all finite z. Let t be a real parameter, $t \geqslant 0$, and let ζ vary only on the unit circle $|z| = 1$. Then, for each $t \geqslant 0$,

$$F_t(\zeta) = \frac{f(t\zeta)}{|f(t\zeta)|}$$

is a continuous map $S^1 \to S^1$, and since any two of these maps are obviously homotopic, they all have the same degree. Because F_0 is a constant map, we must therefore have $D(F_t) = 0$ for all $t \geqslant 0$. We now show that this is impossible. For, since $\eta(\zeta t) \to 0$ as $t \to \infty$, there is a t_0 such that

$$\left| \frac{1 + \eta(\zeta t_0)}{|1 + \eta(\zeta t_0)|} - 1 \right| < 2 \qquad \text{for all } \zeta;$$

thus $|F_{t_0}(\zeta) - \zeta^n| < 2$ for all ζ; consequently (XV, **1.2**) F_{t_0} is homotopic to the map $\zeta \to \zeta^n$. Since the latter has degree $n \neq 0$, **1.1** gives the desired contradiction.

(b) Given a subset $A \subset E^{n+1}$ and $(n + 1)$ continuous real-valued functions $\varphi_1, \cdots, \varphi_{n+1}$ defined on A, this situation can be interpreted in two ways:

(i). As a map $\varphi: A \to E^{n+1}$, where $\varphi(a) = (\varphi_1(a), \cdots, \varphi_{n+1}(a))$.

(ii). As a continuous vector field $\mathbf{\Phi}$ on A, where at each $a \in A$ we have the vector $\mathbf{\Phi}(a)$ with components $(\varphi_1(a), \cdots, \varphi_{n+1}(a))$.

These two viewpoints are obviously equivalent; we call φ the map associated with the vector field $\mathbf{\Phi}$.

A continuous vector field is *nonvanishing* if it has no zero vector. With each continuous nonvanishing vector field $\mathbf{\Phi}$ on A, we can construct a map $c_\varphi: A \to S^n$ by $a \to \varphi(a)/|\varphi(a)|$ which involves only the directions and not the lengths of the vectors in $\mathbf{\Phi}$. In case $A = S^n$, the degree of c_φ is called the *characteristic* of the vector field $\mathbf{\Phi}$.

Ex. 1 If $\mathbf{\Phi}$ is a continuous nonvanishing vector field on V^{n+1}, then $\mathbf{\Phi} \mid S^n$ has characteristic 0. For, $c_\varphi: S^n \to S^n$ is extendable to a $\tilde{c}: V^{n+1} \to S^n$.

Ex. 2 The field of outward-drawn normals on S^n has characteristic 1; by **1**, Ex. 4, the field of inward-drawn normals on S^n has characteristic $(-1)^{n+1}$, since c_φ is the antipodal map.

Ex. 3 If $\mathbf{\Phi}$, $\mathbf{\Psi}$ are two continuous nonvanishing vector fields on S^n such that $\mathbf{\Phi}(x)$ and $\mathbf{\Psi}(x)$ are *not* opposed for each $x \in S^n$, then $\mathbf{\Phi}$ and $\mathbf{\Psi}$ have the same characteristic. For, $c_\varphi(x)$ and $c_\psi(x)$ are never antipodal (*cf.* XV, **1.2**).

We now derive some of the easier facts about vector fields on balls and spheres.

3.2 Each nonvanishing continuous vector field on V^{n+1} must contain, on S^n, at least one outward- and one inward-drawn normal.

Proof: By Ex. 1, $\Phi \mid S^n$ has characteristic 0; by Ex. 2, $\Phi \mid S^n$ and the field of inward (outward) normals have differing characteristics; by Ex. 3, Φ must contain an outward and an inward normal.

3.3 Theorem (H. Poincaré and L. E. J. Brouwer) Every continuous nonvanishing vector field on an *even-dimensional* S^{2n} must contain at least one normal vector. In particular, there can be no continuous nonvanishing vector field of tangential directions on any S^{2n}.

Proof: If n is even, Ex. 2 shows that the inward and outward normal fields have different characteristics. Since any vector field must therefore have characteristics differing from at least one of these two fields, the result follows from Ex. 3.

3.4 Corollary Each $f: S^{2n} \to S^{2n}$ either has a fixed point or sends a point to its antipode.

Proof: If $f: S^{2n} \to S^{2n}$ has no fixed point, the vectors $\overrightarrow{xf(x)}$ form a continuous nonvanishing vector field and must therefore contain a normal vector.

Ex. 4 Every odd-dimensional sphere has a nonvanishing continuous tangent vector field: for each $x = (x_1, \cdots, x_{2n}) \in S^{2n-1}$ let $\Phi(x)$ be the vector with components $(-x_{n+1}, \cdots, -x_{2n}, x_1, \cdots, x_n)$. Observe that the map associated with this vector field is a mapping $S^{2n-1} \to S^{2n-1}$ that has no fixed point and does not send any point to its antipode.

4. Maps of Spheres into S^n

So far we have discussed maps of S^n into a sphere S^n of the same dimension; for $n \geq 1$, it is easy to see (*cf.* end of **7**) that there are infinitely many homotopy classes. We now consider $[S^k, S^n]$ for $k \neq n$.

In case $k > n > 1$, the number of homotopy classes of maps $S^k \to S^n$ is in general unknown except in relatively few cases; their determinations require methods beyond the scope of this book. To cite some examples that indicate the complexity of the problem even for $n = 2$: it is known that there are exactly two homotopy classes of maps $S^4 \to S^2$, $S^5 \to S^2$, and $S^7 \to S^2$, but that there are twelve classes of maps $S^6 \to S^2$ and infinitely many classes of maps $S^3 \to S^2$; on the other hand, for each $k > 1$, all maps $S^k \to S^1$ are nullhomotopic.

The situation for $k < n$ is simple: Each map $S^k \to S^n$ is nullhomotopic. To prove this, we will first establish the general fact that a continuous $f: S^k \to S^n$ (k, n arbitrary!) is always homotopic to a "piecewise" linear map.

Let $f: S^k \to S^n$ be given, and let T be a triangulation of S^k such that $\delta f(\sigma) < 1$ for each k-simplex $\sigma \in T$. Let $\varphi_f: T \to S^n$ be the associated proper vertex map; we extend φ_f to a map $\lambda_f: S^k \to S^n$ by mapping each spherical k-simplex $\sigma \in T$ barycentrically (VIII, **5**) onto $\varphi_f(\sigma)$.

λ_f is continuous, since it is continuous on each σ and its definitions agree on each $(k - 1)$-face common to two k-simplexes (*cf.* III, **9.4**); λ_f is called the linear T-approximation of f. More generally, any continuous map $\lambda: S^k \to S^n$ that maps each simplex of a triangulation of S^k barycentrically is called a piecewise linear map.

4.1 Let $f: S^k \to S^n$ be given. Then $f \simeq \lambda_f$ for each linear T-approximation λ_f of f.

Proof: We need show only (XV, **1.2**) that $\lambda_f(x)$ and $f(x)$ are never antipodal. Let $x \in \sigma = (p_0, \cdots, p_k)$, where $\sigma \in T$. Since $f(p_i) = \lambda_f(p_i)$ for $i = 0, \cdots, k$ and $\delta f(\sigma) < 1$, it follows that each $\lambda_f(p_i)$ is contained in the ball $B(\lambda_f(p_0), 1)$. Because the ball is convex, it must contain the convex hull of the $\lambda_f(p_i)$ and consequently also $\lambda_f(\sigma)$. This shows that $d(\lambda_f(x), \lambda_f(p_0)) < 1$, and since

$$d(f(x), \lambda_f(x)) \leqslant d(f(x), f(p_0)) + d(\lambda_f(p_0), \lambda_f(x)) < 2,$$

the points $f(x)$, $\lambda_f(x)$ are not antipodal.

4.2 Theorem If $k < n$, then each $f: S^k \to S^n$ is nullhomotopic.

Proof: Let $\lambda_f: S^k \to S^n$ be a linear T-approximation to f. Since λ_f is piecewise linear and $k < n$, it follows that $\lambda_f(S^k)$ lies on finitely many great $S^{n-1} \subset S^n$; thus $\lambda_f(S^k) \neq S^n$, and so (XV, **1.2**) $\lambda_f \simeq 0$.

Ex. 1 Observe that an arbitrary continuous $f: S^k \to S^n$ may be surjective, even if $k < n$, as the existence of Peano curves shows. The role of λ_f is to show that whenever $k < n$, the map f can always be deformed to uncover a point of S^n.

Ex. 2 Let $n > 1$ and $\alpha, \beta: I \to S^n$ be two paths, each starting at $p_0 \in S^n$ and ending at $p_1 \in S^n$. Then $\alpha \simeq \beta$ rel Fr(I). Indeed, define φ: Fr$(I^2) \to S^n$ by $\varphi(t, 0) = \alpha(t)$, $\varphi(t, 1) = \beta(t)$ $\varphi(0, s) = p_0$ $\varphi(1, s) = p_1$; the map φ is evidently continuous. Since $S^1 \cong$ Fr(I^2) we can regard φ as a map $S^1 \to S^n$, so that $\varphi \simeq 0$; φ is therefore extendable to a $\Phi: I^2 \to S^n$, which is the required homotopy.

We have seen (VII, **5.3**) that if X is normal and $A \subset X$ is closed, a continuous $f: A \to S^n$ can always be extended over a nbd $U \supset A$, and **2.2**(1) furnishes an instance in which f is not extendable over X itself. We have two important cases where more information about the nature of the extension is available.

4.3 Lemma (1). Let $A \subset S^n$ be closed. Then any continuous $f: A \to S^n$ can be extended to a continuous $F: S^n \to S^n$.

(2). Let $A \subset S^{n+1}$ be closed and $f: A \to S^n$ be continuous. Removing exactly one arbitrarily chosen point p_ι from each component U_ι of $\mathscr{C}A$, there is an extension

$$F: S^{n+1} - \bigcup_\iota \{p_\iota\} \to S^n.$$

Proof: (1). Let $U \supset A$ be a nbd over which $f\colon A \to S^n$ can be extended; since A is compact, $\alpha = d(A, \mathscr{C}U)$ is positive. Triangulate S^n so that the diameter of each n-simplex is $< \alpha/2$; because any n-simplex intersecting A lies completely in U, letting Q be the union of all such n-simplexes, we have an extension $\bar{f}\colon Q \to S^n$ of f. Map each vertex of a simplex not in Q arbitrarily into S^n, and then map each edge of such a simplex to an arc joining the images of its end points. Letting T^k denote the union of all k-dimensional simplexes of the triangulation, we have an extension $f^1\colon Q \cup T^1 \to S^n$ of f. We proceed inductively and show that whenever $k < n$, an $f^k\colon Q \cup T^k \to S^n$ can be extended to an $f^{k+1}\colon Q \cup T^{k+1} \to S^n$. Indeed, given any $(k+1)$-simplex σ^{k+1}, f^k is defined on its boundary and $f^k \mid \mathrm{Fr}(\sigma^{k+1})\colon \mathrm{Fr}(\sigma^{k+1}) \to S^n$; since $\mathrm{Fr}(\sigma^{k+1})$ is homeomorphic to S^k, and $k < n$, it follows from **4.2** and XV, **1.2**, that $f^k \mid \mathrm{Fr}(\sigma^{k+1})$ is extendable over σ^{k+1}, and using such an extension for each $\sigma^{k+1} \in T^{k+1}$ gives an $f^{k+1}\colon Q \cup T^{k+1} \to S^n$. Since $Q \cup T^n = S^n$, the map f^n is the desired extension of f.

(2). By using a triangulation of S^{n+1}, we arrive as in (1) at an extension $f^n\colon Q \cup T^n \to S^n$, where $Q \supset A$. In each $(n+1)$-simplex σ not belonging to Q, choose a point p_σ, let $\pi_\sigma\colon (\sigma - p_\sigma) \to \mathrm{Fr}(\sigma)$ be the radial projection from p_σ, and define $F(x) = f^n \circ \pi_\sigma(x)$ whenever $x \in \sigma - p_\sigma$. This gives an extension $F\colon S^{n+1} - \bigcup_\sigma \{p_\sigma\} \to S^n$ of f, where exactly one point has been removed from each $(n+1)$-simplex not in Q.

We now show that all the singularities $p_1, \cdots, p_m$ of F that lie in any one component U of $\mathscr{C}A$ can be condensed into any selected point $q \in U$. Since (V, **4.2**, and V, **4**, Ex. 1) U is open, it is (V, **5.6**) path-connected, so that there is a path α in U running from p_1 to q and going through all the singularities $p_1, \cdots, p_m$. We move along this path and replace all the singularities by one at q as follows: Cover $\alpha(I)$ by finitely many balls $V_1, \cdots, V_s$ such that all $\bar{V}_i \subset U$, $p_1 \in V_1$, $q \in V_s$, $V_i \cap V_{i+1} \neq \varnothing$, and no singularity p_i lies on $\bigcup_1^s \mathrm{Fr}(V_k)$. Choose an $x_1 \in V_1 \cap V_2$; it suffices to show how to replace all the singularities in V_1 by a single one at x_1: discard $F|(V_1 - \{p_i | p_i \in V_1\})$ and replace it with the map obtained by projecting onto $\mathrm{Fr}(V_1)$ from x_1 (that is, if π is the projection onto $\mathrm{Fr}(V_1)$ from x_1, then $F|(V_1 - \{p_i | p_i \in V_1\})$ is replaced by the map

$$[F|\mathrm{Fr}(V_1)] \circ \pi\colon (V_1 - x_1) \to S^n.$$

Ex. 3 Observe that a result stronger than that stated in (2) has been obtained: even though $\mathscr{C}A$ may have infinitely many components, we need remove points from only *finitely* many to secure the extension F (because the triangulation of S^{n+1} has but finitely many $(n+1)$-simplexes).

The result **4.3** is the key to the deeper topological properties of E^n, as will be seen in the next chapter. Its formulation for maps of subsets of

Euclidean spaces, rather than of spheres, will be more convenient for our later purposes. To obtain the required version, we first observe that, if $A \subset E^n$ is compact, then $\mathscr{C}A$ has exactly one unbounded component: For, A is necessarily bounded, so there is some $\overline{B(0; r)} \supset A$, and because

$$\mathscr{C}\overline{B(0; r)} \subset \mathscr{C}A$$

is connected, it lies in exactly one component of $\mathscr{C}A$. Using this and the standard equivalence (under stereographic projection) between compact sets in the plane and closed sets on S^n that do not contain the north pole, we have

4.4 Theorem (1). Let $A \subset E^n$ be compact. Then any continuous $f: A \to S^n$ can be extended to an $F: E^n \to S^n$.

(2). Let $A \subset E^{n+1}$ be compact and $f: A \to S^n$ be continuous. By removing exactly one point p_ι from each *bounded* component U_ι of $\mathscr{C}A$, there is an extension $F: E^{n+1} - \bigcup_\iota \{p_\iota\} \to S^n$.

Proof: (1). Let $\mu: (S^n - p_+) \to E^n$ (p_+ = the north pole) be the stereographic projection. The map $f \circ \mu: \mu^{-1}(A) \to S^n$ has an extension F over S^n and $F \circ \mu^{-1}$ is the required extension of f. As for (2), we place at exactly p_+ the singularity of the extension F that lies in the component of $\mathscr{C}\mu^{-1}(A)$ that contains the north pole p_+.

5. Maps of Spaces into S^n

Let X be normal and $A \subset X$ be closed. We have seen in XV, **7.1**, **7.3**, that if A is both a G_δ and a strong nbd deformation retract in X, then for *every* space Z, whenever one map $f: A \to Z$ is extendable over X, so also is each $g: A \to Z$ homotopic to f. In the special case that $Z = S^n$ [or, more generally, that Z is an ANR for normal spaces (VII, **5**)] this result is true *without* such additional restrictions on A.

5.1 Theorem (K. Borsuk) Let X be a normal space such that $X \times I$ is also normal. Let $A \subset X$ be closed and $f_0, f_1: A \to S^n$ be homotopic. If f_0 has an extension $F_0: X \to S^n$, then so also does f_1; in fact, an extension F_1 can be chosen so that $F_1 \simeq F_0$.

Proof: Let $\varphi: f_0 \simeq f_1$ and define $\Phi: X \times 0 \cup A \times I \to S^n$ by

$$\Phi(x, 0) = F_0(x),$$
$$\Phi(a, t) = \varphi(a, t).$$

Φ is continuous, and the problem is to extend Φ over $X \times I$. Since $X \times I$ is normal, VII, **5.3**, assures that Φ has an extension $\bar{\Phi}$ over some nbd $U \supset (X \times 0) \cup (A \times I)$, and because I is compact, there is (XI, **2.6**) a nbd $V \supset A$ such that $V \times I \subset U$. Now find a continuous $p: X \to I$ such that $p(A) = 1$, $p(X - V) = 0$; then $(x, tp(x)) \in X \times 0 \cup V \times I$ for all $(x, t) \in X \times I$, so $\Psi(x, t) = \overline{\Phi}(x, t \cdot p(x))$ is an extension of Φ over $X \times I$ as required.

This leads to the following addition to **4.4**.

5.2 Corollary Let $A \subset E^k$ be closed (k arbitrary) and $f: A \to S^n$. Then f is extendable over E^k if and only if f is nullhomotopic.

Proof: Assume that $f \simeq 0$; then because 0 is extendable over E^k, Borsuk's theorem assures that f is also. Conversely, if f is extendable to an $F: E^k \to S^n$, then (XV, **1**, Ex. 2) F is nullhomotopic, and consequently (XV, **2.2**), so also is $f = F \mid A$.

6. Borsuk's Antipodal Theorem

In this and the next section, we derive some special properties of maps $S^n \to S^n$. Although we will not use them in our work on the topology of E^n, they are of fundamental importance for any such deeper study.

A map $f: S^n \to S^n$ is called antipodal-preserving (or simply antipodal) if $f(-x) = -f(x)$ for each $x \in S^n$; that is, f sends each pair of antipodal points to a pair of antipodal points. For example, the identity map and the map $\alpha: S^n \to S^n$, where $\alpha(x) = -x$, are both antipodal-preserving maps; in fact, $f: S^n \to S^n$ is antipodal if and only if $f \circ \alpha = \alpha \circ f$.

6.1 Theorem (K. Borsuk) Let $n \geqslant 0$, and let $f: S^n \to S^n$ be an antipodal map. Then $D(f)$ is odd; in particular, f is not nullhomotopic.

We will use repeatedly the trivial observation that if f is a barycentric map of the n-simplex σ into E^n, and if $\operatorname{Int} f(\sigma) \neq \varnothing$, then for each $y \in f(\sigma)$ there is exactly one point $x \in \sigma$ such that $f(x) = y$. To prove **6.1**, we need the

Lemma Let $n \geqslant 1$. Let T be a triangulation of V^n and let $g: V^n \to V^n$ be any linear T-map such that 0 does not lie on the boundary of any $g(\sigma)$. Let $\pi: V^n - \{0\} \to S^{n-1}$ be the radial projection, and $\varphi = \pi \circ (g \mid S^{n-1}): S^{n-1} \to S^{n-1}$. Then, if q is the number of points mapped by g to 0, we have $q \equiv D(\varphi) \bmod 2$.

Proof of Lemma: Define a regular $h: V^n \to V^n$ as follows:

$$\begin{aligned} h(tx) &= g(2tx) & (x \in S^{n-1}, \quad 0 \leqslant t \leqslant \tfrac{1}{2}) \\ &= (2 - 2t)g(x) + (2t - 1)\pi \circ g(x) & (x \in S^{n-1}, \quad \tfrac{1}{2} \leqslant t \leqslant 1) \end{aligned}$$

According to **1.3**, it suffices to show that $D_r(h) \equiv q \bmod 2$.

Let $A = V^n - B(0, \frac{1}{2})$; the compact $h(A)$ does not contain the origin, so letting $\varepsilon = d(h(A), 0) > 0$, we can obtain a triangulation $\bar{T}$ of V^n such that:

(1). $\delta h(\sigma) < \varepsilon/4$ for each $\sigma \in \bar{T}$.

(2). $\bar{T}$ is a refinement of T on $\overline{B(0; \frac{1}{2})}$.

(3). If $h(x) = 0$, then x is not on the boundary of any n-simplex of $\bar{T}$.

Let λ be a linear $\bar{T}$ approximation of h. Then

(a). $\lambda(\sigma)$ does not contain 0 whenever $\sigma \subset A$.

Indeed, all the vertices of σ lie in a ball B of radius $\varepsilon/4$ centered at one of them. By choice of ε, B does not contain the origin and, being convex, does contain $\lambda(\sigma)$.

Furthermore, because g is T-linear, (2) implies that

(b). $$\lambda|\overline{B(0, \tfrac{1}{2})} = g,$$

and using (3), it follows that the origin does not lie on the boundary of any $\lambda(\sigma)$. We can therefore calculate the parity of $D_r(h)$ by counting the $\lambda(\sigma)$ containing 0 and, by (a), (b) this number is exactly q.

Proof of Theorem: We proceed by induction. For $n = 0$, the theorem is obvious by **I**, Ex. 5. We now assume the theorem true for $n - 1$, and prove it for n.

It is easy to find a triangulation T of S^n such that:

(i) T is mapped onto itself by the antipodal map $\alpha(x) = -x$; that is, $\alpha(\sigma) \in T$ for each $\sigma \in T$.

(ii) T contains a triangulation of the equatorial

$$S^{n-1} = \{x \in S^n \mid x_{n+1} = 0\}.$$

(iii) $\delta f(\sigma) < 1$ for each $\sigma \in T$.

Let λ be a linear T-approximation of f; (i) assures that λ is also an antipodal-preserving map. We can assume also that:

(iv) The north pole p_+ and the south pole p_- do not lie on the boundary of any $\lambda(\sigma)$; in particular, $\lambda(S^{n-1})$ does not contain either p_+ or p_-.

Indeed, we can alter λ itself by changing slightly its values on antipodal vertices to bring about the first requirement; the second statement then follows from (ii).

Any small deformation required for (iv) does not change the homotopy class of λ (XV, **1.2**). Since $\lambda \simeq f$, it suffices to calculate $D(\lambda)$, and because λ is piecewise linear, (iv) shows that this can be done by counting the number of $\lambda(\sigma)$ containing p_+. Thus the theorem will be proved if we can show that there are an odd number of points mapped by λ into p_+.

Let L^n be the n-ball bounded by the equatorial S^{n-1}, let $P: S^n \to L^n$ be the projection parallel to the x_{n+1}-axis, and let S^n_+ be the northern hemisphere $\{x \in S^n \mid x_{n+1} \geqslant 0\}$. Considering the piecewise linear map $g = P \circ (\lambda \mid S^n_+): S^n_+ \to L^n$, note that S^n_+ is homeomorphic to V^n, and that 0 is not on the boundary of any $P \circ \lambda(\sigma)$; because of (iv), we conclude that

$$\varphi = \pi \circ P \circ (\lambda \mid S^{n-1}): S^{n-1} \to S^{n-1}$$

is a well-defined continuous map. Since φ is clearly antipodal-preserving, the induction hypothesis asserts that φ has odd degree, and by the lemma, we find that g maps an odd number of points in S^n_+ to 0.

We now note $g(x) = 0$ if and only if $\lambda(x) = p_+$ or p_-, that is,

$$\{x \in S^n_+ \mid g(x) = 0\} = \{x \in S^n_+ \mid \lambda(x) = p_+\} \cup \{x \in S^n_+ \mid \lambda(x) = p_-\}.$$

Since λ is antipodal,

$$[x \in S^n_+] \wedge [\lambda(x) = p_-] \Leftrightarrow [-x \in S^n_-] \wedge [\lambda(-x) = p_+],$$

so because of (iv), the number of points $x \in S^n$ such that $\lambda(x) = p_+$ is odd, and the inductive step has been completed.

6.2 Corollary Borsuk's antipodal theorem implies each of the following three equivalent statements:

(1). There is no antipodal map $f: S^n \to S^{n-1}$.

(2). Each continuous $f: S^n \to E^n$ (that is, a "flattening") sends at least one pair of antipodal points to the same point.

(3). (**L. Lusternik** and **L. Schnirelmann**) In each family of $(n + 1)$ closed sets covering S^n, at least one set must contain a pair of antipodal points.

Proof: (**6.1**) $\Rightarrow$ (1). Assume that there were an antipodal

$$f: S^n \to S^{n-1};$$

regarding S^{n-1} as the equator of S^n, we would have a nonsurjective antipodal-preserving $f: S^n \to S^n$; by XV, **1.2**, f would be nullhomotopic, contradicting **6.1**.

(1) $\Rightarrow$ (2). Assume that there were a $g: S^n \to E^n$ such that

$$g(-x) \neq g(x)$$

for each $x \in S^n$. Define $f: S^n \to S^{n-1}$ by

$$f(x) = \frac{g(-x) - g(x)}{|g(-x) - g(x)|}.$$

Then f would be an antipodal map, contradicting (1).

(2) $\Rightarrow$ (3). Let $F_1, \cdots, F_{n+1}$ be $(n + 1)$ closed sets covering S^n, let $\alpha: S^n \to S^n$ be the map $\alpha(x) = -x$, and assume that $\alpha(F_i) \cap F_i = \varnothing$ for $i = 1, \cdots, n$. Since the F_i and the $\alpha(F_i)$ are closed G_δ sets, for each $i = 1, \cdots, n$ there is a continuous $g_i: S^n \to I$ such that $g_i^{-1}(0) = F_i$, $g_i^{-1}(1) = \alpha(F_i)$. Define $g: S^n \to E^n$ by $g(x) = \{g_1(x), \cdots, g_n(x)\}$; by (2) there is an $x_0 \in S^n$ such that $g_i(x_0) = g_i(-x_0)$ for each $i = 1, \cdots, n$, so that $x_0 \mathrel{\overline{\in}} \bigcup_1^n F_i$ and $x_0 \mathrel{\overline{\in}} \bigcup_1^n \alpha(F_i)$. Since $\bigcup_1^{n+1} F_i = \bigcup_1^{n+1} \alpha(F_i) = S^n$, we conclude that $x_0 \in F_{n+1} \cap \alpha(F_{n+1})$, so that x_0 and its antipode belong to F_{n+1}.

(3) $\Rightarrow$ (1). Let $f: S^n \to S^{n-1}$ be any continuous map; we show that it cannot be antipodal. Decompose S^{n-1} into $(n + 1)$ closed sets $A_1, \cdots, A_{n+1}$, each of which has diameter <2; this is possible by, say, projecting the boundary of an n-simplex enclosing the origin onto S^{n-1}. Defining $F_i = f^{-1}(A_i)$, $i = 1, \cdots, n + 1$, there is according to (3) an $x_0 \in S^n$ and an index k such that $x_0 \in F_k \cap \alpha(F_k)$. Thus $f(x_0)$ and $f(-x_0)$ both belong to A_k and so f cannot be antipodal.

We have seen that an easy connectedness argument was sufficient to show that E^1 is not homeomorphic to E^n for any $n \neq 1$, and we have indicated that the general result is deeper.

6.3 Theorem E^n is not homeomorphic to E^m whenever $m \neq n$.

Proof: Let $n > m$ and let $h: E^n \to E^m$ be continuous; since $n - 1 \geqslant m$, we know that $h \mid S^{n-1}: S^{n-1} \to E^m \subset E^{n-1}$ must send an antipodal pair of points to the same point, so that h cannot be bijective.

Another proof based directly on **2.1** will be given in the next chapter.

7. Degree and Homotopy

Homotopic maps have the same degree; the purpose of this section is to establish the converse. The proof given here is based on special cases (**7.2**, **7.3**) of general theorems due to H. Freudenthal.

In this work the following normalization is convenient:

7.1 Let $A \subset S^n$ be a closed proper subset. Then there is a homeomorphism $\beta: S^n \to S^n$ such that $\beta \simeq 1$ and $\beta(A)$ is contained in the interior of $S_+^n = \{x \in S^n \mid x_{n+1} \geqslant 0\}$. Furthermore, $\beta^{-1} \simeq 1$ also.

Proof: We can clearly assume that the south pole p_- is not in A. Let S be a spherical nbd of p_- lying in $\mathscr{C}A$. For each $x \in \operatorname{Fr}(S)$, let p_+xp_- be the arc of great circle from p_+ to p_- through x, and let s_x be its intersection with the equatorial S^{n-1}. Define β by mapping each p_+x linearly onto p_+s_x and xp_- linearly onto s_xp_-; clearly, β satisfies the requirements.

7.2 Lemma Let $n \geqslant 2$, and $f\colon S^n \to \Sigma^n$. Then there exists a $g \simeq f$ and a $\delta \in \Sigma^n$ such that $g^{-1}(\delta)$ is either empty or a single point.

Proof: We can assume that f is piecewise linear. Choosing a $\delta \in \Sigma^n$ not on the boundary of any set $f(\sigma)$, it follows that $f^{-1}(\delta)$ consists of finitely many points $p = p_1, p_2, \cdots, p_k$; by **7.1**, we can assume that all p_i are in the interior of S^n_+ (otherwise, replace f by $f \circ \beta^{-1}$). Regarding S^n_+ as a ball in E^n, let $L \subset S^n_+$ be the set consisting of the $(k-1)$ line segments joining $p = p_1$ to each $p_2, \cdots, p_k$. Since $n \geqslant 2$, the compact set $f(L) \neq \Sigma^n$, so there is an ε-nbd $U \supset L$, $U \subset S^n_+$ such that $f(\bar{U}) \neq \Sigma^n$ also; replacing f by $\beta \circ f$ if necessary, we can assume that

$$f\colon (S^n; \bar{U}) \to (\Sigma^n; \Sigma^n_+).$$

We first deform f to a map φ such that $\varphi^{-1}(\delta) = L$. Let $\mu\colon S^n \to I$ be the function $\mu(x) = \varepsilon^{-1}\min[d(x, L), \varepsilon]$ and define

$$\varphi(x) = \frac{\mu(x)f(x) + (1 - \mu(x))\delta}{|\mu(x)f(x) + (1 - \mu(x))\delta|}.$$

Then φ is continuous and homotopic to f, since $f(x)$, $\varphi(x)$ are never antipodal. We note that $(x \in L) \Rightarrow (\varphi(x) = \delta)$; conversely, if $\varphi(x) = \delta$, then either $f(x) = \delta$ or $\mu(x) = 0$, and in either case, $x \in L$. Thus $\varphi^{-1}(\delta) = L$ as required.

We now show that there is an $h\colon (S^n, L) \to (S^n, p)$ that is a homeomorphism of $S^n - L$ onto $S^n - p$. Define

$$h(x) = \frac{x \cdot \mu(x) + (1 - \mu(x))p}{|x \cdot \mu(x) + (1 - \mu(x))p|}.$$

Then h is continuous, $h\colon (S^n, L) \to (S^n, p)$, and also $h \simeq 1$. Furthermore, because L consists of finitely many "straight" lines, μ is evidently monotone nondecreasing on each ray from p, so that $h \mid S^n - L$ is indeed a homeomorphism as asserted.

Since $h\colon S^n \to S^n$ is surjective, it is an identification; thus φh^{-1} is continuous because it is single-valued. If $H\colon h \simeq 1$, then $\varphi h^{-1} \circ H$ shows that $\varphi h^{-1} \simeq \varphi$. Thus $\varphi h^{-1} \simeq f$, and since $(\varphi h^{-1})^{-1}(\delta) = h\varphi^{-1}(\delta) = p$, the proof is complete.

We have seen in XI, **2**, Ex. 5, that the suspension $S\,S^{n-1} \cong S^n$ and (VI, **5.4**) that a $g\colon S^{n-1} \to S^{n-1}$ can be suspended to give $Sg\colon S^n \to S^n$.

7.3 Lemma Let $n \geq 2$ and let $f: S^n \to \Sigma^n$. Then there exists a continuous $g: S^{n-1} \to S^{n-1}$ such that $f \simeq Sg$.

Proof: Let δ_+ be the north pole of Σ^n. By **7.2**, we can assume that $f^{-1}(\delta_+)$ is either empty or p_+, the north pole of S^+. The case

$$f^{-1}(\delta_+) = \varnothing$$

being trivial, since then $f \simeq 0$, we assume that $f^{-1}(\delta_+) = p_+$. By use of III, **11.2**(1), there are spherical nbds $D \supset p_+, \Delta_+ \supset \delta_+, \Delta_- \supset \delta_-$ such that $f(D) \subset \Sigma^n - \Delta_-$ and $f(\overline{\mathscr{C}D}) \subset \Sigma^n - \Delta_+$. We now perform two deformations in the manner of **7.1**. First, let $s: S^n \to S^n$ be a homeomorphism that pushes Fr(D) down so that Fr$(D) = S^{n-1}$, and second let $\tau: \Sigma^n \to \Sigma^n$ be a deformation that pushes $\Sigma^n - \delta_+ - \delta_-$ so that Fr$(\Delta_+) =$ Fr$(\Delta_-) = \Sigma^{n-1}$. Then $g' = \tau \circ f \circ s^{-1} \simeq f$ and also $g'(S^n_+) \subset \Sigma^n_+$, $g'(S^n_-) \subset \Sigma^n_-$. Let $g = g' \mid S^{n-1}: S^{n-1} \to S^{n-1}$; then for each $x \in S^n$, we have $|g'(x) - Sg(x)| < 2$, so (XV, **1.2**) $g' \simeq Sg$ and the proof is complete.

Let $f: S^1 \to S^1$ be any linear T-map. Regarding S^1 parametrized and oriented as the reals mod 1, f is specified by a piecewise linear continuous $F: E^1 \to E^1$ that satisfies $F(x + 1) - F(x) = D$ for all x, where D is an integer (positive, negative, or zero) independent of x. It is routine to verify that $D = D(f)$. Now let $G: S^1 \to S^1$ be defined by $G(x) = Dx$; then $\Phi(x, t) = tF(x) + (1 - t)G(x)$ is a homotopy of F to G and since $\Phi(x + 1, t) - \Phi(x, t) = D$ for each t, the map Φ represents a homotopy of f to g. With this result and with **1.1**, we therefore have: if $f, g: S^1 \to S^1$, then $D(f) = D(g)$ if and only if $f \simeq g$.

7.4 Theorem (H. Hopf) Let $n \geqslant 1$. Two maps of S^n into itself are homotopic if and only if they have the same degree.

Proof: Because of **1.1**, we need show only that $D(f) = D(g) \Rightarrow f \simeq g$. This is true for $n = 1$ according to the remarks made above. Proceeding by induction, we assume that the theorem is true for $n - 1$, and prove it for n. Let $f_1, f_2: S^n \to \Sigma^n$ have the same degree. By **7.3**, we can assume that $f_i = Sg_i$ $(i = 1, 2)$ for suitable $g_i: S^{n-1} \to S^{n-1}$. Now $f_i \mid S^n_+ : S^n_+ \to \Sigma^n_+$ is a regular map, and it is trivial to verify that $D_r(f_i \mid S^n_+) = D(f_i)$, so from **1.3** we find that $D(g_i) = D(f_i)$. By the induction hypothesis, $g_1 \simeq g_2$, and suspending this homotopy gives $Sg_1 \simeq Sg_2$, which concludes the proof.

In particular, the map $[f] \to D(f)$ of $[S^n, S^n]$ into the set Z of positive and nonpositive integers is bijective; a map $S^n \to S^n$ of degree k can be obtained by suspending $(k - 1)$ times the map $f: S^1 \to S^1$ given by $z \to z^k$.

Problems

Section 1

1. Prove: $[S^n, S^n]$ has infinitely many elements.
2. Let $f, g: S^n \to S^n$. Prove $D(g \circ f) = D(g) \cdot D(f)$.
3. Let $f: S^n \to S^n$ be a homeomorphism. Show that $D(f) = \pm 1$.

Section 2

1. Prove Brouwer's theorem equivalent to the following: Let I^n be the unit n-cube, and let C_i be the $(n-1)$-face $\{x \in I^n \mid x_i = 1\}$, C_i' the opposite face. For each i, let B_i be a closed set separating C_i and C_i': (that is, C_i and C_i' are in different components of $I^n - B_i$). Then $\bigcap_1^n B_i \neq \varnothing$.

   Hint: For each $i = 1, \cdots, n$, let $h_i: I \to E^1$ be the function $h_i(x) = \pm d(x, B_i)$, where the sign is chosen suitably so that for each x, the vector $f(x) = x + h(x)$ points into the cube.
2. Let f be a continuous map of V^n into E^n. Assume that for each $p \in S^{n-1}$, p does not lie on the line $\overrightarrow{0f(p)}$. Prove: There exists a $q \in V^n$ such that $f(q) = q$.
3. Prove the case $n = 0$ of **2.2**(2) directly from the intermediate value theorem.
4. Let A be a retract of V^n. Show that each continuous map of A into itself has a fixed point.

Section 3

1. Let $g_i(x_1, \cdots, x_n)$, $i = 1, \cdots, n$ be continuous real-valued functions on V^n. Assume that $\Sigma g_i^2 \neq 0$ everywhere in V^n. Prove: There exists a $\lambda > 0$ and a $\mu < 0$ such that each one of the systems of n equations $g_i(x_1, \cdots, x_n) = \lambda x_i$ and $g_i(x_1, \cdots, x_n) = \mu x_i$ has solutions on S^n.
2. Let n be odd, and let n continuous functions $g_i(x_1, \cdots, x_n)$ be defined on S^{n-1}. Show that there exists some real λ such that the system of equations
$$g_i(x_1, \cdots, x_n) = \lambda x_i$$
   has a solution on S^{n-1}.

Section 4

1. Let f_1, f_2: $V^k \to S^n$ be two maps such that $f_1 \mid S^{k-1} = f_2 \mid S^{k-1}$. Prove: If $k < n - 1$, then $f_1 \simeq f_2$ rel S^{k-1}.
2. Let σ^m be an m-dimensional simplex and σ^n an n-dimensional simplex. Let $f: \sigma^m \to \sigma^n$ be a barycentric map and $p \in \sigma^n$ be not on any $(n-1)$-face of σ^n. Show that the dimension of $f^{-1}(p)$ is $(m - n)$.

Section 5

1. Let X be a normal space such that $X \times I$ is also normal, and let $f, g: X \to S^n$ be homotopic. Let $x_0 \in X$ be any point and α be any path on S^n running from $f(x_0)$ to $g(x_0)$. Prove: If $n \geqslant 2$, then there is a homotopy $\Delta: f \simeq g$ such that $\Delta(x_0, t) = \alpha(t)$, $0 \leqslant t \leqslant 1$.

2. Let X be a normal space such that $X \times I$ is also normal, and let $f, g: X \to S^n$. Assume that there is an $s \in S^n$ and a nbd $U(s)$ such that (a) $f^{-1}(s) = g^{-1}(s)$, and (b) $g(x) = f(x)$ for each $x \in f^{-1}(U)$. Prove: $f \simeq g \text{ rel } f^{-1}(s)$.

Section 6

1. Prove: There does not exist any continuous $f: V^n \to S^{n-1}$ with the property that $f \mid S^{n-1}$ is antipodal-preserving.
2. Let $f: S^n \to S^n$ be such that $f(-x) \neq f(x)$ for each $x \in S^n$. Prove that f is surjective.
3. Let $f: V^n \to S^{n-1}$ be continuous. Prove that there exists an $x \in S^{n-1}$ such that $f(x) = f(-x)$.
4. Let Φ be a continuous nonvanishing vector field on V^n. Show that there exists an $x \in S^{n-1}$ such that the vector $\Phi(x)$ is parallel to the vector $\Phi(-x)$.
5. Let f_i, $i = 1, \cdots, n$, be n continuous real-valued functions on S^n such that $f_i(-x) \neq -f_i(x)$ for each i and $x \in S^n$. Show that there exists an $x_0 \in S^n$ such that $f_i(x_0) = 0$ for all $i = 1, \cdots, n$.
6. Let V^n be covered by n closed sets. Show that at least one of these sets must contain a pair of antipodal points of S^{n-1}.
7. Prove: If $f: S^n \to S^n$ has even degree, it carries at least one pair of antipodal points to the same point. [*Hint:* Let $h(x) = 2^{-1}(f(x) - f(-x))$ and consider $h(x)/|h(x)|$.]
8. Prove: If $f: S^n \to S^n$ has odd degree, it carries at least one pair of antipodal points to antipodal points.

Section 7

1. Prove that the map $S: [S^m, S^n] \to [S^{m+1}, S^{n+1}]$ given by $S[f] = [Sf]$ is surjective whenever $m \leqslant 2n - 1$. (It is actually bijective for $m \leqslant 2n - 2$, though a proof requires methods beyond the scope of this book.)
2. Let $f, g: S^n \to S^n$. Prove that $f \circ g \simeq g \circ f$.
3. Let $f: S^n \to S^n$. Prove:
   a. If $D(f) \neq 1$, then f sends some point to its antipode.
   b. If $D(f) \neq (-1)^{n+1}$, then f has a fixed point.
4. Let $f: S^n \to S^n$ be antipodal and assume that $D(f) \neq (-1)^{n+1}$. Show that there exists an $x_0 \in S^n$ such that both x_0 and $-x_0$ are fixed points of f.
5. Let $f, g: S^{2n} \to S^{2n}$. Prove that at least one of f, g, $g \circ f$ has a fixed point.
6. Let $f: S^{2n} \to S^{2n}$. Show that either f has a fixed point, or else there exists an $x_0 \in S^{2n}$ such that $f(x_0) = y_0$ and $f(y_0) = x_0$.
7. Let S^n be covered by $(n + 1)$ connected closed sets, $F_1, \cdots, F_{n+1}$. Show that there exists an F_k having the following property: for every real number $d \leqslant 2$, there are points $x, y \in F_k$ such that $|x - y| = d$. [*Hint:* Let F_k be one of the sets containing an antipodal pair, x_0 and $-x_0$; then consider the real-valued function $g(x) = |x - x_0|$ on F_k.]

Topology of E^n

XVII

Let X, Y be two given spaces, and assume that X can be embedded in Y. Although all embeddings of X in Y give homeomorphic subspaces, these subsets of Y may behave differently in some respect; that is, they may not all belong to some previously specified subset of $\mathscr{P}(Y)$. A subset $\mathscr{A} \subset \mathscr{P}(Y)$ is called a positional bound of X rel Y if either every embedding of X into Y belongs to $\mathscr{A}$ or no embedding of X into Y belongs to $\mathscr{A}$. A property P that specifies a positional bound for X rel Y is called a positional invariant of X rel Y.

Ex. 1 For any spaces X, Y, each of the two properties "compact" and "connected" is a positional invariant of X rel Y.

Ex. 2 Let Y be d-complete. For any space X, the property "is a G_δ-set" is a positional invariant of X rel Y (*cf.* XIV, **9.3**).

Ex. 3 Let $X = E^1$, $Y = E^2$. Since under the embedding $x \to (x, 0)$, the complement of E^1 is not connected, whereas it is connected under the embedding $x \to \left(\frac{x}{1 + |x|}, 0\right)$, the property "separates E^2" is not a positional invariant of E^1 rel E^2.

Ex. 4 Let $Y \subset E^1$ be the subspace $\{0\} \cup \{1/n\}$ of E^1 and X be a single point. The embeddings $x \to 0$ and $x \to 1$ show that the property "open in Y" is not a positional invariant of X rel Y.

Using this terminology, our principal objective in this chapter is to prove:

(1). Let X be compact. Then the property "separates E^n" is a positional invariant of X rel E^n.
(2). Let X be arbitrary. Then the property "open in E^n" is a positional invariant of X rel E^n.

The main tools are XVI, **2.1**, **4.4**, and **5.2.**

1. Components of Compact Sets in E^{n+1}

1.1 Definition The set $A \subset E^{n+1}$ *separates* E^{n+1} if $E^{n+1} - A$ is not connected.

Ex. 1 S^n separates E^{n+1}, since $\mathscr{C}S^n = \{x \mid |x| < 1\} \cup \{x \mid |x| > 1\}$ is a union of two nonempty disjoint open sets.

Ex. 2 If the compact A separates E^{n+1}, then $\mathscr{C}A$ may have infinitely many components, as shown by $A = \{0\} \cup \bigcup_{1}^{\infty} \mathrm{Fr}[B(0; 1/n)]$.

1.2 Theorem Let $A \subset E^{n+1}$ be compact. Then:

(1). Each component of $E^{n+1} - A$ is a path-connected open set.
(2). $\mathscr{C}A$ has exactly *one* unbounded component.
(3). The boundary of each component of $\mathscr{C}A$ is contained in A.
(4). If A separates E^{n+1}, but no proper closed subset does so, then the boundary of each component of $\mathscr{C}A$ is exactly A.

Proof: (1) follows from V, **4.2** and **4**, Ex. 1; (2) has been proved in XVI, **4**, Ex. 3 and accompanying text.

Ad (3). Let U be any component of $\mathscr{C}A$, and let $x \in \mathrm{Fr}(U) = \overline{U} - U$. Since U is open, we have $x \,\overline{\in}\, U$, and we prove $x \in A$ by showing that x cannot belong to any other component U_1 of $\mathscr{C}A$. Indeed, if $x \in U_1$, there would be a ball $B(x, \varepsilon) \subset U_1$ [see (1)]; since $x \in \overline{U}$, we would then have $B(x, \varepsilon) \cap U \neq \varnothing$ and, consequently, $U \cap U_1 \neq \varnothing$, a contradiction. Thus $\mathrm{Fr}(U) \subset A$, as asserted.

Ad (4). Let U be any component of $\mathscr{C}A$; since A separates, there is another component V of $\mathscr{C}A$, and because $V \subset E^{n+1} - \overline{U}$, we have $E^{n+1} - \overline{U} \neq \varnothing$. The formula

$$E^{n+1} - \mathrm{Fr}(U) = U \cup (E^{n+1} - \overline{U})$$

displays $E^{n+1} - \mathrm{Fr}(U)$ as a union of two disjoint nonempty open sets, so $\mathrm{Fr}(U)$ separates E^{n+1}. Since $\mathrm{Fr}(U) \subset A$ and is closed, the hypothesis on A requires that $\mathrm{Fr}(U) = A$.

2. Borsuk's Separation Theorem

For any $p \in E^{n+1}$, let $\beta_p: (E^{n+1} - p) \to S^n$ be the mapping

$$x \to \frac{x - p}{|x - p|}.$$

For any set $A \subset (E^{n+1} - p)$, we will call $\beta_p \mid A: A \to S^n$ a Borsuk map of A; these simple maps play an important role in discussions of the separation properties of A.

2.1 Theorem (K. Borsuk) Let $A \subset E^{n+1}$ be compact. Then A separates E^{n+1} if and only if there exists an $f: A \to S^n$ that is *not* nullhomotopic.

Proof: Assume that A separates E^{n+1}; then $\mathscr{C}A$ has at least *one* bounded component U. Selecting any $p \in U$, we will show that the Borsuk map $\beta_p \mid A$ is not nullhomotopic, by proving (XVI, **5.2**) that it cannot be extended to an $F: E^{n+1} \to S^n$. Indeed:

(1) If U is a bounded component of $\mathscr{C}A$ and $p \in U$, then $\beta_p \mid A$ cannot be extended over the closed set $A \cup U$.

For, if there were an extension $F: A \cup U \to S^n$, we draw a ball $B = B(p, R) \supset A \cup U$ and define $g: \bar{B} \to \text{Fr}(B)$ as follows:

$$\begin{aligned} g(x) &= p + R \cdot \frac{x - p}{|x - p|} \qquad && x \in \bar{B} - U \\ &= p + R \cdot F(x) && x \in \bar{U}. \end{aligned}$$

Then g would be continuous, since the intersection of the two closed sets lies in $\text{Fr}(U) \subset A$ and both definitions of g agree on A; but, since g is a retraction $\bar{B} \to \text{Fr}(B)$, this would contradict Brouwer's theorem.

Conversely, assume that A does not separate E^{n+1}; then $\mathscr{C}A$ has exactly one component, which is necessarily unbounded. By XVI, **4.4**, any $f: A \to S^n$ is extendable to an $F: E^{n+1} \to S^n$, and so (XVI, **5.2**) is nullhomotopic.

Ex. 1 The compactness of A is essential: $E^1 \subset E^2$ separates E^2; yet, every $f: E^1 \to S^1$ is nullhomotopic (XV, Ex. 1).

Using XV, **3.1**, Borsuk's theorem has the elegant formulation

2.2 Corollary Let $A \subset E^{n+1}$ be compact. Then A does not separate E^{n+1} if and only if the function space $(S^n)^A$ is path-connected.

Since $(S^n)^A$ is a topological invariant of A, **2.2** can also be stated as

2.3 Corollary Let X be a compact space. Then the property "separates E^{n+1}" is a positional invariant of X rel E^{n+1}.

An immediate application of **2.3** yields

2.4 Theorem (Jordan Separation Theorem) Every homeomorphic image $\mathcal{J}^n$ of S^n in E^{n+1} separates E^{n+1}, and no proper closed subset of $\mathcal{J}^n$ does so. In particular, the entire set $\mathcal{J}^n$ is the complete boundary of each component of $\mathscr{C}\mathcal{J}^n$.

Proof: Since S^n is compact and separates E^{n+1} (**1**, Ex. 1), it follows from **2.3** that so also does each $\mathcal{J}^n \subset E^{n+1}$. Any proper subset F of $S^n \subset E^{n+1}$ does not separate E^{n+1}, since $E^{n+1} - F$ is evidently path-connected. Whenever F is closed, **2.3** shows that no homeomorph of F can separate E^{n+1}.

Remark 1: The general Jordan theorem is **2.4** together with the additional statement that $E^{n+1} - \mathcal{J}^n$ has exactly two components. This additional conclusion is nontrivial, and does not follow simply from the fact that $\mathcal{J}^n$ is the boundary of each residual component. There exist compact sets B that separate E^{n+1} into infinitely many parts, each component of $\mathscr{C}B$ having the entire set B as its complete boundary. In this book, we will prove the general Jordan theorem only for embeddings of S^1 in E^2 (**5**).

Remark 2: The general Jordan theorem follows in obvious fashion from the more specific version of **2.3** due to J. W. Alexander: If $A \subset E^{n+1}$ is compact, then the number of components in $\mathscr{C}A$ is a positional invariant of A rel E^{n+1}. A proof of this requires techniques different from those used in this book (either homology theory, or cohomotopy groups).

Remark 3: The general Jordan theorem immediately leads to the Schoenflies problem: If U is the bounded component of $\mathcal{J}^n \subset E^{n+1}$, is $U \cup \mathcal{J}^n$ always homeomorphic to V^{n+1}? The answer is "yes" for $n = 1$, and "no" for each $n > 1$ (for $n = 2$, Alexander's "horned sphere" H is the usual counterexample, and the $(n - 2)$-fold suspension of H is a counterexample in E^{n+1}). It has recently been proved (M. Brown, B. Mazur, M. Morse) that if the embedding $h: S^n \to E^{n+1}$, regarded as an embedding $S^n \times 0 \to E^{n+1}$, can be extended to an embedding $H: S^n \times [-\varepsilon, \varepsilon] \to E^{n+1}$ for some $\varepsilon > 0$, (that is, if h can be extended to an embedding of a "shell" containing S^n) then the Schoenflies problem for $h(S^n)$ has an affirmative answer.

3. Domain Invariance

3.1 Theorem (Brouwer Domain Invariance Theorem) For any space X, the property "open in E^{n+1}" is a positional invariant of X rel E^{n+1}.

Proof: Let $U \subset E^{n+1}$ be open and $h: U \to E^{n+1}$ be a homeomorphism; we are to prove that $U_1 = h(U)$ is open in E^{n+1}. It suffices to show that for each $x_1 \in U_1$, there is an open (in E^{n+1}!) set W such that

$x_1 \in W \subset U_1$. Let $x = h^{-1}(x_1)$ and let $B = B(x, \varepsilon)$ be a ball such that $\bar{B} \subset U$. Then:

(1). $E^{n+1} - h(\bar{B})$ is connected, because $\bar{B} \cong V^{n+1}$ and V^{n+1} does not separate E^{n+1}.

(2). $h(\bar{B} - \text{Fr}[B]) = h(\bar{B}) - h(\text{Fr}[B])$ is connected, since it is homeomorphic to $B(x, \varepsilon)$.

Because of (1) and (2) the formula

$$E^{n+1} - h(\text{Fr}[B]) = \{E^{n+1} - h(\bar{B})\} \cup \{h(\bar{B} - \text{Fr}[B])\}$$

expresses $E^{n+1} - h(\text{Fr}[B])$ as the union of two nonempty disjoint connected sets; these must therefore be the components of $E^{n+1} - h(\text{Fr}[B])$, and since $h(\text{Fr}[B])$ is compact, each must be an open set (in E^{n+1}!). Setting $W = h(\bar{B} - \text{Fr}[B])$ gives $x_1 \in W \subset U_1$ as required.

3.2 Corollary Let $B \subset E^{n+1}$ be *any* set and $h: B \to E^{n+1}$ be a homeomorphism. Then if x is an interior (boundary) point of B, $h(x)$ is also an interior (boundary) point of $h(B)$.

Proof: For "interior": By **3.1**, $h[\text{Int}(B)] = \text{Int}[h(B)]$. For "boundary": $x \in \text{Fr}(B) \Rightarrow x \in B - \text{Int}(B) \Rightarrow h(x) \in h(B) - \text{Int}[h(B)]$.

By using only a connectedness argument, we were able to show that E^1 is not homeomorphic to E^n for any $n > 1$; with **3.1**, we can now prove

3.3 Theorem (Invariance of Dimension Number) E^m and E^n are not homeomorphic whenever $m \neq n$.

Proof: Let $m > n$. If E^n were homeomorphic to E^m then because of **3.1**, the image of E^n under any embedding in E^m would be open in E^m. However, the image is not open under the embedding

$$(x_1, \cdots, x_n) \to (x_1, \cdots, x_n, 0, \cdots, 0)$$

and this proves the theorem.

4. Deformations of Subsets of E^{n+1}

In this section, we derive some properties of the Borsuk maps and use them to express some intuitively evident statements about subsets that separate E^{n+1}.

4.1 Let $A \subset E^{n+1}$ be compact. Then:

(1). p and q belong to the same component of $\mathscr{C}A$ if and only if $\beta_p \mid A \simeq \beta_q \mid A$.

(2). p belongs to the unbounded component of $\mathscr{C}A$ if and only if $\beta_p \mid A \simeq 0$.

Proof: (1). Assume that p, q are in a common component U; then [**1.2**(1)] there is a continuous $\alpha: I \to U$ such that $\alpha(0) = p$, $\alpha(1) = q$, and the map $\Phi: A \times I \to S^n$ given by $\Phi(a, t) = \dfrac{a - \alpha(t)}{|a - \alpha(t)|}$ shows that $\beta_p \mid A \simeq \beta_q \mid A$. Conversely, assume p, q are not in the same component; at least one (say, p) lies in a bounded component U. If $\beta_p \mid A$ were homotopic to $\beta_q \mid A$, then because $\beta_q \mid A$ can be extended over $E^{n+1} - q$, XVI, **5.1** would yield an extension of $\beta_p \mid A$ over $E^{n+1} - q$ also; in particular, β_p would be extendable over $A \cup U$, which, by statement (1) in the proof of **2.1**, is impossible.

(2). We need show only that $\beta_p \mid A \simeq 0$ for some p belonging to the unbounded component U of $\mathscr{C}A$. Choose r so large that $A \subset B(0; r)$, and let $p \in U \cap \mathscr{C}B(0; r)$; then $\beta_p(A) \subset S^n - \{p/|p|\}$, and therefore $\beta_p \mid A \simeq 0$.

Since E^{n+1} is contractible (XV, **1**, Ex. 1), any $A \subset E^{n+1}$ is also contractible (over E^{n+1}) to a point.

4.2 Let $A \subset E^{n+1}$ be compact, and let p be any point of a bounded component of $\mathscr{C}A$. Then, in any contraction of A over E^{n+1} to a point, A must cross p (that is, A must contain p at some stage of the deformation).

Proof: Let $\Phi: A \times I \to E^{n+1}$ be any contraction of A over E^{n+1} to a point where $\Phi(a, 0) = a$, $\Phi(a, 1) = x_0 \in E^{n+1}$ for each $a \in A$. If $p \bar{\in} \Phi(A \times I)$, then $\dfrac{\Phi(a, t) - p}{|\Phi(a, t) - p|}$ would show $\beta_p \mid A \simeq 0$ which, by **4.1**(2), is impossible.

4.3 Let $A \subset E^{n+1}$ be compact, and let p, q be in distinct components of $\mathscr{C}A$. If A is deformed over E^{n+1} into a set A_1, and if A never crosses either p or q during this deformation, then p, q are still in distinct components of $\mathscr{C}A_1$.

Proof: This is immediate from **4.1**(1) and XV, **6.2**.

4.4 Let $A \subset E^{n+1}$ be compact, and $B \subset A$ be a retract of A. Then $\mathscr{C}B$ cannot have more components than $\mathscr{C}A$.

Proof: We will show that if U_B is any bounded component of $\mathscr{C}B$, then there is some component U_A of $\mathscr{C}A$ contained in U_B. Note first that $U_B - A \neq \varnothing$: For, if $U_B \subset A$, then because $\mathrm{Fr}(U_B) \subset B$, the retraction $r: A \to B$ would give a retraction $r \mid \bar{U}_B: \bar{U}_B \to \mathrm{Fr}(U_B)$, which

(XVI, **2.3**) is impossible. Thus there is a $p \in U_B - A$, and we let U_A be the component of $\mathscr{C}A$ containing p. Since it is clear that

$$(\beta_q \mid A \simeq \beta_p \mid A) \Rightarrow (\beta_q \mid B \simeq \beta_p \mid B),$$

it follows at once that $U_A \subset U_B$, as required.

5. The Jordan Curve Theorem

Two maps of an arbitrary space X into S^1 can be combined by multiplying their values (regarded as complex numbers of modulus 1). Taking advantage of this fact [which evidently makes $(S^1)^X$ into a group], we prove that a special type of separation theorem (**5.2**) is valid for subsets of the plane, and this then yields the general Jordan theorem for E^2.

5.1 Theorem (S. Eilenberg) (1). Let X be any compact metric space and $f: X \to S^1$ be continuous. Then $f \simeq 0$ if and only if there exists a continuous $\varphi: X \to E^1$ such that $f(x) = e^{i\varphi(x)}$ for all $x \in X$.

(2). $f \simeq g$ if and only if $f/g \simeq 0$.

Proof: (1). If f is of the stated form, then $\Phi(x, t) = e^{it\varphi(x)}$ shows that $f \simeq 0$. For the converse, we first make the following observation: If f is of the stated form and if $|g(x) - f(x)| < 2$ for all $x \in X$, then g is also of the stated form. Indeed, $\dfrac{g(x)}{f(x)} \neq -1$, since they are never antipodal, so by defining $\varphi(x)$ to be the length of that oriented arc from $z = 1$ to $\dfrac{g(x)}{f(x)}$ which does not contain $z = -1$, it follows that $\varphi(x)$ is continuous and that $\dfrac{g(x)}{f(x)} = e^{i\varphi(x)}$; thus $g(x) = f(x)\cdot e^{i\varphi(x)}$ and g has the stated form also. We now prove the theorem.

Let $\Phi: 0 \simeq f$, and write $\Phi(x, t) = f_t(x)$. By uniform continuity, there is a $\delta > 0$ such that $|f_t(x) - f_{t'}(x)| < 2$ for all x whenever $|t - t'| < \delta$. Choose a subdivision $0 = t_0 < t_1 < \cdots < t_n = 1$ of I such that $|t_{i+1} - t_i| < \delta$. Then, noting that a constant map has the stated form, our observation permits us to conclude successively that each f_{t_i} has the stated form also, and therefore so also does $f_1 = f$.

(2). Clearly $\Phi: f \simeq g \Rightarrow 1/g\cdot\Phi: f/g \simeq 0$; conversely, if $\Phi: f/g \simeq 0$, we can assume that $\Phi(x, 1) = 1$ for all $x \in X$, and then $g(x)\cdot\Phi(x, t)$ shows $f \simeq g$.

We say that $A \subset E^2$ separates x, y if x and y lie in distinct components of $\mathscr{C}A$.

5.2 Theorem (S. Janiszewski) Let A, B be compact sets in E^2 and let x, y be two points of E^2. Assume: (1) A does not separate x, y; (2) B does not separate x, y; and (3) $A \cap B$ is connected. Then $A \cup B$ does not separate x, y.

Proof: Using the Borsuk maps, let

$$f = \frac{\beta_x \mid A \cup B}{\beta_y \mid A \cup B} : A \cup B \to S^1;$$

we are to show that $f \simeq 0$. The hypotheses give $f \mid A = e^{i\varphi_A}$, $f \mid B = e^{i\varphi_B}$ for suitable $\varphi_A: A \to E^1$, $\varphi_B: B \to E^1$, and noting that we have $(f \mid A)(x_0) = (f \mid B)(x_0)$ for each $x_0 \in A \cap B$, it follows that $\varphi_A(x_0) - \varphi_B(x_0) = 2n(x_0)\pi$ for some integer $n(x_0)$. Since the formula

$$n(x) = \frac{\varphi_A(x) - \varphi_B(x)}{2\pi}$$

shows that $n(x)$ is a continuous integer-valued function on the connected $A \cap B$, we conclude that $n(x)$ has a constant value n on $A \cap B$, and by defining

$$\begin{aligned} \psi(x) &= \varphi_A(x) & x \in A \\ &= \varphi_B(x) + 2n\pi & x \in B, \end{aligned}$$

we obtain a continuous $\psi: A \cup B \to E^1$ such that $f = e^{i\psi}$. Thus $f \simeq 0$, as required.

As a final preliminary we will need

5.3 Let $U \subset E^n$ be open, and call an $x \in \mathrm{Fr}(U)$ accessible from U if there is a path $\alpha: I \to \bar{U}$ starting at x and such that $\alpha(I - \{0\}) \subset U$. Then the accessible points are dense on $\mathrm{Fr}(U)$.

Proof: Let $a \in \mathrm{Fr}(U)$ and let $B(a, \varepsilon)$ be any nbd of a. Choose any $p_0 \in U \cap B(a, \varepsilon)$ and consider the function $|p_0 - x|$ on $\mathrm{Fr}(U) \cap \overline{B(a, \varepsilon)}$; since the latter set is compact, $|p_0 - x|$ attains its minimum at some x_0, and x_0 is evidently accessible by a straight line from p_0.

5.4 Theorem (Jordan Curve Theorem) Let $\mathcal{J}^1$ be a Jordan curve (that is, a homeomorph of S^1) in E^2. Then $E^2 - \mathcal{J}^1$ has exactly two components, each of which has $\mathcal{J}^1$ as its complete boundary.

Proof: According to **2.4**, $E^2 - \mathcal{J}^1$ has at least two components; we need prove only that it has no more than two. Assume that there were three: G_0, G_1, G_2. Select points $a, b \in \mathcal{J}^1$ $(a \neq b)$ accessible from G_2, and let L be an arc joining these two points, which (except for its end points) lies in G_2. Choose also points q_0, q_1 on each of the two arcs $\widehat{ab}$ and $\widehat{ba}$ of $\mathcal{J}^1$, and let

$$A = \text{arc}(aq_0b) \cup L,$$
$$B = \text{arc}(bq_1a) \cup L.$$

Then, for any fixed $x \in G_0$, $y \in G_1$, we find:

(1). A does not separate x, y, since $G_0 \cup G_1 \cup q_1$ is a connected set in $\mathscr{C}A$ containing x and y.

(2). B does not separate x, y, since $G_0 \cup G_1 \cup q_0$ is a connected set in $\mathscr{C}B$ containing x and y.

(3). $A \cap B = L$ is connected.

According to **5.2**, it follows that $A \cup B$ does not separate x and y. But this is impossible, since $\mathcal{J} \subset A \cup B$ and x, y are in distinct components of $\mathcal{J}$. This completes the proof. (This argument is due to C. Kuratowski.)

The affirmative solution for the Schoenflies problem follows immediately from the Riemann mapping theorem of function theory: If U is the bounded component of $\mathcal{J}^1$, then $U \cup \mathcal{J}^1$ is homeomorphic to V^2, and in fact a homeomorphism can be chosen that is conformal at all points of U.

Problems

Section 1

1. Prove: For any closed $A \subset E^{n+1}$, $\mathscr{C}A$ can have at most countably many components.

2. Let $A \subset E^{n+1}$ be closed and U be a component of $\mathscr{C}A$. Prove that $A \cup U$ is closed in E^{n+1}.

3. Let A be closed, and assume that for each component U of $\mathscr{C}A$, we have $\text{Fr}(U) = A$. Prove that no proper closed subset of A separates E^{n+1}.

4. Let $T = S^1 \times S^1$ be the torus. Show that "separates T" is not a positional invariant of S^1 rel T.

Section 2

1. Prove: No embedding of I^n in E^n, $n \geqslant 2$, can separate E^n.

2. Prove: If X is compact, then "separates S^{n+1}" is a positional invariant of X rel S^{n+1}.

3. Let T be the torus $S^1 \times S^1$. Show that there exists a map $f\colon T \to S^2$ that is *not* nullhomotopic, and describe its construction.

Section 3

1. Let $h: S^n \to E^{n+1}$ be an embedding, and assume that h is extendable to a homeomorphism $H: V^{n+1} \to E^{n+1}$. Prove: $E^{n+1} - h(S^n)$ has exactly two components, and one of them is $H[\text{Int}(V^{n+1})]$.
2. Prove: S^n is not homeomorphic to any proper subset of itself.
3. A space M is called locally Euclidean of dimension n if each point has a nbd homeomorphic to E^n. Prove that for any space X, the property "open in M" is a positional invariant of X rel M.

Section 4

1. Let $A \subset E^{n+1}$ be homeomorphic to S^n. Prove that $\mathscr{C}A$ has at most finitely many components. [*Hint*: Use **4.4** and the fact that A is an ANR.]
2. Let $A \subset E^{n+1}$ be compact and $B \subset A$ be a deformation retract of A. Show that $\mathscr{C}B$ and $\mathscr{C}A$ have the same number of components.

Section 5

1. Prove directly from Theorem **5.1** that the identity map $1: S^1 \to S^1$ is not nullhomotopic. [*Hint*: Argue by contradiction.]
2. Prove: If A, B are two compact connected sets in E^2, and if $A \cap B$ is not connected, then $A \cup B$ separates E^2.
3. Let Y be compact metric, and $g: Y \to Z$ a continuous open surjection. Prove that if $(S^1)^Y$ is path-connected, then $(S^1)^Z$ is also path-connected. [*Hint*: *cf.* III, Problem **11.16**.]
4. A space Y is called unicoherent if for every decomposition $Y = A_1 \cup A_2$ into two closed connected sets, the intersection $A_1 \cap A_2$ is also connected. Let Y be a compact metric space. Prove: if $(S^1)^Y$ is path-connected, then Y in unicoherent. [*Hint*: If Y were not unicoherent, then $Y = A_1 \cup A_2$ and $A_1 \cap A_2 = C_1 \cup C_2$ where $C_1 \cap C_2 = \varnothing$ and each C_i is closed. Let g be a Urysohn function on Y which equals 0 on C_1 and equals π on C_2; show that the map

$$f(x) = \begin{cases} e^{ig(x)} & x \in A_1 \\ e^{-ig(x)} & x \in A_2 \end{cases}$$

is not nullhomotopic.]

Remark If Y is a locally connected compact metric space, then the converse is also true; an elementary proof based on **3.1** can be constructed, but it is fairly involved. In particular, a compact locally connected $A \subset E^2$ does not separate E^2 if and only if A is unicoherent.

Homotopy Type

XVIII

In this chapter, we use homotopy to derive a classification of topological spaces different from that obtained by homeomorphism, and we determine some of the simplest features of this classification.

1. Homotopy Type

1.1 Definition Two spaces X, Y are homotopic (written: $X \simeq Y$) if there exists an $f: X \to Y$ and a $g: Y \to X$ such that both $g \circ f \simeq 1_X$ and $f \circ g \simeq 1_Y$.

This relation between spaces is clearly reflexive, symmetric, and transitive; it therefore decomposes the class of all spaces into mutually exclusive subclasses called homotopy types, two spaces belonging to the same homotopy type if and only if they are homotopic.

Ex. 1 $X \cong Y$ implies $X \simeq Y$, since $X \cong Y$ amounts to the existence of maps $f: X \to Y$ and $g: Y \to X$ such that both $f \circ g$ and $g \circ f$ are actually identity maps. The converse implication is not true: It is trivial to verify that X is homotopic to a space consisting of a single point if and only if X is contractible. Thus, for example, E^n, I^∞, and one-point spaces all belong to the same homotopy type.

Further examples of homotopic spaces follow from

1.2 (1). A deformation retract of X belongs to the homotopy type of X.

(2). If a deformation of X contracts a closed $A \subset X$ over itself to a point, then $X \simeq X/A$.

(3). If $X_\alpha \simeq Y_\alpha$ for each $\alpha \in \mathscr{A}$, then $\prod_\alpha X_\alpha \simeq \prod_\alpha Y_\alpha$.

Proof: Ad (1). Let $r: X \to A$ be a retraction such that $r \simeq 1_X$ and let $i: A \to X$ be the inclusion map. Then $r \circ i = 1_A$ and $i \circ r = r \simeq 1_X$.

Ad (2). Let $p: X \to X/A$ be the identification map, and

$$\Phi: (X \times I, A \times I) \to (X, A)$$

the deformation. Writing $\Phi(x, t) = \varphi_t(x)$, we note that because $\varphi_1(A)$ is a single point, the map $\varphi_1 p^{-1}: X/A \to X$ is single-valued, and therefore (VI, **3.2**) is continuous. Furthermore (XII, **4.1**),

$$p \times 1: X \times I \to X/A \times I$$

is also an identification, and since Φ contracts A over itself, the map $(p \circ \Phi)(p \times 1)^{-1}: X/A \times I \to X/A$ is single-valued also, so that that $(p \circ \varphi_t)p^{-1}$ is a (continuous!) homotopy. To prove (2), we now need note only that $\varphi_1 p^{-1} \circ p = \varphi_1 \simeq \varphi_0 = 1_X$ and

$$p \circ \varphi_1 p^{-1} = (p \circ \varphi_1)p^{-1} \simeq (p \circ \varphi_0)p^{-1} = 1_{X/A}.$$

Ad (3). This is immediate from XV, **2.2.**

In work with homotopy type, the notion "inverse of a map" is replaced by "inverse in the sense of homotopy:"

1.3 Definition A map $f: X \to Y$ is called a homotopy equivalence if there exists a $g: Y \to X$ such that both $f \circ g \simeq 1_Y$ and $g \circ f \simeq 1_X$; g is called a homotopy inverse of f. If $\bar{g}: Y \to X$ satisfies only $\bar{g} \circ f \simeq 1_X$ (resp. $f \circ \bar{g} \simeq 1_Y$), $\bar{g}$ is called a left (resp. right) homotopy inverse of f.

With this terminology, $X \simeq Y$ if and only if there exists a homotopy equivalence $f: X \to Y$; we denote "f is a homotopy equivalence" by "$f: X \simeq Y$."

The following result, that a given $f: X \to Y$ has a homotopy inverse whenever it has both some right homotopy inverse and some left homotopy inverse, is very useful in practice.

1.4 Theorem (M. M. Day, R. H. Fox) Let $f: X \to Y$ have a left homotopy inverse g_L and a right homotopy inverse g_R. Then f is a homotopy equivalence.

Proof: We will show that $g = g_L \circ f \circ g_R$ is a homotopy inverse of f. Indeed, $f \circ g = f \circ g_L \circ f \circ g_R \simeq f \circ 1_X \circ g_R = f \circ g_R \simeq 1_Y$, and similarly, $g \circ f = g_L \circ f \circ g_R \circ f \simeq g_L \circ f \simeq 1_X$.

2. Homotopy-Type Invariants

Though most topological invariants are not invariants of homotopy type, in this section we give some that are and which account for the importance of the concept.

2.1 Path-connectedness is a homotopy-type invariant More generally in order that $X \simeq Y$ it is necessary and, if both X, Y are locally path-connected, also sufficient that there be a 1-to-1 correspondence between their path-components, with corresponding components belonging to the same homotopy type.

Proof: Let $f: X \simeq Y$ have homotopy inverse g. To prove the proposition, it suffices to show that for two path-components $A \subset X$, $B \subset Y$, we have $f(A) \subset B$ if and only if $g(B) \subset A$. Let $a \in A$ and $f(a) \in B$; then $g(B)$ is path-connected and contains $g \circ f(a)$; using a homotopy $\Phi: g \circ f \simeq 1_X$, the path $t \to \Phi(a, t)$ joins $g \circ f(a)$ to a, so $g(B) \subset A$. The coverse follows by interchanging g and f. The sufficiency is clear, because in locally path-connected spaces each path-component (V, **5.4**) is both open and closed.

Because of **2.1**, we are justified in considering only path-connected spaces. For the homotopy classes of maps,

2.2 Theorem Let $f: X \simeq Y$. Then for any space Z, the induced maps $f^{\#}: [Y, Z] \to [X, Z]$ and $f_{\#}: [Z, X] \to [Z, Y]$ are always bijective.

Proof: Let g be a homotopy inverse of f. Since $f \circ g \simeq 1_Y$, we find for any $[\varphi] \in [Y, Z]$ that $g^{\#} \circ f^{\#}[\varphi] = [\varphi \circ f \circ g] = [\varphi]$ and therefore that $g^{\#} \circ f^{\#} = 1_{[Y, Z]}$; similarly, $f^{\#} \circ g^{\#} = 1_{[X, Z]}$, so both $f^{\#}$, $g^{\#}$ are bijective and, indeed, $f^{\#} = (g^{\#})^{-1}$. The proof for $f_{\#}$, $g_{\#}$ is similar.

The main significance of homotopy type is that homotopic spaces coincide in their classical algebraic invariants (homology, and fundamental groups). More generally, assume that to each space X there is assigned an algebraic structure (group, ring, or field) $G(X)$ and that:

1. Each $f: X \to Y$ induces a homomorphism $f^+: G(X) \to G(Y)$, which depends only on the homotopy class of f.
2. If $f: X \to Y$ and $g: Y \to Z$, then $(g \circ f)^+ = g^+ \circ f^+$.
3. 1_X^+ is an automorphism of $G(X)$.

The argument in **2.2** shows that if $X \simeq Y$, then $G(X)$ is isomorphic to $G(Y)$; that is, $G(X)$ is an invariant of homotopy type. Since any homology theory satisfying the Eilenberg-Steenrod axioms, as well as the fundamental group (and higher homotopy groups), fulfills the requirements (1)–(3), it follows that these invariants alone are incapable of characterizing the homeomorphism type of a given space. A complete set of algebraic invariants that determines whether or not two polytopes belong to the same homotopy type has been found by M. M. Postnikov; however, for a given polytope there is generally no finite algorithm for calculating the entire set of these invariants.

3. Homotopy of Pairs

Let X, Y be two spaces and $A \subset X$, $B \subset Y$ be arbitrary subsets. We say that the *pair* (X, A) is homotopic to the pair (Y, B), and write $(X, A) \simeq (Y, B)$ if there are maps $f: (X, A) \to (Y, B)$ and $g: (Y, B) \to (X, A)$ such that $f \circ g \simeq 1_Y$, keeping the image of B in B during the entire homotopy, and $g \circ f \simeq 1_X$, keeping the image of A in A during the entire homotopy.

With these restrictions on the deformations, the notions of a pair homotopy equivalence $f: (X, A) \simeq (Y, B)$, and of left and right pair homotopy-inverses of a given $g: (X, A) \to (Y, B)$, are defined in obvious fashion, and the analogue of **1.4** is evidently valid.

By taking $A = B = \varnothing$, the pair concepts reduce to those discussed in **1**; in this sense, the concept of homotopic pairs is more general than that of homotopic spaces.

Observe that if $f: (X, A) \simeq (Y, B)$, then both $f: X \simeq Y$ and $f|A: A \simeq B$; however, $X \simeq Y$ and $A \simeq B$ do not imply that $(X,A) \simeq (Y,B)$, that is, that there is a map accomplishing both these equivalences simultaneously.

Ex. 1 Let $X = Y = S^1 \cup \{0\}$, and let $y_0 = (1, 0) \in S^1$. Then $X \simeq Y$ and $\{0\} \simeq \{y_0\}$; yet there is no pair homotopy-equivalence $f: (X, \{0\}) \to (Y, y_0)$, since any $g: (Y, y_0) \to (X, \{0\})$ must send all S^1 to $\{0\}$.

4. Mapping Cylinder

The notion of mapping cylinder, due to J. H. C. Whitehead, is particularly important in homotopy considerations because it provides a link between the concepts "homotopy of maps" and "homotopy of spaces."

4.1 Definition Let $f: X \to Y$ be continuous. The space $(X \times I) \cup_f Y$ obtained by attaching $X \times I$ to Y by the map $(x, 0) \to f(x)$ is called the mapping cylinder of f and is denoted by $C(f)$.

The space $C(f)$ can be regarded as a "cylinder" with $X(= X \times 1)$ at the top and with its base lying in Y, the "generators" of the cylinder being the line segments joining each $x \in X$ to $f(x) \in Y$. Indeed, let $p: (X \times I) + Y \to C(f)$ be the identification map; from VI, **6.3**, follows that $p \mid Y$ embeds Y homeomorphically as a closed subset of $C(f)$ and that $p \mid X \times]0, 1]$ embeds $X \times]0, 1]$ homeomorphically as an open subset of $C(f)$; in particular the map $i: X \to C(f)$ given by $i(x) = p(x, 1)$ is a homeomorphism of X onto the upper face $X \times 1$ of $C(f)$.

We will use the notation $\langle x, t \rangle$ for $p(x, t)$ and $\langle y \rangle$ for $p(y)$; observe (VI, **6.4**) that a pair of continuous maps $\varphi: X \times I \to Z$, $\psi: Y \to Z$ satisfying $\varphi(x, 0) = \psi(f(x))$ for each $x \in X$, determines a continuous $(\varphi, \psi): C(f) \to Z$ by

$$(\varphi, \psi)\langle x, t \rangle = \varphi(x, t),$$
$$(\varphi, \psi)\langle y \rangle = \psi(y),$$

and similarly for homotopies (XV, **2.2**(4)).

The "collapsing map" $\rho: C(f) \to Y$ is defined by

$$\rho\langle x, t \rangle = f(x),$$
$$\rho\langle y \rangle = y,$$

and is evidently continuous. Letting $j: Y \to C(f)$ be the homeomorphism $p \mid Y$, then $j \circ \rho$ is precisely the collapsing of $C(f)$ onto its base.

Thus, with each mapping cylinder, we have the diagram

$$\begin{array}{ccc} & \overset{i}{\nearrow} & C(f) \\ X & & \rho \downarrow \uparrow j \\ & \underset{f}{\searrow} & Y \end{array}$$

where $f = \rho \circ i$.

4.2 The map $\rho: C(f) \to Y$ is a homotopy equivalence, and has j as homotopy inverse. In particular, $C(f)$ belongs to the homotopy type of its base, Y.

Proof: It is evident that $\rho \circ j = 1_Y$; to show $j \circ \rho \simeq 1$, let

$$\varphi: (X \times I) \times I \to C(f)$$

be the continuous map $\varphi[(x, t), \tau] = \langle x, t(1 - \tau) \rangle$ and $\psi: Y \times I \to C(f)$ the continuous map $\psi(y, t) = \langle y \rangle$. Since for each $\tau \in I$ we have $\varphi[(x, 0), \tau] = \langle x, 0 \rangle = \langle f(x) \rangle = \psi[f(x), \tau]$, it follows (XV, **2.2**) that $(\varphi, \psi): C(f) \times I \to C(f)$ is continuous and clearly $(\varphi, \psi): 1 \simeq j \circ \rho$.

The argument in **4.2** shows that in $C(f)$ itself, the base $j(Y)$ is a strong deformation retract of $C(f)$.

Remark: From **4.2** and the diagram for mapping cylinders, we observe that any map $f: X \to Y$ can be factored as $X \xrightarrow{i} C(f) \xrightarrow{\rho} Y$, where i is an injection and ρ is a homotopy equivalence; identifying Y with $j(Y)$, we have $j \circ f \simeq j \circ \rho \circ i \simeq i$, so we can conclude that any map f is homotopic to an *embedding* map into a suitable space belonging to the homotopy type of Y. Thus, for work with homotopy, there is no loss in generality to assume that *all* maps are injective.

The connection between homotopy of maps and of spaces is provided by

4.3 Theorem Let $f_0, f_1: X \to Y$ be homotopic. Then the pair $[C(f_0), X]$ is homotopic to the pair $[C(f_1), X]$.

Proof: Let $\Phi: f_0 \simeq f_1$; for each $x \in X$, the map $t \to \Phi(x, t)$ $(0 \leqslant t \leqslant 1)$ provides an arc in Y joining $f_0(x)$ to $f_1(x)$. We define $F: C(f_0) \to C(f_1)$ to be the identity map on the base, and to map each line segment $\langle x, t\rangle$ in $C(f_0)$ linearly on the arc in $C(f_1)$ consisting of the segment $\langle x, t\rangle$ plus the above arc in the base joining $f_1(x)$ to $f_0(x)$. Precisely, letting $p_i: (X \times I) + Y \to C(f_i)$, $i = 0, 1$, be the identification maps,

$$\begin{aligned} Fp_0(y) &= p_1(y) \\ Fp_0(x, t) &= p_1(x, 2t - 1) && 1 \geqslant t \geqslant \tfrac{1}{2} \\ &= p_1[\Phi(x, 2t)] && \tfrac{1}{2} \geqslant t \geqslant 0. \end{aligned}$$

This is continuous, since

$$Fp_0(x, 0) = p_1[\Phi(x, 0)] = p_1[f_0(x)] = Fp_0[f_0(x)].$$

We define a continuous $G: C(f_1) \to C(f_0)$ similarly by

$$\begin{aligned} Gp_1(y) &= p_0(y) \\ Gp_1(x, t) &= p_0(x, 2t - 1) && 1 \geqslant t \geqslant \tfrac{1}{2} \\ &= p_0[\Phi(x, 1 - 2t)] && \tfrac{1}{2} \geqslant t \geqslant 0. \end{aligned}$$

Thus $G \circ F$ maps $p_0(Y)$ identically, and sends each "generator" $\langle x, t\rangle$ semilinearly on the arc consisting of $\langle x, t\rangle$ plus the arc from $f_0(x)$ to $f_1(x)$ plus the retracing of that arc from $f_1(x)$ to $f_0(x)$. We perform a deformation that uniformly shrinks the retraced arcs over themselves and simultaneously alters the parametrization to the identity, as follows:

$$\begin{aligned} \Delta[\langle y\rangle, s] &= \langle y\rangle \\ \Delta[\langle x, t\rangle, s] &= \left\langle x, \frac{4t - 3s}{4 - 3s}\right\rangle && 1 \geqslant t \geqslant \tfrac{3}{4}s \\ &= \langle \Phi(x, 3s - 4t)\rangle && \tfrac{3}{4}s \geqslant t \geqslant \tfrac{1}{2}s \quad (0 \leqslant s \leqslant 1) \\ &= \langle \Phi(x, 2t)\rangle && \tfrac{1}{2}s \geqslant t \geqslant 0. \end{aligned}$$

This is continuous, since for each $s \in I$ we have

$$\Delta[\langle x, 0\rangle, s] = \langle \Phi(x, 0)\rangle = \langle f_0(x)\rangle = \Delta[\langle f(x_0)\rangle, s].$$

Furthermore, $\Delta: 1 \simeq G \circ F$, and indeed the points of $X(= X \times 1)$ are kept fixed throughout the entire homotopy. In similar fashion, we verify $F \circ G \simeq 1$, and the theorem is proved.

5. Properties of X in $C(f)$

Throughout this section, $f: X \to Y$ will be a fixed given map. It was seen that $j(Y)$ is a deformation retract of $C(f)$; we now determine the properties of $i(X)$ in $C(f)$. To cut down on symbolism, $i(X)$ is identified with X.

5.1 $[C(f) \times 0] \cup [X \times I]$ is always a strong deformation retract of $C(f) \times I$.

Proof: Note that the map $\varphi: C(f) \to I$ given by $\varphi\langle x, t\rangle = 1 - t$, $\varphi\langle y\rangle = 1$ is continuous and shows X is a zero-set in $C(f)$. Since X is a strong deformation retract of the halo $p(X \times]0, 1])$ of x in $C(f)$, the proposition follows from XV, **7.4**.

5.2 X is a retract of $C(f)$ if and only if f has a left homotopy inverse.

Proof: Assume $\Phi: g \circ f \simeq 1_X$ for some $g: Y \to X$. Define a map $r: C(f) \to X$ by

$$r\langle x, t\rangle = \langle \Phi(x, t), 1\rangle$$

$$r\langle y\rangle = \langle g(y), 1\rangle.$$

This is easily verified to be continuous and to be a retraction of $C(f)$ onto X. Conversely, if $r: C(f) \to X$ is a retraction, define $g = r \mid j(Y)$; then $\Phi(x, t) = r\langle x, t\rangle$ is continuous, and $\Phi: g \circ f \simeq 1_X$.

5.3 $C(f)$ is deformable into X if and only if f has a right homotopy inverse.

Proof: Assume $\Phi: f \circ g \simeq 1_Y$ for some $g: Y \to X$. We first note that $Y \subset C(f)$ is then deformable over $C(f)$ into X: deform $1: Y \to Y$ to $f \circ g: Y \to Y$ and then set

$$\Delta(\langle y\rangle, t) = \langle g(y), t\rangle,$$
$$\Delta(\langle y\rangle, 0) = \langle f \circ g(y)\rangle.$$

Thus, preceding this with the deformation retraction of $C(f)$ onto its base, we obtain a deformation of $C(f)$ into X. The converse follows easily, by reversing this argument.

5.4 Theorem X is a deformation retract of $C(f)$ if and only if f is a homotopy equivalence.

Proof: Immediate from **1.4**, **5.2**, **5.3**, and XV, **6.4**.

This result can be regarded as reducing the concept of homotopy type to that of deformation retraction: For, it states in particular that two spaces X, Y belong to the same homotopy type if and only if they can be embedded in a single space Z in such a way that each is a deformation retract of Z.

6. Change of Bases in $C(f)$

Let $f: X \to Y$ be given. In this section, we consider the effect that a change in X, Y (and so also f) has on the homotopy behavior of the pair $[C(f), X]$. We first change the lower base:

6.1 Given $X \xrightarrow{f} Y \xrightarrow{\lambda} Z$, there is an induced continuous map

$$\lambda_+ : (C(f), X) \to (C(\lambda \circ f), X).$$

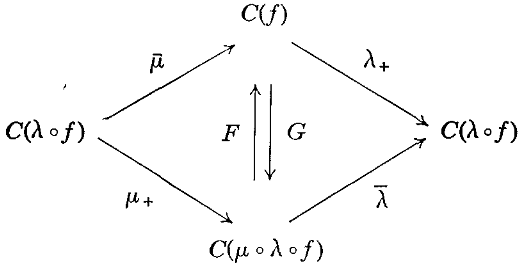
If λ has a left homotopy inverse, then λ_+ has a pair left homotopy inverse; and if λ is a homotopy equivalence, then λ_+ is a pair homotopy equivalence.

Proof: The demonstration is analogous to that in **4.3**. The map λ_+ is defined by mapping each segment $\langle x, t\rangle$ linearly on the segment $\langle x, t\rangle \subset C(\lambda \circ f)$, and sending each $\langle y\rangle$ to $\langle \lambda(y)\rangle$. Assume $\Phi: 1 \simeq \mu \circ \lambda$; define $\bar{\mu}: [C(\lambda \circ f), X] \to [C(f), X]$ by mapping each segment $\langle x, t\rangle$ linearly on the arc [segment $\langle x, t\rangle$ plus arc $\Phi(f(x), t)$], and send each $\langle z\rangle$ to $\langle \mu(z)\rangle$. We have $\bar{\mu} \circ \lambda_+ \simeq 1$, keeping $X(= X \times 1)$ pointwise fixed during the entire deformation, as is shown by

$$\begin{aligned} \Delta[\langle x, t\rangle, s] &= \left\langle x, \frac{2t - s}{2 - s}\right\rangle && (1 \geq t \geq \tfrac{1}{2}s) \\ &= \langle \Phi(f(x), s - 2t)\rangle && (\tfrac{1}{2}s \geq t \geq 0) \end{aligned} \qquad (0 \leq s \leq 1)$$

$$\Delta[\langle x\rangle, s] = \langle \Phi(x, s)\rangle.$$

Assume now that μ is a homotopy inverse of λ. We then have the diagram



where F, G are the homotopy equivalences in **4.3** constructed by using the homotopy $\Phi(f(x), t)$ of f to $\mu \circ \lambda \circ f$. It follows directly from the definitions that $\bar{\mu} = F \circ \mu_+$; consequently,

$$(\bar{\lambda} \circ G) \circ \bar{\mu} = \bar{\lambda} \circ G \circ F \circ \mu_+ \simeq \bar{\lambda} \circ \mu_+ \simeq 1.$$

Since also $\bar{\mu} \circ \lambda_+ \simeq 1$, Theorem **1.4** shows that $\bar{\mu}$ is a homotopy equivalence, with homotopy inverse $\bar{\lambda} \circ G \circ \bar{\mu} \circ \lambda_+ \simeq \lambda_+$, and the proof is complete.

Replacing the top base of $C(f)$, we find

6.2 Given $Z \xrightarrow{\mu} X \xrightarrow{f} Y$, there is an induced continuous map

$$\mu^+ : [C(f \circ \mu), Z] \to [C(f), X].$$

If μ has a right homotopy inverse, then μ^+ has a pair right homotopy inverse; and if μ is a homotopy equivalence, then μ^+ is a pair homotopy equivalence.

Proof: The map μ^+ is determined by sending each segment $\langle z, t \rangle$ linearly onto the segment $\langle \mu(z), t \rangle$ and each $\langle y \rangle$ to $\langle y \rangle$; if $\Phi: 1 \simeq \mu \circ \lambda$, then $\tilde{\lambda}: [C(f), X] \to [C(f \circ \mu), Z]$ is defined by mapping each segment $\langle x, t \rangle$ linearly onto the arc [segment $\langle \lambda(x), t \rangle$ plus arc $f \circ \Phi(x, t)$]. The rest of the proof follows as in **6.1**.

6.3 Theorem (J. H. C. Whitehead) Let $P \xrightarrow{\mu} X \xrightarrow{f} Y \xrightarrow{\lambda} Q$ be given. If μ has a right homotopy inverse and λ has a left homotopy inverse, then there is a map $\varphi: (C(f), X) \to (C(\lambda \circ f \circ \mu), P)$ that has a pair left homotopy inverse; and if λ, μ are homotopy equivalences, φ is a pair homotopy equivalence.

Proof: We have $\mu^+ : (C(f \circ \mu), P) \to (C(f), X)$ and also

$$\lambda_+ : (C(f \circ \mu), P) \to (C(\lambda \circ f \circ \mu), P).$$

If r is a right homotopy inverse of μ, then defining $\tilde{r}$ as in the proof of **6.2**, we find that $\varphi = \lambda_+ \circ \tilde{r}$ has the required properties.

Remark: A pair (Y, B) is said to *dominate* a pair (X, A) if there exists an $f: (X, A) \to (Y, B)$ that has a left homotopy inverse. With this terminology, **6.3** becomes: If P dominates X and Q dominates Y, then the pair $[C(\lambda \circ f \circ \mu), P]$ dominates $[C(f), X]$.

We use **6.1** in another way to derive

6.4 Theorem (S. Eilenberg, M. Shiffman) Let $f: X \to Y$ be a homotopy equivalence, and attach V^n to X by $\alpha: S^{n-1} \to X$. Then, by attaching V^n to Y with $\beta = f \circ \alpha: S^{n-1} \to Y$, we have $X \cup_\alpha V^n \simeq Y \cup_\beta V^n$

Proof: By **6.1** we have that $f_+ : [C(\alpha), S^{n-1}] \to [C(f \circ \alpha), S^{n-1}]$ is a homotopy equivalence. Let g be a homotopy inverse of f; then $f_+ \circ \bar{g} \simeq 1$ and $\bar{g} \circ f_+ \simeq 1$, keeping $S^{n-1} \times 1$ pointwise fixed. Thus, attaching V^n to each mapping cylinder by the identity map $S^{n-1} \to S^{n-1} \times 1$, we still have the resulting spaces homotopic. Since these new spaces are obtained by attaching $(V^n \times 1) \cup (S^{n-1} \times I)$ to X and Y by $\alpha \mid S^{n-1} \times 0$ and $f \circ \alpha \mid S^{n-1} \times 0$, respectively, and since $[(V^n \times 1) \cup (S^{n-1} \times I), S^{n-1} \times 0]$ is homeomorphic to (V^n, S^{n-1}), the proof is complete.

Ex. 1 Theorem **6.4** has many applications. To indicate only one of them, we construct models for the homotopy types of certain spaces (the abstract situation we describe occurs in the Morse critical point theory). Assume that a space X is decomposed as $A_0 \subset B_1 \subset A_1 \subset B_2 \subset \cdots \subset A_n = X$, where (1) A_0 is contractible, and for $i \geqslant 1$, both (2) $B_i \simeq A_{i-1}$ and (3) $(\overline{A_i - B_i}, \mathrm{Fr}(B_i)) \cong (V^{n_i}, S^{n_i - 1})$ for each i. Then, by (1), A_0 is homotopic to a space M_0 consisting of a single point; by (2), we have $B_1 \simeq M_0$; by (3), we can attach V^{n_1} to M_0 and get a space $M_1 \simeq A_1$. By proceeding in this manner, the space X is seen to belong to the homotopy type of n balls of varying dimensions suitably attached to each other.

Problems

Section 1

1. Let $Y \simeq Z$ and let X be locally compact. Prove $Y^X \simeq Z^X$.
2. Let $X \simeq Y$ and let X be locally compact. Prove $Z^Y \simeq Z^X$.
3. For any space X, show that $X \simeq X \times I \simeq X \times E^n \simeq X \times I^\infty$.

4. Give an example to show that $(A \subset X) \wedge (A \simeq X)$ does not imply that A is a deformation retract of X, even though A is closed.
5. Prove: Being a homotopy equivalence is a homotopy class invariant; that is, if $f: X \simeq Y$ and $g \simeq f$, then $g: X \simeq Y$.
6. Let $f: X \simeq Y$ and $g: Y \simeq Z$ have homotopy inverses f', g', respectively. Show that $g \circ f$ has $f' \circ g'$ as homotopy inverse.
7. A space Y is called weakly locally contractible if each $y \in Y$ has a nbd U deformable over Y to y. Show that if $X \simeq Y$ and if Y is weakly locally contractible, then so also is X.

Section 5

1. Let $f: X \to Y$, and $h: X \to Z$. Prove: There exists a $g: Y \to Z$ such that $g \circ f \simeq h$ if and only if h can be extended to a map $C(f) \to Z$.
2. Let $g: Y \to Z$ and $h: X \to Z$. Prove: There exists an $f: X \to Y$ such that $g \circ f \simeq h$ if and only if h is homotopic over $C(g)$ to a map of X into Y.
3. Let $f: X \to Y$ be extendable to a map $F: TX \to TY$. Let $Tf: TX \to TY$. Prove that $Tf \simeq F$ rel X.

Section 6

1. Let the diagram

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\psi} \\ Z & \xrightarrow[g]{} & W \end{array}$$

be commutative. Show that there exists a continuous $\lambda: C(f) \to C(g)$ such that the diagram

$$\begin{array}{ccccc} X & \xrightarrow{i} & C(f) & \xrightarrow{\rho} & Y \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\lambda} & & \downarrow{\scriptstyle\psi} \\ Z & \xrightarrow[i']{} & C(g) & \xrightarrow[\rho']{} & W \end{array}$$

is commutative in each square.
2. Let A be a strong deformation retract of X, and let Y be an arbitrary space. Prove:
   a. If X is attached to Y by a continuous $f: A \to Y$, then Y is a strong deformation retract of $X \cup_f Y$.

   Using this result, show that
   b. If $X \times 1 \cup A \times I$ is attached to Y by $(a, 0) \to f(a)$, then
$$[X \times 1 \cup A \times I] \cup_f Y \simeq X \cup_f Y.$$

   (One can use this result to obtain an evident extension of the model construction described in **6**, Ex. 1.)

Path Spaces; H-Spaces

XIX

In this chapter, we consider structures consisting of a space Y together with a "multiplication" $\Gamma: Y \times Y \to Y$ that has a simple homotopy property. Such structures include, as special cases, both topological groups and path spaces.

1. Path Spaces

Recall that [V, **5**] a path in Y is a continuous *mapping* of the unit interval I into Y, rather than a continuous *image* of I in Y; that is, a path in Y is an element of Y^I. The path $\alpha \in Y^I$ is said to start at the point $\alpha(0) \in Y$ and to end at the point $\alpha(1) \in Y$; a closed path, or loop, at $y_0 \in Y$ is a path starting and ending at y_0. The constant path $\eta(I) = y_0$ is called the null path, or loop, at y_0.

1.1 Definition The *inverse* of a path $\alpha \in Y^I$ is the path $\alpha^{-1} \in Y^I$ defined by the rule $\alpha^{-1}(t) = \alpha(1 - t)$, $(0 \leqslant t \leqslant 1)$. The product of two paths α, β, written $\alpha * \beta$, is defined only in case $\alpha(1) = \beta(0)$, and is the path

$$\alpha * \beta(t) = \begin{cases} \alpha(2t) & 0 \leqslant t \leqslant \frac{1}{2} \\ \beta(2t - 1) & \frac{1}{2} \leqslant t \leqslant 1. \end{cases}$$

We shall always take Y^I with the c-topology. Given $a, b \in Y$, the subspace of Y^I consisting of all paths starting at a and ending at b is denoted by $\Omega(Y; a, b)$; it is clear that $\Omega(Y; a, b)$ is nonempty if and only if a, b belong to a common path component of Y. In case $a = b$, $\Omega(Y; a, a)$ is written simply $\Omega(Y; a)$ and is called the loop space of Y based at a.

One advantage of the c-topology in Y^I is that with no restrictions on Y, the path operations of inverse and product are continuous, when considered as functions defined on the path spaces; we now establish this and several other results.

1.2 The map $\alpha \to \alpha^{-1}$ of $\Omega(Y; a, b) \to \Omega(Y; b, a)$ is a homeomorphism.

Proof: The map is clearly bijective. Now let $p: I \to I$ be the homeomorphism $p(t) = 1 - t$; by XII, **2.1**, the induced map $p^+: Y^I \to Y^I$ is continuous and is precisely the map $\alpha \to \alpha^{-1}$. Similarly, the map $\alpha^{-1} \to \alpha$ is continuous.

The following proposition is very useful.

1.3 Let $p: I \to I$ be continuous and such that $p(0) = 0$, $p(1) = 1$ (p is called a "parameter transformation"). Then the map $\alpha \to \alpha \circ p$ of $\Omega(Y; a, b)$ in itself is homotopic to the identity map.

Proof: Note first that $p \simeq 1 \operatorname{rel} \operatorname{Fr}(I)$, as $p_s(t) = (1 - s)p(t) + st$ $(0 \leqslant s \leqslant 1)$ shows. By XV, **3.2**, it follows that the induced map $p^+: Y^I \to Y^I$ is homotopic to the identity map, and since $p_s^+ \mid \Omega(Y; a, b)$ maps $\Omega(Y; a, b)$ in itself, the assertion follows.

For the product operation,

1.4 (1). The mapping $(\alpha, \beta) \to \alpha * \beta$ of

$$\Omega(Y; a, b) \times \Omega(Y; b, c) \to \Omega(Y; a, c)$$

is continuous.

(2). If $\eta \in \Omega(Y; b)$ is the null path, the map $\alpha \to \alpha * \eta$ of $\Omega(Y; a, b)$ in itself is homotopic to the identity map, and so also is the map $\beta \to \eta * \beta$ of $\Omega(Y; b, c)$ in itself.

Proof: Ad (1). This will follow from XII, **3.1**, by showing that the associated map $\Omega(Y; a, b) \times \Omega(Y; b, c) \times I \to Y$ is continuous. This latter map decomposes into

$$\begin{aligned} &(\alpha, \beta, t) \to (\alpha, t) \to (\alpha, 2t) \to \alpha(2t) && 0 \leqslant t \leqslant \tfrac{1}{2} \\ &(\alpha, \beta, t) \to (\beta, t) \to (\beta, 2t - 1) \to \beta(2t - 1) && \tfrac{1}{2} \leqslant t \leqslant 1. \end{aligned}$$

In each case, the first map is a projection and therefore continuous; the second maps are clearly continuous, and the last maps, being evaluation maps, are also continuous (*cf.* XII, **2.4**). Since the two maps coincide on $[\Omega(Y; a, b) \times \Omega(Y; b, c)] \times \frac{1}{2}$, the result follows.

Ad (2). Let p be the parameter transformation $p(t) = \min[1; 2t]$; since $\alpha \to \alpha * \eta$ is the map $\alpha \to \alpha \circ p$, the result follows from **1.3**. The second part is proved similarly.

For associativity and inversion we have

1.5 (1). The maps $(\alpha, \beta, \gamma) \to (\alpha * \beta) * \gamma$ and $(\alpha, \beta, \gamma) \to \alpha * (\beta * \gamma)$ of $\Omega(Y; a, b) \times \Omega(Y; b, c) \times \Omega(Y; c, d) \to \Omega(Y; a, d)$ are homotopic.

(2). The map $\alpha \to \alpha * \alpha^{-1}$ of $\Omega(Y; a, b) \to \Omega(Y; a)$ is nullhomotopic, and so also is the map $\alpha \to \alpha^{-1} * \alpha$.

Proof: Ad (1). Let $R(\alpha, \beta, \gamma) = \alpha * (\beta * \gamma)$ and observe that there is a parameter transformation p such that $(\alpha * \beta) * \gamma = [\alpha * (\beta * \gamma)] \circ p$ for all α, β, γ. From **1.3** we find that $p^+ \circ R \simeq R$, and the proof is complete.

Ad (2). Define a homotopy

$$p_s(t) = \begin{cases} 2t(1-s) & 0 \leqslant t \leqslant \frac{1}{2} \\ 2(1-t)(1-s) & \frac{1}{2} \leqslant t \leqslant 1 \end{cases} \qquad 0 \leqslant s \leqslant 1,$$

showing that $p_0 \simeq 0$, and note that $\alpha \to \alpha * \alpha^{-1}$ is the map $p_0^+ : Y^I \to Y^I$. By XV, **3.2**, p_s^+ shows $p_0^+ \simeq 0$; since $p_s(0) = p_s(1) = 0$ for each s, it follows that $p_s^+ \mid \Omega(Y; a, b)$ maps $\Omega(Y; a, b)$ into $\Omega(Y; a)$ for each s, so the proof is complete.

The path space $\Omega(Y; a, b)$ is in general not path-connected. We call two paths $\alpha, \beta \in \Omega(Y; a, b)$ *equivalent* (written: $\alpha \sim \beta$) if they belong to the same path component. Stated directly in terms of the maps α, β:

1.6 $\alpha \sim \beta$ if and only if $\alpha \simeq \beta \operatorname{rel} \operatorname{Fr}(I)$; that is, if and only if α can be deformed to β without ever moving the end points.

Proof: If $I \to \Omega(Y; a, b)$ is a path joining α to β, the associated map shows $\alpha \simeq \beta \operatorname{rel} \operatorname{Fr}(I)$; conversely, a homotopy rel $\operatorname{Fr}(I)$ gives a path joining α to β (cf. XV **3.1**).

Since homotopic maps must send any given point to a common path component of the range space, the results **1.4** and **1.5** imply that $(\alpha * \beta) * \gamma \sim \alpha * (\beta * \gamma)$, that $\alpha * \alpha^{-1} \sim \eta$, that $\alpha^{-1} * \alpha \sim \eta$, that $\alpha * \eta \sim \alpha$, and that $\eta * \alpha \sim \alpha$; it is trivial to verify that if $\alpha \sim \bar{\alpha}$ and $\beta \sim \bar{\beta}$, then $\alpha * \beta \sim \bar{\alpha} * \bar{\beta}$.

Let $\alpha \in \Omega(Y; a, b)$ be a fixed path. For any given $y \in Y$, the path α induces a "transition map" $\alpha_R: \Omega(Y; y, a) \to \Omega(Y; y, b)$ by defining $\alpha_R(\gamma) = \gamma * \alpha$. Similarly, α induces a transition map $\alpha_L: \Omega(Y; b, y) \to \Omega(Y; a, y)$ by $\alpha_L(\gamma) = \alpha * \gamma$. According to **1.4**, transition maps are continuous; they have the following important properties

1.7 (a). Each α_R [resp. α_L] is a homotopy equivalence with homotopy inverse $(\alpha^{-1})_R$ [resp. $(\alpha^{-1})_L$].
(b). $\alpha_R \simeq \beta_R$ [resp. $\alpha_L \simeq \beta_L$] if and only if $\alpha \sim \beta$.

Proof: We prove this for the maps α_R; that for the α_L is similar.

Ad (b). Let $\Phi: I \to \Omega(Y; a, b)$ be a path joining α to β; by **1.4**(1), the map $(\gamma, t) \to \gamma * [\Phi(t)]$ is continuous and shows that $\alpha_R \simeq \beta_R$. Conversely, assume $\alpha_R \simeq \beta_R$, or even less, that there is some γ such that $\alpha_R(\gamma)$ and $\beta_R(\gamma)$ belong to a common path component. Then $\gamma * \alpha \sim \gamma * \beta$ therefore $\gamma^{-1} * (\gamma * \alpha) \sim \gamma^{-1} * (\gamma * \beta)$, and consequently $\alpha \sim \beta$.

Ad (a). Note that $(\alpha^{-1})_R \circ \alpha_R = (\alpha * \alpha^{-1})_R$; since $\alpha * \alpha^{-1} \sim \eta$, it follows from **1.4**(2) and from what we have just proved that $(\alpha^{-1})_R \circ \alpha_R \simeq 1$. Similarly, $\alpha_R \circ (\alpha^{-1})_R \simeq 1$, which completes the proof.

In what follows, we are concerned primarily with loop spaces. Each path $\alpha \in \Omega(Y; a, b)$ induces a transition map $\alpha^+: \Omega(Y; a) \to \Omega(Y; b)$ given by $\alpha^+(\lambda) = \alpha^{-1} * (\lambda * \alpha)$; this is continuous, and noting that $\alpha^+ = (\alpha^{-1})_L \circ \alpha_R$, it follows from **1.7** that

1.8 (1). Each α^+ is a homotopy equivalence with homotopy inverse $(\alpha^{-1})^+$.
(2). $\alpha^+ \simeq \beta^+$ if and only if $\alpha \sim \beta$.

2. *H*-Structures

The loop spaces of the preceding section are topological spaces in which a continuous multiplication, the composition of two paths into one, is defined. A topological group is another example of a space in which a continuous composition law is defined. We now consider a class of such structures that contains both the loop spaces and the topological groups.

2.1 Definition An *H*-structure ("Hopf", or "Homotopy" structure) is a couple (Y, Γ) consisting of a space Y and a continuous $\Gamma: Y \times Y \to Y$ which has the following property: There exists a point $e \in Y$ such that maps $y \to \Gamma(y, e)$ and $y \to \Gamma(e, y)$ are both homotopic to the identity map of Y.

The map Γ is called the composition law of the H-structure (Y, Γ), and Y is the carrier of the H-structure; $\Gamma(a, b)$ is written $a \cdot b$. When it is not necessary to specify Γ explicitly, we say simply that "Y is an H-structure".

Ex. 1 Every topological group is an H-structure.

Ex. 2 Given any Y, the loop space $\Omega(Y; y_0)$, with the composition $\Gamma(\alpha, \beta) = \alpha * \beta$, is an H-structure (*cf.* **1.4**). Γ is called the natural composition law in $\Omega(Y; y_0)$.

Ex. 3 Let Y be any contractible space, and $\Gamma\colon Y \times Y \to Y$ any continuous map. Then (Y, Γ) is an H-structure, since any two maps of Y in itself are homotopic. In particular, the unit interval I with the composition law $\Gamma(t_1, t_2) = |t_1 - t_2|$ is an H-structure. Observe (1) that the composition law is not required to be associative, and (2) that a given space may carry many H-structures.

Ex. 4 Let Y be any discrete space. Selecting any $e \in Y$, and defining Γ in any way compatible with $\Gamma(e, y) = \Gamma(y, e) = y$, we have that (Y, Γ) is an H-structure.

The point e in the definition is not unique in general:

2.2 Let $P = \{e \mid$ both of the maps $y \to \Gamma(y, e)$ and $y \to \Gamma(e, y)$ are homotopic to $1_Y\}$. Then P is a path component of Y, called the principal component of the H-structure (Y, Γ).

Proof: Choose any $e_0 \in P$ and let C be the path component of e_0. Then (1) $C \subset P$: Let $e \in C$; choosing a path α joining e to e_0, the continuous map $(y, t) \to \Gamma(y, \alpha(t))$ shows that $y \to \Gamma(y, e)$ is homotopic to $y \to \Gamma(y, e_0)$, and therefore also homotopic to 1_Y. A similar argument shows $y \to \Gamma(e, y)$ to be homotopic to 1_Y, and therefore $e \in P$. (2) $P \subset C$: Let $e \in P$. Then for each $y \in Y$, the two points $\Gamma(y, e)$, y as well as the two points y, $\Gamma(e_0, y)$ must belong to common path components (the homotopy of $y \to \Gamma(y, e)$ to the identity provides a path joining $\Gamma(y, e)$ to y). Thus $\Gamma(e_0, e)$, e_0 and also e, $\Gamma(e_0, e)$ belong to common path components, and therefore $e \in C$.

An H-structure (Y, Γ) induces H-structures carried by suitable spaces related to Y; we consider here three important H-structures that are so induced.

2.3 Let (Y, Γ) be an H-structure. Then its principal component P, with composition law $\Gamma \mid P \times P$, is an H-structure, called the induced H-structure in P.

Proof: This is an immediate consequence of **2.2**, since the composition of two elements belonging to the principal component is again contained in the principal component.

Define Comp Y to be the *discrete* space of path components of Y. We shall denote the path component containing $y \in Y$ by $[y]$.

2.4 Let (Y, Γ) be an H-structure. Then Comp Y, with the composition law $([x], [y]) \to [\Gamma(x, y)]$ is an H-structure, called the induced H-structure in Comp Y.

Proof: Note first that the composition law is a well-defined function: as in (1) of the proof of **2.2**, the path component containing $x \cdot y$ is easily seen to depend only on the path component of x and that of y. Because $[e] \cdot [y] = [y] \cdot [e] = [y]$, the proposition is proved.

Ex. 5 For the natural H-structure $\Omega(Y; y_0)$, the induced H-structure Comp $\Omega(Y; y_0)$ is actually a (generally non-Abelian) group called the fundamental group of Y at y_0. Although this follows immediately from the work in **I**, we prefer to obtain it as a consequence of general results proved later.

2.5 Let (Y, Γ) be an H-structure. Then for each locally compact space X, the space Y^X with composition law $(f, g) \to \Gamma \circ (f \times g)$ is an H-structure, called the induced H-structure in Y^X.

Proof: We first show the composition map $\hat{\Phi}$: $Y^X \times Y^X \to Y^X$ to be continuous. Because X is locally compact, the continuity of $\hat{\Phi}$ will follow once we show that the associated map Φ is continuous. Observe that Φ can be factored so that $(f, g, x) \to (f(x), g(x)) \to \Gamma(f(x), g(x))$. In this factorization, the last map is evidently continuous, and the continuity of the first one follows by noting that its projection on each factor of $Y \times Y$ is an evaluation map that, for locally compact X, is continuous. Thus $\hat{\Phi}$ is continuous, as required. Now let $\bar{e}$ be the constant map $\bar{e}(x) = e$ and let $\Gamma_e(y) = \Gamma(y, e)$; then $\hat{\Phi}(f, \bar{e}) = \Gamma_e \circ f$, and because $\Gamma_e \simeq 1_Y$, it follows from XV, **3.2** that $f \to f \cdot \bar{e}$ is homotopic to the identity map of Y^X. Similarly, $g \to \bar{e} \cdot g$ is homotopic to the identity map, and the proof is complete.

3. H-Homomorphisms

Let X, Y be two H-structures. A continuous map $f: X \to Y$ is called an H-homomorphism whenever the two maps $(x, x') \to f(x) \cdot f(x')$ and $(x, x') \to f(x \cdot x')$ of $X \times X$ into Y are homotopic. It is simple to verify that if $f: X \to Y$ and $g: Y \to Z$ are H-homomorphisms, so also is $g \circ f$. Whenever X and Y are discrete, an H-homomorphism is evidently a homomorphism in the usual sense of algebra, and the prefix "H-" will be omitted.

An H-homomorphism $f\colon X \to Y$ is called an H-isomorphism if there exists an H-homomorphism $g\colon Y \to X$ such that both $g \circ f \simeq 1_X$ and $f \circ g \simeq 1_Y$; in this event, the H-structures are called H-isomorphic. Note that an H-isomorphism is necessarily a homotopy equivalence; the converse need not be true. Observe also that the concepts of H-homomorphism (H-isomorphism) are homotopy class invariants: If one map in a homotopy class is an H-homomorphism (H-isomorphism) so also are all maps in that class.

Ex. 1 Let $a, b \in Y$, and let α be any path from a to b. The transition map $\alpha^+ : \Omega(Y; a) \to \Omega(Y; b)$ is evidently an H-homomorphism, so by **1.8**, α^+ is an H-isomorphism.

Ex. 2 Let $f\colon (X, a) \to (Y, b)$ be continuous. Then f induces an H-homomorphism $f_+ : \Omega(X; a) \to \Omega(Y; b)$ by setting $f_+(\alpha) = f \circ \alpha$; the continuity of f_+ follows from XII, **2.1**, and verification that f_+ is an H-homomorphism is trivial. If f is a homotopy equivalence, then f_+ is an H-isomorphism. For, let g be a homotopy inverse of f, and let $\Phi\colon g \circ f \simeq 1$. For each s, $0 \leqslant s \leqslant 1$, let α_s be the path $t \to \Phi[a, s + (1 - s)t]$ and define $G\colon \Omega(Y; b) \to \Omega(X; a)$ by

$$G(\mu) = [\alpha_1]^+ \circ g_+(\mu);$$

G is continuous and an H-homomorphism. For each $\lambda \in \Omega(X; a)$, let

$$\lambda_s(t) = \Phi(\lambda(t), s)$$

and define a homotopy $\Psi\colon \Omega(X; a) \times I \to \Omega(X; a)$ by $(\lambda, s) \to \alpha_s^{-1} * \lambda_s * \alpha_s$; then Ψ is continuous, and $\Psi(\lambda, 0) = G \circ f_+(\lambda)$, so using **1.4**(2), we conclude that $G \circ f_+ \simeq 1$. In similar fashion, $f_+ \circ G \simeq 1$, and the assertion follows.

Ex. 3 Let Y be an H-structure, let X, Z be locally compact spaces, and $f\colon X \to Z$ be continuous. The induced continuous $f^+ \colon Y^Z \to Y^X$ is an H-homomorphism of the induced H-structures in the function spaces. Furthermore, if f is a homotopy equivalence, then f^+ is an H-isomorphism.

For any two spaces X, Y, a continuous $f\colon X \to Y$ induces a map f_C: Comp $X \to$ Comp Y by $f_C[x] = [f(x)]$.

3.1 If X, Y are H-structures, and $f\colon X \to Y$ is an H-homomorphism (H-isomorphism), then f_C is also a homomorphism (isomorphism) of the induced H-structures in Comp X, Comp Y. Furthermore, homotopic H-homomorphisms induce the *same* homomorphism.

The trivial proof is omitted.

Let (Y, Γ), (Y, Γ') be two H-structures on the same space. Note that the identity map 1_Y is an H-isomorphism if and only if $\Gamma \simeq \Gamma'$; in this case we say Γ, Γ' are *equivalent* H-structures on Y. It is obvious that

3.2 Equivalent H-structures on Y induce the *same* H-structure in Comp Y.

Ex. 4 All H-structures on a contractible space are equivalent.

Ex. 5 Distinct H-structures on a discrete space (with more than one point) are never equivalent. It is true, but much more difficult to prove, that there exist *path-connected* spaces having at least two nonequivalent H-structures.

4. H-Spaces

It is not true that every space can carry an H-structure; spaces that do admit an H-structure are called H-spaces.

Ex. 1 Every group space, and every loop space, is an H-space; each contractible space is an H-space.

Ex. 2 S^0, S^1, and S^3 are group spaces (integers mod 2, complex numbers of norm 1, and quaternions of norm 1, respectively). S^7 is an H-space (Cayley numbers of norm 1), but not a group space. It has been proved by J. F. Adams that, among the spheres, S^0, S^1, S^3, and S^7 are the only ones that are H-spaces.

Ex. 3 Any space with a non-abelian fundamental group (a figure 8, for example) cannot be an H-space, as we shall see in **8.4.**

The notion of H-space is significant only for path-connected spaces, because

4.1 In order that Y be an H-space, it is necessary and, if Y is locally path-connected, also sufficient, that one path component of Y be an H-space.

Proof: Necessity follows from **2.2.** Sufficiency: If P is a path component carrying an H-structure, define a composition in Y by preserving that on P, setting $y \cdot y' = y' \cdot y = p_0 \in P$ whenever $y, y' \bar{\in} P$, and $p \cdot y = y \cdot p = y$ whenever $y \bar{\in} P$, $p \in P$. This composition is continuous, since in locally path-connected spaces each path component is both open and closed, and defines an H-structure on Y.

The concept of H-space is also a homotopy type invariant:

4.2 Let (Y, Γ) be an H-structure, X any space, and $f: X \to Y$ a homotopy equivalence with homotopy inverse g. Define a composition law in X by $\Gamma_0(x, x') = g \circ \Gamma(f(x), f(x'))$. Then (X, Γ_0) is an H-structure and both f, g are H-isomorphisms.

Proof: We first show that Γ_0 determines an H-structure on X. Note first that from $f \circ g \simeq 1_Y$, it follows that $f(X)$ contains points from each path component of Y, so we can choose an $e \in X$ such that $f(e)$ belongs to the principal component of Y. Define $h(y) = y \cdot f(e)$; then

(**2.2**) we have $h \simeq 1_Y$ and the map $x \to \Gamma_0(x, e)$ is precisely the map $g \circ h \circ f \simeq g \circ f \simeq 1_X$. Similarly, $x \to \Gamma_0(e, x)$ is homotopic to 1_X and (X, Γ_0) is an H-structure. That f, g are H-homomorphisms follows at once because $f \circ g \simeq 1$.

Note that the H-structure imposed on X depends on the particular homotopy equivalences that are used: If f', g' is another pair, and if $f \simeq f'$ (so that $g \simeq g'$ also), then the H-structures imposed on X by the pairs (f, g) and (f', g') are easily seen to be equivalent.

Observe that to show (X, Γ_0) is an H-structure, we require only the hypotheses (1) $g \circ f \simeq 1$, and (2) $f(X)$ contains points of the principal component of Y; in this case, however, it may not be true that f and g are H-homomorphisms. Since the hypothesis (2) is always satisfied whenever Y is path-connected, we have

4.3 Every space dominated by a path-connected H-space is also an H-space.

Ex. 4 A retract of a path-connected H-space Y is also an H-space, since Y dominates each of its retracts.

Ex. 5 $\Omega(Y; a, b)$ is an H-space: for, by using **1.7**, we find that $\Omega(Y; a, b) \simeq \Omega(Y; a)$.

Ex. 6 A homotopy type containing a group (or H-) space is composed entirely of H-spaces. The homotopy type of S^7 consists only of H-spaces and H. Samelson has shown that it contains no group space.

5. Units

By a unit of an H-structure Y is meant an element e such that $e \cdot y = y \cdot e = y$ for all $y \in Y$. There can be no more than one unit, since for any two units e, e', we would have $e = e \cdot e' = e'$. Not every H-structure has a unit: For example, the natural H-structure $\Omega(Y; y_0)$ has no unit. However, it is trivial to verify

5.1 Any H-structure on a discrete space has a unit. In particular, the induced H-structure on Comp Y always has a unit.

The requirement that an H-structure have a unit is not an essential restriction. To see this, we need

5.2 Lemma Let Y be an H-structure, let e be a point in the principal component, and let R be a homotopy of the map $y \to e \cdot y$ to the identity, $R(y, 0) = e \cdot y$. Then there exists a homotopy L of the map $y \to y \cdot e$ to the identity such that the two paths $L_e(t) = L(e, t)$ and $R_e(t) = R(e, t)$ from $e \cdot e$ to e are equivalent.

Proof: Let l be any homotopy of $y \to y \cdot e$ to 1_Y, where $l(y, 0) = y \cdot e$. We first show that there is a loop α at e such that

$$R_e \sim (e \cdot \alpha) * l_e.$$

Let $\bar{\alpha} = R_e * l_e^{-1}$. Regarding $\bar{\alpha}$ as a map $\bar{\alpha}\colon (S^1, p_0) \to (Y, e \cdot e)$ and noting that $(S^1 \times 0) \cup (p_0 \times I)$ is a retract of $S^1 \times I$, it follows that we can deform $\bar{\alpha}$ to a map $\alpha\colon (S^1, p_0) \to (Y, e)$ in such a way that the image of p_0 traces R_e during this deformation. The homotopy $R(\alpha(t), 1 - s)$, $0 \leqslant s \leqslant 1$, transforms α to the path $e \cdot \alpha$. During these two deformations, the image of p_0 traces the path $R_e * R_e^{-1}$, which is equivalent to a null path, so that (XV, **8.3**) $\bar{\alpha}$ can be deformed to $e \cdot \alpha$ without ever changing the image of p_0; that is, $\bar{\alpha} \sim e \cdot \alpha$. Thus

$$R_e \sim R_e * l_e^{-1} * l_e \sim \bar{\alpha} * l_e \sim (e \cdot \alpha) * l_e.$$

Now define

$$L(y, t) = \begin{cases} y \cdot \alpha(2t) & 0 \leqslant t \leqslant \frac{1}{2} \\ l(y, 2t - 1) & \frac{1}{2} \leqslant t \leqslant 1. \end{cases}$$

For $t = \frac{1}{2}$, we have $y \cdot \alpha(1) = y \cdot e = l(y, 0)$, so that L is continuous. L is evidently a homotopy of the map $y \to y \cdot e$ to 1_Y, and $L_e = (e \cdot \alpha) * l_e \sim R_e$, completing the proof.

5.3 Theorem Let Y be an H-structure and e any point of the principal component. Let Y' be the subspace $(Y \times 0) \cup (e \times I) \subset Y \times I$. Then:

(1). The H-structure on $Y(= Y \times 0)$ can be enlarged to an H-structure with unit on Y'.
(2). The deformation retraction of Y' collapsing I to e is an H-isomorphism.

Proof: Ad (1). Let $y \in Y$, $s, t \in I$. To extend the H-structure, we need define only $y \cdot s$, $t \cdot y$, and $t \cdot s$. Using the homotopies R, L of the lemma, define $y \cdot s = L(y, s)$, $t \cdot y = R(y, t)$. It remains to define $t \cdot s$. For this, consider the map $\Phi\colon \mathrm{Fr}(I^2) \to Y'$ given by $\Phi(t, 1) = \Phi(1, t) = t$, $\Phi(0, s) = L(e, s)$, $\Phi(t, 0) = R(e, t)$. On $\mathrm{Fr}(I^2)$, this map is equivalent to $L_e * R_e^{-1}$, which is nullhomotopic; consequently, Φ is extendable to a continuous $\Psi\colon I^2 \to Y'$. We define $t \cdot s = \Psi(t, s)$. It is easy to verify that this composition determines an H-structure on Y' and has $t = 1$ as unit.

Ad (2). It is clear that the collapsing map $r\colon Y' \to Y$ is a homotopy equivalence with homotopy inverse the inclusion $i\colon Y \to Y'$. Since i is evidently an H-homomorphism, we need prove only that r is one also.

A routine calculation shows that the map $\Delta\colon Y' \times Y' \times I \to Y'$, given by

$$\begin{aligned} \Delta(y, \bar{y}; \tau) &= y \cdot \bar{y}, \\ \Delta(y, s; \tau) &= L(y, s\tau), \\ \Delta(t, y; \tau) &= R(y, t\tau), \\ \Delta(t, s; \tau) &= r \circ \Psi(t\tau, s\tau), \end{aligned}$$

is consistently defined and therefore is continuous. It deforms $r(y') \cdot r(y'')$ to $r(y' \cdot y'')$.

Observe that the H-structure on Y' depends on the homotopies L, R that are used, and also on Ψ; however, **5.3**(2) assures that the H-structure on Y' is, up to an H-isomorphism, completely determined by the given H-structure on Y.

6. Inversion

An H-structure Y is said to admit inversion if there exists a continuous $\varphi\colon Y \to Y$ having the following property: the map $y \to y \cdot \varphi(y)$ is homotopic to a constant map that sends Y to a point of the principal component. An H-structure together with such a map is called an H-structure with inversion, and we write $\varphi(y) = y^{-1}$.

Ex. 1 If an H-structure on a discrete space admits inversion, then $y \cdot \varphi(y) = e$ for every y, where e is the unit. It need not be true that $\varphi(y) \cdot y = e$, as simple examples show.

Ex. 2 Every H-structure on a contractible space admits inversion.

Ex. 3 The natural H-structure $\Omega(Y; y_0)$ admits inversion, since $\alpha \to \alpha * \alpha^{-1}$ is nullhomotopic.

Ex. 4 If Y is an H-structure with inversion, then for any locally compact X, the induced H-structure on Y^X also admits inversion: we set $\Phi(f) = \varphi \circ f$.

6.1 If an H-structure Y admits inversion, then the induced H-structure on Comp Y also admits inversion.

Proof: We need only set $\Phi[x] = [\varphi(x)]$.

The notion of an H-homomorphism f of H-structures is extended to H-structures with inversion by including the requirement that the two maps $y \to f(y^{-1})$ and $y \to (f(y))^{-1}$ be homotopic.

The possession of H-structures with inversion is a homotopy class invariant:

6.2 Let X, Y be H-structures, and $f\colon X \to Y$ an H-isomorphism. If the H-structure Y admits inversion, then inversion can be defined for the H-structure X so that f is an H-isomorphism in the extended sense given above.

Proof: If $g: Y \to X$ is an H-isomorphism that is a homotopy inverse for f, define $x^{-1} = g(f(x)^{-1})$; the completion of the proof is left for the reader.

7. Associativity

An H-structure Y is H-associative if the two maps $(x, y, z) \to x \cdot (y \cdot z)$ and $(x, y, z) \to (x \cdot y) \cdot z$ of $Y \times Y \times Y$ into Y are homotopic.

Ex. 1 All H-structures on a contractible space are H-associative.
Ex. 2 The natural H-structure $\Omega(Y; y_0)$ is H-associative (*cf.* **1.5**).
Ex. 3 If Y is H-associative, the induced H-structure in Y^X is also H-associative.

7.1 (a). If an H-structure Y is H-associative, then the induced H-structure on Comp Y is associative.
(b). If two H-structures are H-isomorphic, and one is H-associative, so also is the other.

The simple proofs are omitted.

The main result is

7.2 Theorem If the H-structure Y is H-associative and admits inversion, then:

(1). All path components of Y belong to the same homotopy type.
(2). The induced H-structure on Comp Y is in fact a group structure.

Proof: Ad (1). Let P be the principal component and C any path component. Define $f: P \to C$ by choosing a definite $c_0 \in C$ and setting $f(y) = y \cdot c_0$. It is easy to see that because of the H-associativity, f is a homotopy equivalence with homotopy inverse $y \to y \cdot c_0^{-1}$.

Ad (2). By **5.1**, **6.1**, and **7.1**, Comp Y has a (two-sided) unit E, is associative, and each element B has a right inverse B^{-1}. A purely algebraic consequence is that B^{-1} is also a left inverse of B: in the expression $D = B^{-1} \cdot B \cdot B^{-1} \cdot (B^{-1})^{-1}$, combine the last two terms to get $D = B^{-1} \cdot B$; on the other hand, combining by pairs from the middle gives $D = E$. Thus Comp Y is a group.

7.3 Corollary If Y is an H-associative H-structure with inversion, then for any locally compact space X, all the path components of Y^X belong to a common homotopy type. Furthermore, with the induced H-structure in Y^X, Comp Y^X is a group.

Proof: Immediate from Ex. 3 and **6**, Ex. 4.

7.4 Corollary Let Y be any path-connected space. Then:

(a). For each $y_0 \in Y$, all the path components of $\Omega(Y; y_0)$ belong to a common homotopy type.
(b). Comp $\Omega(Y; y_0)$ is always a group, called the fundamental (or: first homotopy) group at y_0.
(c). The fundamental groups at distinct points are isomorphic (so we can speak of *the* fundamental group).
(d). The fundamental group is a homotopy type invariant.

Proof: These are all immediate consequences of **7.2**, in view of **6**, Ex. 2; **3**, Ex. 2; and **3.1**.

Let $T_0(Y)$ denote the homotopy type of the path-connected Y. Let Y_1 be any path component of $\Omega(Y; y_0)$ and let $T_1(Y)$ be the homotopy type of Y_1; $T_1(Y)$ depends only on the homotopy type of Y. Repeating this construction, starting now from Y_1, gives a path-connected space Y_2 belonging to a homotopy type $T_2(Y)$, which depends only on that of Y. In this way, we obtain a sequence

$$Y, Y_1, Y_2, \cdots, Y_n, \cdots$$

of path-connected spaces belonging, respectively, to homotopy types

$$T_0(Y), T_1(Y), T_2(Y), \cdots, T_n(Y), \cdots,$$

which are uniquely determined by the homotopy type of Y. $T_n(Y)$ is called the nth derived homotopy type of Y; the fundamental group of Y_n is called the $(n + 1)$st homotopy group of Y. Note that each Y_i, $i \geqslant 1$, is an H-space.

8. Path Spaces on H-Spaces

Let (Y, Γ) be an H-structure. Its conjugate H-structure is (Y, Γ^+), where $\Gamma^+(x, y) = \Gamma(y, x)$. A given H-structure is called H-abelian if it is equivalent to its conjugate H-structure.

8.1 (1). An H-abelian H-structure Y induces an abelian H-structure on Comp Y.
(2). An H-structure H-isomorphic to an H-abelian H-structure is itself H-abelian.

Proof: (1). is trivial; (2) follows by noting that whenever two H-structures are H-isomorphic, so also are their conjugate structures.

Let Y be an H-structure *with* unit e; then the composition law induced on the function space Y^I maps $\Omega(Y; e)$ in itself, so we find now that $\Omega(Y; e)$ carries two H-structures: (a) this induced H-structure, and (b) the natural H-structure.

8.2 If Y is an H-structure with unit, the induced H-structure on $\Omega(Y; e)$ is equivalent to the natural H-structure $\Omega(Y; e)$.

Proof: The induced H-structure has composition law $(\alpha, \beta) \to \alpha \cdot \beta$, where $\alpha \cdot \beta$ is the path $\alpha(t) \cdot \beta(t)$. Choose the parameter transformations $p_0(t) = \max[0, 2t - 1]$ and $p_1(t) = \min[1, 2t]$. Since each $p_i \simeq 1$, and since e is a unit, the map $(\alpha, \beta) \to \alpha \cdot \beta$ is homotopic to the map $(\alpha, \beta) \to (\alpha \circ p_1) \cdot (\beta \circ p_0)$. Noting that $(\alpha \circ p_1) \cdot (\beta \circ p_0) = \alpha * \beta$ completes the proof.

8.3 Theorem Let Y be a path-connected H-space. Then the natural H-structure $\Omega(Y; y_0)$ is always H-abelian.

Proof: There is no loss in generality to assume that Y carries an H-structure with unit e_0. Indeed, given an H-structure on Y, there is (**5.3**) an H-isomorphism of Y with an H-structure Y' having a unit, 1. By **3**, Ex. 2, $\Omega(Y, e)$ is H-isomorphic to $\Omega(Y'; e)$, which, by **3**, Ex. 1, is H-isomorphic to $\Omega(Y', 1)$; using **8.1**(2), it would then suffice to show that $\Omega(Y'; 1)$ is H-abelian.

Then let e_0 be the unit, let S be the induced H-structure $\Omega(Y; e_0)$, and let S^+ be that induced by the conjugate H-structure in Y; clearly S and S^+ are conjugate H-structures carried by $\Omega(Y; e_0)$. Let N be the natural H-structure on $\Omega(Y, e_0)$ and let N^+ be the conjugate H-structure. Using **8.2**, we have that N is equivalent to S, and therefore N^+ is equivalent to S^+. However, by **8.2** again, N must also be equivalent to S^+; thus N is equivalent to N^+, and the proof is complete.

By **8.1**(1) we immediately obtain

8.4 Corollary The fundamental group of any path-connected H-space is abelian.

Note that, since the derived homotopy types of any space are always represented by H-spaces, it follows that, for all $n \geqslant 2$, the nth homotopy group of any space is abelian.

Problems

Section 1

1. Let a, a', b, b' be any four (not necessarily distinct) points lying in a single path component of Y. Show that $\Omega(Y; a, a') \simeq \Omega(Y; b, b')$.
2. Prove that any two path components of $\Omega(Y; a, b)$ belong to the same homotopy type.
3. Let Y, Z be path-connected spaces belonging to the same homotopy type. Prove that any path component of $\Omega(Y; a, b)$ belongs to the same homotopy type of any path component of $\Omega(Z; c, d)$.
4. Let $y_0 \in Y$ and let $p_0 \in S^1$ be fixed points, and let $\Lambda \in Y^{S^1}$ be the subspace of all maps sending p_0 to y_0. Prove Λ is homeomorphic to $\Omega(Y; y_0)$.
5. For any two subsets of A, B of Y, let

$$\Omega(A, B) = \{\alpha \in Y^I \mid \alpha(0) \in A, \alpha(1) \in B\}.$$

Assume that A is contractible over Y to a point y_0. Prove that

$$\Omega(A, B) \simeq A \times \Omega(y_0; B).$$

Section 2

1. What is the principal component of the natural H-structure $\Omega(Y, y_0)$?
2. Let Y be a topological group and X a locally compact space. Prove:
   a. The induced H-structure Y^X is also a topological group.
   b. The induced H-structure Comp Y^X is also a topological group.
   c. Show that the principal component P of Y^X is a normal subgroup and that Comp Y^X is isomorphic to Y^X/P.

Section 5

1. Let Y be a Hausdorff space, and $y_0 \in Y$. Let $A \subset Y \times Y$ be the closed subset $(Y \times y_0) \cup (y_0 \times Y)$ and let $f\colon A \to Y$ be the map $f(y, y_0) = f(y_0, y) = y$. Prove: Y carries an H-structure with unit y_0 if and only if Y is a retract of $(Y \times Y) \cup_f Y$.

Section 6

1. In the extension of Y to Y' described in **5.3**, show that whenever Y admits inversion, the inversion can be chosen in Y' so that $1^{-1} = 1$.

Section 7

1. Let Y be any path-connected space. For each integer $n \geq 1$, let I^n be the unit n-cube and $\Omega_n(Y, y_0)$ the subspace of Y^{I^n} consisting of all

$$f\colon (I^n, \mathrm{Fr}(I^n)) \to (Y, y_0).$$

Let $k_n(I^n) = y_0$ be the constant map.

a. Prove that $\Omega(\Omega_n(Y; y_0); k_n)$ is homeomorphic to $\Omega_{n+1}(Y; y_0)$. [Represent I^{n+1} as $I^n \times I$ and define $h: \Omega_{n+1} \to \Omega(\Omega_n)$ by $h[f(x, t)] = [f(x)](t)$ $(x, t) \in I^n \times I$.]

b. Prove that any path component of $\Omega_n(Y, y_0)$ belongs to the homotopy type $T_n(Y)$.

2. Let Y be an arbitrary space, and let a, b, be two points of Y. Let X be a locally compact space, and denote the poles of its suspension SX by p_+ and p_-. Let $\{SX, Y\}$ be the space of all maps of SX into Y sending p_+ to a and p_- to b. Show that $\{SX, Y\} \cong [\Omega(Y; a, b)]^X$, and that consequently $\{SX, Y\}$ has an H-structure such that Comp $\{SX, Y\}$ is a group. Describe a composition operation in $\{SX, Y\}$.
3. Show that the fundamental group of S^1 is an infinite cyclic group. [*Hint:* see XVI, **7.4**.]
4. Prove that the fundamental group of S^n, $n \geq 2$, is trivial.
5. Show that the fundamental group of any contractible space is trivial.

Fiber Spaces

XX

In this chapter, we will consider only the covering homotopy property in fiber structures, and some of its more immediate consequences. Our aim is to establish that, under fairly general conditions, the local validity of the covering homotopy property implies its validity in the large.

1. Fiber Spaces

1.1 Definition A fiber structure is a triple (E, p, B) consisting of two spaces E, B and a continuous surjection $p\colon E \to B$.

The space E is called the total (or fibered) space; p is termed the projection, and B is the base space. We refer to (E, p, B) as a fiber structure over B, and for each $b \in B$, the set $p^{-1}(b)$ is called the fiber over b.

Let (E, p, B) be a fiber structure, let X be an arbitrary space, and let $g\colon X \to B$ be continuous, so that we have the diagram

$$\begin{array}{ccc} & & E \\ & & \downarrow{\scriptstyle p} \\ X & \xrightarrow[g]{} & B \end{array}$$

A continuous $\tilde{g}: X \to E$ such that $p \circ \tilde{g} = g$ is called a lifting of g into E (or, more simply, a covering of g). In particular, a lifting σ of the identity map $1: B \to B$ into E is called a cross section; since $p \circ \sigma = 1$, every cross section is evidently injective.

Ex. 1 A map $g: X \to B$ need not be liftable into E. Let (E^1, p, S^1) be the fiber structure with projection $p(x) = e^{ix}$ and let $X = S^1$. It is easy to see that the identity map $g: X \to S^1$ cannot be lifted into E^1 [for example, this follows directly from XVI, **6.2**(2)]. In particular, this fiber structure does not have a cross section.

The general question that we will consider is the following: if $g: X \to B$ can in fact be lifted to a $\tilde{g}: X \to E$, then can every homotopy of g be covered by a homotopy of $\tilde{g}$? Fiber structures in which this can always be done are very important in modern topology.

Ex. 2 If $g: X \to B$ can be lifted to $\tilde{g}: X \to E$, then a homotopy of g may not be liftable to a homotopy of $\tilde{g}$. Let $E = (I \times 0) \cup (0 \times I)$ and let (E, p, I) be the fiber structure with projection $p(x, y) = x$. Let X be any space, let $g: X \to I$ be the constant map $g(X) = 0$, and lift g to the constant map $\tilde{g}$, where $\tilde{g}(X) = (0, 1)$. Then the homotopy $\Phi(X, t) = t$ of g cannot be covered by a homotopy of $\tilde{g}$.

Let $\tilde{\varphi}: X \times I \to E$ be a homotopy covering φ. We say that $\tilde{\varphi}$ is stationary with φ if for each $x_0 \in X$ such that $\varphi(x_0, t)$ is constant as a function of t, the function $\tilde{\varphi}(x_0, t)$ is also constant.

1.2 Definition (1). A fiber structure (E, p, B) is called a fiber space (or fibration) for a class $\mathscr{A}$ of spaces if the following condition (called the covering homotopy condition) is satisfied: For each $X \in \mathscr{A}$, each continuous $f: X \times 0 \to E$ and each homotopy $\varphi: X \times I \to B$ of $p \circ f$, there exists a homotopy $\tilde{\varphi}$ of f covering φ. The fibration is *regular* if $\tilde{\varphi}$ can always be selected to be stationary with φ.

(2). A fiber space for the class of all spaces is called a Hurewicz fibration.

Ex. 3 Let $p: Y \times Z \to Y$ be the natural projection $p(y, z) = y$. Then $(Y \times Z, p, Y)$ is a regular Hurewicz fibration. In fact, let $p_Z: Y \times Z \to Z$ be the projection; then for any continuous $f: X \times 0 \to Y \times Z$ and homotopy $\varphi: X \times I \to Y$ of $p \circ f$, the map $\tilde{\varphi}(x, t) = (\varphi(x, t), p_Z \circ f(x, 0))$ is a covering homotopy, and is stationary with φ.

Ex. 4 Let B be any space, let Z be a normal locally compact space, and let $z_0 \in Z$ be a G_δ and strong nbd deformation retract of Z. Let $\omega: B^Z \to B$ be the evaluation map $\omega(g) = g(z_0)$. Then (B^Z, ω, B) is a Hurewicz fibration. Indeed, let $f: X \times 0 \to B^Z$ be continuous, and let $\varphi: X \times I \to B$ be a homotopy of $\omega \circ f$. Since Z is locally compact, the associated map $\hat{f}: X \times 0 \times Z \to B$ is continuous. Let $L = X \times [(0 \times Z) \cup (I \times \{z_0\})]$, and define $\mu: L \to B$ by

$$\mu(x, 0, z) = \hat{f}(x, 0, z),$$
$$\mu(x, t, z_0) = \varphi(x, t),$$

which is continuous. Because $(0 \times Z) \cup (I \times \{z_0\})$ is a retract of $I \times Z$, we find that L is a retract of $X \times I \times Z$, so that μ is extendable to a $\psi: X \times I \times Z \to B$; the associated map $\hat{\psi}: X \times I \to B^Z$ is the required covering homotopy.

Ex. 5 If (E, p, B) is a fibration for the class $\mathscr{A}$ of spaces, then for each $C \subset B$, the fiber structure $(p^{-1}(C), p \mid p^{-1}(C), C)$ is also a fibration for the class $\mathscr{A}$ of spaces.

We derive some immediate consequences of the covering homotopy condition.

1.3 Let (E, p, B) be a fibration for the class $\mathscr{A}$ of spaces, and let $X \in \mathscr{A}$.

(1). If $f: X \to B$ is nullhomotopic, then f can be lifted into E. In particular, if $B \in \mathscr{A}$ and if B is contractible, then a cross section always exists.

(2). A $\tilde{g}: X \to E$ can be deformed into a single fiber if and only if $p \circ \tilde{g}$ is nullhomotopic.

Proof: (1). Let $\varphi: X \times I \to B$ be a nullhomotopy of f, with $\varphi(X, 0) = b_0$. Since the map $X \to b_0$ can be lifted, so can the homotopy φ, and the map $x \to \tilde{\varphi}(x, 1)$ is a lifting of f into E. The proof of (2) is entirely analogous.

For any two spaces, X, Y, a surjective $f: X \to Y$ is called irreducible whenever all maps homotopic to f are also surjective. Obviously, if Y has more than one point, an irreducible map is never nullhomotopic; in the particular case $Y = S^n$, the irreducible maps are exactly those maps that are not nullhomotopic.

1.4 Let (E, p, B) be a fibration for the class $\mathscr{A}$ of spaces, and let $X \in \mathscr{A}$. If $f: X \to E$ is irreducible, then so also is $p \circ f$. In particular, if $E \in \mathscr{A}$ and if $1: E \to E$ is irreducible, then so also is $p: E \to B$ and $1: B \to B$.

Proof: A homotopy of $p \circ f$ that frees a point of B would be covered by a homotopy of f that frees an entire fiber, so the first assertion is proved. To establish the remaining ones, we need observe only that $1_E: E \to E$ is a lifting of $p: E \to B$, so that a homotopy of p (or of 1_B) can be covered by a homotopy of 1_E.

1.5 Let (E, p, B) be a regular fibration for the class $\mathscr{A}$ of spaces, and let $E \in \mathscr{A}$. If $C \subset B$ is a strong deformation retract of B, then $p^{-1}(C)$ is a strong deformation retract of E.

Proof: Let $\varphi: B \times I \to B$ be a strong deformation retraction; we write $\varphi(b, t) = \varphi_t(b)$, and let $\varphi_0 = 1_B$. Since 1_E covers $\varphi_0 \circ p$, there is a homotopy $\tilde{\varphi}_t$ of 1_E covering $\varphi_t \circ p$, and which is stationary with $\varphi_t \circ p$. Thus, $\tilde{\varphi}_1(E) \subset p^{-1}(C)$ and $\tilde{\varphi}_t(e) = 1_E(e) = e$ for each $e \in p^{-1}(C)$ and $t \in I$, so $\tilde{\varphi}_t$ is the required deformation retraction.

Let (E, p, B) be a fiber structure, let X be any space, and let $g: X \to B$ be any continuous map into the base B. Let $E(g) \subset E \times X$ be the subspace $E(g) = \{(e, x) \mid p(e) = g(x)\}$ of the cartesian product, and let $q: E(g) \to X$ be the projection $q(e, x) = x$. The fiber structure $(E(g), q, X)$ is called the fiber structure over X induced by the map $g: X \to B$; intuitively, $E(g)$ is formed by placing over x the fiber found over $g(x)$. Letting $\pi: E(g) \to E$ be the projection $\pi(e, x) = e$, it is immediate from the definitions that the diagram

$$\begin{array}{ccc} E(g) & \xrightarrow{\pi} & E \\ q\downarrow & & \downarrow p \\ X & \xrightarrow[g]{} & B \end{array}$$

is commutative.

1.6 Let (E, p, B) be a fibration for the class $\mathscr{A}$ of spaces, let X be an arbitrary space, and let $g: X \to B$ be any continuous map. Then $(E(g), q, X)$ is also a fibration for the class $\mathscr{A}$ of spaces.

Proof: Let $Y \in \mathscr{A}$, let $f: Y \times 0 \to E(g)$ be continuous, and let $\varphi: Y \times I \to X$ be any homotopy of $q \circ f$. Then $g \circ \varphi: Y \times I \to B$ is a homotopy of $g \circ q \circ f = p \circ \pi \circ f$, so there is a homotopy $\Phi: Y \times I \to E$ of $\pi \circ f$ covering $g \circ \varphi$. The continuous map $\tilde{\varphi}: Y \times I \to E \times X$ given by $\tilde{\varphi}(y, t) = (\Phi(y, t), \varphi(y, t))$ actually maps $Y \times I$ into $E(g)$, since $p \circ \Phi(y, t) = g \circ \varphi(y, t)$ for all $(y, t) \in Y \times I$, and is evidently a homotopy of f covering φ.

2. Fiber Spaces for the Class of All Spaces

In this section, we obtain an intrinsic characterization of the Hurewicz fibrations, and use this to derive some of their special properties.

Let (E, p, B) and $f: X \times 0 \to E$ be given, and let $\varphi: X \times I \to B$ be a homotopy of $p \circ f$. For each $x \in X$, the map $t \to \varphi(x, t)$ defines a path

φ_x in B. The problem of finding a covering homotopy is essentially that of lifting each path φ_x to a path in E starting at $f(x, 0)$ in such a way that the family $\{\varphi_x \mid x \in X\}$ is lifted "continuously" into E. Thus, we may expect to have an intrinsic characterization of the Hurewicz fibrations if we require that the family of all paths in B can be "continuously" lifted in a similar manner.

To state this lifting process precisely, we use the c-topology in both B^I and E^I, and make the

2.1 Definition Let (E, p, B) be a fiber structure, and let $\Delta \subset E \times B^I$ be the subspace $\Delta = \{(e, \alpha) \mid p(e) = \alpha(0)\}$ of the cartesian product. A lifting function for (E, p, B) is a continuous map $\lambda: \Delta \to E^I$ such that $\lambda(e, \alpha)[0] = e$ and $p \circ \lambda(e, \alpha)[t] = \alpha[t]$ for each $(e, \alpha) \in \Delta$ and $t \in I$. We say that λ is regular if $\lambda(e, \alpha)$ is a constant path whenever α is a constant path.

A lifting function therefore associates with each $e \in E$ and each path α in B starting at $p(e)$, a path $\lambda(e, \alpha)$ in E starting at e and covering α. Since the c-topology is used in E^I, the continuity of λ is equivalent to that of the associated $\hat{\lambda}: \Delta \times I \to E$.

Ex. 1 In the fiber structure $(Y \times Z, p, Y)$ of **1**, Ex. 3, the map $\lambda[(y, z), \alpha](t) = (\alpha(t), z)$ is a regular lifting function.

2.2 Theorem (M. L. Curtis; W. Hurewicz) The fiber structure (E, p, B) is a (regular) Hurewicz fibration if and only if a (regular) lifting function exists.

Proof: Sufficiency. Assume that (E, p, B) has a lifting function. Let $f: X \times 0 \to E$ be given, and let $\varphi: X \times I \to B$ be a homotopy of $p \circ f$. For each $x \in X$, let φ_x be the path $t \to \varphi(x, t)$. Then $\tilde{\varphi}(x, t) = \lambda[f(x, 0), \varphi_x](t)$ is the required covering homotopy, and is stationary with φ whenever λ is regular.

Necessity. Let $X = \Delta$, and let $\varphi: \Delta \times I \to B$ be the map $\varphi[(e, \alpha), t] = \alpha(t)$. Let $f: \Delta \times 0 \to E$ be given by $f[(e, \alpha), 0] = e$. Since $p(e) = \alpha(0)$ for each $(e, \alpha) \in \Delta$, we have $p \circ f = \varphi \mid \Delta \times 0$. Letting $\tilde{\varphi}: \Delta \times I \to E$ be a homotopy covering φ, the associated map $\lambda: \Delta \to E^I$ is a lifting function, which is regular whenever $\tilde{\varphi}$ is stationary with φ.

2.3 Corollary. Let (E, p, B) be a fibration for the class of metric spaces. If both E and B are metric spaces, then (E, p, B) is a Hurewicz fibration.

Proof: In this case, $\Delta \subset E \times B^I$ is a metric space, and the proof of the "necessity" in **2.2** shows that a lifting function exists.

2.4 Corollary. Let (E, p, B) be a Hurewicz fibration. If B is a metric space, then (E, p, B) is a regular Hurewicz fibration.

Proof: Let $\lambda: \Delta \to E^I$ be a lifting function for (E, p, B). Choose a metric d for B such that $\delta(B) < 1$, and for each $\alpha \in B^I$ let $d(\alpha) = \delta[\alpha(I)]$. Define

$$\bar{\alpha}(t) = \alpha[t/d(\alpha)] \qquad t < d(\alpha)$$
$$= \alpha[1] \qquad t \geq d(\alpha),$$

and let $\hat{\lambda}(e, \alpha)[t] = \lambda(e, \bar{\alpha})[t \cdot d(\alpha)]$. Then $\hat{\lambda}$ is a regular lifting function.

In the remainder of this section, we establish some special properties of Hurewicz fibrations that illustrate the use of lifting functions.

A Hurewicz fibration always has an *extended* lifting function Λ which, with each $e \in E$, each $s \in I$ and each path α in B going through $p(e)$ at time $t = s$, associates a path in E going through e and covering α:

2.5 Let (E, p, B) be a Hurewicz fibration. Let $D \subset E \times B^I \times I$ be the subspace $D = \{(e, \alpha, s) \mid p(e) = \alpha(s)\}$. Then there exists a continuous $\Lambda: D \to E^I$ such that $\Lambda(e, \alpha, s)[s] = e$ and

$$p \circ \Lambda(e, \alpha, s)[t] = \alpha[t]$$

for each $(e, \alpha, s) \in D$, $t \in I$. If (E, p, B) is regular, then Λ can be selected to send constant paths to constant paths.

Proof: Let λ be a lifting function. Given $(e, \alpha, s) \in D$, define

$$\alpha_s(t) = \begin{cases} \alpha(s - t) & 0 \leq t \leq s \\ \alpha(0) & s \leq t \leq 1 \end{cases} \qquad \alpha^s(t) = \begin{cases} \alpha(s + t) & 0 \leq t \leq 1 - s \\ \alpha(1) & 1 - s \leq t \leq 1 \end{cases}$$

to get two paths starting at $\alpha(s)$; lift each of them to start at e and reparametrize, by setting

$$\Lambda(e, \alpha, s)[t] = \begin{cases} \lambda(e, \alpha_s)[s - t] & 0 \leq t \leq s \\ \lambda(e, \alpha^s)[t - s] & s \leq t \leq 1 \end{cases}$$

to obtain the desired lifting function. To verify that Λ is continuous, decompose it into

$$(e, \alpha, s, t) \to (e, \alpha_s, s, t) \to (e, \alpha_s, s - t) \to \lambda(e, \alpha_s)[s - t] \qquad 0 \leq t \leq s,$$
$$(e, \alpha, s, t) \to (e, \alpha^s, s, t) \to (e, \alpha^s, t - s) \to \lambda(e, \alpha^s)[t - s] \qquad s \leq t \leq 1.$$

The maps $\alpha \to \alpha_s$ and $\alpha \to \alpha^s$ of B^I into itself are each continuous (XII, **2.1**) so that the first map in each string is continuous; since the remaining maps in each string are continuous, and since the resulting maps agree for $s = t$, the continuity of Λ follows. The second part of the proposition is evident from the definition of Λ.

2.6 Let (E, p, B) be a regular Hurewicz fibration. Let X be any space, let $A \subset X$ be a closed subset, and let $\tilde{\varphi}: (X \times 0) \cup (A \times I) \to E$ be a continuous map such that the image of each $a \times I$ lies in a single fiber (that is, $p \circ \tilde{\varphi}(a, t) = p \circ \tilde{\varphi}(a, 0)$ for each $a \in A$ and $t \in I$). If $\tilde{\varphi}$ can be extended over $X \times I$, then $\tilde{\varphi}$ can also be extended over $X \times I$ in such a way that the image of each $x \times I$ lies in a single fiber.

Proof: Let $\Phi: X \times I \to E$ be an extension of $\tilde{\varphi}$. For each $(x, t) \in X \times I$, let $\alpha(x, t)$ be the path $s \to p \circ \Phi[x, (1 - s)t]$ which starts at $p \circ \Phi(x, t)$ and runs to $p \circ \Phi(x, 0)$. Let λ be a regular lifting function, and define

$$\tilde{\Phi}(x, t) = \lambda[\Phi(x, t), \alpha(x, t)](1).$$

The continuous map $\tilde{\Phi}$ is the desired extension of $\tilde{\varphi}$: indeed, we have

$$p \circ \tilde{\Phi}(x, t) = \alpha(x, t)(1) = p \circ \Phi(x, 0)$$

for all $(x, t) \in X \times I$, and if $a \in A$, then $\alpha(a, t)$ is a constant path, so that the regularity of λ gives

$$\tilde{\Phi}(a, t) = \lambda[\Phi(a, t), \alpha(a, t)](1) = \Phi(a, t) = \tilde{\varphi}(a, t).$$

Given two fiber structures (E, p, B) and (L, q, B) over the same base B, a continuous $f: E \to L$ is called a fiber (or fiber-preserving) map whenever $q \circ f = p$. Two fiber maps $f, h: E \to L$ are fiber-homotopic if there is a homotopy $\Phi: f \simeq h$ such that the homotopy path of each $e \in E$ lies in a single fiber. The fiber structures (E, p, B) and (L, q, B) are fiber-homotopy equivalent if there are fiber maps $f: E \to L$ and $g: L \to E$ such that $f \circ g$ is fiber-homotopic to 1_L and $g \circ f$ is fiber-homotopic to 1_E. For Hurewicz fibrations, **1.3**(1) can be considerably improved:

2.7 Theorem (E. Fadell; J. Feldbau) Let (E, p, B) be a regular Hurewicz fibration. Assume that B is contractible to a point $b_0 \in B$, and let $F = p^{-1}(b_0)$. Then (E, p, B) is fiber-homotopy equivalent to the fiber structure $(B \times F, q, B)$, where q is the projection on the first factor.

Proof: Let $\varphi: B \times I \to B$ be a contraction to b_0, where $\varphi \mid B \times 0 = 1_B$ and $\varphi(B, 1) = b_0$. For each $b \in B$ and $s \in I$, let $\alpha(b, s)$ be the path $t \to \varphi(b, st)$ and let $\alpha^*(b, s)(t) = \varphi[b, s(1 - t)]$. Define $f: E \to B \times F$ by

$$f(e) = \{pe, \lambda[e, \alpha(pe, 1)](1)\},$$

where λ is a regular lifting function, and let $g: B \times F \to E$ be the map

$$g(b, x) = \lambda[x, \alpha^*(b, 1)](1).$$

It is clear that f and g are fiber maps. To show that $g \circ f$ is fiber-homotopic to 1_E, let $\Phi: E \times I \to E$ be the continuous map

$$\Phi(e, s) = \lambda\{\lambda[e, \alpha(pe, s)](1), \alpha^*(pe, s)\}(1).$$

Then $\Phi(e, 0) = e$ because λ is regular; $\Phi(e, 1) = g \circ f(e)$; and for each t the map $e \to \Phi(e, t)$ is a fiber map. It is equally simple to see that $f \circ g$ is fiber-homotopic to $1_{B \times F}$ and the proof is complete.

Remark 1: The theorem is true without the requirement of regularity; the reader will easily show that in this case the map $e \to \Phi(e, 0)$ is fiber-homotopic to the identity map.

Remark 2: The proof of **2.7** can be used to obtain the more general result: If (E, p, B) is a Hurewicz fibration, and if any point $b_0 \in B$ has a nbd U that is deformable over B into b_0, then $(p^{-1}(U), p \mid p^{-1}(U), U)$ is fiber-homotopy equivalent to $(U \times p^{-1}(b_0), q, U)$, where q is the natural projection.

2.8 (E. H. Spanier; J. H. C. Whitehead) Let (E, p, B) be a Hurewicz fibration. If a fiber $F = p^{-1}(b_0)$ is contractible over E to a point $e_0 \in F$, then F is an H-space.

Proof: Let $\Phi: F \times I \to E$ be the contraction, where $\Phi(x, 1) = x$ for each $x \in F$, and $\Phi(F, 0) = e_0$. For each $x \in F$, let $\alpha(x)$ be the path $t \to p \circ \Phi(x, t)$ in B; the path $\alpha(x)$ is a loop at b_0. Define a composition operation $\Gamma: F \times F \to F$ by

$$\Gamma(x, y) = \lambda[e_0, \alpha(x) * \alpha(y)](1),$$

where λ is a lifting function. Then Γ is continuous (XIX, **1.4**) and (F, Γ) is an H-structure. To see that the maps $x \to \Gamma(x, e_0)$ and $x \to \Gamma(e_0, x)$ are homotopic to the identity, a parameter transformation shows it suffices to prove that the map $\gamma(x) = \lambda[e_0, \alpha(x)](1)$ is homotopic to the identity. To do this, let

$$\alpha^s(x)[t] = \begin{cases} \alpha(x)[s + t] & 0 \leqslant t \leqslant 1 - s \\ \alpha(x)[1] & 1 - s \leqslant t \leqslant 1 \end{cases}$$

and define

$$H(x, s) = \lambda[\Phi(x, s), \alpha^s(x)](1 - s).$$

Then $H: F \times I \to F$ is continuous, and $H: \gamma \simeq 1_F$. This completes the proof.

3. The Uniformization Theorem of Hurewicz

In this section we will show that under certain conditions, the property of a fiber structure to be a Hurewicz fibration is essentially a local matter. To make this statement precise, we first require

3.1 Definition Let X be an arbitrary space. An open $U \subset X$ is called a cozero set if there is a continuous $c: X \to I$ such that $c^{-1}(0) = X - U$.

With this, the statement of the uniformization theorem becomes

3.2 Theorem (W. Hurewicz) Let (E, p, B) be a fiber structure, and assume that there is a nbd-finite covering $\{U_\beta \mid \beta \in \mathscr{B}\}$ of B by cozero sets such that each $(p^{-1}(U_\beta), p \mid p^{-1}(U_\beta), U_\beta)$ is a (regular) Hurewicz fibration. Then (E, p, B) is a (regular) Hurewicz fibration.

To prove this theorem, we will need two lemmas about cozero sets in arbitrary spaces.

3.3 Lemma (1). The intersection of finitely many cozero sets is a cozero set.

(2). The union of any nbd-finite family of cozero sets is a cozero set.

(3). Let U be a cozero set in X and let $A \subset I$ be compact. Then the subbasic open set $(A, U) \subset X^I$ is a cozero set.

Proof: The proof of (1) is trivial.

(2). Let $\{U_\beta \mid \beta \in \mathscr{B}\}$ be a nbd-finite family of cozero sets. For each $\beta \in \mathscr{B}$, choose a continuous $c_\beta: X \to I$ such that $c_\beta^{-1}(0) = X - U$, and let $c(x) = \sup\{c_\beta(x) \mid \beta \in \mathscr{B}\}$. Then $c: X \to I$ is continuous, since each point has a nbd on which the sup is taken over a finite subset $\mathscr{B}' \subset \mathscr{B}$, and c shows that $\bigcup \{U_\beta \mid \beta \in \mathscr{B}\}$ is a cozero set.

(3). Let $c: X \to I$ be a continuous function such that $c^{-1}(0) = X - U$, and let $\omega: X^I \times I \to X$ be the evaluation map. Then ω is continuous and consequently the map $b: X^I \times I \to I$, given by $b(f, t) = c[\omega(f, t)] = c\,[f(t)]$ is also continuous. Let

$$\mu(f) = \inf\{b(f, t) \mid t \in A\};$$

since A is compact, we have $\mu(f) \neq 0$ if and only if $f \in (A, U)$, and it remains to show that μ is continuous. By III, **10**(4), μ is upper semi-continuous, so we need prove only that $\{f \mid \mu(f) \leqslant s\}$ is closed for each $s \in I$. To do this, note that because A is compact, we have $\mu(f) \leqslant s$ if and only if there is some $t_0 \in I$ such that $b(f, t_0) \leqslant s$; therefore, if we let $\pi: X^I \times I \to X^I$ be the natural projection, it follows that

$$\{f \mid \mu(f) \leqslant s\} = \pi\{b^{-1}([0, s])\}.$$

Since $b^{-1}([0, s])$ is closed, and since a projection parallel to a compact factor is a closed map, the continuity of μ follows.

3.4 Lemma. Let X be an arbitrary space, and let $\mathfrak{U} = \{U_\beta \mid \beta \in \mathscr{B}\}$ be a covering of X by cozero sets. Assume that $\mathfrak{U}$ can be decomposed into a countable collection $\mathfrak{U}_k = \{U_{k,\beta_k} \mid \beta_k \in \mathscr{B}_k\}$, $k \in Z^+$, of nbd-finite families. Then $\mathfrak{U}$ has a nbd-finite refinement by cozero sets.

Proof: Let $T_k = \bigcup \{U_{k,\beta_k} \mid \beta_k \in \mathscr{B}_k\}$; by **3.3**, each T_k is a cozero set. We let $c_k: X \to I$ be a characterizing function for T_k. For each $k \in Z^+$, define

$$V_k = T_k \cap \bigcap_{n=1}^{k-1} \{x \in X \mid c_n(x) < 1/k\}.$$

Each V_k is a cozero set: indeed, $c(x) = \max\{0, (1/k) - c_n(x)\}$ shows $\{x \in X \mid c_n(x) < 1/k\}$ to be a cozero set, and we apply **3.3**(1).

The family $\{V_k \mid k \in Z^+\}$ is an open covering: given $x \in X$, let k be the first index for which $x \in T_k$; then $x \bar{\in} T_n$ for all $n < k$, so $c_n(x) = 0$ for all $n < k$, and therefore $x \in V_k$.

The covering $\{V_k \mid k \in Z^+\}$ is nbd-finite: Given $x \in X$, we have $x \in V_n \subset T_n$ for some n. Let $k \in Z^+$ be the smallest integer such that $c_n(x) > 1/k$ and consider the nbd $G = \{x \mid c_n(x) > 1/k\}$ of x. If $r \geqslant \max(k, n)$, then $V_r \subset \{x \mid c_n(x) < 1/r\}$ and, since $1/r \leqslant 1/k$, we have $G \cap V_r = \varnothing$. Thus, G meets at most finitely many sets V_ι.

It now follows that $\{V_k \cap U_{k,\beta_k} \mid k \in Z^+, \beta_k \in \mathscr{B}_k\}$ is a nbd-finite cozero refinement of $\mathfrak{U}$ and the lemma is proved.

Proof of Theorem 3.2: We shall construct a lifting function for (E, p, B).

For each finite set $\{\beta_1, \cdots, \beta_n\}$ of indices, let $W(\beta_1, \cdots, \beta_n) \subset B^I$ be the set of all paths α such that

$$\alpha(t) \in U_{\beta_k} \quad \text{for} \quad (k-1)/n \leqslant t \leqslant k/n.$$

Letting $I_\iota = [(i-1)/n, i/n]$, we have $W(\beta_1, \cdots, \beta_n) = \bigcap_1^n (I_\iota, U_{\beta_\iota})$, so that each $W(\beta_1, \cdots, \beta_n)$ is a cozero set (*cf.* **3.3**) in B^I. For each $n \in Z^+$, let $\mathfrak{W}_n = \{W(\beta_1, \cdots, \beta_n) \mid (\beta_1, \cdots, \beta_n) \in \prod_1^n \mathscr{B}\}$; then $\mathfrak{W} = \bigcup_n \mathfrak{W}_n$ is evidently an open covering of B^I and, furthermore, each family $\mathfrak{W}_n$ is nbd-finite: given $\alpha \in B^I$, the compact $\alpha(I) \subset B$ has a nbd V meeting at most finitely many U_β (*cf.* XI, **1.5**), so the nbd (I, V) of α can meet at most finitely many sets of $\mathfrak{W}_n$. According to **3.4**, $\mathfrak{W}$ has a nbd-finite refinement $\{W_\mu \mid \mu \in \mathscr{M}\}$ by cozero sets.

For each W_μ, let $D_\mu \subset E \times W_\mu \times I$ be the subspace

$$D_\mu = \{(e, \alpha, s) \mid p(e) = \alpha(s)\}.$$

We are first going to construct a continuous $\lambda_\mu: D_\mu \to E^I$ such that

$$\lambda_\mu(e, \alpha, s)[s] = e,$$
$$p \circ \lambda_\mu(e, \alpha, s)[t] = \alpha[t],$$

for each $(e, \alpha, s) \in D_\mu$ and $t \in I$; we will then fit the maps λ_μ together to produce a lifting function.

Given W_μ, select some $W(\beta_1, \cdots, \beta_n) \supset W_\mu$ and for each $i = 1, \cdots, n$, let Λ_i be an extended lifting function (*cf.* **2.5**) for

$$(p^{-1}(U_{\beta_i}), p \mid p^{-1}(U_{\beta_i}), U_{\beta_i}).$$

For $i = 1, \cdots, n$ let I_i be the closed interval $[(i - 1)/n, i/n]$, and for each $\alpha \in W_\mu$, let α_i be the path that agrees with $\alpha(t)$ for $t \in I_i$ and is constant elsewhere.

Given $(e, \alpha, s) \in D_\mu$, we define $\lambda_\mu(e, \alpha, s)$ to be the path ω obtained in the following manner: Determine k so that $s \in I_k$; then, because $p(e) = \alpha(s)$, so that $e \in p^{-1}(U_{\beta_k})$, we set

$$\begin{aligned} \omega[t] &= \Lambda_k(e, \alpha_k, s)[t] && t \in I_k \\ &= \Lambda_{k-1}(\omega[(k - 1)/n], \alpha_{k-1}, (k - 1)/n)[t] && t \in I_{k-1} \\ &= \Lambda_{k+1}(\omega[k/n], \alpha_{k+1}, k/n)[t] && t \in I_{k+1} \\ &\cdots \end{aligned}$$

and radiate outward. The path $\omega = \lambda_\mu(e, \alpha, s)$ is uniquely and unambiguously defined; by **2.5** and the definition of ω it follows that the map λ_μ is continuous, and it clearly satisfies the requirements.

Now, let $\Delta = \{(e, \alpha) \mid p(e) = \alpha(0)\} \subset E \times B^I$, and select a definite open covering $\{U\}$ of B^I such that each U meets at most finitely many sets W_μ. For each U, let

$$\Delta_U = \Delta \cap (E \times U) \subset E \times B^I;$$

we will match the λ_μ on each set Δ_U.

Well-order the indexing set $\mathscr{M}$ and for each $\mu \in \mathscr{M}$ choose a definite characterizing function $c_\mu: W_\mu \to I$ for the cozero set W_μ. Given $U \in \{U\}$, let $\mu_1 < \mu_2 < \cdots < \mu_n$ be the indices of all the sets W_μ meeting U, and define n continuous real-valued functions on U by

$$t_r(\alpha) = \sum_{i=1}^{r} c_{\mu_i}(\alpha) \Big/ \sum_{i=1}^{n} c_{\mu_i}(\alpha) \qquad r = 1, 2, \cdots, n.$$

Since we have $p(e) = \alpha(0)$ for each $(e, \alpha) \in \Delta_U$, we define $\lambda_U: \Delta_U \to E^I$ by

$$\begin{aligned} \lambda_U(e, \alpha)[t] &= \lambda_{\mu_1}(e, \alpha, 0)[t] && 0 \leqslant t \leqslant t_1(\alpha) \\ &= \lambda_{\mu_2}(\lambda_U(e, \alpha)[t_1(\alpha)], \alpha, t_1(\alpha))[t] && t_1(\alpha) \leqslant t \leqslant t_2(\alpha) \\ &\cdots \end{aligned}$$

Due to the continuity of the λ_{μ_i} and the t_r, each λ_U is continuous.

Now observe that $t_{\iota-1}(\alpha) = t_\iota(\alpha)$ whenever $\alpha \bar{\in} W_{\mu_\iota}$. It follows that the value of λ_U at $(e, \alpha) \in \Delta_U$ really depends only on the sets W_μ that contain α. In fact, if $\xi_1 < \xi_2 < \cdots < \xi_s$ are the indices of the sets W_μ that contain α, then letting

$$q_r(\alpha) = \sum_{\iota=1}^{r} c_{\xi_\iota}(\alpha) \Big/ \sum_{\iota=1}^{s} c_{\xi_\iota}(\alpha) \qquad r = 1, 2, \cdots, s,$$

we find from the definition of λ_U that

$$\begin{aligned} \lambda_U(e, \alpha)[t] &= \lambda_{\xi_1}(e, \alpha, 0)[t] && 0 \leqslant t \leqslant q_1(\alpha) \\ &= \lambda_{\xi_2}(\lambda_U(e, \alpha)[q_1(\alpha)], \alpha, q_1(\alpha))[t] && q_1(\alpha) \leqslant t \leqslant q_2(\alpha) \\ &\cdots \end{aligned}$$

We conclude from this observation that, if $(e, \alpha) \in \Delta_U \cap \Delta_V$, then $\lambda_U(e, \alpha) = \lambda_V(e, \alpha)$. Thus, since $\{\Delta_U \mid U \in \{U\}\}$ is an open covering of Δ, and since $\lambda_U \mid \Delta_U \cap \Delta_V = \lambda_V \mid \Delta_U \cap \Delta_V$ for all $U, V \in \{U\}$, we can apply III, **9.4**: the mapping $\lambda: \Delta \to E^I$ given by

$$\lambda \mid \Delta_U = \lambda_U$$

is continuous. λ is clearly a lifting function.

If each $(p^{-1}(U_{\beta_\iota}), p \mid p^{-1}(U_{\beta_\iota}), U_{\beta_\iota})$ is regular, then the λ_μ can all be selected to lift constant paths to constant paths (*cf.* **2.5**) and the lifting function λ that we have constructed will be regular. The proof of **3.2** is complete.

3.5 Definition A fiber structure (E, p, B) is called a local (regular) Hurewicz fibration if each $b \in B$ has a nbd U such that $(p^{-1}(U), p \mid p^{-1}(U), U)$ is a (regular) Hurewicz fibration.

The uniformization theorem allows us to conclude that a fiber structure is a fibration for a suitable class of spaces by simply verifying that it is a local Hurewicz fibration:

3.6 Corollary Let the base B be paracompact. Then a fiber structure (E, p, B) is a (regular) Hurewicz fibration if and only if it is a local (regular) Hurewicz fibration.

Proof: It is clear that every (regular) Hurewicz fibration (E, p, B) is a local (regular) Hurewicz fibration, so only the converse requires proof. Since B is paracompact, the covering of B by the nbds U of the definition has a nbd-finite refinement $\{V_\beta\}$ which we shrink to get $\{\overline{W}_\beta\}$. For each index β, let $c_\beta: X \to I$ be a Urysohn function such that $c_\beta(\overline{W}_\beta) = 1$ and $c_\beta(\mathscr{C} V_\beta) = 0$. The covering of B by the open sets $U_\beta = c_\beta^{-1}[I - \{0\}]$ satisfies the requirements of the theorem.

3.7 Corollary Let the base B be arbitrary. Then a local (regular) Hurewicz fibration is always a (regular) fiber space for the class of paracompact spaces.

Proof: Let X be paracompact, let $f: X \times 0 \to E$ be continuous, and let $\varphi: X \times I \to B$ be a homotopy of $p \circ f$. Let $(E(\varphi), q, X \times I)$ be the fiber structure over $X \times I$ induced by φ; it follows from **1.6** that $(E(\varphi), q, X \times I)$ is also a local (regular) Hurewicz fibration and, since $X \times I$ is paracompact (XI, **5.4**), it is therefore a (regular) Hurewicz fibration. Now let $F: X \times 0 \to E(\varphi) \subset E \times X \times I$ be the map $F(x, 0) = (f(x, 0), x, 0)$; then $q \circ F = 1 \mid X \times 0$, so that F extends to a continuous $F': X \times I \to E(\varphi)$ covering $1: X \times I \to X \times I$. Letting $\pi: E(\varphi) \to E$ be the natural projection, we find that $\pi \circ F'$ is a homotopy of f covering φ, and the proof is complete.

4. Locally Trivial Fiber Structures

The fiber structures (E, p, B) that occur most frequently in many important contexts are those in which each point of B has a nbd over which the projection behaves roughly as the projection of a cartesian product on one of its factors. We will call such structures locally trivial, although this is not the most commonly used terminology. Precisely,

4.1 Definition A fiber structure (E, p, B) is called locally trivial if for each $b \in B$ there is a nbd U containing b and a continuous map $\sigma_U: U \times p^{-1}(U) \to E$ such that

(1). $p \circ \sigma_U(b, e) = b$ for all $(b, e) \in U \times p^{-1}(U)$,

(2). $\sigma_U(p(e), e) = e$ for each $e \in p^{-1}(U)$.

The maps σ_U are called "slicing maps" and the nbds U are termed "slicing nbds"; for each $e \in p^{-1}(U)$, the map $b \to \sigma_U(b, e)$ provides a "slice" in $p^{-1}(U)$ going through e. Note that no consistency condition is imposed on the slicing maps: the maps $b \to \sigma_U(b, e_0)$ and $b \to \sigma_V(b, e_0)$ are not required to agree on $U \cap V$.

Ex. 1 Let E be the triangle $\{(x, y) \mid 0 \leqslant y \leqslant x \leqslant 1\} \subset E^2$ and let $p: E \to I$ be given by $p(x, y) = x$. Then (E, p, I) is a locally trivial fiber structure; a slicing function can be defined on all I by

$$\sigma[x', (x, y)] = (x', x') \quad \text{if } x' \leqslant y$$
$$= (x', y) \quad \text{if } x' \geqslant y.$$

Ex. 2 Let R be the equivalence relation in $I \times I$ generated by $(0, y) \sim (1, 1 - y)$ and let S be the equivalence relation in I generated by $0 \sim 1$. The projection $p: I \times I \to I$ given by $p(x, y) = x$ is relation-preserving; passing to the quotient gives a continuous $p_*: (I \times I)/R \to I/S$, and it is easy to verify that $[(I \times I)/R, p_*, I/S]$ is a locally trivial fiber structure. The total space is the Möbius band, and the base space is S^1.

We now consider the covering homotopy property in locally trivial fiber structures. It is evident that a locally trivial fiber structure is a local regular Hurewicz fibration: if U is a slicing nbd, the function $\lambda(e, \alpha)[t] = \sigma_U[\alpha(t), e]$ is a regular lifting function for

$$(p^{-1}(U), p \mid p^{-1}(U), U).$$

As an immediate consequence of **3.6** and **3.7** we therefore have

4.2 Theorem Let (E, p, B) be a locally trivial fiber structure. Then

(1). (E, p, B) is always a regular fiber space for the class of paracompact spaces.

(2). If B is paracompact, then (E, p, B) is a regular Hurewicz fibration.

Under certain conditions on the base, a local regular Hurewicz fibration is necessarily locally trivial. Call a space B locally equiconnected if for each $b \in B$ there is a nbd U of b and a continuous $\mu_U: U \times U \times I \to B$ satisfying

$$\mu_U(a, c, 0) = a \quad \text{and} \quad \mu_U(a, c, 1) = c$$

for all $(a, c) \in U \times U$ and

$$\mu_U(a, a, t) = a$$

for all $a \in U$, $t \in I$. The map μ_U provides a continuous family of paths in B joining pairs of points of U; the nbd U is called an equiconnected nbd and μ_U a connecting function. We remark that the class of locally equiconnected spaces includes the class of metric ANR spaces.

Let B be locally equiconnected, let $b \in B$, and let U be any equiconnected nbd of b. Then for each nbd V of b there is a nbd W of b such that $\mu_U(W \times W \times I) \subset V$: indeed, since $b \times b \times I \subset \mu_U^{-1}(V)$ and I is compact, XI, **2.6** shows there is a G open in $U \times U$ and containing $b \times b$ such that $G \times I \subset \mu_U^{-1}(V)$; since G is therefore also open in $B \times B$, the desired conclusion follows.

4.3 Theorem Let B be a locally equiconnected space. Then a fiber structure (E, p, B) is a regular local Hurewicz fibration if and only if it is locally trivial.

Proof: We need prove only that if (E, p, B) is a regular local Hurewicz fibration, then (E, p, B) is locally trivial. Let $b \in B$ be given, let V be a nbd of b such that $(p^{-1}(V), p \mid p^{-1}(V), V)$ is a regular Hurewicz fibration, and let λ_V be a regular lifting function. By our remarks above, we can find a nbd U of b and a connecting function μ such that $\mu(U \times U \times I) \subset V$. For $(b, e) \in U \times p^{-1}(U)$ let $\alpha(pe, b)$ be the path $t \to \mu(pe, b, t)$ and let $\sigma_U(b, e) = \lambda_V[e, \alpha(pe, b)](1)$. It is trivial to verify that σ_U is a slicing map, and the proof is complete.

4.4 Corollary Let B be paracompact and locally equiconnected. Then a fiber structure (E, p, B) is a regular Hurewicz fibration if and only if it is locally trivial.

Proof: This is immediate from **4.2** and **4.3**.

Ex. 3 (Stiefel fiber structures). Let $V_{n,m}$ be the family of all ordered sets $(w_1, \cdots, w_m)$ of $m < n$ mutually orthogonal unit vectors at the origin of E^n. Using a fixed orthogonal coordinate system, each $w \in V_{n,m}$ is represented by a matrix $[w]$ of m rows and n columns such that $[w]\cdot[w]^T = I_m$ (where $[w]^T$ is the transpose of $[w]$ and I_m is the unit $m \times m$ matrix). The set $V_{n,m}$ topologized as a subset of E^{nm} is called a Stiefel manifold; in particular, it is a compact metric space.

For given $0 < k < m < n$, define $p: V_{n,m} \to V_{n,k}$ by

$$p(w_1, \cdots, w_k, \cdots, w_m) = (w_1, \cdots, w_k).$$

The map p is a continuous surjection, and each fiber is evidently homeomorphic to $V_{n-k,m-k}$.

We now show that $(V_{n,m}, p, V_{n,k})$ is a regular Hurewicz fibration, by proving that it is a locally trivial fiber structure (*cf.* **4.2**(2)). First working in $V_{n,k}$, observe that because $H(u, v) = \det([u][v]^T)$ is a continuous real-valued function on the compact $V_{n,k} \times V_{n,k}$, and because $H(u, u) \equiv 1$, there is an $\varepsilon > 0$ such that $H(u, v) \neq 0$ whenever $d(u, v) < \varepsilon$. It follows from this that if $d(u, pw) < \varepsilon$, then the vectors $u_1, \cdots, u_k, w_{k+1}, \cdots, w_m$ are linearly independent: for if $\sum_1^k a_i u_i + \sum_{k+1}^m b_j w_j = 0$, then $\sum_1^k a_i(u_i \cdot w_s) = 0$ for each $s = 1, \cdots, k$, and since $H(u, pw) \neq 0$, we conclude first that all $a_i = 0$, then that all $b_j = 0$. With this established, let $u \in V_{n,k}$, let $U = B(u, \varepsilon)$, and define $\sigma_U: U \times p^{-1}(U) \to V_{n,m}$ by

$$\sigma_U[(u_1, \cdots, u_k), (w_1, \cdots, w_m)] = (u_1, \cdots, u_k, \hat{w}_{k+1}, \cdots, \hat{w}_m)$$

where $u_1, \cdots, u_k, \hat{w}_{k+1}, \cdots, \hat{w}_m$ are the vectors $u_1, \cdots, u_k, w_{k+1}, \cdots, w_m$ orthogonalized by the Gram-Schmidt process in the order that they are written. It is simple to verify that σ_U is a slicing map; since u is arbitrary, $(V_{n,m}, p, V_{n,k})$ is locally trivial and is therefore a regular Hurewicz fibration.

Note that $V_{n,1}$ is homeomorphic to S^{n-1} and that $V_{n,2}$ can be regarded as the space of all unit tangent vectors to S^{n-1} (by translating the second vector to the end point of the first); with these interpretations, the projection $p: V_{n,2} \to S^{n-1}$ sends each tangent vector to the point at which it acts, and $(V_{n,2}, p, S^{n-1})$ has a cross section if and only if n is even (XVI, **3.3**, Ex. 4). Similarly, $V_{n,k}$ can be regarded as the space of all orthogonal tangent $(k-1)$-frames on S^{n-1}. Furthermore, $V_{n,n-1}$ can also be identified with the group $SO(n)$ of all orthogonal $n \times n$ matrices having determinant $+1$, since for each $v \in V_{n,n-1}$ there is a unique nth row that can be added to $[v]$ to obtain such a matrix.

As the above interpretations suggest, Stiefel fiber structures are important in homotopy theory. We illustrate only one of their many other uses by providing the following theorem of T. Wazeski:

Let X be a contractible space, and let M be a $k \times n$ matrix of continuous real-valued functions on X that has rank k at each $x \in X$. Then M can be enlarged to an $n \times n$ matrix $\tilde{M}$ of continuous real-valued functions on X that is non-singular at each $x \in X$.

Proof: Due to the properties of the orthogonalization process, there is no loss in generality to assume that $M \cdot M^T \equiv I_k$. In an evident manner, the matrix M then determines a continuous map $\mu: X \to V_{n,k}$. Since $(SO(n), p, V_{n,k})$ is a Hurewicz fibration, and since X is contractible, it follows from **1.3**(1) that μ can be lifted to a continuous $\tilde{\mu}: X \to SO(n)$; the matrix $\tilde{M}$ corresponding to the map $\tilde{\mu}$ is the required enlargement. This proof is due to B. Eckmann.

Ex. 4 We will describe the locally trivial fiber structures $S^3 \to S^2$, $S^7 \to S^4$ and $S^{15} \to S^8$ due to H. Hopf.

For $q = 1, 2, 4, 8$, let $\mathfrak{R}_q$ be the algebra of, respectively, the real numbers, complex numbers, quaternions, and Cayley numbers. These are all division algebras; but the multiplication in $\mathfrak{R}_8$ is not associative. (J. F. Adams has shown that these are the only real division algebras.)

For each fixed q, let $u_1, \cdots, u_q$ be units for $\mathfrak{R}_q$ and regard E^{2q} as a two-dimensional $\mathfrak{R}_q$-space, by associating with each $(x_1, \cdots, x_q, y_1, \cdots, y_q) \in E^{2q}$ the ordered pair $(X, Y) \in \mathfrak{R}_q \times \mathfrak{R}_q$ where $X = \sum_{i=1}^{q} x_i u_i$, $Y = \sum_{i=1}^{q} y_i u_i$. The $\mathfrak{R}_q$-lines, $Y = AX$ and $X = 0$, correspond to q-dimensional hyperplanes in E^{2q} through the origin. No two of these lines have points in common other than the origin: This is clear for a line $Y = AX$ and the line $X = 0$, so we need verify it only for two lines $Y = AX$, $Y = A'X$, where $A \neq A'$. If $Y = AX = A'X$, then $(A - A')X = 0$ and, since $\mathfrak{R}_q$ is a division algebra, we conclude first that $X = 0$ and then that $Y = 0$.

Now let S^{2q-1} be the unit sphere in E^{2q}. Each $\mathfrak{R}_q$-line through the origin intersects S^{2q-1} in a great S^{q-1} and it follows from the above remark that these S^{q-1} form a pairwise disjoint closed covering of S^{2q-1}. Let R be the equivalence relation generated by this covering, and let $p: S^{2q-1} \to S^{2q-1}/R$ be the projection on the quotient space. It is easy to see that S^{2q-1}/R is homeomorphic to S^q: indeed, regarding S^q as the one-point compactification of E^q, the map $S^{2q-1}/R \to S^q$ that associates with the equivalence class determined by $Y = AX$ the point $A \in E^q$, and with that determined by $X = 0$ the point ∞, can be verified to be a homeomorphism.

For $q = 1, 2, 4, 8$, we therefore have a fiber structure (S^{2q-1}, p, S^q) in which the fibers are great S^{q-1}; these are the Hopf fiber structures, and we now show them to be locally trivial.

For each $b \in S^q$, let $L(b)$ be the $\mathfrak{R}_q$-plane of the great S^{q-1} that maps to b. For $e \in S^{2q-1}$, let $\pi(b, e)$ be the orthogonal projection of e into $L(b)$, and denote its distance from the origin by $|\pi(b, e)|$. Then $|\pi(b, e)|$ is uniformly continuous on $S^q \times S^{2q-1}$ and because $|\pi(pe, e)| = 1$, there is an $\varepsilon > 0$ such that $|\pi(b, e)| > 0$

whenever $|b - p(e)| < \varepsilon$. For the ball U of radius ε centered at b, define $\sigma_U: U \times p^{-1}(U) \to S^{2q-1}$ by

$$\sigma_U(b, e) = \pi(b, e)/|\pi(b, e)|.$$

It is easy to verify that σ_U is a slicing function.

We give one simple application of Ex. 4. The projections $p: S^{2q-1} \to S^q$ in the Hopf fiber structures are called Hopf maps.

4.5 Theorem (H. Hopf) For $q = 1, 2, 4, 8$, the Hopf maps $p: S^{2q-1} \to S^q$ are not nullhomotopic.

Proof: The Hopf fiber structures (S^{2q-1}, p, S^q) are locally trivial so, by **4.2**, they are Hurewicz fibrations. Since the identity map $1: S^{2q-1} \to S^{2q-1}$ is irreducible, it follows from **1.4** that $p: S^{2q-1} \to S^q$ is also irreducible, and therefore is not nullhomotopic.

Problems

Section 1

1. Let B be a path-connected space, and let $P(B, b_0) \subset B^I$ be the space of all paths in B starting at b_0. Let $p: P(B, b_0) \to B$ be the map $p(\alpha) = \alpha(1)$. Show that $(P(B, b_0), p, B)$ is a Hurewicz fibration. [Note that $p^{-1}(b_0) = \Omega(B, b_0)$.]
2. Let (E, p, B) and (L, q, C) be Hurewicz fibrations. Show that $(E \times L, p \times q, B \times C)$ is also a Hurewicz fibration.
3. Let (E, p, B) be a Hurewicz fibration, and assume that some fiber $p^{-1}(b)$ is path-connected. Prove: E is path-connected if and only if B is path-connected.
4. Let (E, p, B) be a fiber structure, and let $g: X \to B$. Prove:
   a. g is liftable into E if and only if the induced fiber structure $(E(g), q, X)$ has a cross section.
   b. There is a one-to-one correspondence between the liftings of g into E and the cross sections in $E(g)$.
5. Let (E, p, B) and $q: X \to B$ be given. Show that $(E(g), q, X)$ has the following property: For any fiber structure (L, s, X) over X, and any commutative diagram

$$\begin{array}{ccc} L & \xrightarrow{\pi'} & E \\ {\scriptstyle s}\downarrow & & \downarrow{\scriptstyle p} \\ X & \xrightarrow[g]{} & B \end{array}$$

there exists a (unique) continuous $\varphi: L \to E(g)$ such that $\pi' = \pi \circ \varphi$ and such that the diagram

$$\begin{array}{ccccc} L & \xrightarrow{\varphi} & E(g) & \xrightarrow{\pi} & E \\ {\scriptstyle s}\downarrow & & \downarrow{\scriptstyle q} & & \downarrow{\scriptstyle p} \\ X & \xrightarrow[1]{} & X & \xrightarrow[g]{} & B \end{array}$$

commutes in each square.

Section 2

1. Let (E, p, B) be a fiber structure, and let $A \subset B$. Let $i: A \to B$ be the inclusion map, and let $(E(i), q, A)$ be the induced structure. The projection $\pi: E(i) \to E$ is a fiber map. Prove: $\pi: E(i) \cong p^{-1}(A)$.

2. Let (E, p, B) be a fiber structure, let $g: X \to B$ be continuous, and let $(E(g), q, X)$ be the induced structure. Given a continuous $h: Y \to X$, let $(E(g \circ h), q_1, Y)$ be the fiber structure induced by $g \circ h: Y \to B$ and let $(E(g)(h), q_2, Y)$ be that induced by h. Prove that the map $(y, e) \to (y, [h(y), e])$ is a fiber-preserving homeomorphism of $E(g \circ h)$ onto $E(g)(h)$. [This is called the "canonical homeomorphism".]

3. Let (E, p, B) be a Hurewicz fibration, and let $f, g: X \to B$ be homotopic. Prove that the induced structure $(E(g), q, X)$ is fiber-homotopy equivalent to $(E(f), q_1, X)$.

4. Let (E, p, B) be a fibration for the class $\mathscr{A}$ of spaces. Prove that if Δ belongs to $\mathscr{A}$, then (E, p, B) is a Hurewicz fibration.

5. Let (E, p, B) be a Hurewicz fibration. Let λ be a lifting function, and define $\mu: E^I \to E^I$ by $\mu(\alpha) = \lambda[\alpha(0), p \circ \alpha]$. Prove: $\mu \simeq 1$ and a homotopy Φ can be selected so that $p \circ \Phi(\alpha, s)[t] = p \circ \alpha[t]$ for all α, s, t.

6. Let $(E, p, B \times I)$ be a Hurewicz fibration. Show that $(E, p, B \times I)$ is fiber-homotopy equivalent to $(X \times I, q, B \times I)$ where $X = p^{-1}(B \times 0)$ and $q(x, t) = (p(x), t)$.

7. Let (E, p, B) be a Hurewicz fibration, and assume that B is path-connected. Prove that all the fibers belong to the same homotopy type.

8. Let Z be a locally compact normal space, and let $z_0 \in Z$ be a G_δ and strong deformation retract of Z. Let $b_0 \in B$ be fixed, and let $F = \{f \in B^Z \mid f(z_0) = b_0\}$. Prove that F is an H-space.

Section 3

1. Let (E, p, B) be a Hurewicz fibration, where B is paracompact and E is metric. Assume that each $b \in B$ has a nbd U deformable over B into b. Prove that (E, p, B) is a regular Hurewicz fibration. [Use Remark 2 of **2.7** to see that, if U is deformable over B into b, U is metrizable.]

Section 4

1. Let (E, p, B) be a locally trivial fiber structure. Show that $p: E \to B$ is an open mapping (and that therefore p is an identification).

2. Let (E, p, B) be a locally trivial fiber structure, and let $g: X \to B$ be continuous. Show that the induced structure $(E(g), q, X)$ is also locally trivial.

Appendix One

Vector Spaces; Polytopes

A. Vector Spaces

1. A set L together with a map $\alpha: L \times L \to L$ (we write $\alpha(a, b) = a + b$) is called an abelian group whenever:

1. $a + (b + c) = (a + b) + c$ for all a, b, c.
2. $a + b = b + a$ for all a, b.
3. There exists an element $0 \in L$ such that $a + 0 = a$ for every a.
4. For each a there is an element a' such that $a + a' = 0$.

It is simple to verify that 0 is unique, as also is the inverse a' of each $a \in L$; a' is written $-a$, and $a + (-b)$ is written $a - b$.

Let Λ be the set of real numbers with the usual addition and multiplication. An abelian group L together with a map $m: \Lambda \times L \to L$ [we write $m(\lambda, a) = \lambda a$] is called a real vector space whenever, for all λ, a, b:

(1). $\lambda(a + b) = \lambda a + \lambda b.$
(2). $(\lambda + \mu)a = \lambda a + \mu a.$
(3). $\lambda(\mu a) = (\lambda\mu)a.$
(4). $1a = a.$

Ex. 1 The elements of E^n, and also of any Hilbert space $l^2(\aleph)$, with the operations $\{x_\alpha\} + \{y_\alpha\} = \{x_\alpha + y_\alpha\}$, $\lambda\{x_\alpha\} = \{\lambda x_\alpha\}$, become real vector spaces. More generally, let $\{u_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of objects, and let $L(\mathscr{A})$ be the set of all formal sums $\sum_\alpha \lambda_\alpha u_\alpha$ in which all but at most finitely many $\lambda_\alpha = 0$. With the conventions $0 \cdot u_\alpha = 0$ and $1 \cdot u_\alpha = u_\alpha$ for each u_α, and with the operations

$$\lambda \sum_\alpha \lambda_\alpha u_\alpha = \sum_\alpha (\lambda\lambda_\alpha) u_\alpha,$$

$$\sum_\alpha \lambda_\alpha u_\alpha + \sum_\alpha \mu_\alpha u_\alpha = \sum_\alpha (\lambda_\alpha + \mu_\alpha) u_\alpha,$$

the set $L(\mathscr{A})$ is a real vector space, called the real vector space spanned by the u_α.

Two vector spaces are *isomorphic* if there is a bijection $\varphi: L \to L'$ such that $\varphi(\lambda a + \mu b) = \lambda\varphi(a) + \mu\varphi(b)$ for all λ, μ, a and b; that is, φ preserves the two operations.

Let $a, b \in L$; the set $\{c \mid c = \lambda b + (1 - \lambda)a;\ 0 \leqslant \lambda \leqslant 1\}$ is called the segment (or line) joining a to b. A subset $C \subset L$ is *convex* if for each pair $a, b \in C$, the segment joining them also lies in C. Clearly, L is convex, and so also is the intersection of any family of convex sets; thus, we have justified existence, uniqueness, and convexity in the

1.1 Definition Let $A \subset L$ be any set. The convex hull $H(A)$ of A is the intersection of all convex sets containing A.

We now express $H(A)$ directly in terms of A. For each finite set $\mathscr{F} = \{a_1, \cdots, a_n\} \subset A$, let

$$\sigma(\mathscr{F}) = \left\{y \in L \mid y = \sum_1^n \lambda_i a_i;\quad \lambda_i > 0;\quad \sum_1^n \lambda_i = 1\right\}.$$

$\sigma(\mathscr{F})$ is called the open simplex spanned by $a_1, \cdots, a_n$, and is also denoted simply by $(a_1, \cdots, a_n)$. We have

1.2 Let $A \subset L$. Then $H(A) = \bigcup \{\sigma(\mathscr{F}) \mid \mathscr{F} \subset A \text{ is finite}\}$.

Proof: Let $\Sigma = \bigcup \{\sigma(\mathscr{F}) \mid \mathscr{F} \subset A \text{ is finite}\}$. We remark first that $A \subset \Sigma$, since $(a) = a$ for each $a \in A$.

(1). $H(A) \subset \Sigma$. We need observe only that Σ is convex: Indeed, if $x \in (a_1, \cdots, a_n)$ and $y \in (a_1', \cdots, a_s')$, the line joining x to y lies in $(a_1, \cdots, a_n, a_1', \cdots, a_s') \subset \Sigma$.

(2). $\Sigma \subset H(A)$. We will show that each $\sigma(\mathscr{F}) \subset H(A)$; this is done by induction on the number n of elements in $\mathscr{F}$. For $n = 1$, the assertion is true, as has been remarked. Assuming its truth for n, we prove it for $n + 1$. Let

$$y_0 = \sum_1^{n+1} \lambda_i a_i$$

be any point in $(a_1, \cdots, a_{n+1})$; since

$$\mu = \sum_1^n \lambda_i \neq 0,$$

we let

$$y_1 = \sum_1^n \frac{\lambda_i}{\mu} a_i \in (a_1, \cdots, a_n).$$

Because $a_{n+1} \in A \subset H(A)$ and, by the induction hypothesis, also $y_1 \in H(A)$, it follows that all points on the line segment joining y_1 and a_{n+1} must lie in $H(A)$. Noting that $y_0 = \mu y_1 + (1 - \mu)a_{n+1}$ is such a point completes the inductive step.

A finite set of points $a_1, \cdots, a_n$ in L is called linearly independent if the condition $\lambda_1 a_1 + \cdots + \lambda_n a_n = 0$ is satisfied only by $\lambda_1 = \cdots = \lambda_n = 0$; an arbitrary family is linearly independent if each finite subfamily is linearly independent. A maximal linearly independent set $\{a_\alpha \mid \alpha \in \mathscr{A}\}$ in L is called a *basis* for L; each element of L can then be written in one and only one way as a *finite* sum $\sum_\alpha \lambda_\alpha a_\alpha$. In II, **2.5**, we proved that every vector space has a basis; it is not difficult to show (also using Zorn's lemma) that all bases for a given L have the same cardinal number; if this is a finite cardinal n, then L is called n-dimensional.

Any two real vector spaces having equipotent bases are isomorphic; since the vector space $L(\mathscr{A})$ in Ex. 1 has the set $\{u_\alpha \mid \alpha \in \mathscr{A}\}$ as a basis, it follows that any vector space is, up to an isomorphism, simply $L(\mathscr{A})$ for an $\mathscr{A}$ of suitable cardinal, and that the n-dimensional ones are isomorphic to E^n.

A linear subspace (or linear manifold) M in L is a subset with the property: If $a, b \in M$, then $\lambda a + \mu b \in M$ for all λ, μ. A linear subspace is itself a vector space. If $a_1, \cdots, a_n$ are any n elements of L, then the smallest linear manifold (that is, intersection of all linear subspaces) containing $a_1, \cdots, a_n$ is

$$\left\{\sum_1^n \lambda_i a_i \mid \lambda_i \text{ arbitrary}\right\};$$

it is called the linear subspace spanned by $a_1, \cdots, a_n$, and is n-dimensional if and only if the points $a_1, \cdots, a_n$ are linearly independent.

A k-flat (or linear k-variety) is a set $a + M$, where M is a k-dimensional linear manifold in L; it corresponds to a k-plane not necessarily through the origin. The smallest flat containing $(k + 1)$ points $a_0, \cdots, a_k$ is

$$\left\{\sum_0^k \lambda_i a_i \,\middle|\, \sum_0^k \lambda_i = 1\right\},$$

called the flat spanned by $a_0, \cdots, a_k$; it is a k-flat if and only if the k vectors $a_1 - a_0, \cdots, a_k - a_0$ are linearly independent.

A set of points $\{a_\alpha \mid \alpha \in \mathscr{A}\}$ in an n-dimensional vector space is in *general position* if, for each $k < n$, no $k + 2$ of them lie on a k-flat. For the vector space E^n, we have used the following result in the text:

1.3 Given any sequence $\{a_i \mid i \in Z^+\}$ of points in E^n, and any sequence $\{\varepsilon_i \mid i \in Z^+\}$ of positive numbers, there exists a sequence of points $\{p_i \mid i \in Z^+\}$ in general position such that $d(a_i, p_i) < \varepsilon_i$ for each i.

Proof: We proceed by induction. Choose $p_1 = a_1$, and assume that $p_1, \cdots, p_s$ have been selected to be in general position. Consider all k-flats, $k < n$, spanned by families of the p_i. Since there are only finitely many such flats, there is a point $p_{s+1} \in B(a_{s+1}, \varepsilon_{s+1})$ not on any of them; then $p_1, \cdots, p_{s+1}$ are in general position and the inductive step is complete.

2. Definition A linear topological space is a vector space L equipped with a Hausdorff topology such that the two maps $\alpha: L \times L \to L$ and $m: \Lambda \times L \to L$ (Euclidean topology on Λ) are continuous.

In this definition, we have assumed that the topology is Hausdorff; it is not difficult to show that even if it is assumed to be only T_1 (*cf.* VII, **1**, Ex. 3), then it is necessarily regular. With their metric topologies, E^n and all $l^2(\aleph)$ are linear topological spaces.

Since both $x \to x + a$ and its inverse $x \to x - a$ are continuous, each is a homeomorphism of L on itself, and therefore the translation $U + a$ of any open U is also open; similarly, for any $\lambda \neq 0$ and open V, the set λV is also open.

2.1 Theorem (A. Tychonoff) Every n-dimensional linear subspace M of a linear topological space L is topologically isomorphic to the Euclidean n-space E^n (that is, there is a $\varphi: E^n \to M$ that is simultaneously an algebraic isomorphism and a topological homeomorphism).

Proof: Let $a_1, \cdots, a_n$ be a basis for M and define $\varphi: E^n \to M$ by

$$\varphi(\lambda_1, \cdots, \lambda_n) = \sum_1^n \lambda_i a_i.$$

This is an isomorphism of E^n onto M, and it is continuous, since the algebraic operations are continuous in L. It remains only to prove that φ is an open map, and since open sets are invariant under translation, we need show only that the image of each ball $B = B(0; \varepsilon)$ is open. To this end, note that $\text{Fr}(B)$ is compact and does not contain 0; thus the compact set $\varphi[\text{Fr}(B)]$ does not contain the origin ω of L, so there is a nbd $W \supset \omega$ such that $\varphi[\text{Fr}(B)] \cap W = \varnothing$. By continuity of the multiplication, there is a $\delta > 0$ and a nbd V of ω such that $\lambda V \in W$ for all $|\lambda| \leqslant \delta$; we now show that $\delta V \subset \varphi(B)$. Indeed, if some $u = \delta v_0$ were not in $\varphi(B)$, then $\varphi^{-1}(u) \bar{\in} B$, so we could find a $\rho \leqslant 1$ such that $\rho \cdot \varphi^{-1}(u) = \varphi^{-1}(\rho u) \in \text{Fr}(B)$. Then $\rho u = \rho\, \delta v_0 \in \varphi[\text{Fr}(B)]$, and since $|\rho\delta| \leqslant \delta$, also $\rho\, \delta v_0 \in W$, which is impossible.

A linear topological space L is locally convex if for each $a \in L$ and nbd $U(a)$ there is a convex nbd V such that $a \in V \subset U(a)$. To prepare for Tychonoff's fixed-point theorem, we make the trivial general observation:

Lemma. Let X be Hausdorff and $f: X \to X$ be continuous. Then f has a fixed point if and only if for each open covering $\{W_\alpha\}$ of X there is at least one $x \in X$ such that both x and $f(x)$ belong to a common W_α.

Proof: "Only if" is trivial. "If": assume that f has no fixed point. For each $x \in X$ find nbds $W(x)$, $U(f(x))$ such that $W \cap U = \varnothing$, and $f(W) \subset U$; then $\{W(x)\}$ does not have the stated property.

2.2 Theorem (A. Tychonoff) Let L be a locally convex linear topological space and let C be a compact convex subset of L. Then each continuous $f: C \to C$ has a fixed point.

Proof: Let $\{W_\alpha\}$ be any open covering of C; we show that the property of the lemma is satisfied. It is clearly no restriction to assume that each W_α is convex. Let $\{U_\beta\}$ be a star refinement (*cf.* VIII, **3.2**), which we reduce to a finite covering $U_1, \cdots, U_n$, and let $V_k = f^{-1}(U_k)$, $k = 1, \cdots, n$. For each k, choose an $x_k \in U_k \cap C$, and let H be the convex hull of $\{x_1, \cdots, x_n\}$; by **2.1**, H is homeomorphic to the unit s-ball V^s for some $s \leqslant n - 1$. Subdivide H into s-dimensional simplexes $\{\sigma\}$ so small that each $\bar{\sigma}$ is contained in some set V_k, and let $y_1, \cdots, y_q$ be the vertices. Define F on the vertices by

$$F(y_\iota) = x_k,$$

where x_k is the selected point in any U_k containing $f(y_\iota)$. Extend F linearly over each simplex to get a continuous $F: H \to H$.

For each $x \in H$, $f(x)$ and $F(x)$ lie in a common W_α. Indeed, if $x \in \bar{\sigma} \subset V_{\iota_0}$, then $f(x) \in f(\bar{\sigma}) \subset U_{\iota_0}$; furthermore, since the images

$$f(y_0), \cdots, f(y_s)$$

of the vertices of $\bar{\sigma}$ are all in U_{ι_0}, it follows that $f(y_0), \cdots, f(y_s)$ all lie in $\operatorname{St} U_{\iota_0} \subset W_\alpha$. Then $f(x) \in W_\alpha$, and since W_α is convex, also $F(x) \in W_\alpha$.

This established, we observe that by Brouwer's fixed point theorem, F has a fixed point; thus f satisfies the condition of the lemma, and the theorem is proved.

3. A vector space L is normed if there exists a $p: L \to \Lambda$ such that:

(1). $p(a) \geqslant 0$ for all $a \in L$.
(2). $p(a) = 0$ if and only if $a = 0$.
(3). $p(a + b) \leqslant p(a) + p(b)$ for all a, b.
(4). $p(\lambda a) = |\lambda| p(a)$ for all λ, a.

We write $p(a) = \|a\|$, and call $\|a\|$ the norm of a.

3.1 If a vector space L is normed, then:

(a). $d(a, b) = \|a - b\|$ is a metric in L.
(b). With the metric topology, L becomes a locally convex linear topological space.

Proof: Verification that $d(a, b)$ is a metric is trivial. To see that L with the metric topology is a linear topological space, note that

$$\begin{aligned} d(a + b, a_0 + b_0) &= \|(a + b) - (a_0 + b_0)\| \\ &\leqslant \|a - a_0\| + \|b - b_0\| \\ &= d(a, a_0) + d(b, b_0) \\ d(\lambda a, \lambda_0 a_0) &= \|\lambda a - \lambda a_0 + \lambda a_0 - \lambda_0 a_0\| \\ &\leqslant |\lambda|\, d(a, a_0) + |\lambda - \lambda_0|\, \|a_0\|. \end{aligned}$$

The maps $\alpha: L \times L \to L$ and $m: \Lambda \times L \to L$ are therefore continuous. Each ball $B(c; r)$ is convex, since if $a, b \in B(c, r)$, then for all $0 \leqslant \lambda \leqslant 1$, we find

$$d(c, \lambda a + (1 - \lambda)b) = \|\lambda(c - a) + (1 - \lambda)(c - b)\| < r.$$

Consequently, the linear topological space is locally convex.

A complete normed linear topological space is called a Banach space. Thus, with the evident norms, E^n and all $l^2(\aleph)$ are Banach spaces.

Tychonoff's fixed-point theorem is not immediately applicable in analysis because the domain is required to be compact, a situation rarely met in practice. It assumes a more practical form in Banach spaces. The convex closure $\widehat{B}$ of a set B is the intersection of all closed convex sets containing B; in Banach spaces, it is not hard to show that $\widehat{B}$ is compact whenever B is compact.

3.2 Theorem (J. Schauder) Let A be a closed convex set in a Banach space and $f: A \to A$ be continuous. If $\overline{f(A)}$ is compact, then f has a fixed point.

Proof: Since $\overline{f(A)}$ is compact, we have that $\widehat{f(A)}$ is compact; since A is closed and convex, we find from $\overline{f(A)} \subset \overline{A} = A$ that $\widehat{f(A)} \subset \widehat{A} = A$. Thus $f \mid \widehat{f(A)}: \widehat{f(A)} \to \widehat{f(A)}$, and so, Tychonoff's theorem applies.

4. By taking Tychonoff's theorem, **2.1**, as the basic requirement that a topology in a real vector space must satisfy, we are led to a class of spaces broader than the linear topological spaces.

4.1 Definition An affine space is a real vector space with any topology that induces the Euclidean topology on its finite-dimensional flats.

Clearly, every linear topological space is an affine space. In the text, we are primarily concerned with vector spaces topologized as *affine* spaces rather than as linear topological spaces, and more particularly, with the largest topology that does this:

4.2 Definition The finite topology in a real vector space is the weak topology determined by the Euclidean topology on each finite-dimensional linear subspace.

Equivalently, the finite topology is the weak topology determined by the Euclidean topology on each finite-dimensional flat.

We remark that all vector spaces with the finite topology are completely and perfectly normal; furthermore, they are always paracompact, and *every* subset (not only the closed subsets) is also paracompact.

We next consider the continuity of the algebraic operations in affine spaces. Addition and multiplication are evidently continuous when confined to points varying on finite dimensional flats. However,

4.3 An affine space need not be a linear topological space. In fact, if a vector space has a basis of cardinal $\geqslant 2^{\aleph_0}$, then with the finite topology it is neither a linear topological space nor a locally convex space.

Proof: The following proof is simply an immediate adaptation of C. H. Dowker's example (VI, **8**, Ex. 5). Let L be a vector space having a basis $\mathscr{B}$ such that $\aleph(\mathscr{B}) \geqslant 2^{\aleph_0}$. Let $\{u_n\}$ be a set of basis vectors in fixed 1-to-1 correspondence with the positive integers, and let $\{u_f\} \subset \mathscr{B} - \{u_n \mid n \in Z^+\}$ be a subset in a fixed 1-to-1 correspondence with the set $\mathscr{M}$ of all maps f of the positive integers into themselves. Since $\aleph(\{u_f\}) = 2^{\aleph_0}$ and $\aleph(\{u_n\}) = \aleph_0$, two such subsets of $\mathscr{B}$ certainly exist.

For each n and f, let

$$a_{n,f} = \frac{1}{f(n)} u_n + \frac{1}{f(n)} u_f.$$

Clearly, each $a_{n,f} \neq 0$. Let

$$A = \{a_{n,f} \mid (n,f) \in Z^+ \times \mathscr{M}\};$$

since each $a_{n,f}$ lies in the linear subspace spanned by the linearly independent u_n, u_f, a finite-dimensional linear subspace can contain at most finitely many members of A. Thus the intersection of A with each finite-dimensional linear subspace is closed in the Euclidean topology of the linear subspace, and therefore A is a closed set in the finite topology of L.

Consider now the nbd $U = \mathscr{C}A$ of the origin. We will prove that addition is not continuous at 0 by showing that there is no nbd W of 0 such that $W + W \subset U$. Indeed, given any W, then for each $u_\alpha \in \mathscr{B}$, we can determine a real number $\lambda_\alpha > 0$ such that $\lambda u_\alpha \in W$ for all $0 \leqslant \lambda < \lambda_\alpha$. Define now a map $\varphi: Z^+ \to Z^+$ by $\varphi(n) = \max[n, (1/\lambda_n)] + 1$; then $\varphi \in \mathscr{M}$, so it corresponds to some u_φ, and we choose $\bar{n}$ such that $\varphi(\bar{n}) > 1/\lambda_\varphi$. It follows that both

$$\frac{1}{\varphi(\bar{n})}\, u_n \quad \text{and} \quad \frac{1}{\varphi(\bar{n})}\, u_\varphi$$

are in W; yet, their sum $a_{n,\varphi}$ is not in U.

Neither is the space L locally convex: If $F = \frac{1}{2}\cdot A$, then $V = \mathscr{C}F$ is a nbd of 0 containing no convex nbd of 0. For, as we have seen, any nbd W of the origin contains some $\dfrac{1}{f(n)}\, u_n$ and $\dfrac{1}{f(n)}\, u_f$, but the mid-point of the segment containing them is not in V.

Though vector spaces with the finite topology need not be locally convex, they exhibit a behavior analogous to local convexity in certain instances:

4.4 Let X be either locally compact, or first countable, and let L be any vector space with the finite topology. If $f: X \to L$ is continuous, then each $x \in X$ has a nbd $U(x)$ mapped into some finite-dimensional flat.

Proof: Assume that X is first countable and that the assertion is false at x_0. Let $\{V_n \mid n \in Z^+\}$ be a nbd basis at x_0 (where $V_1 \supset V_2 \supset \cdots$); by induction we construct a sequence $\{x_n\}$, with $x_n \in V_n$, as follows: Choose $x_1 \in V_1$ such that $f(x_1) \neq f(x_0)$, and having selected $x_1, \cdots, x_{n-1}$, let $x_n \in V_n$ be such that $f(x_n)$ is not in the linear subspace spanned by $f(x_1), \cdots, f(x_{n-1})$. Then

$$A = \bigcup_1^\infty f(x_n)$$

is closed in L (in fact, it is discrete), since each finite-dimensional linear subspace contains at most finitely many points of A. Since $x_n \to x_0$, each nbd of x_0 has points with images not in the nbd $L - A$ of $f(x_0)$, which contradicts the assumed continuity of f at x_0 and completes the proof. The demonstration is similar for locally compact X: Each nbd with compact closure maps into a finite dimensional linear subspace.

This leads to the class of spaces described in the text (IX, **6**):

4.5 Definition An affine space is of type m if for each first countable space X and every continuous $f: X \to L$, the following is true: For each $x \in X$ and nbd $W \supset f(x)$, there exists a nbd $U \supset x$ and a convex set $C \subset L$ such that $f(U) \subset C \subset W$.

Thus all locally convex linear topological spaces (and also all vector spaces with the finite topology) are of type m.

B. Polytopes

5. By a triangulation of a set X is meant a family $u = \{\bar{\sigma}_\alpha \mid \alpha \in \mathscr{A}\}$ of closed geometric simplexes such that:

(1). $\bigcup \{\bar{\sigma}_\alpha \mid \alpha \in \mathscr{A}\} = X$.
(2). Each face of a $\bar{\sigma}_\alpha \in u$ also belongs to u.
(3). For each $(\alpha, \beta) \in \mathscr{A} \times \mathscr{A}$, $\bar{\sigma}_\alpha \cap \bar{\sigma}_\beta$ is either empty or a face of both $\bar{\sigma}_\alpha$ and $\bar{\sigma}_\beta$.

We do not require the dimensions of the simplexes to have a finite upper bound, and we allow the set of simplexes incident with any given one to have any finite or transfinite cardinal number. A set X together with a definite triangulation u is called a geometric complex, and is denoted by (u, X).

If (u, X) is any geometric complex, the topological space consisting of the set X together with the weak topology in X, determined by the Euclidean topology on each closed $\bar{\sigma} \in u$, is called a polytope and is denoted by $X(u)$. The n-simplexes of u are called the n-cells of $X(u)$.

As in VIII, **5.2** *et seq.*, every polytope $X(u)$ possesses a (homeomorphic) model that has its vertices at the unit points in a vector space with finite topology. $X(u)$ and each of its models are homeomorphic in such a way that cells are mapped linearly onto cells.

By a subdivision of a given geometric complex (u, X) is meant a geometric complex (u_1, X) such that:

(a). Each closed simplex of u_1 is contained in a closed simplex of u.
(b). Each closed simplex of u is the union of at most finitely many closed simplexes of u_1.

It is trivial to verify that the polytope $X(u_1)$ is then homeomorphic to the polytope $X(u)$.

The barycentric subdivision (u', X) of (u, X) is defined as follows:

(i). Vertices: the barycenter p_σ of each open simplex σ.
(ii). Simplexes: $(p_{\sigma_1}, \cdots, p_{\sigma_n})$ is a simplex of u' if and only if in the finite sequence $\sigma_1, \cdots, \sigma_n$, each simplex is a proper face of its successor.

A generalization is the barycentric subdivision mod Q, where Q is a subcomplex of (u, X); here, the vertices are those of the $\bar{\sigma} \in u$ and the barycenters of all open simplexes not in Q (so that Q is not subdivided).

Let $X(u)$ be a polytope. With each subdivision (v, X) of (u, X) is associated a covering of the space $X(u)$ by the (open) stars of the vertices of $X(v)$. We now show that "arbitrarily fine" coverings can be obtained in this way.

5.1 Theorem (J. H. C. Whitehead) Let $X(u)$ be a polytope. Given any open covering of $X(u)$, there exists a subdivision (v, X) of (u, X) such that each closed vertex star of $X(v)$ is contained in some set of the given covering.

Proof: Denote by $X^{(k)}$ the union of all closed k-cells of $X(u)$ [that is, the k-skeleton of $X(u)$]. Assume that there is a subdivision X_1^{n-1} of $X^{(n-1)}$ that satisfies the requirements of the theorem. With each vertex p' of X_1^{n-1}, associate a definite open set $U(p')$ of the given open covering such that the closed star of p' in X_1^{n-1} is contained in $U(p')$. Let $\bar{\sigma}^n$ be a closed n-cell of $X(u)$ and select an $\varepsilon > 0$ such that (1) each subset of $\bar{\sigma}^n$ having diameter $< 2\varepsilon$ is contained in some set of the given covering of $X(u)$; and (2) for each vertex p' on $Q = X_1^{n-1} \cap \mathrm{Fr}(\sigma^n)$, the ε-nbd of the closed star of p' in Q is contained in the selected $U(p')$ (*cf.* XI, **4.5**).

Draw the cone having apex at the barycenter of σ^n over each $(n-1)$-cell on Q to obtain a subdivision of $\bar{\sigma}^n$. Perform repeated barycentric subdivisions mod Q until each cell of the subdivision satisfies either that (a) its diameter is $< \varepsilon$, or (b) that it lies completely in an ε-nbd of a cell on the boundary. Repeating for each n-cell of $X(u)$ clearly yields a subdivision of $X^{(n)}$ that does not alter the given subdivision of $X^{(n-1)}$ and which is easily seen to verify the requirements of the theorem. The subdivision as required of the one-skeleton of $X(u)$ being trivial, the theorem follows by induction.

5.2 Theorem (J. Dugundji) Every polytope is paracompact.

Proof: Let $\{U\}$ be an open covering of $X(u)$; since there is a subdivision such that each open vertex star lies in a member of $\{U\}$, the problem reduces to showing that the open covering of a polytope by its vertex stars has a nbd-finite refinement.

Analogously as in VIII, **5.2**, we embed the geometric complex (u, X) in a suitable generalized Hilbert space $l^2(\aleph)$ so that its vertices are at the unit points of $l^2(\aleph)$. X, with the topology as a subspace of $l^2(\aleph)$ is a metric space, denoted by X_m. The "identity map" $i: X(u) \to X_m$ is continuous, since it is so on each closed cell of $X(u)$; i is bijective and carries vertex stars to vertex stars. Since the vertex stars in X_m are easily seen to be open sets, and since metric spaces are paracompact, the open covering of X_m by its vertex stars has a nbd-finite refinement $\{V\}$. But then $\{i^{-1}(V)\}$ is a nbd-finite refinement of the covering of $X(u)$ by its vertex stars, completing the proof.

Appendix Two

Direct and Inverse Limits

Direct and inverse limits of spaces are frequently used in modern topology. In this Appendix, we shall define these two "limiting" processes and derive some of their simpler properties.

1. Direct Limits

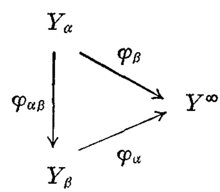
1.1 Definition Let $\mathscr{A}$ be a directed set (*cf.* X, **1.2**) and let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces indexed by $\mathscr{A}$. For each pair of indices α, β satisfying $\alpha \prec \beta$, assume that there is given a continuous map $\varphi_{\alpha\beta}: Y_\alpha \to Y_\beta$, and that these maps satisfy the condition: if $\alpha \prec \beta \prec \gamma$, then $\varphi_{\alpha\gamma} = \varphi_{\beta\gamma} \circ \varphi_{\alpha\beta}$. Then this family $\{Y_\alpha; \varphi_{\alpha\beta}\}$ of spaces and maps is called a direct (or inductive) spectrum over $\mathscr{A}$, with spaces Y_α and connecting maps $\varphi_{\alpha\beta}$.

The image of a $y_\alpha \in Y_\alpha$ under any connecting map is termed a successor of y_α. Each direct spectrum $\{Y_\alpha; \varphi_{\alpha\beta}\}$ yields a limit space in the following way: Let $D = \Sigma\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be the free union of the spaces, and call two elements $y_\alpha \in Y_\alpha$, $y_\beta \in Y_\beta$ in D equivalent whenever they have a common successor in the spectrum. This relation, R, is in fact an equivalence relation in D; that it is reflexive and symmetric is obvious. To see that R is transitive, assume that y_α, y_β have a common successor in Y_ρ and that y_β, y_γ have one in Y_σ; then, because $\mathscr{A}$ is directed, there is an index τ such that $\rho \prec \tau$, $\sigma \prec \tau$, and the successor of y_β in Y_τ is evidently a successor of both y_α and y_γ.

1.2 Definition Let $\{Y_\alpha, \varphi_{\alpha\beta}\}$ be a direct spectrum. The quotient space $\sum_\alpha Y_\alpha/R$ is called the direct (or inductive) limit space of the spectrum, and is denoted by Y^∞. (Other notations commonly used are $\underrightarrow{\text{Lim}}\, Y_\alpha$ and $\text{Lim Dir}_{\mathscr{A}}\, Y_\alpha$.)

Let $p: \sum_\alpha Y_\alpha \to Y^\infty$ be the projection; its restriction $p \mid Y_\alpha$ is denoted by φ_α and is called the canonical map of Y_α into Y^∞. Note that each φ_α is continuous and that $G \subset Y^\infty$ is open (closed) if and only if $\varphi_\alpha^{-1}(G)$ is open (closed) in Y_α for each $\alpha \in \mathscr{A}$. Clearly, $Y^\infty = \bigcup\{\varphi_\alpha(Y_\alpha) \mid \alpha \in \mathscr{A}\}$, so that Y^∞ is not empty whenever at least one $Y_\alpha \neq \varnothing$. For the relations between the φ_α and the $\varphi_{\alpha\beta}$, we have

1.3 (1). Whenever $\alpha \prec \beta$, the diagram



is commutative.

(2). For any α, β we have $\varphi_\alpha(y_\alpha) = \varphi_\beta(y_\beta)$ if and only if there is some γ such that $\alpha, \beta \prec \gamma$ and $\varphi_{\alpha\gamma}(y_\alpha) = \varphi_{\beta\gamma}(y_\beta)$.

Proof: (1). For each $y_\alpha \in Y_\alpha \subset D$, we have y_α equivalent to $\varphi_{\alpha\beta}(y_\alpha)$, so $\varphi_\alpha(y_\alpha) = p(y_\alpha) = p \circ \varphi_{\alpha\beta}(y_\alpha) = \varphi_\beta \circ \varphi_{\alpha\beta}(y_\alpha)$.

(2). $[\varphi_\alpha(y_\alpha) = \varphi_\beta(y_\beta)] \Leftrightarrow [y_\alpha$ is equivalent to y_β in $D] \Leftrightarrow [y_\alpha, y_\beta$ have a common successor$]$.

We call $y_\alpha \in Y_\alpha \subset D$ a representative of $y \in Y^\infty$ whenever $\varphi_\alpha(y_\alpha) = y$; it is clear that a given $y \in Y^\infty$ need not have a representative in each Y_α. However, if finitely many $y_1, \cdots, y_n \in Y^\infty$ are given, it is always possible to find at least one Y_σ that contains representatives for all the given y_i: this follows by induction, once it has been proved for $n = 2$ and in this case, choose representatives $y_\alpha \in Y_\alpha$, $y_\beta \in Y_\beta$ of y_1, y_2, respectively, find a γ such that $\alpha, \beta \prec \gamma$, and note that $\varphi_{\alpha\gamma}(y_\alpha)$, $\varphi_{\beta\gamma}(y_\beta)$ are the desired representatives in Y_γ.

Ex. 1 Let $\mathscr{A}$ be any directed set and Y be any space. The direct spectrum $\{Y_\alpha; \varphi_{\alpha\beta}\}$ in which each $Y_\alpha = Y$ and each $\varphi_{\alpha\beta} = 1$ is called the trivial direct spectrum over $\mathscr{A}$ with space Y. It is clear that $Y^\infty \cong Y$. In particular, any given space is the direct limit space of a suitable direct spectrum over any preassigned $\mathscr{A}$.

Ex. 2 Let $\mathscr{A}$ be the directed set of positive integers in their natural order. Define a spectrum $\{Y_n; \varphi_{nk}\}$ over $\mathscr{A}$ by taking each $Y_n = S^1$ and $\varphi_{n,k}(z) = z^{2^{k-n}}$ when $n \leqslant k$. Then Lim Dir Y_n is not a Hausdorff space. Indeed, for the point $y = \varphi_1(1) \in Y^\infty$, we find that $\varphi_1^{-1}(y) = \{$all 2^s roots of unity, all $s \geqslant 0\}$, which is a countable dense set in Y_1. Since $\varphi_1^{-1}(y)$ is not closed in Y_1, the point y is not closed in Y^∞ so Y^∞ is not Hausdorff. It is not difficult to verify that in fact Y^∞ carries the indiscrete topology.

Ex. 3 Let X be any space and fix an $x_0 \in X$. Let $\mathscr{A}$ be the set of all nbds of x_0, directed by $U \prec V \Leftrightarrow U \supset V$. For each $U \in \mathscr{A}$, let $C_U(Z)$ be the discrete set of all continuous maps of U into a fixed space Z, and for $U \prec V$, define $\varphi_{UV}(f) = f \mid V$. Then $\{C_U(Z); \varphi_{UV}\}$ is a direct spectrum, and the set Lim Dir $C_U(Z)$ is called the set of germs of continuous maps of a nbd of x_0 into Z. By **1.3**(2), two maps defined on nbds U, V of x_0 represent the same germ at x_0 if and only if they coincide on some nbd $W \subset U \cap V$ of x_0.

We now introduce the concept "mapping of direct spectra."

1.4 Definition Let $\{Y_\alpha; \varphi_{\alpha\beta}\}$, $\{Z_\alpha; \psi_{\alpha\beta}\}$ be two direct spectra over the same directed set $\mathscr{A}$. Assume that for each $\alpha \in \mathscr{A}$, there is given a map $h_\alpha\colon Y_\alpha \to Z_\alpha$ and that whenever $\alpha \prec \beta$, the diagram

$$\begin{array}{ccc} Y_\alpha & \xrightarrow{h_\alpha} & Z_\alpha \\ \varphi_{\alpha\beta}\big\downarrow & & \big\downarrow\psi_{\alpha\beta} \\ Y_\beta & \xrightarrow[h_\beta]{} & Z_\beta \end{array}$$

is commutative. Then the family $\{h_\alpha \mid \alpha \in \mathscr{A}\}$ is called a map of the direct spectrum $\{Y_\alpha; \varphi_{\alpha\beta}\}$ into $\{Z_\alpha; \psi_{\alpha\beta}\}$. The map is called continuous whenever each h_α is continuous.

The notion of a map of a space Y into a direct spectrum $\{Z_\alpha; \psi_{\alpha\beta}\}$ over $\mathscr{A}$ is obtained from **1.4** by regarding Y as the trivial direct spectrum over $\mathscr{A}$; similarly, we obtain the meaning of "map of a direct spectrum into a space." Each map of spectra leads to a unique map of the limit spaces:

1.5 Theorem Let $\{h_\alpha\}\colon \{Y_\alpha, \varphi_{\alpha\beta}\} \to \{Z_\alpha; \psi_{\alpha\beta}\}$ be a continuous map. Then there exists one, and only one, continuous map $h^\infty\colon Y^\infty \to Z^\infty$ having the property that for each $\alpha \in \mathscr{A}$, the diagram

$$\begin{array}{ccc} Y_\alpha & \xrightarrow{h_\alpha} & Z_\alpha \\ \varphi_\alpha\big\downarrow & & \big\downarrow\psi_\alpha \\ Y^\infty & \xrightarrow[h^\infty]{} & Z^\infty \end{array}$$

is commutative. Furthermore:

(1). If each h_α is injective, so also is h^∞.
(2). If each h_α is surjective, so also is h^∞.
(3). If each h_α is an identification, so also is h^∞.
(4). If each h_α is a homeomorphism, so also is h^∞.

Proof: Define $h: \Sigma Y_\alpha \to \Sigma Z_\alpha$ by $h \mid Y_\alpha = h_\alpha$; this is clearly continuous, and because the h_α commute with the connecting maps, h is also relation preserving. Passing to the quotient (VI, **4.3**) yields a unique, continuous $h^\infty: Y^\infty \to Z^\infty$ such that the diagram

$$\begin{array}{ccc} \Sigma Y_\alpha & \xrightarrow{\;h\;} & \Sigma Z_\alpha \\ {\scriptstyle p}\downarrow & & \downarrow{\scriptstyle q} \\ Y^\infty & \xrightarrow[h^\infty]{} & Z^\infty \end{array}$$

commutes, and the asserted property of h^∞, as well as unicity, follows at once (I, **7.7**).

Ad (1). Let $h^\infty(y) = h^\infty(y')$. Choose representatives $y_\alpha \in Y, y'_\beta \in Y_\beta$ for y, y', respectively; since

$$h^\infty(y) = h^\infty \circ \varphi_\alpha(y) = \psi_\alpha \circ h_\alpha(y_\alpha) = h^\infty(y') = \psi_\beta \circ h_\beta(y'_\beta),$$

we find from **1.3**(2) that there is a γ, such that $\alpha, \beta \prec \gamma$ and $\psi_{\alpha\gamma} \circ h_\alpha(y_\alpha) = \psi_{\beta\gamma} \circ h_\beta(y'_\beta)$. By commutativity, $h_\gamma \circ \varphi_{\alpha\gamma}(y_\alpha) = h_\gamma \circ \varphi_{\beta\gamma}(y'_\beta)$ so because h_γ is injective, we get $\varphi_{\alpha\gamma}(y_\alpha) = \varphi_{\beta\gamma}(y'_\beta)$, and by **1.3**(2), that $y = y'$.

Ad (2). If each h_α is surjective, then

$$Z^\infty = \bigcup_\alpha \psi_\alpha(Z_\alpha) = \bigcup_\alpha \psi_\alpha \circ h_\alpha(Y_\alpha) = \bigcup_\alpha h^\infty \circ \varphi_\alpha(Y_\alpha) = h^\infty(Y^\infty).$$

Ad (3). Since the map $h: \Sigma Y_\alpha \to \Sigma Z_\alpha$ is an identification, the result follows from VI, **4.3**.

Ad (4). This is an obvious consequence of (1)–(3).

The map h^∞ is called the direct limit map induced by the map $\{h_\alpha \mid \alpha \in \mathscr{A}\}$; it is also denoted by $\underrightarrow{\text{Lim}}\, h_\alpha$, and Lim Dir h_α.

1.6 Corollary (Transitivity) Let $\{g_\alpha\}: \{X_\alpha; \omega_{\alpha\beta}\} \to \{Y_\alpha; \varphi_{\alpha\beta}\}$ and $\{h_\alpha\}: \{Y_\alpha; \varphi_{\alpha\beta}\} \to \{Z_\alpha; \psi_{\alpha\beta}\}$ be maps. Then

$$\{h_\alpha \circ g_\alpha\}: \{X_\alpha; \omega_{\alpha\beta}\} \to \{Z_\alpha; \psi_{\alpha\beta}\}$$

is also a map, and

$$\text{Lim Dir}\,(h_\alpha \circ g_\alpha) = \text{Lim Dir}\, h_\alpha \circ \text{Lim Dir}\, g_\alpha.$$

Proof: It is trivial to verify that $\{h_\alpha \circ g_\alpha\}$ is indeed a map. Now note that $\psi_\alpha \circ (h_\alpha \circ g_\alpha) = h^\infty \circ \varphi_\alpha \circ g_\alpha = h^\infty \circ g^\infty \circ \omega_\alpha$ for each $\alpha \in \mathscr{A}$; by the unicity statement in **1.5**, we must therefore have $h^\infty \circ g^\infty = \underrightarrow{\text{Lim}}\, \{h_\alpha \circ g_\alpha\}$.

Let $\{Y_\alpha; \varphi_{\alpha\beta}\}$ be a direct spectrum over $\mathscr{A}$; we now consider how to cut down $\mathscr{A}$ without affecting Y^∞. It is clear that if $\mathscr{B} \subset \mathscr{A}$ is a directed subset, then $\{Y_\alpha; \varphi_{\alpha\beta} \mid \alpha, \beta \in \mathscr{B}\}$ is a direct spectrum over $\mathscr{B}$, and obvious examples show that with different choices of $\mathscr{B}$, we generally get different spaces $\text{Lim Dir}_{\mathscr{B}}\, Y_\beta$. Nevertheless, there is a certain type of subset $\mathscr{B}$ to which we can always restrict without affecting Y^∞.

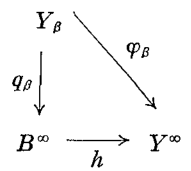
Call $\mathscr{B} \subset \mathscr{A}$ cofinal in $\mathscr{A}$ if $\forall\, \alpha \in \mathscr{A}\ \exists\, \beta \in \mathscr{B}$: $\alpha \prec \beta$. A cofinal $\mathscr{B} \subset \mathscr{A}$ is also a directed subset of $\mathscr{A}$: given $\beta, \beta' \in \mathscr{B}$, find $\gamma \in \mathscr{A}$ such that $\beta, \beta' \prec \gamma$, and use cofinality to obtain $\beta'' \in \mathscr{B}$ such that $\gamma \prec \beta''$.

1.7 Theorem Let $\mathscr{B}$ be cofinal in $\mathscr{A}$. Then

$$\text{Lim Dir}_{\mathscr{B}}\, Y_\beta \cong \text{Lim Dir}_{\mathscr{A}}\, Y_\alpha,$$

for any direct spectrum $\{Y_\alpha; \varphi_{\alpha\beta}\}$ over $\mathscr{A}$.

Proof: Let $B^\infty = \text{Lim Dir}_{\mathscr{B}}\, Y_\beta$ and let $q_\beta\colon Y_\beta \to B^\infty$ be the canonical map. Let $\varphi_\beta\colon Y_\beta \to Y^\infty$ be the canonical map of the $\mathscr{A}$-spectrum. It is evident that $\{\varphi_\beta \mid \beta \in \mathscr{B}\}$ is a continuous map of the $\mathscr{B}$-spectrum into Y^∞ so **1.5** gives a continuous $h\colon B^\infty \to Y^\infty$ such that the diagram



commutes for each $\beta \in \mathscr{B}$.

(1). h is injective. Let $h(b) = h(b')$; as previously remarked, we can always find representatives b_β, b'_β, for b, b', lying in some one Y_β, $\beta \in \mathscr{B}$. Since $\varphi_\beta(b_\beta) = h \circ q_\beta(b_\beta) = h(b) = h(b') = \varphi_\beta(b'_\beta)$, **1.3**(2) shows that b_β, b'_β, have a common successor in some Y_α, $\alpha \in \mathscr{A}$; since $\mathscr{B}$ is cofinal, b_β, b'_β, have a common successor in some Y_γ, $\gamma \in \mathscr{B}$, so they represent the same element in B^∞ and therefore $b = b'$.

(2). h is surjective. We first note that

$$\bigcup \{\varphi_\alpha(Y_\alpha) \mid \alpha \in \mathscr{A}\} = \bigcup \{\varphi_\beta(Y_\beta) \mid \beta \in \mathscr{B}\},$$

since given any $\varphi_\alpha(Y_\alpha)$, the cofinality gives some $\beta \in \mathscr{B}$, $\alpha \prec \beta$, and then $\varphi_\alpha(Y_\alpha) = \varphi_\beta \circ \varphi_{\alpha\beta}(Y_\alpha) \subset \varphi_\beta(Y_\beta)$. The required surjectivity now follows from $Y^\infty = \bigcup_\alpha \varphi_\alpha(Y_\alpha) = \bigcup_\beta \varphi_\beta(Y_\beta) = \bigcup_\beta h \circ q_\beta(Y_\beta) = h[\bigcup_\beta q_\beta(Y_\beta)] = h(B^\infty)$.

(3). h is an open map. Let $U \subset B^\infty$ be open; we are to show that $\varphi_\alpha^{-1}[h(U)]$ is open in Y_α for each $\alpha \in \mathscr{A}$. First, this is true for each $\beta \in \mathscr{B}$, since the bijectivity of h gives $\varphi_\beta^{-1}[h(U)] = q_\beta^{-1}h^{-1}[h(U)] = q_\beta^{-1}(U)$, and q_β is continuous. It now follows for each $\alpha \in \mathscr{A}$, since by choosing

$\beta \in \mathscr{B}$ such that $\alpha \prec \beta$, we have $\varphi_\alpha^{-1}[h(U)] = \varphi_{\alpha\beta}^{-1} \circ \varphi_\beta^{-1}[h(U)]$, and $\varphi_{\alpha\beta}$ is continuous.

The theorem therefore is proved.

This theorem frequently permits comparison of direct limit spaces when the direct spectra are over different directed sets. As an example: Let $\{Y_\alpha; \varphi_{\alpha\beta}\}$, $\{Z_\sigma; \psi_{\sigma\tau}\}$ be direct spectra over $\mathscr{A}$, $\mathscr{S}$, respectively. Assume there is a cofinal $\mathscr{B} \subset \mathscr{A}$ and a relation-preserving map r of $\mathscr{B}$ into $\mathscr{S}$ such that $r(\mathscr{B})$ is cofinal in $\mathscr{S}$. Then a continuous map

$$\{h_\beta \mid \beta \in \mathscr{B}\}: \{Y_\beta; \varphi_{\alpha\beta}\} \to \{Z_\sigma;\ Y_{\sigma\tau} \mid \sigma, \tau \in r(\mathscr{B})\}$$

induces a continuous $h\colon Y^\infty \to Z^\infty$, which is a homeomorphism whenever each h_β is. The simple proof of this result is left for the reader.

Direct limits behave poorly for subspaces and for cartesian products, in a sense that will be made precise for each case.

1.8 (Subspaces) Let $\{Y_\alpha; \varphi_{\alpha\beta}\}$ be a direct spectrum over $\mathscr{A}$. For each $\alpha \in \mathscr{A}$, let $A_\alpha \subset Y_\alpha$ and assume that $\varphi_{\alpha\beta}(A_\alpha) \subset A_\beta$ whenever $\alpha \prec \beta$. Then $\{A_\alpha; \varphi_{\alpha\beta} \mid A_\alpha\}$ is a direct spectrum over $\mathscr{A}$, and there is a continuous injection $h\colon A^\infty \to Y^\infty$.

Proof: Let $i_\alpha\colon A_\alpha \to Y_\alpha$ be the inclusion. Since $\{i_\alpha\}$ is obviously a map $\{A_\alpha; \varphi_{\alpha\beta} \mid A_\alpha\} \to \{Y_\alpha; \varphi_{\alpha\beta}\}$, **1.5** yields a continuous $h\colon A^\infty \to Y^\infty$, which is injective because each i_α is.

By the "poor behavior" of subspaces we mean that, in general, A^∞ is not homeomorphic to the subspace $h(A^\infty) \subset Y^\infty$.

Ex. 4 Let Y be the set $\{0\} \cup \{x \mid |x| > 1\}$ of real numbers taken with the order topology (*cf.* III, **7**, Problem 10). Let $\mathscr{A}$ be the directed set of positive integers, and take $\{Y_n; \varphi_{nk}\}$ to be the trivial spectrum over $\mathscr{A}$ with space Y. For each n, let $A_n \subset Y_n$ be the subspace $\{0\} \cup \{x \mid |x| > 1 + 1/n\}$; then $h\colon A^\infty \to Y^\infty$ is in fact bijective and satisfies $h\{0\} = \{0\}$; however, $\{0\}$ is open in A^∞, whereas it is not open in Y^∞.

1.9 (Cartesian Products) Let $\{Y_\alpha; \varphi_{\alpha\beta}\}$ and $\{Z_\sigma; \psi_{\sigma\tau}\}$ be direct spectra over directed sets $\mathscr{A}$, $\mathscr{S}$, respectively.

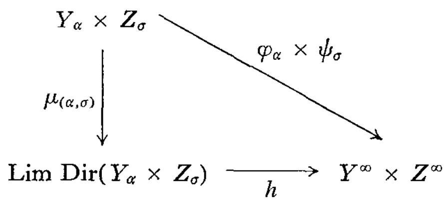
(1). Preorder $\mathscr{A} \times \mathscr{S}$ by $(\alpha, \sigma) \prec (\beta, \tau)$ if both $\alpha \prec \beta$ and $\sigma \prec \tau$. With this preordering, $\mathscr{A} \times \mathscr{S}$ is a directed set.

(2). For each $(\alpha, \sigma) \prec (\beta, \tau)$, let $\mu_{(\alpha,\sigma),(\beta,\tau)}\colon Y_\alpha \times Z_\sigma \to Y_\beta \times Z_\tau$ be the continuous map $\varphi_{\alpha\beta} \times \psi_{\sigma\tau}$. Then the family of spaces $\{Y_\alpha \times Z_\sigma \mid (\alpha, \sigma) \in \mathscr{A} \times \mathscr{S}\}$, together with the maps $\mu_{(\alpha,\sigma),(\beta,\tau)}$, form a direct spectrum.

(3). There is a continuous bijection

$$h\colon \operatorname{Lim\,Dir}_{\mathscr{A} \times \mathscr{S}}\{Y_\alpha \times Z_\sigma\} \to \operatorname{Lim\,Dir}_{\mathscr{A}} Y_\alpha \times \operatorname{Lim\,Dir}_{\mathscr{S}} Z_\sigma.$$

Proof: Since (1) and (2) are trivial, we prove only (3). For each (α, σ), let $h_{(\alpha,\sigma)}: Y_\alpha \times Z_\sigma \to Y^\infty \times Z^\infty$ be the continuous map $\varphi_\alpha \times \psi_\sigma$. The family $\{h_{(\alpha,\sigma)} \mid (\alpha, \sigma) \in \mathscr{A} \times \mathscr{S}\}$ is readily verified to be a map of the spectrum into $Y^\infty \times Z^\infty$, so with the continuous $h = \text{Lim Dir } h_{(\alpha,\sigma)}$, we obtain for each (α, σ) the commutative diagram



To show that h is bijective, it suffices (I, **6.9**) to construct a map

$$g: Y^\infty \times Z^\infty \to \text{Lim Dir}(Y_\alpha \times Z_\sigma)$$

(g is not asserted to be continuous!) such that both $h \circ g = 1$ and $g \circ h = 1$. We define g as follows:

Given $(y, z) \in Y^\infty \times Z^\infty$, choose any representative y_α of y, z_σ of z, and set $g(y, z) = \mu_{(\alpha,\sigma)}(y_\alpha, z_\sigma)$. g is uniquely defined: if the selected representatives were y_β, z_τ, then y_α, y_β (resp. z_σ, z_τ) have a common successor y_γ (resp. z_ρ); since $y_\gamma \times z_\rho$ is clearly a successor of both $y_\alpha \times z_\sigma$ and $y_\beta \times z_\tau$, we have [**1.3**(2)] that $\mu_{(\alpha,\sigma)}(y_\alpha, z_\sigma) = \mu_{(\beta,\tau)}(y_\beta, z_\tau)$.

It is now evident that both $h \circ g = 1$ and $g \circ h = 1$, so the proof is complete.

By the "poor behavior" of cartesian products we mean that in general h is not a homeomorphism. Indeed, it may fail to be a homeomorphism even though $\mathscr{A} = \mathscr{S} =$ positive integers, each Y_ι is a topological group, and each $Y_\iota = Z_\iota$. This is one reason that, when direct spectra of topological groups are considered, each group is usually taken with the discrete topology.

1.10 Remark The notions of direct spectrum and of direct limit can be extended to include the case where more than one connecting map between pairs of spaces is allowed. This extension is as follows:

Let $\mathscr{A}$ be a directed set, and $[Y_\alpha \mid \alpha \in \mathscr{A}\}$ a family of spaces indexed by $\mathscr{A}$. For each pair of elements α, β such that $\alpha \prec \beta$, let $(\varphi^\mu_{\alpha\beta} \mid \mu \in M(\alpha, \beta)\}$ be a given nonempty family of continuous maps $\varphi^\mu_{\sigma\beta}: Y_\alpha \to Y_\beta$, which we call connecting maps [the cardinal of each $M(\alpha, \beta)$ need not be finite, and may vary with (α, β)]. Assume that the family of all connecting maps satisfies the following two conditions:

(1). If $\alpha \prec \beta \prec \gamma$, then each $\varphi^\lambda_{\beta\gamma} \circ \varphi^\mu_{\alpha\beta}$ is also a connecting map.

(2). Given $\alpha \prec \beta$, any two $\varphi^\mu_{\alpha\beta}, \varphi^\lambda_{\alpha\beta}$, and any $y_\alpha \in Y_\alpha$, there exists a γ such that $\beta \prec \gamma$, and a $\varphi^\sigma_{\beta\gamma}$, such that $\varphi^\sigma_{\beta\gamma} \circ \varphi^\mu_{\alpha\beta}(y_\alpha) = \varphi^\sigma_{\beta\gamma} \circ \varphi^\lambda_{\alpha\beta}(y_\alpha)$.

Then $\{Y_\alpha; \varphi^\mu_{\alpha\beta}\}$ is called a generalized direct spectrum.

The construction of Lim Dir Y_α proceeds formally as before (successor of y_α means image under any connecting map), and with evident modifications in the definitions, all the results **1.5–1.9** are valid.

2. Inverse Limits

2.1 Definition Let $\mathscr{A}$ be a preordered set and $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces indexed by $\mathscr{A}$. For each pair of indices α, β satisfying $\alpha \prec \beta$, assume that there is given a continuous map $\mu_{\beta\alpha}: Y_\beta \to Y_\alpha$ and that these maps satisfy the following condition: If $\alpha \prec \beta \prec \gamma$, then $\mu_{\gamma\alpha} = \mu_{\beta\alpha} \circ \mu_{\gamma\beta}$. Then the family $\{Y_\alpha; \mu_{\beta\alpha}\}$ is called an inverse (or projective) spectrum over $\mathscr{A}$ with spaces Y_α and connecting maps $\mu_{\beta\alpha}$.

Observe that, in contrast with **1.1**, we do not here require that the indexing set be directed. The fact that the maps "go in the direction opposite to the order" is obviously irrelevant unless further restrictions are imposed on the ordering of $\mathscr{A}$, and is done simply for later convenience.

Each inverse spectrum yields a limit space:

2.2 Definition Let $\{Y_\alpha; \mu_{\beta\alpha}\}$ be an inverse spectrum over $\mathscr{A}$. Form $\prod \{Y_\alpha \mid \alpha \in \mathscr{A}\}$, and for each α, let p_α be its projection onto the αth factor. The subspace

$$\{y \in \prod_\alpha Y_\alpha \mid \forall\ \alpha, \beta: [\alpha \prec \beta] \Rightarrow [p_\alpha(y) = \mu_{\beta\alpha} \circ p_\beta(y)]\}$$

is called the inverse (or projective) limit space of the spectrum and is denoted by Y_∞ (other notations: $\underleftarrow{\text{Lim}}\ Y_\alpha$; Lim $\text{Inv}_{\mathscr{A}}\ Y_\alpha$).

According to this definition, a point $y = \{y_\alpha\} \in \prod_\alpha Y_\alpha$ belongs to Y_∞ whenever its coordinates "match" in the sense that if $\alpha \prec \beta$, then $y_\alpha = \mu_{\beta\alpha}(y_\beta)$. Since $\alpha \prec \alpha$ for each $\alpha \in \mathscr{A}$, it follows that each coordinate y_α must actually belong to the subspace $A_\alpha = \{x \in Y_\alpha \mid \mu_{\alpha\alpha}(x) = x\}$ of Y_α. In fact, $\{A_\alpha; \mu_{\beta\alpha} \mid A_\beta\}$ is itself an inverse spectrum over $\mathscr{A}$, since if $a_\beta \in A_\beta$, then for each pair $\alpha \prec \beta$ the formula $\mu_{\alpha\alpha} \circ \mu_{\beta\alpha}(a_\beta) = \mu_{\beta\alpha}(a_\beta)$ shows $\mu_{\beta\alpha}(a_\beta) \in A_\alpha$, and it is easy to see that the two subspaces A_∞ and Y_∞ of $\prod_\alpha Y_\alpha$ are the same.

The space Y_∞ is certainly empty whenever at least one $\mu_{\alpha\alpha}$ has no fixed points; however, it can also be empty even though each $Y_\alpha \neq \varnothing$, each $\mu_{\beta\alpha}$ is surjective, and each $\mu_{\alpha\alpha}$ is the identity map. Clearly, Y_∞ need not be a discrete space, even though each Y_α is discrete (*cf.* IV, **1**, Ex. 3). Since Y_∞ is a subspace of $\prod_\alpha Y_\alpha$, we find at once that Y_∞ is resp. Hausdorff, regular, completely regular whenever each Y_α is resp. Hausdorff, regular, completely regular.

The elements of Y_∞ are called threads; note that each thread has a unique representative in each Y_α, but that an element of Y_α may represent many threads. The restriction $p_\alpha \mid Y_\infty \colon Y_\infty \to Y_\alpha$ is denoted by μ_α and is called the canonical map of Y_∞ into Y_α. It is evidently continuous, and two threads x, y are the same if and only if $\mu_\alpha(x) = \mu_\alpha(y)$ for every $\alpha \in \mathscr{A}$.

2.3 (1). Whenever $\alpha \prec \beta$, the diagram

$$\begin{array}{ccc} Y_\infty & \xrightarrow{\ \mu_\beta\ } & Y_\beta \\ & {\scriptstyle \mu_\alpha} \searrow \quad \swarrow {\scriptstyle \mu_{\beta\alpha}} & \\ & Y_\alpha & \end{array}$$

is commutative

(2). If $\mathscr{A}$ is a directed set, then the sets $\{\mu_\alpha^{-1}(U) \mid \text{all } \alpha, \text{ all open } U \subset Y_\alpha\}$ form a basis for Y_∞.

Proof: (1) is obvious.

(2). Let $x \in V$, where V is open in Y_∞. Since Y_∞ is a subspace of $\prod_\alpha Y_\alpha$, there are finitely many open $U_{\alpha_i} \subset Y_{\alpha_i}$, $i = 1, \cdots, n$ such that $x \in \langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle \cap Y_\infty \subset V$. We are to show (III, **2.2**) that for some suitable α and open $U \subset Y_\alpha$, $x \in \mu_\alpha^{-1}(U) \subset \langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle \cap Y_\infty$. Because $\mathscr{A}$ is directed, we first choose α so that $\alpha_1, \cdots, \alpha_n \prec \alpha$, and then define $U = \bigcap_1^n \mu_{\alpha\alpha_i}^{-1}(U_{\alpha_i})$, which is open in Y_α. Now, according to (1), we have

$$\mu_\alpha^{-1}(U) = \bigcap_1^n \mu_\alpha^{-1}\,\mu_{\alpha\alpha_i}^{-1}(U_{\alpha_i}) = \bigcap_1^n \mu_{\alpha_i}^{-1}(U_{\alpha_i}),$$

so that a $y \in Y_\infty$ belongs to $\mu_\alpha^{-1}(U)$ if and only if its α_ith coordinate lies in U_{α_i} for each $i = 1, \cdots, n$; consequently,

$$y \in \mu_\alpha^{-1}(U) \subset \langle U_{\alpha_1}, \cdots, U_{\alpha_n} \rangle \cap Y_\infty,$$

as required.

Ex. 1 Let $\mathscr{A}$ be any preordered set and Y any space; $\{Y_\alpha; \mu_{\beta\alpha}\}$ is called the trivial inverse spectrum over $\mathscr{A}$ with space Y if each $Y_\alpha = Y$ and each $\mu_{\beta\alpha}$ is the identity map. Clearly, $Y_\infty \cong Y$ for this spectrum.

Ex. 2 Let $\mathscr{A}$ be the directed set of positive integers, $Y_n = Z$ for each n, and $\mu_{nk}(z) = 2^{n-k}z$ for $k \leqslant n$. Then Y_∞ has exactly one element, the thread $\{0, 0, 0, \cdots\}$.

Ex. 3 Let $\{Y_n \mid n \in Z^+\}$ be a descending family of subspaces of a given space Y. For $k \leqslant n$, let μ_{nk} be the inclusion map. The reader can verify that

$$Y_\infty \cong \bigcap_1^n Y_i.$$

Note in particular that Y_∞ can be empty even though each $Y_n \neq \varnothing$ and each μ_{nn} is the identity map.

We now consider the position of Y_∞ in $\prod_\alpha Y_\alpha$.

2.4 Let $\{Y_\alpha; \mu_{\beta\alpha}\}$ be an inverse spectrum over $\mathscr{A}$.

(1). If each Y_α is Hausdorff, then Y_∞ is closed in $\prod_\alpha Y_\alpha$.

(2). If each Y_α is compact, then Y_∞ is compact (but possibly empty!).

(3). If (a) $\mathscr{A}$ is a directed set,
(b) each Y_α is compact and nonempty,
(c) $\forall\, \alpha \in \mathscr{A}: \{x \in Y_\alpha \mid \mu_{\alpha\alpha}(x) = x\} \neq \varnothing$,

then Y_∞ is not empty.

Proof: Ad (1). Let $y = \{y_\alpha\} \in (\prod_\alpha Y_\alpha) - Y_\infty$. Then $\mu_{\beta\alpha}(y_\beta) \neq y_\alpha$ for some pair $\alpha \prec \beta$. Because Y_α is Hausdorff and $\mu_{\beta\alpha}$ is continuous, we can find nbds $U_\alpha(y_\alpha)$, $U_\beta(y_\beta)$ such that $U_\alpha \cap \mu_{\beta\alpha}(U_\beta) = \varnothing$, and then $\langle U_\alpha, U_\beta \rangle$ is a nbd of y not meeting Y_∞.

(2) is an immediate consequence, since $\prod_\alpha Y_\alpha$ is compact (XI, **1.4**).

Ad (3). For each fixed $\beta \in \mathscr{A}$, let

$$S_\beta = \{y \in \prod_\alpha Y_\alpha \mid \forall\, \alpha: (\alpha \prec \beta) \Rightarrow p_\alpha(y) = \mu_{\beta\alpha} \circ p_\beta(y)\}.$$

S_β is not empty: Choose any y_β such that $\mu_{\beta\beta}(y_\beta) = y_\beta$; then for each α satisfying $\alpha \prec \beta$ let $y_\alpha = \mu_{\beta\alpha}(y_\beta)$, and finally use I, **9.3**, to obtain an element of S_β. As in (1), each S_β is verified to be closed in $\prod_\alpha Y_\alpha$. Observe that if $\beta \prec \gamma$, then $S_\gamma \subset S_\beta$: Given $y \in S_\gamma$, then for each α such that $\alpha \prec \beta$, we have $\mu_{\beta\alpha} \cdot p_\beta(y) = \mu_{\beta\alpha} \cdot \mu_{\gamma\beta} p_\gamma(y) = \mu_{\gamma\alpha} p_\gamma(y) = p_\alpha(y)$, which shows that $y \in S_\beta$. It follows from this observation that the family $\{S_\alpha \mid \alpha \in \mathscr{A}\}$ has the finite intersection property: Given $S_{\alpha_1}, \cdots, S_{\alpha_n}$, then because $\mathscr{A}$ is directed, there is some β, such that $\alpha_1, \cdots, \alpha_n \prec \beta$, so

$$\varnothing \neq S_\beta \subset \bigcap_1^n S_{\alpha_i}.$$

Now, since $\prod_\alpha Y_\alpha$ is compact, we can conclude that $\bigcap \{S_\alpha \mid \alpha \in \mathscr{A}\} \neq \varnothing$ (XI, **1.3**), and since each element of this intersection is evidently a member of Y_∞, the proof is complete.

Mappings of inverse spectra are defined in analogy to **1.4**. If $\{X_\alpha; \lambda_{\beta\alpha}\}$ and $\{Y_\alpha; \mu_{\beta\alpha}\}$ are two inverse spectra over the same preordered set $\mathscr{A}$, a continuous map $\{h_\alpha\}: \{X_\alpha; \lambda_{\beta\alpha}\} \to \{Y_\alpha; \mu_{\beta\alpha}\}$ is by definition simply a collection of continuous maps $h_\alpha: X_\alpha \to Y_\alpha$, one for each α, having the property that whenever $\alpha \prec \beta$, the diagram

$$\begin{array}{ccc} X_\beta & \xrightarrow{h_\beta} & Y_\beta \\ {\scriptstyle\lambda_{\beta\alpha}}\big\downarrow & & \big\downarrow{\scriptstyle\mu_{\beta\alpha}} \\ X_\alpha & \xrightarrow[h_\alpha]{} & Y_\alpha \end{array}$$

is commutative. By specializing one or the other of the spectra to be a trivial spectrum, we obtain the definition of maps of spaces into spectra, and conversely.

2.5 Theorem Let $\{h_\alpha\}: \{X_\alpha; \lambda_{\beta\alpha}\} \to \{Y_\alpha; \mu_{\beta\alpha}\}$ be a continuous map of inverse spectra. Then there exists one, and only one, continuous $h_\infty: X_\infty \to Y_\infty$ having the property that for each $\alpha \in \mathscr{A}$, the diagram

$$\begin{array}{ccc} X_\infty & \xrightarrow{h_\infty} & Y_\infty \\ {\scriptstyle\lambda_{\alpha}}\big\downarrow & & \big\downarrow{\scriptstyle\mu_{\alpha}} \\ X_\alpha & \xrightarrow[h_\alpha]{} & Y_\alpha \end{array}$$

is commutative. Furthermore

(1). If each h_α is injective, then so also is h_∞.
(2). If $\mathscr{A}$ is directed, and if each $h_\alpha \circ \lambda_\alpha$ is surjective, then $h_\infty(X_\infty)$ is dense in Y_∞.

Proof: Define $h: \prod_\alpha X_\alpha \to \prod_\alpha Y_\alpha$ by $h = \prod_\alpha h_\alpha$; the map h is continuous (IV, **2.5**) and we set $h_\infty = h \mid X_\infty$. Then h_∞ actually maps X_∞ into Y_∞; that is, if $\{x_\alpha\}$ is a thread, so also is $\{h_\alpha(x_\alpha)\}$: For, whenever $\alpha \prec \beta$, we have $x_\alpha = \lambda_{\beta\alpha}(x_\beta)$, and so $\mu_{\beta\alpha} \cdot h_\beta(x_\beta) = h_\alpha \circ \lambda_{\beta\alpha}(x_\beta) = h_\alpha(x_\alpha)$, showing that $\{h_\alpha(x_\alpha)\}$ is indeed a thread. It is trivial to show that the asserted diagram commutes and that h_∞ is unique.

Ad (1). Let $x = \{x_\alpha\}$, $x' = \{x'_\alpha\}$. The condition $h_\infty(x) = h_\infty(x')$ means that $h_\alpha(x_\alpha) = h_\alpha(x'_\alpha)$ for each α; since each h_α is injective, this says that $x_\alpha = x'_\alpha$ for each α; that is, $x = x'$.

Ad (2). Let $\mu_\alpha^{-1}(U)$ be any nonempty basic open set in Y_∞. Since $h_\alpha \circ \lambda_\alpha$ is surjective, $\lambda_\alpha^{-1} \circ h_\alpha^{-1}(U_\alpha) \neq \varnothing$; consequently,

$$h_\infty^{-1} \circ \mu_\infty^{-1}(U_\alpha) \neq \varnothing.$$

Each open set in Y_α therefore contains points of $h_\infty(X_\infty)$, and the assertion is proved.

The following example shows that in contrast to the case for direct limits, h_∞ may not inherit surjectivity from that of all the h_α, even if all the λ_α are surjective.

Ex. 4 Let $\mathscr{A}$ be the directed set of integers, and let $\{X_k; \lambda_{nk}\}$ be the trivial spectrum over $\mathscr{A}$ with space E^1. The spaces $Y_n = \{y \mid 0 \leqslant y \leqslant n\}$ together with the connecting maps $\mu_{nk}(y) = \min[y, k]$ for $k \leqslant n$, form an inverse spectrum $\{Y_k, \mu_{nk}\}$. Defining $h_k(x) = \min[x, k]$ gives a continuous map $\{X_k; \lambda_{nk}\} \to \{Y_k; \mu_{nk}\}$. All the hypotheses of (2) are satisfied, yet the point $\{1, 2, \cdot \cdot, n \cdot\} \in \prod_1^\infty Y_n$ belongs to Y_∞ and is not in $h_\infty(X_\infty)$.

The map h_∞ of **2.5** is called the inverse limit map induced by $\{h_\alpha\}$, and is also denoted by $\underleftarrow{\text{Lim}}\, h_\alpha$ or Lim $\text{Inv}_{\mathscr{A}}\, h_\alpha$. It is easy to prove the transitivity of the limit maps, as in **1.6**. This leads to

2.6 Corollary With the notations in **2.5**, if each $h_\alpha\colon X_\alpha \cong Y_\alpha$, then $h_\infty\colon X_\infty \cong Y_\infty$.

Proof: $\{h_\alpha^{-1}\}$ is evidently a continuous map $\{Y_\alpha; \mu_{\beta\alpha}\} \to \{X_\alpha; \lambda_{\beta\alpha}\}$; letting $g = \text{Lim Inv}\{h_\alpha^{-1}\}$, we find from the transitivity that $g \circ h_\infty = 1$ and $h_\infty \circ g = 1$, so h_∞ is a homeomorphism (III, **12.3**).

Given an inverse spectrum over $\mathscr{A}$, restriction of the indices to a preordered subset $\mathscr{B} \subset \mathscr{A}$ gives an inverse spectrum over $\mathscr{B}$. As with direct spectra, we have

2.7 Theorem Let $\mathscr{A}$ be a directed set, and $\mathscr{B} \subset \mathscr{A}$ cofinal. Then for every inverse spectrum $\{Y_\alpha; \mu_{\beta\alpha}\}$ over $\mathscr{A}$,

$$\text{Lim Inv}_{\mathscr{A}}\, Y_\alpha \cong \text{Lim Inv}_{\mathscr{B}}\, Y_\beta.$$

Proof: Let $B_\infty = \text{Lim Inv}_{\mathscr{B}}\, Y_\beta$ and let $q_\beta\colon B_\infty \to Y_\beta$ be the canonical map. Let $\mu_\beta\colon Y_\infty \to Y_\beta$ be the canonical map of the $\mathscr{A}$-spectrum. Clearly, $\{\mu_\beta \mid \beta \in \mathscr{B}\}$ is a continuous map of Y_∞ into the

$\mathscr{B}$-spectrum, so we obtain a continuous $h: Y_\infty \to B_\infty$ such that the diagram

$$\begin{array}{ccc} Y_\infty & \xrightarrow{h} & B_\infty \\ & {}_{\mu_\beta}\searrow & \downarrow{\scriptstyle q_\beta} \\ & & Y_\beta \end{array}$$

commutes for each $\beta \in \mathscr{B}$.

(1). h is injective. If $h(y) = h(y')$, then $\mu_\beta(y) = q_\beta \circ h(y) = \mu_\beta(y')$ for each $\beta \in \mathscr{B}$. To show $y = y'$, we must prove that $\mu_\alpha(y) = \mu_\alpha(y')$ for each $\alpha \in \mathscr{A}$. We do this by noting that for each $\alpha \in \mathscr{A}$, cofinality gives a $\beta \in \mathscr{B}$ with $\alpha \prec \beta$, so $\mu_\alpha(y) = \mu_{\beta\alpha} \circ \mu_\beta(y) = \mu_{\beta\alpha} \circ \mu_\beta(y') = \mu_\alpha(y')$.

(2). h is surjective. Let $\{b_\beta\} = b$ be any thread in B_∞. We use this to define a thread in Y_∞ as follows:

For each $\alpha \in \mathscr{A}$, choose a $\beta \in \mathscr{B}$ such that $\alpha \prec \beta$, and let $y_\alpha = \mu_{\beta\alpha}(b_\beta)$.

y_α is independent of the $\beta \in \mathscr{B}$ that is used: For, if $\beta, \gamma \in \mathscr{B}$ satisfy $\alpha \prec \beta, \alpha \prec \gamma$, we can find a $\lambda \in \mathscr{B}$ with $\beta, \gamma < \lambda$, and since $\{b_\beta\}$ is a thread in B_∞, $\mu_{\beta\alpha}(b_\beta) = \mu_{\beta\alpha} \circ \mu_{\lambda\beta}(b_\lambda) = \mu_{\lambda\alpha}(b_\lambda) = \mu_{\lambda\alpha} \circ \mu_{\lambda_\gamma}(b_\lambda) = \mu_{\gamma\alpha}(b_\gamma)$.

The set $\{y_\alpha \mid \alpha \in \mathscr{A}\}$ is a thread in Y_∞: Given any $\alpha, \gamma \in \mathscr{A}$ with $\alpha \prec \gamma$, cofinality gives a β with $\gamma \prec \beta$, and

$$y_\alpha = \mu_{\beta\alpha}(b_\beta) = \mu_{\gamma\alpha} \circ \mu_{\beta\gamma}(b_\beta) = \mu_{\gamma\alpha}(y_\gamma).$$

If $\beta \in \mathscr{B}$, then $y_\beta = b_\beta$: For, since $\{b_\beta\}$ is a thread, $\mu_{\beta\beta}(b_\beta) = b_\beta$; that is, $y_\beta = b_\beta$.

Now, for the element $y = \{y_\alpha\} \in Y_\infty$, we clearly have $h(y) = b$, and so h is surjective.

(3). h is open. We first show that $\{\mu_\beta^{-1}(U) \mid \text{all } \beta \in \mathscr{B}, \text{all } U \text{ open in } Y_\beta\}$ is a basis for Y_∞. In view of **2.3**(2), it is enough to show that for any $\alpha \in \mathscr{A}$ and open $U \subset Y_\alpha$, there is a $\beta \in \mathscr{B}$ and $U_\beta \subset Y_\beta$ such that $\mu_\alpha^{-1}(U) = \mu_\beta^{-1}(U_\beta)$; this follows from $\mu_\alpha^{-1}(U) = \mu_\beta^{-1} \circ \mu_{\beta\alpha}^{-1}(U)$ and continuity of $\mu_{\beta\alpha}$. A proof that h is open is now immediate: For each member of this basis, and because h is bijective,

$$h[\mu_\beta^{-1}(U_\beta)] = h[h^{-1} \circ q_\beta^{-1}(U_\beta)] = q_\beta^{-1}(U_\beta).$$

The theorem is proved.

For subspaces, the behavior of inverse limits is more satisfactory than that of direct limits:

2.8 Let $\mathscr{A}$ be a preordered set and $\{Y_\alpha; \mu_{\beta\alpha}\}$ be an inverse spectrum over $\mathscr{A}$. For each $\alpha \in \mathscr{A}$, let $A_\alpha \subset Y_\alpha$ and assume that $\varphi_{\beta\alpha}(A_\beta) \subset A_\alpha$ whenever $\alpha \prec \beta$. Then $\{A_\alpha; \mu_{\beta\alpha} \mid A_\beta\}$ is an inverse spectrum, and A_∞ is homeomorphic to the subspace $Y_\infty \cap \prod \{A_\alpha \mid \alpha \in \mathscr{A}\}$ of Y_∞.

Proof: Let $i_\alpha = A_\alpha \to Y_\alpha$ be the inclusion; since $\{i_\alpha\}$ is a continuous map of $\{A_\alpha; \mu_{\beta\alpha} \mid A_\beta\}$ into $\{Y_\alpha; \mu_{\beta\alpha}\}$, we obtain by **2.5** an injective $i_\infty: A_\infty \to Y_\infty$. The remainder of the proof is left for the reader.

This result implies that in any inverse spectrum $\{Y_\alpha; \mu_{\beta\alpha}\}$, there is no loss of generality to assume that each connecting map and each canonical map is surjective: For, we replace each Y_α by $\mu_\alpha(Y_\infty) = A_\alpha$, note that $\{A_\alpha; \mu_{\beta\alpha} \mid A_\beta\}$ is also an inverse spectrum in which the connecting and canonical maps are surjective, and finally that $A_\infty \cong Y_\infty$.

There is a duality between inverse and direct limits that involves function spaces, and which we now consider. Let $\mathscr{A}$ be a directed set, and $\{Y_\alpha; \varphi_{\alpha\beta}\}$ a direct spectrum over $\mathscr{A}$. For any space Z, each connecting map $\varphi_{\alpha\beta}: Y_\alpha \to Y_\beta$ induces a map $\varphi^+_{\beta\alpha}\ Z^{Y_\beta} \to Z^{Y_\alpha}$, where $\varphi^+_{\beta\alpha}(f) = f \circ \varphi_{\alpha\beta}$. The $\varphi^+_{\beta\alpha}$ will certainly be continuous if the function spaces carry the topology of pointwise convergence (XII, **9**) and it is immediate that $\{Z^{Y_\alpha}; \varphi^+_{\beta\alpha}\}$ is then an inverse spectrum over $\mathscr{A}$. The duality principle is formalized in

2.9 Theorem Let $\mathscr{A}$ be a directed set. Then for any direct spectrum $\{Y_\alpha; \varphi_{\alpha\beta}\}$ over $\mathscr{A}$ and any space Z,

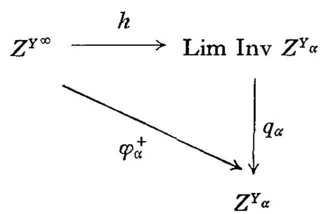
$$\text{Lim Inv}\{Z^{Y_\alpha}; \varphi^+_{\beta\alpha}\} \cong Z^{\text{Lim Dir}\{Y_\alpha; \varphi_{\alpha\beta}\}},$$

all function spaces carrying the topology of pointwise convergence.

Proof: Let $\varphi_\alpha: Y_\alpha \to Y^\infty$ be the canonical map. Since

$$\{\varphi^+_\alpha\}: Z^{Y^\infty} \to \{Z^{Y_\alpha}; \varphi^+_{\beta\alpha}\}$$

is evidently a continuous map, we have for each $\alpha \in \mathscr{A}$ a commutative diagram



where $h = \text{Lim Inv}\,\varphi^+_\alpha$, q_α is the canonical map, and all maps are continuous. Since $q_\alpha \circ h(f) = \varphi^+_\alpha(f)$, the map h can be described as sending each $f \in Z^{Y^\infty}$ to the thread $\{f \circ \varphi_\alpha\}$.

Conversely, given any thread $t = \{f_\alpha\} \in \text{Lim Inv}\, Z^{Y_\alpha}$, we define a map $f_t: Y^\infty \to Z$ as follows:

For each $y \in Y^\infty$, choose any representative $y_\alpha \in Y_\alpha$ and set $f_t(y) = f_\alpha(y_\alpha)$.

f_t is uniquely defined: If y_β were any other representative of y, then $\varphi_{\alpha\gamma}(y_\alpha) = \varphi_{\beta\gamma}(y_\beta)$ for a suitable γ so $f_\alpha(y_\alpha) = \varphi_{\gamma\alpha}^+(f_\gamma)(y_\alpha) = f_\gamma \circ \varphi_{\alpha\gamma}(y_\alpha) = f_\gamma \circ \varphi_{\beta\gamma}(y_\beta) = f_\beta(y_\beta)$. Furthermore, f_t is continuous on Y^∞: For, if U is open in Z, to show $f_t^{-1}(U)$ open in Y^∞ requires that we prove $\varphi_\alpha^{-1}[f_t^{-1}(U)]$ open in each Y_α, and this is evident, since $f_t \circ \varphi_\alpha = f_\alpha$.

The correspondence $t \to f_t$ thus defines a map G: Lim Inv $Z^{Y_\alpha} \to Z^{Y^\infty}$. Evidently, both $G \circ h = 1$ and $h \circ G = 1$, so both h and G are bijective. To show that h is a homeomorphism, we need prove only that G is continuous, and for this we simply verify that if (y_0, V) is any subbasic open set in Z^{Y^∞}, and if $y_\alpha \in Y_\alpha$ is any representative of y_0, then G maps the open set $q_\alpha^{-1}[(y_\alpha, V)]$ onto exactly (y_0, V). The proof is complete.

We remark that if $\{Y_\alpha; \mu_{\beta\alpha}\}$ is an inverse spectrum over the directed set $\mathscr{A}$, then it is *not* in general true that Lim Dir$\{Z^{Y_\alpha}; \mu_{\alpha\beta}^+\} \cong Z^{\text{Lim Inv}\{Y_\alpha;\mu_{\beta\alpha}\}}$.

2.10 Remark Again the notion of an inverse spectrum can be extended to cover the case where more than one map between pairs of spaces is allowed. This extension is as follows:

Let $\mathscr{A}$ be a preordered set and $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be a family of spaces indexed by $\mathscr{A}$. For each pair of elements α, β such that $\alpha \prec \beta$, let $\{\varphi_{\beta\alpha}^\mu \mid \mu \in M(\alpha, \beta)\}$ be a given nonempty family of continuous maps $\varphi_{\beta\alpha}^\mu: Y_\beta \to Y_\alpha$, which we will call connecting maps. The cardinality of $M(\alpha, \beta)$ need not be finite and may vary with (α, β). Assume that whenever $\alpha \prec \beta \prec \gamma$, then each $\varphi_{\beta\alpha}^\lambda \circ \varphi_{\gamma\beta}^\mu$ is also a connecting map. Then $\{Y_\alpha; \varphi_{\beta\alpha}^\mu\}$ is called a generalized inverse spectrum over $\mathscr{A}$.

The limit space is that subspace of $\prod_\alpha Y_\alpha$ defined by the condition

$$\forall\, \alpha, \beta \in \mathscr{A}: (a \prec \beta) \Rightarrow p_\alpha(y) = \varphi_{\beta\alpha}^\mu \circ p_\beta(y) \text{ for all } \mu \in M(\alpha, \beta).$$

With evident modifications, all the results **2.4 2.8** are valid for generalized inverse spectra; and with the generalized direct spectra of **1.10**, the duality theorem is also true.

Problems

1. Let $\{Y_\alpha; \varphi_{\alpha\beta}\}$ be a direct spectrum over $\mathscr{A}$. Prove:

a. If each $\varphi_{\alpha\beta}$ is injective, so also is each canonical φ_α.

b. If each $\varphi_{\alpha\beta}$ is surjective, so also is each canonical φ_α.

c. For each α let $Y'_\alpha = \varphi_{\alpha\alpha}(Y_\alpha)$ and $\psi_{\alpha\beta} = \varphi_{\alpha\beta} \mid Y'_\alpha$. Show that $\{Y'_\alpha; \psi_{\alpha\beta}\}$ is a direct spectrum and that there is a continuous bijection $Y^\infty \to (Y')^\infty$.

d. For each α let $Y'_\alpha = Y_\alpha/K(\varphi_\alpha)$ (*cf.* VI, **7**). Show that each $\varphi_{\alpha\beta}$ is relation-preserving, and let $\varphi'_{\alpha\beta}$ be the maps obtained by passing to the quotient. Show that $\{Y'_\alpha; \varphi'_{\alpha\beta}\}$ is a direct spectrum and that there is a continuous bijection $G: Y^\infty \to (Y')^\infty$.

e. Let $\{h_\alpha\}: \{Y_\alpha; \varphi_{\alpha\beta}\} \to \{Z_\alpha; \psi_{\alpha\beta}\}$ be continuous and $G_\alpha \subset Y_\alpha \times Z_\alpha$ be the graph of h_α. Let $\lambda_{\alpha\beta} = \varphi_{\alpha\beta} \times \psi_{\alpha\beta} \mid G_\alpha$. Show that $\{G_\alpha; \lambda_{\alpha\beta}\}$ is a direct spectrum and that there is a continuous bijection $G^\infty \to$ the graph of $\underrightarrow{\text{Lim}}\, h_\alpha$.

2. For each $n \in Z^+$, let $Y_n = [-1, +1]$. If $k \geqslant n$, define $\varphi_{nk}: Y_n \to Y_k$ by $\varphi_{nk}(y) = 2^{n-k}y$. Show that Lim Dir $\{Y_n, \varphi_{nk}\}$ is homeomorphic to E^1.
3. Let Y, Z be two spaces, and for each $y \in Y$, let $G(y) = \underrightarrow{\text{Lim}}\, C_{U(y)}(Z)$ be the set of germs of continuous maps of a nbd of y into Z (*cf.* I, Ex. 3). For each $f \in C_U(Z)$ and each $y_0 \in U$, let $y_0(f)$ be the canonical image of f in $G(y_0)$ and let $(f, C_U(Z)) = \{y_0(f) \mid y_0 \in U\}$. Topologize $G = \sum \{G(y) \mid y \in Y\}$ by using all sets $(f, C_U(Z))$ as subbasis, and let $p: G \to Y$ be the map sending each $G(y)$ to y. Prove: p is continuous, open, and each $g \in G$ has a nbd U such that $p \mid U: U \cong p(U)$.
4. Let $\{Y_\alpha; \mu_{\beta\alpha}\}$ be an inverse spectrum over the directed set $\mathscr{A}$. Assume that each Y_α is compact and that each $\mu_{\alpha\alpha}$ = identity map. Prove:

   a. If U is an open set in Y_α, and if $\mu_\alpha(Y_\infty) \subset U$, then there exists a β, such that $\alpha \prec \beta$ and $\mu_{\beta\alpha}(Y_\beta) \subset U$.

   b. For each α, $\mu_\alpha(Y_\infty) = \bigcap_\beta \{\mu_{\beta\alpha}(Y_\beta) \mid \alpha \prec \beta\}$.

   c. If each Y_α is connected, then Y_∞ is connected.
5. Let $\{h_\alpha\}: \{Y_\alpha; \mu_{\beta\alpha}\} \to \{Z_\alpha; \lambda_{\beta\alpha}\}$ be a continuous map of inverse spectra. Let $G_\alpha \subset Y_\alpha \times Z_\alpha$ be the graph of h_α and $g_{\beta\alpha} = \mu_{\beta\alpha} \times \lambda_{\beta\alpha} \mid G_\beta$. Prove that $\{G_\alpha; g_{\beta\alpha}\}$ is an inverse spectrum and that G_∞ is homeomorphic to the graph of Lim Inv h_α.
6. Let $\{Y_\alpha \mid \alpha \in \mathscr{A}\}$ be any family of spaces, and partially order the family of all finite subsets $\mathscr{F} \subset \mathscr{A}$ by inclusion.

   a. For $\mathscr{F} \subset \mathscr{F}'$, let $\mu_{\mathscr{F}'\mathscr{F}}: \prod_\alpha \{Y_\alpha \mid \alpha \in \mathscr{F}'\} \to \prod_\alpha \{Y_\alpha \mid \alpha \in \mathscr{F}\}$ be the projection. Show Lim $\text{Inv}_{\mathscr{F}}$ $[\prod \{Y_\alpha \mid \alpha \in \mathscr{F}\}] \cong \prod_\alpha Y_\alpha$.

   b. Choose a fixed point $\{b_\alpha^0\} \in \prod_\alpha \{Y_\alpha \mid \alpha \in \mathscr{A}\}$, and for $\mathscr{F} \subset \mathscr{F}'$ let

   $$\varphi_{\mathscr{F}\mathscr{F}'}: \prod \{Y_\alpha \mid \alpha \in \mathscr{F}\} \to \prod \{Y_\alpha \mid \alpha \in \mathscr{F}'\}$$

   be the homeomorphism onto the slice in $\prod \{Y_\alpha \mid \alpha \in \mathscr{F}'\}$ through

   $$\{b_\alpha^0 \mid \alpha \in \mathscr{F}'\}$$

   parallel to $\prod \{Y_\alpha \mid \alpha \in \mathscr{F}\}$. Prove Lim $\text{Dir}_{\mathscr{F}}$ $[\prod \{Y_\alpha \mid \alpha \in \mathscr{F}\}] \cong PY_\alpha$ (see VI, 8, Prob. 5).

Index

C

D

E

F

G

H

I

M

N

O

P

Q

R

S

W

Z